PROCEEDINGS OF THE INTERNATIONAL CONGRESS OF MATHEMATICIANS

HELSINKI 1978

VOLUME 1

HELSINKI 1980

ISBN 951-41-0352-1

Editor: Olli Lehto
Department of Mathematics
University of Helsinki
Hallituskatu 15
SF-00100 Helsinki 10

Printed in Hungary

The Proceedings of the International Congress of Mathematicians held in Helsinki, August 15—23, 1978, is in two volumes. Volume 1 contains an account of the Congress, the list of members, presentations of the works of the Fields medallists, the plenary one-hour addresses, and the invited addresses in sections 1—5. Volume 2 contains the invited addresses in sections 6—19. A complete index is included in both volumes.

Abstracts of short communications and poster sessions, which were distributed to members during the Congress, are not included.

The American Mathematical Society prepared the manuscripts for printing and took care of proofreading. The setting, printing and binding were done in Hungary.

Helsinki, May 1979

Olli Lehto

Contents

Organization of the Congress

The recommendation to hold the 1978 International Congress of Mathematicians in Helsinki was made by the Site Committee of the International Mathematical Union during the Vancouver Congress in August 1974. Final decision was taken a few days later when the Congress at its closing session accepted the invitation to Helsinki, which was presented by Professor Rolf Nevanlinna on behalf of the Finnish National Committee for Mathematics.

Before Finland offered to be host for the ICM 78, the support of the Finnish Government had been secured. The Congress was particularly honoured by the fact that Dr. Urho Kekkonen, President of Finland, consented to be its Patron.

The scientific program was the responsibility of the International Mathematical Union through the Consultative Committee, whose members were Professors A. Borel (chairman), J. F. Adams, S. Chern, Y. Kawada, O. Lehto, I. S. Louhivaara, B. Malgrange, S. M. Nikolskii and C. Olech. After preparatory work which took about a year, the Committee decided in June 1976 to divide the mathematical program into 19 sections and appointed the cores of the panels for these sections. The panels completed themselves and submitted their suggestions, whereupon the Consultative Committee in October 1977 chose 17 mathematicians to give one-hour plenary addresses and 121 to give 45-minute addresses in sections. Two more names were added later. Of the 136 who accepted the invitation, 119 were present at the Congress.

The Fields Medals Committee, consisting of Professors D. Montgomery (chairman ex officio), L. Carleson, M. Eichler, I. M. James, J. Moser, J. V. Prohorov, B. Szőkefalvi-Nagy and J. Tits, arrived at its decisions in early 1978.

Other preparations of the Congress were in the hands of the Finnish Organizing Committee. Its chairman was Olli Lehto who took direct responsibility of all arrangements. With time, the number of mathematicians involved in the organization increased, and the Committee split into several project groups. In these a great amount of work was performed by Heikki Apiola, Elja Arjas, Heikki Bonsdorff, Timo Erkama, Heikki Haario, Matti Lehtinen, Olli Lokki, Ilppo Simo Louhivaara, Olli Martio, Marjatta Näätänen, Rolf Nevanlinna, Seppo Rickman, Arto Salomaa, Jukka Sarvas, Onerva Savolainen, and others. In all, about 100 Finnish mathematicians gave some assistance to the Congress. A lot of voluntary work was also done by the staff of the University of Helsinki. A small Congress Bureau with salaried staff was set up in late 1975. It expanded considerably during the half year preceding the Congress and assumed then a rapidly increasing share of the arrangements. A particularly important role was played by Leena Kahlas and Tuulikki Mäkeläinen, both engaged from the very beginning. Accommodation was handled by Area Travel Agency Ltd., which saved the organizers a great deal of work.

The Congress had several sources of revenue: (1) annual grants in 1974—78 from the Ministry of Education of Finland; (2) a subvention from the International Mathematical Union; (3) donations, both in the form of direct grants or of facilities placed at its disposal without charge (a list of donors is given on p. 14); (4) membership fees, which were $60 for ordinary members registered before 15 May 1978, $70 after that, and $35 for accompanying members.

The International Mathematical Union gave travel grants to young mathematicians from developing countries or from countries with currency difficulties. The Congress waived their fees and, thanks to aid from the Finnish Government, was able to give them free accommodation.

A short preliminary announcement about the Helsinki Congress was sent out in the autumn of 1976 to all countries of the world where some mathematical organization could be located. The First Announcement was dispatched in July 1977 to these same addresses with the request that copies of it be further distributed among the mathematical community of the respective country. Those mathematicians wishing to receive the Second Announcement were asked to return an attached form. The Second Announcement contained detailed information about the Congress, instructions about short communications and poster sessions, and the registration form. Its mailing began in November 1977, and it was sent in all in 7000 copies, mostly to individual addresses. Registration started immediately in November, reached its peak in May 1978 and continued even during the Congress. The Third Announcement, which contained a list and a rough schedule of invited lectures, was prepared as soon as the organizers had received answers from all invited speakers. Its mailing to registered members started in April 1978.

There were 3038 registered ordinary members and over 900 accompanying members. Of these all were not present; on the other hand, lectures and seminars were also attended by a number of non-registered mathematicians.

Mathematical activities of the Congress took place in the centre of Helsinki. Reports on the work of the Fields medallists and the plenary one-hour addresses were given in Finlandia Hall, where the opening and the closing sessions were also held. All other mathematical events took place at the University of Helsinki, which also housed the Congress Bureau.

In addition to the invited lectures, about 500 ten-minute communications were given and 40 mathematicians spoke about their work in poster sessions. Abstracts of these were published in a book distributed to members at registration in Helsinki. Unofficial mathematical activities also included a three-day symposium organized by the International Commission on Mathematical Instruction, a number of spontaneous seminars, and special sessions and films in two evenings. A book exhibition organized by Suomalainen Kirjakauppa was open throughout the Congress.

The City of Helsinki showed hospitality to all participants. Since the City Hall was not big enough for all to attend at the same time, two receptions were held, in the evenings of August 16 and 17.

The Organizing Committee arranged various social events. An open-air gathering featuring Finnish folklore took place on the island of Seurasaari on August 19. It was attended by well over 3000 people. On Sunday, August 20, the members were able to choose between two excursions. One was a four-hour cruise in the Gulf of Finland, the other a twelve-hour visit to Turku, the old capital of Finland. Both events were attended by about 1500 persons. Two passenger ships were needed for the cruise and 36 buses for the trip to Turku. Two piano concerts were also arranged. One by Minna Pöllänen was in Temppeliaukio Church on August 17, the other by Andrei Gavrilov in Finlandia Hall on August 22.

Opening Ceremonies

The opening ceremonies of the Helsinki Congress took place in Finlandia Hall on August 15, 1978, at 9.30. After a musical performance by the Helsinki Philharmonic Orchestra, Professor Deane Montgomery, President of the International Mathematical Union, opened the proceedings by proposing that Professor Olli Lehto be elected President of the Congress by acclamation. Following his election, Professor Lehto gave his presidential address to the Congress.

It is my pleasant duty to declare the 1978 International Congress of Mathematicians opened. This is a great moment for the Finnish mathematical community on whose behalf I would like to cordially welcome all our foreign guests. This is a gathering of one huge mathematical family and not of delegations or representatives of countries, but just to illustrate its world-wide character, I mention the fact that there are participants here from 83 different countries.

Organizing a meeting of this magnitude would not have been possible without the support of the Finnish government. We greatly appreciate the fact that the President of Finland, Dr. Urho Kekkonen accepted to be the patron of the Congress, and only regret that, owing to his absence from Finland at the moment, he could not personally attend these opening ceremonies as he had intended to do.

At a very early stage, even before the IMU, the International Mathematical Union, had made the final decision in favour of Finland, the Ministry of Education explicitly promised to back the Congress in case it was to be held here. Its financial support was then indispensable throughout the early period of preparations. I have the pleasure to extend our thanks to Mr. Jaakko Numminen, Secretary General of the Ministry of Education, who in the absence of the Minister represents the Finnish Government here.

The Congress will enjoy the hospitality of the City of Helsinki. The City has the reputation of being friendly to scientific conferences, but I guess that this time it took more than the customary friendliness when the size of our Congress was revealed. I am pleased to thank Mr. Teuvo Aura, the Lord Mayor of Helsinki, on behalf of the Congress.

We would scarcely have ventured to take the responsibility for an ICM without hints from the IMU that Finland would not be a disagreeable choice to its Site Committee. This we interpreted as a kind of appreciation of the mathematical research carried out in this country. Now Finland has the privilege of possessing, since well over half a century, an unparalleled mathematical ambassador: I am speaking of course of Professor Rolf Nevanlinna. I propose that Professor Nevanlinna be elected Honorary President of this Congress.

The proposal was warmly accepted by the Congress. and Professor Rolf Nevanlinna was elected Honorary President by acclamation.

Professor Nevanlinna in his opening address of the Stockholm Congress in 1962 and Professor Coxeter four years ago in Vancouver both pointed out the unique role of the ICM's as the only meetings where surveys are presented over the whole range of mathematics. They emphasized the great importance of this tendency for unification at a time when ever-expanding research threatens to lead our science to a dangerous ramification. These views are in full agreement with the instructions given by the IMU to the Consultative Committee for planning and composing the mathematical program of the ICM's.

A careful analysis of the reasons for holding ICM's not only serves as a motivation for the fairly difficult and expensive organization. It is also required if we wish to preserve the present character of these congresses. The mathematicians form a big active group, and it is only natural to try to associate all sorts of activities with a gathering as important as an ICM. No matter how important these activities are as such, and some clearly are, like promoting mathematics in developing countries and various questions related to the teaching of mathematics, at an ICM they can only play a secondary role, subjected to the official mathematical program.

Since the Stockholm Congress 1962, the official mathematical program results from international collaboration governed by detailed rules issued by the IMU. Well over a hundred of the world's leading mathematicians are involved in the work, the panels make proposals about invited speakers, and the Consultative Committee creates the final list. In my opinion, this international cooperation, which goes on for over two years in each four-year period, is very important for our science as such, and I cannot see any essentially better procedure for a neutral and authoritative appraisal of current mathematical research.

This time the Consultative Committee had seven members appointed by the IMU, namely Professor Borel as chairman and Professors Adams, Chern, Kawada, Malgrange, Nikolskii and Olech. Besides, there were two members from the host country. The committee eventually reached almost all its decisions unanimously. Its foreign members also went far beyond their liabilities in giving unobtrusively many valuable pieces of advice to the Organizing Committee. This applies in particular to its chairman, Professor Borel. May I propose a vote of thanks to the Consultative Committee.

The Organizing Committee spent about one year in trying to solve as best it could the not quite trivial problem of informing all mathematicians of the world about the Congress. That the final result was quite satisfactory was largely due to the help obtained from many institutions and individuals. The American Mathematical Society widely advertised the Congress, and more locally, the same was done by several other mathematical societies and national committees.

We were particularly lucky in that the newly established African Mathematical Union, under the leadership of its President, Professor Hogbe-Nlend, practically eliminated our problems with Africa. In Latin America, good results were due to the personal efforts of Professor D'Ambrosio. Much to our pleasure, there are members in this Congress from a higher number of countries than ever before.

The work of the Organizing Committee has been carried out at the University of Helsinki, where also a considerable part of the activities of the Congress will take place. The University has given assistance in so many direct and indirect ways that I am certainly unable to count them all, and the palpably friendly attitude of the administration has made the period of preparations very pleasant to the organizers. I would like to thank the Rector of the University, Professor Ernst Palmén, and the whole central administration.

We had no high hopes when we started fund-raising for the Congress. But in spite of hard times and the fact that after its divorce from computer science, mathematics has virtually no direct contacts here with industry, the end result is very good. The long list of donors is printed in the Program Book, and there is no exaggeration in the text which says that their contributions have been essential for the Congress. I hope this generosity means that society at large still esteems the intrinsic value of mathematics and understands that useful applications are possible only if backed and connected by theories on a more abstract level.

After this, Mr. Jaakko Numminen, Secretary General of the Ministry of Education, representing the Finnish Government, Mr. Teuvo Aura, the Lord Mayor of Helsinki, and Professor Ernst Palmén, Rector of the University of Helsinki, gave short addresses welcoming members of the Congress to Finland.

Professor Montgomery, chairman of the Fields Medals Committee, then presented the following report.

The medals presented at each International Congress of Mathematicians were first proposed by Professor J. C. Fields, who was President of the Congress held in Toronto in 1924. The fund for the medals was obtained from funds remaining after the financing of the Toronto Congress. The proposal was accepted in 1932 and the first two medals were given in 1936. Each medal carries with it a cash prize of 1500 Canadian dollars.

As usual, the Executive Committee of the International Union appointed a committee to select the medalists for this Congress. The Committee consisted of Professors L. Carleson, M. Eichler, I. James, J. Moser, J. V. Prohorov, B. Szőkefalvi-Nagy, J. Tits, and myself as Chairman. The Committee decided to follow the well-established tradition of considering only people of age 40 or under. Even with this limitation, the list of those seriously considered numbered several dozen.

After much deliberation and consultation and after considering advice from many outside the Committee, the Committee has selected four individuals for the award. They are, in alphabetical order, P. Deligne, C. Fefferman, G. A. Margulis, D. Quillen. I offer them our warm congratulations.

Information has been received that, unfortunately, G. A. Margulis is unable to be present, so his award will be presented to him later. I now ask Professor Rolf Nevanlinna, Honorary President of the Congress, to come forward to give the medals to the other three.

When Professors Deligne, Fefferman and Quillen had received the prizes from Professor Nevanlinna, it was announced that after the opening session, Professor N. Katz would speak on the work of Deligne, Professor L. Carleson on Fefferman, Professor J. Tits on Margulis and Professor I. M. James on Quillen.

The opening session ended with the Helsinki Philharmonic Orchestra playing Finlandia by Jean Sibelius, and the National Anthem.

Closing Ceremonies

The closing session of the Helsinki Congress took place in Finlandia Hall on August 23, 1978, at 15.00. Professor J. W. S. Cassels, Vice-President of the International Mathematical Union, presented the following report:

It is traditional that the General Assembly of the International Mathematical Union should be held immediately before the Congress and that the President of the Union should report briefly on it at this closing session. Unfortunately Professor Montgomery is unable to be here and so I have been asked to take his place.

First, I must report the names of the members of the new Executive Committee who were elected to hold office for four years from 1 January 1979. They are:

President	Professor L. Carleson
Vice-Presidents	Professor M. Nagata
	Professor J. V. Prohorov
Secretary	Professor J. L. Lions
Members-at-large	Professor E. Bombieri
	Professor J. W. S. Cassels
	Professor M. Kneser
	Professor O. Lehto
	Professor C. Olech

In addition, Professor D. Montgomery, who becomes Past-President, will remain a member, though without a vote.

I can give only a brief informal report on the work of the General Assembly. A fuller account will appear in the Bulletin of the IMU which is sent to all National Adhering Organizations. The General Assembly had before it the report of the outgoing Executive Committee. Amongst other things, they listed the 19 Symposia and Conferences which have been co-sponsored by the Union during the past four years. The direct financial aid which the Union gives is necessarily small but it also helps with advice on the scientific programme and in other ways: and on occasion the moral support is also useful. Two of the meetings were jointly sponsored with the Physics Union, IUPAP, and two with the Mechanics Union, IUTAM.

A problem of continued concern to the Union is that some mathematicians are prevented from attending meetings sponsored by the Union. This can happen in two ways. The first is that mathematicians may be refused entry by the country in which the meeting is held: this has caused difficulties in the past to our Union but is not, we hope, now a great problem. The other way in which mathematicians may be prevented from attending is that their own country may refuse permission to attend. This is a continuing problem, as the present Congress has again demonstrated. These problems are not, of course, peculiar to our own Union but are common to the scientific community and have greatly occupied the attention and energies of ICSU (International Council of Scientific Unions). The General Assembly endorsed the stand of ICSU on this important matter and requested the incoming President to report on the situation to the next General Assembly.

Another problem which our Union shares with most of the Unions of the ICSU family is that the Peoples' Republic of China is not a member. There are difficult issues here, to which ICSU and its Unions have devoted much attention and which I shall not go into now. It is fair to say that there is a general wish for the Peoples' Republic of China to become a member, but only if this can happen in a way which does not impair the principles on which our Union is based. The General Assembly urged the new Executive Committee to take positive steps to this end.

The General Assembly also considered the organization of our Congresses and, in particular, the machinery for the selection of speakers. The organization of a Congress, other than its scientific programme, is in the hands of the Organizing Committee, which is appointed by the host country — and, in parenthesis, may I say how much we admire the excellent job the Finns have done this time. The selection of speakers is, however, in the hands of the so-called Consultative Committee, which is appointed jointly by the Executive of the Union and the Organizing Committee, with a chairman appointed by the President of the Union. This Committee seeks the advice of a large number of subject panels with a wide international membership. Some dissatisfaction was expressed by delegations at the working of this system, but many declared themselves happy with it. It was agreed that National Committees should be asked to make suggestions to the Executive Committee as to how it might be yet further improved and that the Executive should report if it felt that changes were desirable.

The General Assembly paid much attention to the fostering of mathematics in developing countries. In the first place, I should report that the Executive Committee, following the precedent for Vancouver set up a fund to give travel grants to well-qualified young mathematicians from developing countries, and also from countries where there are severe monetary restrictions, so as to enable them to attend this Congress. The money for this came mainly from the Union's own resources but we also received subventions from UNESCO, COSTED and the computer firm ICL, for which we are most grateful. Enquiries were made on a wide basis and nearly 50 young mathematicians were helped in this way: there were more deserving cases where the Committee administering the fund would have wished to help had more resources been available. There was one very welcome innovation. The Organizing Committee waived the congress fee for the grantees and together with the Government of Finland provided accommodation for them at no cost. The General Assembly welcomed this generous action and agreed to commend it to the Organizing Committees of future Congresses.

In this connection I should like also to mention the Union's Fellowships which were established some time ago to assist mathematicians from developing countries to work at institutions elsewhere. The Italian government made a generous grant which was announced to the General Assembly four years ago at Vancouver. Disappointingly little use has been made of these Fellowships. I must hasten to add that funds are limited and so all proposals have to be submitted to the most rigorous examination.

This is the appropriate place to mention the Commission on Exchange. It was initially set up to foster visits in general by mathematicians from one country to another. However in many cases existing channels work well without the intervention of the Union and so the emphasis has moved towards visits to and from developing countries. This has been in particular the case in the last four years under the energetic chairmanship of Professor Coleman. Particularly noteworthy is that the Commission has obtained generous support from the Canadian Government for the All-African Mathematical Conference which was held at Rabat in 1976, when the African Mathematical Union was founded, and also for a forthcoming conference on pre-university mathematics in Africa.

So much for what has happened so far in relation with development: now for the future. The General Assembly decided to recognize the importance of mathematics for development by replacing the old Commission on Exchange with a new Commission with new terms of reference. The new Commission will be called the Commission on Development and Exchange and the new chairman will be Professor Hogbe-Nlend. Further, it was decided to make a special appeal for contributions to finance development activities. It is hoped that member countries will subscribe generously and several delegations were already in a position to make promises of support.

Last but not least I come to our other Commission: ICMI (International Commission on Mathematical Instruction) which has, in fact, a history antedating that of our Union. Under the able guidance of its chairman, Professor Iyanaga, and its secretary, Professor Kawada, it has continued and expanded its valuable role. Its activities are recorded in some detail in ICMI's own Bulletin and in a more summary form in the Bulletin of IMU. They are too varied to be described here: it is sufficient to remind you of the successful conference at Karlsruhe two years ago. The new chairman will be Professor Whitney.

In conclusion, I am sure that you will join me in wishing the new Executive Committee and the new Commissions all success in the coming four years.

Professor O. Lehto, as a member of the Committee to select the site for the 1982 Congress, then invited Professor K. Urbanik to speak on behalf of the Polish National Committee for Mathematics.

Professor Urbanik spoke as follows:

On behalf of the Polish National Committee of Mathematics I have the honour to invite you to the next International Congress of Mathematicians in Warsaw.

Poland, the home country of Banach, is eager to receive the world-wide mathematical community. For a long time Polish mathematicians have carried deep in their hearts the desire to organize an international congress and we are very happy that we shall now have this opportunity.

We are all well aware that it is going to be a difficult task to organize such a big meeting, the more so as the memory of this splendid Helsinki congress will still be fresh. However, taking into account the help of the International Mathematical Union and the promised support of the Polish Academy of Sciences, we feel optimistic.

Hoping that you will accept our invitation, I welcome all of you to the next International Congress of Mathematicians to be held in August 1982 in Warsaw.

The invitation was accepted by acclamation.

Speaking on behalf of the members of the Congress, Professors K. Chandrasekharan and B. Szőkefalvi-Nagy expressed their thanks to the Finnish hosts.

In his reply, Professor Lehto thanked the members of the Congress, particularly all speakers and the 140 chairmen of the various sessions. He passed on the words of thanks of Professors Chandrasekharan and Szőkefalvi-Nagy to the Finnish mathematicians who had participated in the arrangements and to the members of the Congress Bureau, and closed his address as follows:

Our organizational task was greatly facilitated by the wealth of advice and material we received from the organizers of the Vancouver Congress. We in turn are more than willing to pass our experience, if it is requested, to our Polish colleagues. I wish best success to the ICM 82, and declare the 1978 International Congress of Mathematicians closed.

Donors

The Finnish Government, Ministry of Education
International Mathematical Union
The City of Helsinki
The University of Helsinki
The Helsinki University of Technology
The Finnish Academy of Science and Letters
The Finnish Post Office

Aaltosen Tehtaat Oy
A. Ahlström Osakeyhtiö
Oy Algol Ab
Oy Alko Ab
American Mathematical Society
Enso-Gutzeit Osakeyhtiö
Oy Musiikki Fazer Musik Ab
Finnair
Helsingin Osakepankki
Helsingin Säästöpankki
Huhtamäki Oy
Oy International Business Machines Ab
Kansallis-Osake-Pankki
Oy Kaukas Ab
Kone Osakeyhtiö
Kymi Kymmene
Neste Oy
Oy Nokia Ab
Osuuspankkien Keskuspankki Oy
Outokumpu Oy
Oy Paragon Ab
Paraisten Kalkki Oy — Pargas Kalk Ab
Postipankki
Rauma Repola Oy

Springer-Verlag
Suomen Aktuaariyhdistys
- Aura-yhtiöt
- Eläke-Varma
- Fennia
- Ilmarinen
- Kalervo
- Kaleva
- Kansa-yhtiöt
- Pohja-Yhtymä
- Pohjola
- Sampo-Tarmo
- Suomi-Salama
- Teollisuusvakuutus
- Verdandi
- Yrittäjäin Vakuutus

Suomen Yhdyspankki Oy
Säästöpankkien Keskus-Osake-Pankki
Tukkukauppojen Oy
Vaasan Höyrymylly Osakeyhtiö
Vaisala Oy
Oy Volvo-Auto Ab
Oy Wärtsilä Ab
Jenny ja Antti Wihurin rahasto

Invited Addresses

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

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

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

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

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

Invited one-hour plenary addresses

Invited 45-minute addresses in sections

Section 1. Mathematical logic and foundations of mathematics

Section 2. Algebra

Section 3. Number theory

Section 4. Geometry

Section 5. Topology

Section 6. Algebraic geometry

Section 7. Lie groups, algebraic groups, automorphic functions

Section 8. Real and functional analysis

Section 9. Complex analysis

Section 10. Operator algebras and group representations

Section 15. Mathematical physics and mechanics

Section 16. Numerical analysis

Section 17. Discrete mathematics and mathematical aspects of computer science

Section 18. Mathematics in the social and biological sciences

Section 19. History and Education

Congress Members

The names of the countries are as in the World Directory of Mathematicians.

Abdel Karim, Rahmi I. I., Saudi Arabia
Abdelmalek, Nabih N., Canada
Abdel-Megied, Mohamed, Saudi Arabia
Abe, Eiichi, Japan
Abe, Michio, Japan
Åberg, Hans, Sweden
Abibi, Daniel, Congo
Ablowitz, Mark, USA
Abrahamsson, Leif, Sweden
Abramczuk, Wojciech, Sweden
Abramovich, Shoshana, Israel
A'Campo, Norbert, France
Ackermans, Stan T. M., Netherlands
Acuña-Ortega, Osvaldo, Costa Rica
Aczél, János, Canada
Adamowicz, Zofia, Poland
Adams, John Frank, UK
Addington, Susan L., USA
Ademaj, Eshref, Yugoslavia
Adeniran, Tinuoye M., Nigeria
Afghani, Mohammad Hossein, Iran
Aghili, Feridoun S., France
Agou, Simon, France
Agunwamba, Christian C., Nigeria
Ahlfors, Lars V., USA
Ahmad, Munir, Saudi Arabia
Ahmad, Shair, USA
Ahonen, Hannu T., Finland
Aigner, Alexander, Austria
Ainola, Leo, USSR
Aizicovici, Sergiu, Romania
Akemann, Charles, USA
Aki, Kunio, Japan
Akimoto, Yuko, Japan
Akiyama, Jin, Japan
Akkar, Mohammed, Morocco
Akutowicz, Edwin, France
Alameddine, Ahmad F., Saudi Arabia
Al-Ani, Marouf M., Iraq
Al-Ani, Sabri, Iraq
Alastalo, Hannu, Finland
Albright, Hugh N., USA
Alcalá del Olmo Pérez, Angel, France
Alcantara, Julio, Peru
Alder, Henry L., USA
Alder, Michael D., Australia
Al-Dhahir, Nawar A. R., UK
Aleksandrov, Igor, USSR
Alfsen, Erik M., Norway
Al-Ghurabi, Salem Ismail, Iraq
Alhussaini, Ata, Canada
Alimpić, Branka, Yugoslavia
Alisbah, Orhan H., Turkey
Allen, Richard J., USA
Allison, Bruce, Canada
Almgren, Frederick J., Jr., USA
Al-Moajil, Abdullah H., Saudi Arabia
Alperin, Jonathan, USA
Alperin, Roger, USA
Al-Salihi, Rashid A. R., Iraq
Alsina, Claudi, Spain
Alvarado, William, Costa Rica
Amara, Mohamed, Tunisia
Amice, Yvette, France
Amir, Dan, Israel
Amitsur, Shimshon A., Israel
Ancona, Alano, France
Ancona, Vincenzo, Italy
Andenaes, Per Roar, Norway
Andersen, Kurt Munk, Norway
Anderson, Douglas R., USA
Anderson, Frank W., USA
Anderson, Glen D., USA
Anderson, Joel, USA
Anderson, Richard D., USA
Ando, Shiro, Japan
Andreian Cazacu, Cabiria, Romania
Andrews, Fred C., USA
Andronikof, Emmanuel, France
Andrushkiw, Roman, USA
Aniansson, Jockum, Sweden
Ansorge, Rainer, BRD
Antosik, Piotr, Poland
Anvari, Morteza, Iran
Apéry, Roger, France
Apiola, Heikki K., Finland

Apostol, Constantin, Romania
Appell, Jürgen, BRD
Araki, Huzihiro, Japan
Araki, Shoro, Japan
Arbarello, Enrico, Italy
Archbold, Robert John, UK
Archinard, Gabriel, Switzerland
Arena, Orazio, Italy
Arens, Richard F., USA
Arjas, Elja, Finland
Armiñjon, Paul, Canada
Armitage, David H., UK
Armsen, Moritz, BRD
Aroca, Jose M., Spain
Arosio, Alberto, Italy
Arrowsmith, David Keith, UK
Arsene, Grigore, Romania
Artemiadis, Nicolas, Greece
Artola, Michel, France
Artzy, Rafael, Israel
Asada, Teruko, Japan
Asano, Hiroshi, Japan
Aschbacher, Michael, USA
Ash, Avner, USA
Ash, J. Marshall, USA
Ashman, Charles John, Australia
Ashour, Attia A., Egypt
Assadi, Amir Hossein, Iran
Astala, Kari O., Finland
Atiyah, Michael Francis, UK
Aubert, Karl Egil, Norway
Aubert, Pierre-Louis, Switzerland
Aulaskari, Rauno, Finland
Ault, John Clifford, UK
Aumann, Robert J., Israel
Aupetit, Hubert, France
Aurora, Silvio, USA
Avila-Murillo, Fernando, Mexico
Axelsson, Reynir, Iceland
Ayala-Zazueta, Ignacio, Mexico
Aytuna, Aydin, Turkey
Azari, Abdolrahman S., Iran

Ba, Boubakar, Niger
Baas, Nils Andreas, Norway
Baayen, P. C., Netherlands
Babai, László, Hungary
Babu, G. Jogesh, India
Bachelis, Gregory, USA
Backelin, Jörgen, Sweden
Backhouse, Nigel Bristow, UK
Bäckström, Karl-Johan, Sweden
Bäckström, Lars, Sweden
Badji, Cherif, Senegal
Baernstein, Albert, II, USA
Bailey, G. H., UK
Bailey, Rosemary Anne, UK
Baillette, Aimée, France
Baily, Walter L., Jr., USA
Bajpai, Avinash Chandra, UK
Bajpai, Shri Dhar, Nigeria
Baker, Alan, UK
Baker, Bennett Mitchell, USA
Bakken, Ivar, Norway
Balabane, Mikhael, France
Baldoni Silva, Maria Welleda, Italy
Ballew, Ande C., USA
Balogh, Zoltán, Hungary
Bamba, Kante Siaka, France
Banchoff, Thomas, USA
Banulescu, Marta, Romania
Barbance, Christiane, France
Barker, Simon Ross, UK
Barlotti, Adriano, Italy
Barnard, Anthony David, UK
Barner, Martin, BRD
Barnes, Frank W., Kenya
Barone, Enzo, Italy
Barou, Geneviève, France
Barre, Raymond, France
Barrucand, Pierre, France
Barsky, Daniel, France
Barth, Karl F., USA
Bartle, Robert G., USA
Bartoszewicz, Jarosław, Poland
Bass, Hyman, USA
Bassotti, Lucilla, Italy
Bastero Eleizalde, Jesús, Spain
Batadila, Raphael, Congo
Batle Nicolau, Nadal, Spain
Batty, C. J. K., UK
Bauer, Heinz, BRD
Bauer, Walter, Austria
Baum, John D., USA
Bavinck, Herman, Netherlands
Baxhaku, Shaban, Albania
Baxter, J. P., Australia
Bayoumi, Aboubakr El-S. A., Sweden
Beard, James, New Zealand
Beauville, Arnaud, France
Beauzamy, Bernard, France
Beck, István, Norway
Becker, Gerhard, BRD
Becker, Hans J., BRD
Becker, Helmut, BRD
Becker, Jochen, BRD
Beckert, Herbert, DDR
Beckmann, William, USA
Beckner, William, USA
Beetham, Michael John, UK
Behboodian, Javad, Iran
Behforouz, Ahmad, Iran
Behnam, Ebrahim, Iran
Behncke, Horst, BRD
Behrendt, Willi, BRD
Behzad, Mehdi, Iran
Bejancu, Aurel, Romania
Beker, Henry J., UK
Bell, David John, UK
Bell, Howard E., Canada
Bellaiche, André, France
Bellomo, Nicola, Italy
Bellot Rosado, Francisco, Spain
Beltrametti, Mauro, Italy
Benabou, Jean, France
Benedicks, C. O. Michael, Sweden
Bengel, Gunter, BRD
Ben-Israel, Adi, USA
Benning, Hans Peter, BRD
Benzaghou, B., Algeria
Berard Bergery, Lionel, France
Bercov, Ronald, Canada
Berg, Gunnar, Sweden
Berger, Mel S., USA
Berger, Thomas Robert, USA
Bergh, Jöran, Sweden
Bergman, George Mark, USA
Berkovitz, Leonard D., USA
Berman, Joel, USA
Bernard, Alain, France
Bernat, Pierre, France
Berndt, Rolf, BRD
Berndtsson, Bo Göran, Sweden
Berrick, Jon, UK
Bers, Lipman, USA

Berthier, Anne Marie, France
Bertin, Emile, Netherlands
Bertin, Jean-Etienne, France
Bertin, Marie-José, France
Bertrand, Daniel, France
Besenfelder, Hans-J., BRD
Bessaga, Czesław, Poland
Beyon, Gerard V., Netherlands
Bezuszka, Stanley, USA
Bicadze, A. V., USSR
Bierbaum, Fritz, BRD
Bierstone, Edward, Canada
Biggs, Richard G., Canada
Billotti, Joseph E., USA
Bing, R. H., USA
Bingen, Franz, Belgium
Bingener, Jürgen, BRD
Birch, Bryan John, UK
Birman Joan, USA
Birnbaum, Z. William, USA
Bismut, Jean-Michel, France
Bix, Michael, USA
Bjon, Sten Gustav, Finland
Björck, Göran, Sweden
Björk, Anne, Sweden
Björnsson, Ottó, Iceland
Blackburn, Norman, UK
Blåfield, Eero Antero, Finland
Blanc, Philippe G. A., France
Blancheton, Eliane, France
Blank, Brian E., Canada
Blanton, John D., USA
Blasius, Don M., USA
Blass, Ándreas, USA
Blatter, Christian, Switzerland
Blersch, Hans, BRD
Blevins, Donald, USA
Bloch, Spencer, USA
Block, Richard E., USA
Bloom, Walter R., Australia
Blum, Lenore, USA
Boardman, Elizabeth M., UK
Boboc, Nicu, Romania
Bobrowski, Dobiesław, Poland
Bočkarev, S. V., USSR
Böddeker, Werner, BRD
de Boer, Jan Hendrik, Netherlands
Boers, Arie Hendrik, Netherlands
Boettcher, Heinz, BRD
Boffi, Vinicio, Italy
Bogojavlenskii, O. I., USSR
Bogomolov, F. A., USSR
Böhme, Reinhold, BRD
Bohnké, Georges, France
Bohun-Chudyniv, Boris, USA
Bohun-Chudyniv, Wolodymyr, USA
Boileau, André, Canada
Bojarski, Bogdan, Poland
Boju, Valentin, Romania
Bokowski, Jürgen, BRD
Boman, Jan, Sweden
Bona, Jerry, USA
Bongaarts, Peter J. M., Netherlands
Bonsdorff, Eero, Finland
Bonsdorff, Heikki, Finland
Boo, Per-Anders, Sweden
Booss, Bernhelm, Denmark
Borcea, Ciprian Sorin, Romania
Borel, Armand, USA
Börger, Reinhard, BRD
Borgersen, Hans Erik, Norway
Böröczky, Károly, Hungary
Boron, Leo F., USA
Borovkov, A. A., USSR
Borucka-Cieślewicz, Anna, Poland
Bosák, Juraj, Czechoslovakia
Bose, Raj Chandra, USA
Bothner, Albrecht E., BRD
Bouche, Liane, France
Bouhoyi, Hilaire, Congo
Bouraoui, Lazhar, France
Bourgin, David G., USA
Bourguignon, Jean Pierre, France
Boutet de Monvel, Louis, France
Boutot, Jean-François, France
Bouvier, Alain-Jean, France
Bowden, Leon, Canada
Božović, Nataša, Yugoslavia
Brackx, Freddy F., Belgium
Braconnier, Jean F., France
Brand, Ronald, Netherlands
Brändli, Emil R., Switzerland
Brannan, David A., UK
Branner-Jørgensen, Bodil, Denmark
Bratteli, Ola, Norway
Brattström, Gudrun B., Sweden
Braun, Hel, BRD
Braun, Robert B., BRD
Bredimas, Antoine, France
Brehm, Ulrich, BRD
Brennan, James E., USA
Bresqvar, Anna Maria, Italy
Bressan, Aldo, Italy
Brewer, Dennis W., USA
Brezuleanu, Alexandru, Romania
Bridges, Douglas S., UK
de Brieder, Jan Eduard, Netherlands
Briem, Eggert, Iceland
Brieskorn, Egbert, BRD
Brillard, Alain, France
Brisley, Warren, Australia
Britton, John Leslie, UK
Brkić, Jagoda, Yugoslavia
Broman, Arne, Sweden
Brooks, Foster, USA
Bros, Jacques, France
Browder, Felix, USA
Browder, William, USA
Brown, Aldric L., UK
Brown, Clifford, BRD
Brown, Gordon, USA
Brown, Kenneth S., USA
Brown, Leon, USA
Brown, Mark Robbin, USA
Bruckner, Andrew, USA
Bruhn, Marion, BRD
de Bruin, Marcel G., Netherlands
Bruna, Joaquim, Spain
Brunel, Antoine, France
Brüning, Jochen, BRD
Bruno, A. D., USSR
Brunotte, Horst, BRD
Brunovský, Pavol, Czechoslovakia
Bruns, Gunter W., Canada
Brunstad, Solmund, Norway
Bryant, John L., USA
Bryant, Roger Michael, UK
Brzezinski, Juliusz, Sweden
Buchner, Michael A., USA
Buchweitz, Ragnar-Olaf, BRD
Bucur, Gheorghe, Romania

Budach, Lothar, DDR
Budinský, Bruno, Czechoslovakia
Buianouckas, Francis, USA
Bui Doan, Khanh, France
Bulatovic, Jelena, Yugoslavia
Bulgarelli, Ulderico P., Italy
Bullen, Peter, Canada
Bumby, Richard T., USA
Buraczewski, Adam, Poland
Burau, Werner, BRD
Burbea, Jacob, USA
Burckel, Robert, USA
Burde, Gerhard, BRD
Bureau, Florent J., Belgium
Burgess, C. E., USA
Burkhardt, Hugh, UK
Burlet, Oskar, Switzerland
Burnat, Marek, Poland
Burns, Keith Howard, Australia
Butković, Davor, Yugoslavia
Büyükyenerel, Gültekin, Turkey
Byatt-Smith, John Graham UK

Caballero, Emilia, Mexico
Cagnac, Francis, France
Cain, Bryan E., USA
Çakar, Öner, Turkey
Calais, Josette, France
Calder, Allan, UK
Calderón, Alberto P., USA
Calleb, Benjamin Oyuke, Kenya
Calloway, Jean M., USA
Cámera, Gerardo Agustín, Venezuela
Camina, Alan Robert, UK
Campbell, Colin M., UK
Campbell, Douglas M., USA
Campbell, L. Lorne, Canada
Campillo Lopez, Antonio, Spain
Čanak, Bosiljka, Yugoslavia
Čanak, Miloš, Yugoslavia
Čangalović, Mirjana, Yugoslavia
Cannon, James W., USA
Cano-Garcés, Jesús, Mexico
Cappell, Sylvain, USA
Caraman, Petru, Romania
Carbone, Luciano, Italy
Carleson, Lennart, Sweden
Carlsson, Gunnar E., USA
Carlsson, Renate, BRD
do Carmo, Manfredo, Brazil
Carmona, Jacques, France
Carmona, René, France
Carss, Marjorie C., Australia
Carter, James W., USA
Carter, Michael R., New Zealand
Carton, Edmond D. J., Belgium
Cartwright, Mary L., UK
Casselman, William, Canada
Cassels, John W. S., UK
Cassidy, Charles, Canada
Casson, Andrew John, UK
Casulli, Vincenzo, Italy
Catana, Viorel, Romania
Cattaneo, Carlo, Italy
Cattaneo Gasparini, Ida, Italy
Ceausescu, Zoia-Elena, Romania
Cegrell, Urban, Sweden
Çelebi. A. Okay, Turkey
Celli, Andrea, Italy
Cerf, Jean, France
Cerqueira, Maria Helena Lanat Pedreira de, Brazil
Chacron, Maurice, Canada
Chambon, Claudie, France
Chandler, Bruce, USA
Chandrasekharan, Komaravolu, Switzerland
Chang, Guillermo, Venezuela
Charatonik, Janusz Jerzy, Poland
Charbonnel, Anne-Marie, France
Chatelet, Gilles, France
Chatterji, S. D., Switzerland
Chaves, Manuel, Portugal
Chazarain, Jacques, France
Chébli, Houcine, Tunisia
Chee, P. S., Malaysia
Cheeger, Jeff, USA
Cheney, Ward, USA
Cheng, Charles, USA
Cheniot, Denis, France
Chenon, René, France
Cherlin, Gregory, USA
Chern, Shiing S., USA
Chétivaux, Françoise, France
Cheung, Yiu-Sing, France
Chevalier, Michel V., France
Chevallier, Benoît, France
Chew, Kim-Peu, Malaysia
Chi, Dong Pyo, Korea
Chinn, William G., USA
Choi, Man-Duen, Canada
Chong, Chi-Tat, Singapore
Chong, Frederick, Australia
Choodnovsky, David
Choodnovsky, Gregory
Choquet, Gustave, France
Christ, Lily E., USA
Christiansen, Bent, Denmark
Ciesielski, Zbigniew, Poland
Ciliberto, Carlo, Italy
Ciliberto, Ciro, Italy
Çinlar, Erhan, USA
Cioranescu, Doina-Maria, France
Cîrnu, Mircea, Romania
Claesson, Tomas, Sweden
Clancey, Kevin, USA
Clarke, Frank H., Canada
Clarke, Gerard John, UK
Cleja, Sanda, Romania
Clément, Philippe, Netherlands
Clements, John C., Canada
Clozel, Laurent Y., France
Coates, John H., Australia
Coatmélec, Christian, France
Coelho, Michel, France
Coeure, Gérard, France
Cohen, Daniel I. A., USA
Cohen, David Mordechai, USA
Cohen, Frederick R., USA
Cohen, Gérard, France
Cohen, Henri J., France
Cohen, Marshall M., USA
Cohen, Moses E., USA
Cohn, Harvey, USA
Cohn, Richard M., USA
Cole, Nancy, USA
Coleman, Albert John, Canada
Collino, Alberto, Italy
Collins, Peter J., UK
Colliot-Thélène, Jean-Louis, France
Colojoara, Ion, Romania
Colombini, Ferruccio, Italy
Combes, François, France

Combes, Jean, France
Comfort, W. Wistar, USA
Compta, Albert, Spain
Conduché, Daniel, France
Conlan, James, Canada
Connell, Ian, Canada
Connelly, Robert, USA
Connes, Alain, France
Connolly, Dennis M., Canada
Connolly, Frank, USA
Constantinescu, Adrian, Romania
Conte, Alberto, Italy
Contou-Carrère, Carlos, France
Conway, J. H., UK
Cooney, Thomas J., USA
Coray, Daniel, Switzerland
Cornalba, Maurizio, Italy
Cornea, Aurel, Romania
Costabel, Martin, BRD
Costinescu, Olga, Romania
Cotlar, Mischa, Venezuela
Coudrais-Zeghers, Joël, France
Couëtte, André, France
Cougnard, Jean, France
Cox, David, France
Crăciun, Constantin, Romania
Crapo, Henry, Canada
Craw, Ian G., UK
Croft, Hallard Thomas, UK
Crumeyrolle, Albert, France
Csákány, Béla, Hungary
Császár, Ákos, Hungary
Csetényi, Arthur, Hungary
Csiszár, Imre, Hungary
Cullen, Helen Frances, USA
Cushing, Jim, USA
Cushman, Richard, Netherlands
Cusick, Thomas W., USA
Cutler, Doyle, USA
Cycon, Hans L., BRD
Czechowski, Tadeusz, Poland

Dacunha-Castelle, Didier, France
Dague, Philippe, France
Dalaud, Levry Bogard, Ivory Coast
Dale, Knut Th., Norway
Dałek, Krystyna, Poland
Dalla Volta, Vittorio, Italy
D'Ambrosio, Ubiratan, Brazil
Dancer, Edward Norman, Australia
Dancis, Jerome, USA
Dankel, Thad, USA
Dastrange, Nasser, Iran
Datta, Dilip Kumar, USA
Davarashvili, Margaret, Israel
Daverman, Robert J., USA
Davis, Martin D., USA
Davis, Michael, USA
Dawson, Donald A., Canada
Dazord, Jean, France
de Bartolomeis, Paolo, Italy
de Boor, Carl, USA
Deddens, James A., USA
Dedecker, Paul, Belgium
Dehesa, Jesus S., Spain
Delande, Georges, Belgium
Delange, Hubert, France
Delavault, Huguette, France
Delessert, Andre, Switzerland
Deligne, Pierre, Belgium
Dellacherie, C., France
Delporte, Jean, France
Delver, Robert, Canada
De Marr, Ralph Elgin, USA
Dencker, Nils, Sweden
Dennis, R. Keith, USA
Depaix, Michel, France
Derdziński, Andrzej, Poland
Derrick, John, UK
Derriennic, Marie Madeleine, France
Deschamps, Mireille, France
Deshouillers, Jean-Marc, France
Desq, Roger, France
Dexeus, Josefina, Spain
Dey, Suhrit K., USA
Deza, Mikhail, France
Dia, Galaye, Senegal
Diaz López, Julia Maria, Spain
Dienst, Karl Jakob, BRD
Diers, Yves, France
Digernes, Trond, Norway
Digne, François, France
van Dijk, Gerrit, Netherlands
Dimca, Alexandru, Romania
Dinh Van Huynh, Vietnam
Dionne, Philippe-Auguste, France
DiPerna, Ronald, USA
Dixmier, Jacques, France
Dixmier, Suzanne, France
Dixon, John D., Canada
Djedour, M., Algeria
Djobidja, Maurice, Congo
Djoković, Dragomir, Canada
Dlab, Vlastimil, Canada
Dłotko, Tadeusz, Poland
Dobber, Eelkje, Netherlands
Dodds, Brian P., UK
Dodziuk, Jozef, USA
Dolbeault, Pierre, France
Dolbeault, Simone, France
Dolecki, Szymon, Poland
Doležal, Vladimír, Czechoslovakia
Domar, Yngve, Sweden
Domzał, Krystyna S., Poland
Donaldson, J. A., USA
Donoghue, William F., Jr. USA
Doppel, Karl, Austria
Dörfler, Willibald, Austria
Dorfmeister, Josef, BRD
Dorn, Gernot, BRD
Dotlić, Branimir, Yugoslavia
Douady, Régine, France
Douglas, Ronald G., USA
Dowling, Michael L., UK
Dowson, H. R., UK
Doyen, Jean, Belgium
Doyle, Peter G., USA
Drasin, David, USA
Draxl, Peter, BRD
Drinfeld, V. G., USSR
Drinkwater-Roberts, Anne, USA
Drobot, Stefan, USA
Dryja, Maksymilian, Poland
Dubejko, Maciej, Poland
Dubinsky, Ed, USA
Dubnicki, Walerian, Poland
Dubois, Donald, USA
Dubois, Eugene, France
Dubois, Jacques, Canada
Du Bois, Philippe, France
Dubson, Alberto Samuel, Argentina

Duff, George F. D., Canada
Dugac, Pierre, France
Dugdale, Jack Keith, UK
Duggal, Krishan Lal, Canada
Duhaut, Guy, France
Duncan, Tyrone Edward, USA
Dunkels, Andrejs, Sweden
Dunwoody, Martin John, UK
du Plessis, Andrew Allan, Denmark
du Plessis, Nicolaas, UK
Dupont, Johan L., Denmark
Durfee, Alan H., USA
Durrett, Richard, USA
Dvoretzky, Aryeh, Israel
Dwilewicz, Roman, Poland
Dwork, Bernard, USA
Dyckhoff, Roy, UK
Dynkin, Eugene, USA
Dzinotyiweyi, H. A. M., Netherlands
Džuraev, A. D., USSR

Earle, Clifford, J., USA
Eastwood, Michael George, UK
Eberhard, Walter, BRD
Eberlein, Ernst Wilhelm, BRD
Ebin, David G., USA
Eckert, Ernest J., Denmark
Eckhoff, Jürgen, BRD
Eckmann, Beno, Switzerland
Edamatsu, Takasi, Japan
Edmonds, Allan, USA
Edwards, David Albert, UK
Edwards, Harold M., USA
Edwards, Robert, USA
Eells, James, UK
Efimov, A. V., USSR
Egghe, Leo, Belgium
Egorov, J. V., USSR
Eguchi, Kazuo, Japan
Ehrlich, Karel, USA
Einsele, Carl, Switzerland
Eisele, Carolyn, USA
Eising, Jens, Denmark
Ejike, Uwadiegwu, Nigeria
Ekeland, Ivar, France
Eklof, Paul C., USA
Ekong, Samuel D., France
Elfving, Gustav, Finland
El-Gendi, Salah E., Egypt
Eliaš, Jozef, Czechoslovakia
Elíasson, Halldór I., Iceland
Elie, Laure, France
Eljoseph, Nathan, Israel
El Kolli, Amar, Algeria
Ellingsrud, Geir, Norway
Elliott, George A., Denmark
Elliott, Peter, USA
Elsner, Jürgen, BRD
El Tom, Mohamed, Sudan
Elzein, Fouad, France
Emamirad, Hassan Ali, Iran
Emsalem, Jacques, France
Endo, Mikihiko, Japan
Eneduanya, Sylvanus Ndulue, Nigeria
Engenes, Hans, Norway
Ennola, Veikko, Finland
Epheser, Helmut, BRD
Epstein, Mordechai, Israel
Erbach, David W., USA
Erdös, Paul, Hungary
Eriksson, Folke, Sweden
Erkama, Timo, Finland
Erkip, Albert K., Turkey
Erlander, Sven, Sweden
Erlandsson, Thomas, Sweden
Ernest, John A., USA
Ernst, Bruno, BRD
Ernvall, Reijo G., Finland
Eršov, J. L., USSR
Eslinger, Robert, USA
Esnault, Hélène, France
Espelie, Solveig, USA
Etcheberry, Alain, Venezuela
Evans, David Emrys, Denmark
Everitt, William Norrie, UK
Ewert, Janina, Poland
Ezeilo, James O. C., Nigeria
Ezin, Jean-Pierre, Benin

Fack, Thierry, France
Fantodji, C. Pierre, Ivory Coast
Faour, Nazih S., Kuwait
Farkas, Hershel, Israel
Farkas, Miklós, Hungary
Faulstich, Karin, BRD
Faxén, Birger, Sweden
Fefferman, Charles, USA
Fefferman, Robert, USA
Fegan, Howard David, UK
Fein, Burton, USA
Feinberg, Jerry M., USA
Fell, James Michael, USA
Fellman, Johan, Finland
Fenyö, István S., Hungary
Férentinou-Nicolacopoulou, Jeanne, Greece
Ferguson, Le Baron O., USA
Fernandez, Cesar, BRD
Ferrarese, Giorgio, Italy
Feshbach, Mark, USA
Fetter, Helga, Mexico
Fidelis, Eugeniusz, Poland
Fiedler, Miroslav, Czechoslovakia
Fieldhouse, David, Canada
Figà-Talamanca, Alessandro, Italy
Figiel, Tadeusz, Poland
De Figueiredo, Djairo, Brazil
Filippenko, Ivan Vladimir, USA
Finkel, David, USA
Fischer, Bernd, BRD
Fischer, Gerd, BRD
Fischer, H. Wolfgang, BRD
Fishman, Karl, Israel
Fiskaali, Juhani, Finland
Flanders, Harley, USA
Fleckinger, Jacqueline, France
Flegg, Henry G., UK
Flenner, Hubert, BRD
Fletcher, Beryl, UK
Fletcher, Peter, USA
Fletcher, Trevor James, UK
Flexor, Marguerite, France
Fliess, Michel, France
Flondor, Paul, Romania
Flores-Espinoza, Ruben, Mexico
Floyd, William J., USA
Foiaş, Ciprian Ilie, Romania
Foltynska, Izabela, Poland
Forster, Otto, BRD
Forster, Walter, UK
Fourman, Michael Paul, UK
Fox, Ralph, UK
Fragozo Robles, Arturo, Mexico
Françoise, Jean-Pierre, France

Frank, Evelyn, USA
Frank, Günter, BRD
Frank, Leonid S., Netherlands
Frankel, Robert, USA
Franks, John M., USA
Fredriksson, Einar, Netherlands
Freedman, Haya, UK
Frei, Gunther, Canada
Fremlin, David Heaver, UK
Fresnel, Jean, France
Fried, David, USA
Fried, Ervin, Hungary
Friedrichs, Kurt O., USA
Fritsch, Rudolf, BRD
Frock, Beatrice, Sweden
Fröhlich, Albrecht, UK
Fröhlich, Jürg M., France
Fuchs, Wolfgang Heinrich, USA
Fuchssteiner, Benno, BRD
Fuglede, Bent, Denmark
Fuji'i'e, Tatsuo, Japan
Fujisaki, Masahiko, Japan
Fukuda, Isamu, Japan
Fukushima, Masatoshi, Japan

Gackstatter, Fritz, BRD
Gaeta, Federico, Spain
Gaffney, Matthew, USA
Gaffney, Terence, USA
Gagnon, François P., Canada
Galajda, Pavel, Czechoslovakia
Galbraith, Alan Stuart, USA
Galperin, Efim A., Canada
Gapaillard, Christiane, France
Gapaillard, Jacques, France
Garançon, Maurice, Canada
Garberson, John D., USA
Garcia Pradillo, Julio, Spain
Garcia Roig, Jaime Luis, Spain
Gårding, Lars, Sweden
Gardner, Robert B., USA
Garg, Krishna M., Canada
Garling, David J. H., UK
Garnir, Henri G., Belgium
Garsia, Adriano M., USA
Gashi, Emrush, Yugoslavia
Gaşpar, Dumitru, Romania
Gatesoupe, Michel, France
Gatteschi, Luigi, Italy
Gauld, David Barry, New Zealand
Gauthier, Paul M., Canada
Gauthier, Yvon, Canada
Gautrin, Henri-François, Canada
van der Geer, Gerard, Netherlands
Gehring, Frederick W., USA
Geivaerts, Marcel, Belgium
Gelfgren, Jan K., Sweden
Geoghegan, Ross, USA
Georgijević, Dušan, Yugoslavia
Gérardin, Paul, France
Gerdes, Paulus, Mozambique
Germain, Gilles, France
Germay, Noël, Belgium
Germer, Henning, BRD
Gerner, Helmut, BRD
Gervais, Jean-Jacques, Canada
Gewalter, Lutz, BRD
Ghaffari, Abolghassem, USA
Giacardi, Livia, Italy
Gieseker, David, USA
Gilbarg, David, USA
Gilfeather, Frank, USA
Gillis, Joseph, Israel
Gilman, Robert, USA
Gilmer, Patrick M., USA
Giraud, Jean, France
Gitler, Samuel, Mexico
Giusti, Enrico, Italy
Giusti, Marc, France
Gjone, Gunnar, Norway
Glashoff, Klaus, BRD
Głazek, Kazimierz, Poland
Glicksberg, Irving, USA
Glikson, Aleksander, Australia
Globevnik, Josip, Yugoslavia
Glocke, Theo, DDR
Glötzl, Erhard, Austria
Godet-Thobie, Christiane, France
Godini, Gliceria, Romania
Goebel, Kazimierz, Poland
Göktürk, Zerrin, Turkey
Golan, Jonathan S., Israel
Goldfeld, Dorian, USA
Goldman, Jay, USA
Goldman, William M., USA
Goldys, Dorothea, USA
Golumbic, Martin, USA
Gomez-Larrañaga, Jose-Carlos, Mexico
Gommers, Antonius M. M., Netherlands
Gončar, A. A., USSR
Gondard, Danielle, France
Gondran, Michel, France
Gonshor, Harry, USA
Gonzalez, Eduardo, Italy
Gonzalez Lopez, Francisca, Spain
Gonzalez Sprinberg, Gerardo, France
Good, Anton, Switzerland
Gopalsamy, Kondalsamy, Australia
Gordon, Cameron McAllan, USA
Gorenstein, Daniel, USA
Gorostiza, Luis G., Mexico
Gottlieb, Daniel Henry, USA
Gould, Gerald G., UK
Gould, Sydney H., China-Taiwan
Gowrisankaran, Kohur, Canada
de Graaf, Johannes, Netherlands
Graf, Siegfried, BRD
Graffi, Sandro, Italy
Graham, Ian, Canada
Graham, Victor W., Ireland
Gramain, André, France
Granirer, Edmond E., Canada
Ganlund, Seppo, Finland
Grant, Douglass L., Canada
Gràsser, Horst-Siegfried Paul, South Africa
Grassman, Eckhard G., Canada
Grasu, Constantin, Romania
Gray, Alfred, USA
Gray, Mary, USA
Grayson, Daniel R., USA
Green, Edward L., USA
Green, Judy, USA
Green, Paul S., USA
Greenberg, Ralph, USA
Gregorio, Carlos-Germán, Argentina
Greguš, Michal, Czechoslovakia

Greig, Dorothy Margaret, UK
Greim, Peter, BRD
Grenander, Ulf, Sweden
Grennberg, Anders, Sweden
Greuel, Gert-Martin, BRD
Greve, Georg, BRD
Griess, Robert, USA
Griffiths, H. Brian, UK
Griffiths, Phillip A., USA
Grimaldi, Anna, Italy
Grimmer, Ronald, USA
Grioli, Giuseppe, Italy
Gripenberg, Gustaf, Finland
Grispolakis, Joachim, Canada
Groenewegen, G. L. M., Netherlands
Gromov, Mikhael, USA
Gronau, Detlef, Austria
Grossman, Jerrold W., USA
Grosswald, Emil, USA
Grotemeyer, Karl Peter, BRD
Grotz, Wolfgang, BRD
Grubb, Gerd, Denmark
Gruber, Manfred A., BRD
Gruber, Peter M., Austria
Grünbaum, Branko, USA
Gualtierotti, Antonio F., Switzerland
Guerindon, Jean, France
Guerra Vazquez, Francisco, Cuba
Guessarian, Iréne, France
Guest, Martin Andrew, UK
Guichardet, Alain, France
Guillemin, Victor, USA
Guiro, Abdoulaye, Senegal
Gundlach, Karl-Bernhard
Gündüzalp, Yavuz, Turkey
Gunnarsson, Thomas E. W., Sweden
Gupta, Chander Kanta, Canada
Gupta, Narain, Canada
Gupta, Shiv Kumar, India
Gurarie, David, Israel
Gussi, Gheorghe, Romania
Gustafson, Sven-Åke, Sweden
Gvozdenović, Dušan, Yugoslavia
Gyires, Béla, Hungary
Györy, Kálmán, Hungary
Haahti, Heikki, Finland
Haapasalo, Lenni I., Finland
Haario, Heikki E., Finland
Haber, Seymour, USA
Haefliger, Andre, Switzerland
Hag, Kari, Norway
Hag, Per, Norway,
Haghany, Ahmad, Iran
Haghany, Ghodsieh, Iran
Haifawi, Mahmoud, Turkey
Hajarnavis, Charudatta R., UK
Hajduk, Bogusław, Poland
Hajnal, András, Hungary
Haken, Armin, USA
Haken, Wolfgang, USA
Halberstam, Heini, UK
Haliste, Kersti, Sweden
Hall, Marshall, USA
Halpern, James D., USA
Hämäläinen, Raimo P., Finland
Hamidi, Massih, Iran
Hanner, Olof, Sweden
Hannsgen, Kenneth B., USA
Hanouzet, Bernard, France
Hansen, Sönke, BRD
Harary, Frank, USA
Hardie, Keith A., South Africa
Hardorp, Detlef, USA
Hardy, Kenneth, Canada
Harer, John L., USA
Harju, Tero J., Finland
Harlamov, V. M., USSR
Harlow, Donald William, UK
Harrington, Leo, USA
Harris, Michael, USA
Harrison, Kenneth J., Australia
Harrold, Orville G., USA
Hart, Neal, Botswana
Härtig, Klaus, DDR
Hartshorne, Robin, USA
Haruki, Hiroshi, Canada
Harvey, Josephat Martin, Rhodesia
Harzallah, Khelifa, Tunisia
Harzheim, Egbert, BRD
Hashiguchi, Masao, Japan
Hassard, Brian David, USA
Hastings, Harold M., USA
Hasumi, Morisuke, Japan
Hatcher, Allen E., USA
Hatori, Tsukasa, Japan
Hattori, Akio, Japan
Haudidier, Dominique, France
Haugazeau, Yves, France
Hausmann, Jean-Claude, Switzerland
Hawkes, John, UK
Hawkes, Trevor Ongley, UK
Hawthorne, William, Nigeria
Hayasaka, Tomohide, Japan
Hayashi, Hiroshi, Japan
Hayashida, Tsuyoshi, Japan
Haydon, Richard, UK
Hazewinkel, Michiel, Netherlands
Hecht, Henryk, USA
Hedberg, Lars Inge, Sweden
Hedberg, Torbjörn, Sweden
Heden, Olof, Sweden
Heersink, Rudolf, Austria
Heiede, Torkil, Denmark
Heikkilä, Seppo V., Finland
Heil, Erhard, BRD
Heil, Wolfgang H., USA
Heinrich, Elkedagmar, BRD
Heinrich, Stefan, DDR
Heintze, Ernst, BRD
Heinze, Joachim, BRD
Hejcman, Jan, Czechoslovakia
Heldermann, Norbert C., BRD
Helffer, Bernard, France
Hell, Pavol, USA
Hellberg, H. Stefan, Sweden
Hellerstein, Simon, USA
Helm, Werner, E., BRD
Helmes, Kurt, BRD
Helminck, G. F., Netherlands
Helversen-Pasotto, Anna, France
Henč, Damir, Yugoslavia
Hendericks, Richard, Netherlands
Hengartner, Walter, Canada
Henlein, L., USA
Hennekemper, Wilhelm, BRD
Henrichs, Rolf Wim, BRD
Herbut, Fedor, Yugoslavia
Herceg, Dragoslav, Yugoslavia
Herdman, Jerry L., USA
Herman, Gabor, USA

Herman, Michael Robert, France
Hermann, Peter, BRD
Herrero, Domingo, Venezuela
Herrmann, Oskar, BRD
Hersee, John William, UK
Herz, Jean-Claude, France
Herzog, Jürgen, BRD
Heywood, John G., Canada
Higashiyama, Teiko, Japan
Higgins, John C., USA
Higgs, Denis, Canada
Higuchi, Teiichi, Japan
Hill, C. Denson, USA
Hill, Theodore P., USA
Hiller, Howard, USA
Hilsum, Michel, France
Hinz, Jurgen, BRD
Hirashima, Hideharu, Japan
Hirakowa, Hiroshi, Japan
Hirota, Masaharu, Japan
Hirvonen, Kullervo, Finland
Hitchcock, Anthony Gavin, Rhodesia
Hitchin, Nigel James, UK
Hitotumatu, Sin, Japan
Hlawka, Edmund, Austria
Hoang Tuy, Vietnam
Hochster, Melvin, USA
Hocquenghem, Alexis, France
Ho Duc Viet, Vietnam
Hogbe-Nlend, Henri, Cameroon
Högnäs, Göran, Finland
Höhn. Werner, BRD
Holenda, Jiří, Czechoslovakia
Holland, Charles J., USA
Hollingsworth, John G., USA
Holm, Per, Norway
Holmåker, Kjell, Sweden
Holmann, Harald, Switzerland
Holme, Anne Berit Lunde, Norway
Holme, Audun, Norway
Holmes, Charles, USA
Holmström, Lasse, Finland
Holsti, Niklas, Finland
Holtzman, David N., Netherlands
Honary, Bahman, Iran
Hong, Imsik, Japan
Honkapohja, Matti, Finland
Honma, Toshihiro, Japan
Hood, Rodney T., USA
Hoppe, Ronald, BRD
Hörmander, Lars V., Sweden
Horn, Andreas, BRD
Horneffer, Klaus, BRD
Horst, Reiner, BRD
Houillot, Josette, France
Hovanskii, Askold G., USSR
Hovinen, Esa, Finland
Hsia, John S., USA
Hsiang, Wu-chung, USA
Hsiao, Hong-Jen, China-Taiwan
Hsu, Tao-ning, China-Taiwan
Huard, Pierre, France
Hubbard, John H., USA
Huber, Alfred, Switzerland
Hudson, Donna L., USA
Hueber, Hermann, BRD
Huebschmann, Johannes, BRD
Huet, Patrick, France
Hulek, Klaus W., BRD
Humke, Paul D., USA
Hunter, John, UK
Hurley, Donal, Ireland
Hurrelbrink, Jürgen, BRD
Husain, Taqdir, Canada
Husch, Laurence, USA
Hutchinson, John Edward, Australia
Hveberg, Steinar, Norway
Hyers, Donald H., USA
Hyllengren, Anders, Sweden
Hyvärinen, Lassi P., Finland
Hyvönen, Jaakko, Finland
Hyyrynen, Pertti, Finland

Iagolnitzer, David, France
Iarrobino, Anthony A., Jr., USA
Ibisch, Horst, France
Ibrahim, Edward Mikhail, Egypt
Ichikawa, Kaoru, Japan
Iio, Tsutomu, Japan
Ikegami, Teruo, Japan
Iliadis, Stavros, Greece
Iliev, Ljubomir Georgiev, Bulgaria
Il'jašenko, J. S., USSR
Illman, Sören, Finland
Illusie, Luc, France
Imai, Masataka, Japan
Imai, Tyuiti, Japan
Imam, Magdy Mostafa, Egypt
Im Hof, Hans-Christoph, Switzerland
Infantozzi, Carlos Alberto, Uruguay
Inkeri, Kustaa, Finland
Inoue, Takakazu, Japan
Ion, Patrick D. F., UK
Iriarte de Puyau, Delia Esther, Argentina
Irmisch, Robert, BRD
Irwin, Michael Charles, UK
Isbell, John, USA
Ishida, Mitsuko, Japan
Ishihara, Ikuo, Japan
Ishihara, Shigeru, Japan
Isikawa, Osamu, Japan
Isobe, Kiro, Japan
Itagaki, Yoshio, Japan
Itô, Kiyosi, Japan
Ito, Takashi, USA
Ito, Yoshihiko, Japan
Ito, Yoshitada, Japan
Iversen, Birger, Denmark
Ivrii, V. A., USSR
Iwanaga, Yasuo, Japan
Iwaniec, Henryk, Poland
Iwaniec, Tadeusz, Poland
Iwanowski, Peter, BRD
Iwiński, Tadeusz W., Poland
Iyahen, Sunday O., Nigeria
Iyanaga, Shokichi, Japan
Izaddoustdar, Nourooz, Iran
Izuchi, Keiji, Japan

Jablonskii, S. V., USSR
Jackowski, Stefan, Poland
Jackson, Howard L., Canada
Jackson, Michael, USA
Jacob, Niels, BRD
Jacobsen, Edward
Jaegermann, Michał, Poland
Jafarian, Ali A., Iran
Jaffe, Arthur, USA
Jager, Thomas, USA
Jagers, Albertus A., Netherlands

Jahren, Bjørn, Norway
Jain, S. K., USA
Jakubczyk, Bronisław, Poland
Jakubowski, Zbigniew J., Poland
Jalali, Morteza, Iran
James, David M., USA
James, Donald G., USA
James, Gordon Douglas, UK
James, Ioan, UK
Janković, Zlatko, Yugoslavia
Jankowski, Wiktor, Poland
Jansen, Karl-Heinz, BRD
Janssen, Gerhard, BRD
Janssen, Klaus, BRD
Järnefelt, Gustaf, Finland
Jarraud, Pierre, France
Järvi, Pentti K., Finland
Jauhiainen, Jorma, Finland
Jauho, Leena-Maija, Finland
Jawerth, Björn, Sweden
Jaworowski, Jan, USA
Jaworski, John, UK
Jech, Thomas, USA
Jeltsch, Rolf, Switzerland
Jensen, Christian U., Denmark
Jeśmanowicz, Leon, Poland
Jetten, Antonius, Netherlands
Jetter, Kurt, BRD
Jeurnink, G. A. M., Netherlands
Joffe, Anatole, Canada
Johannson, Klaus, BRD
Johanson, Arnold, USA
Johansson, Bo I., Sweden
Johnsen, Eugene C., USA
Johnsen, Karsten, BRD
Johnson, David L., UK
Johnson, Jerry, USA
Johnson, Raymond, USA
Johnson, Roy A., USA
Johnson, Russell, USA
Johnstone, Iain Murray, Australia
Jones, J. D. S., UK
Joosten, Harry P. G., Netherlands
Joris, Henri, Switzerland
Józefiak, Tadeusz, Poland
Juhász, István, Hungary
Julliot, Denis, France
Jussila, Olli K., Finland
Justice, James H., USA
Jutila, Matti, Finland

Kaapke, Jürgen, BRD
Kac, Victor, USA
Kącki, Edward, Poland
Kadison, Richard V., USA
Kadoyama, Makoto, Japan
Kaftal, Victor, Israel
Kahanpää, Lauri V., Finland
Kähler, Erich, BRD
Kahn, Peter J., USA
Kahramaner, Suzan, Turkey
Kaijser, Sten, Sweden
Kaijser, Thomas, Sweden
Kajiwara, Joji, Japan
Kalka, Morris, USA
Källén, Anders, Sweden
van der Kallen, Wilberd L. J., Netherlands
Kalman, John Arnold, New Zealand
Kambayashi, Tatsuji, USA
Kamber, Franz W., USA
Kametani, Shunji, Japan
Kaminski, Michael, Israel
Kamowitz, Herbert, USA
Kamthan, Pawan K., India
Kaneko, Tetsuo, Japan
Kanemaki, Shoji, Japan
Kangasaho, Jukka, Finland
Kankaanpää, Jouni K., Finland
Kannappan, P. L., Canada
Kano, Takeshi, Japan
Kantola, Pertti, Finland
Kantor, Jean-Michel, France
Kantorovič, Leonid, USSR
Kantowski, Eleanore L., USA
Kaplan, Wilfred, USA
Kapoulitsas, George, Greece
Karamzadeh, O. A. S., Iran
Karhumäki, Juhani, Finland
Karkar, Mohamed T., France
Kärkkäinen, Petter O., Finland
Karlberg, Lennart, Sweden
Karlsson, Johan, Sweden
Karoubi, Max, France
Karras, Ulrich, BRD
Karrer, Guido, Switzerland
Karwot, Izydor
Kashimoto, Masahiko, Japan
Kashiwara, Masaki, Japan
Kátai, Imre, Hungary
Katase, Kiyoshi, Japan
Kato, Masahide, Japan
Katsura, Toshiyuki, Japan
Katz, Nicholas M., USA
Kaufmann-Buhler, Walter, USA
Kaul, Saroop Krishan, Canada
Kausto, Risto, Finland
Kawada, Yukiyoshi, Japan
Kawai, Sóichi, Japan
Kawamata, Yujiro, Japan
Kawamura, Shinzó, Japan
Kaya, Rüstem, Turkey
Kearsley, Mary J., UK
Kearton, Cherry, UK
Keedwell, Anthony Donald, UK
Keener, Lee Lanam, Canada
Kegel, Otto H., BRD
Kehayopulu, Niovi, Greece
Keller, H. Heinrich, Switzerland
Kellogg, Franklin, USA
Kempf, George, USA
Kenderov, Petar S., Bulgaria
Kenku, Monsur A., Nigeria
Kenmotsu, Katsuei, Japan
Kenney, Margaret, USA
Kennison, Lawrence S., USA
Keranto, Tapio O., Finland
Kerckhoff, Steven P., USA
Kerkyacharian, Gérard, France
Kerre, Etienne, Belgium
Kersten, Ina, BRD
Kerzman, Norberto, USA
Kesavan, Srinivasan, India
Kestenband, Barbu, USA
Ketonen, Jussi A., USA
Ketonen, Timo O. J., Finland
Keune, Frans Johan, Netherlands
Khairy, Amin Sedrac Mikhail, Egypt
Khalgui, Mohamed Salah, France
Khalifah, Mohammed, Egypt
Khan, M. Nazir, Pakistan

Khosrovshahi, Golamreza B., Iran
Khuri, Soumaya, Lebanon
Kiernan, Peter, Canada
Kiikka, Maire E., Finland
Kilambi, Shrinivasacharyulu, Canada
Kilpi, Yrjö O., Finland
Kiltinen, John O., USA
Kinnunen, Seppo, Finland
Kirby, Robion, USA
Kirillov, A. A., USSR
Kiselman, Christer O., Sweden
Kitaoka, Yoshiyuki, Japan
Kivelä, Simo K., Finland
Kiyek, Karlheinz, BRD
Kiyohara, Mineo, Japan
Klamkin, Murray S., Canada
Klein, Joseph, France
Klein, Peter, BRD
Kleine Büning, Hans, BRD
Kleinfeld, Erwin, USA
Kleisli, Heinrich Switzerland
Klemola, Tapio, Canada
Kleppe, Jan O., Norway
Klingen, Jutta, BRD
Klingen, Norbert, BRD
Kluvánek, Igor, Australia
Knapp, Anthony W., USA
Knauer, Ulrich, BRD
Kneser, Martin, BRD
Knežević, Julka, Yugoslavia
Knight, Frank, USA
Knight, Julia, USA
Knight, William, USA
Knill, Ronald, USA
Knödel, Walter, BRD
Knoebel, R. Arthur, USA
Kobayashi, Mitsuko, Japan
Kobayashi, Ugoro, Japan
Koblitz, Neal, USA
Koçak, Cevdet, Turkey
Kochman, Stanley, Canada
Kofner, Jacob, Israel
Koh, Eusebio Legarda, Canada
Koh, Khee-Meng, Singapore
Kohler-Jobin, Maria-Thérèse, Switzerland
Kohn, Joseph J., USA
Koizumi, Shoji, Japan
Kolář, Ivan, Czechoslovakia
Kolar-Anić, Ljiljana, Yugoslavia
Kolchin, Ellis R., USA
Kolodner, Ignace I., USA
Komatu, Yûsaku, Japan
Komuro, Hiroaki, Japan
Kondô, Motokiti, Japan
König, Gerhard, BRD
König, Heinz J., BRD
Königs, Robert Paul, BRD
Könkkölä, Kalle, Finland
Koornwinder, Tom H., Netherlands
Kopeć, Jósef, Poland
Kopfermann, Klaus, BRD
Kopociński, Bolesław, Poland
Koponen, Pirkko, Finland
Koprinsky, Stephan Ts., Bulgaria
Koranyi, Adam, USA
Korenblum, Boris, Israel
Korte, Bernhard, BRD
Kortram, Ronald, Netherlands
Kosačevskaja, Elena, USSR
Koschorke, Ulrich, BRD
Koshi, Shozo, Japan
Koski, Timo, Finland
Kosmann-Schwarzbach, Yvette, USA
Kosniowski, Czes, UK
Kostalskii, N. B., USSR
Kostrikin, Aleksej, USSR
Koszul, Jean-Louis, France
Kôta, Osamu, Japan
Kotman, Larry, USA
Kovačević, Ilija, Yugoslavia
Kozłowska, Grażyna, Poland
Kra, Irwin, USA
Kraines, David, USA
Krajňáková, Dorota, Czechoslovakia
Kraljević, Hrvoje, Yugoslavia
Krasner, Marc, France
Krasovskii, N. N., USSR
Krčmar, Emina, Yugoslavia
Krem, Alajos, Hungary
Kremer, Wilhelm, Sweden
Krempa, Jan, Poland
Krengel, Ulrich, BRD
Krgović, Dragica, Yugoslavia
Kristensen, Leif, Denmark
Krogstad, Harald E., Norway
Królikowski, Andrzej, Poland
Krych, Michał, Poland
Krylov, N. V., USSR
Krzywicki, Andrzej, Poland
Krzyż, Jan, Poland
Kubala, Jolanta Maria, Poland
Kubilius, Jonas, USSR
Kućmińska, Bogusława, Poland
Kudrjavcev, L. D., USSR
Kueviakoe, Assingbon David, Togo
Kufner, Alois, Czechoslovakia
Kuga, Michio, USA
Kühn, Burkhard, BRD
Kühnel, Wolfgang, BRD
Kuijk, Willem, Belgium
Kuiper, Nicolaas H., France
Kümmich, Frieder, BRD
Künzle, Hans Peter, Canada
Kuran, Ülkü, UK
Kurepa, Duro, Yugoslavia
Kuribayashi, Akikazu, Japan
Kuroda, Toru, Japan
Kurola, Osmo E. O., Finland
Kurss, Herbert, USA
Kusterer, Roland, BRD
Kuttner, Brian, UK
Kutzko, Philip, USA
Kuusalo, Tapani, Finland
Kwapisz, Marian, Poland

Laakso, Hannu V., Finland
Laasonen, Pentti, Finland
Laatio, Kaisa, Finland
Labelle, Gilbert, Canada
Labrèche, Martine, Canada
Labuda, Iwo, Poland
Labute, Jean-Paul, Canada
Labute, Tamara, Canada
Lacher, Robert C., USA
Lackschewitz, Christina B., BRD
Lacomba, Ernesto A., Mexico
Lacomblez, Chantal, France
Laczkovich, Miklós, Hungary
Laffey, Thomas J., Ireland
Lafon, Jean-Pierre, France
Lafon-Augé, Monique, France
Lagarias, Jeffrey C., USA
Lahtinen O. Aatos, Finland

Lai, Hang-Chin, China-Taiwan
Laine, Ilpo, Finland
Laine, Riitta A., Finland
Laird, Philip G., Australia
Laitinen, Erkki J., Finland
Lakshmikantham, Vangipuram, USA
Laksov, Dan, Norway
Lambert, Pol V. A. J., Belgium
Lamprecht, Erich, BRD
Lance, Christopher, UK
Landes, Rüdiger, BRD
Landman, Alan, USA
Landstad, Magnus B., Norway
Lang, Charles Alfred, UK
Lange, Horst R., BRD
Lange, Lester H., USA
Langevin, Michel, France
Langevin, Remi, France
Langlands, Robert P., USA
Lannes, Jean E., France
Lappalainen, Urpo J., Finland
Lara-Carrero, Lorenzo, Venezuela
Larman, David G., UK
Larmore, Lawrance L., USA
Larson, David R., USA
Lashof, Richard, USA
Lassner, Gerd, USSR
Latvamaa, Esko P. T., Finland
Lau, Anthony To-Ming, Canada
Laudal, Olav Arnfinn, Norway
Launonen, Eero, Finland
Laurence, Peter, USA
Lavasany, Mohammad-Taghi, Iran
Law, Alan G., Canada
Lawrence, John William, Canada
Lawrynowicz, Julian, Poland
Lax, Peter D., USA
Lax, Robert, USA
Laxton, Ronald R., UK
Laywine, Charles, Canada
Lazar, Aldo, Israel
Lé Dung Trang, Vietnam
Lebaud, C., France
Lebelle, Pierre, Netherlands
Leborgne, Daniel, France
Lebow, Arnold, USA
Lebrun, Christiane, Belgium
Lecacheux, Odile, France
Lech, Christer, Sweden
Lechicki, Alojzy, Poland
Ledgard, Ronald Clive, UK
Lee, Ronnie, USA
Leela, Srinivasa, USA
Leenders, Chrit, Netherlands
Leestma, Sanford, USA
Lefton, Phyllis, USA
Legrand, Denise, France
Legrand, Solange, France
Lehman, Eric, France
Lehmer, Derrich H., USA
Lehner, Ellen, USA
Lehner, Guydo, USA
Lehti, Raimo, Finland
Lehtinen, Erkki, Finland
Lehtinen, Markku S., Finland
Lehtinen, Matti, Finland
Lehtinen, Pekka I., Finland
Lehto, Olli, Finland
Lehto, Pertti J., Finland
Lehtonen, Ari T. E., Finland
Lehtonen, Tapani H., Finland
Leicht, Johann B., BRD
Leindler, László, Hungary
Leino, Jarkko, Finland
Leino, Päivi, S. I., Finland
Leinonen, Jorma K., Finland
Leipälä, Raija, Finland
Leipälä, Timo, Finland
Leitman, Marshall, USA
Lejeune-Jalabert, Monique, France
Lekkerkerker, Cornelis G., Netherlands
Lellahi, S. Kazem, Iran
Lelong, France, France
Lelong, Pierre, France
Lemaire, Luc, Belgium
Le Méhauté, Alain, France
Lenagan, Thomas H., UK
Lenzing, Helmut, BRD
Leontiev, A. F., USSR
Lepistö, Timo, Finland
Lepowsky, James, USA
Lesaffre, Brigitte, France
Lesieur, Léonce, France
Lesley, F. David., USA
Leszczyński, Bogumił, Poland
Letizia, Maurizio, Italy
Leung, Kam W., Hong Kong
Lê-van Thiêm, Vietnam
LeVeque, William, USA
Levesque, Claude, Canada
Levin, Jacob J., USA
Levoni, Sergio, Italy
Lewin, Mordechai, Israel
Lewowicz, Jorge, Venezuela
Libera, Richard, USA
Libermann, Paulette, France
Lickorish, W. B. R., UK
Lidl, Rudolf, Australia
Lieb, Ingo, BRD
von Lienen, Horst, BRD
Liess, Otto, Romania
Lih, Ko-Wei, China-Taiwan
Lima, Åsvald, Norway
Lin, James, USA
Lindberg, Per A., Sweden
Lindner, Curt, USA
Lindqvist, Carin E., Finland
Lindqvist, L. Peter, Finland
Lindroos, Markku T., Finland
Lindström, Berndt, Sweden
Lindström, Kauko V. K., Finland
Linna, Matti, Finland
Lions, Jacques Louis, France
Lipiński, Jan S., Poland
Lister, Timothy Charles, UK
Litherland, Richard Andrew, UK
Liulevicius, Arūnas, USA
Lluis, Emilio, Mexico
Loday, Jean-Louis, France
Lofquist, George W., USA
Löfwall, Clas, Sweden
Lohre, Eckhard, BRD
Lohwater, A. J., USA
Loi, Maurice, France
Lokki, Olli, Finland
Lolli, Gabriel, Italy
Londen, Stig-Olof, Finland
London, David, Israel
Long, Frederick W.. UK
Longstaff, William E., Australia
Lonka, Harri, Finland
Lønsted, Knud, Denmark

Looijenga, Eduard J. N., Netherlands
Lootgieter, Jean-Claude, France
Lopez de Medrano, Santiago, Mexico
Lopez Lagomasino, Guillermo, Cuba
Lorch, Lee, Canada
Lord, Graham, Canada
Lord, Harriet, Canada
Lorimer, Peter James, New Zeland
Lösch, Friedrich, BRD
Löttgen, Ulrich, BRD
Louhivaara, Ilppo Simo, Finland
Loukankine, Guennadi, USSR
Lounesto, Pertti, Finland
Lovász, László, Hungary
Love, Eric Russell, Australia
Lübcke, Bernd, BRD
Lubkin, Saul, USA
Lubotzky, Alexander, Israel
Ludwig, Garry, Canada
Luh, Wolfgang, BRD
Luhahi, Lahi, Zaire
Lukacs, Eugene, USA
Lukankin, G. L., USSR
Lukes, Dahlard L., USA
Lukić, Mirko, Yugoslavia
Lumiste, I. G., USSR
Luna, Domingo, France
Luoma, Martti J., Finland
Lupanov, O. B., USSR
Lutterodt, Clement Henry, Ghana
Lutz, Elisabeth, France
Luukkainen, Jouni, Finland
Lyons, Terence J., UK

Määttä, Eljas T., Finland
Macbeath, Alexander Murray, UK
MacDonald, John L., Canada
MacDonnell, John J., USA
MacDonnell, Joseph F., USA
Machado, Armando, Portugal
Macintyre, Angus J., USA
Macphail, Moray S., Canada
MacPherson, Robert D., USA
Madsen, Anders, Denmark
Madsen, Ib Henning, Denmark
Maebashi, Toshiyuki, Japan
Maeda, Fumi-Yuki, Japan
Magarian, Elizabeth Ann, USA
Magenes, Enrico, Italy
Magid, Andy R., USA
Magnus, Arne, USA
Magyarosi, Margareta, Romania
Maher, Mary Anne, USA
Mahony, Louis M., USA
Mailhos, Line, France
Maison, Kofi, Ghana
Majcher, Genowefa, Poland
Makanin, G. S., USSR
Makary, Farouk, Yemen
Mäkelä, Matti J., Finland
Mäkeläinen, Timo O. T., Finland
Mäkinen, Jukka, Finland
Makino, Masahiro, Japan
Mąkowski, Andrzej, Poland
Malaty, George
Malgrange, Bernard, France
Malik, M. A., Canada
Mallet-Paret, John, USA
Malliavin, Marie-Paule, France
Malliavin, Paul, France
Malm, Kaj, Finland
Maltsiniotis, Georges, France
Mamourian, Ahmad, Iran
Mandelbaum, Richard, USA
Manin, J. I., USSR
Manninen, Jouko, Finland
Manolache, Nicolae, Romania
Manolov, Spas, Bulgaria
Manwell, Alfred Raymond, Rhodesia
Marasigan, José, Philippines
Marcellan Español, Francisco, Spain
Marchionna, Ermanno, Italy
Marchionna-Tibiletti, Cesarina, Italy
Marcinkowska, Hanna, Poland
Marcum, Howard J., USA
Mardešić Sibe, Yugoslavia
Marechal, Odile, France
Marino, Mario, Italy
Márki, László, Hungary
Marković, Zoran, Yugoslavia
Markušević, A. I., USSR
Marle, Charles-Michel, France
Marmaridis, Nikolaos, Greece
Marot, Jean, France
Marquetty, Antoine Eugene, France
Marry, Pierre, France
Martens, Henrik H., Norway
Marti, Jürg T., Switzerland
Martikainen, Olli E., Finland
Martin, Donald A., USA
Martin, George E., USA
Martin, Robert P., USA
Martinez-Villa, Roberto, Mexico
Martio, Olli, Finland
Masani, Pesi R., USA
Mashour, Ali Salem, Egypt
Mason, Geoffrey, USA
Mason, John Charles, UK
Massé, Jean-Claude, Canada
Masser, David William, UK
Massy, Richard, France
Matala-aho, Tapani, Finland
Mathias, A. R. D., UK
Mathieu, Philippe, Belgium
Matlis, Eben, USA
Matsuda, Michihiko, Japan
Matsuda, Satoshi, Japan
Matsumoto, Kōzi, Japan
Matsumoto, Yukio, Japan
Matsuo, Yoshitomo, Japan
Matsuzaki, Nagi, Japan
Matthes, Klaus, DDR
Mattila, Jussi Antero, Finland
Mattila, Kirsti, Finland
Mattila, Pertti, Finland
Matumoto, Takao, Japan
Maude, Ronald, UK
Mäulen, Jürgen, BRD
Maumary, Serge, Switzerland
Maurel-Chaleyat, Mireille, France
Mäurer, Helmut, BRD
Mayer-Lindenberg, Fritz, BRD
Mazen, Henrietta, USA
McCleary, John, USA
McCoy, Peter A., USA
McCrea, Michael M., USA
McCrory, Clint, USA

McCulloh, Leon R., USA
McDonough, Thomas, UK
McDowell, Kenneth, Canada
McEvoy, Brian, Ireland
McFaden, Harold H., USA
McGehee, Richard P., USA
McKay, John, Canada
McKean, Henry P., USA
McKilligan, Sheila, UK
McMinn, Trevor J., USA
Mebkhout, Zoghman, France
Megyesi, László, Hungary
Mehri, Bahman, Iran
Mehrtens, Herbert, BRD
Mejzler, David, Israel
Meldrum, J. D. P., UK
Melin, Anders, Sweden
Melrose, Richard Burt, USA
Melter, Robert, USA
Meltzow, Otto, BRD
Memauri, Haghdad, Iran
Mendoza, Leonardo, Venezuela
Meneghello, Luiz Virgilio, Brazil
Mengersen, Ingrid, BRD
Menikoff, Arthur, USA
Menke, Klaus, BRD
Mennicken, Reinhard, BRD
Merikoski, Jorma K., Finland
Meril, Alex, France
Merindol, Jean-Yves, France
Merten, Knut, BRD
Messing, William, USA
Metsänkylä, Tauno, Finland
Metzger, Pierre David, France
Metzler, Wolfgang, BRD
Meurman, Arne, USA
Meyer, Gottfried P., BRD
Meyer, Yves, France
Meyn, Klaus-Hermann, BRD
Michael, Ernest, USA
Micheletti, Anna Maria, Italy
Michiwaki, Yoshimasa, Japan
Mickelsson, Jouko A., Finland
Midura, Stanisław, Poland
Miettinen, Seppo K., Finland
Mikkola, Erkki T., Finland
Mikolás, Miklós, Hungary
Mikulska, Małgorzata, Poland
Miles, E. P., Jr. USA
Miles, Joseph B., USA
Milić, Svetozar, Yugoslavia
Miličić, Dragan, Yugoslavia
Miller, D. D., USA
Miller, John Boris, Australia
Miller, John J. H., Ireland
Miller, Richard Keith, USA
Miller, Sanford S., USA
Millett, Kenneth, USA
Milnes, Paul, Canada
Miloslavsky, George C., USA
Mimura, Yukio, Japan
Minc, Piotr, Poland
Minea, Gheorghe, Romania
Miniowitz, Ruth, Israel
Mirković, Branislav, Yugoslavia
Mishra, Sudhakara, Canada
Mišík, Ladislav, Czechoslovakia
Misiurewicz, Michał, Poland
Mislin, Guido, Switzerland
Mitchell, Josephine, USA
Mitchell, W. J. Rae, UK
Mitsui, Takayoshi, Japan
Mittelmeyer, Günter, BRD
Mittenthal, Lothrop, USA
Miwa, Megumu, Japan
Miyaoka, Yoichi, Japan
Miyatake, Osamu, Japan
Mizohata, Shigeru, Japan
Mizumura, Hideo, Japan
Moghaddam, Mohammad R. R., Iran
Mohammed, Isa Baba, Nigeria
Mohammed, Salah-Eldin Ahmed, Sudan
Mok, Kam-Ping, Hong Kong
Mokobodzki, Gabriel, France
Molęda, Alicja, Poland
Möller, Hans Michael, BRD
Möller, Manfred, BRD
Molzon, Robert, USA
Monegato, Giovanni, Italy
Montgomery, Deane, USA
Mooney, John William, UK
Moore, John Douglas, USA
Moori, Jamshid, Iran
Morales, Rodolfo, Mexico
Moreno, Carlos J., USA
Moreno, Jose Miguel, Spain
Morgan, Kathryn, USA
Morimoto, Mitsuo, Japan
Morton, Hugh Reynolds, UK
Moser, Jürgen, USA
Moser, Louise E., USA
Mosevich, Jack, Canada
Mossino, Jacqueline, France
Mostefai, Abdelhafid, Algeria
Mostow, George, USA
Moszyńska, Maria, Poland
Motchane, Leon, France
Moulin Ollagnier, Jean, France
Moussaoui, M., Algeria
Mrdjenović, Branko, Yugoslavia
Mrmak, Mirjana, Yugoslavia
Mršević, Mila, Yugoslavia
Mubenga, Kampotu, Zaire
Mugambi, Paulo Edward, Uganda
Muldoon, Martin E., Canada
Mulla, Fuad Shakir, Kuwait
Mullen, Gary L., USA
Müller, Helmut, BRD
Müller, Karl-Heinz, Italy
Müller-Römer, Peter, BRD
Munn, W. Douglas, UK
Mura, Roberta, Canada
Murase, Ichiro, Japan
Murasugi, Kunio, Canada
Murphy, Gerard John, Ireland
Murphy, Thomas Noel, Canada
Murre, Jacob P., Netherlands
Musiela, Marek, Poland
Mustonen, Vesa, Finland
Muto, Tsukasa, Japan
Muto, Yoshio, Japan
Mwene, Benjamin, Zambia
Mycielski, Jan, USA
Myjak, Józef, Poland
Myrberg, Lauri, Finland
Myyryläinen, Jukka K., Finland

Näätänen, Marjatta, Finland
Nagamachi, Shigeaki, Japan
Nagata, Masayoshi, Japan
Nagy, Josef, Czechoslovakia
Nakamura, Koshiro, Japan
Nakanishi, Shizu, Japan
Nakano, Kazumi, USA
Näkki, Raimo T., Finland
Nakoulima, Ousseynou, Senegal

Namikawa, Yukihiko, Japan
Narayanaswami, Pallasena P., Canada
Nart, Enric, Spain
Narushima, Hiroshi, Japan
Nashed, M. Z., USA
Nastasescu, Constantin, Romania
Navarro Aznar, Vicente, Spain
Naylor, Vernon, UK
Ndumu, Martin Ngu, Cameroon
Neftidji, Phaedra, Greece
Negishi, Aiko, Japan
Neher, Erhard, BRD
Neittaanmäki, Pekka J., Finland
Nel, Louis Daniel, Canada
Nelimarkka, Esa E., Finland
Néron, André, France
Ness, Linda, USA
Netanyahu, Elisha, Israel
Neuberger, John W., USA
Neubrunn, Tibor, Czechoslovakia
Neumann, Berhard Hermann, Australia
Neuvonen, Timo, Finland
Nevanlinna, Olavi, Finland
Nevanlinna, Rolf, Finland
Nevanlinna, Veikko, Finland
Neveu, Jacques, France
Newman, Michael F., Australia
Newman, Morris, USA
Nghiem Xuan Hai, Vietnam
Nguyen Dinh Tri, Vietnam
Nguyen Huu Anh, Vietnam
Nguyen Van Thu, Vietnam
Nianias, George, Greece
Nicolas, Anne-Marie, France
Nicolas, Jean-Louis, France
Niculescu, Gheorghe, Romania
Nielsen, Aage, Denmark
Nielsen, H. Andreas, Denmark
Nielsen, Niels Jørgen, Denmark
Nielsen, Ole A., Canada
Niemi, Antti J., Finland
Niemi, Eero, Finland
Niemi, Hannu O., Finland
Niemi Seppo J., Finland
Nieminen, Toivo, Finland
Niiniluoto, Ilkka, Finland
Nijenhuis, Albert, USA
Nikišin, E. M., USSR
Nikolić-Despotović, Danica, Yugoslavia
Nikolskii, N. K., USSR
Nillsen, Rodney V., Australia
Nilsson, Lennart, Sweden
Nilsson, Nils, Sweden
Nishikawa, Seiki, Japan
Nishimiya, Han, Japan
Nishimura, Hiroshi, Japan
Nishimura, Toshio, Japan
Nishio, Kazuhiro, Japan
Niskanen, Vesa A., Finland
Niss, Mogens, Denmark
Nitá, Constantin, Romania
Nitsche, Joachim A., BRD
Niven, Barbara Susanna, Australia
Njåstad, Olav, Norway
Noakes, John Lyle, Australia
Nodari Sutti, Carla, Italy
Nohel, John A., USA
Nolte, Wolfgang, BRD
Nordlander, Arne, Sweden
Nordon, Jean, France
Norlamo, Pekka, Finland
Norros, Ilkka, Finland
Norstad, Folke, Sweden
Nouri-Moghadam, Mohammad R., Iran
Novák, Břetislav, Czechoslovakia
Nováková, Eva, Czechoslovakia
Noverraz, Philippe, France
Novikov, S. P., USSR
Nowak, Sławomir, Poland
Nozaki, Yasuo, Japan
Nualart, David, Spain
Numakura, Katsumi, Japan
Nussbaum, A. Edward, USA
Nyberg, Kaisa, Finland

Obayashi, Noboru, Japan
Oberdoerffer, Jacques, France
O'Connor, M. Lesley, USA
Odlyzko, Andrew, USA
Odoni, Robert W. K., UK
Oeljeklaus, Eberhard, BRD
Oesterlé, Joseph, France
Ogawa, Roy H., USA
Ogawara, Masami, Japan
Ohkuma, Tadashi, Japan
Ohno, Katsuhisa, Japan
Ohtsuka, Makoto, Japan
Ohwaki, Shin-ichi, Japan
Oikkonen, Juha M. R., Finland
Oja, Kirsti, Finland
Ojala, Ossi, Finland
Ojanguren, Manuel, Switzerland
Okabayashi, Shigeyoshi, Japan
Okazaki, Masako, Japan
Okiljević, Blažo, Yugoslavia
Okoh, Frank, Nigeria
Okubo, Tanjiro, Canada
Olaofe, Gabriel Oluremi, Nigeria
Olaofe, Kathleen, Nigeria
Olech, Czesław, Poland
Olesen, Dorte, Denmark
Oliker, Vladimir, USA
Oliver, Robert, USA
Olson, Loren David, Norway
Olsson, Jørn Børling, BRD
Olszowicz, Anna, Poland
Olubummo, Adegoke, Nigeria
Olver, Peter John, USA
O'Mahony Tamaschke, Rosalie M., USA
O'Neil, Richard, USA
Oppenheim, Joseph, USA
Orhon, Mehmet, Turkey
Orlik, Peter, USA
Orlov, Konstantin, Yugoslavia
Osborn, J. Marshall, USA
Osgood, Brad G., USA
Oshima, Toshio, Japan
Osner, Henry J., USA
Osofsky, Barbara L., USA
Osório, Vasco, BRD
Osserman, Robert, USA
Ostaszewski, Adam Jerzy, UK
Ostianu, N. M., USSR
Ostoja-Łojasiewicz, Stanisław, Poland
Ostogorski, Tatjana, Yugoslavia
Otsuka, Kayo, Japan

Otsuki, Tominosuke, Japan
Otte, Michael, BRD
Ozawa, Koichiro, Japan

Paakkanen, Markku T., Finland
Paatero, Veikko, Finland
de Paepe, Peter J., Netherlands
Pagani, Carlo Domenico, Italy
de Pagter, Bernardus, Netherlands
Päivärinta, Lassi, Finland
Palis, Jacob, Jr., Brazil
Palka, Bruce P., USA
Pallaschke, Diethard, BRD
Palleschi, Marino, Italy
Palmore, Julian, USA
Paluszny, Marco, Venezuela
Pan, Ting-Kwan, USA
Pankiewicz, Grazyna, Poland
Pankiewicz, Stanislawa, Poland
Panzone, Rafael, Argentina
Papacostas, Georges, Greece
Papakonstantinou, Vassilios, Greece
Papini, Pier Luigi, Italy
Pardon, William, USA
Parimala, Raman, India
Parker, Richard Alan, UK
Parks, James M., USA
Pars, Leopold Alexander, UK
Paschke, William L., USA
Patetta, Nicolas, Argentina
Patkowska, Hanna, Poland
Patronis, Anastasios, Greece
Patterson, Edward McWilliam, UK
Paunić, Djura, Yugoslavia
Pavel, Monique, France
Pavel, Nicolae, Romania
Paveri-Fontana, Stefano L., Italy
Pears, Alan Robson, UK
Pedersen, Erik Kjaer, Denmark
Peetre, Jaak, Sweden
Pellaumail, Jean, France
Peltola, Veli-Pirkka, Finland
Penrose, Roger, UK
Pentikäinen, Teivo, Finland
Penttinen, Markku J., Finland
Peremans, Wouter, Netherlands
Perko, Aarni, Finland
Perović, Miodrag, Yugoslavia
Perraud, Jean, France
Persson, Jan, Norway
Persson, Lars Erik, Sweden
Persson, Ulf A., USA
Peschl, Ernst, BRD
Pesonen, Martti, Finland
Petermann, Eike, Sweden
Peters, Klaus, BRD
Peters, Meinhard, BRD
Petersen, Karl, USA
Peterson, Franklin P., USA
Petiau, Gérard, France
Petrescu, Ioana, Romania
Petryshyn, W. V., USA
Pettersson, Rolf, Sweden
Pézennec, Jean, France
Pfluger, Albert J., Switzerland
Pflugfelder, Hala, USA
Pham, Mau Quan, France
Piatetski-Shapiro, Ilya, Israel
Picard, Colette, France
Piccard, Sophie, Switzerland
Pichaud, Joëlle, France
Pieczyski, Bogusław, Poland
Piene, Ragni, Norway
Pier, Jean-Paul, Luxembourg
Pierce, Richard S., USA
Pierce, Stephen, Canada
Piersma, H. J. D., Netherlands
Pietrocola, Norma, Argentina
Pietsch, Albrecht, DDR
Pihko, Jukka A., Finland
Pijls, Henk, Netherlands
Pilika, Petraq, Albania
Pimsner, Mihai, Romania
Pinchon, Didier, France
Pinkham, Henry C., USA
Piranian, George, USA
Pitcher, Everett, USA
Pitkänen, Paavo, Finland
Pitkäranta, M. Juhani, Finland
Pitt, David Hal, UK
Pittie, Harsh V., USA
Platonov, V. P., USSR
Platt, Czesław, Poland
Pleasants, Peter A. B., UK
Pleijel, Åke V. C., Sweden
Pleśniak, Wiesław, Poland
Płonka, Ernest E., Poland
Plucińska, Agnieszka, Poland
Plymen, Roger John, UK
Pogorelov, Alexei V., USSR
Polexe, Raluca, Romania
Polombo, Albert, France
Pondělíček, Bedřich, Czechoslovakia
Pontrjagin, L. S., USSR
Pop, Florian, Romania
Pop, Horia Calin, Romania
Popa, Eugen, Romania
Popa, Irina, Romania
Popa, Nicolae, Romania
Popa, Sorin, Romania
Popov, Koviljka, Yugoslavia
Porru, Giovanni, Italy
Porte, Jean, France
Porteous, Hugh L., UK
Porubský, Štefan, Czechoslovakia
Post, Karel A., Netherlands
Poston, Timothy, Switzerland
Potra, Florian, Romania
Poulsen, Ebbe Thue, Denmark
Pourcin, Geneviève, France
Povilaitis, Sigitas, Hungary
Poznjak, E. G., USSR
Praciano-Pereira, Tarcisio, Sweden
Pratt, Stephen Philip, UK
Pressman, Irwin, Canada
Preston, Christopher, BRD
Price, Kenneth H., USA
Priddy, Stewart B., USA
Pride, Stephen James, UK
Primc, Mirko, Yugoslavia
Procesi, Claudio, Italy
Protasi, Marco, Italy
Proulx, Ronald J., USA
Proulx, Viera Krňanová, USA
Prucnal, Tadeusz, Poland
Prudnikov, A. P., USSR
Prüss, Jan, BRD
Przymusiński, Teodor C., Poland
Przytycki, Feliks Marian, Poland
Pucci, Carlo, Italy
Pucci, Patrizia, Italy
Puczyłowski, Edmund, Poland

Puech-Raoult, Annie, France
Puel, Jean-Pierre, France
Puerta Sales, Fernando, Spain
Puranen, Juha, Finland
Purmonen, Veikko T., Finland
Pursiheimo, Ulla, Finland
van der Put, Marius, Netherlands
Putcha, Mohan, USA
Puts, Pierre Jacques, Netherlands
Puttaswamy, Tumkur, USA
Puyau, Hermes Augosto, Argentina
Puzynin, Igor, USSR

Quarteroni, Alfio, Italy
Quillen, Daniel, USA
Quinn, Frank, USA

Rabinowitz, Paul H., USA
Radjabalipour, Mehdi, Iran
Radjavi, Heydar, Canada
Radosavljević-Nikolić, Mirjana, Yugoslavia
Radoux, Christian, Belgium
Radu, Nicolae, Romania
Rahman, Qazi Ibadur, Canada
Raimi, Ralph A., USA
Rajala, Jussi A. I., Finland
Rajaona, Pierre, Madagascar
Ramanan, Sundara, India
Ramanathan, Annamalai, India
Rankin, Robert A., UK
Rankin, Stuart A., Canada
Ranne, Paavo Antero, Finland
Rao, D. Ramakrishna, Iran
Rao, Veldanda Venugopal, Canada
Rašajski, Borivoj, Yugoslavia
Rassias, George, Greece
Rassias, Themistocles, Greece
Ratajski, Henryk, Poland
Ratner, Marina, USA
Rauszer, Cecylia, Poland
Rautenberg, Wolfgang, BRD
Ravaska, Timo, Finland
Ravenel, Douglas C., USA
Raviart, Pierre-Arnaud, France
Raynaud, Michel, France
Razen, Reinhard, Austria
Razzaghi, Mohsen, Iran
Rebhuhn, Deborah, USA
Recamán Santos, Bernardo, Colombia
Recillas-Juarez, Félix, Mexico
Rees, David, UK
Rees, Joan Sybil, UK
Rees, Susan Mary, UK
Rehm, Hans P., BRD
Reich, Axel, BRD
Reich, Edgar, USA
Reichert, Marianne, BRD
Reid, Miles, UK
Reilly, Ivan L., New Zealand
Reimann, Hans Martin, Switzerland
Reimann, József, Hungary
Reiten, Idun, Norway
Rejto, Peter A., USA
Rempała, Ryszarda, Poland
von Renteln, Michael, BRD
Reutter, Fritz, BRD
Révész, Pál, Hungary
Revuz, André, France
Rhin, Georges, France
Ribes, Luis, Canada
Richard, Ubaldo, Italy
Richmond, Bruce, Canada
Rickard, John A., Australia
Rickman, Seppo, Finland
Rieffel, Marc A., USA
Riemenschneider, Oswald, BRD
Riera Madurell, Teresa, Spain
Riesel, Hans, Sweden
Rigby, John F., UK
Riihentaus, Juhani, Finland
Risito, Corrado, Italy
Risnes, Martin, Norway
Risto, Ilić, Yugoslavia
Ritter, Gunter, BRD
Rivet, Roger, France
Rizza, Giovanni Battista, Italy
Robert, Didier, France
Roberts, Paul, USA
Robertson, Alexander P., Australia
Robin, Guy, France
Robinson, Christopher Alan, UK
Robinson, Clark, USA
Rocha, Mariaugusta, Brazil
Röck, Hans, BRD
Rogalski, Marc, France
Roger, Claude, France
Rogers, Claude Ambrose, UK
Roiter, A. V., USSR
Roman, Tiberiu, Romania
Ronveaux, André, Belgium
Rooney, Paul George, Canada
Roos, Guy, France
Roos, Jan-Erik, Sweden
Rosa, Alexander, Canada
Rosati, Mario, Italy
Roseblade, James Edward, UK
Rosen, Jay, USA
Rosen, Kenneth H., USA
Rosen, William G., USA
Rosenberg, Ivo G., Canada
Rosenberger, Gerhard, BRD
Rosenblatt-Roth, Millu, USA
Rosenknop, Jon, BRD
Rosinger, Elemer E., Israel
Rossowska, Michalina, Poland
van Rossum, Herman, Netherlands
Rostén, Sigrid S., Finland
Rota, Gian-Carlo, USA
Roth, Klaus Friedrich, UK
Roth, Walter, BRD
Rothberger, Fritz, Canada
Rotkiewicz, Andrzej, Poland
Rottenberg, Reuven Robert, Israel
Roussas, George, Greece
Rousseau, Guy, France
Rousseau-Egele, Jacques, France
Rowley, Christopher A., UK
Royer, Gilles, France
Rozenberg, Grzegorz, Belgium
Rózsa, Pál, Hungary
Rubel, Lee, USA
Rubinstein, Zalman, Israel
Ruderman, Harry D., USA
Rudin, Mary Ellen, USA
Rudin, Walter, USA
Ruggeri, Tommaso, Italy
Ruh, Ernst A., BRD
Ruiz-Claeyssen, Julio, Peru
Rumak, Tadeusz, Poland
Rung, Donald C., USA

Ruohonen, Keijo, Finland
Ruohonen, Matti J., Finland
Rüpprich, Bernward, BRD
Ruskeepää, Heikki, Finland
Russell, Dennis C., Canada
Ruszczyk, Henryk, Poland
Rutter, John W., UK
Ryan, Frank, USA
Ryavec, Charles, USA
Ryff, John, USA

Saarela, P. Juhani, Finland
Saari, Donald, USA
Saarimäki, Mikko M., Finland
Sabbah, Abdalla Saied, Egypt
Sacks, Jonathan, USA
Sadosky, Cora, Venezuela
Sadr-Ghadar-Ghadr, Jamaledin, Iran
Saff, E. B., USA
Sagraloff, Boris, BRD
Saintloup, Bernard, France
Sakagami, Eisho, Japan
Sakai, Fumio, Japan
Sakai, Makoto, Japan
Sakai, Shôichirô, USA
Salagean, Grigore, Romania
Salamon, Simon Montague, UK
Salberger, Per, Sweden
Salenius, Tauno, Finland
Sallinen, Pekka, Finland
Salminen, Erland, Finland
Salo, Seppo K., Finland
Salomaa, Arto, Finland
Salomonsen, Hans Anton, Denmark
Salovaara, Sampo, Finland
Salvadori, Anna, Italy
Salzer, Herbert E., USA
Samarskii, A. A., USSR
Sanchez Fernandez, Carlos, Cuba
Sancho San Román, Juan, Spain
Sands, Arthur David, UK
Sankaran, Subramanian, UK
Sano, Shigeru, Japan
Sansuc, Jean-Jacques, France
Santagati, Giuseppe, Italy
Sanz Serna, Jesus Maria, Spain
Sapagovas, Mifodij P., USSR
Saphar, Pierre, Israel
Šapkarev, Ilija, Yugoslavia
Saranen, Jukka A., Finland
Saranpää, Eino K., Finland
Sard, Arthur, USA
Särkijärvi, Jyri J., Finland
Sarvas, Jukka, Finland
Sasaki, Usa, Japan
Sastri, Chelluri C. A., Canada
Satake, Ichiro, USA
Sato, Isuke, Japan
Saut, Jean-Claude, France
Savolainen, Onerva, Finland
Sawaki, Sumio, Japan
Sawashima, Ikuko, Japan
Saxena, Subhash C., USA
Saxl, Jan, UK
Sburlan, Silviu, Romania
Scarpalezos, Dimitris, France
Schaal, Werner G. H., BRD
Schafer, Alice T., USA
Schafer, Richard D., USA
Schapira, Pierre, France
Schaps, Mary, Israel
Scharlemann, Martin, USA
Schatzman, Michelle, France
Scheerer, Hans, BRD
Schiek, Helmut, BRD
Schinzel, Andrzej, Poland
Schipp, Ferenc, Hungary
Schirmeier, Horst, BRD
Schirmeier, Ursula, BRD
Schirmer, Helga, Canada
Schlichting, Günter, BRD
Schlöglmann, Wolfgang, Austria
Schlomiuk, Norbert, Canada
Schlosser-Haupt, Silke, BRD
Schmeelk, John, USA
Schmeisser, Gerhard C., BRD
Schmid, Wilfried, USA
Schmidt, Asmus L., Denmark
Schmidt, Klaus, UK
Schmidt, Rita M., BRD
Schneider, Hans, USA
Schneider, Manfred, BRD
Schneider, Michael, BRD
Schnitzpan, Daniel, France
Schober, Glenn, USA
Schochet, Claude, USA
Schoemaker, George, Netherlands
Schoenfeld, Lowell, USA
Scholz, Werner, BRD
Schori, Richard, USA
Schott, René, France
Schottenloher, Martin, BRD
Schreiber, Bertram M., USA
Schreiber, Michel, France
Schroeder, Ludwig, BRD
Schroll, Joachim, BRD
Schuder, Werner, BRD
Schumacher, Barbara, BRD
Schuster, Hans Werner, BRD
Schwänzl, Roland, BRD
Schwartz, Jean-Marie, France
Schwarz, Binyamin, Israel
Schwarz, Friedhelm, BRD
Schwarz, Gerald W., USA
Schwarz, Štefan, Czechoslovakia
Schweitzer, Paul A., Brazil
Scourfield, Eira J., UK
Scozzafava, Romano, Italy
Šedivý, Jaroslav, Czechoslovakia
Segal, Jack, USA
Seibert, Peter, Venezuela
Seibt, Peter, BRD
Seifert, George, USA
Seikkala, Seppo A., Finland
Sekizawa, Masami, Japan
Selenius, Clas-Olof, Sweden
Selfridge, John L., USA
Sell, George R., USA
Semadeni, Zbigniew, Poland
Sen, D. K., Canada
Senechal-Rozoy, Brigitte, France
Seppälä, Mika, Finland
Seppälä, Veikko, Finland
Serapioni, Raul, Italy
Seshadri, C. S., India
Seth, Bhoj Raj, India
Shaffer, Dorothy Browne, USA
Shahshahani, Siavash, Iran
Shaki, Ahiezer, Israel
Shaneson, Julius L., USA
Shantaram, R., USA
Shapiro, Daniel B., USA
Shastri, Anant R., India

Shea, Daniel, USA
Shekoury, Raymond N., Iraq
Shelstad, Diana Frost, USA
Shen, Chung Yi, Canada
Shephard, Geoffrey C., UK
Shibata, Katsuyuki, Switzerland
Shih, Kung-sing, China-Taiwan
Shima, Kazuhisa, Japan
Shimura, Goro, Japan
Shintani, Takuro, Japan
Shioda, Tetsuji, Japan
Shiohama, Katsuhiro, Japan
Shiu, Peter Man-Kit, UK
Shiue, J.-S., China-Taiwan
Shizuma, Ryoji, Japan
Shklov, N., Canada
Shultz, Frederic, USA
Sibagaki, Wasao, Japan
Sibley, David, USA
Sibuya, Yasutaka, USA
Siciak, Jozef, Poland
Siddiqi, Jamil Ahmad, Canada
Siddiqi, Rafat Nabi, Kuwait
Siebenmann, Laurent, France
Siebert, Gert, BRD
Siersma, Dirk, Netherlands
Silberstein, J. P. O., Australia
Silva, Alessandro, Italy
Silvennoinen, Risto P., Finland
Silvia, Evelyn, USA
Simader, Christian G., BRD
Simionescu, Corneliu, Romania
Simmons, Harold, UK
Simon, Udo, BRD
Sims, Charles C., USA
Singer, Ivan, Romania
Singh, Balbir, USA
Singh, Shanker, India
Singmaster, David, UK
Sinnott, Warren M., USA
Sion, Maurice, Canada
Širjaev, A. N., USSR
Šišolák, František, Czechoslovakia
Siu, Man-Keung, Hong Kong
Siu, Yum-Tong, USA
Sjöberg, Boris, Finland
Sjödin, Tord, Sweden
Sjöstrand, Johannes, Sweden
Skalba, Justine, USA
Skau, Christian F., Norway
Skjelbred, Tor, Norway
Skoda, Henri, France
Skóvnik, Krystyna Anna, Poland
Skowron, Andrzej, Poland
Skowroński, Andrzej, Poland
Skucha, Jerzy A., Poland
Skupien, Zdzisław, Poland
Skwarczynski, Maciej, Poland
Słezion, Dymitr, Poland
Sloan, Alan, USA
Słodkowski, Zbigniew, Poland
Słowikowski, Wojciech, Denmark
Smale, Stephen, USA
Smith, Anthony G., UK
Smith, Edwin M. R., USA
Smith, Frank D., Jr. USA
Smith, Richard F., USA
Smythe, Robert T., USA
Snaith, Victor P., Canada
Snow, Donald R., USA
Soare, Robert, USA
Soares, Rui, Portugal
Sobel, Milton, USA
Sofonea, Liviu, Romania
Sointu, Marjatta H. E., Finland
Solian, Alexandru, Romania
Solomon, Herbert, USA
Sontag, Alexia, USA
Sorjonen, Pekka, Finland
Sorvali, Tuomas, Finland
Sós, Vera T., Hungary
Soto Andrade, Jorge, Chile
Soulé, Christophe, France
Spagnolo, Sergio, Italy
Spathis, Nicolas, France
Spathis, Spyridon, France
Spathis, Vassiliki, France
Spector, René, France
Speer, Eugene R., USA
Sperner, Emanuel, BRD
Spiegel, Wolfgang, BRD
Spinelli, Giancarlo, Italy
Srinivasan, Bhama, USA
Stacey, Peter J., Australia
Stackelberg, Olaf P., USA
Staffans, Olof, Finland
Staffeldt, Ross E., USA
Stagnaro, Ezio, Italy
Stakgold, Ivar, USA
Stallmann, Friedemann W., USA
Stangler, Sieghart Stefan, BRD
Stanković, Bogoljub, Yugoslavia
Stanojević, Časlav V., USA
Starbird, Michael P., USA
Stavrinos, Panayotis, Greece
Stefánsson, Jón Ragnar, Iceland
Steffens, Karsten, BRD
Stegbuchner, Hans, Austria
Stegeman, Jan D., Netherlands
Steinberg, Robert, USA
Steinby, Magnus, Finland
Steiner, Hans-Georg, BRD
Steiner, Richard J., UK
Sten, Björn T. I., Sweden
Stenij, S. E., Finland
Stensholt, Eivind, Norway
Stenström, Bo, Sweden
Stetkaer, Henrik, Denmark
Stetter, Hans J., Austria
Steurich, Manfred, BRD
Stipanić, Ernest, Yugoslavia
Stocke, Britt-Marie, Sweden
von Stockert, Etzel Ritter, Brazil
Stoica, Lucretiu, Romania
Stojaković, Zoran, Yugoslavia
Stojanović, Miroslava, Yugoslavia
Stoltzfus, Neal W., USA
Stone, Marshall H., USA
Størmes, Erling, Norway
Storvick, David Arne, USA
Strampp, Walter, BRD
Strankmann, Manfred, BRD
Stratila, Serban, Romania
Straume, Eldar, Norway
Straus, Ernst G., USA
Strauss, Dona, UK
Strebel, Kurt, Switzerland
Strecker, George, USA
Streit, Roy, USA
Stroeker, Roelof J., Netherlands
Strömbeck, C. Peter M., Sweden
Strömberg, Jan-Olov, Sweden

Strommer, Julius, Hungary
Strooker, Jan R., Netherlands
Strubecker, Karl, BRD
Strübig, Gebhard, BRD
Strübing, Wolf, BRD
Stümke, Bernd Ulrich, BRD
Sturm, Jacob, USA
Styan, George P. H., Canada
Suetsugu, Kazunori, Japan
Suguri, Tsuneo, Japan
Sun, Hugo Sui-Hwan, USA
Sund, Terje, Norway
Sung, Chen-Han, China-Taiwan
Sunley, Judith S., USA
Sunouchi, Haruo, Japan
Sunyach, Christian, France
Suomela, Pentti, Finland
Suominen, Kalevi, Finland
Surányi, János, Hungary
Suslin, A. A., USSR
Sussmann, Hector Jose, USA
Sutherland, Colin Eric, USA
Suyama, Yukio, Japan
Suzuki, Haruo, Japan
Suzuki, Hitoshi, Japan
Suzuki, Osamu, Japan
Suzuki, Toshio, Japan
Svešnikov, A. G., USSR
Svetlichny, George, Brazil
Swanson, David, USA
Swanson, Richard C., USA
Swiatak, Halina, Canada
Sychowicz, Andrzej, Poland
von Sydow, Björn, Sweden
Sysikaski, Pirjo, Finland
Syvähuoko, Olli, Finland
Szabados, József, Hungary
Szafraniec, Franciszek Hugon, Poland
Szarek, Stanisław J., Poland
Szczepanski, Susan, USA
Szekeres, George, Australia
Szendrei, Ágnes, Hungary
Szendrei, János, Hungary
Szeto, George, USA
Szlanka, Imre, Hungary
Szmydt, Zofia, Poland
Szőkefalvi-Nagy, Béla, Hungary
Szpiro, Lucien, France
Taari, Ossi J., Finland
Taft, Earl J., USA
Taiani, Geraldine, USA
Taira, Kazuaki, Japan
Takacs, Roland E., Austria
Takahashi, Masayuki, Japan
Takahashi, Motoo, Japan
Takahashi, Reiji, France
Takahashi, Shin-ei, Japan
Takahashi, Shuichi, Canada
Takahashi, Yukichi, Japan
Takasaki, Teiji, Japan
Takayanagi, Yukisada, Japan
Takeno, Hyoichiro, Japan
Takesaki, Masamichi, USA
Takeshita, Akira, Japan
Takeuchi, Kensuke, Japan
Takeuchi, Tsuneo, Japan
Takeuti, Gaisi, USA
Talenti, Giorgio, Italy
Tamanini, Italo, Italy
Tamm, Martin, Sweden
Tammi, Olli, Finland
Tamminen, Markku I., Finland
Tamthai, Mark, Thailand
Tan, Sin-Leng, Malaysia
Tanahashi, Katsumi, Japan
Tanaka, Hiroshi, Japan
Tang, C. Y., Canada
Tanimoto, Taffee T., USA
Tanno, Shukichi, Japan
Tanskanen, Asko, Finland
Tao, Yoko, Japan
Tapia-Recillas, Horacio, Mexico
Tappola, Heikki A. J., Finland
Taqqu, Murad S., USA
Targonski, György, BRD
Targonski, Jolán, BRD
Tartakoff, David S., USA
Tartar, Luc C., France
Tashiro, Yoshiaki, Japan
Taubert, Klaus, BRD
Taylor, Andrew W., UK
Taylor, Donald E., Australia
Taylor, Jean E., USA
Taylor, Laurence R., USA
Taylor, Samuel James, UK
Teicher, Mina, Israel
Teitler, D. R. M., Belgium
Tejumola, H. O., Nigeria
Temam, Roger, France
Tengstrand, Anders, Sweden
Tennison, Barry R., UK
Teodorascu, Alice, Romania
Terada, Fumiyuki, Japan
Te Riele, Hermanus J. J., Netherlands
Terrier, Jean-Marc, Canada
Textorius, Björn, Sweden
Thedy, Armin, BRD
Theis, Antoinette, France
Thevenaz, Jacques, Switzerland
Thiam, Magatte, Senegal
Thiam, Sakhir, Senegal
Thiemann, Christoph A., BRD
Thiry, Yves, France
Tholen, Walter, BRD
Thomas, Barbara Smith, USA
Thomas, G. Erik F., Netherlands
Thomas, P. Emery, USA
Thomason, Robert W.,USA
Thomée, Vidar, Sweden
Thomsen, Momme Johs, BRD
Thomson, Brian S., Canada
Thron, Wolfgang J., USA
Thurston, William, USA
Tienari, Juha M. J., Finland
Tienari, Martti J., Finland
Tietäväinen, Aimo, Finland
Tietz, Horst, BRD
Tignol, Jean-Pierre, Belgium
Tigoiu, Victor, Romania
Tijdeman, Robert, Netherlands
van Tilborg, Henk, Netherlands
Tillier, André, France
Timoney, Richard M., USA
Timourian, James G., Canada
Tirri, Henry R., Finland
Tits, Jacques, France
Tobin, Seán, Ireland
Toda, Hiroshi, Japan
Toet, Arie J., Netherlands
Toffin, Philippe, France
Toivonen, Pertti T., Finland
Toland, John F., UK
Tolsted, Elmer B., USA
Tomassini, Giuseppe, Italy

Tomida, Masamichi, Japan
Tomiyama, Jun, Japan
Tomppo, Erkki O., Finland
Tomter, Per, Norway
Tonge, Andrew M., UK
Tonietti, Tito, Italy
Toppila, Sakari, Finland
Topuzoglu, Alev, UK
Toralballa, Leopoldo V., USA
Torgerson, Frances L., USA
Törnig, Willi, BRD
Törnqvist, Leo W., Finland
Torp, Hans, Norway
Tóth, Ferenc László, Hungary
Touré, Saliou, Ivory Coast
Towber, Jacob, USA
Towers, David A., UK
Traczyk, Tadeusz, Poland
Traore, Sekou, Congo
Trauber, Philip, USA
Treger, Robert, Israel
Triebel, Hans, DDR
Trillas, Enric, Spain
Trimeche, Khalifa, Tunisia
Tromba, Anthony, USA
Tronel, Gérard, France
Trotman, D. J. A., UK
Trotter, Hale F., USA
Troutman, John, USA
Truffault, Bernard, France
Tsagas, Grigorios, Greece
Tsangaris, Panayiotis, Greece
Tsubota, Etuko, Japan
Tsuchikura, Tamotsu, Japan
Tsuchiya, Fumihide, Japan
Tsujishita, Toru, Japan
Tsutsumi, Masayoshi, Japan
Tubaro, Luciano, Italy
Tukia, Pekka, Finland
Turku, Halil, Yugoslavia
Turner, Nura Dorothea, USA
Turpin, Philippe, France
Turquette, Atwell, USA
Turtiainen, Yrjö, Finland
Tyc, Andrzej, Poland
Tylikowski, Andrzej, Poland
Tymchatyn, Edward D., Canada

Uchiyama, Moritsune, Japan
Uchiyama, Saburô, Japan
Ueno, Kenji, Japan
Uetake, Tsuneo, Japan
Ugrin-Šparac, Dimitrije, Yugoslavia
Uhlenbeck, Karen K., USA
Uhlenbrock, Dietrich A., BRD
Uhlig, Dietmar, DDR
Ujiie, Katsumi, Japan
Ukkonen, Esko, Finland
Ullman, Joseph L., USA
Ullrich, Christian, BRD
Ulmer, Friedrich, Switzerland
Ungar, Šime, Yugoslavia
Unger, Heinz, BRD
Upmeier, Harald, BRD
Urbanik, Kazimierz, Poland
Usui, Sampei, Japan

Vaaler, Jeffrey, USA
Väänänen, Jouko, Finland
Väänänen, Keijo O., Finland
Vahdati, Ali-Naghi, Iran
Vaillancourt, Rémi, Canada
Väisälä, Jussi, Finland
Väisänen, Mikko, Finland
Vaisman, Izu, Israel
Vala, Klaus, Finland
Valabrega Gibellato, Elda, Italy
Vallée, Robert, France
Valusescu, Ilie, Romania
Vamos, Peter, UK
Van Buskirk, J. M., USA
Van Casteren, Johannes, Belgium
Van Cutsem, Bernard, France
Van der Poorten, Alfred J., Australia
Van Dooren, René, Belgium
Van Geel, J. M. H., Belgium
Vanhamme, Willy, Belgium
Vanninathan, Muthusamy, India
Van Praag, Paul, Belgium
Vanstone, James Ray, Canada
Varadarajan, V. S., USA
Varadhan, Srinivasa, USA
Varga, Tamás, Hungary
Vasco, Carlos E., Colombia
Vasconcelos, Wolmer V., USA
Vasiu, Angela, Romania
Vaughan, Robert Charles, UK
Vélu, Jacques, France
Venema, Gerard A., USA
Venema, Stephanie, Netherlands
Venkataraman, Rangachari, Canada
Venkatesan, Pillapakkam S., Nigeria
Vepsäläinen, Seppo, Finland
Vera, Edgar, Peru
Verdier, Jean-Louis, France
Vernier Piro, Stella, Italy
Veron, Laurent, France
Vesterinen, Markku, Finland
Vicente, Jose Luis, Spain
Vietsch, Willem K., Netherlands
Villamayor, Orlando E., Argentina
Villarquide Lourido, Juan A., Spain
Vilonen, Kari K., Finland
Vinti, Calogero, Italy
Viola, Tullio, Italy
Vira, Juhani, Finland
Virkkunen, Juhani, Finland
Visintin, Augusto, Italy
Visti, Heikki Matias, Finland
Vitter, Albert Leopold, III, USA
Vituškin, A. G., USSR
Viviente, José Luis, Spain
Vladimirov, Vasilii, USSR
Vogel, Pierre, France
Vogel, Walter O., BRD
Voiculescu, Dan-Virgil, Romania
Voivalin, Matti, Finland
Volkmann, Bodo, BRD
Volkmann, Peter, BRD
Volpe, Carolyn C., USA
Volpe, Celso, Brazil
Vorst, Ton, Netherlands
Vought, Eldon, USA
Voutilainen, Raimo K., Finland
Vrabec, Jana, Yugoslavia
Vrabec, Jože, Yugoslavia
de Vries, Hendrik, Netherlands
de Vries, Jan, Netherlands

Vučkovič, Slobodanka, Yugoslavia
Vujičić, Milan, Yugoslavia
Vuorinen, Matti, Finland
Vysokolán, Stephan, Paraguay

Waadeland, Haakon, Norway
van der Wall, Robert, Netherlands
Wada, Hidekazu, Japan
Wagner, Gerold, BRD
Wagoner, John B., USA
Wake, Graeme Charles, New Zeland
Walker, Peter Leslie, UK
Walter, Jan, Netherlands
Wakulicz, Andrzej, Poland
Wall, Charles T. C., UK
Wall, G. Elliott, Australia
Wallach, Nolan R., USA
Wallbank, Sarah Annie, UK
Wallen, Lasse, Finland
Walsh, Bertram, USA
Warga, Jack, USA
Wasén, Rolf H., Sweden
Wasow, Wolfgang R., USA
Watanabe, Hiroshi, Japan
Watanabe, Mineko, Japan
Watanabe, Shinzo, Japan
Waterman, Daniel, USA
Webb, Glenn F., USA
Webb, Ursula H. M., UK
Webber, David Brian, UK
Weck, Sibylle, BRD
Wedell, Horst, BRD
Weglorz, Bogdan Zbigniew, Poland
Wegner, Gerd, BRD
Weibel, Charles, USA
Weil, André, USA
Weinert, Hans J., BRD
Weinstein, Alan, USA
Weintraub, Steven Howard, USA
Weinzweig, A. I., USA
Weisfeiler, Boris, USA
Weiss, Gary, USA
Weit, Yitzhak, Israel
von Weizsäcker, Heinrich, BRD
Wejntrob, Leon, Israel
Wells, R. O., Jr., USA
Wentzell, A. D., USSR
Werner, Helmut, BRD
Wesselius, Willem, Netherlands
West, James E., USA
West, T. T., Ireland
Weston, Kenneth, USA
Westzynthius, Erik, Finland
Wettstein, Brigitte, Switzerland
Wheeler, David H., Canada
Whitman, Andrew P., USA
Whitney, Hassler, USA
Whittaker, James V., Canada
Widdra, Wolfhardt, BRD
Widman, Kjell-Ove, Sweden
Wilhelmi, Gerhard, BRD
Wilkerson, Clarence W., USA
Wille, Rudolf, BRD
Williams, Bruce, USA
Williams, Daniel Arthur, III, USA
Williams, Graham Dudley, UK
Williams, James P., USA
Williams, Janet, USA
Williams, Robert Fones, USA
Williamson, John H., UK
Wills, Jörg, BRD
Wilson, John Stuart, UK
Wilson, P. M. H., UK
Wilson, Robert Ross, USA
Wilson, Robin J., UK
Wilson, Theodore, USA
Winkler, Jörg, BRD
Winter, David J., USA
Wisbauer, Robert, BRD
Wissner, Heinz W., BRD
Witte, Jürgen, BRD
Witting, Herman, BRD
Wittstock, Gerd, BRD
Włodarski, Lech, Poland
Wohlfahrt, Klaus, BRD
Woivalin, Pentti, Finland
Wójcik, Aleksander R., Poland
Wojdyło, Bolesław, Poland
Wojtaszczyk, Przemysław, Poland
Wojtkowiak, Zdzisław, Poland
Wojtynski, Wojciech L., Poland
Wold, Anna-Lisa, Sweden
Wolf, Rudolf, BRD
Wolff, C. Henrik E., Finland
Wolffhardt, Klaus, BRD
Wolff von Gudenberg, Jürgen, BRD
Wollny, Wolfgang, BRD
Wolter, Franz-Erich, BRD
Wong, Bun, Hong Kong
Wong, Chi Song, Canada
Wong, Raymond Y., USA
Woźniak, Czesław, Poland
Wraith, Gavin Charles, UK
Wright, David, USA
Wu, Pei Yuan, China-Taiwan
Wu, Yel-Chiang, USA
Wulfsohn, Aubrey, UK
Wussing, Hans, DDR
Wüst, Rainer BRD

Xuan Loc Nguyen, Vietnam

Yaku, Takeo, Japan
Yamaguchi, Michinari, Japan
Yamamoto, Atsuko, Japan
Yamashima, Shigeho, Japan
Yamazaki, Seishi, Japan
Yano, Kentaro, Japan
Yau, Shing-Tung
Yeh, Neng-Che, China-Taiwan
Yger, Alain, France
Ylinen, Kari, Finland
Yoshizawa, Taro, Japan
Yuan, Shuen, USA
Yui, Noriko, Denmark

Zachariou, Andreas, Greece
Zafarani, Jafar, Iran
Zagier, Don, USA
Zahorski, Zysmunt, Poland
Zdravkovska, Smilka, Yugoslavia
Zéboulon, Hélyette, France
Żelazko, Wieslaw, Poland
Zelinka, Bohdan, Czechoslovakia
Zeller, Karl, BRD
Zelmer, Graham Keith, Canada
Zemánek, Jaroslav, Czechoslovakia
Zhidkov, Yevgeni, USSR

Ziegler, Zvi, Israel
Žilinková, Júlia, Czechoslovakia
Ziller, Wolfgang, BRD
Zimmer, Robert J., USA
Zimmermann, Bernd, BRD
Zimmermann, Birge, BRD
Zippo, Shiro, Japan
Zirilli, Francesco, Italy
Žižčenko, A. B., USSR
Zizi, Khelifa, France
Zougdani, Hassan, Libya
Zucker, Steven, USA
Zuckerman, Gregg J., USA
Zuker, Michael, Canada
Zulauf, Achim, New Zealand
Zvieli, Dvora, Israel
Zygmund, Antoni, USA

Fields Medallists

On the decision of the Fields Medals Committee, the works of the Fields medallists were presented as follows:

N. M. KATZ: The work of Pierre Deligne
L. CARLESON: The work of Charles Fefferman
J. TITS: The work of Gregori Aleksandrovitch Margulis
I. M. JAMES: The work of Daniel Quillen

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Work of Pierre Deligne

N. M. Katz

My purpose here is to convey to you some idea of the scope and the depth of the work for which we are today honoring Pierre Deligne with the Fields Medal. Deligne's work centers around the remarkable relations, first envisioned by Weil, which exist between the cohomological structure of algebraic varieties over the complex numbers, and the diophantine structure of algebraic varieties over finite fields.

I. The Weil conjectures. Let us first consider an algebraic variety Y over a finite field $\boldsymbol{F}_q$. For each integer $n \geqslant 1$ there is a unique field extension $\boldsymbol{F}_{q^n}$ of degree n over $\boldsymbol{F}_q$. We denote by $Y(\boldsymbol{F}_{q^n})$ the (finite) set of points of Y with coordinates in $\boldsymbol{F}_{q^n}$, and by $\# Y(\boldsymbol{F}_{q^n})$ the cardinality of this set. The zeta function of Y over $\boldsymbol{F}_q$ is the formal series defined by

$$Z(Y/\boldsymbol{F}_q, T) = \exp\left(\sum_{n \geqq 1} \frac{T^n}{n} \# Y(\boldsymbol{F}_{q^n})\right).$$

Knowledge of the zeta function is equivalent to knowledge of the numbers $\{\# Y(\boldsymbol{F}_{q^n})\}$.

After the pioneering work of E. Artin, W. K. Schmid, H. Hasse, M. Deuring and A. Weil on the zeta functions of curves and abelian varieties, Weil in 1949 made the following conjectures about the zeta function of a projective non-singular n-dimensional variety Y over a finite field $\mathbf{F}_q$.

(1) The zeta function is a rational function of T, i.e. it lies in $\boldsymbol{Q}(T)$.

(2) There exists a factorization of the zeta function as an alternating product of polynomials $P_0(T), \ldots, P_{2n}(T)$,

$$Z(Y/\boldsymbol{F}_q, T) = \frac{P_1(T)P_3(T)\ldots P_{2n-1}(T)}{P_0(T)P_2(T)\ldots P_{2n}(T)}$$

of the form

$$P_i(T) = \prod_{j=1}^{b_i} (1 - \alpha_{i,j} T),$$

such that the map $\alpha \mapsto q^n/\alpha$ carries the $\alpha_{i,j}$ bijectively to the $\alpha_{2n-i,j}$.

(3) The polynomials $P_i(T)$ lie in $\mathbf{Z}[T]$, and their reciprocal roots $\alpha_{i,j}$ are algebraic integers which, together with all their conjugates, satisfy

$$|\alpha_{i,j}| = q^{i/2}.$$

This is the "Riemann Hypothesis" for varieties over finite fields.

(4) If Y is the "reduction mod p" of a projective smooth variety $\boldsymbol{Y}$ in characteristic zero, then the degree b_i of P_i is the ith Betti number of $\boldsymbol{Y}$ as complex manifold.

Underlying these conjectures was Weil's belief in the existence of a "cohomology theory", with a coefficient field of characteristic zero, for varieties over finite fields. In this theory, the polynomial $P_i(T)$ would be the "inverse" characteristic polynomial $\det(1-TF)$ of the "Frobenius endomorphism" acting on H^i. Conjectures (1) and (2) would then follow from a Lefschetz trace formula for F and its iterates, and from a suitable form of Poincaré duality. Conjecture (4) would follow if the cohomology of the "reduction mod p" Y of a projective smooth variety $\boldsymbol{Y}$ in characteristic zero were (essentially) equal to the topological cohomology of $\boldsymbol{Y}$ as complex manifold.

The next years saw the systematic introduction of sheaf-theoretic and cohomological methods into algebraic geometry. By the mid-1960s, M. Artin and A. Grothendieck had developed the étale cohomology theory of arbitrary schemes, along the lines foreseen in Grothendieck's 1958 Edinburgh address. For each prime number l, this gives a cohomology theory, "l-adic cohomology", with coefficients in the field $\mathbf{Q}_l$ of l-adic numbers, which is adequate to give parts (1), (2) and (4) of the Weil conjectures for projective smooth varieties over finite fields of characteristic $p \neq l$. In their theory, the cohomology of the "reduction mod p" Y of a projective smooth $\boldsymbol{Y}$ in characteristic zero is just the singular cohomology, with coefficients in $\mathbf{Q}_l$, of "$\boldsymbol{Y}$ as complex manifold". In the case of curves and abelian varieties, these constructions agree with those already given by Weil.

For a given projective smooth $Y/\mathbf{F}_q$, we now have, for *each* $l \neq p$, a factorization of the zeta function as an alternating product of l-adic polynomials $P_{i,l}(T)$. There is, however, no assurance that the $P_{i,l}$ have coefficients in $\mathbf{Q}$ rather than in $\mathbf{Q}_l$, much less that their reciprocal zeros are algebraic integers with the predicted absolute values. Of course, if one could prove directly that the reciprocal zeros of $P_{i,l}$ were algebraic integers which, together with all their conjugates, had the correct absolute value $q^{i/2}$, then the polynomials $P_{i,l}$ could be described intrinsically in terms of the zeros and poles of the zeta function itself, and hence would have rational coefficients independent of l. But how could one even introduce archimedean considerations into the l-adic theory without first knowing the rationality of the cofficients of the $P_{i,l}$?

Let me now try to indicate the brilliant synthesis of ideas involved in Deligne's solution of these problems.

Initially, he tries to prove à priori that the $P_{i,l}$ have rational coefficients independent of l. The idea is to proceed by induction on the dimension of Y. If Y is n-dimensional, then Poincaré duality and the fact that the zeta function is itself rational and independent of l reduce us to treating the polynomials $P_{i,l}$ for $i \leqslant n-1$. Now let Z be a smooth hyperplane section of Y. The "weak" Lefschetz theorem assures us that Y and Z have the same $P_{i,l}$ for $i \leqslant n-2$, and that the $P_{n-1,l}$ for Y divides that for Z. This alone is enough to show inductively that the reciprocal zeroes of the $P_{i,l}$ are algebraic integers.

In order to go further, and show that the $P_{i,l}$ actually have rational coefficients independent of l, the idea is to show that $P_{n-1,l}$ for Y is a generalized "greatest common divisor" of the $P_{n-1,l}$ of all possible smooth hyperplane sections. Unfortunately, this "g.c.d." argument, which itself depends on the full strength of the monodromy theory of Lefschetz pencils, works only when Y satisfies the "hard" Lefschetz theorem (existence of the "primitive decomposition" on its cohomology), otherwise the "g.c.d." will be too big at some stage of the induction. But Deligne will later prove the hard Lefschetz theorem in arbitrary characteristic as a *consequence* of the Weil conjectures. What is to be done?

With characteristic daring, Deligne simply ignores the preliminary problem of establishing independence of l. Fixing one $l \neq p$, he turns to a direct attack on the absolute values of the algebraic integers which occur as the reciprocal roots of the $P_{i,l}$.

Consider a smooth projective even dimensional Y, and a Lefschetz pencil Z_t of hyperplane sections, "fibering" Y over the t-line. Factor the $P_{n-1,l}$ of each Z_t as the product of the "g.c.d." of all of them, and of the "variable" part. Deligne shows à priori that these "variable" parts are each polynomials with rational coefficients whose reciprocal zeroes all satisfy the Riemann Hypothesis

$$|\alpha_{n-1,\text{variable}}| = q^{(n-1)/2}.$$

Deligne's proof of this is simply spectacular; no other word will do. He first uses a theorem of Kazdan–Margoulis, according to which the monodromy group of a Lefschetz pencil of odd fibre dimension is "as big as possible", to establish the rationality of the coefficients of the "variable" parts. Then he considers the L-function over the t-line whose Euler factors are the reciprocals of the "variable" parts. This L-function has rational Dirichlet coefficients. Deligne realizes that Rankin's method of estimating Ramanujan's function $\tau(n)$ by "squaring" might be applied in this context to estimate the reciprocal poles of the individual Euler factors (i.e. the reciprocal zeroes of the "variable" parts!). The problem is to control the *poles* of all the L-functions obtained from this one by passing to *even* tensor powers ("squaring"). Deligne gains this control by ingeniously combining Grothendieck's cohomological theory of such L-functions, the Kazdan–Margoulis theorem, and the classical invariant theory of the symplectic group!

Once he has this à priori estimate for the variable parts of the $P_{n-1,t}$ of the hyperplane sections Z_t, a Leray spectral sequence argument shows that in the $P_{n,t}$ for Y itself, all the reciprocal zeroes are algebraic integers which, together with all their conjugates, satisfy the apparently too weak estimate

$$|\alpha_{n,j}| \leq q^{(n+1)/2} \quad \text{(instead of } q^{n/2}\text{)}.$$

But this estimate is valid for Y of any *even* dimension n. The actual Riemann hypothesis for any projective smooth variety X follows by applying this estimate to all the even cartesian powers of X.

II. Consequences for number theory. That there are many spectacular consequences for number theory comes as no surprise. Let us indicate a few of them.

(1) Estimation of $\# Y(F_q)$ when Y has a "simple" cohomological structure. For example, if Y is a smooth n-dimensional hypersurface of degree d, we get

$$|\# Y(F_q)-(1+q+\ldots+q^n)| \leq \left(\frac{(d-1)^{n+2}+(-1)^{n+2}(d-1)}{d}\right) q^{n/2}.$$

(2) Estimates for exponential sums in several variables, e.g

(a) if f is a polynomial over F_p in n variables of degree d prime to p, whose part of highest degree defines a nonsingular projective hypersurface, then

$$\left|\sum_{x_i \in F_p} \exp\left(\frac{2\pi i}{p} f(x_1, \ldots, x_n)\right)\right| \leq (d-1)^n p^{n/2};$$

(b) "multiple Kloosterman sums":

$$\left|\sum_{x_i \in F_p^\times} \exp\left(\frac{2\pi i}{p}\left(x_1+\ldots+x_n+\frac{1}{x_1\ldots x_n}\right)\right)\right| \leq (n+1)\cdot p^{n/2}.$$

(3) The Ramanujan–Petersson conjecture. Already in 1968 Deligne had combined techniques of l-adic cohomology and the arithmetic moduli of elliptic curves with earlier ideas of Kuga, Sato, Shimura and Ihara to reduce this conjecture to the Weil conjectures. Thus if $\sum a(n)q^n$ is the q-expansion of a normalized $(a(1)=1)$ cusp form on $\Gamma_1(N)$ of weight $k \geq 2$ which is a simultaneous eigenfunction of all Hecke operators, then

$$|a(p)| \leq 2p^{(k-1)/2} \quad \text{for all primes} \quad p \nmid N.$$

III. Cohomological consequences; weights. As Grothendieck foresaw in the 1960s with his "yoga of weights", the truth of the Weil conjectures for varieties over finite fields would have important consequences for the cohomological structure of varieties over the complex numbers. The idea is that any reasonable algebro-geometric situation over C is actually defined over a subring of C which, as a ring, is finitely generated over Z. Reducing modulo a maximal ideal $\mathfrak{m}$ of this ring, we find a situation over a finite field, and the corresponding Frobenius endomorphism $F(m)$ operating on this situation. This Frobenius operates by functoriality on the l-adic cohomology, which is none other than the singular cohomology, with $\mathbf{Q}_l$-coefficients, of our original situation over $\mathbf{C}$. This natural operation of Frobenius imposes a previously unsuspected structure on the cohomology of complex algebraic varieties, the so-called "weight filtration", or filtration by the magnitude of the eigenvalues of Frobenius.

In a remarkable tour de force in the late 1960's and early 1970's, Deligne developed, independently of the Weil conjectures, a complete theory of the weight filtration of complex algebraic varieties, by making systematic use of Hironaka's resolution of singularities, of the notion of differential forms with "logarithmic poles" (i.e. products of dt/t's), and of his own earlier work on cohomological descent. The resulting theory, which Deligne named "mixed Hodge theory", should be seen as a far-reaching generalization of the classical theory of "differentials of the second kind" on algebraic varieties, as well as of "usual" Hodge theory.

Consider, for example, a smooth affine variety U over $\mathbf{C}$. By one of Hironaka's fundamental results, we can find a projective smooth variety X and a collection of smooth divisors D_i in X which cross transversally, such that $U \cong X - \cup D_i$. Deligne shows that the Leray spectral sequence, in rational cohomology, of the inclusion map $U \hookrightarrow X$, degenerates at E_3, and that the filtration it defines on the cohomology of U is independent of the choice of the compactification. This is the desired weight filtration; its smallest filtrant is the image of $H^\bullet(X)$ in $H^\bullet(U)$, i.e. the space of "differentials of the second kind on U".

One of the many applications of this theory is to the global monodromy of families of projective smooth varieties. Given a projective smooth map $X \to S$ of smooth complex varieties, and a point $s \in S$, the fundamental group $\pi_1(S, s)$ acts on each of the cohomology groups $H^i(X_s, \mathbf{C})$ of the fibre X_s. Deligne shows that these representations are all completely reducible, and that each of their isotypical components, especially the space of invariants, is stable under the Hodge decomposition into (p, q)-components.

By means of an extremely ingenious and difficult argument drawing upon Grothendieck's cohomological theory of L-functions and the ideas of the Hadamard–de la Vallée Poussin proof of the prime number theorem, Deligne later established an l-adic analogue of this theorem of complete reducibility for l-adic "local systems" in characteristic p over open subsets of the projective t-line $\mathbf{P}^1$, provided that all of the "fibres" of the local system satisfy the Riemann Hypothesis (with a fixed power of $\sqrt{q}$). Once he had *proven* the Riemann Hypothesis for varieties over finite fields, he could apply this theorem to the local system coming from a Lefschetz pencil on a projective smooth variety over a finite field. The resulting complete reducibility is easily seen to imply the hard Lefschetz theorem. This theorem, previously known only over $\mathbf{C}$, and there by Hodge's theory of harmonic integrals, is thus established, in all characteristics as a consequence of the Weil conjectures.

Other outgrowths of Deligne's work on weights include his theory of differential equations with regular singular points on smooth complex varieties of arbitrary dimension, which yields a new solution of Hilbert's 21st problem, and his affirmative solution of the "local invariant cycle problem" in the local monodromy theory of families of projective smooth varieties.

Still another application of "weights" is to homotopy theory. Deligne, Griffiths, Morgan and Sullivan jointly apply mixed Hodge theory to prove that the homotopy theory ($\otimes Q$) of a projective smooth complex variety is a "formal consequence" of its cohomology.

IV. Other work. We have passed over in silence a considerable body of Deligne's work, which alone would be sufficient to mark him as a truly exceptional mathematician; duality in coherent cohomology, moduli of curves (jointly with Mumford), arithmetic moduli (jointly with Rapoport), the Ramanujan–Petersson conjecture for forms of weight one (jointly with Serre), the Macdonald conjecture (jointly with Lusztig), local "roots numbers", p-adic L-functions (jointly with Ribet), motivic L-functions, "Hodge cycles" on abelian varieties, and much more.

I hope that I have conveyed to you some sense, not only of Deligne's accomplishments, but also of the combination of incredible technical power, brilliant clarity, and sheer mathematical daring which so characterizes his work.

PRINCETON UNIVERSITY
PRINCETON, N.J. 08540, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Work of Charles Fefferman

Lennart Carleson

There was a period, in the 1940s and 1950s, when classical analysis was considered dead and the hope for the future of analysis was considered to be in the abstract branches, specializing in generalization. As is now apparent, the rumour of the death of classical analysis was greatly exaggerated and during the 1960s and 1970s the field has been one of the most successful in all of mathematics. Briefly, I think that one can say that the reasons for this are the unification of methods from harmonic analysis, complex variables and differential equations, the discovery of the correct generalizations to several variables and finally the realization that in many problems complications cannot be avoided and that intricate combinatorial arguments rather than polished theories often are in the centre.

This general description of classical analysis also summarizes the work of Charles Fefferman. In an eminent way he masters these techniques and has contributed to the success of our common field and it is with real joy and pride, as a friend and as co-worker in the field, I shall try to sketch some lines in the development with emphasis on certain of Fefferman's many contributions.

It is natural to start with the Hardy spaces H^p, i.e. functions $f(z)$ holomorphic in $|z|<1$ and belonging L^p on the boundary of $|z|=1$. Through the work of Marcel Riesz we know that H^p is the dual of H^q for conjugate exponents p and q for $1<p<\infty$, and the theory becomes similar to the L^p, L^q-theory. There was, however, no analogy to the L^1, L^∞-duality and special methods were necessary for every situation for H^1. It was therefore a great sensation when Fefferman in 1971 showed that the dual of H^1 was a space that had been used a few years earlier by John and Nirenberg, the space BMO of functions of bounded mean oscillation. This is the space of functions which on every interval differs in the mean from its mean value by a bounded quantity. A canonical non-bounded example is the logarithm of the absolute value. Many problems for H^1 now become concrete, constructive problems for this class. As a simple illustration of the force of the method, consider Hardy's theorem that if $f(z)=\sum_0 c_n z^n \in H^1$ then $\sum |c_n| n^{-1} < \infty$. Dually this means that

$$\sum \varepsilon_n n^{-1} e^{in\theta} \in \mathrm{BMO}, \ |\varepsilon_n| = 1,$$

and this is essentially trivial to verify.

The idea, however, carries much further. The result from 1971 by Gundy, Burkholder and Silverstein that a harmonic function $u(z)$ in $|z|<1$ is the real part of an H^1-function if and only if

$$\sup_{z \in V_\theta} u(z) \in L^1$$

where V_θ is the Stolz angle at $e^{i\theta}$, also gets a natural explanation as a representation problem for BMO. The interesting result appears that we need only take the sup over the radius. It is clear how this generalizes to several dimensions and we have in this theory one of the most rapidly expanding branches of analysis. In particular, I should like to mention the recent theory of Muckenhaupt, Wheeden and others, where also Fefferman has contributed essentially, generalizing the L^p-theory of conjugate functions to weighted L^p-spaces. The culmination is Calderón's recent work on singular integrals on C^1-curves which you will hear more about during the congress.

In the centre of this development is the theory of singular integrals and different maximal versions of these integrals. In particular, the maximal partial sum operator for a Fourier series is essentially such a maximal operator

$$S^*(f) = \sup_n S_n(f)(x) = \sup_n \int \frac{f(t)e^{-int}}{x-t}\, dt.$$

Fefferman has given a direct combinatorial proof in the spirit of Kolmogorov that $S_{n(x)}(x)$ for arbitrary choice of $n(x)$ is uniformly bounded on L^2 and hence a new proof of the a.e. convergence of the Fourier series of a continuous function. We have of course similar formulas in several variables. It is remarkable that Fefferman was the first to find a counterexample showing that no similar result holds for rectangular partial sums in several variables, even if we make strong restrictions on the ratio of the sides of the rectangles. This is a result that should have been proved 100 years ago!

We have seen here how the interplay between ideas in real and complex analysis has given striking and deep new results and that singular integrals and Fourier analysis were the main tools. These tools are also, as you all know, closely tied to partial differential equations with constant coefficients because of the algebraic way in which Fourier transform reflects derivations. In a similar way one can also treat differential equations with variable coefficients — the idea is to introduce in the Fourier transform a function p — the symbol — which also depends on the space variables:

$$(Pu)(x) = \iint e^{i(x-y)\cdot\xi} p(x, \xi, y) u(y)\, dy\, d\xi.$$

We can make the theory still more general by replacing $(x-y)\cdot\xi$ in the exponent by more general functions. This will be important for us later.

The theory becomes highly technical and a careful classification of symbols p in terms of estimates of derivatives is necessary. In joint work with R. Beals, Fefferman has introduced a new weighted classification. The main application is a new proof of a result of Nirenberg and Trèves for the local solvability of a partial differential equation of principal type. Using these methods Fefferman and Phong recently showed a best possible version of the sharp Gårding inequality, i.e. if $p(x, \xi)\geqslant 0$ and is at most of second order in ξ then

$$(p(x, D)u, u) \geqslant -C\|u\|_2^2.$$

There is a natural connection back from partial differential equations to complex analysis in the classical Cauchy–Riemann equations. In several variables, these are really a system of equations and an important difference between one and several complex variables is that certain of these equations $\overline{\partial_b} f=0$ also make sense on the boundary of the domain Ω. In particular, if f is given on the boundary there is a natural L^2-projection on solutions of these equations. This projection is realized by a kernel, the Szegö kernel. Similarly, the projection corresponding to L^2 for the volume of Ω is given by the Bergman kernel $K(z, \zeta)$. Clearly, in the centre of interest, we have the regularity of K as the points approach the boundary. It was shown by Kerzman that for strictly pseudo convex domains, the singularities appear as $z\to\zeta$ and the case $z=\zeta$ is particularly interesting.

If Ω is a strictly pseudo convex domain given by a smooth plurisubharmonic function ψ so that $\Omega: \psi<0$, then Hörmander proved in 1965 that for $z_0\in\partial\Omega$

$$\lim_{z\to z_0} \psi(z)^{n+1}K(z, z) = \frac{n!}{\pi^n} \det \begin{pmatrix} \frac{\partial\psi}{\partial z_1} & & \\ \vdots & \tilde{\Delta}\psi = \left(\frac{\partial^2\psi}{\partial z_\nu \partial \bar{z}_\mu}\right) & \\ \frac{\partial\psi}{\partial z_n} & & \\ \psi & \frac{\partial\psi}{\partial \bar{z}_1} \cdots \frac{\partial\psi}{\partial \bar{z}_n} & \end{pmatrix} = \frac{n!}{\pi^n} L(\psi).$$

By a direct very ingenious construction Fefferman (1974) obtained a complete asymptotic formula which to everybody's surprise contained a logarithmic singularity:

$$(-\psi)^{n+1}K(z, z) = F(z)+G(z)(-\psi)^{n+1}\log(-\psi)$$

with F and G smooth. As was shown later by Boutet de Monvel and Sjöstrand this singularity can be best understood in the context of Fourier integral operators. Let $\psi(z, \zeta)$ be a convenient, explicit continuation of $\psi(z, z)=\psi(z)$ from the diagonal in Ω to $C^n\times C^n$. Then $K(z, \zeta)$ is essentially a Laplace transform of the type

$$\int_0^\infty e^{t\psi(z,\zeta)} k(z, \zeta; t)\, dt$$

where k has a singularity at $t=\infty$ of the type t^n which just produces singularities of Fefferman's type at the boundary where $\psi=0$.

Fefferman's interest in the Bergman kernel originated in his desire to show the regularity of biholomorphic mappings up to the boundary of smooth domains. The metric

$$ds^2 = \sum \frac{\partial^2 \log K}{\partial z_\nu \, \partial \bar{z}_\mu} dz_\nu \, d\bar{z}_\mu$$

is invariant under the mapping and its geodesics G are therefore natural bases for a geometric description of the correspondence between the boundaries of the domains, *i.e.* if we can show that G approaches a definite boundary point and that directions correspond smoothly to the boundary. Since we have a precise knowledge of the behavior of the metric at $\partial\Omega$ all becomes concrete differential geometric problems. The difficulties are however very serious because of the singular behavior of K at $\partial\Omega$ but were mastered by Fefferman in a remarkable way.

Through Fefferman's result we now have a foundation for a discussion of the mapping on the boundary. The fundamental problem is to classify domains which are biholomorphically equivalent or locally so. We shall only consider the local problem and then need a set of local invariants. In the case $n=2$ all local invariants were formed already by Eli Cartan and Chern and Moser gave a complete theory. In particular they found certain invariant curves; the chains. You will hear more about this in Jürgen Moser's lecture. Fefferman has given a differential geometric description of the Chern–Moser-chains derived from certain geodesics for a metric, which is related to the Bergman kernel function. He has also started the big program to find algebraic descriptions of the local invariants. In all probability we are here at the beginning of a completely new theory in several complex variables.

I should like to finish by pointing to an alternative approach which is very attractive to a classical analyst. The most important tool in one complex variable is the harmonic functions. The class is invariant under conformal mappings because $\Delta(u \circ f) = |f'|^2 \Delta u$. The natural analogue in several variables is the Hessian determinant

$$|\tilde{\Delta}| u = \det\left(\frac{\partial^2 u}{\partial z_\nu \, \partial \bar{z}_\mu}\right).$$

We have already seen $\tilde{\Delta}$ in Hörmander's formula Lu for $K(z, z)$ and actually $\tilde{\Delta}$ and L are related by the change $u \to \log u$. The study of the equation $\tilde{\Delta} u = \varphi$ has started but basic estimates are still missing. Fefferman has made the important observation that the equation $Lu=1$, $u=0$ on $\partial\Omega$, can be solved approximately, again with regularity until the critical singularities enter for the nth derivative. I am sure that we see here another example of the beginning of an important theory.

I hope this brief survey has convinced you of the vitality of classical analysis and of the great contributions of Charles Fefferman.

INSTITUT MITTAG-LEFFLER
(18262) DJURSHOLM, SWEDEN

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Work of Gregori Aleksandrovitch Margulis

J. Tits

The work of Margulis belongs to combinatorics, differential geometry, ergodic theory, the theory of dynamical systems and the theory of discrete subgroups of real and p-adic Lie groups. In this report, I shall concentrate on the last aspect which covers his main results.

1. Discrete subgroups of Lie groups. The origin. Discrete subgroups of Lie groups were first considered by Poincaré, Fricke and Klein in their work on Riemann surfaces: if M is a Riemann surface of genus $\geqslant 2$, its universal covering is the Lobatchevski plane (or Poincaré half-plane), therefore the fundamental group of M can be identified with a discrete subgroup Γ of $\mathrm{PSL}_2(\boldsymbol{R})$; the problem of uniformization and the theory of differentials on M lead to the study of automorphic forms relative to Γ.

Other discrete subgroups of Lie groups, such as $\mathrm{SL}_n(\boldsymbol{Z})$ (in $\mathrm{SL}_n(\boldsymbol{R})$) and the group of "units" of a rational quadratic form (in the corresponding orthogonal group) play an essential role in the theory of quadratic forms (reduction theory) developed by Hermite, Minkowski, Siegel and others. In constructing a space of moduli for abelian varieties, Siegel was led to consider the "modular group" $\mathrm{Sp}_{2n}(\boldsymbol{Z})$, a discrete subgroup of $\mathrm{Sp}_{2n}(\boldsymbol{R})$.

The group $\mathrm{SL}_n(\boldsymbol{Z})$, the group of units of a rational quadratic form and the modular group are special instances of "arithmetic groups", as defined by A. Borel and Harish-Chandra. A well-known theorem of those authors, generalizing classical results of Fricke, Klein, Siegel and others, asserts that if Γ is an arithmetic subgroup of a semi-simple Lie group G, then the volume of G/Γ (for any G-invariant measure) is finite; we say that Γ has *finite covolume* in G. The same holds for

$G=\mathrm{PSL}_2(\boldsymbol{R})$ if Γ is the fundamental group of a Riemann surface "with null boundary" (for instance, a compact surface minus a finite subset).

Already Poincaré wondered about the possibility of describing all discrete subgroups of finite covolume in a Lie group G. The profusion of such subgroups in $G=\mathrm{PSL}_2(\boldsymbol{R})$ makes one at first doubt of any such possibility. However, $\mathrm{PSL}_2(\boldsymbol{R})$ was for a long time the only simple Lie group which was known to contain non-arithmetic discrete subgroups of finite covolume, and further examples discovered in 1965 by Makarov and Vinberg involved only few other Lie groups, thus adding credit to conjectures of Selberg and Pyatetski–Shapiro to the effect that "for most semisimple Lie groups" discrete subgroups of finite covolume are necessarily arithmetic. Margulis' most spectacular achievement has been the complete solution of that problem and, in particular, the proof of the conjectures in question.

2. The noncocompact case. Selberg's conjecture. Let G be a semisimple Lie group. To avoid inessential technicalities, we assume that G is the group of real points of a real simply connected algebraic group $\mathscr{G}$ which we suppose embedded in some $\mathrm{GL}_n(\boldsymbol{R})$, and that G has no compact factor. Let Γ be a discrete subgroup of G with finite covolume and irreducible in the sense that its projection in any nontrivial proper direct factor of G is nondiscrete. Suppose that the real rank of G is $\geqslant 2$ (this means that G is not a covering group of the group of motions of a real, complex or quaternionic hyperbolic space or of an "octonionic" hyperbolic plane) and that G/Γ *is not compact.* Then, Selberg's conjecture asserts that Γ is arithmetic which, in this case, means the following: there is a base in $\boldsymbol{R}^n$ with respect to which $\mathscr{G}$ is defined by polynomial equations with rational coefficients and such that Γ is commensurable with $\mathscr{G}(\boldsymbol{Z})=G\cap\mathrm{GL}_n(\boldsymbol{Z})$ (i.e. $\Gamma\cap\mathscr{G}(\boldsymbol{Z})$ has finite index in both Γ and $\mathscr{G}(\boldsymbol{Z})$). Selberg himself proved that result in the special case where G is a direct product of (at least two) copies of $\mathrm{SL}_2(\boldsymbol{R})$.

A first important step toward the understanding of noncompact discrete subgroups of finite covolume was the proof by Každan and Margulis [2] of a related, more special conjecture of Selberg: *under the above assumptions* (except that no hypothesis is made on $\mathrm{rk}_{\boldsymbol{R}}G$), Γ *contains nontrivial unipotent elements of* G (i.e. elements all of whose eigenvalues are 1). This was a vast generalization of results already known for $\mathrm{SL}_2(\boldsymbol{R})$ and products of copies of $\mathrm{SL}_2(\boldsymbol{R})$ (Selberg); in view of a fundamental theorem of Borel and Harish–Chandra ("Godement's conjecture"), it had to be true if Γ was to be arithmetic. Let us also note in passing another remarkable byproduct of Každan–Margulis' method: *given* G, *there exists a neighborhood* W *of the identity in* G *such that for every* Γ *(cocompact or not), some conjugate of* Γ *intersects* W *only at the identity*; in particular, *the volume of* G/Γ *cannot be arbitrarily small* (for a given Haar measure in G). For $G=\mathrm{SL}_2(\boldsymbol{R})$, the last assertion had been proved by Siegel, who had also given the exact lower bound of vol (G/Γ) in that case. A. Borel reported on those results of Každan and Margulis at Bourbaki Seminar [26].

The existence of unipotent elements in Γ was giving a hold on its structure. In

[6], Margulis announced, among others, the following result which was soon recognized by the experts as a crucial step for the proof of Selberg's conjecture:

in the space of lattices in $\mathbf{R}^n$, *the orbits of a one-parameter unipotent subsemigroup of* $\mathrm{GL}_n(\mathbf{R})$ *"do not tend to infinity" (in other words, a closed orbit is periodic).*

For a couple of years, Margulis' proof remained unpublished and every attempt by other specialists to supply it failed. When it finally appeared in [9], the proof came as a great surprise, both for being rather short and using no sophisticated technique: it can be read without any special knowledge and gives a good idea of the extraordinary inventiveness shown by Margulis throughout his work.

Using unipotent element, it is relatively easy to show that, G and Γ being as above, there is a $\mathbf{Q}$-structure $\mathscr{G}$ on G such that $\Gamma \subset \mathscr{G}(\mathbf{Q})$. The main point of Selberg's conjecture is then to show that the matrix coefficients of the elements of Γ have bounded denominators. In [15], Margulis announced a complete proof of the conjecture and gave the details under the additional assumption that the $\mathbf{Q}$-rank of $\mathscr{G}$ is at least 2. Another proof under the same restriction was given independently by M. S. Raghunathan. The much more difficult case of a $\mathbf{Q}$-rank one group is treated by Margulis in [19], by means of a very subtle and delicate analysis of the set of unipotent elements contained in Γ. The main techniques used in [15] and [19] are those of algebraic group theory and p-adic approximation.

3. The cocompact case. Rigidity. Margulis was invited to give an address at thf Vancouver Congress, no doubt with the idea that he would expose his solution oe Selberg's conjecture. Instead, prevented (as this time) from attending the Congress, he sent a report on completely new and totally unexpected results on the cocompact case [18].

That case, about which nothing was known before, presented two great additional difficulties which nobody knew how to handle. On the one hand, if G/Γ is compact, Γ contains no unipotent element, so that the main technique used in the other case is not available. But there is another basic difficulty in the very notion of arithmetic group: let G, Γ be as in § 2 except that G/Γ is no longer assumed to be non-compact; then Γ is said to be arithmetic if there exist an algebraic linear semi-simple simply connected group $\mathscr{H}$ defined over $\mathbf{Q}$ and a homomorphism $\alpha: \mathscr{H}(\mathbf{R}) \to G$ with compact kernel such that Γ is commensurable with $\alpha(\mathscr{H}(\mathbf{Z}))$. The point is chat in the non-cocompact case, α is necessarily an isomorphism. In the general case, there is *a priori* no way of knowing what $\mathscr{H}$ will be (in fact, for a given G, $\mathscr{H}$ tan have an arbitrarily large dimension). A conjecture, more or less formulated by Pyatetski-Shapiro at the 1966 Congress in Moscow, to the effect that also in the cocompact case, assuming again $\mathrm{rk}_{\mathbf{R}}\, G \geqslant 2$, Γ had to be arithmetic, was certainly more daring at the time and seemed completely out of reach. It was the proof of that conjecture that Margulis sent, without warning, to the Vancouver Congress.

Arithmetic subgroups of Lie groups are in some sense "rigid"; intuitively, this follows from the impossibility to alter an algebraic number continuously without destroying the algebraicity. On the other hand, theorems of Selberg, Weil and Mostow showed that in semi-simple Lie groups different from $\mathrm{SL}_2(\boldsymbol{R})$ (up to local isomorphism) cocompact discrete subgroups are rigid, and Selberg had observed that rigidity implies a "certain amount of arithmeticity": in fact, it is readily seen to imply that Γ is contained in $\mathscr{G}(K)$ for some algebraic group $\mathscr{G}$ and some number field K. As before, the crux of the matter is the proof that the matrix coefficients of the elements of Γ have bounded denominators. This is achieved by Margulis through a "superrigidity" theorem which, for groups of real rank at least 2, is a vast generalization of Weil's and Mostow's rigidity theorems:

Assume $\mathrm{rk}_{\boldsymbol{R}}\, G \geqslant 2$, *let* F *be a locally compact nondiscrete field and let* $\varrho\colon \Gamma \to \mathrm{GL}_n(F)$ *be a linear representation such that* $\varrho(\Gamma)$ *is not relatively compact and that its Zariski-closure is connected*; *then* $F=\boldsymbol{R}$ *or* $\boldsymbol{C}$ *and* ϱ *extends to a rational representation of* $\mathscr{G}$.

The proof of this theorem is relatively short (considering the power of the result), but is a succession of extraordinarily ingenious arguments using a great variety of very strong techniques belonging to ergodic theory (the "multiplicative ergodic theorem"), the theory of unitary representations, the theory of functional spaces (spaces of measurable maps), algebraic geometry, the structure theory of semi-simple algebraic groups, etc. In 1975–1976, I devoted my course at the Collège de France to those results of Margulis; I believe that I learned more mathematics during that year than in any other year of my life. A summary of the main ideas of that beautiful piece of work is given in [**27**].

Another, quite different proof of the superrigidity theorem and its application to arithmeticity (both in the cocompact and the noncocompact case)—using the work of H. Furstenberg—can be found in [**20**].

4. Other results.

4.1. *S-arithmetic groups.* Let K be a number field, S a finite set of places of K including all places at infinity, $\mathfrak{o}$ the ring of elements of K which are integral at all finite places not belonging to S, $\mathscr{H}\subset\mathscr{G}\mathscr{L}_n$ a simply connected semisimple linear algebraic group defined over K and $\mathscr{H}(\mathfrak{o})=\mathscr{H}\cap\mathrm{GL}_n(\mathfrak{o})$. Then, $\mathscr{H}(\mathfrak{o})$ injects as a discrete subgroup of finite covolume in the product $H=\prod_{v\in S}\mathscr{H}(K_v)$, where K_v denotes the completion of K at v.

(Example: if $\mathfrak{o}$ is the ring of rational numbers whose denominator is a power of 2, $\mathrm{SL}_n(\mathfrak{o})$ is a discrete subgroup of finite covolume of $\mathrm{SL}_n(\boldsymbol{R})\times\mathrm{SL}_n(\boldsymbol{Q}_2)$). Now let G be a direct product of simply connected semi-simple real or p-adic Lie groups. A discrete subgroup Γ of G is called S-arithmetic if there exist K, S, $\mathscr{H}$ as above and a homomorphism $\alpha\colon H\to G$ with compact kernel such that Γ and $\alpha(\mathscr{H}(\mathfrak{o}))$ are commensurable. All results of [**18**], stated above for ordinary Lie groups and arithmetic groups, are in fact proved by Margulis in the more general framework described here. In particular, he shows that *if* G *is as above, if the* *rank of G (i.e. the sum of the relative ranks of its factors) is at least 2 and if Γ is a discrete subgroup of finite covolume in G, which is irreducible (as defined in n°2), then Γ is S-arithmetic.*

4.2. *"Abstract" isomorphisms.* The very general and powerful superrigidity theorem (in the framework of 4.1) has far-reaching consequences besides the arithmeticity. For instance, it enables Margulis to solve almost completely the problem of "abstract isomorphisms" between groups of points of algebraic simple groups over number fields or arithmetic subrings of such fields; his result embodies, in the arithmetic case, all those obtained before on that problem by Dieudonné, O'Meara and his school, A. Borel and me, and goes considerably further.

4.3. *Normal subgroups.* Let G be as in 4.1 and let Γ be an irreducible discrete subgroup of G of finite covolume. Margulis was able to show (cf. [27]) that *if* $\mathrm{rk}\, G \geqslant 2$, *then every noncentral normal subgroup of Γ has finite index.* (In fact, the conditions of Margulis' theorem are more general: under suitable hypotheses, G is allowed to have factors defined over locally compact local fields of finite characteristic.) So far, the only results known in that direction—results of Mennicke, Bass, Milnor, Serre, Raghunathan—were connected with the congruence subgroup problem and valid only in the cases where that problem has a positive solution.

4.4. *Action on trees.* In a paper which appeared in the Springer Lecture Notes, no 372, Serre showed that the group of integral points of a simple Chevalley group-scheme of rank $\geqslant 2$ cannot act without fixed point on a tree; this also means that such a group is not an amalgam in a nontrivial way. Serre points out that his method of proof does not extend to congruence subgroups and asks whether the result generalizes to such subgroups or to other arithmetic groups. With his own methods, Margulis was able to solve at once the problem in its widest generality: *if G is as in* 4.1, *of rank at least* 2, *and if Γ is an irreducible discrete subgroup of finite covolume in G, then Γ cannot act without fixed point on a tree.*

5. Conclusion. Margulis has completely or almost completely solved a number of important problems in the theory of discrete subgroups of Lie groups, problems whose roots lie deep in the past and whose relevance goes far beyond that theory itself. It is not exaggerated to say that, on several occasions, he has bewildered the experts by solving questions which appeared to be completely out of reach at the time. He managed that through his mastery of a great variety of techniques used with extraordinary resources of skill and ingenuity. The new and most powerful methods he has invented have already had other important applications besides those for which they were created and, considering their generality, I have no doubt that they will have many more in the future.

I wish to conclude this report by a nonmathematical comment. This is probably neither the time nor the place to start a polemic. However, I cannot but express my deep disappointment—no doubt shared by many people here—in the absence of Margulis from this ceremony. In view of the symbolic meaning of this city of Helsinki,[1] I had indeed grounds to hope that I would have a chance at last to meet a mathematician whom I know only through his work and for whom I have the greatest respect and admiration.

References

Published work of G. A. Margulis

1. *Positive harmonic functions on nilpotent groups,* Dokl. Akad. Nauk S.S.S.R. **166** (1966), 1054—1057.

2. (with D. A. Každan) *Proof of a conjecture of Selberg,* Mat. Sb. **75** (1968), 163—168.

3. *Appendix to "Certain smooth ergodic systems" by D. V. Anosov and Ya. G. Sinaĭ,* Uspehi Mat. Nauk **22** (1967), 169—171.

4. *Discrete subgroups of real semi-simple Lie groups,* Mat. Sb. **80** (1960), 600—615.

5. *Certain applications of ergodic theory to the investigation of manifolds of negative curvature,* Funkcional. Anal. i Priložen, **3** (4) (1969), 89—90.

6. *On the problem of arithmeticity of discrete groups,* Dokl. Akad. Nauk S.S.S.R. **187** (1969), 518—520.

7. *Certain measures that are connected with U-flows,* Funkcional. Anal. i Priložen **4** (1) (1970), 62—76.

8. *Isometry of closed manifolds of constant negative curvature having the same fundamental group,* Dokl. Akad. Nauk S.S.S.R. **192** (1970), 736—737.

9. *On the action of unipotent groups in the space of lattices,* Mat. Sb. **86** (1971), 552—556.

10. *Nonuniform lattices in semi-simple algebraic groups, Lie groups and their representations,* Summer school of the Bolyai Janos Math. Soc. (Budapest, 1971), Akademiai Kiado, Budapest, 1975.

11. *Correspondence of boundaries in biholomorphic mappings,* Tezis dokladov Vsesoyuznoi konferencii po teorii funkcii kompleksnogo peremennogo (1971).

12. *Metric questions in the theory of U-systems,* in Devyataya Letnyaya Mat. Škola, Kiev. Inst. Mat. Akad. Nauk S.S.S.R., 1972, pp. 342—348.

13. *Arithmeticity of nonuniform lattices,* Funkcional. Anal. i Priložen **7** (3) (1973), 88—89.

14. *Explicit constructions of dilatators,* Probl. Peredači Informacii **9** (4) (1973), 71—80.

15. *Arithmetic properties of discrete groups,* Uspehi Mat. Nauk S.S.S.R. **29** (1974), 49—98.

16. *Arithmeticity and finite dimensional representations of uniform lattices,* Funkcional. Anal. Priložen **8** (3) (1974), 77—78.

17. *Probabilistic characteristics of graphs with high connectivity,* Probl. Peredači Informacii **10** (2) (1974), 101—108.

18. *Discrete groups of motions of manifolds with nonpositive curvature,* Proc. Internat. Congr. Math. (Vancouver, 1974), vol. 2, 1975, pp. 21—34.

19. *Arithmeticity of nonuniform lattices in weakly noncompact groups,* Funkcional. Anal. Priložen **9** (1) (1975), 35—44.

20. *Arithmeticity of irreducible lattices in semisimple groups of rank greater than* 1, Appendix to the Russian translation of M. Raghunathan's book: Discrete Subgroups of Lie Groups, Moscow, 1977, pp. 277—313.

[1] The address was delivered in Finlandia Hall, where the 1975 Helsinki Agreements were concluded.

21. (with G. A. Soifer), *Criteria for the existence of maximal subgroups of infinite index in finitely generated linear groups*, Dokl. Akad. Nauk S.S.S.R. **234** (1977), 1261—1264.

22. *Cobounded subgroups in algebraic groups over local fields*, Funkcional. Anal. i Priložen **11** (1977), 45—57.

23. *Factor groups of discrete subgroups and measure theory*, Funkcional. Anal. i Priložen **12** (4) (1978), 64—76.

24. *Factor groups of discrete subgroups*, Dokl. Akad. Nauk S.S.S.R. **242** (1978) 533—536.

25. *Finiteness of factor groups of discrete subroups*, Funkcional. Anal. i Priložen (to appear).

Other references

26. A. Borel, *Sous-groupes discrets de groups semi-simples* (d'après D. A. Kajdan et G. A. Margoulis), Sém. Bourbaki, 1969, Lecture Notes in Math., vol. 179, Springer-Verlag, Berlin, pp. 199—215.

27. J. Tits, *Travaux de Margulis sur les sous-groupes discrets de groupes de Lie*, Sém. Bourbaki, 1976, Lecture Notes in Math., vol. 567, Springer-Verlag, Berlin, pp. 174—190.

Collège de France
75231 Paris Cedex 05, France

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Work of Daniel Quillen

I. M. James

D. G. Quillen's contributions to algebra are outstanding in their inventiveness, conceptual richness and technical virtuosity. He is the prime architect of the higher algebraic K-theory, and this is perhaps his finest achievement. But before dealing with this I would like to say a few words about some of his other contributions to mathematics.

I will begin with the cohomology of groups. First consider the mod p cohomology ring of a finite group. It was long conjectured that the dimension of this ring, in the sense of commutative algebra, was the same as the rank of a maximal abelian p-subgroup. Quillen established the conjecture by a method which was very original, elegant and powerful. An essential point was that, to prove a purely algebraic result, he enlarged the context by studying a more general problem which brought in topology by having the group act on a space.

Among the most interesting groups, from the cohomological point of view, are the classical groups over finite fields. These occupy a central position in Quillen's work and he has obtained much important information about them. Moreover he has shown how such purely algebraic information can be used to prove deep results in topology.

An example of this is Quillen's proof of the Adams conjecture in topological K-theory. Suppose we consider real vector bundles over a compact Hausdorff space X. These bundles generate the Grothendieck group $\mathrm{KO}(X)$, and this group admits a family of operators, called the ψ-operations. In homotopy theory it is important to know when a bundle, or rather its associated sphere-bundle, is trivial in the sense of fibre homotopy type. Bundles which are trivial in this sense generate a subgroup of $\mathrm{KO}(X)$. Adams, around 1965, made the bold conjecture that this subgroup could be constructed purely algebraically. Specifically he gave a formula for constructing elements of the subgroup in terms of ψ-operations. Adams verified his conjecture for combinations of line bundles and plane bundles.

Despite much effort it was not until 1970 that the conjecture became a theorem. Quillen's proof introduced completely new methods into the subject, based on Brauer's work in representation theory. He begins with the case where the structural group of the bundle can be reduced from the full general linear group to a finite subgroup. Using the Brauer induction theorem he reduces this case to that of bundles with rank not greater than two, where Adams' methods suffice. Then he shows that the finite case exercises enough control over the general case to solve the problem. This was a real tour de force since it involved the use of the Frobenius map $x \to x^p$ in characteristic p (which corresponds to the operator ψ^p) to solve a difficult problem in topology. Another proof was given by Sullivan about the same time, but it was not until several years later that Becker and Gottlieb found a proof using conventional algebraic topology.

Another of Quillen's contributions to topology relates to the cobordism theory of Thom. Here manifolds are classified modulo boundaries—those manifolds which bound manifolds with boundary. The classification of manifolds in this sense can be reduced to a homotopy-theoretic problem. Earlier methods of solving this problem required difficult calculations but Quillen discovered an entirely new approach, involving formal groups, which bypassed the difficulties in a spectacular manner and gave a beautiful and satisfying conclusion to this important theory.

Now for one of Quillen's most recent achievements. About twenty years ago Serre showed that vector bundles over a space X could be interpreted as projective modules over the ring of continuous realvalued functions on X. Trivial bundles correspond to free modules. On the basis of this connection Serre formulated the plausible conjecture that all projective modules over a polynomial ring are free. The experts in commutative algebra found this an irresistible challenge but despite great efforts progress towards establishing the conjecture was very limited. Then, a couple of years ago, Quillen came up with a simple direct proof of a few pages, disposing of the whole problem. About the same time the Russian mathematician Suslin also found a proof.

I hope I haven't been placing too much emphasis on Quillen's phenomenal successes in solving outstanding problems. These are not isolated achievements but appear rather as spectacular features of the new breed of mathematics he has been creating. It is algebraic K-theory, more than anything else, which occupies the central position in his work.

The development of K-theory first in algebraic geometry and then in topology led many to search for a more general algebraic theory which would apply to any ring and especially to the ring of integers. This would be cohomological in form, with groups in all non-negative dimensions. The right definition in dimension zero was well-known, and various proposals were made for the definition in dimension one. However it was Quillen, in 1972, who found the general definition and thereby opened up this major new area of research, of which the full potential has yet to be realized.

Quillen's definition involves a simple but novel idea, known as the plus construction, which has since found many applications. For the construction one needs a perfect normal subgroup E of the fundamental group $\pi_1(X)$, where X is a complex. By attaching 2-cells and 3-cells to X, in a certain way, Quillen constructs a complex X^+ such that the inclusion $X \to X^+$ induces an isomorphism in homology while killing precisely E in the fundamental group. In the application X is the classifying space $\mathrm{BGL}(A)$ of the stable general linear group $\mathrm{GL}(A)$ of a given ring A. The commutator subgroup $E(A)$ of $\mathrm{GL}(A)$ is known to be perfect. Quillen applies his construction, using the corresponding subgroup of the fundamental group of the classifying space, and obtains a space $\mathrm{BGL}(A)^+$. The higher K-groups of A are defined to be the higher homotopy groups $\pi_i(\mathrm{BGL}(A)^+)$. It required extraordinary insight to realize that this gave the correct formulation.

Quillen's work has had a great influence on the thinking and perceptions of the present generation of topologists and algebraists. His papers are not merely informative, but edifying and exciting to read. They bring into clear view a mathematical landscape of great richness and beauty that many others have vainly striven to approach.

University of Oxford
Oxford OX1 3LB, England

One-hour Plenary Addresses

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Quasiconformal Mappings, Teichmüller Spaces, and Kleinian Groups

Lars V. Ahlfors*

I am extremely grateful to the Committee to select hour speakers for the great honor they have bestowed on me, and above all for this opportunity to address the mathematicians of the whole world from the city of my birth. The city has changed a great deal since my childhood, but I still get a thrill each time I return to this place that holds so many memories for me. I assure you that today is even a more special event for me.

I have interpreted the invitation as a mandate to report on the state of knowledge in the fields most directly dominated by the theory and methods of quasiconformal mappings. I was privileged to speak on the same topic once before, at the Congress in Stockholm 1962, and it has been suggested that I could perhaps limit myself to the developments after that date. But I feel that this talk should be directed to a much wider audience. I shall therefore speak strictly to the non-specialists and let the experts converse among themselves at other occasions.

The whole field has grown so rapidly in the last years that I could not possibly do justice to all recent achievements. A mere list of the results would be very dull and would not convey any sense of perspective. What I shall try to do, in the limited time at my disposal, is to draw your attention to the rather dramatic changes that have taken place in the theory of functions as a direct result of the inception and development of quasiconformal mappings. I should also like to make it clear that I am not reporting on my own work; I have done my share in the early stages, and I shall refer to it only when needed for background.

---

* This work has been supported by the National Science Foundation of the United States under Grant number MCS77 07782

1. Historical remarks. In classical analysis the theory of analytic functions of complex variables, and more particularly functions of one variable, have played a dominant role ever since the middle of the nineteenth century. There was an obvious peak around the turn of the century, centering about names like Poincaré, Klein, Picard, Borel, Hadamard. Another blossoming took place in the 1920s with the arrival of Nevanlinna theory. The next decade seemed at the time as a slackening of the pace, but this was deceptive; many of the ideas that were later to be fruitful were conceived at that time.

The war and the first post-war years were of course periods of stagnation. The first areas of mathematics to pick up momentum after the war were topology and functions of several complex variables. Big strides were taken in these fields, and under the leadership of Henri Cartan, Behnke, and many others, the more-dimensional theory of analytic functions and manifolds acquired an almost entirely new structure affiliated with algebra and topology. As a result of this development the gap between the conservative analysts who were still doing conformal mapping and the more radical ones involved with sheaf-theory became even wider, and for some time it looked as if the one-dimensional theory had lost out and was in danger of becoming a rehash of old ideas. The gap is still there, but I shall try to convince you that in the long run the old-fashioned theory has recovered and is doing quite well.

The theory of quasiconformal mappings is almost exactly fifty years old. They were introduced in 1928 by Herbert Grötzsch in order to formulate and prove a generalization of Picard's theorem. More important is his paper of 1932 in which he discusses the most elementary but at the same time most typical cases of extremal quasiconformal mappings, for instance the most nearly conformal mapping of one doubly connected region on another. Grötzsch's contribution is twofold: (1) to have been the first to introduce non-conformal mappings in a discipline that was so exclusively dominated by analytic functions, (2) to have recognized the importance of measuring the degree of quasiconformality by the maximum of the dilatation rather than by some integral mean (this was recently pointed out by Lipman Bers).

Grötzsch's papers remained practically unnoticed for a long time. In 1935 essentially the same class of mappings was introduced by M. A. Lavrentiev in the Soviet Union whose work was connected more closely with partial differential equations than with function theory proper. In any case, the theory of quasiconformal mappings, which at that time had also acquired its name, slowly gained recognition, originally as a useful and flexible tool, but inevitably also as an interesting piece of mathematics in its own right.

Nevertheless, quasiconformal mappings might have remained a rather obscure and peripheral object of study if it had not been for Oswald Teichmüller, an exceptionally gifted and intense young mathematician and political fanatic, who suddenly made a fascinating and unexpected discovery. At that time, many special extremal problems in quasiconformal mapping had already been solved, but these were isolated results without a connecting general idea. In 1939 he presented to the

Prussian Academy a now famous paper which marks the rebirth of quasiconformal mappings as a new discipline which completely overshadows the rather modest beginnings of the theory. With remarkable intuition he made a synthesis of what was known and proceeded to announce a bold outline of a new program which he presents, rather dramatically, as the result of a sudden revelation that occurred to him at night. His main discovery was that the extremal problem of quasiconformal mapping, when applied to Riemann surfaces, leads automatically to an intimate connection with the holomorphic quadratic differentials on the surface. With this connection the whole theory takes on a completely different complexion: A problem concerned with non-conformal mappings turns out to have a solution which is expressed in terms of holomorphic differentials, so that in reality the problem belongs to classical function theory. Even if some of the proofs were only heuristic, it was clear from the start that this paper would have a tremendous impact, although actually its influence was delayed due to the poor communications during the war. In the same paper Teichmüller lays the foundations for what later has become known as the theory of Teichmüller spaces.

2. Beltrami coefficients. It is time to become more specific, and I shall start by recalling the definition and main properties of quasiconformal (q.c.) mappings. To begin with I shall talk only about the two-dimensional case. There is a corresponding theory in several dimensions, necessarily less developed, but full of interesting problems. One of the reasons for considering q.c. mappings, although not the most compelling one, is precisely that the theory does not fall apart when passing to more than two dimensions. I shall return to this at the end of the talk.

Today it can be assumed that even a non-specialist knows roughly what is meant by a q.c. mapping. Intuitively, a homeomorphism is q.c. if small circles are carried into small ellipses with a bounded ratio of the axes; more precisely, it is K-q.c. if the ratio is $\leqslant K$. For a diffeomorphism f this means that the complex derivatives $f_z=\frac{1}{2}(f_x-if_y)$ and $f_{\bar z}=\frac{1}{2}(f_x+if_y)$ satisfy $|f_{\bar z}|\leqslant k|f_z|$ with $k=(K-1)/(K+1)$.

Already at an early stage it became clear that it would not do to consider only diffeomorphisms, for the class of diffeomorphisms lacks compactness. In the beginning rather arbitrary restrictions were introduced, but in time they narrowed down to two conditions, one geometric and one analytic, which eventually were found to be equivalent. The easiest to formulate is the analytic condition which says that f is K-q.c. if it is a weak L^2-solution of a *Beltrami equation*

$$f_{\bar z} = \mu f_z \tag{1}$$

where $\mu=\mu_f$, known as a Beltrami coefficient, is a complex-valued measurable function with $\|\mu\|_\infty\leqslant k$.

The equation is classical for smooth μ, but there is in fact a remarkably strong existence and uniqueness theorem without additional conditions. If μ is defined in the whole complex plane, with $|\mu|\leqslant k<1$ a.e., then (1) has a homeomorphic solution which maps the plane on itself, and the solution is unique up to conformal mappings. Simple uniform estimates, depending only on k, show that the class of K-q.c. mappings is compact.

It must be clear that I am condensing years of research into minutes. The fact is that the post-Teichmüller era of quasiconformal mappings did not start seriously until 1954. In 1957 I. N. Vekua in the Soviet Union proved the existence and uniqueness theorem for the Beltrami equation, and in the same year L. Bers discovered that the theorem had been proved already in 1938 by C. Morrey. The great difference in language and emphasis had obscured the relevance of Morrey's paper for the theory of q.c. mappings. The simplest version of the proof is due to B. V. Boyarski who made it a fairly straightforward application of the Calderón-Zygmund theory of singular integral transforms.

As a consequence of the chain rule the Beltrami coefficients obey a simple composition law:

$$\mu_{g\circ f^{-1}} = \left[\frac{\mu_g - \mu_f}{1-\bar{\mu}_f \mu_g}\,(f_z/|f_z|)^2\right]\circ f^{-1}.$$

The interesting thing about this formula is that for any fixed z and f the dependence on $\mu_g(z)$ is complex analytic, and a conformal mapping of the unit disk on itself. This simple fact turns out to be crucial for the study of Teichmüller space.

3. Extremal length. The geometric definition is conceptually even more important than the analytic definition. It makes important use of the theory of extremal length, first developed by A. Beurling for conformal mappings. Let me recall this concept very briefly. If L is a set of locally rectifiable arcs in R^2, then a Borel measurable function $\varrho\colon R^2 \to R^+$ is said to be *admissible* for L if $\int_\gamma \varrho\, ds \geqslant 1$ for all $\gamma \in L$. The *module* $M(L)$ is defined as $\inf \int \varrho^2 dx$ for all admissible ϱ; its reciprocal is the *extremal length* of L. It is connected with q.c. mappings in the following way: If f is a K-q.c. mapping (according to the analytic definition), then $M(fL) \leqslant KM(L)$. Conversely, this property may be used as a geometric definition of K-q.c. mappings, and it is sufficient that the inequality hold for a rather restrictive class of families L that can be chosen in various ways. This definition has the advantage of having an obvious generalization to several dimensions.

Inasmuch as extremal length was first introduced for conformal mappings, its connection with q.c. mappings, even in more than two dimensions, is another indication of the close relationship between q.c. mappings and classical function theory.

4. Teichmüller's theorem. The problem of extremal q. c. mappings has dominated the subject from the start. Given a family of homeomorphisms, usually defined by some specific geometric or topological conditions, it is required to find a mapping f in the family such that the maximal dilatation, and hence the norm $\|\mu_f\|_\infty$ is a minimum. Because of compactness the existence is usually no problem, but the solution may or may not be unique, and if it is there remains the problem of describing and analyzing the solution.

It is quite obvious that the notion of q.c. mappings generalizes at once to mappings from one Riemann surface to another, each with its own conformal structure, and that the problem of extremal mapping continues to make sense. The Beltrami coefficient becomes a Beltrami differential $\mu(z)d\bar{z}/dz$ of type $(-1, 1)$. Note that $\mu(z)$ does not depend on the local parameter on the target surface.

Teichmüller considers topological maps $f: S_0 \to S$ from one compact Riemann surface to another. In addition he requires f to belong to a prescribed homotopy class, and he wishes to solve the extremal problem separately for each such class. Teichmüller asserted that there is always an extremal mapping, and that it is unique. Moreover, either there is a unique conformal mapping in the given homotopy class, or there is a constant $k, 0<k<1$, and a *holomorphic quadratic differential* $\varphi(z)dz^2$ on S_0 such that the Beltrami coefficient of the extremal mapping is $\mu_f=k\bar{\varphi}/|\varphi|$. It is thus a mapping with constant dilatation $K=(1+k)/(1-k)$. The inverse f^{-1} is simultaneously extremal for the mappings $S \to S_0$, and it determines an associated quadratic differential $\psi(w)dw^2$ on S. In local coordinates the mapping can be expressed through

$$\sqrt{\psi(w)}\,dw = \sqrt{\varphi(z)}\,dz + k\sqrt{\bar{\varphi}}(z)\,d\bar{z}.$$

Naturally, there are singularities at the zeros of φ, which are mapped on zeros of ψ of the same order, but these singularities are of a simple explicit nature. The integral curves along which $\sqrt{\varphi}\,dz$ is respectively real or purely imaginary are called horizontal and vertical trajectories, and the extremal mapping maps the horizontal and vertical trajectories on S_0 on corresponding trajectories on S. At each point the stretching is maximal in the direction of the horizontal trajectory and minimal along the vertical trajectory.

This is a beautiful and absolutely fundamental result which, as I have already tried to emphasize, throws a completely new light on the theory of q.c. mappings. In his 1939 paper Teichmüller gives a complete proof of the uniqueness part of his theorem, and it is still essentially the only known proof. His existence proof, which appeared later, is not so transparent, but it was put in good shape by Bers; the result itself was never in doubt. Today, the existence can be proved more quickly than the uniqueness, thanks to a fruitful idea of Hamilton. Unfortunately, time does not permit me to indicate how and why these proofs work, except for saying that the proofs are variational and make strong use of the chain rule for Beltrami coefficients.

5. Teichmüller spaces. Teichmüller goes on to consider the slightly more general case of compact surfaces with a finite number of punctures. Specifically, we say that S is of finite type (p, m) if it is an oriented topological surface of genus p with m points removed. It becomes a Riemann surface by giving it a conformal structure. Following Bers we shall define a conformal structure as a sense-preserving topological mapping σ on a Riemann surface. Two conformal structures σ_1 and σ_2 are equivalent if there is a conformal mapping g of $\sigma_1(S)$ on $\sigma_2(S)$ such that $\sigma_2^{-1} \circ g \circ \sigma_1$ is homotopic to the identity. The equivalence classes $[\sigma]$ are the points of the Teichmüller space $T(p, m)$, and the distance between $[\sigma_1]$ and $[\sigma_2]$ is defined to be

$$d([\sigma_1], [\sigma_2]) = \log \inf K(f)$$

where $K(f)$ is the maximal dilatation of f, and f ranges over all mappings homotopic to $\sigma_2 \circ \sigma_1^{-1}$. It is readily seen that the infimum is actually a minimum, and that the extremal mapping from $\sigma_1(S)$ to $\sigma_2(S)$ is as previously described, except that the quadratic differentials are now allowed to have simple poles at the punctures.

With this metric $T(p, m)$ is a complete metric space, and already Teichmüller showed that it is homeomorphic to $R^{6p-6+2m}$ (provided that $2p-2+m>0$).

Let f be a self-mapping of S. It defines an isometry $\tilde{f}$ of $T(p, m)$ which takes $[\sigma]$ to $[\sigma \circ f]$. This isometry depends only on the homotopy class of f and is regarded as an element of the *modular group* Mod (p, m). It follows from the definition that two Riemann surfaces $\sigma_1(S)$ and $\sigma_2(S)$ are conformally equivalent if and only if $[\sigma_2]$ is the image of $[\sigma_1]$ under an element of the modular group. The quotient space $T(p, m)/\text{Mod}\,(p, m)$ is the Riemann space of algebraic curves or moduli. The Riemann surfaces that allow conformal self-mappings are branch-points of the covering.

6. Fuchsian and quasifuchsian groups. The universal covering of any Riemann surface S, with a few obvious exceptions, is conformally equivalent to the unit disk U. The self-mappings of the covering surface correspond to a group G of fractional linear transformations, also referred to as Möbius transformations, which map U conformally on itself. More generally one can allow coverings with a signature, that is to say regular covering surfaces which are branched to a prescribed order over certain isolated points. In this case G includes elliptic transformations of finite order. It is always discrete.

Any discrete group of Möbius transformations that preserves a disk or a half-plane, for instance U, is called a Fuchsian group. It is a recent theorem, due to Jørgensen, that a nonelementary group which maps U on itself is discrete, and hence Fuchsian, if and only if every elliptic transformation in the group is of finite order. As soon as this condition is fulfilled the quotient U/G is a Riemann surface S, and U appears as a covering of S with a signature determined by the orders of the elliptic transformations. The group acts simultaneously on the exterior U^* of U, and $S^* = U^*/G$ is a mirror image of S. G is determined by S up to conjugation.

A point is a limit point if it is an accumulation point of an orbit. For Fuchsian groups all limit points are on the unit circle; the set of limit points will be referred to as the *limit set* $\Lambda(G)$. Except for some trivial cases there are only two alternatives: either Λ is the whole unit circle, or it is a perfect nowhere dense subset. With an unimaginative, but classical, terminology Fuchsian groups are accordingly classified as being of the first kind or second kind.

If S is of finite type, then G is always of the first kind; what is more, G has a fundamental region with finite noneuclidean area. Consider a q.c. mapping $f\colon S_0 \to S$ with corresponding groups G_0 and G. Then f lifts to a mapping $f\colon U \to U$ (which we continue to denote by the same letter), and if $g_0 \in G_0$ there is a $g \in G$ such that $f \circ g_0 = g \circ f$. This defines an isomorphism $\theta\colon G_0 \to G$ which is uniquely determined, up to conjugation, by the homotopy class of f. Moreover, f extends to a homeomorphism of the closed disks, and the boundary correspondence is again determined uniquely up to normalization. The Teichmüller problem becomes that of finding f with given boundary correspondence and smallest maximal dilatation. The extremal mapping has a Beltrami coefficient $\mu = k\bar{\varphi}/|\varphi|$ where φ is an invariant quadratic differential with respect to G_0.

Incidentally, the problem of extremal q.c. mappings with given boundary values makes sense even when there is no group, but the solution need not be unique. The questions that arise in this connection have been very successfully treated by Hamilton, K. Strebel, and E. Reich.

For a more general situation, let $\mu d\bar{z}/dz$ be any Beltrami differential, defined in the whole plane and invariant under G_0 in the sense that $(\mu \circ g_0)\bar{g}_0'/g_0' = \mu$ a.e. for all $g_0 \in G$. Suppose f is a solution of the Beltrami equation $f_{\bar{z}} = \mu f_z$. It follows from the chain rule that $f \circ g_0$ is another solution of the same equation. Therefore $f \circ g_0 \circ f^{-1}$ is conformal everywhere, and hence a Möbius transformation g. In this way μ determines an isomorphic mapping of G_0 on another group G, but this time G will in general not leave U invariant. For this reason G is a Kleinian group rather than a Fuchsian group. It has two invariant regions $f(U)$ and $f(U^*)$, separated by a Jordan curve $f(\delta U)$. The surfaces $f(U)/G$ and $f(U^*)/G$ are in general not conformal mirror images.

The group $G = fG_0f^{-1}$ is said to be obtained from G_0 by q.c. deformation, and it is called a quasifuchsian group. Evidently, quasifuchsian groups have much the same structure as fuchsian groups, except for the lack of symmetry. The curve that separates the invariant Jordan regions is the image of the unit circle under a q.c. homeomorphism of the whole plane. Such curves are called quasicircles. It follows by a well-known property of q.c. mappings that every quasicircle has zero area, and consequently the limit set $\Lambda(G)$ has zero two-dimensional measure.

Strangely enough, quasicircles have a very simple geometric characterization: A Jordan curve is a quasicircle if and only if for any two points on the curve at least one of the subarcs between them has a diameter at most equal to a fixed multiple of the distance between the points. It means, among other things, that there are no cusps.

7. The Bers representation. There are two special cases of the construction that I have described: (1) If μ satisfies the symmetry condition $\mu(1/\bar{z})\bar{z}^2/z^2 = \bar{\mu}(z)$, then G is again a Fuchsian group and f preserves symmetry with respect to the unit circle. (2) If μ is identically zero in U and arbitrary in U^*, except for being invariant with respect to G_0, then f is conformal in U, and $f(U)/G$ is conformally equivalent to $S = U/G$, while $f(U^*)/G$ is a q.c. mirror image of S.

I shall refer to the second construction as the Bers mapping. Two Beltrami differentials μ_1 and μ_2 will lead to the same group G and to homotopic maps f_1, f_2 if and only if $f_1=f_2$ on ∂U (up to normalizations). When that is the case we say that μ_1 and μ_2 are equivalent, and that they represent the same point in the Teichmüller space $T(G_0)$ based on the Fuchsian group G_0.

In other words the equivalence classes are determined by the values of f on the unit circle. These values obviously determine $f(U)$, and hence f, at least up to a normalization. One obtains strict uniqueness by passing to the Schwarzian derivative $\varphi=S_f$ defined in U (recall that $S_f=(f''/f')'-\frac{1}{2}(f''/f')^2$). From the properties of the Schwarzian it follows that $\varphi(gz)g'(z)^2=\varphi(z)$ for all $g\in G_0$. Furthermore, by a theorem of Nehari $|\varphi(z)|(1-|z|^2)^2$ is bounded (actually $\leqslant 6$). Thus φ belongs to the Bers class $B(G_0)$ of bounded quadratic differentials with respect to the group G_0. The Bers map is an injection $T(G_0)\to B(G_0)$.

It is known that the image of $T(G_0)$ under the Bers map is *open*, and as a vector space $B(G_0)$ has a natural complex structure. The mapping identifies $T(G_0)$ with a certain open subset of $B(G_0)$ which in turn endows $T(G_0)$ with its own complex structure. If S is of type (p, m) the complex dimension is $3p-3+m$. The nature of the subset that represents $T(p, m)$ in C^{3p-3+m} is not well known. For instancel it seems to be an open problem whether $T(1,1)$ is a Jordan region in C.

The case where $G=I$, the identity group, is of special interest because it is so closely connected with classical problems in function theory. An analytic function φ, defined on U, will belong to $T(I)$ if and only if it is the Schwarzian S_f of a schlicht (injective) function on U with a q.c. extension to the whole plane. The study of such functions has added new interest to the classical problems of schlicht functions.

To illustrate the point I would like to take a minute to tell about a recent beautiful result due to F. Gehring. Let $\boldsymbol{S}$ denote the space of all $\varphi=Sf$, f analytic and schlicht in U, with the norm $\|\varphi\|=\sup(1-|z|^2)^2|\varphi(z)|$, and let $\boldsymbol{T}=T(I)$ be the subset for which f has a q.c. extension. Gehring has shown (i) that $\boldsymbol{T}=\operatorname{Int}\boldsymbol{S}$, (ii) the closure of $\boldsymbol{T}$ is a proper subset of $\boldsymbol{S}$. To prove the second point, which gives a negative answer to a question raised by Bers maybe a dozen years ago, he constructs, quite explicitly, a region with the property that no small deformation, measured by the norm of the Schwarzian, changes it to a Jordan region, much less to one whose boundary is a quasicircle. I mention this particular result because it is recent and because it is typical for the way q.c. mappings are giving new impulses to the classical theory of conformal mappings.

In the finite dimensional case $T(p, m)$ has a compact boundary in $B(G_0)$. It is an interesting and difficult problem to find out what exactly happens when φ approaches the boundary. The pioneering research was carried out by Bers and Maskit. They showed, first of all, that when φ approaches a boundary point the holomorphic function f will tend to a limit which is still schlicht, and the groups G tend to a limit group which is Kleinian with a single, simply connected invariant region. Such groups were called B-groups (B stands either for Bers or for boundary) in the belief that any such group can be obtained in this manner. It can happen that the invariant simply connected region is the whole set of discontinuity; such groups are said to be *degenerate*. Classically, degenerate groups were not known, but Bers proved that they must exist, and more recently Jørgensen has been able to construct many explicit examples of such groups.

Intuitively, it is clear what should happen when φ goes to the boundary. We are interested to follow the q.c. images $f(U^*)$. In the degenerate case the image disappears completely. In the nondegenerate case the fact that one approaches the boundary must be visible in some way, and the obvious guess is that one or more of the closed geodesics on the surface is being pinched to a point. In the limit $f(S^*)$ would either be of lower genus or would disintegrate to several pieces, and one would end up with a more general configuration consisting of a "surface with nodes", each pinching giving rise to two nodes.

A lot of research has been going on with the intent of making all this completely rigorous, and if I am correctly informed these attempts have been successful, but much remains to be done. This is the general trend of much of the recent investigations of Bers, Maskit, Kra, Marden, Earle, Jørgensen, Abikoff and others; I hope they will understand that I cannot report in any detail on these theories which are still in *status nascendi*.

In a slightly different direction the theory of Teichmüller spaces has been extended to a study of the so-called universal Teichmüller curve, which for every type (p, m) is a fiber-space whose fibers are the Riemann surfaces of that type. A special problem is the existence, or rather non-existence, of holomorphic sections.

The Bers mapping is not concerned with extremal q.c. mappings, and it is rather curious that one again ends up with holomorphic quadratic differentials. The Bers model has a Kählerian structure obtained from an invariant metric, the Petersson-Weil metric, on the space of quadratic differentials. The relation between the Petersson-Weil metric and the Teichmüller metric has not been fully explored and is still rather mystifying.

8. Kleinian groups. I would have preferred to speak about Kleinian groups in a section all by itself, but they are so intimately tied up with Teichmüller spaces that I was forced to introduce Kleinian groups somewhat prematurely. I shall now go back and clear up some of the terminology.

It was Poincaré who made the distinction between Fuchsian and Kleinian groups and who also coined the names, much to the displeasure of Klein. He also pointed out that the action of any Möbius transformation extends to the upper half space, or, equivalently, to the unit ball in three-space. Any discrete group of Möbius transformations is discontinuous on the open ball. Limit points are defined as in the Fuchsian case; they are all on the unit sphere, and the limit set Λ may be regarded either as a set on the Riemann sphere or in the complex plane. The elementary groups with at most two limit points are usually excluded, and in modern terminology a Kleinian group is one whose limit set is nowhere dense and perfect. A Kleinian group may be looked upon as a Fuchsian group of the second kind in three dimensions. As such it cannot have a fundamental set with finite non-euclidean volume. Therefore, the relatively well developed methods of Lie group theory which require finite Haar measure are mostly not available for Kleinian groups. However, the important method of Poincaré series continues to make sense.

Let G be a Kleinian group, Λ its limit set, and Ω the set of discontinuity, that is to say the complement of Λ in the plane or on the sphere. The quotient manifold Ω/G inherits the complex structure of the plane and is thus a disjoint union of Riemann surfaces. It forms the boundary of a three-dimensional manifold $M(G)=$ $=B(1)\cup\Omega/G$.

What is the role of q.c. mappings for Kleinian groups? For one thing one would like to classify all Kleinian groups. It is evident that two groups that are conjugate to each other in the full group of Möbius transformations should be regarded as essentially the same. But as in the case of quasifuchsian groups two groups can also be conjugate in the sense of q.c. mappings, namely if $G'=fGf^{-1}$ for some q.c. mapping of the sphere. In that case G' is a q.c. deformation of G, and such groups should be in the same class.

But this is not enough to explain the sudden blossoming of the theory under the influence of q.c. mappings. As usual, linearization pays off, and it has turned out that infinitesimal q.c. mappings are relatively easy to handle. An infinitesimal q.c. mapping is a solution of $f_{\bar{z}}=\nu$ where the right-hand member is a function of class L^∞. This is a non-homogeneous Cauchy-Riemann equation, and it can be solved quite explicitly by the Pompeiu formula, which is nothing else than a generalized Cauchy integral formula. In order that f induce a deformation of the group ν must be a Beltrami differential, $\nu\in\operatorname{Bel} G$, this time with arbitrary finite bound. There is a subclass N of trivial differentials that induce only a conformal conjugation of G, and the main theorem asserts that the dual space of $\operatorname{Bel} G/N$ can be identified with the space of quadratic differentials on $\Omega(G)/G$ which are of class L^1.

This technique is particularly successful if one looks only at finitely generated groups. In that case the deformation space is finite dimensional, so that there are only a finite number of linearly independent integrable quadratic differentials. This result led me to announce, somewhat prematurely, the so-called *finiteness theorem:* If G is finitely generated, then $S=\Omega(G)/G$ is a finite union of Riemann surfaces of finite type. I had overlooked the fact that a triply punctured square carries no quadratic differentials. Fortunately, the gap was later filled by L. Greenberg, and again by L. Bers who extended the original method to include differentials of higher order. With this method Bers obtained not only an upper bound for the number of surfaces in terms of the number of generators, but even a bound on the total Poincaré area of S.

It was not unreasonable to expect that finitely generated Kleinian groups would have other simple properties. For instance, since a finitely generated Fuchsian group has a fundamental polygon with a finite number of sides one could hope that every finitely generated Kleinian group would have a finite fundamental polyhedron. All such hopes were shattered when L. Greenberg proved that a degenerate group in the sense of Bers and Maskit can never have a finite fundamental polyhedron. Groups with a finite fundamental polyhedron are called *geometrically finite*, and it has been suggested that one should perhaps be content to study only geometrically finite groups. With his constructive methods that go back to Klein, Maskit has been able to give a complete classification of all geometrically finite groups, and Marden has used three-dimensional topology to study the geometry of the three-manifold. These are very farreaching and complicated results, and it would be impossible for me to try to summarize them even if I had the competence to do so.

9. The zero area problem. An interesting problem that remains unsolved is the following: Is it true that every finitely generated Kleinian group has a limit set with twodimensional measure zero?

The most immediate reason for raising the question is that it is easy to prove the corresponding property for Fuchsian groups of the second kind, two-dimensional measure being replaced by one-dimensional. How does one prove it? If the limit set of a Fuchsian group has positive measure one can use the Poisson integral to construct a harmonic function on the unit disk with boundary values 1 a.e. on the limit set and 0 elsewhere. If the group is finitely generated the surface must have a finitely generated fundamental group, and it is therefore of finite genus and connectivity. The ideal boundary components are then representable as points or curves. If they are all points the group would be of the first kind, and if there is at least one curve the existence of a nonconstant harmonic function which is zero on the boundary violates the maxium principle. Therefore the limit set must have zero linear measure. The proof is thus quite trivial, but it is trivial only because one has a complete classification of surfaces with finitely generated fundamental group.

For Kleinian groups it is easy enough to imitate the construction of the harmonic function, which this time has to be harmonic with respect to the hyperbolic metric of the unit ball. If the group is geometrically finite this leads rather easily to a proof of measure zero. For the general case it seems that one would need a better topological classification of three-manifolds with constant negative curvature. It is therefore not suprising that the problem has come to the attention of the topologists, and I am happy to report that at least two leading topologists are actively engaged in research on this problem. I believe that this pooling of resources will be very fruitful, and it would of course not be the first time that analysis inspires topology, and vice versa.

Some time ago W. Thurston became interested in a topological problem concerning foliations of surfaces, and he proved a theorem which is closely related to Teichmüller theory. I have not seen Thurston's work, but I have seen Bers' interpretation of it as a new extremal problem for self-mappings of a surface. It is fascinating, and I could and perhaps should have talked about it in connection with the Teichmüller extremal problem, but I am a little hesitant to speak about things that are not yet in print, and therefore not quite in the public domain. Nevertheless, since many exciting things have happened quite recently in this particular subject, I am taking upon myself to report very informally on some of the newest developments, including some where I have to rely on faith rather than proofs.

Thurston has now begun to apply his remarkable geometric and topological intuition and skill to the problem of zero measure. I certainly do not want to preempt him in case he is planning to talk about it in his own lecture, and I have seen only glimpses of his reasoning, but it would seem that he can prove zero area for all groups that are limits, in one sense or another, of geometrically finite groups. This would be highly significant, for it would show that all groups on the boundary of Teichmüller space have limit sets with zero measure. It would neither prove nor disprove the original conjecture, but it would be a very big step. Personally, I feel that a definitive solution is almost imminent.

Very recently there was a highly specialized conference on Riemann surfaces in the United States, and there was an air of excitement caused not only by what Thurston had done and was doing, but also by the presence of D. Sullivan who had equally fascinating stories to tell. Sullivan, too, has worked hard on the area problem, and he has come up with a by-product that does not solve the problem, but is extremely interesting in itself. He applies the powerful tool of what has been called topological dynamics. If a transformation group acts on a measure space, the space splits into two parts, a dissipative part with a measurable fundamental set, and a recurrent part whose every measurable subset meets infinitely many of its images in a set of positive measure. This powerful theorem, which goes back to E. Hopf, does not seem to have been familiar to those who have approached Kleinian group from the point of view of q.c. mappings. The dissipative part of a finitely generated group is the set of discontinuity, and nothing more; this is a known theorem. The recurrent part is the limit set, and it is of interest only if it has positive measure. But even if the area conjecture is true Sullivan's work remains significant for groups whose limit set is the whole sphere.

Sullivan has several theorems, but the one that has captured my special interest because I understand it best asserts that there is no invariant vector field supported on the limit set. If the limit set is the whole sphere there is no invariant vector field, period. In an equivalent formulation, the limit set carries no Beltrami differential. It was known before that there are only a finite number of linearly independent Beltrami differentials on the limit set of a finitely generated Kleinian group, but that there are none was a surprise to me, and Sullivan's approach gives results even for groups that are not finitely generated. Sullivan's results, taken as a whole, give a new outlook on the ergodic theory of Kleinian groups. They are related to, but go beyond the results of E. Hopf which were already considered deep and difficult, and as a corollary Sullivan obtains a strengthening of Mostow's rigidity theorem. I cannot explain the proofs beyond saying that they are very clever and show that Sullivan is not only a leading topologist, but also a strong analyst.

10. Several dimensions. In the remaining time I shall speak briefly about the generalizations to more than two dimensions. There are two aspects: q.c. mappings *per se,* and Kleinian groups in several dimensions.

The foundations for q.c. mappings in space are essentially due to Gehring and J. Väisälä, but very important work has also been done in the Soviet Union and Roumania. I have already mentioned, in passing, that correct definitions can be based on modules of curve families, and the modules give the only known workable technique. Otherwise, the difficulties are enormous. It is reasonably clear that the Beltrami coefficient should be replaced by a matrix-valued function, but this function is subject to conditions that were known already to H. Weyl, but which are so complicated that nobody has been able to put them to any use. Very little is known about when a region in n-space is q.c. equivalent to a ball, and there is not even an educated guess what Teichmüller's theorem should be replaced by. On the positive side one knows a little bit about boundary correspondence.

In two dimensions there is not much use for mappings that are locally q.c. but not homeomorphic, for by passing to Riemann surfaces they can be replaced by homeomorphisms. In several dimensions the situation is quite different, and there has been rapid growth of the theory of so-called quasiregular mappings from one n-dimensional space to another. It has been developed mostly in the Soviet Union and Finland, and this is perhaps a good opportunity to congratulate the young Finnish mathematicians to their success in this area. In the spirit of Rolf Nevanlinna they have even been able to carry over parts of the value distribution theory to quasiregular functions. In fact, less than a month ago I learned that Rickman has succeeded in proving a generalization of Picard's theorem that I know they have been looking for for a long time. It is so simple that I cannot resist quoting the result: There exists $q=q(n, K)$ such that any K-q.c. mapping $f\colon R^n \to R^n - \{a_1, \ldots, a_q\}$ is constant. (They believe that the theorem is true with $q=2$.)

As for Kleinian groups, they generalize trivially to any number of dimensions, and the distinction between Fuchsian and Kleinian groups disappears. Some properties that depend purely on hyperbolic geometry will carry over, but they are not the ones that use q.c. mappings. However, infinitesimal q.c. mappings have an interesting counterpart for several variables. There is a linear differential operator that takes the place of $f_{\bar{z}}$, namely $Sf = \frac{1}{2}(Df + Df') - (1/n) \operatorname{tr} Df \cdot 1_n$ which is a symmetric matrix with zero trace. It has the right invariance, and the conditions under which the Beltrami equation $Sf = \nu$ has a solution can be expressed as a linear integral equation. The formal theory is there, but it will take time before it leads to tangible results.

My survey ends here. I regret that there are so many topics that I could not even mention, and that my report has been so conspicuously insufficient as far as research in the Soviet Union is concerned. I know that I have not given a full picture, but I hope that I have given you an idea of the extent to which q.c. mappings have penetrated function theory.

References

The following surveys have been extremely helpful to the author:

Lipman Bers, *Quasiconformal mappings, with applications to differential equations, function theory and topology*, Bull. Amer. Math. Soc. (6) **83** (1977), 1083—1100.

L. Bers and I. Kra (Editors), *A crash course on Kleinian groups*, Lecture Notes in Math., vol. 400, Springer-Verlag, Berlin and New York, 1974.

W. J. Harvey (Editor), *Discrete groups and automorphic functions*, Proc. an Instructional Conf. organized by the London Math. Soc. and the Univ. of Cambridge, Academic Press, London, New York, San Francisco, 1977.

Harvard University
Cambridge, Massachusetts 02138, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Commutators, Singular Integrals on Lipschitz Curves and Applications

A. P. Calderón

The topics discussed in this lecture had their origin in the theory of linear partial differential equations. In order to explain how the problem of the so-called commutators and that of the Cauchy integral on Lipschitz curves arose, I will recall and analyze some of the modern methods employed in the theory of linear partial differential equations, and in particular that of the pseudodifferential operators which became widely used in the last decade.

Let us consider the basic idea of the method of pseudo-differential operators. Every linear partial differential operator is a sum of monomial operators

$$a(x)\left(\frac{\partial}{\partial x}\right)^{\alpha}, \tag{1}$$

and the operator $(\partial/\partial x)^{\alpha}$ applied to the function $f(x)$ can be thought of as multiplication of the Fourier transform $\hat{f}(\xi)$ of f by the function $(-i\xi)^{\alpha}$, that is,

$$a(x)(\partial/\partial x)^{\alpha} = a(x)K_{\alpha}, \quad \text{where} \quad (K_{\alpha}f)\hat{} = \hat{f}(\xi)(-i\xi)^{\alpha}.$$

Consequently, a linear differential operator L can be expressed as

$$\begin{gathered} Lf = \sum a_{\alpha}(x)K_{\alpha}f, \ (K_{\alpha}f)\hat{} = \hat{f}(\xi)(-i\xi)^{\alpha}, \quad \text{or} \\ Lf = \frac{1}{(2\pi)^{n}}\int \sum a_{\alpha}(x)(-i\xi)^{\alpha}e^{-ix\cdot\xi}\hat{f}(\xi)\,d\xi. \end{gathered} \tag{2}$$

Now, pseudo-differential operators are obtained by replacing the function

$$\sum a_{\alpha}(x)(-i\xi)^{\alpha} \tag{3}$$

in the preceding expression, which in the case of differential operators is a polynomial in ξ, by more general functions $p(x, \xi)$ in such a way that the resulting class of operators be closed under composition, adjunction, inversion if possible, etc. One should observe here that if the class is to be closed under composition, differential operators contained in it should be freely composable. As is well known, a differential operator can be freely composed with itself only if its coefficients are infinitely differentiable. Thus, classes of pseudo-differential operators which are closed under composition cannot possibly contain differential operators with non-smooth coefficients.

Another method, which preceded chronologically the one above and avoids this obstacle, is that of the singular integral operators. It consists in writing the polynomial in (3) as

$$\frac{1}{(2\pi)^n} \sum a_\alpha(x)(-i\xi)^\alpha = [q(x, \xi)+r(x, \xi)]\varphi(\xi)^m$$

where m is the degree of the polynomial, $\varphi(\xi)$ is a positive infinitely differentiable function such that $\varphi(\xi)=|\xi|$ if $|\xi|\geqslant 1$, and

$$q(x, \xi) = |\xi|^{-m} \sum_{|\alpha|=m} a_\alpha(x)(-i\xi)^\alpha.$$

Then if

$$Kf = \int q(x, \xi)e^{-ix\cdot\xi}\hat{f}(\xi)\,d\xi+Sf, \tag{4}$$
$$Sf = \int r(x, \xi)e^{-ix\cdot\xi}\hat{f}(\xi)\,d\xi,$$

we have

$$Lf = K\Lambda^m f, \quad \text{where} \quad (\Lambda f)\hat{} = \varphi(\xi)\hat{f}(\xi). \tag{5}$$

The function $q(x, \xi)$ is homogeneous of degree zero in ξ, and, as is readily verified, the operators S and $S(\partial/\partial x)$ are bounded in L^2, or more generally in L^p, $1<p<\infty$, provided the coefficients $a_\alpha(x)$ are bounded. Now the operators K are generalized in the following manner: one replaces $q(x, \xi)$ by a function which is homogeneous of degree zero in ξ and bounded but otherwise arbitrary, and S by any operator with the properties described above. Evidently, $q(x, \xi)$ cannot be assumed to be more regular, as a function of x, than the coefficients of the differential operators we want to be included in the theory. On the other hand, if one considers general differential operators whose coefficients have a certain degree of regularity, it seems reasonable to exclude those whose terms of highest order have coefficients not satisfying at least a Lipschitz condition. This becomes clear if one considers the case of first order operators. If one allows the coefficients not to satisfy a Lipschitz condition there can arise pathologies such as the nonuniqueness of trajectories of the associated vector fields. This suggests restricting the generalization to the operators K in (4) to those for which the function $q(x, \xi)$ is bounded homogeneous of degree zero and regular in ξ and Lipschitzian in x. Then every differential operator whose coefficients are bounded and Lipschitzian for the highest order terms and merely bounded for the remaining ones can be represented as in (5) with such a $q(x, \xi)$. However, in order that this description be useful the operators

K in (4) thus generalized should form an algebra under composition. This is indeed the case, and this algebra becomes an instrument which allows us to manipulate effectively general linear differential operators and obtain for them results on existence, uniqueness, a priori estimates, etc. Even in the case of operators with smooth coefficients this allows us to obtain estimates which depend only on the bounds of the coefficients and the bounds of the first order derivatives of the coefficients of highest order terms. But let us return to the generalized operators K as in (4). The problem of showing that the composition of two such operators is one of the same kind can be reduced without great difficulty to the following problem: let A be, in the case of one variable, the operator multiplication by the bounded Lipschitzian function $a(x)$ and Hf the Hilbert transform of f. As is well known, this transform can be expressed as follows

$$(Hf)(x)=\frac{i}{2}\int_{-\infty}^{+\infty} sg\xi e^{-ik\xi}\hat{f}(\xi)\,d\xi$$

and this makes it clear that A, H and AH are operators of the type of the generalized K, and the simplest of their kind. In order to show that HA is of the same type, since

$$HA = AH+(HA-AH),$$

it would suffice to show that $(HA-AH)D$, $D=d/dx$, is bounded in L^p, $1<p<\infty$. This was done in 1965 in [4] with the aid of the theory of analytic functions and a result closely related to an old conjecture of Littlewood. If we denote now by $C_a(K)$ the commutator of K and A, that is $AK-KA$, then

$$(AH-HA)D = C_a(H)D = C_a(HD)-HC_a(D)$$

and since the operator $C_a(D)$ is multiplication by $a'(x)$, which is a bounded function if $a(x)$ is Lipschitzian, $HC_a(D)$ is bounded in L^p and the continuity of $C_a(H)D$ is equivalent to that of $C_a(HD)$. Now, it is easy to see that

$$C_a(HD)f=\text{p.v.}\int_{-\infty}^{+\infty}\frac{(-1)}{x-y}\left[\frac{a(x)-a(y)}{x-y}\right]f(y)\,dy. \tag{6}$$

The integral on the right, which in the case $a(x)=x$ reduces to the Hilbert transform, is the one studied in [4] and is the so-called first commutator. Thus, its role in the theory of partial differential equations becomes apparent.

Next let us consider some generalizations of (6) whose interest we will explain later. The first one

$$C_a^m(HD^m)f=\text{p.v.}\int_{-\infty}^{+\infty}\frac{(-1)^m m!}{x-y}\left[\frac{a(x)-a(y)}{x-y}\right]^m f(y)\,dy \tag{7}$$

is the so-called mth commutator. This equality is not evident but also not difficult to prove. Aside from the intrinsic interest of the left-hand side of (7) and the analogy of the right-hand sides of (6) and (7), the integral in (7) is a special case of

$$\text{p.v.} \int_{-\infty}^{+\infty} \frac{1}{x-y} F\left[\frac{a(x)-a(y)}{x-y}\right] f(y)\, dy \tag{8}$$

where F is an analytic function of its argument. Several classical integrals are special cases of (8). Let Γ be the graph of the real valued function $a(x)$, $x \in \boldsymbol{R}$, that is, the range of the function $x+ia(x)$ in the complex plane, and let us regard the function $f(x)$ as a function on Γ. Consider now the Cauchy integral of this function on Γ

$$G(z) = \frac{1}{2\pi i} \int_{-\infty}^{+\infty} \frac{f(y)[1+ia'(y)]}{z-[y+ia(y)]}\, dy.$$

Then the limit of $G(z)$ when z approaches $x+ia(x)$ from above Γ and non-tangentially, if it exists, is given by

$$\frac{1}{2} f(x) + \frac{1}{2\pi i}\, \text{p.v.} \int_{-\infty}^{+\infty} \frac{1}{x-y} \left[1+i\,\frac{a(x)-a(y)}{x-y}\right]^{-1} f(y)[1+ia'(y)]\, dy, \tag{9}$$

and this integral, except for the presence of the factor $[1+ia'(y)]$ which can be incorporated in the function $f(y)$, is of the form (8). Thus, the study of the behaviour at the boundary of analytic functions given by integrals of the Cauchy type reduces to the study of an integral of the type (8).

On the other hand, let us consider the derivative at the point $x+ia(x)$ and in the direction of the vector $(a'(x), -1)$ of the logarithmic potential of the mass distribution $f(y)\, dy$ on Γ. This derivative, if it exists, is given by the following expression

$$\pm \frac{1}{2} f(x) + \frac{1}{2\pi}\, \text{p.v.} \int_{-\infty}^{+\infty} \frac{(x-y)a'(x)-a(x)+a(y)}{(x-y)^2+[a(x)-a(y)]^2} f(y)\, dy \tag{10}$$

which is also essentially of the form (8). Similarly, the value on Γ of the potential of a double layer distributed on Γ is given by the transpose of the preceding expression. As is well known, these potentials are used to obtain and represent solutions of boundary value problems for the Laplace equation such as the Dirichlet problem, the Neumann problem, etc. The Neumann problem, for example, reduces to the integral equation obtained by equating (10) to the given function on the boundary Γ. The applicability of this method depends on Γ being such that the resulting integrals have reasonable continuity properties. However, it depends less on the specific type of boundary value problem or equation under consideration than the methods using potential theory which are inapplicable to the Neumann problem, for example. For this reason it is natural to expect that the study of the integral (8) and its generalizations will yield an effective tool for the treatment of boundary value problems for elliptic equations in domains with nonregular boundaries. In fact, some interesting results have already been obtained.

Having justified the interest of the integral (8), let us see what can be said about it. In the first place we observe that if one develops in power series the function F in (8), that integral appears as a series of integrals as in (7). Thus, it would suffice in principle to study these, which are apparently simpler. As was mentioned before, the case $m=1$ of (7) could be treated by means of a technique based on the use of analytic functions. Unfortunately, this method fails utterly if $m \geqslant 2$, and this case resisted all efforts to extend to it the results known for $m=1$ until 1975, when R. Coifman and Y. Meyer [11] settled the case $m=2$ with an entirely different approach. They succeeded by using simultaneously the Fourier and the Mellin transforms and certain real variable methods of M. Cotlar and C. P. Calderón. Soon afterwards they extended their results to all m, and more recently and by using a generalization of the theory of the function g of Littlewood and Paley, they obtained the continuity in L^p, $1<p<\infty$, of the left-hand side of (7) with HD^m replaced by a pseudo-differential operator in $S^m_{1,0}$ in several variables. Unfortunately, the estimates for the norms of the operators (7) obtained by these methods do not allow to sum the series resulting from the power series expansion of the function F in (8). However, last year it was observed, [5], that the technique of analytic functions used in the treatment of (6), strengthened by the results on weighted inequalities between a function, its maximal function and area function of B. Muckenhoupt, R. P. Gundy and R. L. Wheeden, [20], and certain results on conformal mapping, is applicable to the Cauchy integral in (9). It was already known that from results on the Cauchy integral there follow corresponding results on the integral in (8). Specifically, it was shown that the integral in (8) represents a bounded operator in L^p, $1<p<\infty$, provided that $\|a'\|_\infty<\varrho\alpha$, where ϱ is the radius of a disc centered at the origin where F is analytic, and α is an absolute positive constant. By means of the so-called rotation method one can extend this result to functions of several variables and prove that, for example, if $k(x,z)$, $x, z\in \boldsymbol{R}^n$, is bounded, homogeneous of degree $-n$ and even (odd) in z, F is odd (even) and analytic in a disc of radius ϱ centered at the origin, and $a(x)$ is Lipschitzian and such that $\|\nabla a\|_\infty<\varrho\alpha$, then the operator

$$\text{(11)} \qquad \text{p.v.} \int_{\boldsymbol{R}^n} k(x, x-y) F\left[\frac{a(x)-a(y)}{|x-y|}\right] f(y)\, dy$$

is well defined and continuous in L^p, $1<p<\infty$. Later on we shall outline the proof of the results of R. Coifman and Y. Meyer as well as those on this last integral.

Before proceeding to describe some of the applications of the foregoing results, I would like to mention still another result on commutators due to R. Coifman, R. Rochberg and G. Weiss, [17], which is of a different character. Let $k(x)$, $x\in\boldsymbol{R}^n$, be homogeneous of degree $-n$, of mean value zero on $|x|=1$, and sufficiently regular in $x\neq 0$, and K the operator convolution with k, then if $a(x)$ is of bounded mean oscillation the operator

$$C_a^m(K)f = \text{p.v.}\int (a(x)-a(y))^m k(x-y)f(y)\,dy$$

is bounded in L^p, $1<p<\infty$.

Now let us turn to applications. Let Γ be a simple rectifiable arc in the complex plane. Then the function

$$G(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(w)}{w-z}\,dw,$$

where $f(w)$ is a function on Γ which is integrable with respect to arc length, has a limit almost everywhere in Γ as z approaches nontangentially a point of Γ. In the case of several variables one has similar results about double layer potentials and derivatives of single layer potentials of functions defined on graphs of functions which are of bounded variation in the sense of Tonelli. This gives an affirmative answer to old problems about the existence of such limits.

Another application is the following result due to D. E. Marshall (personal communication) which confirms an old conjecture of A. Denjoy (C. R. Acad Sci. Paris **149** (1909), 258—260): the analytic capacity $\gamma(E)$ of a compact subset E of a rectifiable arc in the complex plane is zero if and only if its one-dimensional Hausdorff measure vanishes.

Finally, I will mention some applications to the theory of partial differential equations which motivated the study of our subject. In the first place, on the basis of the preceding results it is possible to construct algebras of singular integral operators [6] which allow to extend automatically to equations with bounded coefficients and terms of highest order with bounded Lipschitzian coefficients the results on the uniqueness of the Cauchy problem and the existence and uniqueness of solutions of totally hyperbolic systems obtained in [6] and [7]. On the other hand, results such as the ones obtained by E. Fabes, M. Jodeit and N. M. Riviere for the Laplace equation, [9], with the method of the integral equations on the boundary described earlier, are surely also obtainable for much more general elliptic systems. Let us see what these results are. Let Ω be a domain in $\mathbf{R}^n$ with boundary $\partial\Omega$ of class C^1. Let N_y be the interior normal unit vector at the point y of $\partial\Omega$, and Λ_y a cone with vertex at y, with fixed height and aperture, and except for its vertex, entirely contained in Ω. Then, in the case of the Dirichlet problem, one has the following: if $g(y)$ is a function in $L^p(\partial\Omega)$, $1<p<\infty$, there is a unique function $u(x)$, harmonic in Ω, and such that

$$u^*(y) = \sup\{|u(x)|\,|x\in\Lambda_y\}\in L^p(\partial\Omega),\ \lim_{x\to y}\{u(x)|x\in\Lambda_y\} = g(y) \quad \text{p.p.}$$

This result was obtained for the first time by B. E. J. Dahlberg with different methods which show that if $p\geq 2$ the same holds even if $\partial\Omega$ is merely Lipschitzian. If in addition $g(y)\in L_1^p(\partial\Omega)$, then $|\nabla u|^*(y)$, whose definition is similar to that of $u^*(y)$, also belongs to $L^p(\partial\Omega)$. On the other hand, in the case of the Neumann problem, one has that if $g(y)\in L^p(\partial\Omega)$ there exists a harmonic function $u(x)$, which is unique up to an additive constant, such that $|\nabla u|^*(y)$ is in $L^p(\partial\Omega)$ and

$$\lim_{x\to y}\{\nabla u(x)\cdot N_y | x\in\Lambda_y\} = g(y) \quad \text{p.p.}$$

These results are also valid if $\partial\Omega$ is merely Lipschitzian, provided that the local oscillation of N_y does not exceed a constant which depends on p but not on Ω, or only on certain global properties of Ω.

An interesting consequence of the preceding results is the following. Given p, $1<p<\infty$, there is a positive ε such that if the local oscillation of the normal N_y to the boundary $\partial\Omega$ of a Lipschitzian domain Ω is less than ε, then every harmonic measure on $\partial\Omega$ is absolutely continuous with respect to surface area and has a density in $L^p(\partial\Omega)$.

The method of R. Coifman and Y. Meyer. We shall now outline the elegant way in which these authors treat the problem of the commutators by its reduction to the continuity of certain multilinear operators. We shall confine ourselves to the bilinear case where the ideas and techniques they employ are already apparent.

THEOREM 1. *Let $\hat{\varphi}^{(1)}(\xi)$ and $\hat{\varphi}^{(2)}(\xi)$ be two infinitely differentiable functions with compact support in $\boldsymbol{R}^n$ such that at least one of them vanishes in a nelghborhood of the origin. Let $\varphi_t(x)=t^{-n}\varphi(x/t)$, $t>0$, where $\varphi(x)$ is the inverse Fourier transform of $\hat{\varphi}(\xi)$. Let*

$$g=\int_0^\infty (f_1*\varphi_t^{(1)})(f_2*\varphi_t^{(2)})m(t)\frac{dt}{t}$$

where $m(t)$ is a bounded function. Then

$$\|g_2\| \leqslant c\|f_1\|_2\|f_2\|_\infty\|m\|_\infty$$

where c depends only on the functions $\varphi^{(j)}(x)$.

In order to show this let us assume first that both functions $\hat{\varphi}^{(j)}(\xi)$ vanish near the origin, that is, they both have support in $0<a\leqslant|\xi|\leqslant b$. It is easy to see that if $\hat{\eta}(\xi)$ has compact support and equals 1 in $\xi\leqslant 2b$, and $\hat{\eta}(\xi)=\hat{\eta}(-\xi)$, then

$$g=\int_0^\infty \eta_t*[(f_1*\varphi_t^{(1)})(f_2*\varphi_t^{(2)})]m(t)\frac{dt}{t}$$

and

$$\int hg\,dx=\int\int_0^\infty (h*\eta_t)(f_1*\varphi_t^{(1)})(f_2*\varphi_t^{(2)})m(t)\frac{dt}{t}dx. \tag{12}$$

If we assume now that f_2 is bounded then $|(f_2*\varphi_t^{(2)})|^2(dx\,dt/t)$ is a Carleson measure, that is, if Q is a cube in $\boldsymbol{R}^n\times\{t\geqslant 0\}$ with base B in $t=0$, and denote this measure by μ, then $\mu(Q)\leqslant c\|f_2\|_\infty^2|B|$. Hence, as is well known,

$$\int |(h*\eta_t)|^2\,d\mu\leqslant c\|h\|_2^2\|f_2\|_\infty^2.$$

On the other hand, by taking Fourier transforms one verifies readily that

$$\int_0^\infty \int |(f_1 * \varphi_t^{(1)})|^2 \frac{dx\,dt}{t} \leq c\|f_1\|_2^2$$

and using these inequalities in estimating (12) the desired result follows.

In case one of the functions $\hat{\varphi}^{(i)}(\xi)$, call it simply $\hat{\varphi}(\xi)$, does not vanish near the origin while the other vanishes in $|\xi| \leq a$, one decomposes $\hat{\varphi} = \hat{\Phi} + \hat{\varrho}$, where $\hat{\Phi}$ vanishes in $|\xi| \leq a/8$ and $\hat{\varrho}$ has support in $|\xi| \leq a/4$. The contribution of $\hat{\Phi}$ is treated exactly as is the preceding case. The contribution of $\hat{\varrho}$ is treated essentially the same way, with the difference that now one chooses $\hat{\eta}(\xi)$ so that it equals 1 in $a/4 \leq |\xi| \leq 2b$ and vanishes near the origin, and then (12) can again be estimated as above.

Remark. A closer examination of the preceding argument shows that if $\hat{\varphi}^{(j)}(\xi)$ is replaced by $e^{iu_j\cdot\xi}\hat{\varphi}^{(j)}(\xi)$, then

$$\|g\|_2 \leq c(1+|u_1|)^{n/2}(1+|u_2|)^{n/2}\|f_1\|_2\|f_2\|_\infty\|m\|_\infty$$

where again c depends only on the function $\varphi^{(j)}(x)$.

Theorem 2. *Let $p(x, \xi)$, $x=(x_1, x_2)\in R^n\times R^n$, $\xi=(\xi_1, \xi_2)\in R^n\times R^n$, be a symbol of class $S^0_{1,0}$ and $p(x, D)$ the corresponding pseudo-differential operator. Then if $f(x)=f_1(x_1)f_2(x_2)$ and g is the restriction to the diagonal $x_1=x_2$ of $p(x, D)f$ we have*

$$\|g\|_2 \leq c\|f_1\|_\infty\|f_2\|_2.$$

For the sake of simplicity we shall only consider the case where p is independent of x and vanishes near the origin. We shall show that g is a convergent integral of operators like those in the preceding theorem. First we take two infinitely differentiable functions with compact support $\hat{\varphi}(\xi)$ and $\hat{\psi}(\xi)$, $\xi\in R^n$, one of which vanishes near the origin, and such that

$$r(\xi_1, \xi_2) = \int_0^\infty [\hat{\varphi}(t\xi_1)^2\hat{\psi}(t\xi_2)^2 + \hat{\psi}(t\xi_1)^2\hat{\varphi}(t\xi_2)^2]\frac{dt}{t} \neq 0,\ \ |\xi_1|+|\xi_2| > 0.$$

Then if

$$p(\xi_1, \xi_2) = q(\xi_1, \xi_2)r(\xi_1, \xi_2)$$

and

$$p_t(\xi_1, \xi_2) = q(\xi_1/t, \xi_2/t)[\hat{\varphi}(\xi_1)\hat{\psi}(\xi_2) + i\hat{\psi}(\xi_1)\hat{\varphi}(\xi_2)],$$

$$p_t(\xi_1, \xi_2) = \int e^{i(\xi_1\cdot u_1)+i(\xi_2\cdot u_2)}m(t, u_1, u_2)\,du_1\,du_2,$$

we have $|m(t, u_1, u_2)| \leq c_k(1+|u_1|+|u_2|)^{-k}$, $\forall k$. Now, if $x\in R^n$, then

$$g(x) = \frac{1}{(2\pi)^n}\int p(\xi_1, \xi_2)e^{-ix\cdot(\xi_1+\xi_2)}\hat{f}_1(\xi_1)\hat{f}_2(\xi_2)\,d\xi_1\,d\xi_2$$

and replacing p by rq and using the preceding identities we obtain

$$g = \frac{1}{(2\pi)^n} \int_0^\infty \int e^{i(\xi_1 t\cdot u_1 + \xi_2 t\cdot u_2 - x\cdot\xi_1 - x\cdot\xi_2)} (\hat\varphi(t\xi_1)\hat\psi(t\xi_2) - i\hat\psi(t\xi_1)\hat\varphi(t\xi_2))$$

$$\cdot \hat f_1(\xi_1)\hat f_2(\xi_2) m(t, u_1, u_2)\, du_1\, du_2\, d\xi_1\, d\xi_2 \frac{dt}{t}$$

$$= \int_0^\infty \int ((\varphi_t^{u_1} * f_1)(\psi_t^{u_2} * f_2) - i(\psi_t^{u_1} * f_1)(\varphi_t^{u_2} * f_2)) m(t, u_1, u_2) \frac{dt}{t}$$

where φ^u is the inverse Fourier transform of $e^{iu\cdot\xi}\hat\varphi(\xi)$. Now an application of the preceding theorem yields the desired result.

THEOREM 3. *Let $p(x, D)$ be a pseudo-differential operator of type $S^0_{1,0}$ and $a(x)$ a function with bounded derivatives. Let C_a denote the commutator of multiplication by $a(x)$ and*

$$g = [C_a p(x, D)]\, \partial f/\partial x_1;$$

then

$$\|g\|_2 \le c\|f\|_2 \|\nabla a\|_\infty .$$

We shall assume again that p is independent of x. Then, as is readily verified

$$g = \frac{1}{(2\pi)^n} \int \xi_1 [p(\xi) - p(\xi+\eta)] e^{-ix\cdot(\xi+\eta)} \hat f(\xi) \hat a(\eta)\, d\xi\, d\eta.$$

Now we decompose

$$\xi_1[p(\xi) - p(\xi+\eta)] = \sum_1^n q_j(\xi, \eta)\eta_j + p(\xi) r_j(\xi, \eta)\eta_j - p(\xi+\eta) s_j(\xi, \eta)\eta_j$$

where the q_j are functions in the class $S^0_{1,0}$ multiplied by homogeneous functions of degree zero, and the r_j and s_j are homogeneous of degree zero and infinitely differentiable away from the origin. The preceding theorem applies to these functions thought of as symbols. If we denote now by $B(u)$ (f_1, f_2) the bilinear operator of the preceding theorem associated with the symbol $u(\xi, \eta)$. Then the following identities are readily verified

$$B(q_j\eta_j)(f, a) = B(q_j)(f, \partial a/\partial x_j),$$

$$B(p(\xi) r_j(\xi, \eta)\eta_j)(f, a) = B(r_j)(p(D)f, \partial a/\partial x_j),$$

$$B(p(\xi+\eta) s_j(\xi, \eta)\eta_j)(f, a) = p(D) B(s_j)(f, \partial a/\partial x_j)$$

and an application of the preceding theorem yields the desired result.

The method of the Cauchy integral. We will see now how the study of the integral (11) reduces to that of (8), this to that of (9), and in turn, this one to that of an integral similar to the one which appears in (6). Unfortunately the study of the latter is too complicated to allow a brief description and we are compelled to refer the reader to the literature in this respect (see [4] and [5]).

Let us write the integral (8) as

$$g(s) = \text{p.v.} \int_{-\infty}^{+\infty} t^{-1} F\left[\frac{a(s)-a(s-t)}{t}\right] f(s-t)\, dt$$

and let us assume that for a given function F and a positive number M we have $\|g\|_p \leq c\|f\|_p$ whenever the function $a(t)$ satisfies the condition $|a(s)-a(t)| \leq M|s-t|$. We also write the integral in (11) as

$$g(x) = \text{p.v.} \int_{R^n} k(x, y) F\left[\frac{a(x)-a(x-y)}{|y|}\right] f(x-y)\, dy,$$

and, if v denotes a unit vector and t a real variable, we define

$$g(x, v) = \frac{1}{2} k(x, v)\, \text{p.v.} \int_{-\infty}^{+\infty} t^{-1} F\left[\frac{a(x)-a(x-t)}{t}\right] f(x-t)\, dt.$$

Given our assumptions on the parities of k and F, integration with respect to v on the unit shpere Σ yields

$$\int_{\Sigma} g(x, v)\, d\sigma = g(x).$$

Now, on account of our assumptions on the one-dimensional case and boundedness of $k(x, y)$ we have

$$\int_{-\infty}^{+\infty} |g(x+tv, v)|^p\, dt \leq c \int_{-\infty}^{+\infty} |f(x+tv)|^p\, dt$$

and integrating with respect to x on a hyperplane perpendicular to v we obtain $\|g(x, v)\|_p^p \leq c\|f\|_p^p$. Integrating $g(x, v)$ with respect to v on the unit sphere Σ and using Minkowski's integral inequality we find that $\|g\|_p \leq c\|f\|_p$. Next let us see how the integral (8) reduces to that in (9). Let L denote the operator represented by the integral (8) and

$$A_z f = \text{p.v.} \int_{-\infty}^{+\infty} [s-t-z^{-1}(a(s)-a(t))]^{-1} f(s)\, ds.$$

Then, if $F(z)$ is analytic in $|z| \leq \varrho = \sup |(a(s)-a(t))/(s-t)|$, we have

$$(Lf)(t) = \frac{1}{2\pi i} \int_{|z|=\varrho} F(z)(A_z f)(t) \frac{dz}{z}.$$

Now, the integral defining A_z is not of the form (9) because there the function $a(x)$ is real, while in general this is not the case in A_z. However, introducing the new variables $\bar{t} = t - ua(t)$, $\bar{s} = s - ua(s)$, where $z^{-1} = u + iv$, A_z takes the form of the integral (9), except for the absence of the factor $(1+ia'(s))$ in the integrand, which is irrelevant. This makes it possible to estimate the operator L by means of integrals of the form (9).

Finally, we will show how (9) can be estimated by means of integrals similar to that in (6). For this purpose let $z(\lambda)=t+i\lambda a(t)$, $w(\lambda)=s+i\lambda a(s)$, and consider the operator

$$A(\lambda)f = \text{p.v.} \int_{-\infty}^{+\infty} \frac{f(s)\,ds}{z(\lambda)-w(\lambda)}.$$

This operator has essentially the form (9) for each real value of λ, and for $\lambda=0$ it reduces to the Hilbert transform. Differentiating with respect to λ one obtains the operator

$$B(\lambda)f = \frac{d}{d\lambda}A(\lambda)f = \text{p.v.} \int_{-\infty}^{+\infty} \frac{-i}{z(\lambda)-w(\lambda)} \left[\frac{a(t)-a(s)}{z(\lambda)-w(\lambda)}\right] f(s)\,ds,$$

whose analogy with (6) is apparent. As was said before, the method used in the study (6) can be applied to $B(\lambda)$ and in this manner one obtains

$$d\|A(\lambda)\|/d\lambda \leqslant \|B(\lambda)\| \leqslant c[1+\|A(\lambda)\|]^2,$$

where the norms denote operator norms in L^2 and c denotes a constant depending on $\|a'\|_\infty$. From this differential inequality and the fact that $A(0)$ is the Hilbert transform and consequently $\|A(0)\|=\pi$, there follows that

$$\|A(1)\| \leqslant 2\pi(1-\|a'\|_\infty \alpha^{-1})^{-1}-\pi, \quad \alpha>0,$$

where α is an absolute constant. This result can be extended to L^p, $1<p<\infty$, by means of well known techniques.

Problems. There still are some basic unresolved problems in the subject we have been discussing. Consider the integral in (9). Are the results obtained so far about it also valid without restrictions on the norm $\|a'\|_\infty$? It is natural to expect an affirmative answer to this question. However, since the integral depends on a in a non-linear fashion, a negative answer cannot be ruled out. More generally we may ask which are the function spaces in which the operator given by the integral (8) is continuous with the only condition that the quotient $(a(x)-a(y))/(x-y)$ remain in a compact subset of the domain of analyticity of F. A clarification of this question would be very important in the study of boundary value problems for elliptic equations in general Lipschitzian domains. The methods employed so far seem to be insufficient for the treatment of these problems.

References

For further references to the Russian literature on the subject of Cauchy integrals the reader should consult the survey article [22] by V. P. Havin.

1. B. Bajsanski and R. Coifman, *On singular integrals*, Proc. Sympos. Pure Math., vol. 10, Amer. Math. Soc., Providence, R. I., 1967, pp. 1—17.

2. ——— *Pointwise estimates for commutators of singular integrals*, preprint.

3. Bui Doan Khanh, *Intégrales singulières, commutateurs et la fonction $f^\#$*, preprint.

4. A. P. Calderón, *Commutators of singular integral operators*, Proc. Nat. Acad. Sci. U.S.A. **53** (1965), 1092—1099.

5. ——— *Cauchy integrals on Lipschitz curves and related operators*, Proc. Nat. Acad. Sci. U.S.A. **74** (1977), 1324—1327.

6. ——— *Algebras of singular integral operators*, Proc. Sympos. Pure Math., vol. 10, Amer. Math. Soc., Providence, R. I., 1967, pp. 18—55.

7. ——— *Uniqueness in the Cauchy problem for partial differential equations*, Amer. J. Math. **80** (1958), 16—36.

8. ——— *Integrales singulares y sus aplicaciones a ecuaciones diferenciales hiperbólicas*, Cursos y seminarios de Matemática, Fasc. 3, Univ. Buenos Aires, 1960.

9. A. P. Calderón, C. P. Calderón, E. Fabes, N. M. Riviere and M. Jodeit, *Applications of the Cauchy integral on Lipschitz curves*, Bull. Amer. Math. Soc. **84** (1978), 287—290.

10. C. P. Calderón, *On commutators of singular integrals*, Studia Math. **53** (1975), 139—174.

11. R. Coifman and Y. Meyer, *Le Théorème des commutateurs de Calderón*, Sem. Anal. Harm. Orsay, 1974—1975, pp. 37—46.

12. ——— *On commutators of singular integrals and bilinear integrals*, Trans. Amer. Math. Soc. **212** (1975), 315—331.

13. ——— *Commutateurs d'intégrales singulières*, Sem. Anal. Harm. Orsay, No. **211**, 1976.

14. ——— *Commutateurs d'intégrales singulières et opérateurs multilineaires*, preprint, Orsay 1977.

15. ——— *Opérateurs pseudo-differentiels et le théorème de Calderón*, Sem. Anal. Harm. Orsay, 1976—1977, pp. 28—40.

16. ——— *Multilinear pseudo-differential operators and commutators*, preprint.

17. R. Coifman, R. Rochberg and G. Weiss, *Factorization theorems for Hardy spaces in several variables*, Ann. Math. **103** (1976), 611—635.

18. J. Cohen and J. A. Gosselin, *On multilinear singular integrals in R^n*, preprint.

19. M. Cotlar, *Condiciones de continuidad de operadores potenciales y de Hilbert*, Cursos y seminarios de Matemática, Fasc. 2, Univ. Buenos Aires, 1959.

20. R. P. Gundy and R. L. Wheeden, *Weighted integral inequalities for the non-tangential maximal function, Lusin area integral, and Walsh-Paley series*, Studia Math. **49** (1975), 107—124.

21. V. P. Havin, *On the continuity in L^p of an integral operator with Cauchy kernel*, Vestnik Leningrad Univ. **7** (1967), 103—108 (Russian).

22. ——— *Boundary properties of integrals of Cauchy type and harmonic conjugate functions in domains with rectifiable boundary*, Mat. Sb. (N. S.) **68 (110)** (1965), 499—517.

23. S. Janson, *Mean oscillation and commutators of singular integral operators*, Uppsala Univ. Report No. 5, 1977.

UNIVERSIDAD DE BUENOS AIRES
BUENOS AIRES, ARGENTINA

UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

von Neumann Algebras

A. Connes

For every selfadjoint operator T in the Hilbert space H,[1] $f(T)$ makes sense not only in the obvious case where f is a polynomial but also if f is just measurable, and if $f_n(x)\to f(x)$ for all $x\in\boldsymbol{R}$ (with (f_n) bounded) then $f_n(T)\to f(T)$ weakly, i.e. $\langle f_n(T)\xi,\eta\rangle\to\langle f(T)\xi,\eta\rangle\ \forall\xi,\eta\in H$. Moreover the set $\{f(T),\ f$ measurable$\}$ is the set of all operators S in H invariant under all unitary transformations of H which fix T. More generally, if (T_i), $i=1,\ldots,k$, are operators in H then the weak closure of the set of polynomials in T_i, T_i^* is the space of all operators in H invariant under all the unitaries fixing the T_i, as follows from the bicommutation theorem of von Neumann (1929):

A subset M of $L(H)$ is the commutant of a subgroup G of the unitary group $U(H)$ iff it is a weakly closed * subalgebra of $L(H)$ (containing the identity 1).

Such an algebra is called a *von Neumann algebra* (or ring of operators). Any commutative one is of the form $\{f(T), f$ measurable$\}$ for a selfadjoint T, and hence is the algebra of essentially bounded measurable functions: L^∞ (Spectrum T, Spectral measure T). In general the center of M is a commutative von Neumann algebra and hence an $L^\infty(X,\mu)$ for some measure space X, then $M=\{(T(x))_{x\in X},\ T(x)\in M(x)\ \forall x\in X\}$ is the algebra of all essentially bounded measurable sections of a family $M(x)$, $x\in X$, of von Neumann algebras with trivial centers, i.e. *factors*. If $M=\pi(G)'$ is, to start with, the commutant of the unitary representation π of the group G, by the above decomposition, π becomes the direct integral of *factor*

[1] With infinite countable orthonormal basis.

representations π_x, i.e. representations with $\pi_x(G)'$ a factor. As subrepresentations of π correspond bijectively to selfadjoint idempotents of $M=\pi(G)'$, to say that $\pi_x(G)'$ is a factor means that any two subrepresentations of π_x have a common subrepresentation. In finite dimension this says that π_x is a multiple of an irreducible subrepresentation, i.e. that $\pi_x(G)'$ is $M_n(C)$, with n=multiplicity of π_x, but in infinite dimension it is not always true that π_x has an irreducible subrepresentation, or equivalently that a factor always has a minimal projection. In fact it does iff it arises from an honest factorization of H as a tensor product: $H=H_1\otimes H_2$ with $M=\{T\otimes 1,\ T\in L(H_1)\}$. Murray and von Neumann discovered the existence of factors M not coming from the above trivial factorizations of H, and translating in terms of projections in M (i.e. selfadjoint idempotents $e=e^2=e^*\in M$) the comparison of subrepresentations they obtained the following *multiplicity theory:*

THEOREM *Let M be a factor, then there exists a unique (up to normalization) injection of equivalence classes of projections of M in* $[0,+\infty]$ *such that:*

$$\dim_M(e+f)=\dim_M(e)+\dim_M(f)\quad \textit{whenever}\quad e\perp f,$$[2]

and its range is

$\{0, 1, \ldots, n\}$ *then M is of type* I_n,

$\{0, 1, \ldots, \infty\}$ *then M is of type* I_∞,

$[0, 1]$ *then M is of type* II_1,

$[0, +\infty]$ *then M is of type* II_∞,

$\{0, +\infty\}$ *then M is of type* III.

The simplest example of a factor not of type I is the group algebra of an infinite discrete group Γ such that the normal subgroup of finite classes is trivial. One lets $R(\Gamma)$ be generated in $l^2(\Gamma)$ by the right translations, it is the commutant of the left translations, and is a factor. If ξ is the basis vector associated in $l^2(\Gamma)$ to the unit of Γ then the functional $\mathrm{Trace}_\Gamma(A)=\langle A\xi,\xi\rangle$ on $R(\Gamma)$ satisfies:

$$\mathrm{Trace}_\Gamma(AB)=\mathrm{Trace}_\Gamma(BA)\quad \forall A, B,$$

$$\mathrm{Trace}_\Gamma(1)\quad =1$$

which is impossible if M was of type I_∞, i.e. isomorphic to $L(H_1)$ since every $A\in L(H_1)$ is a finite sum of commutators. What is amazing in case II_1 (or II_∞) is that the relative dimension of projection $e\in M$ (or equivalently the relative multiplicity of subrepresentations of π) can be *any real number* α, even irrational, in $[0, 1]$. Moreover, if one defines for any selfadjoint $T\in M$, its relative trace by $\mathrm{Trace}_M(T)=\int\lambda\dim_M(dE_\lambda)$ (where $E_\lambda=1_{]-\infty,\lambda]}(T)$ is the spectral resolution of T), then, while it is easy to check that $\mathrm{Trace}_M(TT^*)=\mathrm{Trace}_M(T^*T)\geqslant 0\ \forall T\in M$,

[2] I.e. $ef=fe=0$.

the *additivity of the trace,* $\mathrm{Trace}_M(T_1+T_2)=\mathrm{Trace}_M(T_1)+\mathrm{Trace}_M(T_2)$, $\forall T_1, T_2$ was another striking result of Murray and von Neumann.

Around 1940 Gelfand and Naimark discovered a remarkable class of infinite dimensional algebras over $\boldsymbol{C}$. Among all $*$ algebras over $\boldsymbol{C}$ the C^* algebras are characterized by the very simple condition [**1**]:

$$\|x\| = \sqrt{\text{Spectral Radius } x^*x} \text{ is a complete norm.}$$

The commutative ones (with unit) are canonically isomorphic to the algebra of continuous functions on their compact spectrum. Every normed closed $*$ subalgebra of $L(H)$ is a C^* algebra and conversely every C^* algebra has a faithful representation in a Hilbert space. If $A=C(X)$ is a commutative C^* algebra and π a representation of A in H then each coefficient $f\to\langle\pi(f)\xi, \xi\rangle$ is a positive linear functional on $C(X)$, i.e. a Radon measure on X. In the noncommutative situation, positive linear functionals (i.e. elements φ of A^* with $\varphi(x^*x)\geq 0$) always exist in profusion (thanks to the convexity of $\{x^*x, x\in A\}$) and each determines a Hilbert space: the completion H_φ of A with the scalar product $\langle x, y\rangle_\varphi=\varphi(y^*x)$ and a representation π_φ of A in H_φ by left multiplication. This extends the usual construction of $L^2(X, \mu)$ for a Radon measure μ on the compact space X, and as in the commutative case the integral extends from continuous functions to measurable functions, i.e. here to the von Neumann algebra $\pi_\varphi(A)''$ generated by A in H_φ.

As an example let us describe the non commutative analogue of the construction of the probability space associated with the experiment of coin tossing. Instead of the Radon measure μ on the cantor set, $X=\prod_1^\infty X_\nu$, $X_\nu=\{a, b\}$, defined by

$$\mu(f_1\otimes f_2\otimes\ldots\otimes f_k\otimes 1) = \prod_1^k \mu(f_i)$$

one considers on the C^* algebra A, inductive limit of the $\bigotimes_1^k M_2(\boldsymbol{C})$, the positive linear functional Ψ such that

$$\Psi(x_1\otimes x_2\otimes\ldots\otimes x_k\otimes 1) = \prod_1^k \varphi(x_i)$$

where φ is a positive linear functional on $M_2(\boldsymbol{C})$ with $\varphi(1)=1$ (such a φ is called a state, because it corresponds to a state of a quantum mechanical system with $M_2(\boldsymbol{C})$ as algebra of observables). Up to unitary equivalence φ is always of the form

$$\varphi_\lambda(x) = \left(\frac{\lambda}{1+\lambda}\right)x_{11}+\left(\frac{1}{1+\lambda}\right)x_{22} \quad \forall x = [x_{ij}]\in M_2(\boldsymbol{C}).$$

The corresponding von Neumann algebras $R_\lambda=(\pi_{\varphi_\lambda}(A))''$ are factors of type III and R. Powers (motivated by quantum field theory) proved in 1967 that they *are mutually nonisomorphic.* Previously only finitely many nontype I factors were known. The problem of classification of von Neumann algebras up to *spatial* *isomorphism* (i.e. as pairs (H, M)) was since the beginning of the theory reduced to the problem of *algebraic isomorphism.* (If M is a factor, then the isomorphisms of M with von Neumann algebras in H are parametrized up to equivalence by an integer $n \in \{1, \dots \infty\}$ in the type I case, a real $\lambda \in]0, +\infty]$ in the type II case and are all equivalent in the type III case.) Moreover an abstract * algebra M is a von Neumann algebra iff (1) it is a C^* algebra (2) as a Banach space it is a dual [31]. Moreover the predual of a C^* algebra M is unique, if it exists, and is the space of σ-additive linear functionals φ on M (i.e. $\varphi(\sum E_\alpha) = \sum \varphi(E_\alpha)$ for any family of pairwise orthogonal projections). A *foliated manifold* $\mathfrak{f}$ gives rise in a natural way to such an abstract von Neumann algebra $R(\mathfrak{f})$. Let Ω be the set of leaves of $\mathfrak{f}$, a random operator $T = (T_f)_{f \in \Omega}$ is a bounded measurable family of operators, T_f acting in $L^2(f)$ for all f. Sums, product and * are defined pointwise, and as in usual measure theory, one neglects any set of leaves whose union in V is negligible (here for the smooth measure class) and any random operator T with $T_f = 0$ for almost all leaves. Thus $R(\mathfrak{f})$ plays the role of the algebra of all bounded operators in "L^2 (generic leaf of $\mathfrak{f}$)". It is not of type I in general, it is a factor iff $\mathfrak{f}$ is ergodic (i.e. any measurable function on V, constant on the leaves, is a.e. constant), and can be of type II_∞ or III. If Λ is a holonomy invariant transverse measure for $\mathfrak{f}$ one can give a meaning to $\varphi(T) = \int \operatorname{Trace}(T_f)\, d\Lambda(f)$ for every positive random operator T,[3] and this defines on the von Neumann algebra M of random operators (modulo equality Λ almost everywhere) a functional φ satisfying:

(1) φ is a *weight* on M i.e. φ is a linear map from M_+ to $[0, +\infty]$, $\varphi(\operatorname{Sup} T_\alpha) = \operatorname{Sup} \varphi(T_\alpha)$ for any increasing bounded family, and there are enough T with $\varphi(T) < \infty$ to generate M.

(2) φ is *faithful:* $\varphi(T) > 0$, $\forall T > 0$ in M.

(3) φ is a *trace* i.e. is unitarily invariant, $\varphi(UTU^{-1}) = \varphi(T)$.

Here (3) is the translation of the holonomy invariance of Λ.

Every von Neumann algebra M has a faithful weight; those which possess a faithful trace are called *semifinite.* The *additivity of the trace* of Murray and von Neumann shows that a factor fails to be semifinite iff it is of type III. Around 1950, Dixmier and Segal showed many important consequences of semifiniteness. One can define, as in usual integration theory, the L^p spaces by the norms

$$\|x\|_p = (\operatorname{Trace}_M |x^p|)^{1/p} \quad \text{where } x \in M,\ |x| = \sqrt{x^* x}.$$

Then L^1 is the predual M_*, and the representation π of M by left multiplication in L^2 satisfies the commutation theorem:

$$\pi(M)' = J\pi(M)J,\ \ J : L^2 \to L^2,\ J^2 = 1$$

where J is the isometric involution $x \to x^*$ in L^2. As a corollary one gets the commutation theorem for tensor products $((M_1 \otimes M_2)' = M_1' \otimes M_2')$ for M_1 and

[3] This can be finite even if Trace $T_f = +\infty$ for all $f \in \Omega$, see [8] for more details.

M_2 *semifinite* and for any *unimodular* locally compact group G the fact that the right regular representation generates the von Neumann algebra $R(G)$ of left invariant operators in $L^2(G)$. The natural weight $\varphi_G(f)=f(e)$ (e the unit of G) on the convolution algebra $R(G)$ is a trace iff G is unimodular. J. Dixmier obtained the above result also for *nonunimodular* G, and it was Tomita who succeeded in proving the two other results (existence of (π, J) and commutation theorem for tensor products) for arbitrary von Neumann algebras — this theory, once supplemented by the general theory of weights (Takesaki, Combes, Pedersen, Haagerup) can be summarized as follows:

Instead of a trace, one starts with a faithful weight φ on M. The lack of tracial property for φ creates two natural scalar products $\varphi(x^*x)$ and $\varphi(xx^*)$ and hence a positive (unbounded) operator Δ_φ in the Hilbert space H_φ of the first scalar product. In the group algebra situation H_φ is identical with $L^2(G)$ and Δ_φ is the multiplication by the module Δ_G of G. In this special case, since Δ_G is a homomorphism (from G to $\mathbf{R}_+^*$) it follows that the one parameter group of unitaries Δ_φ^{it} normalizes $R(G)$. The most remarkable result of Tomita is that this is a general fact:

THEOREM. *Let M act in H_φ by left multiplications, then* $\Delta_\varphi^{it} M \Delta_\varphi^{-it} = M$ $\forall t \in \mathbf{R}$.

This result became central when Takesaki discovered that the corresponding one parameter group of automorphisms of M $(\sigma_t^\varphi(x)=\Delta_\varphi^{it} x \Delta_\varphi^{-it}\ \forall t\in \mathbf{R})$ is characterized (in its link with φ) by an algebraic form[4] of the condition long known in quantum statistical physics as the Kubo Martin Schwinger condition. (If A is the algebra of observables, φ a statistical state, and σ_t the time evolution, a group of automorphisms of A, then (φ, σ) satisfies the Kubo Martin Schwinger condition at inverse temperature β iff $\varphi(x\sigma_{-i\beta}(y))=\varphi(yx)$ $\forall x, y\in A$. When $A=L(H)$ and $\sigma_t(x)=e^{itH}xe^{-itH}$ where H is the hamiltonian, the unique φ satisfying this condition is the Gibbs state $x \to \mathrm{Trace}\,(e^{-\beta H}x)/\mathrm{Trace}\,(e^{-\beta H})$.) After the discovery of Powers in 1967 of the non isomorphism of the factors R_λ, $\lambda\in]0, 1[$, Araki and Woods analyzed the infinite tensor products of finite dimensional factors by means of two invariants, computable in terms of the eigenvalue list,

$$r_\infty(M) = \{\lambda\in \mathbf{R}_+^* \mid M\otimes R_\lambda \text{ is isomorphic to } M\},$$
$$\varrho(M) = \{\lambda\in \mathbf{R}_+^* \mid M\otimes R_\lambda \text{ is isomorphic to } R_\lambda\}.$$

My point of departure was the existence of simple formulae relating, in the special case considered by Araki and Woods, those invariants and the Tomita—Takesaki theory, namely:

$$r_\infty(M) = \bigcap_\varphi \mathrm{Sp}\,\Delta_\varphi,\quad \varrho(M) = \{e^{2\pi/T}, T\in \bigcup_\varphi \mathrm{Ker}\,\sigma^\varphi\}.$$

This suggested that one ought to study for their own sake the invariants $S(M)=$

---

4 Identified by Haag Hugenholtz and Winnink in 1966.

$\bigcap_\varphi \operatorname{Sp} \Delta_\varphi$ and $T(M)=\bigcup_\varphi \operatorname{Ker} \sigma^\varphi$. The first question was computability. In the semifinite case, all weights φ are of the form $\varphi(x)=\operatorname{Trace}_M(\varrho x)$ where ϱ is a positive operator and the spectrum of Δ_φ is the closure of the set of ratios $\lambda_1/\lambda_2, \lambda_i \in \operatorname{Spectrum} \varrho$ while $\sigma_t(x)=\varrho^{it} x \varrho^{-it}$ for $x \in M$, taking $\varphi=\operatorname{Trace}_M$, $S(M)=\{1\}$, $T(M)=\boldsymbol{R}$. In the type III case, the one parameter group σ^φ is never inner but the following result solved completely the problem of computability of S and T.

THEOREM. *Let M be a von Neumann algebra,* Aut M *its automorphism group,* ε: Aut $M \to$ Out $M=$ Aut $M/$Int M *the canonical quotient map, and φ a weight on M. Then a one parameter group of automorphisms of M, $(\alpha_t)_{t\in \boldsymbol{R}}$ is of the form σ^Ψ for a suitable Ψ iff $\varepsilon(\alpha_t)=\varepsilon(\sigma_t^\varphi)$ $\forall t \in \boldsymbol{R}$.*

In particular together with a type III factor there is a canonical homomorphism $\delta: \boldsymbol{R} \to \operatorname{Out} M$, with $\delta(t)=\varepsilon(\sigma_t^\varphi)$ for *any* weight φ. Moreover with a suitable notion of spectrum for δ one has:

$$S(M) = \operatorname{Spectrum} \delta, \; T(M) = \operatorname{Kernel of } \delta.$$

In particular both are subgroups (of $\boldsymbol{R}_+^*$ and $\boldsymbol{R}$). As S is closed and as closed subgroups of $\boldsymbol{R}_+^*$ form a compact interval [0, 1], one gets a finer classification of type III factors:

M is of type III_λ, $\lambda \in]0, 1[$ if $S(M)=\{0\}\cup\lambda^{\boldsymbol{Z}}$,
M is of type III_0 if $S(M)=\{0, 1\}$,
M is of type III_1 if $S(M)=[0, +\infty[$.

In the case of foliations the invariant S of the von Neumann algebra coincides with the *ratio set* introduced by W. Krieger in ergodic theory as a generalization of the Araki-Woods ratio set.

Roughly speaking to evaluate the ratio set of a foliation, one travels on the generic leaf from the point a to a point b which is close to a in V (but at any distance on the leaf) and one compares a unit of transversal volume in a with its transformed under holonomy at b; the set of all essential such ratios coincides with S, and is thus a natural obstruction to the existence of a holonomy invariant choice of unit of volume in the transverse bundle.

Exactly as in noncommutative algebra where one uses the cross product of an algebra by a group of automorphisms, one defines the cross product of a factor N by an automorphism θ (it is characterized as a von Neumann algebra M generated by N and a unitary U with $UxU^*=\theta(x)$ $\forall x \in N$, so that the equality $\sigma_t(x)=x$ $\forall x \in N$, $\sigma_t(U)=e^{it}U$ defines an automorphism of M for all $t \in \boldsymbol{R}$).

The general theory of factors of type III_λ, $\lambda \in]0, 1[$ is summarized as follows:

(a) Let N be a factor of type II_∞ and θ an automorphism with $\bmod(\theta)=\lambda$ (i.e. $\operatorname{Trace}_N \circ \theta=\lambda \operatorname{Trace}_N$); then the cross product $N \otimes_\theta \boldsymbol{Z}$ is a factor of type III_λ.

(b) Any factor of type III_λ is of the form (a), and in a unique way (i.e. if (N_i, θ_i) give the same M there exists an isomorphism $N_1 \to N_2$ carrying θ_1 on θ_2.

In case III_0 we proved an analogue discrete description but the definitive understanding and solution of the III_1 case is contained in the following result of Takesaki:

Any factor of type III is of the form $N \otimes_\theta \boldsymbol{R}_+^*$ where N is a von Neumann algebra of type II_∞ (i.e. in its central decomposition $N = \{(x(u))_{u\in A}, x(u)\in N(u)\ \forall u \in A\}$ every $N(u)$ is a factor of type II_∞) and where for some trace τ on N one has $\tau \circ \theta_\lambda = \lambda\tau$. Moreover this decomposition is unique as above, and:

The restriction of θ_λ to A defines an ergodic flow $F(M)$, which is an invariant of M. This flow has a very natural interpretation as an abstract flow of weights on M.

One has $S(M) = \{\lambda, F_\lambda = \mathrm{id}\}$; when M is of type III_1 it follows that N is a factor so one gets the analogue of (a), (b) with the group $\boldsymbol{Z}$ replaced by $\boldsymbol{R}$.

In the III_λ case, $N = \{(x(u))_{u\in S^1}, x(u)\in N(u), \forall u \in S^1\}$ so that N "fibers over a circle", and the θ of (a), (b) is θ_λ. The above structure theorem for factors of type III_λ reduces the problem of classification in this case to

(1) Classify factors of type II_∞.

(2) Given a factor of type II_∞, N, classify (up to conjugacy) its automorphism with module λ, $\lambda \in]0, 1[$.

Every factor of type II_∞ is the tensor product of a factor of type II_1 by the type I_∞ factor. In the last of their papers, Murray and von Neumann had shown that, though there exists more than one factor of type II_1 (they exhibited 2, in 1968 D. MacDuff constructed a continuum of them) there is among them, only one having the following approximation property: $\forall$ finite subset F of N, $\forall \varepsilon > 0$, $\exists$ a finite dimensional * subalgebra K with distance $(x, N) < \varepsilon$, $\forall x \in F$ (where the distance is in the hilbert space L^2 of the trace_N). As any other factor of type II_1 contains this hyperfinite one, it was hence natural to think it is the simplest of all and to consider problem (2) in this case. The answer is the following:

For $\lambda \in]0, 1[$ there is, up to conjugacy, only one automorphism of $R_{0,1} = R \otimes \mathrm{I}_\infty$ with module λ.

If $1/\lambda$ is an integer n, one can construct θ_λ as the shift on $R_{0,1}$ built as an infinite tensor product of $n \times n$ matrices. As another example, if T is the Anosov diffeomorphism of the 2 torus $\boldsymbol{R}^2/\boldsymbol{Z}^2$ defined by the matrix $\left[\begin{smallmatrix}1 & 1\\ 1 & 2\end{smallmatrix}\right]$ then T defines an automorphism of its stable foliation, and hence of the corresponding factor which is $R_{0,1}$, this automorphism has module λ where (λ, λ^{-1}) are the eigenvalues of the above matrix. A crucial motivation in the proof of the above theorem is that, since the study of automorphisms of abelian von Neumann algebras is equivalent to ergodic theory of a single transformation, one would expect many results of this theory to have an analogue in the non abelian situation. This turns out to be the case in particular for the Rokhlin tower theorem.

There is however a striking difference with usual ergodic theory, the existence of a *complex valued* invariant for periodic automorphisms. If N is a factor, it can happen for $\theta \in \mathrm{Aut}\, N$ that θ^k is inner for some $k > 0$, but that no automorphism θ', $\varepsilon(\theta') = \varepsilon(\theta)$ satisfies $\theta'^k = 1$, the resulting obstruction is a kth root of 1 in $\boldsymbol{C}$, $\gamma(\theta)$

which is invariant by multiplication of θ by an inner automorphism. This happens when $N=R$, every pair (k, z), $k>0$, $z\in C$, $z^k=1$ appears from a $\theta\in \operatorname{Aut} R$ and moreover the pair (k, z) is the only invariant of $\varepsilon(\theta)\in \operatorname{Out} R$, in other words the group $\operatorname{Out} R=\operatorname{Aut} R/\operatorname{Int} R$ has only countably many conjugacy classes parametrized by (k, z). As a corollary one gets that $\operatorname{Int} R$ is the only normal subgroup of $\operatorname{Aut} R$.

Elaborating on the existence of this complex valued invariant, we showed that not all factors (even of type II_1) are antiisomorphic to themselves.

In general if N is a factor of type II_∞ one has a lot of non conjugate automorphisms with the same module $\lambda\in]0, 1[$; it was thus very natural to decide when, given a factor M of type III_λ, the corresponding factor of type II_∞ is $R_{0,1}$. If one knows that it is $R_{0,1}$ then by the above theorem one knows M is isomorphic to Powers factor R_λ.

As seen above R is characterized, among factors of type II_1, by the approximation property of Murray and von Neumann. In the general (non II_1) case, a factor M is called *approximately finite dimensional*[5] when:

$\forall F$ finite subset of $M, \forall *$ strong neighborhood V of 0

$\exists K$ finite dimensional $*$ subalgebra with $K+V$.

As an elaboration on Glimm's theorem characterizing C^*-algebra with only type I representations, it follows from the work of O. Marechal [22] and Elliott-Woods [13] that for any approximately finite dimensional factor M (not of type I_n, $n<\infty$, or II_1) and any C^*-algebra A not of type I, there is a representation π of A which generates M as a von Neumann algebra. Thus as soon as one goes beyond type I C^*-algebras one meets this whole class of factors. Moreover if A is the C^*-algebra corresponding to the "non commutative Cantor set" i.e. $A=\bigotimes_1^\infty M_2(C)$, then for any representation of A, $\pi(A)$ is approximately finite dimensional.

This obviously raises two questions:

(α) Classify the approximately finite dimensional factors.

(β) Characterize the C^*-algebras which generate only approximately finite dimensional factors.

In 1968 after trying to characterize $R_{0,1}$ (among factors of type II_∞) by the approximation property above, V. Ya. Golodets succeeded in showing that this class is stable under crossed products by abelian groups. It follows in particular that if M is of type III_λ, and M is AFD then the associated II_∞ also is AFD. This indicated the interest of the problem: is $R_{0,1}$ unique among AFD of type II_∞. The difficulty is that while any II_∞ is $\mathrm{II}_1\otimes \mathrm{I}_\infty$ it is very difficult to see what property inherited by the II_1 would force it to be isomorphic to R.

In fact the characterization of R, of Murray and von Neumann involves * subalgebras, and hence has still some descriptive flavor. The second factor of type II_1 which they discovered was distinguished by "property Γ" which they considered

[5] In short AFD.

technical; this property had no reason to characterize R since for any N, $N\otimes R$ possesses it, and in fact in 1962 J. T. Schwartz distinguished between N, R and $N\otimes R$. But in doing that, he found another property of R which was the germ of many later developments.

Property P. M in H has property P iff for any bounded $T\in L(H)$, the norm closed convex hull of the uTu^*, u unitary of M, intersects M'.

He proved that among N, R and $N\otimes R$ only R has property P, and moreover that the group algebra $R(\Gamma)$ of a discrete group has property P iff Γ is amenable. Any AFD factor possesses property P, but it is not clear from the definition that if $M=N\otimes Q$ then N has property P if M does. In fact the most important consequence of property P is the existence of a projection of norm one E from $L(H)$ to M', with $E(1)=1$. By a result of J. Tomiyama any such projection satisfies $E(aTb)=aE(T)b$ $\forall a, b\in M'$, $\forall T\in L(H)$, and the existence of such a projection of $L(H)$ on M is independent of the choice of representation. The family of von Neumann algebras satisfying it has the following remarkable stability properties:

(1) It is a monotone class (under decreasing intersections and weak closure of ascending unions).

(2) It is stable under commutant.

(3) Stable under cross products by amenable groups.

(4) Stable under tensor product.

The name used to qualify this class is *injectivity*, since it characterizes, thanks to a noncommutative version of the Hahn-Banach theorem due to W. Arveson, those von Neumann algebras which are injective objects in the category of C^* algebras, with completely positive maps as morphisms. As shown by Choi and Effros, it is also equivalent to the existence of a solution in M of the equation $y\otimes a\leqslant b$ (where $a\in M_n(C)$, $a=a^*$ and $b\in M\otimes M_n(C)$ are given) as soon as a solution exists in $L(H)$. This is very useful because it allows us to treat direct integrals:

(5) $M=\{(x(s))_{s\in A}, x(s)\in M(s)\ \forall s\in A\}$ is injective iff almost all $M(s)$ are injective.

So let M be an injective von Neumann algebra, (5) and the reduction theory of von Neumann allow to assume that M is a factor, then the corresponding von Neumann algebra of type II_∞ is injective by (3) and again by (5) one can reduce to analysing injective factors of type II_∞ and finally of type II_1, writing $M=N\otimes I_\infty$. Then N is injective of type II_1 and by Tomiyama's theorem any projection of norm one $E: L(H)\to N$, with $E(1)=1$ satisfies $E(aTb)=aE(T)b$, $\forall a, b\in N$, $\forall T\in L(H)$. It follows that $\varphi=\mathrm{Trace}_N\circ E$ is a state on $L(H)$ *invariant under all unitaries of* N. We call such a state an *hypertrace*. In 1960 M. Takesaki had shown that if A_1, A_2 are simple C^* algebras then $A_1\otimes A_2$ is also simple (here $A_1\otimes A_2$ acts in $H_1\otimes H_2$ if A_1 and A_2 act in H_1, H_2), his proof involved a characterization of the norm on the algebraic tensor product $A_1\odot A_2$ coming from the representation in $H_1\otimes H_2$ as the least of all possible C^* norms on $A_1\odot A_2$. The corresponding completion $A_1\otimes_{\min} A_2$ is called the minimal tensor product of A_1 and A_2. He showed moreover that (as in Grothendieck theory for locally convex spaces) for certain C^*

algebras (the *nuclear* ones by definition), only one C^* norm exists on $A \odot B$ for arbitrary C^* algebras B ([34], [31]). In 1972 Effros and Lance discovered that some factors (all the Araki—Woods factors at the time) give very good factorizations of $L(H)$ inasmuch as the natural map η from $M \odot M'$ in $L(H)$ given by $\eta(\sum a_i \otimes b_i) = \sum a_i b_i$ is not only an injective homomorphism but is an *isometry* from $M \otimes_{\min} M'$ to the C^* algebra $C^*(M, M')$ generated by M and M' in $L(H)$. They called this remarkable property *semidiscreteness* and proved semidiscreteness $\Rightarrow$ Injectivity. So we get

$$\begin{tikzcd}[column sep=small]
\text{Approximately finite dimensional} \arrow[dr, Rightarrow] & & & \text{Semidiscrete} \arrow[ddl, Rightarrow] \\
& \text{Property P} \arrow[dr, Rightarrow] & & \\
& & \text{Injective} &
\end{tikzcd}$$

In fact, these properties are all equivalent.

Assume first that N is a factor of type II_1 and is injective, the existence of an hypertrace on N implies that it is semidiscrete; then Takesaki's theorem shows that $C^*(N, N')$ is *simple* and hence that it cannot contain a nonzero compact operator in H. The following dichotomy then shows that N has property Γ. Let N be a factor of type II_1 in H; then N has property Γ or $C^*(N, N')$ contains all compact operators. (This was suggested by fine computations of C. Akemann and P. Ostrand showing that for the group algebra of free groups $C^*(N, N')$ contains all compact operators.)

Now N has property Γ iff the group $\operatorname{Int} N$ is not closed in $\operatorname{Aut} N$ (where $\operatorname{Aut} N$ is gifted with its natural topology: $\theta_\alpha \to \theta$ iff $\theta_\alpha(x) \to \theta(x)$ strongly for any $x \in N$). Moreover in general the closure of $\operatorname{Int} N$ is characterized in terms of $C^*(N, N')$ by the existence of an extension $\hat{\theta}$ of θ on N which is identity on N'. As in our case $C^*(N, N')$ is $N \otimes_{\min} N'$ we see that $\operatorname{Aut} N = \overline{\operatorname{Int}} N \not\subset \operatorname{Int} N$.

The next step is to show that $N \otimes R$ is isomorphic to N. A remarkable result of D. MacDuff asserts that this is true as soon as N has a central sequence which is not hypercentral, which once translated in terms of automorphisms implies that

$$\overline{\operatorname{Int}}\, N \not\subset \operatorname{ct} N \Rightarrow N \sim N \otimes R$$

where $\operatorname{ct} N$ is the normal subgroup of all automorphisms θ of N which are trivial on central sequences. Here one has $\operatorname{ct} N = \operatorname{Int} N$ because if $\theta \in \operatorname{ct} N$ then $\theta \otimes 1 \in \operatorname{ct}(N \otimes N)$ (this is due to a characterization of ct using $C^*(N, N')$) and as the symmetry $\sigma_N(x \otimes y) = y \otimes x$ in $N \otimes N$ is in $\overline{\operatorname{Int}}\, N \otimes N$ one has $\theta \otimes \theta^{-1}$ inner (and hence θ inner), because $\varepsilon(\mathrm{ct})$ and $\varepsilon(\overline{\mathrm{Int}})$ always commute and $\varepsilon(\theta \otimes \theta^{-1}) = [\varepsilon(\theta \otimes 1), \varepsilon(\sigma_N)]$. From the properties $N \sim N \otimes R$ and $\sigma_N \in \overline{\operatorname{Int}}$ one finally deduces the approximation property of Murray and von Neumann. This can be very simply seen if one assumed N to be a subfactor of R but for the general case one uses the existence of an isomorphism of N with a subfactor of the ultraproduct R^ω where ω is a free ultrafilter, which in turn follows from the analogue of the Day—Namioka proof of Følner's characterization of amenable groups. The role of the invariant mean is played by the hypertrace and $L(H)$ replaces $l^\infty(\Gamma)$ where Γ is the discrete group. Among those proofs the most technical are those relating properties of automorphisms (like $\theta\in\overline{\operatorname{Int} N}$) with properties of $C^*(N, N')$ (like the existence of $\hat{\theta}$). They involve an exhaustion method, allowing to pass from some infinitesimal information to a global one, and a probabilistic way of taking the polar decomposition of an operator (the usual way $x\to u(x)|x|$ being too discontinuous), based on the inequality $\int \|E^a(h^2)-E^a(k^2)\|_2^2\, da \leq \|h-k\|_2 \|h+k\|_2$ where E^a is the spectral projection $1_{[a,+\infty[}$. So we have now that all injective factors of type II_1 are isomorphic to R. As an immediate corollary, since all von Neumann subalgebras of R are also injective, one gets their complete classification up to isomorphism. It follows that R *is the only factor contained in all others,* which fully justifies the original belief of Murray and von Neumann that it is the simplest. Also if Γ is a *discrete amenable group* then its group algebra is isomorphic to R as soon as $\{g\in\Gamma$, class of g finite$\}=\{e\}$. If M is injective of type II_∞ then it is isomorphic to $R_{0,1}$. It follows that if G is an arbitrary *connected* locally compact group then the non type I part of its group algebra $R(G)$ is of the form $A\otimes R_{0,1}$ where A is an *abelian* von Neumann algebra. Moreover the type III theory, allows to deduce from that, that the above 4 properties are equivalent in general. We thus have only one class which has, on the one hand, the nice characterization seen after Glimm's theorem, and on the other all the stability properties of the injective.

Furthermore in their work on C^* tensor products, Effros and Lance had shown that (1) all representations of nuclear C^* algebras generate injective von Neumann algebras (2) that if all representations of a C^* algebra are semidiscrete, then the C^* algebra is nuclear. Hence the C^* algebras satisfying condition (3) are exactly the nuclear ones (as a corollary $C^*(G)$ is nuclear for G locally compact connected). Let us mention also that for foliations the injectivity of the associated von Neumann algebra is equivalent to the *amenability* of the foliation, a remarkable and very useful property developed by Zimmer for ergodic group actions. (For instance the action of the fundamental group Γ of a compact Riemann surface V on the natural Poisson boundary $\partial\tilde{V}$ of its covering space $\tilde{V}$ is always amenable ergodic and is often of type III_1.) Let us turn now to injective factors of type III. If M is of type III_λ, $\lambda\in]0,1[$ then M is isomorphic to Powers factor R_λ. Wolfgang Krieger has shown in 1973 that for factors associated to a single ergodic transformation of a measure space, the flow of weights is a complete invariant and can be any ergodic flow. It follows from a very powerful cohomological lemma in his proof, and from the discrete decomposition of factors of type III_0, that any injective factor of type III_0 arises from a single ergodic transformation of a measure space and is thus one of

Krieger's factors. Thus in the III_0 case, the classification problem is transferred to ergodic theory: there are as many injective factors of type III_0 as ergodic (non-transitive) flows. There is only one injective factor M with $r_\infty(M)=[0, +\infty]$, it is the Araki—Woods factor R_1, arising as algebra of local observables in the free field, but it is still unknown if it is the only injective of type III_1, (i.e. if $r_\infty(M)=S(M)$ for any injective). This factor R_1 is associated to the Anosov foliation of the geodesic flow of a Riemann surface of genus >1. We have used foliations above to illustrate the general theory by examples but von Neumann algebras can be very useful for the study of foliations per se. Ruelle and Sullivan have shown how, for an oriented foliation $\mathfrak{f}$ of the compact manifold V (i.e. the subbundle F of TV, tangent to $\mathfrak{f}$ is oriented), the holonomy invariant transverse measures Λ correspond exactly to *closed currents* C, "positive in the leaf direction". For such a measured foliation it is natural to define the Euler characteristic as $\langle e(F), [C]\rangle$, the Euler class of the bundle F evaluated on the cycle $[C]$ created by the current C. Now von Neumann algebras allow to define the Betti numbers

$$\beta_i = \int \dim H^i(f)\, d\Lambda(f) < \infty,^{6}$$

where $H^i(f)$ is the space of square integrable harmonic forms on f (with respect to some Euclidean structure on F, of which β_i turns out to be independent). One has then $\sum(-1)^i \beta_i = \langle e(F), [C]\rangle$. As β_0 is the measure of the set of compact leaves with finite holonomy, one gets that for 2 dimensional foliations without such leaves, the mean curvature of leaves is negative. The above formula is a special case of an index theorem computing for elliptic differential operators on $\mathfrak{f}$ the scalar

$$\int \mathrm{Dim}\,(\mathrm{Ker}\, D_f)\, d\Lambda(f) - \int \mathrm{Dim}\,(\mathrm{Ker}\, D_f^*)\, d\Lambda(f)$$

as $\mathrm{Ch}\, D\cdot\tau(F\otimes C)\,[C]$, where $\mathrm{Ch}\, D\in H^*(V, \mathbf{Q})$ is the chern character of the symbol of D, $\tau(F\otimes C)\in H^*(V, \mathbf{Q})$ the Todd class of $F\otimes C$ and $[C]\in H_p(V, \mathbf{R})$ the homology class of the Ruelle—Sullivan current.

Bibliography

1. H. Araki and G. Elliott, *On the definition of C^* algebras*, Publ. Res. Inst. Math. Sci. **9** (1973), 93—112.

2. M. Choi and E. Effros, *Injectivity and operator algebras* (preprint).

3. F. Combes *Poids associè à une algebre hilbértienne á gauche*, Compositio Math. **23** (1971), 49—77.

4. A. Connes, *Une classification des facteurs de type III*, Ann. Sci. École Norm. Sup. **6** (1973), 133—252.

5. ——— *Outer conjugacy classes of automorphisms of factors*, Ann. Sci. École Norm. Sup. **8** (1975), 383—420.

[6] Even though it may happen that $\dim H^i(f) = +\infty$, $\forall f\in\Omega$.

6. A Connes and M. Takesaki, *The flow of weights on factors of type III,* Tôhoku Math. J. **29** (1977), 473—575.

7. A. Connes, *Classification of injective factors,* Ann. of Math. **104** (1976), 73—115.

8. *Sur la theorie de l'integration non commutative* (preprint).

9. J. Dixmier, *Algebres quasi unitaires,* Comment. Math. Helv. **26** (1952), 275—322.

10. *Les algebres d'operateurs dans l'espace hilbertien,* 2nd ed., Gauthier-Villars, Paris.

11. *Les C^* algebres et leurs representations,* 2nd ed., Gauthier-Villars, Paris.

12. E. Effros and C. Lance, *Tensor product of operator algebras* Advances in Math. **25** (1977), 1—39.

13. G. Elliott and E. J. Woods, *The equivalence of various definitions of hyperfiniteness of a properly infinite von Neumann algebra* (preprint).

14. I. M. Gelfand and M. Naimark, *On the imbedding of normed rings into the ring of operators in Hilbert space,* Mat. Sb. **12** (1943), 197—213.

15. J. Glimm, *Type* I *C^* algebras,* Ann. of Math. **73** (1961), 572—612.

16. V. Ya. Golodets, *Crossed products of von Neumann algebras,* (Uspehi Math. Nauk. **26** (1971), 3—50.

17. U. Haagerup, *Normal weights on W^* algebras,* J. Functional Anal. **19** (1975), 302—317.

18. W. Krieger, *Ergodic flows and the isomorphism of factors,* Math. Ann. **223** (1976), 19—70.

19. D. Mc. Duff, *Uncountably many* II_1 *factors,* Ann. of Math. **90** (1969), 372—377.

20. *Central sequences and the hyperfinite factor,* Proc. London Math. Soc. **21** (1970), 443—461.

21. G. Mackey, *Ergodic theory, group theory and differential geometry,* Proc. Nat. Acad. Sci. U.S.A. **50** (1963), 1184—1191.

22. O. Marechal, *Une remarque sur un theorème de Glimm,* Bull. Sci. Math. **99** (1975), 41—44.

23. F. J. Murray and J. von Neumann, *On rings of operators,* Ann. of Math. **37** (1936), 116—229.

24. *On rings of operators* II, Trans. Amer. Math. Soc. **41** (1937), 208—248.

25. *On rings of operators* IV, Ann. of Math. **44,** (1943), 716—808.

26. J. von Neumann, *On a certain topology for rings of operators,* Ann. of Math. **37** (1936), 111—115.

27. *On rings of operators* III, Ann. of Math. **41** (1940), 94—161.

28. G. Pedersen, *C^* integrals,* Thesis, Copenhague University, 1971.

29. R. T. Powers, *Representations of uniformly hyperfinite algebras and their associated von Neumann rings,* Ann. of Math **86** (1967), 138—171.

30. D. Ruelle and D. Sullivan, *Currents, flows and diffeomorphisms,* Topology **14** 319—327.

31. S. Sakai *C^* and W^* algebras,* Ergebnisse der Mathematik Und ihrer Grenzgebite, vol. **60** Springer-Verlag, Berlin and New York, 1971.

32. J. T. Schwartz, *Two finite, non hyperfinite, non isomorphic factors,* Comm. Pure. Appl. Math. **16** (1963), 111—120.

33. I. Segal, *A non commutative extension of abstract integration,* Ann. of Math. **57** (1953), 401—457, correction 595—596.

34. M. Takesaki, *On the cross norm of the direct product of C^* algebras,* Tôhoku Math. J. **16** (1964), 111—122.

35. *Tomita's theory and its applications,* Lecture Notes in Math., vol. **128,** Springer-Verlag, Berlin and New York, 1970.

36. *Duality for crossed products and the structure of von Neumann algebras of type* III, Acta. Math. **131** (1973), 249—310.

University of Paris VI
Paris, France

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Topology of Manifolds and Cell-Like Maps

Robert D. Edwards

1. Introduction. The focus of this expository article will be on the notion of a cell-like set and a cell-like map (definitions below). It will be discussed how these notions arise naturally in the study of certain problems in topology, and how some solutions to these problems have been achieved. It should be emphasized at the outset that the problems discussed here are all *topological* in nature, and so in particular there will be a minimum of extra global structure on the various spaces at hand.

The principal questions to be discussed (and motivated) are the following:

I. POINT-LIKE QUESTION. Which compact subsets of the m-sphere have the property that their complements are homeomorphic to euclidean m-space R^m? (Such subsets are called *point-like,* for the natural reason.)

II. POLYHEDRAL MANIFOLD QUESTION. When is a polyhedron a topological manifold? In particular, are there any "unexpected" examples of such polyhedra? (i.e., examples which do not locally polyhedrally embed in the euclidean space of the same dimension.)

III. MANIFOLD FACTOR QUESTION. When is a space X a factor of a manifold, i.e., when is it the case that $X \times Y$ is a manifold for some space Y? (Usually Y is taken itself to be some euclidean space.)

2. Definitions. All spaces throughout are locally compact separable metric (except in §11, where local compactness is dropped). A manifold will always be understood to be a topological manifold, either finite dimensional, or else modelled on the hilbert cube I^∞ (which is the countably infinite topological product of the interval $[-1, 1]$ with itself). Precisely stated, then, a *(topological) manifold* is a separable metric space each point of which has a neighborhood homeomorphic either to the m-cell I^m or to I^∞. Our manifolds will always be connected.

Later we will be talking about the notion of an *absolute neighborhood retract* (ANR) which we will take to mean a (locally compact separable metric) space which can be embedded as a closed subset of $I^\infty \times [0, \infty)$ (recall any locally compact separable metric space can be so embedded) in such a manner that some neighborhood U of the image retracts to the image, i.e., there is a map $r: U \to$image such that $r|$image$=$identity. A basic fact about ANR's is that this retraction property is independent of the embedding chosen: if it holds for one closed embedding, it holds for any closed embedding (see e.g. [Hu]; the embeddings even need not be closed). A finite dimensional ANR is called a *euclidean neighborhood retract* (ENR) because it can be embedded as a closed subset of euclidean space so as to have this retraction property.

The fundamental notion in this article is that of *cell-likeness* which, as we will see, broadens a bit the notion of contractibility. A *cell-like* space is a compact metric space C having the following property of a cell: there exists an embedding of C into the hilbert cube I^∞ such that

$(*)$ for any neighborhood U of C in I^∞, C is null-homotopic in U

(examples and properties are given in the next section).

A *proper* map is a map such that the preimage of each compact subset is compact. A *cell-like* map is a proper surjection such that each point-inverse is cell-like. A *near-homeomorphism* is a proper surjection which can be *approximated arbitrarily closely by homeomorphisms*. On compact spaces, this means ordinary uniform approximation. On non-compact, locally compact spaces (which are only of secondary interest in this article) we take this to mean "majorant approximable" by homeomorphisms, i.e., given $f: X \to Y$ and given any majorant map $\varepsilon: X \to (0, \infty)$, there should exist a homeomorphism $h: X \to Y$ such that for each $x \in X$, $\operatorname{dist}(f(x), h(x)) < \varepsilon(x)$. Finally, $X \approx Y$ denotes that X is homeomorphic to Y.

3. Examples and properties of cell-like compacta. There are two basic comments concerning the definition of *cell-like* which should be recalled at this time.

REMARK. (1) Cell-likeness is an intrinsic property of the compact metric space C. That is, if property $(*)$ holds for one embedding $i: C \hookrightarrow I^\infty$, then it holds for any embedding $j: C \hookrightarrow I^\infty$.

(2) In the definition of cell-like, and in (1), if the hilbert cube I^∞ is replaced by any ANR, the statements remain true.

Both remarks are a simple consequence of the map extension property of ANR's. For example, if $i: C \hookrightarrow I^\infty$ is a given embedding satisfying $(*)$, and if $j: C \hookrightarrow W$ is any embedding of C into an ANR W, and U is any neighborhood of $j(C)$ in W, then the fact that W is an ANR implies that there exists some neighborhood V of $i(C)$ in I^∞ and a map $f: V \to U$ extending $ji^{-1}: i(C) \to j(C)$. So if $\alpha_t: C \to V$,

$0 \leq t \leq 1$, is a homotopy provided by $(*)$ such that $\alpha_0 = i$ and $\alpha_1(C) = \text{point}$, then $f\alpha_t : C \to U$ provides the desired null-homotopy of C in U.

The simplest examples of cell-like spaces are of course cells, that is, spaces which are homeomorphic to the closed unit ball in some euclidean space. More generally, any contractible compactum is cell-like. In fact, we have

REMARK. Suppose C is a compact ANR. Then C is cell-like $\Leftrightarrow$ C is contractible.

PROOF. To establish the implication $\Rightarrow$, suppose that $C \hookrightarrow I^\infty$ and let $r: U \to C$ be a retraction of a neighborhood, and let $\alpha_t: C \to U$ be a null-homotopy of C in U. Then $r\alpha_t: C \to C$ provides a contraction of C.

Thus the notion of cell-likeness can be regarded as a generalization of the notion of contractibility, and this notion is most useful for non-ANR's. An example of a noncontractible (hence non-ANR) cell-like compactum is the following planar *wedge* ($\equiv$ one-point-union) of two cones on cantor sets.

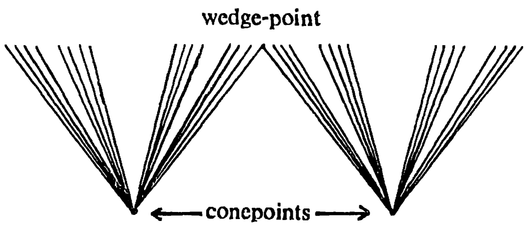


Figure 1. The wedge of two cones on cantor sets

One can construct many more interesting examples using the following

REMARK (Operations preserving cell-likeness).

(1) A countable null ($\equiv$ diameters tending to 0) wedge of cell-like spaces is cell-like.

(2) A product (finite or countable) of cell-like spaces is cell-like.

(3) The intersection of a countable nested collection of cell-like spaces is cell-like.

Regarding (3), note that the cell-like set pictured above is an intersection of 2-cell neighborhoods (when regarded as a subset of the plane).

In general, a cell-like space embedded in R^m (or S^m or any manifold) is said to be *cellularly embedded* (or *cellular*) if it has arbitrarily small neighborhoods homeomorphic to cells. This notion of cellularity definitely depends on the embedding. For example, the following picture (from [F—A]) shows an arc in R^3 which is not cellularly embedded there.

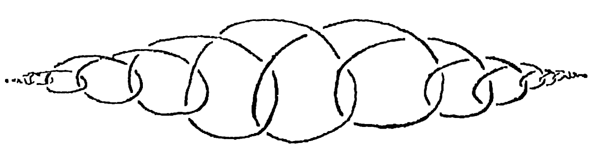


Figure 2. The Artin-Fox wild arc in R^3

Another example of a cell-like, noncellular subset of R^3 is the familiar horned ball of Antoine and Alexander, pictured here (the closed bounded region really is homeomorphic to a ball).

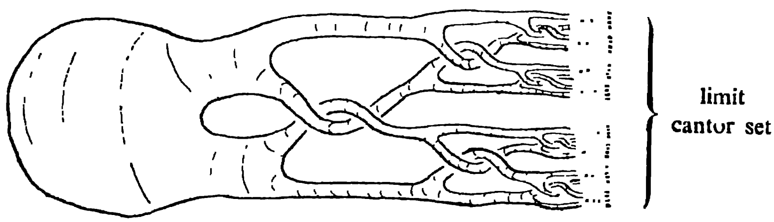


Figure 3. The Antoine-Alexander horned ball in $\mathbf{R}^3$

These examples show that cells can admit noncellular embeddings in euclidean space. In fact, any finite dimensional cell-like set (except a point) can be noncellularly embedded in any euclidean space of greater than twice its dimension. But more importantly, every finite dimensional cell-like set admits a cellular embedding in some euclidean space. In fact, if C is cell-like and $C \hookrightarrow R^m$, then $C \hookrightarrow R^{m+1}$ is cellular (deducible from [McM]).

This discussion leads to an answer to Introductory Question I (provided by M. Brown).

THEOREM. *A compact subset C of S^m is point-like (that is, $S^m - C \approx R^m$) $\Leftrightarrow$ C is cellular in S^m.*

SKETCH OF PROOF. The easier implication is $\Rightarrow$, for letting rB^m be an arbitrarily large ball in R^m, then $S^m - h(\operatorname{int} rB^m)$ is an arbitrarily small ball neighborhood of C in S^m (where $h: R^m \xrightarrow[\approx]{} S^m - C$ is the hypothesized homeomorphism). (Note: It is far from obvious that $S^m - h(\operatorname{int} rB^m)$ is a ball; in fact, this amounts to the Annulus Conjecture, which has been established in almost all dimensions through the efforts of many, especially R. Kirby, but remains unresolved for $m=4$. However, the above proof can be modified so as to circumvent this issue.) To prove the implication $\Leftarrow$, one first establishes, using the converse of the preceding argument, that any compact subset of $S^m - C$ lies in the interior of an m-cell in $S^m - C$. Then one shows that any space having this property is in fact homeomorphic to R^m. Details of these arguments are in [$\mathbf{Br}_1$] and [$\mathbf{Br}_2$].

Still, this theorem begs the question somewhat. How does one recognize a cellular subset of S^m? In the 2-sphere (or the plane) this is comparatively simple. A compact subset C of S^2 is cellular (and hence point-like) $\Leftrightarrow$ both C and $S^2 - C$ are nonempty and connected (i.e., C is a nonseparating continuum). This is an instructive exercise in plane topology. In higher dimensions, since being cellular always implies being cell-like, let us assume we can recognize when a subset $C \subset S^m$

is cell-like (inasmuch as this property is basically a homotopy property, C is often presented as a cell-like subset by the problem at hand). Given that C is cell-like, how does one tell whether it is cellular? (Remember the preceding examples.) A natural and useful condition to verify is whether $S^m - C$ is "simply-connected at infinity", i.e., whether given any neighborhood U of C in S^m, there exists a smaller neighborhood V of C such that the homomorphism $\pi_1(V-C)\to\pi_1(U-C)$ is trivial. This condition clearly is necessary when $m \geq 3$. It turns out that this *cellularity criterion* is also sufficient, at least when $m \neq 4$ [**McM**].

This result may be regarded as one of the prototypal theorems in the study of tame versus wild embeddings of compacta in manifolds, a subject which has developed into a very coherent theory during the last twenty years (e.g. see the brief surveys in [**La**$_2$] and [**Ed**$_1$]).

4. Examples and properties of cell-like maps. We now move on to cell-like maps. Probably the most useful characterization of a cell-like map is the following.

PROPOSITION (Homotopy characterization of cell-like maps). *Suppose $f\colon X\to Y$ is a proper surjection of* ANR's. *Then f is cell-like $\Leftrightarrow$ for each open subset U of Y, the restriction $f|\colon f^{-1}(U)\to U$ is a (proper) homotopy equivalence.*

Note: The parenthetical word (proper) can be inserted for $\Rightarrow$, and deleted for $\Leftarrow$, to provide the strongest statements.

To understand this proposition one should initially assume that X and Y are as nice as possible, e.g. manifolds. Consider first the implication $\Rightarrow$. As a special argument, consider establishing that $f_*\colon \pi_1(X)\to\pi_1(Y)$ is onto (which clearly it must be, if f is to be a homotopy equivalence). To verify this surjectivity, one takes a loop $\alpha\colon S^1\to Y$, and attempts to find a loop $\beta\colon S^1\to X$ such that $f\beta$ is homotopic to α (basepoints suppressed here). This is achieved by partitioning S^1 very finely, say by $\theta_1, \ldots, \theta_n$, and then arbitrarily choosing a point $\beta(\theta_i)\in f^{-1}(\theta_i)$ for each i. Now one attempts to join adjacent $\beta(\theta_i)$'s with paths which map under f to paths of small diameter. If the partition $\{\theta_i\}$ was chosen sufficiently fine, then each adjacent pair $\{\beta(\theta_i), \beta(\theta_{i+1})\}$ lies very close to some point-inverse $f^{-1}(y)$. Hence, by the definition of cell-likeness, $\beta(\theta_i)$ and $\beta(\theta_{i+1})$ can be joined by a path lying near $f^{-1}(y)$, since the null-homotopy of $f^{-1}(y)$ can be assumed to carry along a nearby neighborhood. Stringing these paths together gives a map $\beta\colon S^1\to X$ with the property that $f\beta$ is (pointwise) close to α, and hence homotopic to α. Thus f_* is surjective on π_1. This argument is a classical lifting argument which occurs over and over again topology. The full implication $\Rightarrow$ is a straightforward generalization.

To prove the implication $\Leftarrow$, we use the hypothesis that $y\in Y$ has a (arbitrarily small) contractible neighborhood U (assuming still for simplicity that Y is a manifold). Hence by the hypothesis, $f^{-1}(U)$ is a (arbitrarily small) contractible neighborhood of $f^{-1}(y)$. The general case, where Y is merely an ANR, is only slightly more complicated.

Cell-like maps are to be regarded as generalizations of homeomorphisms. This is a recurring theme. One important advantage that cell-like maps have over homeomorphisms is that (unlike homeomorphisms) they are closed under the operation of taking limits.

PROPOSITION (Operations with cell-like maps of ANR's). (1) *If a proper map* $f: X \to Y$ *of ANR's is approximable (as in* §2) *by cell-like maps, then* f *is cell-like.* (2) *The composition of cell-like maps of ANR's is cell-like.*

The proof is an interesting exercise, using the preceding proposition (one should assume X and Y are manifolds, at least initially). See [$\mathbf{La_1}$] and [$\mathbf{La_2}$].

We close this section by mentioning a classical theorem of R. L. Moore (as refined by Roberts–Steenrod and Youngs). The theorem essentially describes all possible cell-like maps defined on surfaces. (Those defined on 1-manifolds clearly are just those maps having point and interval point-inverses.)

THEOREM ([**Mo**], [**R–S**] and [**Yo**]). *Suppose* $f: M^2 \to Y$ *is a cell-like map defined on a closed surface* M^2 (*i.e. each* $f^{-1}(y)$ *is connected, and* $M^2 - f^{-1}(y)$ *is connected and has genus equal that of* M^2). *Then* Y *is also a surface, and* f *is approximable by homeomorphisms.*

The proof is a tour de force in plane topology.

5. Cell-like maps as limits of homeomorphisms. The preceding proposition implies that a near-homeomorphism is a cell-like map. When is the converse true? (Certainly not always, e.g. the map interval $\to$ point.)

It turns out to be natural to restrict attention to the case where the source is a manifold-without-boundary (possibly even a hilbert cube manifold). If in addition the target is also assumed to be a manifold, then we have the following fundamental answer.

THEOREM (for $m \leqq 2$ see above; m=3 Armentrout [$\mathbf{Ar_2}$]; $5 \leqslant m < \infty$ Siebenmann [**Si**]; $m = \infty$ Chapman [$\mathbf{Ch_1}$]; $m=4$ unknown). *Suppose* $f: M^m \to N^m$ *is a cell-like map (read cellular if* $m=3$) *of m-manifolds-without-boundary,* $m \neq 4$. *Then* f *is approximable by homeomorphisms.*

We now arrive at the focal point of this article, which is: What happens in the above theorem if N is not at the outset assumed to be a manifold (but M is)? Does the rest of the data (namely, that N is a cell-like image of a manifold) necessarily imply that N is a manifold? If $\dim M \leqslant 2$, then N is necessarily a manifold, by the Moore-et-al Theorem. But if $\dim M \geqslant 3$, then in fact N need not be a manifold, e.g., let $N = S^3$/Fox–Artin arc (see §3). So the problem becomes that of finding good conditions which ensure that N is a manifold.

Starting in the 1950's, a great deal of energy was put into understanding various special but important cases of this question. Most of the energy and insight was provided by R. H. Bing, whose pioneering work opened up the area and established a viable theory. Progress in the area was steady and remarkable, mostly at first in dimension 3, and later in higher dimensions. Here we will pass over all of these efforts, to concentrate in the next few sections on only the most recent developments.

6. The Approximation Problem. We cast our problem in the following form.

APPROXIMATION PROBLEM: *Suppose $f: M \to X$ is a cell-like map from a topological m-manifold-without-boundary onto an ANR X (possibly $m=\infty$ here, i.e., M may be a hilbert cube manifold). Find natural and useful conditions on X which guarantee that f be approximable by homeomorphisms.*

In light of the Moore–Armentrout–Siebenmann–Chapman Theorem above, this can be regarded as asking for conditions which guarantee that X be a manifold, at least if $\dim X \neq 4$. However, the reason for formulating it as an approximation problem will become clear, especially in §9.

The assumption here that X be an ANR is one of unfortunate necessity. If $m=\infty$, there exists an example of a cell-like map of the hilbert cube onto a non-ANR **[Ta]**. If $m<\infty$, it is not known whether such an X as in this problem need be an ANR. (This is a significant unresolved question; it is known to be equivalent to whether X is finite dimensional. In fact, this question is known to be equivalent to that of whether a compact metric space of finite cohomological dimension necessarily has finite (covering) dimension **[Ed$_5$]**.) At any rate, in most of the interesting situations which arise X is independently known to be an ANR.

7. The Approximation theorem in finite dimensions ≥ 5. In the next three sections, we restrict attention to finite dimensions.

In past years it was most often a certain special case of the Approximation Problem which was examined (as a rule), namely the stabilized case, where one asks whether $f: M \times R^1 \to X \times R^1$ is approximable by homeomorphisms. Progress on this special question was slow but steady, working in general on f's with increasingly pathological singularities. For example, one of the cases here which took a long time to resolve, eventually affirmatively, was the case where f has only one single nontrivial point-inverse.

A key step in recent years was made by J. Cannon, who turned the focus back to the pure, unstabilized question, by introducing in **[Ca$_2$]** as a workable X-condition the disjoint disc property. A space X has the *disjoint disc property* if, given any two maps $f_1, f_2: B^2 \to X$, there are arbitrarily close maps $g_1, g_2: B^2 \to X$ which have disjoint images. That is, two maps of the 2-disc into X can be general positioned apart. (Interestingly, a version of the disjoint disc property was used by Bing in his fundamental paper **[Bi$_3$]**.)

Cannon showed in **[Ca$_2$]** that, with regard to the Approximation Problem, if X has the disjoint disc property, and if X is already known to be a manifold except on a codimension ≥ 3 subset, then in fact f is approximable by homeomorphisms. This in turn inspired the following generalization.

Approximation Theorem [$\mathbf{Ed_3}$]. *Suppose $f: M^m \to X$ is a cell-like map from a topological m-manifold-without-boundary M onto an ANR X, and suppose $5 \leqslant m < \infty$. Then f is approximable by homeomorphisms $\Leftrightarrow$ X has the disjoint disc property.*

The proof is outlined in §9. For one thing, this theorem provides another proof of the Siebenmann Approximation Theorem (§5), inasmuch as that is not an ingredient.

8. Applications of the Approximation Theorem to the introductory questions II and III. The Polyhedral Manifold Question naturally arose in early attempts to understand triangulations of topological manifolds (e.g. see the historical discussions in [$\mathbf{Ca_1}$, §2] and [$\mathbf{Ed_2}$, Introduction]). Given a simplicial complex K which is topologically a manifold-without-boundary (i.e., K is topologically homogeneous), it is not too hard to establish via basic algebraic topology that K has the following two properties.

(1) For any (closed) simplex σ of K, the homology groups of the link of σ in K $\left(=lk(\sigma, K) \equiv \text{the collection } \{\tau\} \text{ of all closed simplexes such that } \tau \cap \sigma = \emptyset\right.$ and the span $\tau * \sigma$ is a simplex of $K\big)$ coincide with the homology groups of some sphere (in fact a sphere of dimension $\dim K - \dim \sigma - 1$), and

(2) In addition, if σ is a vertex and $\dim K > 2$, then $lk(\sigma, K)$ is simply connected.

Are these conditions sufficient to guarantee that K be a topological manifold? Yes, if $\dim K \leqslant 3$, but the situation becomes less clear in higher dimensions. The essence of this question turns out to be the:

Multiple Suspension Question. Suppose H^m is a homology m-sphere (defined below). For some $l \geqslant 2$, is it true that the lth suspension of H^m, $\Sigma^l H^m$ (which is the same as the join $S^{l-1} * H^m$), is topologically a manifold?

(If so, it is known to be a $(m+l)$-sphere, since it is necessarily covered by two coordinate patches. Hence, if the answer is yes for some l, it remains yes for any greater l.)

A *homology m-sphere* can be taken to be a topological m-manifold-without-boundary whose homology groups coincide with those of S^m. The $l=1$ case is passed over, because if H^m is not simply-connected (e.g., Poincaré's famous homology 3-sphere which has for fundamental group the 120-element binary dodecahedral group), then $\Sigma^1 H^m$ cannot possibly be a manifold at the two suspension points. The relation of this question to the preceding question is that H^m can be taken to be the link of some simplex σ in K, and $l = \dim \sigma + 1$, in which case a neighborhood of the open simplex $\mathring{\sigma}$ in K is homeomorphic to an open subset of $\Sigma^l H^m$ containing part of the suspension $(l-1)$-sphere.

An affirmative answer to the Multiple Suspension Question for any nonsimply-connected triangulated homology sphere would provide a non-combinatorial triangulation of a sphere, i.e., a triangulation which cannot be locally polyhedrally embedded in the euclidean space of the same dimension (c.f. Introductory Question II).

How does this tie in with the earlier discussion of cell-like maps? The connection is that it can be shown (without too much trouble, at least in most cases) that given any homology m-sphere H^m, and any $l \geqslant 2$, then there is a cell-like map $f\colon S^{m+l} \to \Sigma^l H^m$ from the $(m+l)$-sphere onto $\Sigma^l H^m$. Hence, in view of the earlier discussion, the Multiple Suspension Question boils down to whether this map f is approximable by homeomorphisms.

The Approximation Theorem of the preceding section provides an answer. The point is, the target space $\Sigma^l H^m$ has the disjoint disc property whenever $l \geqslant 2$ (recall without loss $m \geqslant 3$). One way to verify this is to show that given any map $f\colon B^2 \to \Sigma^l H^m$, there is an arbitrarily close map such that $f^{-1}(\Sigma^{l-1})$ has all of its components of arbitrarily small diameter, where Σ^{l-1} denotes the suspension $(l-1)$-sphere. Hence, if one started with two maps $f_1, f_2\colon B^2 \to \Sigma^l H^m$, then one could find nearby maps $g_1, g_2\colon B^2 \to \Sigma^l H^m$ such that $g_1(B^2) \cap g_2(B^2) \cap \Sigma^{l-1} = \emptyset$, by arranging the images of these components to be points and moving them to be disjoint. Then it is merely a matter of applying general position in the manifold $\Sigma^l H^m - \Sigma^{l-1} \approx H^m \times R^l$ to achieve that $g_1(B^2) \cap g_2(B^2) = \emptyset$.

Historically, the Multiple Suspension Question was answered affirmatively in almost all cases by the work described in [**Ed**$_2$], and it was completely settled by the subsequent work in [**Ca**$_2$]. The Approximation Theorem [**Ed**$_3$] came later. For the Multiple Suspension question itself, proofs have been improved now to the point where they are quite succinct (e.g. see [**Ca**$_3$]).

Regarding Introductory Question II, there is now a very satisfactory answer which follows as a consequence of the preceding work (excluding the unknown dimension 4).

POLYHEDRAL-TOPOLOGICAL MANIFOLD CHARACTERIZATION THEOREM. *A simplicial complex K of dimension $\neq 4$ is topologically a manifold-without-boundary if and only if conditions* (1) *and* (2) *above hold.*

Note that this theorem includes as a special case an affirmative answer to the Multiple Suspension Question, for all $l \geqslant 2$

It is worth pointing out that the related question of whether a given topological manifold is homeomorphic to some simplicial complex (i.e., the triangulation problem, in its broader form) is now known, as a consequence of the preceding work and the work of Galewski–Stern and Matumoto, to rest entirely on the question of whether certain homology 3-spheres bound acyclic 4-dimensional manifolds (see [G–S] or [**Ma**]). This problem is smack in the middle of an active area of research in low dimensional manifold topology.

Regarding Introductory Question III, the following corollary to the Approximation Theorem offers some insight.

COROLLARY. *Suppose an ANR X is a cell-like image of some topological manifold-without-boundary. Then $X \times R^2$ is itself a topological manifold.*

This follows by an argument of R. Daverman, who shows that $X \times R^2$ has the disjoint disc property (assuming without loss that $\dim X \geqslant 3$; observe that trivially $X \times R^5$ has the disjoint disc property, by applying general position in the R^5 coordinate). The corollary remains unresolved with R^2 replaced by R^1 (even if $\dim X \geqslant 4$, i.e., it is unknown whether $X \times R^1$ has the disjoint disc property).

Actually, with regard to Question II, there is a very natural and appealing:

CONJECTURE: A space X is a finite-dimensional-manifold factor $\Leftrightarrow$ X is an ENR homology manifold.

Being a homology m-manifold means that $H^*(X, X-x; \mathbf{Z}) \approx H^*(R^m, R^m - 0; \mathbf{Z})$ for each point $x \in X$. That a manifold factor X has this property is a straightforward consequence of Alexander duality.

There is further discussion of this conjecture in §12.

9. Sketch of the proof of the Approximation Theorem. The purpose of this section is to give some indication of the ideas, and their history, that go into the proof. We confine ourselves to the case where M and X are compact.

The first important point to make is that the proof uses the Bing Shrinking Criterion, which is a tool introduced by Bing almost three decades ago in [$\mathbf{Bi}_1$] for showing that a map is approximable by homeomorphisms. We discuss it only for compact spaces. This theorem was first stated in the following form by L. McAuley.

SHRINKING THEOREM. *A surjective map $f: X \to Y$ of compact metric spaces is approximable by homeomorphisms $\Leftrightarrow$ the following Bing Shrinking Criterion holds. Given any $\varepsilon > 0$, there is a homeomorphism $h: X \to X$ such that*

(1) $\operatorname{dist}(fh, f) < \varepsilon$ *and*

(2) *for each* $y \in Y$, $\operatorname{diam} h(f^{-1}(y)) < \varepsilon$.

Our proof will make use of the implication $\Leftarrow$. The reverse implication is mentioned only for completeness; it is quickly proved by letting $h = g_0^{-1} g_1$ for two successively chosen homeomorphisms g_0, g_1 approximating f. Concerning the implication $\Leftarrow$, it is worth presenting here a slick baire category proof (which is not the way the proof was originally discovered and developed). In the baire space $\mathscr{C}(M, X)$ of maps from M to X, with the uniform metric topology, let $\mathscr{E}$ be the closure of the set $\{fh \mid h: M \to M$ is a homeomorphism$\}$. The Bing Shrinking Criterion amounts to saying that for any $\varepsilon > 0$, the open subset of ε-maps in $\mathscr{E}$ ($\equiv$ maps having all point-inverses of diameter $< \varepsilon$), denoted $\mathscr{E}_\varepsilon$, is dense in $\mathscr{E}$. Hence $\mathscr{E}_0 \equiv \bigcap_{\varepsilon > 0} \mathscr{E}_\varepsilon$ is dense in $\mathscr{E}$, since $\mathscr{E}$ is a baire space. Since $\mathscr{E}_0$ consists of homeomorphisms, this shows that $f \in \mathscr{E}$ is approximable by homeomorphisms.

As a consequence of this discussion we see that, in order to prove the Approximation Theorem, it suffices to construct, for any given $\varepsilon > 0$, a homeomorphism $h: M \to M$ as described in the Shrinking Theorem.

The next basic point about the proof is that it proceeds more or less by induction. The idea is to filter the target X as

$$X = X^{(m)} \supset X^{(m-1)} \supset \ldots \supset X^{(1)} \supset X^{(0)} \supset X^{(-1)} = \emptyset,$$

where each $X^{(i)}$ is σ-compact ($\equiv$ a countable union of compacta) and $\dim(X^{(i)}-X^{(i-1)})=0$ (hence $\dim X^{(i)}=i$). Such a filtration is easy to find, and is common in dimension theory arguments. J. Cannon is probably most responsible for introducing the filtration method of argument into "shrinking theory".

Given the filtration, the idea is to take the given cell-like map f, and to approximate it by successively better cell-like maps $\{f_i\}$. Each f_i will have the property that it is 1-1 *over* $X^{(i)}$ (that is, the restriction $f_i|f_i^{-1}(X^{(i)})$ is 1-1). Of course, when we reach $i=m$ we are done.

Given a map $f\colon M\to X$, the *singular set* is the set $S(f)=\cup\{x\in X|f^{-1}(x)$ contains more than one point$\}$. Observe that $S(f)$ is σ-compact, because $S(f)=$ $=\bigcup_{\varepsilon>0}\{x\in X|\operatorname{diam} f^{-1}(x)\geqq\varepsilon\}$.

In going from $X^{(i-1)}$ to $X^{(i)}$, we make use of the following:

0-Dimensional Approximation Proposition. *Suppose $f\colon M\to Y$ is a cell-like map such that $S(f)$ is 0-dimensional (i.e., each compact subset of $S(f)$ is totally disconnected) and $S(f)$ is π_1-negligible in Y, that is, for each open set U in Y, $\pi_1(U-S(f))\to\pi_1(U)$ is an isomorphism. Then f is approximable by homeomorphisms.*

Discussion of Proof. One should consider first the simplest possible case, where $S(f)$ is a single point, say y. In this case $f^{-1}(y)$ satisfies the "cellularity criterion" (see §3), and so is cellular in M. Hence Y $(\approx M/\{f^{-1}(y)\sim\text{point}\})$ is a manifold, and f is approximable by homeomorphisms.

The model case of the Proposition, to which the general case is readily reduced, is the "countable null" case, in which $S(f)$ is countable (say $S(f)=\{y_1, y_2, \dots\}$), and $\operatorname{diam} f^{-1}(y_i)\to 0$ as $i\to\infty$. In this case the $f^{-1}(y_i)$'s can appear to be quite tangled together in M, but in reality they are not (e.g., it turns out that the preimages of any two subsets of $S(f)$ whose closures in Y are disjoint can be separated by a locally smooth $(m-1)$-sphere in $M-f^{-1}(S(f))$. In order to show that this f is approximable by homeomorphisms it suffices, according to the Shrinking Theorem, to find a homeomorphism $h\colon M\to M$, with fh close to f, such that each $h(f^{-1}(y_i))$ has small diameter. Inasmuch as there are only finitely many $f^{-1}(y_i)$'s bigger than any given size, this at first may seem an easy matter, but the difficulty is that in shrinking small a given $f^{-1}(y_i)$, one may inadvertently stretch larger some of the nearby $f^{-1}(y_j)$'s. To find the desired homeomorphism h, we generalize a 1950's argument of Bing, who (implicitly) in **[Bi$_2$]** constructed h for the case where each $f^{-1}(y_i)$ is geometrically a cone in some coordinate patch covering it. With a little bit of work, one can show that our more general $f^{-1}(y_i)$'s have sufficiently good conelike structure (after all, they are almost contractible) so that the Bing program can be made to succeed.

In order to be able to apply this Proposition, we assume an additional condition on the $X^{(i)}$'s, namely, that $X^{(m-3)}$ is π_1-negligible in X, and similarly that $X-X^{(2)}$

is π_1-negligible in X. This is exactly the point where the disjoint disc property of X is used.

Given these tools, the proof of the Approximation Theorem can be summarized quickly as follows. (Note: this is the baire category version of the original argument, and so the Shrinking Theorem appears here only implicitly.) Let $\mathscr{C}(M, X; X_0)$ denote the space of cell-like maps from M onto X which are 1-1 over $X_0 \subset X$, provided with the uniform metric topology (recall M, X are compact). If X_0 is σ-compact, then $\mathscr{C}(M, X; X_0)$ is a G_δ (hence baire) subspace of the baire space $\mathscr{C}(M, X)$, for if X_0 is compact, the set of maps in $\mathscr{C}(M, X)$ which are ε-maps over X_0 is open in $\mathscr{C}(M, X)$.

Our goal is to show that $\mathscr{C}(M, X; X)$ (= the homeomorphisms from M to X) is dense in $\mathscr{C}(M, X; \emptyset)=\mathscr{C}(M, X)$. To achieve this, it suffices to show that each $\mathscr{C}(M, X; X^{(i)})$ is dense in $\mathscr{C}(M, X; X^{(i-1)})$. Write $X^{(i)}=\bigcup_{j=1}^{\infty} X_j^{(i)}$, where each $X_j^{(i)}$ is compact. Then $\mathscr{C}(M, X; X^{(i)})=\bigcap_{j=1}^{\infty} \mathscr{C}(M, X; X^{(i-1)} \cup X_j^{(i)})$. By the baire property it suffices to show that for each j (and each i), $\mathscr{C}(M, X; X^{(i-1)} \cup X_j^{(i)})$ is dense in $\mathscr{C}(M, X; X^{(i-1)})$. This in turn is a straightforward application of the 0-Dimensional Approximation Proposition, thus: Given $g \in \mathscr{C}(M, X; X^{(i-1)})$, factor g as

$$g: M \underset{g_0}{\rightarrow} Y \underset{g_1}{\rightarrow} X,$$

where Y and g_i are defined by declaring that the nontrivial point-inverses of g_0 are precisely the nontrivial point-inverses of g which lie over $X_j^{(i)}$. That is, Y is the quotient space $Y=M/\{g^{-1}(x) \sim \text{point} \mid x \in X_j^{(i)}\}$. Then the quotient map $g_0\colon M \to Y$ has 0-dimensional singular set $S(g_0)$ which is π_1-negligible in Y, since either $g_1(S(g_0)) \subset X^{(m-3)}$ or else $g_1(S(g_0)) \subset X - X^{(2)}$. Now by the proposition g_0 is approximable by a homeomorphism, h_0 say. Then $g_1 h_0 \in \mathscr{C}(M, X; X^{(i-1)} \cup X_j^{(i)})$ and it approximates g.

10. Cell-like maps on hilbert cube manifolds. The preceding sections concentrated on cell-like maps of finite dimensional spaces. As already noted, the Approximation Problem (see §6) makes perfectly good sense even in the hilbert cube manifold setting. We repeat it here.

Approximation Problem: *Suppose $f\colon M \to X$ is a cell-like map from a hilbert cube manifold M onto an ANR X. Find natural and useful conditions on X which guarantee that f be approximable by homeomorphisms.*

Geometric topologists have recognized for several years now, thanks largely to the work of T. Chapman, that finite dimensional manifold questions often have worthwhile hilbert cube manifold analogues. These analogues are usually more pristine and more tractable, largely because of the homogeneity of the hilbert cube ($\equiv$ for all $x, y \in I^\infty$, there exists a homeomorphism of I^∞ carrying x to y; in particular, the hilbert cube has no "boundary") and the stability of the hilbert cube ($\equiv I^\infty \times I^\infty \approx I^\infty$). A specific example of such a problem is the stabilization problem for cell-like maps (which is a special case of the Approximation Problem): If $f\colon M\to X$ is a cell-like map from a hilbert cube manifold onto an ANR X, is it true that the stabilized map $f\times\mathrm{id}\,(I^\infty)\colon M\times I^\infty\to X\times I^\infty$ is approximable by homeomorphisms? This question gained significance after R. Miller established in 1974 **[Mi]** that for any ANR X, the product $X\times[0,\infty)$ is a cell-like image of a hilbert cube manifold (following which J. West **[We]** showed how to eliminate the $[0,\infty)$ factor). Consequently, to establish the longstanding Borsuk conjecture that an ANR crossed with the hilbert cube becomes a hilbert cube manifold, it became sufficient (and necessary, by Chapman's Approximation Theorem (§5)) to establish the above stabilization problem. This was accomplished in 1975 **[Ed$_4$]**, by making use of a Bing Shrinking Criterion argument (theretofore unexploited in infinite dimensional topology).

The following year H. Toruńczyk extended this work in striking fashion, to provide an attractive answer to the Approximation Problem. Completely independently of Cannon, Toruńczyk hit upon the disjoint cells property: A space X has the *disjoint cells property* if, given any two maps from an n-cell (n arbitrary) into X, there are two arbitrarily close maps having disjoint images. (For ANR's, this property has many interesting equivalent formulations, e.g., there exist two maps $i, j\colon X\to X$, each arbitrarily close to $\mathrm{id}\,(X)$, such that $i(X)\cap j(X)=\emptyset$.) Clearly for X to be a hilbert cube manifold this is a necessary condition. Toruńczyk established its sufficiency, again using a Bing Shrinking Criterion argument.

APPROXIMATION THEOREM (H. Toruńczyk **[To$_1$]**). *Suppose $f\colon M\to X$ is a cell-like map from a hilbert cube manifold M onto an ANR X. Then f is approximable by homeomorphisms $\Leftrightarrow$ X has the disjoint cells property.*

In light of the Miller–West theorem, one can drop the map f from the theorem, and assert the following.

I^∞-MANIFOLD CHARACTERIZATION THEOREM (H. Toruńczyk **[To$_1$]**). *An ANR X is a hilbert cube manifold $\Leftrightarrow$ X has the disjoint cells property.*

The significance of this theorem is in its applications, one of which is a satisfying proof of the following old conjecture.

COROLLARY (Schori–West [S–W], Curtis–Schori [C–S]). *Suppose X is a metric continuum ($\equiv$ compact and connected). Let 2^X [resp. $C(X)$] denote the space, provided with the hausdorff metric, of all closed [resp. closed and connected] subsets of X. Then*

(1) *2^X is homeomorphic to the hilbert cube $\Leftrightarrow$ X is locally connected, that is, X is a peano continuum, and*

(2) *$C(X)$ is homeomorphic to the hilbert cube $\Leftrightarrow$ X is a nondegenerate peano continuum and X contains no free arcs.*

The classical case of part (1) of this conjecture, solved by Schori–West, is the case $X=I$. It is a pleasant exercise to verify that 2^I has the disjoint cells property.

11. Analogues in hilbert space topology. In this section we mention briefly some very recent additional work of H. Toruńczyk [To$_2$], which grew out of his preceding work and some earlier work of his on non-locally-compact ANR's. In this section, all spaces are (possibly non-locally-compact) separable complete metric spaces.

The appropriate model manifold in the non-locally-compact infinite dimensional setting is hilbert space l_2 (here we use only its topological structure and so we could as well use its homeomorph $R^\infty \equiv \times_\infty R^1$, as established by R. D. Anderson). Toruńczyk found that in hilbert space topology the appropriate analogue of the disjoint cells property is the following: A space X has the *discrete cells property* if, given any map $f: D \to X$, where $D = +_{n=0}^{\infty} D^n$ is the disjoint union of n-cells $(0 \leq n < \infty)$, then there is an arbitrarily (uniformly) close map $g: D \to X$ so that the images of the D^n's comprise a disjoint, discrete (hence closed) collection of compacta in X. Using this, one has the

APPROXIMATION THEOREM (Toruńczyk). *Suppose $f: M \to X$ is a map from a hilbert space manifold M onto an ANR X. Then f is approximable by homeomorphisms $\Leftrightarrow$ f is a fine homotopy equivalence, and X has the discrete cells property.*

Here "approximable by homeomorphisms" means that given any $\varepsilon: X \to (0, \infty)$, there exists a homeomorphism $h: M \to X$ such that for all $z \in M$, $\operatorname{dist}(f(z), h(z)) < \varepsilon(f(z))$. The phrase "$f$ is a fine homotopy equivalence" means that given any $\varepsilon: X \to (0, \infty)$, there exists a map $g: X \to M$ such that fg is homotopic to $\operatorname{id}(X)$ by a homotopy whose motion is limited by ε, and similarly gf is homotopic to $\operatorname{id}(M)$ by a homotopy whose motion is limited under f by ε. We note that in this theorem f is not assumed to be any special kind of map (e.g. neither proper, nor closed, nor even a quotient map), so that for example the theorem applies to the (known) case of the projection map $l_2 \times l_2 \to l_2$. The above theorem is proved via a Bing shrinking argument.

Combining the above with his earlier proof that an ANR crossed with hilbert space becomes a hilbert space manifold, Toruńczyk obtained the impressive

HILBERT SPACE MANIFOLD CHARACTERIZATION THEOREM (Toruńczyk). *An ANR X is a hilbert space manifold $\Leftrightarrow$ X has the discrete cells property.*

An example of an interesting corollary of this theorem is the following.

COROLLARY. *A countably infinite product of AR's, infinitely many of which are noncompact, is homeomorphic to hilbert space l_2.*

Recall an AR (absolute retract) is nothing more than a contractible ANR. Toruńczyk extended these results to frechet manifolds of higher weights, too.

12. Characterizing topological manifolds. How does one characterize a finite dimensional topological manifold? (compare the nice infinite dimensional characterizations in §§10, 11). This is the kind of question to which there can be many useful answers. One conjecture in particular that is appealing, and ties in very strongly with the material in this article, has been enunciated by J. Cannon.

MANIFOLD CHARACTERIZATION CONJECTURE. Fix $5 \leqslant m < \infty$. A space X is an m-manifold-without-boundary $\Leftrightarrow X$ is an ENR homology m-manifold having the disjoint disc property.

This conjecture is a bit stronger than the one at the end of §8 (see the definitions there). Again, the forward implication above is well known. Interestingly, an affirmative solution of this conjecture was announced recently by F. Quinn.

13. Dimensions 3 **and** 4. Almost all of the results above exclude dimension 4, and many exclude dimension 3 as well. This is in part due to the lack of an appropriate analogue of the disjoint disc property in these dimensions, and also in part due to our ignorance of the topology of manifolds in these dimensions (particularly dimension 4). What is a good conjecture to make for the Approximation Problem, in either dimensions 3 or 4?

Dimension 4 is particularly bewildering. The difficulties there tie in with the difficulties already encountered by smooth manifold topologists working in that dimension on handlebody-structure related problems (e.g. surgery and s-cobordism theorems). As an example, it is not even known whether a cell-like surjection $f\colon M^4 \to N^4$ of closed 4-manifolds having exactly one non-trivial point inverse is approximable by homeomorphisms (this amounts to asking whether the nontrivial point-inverse is cellular in M^4). Or whether the cellularity criterion (§3) works in dimension 4. Questions such as these are in need of answers.

References

Some other articles which amplify the material of this article, and provide additional details, history and references, are [Ar_1], [Ca_1], [La_2], [McA, §§11, 12, 13]; [$\mathrm{Ca}_{2,3}$], [$\mathrm{Ed}_{2,3}$], [$\mathrm{To}_{1,2}$].

[An] R. D. Anderson, *Homeomorphisms on infinite-dimensional manifolds,* Actes, Congrès International des Mathématiciens, Nice 1970, Tome 2, 13—18.

[Ar_1] S. Armentrout, *Monotone decompositions of E^3*, Topology Seminar (Wisconsin, 1965), Ann. of Math. Studies, no. 60, Princeton Univ. Press, Princeton, N. J., 1966. pp. 1-25.

[Ar_2] ——, *Cellular decompositions of 3-manifolds that yield 3-manifolds,* Mem. Amer. Math. Soc. **107**, (1971).

[Bi_1] R. H. Bing, *A homeomorphism between the 3-sphere and the sum of two solid horned spheres,* Ann. of Math. **56** (1952), 354—362.

[Bi_2] ——, *Upper semicontinuous decompositions of E^3*, Ann. of Math. **65** (1957), 363—374.

[Bi_3] ——, *A decomposition of E^3 into points and tame arcs such that the decomposition space is topologically different from E^3*, Ann. of Math. **65** (1957), 484—500.

[Bi_4] ——, *The cartesian product of a certain non-manifold and a line is E^4*, Ann. of Math. **70** (1959), 399—412.

[Br$_1$] M. Brown, *A proof of the generalized Schoenflies theorem*, Bull. Amer. Math. Soc. **66** (1960) 74—76.

[Br$_2$] M. Brown, *The monotone union of open n-cells is an open n-cell*, Proc. Amer. Math. Soc. **12** (1961), 812—814.

[Ca$_1$] J. W. Cannon, *The recognition problem: What is a topological manifold?* Bull. Amer. Math. Soc. **84** (1978), 832—866.

[Ca$_2$] J. W. Cannon, *Shrinking cell-like decompositions of manifolds: codimension three*, Ann. of Math. **110** (1979), 83—112.

[Ca$_3$] *The characterization of topological manifolds of dimension $n \geqq 5$*, these Proceedings Helsinki, 1978.

[Ch$_1$] T. Chapman, *Cell-like mappings of Hilbert cube manifolds: applications to simple homotopy theory*, Bull. Amer. Math. Soc. **79** (1973), 1286—1291.

[Ch$_2$] *Topological classification of simple homotopy equivalences*, Proceedings of the International Congress of Mathematicians, Vancouver, 1974, Vol. 1, 509—513.

[C-S] D. W. Curtis and R. M. Schori, *Hyperspaces of polyhedra are Hilbert cubes*, Fund. Math. **99** (1978), 189—197. Announceg in Bull. Amer. Math. Soc. **80** (1974), 927—931.

[Ed$_1$] R. D. Edwards, *Demension theory*, Geometric Topology (Proceedings of the Park City Conference), L. C. Glaser and T. B. Rushing, eds., Lecture Notes in Math., no. 438, Springer-Verlag, Berlin and New York, 1975.

[Ed$_2$] *Suspensions of homology spheres*, Annals of Math., (to appear.)

[Ed$_3$] *Approximating certain cell-like maps by homeomorphisms*, preprint.

[Ed$_4$] *The ANR Theorem*, Chapter 14 in T. A. Chapman, Lectures on Hilbert Cube Manifolds, Regional Conference Series in Mathematics no 28, Amer. Math. Soc. 1976.

[Ed$_5$] *On the relation between cohomological dimension and covering dimension*, preprint.

[F-A] R. H. Fox and E. Artin, *Some wild cells and spheres in three-dimensional space*, Ann. of Math. **49** (1948), 979—990.

[G-S] D. E. Galewski and R. J. Stern, *Classification of simplicial triangulations of topological manifolds*, Ann. of Math. (to appear). (Announced in Bull. Amer. Math. Soc. **82** (1976), 916—918.)

[Hu] S.-T. Hu, *Theory of retracts*, Wayne State Univ. Press, 1965.

[La$_1$] R. C. Lacher, *Cell-like mappings*, I and II, Pacific J. Math. **30** (1969), 717—731; **35** (1970), 649—660.

[La$_2$] R. C. Lacher, *Cell-like mappings and their generalizations*, Bull. Amer. Math. Soc. **83** (1977), 495—552.

[Ma] T. Matumoto, *Variétés simpliciales d'homologie et variétés topologiques métrisables*, Thesis, University of Paris, Orsay, 1976.

[McA] L. F. McAuley, ed. *Some fundamental theorems and problems related to monotone mappings*, Proc. First Conf. on Monotone Mappings and Open Mappings, SUNY at Binghamton, Binghamton, N. Y., 1971. pp. 1—36.

[McM] D. R. Mc Millan, Jr., *A criterion for cellularity in a manifold*, Ann. of Math. **79** (1964), 327—337.

[Mi] R. T. Miller *Mapping cylinder neighborhoods of some ANR's*, Ann. of Math. **103** (1976), 417—427. Announced in Bull. Amer. Math. Soc. **79** (1973), 1286—1291.

[Mo] R. L. Moore, *Concerning upper semi-continuous collections of continua*, Trans. Amer. Math Soc. **27** (1925), 416—428.

[R-S] J. H. Roberts and N. E. Steenrod, *Monotone transformations of 2-dimensional manifolds* Ann. of Math. **39** (1938), 851—862.

[S-W] R. M. Schori and J. E. West, *The hyperspace of the closed unit interval is a Hilbert cube* Trans. Amer. Math. Soc. **213** (1975), 217—235. Announced in Bull. Amer. Math. Soc. **81** (1975), 187—188.

[Si] L. C. Siebenmann, *Approximating cellular maps by homeomorphisms*, Topology **11** (1973) 271—294.

[Ta] J. L. Taylor, *A counterexample in shape theory*, Bull. Amer. Math. Soc. **81** (1975), 629—632.

[To$_1$] H. Toruńczyk, *On CE-images of the Hilbert cube and characterization of Q-manifolds*, preprint.

[To$_2$] ——, *Characterizing Hilbert space topology*, Institute of Mathematics, Polish Academy of Sciences, Preprint 143, 1978.

[We$_1$] J. E. West, *Mapping Hilbert cube manifolds to ANR's: a solution of a conjecture of Borsuk*, Ann. of Math. **106** (1977), 1—18. Announced in Bull. Amer. Math. Soc. **81** (1975), 163—165.

[We$_2$] J. E. West, *Hilbert cube manifolds—meeting ground of geometric topology and absolute neighborhood retracts*, these Proceedings, Helsinki, 1978.

[Yo] J. W. T. Youngs, *Homeomorphic approximations to monotone mappings*, Duke Math. J. **15** (1948), 87—94.

UNIVERSITY OF CALIFORNIA AT LOS ANGELES
LOS ANGELES, CALIFORNIA 90024

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Classification of Finite Simple Groups

Daniel Gorenstein

My aim in this lecture will be to try to convince you that the classification of the finite simple groups is nearing its end. This is, of course, a presumptuous statement, since one does not normally announce theorems as "almost proved". But the classification of simple groups is unlike any other single theorem in the history of mathematics, since the final proof will cover at least 5,000 journal pages. Moreover, at the present time, perhaps 80% of these pages exist either in print or in preprint form. One obtains a better perspective of the subject if instead of thinking of the classification as a single theorem, one views it as an entire field of mathematics—the *structure of finite groups*. Then when I say that there are some 4,000 pages in print, proving many general and specific results about simple groups, it should sound entirely reasonable, since one can make the same claim concerning many areas of mathematics. Thus my task is really to convince you that we have established so many results about simple groups and have developed sufficient techniques for completing the classification.

There are other reasons for skepticism besides my premature announcement of the impending completion of the classification. Indeed, to the nonspecialist, simple group theory appears to be in a rather chaotic state. Strange sporadic simple groups dot the landscape—26 at last count; and they appear to be widely unrelated to each other. The five Mathieu groups, 100 years old, examples of highly transitive permutation groups, the four groups of Janko, each arising from the study of centralizers of involutions, the three Conway groups, determined from the automorphisms of a certain integral lattice in 24-dimensional Euclidean space, etc. And now there comes the Fischer-Griess monster, of order over 10^{53}; to be precise:

$$2^{46} \quad 3^{20} \quad 5^{9} \quad 7^{6} \quad 11^{2} \quad 13^{3} \quad 17 \quad 19 \quad 23 \quad 29 \quad 31 \quad 41 \quad 47 \quad 59 \quad 71.$$

And to add to the confusion, we don't even know whether the group exists! If it does, it involves, inside of itself, approximately 20 of the 26 sporadic groups. But whatever the case may be, it is clear that at present we have no coherent explanation of these sporadic groups. How then will it be possible to classify the simple groups in the face of this reality?

There is another troubling problem. Who will accept a 5,000 page proof when it exists? For it seems humanly impossible to avoid local errors in very long papers, and there is no doubt that there are many such errors in the existing 4,000 pages. Most of us have been rushing ahead towards the finish line with little time to look backwards; but it is clear that the first major "postclassification" problem will be a reexamination of the entire proof.

The fact is that the chaos in the subject is apparent rather than real. In searching for new simple groups, any plausible direction is worth exploring. It is much like experimental science and there is an element of the haphazard about the whole process. One can compare the discovery of a new simple group with that of an elementary particle in physics.

This is quite the opposite of the idea of classification, which implies something *systematic*. If one studies all simple groups G with some property X: for example,

Property X: (a) G has odd order.
(b) G has order at most 1,000,000.
(c) G has abelian Sylow 2-subgroups.
(d) The normalizer of every nontrivial solvable subgroup of G is solvable.

Then the analysis must uncover every simple group having the specified property X. The major thrust of simple group theory during the past 25 years has been the development of methods which enable us to determine all simple groups with some such property X. Each of the above listed four problems has indeed been solved. The first is, of course, the celebrated Feit-Thompson theorem which asserts that all groups of odd order are solvable—equivalently, that every (nonabelian) simple group has even order. In fact, it is this landmark theorem which started the whole show!

Primarily the methods for dealing with such general classification problems are *internal*. They involve the study of the proper subgroup structure of the simple group G under investigation. This point of view is fundamental. These methods have as their goal the following objective:

Prove that the internal structure of G closely resembles that of some known simple group G^*.

In the extreme, this can be taken to mean that G and G^* have identical lattices of proper subgroups. However, in practice, one does not require such complete similarity. Often it is entirely sufficient for a *single* subgroup of G to resemble the corresponding subgroup of G^*—for example, the centralizer of an *involution* (i.e., an element of order 2 in G).

We must emphasize that this internal resemblance of G to G^* may have nothing whatsoever to do with the way that the group G^* was initially discovered. For example, suppose one of Conway's groups C has the specified property X. Then the analysis must yield C as a possible answer. But to assert that, say, the centralizer of an involution of G is isomorphic to that of C has no connection with 24-dimensional integral lattices. It says nothing more than that the internal structure of G resembles that of the group C. Naturally we would like to be able to conclude from the given resemblance that G must, in fact, be isomorphic to C.

This leads us to the first major chapter of simple group theory, which must be resolved before one can attempt to prove any classification theorems whatsoever. It is called *Recognition Theory* and concerns the following general question:

If a simple group G has an internal structure closely resembling that of a known simple group G^*, must G be isomorphic to G^*?

If so, we say that the group G^* is *characterized* by the given set of internal conditions.

At the present time, essentially every known simple group possesses such an internal characterization. What are the known simple groups? Obviously I have no time to do more than simply list them here. They are the trivial groups Z_p, the 26 sporadic groups, the alternating groups of degree $n \geqslant 5$, and the so-called groups of Lie type. These last are the finite analogues of the complex Lie groups; thus we have finite analogues of the complex linear, symplectic, and orthogonal groups, and of the five exceptional Lie groups G_2, F_4, E_6, E_7, E_8, as well as finite analogues of the unitary groups. In the finite case, it turns out that there are somewhat more families than in the complex case. But in any event, we have a complete list of the finite simple groups of Lie type.

We can think of the linear groups as the typical example of a group of Lie type.

General linear group: $\mathrm{GL}(n, q)$ is the group of all nonsingular $n \times n$ matrices with coefficients in the Galois field $\mathrm{GF}(q)$ with q elements.

Special linear group: $\mathrm{SL}(n, q)$ is the normal subgroup of $\mathrm{GL}(n, q)$ of matrices of determinant 1.

Projective special linear group: $L_n(q) = \mathrm{PSL}(n, q)$ is the factor group of $\mathrm{SL}(n, q)$ modulo scalar matrices of determinant 1.

Fact: $L_n(q)$ is simple if $n \geqslant 3$ or if $n=2$ and $q \geqslant 4$.

We cannot expect to have an internal characterization of the Fischer-Griess monster, since we do not even know if it exists. The same is true of Janko's most recently discovered fourth group J_4. I should say that the problem here of existence and uniqueness of these two groups will be dealt with by a high-speed computer. There remains perhaps a little more theoretical work to do to set these problems up for the computer. However, the main question will be simply whether the present generation of computers is fast enough to make the required calculations.

Apart from these two groups, every other known simple group with the exception of a single family of groups of Lie type discovered by Rimhak Ree has such an internal characterization. Ultimately the characterizations of the groups of Lie type rests on Tits' geometric descriptions of these groups in terms of apartments and buildings or on the so-called Steinberg presentation in terms of generators and relations.

The Ree groups are a troublesome family. They have no complex analogue and they exist only in characteristic 3.

Ree group $R(q)$*:* order $q^3(q-1)(q^3+1)$, q an odd power of 3. Also $R(q)$ is a doubly transitive permutation group on q^3+1 letters and a subgroup fixing three letters has order 2.

Problem: Prove that the groups $R(q)$ are the only doubly transitive permutation groups of this order satisfying the given conditions.

Let $G(q)$ be an arbitrary such group. With great effort, Thompson has proved the following results:

(1) Associated with any such group $G(q)$ is an automorphism θ of the field GF (q);

(2) If θ^2 is the automorphism: $x \to x^3$, $x \in \mathrm{GF}(q)$, then $G(q) \cong R(q)$;

(3) For each value of q and θ, there is at most one group $G(q)$.

Open Question: Must θ^2 be the cubing map for the group $G(q)$ to exist?

Theoretically, therefore, there may exist new simple groups corresponding to other values of the parameter θ. A recent Ph. D. student of Suzuki has shown with the aid of a computer that no other groups than $R(q)$ exist for $q=3^n$ when $n \leq 29$. In any event, the ambiguity here does not bother us too much—we simply allow for this degree of indeterminacy by speaking of a group of *Ree type* as any group satisfying all the specified conditions.

Likewise we have groups of *monster type*. They are simple groups of the order I have written above and which have the various properties already established for the Fischer-Griess monster. Even though we do not know whether such a group exists, we allow for its existence in our analysis. Similarly we have a group of *type* J_4.

Subject then to these precise indeterminacies, this chapter of simple group theory is complete. This means that we are ready to begin the classification of the finite simple groups. However, we must emphasize that these ambiguities will remain even after our present classification theorem is completed. They should be viewed as isolated problems, which hopefully will eventually be settled.

As the classification has evolved, it has broken down into four major categories, as follows:

A. Nonconnected groups.

B. Groups of component type.

C. Small groups of noncomponent type.

D. The general group of noncomponent type.

In the balance of the talk, I shall attempt to outline the results obtained to date in categories A and B. This is all that time will permit. Fortunately, Michael Aschbacher in his lecture will describe the current state of affairs in categories C and D. Taken together, these two talks should give you a good idea of how close we actually are to completing the classification of the finite simple groups.

Let me explain the term *connectedness.* Given any group X, consider the collection $\mathscr{K}$ of Klein four subgroups of X; i.e., of subgroups of X isomorphic to $Z_2 \times Z_2$. Construct a graph Γ, whose vertices are the elements of $\mathscr{K}$. Connect two vertices A, B of Γ, if A and B commute elementwise; i.e., if $[A, B]=1$. The group X is said to be *connected* if the resulting graph Γ is connected in the usual sense. It is this degree of freedom which is needed to carry out certain general lines of argument.

The meaning of category A is the following: *Determine all nonconnected simple groups.* This chapter of simple group theory has been completed. However, it has taken some 3,000 journal pages to achieve. The proof has been carried out in two major parts:

I. Determine all simple groups which possess a nonconnected Sylow 2-subgroup.
II. Determine all nonconnected simple groups with a connected Sylow 2-subgroup.

Subgroups and homomorphic images of nonconnected groups may be connected, so nonconnectedness is not a good inductive concept. The solution of I has been obtained by treating it as a special case of a more general classification problem which is inductive. This is based on the following proposition.

PROPOSITION. *Let* S *be a nonconnected* 2*-group. If* A *is any subgroup of* S *and* $\bar{A}$ *is any homomorphic image of* A, *then* $\bar{A}$ *does not contain a subgroup isomorphic to* $Z_2 \times Z_2 \times Z_2 \times Z_2 \times Z_2$.

Such a group $\bar{A}$ is known as a *section* of S and so we rephrase the proposition by saying that a group with nonconnected Sylow 2-subgroup has *sectional 2-rank at most* 4. Thus I will be solved if we determine all simple groups of sectional 2-rank at most 4. The advantage of the latter condition is that it is preserved by subgroups and homomorphic images and so can be proved inductively. The resulting theorem will then stand in its own right, independent of whether the full classification is ever achieved. Most of the major results of simple group theory have a similar degree of independence.

I wish to state the sectional 2-rank $\leqslant 4$ theorem in its entirety, for the answer is instructive. You will have to accept the fact that each of the terms I write down stands for some specific groups of family of groups.

THEOREM. *If* G *is a simple group of sectional* 2*-rank at most* 4, *then* G *is isomorphic to one of the groups on the following list:*

I. *Odd characteristic:* $L_2(q)$, $L_3(q)$, $U_3(q)$, $G_2(q)$, ${}^3D_4(q)$, $\mathrm{Psp}\,(4, q)$, $L_4(q)$, $q \equiv 1 \pmod 8$, $U_4(q)$, $q \equiv 7 \pmod 8$, $L_5(q)$, $q \equiv 3 \pmod 4$, $U_5(q)$, $q \equiv 1 \pmod 4$, *or Ree type of characteristic* 3 *(Note the word "type" here).*

II. *Characteristic* 2: $L_2(8)$, $L_2(16)$, $L_3(4)$, $U_3(4)$, *or* $\mathrm{Sz}(8)$.
III. *Alternating:* A_n, $7 \leqslant n \leqslant 11$.
IV. *Sporadic:* M_{11}, M_{12}, M_{22}, M_{23}, J_1, J_2, J_3, Mc, *or* Ly

Thus, apart from certain families of groups of Lie type of odd characteristic of low dimension, there are precisely 19 other groups, half of them sporadic. You can see why the proof of this theorem must be a long one. If we think of each family as a single type of group, then there are some 30 distinct internal structures that can arise, 19 of them corresponding to individual groups. Thus our internal analysis of G must branch off into various directions, so that we can eventually show that G resembles internally one of these 30 types of groups. Each of these branches requires its own analysis.

Of course, groups of odd order correspond to "Case 0" of the theorem, which accounts for 250 pages of the argument!

To avoid repetition, we state the second part of the nonconnectedness theorem as follows:

THEOREM. *If G is a nonconnected simple group of sectional 2-rank at least 5, then G is isomorphic to one of the following groups:* $L_2(2^n)$, $U_3(2^n)$, *or* $\mathrm{Sz}(2^n)$.

Equivalently, G is of Lie type of characteristic 2 and "Lie rank 1." In particular, a Sylow 2-subgroup of G intersects its distinct conjugates only in the identity.

This last statement explains the structure of the graph Γ of G: each Sylow 2-subgroup of G corresponds to a distinct component of Γ.

The effect of having a complete solution to category A is that in all subsequent classification problems, one can assume at the outset that the group G under investigation is connected. I can only give the barest hint of the way this condition is used. Basically it helps us to analyze the *cores* of centralizers of involutions.

For any group X, the *core* of X is the unique largest normal subgroup of X of *odd* order. It is denoted by $O(X)$. By Feit-Thompson, cores are always solvable.

Fact. If G^* is a known simple group and t^* an involution of G^*, then $O(C_{G^*}(t^*))$ is a *cyclic* group.

Hence in studying arbitrary simple groups G and attempting to show that G internally resembles some known simple group, one of the first objectives is to prove that cores of centralizers of involutions of G are necessarily "small" (a cyclic group being a typical example of a small group). The methods we have developed for achieving this goal require G to be connected. This is all I can say here.

To describe the results in category B, I must now define a group of *component* type. To motivate the concept, let us examine briefly the general structure of the centralizer of an involution in a group of Lie type defined over $\mathrm{GF}(q)$, $q=p^n$, p a prime. As we shall see, we obtain quite distinct answers according as p is odd or $p=2$. This is, in fact, to be expected since in the Lie terminology an involution corresponds to a "semisimple" element when p is odd and to a "unipotent" element when $p=2$. We shall illustrate the situation using the groups $\mathrm{GL}(n, q)$.

odd characteristic

$$\textit{Involution } t\colon \begin{pmatrix} -1 & & & & & & \\ & -1 & {\scriptstyle k} & & & & \\ & & \ddots & & & & \\ & & & -1 & & & \\ & & & & 1 & {\scriptstyle n-k} & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \quad \textit{Centralizer } C_t\colon \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}.$$

Here A is nonsingular $k \times k$ and B is nonsingular $(n-k)\times(n-k)$.
Structure $C_t \cong \mathrm{GL}(k, q) \times \mathrm{GL}(n-k, q)$.

characteristic 2

$$\textit{Involution } t\colon \begin{pmatrix} 1 & & & \\ 0 & 1 & 0 & \\ \vdots & & \ddots & \\ 0 & & & \\ 1 & 0 \dots 0 & & 1 \end{pmatrix} \quad \textit{Centralizer } C_t\colon \left\{ \begin{pmatrix} x_{11} & 0 & \dots & 0 \\ x_{21} & & & \vdots \\ \vdots & & A & 0 \\ x_{n1} & x_{n2} & \dots & x_{nn} \end{pmatrix} \right\}.$$

Here A is nonsingular $(n-2)\times(n-2)$, $x_{11} \neq 1$, $x_{nn} \neq 1$.

$$\text{Define } Q = \left\{ \begin{pmatrix} 1 & & & 0 \\ x_{21} & 1 & & \\ & 0 & \ddots & \\ x_{n1} & x_{n2} & \dots & 1 \end{pmatrix} \right\} \quad \text{and} \quad K = \left\{ \begin{pmatrix} 1 & 0 \dots & & 0 \\ 0 & & A & \vdots \\ \vdots & & & 0 \\ 0 & & \dots 0 & 1 \end{pmatrix} \right\}$$

Q is a 2-group, $K \cong \mathrm{GL}(n-2, q)$.

Structure $C_t \sim Q \cdot K$; semidirect product; Q is normal in C_t; K acts faithfully on Q by conjugation.

We see then that when p is odd, the centralizer C_t is a product of groups of Lie type of lower dimension. Actually it is the $\mathrm{SL}(m, q)$ factors we are interested in rather than $\mathrm{GL}(m, q)$, for these are closer to being simple. In this example, each of these factors is normal, since the product is direct. However, in other groups, the centralizer may contain an element interchanging the factors, so these factors will only be what we call *subnormal.*

In general, a subgroup Y of a group X is called *subnormal* if there exists a chain of subgroups $Y = X_n, X_{n-1}, \dots, X_1 = X$ of X with each X_i normal in X_{i-1}.

On the other hand, when $p=2$, C_t has no such normal or subnormal subgroups of Lie type. The subgroup Q is an obstruction to the existence of such subgroups.

Only if one considers the factor group C_t/Q does one obtain a normal subgroup of Lie type.

This dichotomy is fundamental for understanding the general finite simple group, for it leads to a basic subdivision of simple groups into two distinct categories, one reflecting the odd characteristic phenomenon, the other the characteristic 2 phenomenon. To make the definition, we must take into account that the groups SL (m, q), need not be simple. Consider, for example, $X=\mathrm{SL}(2, q)$, q odd, $q \geqslant 5$; then $\left(\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right)$ is an element of X and commutes with every element of X, so is in the *center* of X. Hence certainly X is not simple. It is the factor group $X/\left\langle\left(\begin{smallmatrix}-1 & 0\\ 0 & -1\end{smallmatrix}\right)\right\rangle \cong L_2(q)$ which is simple. Thus X is what we refer to as a *covering* group of a simple group.

The more precise term is given by the following definition.

Definition. A group X is said to be *quasisimple* if X is perfect (i.e., $X=[X, X]$) and X (center of X) is simple.

In the study of simple groups G, we have already observed that the core $O(C_t)$, t an involution of G, also acts as an obstruction to any statement we may wish to make about the structure of the centralizers of an involution. Hence the definition of a group of component type must be formulated in terms of $C_t/O(C_t)$ rather than of C_t itself.

Definition. A group G is said to be of *component* type if for *some* involution t of G, $C_t/O(C_t)$ possesses a quasisimple subnormal subgroup. In the contrary case, G is said to be of *noncomponent* type.

Now we see the meaning of category B and the contrapositive categories C and D.

I shall now state the goal of much of the research of the past ten years. Again to avoid repetition, I shall assume that G has sectional 2-rank at least 5.

Theorem (?). *Let G be a simple group of component type (of sectional 2-rank at least 5) and assume that for some involution t of G, $\bar{C}_t=C_t/O(C_t)$ possesses a quasisimple subnormal subgroup $\bar{L}$ which is a covering group of a known simple group. Then one of the following holds:*

I. *G is of Lie type of odd characteristic (of sectional 2-rank at least 5);*

II. *$G \cong A_n$, $n \geqslant 12$; or*

III. *$G \cong$ one of the following 13 sporadic groups:* HS, ON, He, Suz, Ru, *Conway* .1 *or* .3, *Fischer $M(22)$, $M(23)$, $M(24)'$, the baby monster F_2, Harada's group F_5 (a subgroup of the monster), or G is of monster type.*

The (?) here is to indicate that the proof is not quite complete. At this time, there still exist certain possibilities for $\bar{L}$ for which it has not been established. Here is the present list of open cases:

characteristic 2. $\bar{L} \cong {}^2F_4(2)'$, ${}^2F_4(2^n)$, n odd, $n>1$, $F_4(2)$, Sp $(6, 2)$, $U_6(2)$, $O_8^{\pm}(2)$, or a covering group of Sp $(6, 2)$, $U_6(2)$, $O_8^{\pm}(2)$.

characteristic 3. $\bar{L} \cong U_3(3)$, $U_4(3)$, $L_4(3)$, $G_2(3)$.

sporadic. $\bar{L} \cong$ Conway .2 and Thompson F_3.

Thus there is a single family of groups plus 16 individual possibilities for $\bar{L}$. It should be emphasized that the open list has been steadily shrinking as group theorists tackle the remaining cases. Moreover, the methods for treating these problems are well understood. It is, of course, possible that one or more of these cases may lead to a new simple group. If so, each such new group as well as all of its covering groups would then have to be "plugged in" for $\bar{L}$. The same applies if a new simple group of noncomponent type is discovered in the future. However, what is possible and what is probable are two different matters. The most likely conjecture is that *every* finite simple group is now known and the remaining cases of the component theorem will be finished within approximately a year's time!

In conclusion, I would like to state a magnificent theorem of Aschbacher which characterizes the groups of Lie type of odd characteristic among the groups of component type and which is completely proved.

Suppose, in the above theorem, that the group $\bar{L} \cong \mathrm{SL}(2, q)$, q odd. Then $\bar{L}$ has a center of order 2. The involution $\bar{t}$ is certainly in the center of $\bar{C}_t$ and so $\bar{t}$ is a possible candidate for the involution in the center of $\bar{L}$. If $\bar{t}$ does lie in $\bar{L}$, we say that $\bar{L}$ is an *intrinsic* SL (2, q) and we call the involution t a *classical* involution.

THEOREM. *If G is a simple group which possesses a classical involution (and G has sectional 2-rank at least 5), then G is a group of Lie type of odd characteristic.*

This is a remarkable result because it asserts that the full structure of G is completely determined by a "tiny" piece of information in the centralizer of a single involution. It also shows the fundamental significance of the subgroups SL (2, q) for the structure of a group of Lie type. I think you will agree that the theory of finite simple groups must be quite fully developed for us to be in a position to establish such a powerful conclusion from so little information!

Aschbacher's lecture on groups of noncomponent type will indicate that our results for the groups in categories C and D are rapidly approaching the same degree of finality as presently exists for the groups of component type in category B. In fact, if and when all the present work in progress is completed, there will remain only a very few, essentially isolated, problems of the type described above, to complete the entire classification of finite simple groups (these do not include the Ree group problem and the question of the existence and uniqueness of both the monster and Janko's fourth group, which as I have tried to make clear, may very well remain unresolved after the classification).

RUTGERS UNIVERSITY
NEW BRUNSWICK, NEW JERSEY 08903, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Micro-Local Analysis

Masaki Kashiwara

We mean by micro-local analysis the analysis of functions and systems of differential equations on the cotangent bundle. The role of cotangent bundles in analysis has been recognized for a long time, but the formulation which we treat here started from Sato's introduction of microfunctions around 1970. The book of Sato–Kawai–Kashiwara [10] concerned was the systematic work on micro-local analysis. One of the most remarkable results in that book is the discovery of three types of micro-differential equations: de Rham type $\partial u/\partial x_1=0$, Cauchy–Riemann type $(\partial/\partial x_1+\sqrt{-1}\,\partial/\partial x_2)u=0$ and Lewy–Mizohata type $(\partial/\partial x_1+\sqrt{-1}x_1\,\partial/\partial x_2)u=0$. Any system of differential equations (or more generally micro-differential equations) is micro-locally equivalent to the mixture of these three types at a generic point.

They also proved that the characteristic variety of any system of differential equations is involutive. (See also [9].)

Let $P(x, D)=\sum_\alpha a_\alpha(x)D^\alpha$ be a differential operator defined on an open subset X of C^n. Here, $\alpha=(\alpha_1, \ldots, \alpha_n)$ is an n-tuple of non-negative integers, $|\alpha|=\alpha_1+\ldots+\alpha_n$ and $D^\alpha=\partial^{|\alpha|}/\partial x_1^{\alpha_1}\ldots\partial x_n^{\alpha_n}$. The largest m such that $a_\alpha(x)\neq 0$ for some α with $|\alpha|=m$ is called the order of $P(x, D)$. The function $\sum_{|\alpha|=m}a_\alpha(x)\xi^\alpha$ is called the principal symbol of $P(x, D)$. Here, $(x, \xi)=(x_1, \ldots, x_n, \xi_1, \ldots, \xi_n)$ is the coordinate system of the cotangent bundle T^*X of X and ξ^α means $\xi_1^{\alpha_1}, \ldots, \xi_n^{\alpha_n}$.

Let us consider a system of differential equations

$$\mathfrak{M} : P_1(x, D)u = \ldots = P_N(x, D)u = 0.$$

The common zeroes of the principal symbols of linear combinations $\sum_j A_jP_j$ with differential operators A_j as coefficients is called the characteristic variety of $\mathfrak{M}$.

Let c denote the codimension of the characteristic variety in the cotangent bundle. The characteristic variety being always involutive, the codimension c is equal to or greater than n. The number $n-c$ indicates, roughly speaking, the number of variables on which solutions of $\mathfrak{M}$ depend. For example, in the case of the system of differential equations $\partial u/\partial x_1=\ldots=\partial u/\partial x_c=0$, a solution $u(x)$ depends on the $(n-c)$ variables $x_{c+1},\ldots,x_n$. Hence, when $n=c$, we can expect that the space of solutions of $\mathfrak{M}$ is of finite dimension. In fact, this is a case [3]. We say that $\mathfrak{M}$ is holonomic if $c=n$. This notion is a generalization of the notion of ordinary differential equations into several variables. As we succeeded to study the properties of functions with one variable through their ordinary differential equations, we can expect that the study of holonomic systems of differential equations gives many properties of their solutions.

Let P be an ordinary differential operator of the form $\sum_{j=0}^m a_j(t)(t\,d/dt)^j$ with $a_m(0)\neq 0$. In this case, $t=0$ is called the regular singularity of $Pu=0$. Then, any solution of $Pu=0$ has the form:

$$u(t)=\sum_{j,\nu}\varphi_{j,\nu}(t)t^{\lambda_j}(\log t)^\nu$$

with holomorphic functions $\varphi_{j,\nu}(t)$ defined on a neighborhood of $t=0$. Moreover, λ_j are the solutions of the equation

$$\sum a_j(0)\lambda^j=0.$$

This phenomenon can be generalized to the case of holonomic systems of partial differential equations.

First, we can introduce the notion of regular singularities for holonomic systems (see §4). Let $\mathfrak{M}: P_1(x,D)u=\ldots=P_N(x,D)u=0$ be a holonomic system of differential equations with regular singularities. Then, we can find a nonzero polynomial $b(s)$ of degree m and a linear combination $Q=\sum A_jP_j$ such that Q is written in the form $b(x_1D_1)+x_1\sum_{j+|\alpha|\leq m}g_{j,\alpha}(x)(x_1D_1)^jD'^\alpha$, where $\alpha=(\alpha_2,\ldots,\alpha_n)$ is an $(n-1)$-tuple of nonnegative integers, $|\alpha|=\alpha_2+\ldots+\alpha_n$ and $D'^\alpha=\partial^{|\alpha|}/\partial x_2^{\alpha_2}\ldots\partial x_n^{\alpha_n}$. This implies that any solution of $\mathfrak{M}$ satisfies $Qu=0$. In the case of ordinary differential equation with regular singularities,

$$\sum_j a_j(t)\left(t\frac{d}{dt}\right)^j=\sum_j a_j(0)\left(t\frac{d}{dt}\right)^j$$
$$+t\sum_j t^{-1}(a_j(t)-a_j(0))\left(t\frac{d}{dt}\right)^j.$$

The hyperfunction solution $u(x)$ of the system $\mathfrak{M}$ has an "asymptotic expansion"

$$u(x)=\sum_{j=1}^r\sum_{\nu=0}^N\sum_{k=0}^\infty v_{j,\nu,k}(x')x_1^{\lambda_j+k}(\log x_1)^\nu.$$

Here, $v_{j,\nu,k}(x')$ are hyperfunctions on the variables $x'=(x_2,\ldots,x_n)$. λ_j satisfies

$b(\lambda_j)=0$. Moreover, the hyperfunctions $v_{j,\nu,k}(x')$ satisfy holonomic systems of differential equations with regular singularities.

1. Hyperfunctions and microfunctions [10]. Hyperfunctions are generalized functions obtained by sum of "ideal boundary values" of holomorphic functions, and microfunctions are "singular parts" of hyperfunctions. First, we shall remember the definitions of hyperfunctions and microfunctions.

1.1. *Tangent cone.* Let A and B be two subsets of a differential manifold X. For a point p of X, choosing a local coordinate system around p, we define the subset $C_p(A; B)$ of the tangent vector space T_pX as the totality of $\lim a_n(x_n-y_n)$, where $\{x_n\}$ (resp. $\{y_n\}$) is a sequence of points in A (resp. B) which converges p and $\{a_n\}$ is a sequence of positive numbers.

We define $C(A; B)$ as the union of $C_p(A; B)$.

If B is a closed submanifold of X, then $C(A; B)$ is invariant by translations of tangent vectors of B; we denote by $C_B(A)$ the subset $C(A; B)/TB$ of the normal bundle T_BX of B.

1.2. Let M be an open subset of $\boldsymbol{R}^n$ and let X be a neighborhood of M in $\boldsymbol{C}^n$. The tangent bundle TX (resp. TM) of X (resp. M) is identified with $X\times\boldsymbol{C}^n$ (resp. $M\times\boldsymbol{R}^n$), and the normal bundle T_MX of M is identified with $\sqrt{-1}TM=M\times\sqrt{-1}\boldsymbol{R}^n$. The conormal bundle T_M^*X of M is identified with $\sqrt{-1}\,T^*M$. We shall denote by τ and π the projection from $\sqrt{-1}\,TM$ and $\sqrt{-1}\,T^*M$ onto M, respectively. A subset Ω of $\sqrt{-1}\,TM$ (resp. $\sqrt{-1}\,T^*M$) is called convex or cone if $\Omega\cap\tau^{-1}(p)$ (resp. $\Omega\cap\pi^{-1}(p)$) is convex or cone. An open set U of X is called an *infinitesimal neighborhood* of a subset Ω of $\sqrt{-1}\,TM$ if $C_M(X-U)\cap\Omega=\emptyset$.

We denote by $\tilde{\mathfrak{A}}(\Omega)$ the inductive limit of the space $\mathcal{O}(U)$ of holomorphic functions defined on U, where U runs over the set of infinitesimal neighborhoods of Ω.

Let V be an open subset of M. Let $\mathscr{F}(V)$ be the set of families $\{\Omega_i, \varphi_i\}_{i\in I}$, where Ω_i is a convex open cone of $\sqrt{-1}\,TM$ such that $\tau(\Omega_i)=V$, φ_i is an element of $\tilde{\mathfrak{A}}(\Omega_i)$ and I is a finite index set. We say that two members $\{\Omega_i, \varphi_i\}_{i\in I}$ and $\{\Omega'_j, \varphi'_j\}_{j\in J}$ of $\mathscr{F}(V)$ are equivalent if there are open convex cones $\Omega_{i,j}$ and $\varphi_{i,j}\in\tilde{\mathfrak{A}}(\Omega_{i,j})$ $(i\in I, j\in J)$ satisfying the conditions: $\Omega_{i,j}\supset\Omega_i\cup\Omega'_j$, $\varphi_i=\sum_{j\in J}\varphi_{i,j}|_{\Omega_i}$ and $\varphi'_j=\sum_{i\in I}\varphi_{i,j}|\Omega'_j$.

We denote by $\mathscr{B}(V)$ the set of equivalence classes of $\mathscr{F}(V)$, and the element of $\mathscr{B}(V)$ is called a *hyperfunction* defined on V. For an open convex cone Ω and $\varphi\in\tilde{\mathfrak{A}}(\Omega)$, the hyperfunction corresponding to $\{\Omega, \varphi\}$ is denoted by $b_\Omega(\varphi)$ (or, if φ is defined on an infinitesimal neighborhood U of Ω, denoted by $b_U(\varphi)$, or simply $b(\varphi)$), and called the *boundary value* of φ. If $b_\Omega(\varphi)=0$, then $\varphi=0$. $\mathscr{B}(V)$ has clearly a structure of vector space, and the equivalence class of $\{\Omega_i, \varphi_i\}_{i\in I}$ equals $\sum_i b_{\Omega_i}(\varphi_i)$.

1.3. Let $(x_0, \sqrt{-1}\,\xi_0)$ be a point of the conormal bundle $\sqrt{-1}\,T^*M=M\times\sqrt{-1}\,\boldsymbol{R}^n$ of M, and let u be a hyperfunction defined in a neighborhood of x.

We say that u is *micro-analytic* if u can be expressed in the form: $u=\sum_i b_{\Omega_i}(\varphi_i)$ with open convex cones Ω_i of $\sqrt{-1}\,TM$ and $\varphi_i\in\tilde{\mathfrak{A}}(\Omega_i)$ such that $\tau^{-1}(p)\cap\Omega_i\subset\{\sqrt{-1}\,v\in\tau^{-1}(p)=\sqrt{-1}\,\boldsymbol{R}^n;\ \langle\sqrt{-1}\,v, \sqrt{-1}\,\xi_0\rangle=-\langle v, \xi_0\rangle>0\}$. We denote by $\mathrm{SS}(u)$ the set of points of $\sqrt{-1}\,T^*M$ where u is not micro-analytic, and call it the *singular spectrum* of u.

A real analytic function u defined on V is considered as hyperfunction on V; in fact, u is a restriction of a holomorphic function φ defined on a neighborhood U of V. u corresponds to the hyperfunction $b_U(\varphi)$. A hyperfunction u is real analytic if and only if $\mathrm{SS}(u)$ is contained in the zero section $\{(x, \sqrt{-1}\,\xi)\in\sqrt{-1}\,T^*M;\ \xi=0\}$.

1.4. For any open subset W of $\sqrt{-1}\,T^*M$, we define

$$\mathscr{C}'(W)=\mathscr{B}(V)/\{u\in\mathscr{B}(V);\ \mathrm{SS}(u)\cap W=\emptyset\},$$

where V is an open subset of M which contains $\pi(W)$. This definition does not depend on the choice of V. Let $\mathscr{C}$ be the sheaf on $\sqrt{-1}\,T^*M$ associated with the presheaf $W\mapsto\mathscr{C}'(W)$. If W is an open cone, we have $\mathscr{C}(W)=\mathscr{C}'(W)$. A section of $\mathscr{C}$ is called a *microfunction,* and we denote by sp the homomorphism from $\mathscr{B}(V)$ to $\mathscr{C}(W)$.

A differential operator $P(x, D)$ with real analytic functions as coefficients operates on the sheaf of hyperfunctions and microfunctions in the following way:

$$P\left(\sum b_{\Omega_j}(\varphi_j)\right)=\sum b_{\Omega_j}(P\varphi_j) \quad\text{and}\quad P\,\mathrm{sp}(u)=\mathrm{sp}(Pu).$$

2. Micro-differential operators. We can construct a class of operators, wider than the class of differential operators, which operate on the sheaf of microfunctions.

2.1. Let X be an open subset of $\boldsymbol{C}^n$. The cotangent bundle T^*X of X is identified with $X\times\boldsymbol{C}^n$. For a complex number λ and an open subset Ω of T^*X, we denote by $\mathscr{E}^{\infty(\lambda)}(\Omega)$ the set of sequences $\{P_{\lambda+j}(z, \zeta)\}_{j\in Z}$ of holomorphic functions satisfying the following conditions:

(2.1.1) $P_{\lambda+j}(z, \zeta)$ is a holomorphic function defined on Ω, homogeneous of degree $\lambda+j$ with respect to ζ; i.e.

$$\left(\sum_{\nu=1}^{n}\zeta_\nu\partial/\partial\zeta_\nu\right)P_{\lambda+j}=(\lambda+j)P_{\lambda+j}.$$

(2.1.2) $P_{\lambda+j}(z, \zeta)$ satisfies the following growth conditions:

(2.1.2.1) For any compact subset K of Ω and a positive number ε, there is a positive number $C_{K,\varepsilon}$ such that

$$|P_{\lambda+j}(z, \zeta)|\leqslant\frac{C_{K,\varepsilon}}{j!}\varepsilon^j \quad\text{for any}\quad j\geqslant 0 \text{ and } (z, \zeta) \text{ in } K.$$

(2.1.2.2) For any compact subset K of Ω, there is a positive number R_K such that

$$|P_{\lambda+j}(z, \zeta)|\leqslant(-j)!\,R_K^{-j} \quad\text{for any}\quad j<0 \text{ and } (z, \zeta) \text{ in } K.$$

We denote by $\mathscr{E}(\lambda)(\Omega)$ the set of $\{P_{\lambda+j}\}$ of $\mathscr{E}^{\infty(\lambda)}(\Omega)$ such that $P_{\lambda+j}=0$ for $j>0$. $\mathscr{E}^{\infty(\lambda)}$ and $\mathscr{E}(\lambda)$ are clearly sheaves on T^*X. We define $\mathscr{E}^{(\lambda)}$ as the union of $\mathscr{E}(\lambda+j)$, and set $\mathscr{E}^{\infty}=\mathscr{E}^{\infty(0)}$, $\mathscr{E}=\mathscr{E}^{(0)}$.

$\{P_{\lambda+j}(z,\zeta)\}_{j\in Z}$ will be denoted by $\sum P_{\lambda+j}(z,D)$. The section of $\mathscr{E}^{\infty(\lambda)}$ is called *micro-differential operator*, and the section of $\mathscr{E}(\lambda)$ will be called *micro-differential operator of order* λ.

For a section $\sum P_{\lambda+j}(z,D)$ of $\mathscr{E}(\lambda)$, $P_\lambda(z,\zeta)$ is called *principal symbol* and denoted by $\sigma_\lambda(P)$. Hence, σ_λ is a sheaf homomorphism from $\mathscr{E}(\lambda)$ onto the sheaf $\mathscr{O}(\lambda)$ of holomorphic functions on T^*X, homogeneous of degree λ with respect to ζ.

2.2. We define a product $R=\sum R_{\lambda+\mu+j}(z,D)$ of micro-differential operators $P=\sum P_{\lambda+j}(z,D)$ and $Q=\sum Q_{\mu+j}(z,D)$ by

$$R_{\lambda+\mu+l}(z,\zeta)=\sum_{l=j+k-|\alpha|}\frac{1}{\alpha!}\big(D_\zeta^\alpha P_{\lambda+j}(z,\zeta)\big)\big(D_z^\alpha Q_{\mu+k}(z,\zeta)\big),$$

where $\alpha=(\alpha_1,\dots,\alpha_n)$ is a set of non negative integers, $|\alpha|=\alpha_1+\dots+\alpha_n$ and $D_z^\alpha=\partial^{|\alpha|}/\partial z_1^{\alpha_1}\dots\partial z_n^{\alpha_n}$. By this law of multiplication, we have $(PQ)R=P(QR)$ for $P\in\mathscr{E}^{\infty(\lambda)}$, $Q\in\mathscr{E}^{\infty(\mu)}$ and $R\in\mathscr{E}^{\infty(\varrho)}$. We have $\mathscr{E}(\lambda)\cdot\mathscr{E}(\mu)\subset\mathscr{E}(\lambda+\mu)$. $\mathscr{E}^\infty,\mathscr{E}$ and $\mathscr{E}(0)$ are sheaves of rings with the identity $1=\sum P_j(z,D)$ ($P_j=1$ for $j=0$ and $P_j=0$ for $j\neq 0$). For $P\in\mathscr{E}(\lambda)$ and $Q\in\mathscr{E}(\mu)$, we have $\sigma_{\lambda+\mu}(PQ)=\sigma_\lambda(P)\sigma_\mu(Q)$.

A differential operator $P=\sum_\alpha a_\alpha(z)D^\alpha$ is identified with the micro-differential operator $\sum P_j(z,D)$ with $P_j(z,\zeta)=\sum_{j=|\alpha|}a_\alpha(z)\zeta^\alpha$.

The following lemma shows an importance of principal symbols.

Lemma 2.2.1. *If the principal symbol $\sigma_\lambda(P)$ of $P\in\mathscr{E}(\lambda)$ does not vanish at a point p of T^*X, then P has an inverse R in $\mathscr{E}(-\lambda)$; i.e. $RP=PR=1$.*

The rings $\mathscr{E}$ and $\mathscr{E}(0)$ have nice ring theoretic properties; for example, they are coherent rings, any stalk of them is a noetherian ring, etc.

2.3. Let M be an intersection of R^n and an open subset X of C^n. Then, the conormal bundle $T_M^*X=\sqrt{-1}\,T^*M$ is a closed submanifold of T^*X. Let V be an open subset of $\sqrt{-1}\,T^*M$ and let Ω be an open subset of T^*X containing V. Then, $\mathscr{E}^{\infty(\lambda)}(\Omega)$ operates on $\mathscr{C}(V)$, and we have $1\cdot u=u$, $(PQ)u=P(Qu)$ for $u\in\mathscr{C}(V)$, $P\in\mathscr{E}^{\infty(\lambda)}(\Omega)$, $Q\in\mathscr{E}^{\infty(\mu)}(\Omega)$.

By using this operation of micro-differential operators, the following proposition is easily obtained from Lemma 2.2.1.

Proposition 2.3.1. *For a differential operator P and a hyperfunction u, we have*

$$\mathrm{SS}(u)\subset\sigma(P)^{-1}(0)\cup\mathrm{SS}(Pu).$$

In fact, outside $\sigma(P)^{-1}(0)\cup\mathrm{SS}(u)$, P is invertible as micro-differential operator by Lemma 2.2.1, and hence $P\,\mathrm{sp}(u)=0$ implies $\mathrm{sp}(u)=0$.

2.4. Let $p_0=(z_0,\zeta_0)$ and $p_1=(z_1,\zeta_1)$ be two points of T^*C^n and F a holomorphic map from a neighborhood of p_0 to a neighborhood of p_1 such that

$F(p_0)=p_1$. Assume that F is a homogeneous symplectic transformation, i.e. $F^*(\sum\zeta_j dz_j)=\sum\zeta_j dz_j$. Then, we can construct a sheaf isomorphism $\Phi: F^{-1}(\mathscr{E}^{\infty(\lambda)})\xrightarrow{\sim}\mathscr{E}^{\infty(\lambda)}$ on a neighborhood of p_1 such that $\Phi(PQ)=\Phi(P)\Phi(Q)$ for $P\in\mathscr{E}^{\infty(\lambda)}$ and $Q\in\mathscr{E}^{\infty(\mu)}$, and that $\sigma_\lambda(\Phi(P))$ $\sigma_\lambda(P)\circ F$ for any $P\in\mathscr{E}(\lambda)$. Moreover if p_0 and p_1 are contained in $\sqrt{-1}\,T^*M$ and if F maps $\sqrt{-1}\,T^*M$ into $\sqrt{-1}\,T^*M$, then we can construct a sheaf isomorphism $\Psi: F^{-1}(\mathscr{C})\xrightarrow{\sim}\mathscr{C}$ such that $\Psi(Pu)=\Phi(P)\cdot\Psi(u)$ for any $P\in\mathscr{E}^{\infty(\lambda)}$ and $u\in\mathscr{C}$. We shall call (Φ, F) or (Φ, Ψ, F) the *quantized contact transformation.*

Quantized contact transformations are effectively used in order to transform micro-differential operators into the normal forms. For example, let us consider a micro-differential operator P of order 1 such that $\sigma_1(P)\wedge(\sum\zeta_j dz_j)\neq 0$ and that $\sigma_1(P)|_{\sqrt{-1}\,T^*R^n}$ is real-valued. Then, there is a quantized contact transformation which transforms P into $\sqrt{-1}\partial/\partial z_1$. Hence, we can deduce the properties of P from those of $\partial/\partial z_1$, which is easy to analyze. Thus, we obtain the following

PROPOSITION 2.4.1. *Let P be a micro-differential operator such that the restriction of the principal symbol p of P to $\sqrt{-1}\,T^*\mathbf{R}^n$ is real valued and that the differential of the principal symbol is not parallel to $\sum\zeta_j dz_j$. Then, $P: \mathscr{C}\to\mathscr{C}$ is surjective and the support of microfunction solution of $Pu=0$ is the union of bicharacteristics of P, i.e. integral curves of the Hamiltonian*

$$\sum_j \frac{\partial p}{\partial\zeta_j}\frac{\partial}{\partial z_j}-\frac{\partial p}{\partial z_j}\frac{\partial}{\partial\zeta_j}.$$

3. System of micro-differential equations.

3.1. In this talk, a *system of micro-differential equations* is, by definition, a coherent left $\mathscr{E}$-Module. Let $\mathfrak{M}$ be a coherent $\mathscr{E}$-Module. Then, $\mathfrak{M}$ has locally a free resolution

$$0\leftarrow\mathfrak{M}\leftarrow\mathscr{E}^{N_0}\xleftarrow{P}\mathscr{E}^{N_1},$$

where $P=(P_{ij})$ is an $N_1\times N_0$ matrix of micro-differential operators and $P: \mathscr{E}^{N_1}\to\mathscr{E}^N$ is given by $(Q_1, \dots, Q_{N_1})\mapsto(Q_1, \dots, Q_{N_1})P=(\sum Q_i P_{i1}, \dots, \sum Q_i P_{iN_0})$.

We have $\mathscr{H}om_{\mathscr{E}}(\mathfrak{M}, \mathscr{C})=\mathrm{Ker}\,(\mathscr{C}^{N_0}\xrightarrow{P}\mathscr{C}^{N_1})$, and hence $\mathscr{H}om_{\mathscr{E}}(\mathfrak{M}, \mathscr{C})$ is a sheaf of microfunction solutions of a system of micro-differential equations:

$$\sum_{j=1}^{N_0} P_{ij}u_j=0 \ \ (i=1, \dots, N_1).$$

3.2. The support of a coherent $\mathscr{E}$-module is not an arbitrary subset of T^*M.

THEOREM 3.2.1 ([9], [10]). *The support of a coherent $\mathscr{E}$-module is an involutive analytic subset of T^*X.*

Remember that an analytic subset V of T^*X is called *involutive* if the Poisson bracket

$$\{f, g\}=\sum_j\left(\frac{\partial f}{\partial\zeta_j}\frac{\partial g}{\partial z_j}-\frac{\partial g}{\partial\zeta_j}\frac{\partial f}{\partial z_j}\right)$$

vanishes on V for any holomorphic functions f and g vanishing on V. An involutive analytic subset has always codimension equal or greater than $n=\dim X$. We say an involutive analytic subset V is *Lagrangian* if $\operatorname{codim} V=\dim X$, and a system of micro-differential equations is *holonomic* if its support is Lagrangian.

THEOREM 3.2.2 ([6], [8]). *Let $\mathfrak{M}$ be a holonomic system of micro-differential equations. Then, we have*

(1) *For any point p of $\sqrt{-1}\,T^*\boldsymbol{R}^n$, $\mathcal{H}om_{\mathcal{E}}(\mathfrak{M}, \mathcal{C})_p$ is a finite-dimensional vector space. More generally, so is $\mathcal{E}xt^j_{\mathcal{E}}(\mathfrak{M}, \mathcal{C})_p$ for any j.*

(2) *There is a stratification $\sqrt{-1}\,T^*\boldsymbol{R}^n=\sqcup V_\alpha$ of $\sqrt{-1}\,T^*\boldsymbol{R}^n$ into subanalytic submanifolds V_α such that $\mathcal{E}xt^j_{\mathcal{E}}(\mathfrak{M}, \mathcal{C})|_{V_\alpha}$ is a locally constant sheaf for any α and j.*

A microfunction which satisfies a holonomic system of micro-differential equations is called *holonomic microfunction*. Theorem 3.2.2 suggests that we have a great chance to analyze holonomic microfunctions through their holonomic systems of micro-differential equations.

4. Holonomic system with regular singularities.

4.1. Let $\mathfrak{M}$ be a holonomic system of micro-differential equations and let Λ be the support of $\mathfrak{M}$. Λ is therefore a Lagrangian analytic subset of T^*X. At a generic point p of Λ, Λ is a conormal bundle of a complex submanifold Y of X. Let us take a local coordinate system $x=(x_1, \dots, x_n)$, such that Y is given by $x_1=\dots=x_l=0$ and $p=(0, dx_1)$. Let $\mathcal{L}_{\lambda,m}$ be the holonomic system defined by

$$\mathcal{L}_{\lambda,m}=\mathcal{E}/\mathcal{E}(x_1D_1-\lambda)^m+\mathcal{E}x_2+\dots+\mathcal{E}x_l+\mathcal{E}D_{l+1}+\dots+\mathcal{E}D_n=\mathcal{E}u_{\lambda,m}.$$

Here, $u_{\lambda,m}$ is the modulo class of 1.

THEOREM 4.1.1. *There are complex numbers λ_j and integers m_j $(j=1, \dots, N)$ such that $\mathcal{E}^\infty\otimes_{\mathcal{E}}\mathfrak{M}$ is isomorphic to $\oplus\mathcal{E}^\infty\otimes_{\mathcal{E}}\mathcal{L}_{\lambda_j,m_j}$. $(\lambda_j \bmod \boldsymbol{Z}, m_j)$ are uniquely determined by $\mathfrak{M}$ up to permutation.*

The integer $\sum m_j$ is called the *multiplicity* of $\mathfrak{M}$.

We say that $\mathfrak{M}$ has *regular singularities* at p if $\mathfrak{M}$ is isomorphic to $\bigoplus_j \mathcal{L}_{\lambda_j,m_j}$. In general, we say that $\mathfrak{M}$ has *regular singularities,* if $\mathfrak{M}$ has regular singularities at a generic point of any irreducible component of the support of $\mathfrak{M}$.

Any holonomic system can be transformed into a holonomic system with regular singularities by using micro-differential operators of infinite order. More precisely, we have the following

THEOREM 4.1.2 [7]. *For any holonomic system $\mathfrak{M}$, there exists a holonomic system $\mathfrak{M}'$ with regular singularities such that $\mathcal{E}^\infty\otimes_{\mathcal{E}}\mathfrak{M}$ is isomorphic to $\mathcal{E}^\infty\otimes_{\mathcal{E}}\mathfrak{M}'$. Moreover, $\mathfrak{M}'$ is unique up to isomorphism.*

4.2. Let Y be a complexification of a real analytic submanifold of $\boldsymbol{R}^n$, and let $\mathscr{L}_{\lambda,m}$ be as in § 4.1. Then, any homomorphism from $\mathscr{L}_{\lambda,1}$ into the sheaf of micro-functions $\mathscr{C}$ is given by $u_{\lambda,1}\mapsto c(D_1/\sqrt{-1})^{-\lambda-1}\delta(x_1, \dots, x_l)$ for some complex number c. Let $\mathfrak{M}$ be a holonomic system with multiplicity 1 (this implies that $\mathfrak{M}$ has regular singularities), and suppose that $\mathfrak{M}$ is generated by a section u. Then, $\mathfrak{M}$ is isomorphic to $\mathscr{L}_{\lambda,1}$ for some complex number λ, and the isomorphism from $\mathfrak{M}$ onto $\mathscr{L}_{\lambda,1}$ is given by $u\mapsto Pu_{\lambda,1}$ with a micro-differential operator $P\in\mathscr{E}(k)$ (for some integer k) such that $\sigma_k(P)|_\Lambda\not\equiv 0$. Hence, for any $F\in\mathscr{H}om_{\mathscr{E}}(\mathfrak{M}, \mathscr{C})$, $F(u)$ equals $cP(D_1/\sqrt{-1})^{-\lambda-1}\delta(x_1, \dots, x_l)$ for a complex number c. We define the *principal symbol* $\sigma(F(u))$ of $F(u)$ by

$$\sigma(F(u)) = \frac{1}{(2\pi)^{l/2}}c(\sigma_k(P)|_\Lambda)\xi_1^{-\lambda-1}\sqrt{|d\xi_1 \dots d\xi_l\, dx_{l+1} \dots dx_n|/|dx_1 \dots dx_n|}$$

regarded as a section of $\Omega_\Lambda^{1/2}\otimes(\Omega_M^{1/2})^{\otimes(-1)}$. Here $\Omega_\Lambda^{1/2}$ is the sheaf of half densities on Λ. For a holonomic microfunction which satisfies a holonomic system with regular singularities with multiplicity 1, we can define its principal symbol in this manner (see [2]).

The above observation shows that a solution $F\colon \mathfrak{M}\mapsto\mathscr{C}$ is uniquely determined by the principal symbol of $F(u)$.

5. Asymptotic expansion of holonomic microfunctions.

5.1. Let us consider the situation in § 4 with $l=1$. Then, $F(u)$ is given by $P(x, D)(D_1/\sqrt{-1})^{-\lambda-1}\delta(x_1)$. For the sake of simplicity let us assume that λ is not an integer. Then, we can show that $F(u)$ equals $\mathrm{sp}(\varphi(x)(x_1+\sqrt{-1}\,0)^\mu)$ with $\mu=\lambda-k$ and a real analytic function $\varphi(x)$. We have

$$\sigma(F(u)) = \frac{\sqrt{2\pi}\xi_1^{-\mu-1}}{\Gamma(-\mu)\exp(-\pi i\mu/2)}\varphi(0, x_2, \dots, x_n)\sqrt{|d\xi_1\, dx_2 \dots dx_n|/|dx_1 \dots dx_n|}.$$

Since

$$\varphi(x)(x_1+i0)^\mu = \varphi(0, x_2, \dots, x_n)(x_1+i0)^\mu + \frac{\partial\varphi}{\partial x_1}(0, x_2, \dots, x_n)(x_1+i0)^{\mu+1}+\dots,$$

$\sigma(F(u))$ gives the first coefficient of the power series expansion of $F(u)$ with respect to x_1.

5.2. Let us consider another example. Let $f(x)$ be a real-valued real analytic function. Then, it is known that $\delta(t-f(x))$ has the asymptotic expansion

$$\delta(t-f(x)) \sim \sum_{\nu=0}^{n-1}\sum_{j=1}^{N}\sum_{k=0}^{\infty} a_{j,\nu,k}(x)t^{\lambda_j+k}(\log t)^\nu \quad (t\searrow 0).$$

Here, $a_{j,\nu,k}(x)$ is a distribution related with the residues of the analytic continuation of $f(x)_+^s$ with respect to s.

The above asymptotic expansion means the following: for any compactly supported C^∞-function $\varphi(x)$, we have the asymptotic expansion

$$\int \delta(t-f(x))\varphi(x)\,dx \sim \sum_\nu \sum_j \sum_k \left(\int a_{j,\nu,k}(x)\varphi(x)\,dx\right) t^{\lambda_j+k}(\log t)^\nu.$$

5.3. Let us consider a more general case. Let X be an open subset of C^{1+n}. We shall denote by $(t, x)=(t, x_1, \ldots, x_n)$ the point of C^{1+n}. Let $\mathscr{D}$ be the sheaf of differential operators on X. A left coherent $\mathscr{D}$-module is called a *system of differential equations*. Let us denote by π the projection from T^*X onto X. Then, $\mathscr{E}$ contains $\pi^{-1}\mathscr{D}$ as its subring. For a system $\mathfrak{M}$ of differential equations, the support of $\mathscr{E}\otimes_{\pi^{-1}\mathscr{D}}\pi^{-1}\mathfrak{M}$ is called the *characteristic variety* of $\mathfrak{M}$.

We say that $\mathfrak{M}$ is *holonomic* if so is $\mathscr{E}\otimes_{\pi^{-1}\mathscr{D}}\pi^{-1}\mathfrak{M}$.

Let $\mathfrak{M}$ be a holonomic system of differential equations, u a section of $\mathfrak{M}$, and let F be a $\mathscr{D}$-linear homomorphism from $\mathfrak{M}$ into the sheaf $\mathscr{B}$ of hyperfunctions on $M=X\cap R^{1+n}$.

Suppose that $F(u)$ has an asymptotic expansion

$$F(u) \sim \sum_j \sum_\nu \sum_k a_{j,\nu,k}(x)\,t^{\lambda_j+k}(\log t)^\nu. \tag{5.3.1}$$

We may assume that $\lambda_j-\lambda_{j'}$ is not an integer for $j\neq j'$ and that for any j there is ν satisfying $a_{j,\nu,0}\neq 0$.

First, let us determine λ_j.

THEOREM 5.3.1. *There exist a non zero polynomial $b(s)$ and a differential operator $P(t, x, tD_t, D_x)$ satisfying the following conditions*

(1) $b(tD_t) = tP(t, x, tD_t, D_x)u$,

(2) $P(t, x, tD_t, D_x)$ *has the form* $\sum_{j,\alpha} g_{j,\alpha}(t, x)(tD_t)^j D_x^\alpha$.

This theorem gives λ_j. We have

$$b(\lambda_j) = 0 \quad \text{for any } j. \tag{5.3.2}$$

In fact, by (1) of Theorem 5.3.1, we have

$$\big(b(tD_t)-tP(t, x, tD_t, D_x)\big)\sum_j \sum_\nu \sum_k a_{j,\nu,k}(x)\,t^{\lambda_j+k}(\log t)^\nu = 0. \tag{5.3.3}$$

If $a_{j,\nu_0,0}=0$ for $\nu>\nu_0$ and $a_{j,\nu_0,0}\neq 0$, then the coefficient of $t^{\lambda_j}(\log t)^{\nu_0}$ of the expansion of the left hand side of (5.3.3) is $b(\lambda_j)a_{j,\nu_0,0}(x)$. Hence, we obtain $b(\lambda_j)=0$.

We shall call $b(s)$ the *b-function* of u with respect to the hypersurface $t=0$.

EXAMPLE 5.3.2. $u=(t^2-x^3)^\alpha$, i.e.

$$\left(\frac{1}{2}tD_t+\frac{1}{3}xD_x-\alpha\right)u = (3x^2D_t+2tD_x)u = 0.$$

We have

$$(tD_t)\left(tD_t-2\alpha-\frac{2}{3}\right)\left(tD_t-2\alpha-\frac{4}{3}\right)u=-\frac{8}{27}t^2D_x^3u,$$

and hence $b(s)=s(s-2\alpha-2/3)(s-2\alpha-4/3)$.

EXAMPLE 5.3.3. $u=e^{-1/t}$; i.e. $(t^2D_t-1)u=0$. In this example, $b(s)=1$. In fact, $u=t^2D_tu$. See § 5.4.

EXAMPLE 5.3.4. Let $f(x)$ be a holomorphic function. Let $b(s)$ be the b-function of $u=\delta(t-f(x))$. Then, we have

$$b(tD_t)\delta(t-f(x))=tP(x,tD_t,D_x)\delta(t-f(x)).$$

By multiplying t^s, we obtain

$$b(tD_t-s)t^s\delta(t-f(x))=P(x,tD_t-s-1,D_x)t^{s+1}\delta(t-f(x)).$$

This implies, with the change of variables $(t,x)\mapsto(t+f(x),x)$, the following:

$$b((t+f(x))D_t-s)f(x)^s\delta(t)=P(x,(t+f(x))D_t-s-1,D_x-(df)D_t)f(x)^{s+1}\delta(t).$$

Comparing the coefficients of $\delta(t)$, we obtain

$$b(-s-1)f(x)^s=P(x,-s-2,D_x)f(x)^{s+1}.$$

Hence, $b(-s-1)$ is the b-function of $f(x)$, i.e. $b(-s-1)f(x)^s\in\mathscr{D}[s]f(x)^{s+1}$. See [1], [4].

5.4. We shall micro-localize the situation of § 5.3. Let Λ be a Lagrangian submanifold of T^*X and J_Λ the sheaf of holomorphic functions which vanish on Λ. There is $f\in J_\Lambda\cap\mathcal{O}(1)$ such that $df\equiv\omega\ (=\sum\zeta_j dz_j)$ modulo J_Λ, i.e. $df-\omega\in J_\Lambda\Omega^1$. The function f is determined modulo J_Λ^2. Let $\mathscr{J}_\Lambda$ be the subsheaf of $\mathscr{E}(1)$ defined by $\sigma_1^{-1}(J_\Lambda)$ and $\mathscr{E}_\Lambda$ the subring of $\mathscr{E}$ generated by $\mathscr{J}_\Lambda$. We denote by $\mathscr{E}_\Lambda(m)$ the $\mathscr{E}_\Lambda$-submodule $\mathscr{E}(m)\cdot\mathscr{E}_\Lambda=\mathscr{E}_\Lambda\cdot\mathscr{E}(m)$ of $\mathscr{E}$ for $m\in\boldsymbol{Z}$ and we define $\mathscr{E}_{\Lambda,m}=\mathscr{E}_\Lambda\cap\mathscr{E}(m)=\mathscr{J}_\Lambda^m$. Let $\Phi=\sum\Phi_j(x,D)$ be a micro-differential operator in $\mathscr{J}_\Lambda$ such that $\Phi_1\equiv f\bmod J_\Lambda^2$ and $\Phi_0-\frac{1}{2}\sum_\nu\partial^2\Phi_1/\partial z_\nu\partial\zeta_\nu\equiv 0\bmod J_\Lambda$. Then, Φ is determined modulo $\mathscr{E}_{\Lambda,2}(-1)=\mathscr{J}_\Lambda^2(-1)$. Note that $[\Phi,P]\in\mathscr{E}_\Lambda(-1)$ for any $P\in\mathscr{E}_\Lambda$.

THEOREM 5.4.1. *Let $\mathfrak{M}$ be a holonomic $\mathscr{E}$-module with regular singularities, and let u be a section of $\mathfrak{M}$. Then, there are a nonzero polynomial $b(s)$ and $P\in\mathscr{E}_\Lambda(-1)$ such that $b(\Phi)u=Pu$ and that* $\operatorname{ord}P\leqq\deg b$.

Consider the quotient ring $\mathscr{A}=\mathscr{E}_\Lambda/\mathscr{E}_\Lambda(-1)$. Then this ring is locally isomorphic to the ring of differential operators on Λ homogeneous of degree 0 with respect to ζ. Let ϑ be the modulo class of Φ. Then, the center of $\mathscr{A}$ is the polynomial ring generated by Φ. $\mathscr{A}$ is embedded (locally) into the ring $\mathscr{D}_\Lambda$ of differential operators on Λ.

THEOREM 5.4.2. *Let $\mathfrak{M}$ be a holonomic $\mathscr{E}$-module with regular singularities and $\mathfrak{N}$ a coherent $\mathscr{E}_\Lambda$-sub-module of $\mathfrak{M}$. Then, the $\mathscr{A}$-module $\mathfrak{N}/\mathscr{E}(-1)\mathfrak{N}$ has the following properties.*

(1) *There is a nonzero polynomial $b(s)$ such that $b(\vartheta)(\mathfrak{N}/\mathscr{E}(-1)\mathfrak{N})=0$.*

(2) *The $\mathscr{D}_\Lambda$-module $\mathscr{D}_\Lambda\otimes_{\mathscr{A}}(\mathfrak{N}/\mathscr{E}(-1)\mathfrak{N})$ is a holonomic system of differential equations on Λ with regular singularities.*

5.5. Let X be an open subset of $C^{1+n}=\{(t,x);\ t\in C, x\in C^n\}$ and Λ the conormal bundle of the hypersurface $t=0$. Let $\mathfrak{M}$ be a holonomic system of micro-differential equations with regular singularities defined on a neighborhood of $p_0=(0,\sqrt{-1}\,dt)$, and u a section of $\mathfrak{M}$. Let F be an $\mathscr{E}$-linear homomorphism from $\mathfrak{M}$ into the sheaf $\mathscr{C}$ of microfunctions. Then, $v(t,x)=F(u)$ is a holonomic microfunction.

Let $b(s)$ be the polynomial given in Theorem 5.4.1, $\{\lambda_j\}_{j=1,\ ,N}$ the set of the distinct roots of $b(s)$, and let m_j be the number of the set $\{j';\ j'\neq j,\ \lambda_{j'}-\lambda_j$ is a nonnegative integer$\}$. Then, $v(t,x)$ has the "asymptotic expansion" at p_0:

$$v(t,x)\sim\sum_{j=1}^{N}\sum_{\nu=0}^{m_j}\sum_{k=0}^{\infty}a_{j,\nu,k}(x')(D_t/\sqrt{-1})^{\lambda_j-1/2-k}(\log D_t/\sqrt{-1})^{\nu}\delta(t).$$

Moreover, it is easy to see that $a_{j,\nu,k}(x)$ can be calculated from $a_{j,\nu,k}(x)$'s with $k<\deg b$ by using the equation $b(\Phi)v=Pv$. Each hyperfunction $a_{j,\nu,k}(x)$ satisfies the system of differential equations derived from $\mathfrak{N}/\mathscr{E}(-1)\mathfrak{N}$ and hence $a_{j,\nu,k}(x)$ satisfies a holonomic system of differential equations with regular singularities by Theorem 5.4.2.

EXAMPLE 5.5.1. Let us consider the hyperfunction $u=(t^2-x^3)^\alpha+\dots$ as in Example 5.3.2. This hyperfunction has the meromorphic continuation on α and has poles $\alpha=-5/6-n,\ -1-n,\ -7/6-n$ $(n=0,1,2,\dots)$. This u has the asymptotic expansion:

$$u=c_0\sum_{n=0}^{\infty}(4/27)^n\frac{[\alpha+5/6]_n}{n![1/3]_n}D_x^{3n}\delta(x)(D_t/\sqrt{-1})^{-2\alpha-5/3-2n}\delta(t)$$

$$+c_1\sum_{n=0}^{\infty}(4/27)^n\frac{[\alpha+7/6]_n}{n![4/3]_n}D_x^{3n+1}\delta(x)(D_t/\sqrt{-1})^{-2\alpha-7/3-2n}\delta(t).$$

Here,

$$c_0=-\frac{2^{2\alpha+2/3}}{\sqrt{3\pi}}\Gamma(1/3)\Gamma(\alpha+1)\Gamma(\alpha+5/6)\sin\pi(\alpha+1/6)$$

and

$$c_1=-\frac{2^{2\alpha+4/3}}{\sqrt{3\pi}}\Gamma(2/3)\Gamma(\alpha+1)\Gamma(\alpha+7/6)\sin\pi(\alpha-1/6).$$

$[s]_n$ means $s(s+1)(s+2)\dots(s+n-1)$.

Bibliography

1. I. N. Bernstein, *The analytic continuation of generalized functions with respect to a parameter*, Functional Anal. Appl. **6** (1972), 273—285.

2. L. Hörmander, *Fourier integral operators* I, Acta Math. **127** (1971), 79—183.

3. M. Kashiwara, *On the maximally overdetermined system of linear differential equations* I, Publ. RIMS, Kyoto Univ. **10** (1975), 563—579.

4. ——— *B-functions and holonomic systems*, Invent. Math. **38** (1976), 33—53.

5. ——— *On holonomic systems of linear differential equations* II (to appear).

6. M. Kashiwara, and T. Kawai, *Finiteness theorem for holonomic systems of micro-differential equations*, Proc. Japan Acad. **52** (1976), 341—343.

7. ——— *On holonomic systems of linear differential equations* III, *Systems with regular singularities* (to appear).

8. M. Kashiwara, and P. Schapira, *Micro-hyperbolic systems* (to appear).

9. B. Malgrange, *L'involutivité des caractéristiques des systèmes différentiels et microdifférentiels*, Sém. Bourbaki, Exposé **522** 1977/78.

10. M. Sato, T. Kawai and M. Kashiwara, *Microfunctions and pseudodifferential equations*, Lecture Notes in Math., vol. 287, Springer-Verlag, Berlin and New York, 1973, pp. 265—529.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Control under Incomplete Information and Differential Games

N. N. Krasovskii

Certain problems of control theory under incomplete information may be formalized within the framework of differential games. This report will be devoted to one such formalization developed by the author and his students. The size of this report leaves no opportunity for the discussion of many valuable contributions due to other authors in this field. I would only like to mention that our investigations are related to those of Bellman, Bensoussan, Boltyansky, Breakwell, Chernousko, Elliot, Fleming, Friedman, Gamkrelidze, Ho, Isaacs, Kalton, Lions, Markus, Mischenko, Nikolskii, Pontriagin, Pschenichnyi, Roxin, Varaiya, Young and certain other investigators in adjacent fields.

1. Let us first give an informal description of our problems. We will consider systems formed of the controlled plant, of the controller and of the environment. The current state of the plant is determined by its state variable $x[t]$. The evolution of $x[t]$ is described by a differential equation. The action of the controller on the plant will be named as the control and denoted by a minor u. The action of the environment will be called as the disturbance and denoted by a minor v. The accessible information on the current state of the system will be given by a certain informational variable $y[t]$ that is related in a certain way to $x[t]$ and $v[t]$. In particular it may be $y[t]=x[t]$.

In our case of uncertain information the values of the disturbance $v[\tau]$ are unknown in advance. At time t we are informed only of the domains $Q(\tau)$, $\tau \geqslant t$, that will contain the future values $v[\tau]$. These domains $\{Q(\tau), \tau \geqslant t\}$ may be included in the informational variable $y[t]$.

2. We will consider problems of closed-loop control when the desired law of control U should assign the current action $u[t]$ on the basis of the accessible realization $y[t]$.

Assume a certain functional

$$\gamma = \gamma(x[\cdot]),\ x[\cdot] = \{x[t],\ t_0 \leqslant t \leqslant \vartheta\}, \tag{2.1}$$

for the process to be given. We will say that an optimal law is a law of control U^0 that gives a minimax

$$U^0 : \gamma^0 = \min_{U} \sup_{x[\cdot]} \gamma \tag{2.2}$$

where the minimum is taken over all the admissible laws of control, and the maximum over $x[\cdot]$ is determined by all the possible realizations of the uncertain factors, namely, of the disturbance $v[t]$. It is then convenient to treat the situation as a two-person game. In this game we select the law of control—our strategy U. The realization of the uncertain factors is determined by the second player which is in general a fictitious one.

3. Consider a model problem. Assume the plant to be a heat conductor in the form of a rod $0 \leqslant \xi \leqslant 1$ on the axis ξ. Assume the informational variable $y[t]$ to coincide with the state variable $x[t]$. This is the temperature distribution ζ:

$$y[t] = x[t] = \zeta(t,\cdot) = \{\zeta(t,\xi),\ 0 \leqslant \xi \leqslant 1\}. \tag{3.1}$$

The control u may be an action of heating concentrated at point $\xi = v[t]$, so that we have the standard heat equation

$$\frac{\partial \zeta}{\partial t} = a^2 \frac{\partial^2 \zeta}{\partial \xi^2} + u\delta(\xi - v[t]) \tag{3.2}$$

under certain boundary conditions. Certain restrictions on the control and on the disturbance are also given:

$$|u| \leqslant \lambda,\ v_* \leqslant v \leqslant v^*. \tag{3.3}$$

The following problem is possible. Suppose the initial condition $\{t_0, x_0\}$ is given. Among the admissible functions $U(t, y)$ that generate the control

$$u[t] = U(t, y[t]) \tag{3.4}$$

one is to specify a law of control U^0, that ensures a minmax

$$U^0 : \min_{U} \sup_{v[\cdot]} \max_{0 \leqq \xi \leqq 1} |\zeta(\vartheta, \xi)| \tag{3.5}$$

where ϑ is a given instant of time. (A precise setting of the problem must of course include the description of the admissible functions U, v, ζ, etc.)

4. The problem may be made somewhat more complex. For example assume the coefficient a in the heat equation (3.1) and the current distribution $\zeta(t, \xi)$ to be unknown precisely. However at each time t the value $v[t]$ turns out to be known. Then the informational variable $y[t]$ may be the informational domain

$$y[t] = \{\zeta_*(t, \xi) \leqslant \zeta(t, \xi) \leqslant \zeta^*(t, \xi),\ a_*(t) \leqslant a \leqslant a^*(t),\ v[t]\} \tag{4.1}$$

where ζ^*, ζ_* and a^*, a_* are the results of an observation of the system until time t.

5. Let us return to the general case. The given concept of differential game has been developed for both ordinary and partial differential equations. It is natural that in the finite-dimensional case the results are more accomplished. We will therefore discuss the latter case in more detail. At the end of the report I will demonstrate how these results may be propagated to the case of infinite dimensions.

Assume the plant to be described by ordinary differential equation with given restrictions

$$\dot{x} = f(t, x, u, v),\quad u \in \mathscr{P},\quad v \in Q. \tag{5.1}$$

Here x, u, v are finite-dimensional vectors, the function f is continuous, while $\mathscr{P}$ and Q are compact. Assume the process to start at time t_0 from the state x_0.

The informational variable $y[t]$ may be the history

$$y[t] = x[\cdot]_t = \{x[\tau],\ t_0 \leqslant \tau \leqslant t\} \tag{5.2}$$

of the motion until time t, or $y[t]$ may be the pair $y[t] = \{x[\cdot]_t, v[t]\}$. Very often $y[t]$ may be the state $x[t]$ itself or $y[t]$ may be the pair $\{x[t], v[t]\}$.

6. The motion

$$x[\cdot] = \{x[t],\ t_0 \leqslant t \leqslant \vartheta\} \tag{6.1}$$

will be evaluated by the given functional

$$\gamma = \gamma(x[\cdot]),\quad x[\cdot] \in C[t_0, \vartheta] \tag{6.2}$$

Further on, if there is no additional reservation, the functional γ will be assumed to be continuous in the space C of continuous functions. In many practical problems one may encounter for example the functional

$$\gamma = \min_t \left[\sigma(t, x[t]),\ t \in \theta \subset [t_0, \vartheta]\right]. \tag{6.3}$$

Here $\sigma(t, x)$ is a continuous function and θ is compact.

7. It is well known that one is incapable of presenting a good formalization of minmax problems for γ if one identifies strategies with functions $u(t, x)$ and $v(t, x)$ while treating $x[\cdot]$ as classical solutions of the equation

$$\dot{x} = f\big(t, x, u(t, x), v(t, x)\big). \tag{7.1}$$

Indeed in many cases the optimal strategies $u^0(t, x)$, $v^0(t, x)$ could not be found among the functions that are suitable for a direct integration of equation (7.1).

8. Therefore we will assume the following generalized formalization. Let us name the pair $\{t, x\}$ as the position. Assume that we have selected a certain variety $\mathscr{L}$ of conditional probabilistic measures

$$\mu = \mu_v(B),\ B \subset \mathscr{P},\ v \in Q,\ \mu_v(\mathscr{P}) = 1. \tag{8.1}$$

A positional closed-loop $\mathscr{L}$-strategy U is a function

$$\mu = \mu_v(du|t, x) \tag{8.2}$$

that transforms the positions into measures from $\mathscr{L}$.

9. The strategy U generates Euler splines. These are the continuous solutions of the step-by-step equation

$$x_\Delta[t] = \int\limits_{[\tau_i, t]\times\mathscr{P}\times Q} f(\tau, x_\Delta[\tau], u, v)\mu_v(du|\tau_i, x_\Delta[\tau_i])\nu(d\tau, dv) + x_\Delta[\tau_i],\ \tau_i \leq t \leq \tau_{i+1}, \tag{9.1}$$

along a certain subdivision Δ with increment α:

$$\Delta = \{\tau_i\},\quad \tau_0 = t_0,\quad \tau_m = \vartheta,\quad \alpha = \max_i(\tau_{i+1} - \tau_i). \tag{9.2}$$

Here $\nu(d\tau, dv)$ is any measure

$$\nu = \nu(d\tau, dv)\quad \text{on}\quad [t_0, \vartheta]\times Q,\ \nu([\tau_*, \tau^*)\times Q) = \tau^* - \tau_*. \tag{9.3}$$

Here and further on all the measures are assumed to be Borel measures.

Our main assumption on the function f is that for any initial state $x_\Delta[\tau_i]$ in (9.1) and for any certain control $\eta = \mu_v \times \nu$ the solution of the equation (9.1) on the interval $\tau_i \leq t \leq \tau_{i+1}$ is unique. And all such program motions are assumed to be equibounded for each given position $\{\tau_i, x_\Delta[\tau_i]\}$.

The positional motion $x[t]$ is the limit

$$x[\cdot] = \lim x^{(k)}_{\Delta(k)}[\cdot],\ k \to \infty,\ \alpha^{(k)} \to 0, \tag{9.4}$$

of a certain sequence of Euler splines, that converges in the space C with $\alpha^{(k)}$ tending to zero.

10. In particular, a pure strategy is the measure μ (8.2) concentrated at point $u = u(t, x)$:

$$\mu(du|t, x) = \delta\big(u - u(t, x)\big)\,du.$$

Therefore it may be identified with a function $u(t, x)$. A counterstrategy is identified with a function $u(t, x, v)$: $\mu_v(du|t, x) = \delta\big(u - u(t, x, v)\big)\,du$. A mixed strategy is a function $\mu = \mu(du|t, x)$ weakly Borel in x.

11. Similarly but with substitution of u for v, μ for ν (and vice versa) and with a substitution of set $\mathscr{L}$ for a certain set K of conditional measures

$$\nu = \nu_u(B),\ B \subset Q,\quad u \in \mathscr{P}, \tag{11.1}$$

we may determine the strategies V for the second player.

12. Together with such positional closed-loop strategies we will consider the more general historical closed-loop strategies

$$\mu_v(du|x[\cdot]_t), \quad \nu_u(dv|x[\cdot]_t). \tag{12.1}$$

For a transition to these it is necessary in the former constructions to substitute position $\{t, x\}$ for the history $x[\cdot]_t$ of the motion until time t. If for any value c in which we are interested the inequality $\gamma(x[\cdot])\leqslant c$ is equivalent to the condition $\{\tau, x[\tau]\}\in\mathscr{M}_c$, $\{t, x[t]\}\in\mathscr{N}_c$, $t_0\leqslant t\leqslant\tau(x[\cdot])\leqslant\vartheta$, where $\mathscr{M}_c$ and $\mathscr{N}_c$ are closed sets in the space $\{t, x\}$ we will make use of the positional strategies. In other cases we will make use of the historical strategies with no additional explanation whatever.

13. The starting position $\{t_0, x_0\}$ and the strategy U or the strategy V determine certain boundles $X(t_0, x_0, U)$, $X(t_0, x_0, V)$ of motions $x[\cdot]=\{x[t], t_0\leqslant t\leqslant\vartheta\}$.

Assume the starting position to be given and the classes of strategies to be selected. We will formulate the following two problems that form our differential game.

The first problem is to select an optimal strategy U^0 that gives a minmax

$$U^0\colon \min_U \max_{x[\cdot]} \gamma(x[\cdot]) = c^0(t_0, x_0). \tag{13.1}$$

The second problem is to select an optimal strategy V^0 that gives a maxmin

$$V^0\colon \min_V \max_{x[\cdot]} \gamma(x[\cdot]) = c_0(t_0, x_0). \tag{13.2}$$

14. We will say that the class of $\mathscr{L}$-strategies U and the class of K-strategies V are coordinated (with respect to the function f) if for any possible position $\{t, x\}$ and for any vector s the following equality is true

$$\begin{aligned}&\min_{\mu\in\mathscr{L}} \max_{\nu} \int \langle s\cdot f(t, x, u, v)\rangle\, \mu_v(du)\times\nu(dv)\\ &= \max_{\nu\in K} \min_{\mu} \int \langle s\cdot f(t, x, u, v)\rangle\, \nu_u(dv)\times\mu(du).\end{aligned} \tag{14.1}$$

Here the symbol $\langle s\cdot f\rangle$ denotes a scalar product. In particular the classes {pure strategies—counterstrategies} and {mixed strategies—mixed strategies} are always coordinated.

15. One of the principal results is as follows.

THEOREM. *Assume that the classes of strategies $\{U\}$ and $\{V\}$ are coordinated. Then the differential game has a value*

$$\gamma^0(t_0, x_0) = c^0(t_0, x_0) = c_0(t_0, x_0) \tag{15.1}$$

and it has a saddle point—a pair of optimal strategies

$$\{\mu_v^0(du|t, x), \quad \nu_u^0(dv|t, x)\}. \tag{15.2}$$

In particular, for any function f the game always has a saddle point in the classes of mixed strategies

$$\{\mu^0(du|t, x), \quad \nu^0(dv|t, x)\}. \tag{15.3}$$

16. Assume the functional γ to be only lower semicontinuous. Then there exists an optimal strategy μ^0 and in general, only an optimalizing sequence $v_k^0, k=1, 2, \ldots,$ that approximates the value γ^0. Such an example is given by certain problems with a functional

$$\gamma(x[\cdot]) = \tau(x[\cdot]), \tau(x[\cdot]) = \min(t: \{t, x[t]\} \in \mathscr{M}) \tag{16.1}$$

where $\mathscr{M}$ is a certain given closed set $\mathscr{M}=\{x \in \mathscr{M}(t),\ t_0 \leqslant t \leqslant \vartheta\}$. Naturally in the general case of functional γ the game may have no value γ^0. An example is given already by certain problems with functonal $\gamma(x[\cdot])=\sigma(\tau(x[\cdot]), x[\tau(x[\cdot])])$. Here $\sigma(t, x)$ is a continuous function and the value $\tau(x[\cdot])$ is determined by the equality (16.1). In these cases one may indicate some additional sufficient conditions for the existence of a value of the game and of a saddle point.

17. The essence of the given formal theorem on the saddle point may be clarified with the aid of approximations. Suppose for example that the game is formalized in the pair of classes of positional strategies; {pure strategies—counterstrategies}. Then the pure optimal strategy $u^0(t, x)$ ensures an inequality

$$\gamma(x_\Delta[\cdot]) < \gamma^0(t_0, x_0) + \varepsilon \tag{17.1}$$

for any $\varepsilon > 0$ selected in advance. This is true for any Euler spline described by the equation

$$\begin{gathered} \dot{x}_\Delta = f(t, x_\Delta[t], u^0(\tau_i, x_\Delta[\tau_i]),\ v[t]), \\ \tau_i \leqslant t \leqslant \tau_{i+1},\ v[t] \in Q, \end{gathered} \tag{17.2}$$

provided the increment α of subdivision Δ is sufficiently small: $\tau_{i+1} - \tau_i \leqq \alpha(\varepsilon)$, $\alpha(\varepsilon) > 0$. Here the measurable realization of the variable $v[t]$ is generated by the environment on the basis of the one or the other of its laws. In particular if the disturbance $v[t]$ will be formed on the basis of its optimal counterstrategy $v^0(t, x, u)$ with its own subdivision $\Delta^* = \{\tau_i^*\}$ that means that the motion $x_\Delta[t]$ will also satisfy the equation

$$\begin{gathered} \dot{x}_\Delta = f(t, x_\Delta[t], u[t], v^0(\tau_i^*, x_\Delta[\tau_i^*], u[t])), \\ \tau_i^* \leqslant t \leqslant \tau_{i+1}^*, \end{gathered} \tag{17.3}$$

where $u[t]$ is any measurable realization of the control, then the following inequality will be fulfilled

$$\gamma(x_\Delta[\cdot]) > \gamma^0(t_0, x_0) - \varepsilon \tag{17.4}$$

provided increment α^* of the subdivision Δ^* is sufficiently small: $\tau_{i+1}^* - \tau_i^* \leqslant \alpha^*(\varepsilon)$.

18. Unfortunately these approximations are unstable with respect to minor informational errors $\Delta x[t] = x^*[t] - x[t]$. Indeed, the inequalities (17.1), (17.4) determined above may be destroyed if the actual realizations $x_\Delta[t]$ are determined by the equations

$$\dot{x}_\Delta = f(t, x_\Delta[t], u^0(\tau_i, x_\Delta^*[\tau_i]),\ v[t]) \tag{18.1}$$

or

$$\dot{x}_{\Delta^*} = f(t, x_\Delta[t], u[t], v^0(\tau_i^*, x_{\Delta^*}^*[\tau_i^*], u[t]) \tag{18.2}$$

even in the case of arbitrary small errors Δx. And this occurs not only for our concept. An instability with respect to minor informational errors is typical for many well-known solutions of differential games.

19. However we may offer the following improvement of the approximations. To the primary plant $\mathscr{F}$ we add a certain model $\mathscr{H}$ whose current state may be characterized by a suitable variable $w[t]$. This model may be materialized in the actual circuit of control on some computer. The variable $w[t]$ is a guide that directs the motion $x[t]$ to the desired target. The relation between $x[t]$ and $w[t]$ is constructed on the basis of the stability theory. The motion $x[t]$ is governed by an appropriate substrategy $\mu_v(du|t, x, w)$. The model $\mathscr{H}$ is constructed on the basis of one of the formal models of the game of which we will speak in the sequel. Therefore, if we follow the terminology of a chess game, we will have a game "on two boards". In the plant $\mathscr{F}$ we are playing with nature while in the model $\mathscr{H}$ we are playing with ourselves. When this procedure is implemented we always achieve a stable procedure of control that ensures a suboptimal result (17.1) or (17.4) for the player that uses its guide $w[t]$.

20. Let us now discuss an approximation for the case of mixed strategies. A mixed positional strategy $\mu(du|t, x)$ in approximation generates already a random motion $x_\Delta[t]$ that satisfies a step-by-step equation

$$\dot{x}_\Delta = f(t, x_\Delta[t], u[\tau_i], v[t]), \ \ \tau_i \leqslant t \leqslant \tau_{i+1}. \tag{20.1}$$

Here $u[\tau_i]$ is the result of a random test with probability distribution $\mu(du) = \mu(du|\tau_i, x_\Delta[\tau_i])$ on $\mathscr{P}$. Such a procedure of control based on anoptimal strategy μ^0, ensures the inequality

$$P(\gamma(x_\Delta[\cdot]) < \gamma^0(t_0, x_0) + \varepsilon) > \beta \tag{20.2}$$

for any $\varepsilon > 0$ and $\beta < 1$ selected in advance provided increment α of subdivision Δ is sufficiently small: $\tau_{i+1} - \tau_i \leqslant \alpha(\varepsilon, \beta)$, $\alpha(\varepsilon, \beta) > 0$. Here the symbol $P(\ldots)$ denotes the probability of the respective event.

The disturbance $v[t]$ may be formed in an arbitrary manner that may also allow an appropriate statistical interpretation.

If the disturbance $v[t]$ is formed on the basis of its optimal strategy $\nu^0(dv|t, x)$, with its own subdivision then the following inequality will be ensured

$$P(\gamma(x_\Delta[\cdot]) > \gamma^0(t_0, x_0) - \varepsilon) > \beta \tag{20.3}$$

provided the increment α^* of the subdivision Δ^* will be sufficiently small: $\tau^*_{i+1} - \tau^*_i \leqslant \alpha^*(\varepsilon, \beta)$, $\alpha^*(\varepsilon, \beta) > 0$.

It is important to note that each of the propositions (20.2) and (20.3) is true under the condition that the actions $u[t]$ and $v[t]$ are stochastically independent within minor intervals of time or at least that they are sufficiently weakly correlated. If we have a game with nature, then the a priori given assumption on the disturbance $v[t]$ in equation (20.2) seems to be a completely tolerable independent postulate.

However if we speak of a game between two intelligent players, each of which may select its mixed strategy with its own subdivision then this assumption cannot be taken as an independent postulate. It should then be founded. Indeed it may be well founded if we take that the current realizations $x_\Delta[t]$ are available for each player with sufficiently small errors Δx. Then the optimal strategies in appropriate schemes of control with a guide will ensure the inequalities (20.2) and (20.3) under the condition that the informational errors are sufficiently small and that the increments α and α^* are also sufficiently small. I would like to emphasize that all the approximations described here had been formulated and proved in precise terms. Here however due to a lack of space I was capable of giving only a partial and rather loose presentation of these topics.

21. The proof for the existence of saddle points for our game and the construction of control algorithms for the actual approximations are based on various formalized models of the game. One such model based on the limit motions $x[t]$ has been already described above. Let us now describe some other model. For determinicity we will further restrict ourselves to the case of mixed strategies. Let us consider a formalization based on quasistrategies. A quasistrategy $\mathcal{U}$ for the interval $[\tau, \vartheta]$ is a transformation

$$\{\mu_t(du), \tau \leq t \leq \vartheta\} = \mathcal{U}\{\nu_t(dv), t_* \leq t \leq \vartheta\} \tag{21.1}$$

that transforms conditional stochastic measures $\nu_t(dv)$ onto conditional stochastic measures $\mu_t(du)$. The transformation $\mathcal{U}$ must satisfy a condition of physical realizability. That is for any τ_* from the given interval $[\tau_*, \vartheta]$ the histories of the images μ_t, $t<\tau$, coincide provided the histories of the arguments ν_t, $t<\tau_*$, have already coincided.

The starting position $\{\tau_*, w_*\}$, the quasistrategy and the conditional measure ν_t, $\tau_*<t$, generate a quasimotion $w[t]$ that is a solution of the equation

$$\dot{w} = \int_{\mathcal{P}\times Q} f(t, w, u, v)\mu_t(du)\times\nu_t(dv), \quad w[\tau_*] = w_*, \tag{21.2}$$

where $\mu_t(du)$ is defined by condition (21.1).

The quasistrategy $\mathcal{V}$ is defined similarly with appropriate substitutions.

22. The formal model considered here is formed of two problems. Assume a certain initial history $w[\cdot]_{\tau_*}=\{w[\tau], t_0\leq\tau\leq\tau_*\}$ is given.

The First Problem is to select an optimal quasistrategy $\mathcal{U}^0$ that gives a minmax

$$\mathcal{U}^0\colon \min_{\mathcal{U}} \max_{\nu} \gamma(x[\cdot]) = c^0_*(w[\cdot]_{\tau_*}). \tag{22.1}$$

The Second Problem is to select an optimal quasistrategy $\mathcal{V}^0$ that gives a maxmin

$$\mathcal{V}^0\colon \max_{\mathcal{V}} \min_{\mu} \gamma(x[\cdot]) = c^*_0(w[\cdot]_{\tau_*}). \tag{22.2}$$

It has been proved that this formal game has a value γ_*^0 and a pair of optimal strategies

$$\gamma_*^0(w[\cdot]_{\tau*}) = c_0^* = c_*^0, \ \{\mathscr{U}^0, \mathscr{V}^0\}. \tag{22.3}$$

The main result here is that the value $\gamma^0(t_0, x_0)$ of the initial closed-loop game considered above is equal to the value $\gamma_0^*(w[\cdot]_{\tau_0})$ of this formal game for the same starting position $w[\cdot]_{t_0}=w[t_0]=x_0$.

23. If the optimal quasistrategy $\mathscr{U}^0$ and $\mathscr{V}^0$ has been determined then the closed loop strategy μ^0 and ν^0 is constructed in principle without a great difficulty. Moreover this formal model in terms of quasistrategies is very suitable for the construction of an actual model $\mathscr{M}$ in the control scheme with a guide $w[t]$. However the search for the optimal quasistrategies is complexified by conditions of physical realizability. Let us omit this condition say for the case of quasistrategies for the first player. We will obtain the operators Π:

$$\{\mu_t(du), \tau_* \leqslant t \leqslant \vartheta\} = \Pi\{\nu_t(dv), \tau_* \leqslant t \leqslant \vartheta\} \tag{23.1}$$

that will be named as the programs. Let us formulate the first open-loop problem. This problem is to select an optimal program that gives a minmax

$$\Pi^0: \min_{\Pi} \max_{\nu} \gamma(w(\cdot)) = c_*(w[\cdot]_{\tau*}). \tag{23.2}$$

Here the program motions $w(t)$ are the solutions of the equation (21.2) where $\mu_t(du)$ is determined by condition (22.1).

In general this open-loop problem is not equivalent with respect to the value c to a similar problem for quasistrategies. However one may indicate certain regularity conditions when the equality $c_*^0(w[\cdot]_{\tau_*})=c_*(w[\cdot]_{\tau_*})$ is true. In these cases one may construct a closed-loop strategy $\mu^0(du|t, x)$ on the basis of solving some auxiliary open loop problems (23.2).

24. The solution

$$\{\mu_t^0(du), \tau_* \leqslant t \leqslant \vartheta\} = \Pi^0\{\nu_t^0(dv), \tau_* \leqslant t \leqslant \vartheta\} \tag{24.1}$$

of the open-loop problem (23.2) is determined under certain assumptions by a minmax condition

$$\begin{aligned}&\langle s(t)\cdot \int f(t, w^0(t), u, v)\mu_t^0(du)\times \nu_t^0(dv)\rangle \\ &= \min_{\mu} \max_{\nu} \langle s(t)\cdot \int f(t, w^0(t), u, v)\mu(du)\times \nu(dv)\rangle\end{aligned} \tag{24.2}$$

that corresponds here to the well-known maximum principle of Pontriagin. Here $w^0(t)$ and $s(t)$ are the solutions of certain ordinary differential equations, that very often turn out to be of the Hamiltonian type.

The main point in the regularity conditions is as follows. Let $S(w[\cdot]_{\tau_*})$ denote the set of all vectors $s(\tau_*)$, that correspond to all of the possible optimal solutions.

of the mentioned equation. Then the condition will consist in the feature that for any selection of the vector l the intersection

$$W_l^+ \cap \left(\bigcap_s W_s^-\right) \neq \emptyset, \quad s \in S(w[\cdot]_{\tau*}), \tag{24.3}$$

would be nonvoid. Here the symbols W_s^+ and W_s^- denote the semispaces

$$W_s^+ = \{w\colon \langle s \cdot w\rangle \geqslant \varkappa\},\ W_s^- = \{w\colon \langle s\cdot w\rangle \leqslant \varkappa\}$$
$$\varkappa = \min_\mu \max_\nu \int \langle s \cdot f\rangle \mu(du) \times \nu(dv). \tag{24.4}$$

Beyond such regular cases we have a great deal of very peculiar cases of the games. One may give some classification of these cases, and obtain the regular case of the first rank, of the second rank, etc., until infinity.

25. Thus in the regular cases the problem of synthesizing an optimal closed-loop strategy may be reduced to the solution of auxiliary open loop control problems on the basis of ordinary Hamiltonian equations. We have therefore arrived at a typical procedure of analytical mechanics. Another scheme of solving extremal problems that is also standard for analytical mechanics is related to the Hamilton–Jacobi partial equation. In our case this scheme leads to the partial equations of dynamic programming. Unfortunately the value γ^0 that must be a solution of this equation often turns out to be a nondifferentiable function of the position $\{t, x\}$ or of the history $x[\cdot]_t$ of the motion. It is known however that as a rule a transition to related stochastic games for systems with a minor Wiener noise yield a regularization of the value of the game. Within the given concept this appears in the following way. Again for the sake of determinicity we consider only the case of mixed strategies and for example only for the functional γ of type $\gamma = \sigma(x[\vartheta])$ where time ϑ is given, $\sigma(x)$ is a continuous function.

Consider a plant $\mathscr{H}_\lambda$ described by the Ito equation

$$dw = \int_{\mathscr{P}\times Q} h_\lambda(t, w, u, v)\mu(du \mid t, w)$$
$$\times \nu(dv \mid t, w)\, dt + \lambda\, dz[t] \tag{25.1}$$

and a functional

$$\gamma_\lambda(w[\cdot]) = M\{\sigma_\lambda(w[\vartheta]) \mid t_0, w[t_0] = x_0\} \tag{25.2}$$

where $M\{\ldots\}$ stands for the mean value.

Here $z[t]$ is a nongenerate Wiener process, $\lambda > 0$ is a small parameter, the function $\sigma_\lambda(x)$ is bounded and sufficiently smooth. Moreover in a sufficiently large domain G uniform limit relations $\lim h_\lambda = f$, $\lim \sigma_\lambda = \sigma$, $\lambda \to 0$ are valid.

The mixed strategies are identified with conditional probability measures $\mu(du|t, w)$, $\nu(dv|t, w)$ that are weakly Borel in $\{t, w\}$. The random motions are the weak solutions of equation (25.1).

26. It is known that this game on the minmax-maxmin of the functional γ_λ (25.2) has a value $\gamma^0(t_0, x_0)$ and a saddle point $\{\mu_\lambda^0(du|t, w),\ \nu_\lambda^0(dv|t, w)\}$.

The value $\gamma_\lambda^0(t, x)$ is a smooth function and satisfies the well-known parabolical partial equation

$$\frac{\partial \gamma_\lambda^0}{\partial t}+\frac{\lambda^2}{2}\Delta\gamma_\lambda^0+\min_{\mu}\max_{\nu}\int_{\mathscr{P}\times Q}\left\langle\frac{\partial\gamma_\lambda^0}{\partial x}\cdot h_\lambda\right\rangle\mu(du)\times\nu(dv)=0 \tag{26.1}$$

with boundary condition $\gamma_\lambda^0(\vartheta, x)=\sigma_\lambda(x)$.

27. The principal result consists here in the limit relation

$$\lim_{\lambda\to 0}\gamma_\lambda^0(t_0, x_0) = \gamma^0(t_0, x_0). \tag{27.1}$$

Unfortunately, in general there are no analogous limit relations for the strategies. However if we know the optimal strategies μ_λ^0 and ν_λ^0 for the stochastic game with minor $\lambda>0$, it is always possible to construct a control for the given plant with a stochastic guide $\mathscr{H}_\lambda$ so that for any $\varepsilon>0$ and $\beta<1$ selected in advance the inequality

$$P\big(\sigma(x[\vartheta]) < \gamma^0(t_0, x_0)+\varepsilon\big) > \beta \tag{27.2}$$

would be fulfilled for the first player or the inequality

$$P\big(\sigma(x[\vartheta]) > \gamma^0(t_0, x_0)-\varepsilon\big) > \beta \tag{27.3}$$

for the second player, provided the parameter λ, the informational errors Δx and the increment α are sufficiently small.

28. We will now pass to the discussion of systems with infinite dimensions. First of all, note that with no great difficulty the previous results may be propagated to systems with time lag

$$\dot{x} = f(t, x_{t-\varrho}[\cdot]_t, u, v) \tag{28.1}$$

where f is a functional of the history of the motion $x_{t-\varrho}[\cdot]_t=\{x[\tau], t-\varrho\leqslant\tau\leqslant t\}$. As a specific fact we note that under rather general assumptions the differential games for a functional differential system (28.1) are well approximated by appropriate games for finite-dimensional systems described by ordinary differential equations.

29. Further on the results are propagated to certain parabolical and hyperbolical systems under standard initial and boundary conditions. Here the assumptions on the admissible classes of parameters and spacial and boundary control actions are related to the conventional assumptions of the general theory of parabolical or hyperbolical systems and in particular, to the theory of optimal open-loop control for such systems. Here first of all I have in mind the investigations of the group led by Lions.

Some theorems were proved within the framework of our concept. These theorems concern the existence of saddle points for the respective closed-loop differential games. Respective formal models as well as algorithms for the construction of optimal closed-loop strategies have been developed. In particular these algorithms include those that are based on solving auxiliary open-loop problems. In the actual approximations the motions or Euler splines are understood to be the splines formed by generalized solutions of the corresponding equations for the intervals $\tau_i \leq t \leq \tau_{i+1}$. It is natural that the functional nature of the problem yields a lot of additional difficulties. However under certain assumptions we may overcome them. Here in each case the result depends greatly on the selection of an appropriate functional space. In many cases of parabolical systems for example the approximate motions are considered in the space $\mathscr{L}^{(2)}$ but the respective formal constructions were developed through special transformation in other appropriate spaces, say in the state space $\mathscr{H}^{(-1)}$.

For the parabolical and hyperbolical systems considered here it is also possible to achieve a good approximation by related game theoretic problems for finite-dimensional systems. Such problems of approximation sufficient tolerably may be solved as on the bases of the method of Galerkin as well as within difference schemes.

30. Let us return for example to the model of the heat conducting rod at the beginning of our report. The respective differential game has a value $\gamma^0(t_0, \zeta_0)$ and a saddle point in mixed strategies

$$\{\mu^0(du|t, \zeta(t, \cdot)), \nu^0(dv|t, \zeta(t, \cdot))\}. \tag{30.1}$$

This denotes that in the actual approximation schemes that are based on the equations

$$\frac{\partial \zeta_\Delta}{\partial t} = a^2 \frac{\partial^2 \zeta}{\partial \xi^2} + u[\tau_i]\delta(\xi - v[t]), \quad \tau_i \leq t < \tau_{i+1}, \tag{30.2}$$

or

$$\frac{\partial \zeta_\Delta}{\partial t} = a^2 \frac{\partial^2 \zeta}{\partial \xi^2} + u[t]\delta(\xi - v[\tau_i^*]), \quad \tau_i^* \leq t < \tau_{i+1}^*, \tag{30.3}$$

for the first or second player respectively we have that for any $\varepsilon > 0$ and $\beta < 1$ selected in advance the inequalities

$$P\big(\max_{0 \leq \xi \leq 1} |\zeta_\Delta(\vartheta, \xi)| < \gamma^0(t_0, \zeta_0(t_0, \cdot)) + \varepsilon\big) > \beta \tag{30.4}$$

and

$$P\big(\max_{0 \leq \xi \leq 1} |\zeta_\Delta(\vartheta, \xi)| < \gamma^0(t_0, \zeta_0(t_0, \cdot)) - \varepsilon\big) > \beta \tag{30.5}$$

would be fulfilled provided the increments α and α^* of the subdivisions Δ and Δ^* are sufficiently small. Here $u[\tau_i]$ and $v[\tau_i^*]$ are respectively the results of stochastic tests with distributions

$$\mu(du) = \mu(du|\tau_i, \zeta(\tau_i, \cdot)) \quad \text{on } \mathscr{P} \tag{30.6}$$

and

$$\nu(dv) = \nu(dv|\tau_i^*, \zeta(\tau_i^*, \cdot)) \quad \text{on } Q. \tag{30.7}$$

Here the results and the considerations of above that concern the similar finite-dimensional case still remain true.

31. Let us now discuss the case when the informational variable $y[t]$ is the informational domain $y[t] = G[t]$, $x[t] \in G[t]$ in the space $\{x\}$ that includes the actual state $x[t]$. The problem is now transferred to the one of controlling the evolution of these domains. For the construction of the laws for this evolution a respective theory of differential games of observation has been developed. This theory may be considered as a certain minmax analogy for the statistical filtering theory. The combination of dual closed-loop differential games of control and observation includes sufficiently general theorems for the existence of saddle points as well as certain methods of constructing of optimal strategies of observation and control. However a practical realization of the solutions is achieved here for more or less simple model problems. The investigations are more effective for the case when the domains $G[t]$ are convex. Then the problem may be reduced to differential games in one or another functional space that includes the support function $g[t, l]$ of these domains. In this form the problems may be placed within a sufficiently general framework of differential games for differential evolutionary systems.

References

1. N. N. Krasovskii and A. I. Subbotin, *Positional differential games,* Nauka, Moscow, 1974.

2. N. N. Krasovskii, *Mixed strategies in differential games,* Soviet Math. Dokl. **235** (1977). 519—522.

3. A. B. Kurzhanskii, *Control and observation under conditions of uncertainty,* Nauka, Moscow, 1977.

4. Yu. S. Osipov, *Closed-loop control in parabolic systems,* Prikl. Mat. Meh. **41** (1977), 195—201.

5. V. D. Batuhtin, *A programmed construction for the positional control,* Optim. Techn. IFIP Techn. Conf., Springer-Verlag, Berlin, 1975, 435—439.

6. A. G. Chentsov, *On a problem of tracking,* Soviet Math. Dokl. **226** (1976), 73—76.

7. A. V. Krjažimskii, *On the theory of positional differential games of convergence-evasion,* Soviet. Math. Dokl. **239** (1978), 779—782.

СССР, 620066, г. Свердловск, К—66,
ул. С. Ковалевской, 16
Институт математики и механики
УНЦ АН СССР

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

L-Functions and Automorphic Representations

R. P. Langlands

Introduction. There are at least three different problems with which one is confronted in the study of L-functions: the analytic continuation and functional equation; the location of the zeros; and in some cases, the determination of the values at special points. The first may be the easiest. It is certainly the only one with which I have been closely involved.

There are two kinds of L-functions, and they will be described below: motivic L-functions which generalize the Artin L-functions and are defined purely arithmetically, and automorphic L-functions, defined by data which are largely transcendental. Within the automorphic L-functions a special class can be singled out, the class of standard L-functions, which generalize the Hecke L-functions and for which the analytic continuation and functional equation can be proved directly.

For the other L-functions the analytic continuation is not so easily effected. However all evidence indicates that there are fewer L-functions than the definitions suggest, and that every L-function, motivic or automorphic, is equal to a standard L-function. Such equalities are often deep, and are called reciprocity laws, for historical reasons. Once a reciprocity law can be proved for an L-function, analytic continuation follows, and so, for those who believe in the validity of the reciprocity laws, they and not analytic continuation are the focus of attention, but very few such laws have been established.

The automorphic L-functions are defined representation-theoretically, and it should be no surprise that harmonic analysis can be applied to some effect in the study of reciprocity laws. One recent small success was the proof of a reciprocity law for the Artin L-functions associated to tetrahedral representations of a Galois group and to a few other representations of degree two. It is the excuse for this lecture, but I do not want to overwhelm you with technique from harmonic analysis. Those who care to will be able to learn it at leisure from **[6]**, in which a concerted effort was made to provide an introduction to automorphic representations, and so I forego proofs, preferring instead to review the evolution of our notion of an L-function and a reciprocity law over the past five decades.

Artin and Hecke L-functions. An L-function is, first of all, a function defined by a Dirichlet series with an Euler product, and is therefore initially defined in a right half-plane. I will forbear defining explicitly the best known L-functions, the zeta-functions of Riemann and Dedekind, and the L-functions of Dirichlet, and begin with the more general functions introduced in this century by Hecke **[19]** and by Artin **[2]**. Artin's reciprocity law is the pattern to which all others, born and unborn, are cut.

Although they overlap, the two kinds of L-functions are altogether different in their origins. If F is a number field or, if one likes, a function field, although I prefer to leave function fields in the background, for they will be discussed by Drinfeld [12], then a Hecke L-function is an Euler product $L(s, \chi)$ attached to a character χ of $F^\times \backslash I_F$. I_F is the group of idèles of F. If v is a place of F then $F_v^\times$ imbeds in I_F and χ defines a character χ_v of $F_v^\times$. To form the function $L(s, \chi)$ we take a product over all places of F:

$$L(s, \chi) = \prod_v L(s, \chi_v).$$

If v is archimedean the local factor $L(s, \chi_v)$ is formed from Γ-functions. Here the important point is that whenever v is defined by a prime $\mathfrak{p}$ and χ_v is trivial on the units, as it is for almost all v, then

$$L(s, \chi_v) = \frac{1}{1-\alpha(\mathfrak{p})/N\mathfrak{p}^s}$$

with

$$\alpha(\mathfrak{p}) = \chi_v(\varpi_\mathfrak{p}),$$

$\varpi_\mathfrak{p}$ being a uniformizing parameter at $\mathfrak{p}$. The function $L(s, \chi)$ can be analytically continued and has a functional equation of the form

$$L(s, \chi) = \varepsilon(s, \chi) L(1-s, \chi^{-1}),$$

$\varepsilon(s, \chi)$ being an elementary function **[35]**.

An Artin L-function is associated to a finite-dimensional representation ϱ of a Galois group $\mathrm{Gal}(K/F)$, K being an extension of finite degree. It is defined arithmetically and its analytic properties are extremely difficult to establish. Once again

$$L(s, \varrho) = \prod L(s, \varrho_v),$$

ϱ_v being the restriction of ϱ to the decomposition group. For our purposes it is enough to define the local factor when v is defined by a prime $\mathfrak{p}$ and $\mathfrak{p}$ is unramified in K. Then the Frobenius conjugacy class $\Phi_{\mathfrak{p}}$ in $\mathrm{Gal}(K/F)$ is defined, and

$$L(s, \varrho_v) = \frac{1}{\det\left(I-\varrho(\Phi_{\mathfrak{p}})/N\mathfrak{p}^s\right)} = \prod_{i=1}^{d} \frac{1}{1-\beta_i(\mathfrak{p})/N\mathfrak{p}^s},$$

if $\beta_1(\mathfrak{p}), \ldots, \beta_d(\mathfrak{p})$ are the eigenvalues of $\varrho(\Phi_{\mathfrak{p}})$.

Although the function $L(s, \varrho)$ attached to ϱ is known to be meromorphic in the whole plane, Artin's conjecture that it is entire when ϱ is irreducible and non-trivial is still outstanding. Artin himself showed this for one dimensional ϱ [3], and it can now be proved that the conjecture is valid for tetrahedral ϱ, as well as a few octahedral ϱ. Artin's method is to show that in spite of the differences in the definitions the function $L(s, \varrho)$ attached to a one-dimensional ϱ is equal to a Hecke L-function $L(s, \chi)$, where $\chi=\chi(\varrho)$ is a character of $F^\times\backslash I_F$. He employed all the available resources of class field theory, and went beyond them, for the equality of $L(s, \varrho)$ and $L(s, \chi(\varrho))$ for all ϱ is pretty much tantamount to the Artin reciprocity law, which asserts the existence of a homomorphism from I_F onto the Galois group $\mathrm{Gal}(K/F)$ of an abelian extension which is trivial on $F^\times$ and takes $\bar{\omega}_{\mathfrak{p}}$ to $\Phi_{\mathfrak{p}}$ for almost all $\mathfrak{p}$.

The equality of $L(s, \varrho)$ and $L(s, \chi)$ implies that of $\chi(\bar{\omega}_{\mathfrak{p}})$ and $\varrho(\Phi_{\mathfrak{p}})$ for almost all $\mathfrak{p}$. On close examination both these quantities are seen to be defined by elementary, albeit extremely complicated, operations, and so the reciprocity laws for one-dimensional ϱ, like the quadratic and higher reciprocity laws implicit in them, are ultimately elementary, and can for any ϱ and any given prime $\mathfrak{p}$ be verified by computation. The reciprocity law for tetrahedral ϱ seems, on the other hand, to be of a truly transcendental nature, and must be judged not by traditional criteria but by its success with the Artin conjecture.

Motivic L-functions. If V is a nonsingular projective variety over a number field then, for almost all $\mathfrak{p}$, V has a good reduction over the residue field $F_{\mathfrak{p}}$ at $\mathfrak{p}$ and we can speak of the number $N(n)$ of points with coordinates in the extension of F of degree n. Following Weil [36], we define the zeta-function $Z_{\mathfrak{p}}(s, V)$ by

$$\log Z_{\mathfrak{p}}(s, V) = \sum_{n=1}^{\infty} \frac{1}{n} \frac{N(n)}{N\mathfrak{p}^{ns}}.$$

We owe to the efforts of Dwork, Grothendieck, Deligne, and others the proof that

$$Z_{\mathfrak{p}}(s, V) = \prod_{i=0}^{2d} \prod_{j=1}^{b_i} \left(1 - \frac{\alpha_{ij}(\mathfrak{p})}{N\mathfrak{p}^s}\right)^{(-1)^{i+1}} = \prod_{i=0}^{2d} L_{\mathfrak{p}}^i(s, V)^{(-1)^i}.$$

Here d is the dimension of V, b_i its i^{th} Betti number, and

$$|\alpha_{ij}(\mathfrak{p})| = N\mathfrak{p}^{i/2}.$$

It seems to have been Hasse (cf. [18]) who first proposed, in the case of an elliptic curve, the problem of proving that the L-function $L^i(s, V)$ defined by the Euler product

$$\prod_v L_v^i(s, V)$$

has analytic continuation and functional equation. Of course, a solution of the problem involves a reasonable definition of the local factors at the infinite places and at the finite places at which V does not have good reduction.

Since b^i is generally greater than 1 and $L^i(s, V)$ is an Euler product of degree b^i, it cannot, except in special circumstances, be equal to an $L(s, \chi)$. Sometimes, however, $L^i(s, V)$ can be factored into a product of b^i Euler products of degree 1, each of which is equal to a Hecke L-function. The idea of factoring an L-function into Euler products of smaller degree is very important. It led Artin from the zeta-function of K to the L-functions associated to representations of $\mathrm{Gal}(K/F)$. Allusions to the same idea can be found in the correspondence of Dedekind with Frobenius [9], from which it appears that it was at the origin of the notion of a group character. The factorization can be simply interpreted in the context of the l-adic representations of Grothendieck.

The field K is the function field of an algebraic variety of dimension 0 over F and the zeta-function of K is $L^0(s, V)$. The variety $V_{\bar{F}}$ obtained from V by extension of scalars to the algebraic closure $\bar{F}$ has $[K\colon F]$ points. The Galois group $\mathrm{Gal}(\bar{F}/F)$ acts on these points and hence on the l-adic étale cohomology group $H^0(V_{\bar{F}})$. The zeta-function may be defined in the same way as the Artin L-function except that it is now associated to the representation of $\mathrm{Gal}(\bar{F}/F)$ on $H^0(V_{\bar{F}})$. The function field of $V_{\bar{F}}$ is $K \otimes_F \bar{F}$ and the action of $\mathrm{Gal}(\bar{F}/F)$ on points is induced by its natural action on the second factor of the tensor product. The action of $\mathrm{Gal}(K/F)$ on $V_{\bar{F}}$ induced by its action on the first factor is geometric, because the ground field $\bar{F}$ is fixed, and commutes with $\mathrm{Gal}(\bar{F}/F)$. The group $\mathrm{Gal}(K/F)$ will then act on the cohomology as well, and so to each σ in $\mathrm{Gal}(K/F)$ we associate an operator $T(\sigma)$ on $H^0(V_{\bar{F}})$. If some linear combination of the $T(\sigma)$ is an idempotent E, we can restrict the representation of $\mathrm{Gal}(\bar{F}/F)$ to its range, and employing Artin's procedure attach an L-function $L(s, E)$ to the restriction. Taking a family of such idempotents, orthogonal and summing to the identity, we obtain a factorization of $L^0(s, V)$ or of the zeta-function of K. Since the representation of $\mathrm{Gal}(\bar{F}/F)$ on $H^0(V_{\bar{F}})$ is equivalent to that obtained by lifting the regular representation of $\mathrm{Gal}(K/F)$ to $\mathrm{Gal}(\bar{F}/F)$, and the interplay between the actions of $\mathrm{Gal}(K/F)$ and $\mathrm{Gal}(\bar{F}/F)$ is that between the left and right regular representations, we obtain the factorization of Artin

$$L^0(s, V) = \zeta_K(s) = \prod_\varrho L(s, \varrho)^{\deg \varrho}.$$

The product is taken over all irreducible representations of $\mathrm{Gal}(K/F)$.

For a general variety the function $L^i(s, V)$ is obtained from the representation of $\mathrm{Gal}(\bar{F}/F)$ on the l-adic cohomology group $H^i(V_{\bar{F}})$. The algebraic correspondences of V with itself which are of degree 0 and defined over F will define operators on $H^i(V_{\bar{F}})$ which commute with $\mathrm{Gal}(\bar{F}/F)$. Once again, if some linear combination of these operators is an idempotent E we may introduce $L(s, E)$, hoping that it will have an analytic continuation, and that it will be equal to a Hecke L-function if the range of E has dimension one.

In particular, if we can write the identity as a sum of such idempotents which are orthogonal and of rank one then we can hope to prove that $L^i(s, V)$ is a product of Hecke L-functions, and so has an analytic continuation and a functional equation. The major examples here are abelian varieties of CM-type, the relevant endomorphisms being defined over F. The idempotents are constructed from these endomorphisms. The theorems were proved by Shimura, Taniyama, Weil, and Deuring (cf. [33]).

The functions $L(s, E)$ seem to be the correct, perhaps the ultimate, generalizations of the Artin L-functions. There is no reason to expect that they can be further factored. On close examination, it will be seen that the meaning of E has been left fuzzy. It should be a motive, a problematical notion, which Grothendieck has made precise ([23], [29]). But it cannot be shown to have all the properties desired of it without invoking certain conjectures closely related to the Hodge conjecture. Indeed, if the Hodge conjecture itself turns out to be false the notion will lose much of its geometric appeal. Furthermore there are L-functions arising in the study of Shimura varieties which we would be unwilling to jettison but which have not been shown to be carried by a motive in the sense of Grothendieck. But the notion is indispensable, and if the attendant problems will not yield to a vigorous assault then we have to prepare for a long siege.

If the functions $L(s, E)$ cannot be factored further then the theorems of Artin and Shimura-Taniyama mark the limits of usefulness of the Hecke L-functions in the study of the motivic L-functions. Fortunately the Hecke L-functions can be generalized.

Standard L-functions and the principle of reciprocity. If A is the adèle ring of F then I_F is $\mathrm{GL}(1, A)$, $F^\times$ is $\mathrm{GL}(1, F)$, and a character of $F^\times\backslash I_F$ is nothing but a representation of $\mathrm{GL}(1, A)$ that occurs in the space of continuous functions on $\mathrm{GL}(1, F)\backslash\mathrm{GL}(1, A)$. It is the simplest type of automorphic representation. $\mathrm{GL}(n, A)$ acts on the factor space $\mathrm{GL}(n, F)\backslash\mathrm{GL}(n, A)$ and hence on the space of continuous functions on it. An automorphic representation of $\mathrm{GL}(n, A)$ is basically an irreducible constituent π of the representation on the space of continuous functions, but the topological group $\mathrm{GL}(n, A)$ is not compact and π is, in general, infinite-dimensional, and some care must be taken with the definitions [7]. One can attach to an automorphic representation π of $\mathrm{GL}(n, A)$ an L-function $L(s, \pi)$ which will have an analytic continuation and a functional equation [17]:

$$L(s, \pi) = \varepsilon(s, \pi)L(1-s, \tilde{\pi}),$$

with $\tilde{\pi}$ contragredient to π. It is possible [14] to write π as a tensor product $\pi = \otimes_v \pi_v$, the product being taken over all places of F, and $L(s, \pi)$ is an Euler product $\prod_v L(s, \pi_v)$. At a finite place $v=\mathfrak{p}$

$$L(s, \pi_v) = \prod_{i=1}^{n} \frac{1}{1-\alpha_i(\mathfrak{p})/N\mathfrak{p}^s}$$

is of degree n, and for almost all $\mathfrak{p}$ the matrix

$$A(\pi_v) = \begin{pmatrix} \alpha_1(\mathfrak{p}) & & 0 \\ & \ddots & \\ 0 & & \alpha_n(\mathfrak{p}) \end{pmatrix}$$

is invertible.

Since these L-functions, called standard, come in all degrees, there is no patently insurmountable obstacle to showing that each $L(s, E)$ is equal to some standard L-functions, thereby demonstrating the analytic continuation of $L(s, E)$. But the difficulties to overcome before this general principle of reciprocity is established are enormous, new ideas are called for, and little has yet been done.

If $F=\boldsymbol{Q}$ an automorphic representation of $\mathrm{GL}(2, \boldsymbol{A})$ is an ordinary automorphic form, analytic or non-analytic, in disguise, and the L-functions $L(s, \pi)$ have been with us for almost half a century. They were introduced and studied by Hecke **[20]**, and later defined for nonanalytic forms by Maass **[28]**. Moving from $n=1$ to $n=2$ does not give us much more latitude, but there are two obvious kinds of motivic L-functions of degree two.

If V is an elliptic curve then $L^1(s, V)$ is of degree two and the possibility that it would be equal to a standard L-function was first raised by Taniyama and later by Weil **[37]**, during his re-examination of Hecke's theory. The numerical evidence is good, but no theoretical progress has been made with the problem, except over function fields where it is solved **[10]**.

If ϱ is a two-dimensional representation of $\mathrm{Gal}(K/F)$ then the Artin L-function $L(s, \varrho)$ is of degree two. If ϱ is reducible or dihedral, Artin's theorem can deal with $L(s, \varrho)$. Otherwise the image of $\mathrm{Gal}(K/F)$ in $\mathrm{PGL}(2, \boldsymbol{C})=\mathrm{SO}(3, \boldsymbol{C})$ is tetrahedral, octahedral, or icosahedral. One example of an icosahedral representation with a reciprocity law has been found **[8]**, but no general theorems are available. I will return to the tetrahedral and octahedral below, after the principle of functoriality has been described.

The first successful applications of standard L-functions of degree two to the study of zeta-functions of algebraic varieties were for curves V obtained by dividing the upper half-plane by an arithmetic group, either a congruence subgroup of $\mathrm{SL}(2, \boldsymbol{Z})$ or a group defined by an indefinite quaternion algebra (**[13]**, **[32]**). Here $L^1(s, V)$ is a product of several $L(s, \pi)$ and the situation is similar to that for curves whose Jacobian is of CM-type, except that standard L-functions of degree two replace the Hecke L-functions, which are of degree one. The projections underlying the factorizations are linear combinations of the Hecke correspondences.

It is not surprising that these varieties were handled first, for they are defined by a group, and the mechanism which links their zeta-functions with automorphic L-functions is relatively simple, similar to that appearing in the study of cyclotomic extensions of the rationals. There is a great deal to be learned from the study of these varieties and their generalizations, the Shimura varieties, but there are no Shimura varieties attached to $\mathrm{GL}(n)$ when $n>2$, and we must pass to more general groups.

Automorphic L-functions and the principle of functoriality. If G is any connected, reductive group over a global field an automorphic representation of $G(A)$ is defined as for $\mathrm{GL}(n)$. The study of Eisenstein series led to a plethora of L-functions attached to automorphic representations. The Artin L-functions and the Hecke L-functions are fused in the class of automorphic L-functions, which contains them both, but the general automorphic L-function is in fact a kind of mongrel object, the true generalization of the Artin L-functions being the motivic L-functions and the true generalization of the Hecke L-functions being the standard L-functions.

To define the automorphic L-functions one associates to each connected, reductive group G over F an L-group ${}^LG={}^LG_F$ ([5], [25]), itself an extension

$$1 \to {}^LG^0 \to {}^LG \to \mathrm{Gal}(K/F) \to 1$$

with ${}^LG^0$ a connected, reductive, complex group. K is simply a finite but large Galois extension of F. To each continuous finite-dimensional representation ϱ of ${}^LG$ which is complex-analytic on ${}^LG^0$ and each automorphic representation π of $G(A)$ one attaches an L-function $L(s, \pi, \varrho)$, which is an Euler product of degree equal to the dimension of ϱ. There is evidence to support the hypothesis that each $L(s, \pi, \varrho)$ can be analytically continued to the whole plane as a meromorphic function with few poles and a functional equation.

The representation π is again a tensor product $\pi=\otimes_v \pi_v$ and

$$L(s, \pi, \varrho) = \prod_v L(s, \pi_v, \varrho).$$

For almost all finite v the theory of spherical functions, or, if one prefers, of Hecke operators, attaches to π_v a conjugacy class $\{g_v\}=\{g(\pi_v)\}$ in ${}^LG$ which reduces to the Frobenius class when $G=\{1\}$. The local factor for these places is

$$L(s, \pi_v, \varrho) = \frac{1}{\det\left(1-\varrho(g_v)/N\mathfrak{p}^s\right)}$$

if v is defined by $\mathfrak{p}$. If G is $\mathrm{GL}(n)$ then ${}^LG$ is a direct product $\mathrm{GL}(n, \mathbf{C})\times \times \mathrm{Gal}(K/F)$ and the projection of $\{g(\pi_v)\}$ on the first factor is the class of $A(\pi_v)$. Consequently if ϱ is the projection on the first factor then $L(s, \pi, \varrho)$ is the standard L-function $L(s, \pi)$.

The automorphic L-functions once defined, their resemblance to the Artin L-functions is manifest, and the possibility suggests itself of establishing their analytic continuation by showing that when $G, \pi,$ and ϱ are given there is a representation

π' of $\mathrm{GL}(n, A)$, with $n = \deg \varrho$, such that $\{A(\pi'_v)\} = \{\varrho(g(\pi_v))\}$ for almost all v and

$$L(s, \pi, \varrho) = L(s, \pi').$$

For $G = \{1\}$ this would be the reciprocity law for Artin L-functions.

More generally, if H and G are two connected reductive groups over F and we have a commutative diagram

$$\begin{array}{ccc} {}^LG & \searrow & \\ \big\downarrow{\scriptstyle\varphi} & & \mathrm{Gal}(K/F) \\ {}^LH & \nearrow & \end{array}$$

with φ complex-analytic, then to every automorphic representation π of ${}^LG$ there should be an automorphic representation π' of H which is such that $\{g(\pi'_v)\} = \{\varphi(g(\pi_v))\}$ for almost all v. There is evidence that this is so, although some subtleties must be taken into account. I refer to the phenomenon as the principle of functoriality in the L-group.

Examples. Suppose E is a finite extension of F. Then G is also a group over E and the L-group over E, ${}^LG_E$, is a subgroup of ${}^LG_F$. It is the inverse image of $\mathrm{Gal}(K/E)$ in ${}^LG_F$. The principle of functoriality implies the possibility of making a change of base from F to E and associating to each automorphic representation π of $G(A_F)$ an automorphic representation Π of $G(A_E)$, sometimes called a lifting of π. For almost all places, w, of E the class $\{g(\Pi_w)\}$ must be $\{g(\pi_v)^f\}$ if w divides the place v of F and $f = [E_w : F_v]$.

Ideas of Saito [30] and Shintani [34] allow us to show that base change is always possible when $G = \mathrm{GL}(2)$ and E is a cyclic extension of prime degree [26], and thus, by iteration, a solvable extension. For now extensions of prime degree are enough, and for them it is possible to characterize those Π which are liftings. The Galois group $\mathrm{Gal}(E/F)$ acts on A_E and on $\mathrm{GL}(2, A_E)$ and thus on the set of automorphic representations of $\mathrm{GL}(2, A_E)$. Apart from some trivial exceptions, Π is a lifting if and only if Π is fixed by $\mathrm{Gal}(E/F)$.

Base change is a first step towards a proof of the principle of reciprocity and Artin's conjecture for two-dimensional representations. Suppose, for example, that σ is a tetrahedral representation of $\mathrm{Gal}(\bar{F}/F)$. Then there is a cyclic extension E of F of degree three which is such that the restriction Σ of σ to $\mathrm{Gal}(\bar{F}/E)$ is dihedral. Consequently the principle of reciprocity applies to it and yields an automorphic representation $\Pi = \Pi(\Sigma)$ of $\mathrm{GL}(2, A_E)$. The class of Σ is invariant under $\mathrm{Gal}(E/F)$ and therefore Π is too, and is a lifting. There is precisely one representation π which lifts to Π and has central character $\det \sigma$. It should be $\pi(\sigma)$, the representation whose existence is demanded by the principle of functoriality. At first sight this does not look difficult to show, for the eigenvalues of $\sigma(\Phi_\mathfrak{p})$ and $\{A(\pi_v)\}$, where v is the place defined by $\mathfrak{p}$, differ only by cube roots of unity, but it should be a deep matter. However fortune smiles on us, for we can deduce some interesting theorems without pressing for a full understanding.

There are two ways of proceeding. The one used in [26] has the disadvantage that it does not work for all fields or all tetrahedral representations, but the advantage that it also works for some octahedral representations. It invokes a theorem of Deligne-Serre, characterizing some of the automorphic representations attached to two-dimensional representations of the Galois group. The other (cf. [15]) employs special cases of the principle of functoriality proved by Piatetskii-Shapiro and Gelbart-Jacquet.

One begins with Serre's observation to me that composition of σ with the adjoint representation φ of $GL(2)$ on the Lie algebra of $PGL(2)$ gives a three-dimensional monomial representation ϱ to which, by a theorem of Piatetskii-Shapiro [21], the principle of functoriality applies to yield an automorphic representation $\pi(\varrho)$ of $GL(3, A_F)$. On the other hand, the L-group of $GL(2)$ is a direct product $GL(2, \mathbf{C}) \times \mathrm{Gal}(K/F)$ and that of $GL(3)$ is a direct product $GL(3, \mathbf{C}) \times \mathrm{Gal}(K/F)$. The principle of functoriality should attach to the homomorphism

$$\varphi \times \mathrm{id}\colon GL(2, \mathbf{C}) \times \mathrm{Gal}(K/F) \to GL(3, \mathbf{C}) \times \mathrm{Gal}(K/F)$$

a map φ_* from automorphic representations of $GL(2, A_F)$ to automorphic representations of $GL(3, A_F)$. The existence of φ_* has been proven by Gelbart-Jacquet [16].

If the principle of functoriality is consistent and π is $\pi(\sigma)$ then $\varphi_*(\pi)$ must be $\pi(\varrho)$. Conversely, elementary considerations, which exploit the absence of an element of order six in the tetrahedral group, show that if $\varphi_*(\pi)$ equals $\pi(\varrho)$ then π is $\pi(\sigma)$. That $\varphi_*(\pi)$ equals $\pi(\varrho)$ follows easily from an analytic criterion of Jacquet-Shalika [22].

Even for $GL(2)$ base change for cyclic extensions is not proved without some effort, the principal tools being the trace formula and combinatorics of the Bruhat-Tits building. These are being developed by Arthur [1] and by Kottwitz [23], but our knowledge of harmonic analysis is still inadequate to a frontal attack on the problem of base change for a general group. Nonetheless some progress can be anticipated, although it is not clear how close base change will bring us to the Artin conjecture.

For number fields there has been no other recent progress with the principle of reciprocity. But we could also try to show that a motivic L-function is equal to an automorphic L-function $L(s, \pi, \varrho)$ which is not standard or to a product of such functions. This may not imply the analytic continuation of $L(s, E)$ but can have concrete arithmetic consequences and the proof may direct our attention to important features of the mechanism underlying the principles of reciprocity and functoriality [31].

The immediate examples are the L-functions defined by Shimura varieties [27]. These varieties are a rich source of ideas and problems, but once again we must advance slowly, deepening our understanding of harmonic analysis and arithmetic as we proceed. For the varieties associated to the group over Q obtained by restriction of scalars from a totally indefinite quaternion algebra over a totally real field F, the problems are tractable. In [27] no motives are mentioned, but the zeta-function is expressed as a quotient of products of automorphic L-functions of degree 2^n, where $n=[F:Q]$ is the dimension of the variety. For $n=2$ the analytic continuation and functional equation have been established by Asai [4], and we have the first examples of analytic continuation for motivic L-functions which are of degree four and, apparently, irreducible and not induced.

References

1. J. Arthur, *Eisenstein series and the trace formula*, in **6.**

2. E. Artin, *Über eine neue Art von L-Reihen*, Hamb. Abh. (1923), 89—108.

3. *Beweis des allgemeinen Reziprozitätsgesetzes*, Hamb. Abh. (1927), 353—363.

4. T. Asai, *On certain Dirichlet series associated with Hilbert modular forms and Rankin's method*, Math. Ann. **226** (1977), 81—94.

5. A. Borel, *Automorphic L-functions*, in **6.**

6. A. Borel and W. Casselman, eds., *Proceedings of AMS Summer Institute on L-functions and automorphic representations*, Corvallis, 1977.

7. A. Borel and H. Jacquet, *Automorphic forms and automorphic representations*, in **6.**

8. J. Buhler, *An icosahedral modular form of weight one*, Lecture Notes in Math., vol. 601, Springer-Verlag, Berlin and New York, 1977.

9. R. Dedekind, *Gesammelte Mathematische Werke*, F. Vieweg u. Sohn, Brunswi ck, 1930.

10. P. Deligne, *Les constantes des équations fonctionnelles des fonctions L*, Lecture Notes in Math., vol. 349, Springer-Verlag, Berlin and New York, 1973.

11. P. Deligne and J.-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. **4** (1971), 507—530.

12. V. G. Drinfeld, *Langlands conjecture for* GL (2) *over function fields*, these Proceedings.

13. M. Eichler, *Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion*, Arch. Math. **5** (1954), 355—366.

14. D. Flath, *Decomposition of representations into tensor products*, in **6.**

15. S. Gelbart, *Automorphic forms and Artin's conjecture*, Lecture Notes in Math., vol. 627, Springer-Verlag, Berlin and New York, 1978.

16. S. Gelbart and H. Jacquet, *A relation between automorphic forms on* GL (2) *and* GL (3), Proc. Nat. Acad. Sci. U.S.A. **73** (1976), 3348—3350.

17. R. Godement and H. Jacquet, *Zeta functions of simple algebras*, Lecture Notes in Math., vol. 260, Springer-Verlag, Berlin and New York, 1972.

18. H. Hasse, *Zetafunktion und L-Funktionen zu einem arithmetischen Funktionenkörper vom Fermatschen Typus*, Abh. Deutsch. Akad. Wiss. Berlin Kl. Math. Phys. Tech. **4** (1954).

19. E. Hecke, *Eine neue Art von Zetafunktionen und ihre Beziehungen zur Verteilung der Primzahlen*. I, Math. Z. **1** (1918). II, Math. Z. **6** (1920), 11—51.

20. *Über Modulfunktionen und die Dirichletschen Reihen mit Eulerschen Produktentwicklung*, I, II, Math. Ann. **114** (1937), 1—28, 316—351.

21. H. Jacquet, I. I. Piatetskii-Shapiro and J. Shalika, *Construction of cusp forms on* GL (3), Univ. of Maryland, 1975.

22. H. Jacquet and J. Shalika, *Comparaison des représentations automorphes du groupe linéaire* C. R. Acad. Sci. Paris **284** (1977), 741—744.

23. S. Kleiman, *Motives*, Algebraic Geometry, Oslo, 1970, F. Oort, ed., Wolters-Noordhoff, Groningen, 1972.

24. R. Kottwitz, Thesis, Harvard Univ., 1977.

25. R. Langlands, *Problems in the theory of automorphic forms*, Lecture Notes in Math., vol. 170, Springer-Verlag, Berlin and New York, 1970.

26. ———, *Base change for* GL (2), Notes IAS, 1975.

27. ———, *On the zeta-functions of some simple Shimura varieties*, Canad. J. Math.

28. H. Maass, *Über eine neue Art von nichtanalytischen automorphen Funktionen und die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **121** (1944), 141—183.

29. N. Saavedro-Rivano, *Catégories Tannakiennes*, Lecture Notes in Math., vol. 265, Springer-Verlag, Berlin and New York, 1972.

30. H. Saito, *Automorphic forms and algebraic extensions of number fields*, Lectures in Mathematics, Kyoto Univ., 1975.

31. D. Shelstad, *Notes on L-indistinguishability*, in **6**.

32. G. Shimura, *Correspondences modulaires et les fonctions ζ de courbes algébriques*, J. Math. Soc. Japan **10** (1958), 1—28.

33. G. Shimura and Y. Taniyama, *Complex multiplication of abelian varieties*, Publ. Math. Soc. Japan, Tokyo, 1961.

34. T. Shintani, *On liftings of holomorphic cusp forms*, in **6**.

35. J. Tate, *Fourier analysis in number fields and Hecke's zeta-functions*, Algebraic Number Theory, J. W. S. Cassels and A. Fröhlich, eds., Thompson, Washington, 1967.

36. A. Weil, *Number of solutions of equations in finite fields*, Bull. Amer. Math. Soc. **55** (1949), 497—508.

37. ———, *Über die Bestimmung Dirichletscher Reihen durch Funktionalgleichungen*, Math. Ann. **168** (1967), 149—156.

THE INSTITUTE FOR ADVANCED STUDY
PRINCETON, NEW JERSEY 08540, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Modular Forms and Number Theory

Ju. I. Manin

Introduction. Ladies and gentlemen, look at the emblem of our Congress:

You will easily recognize the design. This is part of the famous "modular configuration" consisting of (a) the Lobachevsky plane modeled on the complex unit disk; (b) the set of fundamental triangles for the theta-group bounded by Lobachevsky lines. To get the whole picture, draw the central triangle and perform consecutive reflections relative to its sides. Of course, the net becomes infinitely thick at the boundary ("Absolut"), and the designer wisely stopped not too far from the center. But the most interesting things happen far away. To understand them one could probe the Absolut as follows.

Choose two points α, β on the unit circle, draw the Lobachevsky line from α to β, and try to calculate the number $n(\alpha, \beta)$ of triangles it intersects on the way. This $n(\alpha, \beta)$ will be infinite unless α, β are cusps. What sort of function is $n(\alpha, \beta)$? The question might not sound particularly deep, but look at a sample of the number-theoretic facts obtained with the help of $n(\alpha, \beta)$ and its more sophisticated versions.

EXAMPLES. (a) Set

$$\Delta(z) = e^{2\pi iz} \prod_{n=1}^{\infty} (1 - e^{2\pi inz})^{24}.$$

Then (see [M2])

$$\Big(\int_0^{i\infty} \Delta(z)\,dz : \int_0^{i\infty} \Delta(z) z^2\,dz : \int_0^{i\infty} \Delta(z) z^4\,dz\Big) = \Big(1 : -\frac{691}{2^2 3^4 5} : \frac{691}{2^3 3^2 5\cdot 7}\Big). \qquad (1)$$

(b) Let p be a prime number. Consider the solutions of the Diophantine equation $p=\Delta\Delta'+\delta\delta'$, where $\Delta>\delta>0, \Delta'>\delta'>0$ or $\Delta=p, 0\leq\delta<p/2, \Delta'=1, \delta'=0$; and let $M_a(p)$ denote the number of such solutions with $\Delta/\delta\equiv\pm a \bmod 11$. Then for $p\neq 2, 11$ we have

$$N(p)/5 = 2M_2(p)+M_3(p)-2M_5(p)-M_4(p), \qquad (2)$$

where $N(p)$ is the number of solutions of the congruence $Y^2+Y\equiv X^3-X^2-10X-20 \bmod p$, including (∞,∞) (see [M4]).

(c) Again let p be a prime number. This time consider the elliptic curve $Y^2+Y=X^3-X$, and let N_p denote the number of its points $\bmod p$. Over Q its group of points is infinite cyclic with generator $(0,0)=x_0$. It follows that $N_p x_0$ reduces to infinity $\bmod p$; hence, the numerator $n(N_p x_0)$ of the X-coordinate of the point $N_p x_0$ is a p-adic unit, and the p-adic logarithm $\log_p n(N_p x_0)$ is well defined and is divisible by p.

Now define a function $\xi: Z/37Z \cup \{\infty\}\to\{0,\pm 1\}$:

$$\begin{aligned}\xi(a) &= 1 \quad \text{for } a\equiv\pm 5, \pm 7, \pm 8 \bmod 37,\\ &= -1 \quad \text{for } a\equiv\pm 14, \pm 15, \pm 16 \bmod 37,\\ &= 0 \quad \text{for the remaining values.}\end{aligned}$$

For each rational number a/b construct the series of its consecutive convergents $a/b=a_n/b_n, a_{n-1}/b_{n-1}, \dots, 0/1,$

$$\det\begin{vmatrix} a_i & a_{i-1}\\ b_i & b_{i-1}\end{vmatrix} = \pm 1,$$

and set

$$x^+(a/b) = \sum_{i=0}^{n-1} \xi(b_i/b_{i+1} \bmod 37).$$

With this notation, the last identity I wish to show you is:

$$\frac{2}{p}\log_p n(N_p x_0) \equiv \sum_{a=1}^{p-1} a \sum_{\varepsilon^{p-1}\equiv 1 \bmod p^2} x^+\Big(\frac{\varepsilon(1+ap)}{p^2}\Big) \mod p. \qquad (3)$$

This is in fact only conjectural and probably quite difficult. It has been numerically checked for $p=5, 7, 11, 13, 23, 29, 31$ by M. Rosenblum (Moscow University), who devised this conjecture, as a particular case of a "p-adic mod p" Birch–Swinnerton–Dyer conjecture, and who is now working on a full p-adic version. The primes

$p=2, 3, 17, 19$ are omitted because 2 and 3 are too small and 17, 19 are supersingular, so that some modifications are needed for such p's.

Explanations. Some explanations are now in order to show how all this is connected with Lobachevsky lines. First of all, change the model and consider the upper half-plane $H=\{z\in C \mid \operatorname{Im} z>0\}$ instead of the unit disk. The boundary is $R\cup\{i\infty\}$, and the cusps are $Q\cup\{i\infty\}$. Instead of the theta-group we will consider

$$\Gamma_0(N)=\left\{\begin{pmatrix} a & b\\ c & d\end{pmatrix}\middle| c\equiv 0 \bmod N\right\}\subset \mathrm{SL}(2, Z),$$

with the standard action $z\mapsto(az+b)/(cz+d)$ on H. Lobachevsky lines are semicircles orthogonal to the real axis (including vertical semilines). Instead of counting the intersections of such a line with triangles, we will proceed more invariantly. For each subgroup $\Gamma\subset\mathrm{SL}(2, Z)$ of finite index, define the Riemann surface $X_\Gamma^0=\Gamma\backslash H$ and its compactification by the (images of the) cusps: $X_\Gamma=X_\Gamma^0\cup P_\Gamma$. The image of the geodesic joining cusps α and β determines a homology class $\{\alpha, \beta\}_\Gamma\in H_1(X_\Gamma, P_\Gamma, Z)$. If $\Gamma\alpha=\Gamma\beta$, then we may even consider $\{\alpha, \beta\}_\Gamma$ as an element of $H_1(X_\Gamma, Z)$, since our geodesic projects onto a loop in X_Γ.

This "modular symbol" $\{\alpha, \beta\}_\Gamma$ is a good substitute for our initial $n(\alpha, \beta)$, since a fundamental domain for Γ is a union of a finite set of triangles, so that $\{\alpha, \beta\}_\Gamma$ also retains some additional information about the geometry of the consecutive intersections.

All of our examples (a), (b), (c) reflect two different ways of calculating $\{\alpha, \beta\}_\Gamma$. The first method—which gives the right sides of our identities—is quite elementary. It is based on two sorts of identities between modular symbols. From the definition it is clear that $\{g\alpha, g\beta\}_\Gamma=\{\alpha, \beta\}_\Gamma$ for all $g\in\Gamma$; hence, for $g\in\mathrm{SL}(2, Z)$ the symbol $\{g\alpha, g\beta\}_\Gamma$ depends only on the coset $g\Gamma$. The second identity is: $\{\alpha_1, \alpha_2\}_\Gamma+\{\alpha_2, \alpha_3\}_\Gamma+\ldots+\{\alpha_k, \alpha_1\}_\Gamma=0$ for an arbitrary sequence of cusps $\alpha_1, \ldots, \alpha_k$. In particular, if a/b, $c/d\in Q$ are Farey neighbors (that is, $ad-bc=1$), then $\{a/b, c/d\}_\Gamma=\{g(i\infty), g(0)\}_\Gamma$ depends only on $\Gamma g=\Gamma\left(\begin{smallmatrix} a & b\\ c & d\end{smallmatrix}\right)$. For $\Gamma=\Gamma_0(N)$ this coset is determined by $(c{:}d) \bmod N$. But for each $a/b\in Q$, $0\leqslant a/b<1$, we can find a sequence of Farey neighbors $a/b=a_n/b_n$, $a_{n-1}/b_{n-1}, \ldots, 0/1$, joining a/b to 0, for example, taking the series of convergents. This means that the class $\{a/b, 0\}_{\Gamma_0(N)}$ can be represented as a sum of classes $\xi(b_i{:}b_{i+1} \bmod N)=\{a_i/b_i, a_{i+1}/b_{i+1}\}_{\Gamma_0(N)}$. So you see how convergents enter into the picture in example (c). In example (b) they appear in disguise: in fact, the family of pairs (Δ, δ) taken from our solutions of $p=\Delta\Delta'+\delta\delta'$ coincides with the family of pairs of consecutive denominators of all convergents to the numbers $0<a/p\leqslant\frac{1}{2}$, $a\in Z$ (Heilbronn's lemma). No convergents can be seen directly in example (a), because it is only a particular case of a series of identities. I will comment upon it later (see **[M2]**, **[MF]**, **[B]**, **[Ma1]**, **[Ma2]**).

To explain the left sides of our identities we have to introduce modular forms at last. Let us dualize and consider $\{\alpha, \beta\}_\Gamma$ as a functional on the 1-cohomology of X_Γ. The (1, 0)-part of the de Rham cohomology is represented by everywhere regular holomorphic differentials ω on X_Γ. So we must calculate the integrals

$$\omega \mapsto \int_{\{\alpha,\beta\}_\Gamma} \omega = \int_\alpha^\beta \varphi(z)\,dz,$$

where $\varphi(z)dz$ is the inverse image of ω on H. Now the Γ-invariance of $\varphi(z)dz$ implies the functional equation

$$\varphi\left(\frac{az+b}{cz+d}\right) = \varphi(z)(cz+d)^2 \quad \text{for all } \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \Gamma;$$

and the regularity of ω means that $\varphi(z)dz$ is a cusp form of weight two in the classical sense.

First consider the integral $\int_{\{0,i\infty\}_\Gamma} \omega = \int_0^{i\infty} \varphi(z)\,dz$. Suppose for simplicity that $\left(\begin{smallmatrix} 1 & 1 \\ 0 & 1 \end{smallmatrix}\right) \in \Gamma$, for example $\Gamma = \Gamma_0(N)$. Then $\varphi(z) = \varphi(z+1)$, and we have the Fourier series $\varphi(z) = \sum_{n=1}^\infty a_n e^{2\pi inz}$ ($a_0 = 0$, since φ is a cusp form). Set $L_\varphi(s) = \sum_{n=1}^\infty a_n n^{-s}$. This series is convergent for $\operatorname{Re} s \gg 0$, and the classical Mellin transform

$$L_\varphi(s) = \frac{(2\pi)^s}{i\Gamma(s)} \int_0^{i\infty} \varphi(z) \left(\frac{z}{i}\right)^{s-1} dz$$

defines the analytic continuation of $L_\varphi(s)$ on the whole s-plane, because the functional equation for $\varphi(-1/Nz)$ $(\Gamma = \Gamma_0(N))$ shows the convergence of the integral for all s. In fact, this functional equation for φ even gives the usual functional equation for $L_\varphi(s)$ relating it to $L_\varphi(1-s)$.

In terms of this L-function we can write

$$\int_0^{i\infty} \varphi(z)\,dz = \int_{\{0,i\infty\}_\Gamma} \omega = \frac{i}{2\pi} L_\varphi(1).$$

It is now easy to explain how prime numbers enter into our picture. The point is that the whole space of modular forms of weight two, say, for $\Gamma_0(N)$, has a basis $\{\varphi_i\}$ such that each L_{φ_i} admits an Euler product with factors $(1 - a_p p^{-s} + p^{1-s})^{-1}$ (if $p \nmid N$). Actually, this space is acted upon by the Hecke operators

$$T_n = \sum_{\substack{d|n \\ b \bmod d}} \begin{pmatrix} nd^{-1} & b \\ 0 & d \end{pmatrix} \in Z[\mathrm{GL}(2, R)],$$

and L_φ has such an Euler product if and only if $T_n\varphi = a_n\varphi$, $n = 1, 2, \ldots,$ where a_n are the Fourier coefficients of φ. So we have

$$\int_0^{i\infty} T_n(\varphi)\,dz = a_n \int_0^{i\infty} \varphi(z)\,dz = a_n \int_{\{0,i\infty\}} \omega = \int_{T_n^*\{0,i\infty\}} \omega.$$

But the conjugate action of T_n^* on the space of all modular symbols is easily calculated with the help of our continued fractions technique. By the way, the same calculations show that $\{\alpha, \beta\}_\Gamma$ as a functional on $H^0(X_\Gamma, \Omega^1)$ lies in $H_1(X_\Gamma, Q)$ even if $\Gamma\alpha \neq \Gamma\beta$.

In the case $\Gamma = \Gamma_0(11)$ the curve X_Γ is defined over Q and has genus one; hence, there is a unique modular form $\varphi(z)$ for Γ normalized by the condition $a_1 = 1$. This $\varphi(z)$ is then automatically an eigen-function for all T_n's, and, after some computation, we get

$$1 - a_p + p = \text{the right side of (2).}$$

Now comes the main point: $1 - a_p + p$ is equal to the number of points on the curve $X_\Gamma \bmod p$, or in other words, $L_\varphi(s)$ is the Hasse–Weil zeta-function of X_Γ. This is a particular instance of a very important general principle to which I will return later. In our example (b) the equation of $X_{\Gamma_0(11)}$ is just $Y^2 + Y = X^3 - X^2 - 10X - 20$; hence we get (2).

The relations (3) arise in a similar setting but are much deeper. We take here $\Gamma = \Gamma_0(37)$. Then X_Γ is of genus two, but has a factor Y of genus one given by the equation $Y^2 + Y = X^3 - X$; and we construct φ and L_φ corresponding to this factor. Again $L_\varphi(s)$ coincides with the Hasse–Weil zeta of Y, but this time $L_\varphi(1) = 0$! This agrees with the Birch–Swinnerton–Dyer conjecture, according to which the rank of the group of rational points of Y should be equal to the order of zero of $L_\varphi(s)$ at $s = 1$. Both numbers are 1 in our case. But the conjecture goes much farther and also predicts the coefficient A in the expansion $L_\varphi(s) = A(s-1) + \dots$. This coefficient involves such oddities as the order of the Shafarevich–Tate group of Y and, moreover, the logarithm of the Néron–Tate height of a generator of $Y(Q)$. At the present time, however, we are unable to compare $A = L'_\varphi(1)$ with the Birch–Swinnerton–Dyer expression for even a single curve. What we can do is: (a) define a p-adic version $L_{\varphi,p}$ of L_φ and (b) calculate $L'_{\varphi,p}(1) \bmod p^n$ in this version in terms of modular symbols. So the question arises whether the Birch–Swinnerton–Dyer conjecture has a reasonable p-adic analog. The main trouble is how to define the p-adic height, and Rosenblum's calculations give at least a strong indication that a reasonable definition actually exists and fits into the p-adic version of the conjecture.

We now turn to example (a). The main difference between (a) and the previous examples stems from the fact that $\Delta(z)$ is a cusp form of weight 12, not 2, for the full modular group $\mathrm{SL}(2, Z)$. In general, a modular form φ of weight $w+2$ satisfies the functional equation

$$\varphi\left(\frac{az+b}{cz+d}\right) = \varphi(z)(cz+d)^{w+2}.$$

So $\varphi(z)dz$ is not an invariant differential, only $\varphi(z)(dz)^{(w+2)/2}$ is. One can still use the continued fractions technique, properly modified. All the integrals

$\int_0^{i\infty} \varphi(z) z^k dz$, $0 \leqslant k \leqslant w$, will enter in our formulas, that is, all the values $L_\varphi(k+1)$, $1 \leqslant k+1 \leqslant w+1$; and, using Hecke operators, we will obtain sufficiently many linear relations between these values to obtain the identity (1).

But it is instructive to enlarge the geometric picture in order to be able to interpret cusp forms of arbitrary weight as honest differential forms. To do this, recall that H parametrizes elliptic curves. More precisely:

(a) Form the direct product $H(1) = H \times C$ and note that its fibre C_z over a point $z \in H$ contains the lattice $Z + Zz$, which, after factorization, determines the elliptic curve $E_z = C_z/(Z+Zz)$; moreover, $E_z \simeq E_{z'}$ iff $gz = z'$ for some $g \in \mathrm{SL}(2, Z)$.

(b) Extend the group $\Gamma \subset \mathrm{SL}(2, Z)$ to the group $\Gamma(1)$, the semidirect product of G and Z^2:

$$\Gamma(1) = \{[g; m, n] | g \in G;\ m, n \in Z\};$$

$$[g'; m', n'][g; m, n] = [g'g; (m', n')g + (m, n)].$$

(c) Extend the action of Γ on H to the following action of $G(1)$ on $H(1)$:

$$[g; m, n](z, u) = \left(\frac{az+b}{cz+d};\ \frac{u+mz+n}{cz+d}\right),\quad z \in H,\ u \in C.$$

This action takes C_z to $C_{g(z)}$ and $Z+Zz$ to $Z+Zg(z)$.

(d) Take $w \geqslant 0$, and form the two fibre products

$$\Gamma(w) = \Gamma(1) \times_{\Gamma} \dots \times \Gamma(1) \quad \text{and} \quad H(w) = H(1) \times_{H} \dots \times H(1) \ (w \text{ times}).$$

This $\Gamma(w)$ acts naturally on $H(w)$.

Now suppose $\varphi(z)$ is a modular form of weight $w+2$ on H for Γ. Then in the natural coordinates $(z, u_1, \dots, u_w)$ on $H(w)$ the differential form

$$\omega_\varphi = \varphi(z)\, dz \wedge du_1 \wedge \dots \wedge du_w$$

is $G(w)$-invariant, and so determines a meromorphic form on $G(w) \backslash H(w)$. If $\varphi(z)$ is a cusp form, then ω_φ extends to a regular algebraic form on a smooth compac- $B_\Gamma(w)$ of $\Gamma(w) \backslash H(w)$. This latter variety is called the Kuga variety of weight tification $w+2$. It is fibered by abelian varieties over X_Γ, at least if Γ is torsion free. Over each of the Lobachevsky lines on H connecting two cusps there lies a family of $(w+1)$-chains on $H(w)$ which, after projecting onto $B_\Gamma(w)$, determines a family of higher modular symbols

$$\{\alpha, \beta;\ m, n\} \in H_{w+1}(B_\Gamma(w), P_\Gamma(w), Z),\ m, n \in Z^w,$$

where $P_\Gamma(w)$ is made up of fibres over cusps. This construction was studied in detail in Shokurov ([Sh1], [Sh2]), and was generalized to Hilbert modular varieties in Manin [M5] (but the definitions given there require some corrections). The L-series $L_\varphi(s)$ is intimately connected with the zeta-function of $B_\Gamma(w)$ for any w.

It is time to finish the first round of explanations. I have reversed the history of our subject and made a mess of its logic in order to show you some recent and beautiful number theory as quickly as possible. Now let me be more systematic and review the fundamental structures. Broadly speaking, we have three large classes of objects—algebraic varieties, modular forms and L-functions—and a lot of interrelations between them, some proved but most conjectural. Most of number theory lies in these interrelations, and I will try to concentrate on them.

L-functions. We now know two great classes of L-functions: "algebro-geometric" and "automorphic" ones. The first is connected with schemes, motives and l-adic representations; the second with Mellin transforms of automorphic forms and with Lie and adelic group representations. Each new equality between two L-functions belonging to the different classes brings a list of concrete number theoretic facts. Instead of stating current conjectures about these identifications, I will try to explain why this is so. The main reason is that by their origin these two classes are equipped with complementary (almost disjoint) lists of properties. Namely, for algebro-geometric L-functions we know:

(a) Euler products with a down-to-earth interpretation of p-factors in terms of solutions of congruences.

(b) Deligne's theorem about the absolute values of the roots of these p-factors, which gives precise bounds for many "remainder terms" in classical asymptotic formulas, and in particular, for trigonometric sums.

(c) l-adic and p-adic cohomological interpretations of those p-factors, which give a good grasp of the divisibility properties of L-series coefficients.

(d) Very important but still rather scarce connections between the global Diophantine properties of the scheme in question and the analytic behavior of its L-function. Here I have in mind such classics as Dirichlet's class-number/regulator formula, Riemann's explicit formula for $\pi(x)$, Hardy–Littlewood's "singular series" for the Waring problem, and also more recent developments such as Tamagawa numbers and the Birch–Swinnerton–Dyer conjecture.

On the other hand, for automorphic L-functions we can usually obtain:

(e) Analytic continuation and functional equation.

(f) Formulas for the values at certain "critical" points. Recently, these values have been used for p-adic interpolation.

(g) Some information about zeros and poles.

(h) Asymptotic behavior and exact formulas for the coefficients based on the behavior of the automorphic form near the boundary.

We list some examples of these identifications of algebro-geometric with automorphic L-functions and their arithmetic consequences.

Riemann's $\zeta(2s)$ is the Mellin transform of $\theta(z)=\sum_{n=1}^{\infty} e^{2\pi i n^2 z}$, which is an automorphic form of weight 1/2. This gives one of the most natural classical proofs of the functional equation for ζ.

Hecke's L-functions with Grössencharakters are Mellin transforms of special modular forms. Hecke used this fact again for deriving functional equations and for obtaining asymptotic formulas for the distribution of prime ideals in "sectors".

The L-functions of modular curves are (products of) Mellin transforms of cusp forms of weight two for congruence subgroups of $\mathrm{SL}(2, Z)$. Eichler first observed this fact and used it to obtain the best remainder term for representations of integers by quaternary quadratic forms. These remainder terms are just the coefficients of cusp forms, and, as soon as we know their connection with congruences and the Hasse–Weil–Riemann hypothesis for curves over finite fields, we obtain the necessary bounds. Taniyama and Weil conjectured that the L-functions of all elliptic curves over Q can be obtained in this way.

Using this identification and taking for granted the Birch–Swinnerton–Dyer and Taniyama–Weil conjectures, the author has been able to derive explicit formulas for the order of the highly mysterious Tate–Shafarevich groups of modular curves over a tower cyclotomic fields, and also a conditional algorithm for calculating the generators of the group of points of an elliptic curve over Q. The algorithm is very practical, can be applied to any particular elliptic curve, and always ends up giving a finite set of points on this curve. The conjectures are used only to prove that this set generates the whole Mordell group (see [M6]).

Ramanujan's τ-function is the sequence of coefficients of $\Delta(z)$ in our example (a). The Mellin transform of $\Delta(z)$ is a divisor of the zeta-function of the Kuga variety $B_{\mathrm{SL}(2,Z)}(10)$. Deligne's proof of the Riemann hypothesis for all varieties over finite fields resulted in the proof of the Ramanujan–Petersson conjecture: $|\tau(p)|<2p^{11/2}$. Using the l-adic representation of this zeta-function, Serre conceptually explained the beautiful Ramanujan congruence $\tau(p)\equiv 1+p^{11} \bmod 691$ and vastly generalized it.

Deligne and Serre proved that all cusp forms of weight 1 which are "new forms of odd character" are inverse Mellin transforms of Artin L-series corresponding to irreducible two dimensional representations of $\mathrm{Gal}(\bar{Q}/Q)$ taking complex conjugation to $\left(\begin{smallmatrix}1 & 0\\ 0 & -1\end{smallmatrix}\right)$. Langlands used this result to prove Artin's conjecture for a new class of Artin L-series (see [DS]).

All of the examples in this list of identifications are embedded into a vast and fascinating program which is called the "Langlands philosophy". Here I cannot do more than pay it lip service; see [B], [MF], [D1]–[D3] for thorough and technical expositions. My aim was only to underline its number theoretic implications, which go beyond the widely advertised and very important "non-commutative class field theory".

Varieties. Algebro-geometric L-functions are defined in terms of varieties and schemes, and their cohomology of various kinds forms the main link between these L-functions and automorphic L-functions. Two points here deserve the particular attention of a number theorist.

(a) The varieties carrying modular forms considered up to now have a very special structure, and are usually families of abelian varieties like Kuga's $B_\Gamma(w)$ or Shimura varieties. The problem of their universality in the sense of Taniyama–Weil–Langlands... is arithmetic, not analytic, since it is deeply connected with the arithmetic of global ground fields. The same varieties have recently been studied in detail from the Diophantine point of view (see **[Ma3]**).

(b) The deepest structure which seems to underlie all number theoretic applications of the theory is that of Grothendieck's "motives". A tentative definition is that a motive is a "twisted direct summand" of an algebraic variety: see **[M7]** and **[De]**. Motives serve as a universal cohomology of algebraic varieties, of which different realizations exist, such as de Rham cohomology, Hodge cohomology, l-adic cohomology, crystalline cohomology, a host of p-adic cohomologies. Each realization gives some number theoretic information, and the underlying motive serves as a link between them. From this point of view, the modular symbols should be considered as a specific structure on the Q- or Z-cohomology of the universal varieties mentioned above. This explains their role as a new source of arithmetic facts.

Finally, I should say that I left a lot of topics untouched. Among the most interesting is the theory of Siegel modular forms, which has recently been studied in depth by Andrianov. I would also like to mention I. M. Gel'fand's suggestion that the ζ-functions of certain special differential operators should have an arithmetic meaning. The first class to consider is that of the Schrödinger operators $-d^2/dx^2+u(x)$ with algebro-geometric potentials $u(x)$ arising as solutions of the Korteweg–deVries equation, for example: $u(x)=2\wp(x)$, where $\wp$ is the Weierstrass function of an elliptic curve over Q. In fact, it seems that the values of this zeta-function at negative integers, which can be calculated explicitly, admit a p-adic interpolation.

References

The dominant moods in modular function theory of one variable are well documented in

[Shi] G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Iwanami and Princeton, 1971.

[MF] *Modular functions of one variable*. I—VI, Lecture Notes in Math., vol. 320, 1973; vol. 349, 1973; vol. 350, 1973; vol. 476, 1975; vol. 601, 1977; vol. 627, 1977, Springer-Verlag, Berlin and New York.

This should be compared with the state of the art in Siegel modular form theory, for which we refer to

[A1] A. N. Andrianov, *Euler products for Siegel modular forms of genus* 2, Uspehi Mat. Nauk **29** (1974), 73—110. (Russian)

[A2] ———, *Euler products of theta transforms for Siegel modular forms of genus n*, Mat. Sb. **105** (1978), 291—341, and the references therein. For Langlands' philosophy, see the (Russian) report.

[B] A. Borel, *Formes automorphes et séries de Dirichlet*, Séminaire Bourbaki, no. 466, 1975, and the references therein. See also Deligne's paper in [MF] and the recent publications:

[D1] V. G. Drinfeld, *Elliptic moduli*, Mat. Sb. **94** (1974), 594—627. (Russian)

[D2] Elliptic moduli II. Mat. Sb. **102** (1977), 182—194. (Russian)

[D3] *A proof of Petersson's conjecture over functional fields*, Funkcional Anal. i Priložen **11** (1977), 11—14. (Russian)

[T] J. B. Tunnell, *On the local Langlands conjecture for* GL(2), Invent. Math. **46** (1978), 179—200.

[DS] P. Deligne and J.-P. Serre, *Formes modulaires de poids* 1, Ann. Sci. École Norm. Sup. **4** (1974), 507—530.

The theory of modular symbols is developed and applied in

[B] B. Birch, *Elliptic curves, a progress report*, Amer. Math. Soc. Conf. Number Theory, Stony Brook, Amer. Math. Soc., Providence, R. I., 1969, pp. 396—400.

[M1] Ju. I. Manin, *Parabolic forms and zeta-functions of modular curves*, Izv. Akad. Nauk SSSR **6** (1972), 19—64.

[M2] *Periods of parabolic forms and p-adic Hecke series*, Math. Sb. **21** (1973), 371—393.

[M3] *Values of p-adic Hecke series at integer points of the critical strip*, Math. Sb. **95** (1974), 357—383. (Russian)

[M4] *Explicit formulas for the eigen-values of Hecke operators*, Acta Arith. **24** (1973), 239—249.

[M5] *Non-archimedean integration and p-adic Jacquet-Langlands L-functions*, Uspehi. Mat. Nauk **31** (1976), 5—54.

[M6] *Cyclotomic fields and modular curves*, Uspehi. Mat. Nauk **26** (1971), 7—71 = Russian Math. Surveys **26** (1971), 7—78.

[Ma1] B. Mazur and P. Swinnerton-Dyer, *Arithmetic of Weil curves*, Invent. Math. **25** (1974), 1—16.

[Ma2] B. Mazur, *p-adic analytic number theory of elliptic curves and abelian varieties over Q*, Proc. Internat, Congr. Math. (Vancouver 1974, vol. I, pp. 369—378.

[Ma3] *Modular curves and the Eisenstein ideal*, Inst. Hautes Études Sci. Publ. Math. **47** (1977), 33—186.

[K1] P. Kurchanov, *Values of Jacquet-Langlands L-Series in integer points* (preprint).

[K2] *Local measures connected with cusp forms of Jacquet-Langlands over CM-fields* (preprint).

[H] K. Hatada, *Eigenvalues of Hecke operators on* SL (2, Z), University of Tokyo preprint, 1978.

[Sh1] V. Shokurov, *Modular symbols of higher weight*, Funkcional Anal. i Priložen, **7** (1975), 55—56. (Russian)

[Sh2] *Periods of cusp forms and Kuga varieties*, Uspehi Mat. Nauk **30** (1975), 183—184. (Russian)

On the subject of congruences, p-adic L-functions and values of L-functions, see also the following articles, which have extensive bibliographies:

[S] J.-P. Serre, *Divisibilité de certaines fonctions arithmétiques*, Enseignement Math. **22** (1976), 227—260.

[Ka] N. M. Katz, *p-adic L-functions for CM-fields*, Harvard preprint, 1978.

[D] P. Deligne, *Valeurs de fonctions L et périedes d'intégrales*, Inst. Hautes Études Sci., preprint 1977.

Motives are explained in

[M7] Ju. I. Manin, *Correspondences, motives and monoidal transformations*, Mat. Sb. **6** (1968), 439—470.

[De] M. Demazure, *Motifs des variétés algébriques*, Séminaire Bourbaki, no. 365, 1971.

Mathematical Institute of the Academy of Science of the USSR
Moscow, U.S.S.R.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Linear Operators and Integrable Hamiltonian Systems

S. P. Novikov

I. Examples of nonlinear wave systems. (KdV, Schrödinger nonlinear equation, Sine–Gordon and others.) The history of the problem: works of the 19th and early 20th centuries; discovery of the inverse scattering method for the solution of KdV (Kruskal, Green, Gardner, Miura, 1967). Lax representation (1968). Hamiltonian formalism of Gardner–Zakharov–Faddeev (1971). Zakharov–Shabat (1971). Ablovitz–Kaup–Newell–Segur cases (1973) and others. Exact solutions (rapidly-decreasing case). Toda lattice (Henon, Flashka, Manakov, 1974). Moser–Calogero particle systems (1975).

II. Periodic and quasiperiodic solutions. Works by the reporter, Dubrovin, Matveev, Its (1974), Lax, McKean and Van Moerbeke (1975). Equations of the zero curvature and KdV. Hamiltonian formalism in stationary problems. Finite-zone periodic potentials, Hamiltonian systems and classical algebraic geometry of the Riemannian surfaces. Baker–Achiezer function. Finite-zone solutions of the KdV equation. Particular cases. Elliptical solutions. Rational solutions. Theorem by Moser, McKean and Airault (1976). Multi-dimensional analogue of the rigid body. Manakov theorem (1976).

III. Multi-dimensional problems. Kadomtsev–Petviashvili equation (two-dimensional KdV). Zakharov–Shabat method (1974). Krichever solutions (1976). Two-dimensional solitons of Petviashvili–Its et al (1977). Rational solutions (1977). Solutions of the reporter and Krichever depending on functional parameters and holomorphic vector bundles over the Riemannian surfaces (1978).

IV. Commutative algebras of operators. One-dimensional case. Works of the 20s (Burchnall, Chaundy, Baker). Krichever theorems (1975–76). Operators of non-rel.

prime orders (1977). Holomorphic vector bundles. Algebras related to the two-dimensional Schrödinger operators (the reporter, Krichever, Dubrovin, 1976). Further problems.

References

1. B. Dubrovin, V. Matveev and S. Novikov, Soviet Mat. Uspehi **1** (1976), (Russian), 55—136.
2. I. Krichever, Soviet Mat. Uspehi **6** (1977), (Russian), 20—31.
3. F. Calogero (editor), *Nonlinear evolution equations solvable by spectral transform*, Putman, New York, 1978 (collection of surveys)

STEKLOV INSTITUTE OF MATHEMATICS
Moscow 177333 USSR

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Complex Geometry of the Natural World

Roger Penrose

This century has seen two major revolutions in physical thought. The first of these, relativity, uprooted earlier ideas of the nature of time and space, and provided us with our present picture of the world as a real differentiable manifold of dimension four, possessing a pseudo-Riemannian metric with a $(+ - - -)$ signature. The second revolution, quantum theory, altered our picture of things yet more radically than did relativity—even to the extent that, as we were told, it became no longer appropriate to form pictures at all, in order to give accurate representations of physical processes on the quantum scale. And, for the first time, the complex field $\boldsymbol{C}$ was brought into physics at a fundamental and universal level, not just as a useful or elegant device, as had often been the case earlier for many applications of complex numbers to physics, but at the very basis of physical law. Thus, the allowable physical states were to form a complex vector space, in fact a Hilbert space. So, on the one hand, we had the real-manifold picture of space-time geometry, and on the other, the complex-vector space view, according to which geometrical pictures were deemed inappropriate.

This conflict has remained with us since the conceptions of these great theories, to the extent that, even now, there is no satisfactory union between the two. Even at the most elementary level, there are still severe conceptual problems in providing a satisfactory interpretation of quantum mechanical observations in a way compatible with the tenets of special relativity. And quantum field theory, which represents the fully special-relativistic version of quantum theory, though it has had some very remarkable and significant successes, remains beset with inconsistencies and divergent integrals whose illeffects have been only partially circumvented. Moreover, the present status of the unification of *general* relativity with quantum mechanics remains merely a collection of hopes, ingenious ideas and massive but inconclusive calculations.

In view of this situation it is perhaps not unreasonable to search for a different viewpoint concerning the role of geometry in basic physics. Broadly speaking, "geometry", after all, means any branch of mathematics in which pictorial representations provide powerful aids to one's mathematical intuition. It is by no means necessary that these "pictures" should refer just to a spatio-temporal ordering of physical events in the familiar way. And since C plays such a basic universal role at the primitive levels of physics at which quantum phenomena are dominant, one is led to expect that the primitive geometry of physics might be complex rather than real. Moreover, the macroscopic geometry of relativity has many special features about it that are suggestive of a hidden complex-manifold origin, and of certain deep underlying physical connections between the normal spatio-temporal relations between things and the complex linear superposition of quantum mechanics.

One of the most striking and elementary of these physical connections appears in the geometry of the spin states of an ordinary spin $\frac{1}{2}\hbar$ particle (such as an electron or proton). Each pure state corresponds to a direction in space, about which we may regard the particle as spinning in a right-handed sense. But these states refer also to the rays through the origin of a 2-complex-dimensional Hilbert space. Thus we have an unambiguous correspondence between the complex directions of the Hilbert 2-space and the real directions of ordinary Euclidean 3-space. The set of complex Hilbert-space directions has a natural structure as a complex-projective (Kähler) 1-manifold CP^1, while the real Euclidean directions constitute the ordinary 2-sphere S^2. The essential identity between these two spaces provides a link between, on the one hand, the fact that it is *complex* linear superposition that occurs in quantum mechanics and, on the other, that ordinary space is 3-dimensional! In terms of the symmetry groups involved, this relation finds expression in the special local isomorphism

$$SU(2) \to SO(3)$$

(which is a 2-1 map).

The group $SO(3)$ refers only to spatial symmetry. According to relativity, time must be brought in also, and the local symmetry about a point in space-time becomes the Lorentz group, whose connected component $O_+^\uparrow(1, 3)$, being a complex group, has an even more intimate relation to C than has $SO(3)$, the relativistic version of the above local isomorphism being

$$SL(2, C) \to O_+^\uparrow(1, 3)$$

(which is again a 2-1 map). We may again think of these groups as acting on CP^1 ($\cong S^2$), but where now we are concerned with *all* holomorphic (i.e. complex-analytic) maps of CP^1 to itself (the Kähler structure being now dropped), giving precisely $O_+^\uparrow(1, 3)$. Indeed, we can (literally!) *see* this CP^1 as the celestial sphere or "sky" of an observer [1]. Two observers who pass one another at the same event O, have their skies related, according to the relativistic transformation law, by a conformal map, i.e. by a holomorphic map, if we regard the sky as CP^1.

In space-time terms this CP^1 is to be regarded as the set of light rays through the point O. Thus it is in the *light-ray* structure of space-time that complex geometry begins to emerge. (The relation to quantum superposition is, for the moment, less evident, but it reappears at a later stage.) Let us explore this further by considering the space $\mathcal{N}$ of light rays in ordinary Minkowski (i.e. flat pseudo-Euclidean $(+\ -\ -\ -)$ signature) 4-space M'. Since $\mathcal{N}$ is 5-real-dimensional it cannot be a complex manifold. Nevertheless, it is very close to being one and has the structure of what is known as a CR-hypersurface [2], [10]. This structure may be viewed as that which is induced on $\mathcal{N}$ by its imbedding as a real hypersurface in a complex manifold—and, in this case, the ambient complex manifold can, in fact [3], [10], be taken to be CP^3. The tangent space D to a point of $\mathcal{N}$ can be thought of as the family of light rays that lie in the immediate neighbourhood of a given light ray. In D are certain real 2-planes, namely those, called *holomorphic directions,* that are spanned by the real and imaginary parts of complex vectors in CP^3. Together, these sweep out a 4-real-dimensional or 2-complex-dimensional subspace of D, called the holomorphic tangent space. This represents light rays that are displaced from the given one by a small amount orthogonal to the direction of motion. A certain 1-real-parameter family of holomorphic directions arises from the system of light cones of the points of the given light ray. This family provides the directions along which the *Levi form* [2] (which measures the holomorphic convexity properties of the hypersurface $\mathcal{N}$ in CP^3) vanishes. The Levi form eigenvalues for $\mathcal{N}$ have signs $(+\ -)$.

The space CP^3, which in this context is referred to as the *projective twistor space* for M', can be constructed in the following way [3], [5], [6], [7]. Let CM' be the standard complexification of M', and let $(z^0, z^1, z^2, z^3) \in C^4$ be standard coordinates for a point $z \in CM'$, the complex metric being $ds^2 = (dz^0)^2 - (dz^1)^2 - (dz^2)^2 - (dz^3)^2$. Now consider the matrix equation

$$\begin{pmatrix} Z^0 \\ Z^1 \end{pmatrix} = \frac{i}{\sqrt{2}} \begin{pmatrix} z^0 + z^1 & z^2 + iz^3 \\ z^2 - iz^3 & z^0 - z^1 \end{pmatrix} \begin{pmatrix} Z^2 \\ Z^3 \end{pmatrix},$$

where the ratios $Z^0 : Z^1 : Z^2 : Z^3$ provide coordinates for the points of CP^3. Holding these ratios fixed, the solutions of this equation give (provided that Z^2 and Z^3 do not both vanish) a complex-2-plane along which the induced metric vanishes. Such a plane is called *totally null,* and the particular ones arising, as above, from *points* of CP^3 are called *α-planes.* There is also another family of totally null planes in CM', called β-planes. These arise in a corresponding way from the various *planes* (CP^2's) in CP^3. If an α-plane (or a β-plane) has one real point in CM' (i.e., in M'), then it has a whole light ray of real points, the light ray then determining the α-plane (and its complex conjugate, the β-plane) uniquely. In this case the Hermitian form

$$2\Sigma(Z^\alpha) \equiv |Z^0 + Z^2|^2 + |Z^1 + Z^3|^2 - |Z^0 - Z^2|^2 - |Z^1 - Z^3|^2$$

vanishes. The equation $\Sigma(Z^\alpha)=0$, in $\boldsymbol{CP}^3$, defines $\mathcal{N}$ (which we take compact $\cong S^2\times S^3$).

Now suppose instead that, in the above matrix equation, z^a is held fixed and Z^α allowed to vary. Then we obtain a 2-complex-dimensional subspace of $\boldsymbol{C}^4$, i.e. a complex *line* ($\boldsymbol{CP}^1$) in $\boldsymbol{CP}^3$. Therefore, points of $\boldsymbol{C}M'$ are represented as lines in $\boldsymbol{CP}^3$. In fact the only lines of $\boldsymbol{CP}^3$ that do not so represent points of $\boldsymbol{C}M'$ are the lines meeting the particular line I in $\mathcal{N}$, given by $Z^2=Z^3=0$. To include these lines, we adjoin extra points to $\boldsymbol{C}M'$ to obtain a *compact* manifold $\boldsymbol{C}M$, this being the standard Klein (or Grassmann) representation [4] of lines in $\boldsymbol{CP}^3$. The *real* points of $\boldsymbol{C}M$ define the *compactified Minkowski space* M ($\supset M'$) and this is represented by the family of lines *in* $\mathcal{N}$. Thus, the light rays through a point O of M' correspond, in $\boldsymbol{CP}^3$, to the points of the line in $\mathcal{N}$ which represents O. Since this line has the holomorphic structure of $\boldsymbol{CP}^1$, we find agreement with the earlier discussion concerning the holomorphic structure of the light cone or sky of an observer at O.

Let $\boldsymbol{PT}^+$ and $\boldsymbol{PT}^-$ denote the portions of $\boldsymbol{CP}^3$ for which $\Sigma(Z^\alpha)$ is, respectively, positive or negative. Then the group of holomorphic self-transformations of $\boldsymbol{PT}^+$ is just the nonreflective conformal group $C_+^\uparrow(1,3)$ on M. This is related to the appropriate pseudo-unitary group on $\boldsymbol{C}^4$ (preserving $\Sigma(Z^\alpha)$) via the 4-1 local isomorphism

$$SU(2,2)\to C_+^\uparrow(1,3).$$

The Poincaré (or inhomogeneous Lorentz) group is a 10-real-parameter subgroup of $C_+^\uparrow(1,3)$ preserving I and a certain scaling for I.

The idea of the *twistor programme* [7] is to re-express the whole of basic physics in terms of the above space $\boldsymbol{CP}^3$ or the space $\boldsymbol{C}^4$ from which it arises. More correctly, however, twistor space should not be regarded as $\boldsymbol{C}^4$ (which is the coordinate space) but as a more directly physically defined space $\boldsymbol{T}$ whose elements (up to phase multipliers) correspond to *massless particles with spin.* The helicity is given by the value $\frac{1}{2}\hbar\Sigma(Z^\alpha)$. Moreover, it is envisaged that when the space-time is curved, or endowed with an electromagnetic field, or a more general gauge field, [8], the twistor space should acquire a more complicated complex structure than that of $\boldsymbol{C}^4$. Some considerable successes have been achieved, as it turns out, using this kind of method to describe such fields, in the cases for which the fields are *self-dual.* The general self-dual solution of Einstein's non-linear vacuum (Ricci flat) equations can, indeed, be completely coded [9] into the structure of a deformed version of a suitable portion of $\boldsymbol{T}$. Some recent ideas lead one to hope that it ought to be possible eventually to remove the self-dual restriction [11], but as yet this has not been satisfactorily achieved. The present situation with regard to the (non-linear) self-dual gauge fields (Yang–Mills fields) is broadly similar [12], [13], [24].

A remarkable feature of these constructions, when they *have* been successful, is that the information is not stored locally. The local structure of each of the deformed spaces is precisely the same as in the particular case when the field to be described is zero! Nevertheless an infinite-dimensional system of inequivalent deformed spaces can be constructed and these correspond precisely to the freedom that one expects (free functions of three variables) for the fields in space-time. In the weak-field limit, we are concerned with infinitesimal deformations of complex structure. These are described in terms of the first holomorphic *sheaf cohomology groups* of portions of the flat twistor space $\boldsymbol{T}$, so it appears that the twistor description of linear space-time fields should be described in terms of such cohomology groups. Indeed, for massless fields of any helicity this turns out to be the case **[14]**, **[10]**, **[15]**. The situation for massive fields is not so clear, but in any case products of twistor spaces are then needed, rather than the single space $\boldsymbol{T}$. Part of the twistor programme is concerned with exploiting the extra freedom that then arises in order to obtain a classification of the numerous observed varieties of elementary particles **[18]**, **[19]**, **[20]**.

One of the major aims of the twistor programme is to find an alternative version of quantum field theory which might, one hopes, avoid some of the difficulties of the conventional formulation. An aspect of this approach is that wave functions of massless fields, are described **[11]**, **[16]**, **[17]** as the elements of the sheaf cohomology groups $H^1(\boldsymbol{PT}^+, \mathcal{O}(n))$ where $\mathcal{O}(n)$ refers to the sheaf of twisted holomorphic functions on $\boldsymbol{PT}^+$, as defined by homogeneous functions on $\boldsymbol{T}$ of degree n. Here, $n=-2s\hbar^{-1}-2$, where s is the helicity of the field. There are many curious features of these descriptions which are only just beginning to be explored. For example it is known **[21]** that infinite-dimensional H^1's only arise here because the boundary $\mathcal{N}$ of $\boldsymbol{PT}^+$ has a Levi form with one non-positive eigenvalue. It is also known **[22]**, **[23]** that such infinite-dimensional H^1's then live *locally* on $\mathcal{N}$ (and, somewhat ironically, arise precisely because of the existence of *insoluble* differential equations of the Hans Lewy type). The space-time interpretation of these local H^1's is still somewhat obscure, but they appear to be concerned with the propagation of non-analytic behaviour along light rays. There is some hope also, that significantly different insights into the somewhat paradoxical nature of quantum-mechanical wave functions might eventually arise from the cohomological formalism that appears forced upon us in twistor descriptions.

References

1. R. Penrose, *The apparent shape of a relativistically moving sphere*, Proc. Cambridge Philos. Soc. **55** (1959), 137.

2. S. S. Chern and J. Moser, *Real hypersurfaces in complex manifolds*, Acta Math., **133** (1974), 219.

3. R. Penrose, *Twistor algebra*, J. Math. Phys. **8** (1967), 345.

4. W. V. D. Hodge and D. Pedoe, *Methods of algebraic geometry*. Vol. II. Cambridge Univ. Press, London, 1951.

5. R. Penrose and M. A. H. MacCallum, Twistor theory, an approach to the quantization of fields and space-time, Phys. Rep. **6C** (1973), 241.

6. R. Penrose, *Twistor theory: its aims and achievements,* Quantum Gravity an Oxford Symposium, C. J. Isham, R. Penrose and D. W. Sciama, eds., Oxford Univ. Press, London, 1975.

7. *The twistor programme,* Rep. Mathematical Phys. **12** (1977), 65.

8. J. C. Taylor, *Gauge theories of weak interactions,* Cambridge Univ. Press, London, 1975.

9. R. Penrose, *Non-linear gravitons and curved twistor space,* General Relativity and Grav. **7** (1976), 31.

10. *Twistor quantization and curved space-time,* Internat. J. Theor. Phys. **1** (1968), 61.

11. *On the twistor description of massless fields,* Complex Manifold Techniques in Theoretical, Physics, D. Lerner and P. Sommers, eds. Pitman Publ. Co., London, 1979.

12. R. S. Ward, *On self-dual gauge fields* Phys. Lett. **61A** (1977), 81.

13. M. F. Atiyah and R. S. Ward, *Instantons and algebraic geometry,* Comm. Math. Phys. (1977)

14. R. Penrose, *Solutions of the zero rest-mass equations,* J. Math. Phys. **10** (1969), 61.

15. Twistor Newsletter **2** (1976); **5** (1977).

16. R. O. Wells, Jr., *Complex manifolds and mathematical physics,* Bull. Amer. Math. Soc. **1** (1979), 296—336.

17. S. Gindikin and G. Henkin, *Transformation de Radon pour la d''-cohomologie des domaines q-linéairement concaves,* C. R. Acad. Sci. Paris Ser. A. **303** (1978).

18. R. Penrose, *Twistors and particles,* Quantum Theory and the Structures of Time and Space, L. Castell, M. Drieschner, C. F. von Weiszacker, eds., Hanser, Munich, 1975.

19. Z. Perjés, *Twistors and particles,* Phys. Rev. D **11** (1975), 2031.

20. *Perspectives of the Penrose theory of particles,* Reports on Math. Phys. **12** (1977).

21. A. Andreotti and H. Grauert, *Théorèmes de finitude pour la cohomologie des espaces complexes,* Bull. Soc. Math. France **90** (1962), 193.

22. A. Andreotti and C. D. Hill, *E. E. Levi convexity and the Hans Lewy problem,* Ann. Scuola. Norm. Sup. Pisa CI. Sci. **26** (1972), 325; 747.

23. C. D. Hill, *Partial differential equations,* Proc. Sympos. Pure Math., vol. 23, Amer. Math. Soc., Providence, R. I., 1971, p. 135.

24. M. F. Atiyah, N. J. Hitchin, V. G. Drinfeld and Yu. I. Manin, *Construction of instantons,* Phys. Lett. **65A** (1978), 185.

Mathematical Institute
Oxford, England

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Representations of Semisimple Lie Groups

Wilfried Schmid

Fifty years ago, at the International Congress in Bologna, Hermann Weyl gave a report on representations of compact groups and, in particular, of compact Lie groups. Most of the important results had just been proved by him and by others, and at the time of his lecture, in 1928, the representation theory of compact Lie groups had become a very appealing subject. To a large extent, Weyl's theory has served as model and inspiration for the work on representations of noncompact, noncommutative groups, which was carried out in the last thirty years. To put the subject of my survey into perspective, I shall begin with a discussion of compact groups.

Initially, G will denote a compact topological group, and $\hat{G}$ its set of isomorphism classes of irreducible unitary representations. To avoid complicated notation, I shall not distinguish between an isomorphism class and its members: each $\pi \in \hat{G}$ s to be thought of as a specific continuous[1] homomorphism

$$\pi\colon G \to U(H_\pi)$$

into the unitary group $U(H_\pi)$ of a specific Hilbert space H_π. The irreducibility of π, i.e. the nonexistence of a proper, closed G-invariant subspace, implies that H_π is finite dimensional. According to the *Peter–Weyl theorem* [23], [30], there exists an isomorphism of Hilbert spaces

$$L^2(G) \cong \bigoplus_{\pi \in \hat{G}} H_\pi \otimes H_\pi^* \quad \text{(Hilbert space direct sum)}, \tag{1}$$

[1] Continuous with respect to the "weak topology" on $U(H_\pi)$ — the weakest topology which makes the functions $T \mapsto (Tu, v)$ continuous, for all u, v in H_π.

which can be described explicitly, and which has the following crucial property: the action of G on $L^2(G)$ induced by left translation corresponds to the action on the left factors H_π, whereas the right translation action corresponds to the dual action on the dual spaces H_π^*.

The statement of the Peter–Weyl theorem already points to the most fundamental reason for studying representation theory: to understand the representations of G is to understand $L^2(G)$, as a left and right G-module. The theorem, an early success of "soft analysis", makes only a rather abstract assertion, however; it does not

(2)
(a) describe the set $\hat{G}$,
(b) give information about the structure of the irreducible representations.

These two problems must be dealt with if the Peter–Weyl theorem is to answer concrete questions about $L^2(G)$.

To get a grasp on the first problem, one associates to each finite dimensional representation π the function

$$\chi_\pi \colon g \mapsto \operatorname{trace} \pi(g), \tag{3}$$

the so-called character of π. As a formal consequence of the Peter–Weyl theorem, χ_π determines π up to isomorphism. In particular, the passage from representations to their characters establishes a one-to-one correspondence

$$\hat{G} \leftrightarrow \text{set of irreducible characters.} \tag{4}$$

It is usually easier to describe $\hat{G}$ indirectly, via this correspondence: characters, as functions on G, are less complicated objects than representations.

For only one large and significant class of compact, noncommutative groups does one understand the two problems (2) reasonably well—namely compact, connected Lie groups. I shall now assume specifically that G belongs to this class. By a simple, but ingenious argument, which combines the Peter–Weyl theorem with basic properties of compact Lie groups, Herman Weyl was able to compute the irreducible characters of any such G. Implicitly the resulting Weyl character formula [27], [29] provides a parametrization of the set $\hat{G}$, and hence a solution of the problem (2a).

As for the second problem, the most useful technique is to study representations of G by analyzing their restriction to a maximal torus $T \subset G$. Any two maximal tori are conjugate, and hence the particular choice of T does not matter. Viewed as representation of T, each $\pi \in \hat{G}$ decomposes into a direct sum of one dimensional representations, which are called the weights of π. Among the weights, one is distinguished by being the "highest", in a certain definite sense; the highest weight occurs with multiplicity one and, most importantly, it characterizes π up to isomorphism. This is the essence of Élie Cartan's *theorem of the highest weight* [26]. Virtually all general structural information about representations of compact Lie groups follows from it, at least indirectly. The theorem can be proved by infinitesimal methods, or alternatively, deduced from Weyl's character formula.

In a nutshell, the Peter–Weyl theorem, the Weyl character formula, and the theorem of the highest weight constitute the fundamentals of the representation theory of compact Lie groups. Taken together, they give a good grasp of $L^2(G)$, and hence also of $L^2(X)$, for any homogeneous space X on which G operates transitively. Indeed, every such homogeneous space X can be represented as a quotient $X=G/U$, with $U=$isotropy subgroup at some point of X. Pulling back functions from G/U to G, one finds

(5) $$L^2(X) = \text{space of right } G\text{-invariants in } L^2(G) \cong \bigoplus_{\pi\in\hat{G}} H_\pi \otimes (H_\pi^*)^U;$$

here $(H_\pi^*)^U$ denotes the subspace of all U-invariant vectors in H_π^*. The description (5) of $L^2(X)$ makes it a simple matter to determine the G-invariant subspaces of $L^2(X)$: they are of the form

(6) $$\bigoplus_{\pi\in\hat{G}} H_\pi \otimes W_\pi,$$

with suitably chosen subspaces $W_\pi \subset (H_\pi^*)^U$; conversely, every direct sum (6) is actually G-invariant.

Let me now consider a linear differential operator D on X, G-invariant and, for simplicity, acting on scalar functions. One may extend D to an unbounded operator on $L^2(X)$, by taking its closure. The kernel of D then becomes an invariant subspace of $L^2(X)$:

(7) $$\operatorname{Ker} D \cong \bigoplus_{\pi\in\hat{G}} H_\pi \otimes W_\pi.$$

In our particular context, the W_π can be identified as the kernels of a family of linear transformations,

(8) $$\begin{gathered} W_\pi = \operatorname{Ker} D_\pi, \\ D_\pi\colon (H_\pi^*)^U \to (H_\pi^*)^U, \end{gathered}$$

which are derived from D in quite an explicit manner. Although D was assumed to be a scalar operator, these remarks apply—mutatis mutandis—also to invariant systems of differential equations.

The preceding discussion, straightforward and formal as it is, should convey one salient point: the Peter–Weyl theorem and its companion statements make it possible, at least in principle, to solve invariant systems of differential equations on homogeneous spaces. Undoubtedly, this connection with the problem of solving invariant differential equations is one of the most important aspects of the representation theory of Lie groups.

To give a concrete example, I shall mention the *Borel–Weil–Bott theorem* [4], [21]. As before, $T\subset G$ denotes a maximal torus. One knows that the quotient G/T can be made into a homogeneous complex manifold—a complex manifold such that G acts, by left translation, as a group of holomorphic mappings. Moreover, each one dimensional representation

$$\sigma\colon T \to C^*$$

gives rise to a homogeneous holomorphic line bundle

$$\mathscr{L}_\sigma \to G/T, \tag{9}$$

i.e. a holomorphic line bundle to which the translation action of G lifts; it is uniquely determined by the requirement that T should operate on the fibre at the identity coset via σ. Since G acts on the bundle (9), it also acts on the cohomology groups

$$H^k(G/T, \mathcal{O}(\mathscr{L}_\sigma)) \tag{10}$$

of the sheaf of germs of holomorphic sections $\mathcal{O}(\mathscr{L}_\sigma)$. The sheaf cohomology groups thus become finite dimensional representation spaces for G—finite dimensional because G/T is compact. The Hodge theorem identifies the cohomology group (10) with the kernel of the (G-invariant) Laplace–Beltrami operator, acting on the $\mathscr{L}_\sigma$-valued $(0, k)$-forms. In particular, the present example fits into the framework of invariant differential equations.

The Borel–Weil–Bott theorem describes the cohomology groups (10): they vanish identically for certain special choices of σ; in all remaining cases, they are nonzero for exactly one integer $k=k(\sigma)$, and the representation of G on this one non-zero cohomology group is irreducible, with a highest weight whose dependence on σ can be made explicit. Every irreducible representation of G arises in this fashion, even with $k(\sigma)=0$; for $k=0$, it should be noted, the group (10) is simply the space of holomorphic sections of the line bundle $\mathscr{L}_\sigma$. Most proofs reduce the Borel–Weil–Bott theorem to the theorem of the highest weight, through arguments in the spirit of (7)–(8). The theorem serves at least two purposes. It provides a realization of every $\pi\in\hat{G}$, on a concrete vector space, with a concrete G-action—in contrast to the Weyl character formula or the theorem of the highest weight, which enumerate the irreducible representations, without giving such a realization. Secondly, it computes certain cohomology groups which are of interest in complex analysis, and which were not understood before the advent of the Borel–Weil–Bott theorem.

The statement of the theorem also suggests another possible approach to the representation theory of compact Lie groups. One can use methods of differential geometry and complex analysis to prove the theorem directly, avoiding any reference to the Weyl character formula and the theorem of the highest weight. An application of the Atiyah–Bott fixed point formula then leads to the character formula, which thus becomes a consequence of the Borel–Weil–Bott theorem. This chain of arguments employs rather heavy machinery and may seem merely a curiosity. I mention it here because in the case of non-compact groups, analogous arguments turn out to be quite efficient.

So much for compact groups! The object of interest shall now be a locally compact group G, unimodular—i.e. the essentially unique left invariant measure is also right invariant— and of type I. The latter is a technical condition, satisfied by all the special classes of groups which are considered in this survey; it insures that G has a "reasonable" representation theory. Again $\hat{G}$ stands for the set of isomorphism classes of irreducible unitary representations. In general, these will be infinite dimensional, since G may not be compact. For the same reason, the Peter–Weyl theorem no longer applies; even the simplest noncompact examples show that $L^2(G)$ cannot be expressed as a direct sum of irreducibles. Its place is taken by the ***abstract Plancherel theorem*** [6], which essentially goes back to von Neumann: $L^2(G)$ decomposes into a Hilbert space direct integral,

$$L^2(G) \cong \int_{\pi \in \hat{G}} H_\pi \otimes H_\pi^* \, d\mu(\pi), \tag{11}$$

with respect to a measure μ on $\hat{G}$, the so-called Plancherel measure.

Just as in the case of the Peter–Weyl theorem, the isomorphism makes the left and right actions of G correspond to the actions on the left and right factors of the integrand $H_\pi \otimes H_\pi^*$. The tensor product sign refers not to the algebraic tensor product, but rather to its completion. The notion of Hilbert space direct integral generalizes the notion of Hilbert space direct sum. For instance if G is compact after all, the measure μ becomes discrete, and the direct integral (11) reduces to the direct sum[2] (1). The best known example of a direct integral, which is not actually a direct sum, is furnished by classical Fourier analysis on the real line: $L^2(\boldsymbol{R})$ may be viewed as a direct integral of a continuous family of one dimensional function spaces, namely those spanned by the unitary characters

$$x \mapsto e^{ixy}, \quad y \in \boldsymbol{R}.$$

These function spaces do not occur in $L^2(\boldsymbol{R})$ discretely, as subspaces, but only "infinitesimally".

Again, the description (11) of $L^2(G)$ raises some immediate questions:

(12)
- (a) what is the set $\hat{G}$?
- (b) what is the Plancherel measure μ?
- (c) what can one say about the structure of the irreducible unitary representations?

Reasonably complete answers exist for only two major classes of noncompact, noncommutative groups—on the one hand, nilpotent Lie groups, and to some extent also solvable groups; on the other, semisimple Lie groups. The techniques which are appropriate in these two cases diverge widely, for quite fundamental reasons. I shall therefore limit the discussion to the semisimple case. The classical matrix groups $\mathrm{Sl}(n, \boldsymbol{R})$, $\mathrm{Sl}(n, \boldsymbol{C})$, $\mathrm{SO}(p, q)$, $\mathrm{SO}(n, \boldsymbol{C})$, $\mathrm{SU}(p, q)$, $\mathrm{Sp}(n, \boldsymbol{R})$, ..., which are of special interest in geometry, number theory, and physics, all fall into

[2] The measure μ, which does not show up in (1), has been absorbed into the particular isomorphism. In fact, the isomorphism (11), and with it the measure μ, are not uniquely determined. There is one natural choice, however.

this class (as do compact Lie groups with finite center)—ample justification for studying semisimple Lie groups in particular detail.

Let then G be a connected semisimple Lie group, and π an irreducible unitary representation of G, on a Hilbert space H_π. Typically H_π is infinite dimensional. The definition of character, which proved so useful in the finite dimensional case, thus loses meaning, at least in its naive form (3): as unitary operators acting on an infinite dimensional space, the operators $\pi(g)$ do not have a trace in any obvious sense.

There exists a way around this difficulty, first discovered by Gelfand and Naimark in their study of the complex classical groups, later fully developed and systematically exploited by Harish-Chandra [10]. It proceeds from the following observation: for every compactly supported C^∞ function f on G, the operator-valued integral

$$\pi(f) = \int_G f(g)\pi(g)\,dg \tag{13}$$

is of trace class. In other words, if one represents $\pi(f)$ by an infinite matrix, relative to any orthonormal basis of H_π, the sum of the diagonal matrix entries converges absolutely. It then follows that the sum does not depend on the particular choice of basis, and one calls this sum the trace. The linear mapping

$$\Theta_\pi : f \mapsto \operatorname{trace} \pi(f), \tag{14}$$

which assigns to every $f \in C^\infty(G)_0$ the trace of the operator $\pi(f)$, turns out to be a distribution in the sense of L. Schwartz. It determines π up to isomorphism and is, by definition, the character of π. If π happens to be a finite dimensional representation, Θ_π is given by integration against the ordinary character. Thus Harish-Chandra's definition of character embraces the usual one.

Because of its definition in terms of a trace, Θ_π remains invariant under all inner automorphisms of G. A slightly more subtle argument, based on the irreducibility of π, shows that every bi-invariant linear differential operator maps Θ_π to a multiple of itself. In shorthand terminology, a distribution with these two properties is an *invariant eigendistribution.*

Distributions are decidedly more complicated objects than functions—more complicated to write down, more complicated to manipulate. At first glance, this appears to be a serious shortcoming of the notion of character in the infinite dimensional case. Fortunately, there is a remedy, Harish-Chandra's *regularity theorem* for invariant eigendistributions [13]-[15], [2]: every invariant eigendistribution, and in particular every character, can be expressed as integration against a locally L^1 function; this function is real-analytic on the complement of a real-analytic subvariety of G. Thus characters turn out to be functions, after all. The regularity theorem plays a crucial role in the representation theory of semisimple Lie groups; without it, the notion of character would be far less useful.

Let me now turn to the problem of describing $\hat{G}$. As a first step, it is helpful to consider a certain subset. An irreducible unitary representation π is said to be *square-integrable* if the Plancherel measure assigns a positive mass to the single point $\pi \in \hat{G}$, i.e. if π contributes discretely to the Plancherel decomposition (11) of $L^2(G)$. The isomorphism classes of all such irreducible, square-integrable representations constitute a subset $\hat{G}_{ds} \subset \hat{G}$, the *discrete series* of G.

To state Harish-Chandra's fundamental results on the discrete series **[1]**, **[16]**, I select a maximal compact subgroup $K \subset G$. Any two of them are conjugate, and this fact makes the particular choice of K unimportant. According to Harish-Chandra's existence criterion, the discrete series of G is nonempty if, and only if, K has the same rank as G—equivalently, if any maximal torus $T \subset K$ is its own centralizer in G, or in more technical language, if G contains a compact Cartan subgroup. Going back to the list of examples, one finds that $\mathrm{Sl}\,(n, \boldsymbol{R})$ has a discrete series only for $n=2$, $\mathrm{Sl}\,(n, \boldsymbol{C})$ never does, $\mathrm{SO}\,(p, q)$ has one precisely when pq is even, and finally $\mathrm{SU}\,(p, q)$ and $\mathrm{Sp}\,(n, \boldsymbol{R})$ always have a discrete series.

In case G satisfies the criterion, and subject to a minor restriction which will be mentioned presently, Harish-Chandra's parametrization of the discrete series establishes a bijection

$$\hat{K}' \leftrightarrow \hat{G}_{ds} \tag{15}$$

between a subset $\hat{K}'$ of $\hat{K}$ and the discrete series $\hat{G}_{ds}$. It assigns to the character χ of a representation in $\hat{K}'$ a discrete series character Θ, whose restriction to K is given by the formula

$$\Theta_K = \pm \frac{\chi}{D}; \tag{16}$$

here D denotes a universal denominator, independent of χ, and itself a linear combination of irreducible characters of K. The Weyl character formula for K identifies $\hat{K}$ with a lattice, divided by the action of a finite linear group. In terms of this description, $\hat{K}'$ corresponds to the complement, in the lattice, of a finite number of hyperplanes. As was remarked already, the parametrization (15) does not apply to an arbitrary semisimple G; however, it does apply to some finite covering of any given G. This restriction is quite innocuous, since the discrete series for G may be viewed as a subset of the discrete series for the covering group.

The discrete series provides a basic repertory of representations, from which others can be constructed. To be more concrete, I shall need the notion of Cartan subgroup. It is most easily defined for a linear semisimple group G: Cartan subgroups are then Abelian, they consist of group elements that can be diagonalized over $\boldsymbol{C}$, and are maximal subgroups with respest to these two properties. One can classify the conjugacy classes of Cartan subgroups **[20]**, **[28]**; in particular, they are finite in number.

To each conjugacy class, Harish-Chandra attaches a series of irreducible unitary representations **[17]**. If there exists a—necessarily unique—conjugacy class of compact Cartan subgroups, the corresponding series is the discrete series. The other series are obtained by an induced representation process, starting from discrete series representations of subgroups of G. In this construction, distinct conjugacy classes of Cartan subgroups lead to non-overlapping series of representations. Although the terminology is by no means standard, I shall call a representation *generic* if it belongs to one of the series, and otherwise *special.* Both types actually occur, unless G is compact, in which case $\hat{G}=\hat{G}_{ds}$.

The crowning achievement of Harish–Chandra's program is a solution of problem (12b). Perhaps the explicit, somewhat complicated description of the Plancherel measure μ [17] matters less than the nature of the answer. To begin with, μ has the set of generic representations as support. Each of the various series is parametrized, roughly speaking, by the product of a lattice with a vector space, divided by the action of a finite linear group. It therefore carries a distinguished measure, namely the one derived from the invariant measures on the two factors. The restriction of μ to the series in question is completely continuous with respect to this distinguished measure. The ratio of the two measures reflects the rate at which the matrix coefficients of the representations in the given series decay at infinity, a fact which is a crucial ingredient of the actual computation of the measure.

Since the Plancherel measure completely disregards the special representations, these become irrelevant as far as the decomposition of $L^2(G)$ is concerned. They are quite important from other points of view, and I shall come back to them later. In any case, $L^2(G)$ is made up of generic representations, which are described, in Harish–Chandra's construction, in terms of their characters. Thus Harish–Chandra's theory accomplishes for semisimple Lie groups what Weyl's theory did for compact Lie groups. I should point out, however, that the technical difficulties are immensely greater. In my very condensed summary, I have broken down Harish–Chandra's program into three major components: the study of characters, the construction of the discrete series, and the determination of the Plancherel measure. Each of these is a large and elaborate edifice.

A brief historical remark: the idea that various series of representations should be attached to the conjugacy classes of Cartan subgroups made its first appearance in the work of Gelfand and his collaborators on the complex classical groups and the real special linear group. In general, it was conjectured—and of course later worked out—by Harish–Chandra. His address at the 1954 Congress in Amsterdam already gives a glimpse, in very rough outline, of his entire program.

To understand the structure of representations of compact Lie groups, one investigates their restrictions to a maximal torus. In the context of semisimple groups, there is a similar device, namely to break up representations under the action of a maximal compact subgroup $K\subset G$. When restricted to K, every $\pi\in\hat{G}$ becomes a direct sum of irreducibles, each occurring with finite multiplicity; in symbolic notation,

(17) $$\pi|_K = \bigoplus_{\tau\in\hat{G}} n_\tau(\pi)\cdot\tau.$$

The analogy with the compact case, as well as examples of low dimensional noncompact groups [3], [5] suggest that the pattern of the multiplicities $n_\tau(\pi)$ is an important invariant of π.

Generic representations either belong to the discrete series, or are constructed, by a well understood procedure, from discrete series representations of subgroups of G. Structural questions about generic representations therefore come down to questions about the discrete series. A *conjecture of Blattner*, now a theorem **[18]**, describes the K-multiplicities $n_\tau(\pi)$, for every discrete series representation π. According to the conjecture, of the various $\tau \in \hat{K}$ which appear in the restriction $\pi|_K$, one is lowest, in an appropriate sense, and occurs with multiplicity one. The *theorem of the lowest K-type*, counterpart to the theorem of the highest weight, asserts that this feature characterizes the discrete series representation π uniquely, among all irreducible representations **[24]**. It can be quite difficult, if not impossible, to check directly whether a representation is square-integrable. The theorem of the lowest K-type provides a useful criterion, of an essentially algebraic nature.

The problem of realizing discrete series representations concretely is closely related to the theorem of the lowest K-type. If G has a discrete series, it contains a compact Cartan subgroup T, which is in particular a maximal torus. Just as in the compact case, the quotient G/T can be turned into a homogeneous complex manifold, noncompact of course, unless G itself is compact. Every character

$$\sigma\colon T \to C^*$$

again determines a homogeneous holomorphic line bundle

$$\mathscr{L}_\sigma \to G/T.$$

By translation, G acts unitarily on $\mathscr{H}(\mathscr{L}_\sigma)$, the Hilbert space of square-integrable, holomorphic sections. Long before the discrete series was fully understood, Harish–Chandra showed that for certain characters σ, $\mathscr{H}(\mathscr{L}_\sigma)$ is nonzero; the resulting representation is then necessarily irreducible and belongs to the discrete series **[12]**. Unfortunately this construction gives only a relatively small part of the discrete series, and for some groups G it even gives nothing at all. To produce a realization of every discrete series representation, one must turn elsewhere.

The first explicit suggestion was made by Langlands: one should consider also the higher[3] L^2-cohomology groups $\mathscr{H}^k(\mathscr{L}_\sigma)$, i.e. the spaces of harmonic, square-integrable, $\mathscr{L}_\sigma$-valued $(0, k)$-forms on G/T. If G/T happens to be compact, the Hodge theorem identifies $\mathscr{H}^k(\mathscr{L}_\sigma)$ with the kth sheaf cohomology group of $\mathscr{L}_\sigma$. Such a simple connection between L^2-cohomology and sheaf cohomology does not exist in general, for noncompact G, but in any case $\mathscr{H}^k(\mathscr{L}_\sigma)$ is a Hilbert space, on which G operates as a group of unitary transformation. Guided by the analogy with the Borel–Weil–Bott theorem, and also by curvature computations of Griffiths, Langlands conjectured that all of the L^2-cohomology groups should vanish for

[3] For $k=0$, $\mathscr{H}^0(\mathscr{L}_\sigma)$ coincides with $\mathscr{H}(\mathscr{L}_\sigma)$.

special choices of σ; in the remaining cases, $\mathscr{H}^k(\mathscr{L}_\sigma)$ was to be nonzero for exactly one integer $k=k(\sigma)$, and G was to act according to a representation of the discrete series. In this manner, one would be able to realize every discrete series representation, but usually not with $k(\sigma)=0$. The conjecture has in fact been proved, by an argument which reduces it to the theorem of the lowest K-type [25].

In order to verify the conjecture, one must somehow exhibit squareintegrable harmonic forms. The proof in [25] relies on Harish–Chandra's construction of the discrete series to overcome this analytic problem. Instead, one can also base a proof on Atiyah's L^2-index theorem, in combination with the Atiyah–Singer index theorem. An argument, showing that Langlands' conjecture accounts for all of the discrete series, goes hand in hand with the existence proof. The outcome[4] is an alternative, more geometrically oriented approach to the main results on the discrete series [1].

I shall conclude with some remarks about special representations. This will also give me an opportunity to touch on some important points which have been passed over so far. To describe what is known, one must look beyond the class of unitary representations. Indeed, techniques which are designed to deal with unitary representations very often lead naturally into questions about nonunitary representations.

A representation π of G on a Banach space is said to be admissible if every $\tau\in\hat{K}$ occurs at most finitely often in the restriction of π to K. Irreducible unitary representations automatically have this property, it remains unknown whether all irreducible Banach representations are admissible. To simplify the terminology, when I speak of an irreducible representation, I shall always mean an irreducible admissible representation on a Banach space.

In the finite dimensional case, there exists a well understood, very useful relationship between representations of the group and those of the Lie algebra. Such a relationship exists also for infinite dimensional representations, but this involves some subtleties. Let π be an irreducible representation, on a Banach space B. The analytic vectors—those vectors $v\in B$ for which

$$g\mapsto\pi(g)v$$

is a real analytic mapping from G to B—form a dense subspace $B_\omega\subset B$. The Lie algebra $\mathfrak{g}$ of G acts on B_ω by differentiation, but the resulting representation of $\mathfrak{g}$ is "too large"; in particular, it fails to be irreducible. The most natural way to pick out an irreducible subrepresentation is to pass to the space of K-finite vectors B_0, i.e. to the linear span of all finite dimensional K-stable subspaces of B. It is contained in B_ω, dense in B, $\mathfrak{g}$-invariant, and algebraically irreducible as $\mathfrak{g}$-module. In this manner, one attaches to each irreducible representation of the group an irreducible representation of the Lie algebra [9].

A side remark is perhaps appropriate at this point. Hermann Weyl repeatedly

[4] For minor technical reasons, the construction in [1] works with harmonic spinors on G/K, rather than L^2-cohomology.

emphasized the fact that his methods were purely global; he studied representations of the group directly, without reference to the Lie algebra. My very brief account of Harish-Chandra's work may have fostered the impression that his approach is equally global. Quite to the contrary, infinitesimal arguments play an important role in many of his proofs.

The definition of character (14) seems to depend on the Hilbert space structure of the representation space. However, with relatively minor modifications, it makes sense also for representations on Banach spaces: every irreducible representation π has a character Θ_π, which is again an invariant eigendistribution. The characters of two irreducible representations π_1, π_2 coincide precisely when the corresponding Lie algebra representations are algebraically isomorphic **[10]**. In this case, π_1 and π_2 are said to be infinitesimally equivalent; loosely speaking, they look alike, except possibly for the topology on the representation spaces.

In the case of unitary representations, infinitesimal equivalence implies that the representations in question are actually isomorphic **[9]**. An irreducible representation on a Banach space can thus be infinitesimally equivalent to at most one unitary representation. If it is, one calls the representation unitarizable. To classify the irreducible unitary representations—or what amounts to the same, the special representations, since one already knows the others—, it suffices to

(18)
(a) classify the irreducible representations on Banach spaces, up to infinitesimal equivalence, and
(b) determine which are unitarizable.

The first of these two problems has been solved by Langlands **[22]**, and in somewhat different form also by Vogan **[31]**. Langlands classifies representations in terms of the growth behavior of their matrix coefficients. This has the effect of reducing the problem to the classification of a smaller class of representations, which was worked out by Knapp–Zuckerman **[19]**. Vogan's classification involves a very detailed study of how representations break up under the action of K. Each irreducible representation has a lowest K-type, which occurs with multiplicity one. The meaning of "lowest", however, is weaker than in the context of the discrete series, and the lowest K-type alone does not suffice to characterize an arbitrary irreducible representation uniquely.

In principle, (18b) is a purely algebraic problem, but apparently a difficult one, and no solution is yet in sight. If a representation is to be unitarizable, it must satisfy certain fairly obvious necessary conditions. This gives at least a partial answer to (18b), and hence some information about special representations—virtually the only information that is known at present.

Why is one interested in the special representations? The problem of decomposing function spaces on quotients of G reveals perhaps the best explanation. Every quotient G/U, by a closed unimodular subgroup $U \subset G$, carries an invariant measure, and so G acts unitarily on $L^2(G/U)$. Whenever U is compact, one can embed $L^2(G/U)$ into $L^2(G)$,

$$L^2(G/U) \hookrightarrow L^2(G), \tag{19}$$

by pulling back functions from G/U to G. In this way, the Plancherel decomposition of $L^2(G)$ leads to a decomposition of $L^2(G/U)$. If U is non-compact, on the other hand, no inclusion (19) exists, and it becomes an entirely separate problem to express $L^2(G/U)$ as a direct integral of irreducible representations. In particular, special representations can and do contribute to the decomposition of $L^2(G/U)$.

The most important case is that of a discrete subgroup $\Gamma \subset G$, such that G/Γ has finite volume; an arithmetic subgroup, for example. One can then use the Selberg trace formula to analyze $L^2(G/\Gamma)$. Although much work is being done in this direction, to understand $L^2(G/\Gamma)$ remains a distant goal.

References

1. M. F. Atiyah and W. Schmid, *A geometric construction of the discrete series for semisimple Lie groups*, Invent. Math. **42** (1977), 1—62.

2. ——— *A new proof of the regularity theorem for invariant eigendistributions on semisimple Lie groups* (to appear)

3. V. Bargman, *Irreducible unitary representations of the Lorentz group*, Ann. of Math. **48** (1947), 568—640.

4. R. Bott, *Homogeneous vector bundles*, Ann. of Math. **66** (1957), 203—248.

5. J. Dixmier, *Représentations intégrables du groupe de De Sitter*, Bull. Soc. Math. France **89** (1961), 9—41.

6. ——— *Les C*-algèbres et leur représentations*, Gauthier-Villars, Paris, 1964.

7. I. M. Gelfand and M. A. Naimark, *Unitäre Darstellungen der klassischen Gruppen*, Akademie Verlag, Berlin, 1957.

8. R. Godement, *A theory of spherical functions* I, Trans. Amer. Math. Soc. **73** (1952), 496—556.

9. Harish-Chandra, *Representations of semisimple Lie groups* I, Trans. Amer. Math. Soc. **75** (1953), 185—243.

10. ——— *Representations of semisimple Lie groups* III, Trans. Amer. Math. Soc. **76** (1954), 234—253.

11. ——— *Representations of semisimple Lie groups*, Proc. Internat. Congr. Mathematicians, Amsterdam, 1954, vol. I, North-Holland, Amsterdam, 1957, 299—304.

12. ——— *Representations of semisimple Lie groups*. IV, V, VI, Amer. J. Math. **77** (1955), 743—777; **78** (1956), 1—41; 564—628.

13. ——— *Invariant distributions on Lie algebras*, Amer. J. Math. **86** (1964), 271—309.

14. ——— *Invariant eigendistributions on a semisimple Lie algebra*, Inst. Hautes Études Sci. Publ. Math. **27** (1965), 5—54.

15. ——— *Invariant eigendistributions on a semisimple Lie group*, Trans. Amer. Math. Soc. **119** (1965), 457—508.

16. ——— *Discrete series for semisimple Lie groups* I, II, Acta Math. **113** (1965), 241—318; **116** (1966), 1—111.

17. ——— *Harmonic analysis on real reductive groups* III, Ann. of Math. **104** (1976), 117—201.

18. H. Hecht and W. Schmid, *A proof of Blattner's conjecture*, Invent. Math. **31** (1975), 129—154.

19. A. W. Knapp and G. Zuckerman, *Classification of irreducible tempered representations of semisimple Lie groups*, Proc. Nat. Acad. Sci. U.S.A. **73** (1976), 2178—2180.

20. B. Kostant, *On the conjugacy of real Cartan subalgebras* I, Proc. Nat. Acad. Sci. U.S.A. **41** (1955), 967—970.

21. ——— *Lie algebra cohomology and the generalized Borel-Weil theorem*, Ann. of Math. **74** (1961), 329—387.

22. R. P. Langlands, *On the classification of irreducible representations of real algebraic groups*, Mimeographed notes, Institute for Advanced Study, Princeton, N. J., 1973.

23. F. Peter and H. Weyl, *Die Vollständigkeit der primitiven Darstellungen einer geschlossenen kontinuierlichen Gruppe*, Math. Ann. **97** (1927), 737—755.

24. W. Schmid, *Some properties of square-integrable representations of semisimple Lie groups*, Ann. of Math. **102** (1975), 535—564.

25. ——— *L^2-cohomology and the discrete series*, Ann. of Math. **103** (1976), 375—394.

26. J.-P. Serre, *Algèbres de Lie semi-simples complexes*, Benjamin, New York, 1966.

27. E. Stiefel, *Kristallographische Bestimmung der Charaktere der geschlossenen Lie'schen Gruppen*, Comment. Math. Helv. **17** (1944), 165—200.

28. M. Sugiura, *Conjugate classes of Cartan subalgebras in real semisimple Lie algebras*, J. Math. Soc. Japan **11** (1959), 374—434.

29. H. Weyl, *Theorie der Darstellung kontinuierlicher halb-einfacher Gruppen durch lineare Transformationen.* I, Math. Z. **23** (1925), 271—309.

30. ——— *Kontinuierliche Gruppen und ihre Darstellungen durch lineare Transformationen*, Proc. Internat. Congr. Mathematicians, Bologna, 1928, vol. I, Nicola Zanichelli, Bologna, 1929, 233—246.

31. D. Vogan, *The algebraic structure of representations of semisimple Lie groups.* I, Ann. of Math. **109** (1979), 1—60.

Harvard University
Cambridge, Massachusetts 02138, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Absolute Continuity and Singularity of Probability Measures in Functional Spaces

A. N. Shiryaev

1. Introduction. The exceptional role played in the probability theory by two random processes—the Wiener and the Poisson ones—naturally poses the following question: What kind of pattern characterises the processes (all of them or a certain class) whose measure is absolutely continuous with respect to the Wiener, and, respectively, the Poisson measure? In more exact terms, let (C, B) be the measurable space of continuous functions $(X_t)_{t\geqq 0}$, $X_0=0$, P—the Wiener measure, and $\tilde{P}$—some other measure on (C, B) which is absolutely continuous with respect to P $(\tilde{P}\ll P)$. The question is now: What is the structure of the random process $\tilde{X}=(X_t, \tilde{P})_{t\geqq 0}$? What is the structure of the Radon–Nikodym derivative $d\tilde{P}/dP=$ density of $\tilde{P}$ with respect to P? If (D, B) is the measurable space of right-continuous piecewise constant functions $(X_t)_{t\geqq 0}$, $X_0=0$, with $\Delta X_t=0$ or $+1$ $(\Delta X_t=X_t-X_{t-})$, P—the Poisson measure, and $\tilde{P}$—another measure, such that $\tilde{P}\ll P$, find the structure of the random process $\tilde{X}=(X_t, \tilde{P})_{t\geqq 0}$ and the derivative $d\tilde{P}/dP$.

Obviously, similar questions arise also for other (not the Wiener or the Poisson) random processes (of a simple structure).

An answer to these and related questions can be obtained within the framework of the general theory of absolute continuity and singularity of the probability measures (ACS), which has made much progress over the past decade. It makes the subject of this paper. It must be stressed that advances in this field are primarily associated with the development of the general theory of random processes, particularly its chapter concerned with the notion of the martingale and its various extensions. "Local martingales, semimartingales, predictability, stochastic integrals with respect to semimartingales and random measures..." that is just a fragment of the list of notions that researchers have been much excited and preoccupied with over the past ten years or so.

2. ACS criteria. The general problem of the absolute continuity and singularity of probability measures is stated as follows. Let $(\Omega, F, F_t)_{t\geqq 0}$ be a measurable space on which is defined a nondecreasing and right-continuous family of σ-algebras F_t such that $F=\bigvee F_t$. Let P and $\tilde{P}$ be two probability measures on (Ω, F) and

$$P_t = P|F_t, \quad \tilde{P}_t = \tilde{P}|F_t$$

be their restriction on F_t.

We shall say that a measure $\tilde{P}$ is locally absolutely continuous with respect to the measure $\tilde{P}$ ($\tilde{P} \ll^{\mathrm{loc}} P$), if $\tilde{P}_t \ll P_t$ for every $t \geqslant 0$. The question is: When does it follow from the local absolute continuity that $\tilde{P} \ll P$ (absolute continuity) or that $\tilde{P} \perp P$ (singularity)?

Let $Z_t = d\tilde{P}_t/dP_t$ be the density of $\tilde{P}_t$ with respect to P_t and $Z_\infty = \overline{\lim}_{t\to\infty} Z_t$. The process $Z=(Z_t, F_t, P)$ is a nonnegative martingale and, therefore, there exists (P-a.s.) $\lim_{t\to\infty} Z_t$ $(=Z_\infty)$. One can show that this limit exists also $\tilde{P}$-a.s.

The following proposition is well known.

Proposition. *Let* $\tilde{P} \ll^{\mathrm{loc}} P$. *Then*

$$\begin{aligned} &\tilde{P} \ll P \Leftrightarrow EZ_\infty = 1, \\ &\tilde{P} \perp P \Leftrightarrow EZ_\infty = 0. \end{aligned} \tag{1}$$

The condition $EZ_\infty = 1$ is equivalent to the assumption that the family $(Z_t)_{t\geqq 0}$ is uniformly integrable with respect to P and it is the condition which is normally used for searching of "convenient" criteria of absolute continuity in a particular situation. The new approach to ACS problems, developed with regard to the general case in [1]–[4], is associated with the idea of expressing the conditions for $\tilde{P} \ll P$ and $\tilde{P} \perp P$ in terms of the measure $\tilde{P}$ (and not P as in the proposition). That simple idea, which is realised in Theorem 1 below, has made it possible to obtain in the form of corollaries many of known results and, moreover, to move further substantially and obtain the effective ACS conditions for a broad class of random processes (§§ 4–7).

Theorem 1. *Let* $\tilde{P} \ll^{\mathrm{loc}} P$. *Then*

$$\begin{aligned} &\tilde{P} \ll P \Leftrightarrow \tilde{P}(Z_\infty < \infty) = 1, \\ &\tilde{P} \perp P \Leftrightarrow \tilde{P}(Z_\infty = \infty) = 1. \end{aligned} \tag{2}$$

The advantage these criteria have over (1) lies with their enabling ACS problem to be reduced to investigating the asymptotic properties of the sequence $(Z_t)_{t\geqq 0}$ at $t\to\infty$ with respect to the measure $\tilde{P}$. By way of illustration, let us show how the well-known absolute continuity/singularity alternative of Kakutani [5] for sequences of independent random variables follows from (2).

Let $P=\mu_1\times\mu_2\times\ldots$, $\tilde P=\tilde\mu_1\times\tilde\mu_2\times\ldots$ be two measures (which are direct products of measures) in the space of real sequences $x=(x_1, x_2, \ldots)$ and they are such that $\tilde\mu_n\ll\mu_n$, $n=1$. By Kakutani's result, then the alternative holds: "either $\tilde P\ll P$, or $\tilde P\perp P$". This follows directly from (2), since in our case

$$Z_n=\prod_{i=1}^{n}\frac{d\tilde\mu_i}{d\mu_i}(x_i)$$

and the $\tilde P$-probability of the event $\{Z_\infty<\infty\}$, by the Kolmogorov's "zero-one" law, has only two values—0 or 1.

From Theorem 1, it is possible also to derive the well-known Hajek–Feldman's [6], [7] dichotomy which asserts that if P and $\tilde P$ are two Gaussian measures in the space of real sequences, then the alternative holds: "they are either equivalent $(\tilde P\sim P)$, or singular $(\tilde P\perp P)$".

The proof of Theorem 1 is simple. Basically, it is a consequence of the Lebesgue decomposition of $\tilde P$ with respect to P, which, assuming $\tilde P\overset{\text{loc}}{\ll}P$, appears as

$$\tilde P(A)=\int_A Z_\infty dP+\tilde P(A\cap\{Z_\infty=\infty\}),\ A\in F. \tag{3}$$

(Concerning the proofs of decomposition (3), see [1]–[4]; in [4], applications o Theorem 1 are also given to the question of the validity of the "absolute continuity-singularity" alternative for Markovian sequences.)

3. "Predictable" ACS criteria. Consider in more detail the "local" density $Z=(Z_t, F_t, P)$. First of all, having complemented, if necessary, the σ-algebra F_0 with sets from F of Q-measure $(Q=\frac{1}{2}(P+\tilde P))$ zero, let us choose a variant of that density with (P-a.s.) regular trajectories (continuous on the right and with left limit). Denote

$$\tau_n=\inf\{t\colon Z_t<1/n\},\ \tau=\lim_n\tau_n$$

and introduce the process

$$M_t=\int_0^t Z_{s-}^{\oplus}\,dZ_s \tag{4}$$

where $a^{\oplus}$ is a pseudo-inversion of a (i.e., $a^{\oplus}=a^{-1}$ when $a\neq0$ and $a^{\oplus}=0$ when $a=0$). The process $M=(M_t, F_t, P)$ is the τ-local martingale (i.e., $M^{\tau_n}=(M_{t\wedge\tau_n}, F_t, P)$ are martingales for any $n\geqslant1$) and, also, $\Delta M_t\geqslant-1$ and

$$Z_t=Z_0+\int_0^t Z_{s-}\,dM_s,\ Z_0=\frac{d\tilde P_0}{dP_0}. \tag{5}$$

This equation has a solution, which is a unique one, in the class of nonnegative local martingales; according to [8], [9], that solution can be written as

$$Z_t=Z_0\exp\left\{M_t-\frac{1}{2}\langle M^C\rangle_t+\sum_{s\leqq t}[\ln(1+\Delta M_s)-\Delta M_s]\right\} \tag{6}$$

where M^c is the continuous part of the τ-local martingale M in its decomposition

$$M = M^c + M^d \tag{7}$$

into the continuous and the pure discontinuous components and $\langle M^c \rangle$ is a characteristic of M^c (i.e., a predictable increasing process for which $(M^c)^2 - \langle M^c \rangle$ is the τ-local martingale). By virtue of (4) and (5), there is a one-one correspondence between the trajectories Z and M (for all $t < \tau$). Also, since $\tilde{P}(\inf Z_t > 0) = 1$ (cf. the Lebesgue decomposition (3)), $\tilde{P}(\tau = \infty) = 1$ and, hence, if we consider Z and M with respect to the measure $\tilde{P}$, there is one-to-one ($\tilde{P}$-a.s.) correspondence between them for every $t \geqslant 0$.

Let now μ be the measure of jumps of the process M:

$$\mu((0, t], \Gamma) = \sum_{s \leqq t} I(\Delta M_s \in \Gamma), \ \ \Gamma \in B(E), \ \ E = R \setminus \{0\}$$

and ν the compensator of that measure, [10], [11]. Then decomposition (7) can be written as

$$M_t = M_t^c + \int_0^t \int_E x \, d(\mu - \nu) \tag{8}$$

where $\int_0^t \int_E x \, d(\mu - \nu)$ is a stochastic integral with respect to the random "martingale" measure $\mu - \nu$ [12], [3]. From (8), it follows that the predictable [9] characteristics of M are $(\langle M^c \rangle, \nu)$, and it is with these notions that the ACS conditions can be formulated in a natural way.

As any τ-local martingale, the process M has the property that the square root of the increasing process

$$[M, M]_t \equiv \langle M^c \rangle_t + \sum_{s \leqq t} (\Delta M_s)^2$$

is τ-locally integrable with respect to the measure P ($[M, M]^{1/2} \in A^+_{\mathrm{loc}(\tau)}(P)$), i.e., there exists a sequence $(\sigma_n)_{n \geqq 1}$ of Markov times, $\sigma_n \uparrow \tau$ and such that

$$E[M, M]^{1/2}_{\sigma_n} < \infty, \quad n \geqslant 1.$$

In particular,

$$\sqrt{\sum_{s \leqq \cdot} (\Delta M_s)^2} \in A^+_{\mathrm{loc}(\tau)}(P), \ \ \langle M^c \rangle \in A^+_{\mathrm{loc}(\tau)}(P).$$

The condition "with a root" is inconvenient to handle, but one can show [3, Lemma 1] that in fact

$$\sqrt{\sum_{s \leqq \cdot} (\Delta M_s)^2} \in A^+_{\mathrm{loc}(\tau)}(P) \Leftrightarrow \sum_{s \leqq \cdot} \frac{(\Delta M_s)^2}{1 + |\Delta M_s|} \in A^+_{\mathrm{loc}(\tau)}(P).$$

Therefore, by virtue of M being a τ-local martingale, the process

$$\langle M^c \rangle + \sum_{s \leqq \cdot} \frac{(\Delta M_s)^2}{1 + |\Delta M_s|} \in A^+_{\mathrm{loc}(\tau)}(P)$$

or, which is the same,

$$\langle M^c\rangle+\frac{x^2}{1+|x|}*\mu\in A^+_{\mathrm{loc}\,(\tau)}(P) \tag{9}$$

where

$$\frac{x^2}{1+|x|}*\mu_t\equiv\int_0^t\int_E\frac{x^2}{1+|x|}\,d\mu=\sum_{s\leqq t}\frac{(\Delta M_s)^2}{1+|\Delta M_s|}.$$

If

$$\frac{x^2}{1+|x|}*\mu\in A^+_{\mathrm{loc}\,(\tau)}(P),$$

then its compensator $[x^2/(1+|x|)]*\nu$ also belongs to the class $A^+_{\mathrm{loc}(\tau)}(P)$.

Hence, the process

$$\boldsymbol{B}(M)\in A^+_{\mathrm{loc}\,(\tau)}(P), \tag{10}$$

where

$$\boldsymbol{B}_t(M)=\langle M^c\rangle_t+\frac{x^2}{1+|x|}*\nu_t. \tag{11}$$

The process $\boldsymbol{B}(M)$ has a fundamental role in the ACS problem since it is its properties that are decisive for the question as to the absolute continuity or singularity of the probability measures P and $\tilde{P}$. If one examines the properties of the process $\boldsymbol{B}(M)$ with respect to the measure P, the utmost one can do is to give the sufficient condition for $\tilde{P}\ll P$. One such condition is as follows ([3, Theorem 12]; see also [13], [14]):

THEOREM 2. *Let* $\tilde{P}\overset{\mathrm{loc}}{\ll}P$ *and* C *be a constant. Then*

$$P(\boldsymbol{B}_\tau(M)\leqslant C)=1\Rightarrow\tilde{P}\ll P. \tag{12}$$

It is noteworthy, however, that in passing over to the measure $\tilde{P}$ one can obtain the necessary and sufficient conditions both for the absolute continuity and the singularity of the probability measures ([1]–[3]):

THEOREM 3. *Let* $\tilde{P}\overset{\mathrm{loc}}{\ll}P$. *Then*

$$\begin{aligned}\tilde{P}(\boldsymbol{B}_\infty(M)<\infty)=1&\Leftrightarrow\tilde{P}\ll P,\\ \tilde{P}(\boldsymbol{B}_\infty(M)=\infty)=1&\Leftrightarrow\tilde{P}\perp P.\end{aligned} \tag{13}$$

Let us well on the main points in the proof of that theorem.

By virtue of Theorem 1, we have only to ascertain that $\tilde{P}$-a.s. $\{Z_\infty<\infty\}=\{\boldsymbol{B}_\infty^{(M)}<\infty\}$. Since $\tilde{P}(Z_\infty>0)=1$, so, by (6), this is equivalent to the proof that $\tilde{P}$-a.s.

$$\{\varphi\to\}=\{\boldsymbol{B}_\infty(M)<\infty\}$$

where

$$\varphi_t=M_t-\frac{1}{2}\langle M^c\rangle_t+\sum_{s\leqq t}[\ln(1+\Delta M_s)-\Delta M_s]$$

and $\{\varphi\to\}$ is the set of those elementary events for which $\lim_{t\to\infty}\varphi_t$ exists and is finite.

If M is a continuous process, then

$$\varphi_t = M_t - \frac{1}{2}\langle M_t\rangle = (M_t - \langle M\rangle_t) + \frac{1}{2}\langle M\rangle_t$$

where the process $N = M - \langle M\rangle$ is with respect to $\tilde{P}$ a continuous local martingale with $\langle N\rangle = \langle M\rangle$ ($\tilde{P}$-a.s.). For such martingales it is known (cf., e.g., [3], [15], and a more general statement below in Theorem 4) that $\{N\to\} = \{\langle N\rangle\to\}$ ($\tilde{P}$-a.s.). Therefore, $\{\varphi\to\} = \{\langle M\rangle\to\}$ ($\tilde{P}$-a.s.), i.e., $\{Z_\infty < \infty\} = \{B_\infty^{(M)} < \infty\}$ ($\tilde{P}$-a.s.).

The general case is somewhat more involved. For considering it, let us introduce the function

$$u(x) = \begin{cases} x, & |x| \le 1, \\ \operatorname{sign} x, & |x| > 1 \end{cases}$$

and pose

$$\varphi_t^u = M_t - \frac{1}{2}\langle M^c\rangle_t + \sum_{s\le t}\left[u(\ln(1+\Delta M_s)) - \Delta M_s\right].$$

It must be noted that, while the increaments $\Delta\varphi_t = \ln(1+\Delta M_t)$ can take on arbitrarily great (modulo) values, the increments $\Delta\varphi_t^u = u(\ln(1+\Delta M_t))$ are such that $|\Delta\varphi_t^u| \le 1$.

By a rather simple analysis one sees that $\tilde{P}$-a.s.

$$\{\varphi^u\to\} = \{\varphi\to\}.$$

So one must only ascertain that $\tilde{P}$-a.s.

$$\{\varphi^u\to\} = \{B_\infty(M) < \infty\}.$$

Transforming φ_t^u, we find that

$$\begin{aligned}
\varphi_t^u &= M_t^c + M_t^d - \frac{1}{2}\langle M^c\rangle_t + \sum_{s\le t}\left[u(\ln(1+\Delta M_s)) - \Delta M_s\right] \\
&= M_t^c - \frac{1}{2}\langle M^c\rangle_t + x*(\mu-\nu)_t + \left[u(\ln(1+x)) - x\right]*\mu_t \\
&= M_t^c - \frac{1}{2}\langle M^c\rangle_t + u(\ln(1+x))*(\mu-\nu)_t + \left[u(\ln(1+x)) - x\right]*\nu_t \qquad (14) \\
&= M_t^c - \frac{1}{2}\langle M^c\rangle_t + u(\ln Y)*(\mu-\tilde{\nu})_t + [Yu(\ln Y) + Y - 1]*\nu_t \\
&= [M_t^c - \langle M^c\rangle_t + u(\ln Y)*(\mu-\tilde{\nu})_t] + \left[\frac{1}{2}\langle M^c\rangle_t + (Yu(\ln Y) + Y - 1)*\nu_t\right] \\
&(= N_t + D_t)
\end{aligned}$$

where we have denoted $Y = 1 + x$ and used the fact that the compensator $\tilde{\nu}$ of the measure μ with respect to $\tilde{P}$ is connected with the compensator ν by the expression ([16], [3])

$$d\tilde{\nu} = Y\,d\nu.$$

At $Y \geqslant 0$, the function $Yu(\ln Y)+Y-1 \geqslant 0$ and, since $\Delta M_t \geqslant -1$, the measure ν can be chosen so that $I(Y<0)*\nu_\infty = 0$. Hence, $D_t \geqslant 0$, and the process N is (with respect to $\tilde{P}$) a locally square-integrable martingale with

$$\langle N\rangle_t = \langle M^c\rangle_t + Yu^2(\ln Y)*\nu_t - \sum_{s\leqq t}\Big[\int_E Yu(\ln Y)\nu(\{s\},dx)\Big]^2. \tag{15}$$

Hence, and from (14), it follows that the process $\varphi^u = N+D$ is a local submartingale with $|\Delta\varphi_t^u| \leqslant 1$.

Let us now make use of the following result ([3], [17]):

Theorem 4. *Let $X=M+A$, where M is a local martingale, A is a non-decreasing process of a locally integrable variation, $A_0=0$ and $|\Delta X_t| \leqslant C$, $t \geqslant 0$. Then (almost surely)*

$$\{A_\infty + \langle M\rangle_\infty < \infty\} = \{X \to\}. \tag{16}$$

From this theorem we find that $\tilde{P}$-a.s.

$$\{D_\infty + \langle N\rangle_\infty < \infty\} = \{\varphi^u \to\}. \tag{17}$$

Since

$$\sum_{s\leqq t}\int_E Yu(\ln Y)\nu(\{s\},dx) = \int_0^t I(\nu(\{s\},E)>0)\,dD_s$$

so

$$\begin{aligned}\{D_\infty+\langle N\rangle_\infty<\infty\} &= \Big\{\frac{3}{2}\langle M^c\rangle_\infty + Yu^2(\ln Y)*\nu_\infty - \Big[\int_0^\infty I(\nu(\{s\},E)>0)\,dD_s\Big]^2 \\ &\qquad + [Yu(\ln Y)+Y-1]*\nu_\infty < \infty\Big\} \\ &= \Big\{\langle M^c\rangle_\infty + [Yu^2(\ln Y)+Yu(\ln Y)+Y-1]*\nu_\infty \\ &\qquad - \Big[\int_0^\infty I(\nu(\{s\},E)>0)\,dD_s\Big]^2 < \infty\Big\}.\end{aligned}$$

But

$$\{D_\infty < \infty\} \subseteqq \Big\{\int_0^\infty I(\nu(\{s\},E)>0)\,dD_s < \infty\Big\}$$

and hence

$$\{D_\infty + \langle N\rangle_\infty < \infty\} = \{\langle M^c\rangle_\infty + [Yu^2(\ln Y)+Yu(\ln Y)+Y-1]*\nu_\infty < \infty\}.$$

At $Y \geqslant 0$,

$$Yu^2(\ln Y)+Yu(\ln Y)+Y-1 \asymp (1-\sqrt{Y})^2$$

and[1] at $x \geqslant -1$,

$$(1-\sqrt{1+x})^2 \asymp \frac{x^2}{1+|x|}.$$

---

[1] $f \asymp g$, if there exist nonzero constants c_1 and c_2 such that for all argument values $c_1 g < f < c_2 g$.

Therefore ($\tilde{P}$-a.s.)

$$\{\varphi \to\} = \{\varphi^u \to\} = \{\langle M^c\rangle_\infty + (1-\sqrt{1+x})^2 * \nu_\infty < \infty\}$$
$$= \left\{\langle M^c\rangle_\infty + \frac{x^2}{1+|x|} * \nu_\infty < \infty\right\}$$

which, together with Theorem, proves Theorem 3.

How is Theorem 3 reformulated in the case of discrete time? Let (Ω, F) be a measurable space, $(F_n)_{n\geq 0}$ a nondecreasing family of σ-algebras such that $F = \bigvee_n F_n$, P and $\tilde{P}$ two probability measures, and P_n and $\tilde{P}_n$ their restrictions on F_n, $\tilde{P}_n \ll P_n$, $n \geq 0$, $Z_n = d\tilde{P}_n/dP_n$, $M_n = \sum_{k=1}^n Z_{k-1}^\oplus \Delta Z_k$, $\alpha_n = 1 + \Delta M_n$, $\mu_n(\Gamma) = I(\Delta M_n \in \Gamma)$, $\nu_n(\Gamma) = P(\Delta M_n \in \Gamma | F_{n-1})$, $\Gamma \in B(E)$. Then $\tilde{P}$-a.s.

$$B_\infty(M) = \sum_{n=1}^\infty \int_E (1-\sqrt{1+x})^2 \nu_n(dx) = \sum_{n=1}^\infty E[(1-\sqrt{1+\Delta M_n})^2 | F_{n-1}]$$
$$= \sum_{n=1}^\infty E[(1-\sqrt{\alpha_n})^2 | F_{n-1}] = 2\sum_{n=1}^\infty E(1-\sqrt{\alpha_n} | F_{n-1}).$$

Therefore, if $\tilde{P}_n \ll P_n$, $n \geq 0$, then

$$\tilde{P} \ll P \Leftrightarrow \tilde{P}\left\{\sum_{n=1}^\infty E(1-\sqrt{\alpha_n}|F_{n-1}) < \infty\right\} = 1,$$
$$\tilde{P} \perp P \Leftrightarrow \tilde{P}\left\{\sum_{n=1}^\infty E(1-\sqrt{\alpha_n}|F_{n-1}) = \infty\right\} = 1.$$

In particular, in the case examined by Kakutani the alternative "either $\tilde{P} \ll P$, or $\tilde{P} \perp P$" holds, and also

$$\tilde{P} \ll P \Leftrightarrow \sum_{n=1}^\infty E(1-\sqrt{\alpha_n}) < \infty,$$
$$\tilde{P} \perp P \Leftrightarrow \sum_{n=1}^\infty E(1-\sqrt{\alpha_n}) = \infty,$$

where $\alpha_n = d\tilde{\mu}_n/d\mu_n$.

We shall now formulate two propositions which follow immediately from Theorem 3 (for more detail, see [3]).

Suppose that the τ-local martingale has this structure:

$$M = \gamma^* \cdot m + [(Y^*-1) + (1-a)^\oplus (\hat{Y}^* - a)] * (\mu - \nu), \tag{18}$$

where μ is an integer random measure (not necessarily a measure of the jumps of M), ν — its compensator, m — the τ-local continuous martingale, γ^* — the predictable process with $\gamma^* \cdot m_{\tau_n} < \infty$, $n \geq 1$, and Y^* — a $\tilde{P}$-predictable process such that $0 \leq Y^*(t, x) < \infty$, $\hat{Y}_t \equiv \int_E Y^*(t, x)\nu(\{t\}, dx) \leq 1$, $a_t \equiv \nu(\{t\}, E) \Rightarrow \hat{Y}^*(t, x) = 1$ and

$$(1-\sqrt{Y^*})^2 * \nu_{\tau_n} + \sum_{s \leq \tau_n} I(a_s < 1)\left(1 - \sqrt{\frac{1-\hat{Y}_s^*}{1-a_s}}\right)^2 (1-a_s) < \infty, \ n \geq 1.$$

(Hereafter everywhere $(H\cdot X)_t$ denotes the stochastic integral $\int_0^t H_s\,dX_s$ with respect to the semimartingale X, [9].)

Now we obtain directly from Theorem 3 this

CONSEQUENCE 1. Let $\tilde{P}\overset{\text{loc}}{\ll} P$ and (18) holds. Then

$$\tilde{P}\ll P \Leftrightarrow \tilde{P}\{B_\infty^* <\infty\}=1,$$

$$\tilde{P}\perp P \Leftrightarrow \tilde{P}\{B_\infty^* =\infty\}=1$$

where

$$B_t^* = (\gamma^*)^2\cdot\langle m\rangle_t + \left(1-\sqrt{Y^*}\right)^2 * \nu_t + \sum_{s\le t} I(a_s<1)\left(1-\sqrt{\frac{1-\hat{Y}_s^*}{1-a_s}}\right)^2(1-a_s). \tag{19}$$

CONSEQUENCE 2. Let $\tilde{P}\overset{\text{loc}}{\ll} P$, $Z_t=d\tilde{P}_t/dP_t$, μ is an integer random measure with the condensator ν, m is the continuous local martingale, and

$$Y^{**}=Z_-^{\oplus}E_\mu^P(Z\mid\tilde{\varphi}),\quad \gamma^{**}=Z_-^{\oplus}\frac{d\langle Z^c, m\rangle}{d\langle m\rangle},$$

$$B_t^{**} = (\gamma^{**})^2\cdot\langle m\rangle_t + \left(1-\sqrt{Y^{**}}\right)^2 * \nu_t + \sum_{s\le t}\left(1-\sqrt{\frac{1-\tilde{a}_s}{1-a_s}}\right)^2(1-a_s) \tag{20}$$

with $a_s=\nu(\{s\},E)$, $\tilde{a}_s=\tilde{\nu}(\{s\},E)$. Then

$$\tilde{P}\ll P \Rightarrow \tilde{P}(B_\infty^{**}<\infty)=1.$$

We shall now consider the ACS problem for the particular classes of random processes with continuous time. The approach presented below and based on a combined application of Theorem 2 and Theorem 3 (with its Consequences 1 and 2) has the merit of affording a uniform view of the various specific cases. According to Theorem 3, the decision as to the absolute continuity or singularity implies the ability of finding, in terms of local characteristics, the conditions for local absolute continuity. The examples discussed below demonstrate the manner in which this is done.

4. Processes with independent increments. These processes are a natural analogue to sequences of independent random variables, so it will be logical to begin at them. Under rather general assumptions (cf., e.g., [9], [3]), the process $X=(X_t, P)$ with independent increments is a semimartingale.

But every semimartingale $X=(X_t, P)$ admits of the canonical decomposition (P-a.s.).

$$X_t = X_0+\alpha_t+m_t+\int_0^t\int_{|x|>1} x\,d\mu+\int_0^t\int_{|x|\le 1} x\,d(\mu-\nu), \tag{21}$$

where α is a predictable process, m is a continuous local martingale, μ is a measurt of the jumps of X, and ν is its compensator. The set (α,β,ν), where $\beta=\langle m\rangle$, is referred to as the triplet of predictable characteristics of the semimartingale X.

When the semimartingale X is a process with independent increments, the triplet (α,β,ν) is nonrandom. What is essential is that for processes with independent increments the triplet defines uniquely the distribution of the probabilities P (in the space of right-continuous and left-hand limited functions), [18].

Let $\tilde{X}=(X_t, \tilde{P})$ be another process with independent increments with the triplet $(\tilde{\alpha}, \tilde{\beta}, \tilde{\nu})$. We desire to find the necessary and sufficient conditions on the triplets (α, β, ν) and $(\tilde{\alpha}, \tilde{\beta}, \tilde{\nu})$ for the absolute continuity and singularity of the probability measures P and $\tilde{P}$.

We shall begin with the necessary conditions. Let $\tilde{P} \ll P$. Then, evidently,

$$\text{(I)} \qquad \tilde{P}_0 \ll P_0.$$

In [16], it has been shown that the condition $\tilde{P} \ll P$ entails the fulfilment of the following condition:

$$\text{(II)} \quad \begin{aligned} &\text{(a)} && d\tilde{\nu} = Y d\nu, \\ &\text{(b)} && \nu(\{t\}, E) = 1 \Rightarrow \tilde{\nu}(\{t\}, E) = 1, \\ &\text{(c)} && \langle \tilde{m} \rangle = \langle m \rangle, \\ &\text{(d)} && \tilde{\alpha}_t - \alpha_t - \int_0^t \int_{|x| \leqq 1} x(Y-1)\, d\nu = \int_0^t \gamma_s \, d\langle m \rangle_s \end{aligned}$$

with

$$Y = Z_-^{\oplus} E_\mu^P(Z|\tilde{\varphi}), \; \gamma = Z_-^{\oplus} \frac{d\langle Z^c, m\rangle}{d\langle m \rangle}$$

where $Z_t = d\tilde{P}_t/dP_t$ and $E_\mu^P(Z|\tilde{P})$ is the "conditional expectation with respect to the measure $\mu(dt, dx)P(d\omega)$ and to the σ-algebra of the $\tilde{\varphi}$-predictable events", [10].

We now introduce the (deterministic) function $(a_s = \nu(\{s\}, E), \tilde{a}_s = \tilde{\nu}(\{s\}, E))$

$$B_t = \gamma^2 \cdot \langle m \rangle_t + (1-\sqrt{Y})^2 * \nu_t + \sum_{s \leqq t} I(a_s < 1) \left(1 - \sqrt{\frac{1-\tilde{a}_s}{1-a_s}}\right)^2 (1-a_s) \tag{22}$$

and formulate the following group of conditions:

$$\text{(III)} \quad \begin{aligned} &\text{(a)} && B_t < \infty, \; t < \infty, \\ &\text{(b)} && B_\infty < \infty, \\ &\text{(c)} && B_\infty = \infty. \end{aligned}$$

The functions B^{**} and B, introduced into (20) and (22), respectively, coincide, and, therefore, by Consequence 2 to Theorem 3, $B_\infty < \infty$.

Thus,

$$\tilde{P} \ll P \Rightarrow \text{I, II, III}_b.$$

The inverse implication holds as well. Indeed, by virtue of conditions I, II, III_b, the process

$$M = \gamma \cdot m + [(Y-1) + (1-a)^{\oplus}(\hat{Y}-a)] * (\mu - \nu) \tag{23}$$

is defined, which is a local martingale with $\Delta M \geqslant -1$. Therefore, there exists a nonnegative solution of the equation $dZ = Z_- dM$. In the case under consideration, the function B is nonrandom and, hence, by virtue of condition $\mathrm{III_b}$ $(B_\infty<\infty)$ and Theorem 2, the family $(Z_t)_{t\geq 0}$ is uniformly integrable with respect to the measure P. Therefore, $EZ_\infty=1$ and the measure P' with $dP'=Z_\infty dP$ is probabilistic.

With respect to the measure P', the process $(X_t)_{t\geq 0}$ also is a process with independent increments. Using the specifics of process (23) and the rules of recalculation of the local characteristics of semimartingales in the case of an absolutely continuous substitution of the measure (cf. Condition II), one can find that the triplet (α', β', ν') of the process $X'=(X_t, P')$ coincides with the triplet $(\tilde{\alpha}, \tilde{\beta}, \tilde{\nu})$. For the processes with independent increments, the triplet defines uniquely the probabilities distribution and, hence, $P'=\tilde{P}$, $d\tilde{P}=Z_\infty dP$, and I, II, $\mathrm{III_b}\Leftrightarrow\tilde{P}\ll P$.

The same method of proof shows that I, II, $\mathrm{III_a}\Leftrightarrow\tilde{P}\overset{\mathrm{loc}}{\ll} P$.

Thus, there takes place the following

THEOREM 5. *If* $X=(X_t, P)$, $\tilde{X}=(X_t, \tilde{P})$ *are two processes with independent increments, then*

(1) I, II, $\mathrm{III_a}\Leftrightarrow\tilde{P}\overset{\mathrm{loc}}{\ll} P$,

(2) *if* $\tilde{P}\overset{\mathrm{loc}}{\ll} P$, *then the alternative "either* $\tilde{P}\ll P$, *or* $\tilde{P}\perp P$*" takes place and, in addition,*

$$\mathrm{III_b} \Leftrightarrow \tilde{P} \ll P,$$
$$\mathrm{III_c} \Leftrightarrow \tilde{P} \perp P.$$

Note also that from this proof it follows as well that the density $Z_t=d\tilde{P}_t/dP_t$ is a solution of the equation $dZ=Z_-dM$, where M is defined by (23).

5. Semimartingales. Let $X=(X_t, P)$ be a semimartingale, i.e., a process admitting of being represented as

$$X_t = X_0+A_t+M_t$$

where M is a local martingale and A is a process of locally bounded variation. Each semimartingale admits of the canonical representation (21) in which the triplet of local characteristics is, generally speaking, random and does not define uniquely the measure P. It is, indeed, because of this non-uniqueness that one has, apart from such conditions as I, II, III of the preceding section, to introduce, in order to formulate the ACS conditions, one more condition

"(IV) The measure $\tilde{P}$ is (σ_n)-unique",

first introduced in [**16**] and meaning that "stopped" triplets $(\tilde{\alpha}_{t\wedge\sigma_n}, \tilde{\beta}_{t\wedge\sigma_n}, \tilde{\nu}((0, t\wedge\sigma_n), dx))$ define uniquely the restrictions $\tilde{P}_{\sigma_n}$ of the measure $\tilde{P}$ on the σ-algebras F_{σ_n}, where $\sigma_n=\inf\{t: B_t>n\}$, and the process B is defined by (22).

THEOREM 6. *Let* $X=(X_t, P)$ *and* $\tilde{X}=(X_t, \tilde{P})$ *be two semimartingales and the condition* IV *be fulfilled. Then*

(1) I, II, $\mathrm{III_a}\Leftrightarrow\tilde{P}\overset{\mathrm{loc}}{\ll} P$,

(2) *if* $\tilde{P}\overset{\mathrm{loc}}{\ll} P$, *then*

$$\mathrm{III_b} \Leftrightarrow \tilde{P} \ll P,$$
$$\mathrm{III_c} \Leftrightarrow \tilde{P} \perp P.$$

The proof of the statement I, II, $\mathrm{III}_a \Leftarrow \tilde{P} \ll^{\mathrm{loc}} P$ is the same as in Theorem 5. The proof of the inverse statement involves a difficulty with the application of Theorem 2 (unlike the case of the processes with independent increments); the difficulty is caused by the "randomness" of the function $\boldsymbol{B}$. However, since $\Delta \boldsymbol{B}_t \leqslant 2$ the random variable $\boldsymbol{B}_{\sigma_n} \leqslant n+2$ and, consequently, the corresponding family of random variables $(Z_{t\wedge\sigma_n})_{t\geqq 0}$ is uniformly integrable (Theorem 2). As in Theorem 5, hence it is derived that then $\tilde{P}_{\sigma_n} \ll P_{\sigma_n}$, $n \geqslant 1$, and, as a consequence to this, $\tilde{P} \ll^{\mathrm{loc}} P$.

As in the preceding case, the density $Z_t = d\tilde{P}_t/dP_t$ is also the solution of the equation $dZ = Z_- dM$ with the process M defined by (23).

6. Semimartingales with Gaussian martingale part. Let $X=(X_t, P)$ and $X=(X_t, \tilde{P})$ be two semimartingales,

$$X_t = X_0 + M_t, \quad X_t = X_0 + \tilde{A}_t + \tilde{M}_t$$

whose martingale parts, M and $\tilde{M}$, are Gaussian martingales. The question of the absolute continuity and singularity of the measures P and $\tilde{P}$ of such (generally, non-Gaussian) processes has been investigated by a great number of authors ([19]–[27]). The method presented above affords the following results.

Introduce the conditions:

(I) $$\tilde{P}_0 \ll P_0,$$

(II)
$$\text{(a)} \quad \tilde{A} = \gamma \cdot \langle \tilde{M} \rangle,$$
$$\text{(b)} \quad \langle \tilde{M}^c \rangle = \langle M^c \rangle,$$
$$\text{(c)} \quad \langle \tilde{M}^d \rangle = \varrho \cdot \langle M^d \rangle, \ \langle M^d \rangle = \varrho^{-1} \cdot \langle \tilde{M}^d \rangle$$

(III)
$$\text{(a)} \quad \tilde{P}(\boldsymbol{B}_t < \infty) = 1, \ t < \infty,$$
$$\text{(b)} \quad \tilde{P}(\boldsymbol{B}_\infty < \infty) = 1,$$
$$\text{(c)} \quad \tilde{P}(\boldsymbol{B}_\infty = \infty) = 1,$$

where
$$\boldsymbol{B}_t = \gamma^2 \cdot \langle \tilde{M} \rangle_t + \sum_{s \leqq t} I(\Delta \langle M \rangle_s > 0)(1 - \varrho_s)^2.$$

THEOREM 7. *The following statements hold:*

(1) I, II, $\mathrm{III}_a \Rightarrow \tilde{P} \ll^{\mathrm{loc}} P$,

(2) I, II, $\mathrm{III}_b \Rightarrow \tilde{P} \ll P$

and if, in addition, $\tilde{P}_0 \sim P_0$ *and*

$$\tilde{E} \exp\left(-\gamma \cdot \tilde{M}_\infty - \frac{1}{2} \gamma^2 \cdot \langle \tilde{M} \rangle_\infty\right) = 1,$$

then $\tilde{P} \sim P$.

THEOREM 8. *Let the process* $\tilde{A}$ *be nonanticipative functional of* $X\,(\tilde{A}_t=\tilde{A}_t(X))$. *Then*

(1) $\mathrm{I}, \mathrm{II}, \mathrm{III}_a \Leftrightarrow \tilde{P} \ll^{\mathrm{loc}} P$,

(2) *if* $\tilde{P} \ll^{\mathrm{loc}} P$, *then*

$$\mathrm{III}_b \Leftrightarrow \tilde{P} \ll P,$$
$$\mathrm{III}_c \Leftrightarrow \tilde{P} \perp P.$$

Consider a special case of these statements (for their proofs, see [3]).

Let a Gaussian martingale M and a semimartingale $X=A+M$ be defined on the probability space (Ω, F, F_t, P). Denote their probabilities distributions by P^M and P^X. Then

$$A = \gamma\cdot\langle M\rangle,\ \gamma^2\cdot\langle M\rangle_\infty < \infty'' \Rightarrow P^X \ll P^M$$

and the conditions

$$A = \gamma\cdot\langle M\rangle,\ \gamma^2\cdot\langle M\rangle_\infty < \infty,\ E\exp\left\{-\gamma\cdot M_\infty - \frac{1}{2}\gamma^2\cdot\langle M\rangle_\infty\right\} = 1''$$

are necessary and sufficient for the equivalence of the measures P^M and P^X. Moreover,

$$dP^M(X)/dP^X = E\left\{\exp\left[-\gamma\cdot M_\infty - \frac{1}{2}\gamma^2\cdot\langle M\rangle_\infty\right]\middle|\, F^x\right\}$$

where $F^x=\sigma\{\omega\colon X_t,\ t\geqslant 0\}$.

As to the case when $A=A(X)$ is a nonanticipative functional of X, the density

$$dP^X/dP^M(X) = \exp\left\{-\gamma\cdot M_\infty - \frac{1}{2}\gamma^2\cdot\langle M\rangle_\infty\right\}. \tag{24}$$

It is noteworthy that for (arbitrary) Gaussian martingale M the nature of the absolute continuity (and singularity) conditions and the expression for the Radon–Nikodym derivative are the same as when M is a Wiener process [27]. Consider this case in a little more detail. If W is a standard Wiener process, then, as follows from the preceding results, every semimartingale $X=A(X)+W$ with process $A(X)$ such that $A_t(X)=\int_0^t \gamma_s\,ds$, $\int_0^\infty \gamma_s^2\,ds<\infty$ (P-a.s.) has the measure P^X absolutely continuous with respect to the Wiener measure. In a certain sense, the inverse result also holds (cf. [28], [27]); it is established with the aid of the preceding results and, among other things, gives an answer to the question put forward in § 1.

THEOREM 9. *Let a continuous random process* $X=(X_t, F_t)$ *and a Wiener process* $W=(W_t, F_t)$ *such that* $P^x \ll P^w$ *be defined on a complete probability space* (Ω, F, F_t, P). *Then there exist such a Wiener process* $\hat{w}=(\hat{w}_t, F_t^x)$ *and a nonanticipative functional* $\gamma=\gamma(x)$ *that* P-a.s.

$$X_t = \int_0^t \gamma_s(X)\,ds + \hat{w}_t,\ t\geqslant 0.$$

7. Multivariate point processes. Let (Ω, F) be a measurable space with a non-decreasing family of σ-algebras $(F_t)_{t\geqq 0}$, $F=\bigvee_{t\geqq 0} F_t$, $(E, \mathscr{E})$ being a Luzin space, Δ an auxiliary point, and $E_\Delta = E\cup\{\Delta\}$, $\mathscr{E}_\Delta=\mathscr{E}\vee\{\Delta\}$. According to [10], [11], a multivariant point process is a term applied to a sequence $(T_n, \xi_n)_{n\geqq 0}$, where T_n are Markov times with the properties

(1) $T_0\equiv 0, \quad T_1>0,$

(2) $T_{n+1}>T_n \quad \text{if} \quad T_n<\infty$

(3) $T_{n+1}=T_n \quad \text{if} \quad T_n=\infty$

and ξ_n are F_{T_n}-measurable random elements with values in $(E_\Delta, \mathscr{E}_\Delta)$, and $X_n=\Delta$ if and only if $T_n=\infty$.

Let $T=\lim_n T_n$ be the point of accumulation.

Every multivariate point process can be conveniently defined by means of the integer random measure μ on $(0, \infty)\, E$:

$$\mu((0, t], \Gamma) = \sum_{n\geqq 1} I(T_n\leqslant t, \xi_n\in\Gamma), \quad \Gamma\in\mathscr{E}.$$

Pose $F_t^\mu=\sigma\{\mu((0,s], B), s\leqslant t, \ B\in\mathscr{E}\}$, $G_t=F_0\vee F_t^\mu$, $G=\bigvee_{t\geqq 0} G_t$ and assume that $F_t=G_t$, $F=G$. Let, further, P and $\tilde{P}$ be two probability measures on (Ω, G) and ν and $\tilde{\nu}$ be compensators of the measure μ with respect to P and $\tilde{P}$, respectively.

Let us formulate the conditions to be used in Theorem 10 below:

(I) $$\tilde{P}_0\ll P_0,$$

(II) (a) $$d\tilde{\nu}=Y d\nu \quad (\tilde{P}\text{-a.s.}),$$

(b) $$\nu(\{t\}, E)=1\Rightarrow\tilde{\nu}(\{t\}, E)=1 \quad (\tilde{P}\text{-a.s.}).$$

Pose

$$B_t=(1-\sqrt{Y})^2 * \nu_t+\sum_{s\leqq t} I(a_s<1)\left(1-\sqrt{\frac{1-\tilde{a}_s}{1-a_s}}\right)^2(1-a_s)$$

where $a_s=\nu(\{s\}, E)$, $\tilde{a}_s=\tilde{\nu}(\{s\}, E)$ and let

(III) (a) $$\tilde{P}(B_t=\infty, \ t<T)=0,$$

(b) $$\tilde{P}(B_T<\infty)=1,$$

(c) $$\tilde{P}(B_T=\infty)=1.$$

THEOREM 10. *For multivariate point processes* $X=\{(T_n, \xi_n), P\}$ *and* $\tilde{X}=\{(T_n, \xi_n), \tilde{P}\}$, *the following statements hold:*

(1) I, II, $\text{III}_a\Leftrightarrow\tilde{P}\overset{\text{loc}}{\ll} P$;

(2) *if* $\tilde{P}\overset{\text{loc}}{\ll} P$, *then*

$$\text{III}_b\Leftrightarrow\tilde{P}\ll P,$$

$$\text{III}_c\Leftrightarrow\tilde{P}\perp P.$$

The theorem is proved according to the same method as Theorems 5 and 6. We shall just mention that in the case under consideration—the multivariate point processes—the compensator defines uniquely the probabilities distribution, bearing an analogy to the case of the processes with independent increments, where the triplets of local characteristics, too, defined these distributions uniquely.

From the proof it follows also that the density $Z_t = d\tilde{P}_t/dP_t$, $t \geqslant 0$, is a solution of the following equation

$$Z_t = d\tilde{P}_0/dP_0 + Z_-[(Y-1)+(1-a)^{\oplus}(\hat{Y}-a)] * (\mu-\nu)_t,$$

where $\hat{Y}_s = \int_E Y(s,x)\,\nu(\{s\},dx)$, $a_s = \nu(\{s\},E)$.

An important special case of the multivariate point processes is the so-called point or counting processes for which $X_n \equiv 1$.

Denote $X_t = \mu((0,t],\{1\})$ $(=\sum_{n\geq 1} I(T_n \leqslant t))$, $A_t = \nu((0,t],\{1\})$, $\tilde{A}_t = \tilde{\nu}((0,t],\{1\})$. Then condition I will be fulfilled in an evident fashion and condition II will appear as

$$\text{(II)} \qquad \begin{aligned} &\text{(a)} \quad \tilde{A}_t = \int_0^t Y_s\,dA_s, \\ &\text{(b)} \quad \Delta A_t = 1 \Rightarrow \Delta\tilde{A}_t = 1 \quad (\tilde{P}\text{-a.s.}) \end{aligned}$$

and the function

$$B_t = \int_0^t (1-\sqrt{Y_s})^2\,dA_s + \sum_{s\leqq t} I(0<\Delta A_s<1)\left(1-\sqrt{\frac{1-\Delta\tilde{A}_s}{1-\Delta A_s}}\right)^2 (1-\Delta A_s).$$

An elementary example of a point process is the Poisson process $X = (X_t, P)$ with parameter equal to one $(A_t = t)$. Now if $\tilde{X} = (X_t, \tilde{P})$ is another point process, then, by Theorem,

$$\tilde{P} \ll P \Leftrightarrow \text{“}\tilde{A}_t = \int_0^t Y_s\,ds,\ \int_0^\infty (1-\sqrt{Y_s})^2\,ds < \infty\ (\tilde{P}\text{-a.s.})\text{”}.$$

From this result, it follows that the question posed in § 1 is answered: If $\tilde{X} = (X_t, \tilde{P})$ is a point process whose measure is absolutely continuous with respect to the Poisson measure, then this process inevitably must have the following structure

$$X_t = \int_0^t Y_s\,ds + M_t$$

where M is a local martingale and the predictable process Y is such that ($\tilde{P}$-a.s.) $\int_0^\infty (1-\sqrt{Y_s})^2\,ds < \infty$.

8. Concluding remarks. Theorems 1 and 3 give general and “predictable” criteria of absolute continuity and singularity for two probability measures one of which is locally absolutely continuous with respect to the other. For the processes with independent increments, semimartingales and multivariant point processes, it has been shown how these criteria are restated in terms of the local characteristics of the processes concerned. As regards the other examples of the efficiency of ACS conditions, the reader is referred to [3]. The corresponding results pertinent to the case of discrete time have been exposed in [1], [4].

References

1. Кабанов Ю. М., Липцер Р. Ш. и Ширяев А. Н., *К вопросу об абсолютной непрерывности и сингулярности вероятностных мер*, Матем. сб. 104, **2** (1977), 227—247.

2. *«Предсказуемые» критерии абсолютной непрерывности и сингулярности вероятностных мер (случай непрерывного времени)*, ДАН СССР, 237, 5 (1977), 1016—1019.

3. *Абсолютная непрерывность и сингулярность локально абсолютно непрерывных вероятностных распределений*, Матем. сб (3) **107** (1978), (1) **108** (1979).

4. H.-J. Engelbert and A. N. Shiryayev, *On absolute continuity and singularity of probability measures*, Banach Center Publications, vol. 6: Mathematical Statistics, 1979.

5. S. Kakutani, *On equivalence of infinite product measures*, Ann. of Math. **49** (1948), 214—224.

6. Гаек, Я. *Об одном свойстве нормальных распределений произвольных стохастических процессов*, Чехосл. матем. журн. **8** (1958), 610—618.

7. J. Feldman, *Equivalence and perpendicularity of Gaussian processes*, Pacific J. Math. **8** (1958), 699—708.

8. C. Doleans-Dade, *Existence and unicity of solutions of stochastic differential equations*, Z.W-theorie **36** (1976), 93—102.

9. P. A. Meyer, *Un cours sur les intégrales stochastiques*, Lectures Notes in Math., vol. 511, Springer-Verlag, Berlin and New York, 1976, pp. 245—400.

10. J. Jacod, *Multivariate point processes: predictable projection, Radon-Nikodym derivatives, representation of martingales*, Z. W-theorie **31** (1975), 235—253.

11. Григелионис Б. И., *Случайные точечные процессы и мартингалы*, Лит. Матем. сб., ХУ, **3** (1975), 101—114.

12. J. Jacod, *Un théorème de représentation pour les martingales discontinues*, Z. W-theorie **34** (1976), 225—244.

13. J. Memin, *Décompositions multiplicatives de semimartingales exponentielles et application*, Sem. Probab., vol. 12, 1978.

14. J. Memin and A. N. Shiryayev, *Un critére prévisible pour l'uniforme intégrabilité des semimartingales exponentielles*, Sem. Probab., vol. 13, 1979.

15. E. Lenglart, *Sur la convergence presque sûre des martingales locales*, C. R. Acad. Sci. Paris Sér. A **284** (1977), 1085—1088.

16. J. Jacod, and J. Memin, *Caractéristiques locales et conditions continuité absolue pour les semimartingales*, Z. W-theorie **35** (1976), 1—37.

17. H.-J. Engelbert and A. N. Shiryayev, *On the sets of convergence of generalized submartingales*, Stochastics, 1979.

18. Григелионис Б. И., *О мартингальной характеризации случайных процессов с независимыми приращениями*, Литовский матем. сб., ХУП, **1** (1977).

19. Скороход А. В., *Случайные процессы с независимыми приращениями*, Изд-во «Наука», М., 1964.

20. Ибрагимов И. А. и Розанов Ю. А., *Гауссовские случайные процессы*, Изд-во «Наука», М., 1970.

21. L. A. Shepp, *Radon—Nikodym derivative of Gaussian measures*, Ann. Math. Statist, **37** (1966), 321—354.

22. Ершов М. П., *Об абсолютной непрерывности мер, отвечающих процессам диффузионного типа*, Теория вероятн. и ее. примен. **1** (1972), 173—178.

23. Гихман И. И. и Скороход А. В., *О плотностях вероятностных мер в функциональных пространствах*, УМН **21** (1966), 83—152.

24. T. Kailath and M. Zakai, *Absolute continuity and Radon—Nikodym derivatives for certain measures relative to Wiener measure*, Ann. Math. Statist. **42** (1971), 130—140.

25. Григелионис Б. И., *О структуре плотностей, соответствующих случайным процессам*, Литовский матем. сб., XIII, **1** (1973), 71—78.

26. S. Orey, *Conditions for the absolute continuity of two diffusions*, Trans. Amer. Math. Soc. **193** (1974), 413—426.

27. Липцер Р. и Ш. Ширяев А. Н., *Статистика случайных процессов*, Изд-во «Наука», М., 1974.

28. M. Hitsuda, *Representation of Gaussian processes equivalent to Wiener process*, Osaka J. Math. **5** (1968), 299—312.

Steklov Mathematical Institute
Moscow V333, U.S.S.R.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

History of Mathematics: Why and How

André Weil

My first point will be an obvious one. In contrast with some sciences whose whole history consists of the personal recollections of a few of our contemporaries, mathematics not only has a history, but it has a long one, which has been written about at least since Eudemos (a pupil of Aristotle). Thus the question "Why?" is perhaps superfluous, or would be better formulated as "For whom?".

For whom does one write general history? for the educated layman, as Herodotus did? for statesmen and philosophers, as Thucydides? for one's fellow-historians, as is mostly done nowadays? What is the right audience for the art-historian? his colleagues, or the art-loving public, or the artists (who seem to have little use for him)? What about the history of music? Does it concern chiefly music-lovers, or composers, or performing artists, or cultural historians, or is it a wholly independent discipline whose appreciation is confined to its own practitioners? Similar questions have been hotly debated for many years among eminent historians of mathematics, Moritz Cantor, Gustav Eneström, Paul Tannery. Already Leibniz had something to say about it, as about most other topics:

"Its use is not just that History may give everyone his due and that others may look forward to similar praise, but also that the art of discovery be promoted and its method known through illustrious examples." [1]

[1] *"Utilissimum est cognosci veras inventionum memorabilium origines, praesertim earum, quae non casu, sed vi meditandi innotuere. Id enim non eo tantum prodest, ut Historia literaria suum cuique tribuat et alii ad pares laudes invitentur, sed etiam ut augeatur ars inveniendi, cognita methodo illustribus exemplis. Inter nobiliora hujus temporis inventa habetur novum Analyseos Mathematicae genus, Calculi differentialis nomine notum..."* (Math. Schr., ed. C. I. Gerhardt, t. V, p. 392).

That mankind should be spurred on by the prospect of eternal fame to ever higher achievements is of course a classical theme, inherited from antiquity; we seem to have become less sensitive to it than our forefathers were, although it has perhaps not quite spent its force. As to the latter part of Leibniz' statement, its purport is clear. He wanted the historian of science to write in the first place for creative or would-be creative scientists. This was the audience he had in mind while writing in retrospect about his "most noble invention" of the calculus.

On the other hand, as Moritz Cantor observed, one may, in dealing with mathematical history, regard it as an auxiliary discipline, meant for providing the true historian with reliable catalogues of mathematical facts, arranged according to times, countries, subject-matters and authors. It is then a portion, and not a very significant one, of the history of techniques and crafts, and it is fair to look upon it entirely from the outside. The historian of the XIX$^{\text{th}}$ century needs some knowledge of the progress made by the railway engine; for this he has to depend upon specialists, but he does not care how the engine works, nor about the gigantic intellectual effort that went into the creation of thermodynamics. Similarly, the development of nautical tables and other aids to navigation is of no little importance for the historian of XVII$^{\text{th}}$ century England, but the part taken in it by Newton will provide him at best with a footnote; Newton as keeper of the Mint, or perhaps as the uncle of a great nobleman's mistress, is closer to his interests than Newton the mathematician.

From another point of view, mathematics may occasionally provide the cultural historian with a kind of "tracer" for investigating the interaction between various cultures. With this we come closer to matters of genuine interest to us mathematicians; but even here our attitudes differ widely from those of professional historians. To them a Roman coin, found somewhere in India, has a definite significance; hardly so a mathematical theory.

This is not to say that a theorem may not have been rediscovered time and again, even in quite different cultural environments. Some power-series expansions seem to have been discovered independently in India, in Japan and in Europe. Methods for the solution of Pell's equation were expounded in India by Bhaskara in the XII$^{\text{th}}$ century, and then again, following a challenge from Fermat, by Wallis and Brouncker in 1657. One can even adduce arguments for the view that similar methods may have been known to the Greeks, perhaps to Archimedes himself; as Tannery suggested, the Indian solution could then be of Greek origin; so far this must remain an idle speculation. Certainly no one would suggest a connection between Bhaskara and our XVII$^{\text{th}}$ century authors.

On the other hand, when quadratic equations, solved algebraically in cuneiform texts, surface again in Euclid, dressed up in geometric garb without any geometric motivation at all, the mathematician will find it appropriate to describe the latter treatment as "geometric algebra" and will be inclined to assume some connection with Babylon, even in the absence of any concrete "historical" evidence. No one asks for documents to testify to the common origin of Greek, Russian and Sanskrit, or objects to their designation as indo-european languages.

Now, leaving the views and wishes of laymen and of specialists of other disciplines, it is time to come back to Leibniz and consider the value of mathematical history, both intrinsically and from our own selfish viewpoint as mathematicians. Deviating only slightly from Leibniz, we may say that its first use for us is to put or to keep before our eyes "illustrious examples" of first-rate mathematical work.

Does that make historians necessary? Perhaps not. Eisenstein fell in love with mathematics at an early age by reading Euler and Lagrange; no historian told him to do so or helped him to read them. But in his days mathematics was progressing at a less hectic pace than now. No doubt a young man can now seek models and inspiration in the work of his contemporaries; but this will soon prove to be a severe limitation. On the other hand, if he wishes to go much further back, he may find himself in need of some guidance; it is the function of the historian, or at any rate of the mathematician with a sense for history, to provide it.

The historian can help in still another way. We all know by experience how much is to be gained through personal acquaintance when we wish to study contemporary work; our meetings and congresses have hardly any other purpose. The life of the great mathematicians of the past may often have been dull and unexciting, or may seem so to the layman; to us their biographies are of no small value in bringing alive the men and their environment as well as their writings. What mathematician would not like to know more about Archimedes than the part he is supposed to have taken in the defense of Syracuse? Would our understanding of Euler's number-theory be quite the same if we merely had his publications at our disposal? Is not the story infinitely more interesting when we read about his settling down in Russia, exchanging letters with Goldbach, getting almost accidentally acquainted with the works of Fermat, then, much later in life, starting a correspondence with Lagrange on number-theory and elliptic integrals? Should we not be pleased that, through his letters, such a man has come to belong to our close acquaintance?

So far, however, I have merely scratched the surface of my theme. Leibniz recommended the study of "illustrious examples", not just for the sake of esthetic enjoyment, but chiefly so that "the art of discovery be promoted". At this point one has to make clear the distinction, in scientific matters, between tactics and strategy.

By tactics I understand the day-to-day handling of the tools at the disposal of the scientist or scholar at a given moment; this is best learnt from a competent teacher and the study of contemporary work. For the mathematician it may include the use of differential calculus at one time, of homological algebra at another. For the historian of mathematics, tactics have much in common with those of the genera- historian. He must seek his documentation at its source, or as close to it as practicable; second-hand information is of small value. In some areas of research one must learn to hunt for and read manuscripts; in others one may be content with published texts, but then the question of their reliability or lack of it must always be kept in mind. An indispensable requirement is an adequate knowledge of the language of the sources; it is a basic and sound principle of all historical research that a translation can never replace the original when the latter is available. Luckily the history of Western mathematics after the XV$^{\text{th}}$ century seldom requires any linguistic knowledge besides Latin and the modern Western European languages; for many purposes French, German and sometimes English might even be enough.

In contrast with this, strategy means the art of recognizing the main problems, attacking them at their weak points, setting up future lines of advance. Mathematical strategy is concerned with long-range objectives; it requires a deep understanding of broad trends and of the evolution of ideas over long periods. This is almost indistinguishable from what Gustav Eneström used to describe as the main object of mathematical history, viz., "the mathematical ideas, considered historically",[2] or, as Paul Tannery put it, "the filiation of ideas and the concatenation of discoveries".[3] There we have the core of the discipline we are discussing, and it is a fortunate fact that the aspect towards which, according to Eneström and Tannery, the mathematical historian has chiefly to direct his attention is also the one of greatest value for any mathematician who wants to look beyond the everyday practice of his craft.

The conclusion we have reached has little substance, to be sure, unless we agree about what is and what is not a mathematical idea. As to this, the mathematician is hardly inclined to consult outsiders. In the words of Housman (when asked to define poetry), he may not be able to define what is a mathematical idea, but he likes to think that when he smells one he knows it. He is not likely to see one, for instance, in Aristotle's speculations about the infinite, nor in those of a number of medieval thinkers on the same subject, even though some of them were rather more interested in mathematics than Aristotle ever was; the infinite became a mathematical idea after Cantor defined equipotent sets and proved some theorems about them. The views of Greek philosophers about the infinite may be of great interest as such; but are we really to believe that they had great influence on the work of Greek mathematicians? Because of them, we are told, Euclid had to refrain from saying that there are infinitely many primes, and had to express that fact differently. How is it then that, a few pages later, he stated that "there exist infinitely many lines"[4] incommensurable with a given one? Some universities have established chairs for "the history and philosophy of mathematics"; it is hard for me to imagine what those two subjects can have in common.

Not so clearcut is the question where "common notions" (to use Euclid's phrase) end and where mathematics begins. The formula for the sum of the first n integers, closely related as it is to the "Pythagorean" concept of triangular numbers, surely

---

[2] *Die mathematischen Ideen in historischer Behandlung* (Bibl. Math. **2** (1901), p.1).

[3] *La filiation des idées et l'enchaînement des découvertes* (P. Tannery, *Oeuvres*, vol. X, p. 166).

[4] *Ὑπάρχουσιν εὐθεῖαι πλήθει ἄπειροι* (Bk. X, Def. 3).

deserves to be called a mathematical idea; but what should we say about elementary commercial arithmetic, as it appears in ever so many textbooks from antiquity down to Euler's potboiler on the same subject? The concept of a regular icosahedron belongs distinctly to mathematics; shall we say the same about the concept of a cube, that of a rectangle, or that of a circle (which is perhaps not to be separated from the invention of the wheel)? Here we have a twilight zone between cultural and mathematical history; it does not matter much where one draws the borderline. All the mathematician can say is that his interest tends to falter, the nearer he comes to crossing it.

However that may be, once we have agreed that mathematical ideas are the true object of mathematical history, some useful consequences can be drawn; one has been formulated by Tannery as follows (*loc. cit.*, (footnote 3), p. 164). There is no doubt at all, he says, that a scientist can possess or acquire all the qualities needed to do excellent work on the history of his science; the greater his talent as a scientist, the better his historical work is likely to be. As examples, he mentions Chasles for geometry; also Laplace for astronomy, Berthelot for chemistry; perhaps he was also thinking of his friend Zeuthen. He might well have quoted Jacobi, if Jacobi had lived to publish his historical work.[5]

But examples are hardly necessary. Indeed it is obvious that the ability to recognize mathematical ideas in obscure or inchoate form, and to trace them under the many disguises which they are apt to assume before coming out in full daylight, is most likely to be coupled with a better than average mathematical talent. More than that, it is an essential component of such talent, since in large part the art of discovery consists in getting a firm grasp on the vague ideas which are "in the air", some of them flying all around us, some (to quote Plato) floating around in our own minds.

How much mathematical knowledge should one possess in order to deal with mathematical history? According to some, little more is required than what was known to the authors one plans to write about;[6] some go so far as to say that the less one knows, the better one is prepared to read those authors with an open mind and avoid anachronisms. Actually the opposite is true. An understanding in depth of the mathematics of any given period is hardly ever to be achieved without knowledge extending far beyond its ostensible subject-matter. More often than not, what

---

[5] Jacobi, as a student, had hesitated between classical philology and mathematics; he always retained a deep interest in Greek mathematics and mathematical history; extracts from his writings on this subject have been published by Koenigsberger in his biography of Jacobi (incidentally, a good model for a mathematically oriented biography of a great mathematician): see L. Koenigsberger, *Carl Gustav Jacob Jacobi*, Teubner, 1904. pp. 385—395 and 413—414.

[6] Such seems to have been Loria's view: "Per comprendere e giudicare gli scritti appartenenti alle età passate, basta di essere esperto in quelle parti delle scienze che trattano dei numeri e delle figure e che si considerano attualmente come parte della cultura generale dell'uomo civile" (G. Loria, *Guida allo Studio della Storia delle Matematiche*, U. Hoepli, Milano, 1946, p. 271).

makes it interesting is precisely the early occurrence of concepts and methods destined to emerge only later into the conscious mind of mathematicians; the historian's task is to disengage them and trace their influence or lack of influence on subsequent developments. Anachronism consists in attributing to an author such conscious knowledge as he never possessed; there is a vast difference between recognizing Archimedes as a forerunner of integral and differential calculus, whose influence on the founders of the calculus can hardly be overestimated, and fancying to see in him, as has sometimes been done, an early practitioner of the calculus. On the other hand, there is no anachronism in seeing in Desargues the founder of the projective geometry of conic sections; but the historian has to point out that his work, and Pascal's, soon fell into the deepest oblivion, from which it could only be rescued after Poncelet and Chasles had independently rediscovered the whole subject.

Similarly, consider the following assertion: logarithms establish an isomorphism between the multiplicative semigroup of numbers between 0 and 1 and the additive semigroup of positive real numbers. This could have made no sense until comparatively recently. If, however, we leave the words aside and look at the facts behind that statement, there is no doubt that they were well understood by Neper when he invented logarithms, except that his concept of real numbers was not as clear as ours; this is why he had to appeal to kinematic concepts in order to clarify his meaning, just as Archimedes had done, for rather similar reasons, in his definition of the spiral.[7] Let us go further back; the fact that the theory of the ratios of magnitudes and of the ratios of integers, as developed by Euclid in Books V and VII of his *Elements,* is to be regarded as an early chapter of group-theory is put beyond doubt by the phrase "double ratio" used by him for what we call the square of a ratio. Historically it is quite plausible that musical theory supplied the original motivation for the Greek theory of the group of ratios of integers, in sharp contrast with the purely additive treatment of fractions in Egypt; if so, we have there an early example of the mutual interaction between pure and applied mathematics. Anyway, it is impossible for us to analyze properly the contents of Books V and VII of Euclid without the concept of group and even that of groups with operators, since the ratios of magnitudes are treated as a multiplicative group operating on the additive group of the magnitudes themselves.[8] Once that point of view is adopted, those books of Euclid lose their mysterious character, and it becomes easy to follow the line which leads directly from them to Oresme and Chuquet, then to Neper and logarithms (cf. NB, pp. 154—159 and 167—168). In doing so, we are of course not attributing the group concept to any of these authors; no more should

[7] cf. N. Bourbaki, *Eléments d'histoire des mathématiques*, Hermann, 1960, pp. 167—168 and 174; that collection of historical essays, extracted from the same author's *Eléments de mathématique* under a rather misleading title, will be quoted henceforth as NB.

[8] Whether or not Euclid believed the group of the ratios of magnitudes to be independent of the kind of magnitudes under study is still a moot point; cf. O. Becker, Quellen u. Studien **2** (1933), 369—387.

one attribute it to Lagrange, even when he was doing what we now call Galois theory. On the other hand, while Gauss had not the word, he certainly had the clear concept of a finite commutative group, and had been well prepared for it by his study of Euler's number-theory.

Let me quote a few more examples. Fermat's statements indicate that he was in possession of the theory of the quadratic forms X^2+nY^2 for $n=1,2,3$, using proofs by "infinite descent". He did not record those proofs; but eventually Euler developed that theory, also using infinite descent, so that we may assume that Fermat's proofs did not differ much from Euler's. Why does infinite descent succeed in those cases? This is easily explained by the historian who knows that the corresponding quadratic fields have an Euclidean algorithm; the latter, transcribed into the language and notations of Fermat and Euler, gives precisely their proofs by infinite descent, just as Hurwitz' proof for the arithmetic of quaternions, similarly transcribed, gives Euler's proof (which possibly was also Fermat's) for the representation of integers by sums of 4 squares.

Take again Leibniz' notation $\int y\,dx$ in the calculus. He insisted repeatedly on its invariant character, first in his correspondence with Tschirnhaus (who showed no understanding for it), then in the *Acta Eruditorum* of 1686; he even had a word for it ("*universalitas*"). Historians have hotly disputed when, or whether, Leibniz discovered the comparatively less important result which, in some textbooks, goes by the name of "the fundamental theorem of the calculus". But the importance of Leibniz' discovery of the invariance of the notation $y\,dx$ could hardly have been properly appreciated before Élie Cartan introduced the calculus of exterior differential forms and showed the invariance of the notation $y\,dx_1\cdots dx_m$, not only under changes of the independent variables (or of local coordinates), but even under "pull-back".[9]

Consider now the debate which arose between Descartes and Fermat about tangents (cf. NB, p. 192). Descartes, having decided, once and for all, that only algebraic curves were a fit subject for geometers, invented a method for finding their tangents, based upon the idea that a variable curve, intersecting a given one C at a point P, becomes tangent to C at P when the equation for their intersections acquires a double root corresponding to P. Soon Fermat, having found the tangent to the cycloid by an infinitesimal method, challenged Descartes to do the same by his own method. Of course he could not do that; being the man he was, he found the answer (*Oeuvres,* II, p. 308), gave a proof for it ("quite short and quite simple", by using the instantaneous center of rotation which he invented for the occasion) and added that he could have supplied another proof "more to his taste and more geometrical" which he omitted "to save himself the trouble of writing it out"; anyway, he said, "such lines are mechanical" and he had excluded them from geometry. This, of course, was the point that Fermat was trying to make; he knew,

[9] Cf. NB, p. 208, and A. Weil, Bull. Amer. Math. Soc. **81** (1975), 683.

as well as Descartes, what an algebraic curve was, but to restrict geometry to those curves was quite alien to his way of thinking and to that of most geometers in the XVIIth century.

Gaining insight into a great mathematician's character and into his weaknesses is an innocent pleasure that even serious historians need not deny themselves. But what else can one conclude from that episode? Very little, as long as the distinction between differential and algebraic geometry has not been clarified. Fermat's method belonged to the former; it depended upon the first terms of a local power-series expansion; it provided the starting point for all subsequent developments in differential geometry and differential calculus. On the other hand, Descartes' method belongs to algebraic geometry, but, being restricted to it, it remained a curiosity until the need arose for methods valid over quite arbitrary groundfields. Thus the point at issue could not be and was not properly perceived until abstract algebraic geometry gave it its full meaning.

There is still another reason why the craft of mathematical history can best be practised by those of us who are or have been active mathematicians or at least who are in close contact with active mathematicians; there are various types of misunderstandings of not infrequent occurrence from which our own experience can help preserve us. We know only too well, for instance, that one should not invariably assume a mathematician to be fully aware of the work of his predecessors, even when he includes it among his references; which one of us has read all the books he has listed in the bibliographies of his own writings? We know that mathematicians are seldom influenced in their work by philosophical considerations, even when they profess to take them seriously; we know that they have their own way of dealing with foundational matters by an alternation between possibly reckless disregard and the most painful critical attention. Above all, we have learnt the difference between original thinking and the kind of routine reasoning which a mathematician often feels he has to spin out for the record in order to satisfy his peers, or perhaps only to satisfy himself. A tediously laborious proof may be a sign that the writer has been less than felicitous in expressing himself; but more often than not, as we know, it indicates that he has been laboring under limitations which prevented him from translating directly into words or formulas some very simple ideas. Innumerable instances can be given of this, ranging from Greek geometry (which perhaps was at last suffocated by such limitations) down to the so-called epsilontic and down to Nicolas Bourbaki, who even once considered using a special sign in the margin to warn the reader about proofs of that kind. One important task of the serious historian of mathematics, and sometimes one of the hardest, is precisely to sift such routine from what is truly new in the work of the great mathematicians of the past.

Of course mathematical talent and mathematical experience are not enough for qualifying as a mathematical historian. To quote Tannery again (*loc. cit.* (footnote 3), p. 165), "what is needed above all is a taste for history; one has to develop a historical sense". In other words, a quality of intellectual sympathy is required, embracing past epochs as well as our own. Even quite distinguished mathematicians may lack it altogether; each one of us could perhaps name a few who resolutely refuse to be acquainted with any work other than their own. It is also necessary not to yield to the temptation (a natural one to the mathematician) of concentrating upon the greatest among past mathematicians and neglecting work of only subsidiary value. Even from the point of view of esthetic enjoyment one stands to lose a great deal by such an attitude, as every art-lover knows; historically it can be fatal, since genius seldom thrives in the absence of a suitable environment, and some familiarity with the latter is an essential prerequisite for a proper understanding and appreciation of the former. Even the textbooks in use at every stage of mathematical development should be carefully examined in order to find out, whenever possible, what was and what was not common knowledge at a given time.

Notations, too, have their value. Even when they are seemingly of no importance, they may provide useful pointers for the historian; for instance, when he finds that for many years, and even now, the letter K has been used to denote fields, and German letters to denote ideals, it is part of his task to explain why. On the other hand, it has often happened that notations have been inseparable from major theoretical advances. Such was the case with the slow development of the algebraic notation, finally brought to completion at the hands of Viète and Descartes. Such was the case again with the highly individual creation of the notations for the calculus by Leibniz (perhaps the greatest master of symbolic language that ever was); as we have seen, they embodied Leibniz' discoveries so successfully that later historians, deceived by the simplicity of the notation, have failed to notice some of the discoveries.

Thus the historian has his own tasks, even though they overlap those of the mathematician and may at times coincide with them. Thus, in the XVII$^{\text{th}}$ century, it happened that some of the best mathematicians, in the absence of immediate predecessors in any field of mathematics except algebra, had much work to do which in our view would fall to the lot of the historian, editing, publishing, reconstructing the work of the Greeks, of Archimedes, Apollonios, Pappos, Diophantos. Even now the historian and the mathematician will not infrequently find themselves on common ground when studying the production of the XIX$^{\text{th}}$ and XX$^{\text{th}}$ centuries, not to mention anything of more ancient vintage. From my own experience I can testify about the value of suggestions found in Gauss and in Eisenstein. Kummer's congruences for Bernoulli numbers, after being regarded as little more than a curiosity for many years, have found a new life in the theory of p-adic L-functions, and Fermat's ideas on the use of the infinite descent in the study of Diophantine equations of genus 1 have proved their worth in contemporary work on the same subject.

What, then, separates the historian from the mathematician when both are studying the work of the past? Partly, no doubt, their techniques, or, as I proposed to put it, their tactics; but chiefly, perhaps, their attitudes and motivations. The historian tends to direct his attention to a more distant past and to a greater variety of cultures; in such studies, the mathematician may find little profit other than the esthetic satisfaction to be derived from them and the pleasures of vicarious discovery. The mathematician tends to do his reading with a purpose, or at least with the hope that some fruitful suggestion will emerge from it. Here we may quote the words of Jacobi in his younger days about a book he had just been reading: "Until now, he said, whenever I have studied a work of some value, it has stimulated me to original thoughts; this time I have come out quite empty-handed".[10] As noted by Dirichlet, from whom I have borrowed this quotation, it is ironical that the book in question was no other than Legendre's *Exercices de calcul intégral,* containing work on elliptic integrals which soon was to provide the inspiration for Jacobi's greatest discoveries; but those words are typical. The mathematician does his reading mostly in order to be stimulated to original (or, I may add, sometimes not so original) thoughts; there is no unfairness, I think, in saying that his purpose is more directly utilitarian than the historian's. Nevertheless, the essential business of both is to deal with mathematical ideas, those of the past, those of the present, and, when they can, those of the future. Both can find invaluable training and enlightenment in each other's work. Thus my original question "Why mathematical history?" finally reduces itself to the question "Why mathematics?", which fortunately I do not feel called upon to answer.

[10] "Wenn ich sonst ein bedeutendes Werk studiert habe, hat es mich immer zu eignen Gedanken angeregt... Diesmal bin ich ganz leer ausgegangen und nicht zum geringsten Einfall inspiriert worden". (Dirichlet, *Werke*, Bd. II, S. 231).

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Role of Partial Differential Equations in Differential Geometry

Shing-Tung Yau

In the study of geometric objects that arise naturally, the main tools are either groups or equations. In the first case, powerful algebraic methods are available and enable one to solve many deep problems. While algebraic methods are still important in the second case, analytic methods play a dominant role, especially when the defining equations are transcendental. Indeed, even in the situation where the geometric object is homogeneous or algebraic, analytic methods often lead to important contributions. In this talk, we shall discuss a class of problems in differential geometry and the analytic methods that are involved in solving such problems.

One of the main purposes of differential geometry is to understand how a surface (or a generalization of it) is curved, either intrinsically or extrinsically. Naturally, the problems that are involved in studying such an object cannot be linear. Since curvature is defined by differentiating certain quantities, the equations that arise are nonlinear differential equations. In studying curved space, one of the most important tools is the space of tangent vectors to the curved space. In the language of partial differential equations, the main tool to study nonlinear equations is the use of the linearized operators. Hence, even when we are facing nonlinear objects, the theory of linear operators is unavoidable. Needless to say, we are then left with the difficult problem of how precisely a linear operator approximates a nonlinear operator.

To illustrate the situation, we mention five important differential operators in differential geometry. The first one, which is probably the most important one, is the Laplace–Beltrami operator. If the metric tensor is given by $\sum_{i,j} g_{ij}\, dx^i \otimes dx^j$, then the operator is given by

$$L(\varphi) = \frac{1}{\sqrt{g}} \sum_{i,j} \frac{\partial}{\partial x^i}\left[g^{ij}\sqrt{g}\,\frac{\partial \varphi}{\partial x^j}\right]$$

where $g = \det(g_{ij})$ and (g^{ij}) is the inverse matrix of (g_{ij}).

The second one is the minimal surface operator and is given by

$$L(\varphi) = \sum_i \frac{\partial}{\partial x^i}\left[(1+|\nabla\varphi|^2)^{-1/2}\frac{\partial \varphi}{\partial x^i}\right]$$

where $|\nabla\varphi|^2 = \sum_i (\partial\varphi/\partial x^i)^2$.

The third one is the Monge–Ampère operator

$$L(\varphi) = \det\left(\frac{\partial^2 \varphi}{\partial x^i\,\partial x^j}\right).$$

The fourth one is the complex Monge–Ampère operator

$$L(\varphi) = \det\left(\frac{\partial^2 \varphi}{\partial z^i\,\partial \bar{z}^j}\right).$$

The fifth one is the Einstein field equation which is a nonlinear hyperbolic system. If $\sum_{i,j} g_{ij}\,dx^i\,dx^j$ is the Lorentz metric to be determined, then the operator involved in the Einstein field equation is

$$L(g_{ij}) = R_{ij} - (R/2)\,g_{ij}$$

where R_{ij} is the Ricci tensor and R is the scalar curvature of the Lorentz metric.

Both the Laplace–Beltrami operator and the minimal surface operator are elliptic. The (real) Monge–Ampère operator is elliptic only at those functions φ where φ is strictly convex and the complex Monge–Ampère operator is elliptic only at those functions φ where φ is strictly plurisubharmonic. All the above operators except the Laplace–Beltrami operator are nonlinear. However, a suitable interpretation shows that the linearized operators of the minimal surface operator and the Monge–Ampère operators are the Laplace–Beltrami operators of certain metrics.

To see how these operators arose in differential geometry, we will discuss one important problem here. Roughly speaking, this problem is to ask how a space is curved globally. In a little more precise form, it can be stated as follows. Given a manifold M, find a necessary and sufficient condition for M to admit a metric with certain curvature properties.

To set up the terminology, we remind the reader of some definitions. From the curvature tensor, one can extract the following quantities. Given a point in the manifold and a two dimensional plane in the tangent space at that point, we can form the *sectional curvature* of the manifold at this plane. Given a point and a tangent at a point, we can form the *Ricci curvature* in this tangent direction by averaging all the sectional curvatures of the two dimensional tangent planes that contain this tangent. Given a point, we can form the *scalar curvature* at this point by simply averaging all the sectional curvatures at this point. It is clear from these definitions that the sectional curvatures give much more information than the others. For example, as the sectional curvature tells us how the manifold curves in every two plane, it gives good control of the behavior of the geodesics of the manifold. The latter depends on the theory of ordinary differential equations. However, in the other cases, the information about geodesics is much less and the theory of partial differential equations must be involved. Thus in this talk we will concentrate only on the scalar curvature and the Ricci curvature. We begin by discussing the general method of obtaining integrability conditions for the existence of metrics with certain curvature conditions.

1. Integrability conditions. The problem of finding complete integrability conditions for the global existence of metrics with certain curvature conditions is rather difficult. However, for a two dimensional surface, this has a satisfactory answer, thanks to the Gauss–Bonnet theorem for compact surfaces and to the Cohn–Vossen inequality for the complete open surfaces. (The recent works of Kazdan–Warner **[32]** gave more precise information on the behavior of the curvature function in two dimensional geometry.)

In higher dimension, the situation is much more complicated partly because the curvature is a tensor and partly because the link between topological invariants and geometric invariants is rather weak at this stage. We list here the major methods that were used to find integrability conditions.

1. Chern's theory of representing Euler class, Pontryagin classes and Chern classes by curvature forms gives the most basic integrability conditions for general manifolds. The celebrated theorem of Atiyah–Singer can be considered as a glorified generalization. Some of their applications will be explained later.

2. Bochner's method of proving vanishing theorems via Hodge theory will remain to be important for a long time. It led to the Kodaira vanishing theorem, L^2 methods in several complex variables, etc.

3. The variational method has been one of the most classical and most important methods in differential geometry. It includes variation of curves, surfaces, maps, etc.

Naturally, these do not exhaust all the methods. However, for all the results that we are going to discuss, they are obtained by suitable combination of the above three methods.

2. Scalar curvature. The simplest problem concerning the scalar curvature is to find those manifolds which admit a complete metric whose scalar curvature has the same sign.

A long time ago, Yamabe **[52]** was interested in deforming a metric conformally to one with constant scalar curvature. The equation that is involved in such a process has the following form

$$\Delta u = \frac{(n-2)}{4(n-1)} Ru - \frac{(n-2)}{4(n-1)} \bar{R} u^{(n+2)/(n-2)}$$

where n is the dimension of the manifold, R and $\bar{R}$ are scalar curvatures of the undeformed and deformed metrics respectively.

As was pointed out by Trudinger [49], Yamabe's method does not seem to work. After Trudinger, there were works by Aubin, Berger, Eliason, Kazdan–Warner, Nirenberg, Moser, etc. An easy consequence of these results is that every compact manifold with dimension greater than two admits a metric with negative scalar curvature. Greene and Wu [27], using another method, proved that every noncompact manifold admits a complete metric with negative scalar curvature. Hence we conclude that in higher dimensions, existence of complete metrics with negative scalar curvature poses no topological restriction on the manifold.

However, complete metrics with nonnegative scalar curvature do give topological information. The first result in this direction is due to Lichnerowicz [35] who proved that for a compact spin manifold with positive scalar curvature, there are no harmonic spinors. Applying the Atiyah–Singer index theorem, the Lichnerowicz vanishing theorem then proves that for a compact spin manifold with positive scalar curvature, the $\hat{A}$-genus is zero. By pursuing these arguments, Hitchin [30] observed that the (mod 2) KO-theory invariant introduced by Milnor is also zero for a compact spin manifold with positive scalar curvature. In particular, any exotic sphere which does not bound a spin manifold admits no metric with positive scalar curvature.

While mathematicians were working on problems related to scalar curvature, it turned out that physicists, from other points of view, were also interested in similar problems.

Let us describe this problem in general relativity in geometric terms. Suppose we are given a Lorentzian metric on a four dimensional manifold. Then under a fairly general condition, one expects to prove the existence of a maximal space-like hypersurface, i.e., a hypersurface which is locally stable under the deformation of the induced area. Usually, we assume that the Lorentzian metric satisfies the weak energy condition so that, by the Gauss curvature equation, the scalar curvature of the above mentioned maximal space-like hypersurface has non-negative scalar curvature.

Since the maximal space-like hypersurface is three dimensional, we are dealing with a three dimensional manifold with nonnegative scalar curvature. On the other hand, it is well known that three dimensional manifolds are parallelizable. Hence, most of the known topological invariants in higher dimension vanish and the consequences derived from the Lichnerowicz theorem and the Atiyah–Singer index theorem provided no information. On the other hand, the above mentioned problem in general relativity does provide us some guideline. It roughly states [26] that, for an isolated physical system, nonnegativity of local mass density implies the nonnegativity of total mass. In mathematical terms, it may be described as follows. Let M be a three dimensional manifold with non-negative scalar curvature. (It is the maximal space-like hypersurface mentioned above.) Suppose M is diffeomorphic to R^3 (the situation described below can be generalized to other three dimensional manifolds) such that the metric has the form $(1+m/2r)^4 ds_0^2+O(1/r^2)$ where ds_0^2 is the standard euclidean metric on R^3, r is the distance from the origin and $O(1/r^2)$ is a tensor which vanishes along with its first two derivatives like $1/r^2$ when r tends to infinity. The number m is called the total mass of the manifold M. The positive mass conjecture in general relativity says that m is nonnegative and is zero iff the metric is euclidean. A special case of the conjecture says that if we have a metric of nonnegative scalar curvature defined on R^3 which is euclidean outside a compact set, then the metric is euclidean everywhere. This last statement has direct bearing to the questions that geometers are considering.

This positive mass conjecture was proved by R. Schoen and myself recently. (The best previous work on the conjecture was a local result due to Choquet–Bruhat and Marsden [**21**]). Our motivation and method comes out from an attempt to understand the topology of three dimensional manifolds with nonnegative scalar curvature. Because of the nature of the topology of three dimensional manifolds, it is important to understand the fundamental group. In this regard, we proved that if the fundamental group of the three dimensional manifold with nonnegative scalar curvature contains a subgroup which is isomorphic to the fundamental group of a compact surface with genus ≥ 1, then the metric is a flat metric. The method of proving this theorem and the above mentioned mass conjecture comes out from the study of the minimal surface equation mentioned in the beginning. It describes a surface in M which locally has minimal area compared with nearby surfaces. The study of such objects has been one of the most important branches in nonlinear elliptic partial differential equations and calculus of variations. (It motivated a new important subject—geometric measure theory—about which Almgren will talk during this Congress.) The reason that it is useful in studying the topology of the manifold is that it tells us how the internal geometry of the manifold behaves. In two dimensions, we can control the topology of the minimal surface, thanks to the work of C. B. Morrey. In higher dimensions, this remains to be studied.

It would be nice to give a criterion for a manifold to admit a metric with positive scalar curvature. However, we do not have a good existence theorem yet. In this regard, we may mention a theorem of B. Lawson and the author [34]. We proved that if a manifold admits a differentiable nonabelian connected compact Lie group action, then the manifold admits a complete metric with positive scalar curvature. (Combining with the above mentioned theorem of Hitchin, we showed that exotic spheres do not admit effective $SU(2)$ action if they do not bound a spin manifold. This gives a theorem in topology and illustrates how curvature can be used to deal with topological problems.) As a generalization of the above work on three dimensional manifold, we mention the following problem. If a compact manifold with nonnegative scalar curvature is covered by the euclidean space topologically, is it a flat manifold?[1]

3. Ricci curvature. As in the case of scalar curvature, the simplest problem concerning the Ricci curvature is to find those manifolds which admit a complete metric whose Ricci curvature has the same sign. Since the Ricci curvature is given by a tensor and the integrability condition is stronger, the problem of existence is considerably harder. The known integrability conditions are not yet complete and we shall only mention a few here.

First of all, Bonnet's theorem tells us that for a compact manifold with positive sectional curvature, the fundamental group must be finite and this was later generalized by Myers [41] for positive Ricci curvature and by Cheeger and Gromoll [14] to the case where we only assume the Ricci curvature to be nonnegative. For a noncompact complete manifold with nonnegative Ricci curvature, there are also conditions on the fundamental group due to Milnor [36], Wolf [51], Schoen and Yau [46]. It seems that a complete manifold with positive Ricci curvature should have a finite fundamental group. But this has never been proved. Metrics with negative Ricci curvature seem to be even harder to understand. For example, in higher dimension, we do not even know whether spheres admit such a metric or not. Only recently, the author [55] was able to produce such a metric on a compact simply-connected manifold. It would be interesting to find some integrability conditions for the existence. It seems possible that for a manifold to admit a metric with negative Ricci curvature, it should admit no effective differentiable nonabelian connected compact Lie group action. It would also be interesting to see whether a compact manifold can admit both a metric with nonnegative scalar curvature and a metric with negative Ricci curvature.

Because of the interest in general relativity, metrics with constant Ricci curvature are of particular importance. For a long time, the only known examples were those manifolds that are acted transitively upon by a compact Lie group. The first necessary condition for the existence was found by M. Berger [6] who proved that for four dimensional Einstein manifolds, i.e., manifolds with constant Ricci curvature, the Euler number must be positive unless they are flat. This inequality of Berger was later generalized by Hitchin [31]. In all these theorems, Chern's representation of the topological invariants by curvature plays a very important role.

For quite a long time, there was no example of nonhomogeneous Einstein manifolds. In particular, it was not known whether there exists a non-flat compact Riemannian manifold with zero Ricci curvature. (This attracted people's attention because of its analogue with the situation in general relativity.) Partly motivated by this

[1] After the Congress, R. Schoen and the author were able to generalize our work on three dimensional manifolds to higher dimensional manifolds. This was also achieved by Gromov and Lawson about the same time. Our works also indicate the possibility of classifying compact simply connected manifolds with positive scalar curvature.

question, Calabi [9] proposed a way to study the Ricci tensor for some special class of manifolds. He observed that in the case of Kähler manifolds, the expression for the Ricci tensor is particularly simple. This observation was based on Chern's representation of the first Chern's class [17] by the curvature form and can be described as follows. Let $\sum_{i,j} g_{i\bar{j}}\, dz^i \otimes d\bar{z}^j$ be a Kähler metric defined on a compact complex manifold. Then the (1,1) form

$$\frac{\sqrt{-1}}{2\pi} \sum_{r,s} \frac{\partial^2}{\partial z^r\, \partial \bar{z}^s} [\log \det (g_{i\bar{j}})]\, dz^r \wedge d\bar{z}^s$$

is closed, globally defined on the manifold and represents the first Chern class. According to Chern [17], this (1,1) form is also the Ricci form of the Kähler metric. Hence for a (1,1) form to be the Ricci form of some Kähler metric, it must be closed and represents the first Chern class. What Calabi asked was whether this is the only integrability condition. This question stimulated a lot of interest partly because it could give a complete understanding of the Ricci tensor of a Kähler manifold and partly because it would create a lot of examples of compact manifolds with zero Ricci curvature. For example, the $K-3$ surface is a compact simply connected manifold with zero first Chern class. Calabi's conjecture immediately shows the existence of a Ricci flat metric on the $K-3$ surface. (The simple-connectivity of the $K-3$ surface guarantees that it does not admit any flat metric.) The equation that is needed to solve Calabi's conjecture has the following form

$$(*) \qquad \det\left(g_{i\bar{j}} + \frac{\partial^2 \varphi}{\partial z^i \partial \bar{z}^j}\right) = e^F \det (g_{i\bar{j}})$$

where φ is the unknown function and F is a smooth function so that $\int_M e^F$ is the volume of M.

Equation $(*)$ is similar to the real Monge–Ampère operator and can be considered as the complex Monge–Ampère equation. In order to make $(*)$ to be elliptic, we have to look for functions φ so that $(g_{i\bar{j}} + \partial^2 \varphi / \partial z^i \partial \bar{z}^j)$ is a positive definite metric.

In order to understand the equation $(*)$, Calabi [11] studied the equation $\det (\partial^2 \varphi / \partial x^i \partial x^j) = 1$ where φ is required to be convex. He tried to prove that if φ is defined over the entire euclidean space, then it is a quadratic polynomial. He generalized Jörgen's theorem [59] from two dimension to dimension ≤ 5. The important ingredient in his paper is the introduction of the quantity $S = \sum \varphi^{ir} \varphi^{js} \varphi^{kt} \varphi_{ijk} \varphi_{rst}$ where (φ^{ij}) is the inverse matrix of (φ_{ij}) and φ_{ijk} is the third derivative of φ with respect to x^i, x^j and x^k. This quantity comes up naturally from affine geometry, a geometry where we want to study quantities invariant under the special linear group. Affine geometry is very natural in dealing with the Monge–Ampère equation because the Monge–Ampère operator is clearly invariant under the special linear group. Indeed, the graph defined by the solution of the equation $\det (\partial^2 \varphi / \partial x^i \partial x^j) = 1$ has a nice affine geometric meaning. It is called the improper affine sphere. The important contribution of Calabi is that he found a nice formula when the linearized operator of $M(\varphi)=\det(\partial^2\varphi/\partial x^i\partial x^j)$ operates on the above quantity S. His formula enables one to estimate, in the interior of the domain, the third derivatives of the solution of the equation $\det(\partial^2\varphi/\partial x^i\partial x^j)=F(x,\varphi)$ assuming that we know the lower order estimates of φ. It turns out that a complex analogue of Calabi's third order quantity exists and that a nice formula (as was shown by Nirenberg) still holds.

In 1971, Pogorelov [45] was able to push Calabi's method to prove that in general, any convex entire solution of the equation $\det(\partial^2\varphi/\partial x^i\partial x^j)=1$ is a quadratic polynomial. One of the main ingredients of Pogorelov was his interior estimate of the second derivatives of the equation $\det(\partial^2\varphi/\partial x^i\partial x^j)=F(x)$. Besides the interior estimate, Pogorelov used a lot of convex geometry to prove the completeness of the affine metric which was the major point left in Calabi's approach. Later, Calabi, Cheng, Nirenberg and the author were able to prove the completeness of a large class of affine metrics. These include also the hyperbolic affine sphere where the equation is given by $\det(\partial^2\varphi/\partial x^i\partial x^j)=(-1/\varphi)^{n+2}$. This last method does not depend on convex geometry. It has direct influence on our later work mentioned below.

Coming back to the equation $(*)$, one notices that Calabi proved that if F is close enough to zero, $(*)$ has a unique solution. Assuming a curvature condition on the Kähler manifold, Aubin [4] indicated a variational method to prove the existence of solution to $(*)$. (It was conjectured, for example, that such a curvature condition would imply that the manifold is the complex projective space. This is not enough for our later applications in geometry. Furthermore, for the Monge–Ampère equation, variational methods are still rather difficult.) In 1976, the author [55], [56] was able to use the continuity method to prove that $(*)$ has a unique solution without any additional assumption. As usual, the basic steps in the proof are giving the a priori estimates of $(*)$ up to the third derivatives. The third order estimate is essentially a consequence of the fundamental contributions of Calabi. The second order estimate is motivated by Pogorelov's work in [45]. However, both these estimates depend on the estimate of $\sup|\varphi|$. This was not known for a long time and was the major difficulty in solving $(*)$. In case the right hand of $(*)$ has the form $e^{\varphi+F}\det(g_{i\bar{j}})$, an estimate of $\sup|\varphi|$ follows trivially from the maximum principle. In [56], the estimate of $\sup|\varphi|$ depends on a delicate and technically very complicated interplay of the maximum principle and the integration method. Later there was a slight simplification of this estimate due to Kazdan [60] and Bourguignon. As a consequence of the solution of $(*)$ and its proof, one can deduce the existence of a (canonical) Kähler Einstein metric on a compact Kähler manifold with zero or negative first Chern class. (In the special case where the right-hand side of $(*)$ is $e^{\varphi+F}\det(g_{i\bar{j}})$, Aubin [4b] independently announced and sketched a proof which depends on the variational method of his previous paper [4a].)

In a way, the solution of $(*)$, which is commonly known as Calabi's conjecture, gives a complete understanding of the Ricci tensor for a compact Kähler manifold.

However, when one thinks deeper, a lot of problems still have to be done in this direction. One may mention that the solution of Calabi's conjectures gives quite a lot of unexpected application in algebraic geometry [55]. The most interesting one is perhaps the uniqueness of the complex structure of the complex projective plane. This comes out from the canonical metrics that we construct on the algebraic manifolds. These metrics generalize the Poincaré metric of algebraic curves. ·One expects that they will be useful in the moduli problem of algebraic geometry. Indeed, two years ago, the author was able to use the metric above to prove that if M is an algebraic manifold of dimension n whose canonical line bundle is ample, then $(-1)^n 2(n+1)c_2 c_1^{n-2} \geqslant (-1)^n n c_1^n$ and equality holds iff M is covered by the complex ball. (For two dimension algebraic surfaces, there were works of Van de Ven, Bogomolov and Miyaoka. It was Miyaoka who found the above precise inequality independently. However, up to now, their algebraic method cannot be generalized to higher dimension and cannot decide what happens when equality holds.) An easy consequence of the theorem is that there is only one Kähler structure on the complex projective space. The Kähler metric with nonnegative Ricci can also be used to deal with problems related to algebraic manifolds. Up to covering problems and the study of complex torus, one can reduce the study of Kähler manifolds with nonnegative first Chern class to the study of simply-connected Kähler manifold with nonnegative first Chern class. In case the Kähler manifold M has zero first Chern class, then one can prove that for any Kähler class ω in $H^{1,1}(M)$, $\omega^{n-2} \cup c_2(M) \geqslant 0$ and that equality holds only if M is covered by the torus. There are also interesting works of S. Kobayashi [58] who showed how to use the Einstein metric to obtain new vanishing theorems. Bourguignon and Koiso were also able to extend the work of Berger–Ebin [61] to study the deformation of Einstein metrics. They generalized the work of Calabi–Vesentini [63] to Kähler manifolds with negative curvature. Since Einstein metrics have nice curvature properties, it may also be used to strengthen the transcendental method of Griffiths in algebraic geometry.

By pushing more the method that the author used above, Cheng and the author were able to prove the existence of complete Kähler Einstein metrics on many non-compact complex manifolds. For example, if D is a divisor with normal crossings in a compact algebraic manifold M so that $c_1(M) - c_1([D]) < 0$ (see [29]), then we can prove the existence of such a metric on $M \setminus D$. We can also prove the existence of a complete Kähler Einstein metric on any bounded pseudoconvex domain with C^2 boundary in a Stein manifold. It may be interesting to know that it is an easy consequence of the Schwarz lemma given by the author [54] that there is at most one complete Kähler Einstein metric with Ricci curvature $= -1$ on any complex manifold. (This fact was also pointed out by H. Wu.) Therefore, even in the case of noncompact manifolds, complete Kähler Einstein metric is canonical and deserves more investigation.

Concerning the Kähler Einstein metric on a smooth bounded domain Ω in C^n, the equation that we propose to solve has the form $\det(\partial^2 u/\partial z^i \partial \bar{z}^j) = e^{(n+1)u}$

and u is required to tend to infinity on $\partial\Omega$. In order to understand the boundary behavior of the metric near $\partial\Omega$, it suffices to study the boundary behavior of the function $v=e^{-u}$.

The function v satisfies another equation of the Monge–Ampère type. This equation was studied by other people, especially C. Fefferman [24] who studied its relation with the asymptotic behavior of the Bergman kernel. A few years ago, he demonstrated how to find the asymptotic behavior of v assuming its existence. He expanded v in terms of power series expansion of the defining function of Ω. His expansion shows that log terms must occur after the $(n+1)^{\text{th}}$ stage of expansion where $n=\dim\Omega$. His recent deep work on computing the coefficient of the Bergman kernel expansion also shows the importance of this function v. Partly inspired by his work, Cheng and the author were able to demonstrate that the actual solution is $C^{n+3/2-\delta}(\bar{\Omega})$ where $\delta>0$ is an arbitrary small constant. The optimal case should be $n+2-\delta$ and we believe our method will give it after suitable modification. In any case, the information that we obtain is enough to give suitable description of the Kähler Einstein metric near $\partial\Omega$.

Finally, let us come to the question of the existence of complete Kähler metrics with zero Ricci curvature. These metrics have considerable interest in general relativity. There are more conditions for the existence of such metrics and the knowledge of them is far less complete than the previous case. We outline here questions that may lead to future progress.

The first question is: Does every four dimensional compact simply-connected Riemannian manifold with zero Ricci curvature admit a Kähler structure? According to an observation of Hitchin, this is true for the $K-3$ surfaces where the author has constructed Ricci flat Kähler metrics. (In fact, it is true if the compact manifold is a spin manifold with nonzero index.)

The second question is: Can every complete Kähler manifold with zero Ricci curvature be compactified in the complex analytic sense? The author [57] proved that such a manifold does not admit any bounded holomorphic function which gives an indication to support the truth of the statement.

The third question is: Suppose $\tilde{M}$ is one of the compactifications of our manifold M. Does the anticanonical line bundle of $\tilde{M}$ admit a holomorphic section which is zero precisely on $\tilde{M}\setminus M$? If the metric on M "grows only polynomially", then one can indeed prove that the volume form of M gives rise to such a section. This is based on a theorem proved by Calabi and the author [53] that complete noncompact Riemannian manifold with nonnegative Ricci curvature has infinite volume.

In any case, the author is able to prove that, for a compact Kähler manifold $\tilde{M}$, if the anticanonical line bundle of $\tilde{M}$ admits a holomorphic section with nonsingular zero locus, then the completion of the zero locus admits a complete Kähler metric with zero Ricci curvature. The assumption that the zero locus is nonsingular seems to be not necessary. In fact, for many negative holomorphic vector bundles over a compact Kähler Einstein manifold whose Chern classes satisfy some relation, the total spaces admit complete Kähler metric with zero Ricci curvature. (For many special bundles, Calabi also discovered these metrics. For the cotangent bundle of CP^1, it was discovered earlier by Eguchi–Hansen and Hitchin. They even know the metric explicitly.) In these cases, when we compactify the total space, the zero section of the anticanonical line bundle has multiplicity greater than one.

In question three, we request $\tilde{M}\setminus M$ to be a divisor because one can use the growth of the volume to prove that none of the components of $\tilde{M}\setminus M$ is a subvariety with co-dimension greater than one. A theorem of Cheeger–Gromoll **[14]** also shows that the divisor $\tilde{M}\setminus M$ is connected unless M is the product of C and other space. One can prove that the plurigenera of $\tilde{M}$ is zero because the positivity of $P_m(M)$ for some $m>0$ would imply the existence of a non-zero (n,n) form $\tilde{V}=(\sqrt{-1})^n f dz^1 \wedge \ldots \wedge dz^n \wedge d\bar{z}^1 \wedge \ldots \wedge d\bar{z}^n$ where $f \geqslant 0$ and $\log f$ is pluriharmonic at points where $f \neq 0$. If dV is the volume form of M, then $\tilde{V}/dV$ defines a function which is L^1-integrable on M. The condition on $\tilde{V}$ and the fact that M has zero Ricci curvature then imply that $\tilde{V}/dV$ is a constant **[53]**. As M has infinite volume, this constant must be zero. This is a contradiction.

Recall that it is a consequence of the Schwarz lemma proved in **[54]** that M and its universal cover admit no bounded holomorphic function. Specialized to two dimensional complex surfaces, one can then use the classification theory to conclude that $\tilde{M}$ must be rational at least when M is simply connected. In any case, we hope the questions asked above will be answered in the near future. An affirmative answer will be very interesting even for complex surfaces.

4. Applications to partial differential equations. Up to now, it seems that we mainly use methods of partial differential equations to deal with problems in geometry. It turns out that the reverse procedure is also the case. Very often the geometric situation motivates the study of certain quantities in differential equations which turns out to be useful. This is true especially for the minimal surface equation and the Monge–Ampère equation. Indeed, one can use the metric mentioned above to treat the Dirichlet boundary valued problem for the Monge–Ampère equation. The procedure does not depend on the concept of generalized solution. For the real Monge–Ampère equation, there were works of Alexandrov **[1]** and Pogorelov **[43]**. Pogolerov **[43]** sketched a proof for the smoothness of the generalized solution in case the right-hand side is independent of the unknown. (In **[16]** Cheng and the author gave a detailed proof of the smoothness in the general case where the right-hand side depends on the unknown. By a different procedure, we were also able to take care of several essential points overlooked in **[43]**.) For the complex Monge–Ampère equation, the best previously known result was due to Bedford and Taylor **[5]** who proved the existence of C^1 generalized solution. (Using a different method, Gaveau **[62]** was able to obtain a generalized solution similar to that of Bedford and Taylor.)

References

1. A. D. Alexandrov, *Dirichlet problem for the equation* Det $\|Z_{ij}\| = \Phi(Z_1, \ldots, Z_n, Z, X_i, \ldots, X_n)$. I, Vestnik Leningrad Univ. Mat. Meh. Astronom **13** (1958), 5—24. (Russian)

2. A. D. Alexandrov and V. A. Zalgaller, *Intrinsic geometry of surfaces,* Math. Monographs, vol. 15, Amer. Math. Soc., Providence R. I., (1967).

3. M. F. Atiyah and I. M. Singer, *The index of elliptic operators.* III, Ann. of Math., **87** (1968), 546—604.

4. T. Aubin, *Mètriques Riemanniennes et Courbure,* J. Diff. Geom. 4, 1970, 383—424. *Equations du type Monge—Ampère sur les variétés Kähleriennes compactes,* C. R. Acad. Sci. Paris **283** (1976), 119—121.

5. E. Bedford and B. A. Taylor, *The Dirichlet problem for a complex Monge-Ampère equation,* Invent. Math. **37** (1976), 1—44.

6. M. Berger, *Sur les variétés d'Einstein compactes,* Comptes Rendus III^e Reunion Math. Expression Latine, Namur 1965.

7. S. Bochner, *Curvature and Betti numbers,* Ann. of Math. **49** (1948), 379—390.

8. D. Brill and S. Deser, *Variational methods and positive energy in general relativity,* Ann. Physics **50** (1968), 548—570.

9. E. Calabi, *The space of Kähler metrics,* Proc. Internat. Congress Math., Amsterdam, (1954), vol. 2, pp. 206—207.

10. ——— *On Kähler manifolds with vanishing canonical class,* Algebraic Geometry and Topology, A Symposium in Honour of S. Lepschitz, Princeton Univ. Press, Princeton, N. J. (1955), pp. 78—89.

11. ——— *Improper affine hyperspheres and a generalization of a theorem of K. Jörgens,* Michigan Math. J. **5** (1958), 105—126.

12. ——— *Complete hyperspheres* I, Ist. Naz. Alta Mat. Symp. Math **10** (1972), 19—38.

13. ——— *On manifolds with non-negative Ricci curvature* II, Notices Amer. Math. Soc. **22** (1975), A205.

14. J. Cheeger and D. Gromoll, *The splitting theorem for manifolds of non-negative Ricci curvature,* J. Differential Geometry **6** (1971), 119—128.

15. S. Y. Cheng and S. T. Yau, *On the regularity of the solution of the n-dimensional Minkowski problem,* Comm. Pure Appl. Math. **29** (1976), 495—515.

16. ——— *On the regularity of the Monge—Ampère equation* det $(\partial^2 u/\partial x^i \partial x^j) = F(x, u)$, Comm. Pure Appl. Math. **30** (1977), 41—68.

17. S. S. Chern, *Characteristic classes of Hermitian manifold,* Ann. of Math. **47** (1946), 85—121.

18. ——— *On curvature and characteristic classes of a Riemannian manifold,* Abh. Math. Sun. Univ. Hamburg **20** (1956), 117—126.

19. ——— *On holomorphic mappings of Hermitian manifolds of the same dimension,* Proc. Sympos Pure Math., vol. 11, Amer. Math. Soc., Providence, R. I., (1968), pp. 157—170.

20. S. S. Chern and J. Moser, *Real hypersurfaces in complex manifolds,* Acta Math. **133** (1974), 219—271.

21. Y. Choquet-Bruhat and J. Marsden, *Solution of the local mass problem in general relativity,* Comm. Physics, **51** (1976), 283—296.

22. J. Ells and S. Sampson, *Harmonic mappings of Riemannian manifolds,* Amer. J. Math. **86** (1964), 109—160.

23. H. Federer, *Geometric measure theory,* Springer-Verlag, New York, 1969.

24. C. Fefferman, *Monge—Ampère equations, the Bergman kernel, and geometry of pseudoconvex domains,* Ann. of Math., **103** (1976), 395—416.

25. A. Fischer and J. Marsden, *Deformations of the scalar curvature,* Duke Math. J. **42** (1975), 519—547.

26. R. Geroch, *General relativity*, Proc. Sympos. Pure Math., vol. 27, part 2, Amer. Math. Soc., Providence, R. I. (1975), pp. 401—414.

27. R. E. Greene and H. Wu, *Whitney's embedding theorem by solutions of elliptic equations and geometric consequences*, Proc. Sympos. Pure Math., vol. 27, Part 2, Amer. Math. Soc., Providence, R. I., (1975), pp. 287—296.

28. P. A. Griffiths, *Hermitian differential geometry, Chern classes and positive bundles*, Global Analysis, Papers in Honour of K. Kodaira, Princeton Univ. Press, Princeton, N. J. (1970), pp. 185—251.

29. ——, *A Schottky—Landau theorem for holomorphic mappings in several complex variable*, Ist. Naz. di Alte Mate. Symposia Math. **10** (1972), 229—243.

30. N. Hitchin, *The space of harmonic spinors*, Advances in Math. **14** (1974), 1—55.

31. ——, *Compact four-dimensional Einstein manifolds*, J. Differential Geometry **9** (1974), 435—441.

32. J. Kazdan and F. Warner, *A direct approach to the determination of gaussian and scalar functions*, Invent. Math. **28** (1975), 227—230.

33. ——, *Curvature functions for compact 2-manifolds*, Ann. of Math. **99** (1974), 14—47.

34. H. B. Lawson and S. T. Yau, *Scalar curvature, non-abelian group actions, and the degree of symmetry of exotic spheres*, Comment. Math. Helv. **49** (1974), 232—244.

35. A. Lichnerowicz, *Spineurs harmoniques*, C. R. Acad. Sci. Paris **257** (1963), 7—9.

36. J. Milnor, *A note on curvature and fundamental group*, J. Differential Geometry **2** (1968), 1—7.

37. C. B. Morrey, *Multiple integrals in the calculus of variations*, Springer-Verlag, New York, 1966.

38. ——, *The problem of plateau on a Riemannian manifold*, Ann. of Math. **49** (1948), 807—851.

39. J. Moser, *On Harnack's theorem for elliptic differential equation*, Comm. Pure Appl. Math. **14** (1961), 577—591.

40. ——, *On a non-linear problem in differential geometry*, Dynamical Systems (M. Peixoto, ed.), Academic Press, New York, 1973.

41. S. Myers, *Riemannian manifolds with positive mean curvature*, Duke Math. J. **8** (1941), 401—404.

42. L. Nirenberg, *The Weyl and Minkowski problems in differential geometry in the large*, Comm. Pure Appl. Math. **6** (1953), 337—394.

43. A. V. Pogorelov, *On the regularity of generalized solutions of the equation* $\det(\partial^2 u/\partial x^i \partial x^j) = \varphi(x^1, \ldots, x^n) > 0$, Dokl. Akad. Nauk SSSR **200** (1971), 534—537.

44. ——, *The Dirichlet problem for the n-dimensional analogue of the Monge—Ampère equation*, Dokl. Akad. Nauk SSR **201** (1971), 790—793.

45. ——, *Improper convex affine hyperspheres*, Dokl. Akad. Nauk SSSR **13** (1972), 240—242.

46. R. Schoen and S. T. Yau, *Harmonic maps and the topology of stable hypersurfaces and manifolds with non-negative Ricci curvature*, Comment Math. Helv. **39** (1977), 333—341.

47. ——, *Existence of incompressible minimal surfaces and the topology of three dimensional manifolds with non-negative scalar curvature*, Ann. of Math. (to appear).

48. J. L. Synge, *On the connectivity of spaces of positive curvature*, Quart. J. Math. Oxford Ser. **7** (1936), 316—320.

49. N. Trudinger, *Remarks concerning the conformal deformation of Riemannian structures on compact manifolds*, Ann. Scuola Norm. Sup. Pisa. **22** (1968), 265—274.

50. A. Van de Ven, *On the Chern numbers of surfaces of general type*, Invent. Math. **36** (1976), 235—293.

51. J. A. Wolf, *Growth of finitely generated solvable groups and convergence of Riemannian manifolds*, J. Differential Geometry **2** (1968), 421—446.

52. H. Yamabe, *On a deformation of Riemannian structures on compact manifolds*, Osaka Math. J. **12** (1960), 21—37.

53. S. T. Yau, *On some function theoretic properties of a complete Riemannian manifold and their applications in geometry,* Indiana Univ. Math. J. **25** (1976), 659—670.

54. *A general Schwarz lemma for Kähler manifolds,* Amer. J. Math. **100** (1978), 197—203.

55. *Calabi's conjecture and some new results in algebraic geometry,* Proc. Nat. Acad. Sci. U.S.A. **74** (1977), 1798—1799.

56. *On the Ricci curvature of a compact Kähler manifold and the complex Monge—Ampère equation* I, Comm. Pure Appl. Math. **31** (1978), 339—411.

57. *Harmonic functions on a complete Riemannian manifold,* Comm. Pure Appl. Math. **28** (1975), 201—228.

58. S. Kobayashi, *The first Chern class and holomorphic symmetric tensor fields* (to appear).

59. K. Jörgens, *Über die Lösungen der Differentialgleichung* $rt - s^2 = 1$, Math. Ann. **127** (1954), 130—134.

60. J. L. Kazdan, *A remark on the preceding paper of Yau,* Comm. Pure Appl. Math. **31** (1978), 412—413.

61. M. Berger and D. Ebin, *Some decompositions of the space of symmetric tensors on a Riemannian manifold,* J. Differential Geometry **3** (1969), 379—392.

62. B. Gaveau, *Méthode de contrôle optimal dans analyse complexe* I, *Résolution d'equation de Monge—Ampère*. J. Functional Analysis **25** (1977).

63. E. Calabi and E. Vesentini, *On compact locally symmetric Kähler manifolds,* Ann. of Math. **71** (1960), 472—507.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305, U.S.A.

Invited Addresses in Sections 1—5

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Nonstandard Number Theory

Angus Macintyre*

0. Introduction. This paper is intended as an outline of a model-theoretic method in diophantine geometry. The foundations, and the most advanced development, are due to Robinson and Roquette [**RR**], [**Roq 1**]. The key idea is to relate the algebraic geometry of a number field K to the nonstandard arithmetic of an enlargement K^*. For example, generic points acquire arithmetical structure. The main success of the method till now is a new proof of the Siegel–Mahler–Lang Theorem [**RR**], and new insights into the basic problem of effectiveness in that theorem [T]. The role of Hilbert's Irreducibility Theorem in diophantine geometry is clarified [**Roq 1**].

I concentrate on the Robinson—Roquette formulation of Weil's theory of distributions [**W**]. My only original contribution is a "covering theorem" relating geometric and arithmetical idèles. This gives another formulation of Weil's theory, based on the heuristic principle that "adèles of K become principal in K^*."

1. Foundations.

1.1. Let Sets be the category of sets and functions. Let I be any set, and Sets^I the product (or functor) category. We have the "diagonal functor" $\Delta: \text{Sets} \to \text{Sets}^I$. Let D be an ultrafilter on I. We have the "collapsing functor" $[/D]: \text{Sets}^I \to \text{Sets}^I/D$.

* Partially supported by grant MCS77—07731 from the National Science Foundation of the United States. The paper was completed at the Mathematics Institute of Warsaw University. The author thanks the administration of that institution for their hospitality and financial support. Most of all, he thanks his Warsaw colleagues, especially Dr. Cecylia Rauszer, for their unfailing kindness.

Let Sets* be the image of the collapsing functor. Sets* is the nonstandard universe corresponding to D. Let $* : \text{Sets} \to \text{Sets}^*$ be $[/D] \circ \Delta$. I write x^* for $*(x)$.

Sets, Sets^I and Sets* are topoi. I say this not to annoy, but simply because I believe that it is mathematically natural to have categorical foundations for nonstandard analysis. The tradition, deriving from Robinson, of converting Sets* locally, to an ε-structure, leads to a gruesome formalism, is quite unnecessary in practice, and can even obscure correct arguments.

With no assumptions on D, $*$ is an elementary map [CK]. If D is suitably good [CK], Sets* has the all-important property of $\varkappa$-saturation for some $\varkappa$ (usually $\varkappa = \omega_1$ suffices). To say Sets* is $\varkappa$-saturated means the following:

If $A_\lambda \to A$ $(\lambda \in \Delta)$ are morphisms, where $\operatorname{card}(\Delta) < \varkappa$, and if for each finite subset Δ_0 of Δ there are morphisms $1 \to_{\psi_\lambda;\Delta_0} A_\lambda$ $(\lambda \in \Delta_0)$ so that

$$\begin{tikzcd} & A_\lambda \arrow[dr] & \\ 1 \arrow[ur, "{\psi_\lambda, \Delta_0}"'] \arrow[dr, "{\psi_\mu, \Delta_0}"'] & & A \\ & A_\mu \arrow[ur] & \end{tikzcd}$$

commutes for λ, μ in Δ_0, then there are morphisms $1 \to A_\lambda$ $(\lambda \in \Delta)$ so that

$$\begin{tikzcd} & A_\lambda \arrow[dr] & \\ 1 \arrow[ur] \arrow[dr] & & A \\ & A_\mu \arrow[ur] & \end{tikzcd}$$

commutes for λ, μ in Δ. (1 is the terminal object of Sets*.)

1.2. For reasons I do not clearly understand, it is often useful to work in the double enlargement Sets**. This is obtained as follows. The objects and morphisms of Sets* are sets, so there is a well-defined category (Sets*)*, obtained by forming $(\text{Sets}^*)^I$ and then collapsing modulo D. Sets** is defined as (Sets*)*, and $** :$ Sets $\to$ Sets** is the obvious composite

$$\text{Set} \xrightarrow{*} \text{Sets}^* \xrightarrow{\text{new}*} (\text{Sets}^*)^*.$$

2. Universal domains. 2.1. Let K be a number field, and K^* its enlargement. All one needs of $*$ is ω_1-saturation. I now show K^* encodes many important infinitistic constructions over K. In this sense K^* is a universal domain for K. However, varieties over K with no points in K have no points in K^*, so my sense of universal domain is not that of algebraic geometry.

2.2. Let S_K be the set of those topologies T on K coming from a nontrivial absolute value $\|\cdot\|: K \to \boldsymbol{R}$. For $T \in S_K$, T^* is a (base for a) topology on K^*, coming from a generalized absolute value $\|\cdot\|^*: K^* \to \boldsymbol{R}^*$. $(S_K)^*$ is a set of (bases for) topologies on K^*, coming from generalized absolute values $\|\ \|: K^* \to \boldsymbol{R}^*$. The standard elements of $(S_K)^*$ are those of the form T^*, $T \in S_K$.

(a) $\boldsymbol{B}_T = \{x \in \boldsymbol{R}^*: |x|^* \leq r \text{ some } r \text{ in } \boldsymbol{R}\}$;

(b) $I = \{x \in R^*: |x|^* \leq r \text{ all } r > 0 \text{ in } R\}$.

The local compactness of R implies that $\boldsymbol{B}/I$ is canonically isomorphic to R. One has a retraction

$$R \to \boldsymbol{B}/I \to R$$

where the leftmost map is the natural inclusion, and the other is the fundamental *standard part map* st_R.

Now let $T \in S_K$. Consider

(a) $\boldsymbol{B}_T = \{x \in K^*: \|x\|^* \in \boldsymbol{B}\}$;

(b) $I_T = \{x \in K^*: \|x\|^* \in I\}$.

($\|\cdot\|$ is any absolute value for T.) Define $\|\cdot\|$ on $\boldsymbol{B}_T/I_T$ by

$$\|x + I_T\| = \mathrm{st}_R(\|x\|^*).$$

Then

THEOREM 1 *$\boldsymbol{B}_T/I_T$ with $\|\cdot\|$ is canonically isometrically isomorphic to K_T, the completion of K at T.*

Next consider:

(a) $\boldsymbol{B}_\infty = \{x \in K^*: x \in \boldsymbol{B}_T$, all standard T, and x is in the unit ball of T for all but finitely many standard $T\}$;

(b) $I_\infty = \bigcap_{T \text{standard}} I_T$.

Then

THEOREM 2 *$\boldsymbol{B}_\infty/I_\infty$ is canonically K-isomorphic to A_K, the ring of K-adèles.*

It is entirely routine to describe nonstandardly the adèlic topology and supplement Theorem 2. Since A_K/K is compact [C] there is an *adelic standard part map* $\mathrm{st}_{\mathrm{ad}}$ giving a retraction

$$\begin{array}{ccccc} (A_K/K) & \to & (A_K/K)^* & \xrightarrow{\mathrm{st}_{\mathrm{ad}}} & (A_K/K) \\ & & \| & & \\ & & (A_K)^*/K^* & & \end{array}$$

3. Product formula, idèles. 3.1. Since Artin—Whaples one knows that the product formula is the axiom for number theory. To formulate it one needs a canonical choice $\|\cdot\|_T$ of absolute value for each T in S_K. That choice is explained measure-theoretically [C] and may also be explained in terms of nonstandard counting.

Let the canonical choice be made as in [C].

Product formula. If $x \in K^\times$, $\prod_{T \in S_K} \|x\|_T = 1$.

This presupposes:

1st discreteness property. If $x \in K^\times$, then $\|x\|_T = 1$ for all but finitely many T in S_K.
Progressively more subtle are:

2nd discreteness property. For each $c > 0$ in $\boldsymbol{R}$ there are only finitely many x in K such that $\|x\|_T \leqslant c$ for all T in S_K.

3rd discreteness propert y (Roth's Theorem). Let S be a finite subset of S_K and let a_T $(T \in S)$ be elem ents of K. Let $k \in \boldsymbol{R}$, $k > 2$. Then there are only finitel many elements x in K s atisfying

$$\prod_{T \in S} \|x - a_T\|_T \leqslant \frac{1}{H(x)^k}$$

(where $H(x) = \prod_{T \in S_K} \max(1, \|x\|_T)$).

These properties are exploited in Sets*, via the principle that a set finite in Sets has no nonstandard "members" in Sets*.

3.2. *Idèles.* The group J_K of K-idèles is the group of invertible elements of A_K' topologized to make inversion continuous [C]. Given our discussion of A_K, it 'i routine to obtain J_K nonstandardly. Note that all K-adèles become *principal* in K^*

For $f \in J_K$ (classical definition) one defines

$$c(f) = \prod_{T \in S_K} \|f(T)\|_T,$$

and

$$J_K^0 = \{f \in J_K \colon c(f) = 1\}.$$

One has:

Compactness property: $K^\times$ is discrete in J_K^0, and $J_K^0/K^\times$ is compact.

Whence, there is an idelic standard part map

$$(J_K^0/K^\times)^* \xrightarrow[\mathrm{st_{id}}]{} J_K^0/K^\times,$$

so that $\mathrm{st_{id}}(\alpha)$ is idèlically-infinitesimally close to α. Let

$$F_K = \{\alpha \in (J_K)^* \colon \|\alpha(T)\|_T \in \boldsymbol{B} \setminus I, \text{ all } T\}.$$

By unravelling the compactness conditions, one has

THEOREM 3 (ROBINSON—ROQUETTE).

$$(J_K^0)^* = K^{*\times}(F_K \cap (J_K^0)^*).$$

That is, elements of $(J_K^0)^*$ are principal, modulo F_K.

Theorem 3 is a key result of nonstandard number theory. Let

$$\mathring{J}_{K^*} = (J_K)^*/F_K.$$

I refer to $\mathring{J}_{K^*}$ as the collapsed K^*-idèles.

3.3. *The formulation in V^{**}.* In my approach to Theorem 3 I freely mixed standard and nonstandard methods. But there is a definite gain in being systematically nonstandard.

A_K, an object in Sets, has been identified as B_∞/I_∞. The latter is not an element of (Sets)* (i.e. it is external) but it is an element of Sets, as is the ultrafilter D inducing $*$. That is, B_∞/I_∞ is an element of Sets defined in terms of D. So one considers $(B_\infty/I_\infty)^*$, which is naturally isomorphic to B^*_∞/I^*_∞, and I hope it is obvious that the latter is naturally isomorphic to $(A_K)^*$ (which I may write as A^*_K or A_{K^*}).

The appropriate picture is:

$$\begin{tikzcd}
\mathrm{Sets} \arrow[r] & \mathrm{Sets}^* \arrow[r] & \mathrm{Sets}^{**} \\
K \arrow[u] \arrow[r, "*"] \arrow[dr] & K^* \arrow[u] \arrow[r, bend left, "(*)^*"] \arrow[r, bend right, "\mathrm{new}^*"'] & K^{**} \arrow[u] \\
& B_\infty \arrow[u] \arrow[r] & (B_\infty)^* \arrow[u]
\end{tikzcd}$$

I must explain $(*)^*$ and new*. For the latter, see 1.2. $(*)^*$ is got by applying the functor $*$ to the map $*\colon K \to K^*$. It is to be stressed that new $^* \neq (*)^*$. However, both maps are elementary.

4. Nonstandard theory of distributions. 4.1. *Embedding function fields in K^*.* Let A be a variety (in affine n-space) defined over K.

LEMMA 4. *Let M be any set. Then A has infinitely many points in $K^n \cap M$ iff A has a nonstandard point in M^*.*

PROOF. Trivial, by ω_1-saturation.

So we come to the (at first glance tenuous) connection between nonstandard analysis and diophantine geometry. For example, let S be a finite subset of S_K, and define O_S, the set of S-integers, as the set of those y in K such that $\|y\|_T \leq 1$ for all $T \in S$. Let A be a *curve* over K. Then:

THEOREM 5 (SIEGEL–MAHLER–LANG). *Suppose A has genus ≥ 1. Then A has only finitely many points in O^n_S.*

For a proof using algebraic geometry and Roth's Theorem, see [L] or [S]–[D].

The connection with Lemma 7 is made by taking $M = O^n_S$. One then has the immediate reformulation:

THEOREM 5*. *Suppose A has genus $\geqslant 1$. Then A has no nonstandard point in O_S^{*n}.*

The nonstandard analysis of Theorem 5 looks at an arbitrary curve A over K, and examines the consequences of the

Assumption. A has a nonstandard point η in O_S^{*n}.

An important consequence, using the fact that A has dimension 1, is that η is a *generic* point of A. But this is a generic point with arithmetical structure. One seeks to prove Theorem 5 by confronting the geometry and the arithmetic of η.

Define $K(\eta)$ as the subfield of K^* generated by the coördinates of η. Then of course $K(\eta)$ is the function field of A over K. One now considers the general case

$$K \xrightarrow{\alpha} L \xrightarrow{\beta} K^* \tag{1}$$

where L is a function field of one variable over K and the diagram is a K-embedding.

L has thus two divisor theories. The first is the classical "geometric" theory, as expounded in [Chev], which studies those absolute values $\|\cdot\|$ on L which take the constant value 1 on $K^\times$. The second is the restriction of the divisor theory on K^* (based on $(S_K)^*$).

Let S_L be the set of the above "geometric topologies" on L. I refer to [Chev] and [S] for all details on the product formula for S_K, and for the L-adèles A_L and the L-idèles J_L.

4.2. *Covering S_L by $(S_K)^*$.* Suppose $T\in(S_K)^*$. $\|\cdot\|_T$ restricts to a "generalized absolute value" $L\to\boldsymbol{R}^*$. If T is nonarchimedean, then the induced topology is actually given by a (Krull) valuation. If the valuation is trivial on $K^\times$ then it is either trivial on L or discrete on L [Chev]. In the latter case the induced topology is in S_L. Our problem is how to handle those T for which the valuation is not trivial on $K^\times$. (There is also the problem of handling the finitely many archimedean T. This will be solved in passing.)

Let $T\in(S_K)^*$. Define

$$W_T(x) = -\log\|x\|_T \quad \text{for} \quad x\in K^{*\times}.$$

Now collapse the additive group $\boldsymbol{R}^*$ by the convex subgroup $\boldsymbol{B}$, and consider

$$v_{T,\boldsymbol{B}}(x) = W_T(x)+\boldsymbol{B}\in\boldsymbol{R}^*/\boldsymbol{B}.$$

If we put $v_{T,\boldsymbol{B}}(0)=\infty>\boldsymbol{R}^*/\boldsymbol{B}$, then we have:

LEMMA 6. *$v_{T,\boldsymbol{B}}$ is a valuation on K^*, trivial on K.*

$v_{T,\boldsymbol{B}}$ induces a topology $T/\boldsymbol{B}$ on L. Let

$$R_L = \{T/\boldsymbol{B}\colon T\in(S_K)^*,\ T/\boldsymbol{B}\ \text{nontrivial on}\ L\}.$$

The next lemma is vital. It depends on the 2nd discreteness property, and the Riemann–Roch Theorem.

LEMMA 7. $R_L=S_L$.

. This is the most that can be said in general. But if genus $(L) \geqslant 1$, much more is true, and this is precisely the Siegel–Mahler–Lang Theorem. Define

$$R_L^{\#} = \{T/\boldsymbol{B} \in R_L \colon T \text{ nonstandard}\}.$$

Then Theorem 5 is equivalent to:

THEOREM 5**. $R_L^{\#} = S_L$ *if* genus $(L) \geqslant 1$.

Apparently all known proofs of Theorem 5 involve considering finite algebraic extensions K_1 of K and the corresponding extensions $L_1 = K_1 \otimes_K L$ of L. Each element of S_L lifts to finitely many elements of S_{L_1} (all conjugate if K_1 is normal over K). Of central importance is the induced conorm embedding $A_L \to A_{L_1}$ (cf. [C]).

So it seems appropriate to construct a link

$$A_L \dashrightarrow (A_K)^*$$

since we already have Lemma 7.

Further motivation, and the key in obtaining a correct version of (2), comes from:

LEMMA 8. *Let* $T_0 \in S_L$. *Then* $\{T \in (S_K)^* \colon T/\boldsymbol{B} = T_0\}$ *is contained in a* **-finite set.*

To get (2), one chases round:

$$\begin{tikzcd}
 & & L^* \arrow[dr, "\beta^*"] & \\
L \arrow[urr, "*"] \arrow[drr, "\beta"'] & & & K^{**} \\
 & & K^* \arrow[uu, "\alpha^*"'] \arrow[ur, "\text{new}^*"] \arrow[ur, bend right, "(*)^*"'] & \\
 & K \arrow[uul, "\alpha"] \arrow[ur, "*"'] & &
\end{tikzcd}$$

Make the obvious definition of B_T, I_T for T in S_L, and then put:

(a) $\boldsymbol{B}_{\infty,L} = \{x \in L^* \colon x \in \boldsymbol{B}_T$, all standard T, and x is in the unit ball of T for all but finitely many standard $T\}$;

(b) $I_{\infty,L} = \bigcap_{T \text{standard}} I_T$.

There is a natural ring embedding (not an isomorphism!)

$$A_L \to \boldsymbol{B}_{\infty,L}/I_{\infty,L} \quad \text{(cf. 2).}$$

The philosophy is that elements of A_L are made principal in L^*.

Now embed $\boldsymbol{B}_{\infty,L}$ in K^{**} via β^*. Make the following definitions:

(c) For T in $(S_K)^{**}$,

$$\Gamma_T = \{\|\beta^*(y^*)\|_T \colon y \in L\}$$

(i.e. set of values of $\|\cdot\|_T$ on L);

(d) For T in $(S_K)^{**}$,

$$\boldsymbol{B}_T^{(L)} = \{x \in K^{**} \colon \|x\|_T \text{ is bounded above by an element of } B^* \cdot \Gamma_T\};$$

(e) $B_T^{(L)} = \{x \in K^{**} : x \in B_T^{(L)}$, all T in $(S_K)^*$, and $x \in B_T$ for all but *-finitely many T in $(S_K)^*\}$;

(f) For T in $(S_K)^{**}$, $I_T^{(L)}$ = set of non units of $B_T^{(L)}$;

(g) $I_\infty^{(L)} = \bigcap_{T \in (S_K)^*} I_T^{(L)}$.

The following is obtained by elementary considerations, unpacking definitions in the covering mechanism:

LEMMA 9. *Via β^*, $B_{\infty,L}$ is sent to $B_\infty^{(L)}$, and $I_{\infty,L}$ is sent to $I_\infty^{(L)}$. So β^* induces an embedding*

$$B_{\infty,L}/I_{\infty,L} \to B_\infty^{(L)}/I_\infty^{(L)}.$$

This is the best we can do towards (2). $B_{\infty,L}/I_{\infty,L}$ naturally contains A_L. And, clearly, $B_\infty^{(L)}/I_\infty^{(L)}$ is similar to $(A_K)^*$, with a distortion factor specific to L.

It will be shown in detail in another publication that Lemma 9 covers the Robinson–Roquette version (in terms of nonstandard divisors) of Weil's theory of distributions. One formulates an idèlic version of Lemma 9, and obtains a commuting diagram

$$\begin{array}{ccc} J_{\infty,L} & \longrightarrow & J_\infty^{(L)} \\ \downarrow & & \downarrow \\ D_L & \longrightarrow & (D_K)^* \text{ (finite divisors)} \end{array}$$

The left map is the natural divisor map.

The top map is induced by β^*.

The bottom map is that of Robinson—Roquette.

The right map is difficult to construct except by strict use of functoriality.

5. Concluding the proof of Theorem 5.** 5.1. The main idea is to exploit the strange analogy between $L \to^\beta K^*$ and a finite algebraic extension $L \to L_1$. One considers possible counterexamples to Theorem 5** (so called exceptional primes) and *exceptional divisors* which are sums of distinct exceptional primes. Using the above analogy, together with Roth's Theorem, one proves

$$\text{degree}\,(A) \leqslant 2 \cdot [L : K(x)]$$

for A exceptional and $x \in L \setminus K$.

5.2. One then deduces that $\deg(A) = 0$, if genus $(L) \geqslant 1$. The strategy of Robinson–Roquette was to find finite unramified extensions L' of L with commuting diagram

$$\begin{tikzcd} L \arrow[r] \arrow[rr, bend right, "\beta"'] & L' \arrow[r, "\beta"] & K^* \end{tikzcd}$$

For a function field M over K, define

$$d_M = \min_{x \in M \setminus K} [M : K(x)].$$

Then by elementary functorial arguments one shows that $\deg(A) \leqslant 2d_{L'}[L':L]$. So one wants to find L' with $d_{L'} \leqslant (1/2)[L':L]$. By an elementary, but lengthy, argument in **[Roq 2]** Roquette showed that if genus $(L) \geqslant 1$ then for sufficiently large n the maximal unramified semiabelian extension of L of exponent n satisfies this inequality. Finally, by using Theorem 3, one shows that these semiabelian extensions are L-embeddable into K^*. This concludes the proof.

6. The connection with Hilbert's irreducibility theorem. In the last part of their proof Robinson and Roquette are treating special cases of the following.

Problem. Given $L \dashrightarrow^{\beta} K^*$ as above, what is the structure of $\mathrm{Alg}(L, K^*)$, the lattice of intermediate fields finite-dimensional over K?

By a result of Gilmore–Robinson **[Roq 1]** Hilbert's Irreducibility Theorem for any K (not necessarily a number field) is equivalent to the existence of a pure transcendental $L=K(x)$ with $\mathrm{Alg}(L, K^*)$ *empty*. In a beautiful paper **[Roq 1]** Roquette shows that this is a *useful* result. Yet again one admires Robinson's vision, that glimpsed these deep ideas over twenty years ago.

Another interesting possibility for $\mathrm{Alg}(L, K^*)$, $L=K(x)$, is given by Roquette **[Roq 1]**. Namely, there may be exactly one extension of each degree n. This is closely connected with the nonstandard analysis of curves of genus 0 **[RR]**.

The general problem must be of extreme difficulty. For, if there is an L so that L has an unramified abelian extension of exponent 2 and dimension 8 in K^*, then Mordell's Conjecture is false, and conversely.

7. Effective estimates. An outstanding problem of number theory is to find effective bounds for the S-integral points on a curve C of genus $\geqslant 1$. It is widely conjectured that such a method exists. For integer points, and curves over Q of genus 1, Baker **[B]** found such a method.

Inspection of the classical proof **[L]** reveals two nodes of noneffectiveness. One is the use of Roth's Theorem, where no effective bounds are presently known. The second is the use of generators for the Mordell–Weil group of the Jacobian. Despite extensive scrutiny of the proof, and advice from leading authorities, I do not see how to eliminate either ineffectiveness.

An astonishing feature of the Robinson–Roquette proof is that there is no use of the generators of the Mordell–Weil group (although there are manoeuvres in the proof reminiscent of the proof of the Mordell–Weil Theorem). Robinson's last mathematical achievement was to see that the new proof reveals that there is an effective bound in the Siegel–Mahler–Lang Theorem, *relative to effective bounds in Roth's Theorem*. This is by no means evident from the Robinson–Roquette proof (still less from mine), but a careful axiomatization of what is used in the proof will yield the result. Takeuti **[T]** verified this claim of Robinson, by showing that the proof could be done in a suitable fragment of nonstandard analysis.

[The use of model theory to give effective estimates in algebra is well established. See **[R]** for the first ideas, and **[D]** for some nice new ones.]

As a matter of fact, number theorists tend to dismiss this relative recursive estimate. I do not understand why. If they are to be taken seriously in conjecturing an absolute effective estimate, why do they not provide, by approved methods, an elimination of the Mordell–Weil Theorem from the classical proof of Siegel–Mahler–Lang?

Postscript. This paper was completed in a mood of deep dismay, after the tragic death of my friend and former student George Loullis. I dedicate this work to his memory, in fond remembrance of his cheerful presence at work and at play.

References

[B] A. Baker, *Effective methods in the theory of numbers,* Proc. Internat. Congr. Mathematicians (Nice 1970), Vol. 1, Gauthier-Villars, Paris 1971, pp. 19—26.

[C] J. W. S. Cassels, *Global fields,* Algebraic Number Theory, Cassels and Fröhlich, Academic, London, 1967, pp. 42—84.

[Chev] C. Chevalley, *Introduction to the theory of algebraic functions of one variable,* New York, 1951.

[CK] C. C. Chang and H. J. Keisler, *Model theory,* North-Holland, Amsterdam, 1973.

[D] L. van den Dries, *Model theory of fields,* doctoral dissertation, Utrecht, 1978.

[L] S. Lang, *Diophantine geometry,* Interscience, New York, 1962.

[RR] A. Robinson and P. Roquette, *On the finiteness theorem of Siegel and Mahler concerning diophantine equations,* J. Number Theory **7** (1975), 121—176.

[Roq 1] P. Roquette, *Nonstandard aspects of Hilbert's Irreducibility Theorem,* Model Theory and Algebra, Saracino and Weisspfenning, eds., Lecture Notes in Math., vol. 498, Springer-Verlag, Berlin and New York, 1975, pp. 231—275.

[Roq 2] *On the division fields of an algebraic function field,* Houston J. Math. **2** (1976), 251—287.

[S] J. P. Serre, *Groupes algebriques et corps de classes,* Hermann, Paris, 1959.

[T] G. Takeuti, *A relatively effective procedure getting all quasiinteger solutions of diophantine equations with positive genuses* (unpublished).

[W] A. Weil, *Arithmetic on algebraic varieties,* Ann. of. Math. **53** (1951), 412—444.

Yale University
New Haven, Connecticut 06520, U.S.A.

University of Warsaw
Warsaw, Poland

Труды Международного Конгресса Математиков
Хельсинки, 1978

Уравнения в Свободной Полугруппе

Г. С. Маканин

1. Пусть Π—свободная полугруппа с конечным алфавитом образующих

(1) $$a_1, \ldots, a_k.$$

Пусть

(2) $$x_1, \ldots, x_n$$

алфавит переменных, значениями которых являются слова в алфавите (1). Уравнением в свободной полугруппе Π называется равенство слов в алфавите (1), (2)

(3) $$\varphi(a_1, \ldots, a_k, x_1, \ldots, x_n) = \psi(a_1, \ldots, a_k, x_1, \ldots, x_n).$$

Список слов $L_1, \ldots, L_n$ в алфавите (1) называется решением уравнения (3), если слова $\varphi(a_1, \ldots, a_k, L_1, \ldots, L_n)$ и $\psi(a_1, \ldots, a_k, L_1, \ldots, L_n)$ тождественно равны.

В настоящем докладе будет указан алгоритм, который по всякому уравнению в свободной полугруппе распознает, имеет оно решение или нет.

Наряду с уравнениями в свободной полугруппе рассматриваются и конечные системы уравнений в свободной полугруппе. Очевидно, что система двух уравнений

$$\begin{cases} \varphi_1 = \psi_1, \\ \varphi_2 = \psi_2, \end{cases}$$

имеет решение тогда и только тогда, когда имеет решение уравнение

$$\varphi_1 a_1 \varphi_2 \varphi_1 a_2 \varphi_2 = \psi_1 a_1 \psi_2 \psi_1 a_2 \psi_2$$

(см. [1], лемма 1.28), так что вопрос о разрешимости систем уравнений в свободной полугруппе легко сводится к вопросу о разрешимости одного уравнения в свободной полугруппе.

2. Начало исследованию уравнений в свободной полугруппе, называемых также уравнениями в словах, положил А. А. Марков в конце 50-х годов в связи с тогда еще не решенной 10-ой проблемой Гильберта. Используя результат Нильсена [2] о том, что матрицы второго порядка с натуральными элементами и определителем, равным единице, образуют свободную полугруппу с двумя образующими

$$a_1 = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad a_2 = \begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix} \tag{4}$$

и с единицей

$$1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{5}$$

А. А. Марков каждому уравнению

$$\varphi(a_1, a_2, x_1, \ldots, x_n) = \psi(a_1, a_2, x_1, \ldots, x_n) \tag{6}$$

в свободной полугруппе поставил в соответствие систему диофантовых уравнений Σ такую, что между решениями уравнения (6) и натуральными решениями системы Σ существует взаимно однозначное соответствие.

Система Σ строится из уравнения (6) следующим образом: Каждой словарной переменной x_i ставится в соответствие матрица

$$x_i = \begin{pmatrix} x_{i,1} & x_{i,2} \\ x_{i,2} & x_{i,4} \end{pmatrix} \quad (i = 1, \ldots, n), \tag{7}$$

где $x_{i,j}$ — неизвестные, принимающие натуральные значения. Матрицы из списков (4) и (7) перемножаются в порядке, указанном словами $\varphi(a_1, a_2, x_1, \ldots, x_n)$ и $\psi(a_1, a_2, x_1, \ldots, x_n)$. Получаются две матрицы

$$\begin{pmatrix} P_1 & P_2 \\ P_3 & P_4 \end{pmatrix}, \quad \begin{pmatrix} Q_1 & Q_2 \\ Q_3 & Q_4 \end{pmatrix},$$

где P_i, Q_i — многочлены с натуральными коэффициентами от неизвестных $x_{i,j}$. Искомая система Σ имеет вид

$$\begin{cases} P_j = Q_j \quad (j = 1, 2, 3, 4), \\ x_{i,1}x_{i,4} - x_{i,2}x_{i,3} = 1 \quad (i = 1, \ldots, n). \end{cases}$$

А. А. Марков предполагал, что будет доказана алгоритмическая неразрешимость уравнений в свободной полугруппе, из чего следовала бы алгоритмическая неразрешимость 10-ой проблемы Гильберта.

Алгоритм, распознающий разрешимость уравнений в свободной полугруппе, построенный докладчиком, показал, что предлагаемый А. А. Марковым подход к 10-ой проблеме Гильберта привести к успеху не может. С другой стороны построенный алгоритм вместе с соответствием, указанным А. А. Марковым, позволяет распознавать разрешимость некоторых систем диофантовых уравнений, а именно, тех систем, которые соответствуют уравнениям в свободной полугруппе.

3. На пути разрешимости уравнений в свободной полугруппе были получены следующие результаты: В 1964 году Ю. И. Хмелевский [3] построил алгоритм, распознающий разрешимость таких систем уравнений, каждое уравнение у которых содержит не более двух переменных. В 1967 году Ю. И. Хмелевский (см. [1]) построил алгоритм, разпознающий разрешимость уравнений с тремя переменными. В 1968 году Ю. В. Матиясевич [4] построил алгоритм, распознающий разрешимость таких систем, в которые каждая переменная входит не более двух раз. И наконец, в 1977 году докладчик [5], [6] построил алгоритм, распознающий разрешимость произвольных уравнений в свободной полугруппе.

4. Центральным понятием в алгоритме, распознающем разрешимость уравнений в свободной полугруппе, является понятие обобщённого уравнения.

Обобщённым уравнением Ω мы называем всякую систему, состоящую из следующих ниже пяти частей.

1. Алфавит коэффициентов:

$$a_1, \dots, a_\omega \quad (\omega > 0). \tag{8}$$

2. Таблица словарных переменных:

$$\begin{matrix} x_1, \dots, x_n, & x_{n+1}, \dots, x_{2n} \\ l_1, \dots, l_\varrho, & r_1, \dots, r_\varrho, t \end{matrix}$$

где $n \geqslant 0$, $\varrho \geqslant 2$. По определению

$$\Delta(p) = \begin{cases} p+n, & \text{если} \quad 1 \leqslant p \leqslant n, \\ p-n, & \text{если} \quad n+1 \leqslant p \leqslant 2n. \end{cases}$$

Переменная $x_{\Delta(p)}$ называется двойником переменной x_p. Каждая пара двойников удовлетворяет *равенству двойников*

$$x_i = x_{i+n} \quad (i = 1, \dots, n). \tag{9}$$

Переменные $l_1, \dots, l_\varrho$ называются границами. Каждой границе соответствует *граничное равенство*

$$l_i r_i = t \quad (i = 1, \dots, \varrho). \tag{10}$$

3. Таблица сравнения границ:

$$0 = \partial(l_1) < \partial(l_2) < \dots < \partial(l_{\varrho-1}) < \partial(l_\varrho) = \partial(t). \tag{11}$$

4. Таблица расположения основ: Обобщенное уравнение Ω содержит функции $\psi(i)$, $\alpha(i)$, $\beta(i)$ с областями определения и значений

$$\begin{gathered} \psi(i)\colon [2n+1, \dots, 2n+m] \to [1, \dots, \omega], \\ \alpha(i), \beta(i)\colon [1, \dots, 2n+m] \to [1, \dots, \varrho], \end{gathered}$$

причем $\psi(i)$ принимает все значения из $[1, \ldots, \omega]$, а $\alpha(i)$ и $\beta(i)$ удовлетворяют условиям $\alpha(i) < \beta(i)$ $(i = 1, \ldots, 2n+m)$.

Переменные $x_1, \ldots, x_n, x_{n+1}, \ldots, x_{2n}$ и коэффициенты $a_{\psi(2n+1)}, \ldots, a_{\psi(2n+m}$ называются *основами* Ω и обозначаются соответственно через

$$w_1, \ldots, w_n, \quad w_{n+1}, \ldots, w_{2n}, \quad w_{2n+1}, \ldots, w_{2n+m}.$$

Каждой основе w_i соответствует *равенство расположения* этой основы

$$l_{\alpha(i)} w_i r_{\beta(i)} = t \quad (i = 1, \ldots, 2n+m), \tag{12}$$

Граница $l_{\alpha(i)}$ называется левой границей основы w_i, а граница $l_{\beta(i)}$ — правой границей. Одна и та же граница может быть и левой и правой.

Некоторые границы $l_{i_1}, \ldots, l_{i_\tau}$ фиксируются и называются *начальными* границами.

Д. 1. **Всякая левая граница является начальной границей.**

Все начальные и все правые границы называются *существенными*. Остальные границы *несущественные*.

5. *Список граничных связей.* Ω содержит некоторый конечный (возможно, пустой) список граничных связей.

Всякая граничная связь имеет вид

$$l_p, x_{\lambda_1}, x_{\lambda_2}, \ldots, x_{\lambda_k}, w_{\lambda_{k+1}}, l_q, \tag{13}$$

где $k > 0$ и $q = \beta(\lambda_{k+1})$. Граница l_p называется исходной границей, граница l_q — заключительной границей.

Д. 2. **Всякая несущественная граница является исходной границей некоторой граничной связи.**

5. Список слов в алфавите (8) $B_1, \ldots, B_k, B_{k+1}$ называется *выпуклым*, если для всякого $i = 2, \ldots, k$ выполнена дизъюнкция

$$\partial(B_{i-1}) \leqslant \partial(B_i) \vee \partial(B_{i+1}) \leqslant \partial(B_i).$$

Таблица слов в алфавите (8)

$$X_1, \ldots, X_n, X_{n+1}, \ldots, X_{2n}, \quad L_1, \ldots, L_\varrho, \quad R_1, \ldots, R_\varrho, T \tag{14}$$

удовлетворяет граничной связи (13), если существуют непустые слова в алфавите (8) $B_1, \ldots, B_k, B_{k+1}, C_1, \ldots, C_k$ такие, что выполнены следующие условия:

1) $X_{\lambda_i} \overline{\underline{\circ}} B_i C_i$ $(i = 1, \ldots, k)$
2) $W_{\lambda_{k+1}} \overline{\underline{\circ}} B_{k+1}$
3) Список $B_1, \ldots, B_k, B_{k+1}$ выпуклый.
4) Если $\lambda_i = \lambda_j \& i \neq j$, то $B_i \overline{\underline{\phi}} B_j$
5) $L_p \overline{\underline{\circ}} L_{\alpha(\lambda_1)} B_1$
6) $L_{\alpha(\Delta(\lambda_i))} B_i \overline{\underline{\circ}} L_{\alpha(\lambda_{i+1})} B_{i+1}$ $(i = 1, \ldots, k-1)$
7) $L_{\alpha(\Delta(\lambda_k))} B_k \overline{\underline{\circ}} L_q$.

Таблица слов (14) называется *решением* обобщенного уравнения Ω, если она удовлетворяет условиям (9), (10), (11), (12) и всем граничным связям уравнения Ω. Число $\partial(T)$ называется длиной решения (14).

6. Алгоритм, распознающий разрешимость уравнений в свободной полугруппе, схематически состоит в следующем:

По заданному уравнению Y в свободной полугруппе строится список обобщенных уравнений $\Omega_1, \ldots, \Omega_r$ такой, что Y имеет решение тогда и только тогда, когда имеет решение некоторое Ω_i. Это сделать несложно, поскольку обобщенное уравнение является по существу уравнением в свободной полугруппе, записанным таким образом, что каждой основе w_i (переменной или коэффициенту) соответствует равенство вида $l_{\alpha(i)} w_i r_{\beta(i)} = t$, указывающее ее расположение в уравнении.

Затем по всякому обобщенному уравнению Ω строится список обобщенных уравнений $\Omega^{(1)}, \ldots, \Omega^{(q)}$, параметры записи которых не превосходят соответствующих параметров записи Ω, и кроме того Ω имеет решение длины d тогда и только тогда, когда некоторое $\Omega^{(i)}$ имеет решение длины меньшей d. Этот результат достигается, грубо говоря, за счет «переноса» слева направо основы w_j обобщенного уравнения Ω в случае, когда Ω содержит расположения

$$l_1 x_i r_{\beta(i)} = t, \quad l_1 w_j r_{\beta(j)} = t, \quad l_h x_{\Delta(i)} r_g = t \quad \text{при} \quad \beta(j) \leqq \beta(i),$$

$1 < h$, и за счет частичного сокращения слева оставшейся в одиночестве основы w_u с расположением $l_1 w_u r_{\beta(u)} = t$.

Затем доказывается, что обобщенных уравнений, участвующих в «дереве» заданного уравнения Y, может быть только конечное число, и указывается, как по некоторым специальным обобщенным уравнениям узнавать, имеют они решение или нет.

При доказательстве конечности «дерева» уравнения Y используется следующий результат В. К. Булитко [7]. Существует вычислимая функция $\gamma(i)$ такая, что всякое уравнение Y в свободной полугруппе, имеющее решение, имеет решение, показатель периодичности которого не превосходит $\gamma(d)$, где d — длина записи Y. (Показателем периодичности списка слов $L_1, \ldots, L_n$ называется максимальное число s такое, что некоторое L_t содержит подслово вида P^s, где P—непустое слово).

7. Полученный докладчиком результат легко обобщается до следующего результата.

Существует алгоритм, позволяющий по любой формуле вида

$$\exists x_1 \ldots \exists x_n \left(\bigvee_{i=1}^{p} \mathop{\&}_{j=1}^{r_i} \varphi_{i,j}(a_1, \ldots, a_k, x_1, \ldots, x_n) = \psi_{ij}(a_1, \ldots, a_k, x_1, \ldots, x_n) \right)$$

определить, истинна она или ложна.

Последний результат смыкается со следующим результатом В. Г. Дурнева [8].

Для всякого $n > 4$ не существует алгоритма, позволяющего по любой формуле вида

$$\exists x_1 \forall x_2 \exists x_3 \ldots \exists x_n \left(\bigvee_{i=1}^{p} \mathop{\&}_{j=1}^{z_i} \varphi_{i,j}(a_1, \ldots, a_k, x_1, \ldots, x_n) = \psi_{i,j}(a_1, \ldots, a_k, x_1, \ldots, x_n) \right)$$

определить, истинна она или ложна.

Литература

1. Ю. И. Хмелевский. *Уравнения в свободной полугруппе*, Труды Мат. Инст. Стеклова **107** (1971).

2. J. Nielsen *Die Gruppe der dreidimensionalen Gittertransformationen*. Danske Vid. Selsk., Math.—Fys. Medd. **5** (1924), no. 12.

3. Ю. И. Хмелевский. *Решение некоторых систем уравнений в словах*, ДАН СССР **156** (1964), № 4.

4. Ю. В. Матиясевич. *Связь систем уравнений в словах и длинах с 10-й проблемой Гильберта*, Зап. Научн. Сем. ЛОМИ АН СССР **8** (1968), 132—144.

5. Г. С. Маканин. *Проблема разрешимости уравнений в свободной полугруппе*, ДАН СССР **233** (1977), № 2.

6. Г. С. Маканин. *Проблема разрешимости уравнений в свободной полугруппе*, Мат. Сб., **103 (145)** (1977), № 2.

7. В. К. Булитко, *Об уравнениях и неравенствах*, Учен. Зап. Мат. Кафедр. Тульский Гос. Пед. Инст, вып. 2 (1970), 242—253.

8. В. Г. Дурнев, *Позитивная теория свободной полугруппы*, ДАН СССР **211** (1973), № 4.

Математический Институт им. В. А. Стеклова
Академии Наук СССР
Москва

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Infinite Games

Donald A. Martin*

1. Introduction. If σ and τ are finite or infinite sequences, $\sigma \prec \tau$ means that τ extends σ. A *tree* T is a collection of finite sequences such that, if $\sigma \in T$ and $\tau \prec \sigma$, then $\tau \in T$. We will be especially concerned with the tree Seq of all finite sequences of natural numbers. If T is a tree, $[T]$ is the set of all infinite sequences x such that, for all $n \in \omega$, $x \upharpoonright n \in T$, where $x \upharpoonright n$ is the σ of length n such that $\sigma \prec x$. [Seq] we identify with ω^ω, the set of all functions from the natural numbers to the natural numbers.

Suppose T is a tree such that every element of T has a proper extension belonging to T. Let $A \subseteq [T]$. We define a game G, *the game with payoff* A, as follows. Two players, I and II, take turns moving as follows:

$$\begin{array}{cccc} z_0 & z_1 & z_2 & z_3 \; \ldots \\ \mathrm{I} & \mathrm{II} & \mathrm{I} & \mathrm{II} \ldots \end{array}$$

Each sequence $\langle z_0, \ldots, z_n \rangle$ must belong to T. I *wins* a play of G just in case the sequence $\langle z_i \colon i \in \omega \rangle \in A$. The notions of *strategy for* I (or II) for G and *winning strategy for* I (or II) for G are defined in the obvious way. G is *determined* if either I or II has a winning strategy for G.

Gale and Stewart [3] introduced the games G and proved, using the axiom of choice, that there is an undetermined game with $T = \text{Seq}$. To consider more restricted games, they put a topology on $[T]$ by letting the basic open sets be those sets of the

* This paper was supported in part by Grant Number MCS 76-05525 from the National Science Foundation of the United States.

forms $\{x: \sigma \prec x\}$ for $\sigma \in T$. Let us say that G is *open, Borel,* etc., just in case the payoff A is open, Borel, etc.

1.1 THEOREM (GALE-STEWART). *All open games are determined.*

After further results by Wolfe [10], Davis [1], and Paris [8], Martin [6] proved the following result.

1.2 THEOREM. *All Borel games are determined.*

The method of proof is to associate, with an $A \subseteq [T]$ of Borel rank α, an $A^* \subseteq [T^*]$ with A^* open and to prove that the game G with payoff A and the game G^* with payoff A^* are equivalent: whoever has a winning strategy for one has a winning strategy for the other. T^* is much bigger than T: if T has size $\aleph_\beta$, then T^* has size roughly $\aleph_{\beta+\alpha}$. Individual moves in G^* represent complex commitments as to how the players will move in an associated play of G. Results of Friedman [2] showed that, even for $T=\text{Seq}$, some kind of appeal to uncountable cardinals would be necessary to prove all Borel games are determined.

If Γ is a class of subsets of ω^ω, let Det (Γ) be the assertion that all games with $T=\text{Seq}$ and payoff in Γ are determined. Recent work (see [7]) has shown that Det (Projective) is a very powerful hypothesis in descriptive set theory. For example, Det (Π^1_2) = Det $(CPCA)$ implies that all Σ^1_3 sets of real numbers are Lebesgue measurable and yields a complete structural theory for levels three and four of the projective hierarchy.

J. Mycielski observed that a result of [1] implies that Det (Π^1_1) is not provable in the usual set theory ZFC. If one assumes *large cardinal axioms,* one gets more determinacy:

1.3 THEOREM (MARTIN [5]). *If a measurable cardinal exists,* Det (Π^1_1).

1.4 THEOREM. *If there are* 2 *(actually* $1\frac{1}{2}$*) supercompact cardinals, then* Det $(A(\Pi^1_1))$ *where* A *is operation* A.

1.5 THEOREM. *If there is a non-trivial iterable elementary embedding of a rank* R_λ *into itself, then* Det (Π^1_2).

L. Harrington has proved the converse of a slightly sharper version of Theorem 1.3. It is known from work of J. Green, Martin, W. Mitchell, J. Simms, and R. Solovay that much stronger hypotheses than those of Theorem 1.3 are needed to prove the conclusions of Theorems 1.4 and 1.5.

The rest of this paper is devoted to sketching the proof of Theorem 1.5. Iterability will be explained in §2 below. The hypothesis of Theorem 1.5 is strictly weaker (barring inconsistency) than the existence of an elementary $j: V \to M$ with M transitive and $j \upharpoonright R_\lambda \neq \text{identity}$ and $j(R_\lambda)=R_\lambda$. Kunen [4] shows that $j(R_{\lambda+1})=R_{\lambda+1}$ impossible.

2. Iterable embeddings. For the rest of this paper let $j: R_\lambda \to R_\lambda$ be an elementary embedding first moving $\varkappa$. Let $\varkappa_0 = \varkappa$ and $\varkappa_{i+1} = j(\varkappa_i)$. It follows by [4] that $\lambda = \sup_i \varkappa_i$ or $\lambda = \sup_i \varkappa_i + 1$. We assume the former.

If $Y \subseteq R_\lambda$, let $j(Y) = \bigcup_i j(Y \cap R_{\varkappa_i})$. Clearly $j(j)$ is an elementary embedding of R_λ into R_λ, first moving $\varkappa_1$. Let $j_0 = j$ and $j_{n+1} = j_n(j_n)$. Let $j_{n,n}$ = identity and $j_{n,m+1} = j_m \circ j_{n,m}$ for $n \leq m$. Here $\circ$ denotes composition. Let $M_i = R_\lambda$, $i = 0, 1, \ldots$. As long as direct limits are well-founded, we can iterate the system $\langle M_i, j_{n,m} \rangle$ to get a system $\langle M_\alpha, j_{\beta,\gamma} \rangle$ for ordinals α, β, γ with $\beta \leq \gamma$, where $j_{\beta,\gamma}: M_\beta \to M_\gamma$ is elementary and each M_α is transitive. When this can be done, we say that j is *iterable*. Set $j_\alpha = j_{\alpha,\alpha+1}$.

2.1 LEMMA. *If $m \leq n$, $j_{m,n} = j_m^{n-m}$.*

PROOF. If $z \in R_\lambda$, $j_q \circ j_q(z) = (j_q(j_q)) \circ j_q(z)$ by elementarity, and this is just $j_{q+1} \circ j_q(z)$ Applying this fact repeatedly yields the lemma.

2.2 LEMMA. *Suppose j is iterable. Suppose $\alpha \leq \beta$ are ordinals and $n \in \omega$.*

$$j_{\alpha,\beta} \circ j_{\alpha+n} = j_{\beta+n} \circ j_{\alpha,\beta}.$$

PROOF. For $z \in M_\alpha$ $(= M_{\alpha+n})$, $j_{\alpha,\beta} \circ j_{\alpha+n}(z) = (j_{\alpha,\beta}(j_{\alpha+n})) \circ j_{\alpha,\beta}(z)$. But $j_{\alpha,\beta}(j_{\alpha+n}) = j_{\beta+n}$.

3. β-embeddings. If β is an ordinal, a *β-embedding* is an elementary embedding

$$k: R_{\alpha+\beta} \to R_{\alpha'+\beta}.$$

first moving $\alpha > \beta$. Set $\nu(k) = \alpha$ and $\nu'(k) = \alpha'$. If k is a β-embedding and $\gamma + 1 < \beta$ define a 0-1 measure μ_γ^k as follows:

$$\mu_\gamma^k(X) = 1 \leftrightarrow k \restriction R_{\nu(k)+\gamma} \in k(X).$$

It is easily checked that μ_γ^k is $\nu(k)$-complete and concentrates in γ-embeddings k' with $\nu(k') < \nu(k)$ and $\nu'(k') = \nu(k)$. The following lemmas are easily verified.

3.1 LEMMA. *If k is a β_2-embedding and $\gamma + 1 < \beta_1 \leq \beta_2$, then $\mu_\gamma^k = \mu_\gamma^{k \restriction R_{\nu(k)+\beta_1}}$.*

3.2 LEMMA. *Let k be a β-embedding and let $\gamma_1 \leq \gamma_2$ and $\gamma_2 + 1 < \beta$. Suppose $\mu_{\gamma_1}^k(X) = 1$. Then*

$$\mu_{\gamma_2}^k\{z:\ z \restriction R_{\nu(z)+\gamma_1} \in X\} = 1.$$

4. A normal form for Π_2^1 sets. For the rest of this paper, let $A \subseteq \omega^\omega$ be a fixed Π_2^1 set. Let Seq* be the collection of nonempty elements of Seq. For $\sigma \in$ Seq, let $\mathrm{lh}(\sigma)$ be the length of σ. Let $\mathrm{Seq}^{*2} = \{\langle \sigma, \tau \rangle: \sigma, \tau \in \mathrm{Seq}^* \,\&\, \mathrm{lh}(\sigma) = \mathrm{lh}(\tau)\}$. The following lemma is just a restatement of Shoenfield's analysis of Π_2^1 sets [9].

4.1 LEMMA. *There is a function $\varrho: \mathrm{Seq}^{*2} \to \omega$ such that $\mathrm{lh}(\sigma) = 1 \to \varrho(\sigma, \tau) = 0$ and $\mathrm{lh}(\sigma) > 1 \to \varrho(\sigma, \tau) + 1 < \mathrm{lh}(\sigma)$ and, for any uncountable cardinal η and any*

$x \in \omega^\omega$, $x \in A$ *if and only if there are* $F_\tau : [\eta]^{\mathrm{lh}(\tau)} \to \eta$ *for* $\tau \in \mathrm{Seq}^*$ *such that, if* $\alpha_0 < \ldots < \alpha_{\mathrm{lh}(\tau)} < \eta$ *and* τ' *is a one-term extension of* τ, *then*

$$F_{\tau'}\{\alpha_0, \ldots, \alpha_{\mathrm{lh}(\tau)}\} < F_\tau\{\alpha_0, \ldots, \widehat{\alpha_{\varrho(x \restriction \mathrm{lh}(\tau'), \tau')}}, \ldots, \alpha_{\mathrm{lh}(\tau)}\}.$$

Here $[\eta]^n$ is the collection of all size n subsets of η.

4.2 Lemma. *If* j *is iterable and* $x \in \omega^\omega$, $x \in A$ *if and only if there is an* $H : \mathrm{Seq}^* \to \lambda$ *such that*

(1) $\mathrm{lh}(\tau) = 1 \to H(\tau) < \varkappa_1$;

(2) *if* τ' *is a one-term extension of* τ, *then* $H(\tau') < j_{\varrho(x \restriction \mathrm{lh}(\tau'), \tau')} H(\tau)$.

Proof. Assume the F_τ exist with $\eta = \varkappa$. If $\mathrm{lh}(\tau) = n$, let

$$H(\tau) = (j_{0,n}(F_\tau))\{\varkappa_0, \ldots, \varkappa_{n-1}\}.$$

(1) is immediate, and applications of Lemma 2.1 yield (2).

Now assume that H exists. If $n = \mathrm{lh}(\tau)$ and $\alpha_0 < \ldots < \alpha_{n-1}$, set

$$F_\tau\{\alpha_0, \ldots, \alpha_{n-1}\} = j_{\alpha_{n-2}+1, \alpha_{n-1}} \circ \ldots \circ j_{\alpha_0+1, \alpha_1} \circ j_{0, \alpha_0}(H(\tau)).$$

Applications of Lemma 2.2 show that, with suitable η, the F_τ are as required.

5. Proof of Theorem 1.5. Let G be the game with payoff A. Let $\tau_0, \tau_1, \ldots,$ be an enumeration of Seq^* such that $\tau_i \prec \tau_j \to i \leq j$. Let G^* be played as follows:

$$\begin{array}{cccccc} (n_0, \alpha_0) & n_1 & (n_2, \alpha_1) & n_3 & (n_4, \alpha_2) & \ldots \\ \mathrm{I} & \mathrm{II} & \mathrm{I} & \mathrm{II} & \mathrm{I} & \ldots \end{array}$$

Let $x(i) = n_i$ and $H(\tau_i) = \alpha_i$. I *wins* G^* just in case H obeys the constraints of Lemma 4.2.

5.1 Lemma. G^* *is determined.*

Proof. G^* is closed.

5.2 Lemma. *If* I *has a winning strategy for* G^*, *then* I *has a winning strategy for* G.

5.3 Lemma. *If* II *has a winning strategy for* G^*, *then* II *has a winning strategy for* G.

Proof. Let s^* be a winning strategy for II for G^*. We define a strategy s for II for G. Let σ be a position in G with II to move. Let $\beta_0, \ldots, \beta_m$ be even ordinals such that, if σ is extended to a position σ^* in G^* by setting $\alpha_i = \beta_i$, then I is not already lost at σ^*. We define an iterated product measure on the set of all sequences $\langle k_0, \ldots, k_m \rangle$, where each k_i is a β_i-embedding with $\nu'(k_i) < \lambda$. To do this we assign to each i a measure space, which may depend upon

$\langle k_0, \ldots, k_{i-1}\rangle$. If $\mathrm{lh}(\tau_i)=1$, then the measure for i is $\mu_{\beta_i}^{j_1 | R_{\varkappa_1+\beta_i+2}}$. If $\tau_{i'}$ is a one-term extension of τ_i, then the measure for i', is

$$\mu_{\beta_i'}^{j_a(k_i)}, \quad \text{where} \quad a = \varrho(\sigma \restriction \mathrm{lh}(\tau_{i'}), \tau_{i'}).$$

Let $s(\sigma)$ be the constant value of $s^*(\sigma^*(k_0, \ldots, k_m))$ for measure one of $\langle k_0, \ldots, k_m\rangle$, where $\sigma^*(k_0, \ldots, k_m)$ is the result of extending σ by setting $\alpha_i = \nu(k_i)$. Lemmas 3.1 and 3.2 imply that $s(\sigma)$ is independent of the choice of the β_i.

Suppose that x is a play of G according to s and that $x \in A$. Let H witness that $x \in A$. Let $\beta_i = 2H(\tau_i)$. Using the β_i to compute s, we can find a sequence $k_0, k_1, \ldots,$ such that, if we set $\alpha_i = \nu(k_i)$, then we extend x to a play of G^* according to s^* which is won by I. This contradiction completes the proof.

References

1. Morton Davis, *Infinite games of perfect information*, Advances in Game Theory, Ann. of Math. Studies No. 52, Princeton Univ. Press, Princeton, N. J., 1964, pp. 85—101.

2. H. Friedman, *Higher set theory and mathematical practice*, Ann. Math. Logic **2** (1971), 326—357.

3. D. Gale and F. M. Stewart, *Infinite games with perfect information*, Contributions to the Theory of Games, Ann. of Math. Studies No. 28, Princeton Univ. Press, Princeton, N. J., 1953, pp. 245—266.

4. K. Kunen, *Elementary embedding and infinitary combinatorics*, J. Symbolic Logic **36** (1971), 407—413.

5. D. A. Martin, *Measurable cardinals and analytic games*, Fund. Math. **66** (1970), 287—291.

6. ——, *Borel determinacy*, Ann. of Math. **102** (1975), 363—371.

7. Y. Moschovakis, *New methods and results in descriptive set theory*, Proceedings of the International Congress of Mathematicians (Vancouver 1974), Volume I, pp. 251—257.

8. J. Paris, ZF $\vdash \Sigma_4^0$ *determinateness*, J. Symbolic Logic **37** (1972), 661—667.

9. J. R. Shoenfield, *The problem of predicativity*, Essays on the Foundations of Mathematics, Magnes Press, Jerusalem, 1961, pp. 132—139.

10. P. Wolfe, *The strict determinateness of certain infinite games*, Pacific J. Math. **5** (1955), 841—847.

University of California
Los Angeles, California 90024, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Recursive Enumerability

Robert I. Soare*

One of the fundamental contributions of mathematical logic has been the precise definition and study of algorithms and the closely associated study of recursively enumerable sets. A subset $A \subseteq \omega$ is *recursive (decidable)* if there is an algorithm for computing its characteristic function c_A and *recursively enumerable* (r.e.) if there is an algorithm for generating its members. Nonrecursive r.e. sets have played a crucial role in undecidability results beginning with Gödel's incompleteness theorem [2] and more recently in number theory and group theory. Matiyasevič showed undecidability of Hilbert's tenth problem by proving that every r.e. set A is Diophantine (namely there is a polynomial $p(x, \vec{y})$ with integral coefficients such that $x \in A$ iff $(\exists \vec{y})[p(x, \vec{y})=0]$), and Boone, Clapham and Fridman each independently proved that every r.e. degree is the degree of the word problem for a finitely presented group (thus generalizing the Boone–Novikov result that the word problem is unsolvable).

For sets $A, B \subseteq \omega$ (the set of nonnegative integers), A is *recursive in (Turing reducible to)* B, written $A \leqslant_T B$, if there is an algorithm for computing c_A given c_B, and $A \equiv_T B$ if $A \leqslant_T B$ and $B \leqslant_T A$. The *degree* of A, $\mathrm{dg}(A)$, is the equivalence class $\{B: B \equiv_T A\}$, $\mathrm{dg}(A) \leqslant \mathrm{dg}(B)$ if $A \leqslant_T B$, and a degree is r.e. if it contains an r.e. set. The classification of r.e. sets was initiated by Post [10] who posed the problem: does there exist more than one nonrecursive r.e. degree? The existence of infinitely many such degrees implies for example that there are infinitely many genuinely different unsolvable word problems for finitely presented groups.

* The author was partially supported by grant MCS 76-07033 A01 from the National Science Foundation of the United States.

Here we describe some recent results on the classification of r.e. sets and their degrees. A detailed survey of current developments and bibliography can be found in [14], and a complete treatment will appear in the monograph [15].

1. The relation of the structure of an r.e. set to its degree. The r.e. sets $\{W_n\}_{n\in\omega}$ form a distributive lattice $\mathscr{E}$ under inclusion whose complemented elements are precisely the recursive sets. The r.e. degrees form an upper semi-lattice $\boldsymbol{R}$ with least element $\mathbf{0}=\mathrm{dg}\,(\emptyset)$ and greatest element $\mathbf{0}'=\mathrm{dg}\,(K)$, where K is the complete r.e. set $\{n: n\in W_n\}$. Post's program for solving his problem was to find a structural property of the complement $\bar{A}$ of an r.e. set A which guarantees incompleteness, namely $\emptyset <_T A <_T K$. More generally the program is to relate the $\mathscr{E}$-structure of an r.e. set to its degree. Post believed that a coinfinite r.e. set A with a complement $\bar{A}$ sufficiently "thin" with respect to containment of r.e. sets would be incomplete. A coinfinite r.e. set A is *simple* if $\bar{A}$ contains no infinite r.e. set. Post constructed simple sets, proved their incompleteness for reducibilities weaker than $\leq_T$, and introduced sets with still thinner complements (*h*-simple and *hh*-simple) to handle T-reducibility.

Let $\mathscr{E}^*$ denote the quotient lattice of $\mathscr{E}$ modulo the ideal $\mathscr{F}$ of finite sets, and A^* the equivalence class of A in $\mathscr{E}^*$. The thinnest possible (infinite) complement is possessed by a *maximal* set M, namely $M\in\mathscr{E}$ such that M^* is a coatom of $\mathscr{E}^*$. Friedberg constructed maximal sets and Yates showed they could be complete (i.e. $M\equiv_T K$). We give a negative answer to Post's program for a much larger class of properties, those invariant under $\mathrm{Aut}\,\mathscr{E}$, the group of automorphisms of $\mathscr{E}$. (A partial positive answer for noninvariant properties is given in [13].)

THEOREM 1.1. *For any nonrecursive r.e. set A there exists $\Phi\in\mathrm{Aut}\,\mathscr{E}$ such that $\Phi(A)\equiv_T K$.*

Meanwhile, the existence of infinitely many nonrecursive r.e. degrees was shown by Friedberg and Muchnik and their classification under the jump operator was carried out by Sacks, Lachlan, Martin and others. For $A\subseteq\omega$, define the *jump* of A, $A'=\{n: n\in W_n^A\}$ where W_n^A is the nth set which is r.e. relative to A. The jump operator is well-defined on degrees, where $\mathrm{dg}\,(A)'=\mathrm{dg}\,(A')$. Let $\boldsymbol{a}^{(n+1)}=(\boldsymbol{a}^{(n)})'$. For each $n\geq 0$ define the subclasses of r.e. degrees,

$$\boldsymbol{H}_n=\{\boldsymbol{d}: \boldsymbol{d}\in\boldsymbol{R} \text{ and } \boldsymbol{d}^{(n)}=\boldsymbol{0}^{(n+1)}\}, \quad\text{and}$$
$$\boldsymbol{L}_n=\{\boldsymbol{d}: \boldsymbol{d}\in\boldsymbol{R} \text{ and } \boldsymbol{d}^{(n)}=\boldsymbol{0}^{(n)}\},$$

where $\boldsymbol{d}^{(0)}=\mathbf{0}$ and $\bar{\boldsymbol{L}}_n=\boldsymbol{R}-\boldsymbol{L}_n$. The degrees in $\boldsymbol{H}_1(\boldsymbol{L}_1)$ are called *high (low)* since they have the highest (lowest) possible jump. An r.e. set A is *high (low)* if $\mathrm{dg}\,(A)\in\boldsymbol{H}_1(\boldsymbol{L}_1)$.

In the opposite direction of Post's approach Martin [8] showed that maximal sets M (and many others with thin complements) more closely resemble *complete* than incomplete sets since $\bar{M}$ dominates every recursive function, and this guarantees that M has high degree. For $\mathscr{C}\subseteq\mathscr{E}$, let $\mathrm{dg}\,(\mathscr{C})=\{\mathrm{dg}\,(W): W\in\mathscr{C}\}$.

THEOREM 1.2 (MARTIN). *Let $\mathcal{M}$ be the class of maximal sets. Then* $\mathrm{dg}(\mathcal{M})=\boldsymbol{H}_1$.

If the sets of high degree resemble complete sets, those of low degree should resemble recursive sets. For $A\in\mathcal{E}$ define the principal filter $\mathcal{L}(A)=\{W\colon W\in\mathcal{E}$ and $A\subseteq W\}$. If R is a coinfinite recursive set then $\mathcal{L}(A)\cong\mathcal{E}$, because $\bar{R}$ is recursively isomorphic to ω.

THEOREM 1.3. *If a coinfinite r.e. set A is low* ($\mathrm{dg}(A)\in\boldsymbol{L}_1$) *then* $\mathcal{L}(A)\cong\mathcal{E}$.

A class $\boldsymbol{C}$ of r.e. degrees is *invariant if* $\boldsymbol{C}=\mathrm{dg}(\mathscr{C})$ for some class $\mathscr{C}\subseteq\mathcal{E}$ invariant under $\mathrm{Aut}\,\mathcal{E}$. Martin asked which other degree classes are invariant besides $\boldsymbol{H}_1$ (and the trivial classes $\boldsymbol{R}, \boldsymbol{L}_0$, and $\bar{\boldsymbol{L}}_0$). In particular, he asked for a classification of $\mathrm{dg}(\mathcal{M}^{\#})$ where $\mathcal{M}^{\#}$ is the class of *atomless* sets, coinfinite r.e. sets with no maximal superset.

THEOREM 1.4 (LACHLAN [4], SHOENFIELD [11]). $\mathrm{dg}(\mathcal{M}^{\#})=\bar{\boldsymbol{L}}_2$.

It is unknown whether the methods of this theorem can be combined with Theorem 13 to replace $\boldsymbol{L}_1$ by $\boldsymbol{L}_2$ in the latter. Besides $\bar{\boldsymbol{L}}_0$, $\boldsymbol{H}_1$, $\bar{\boldsymbol{L}}_2$ it is unknown which classes of the form $\boldsymbol{H}_n$, $\bar{\boldsymbol{L}}_n$ are invariant. The case of $\bar{\boldsymbol{L}}_1$ is particularly interesting. A.s a strong generalization of Theorems 1.1 and 1.2 we conjecture that any invariant degree class $\boldsymbol{C}$ is closed upward and $\boldsymbol{H}_1\subseteq\boldsymbol{C}$.

It was conjectured that *every* nontrivial invariant degree class $\boldsymbol{C}$ is of the form $\boldsymbol{H}_n$ or $\bar{\boldsymbol{L}}_n$. This is refuted by the class $\mathscr{D}$ of *d*-simple sets, coinfinite r.e. sets simple with respect to certain *differences* of r.e. sets (d.r.e. sets). The latter naturally arise in studying the structure and automorphisms of $\mathcal{E}$ because the members of the Boolean algebra $\mathscr{A}$ generated by $\mathcal{E}$ are finite unions of d.r.e. sets.

THEOREM 1.5 (LERMAN–SOARE [7]). *Let* $\boldsymbol{D}=\mathrm{dg}(\mathscr{D})$. *Then* $\boldsymbol{H}_1\subseteq\boldsymbol{D}$ *and* $\boldsymbol{D}$ *splits* $\boldsymbol{L}_1$, *so* $\boldsymbol{D}$ *is not of the form* $\boldsymbol{H}_n$ *or* $\bar{\boldsymbol{L}}_n$ *for any* n.

2. The structure and automorphisms of $\mathcal{E}$. A major goal in studying the structure of $\mathcal{E}$ is to find complete sets of invariants for classifying the orbit of $A\in\mathcal{E}$ under $\mathrm{Aut}\,\mathcal{E}$. (Since every $\Phi\in\mathrm{Aut}\,\mathcal{E}^*$ is induced by some $\Psi\in\mathrm{Aut}\,\mathcal{E}$ one can consider either $\mathcal{E}$ or $\mathcal{E}^*$.) In **[12]** a new method is introduced for generating automorphisms of $\mathcal{E}$ and it is used to prove

THEOREM 2.1. *If A and B are maximal sets then $\Phi(A)=B$ for some $\Phi\in\mathrm{Aut}\,\mathcal{E}$.*

In the proof we as yet know too little of the structure of $\mathcal{E}$ to specify $\Phi(W)$ immediately given $W\in\mathcal{E}$. Rather we attempt to simultaneously enumerate arrays of r.e. sets $\{W_{f(n)}\}_{n\in\omega}$ and $\{W_{g(n)}\}_{n\in\omega}$ such that $\Phi(W_n)=^* W_{f(n)}$ and $\Phi^{-1}(W_n)=^* W_{g(n)}$ (where $X=^* Y$ denotes that the symmetric difference $X\Delta Y$ is finite). For different values of n these requirements generate considerable conflicts which are resolved by a complicated machinery. An immediate corollary of Theorem 2.1 is that for every $k\in\omega$ the group $\mathrm{Aut}\,\mathcal{E}^*$ is k-ply transitive on its coatoms. Hence, if $\mathcal{L}^*(A)$ is finite then the orbit of A is completely determined by the isomorphism type of $\mathscr{L}^*(A)$. When $\mathscr{L}^*(A)$ is infinite this is not necessarily true even when $\mathscr{L}^*(A)$ is particularly well-behaved. We say $X \subseteq Y$ is an *r-maximal major subset* (rm subset) of Y if $Y-X$ is infinite, $Y-X$ is not split into infinite pieces by any recursive set, and for any $W \in \mathscr{E}$, if $W \cup Y = \omega$ then $W \cup X =^* \omega$.

THEOREM 2.2 (LERMAN–SHORE–SOARE [5]). *There are r.e. sets A and B such that $\mathscr{L}^*(A)$ and $\mathscr{L}^*(B)$ are isomorphic to the countable atomless Boolean algebra, but A possesses an rm subset while B does not. (Hence, A and B are not automorphic or even elementarily equivalent.)*

The proof of Theorem 2.2 relies on a new classification (in terms of Δ_3 functions) of which nonrecursive r.e. sets possess rm subsets, a question arising in the extended decision procedure of Theorem 3.1. In view of Theorem 1.3 another candidate for an easily describable orbit is the class of low simple sets. However, Lerman and Soare showed that there are low simple sets which are d-simple and those which are not.

Thus, the Post style classification of A in terms of $\bar{A}$ or $\mathscr{L}(A)$ which has predominated for 30 years is seen to be increasingly inadequate for determining the orbit of A. Rather one must examine properties relating $\bar{A}$ to A such as d-simplicity and rm subsets. These will give necessary conditions and hopefully the automorphism method of Theorem 2.1 (which relies on the fact that maximal sets possess a strong d-simplicity property) will prove the conditions sufficient.

3. The elementary theory of $\mathscr{E}$. One of the most important open questions on $\mathscr{E}$ is the decidability of its elementary theory. Lachlan proved that the theories of $\mathscr{E}$ and $\mathscr{E}^*$ are equi-decidable, and gave a decision procedure for the $\vec{\forall}\vec{\exists}$-sentences of the theory of $\mathscr{E}^*$. Lerman and Soare extended this by adding new relations. The aim is to add enough additional relations to give a decision procedure first for the $\vec{\exists}\vec{\forall}\vec{\exists}$-sentences and then perhaps for all sentences.

Let $\mathscr{A}$ be the Boolean algebra generated by $\mathscr{E}$. Let L be the first order language which has function symbols $\cup$, $\cap$, $'$, and a constant symbol 0 to be interpreted in a Boolean algebra as join, meet, complement, and least element respectively, and which has a unary predicate symbol $E(x)$ to be interpreted over $\mathscr{A}^*$ as "$x \in \mathscr{E}^*$". An $\vec{\forall}\vec{\exists}$-sentence in this language is one of the form $(\forall\vec{x})(\exists\vec{y})\,P(\vec{x}, \vec{y})$ with P quantifier free. Lachlan [3] showed there is an algorithm for deciding which $\vec{\forall}\vec{\exists}$-sentences of L are true in $\mathscr{A}^*$ when quantifiers range over $\mathscr{E}^*$. The next step in the decision procedure for $\mathscr{E}^*$ is to consider the $\vec{\exists}\vec{\forall}\vec{\exists}$-sentences of L. The most reasonable attack seems to be to expand L by adding new predicates. Let L^+ be the result of adding to L the predicates $\mathrm{Max}(x)$ and $Hhs(x)$ to be interpreted in $\mathscr{E}^*$ as "x is maximal" and "x is hh-simple". Many new statements become $\vec{\forall}\vec{\exists}$ in L^+ such as "there exists an atomless hh-simple set with an rm subset", or "there exists an atomless r-maximal set". Thus, in addition to a more complicated version of Lachlan's refinement method new structural theorems were required to prove

THEOREM 3.2 (LERMAN–SOARE [6]). *There is an algorithm for deciding which $\vec{\forall}\vec{\exists}$-sentences of L^+ are true in $\mathscr{A}^*$ when quantifiers range over $\mathscr{E}^*$.*

Further information is given by classifying the elementary theory of intervals $\mathscr{L}(A, B) = \{W: W \in \mathscr{E}$ and $A \subseteq W \subseteq B\}$. Stob [18] has shown that if A is a major subset of B then the $\vec{\forall}\vec{\exists}$-theory of $\mathscr{L}(A, B)$ is decidable and indeed independent of A and B. The next structural theorems to be proved about $\mathscr{E}$ will be those needed for further steps in the decision procedure and will not be merely random facts.

4. Relative enumerability. For any degree $\boldsymbol{b}$ let $\boldsymbol{R}(\boldsymbol{b})$ denote the set of degrees $\boldsymbol{a} \geqslant \boldsymbol{b}$ such that $\boldsymbol{a}$ is r.e. relative to $\boldsymbol{b}$. If $\boldsymbol{b}$ is low then $\boldsymbol{R}(\boldsymbol{b})$ and $\boldsymbol{R}$ have the same maximum element $\boldsymbol{0}'$ but any r.e. degree $\boldsymbol{b} > \boldsymbol{0}$ allows us to obtain new relative r.e. degrees in $\boldsymbol{R}(\boldsymbol{b})$.

THEOREM 4.1. *For every r.e. degree $\boldsymbol{b} > \boldsymbol{0}$ there is a degree $\boldsymbol{a} \in \boldsymbol{R}(\boldsymbol{b}) - \boldsymbol{R}$, and $\boldsymbol{a}$ can be found uniformly in $\boldsymbol{b}$.*

The proof uses a tree of nested strategies construction (as in some recent arguments by Lachlan) to combine the strategies for meeting individual requirements. Since the construction relativizes to any degree $\boldsymbol{d}$ this negatively answers a conjecture of Cooper that for every high degree $\boldsymbol{d}$ any degree $\boldsymbol{a} \geqslant \boldsymbol{0}'$ and r.e. in $\boldsymbol{0}'$ is r.e. in $\boldsymbol{d}$. This raises the question of what special role (if any) $\boldsymbol{0}'$ plays in $\boldsymbol{R}(\boldsymbol{d})$ (in the non-trivial case $\boldsymbol{d} < \boldsymbol{0}'$ and $\boldsymbol{d}' > \boldsymbol{0}'$). Also open is the homogeneity question of which r.e. degrees $\boldsymbol{d}$ satisfy $\boldsymbol{R}(\boldsymbol{d})$ isomorphic (or even elementarily equivalent) to $\boldsymbol{R}$.

References

1. W. W. Boone, *Word problems and recursively enumerable degrees of unsolvability. A sequel on finitely presented groups*, Ann. of Math. (2) **84** (1966), 49—84.

2. K. Gödel, *Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme*. I, Monatsch. Math. Phys. **38** (1931), 173—178.

3. A. H. Lachlan, *The elementary theory of recursively enumerable sets*, Duke Math. J. **35** (1968), 123—146.

4. ——, *Degrees of recursively enumerable sets which have no maximal superset*, J. Symbolic Logic **33** (1968), 431—443

5. M. Lerman, R. A. Shore, and R. I. Soare, *r-maximal major subsets*, Israel J. Math. **31** (1978), 1—18.

6. M. Lerman and R. I. Soare, *A decidable fragment of the elementary theory of the recursively enumerable sets*, Trans. Amer. Math. Soc. (to appear).

7. ——, *d-simple sets, small sets, and degree classes* (to appear).

8. D. A. Martin, *Classes of recursively enumerable sets and degrees of unsolvability*, Z. Math. Logik Grundlagen Math. **12** (1966), 295—310.

9. Y. Matiyasevič, *Diophantine representation of enumerable predicates*, Izv. Akad. Nauk SSSR Ser. Math. **35** (1971), 3—30. (Russian)

10. E. L. Post, *Recursively enumerable sets of positive integers and their decision problems*, Bull. Amer. Math. Soc. **50** (1944), 284—316.

11. J. R. Shoenfield, *Degrees of classes of r.e. sets*, J. Symbolic Logic **41** (1976), 695—696.

12. R. I. Soare, *Automorphisms of the lattice of recursively enumerable sets. Part I: Maximal Sets*, Ann. of Math. **100** (1974), 80—120.

13. ——— *Computational complexity, speedable and levelable sets*, J. Symbolic Logic **42** (1977), 545—563.

14. ——— *Recursively enumerable sets and degrees*, Bull. Amer. Math. Soc. **84** (1978), 1149—1181.

15. ——— *Recursively enumerable sets*, Omega Series, Springer-Verlag, Berlin and New York (to appear).

16. ——— *Automorphisms of the lattice of recursively enumerable sets II. Low sets* (to appear).

17. ——— *Relative enumerability* (in preparation).

18. M. Stob, Doctoral Dissertation, University of Chicago, Chicago, Illinois, 1979.

UNIVERSITY OF CHICAGO
CHICAGO, ILLINOIS 60637, U.S.A

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

A Survey of the Classification Program for Finite Simple Groups of Even Characteristic

Michael Aschbacher

We begin with an example that will illustrate some of the concepts which will be discussed here. Consider the general linear group $G=\mathrm{GL}_n(p^m)$. Recall this is the group of all n-by-n nonsingular matrices over the finite field of order p^m. It will be informative to consider two types of elements in G, the *unipotent* and *semisimple* elements. Here a matrix x is unipotent if $x-I$ is nilpotent and x is semisimple if it is diagonalizable. The unipotent elements are *p-elements*; that is their order is a power of the characteristic p. The semisimple elements are *p′-elements*. Their order is relatively prime to p.

Let us consider the centralizer of a unipotent element u and a semisimple element s in G.

$$s=\begin{pmatrix} aI_r & 0 \\ 0 & bI_{n-r}\end{pmatrix},\quad a\neq b;$$

$$C_G(s)=\left\{\begin{pmatrix} A & 0 \\ 0 & B\end{pmatrix}: A\in \mathrm{GL}_r(p^m),\ B\in \mathrm{GL}_{n-r}(p^m)\right\};$$

$$C_G(s)\cong \mathrm{GL}_r(p^m)\times \mathrm{GL}_{n-r}(p^m);$$

$$u=\begin{pmatrix} I_r & 0 & 0 \\ 0 & I_{n-2r} & 0 \\ I_r & 0 & I_r\end{pmatrix};$$

$$C_G(u)=\left\{\begin{pmatrix} A & 0 & 0 \\ P & B & 0 \\ Q & R & A\end{pmatrix}: P, Q, R \text{ arbitrary},\ A\in \mathrm{GL}_r(p^m),\ B\in \mathrm{GL}_{n-r}(p^m)\right\};$$

$$O_p(C_G(u)) = \left\{ \begin{pmatrix} I & 0 & 0 \\ P & I & 0 \\ Q & R & I \end{pmatrix} : P, Q, R \text{ arbitrary} \right\} = \textit{unipotent radical} \text{ of } C_G(u);$$

$$C(u)/O_p(C(u)) \cong \mathrm{GL}_r(p^m) \times \mathrm{GL}_{n-2r}(p^m) \quad \text{and}$$

$$C(u) \cap C(O_p(C(u))) \leq O_p(C(u)).$$

Here $O_p(H)$ is the largest normal p-subgroup of H.

This illustrates the dicotomy between centralizers of unipotent and semisimple elements. The centralizers of semisimple elements are essentially the product of almost simple subgroups while the centralizers of unipotent elements are dominated by a large unipotent radical. This is made precise in Professor Gorenstein's talk at this Congress.

We will return to this example from time to time.

Most analysis of a finite simple group G concentrates on the *local subgroups* of G. Given a prime p, a *p-local subgroup* of G is the normalizer of a nontrivial p-subgroup of G. The centralizers of elements of prime order are of particular importance. The prime 2 plays a special role.

The fundamental principal underlying this analysis is that when the centralizer H of an element of prime order p is essentially the product of almost simple subgroups then G can be determined from this centralizer. In this talk such elements will be called *semisimple*. On the other hand if $C_H(O_p(H)) \leq O_p(H)$ then $O_p(H)$ is large and its structure is probably difficult to pin down. As a result it is usually impossible in this case to use H to determine G.

With these thoughts in mind and recalling the special role of the prime 2, we are led to the following definition: A finite group G is of characteristic 2-type if $C_G(O_2(M)) \leq O_2(M)$ for each 2-local subgroup M of G.

It is use ful to think of $\mathrm{GL}_n(p^m)$ as the typical finite simple group (even though it is not simple). As our examples indicate, $\mathrm{GL}_n(p^m)$ is of characteristic 2-type when $p=2$ but usually not otherwise. If G is not of characteristic 2-type one shows that some involution is semisimple and hence determines G. This is discussed in Professor Gorenstein's talk.

We wish to classify the simple groups of characteristic 2-type. With the fundamental principal in mind we proceed by attempting to show that some element of odd prime order is semisimple. It seems however that this is feasible only when G is of sufficiently high rank. This rank will be defined in a second. It approximates the Lie rank of G when G is of Lie type. The rank of $\mathrm{GL}_n(2^m)$ is usually $n-1$.

Let p be a prime. The *p-rank* of G, denoted by $m_p(G)$, is the maximum dimension of an abelian p-subgroup of G of exponent p, considered as a vector space over the field of order p. The *2-local p-rank* of G, denoted by $m_{2,p}(G)$, is the maximum p-rank among 2-local subgroups of G. $e(G)$ is the maximum of $m_{2,p}(G)$ as p ranges over all odd primes. $e(G)$ is the *rank* of G.

It turns out that when G has rank at least 3 that certain methods are applicable which do not seem to work for groups of rank at most 2. The latter groups are called *quasithin*. Groups of rank one are called *thin*. The thin groups are precisely the groups in which all 2-locals have cyclic Sylow p-subgroups for each odd prime p. Thin groups were classified by the author in [3] extending work of Janko [**8**] and Thompson [**14**].

THEOREM. *Let G be a thin finite simple group. Then G is $L_2(q)$, q a prime power, $L_3(p), p=1+2^a3^b$, $U_3(p)$, $p=-1+2^a3^b$, p prime, $b=0$ or* 1, $\mathrm{Sz}(2^n)$, $U_3(2^n)$, $L_3(4)$, ${}^2F_4(2)'$, ${}^3D_4(2)$, M_{11}, *or* J_1.

Several mathematicians are at present working toward a classification of the quasithin groups.

Let us now turn to groups of characteristic 2-type and rank at least 3. Let $\mathscr{K}$ be the set of all known simple groups. G is said to be *$\mathscr{K}$-group* if the composition factors of each subgroup of G are in $\mathscr{K}$. We wish to show all simple groups are in $\mathscr{K}$. In a minimal counterexample to such a theorem, all proper subgroups will be $\mathscr{K}$-groups. We consider such a counter example and assume it be of characteristic 2-type. The proof of the following theorem is not yet in final form. The case $e(G)>3$ is due to Gorenstein and Lyons. The case $e(G)=3$ is due to the author.

THEOREM. *Let G be simple of characteristic* 2*-type and rank at least* 3. *Assume all proper subgroups of G are $\mathscr{K}$-groups. Then one of the following holds:*

(1) *G possesses elements of odd prime order which are semisimple.*

(2) *G possesses a* 2*-local M such that $O_2(M)$ has no noncyclic characteristic abelian subgroups.*

(3) *For many odd primes p, G possesses an almost strongly p-embedded subgroup.*

Not much will be said about case (1) here. The fundamental principal suggests this case should be doable.

In our example above if $G=\mathrm{GL}_n(2)$ and u is chosen with $r=1$ (that is u is a transvection) then $C_G(u)$ satisfies the conclusion of case (2). Indeed most of the groups of Lie type over the field of order 2 and about half of the sporadic simple groups have a 2-local of this type. The following theorem classifies such groups. Probably the largest contribution to its proof is from Timmesfeld [**15**]. Other contributions were made by Aschbacher [**1**], [**2**], Dempwolf and S. Wong [**4**] Gorenstein and Harada [**6**], F. Smith [**9**], [**10**], S. Smith [**11**], [**12**], [**13**] and Thompson.

THEOREM. *Let G be a finite simple group possessing a* 2*-local M such that $C_M(O_2(M))\leq O_2(M)$ and $O_2(M)$ has no noncyclic characteristic abelian subgroups. Then G is a group of Lie type over the field of order* 2, $G_2(3)$, $U_4(3)$, $L_4(3)$, $\Omega_8^+(3)$, A_9, $L_2(2^n\pm1)$, M_{11}, M_{12}, M_{24}, HS, He, Sz, J_2, J_3, J_4, Co_2, Co_1, F_{24}, F_5, F_3, F_2, *or* F_1.

Notice here that $Z(M)$ is of order 2, so that M is the centralizer of an involution and $O_2(M)$ is large. Still G can be determined because the structure of $O_2(M)$ is precisely determined by a theorem of P. Hall [7].

I will refer to case (3) as the *uniqueness case*. A subgroup H of G is *strongly p-embedded* if p divides the order of H but not the order of $H \cap H^g$ for $g \in G-H$. H^g is the conjugate of H under g. Almost strong embedding is an approximation of strong embedding I will not define. Groups with a strongly 2-embedded subgroup have been classified by H. Bender but the proof depends on very special properties of the prime 2. For odd primes an approach similar to that used by John Thompson in section 13 of the N-group paper seems to be more profitable.

Just as the prime 2 plays a special role in the analysis of simple groups the prime 3 plays a special role in the uniqueness case. In particular in the generic situation G has a strongly 3-embedded 2-local of high 3-rank (*i.e.* 3-rank at least 4). This subcase has been handled by the author. Work is in progress on the general case.

References

1. M. Aschbacher, *On finite groups in which the generalized Fitting group of the centralizer of some involution is symplectic but not extraspecial,* Comm. Algebra **4** (1976), 595—616.

2. ——— *On finite groups in which the generalized Fitting group of the centralizer of some involution is extraspecial,* Illinois J. Math. **21** (1977), 347—364.

3. ——— *Thin finite simple groups,* J. Algebra **54** (1978), 50—152.

4. U. Dempwolf and S. Wong, *On finite groups whose centralizer of an involution has normal extraspecial and abelian subroups,* I, II, J. Algebra **45** (1977), 247—253; **52** (1978), 210—217.

5. D. Gorenstein, *The classification of finite simple groups,* these Proceedings, pp. 129—137.

6. D. Gorenstein and K. Harada, *Finite groups whose 2-subgroups are generated by at most* 4 *elements,* Mem. Amer. Math. Soc. No. 147, 1974.

7. P. Hall, *A contribution to the theory of groups of prime power order,* Proc. London Math. Soc. **36** (1933), 29—95.

8. Z. Janko, *Nonsolvable finite groups all of whose 2-local subgroups are solvable,* I, J. Algebra **21** (1972), 458—517.

9. F. Smith, *On a result of M. Aschbacher,* Comm. Algebra **51** (1977), 279—288.

10. ——— *On finite groups with large extraspecial* 2*-subgroups,* J. Algebra **44** (1977), 477—487.

11. S. Smith, *A characterization of orthogonal groups over* GF (2) (to appear).

12. ——— *A characterization of finite Chevalley and twisted groups of type* E *over* GF (2) (to appear).

13. ——— *Large extraspecial subgroups of widths* 4 *and* 6 (to appear).

14. J. Thompson, *Nonsolvable finite groups all of whose local subgroups are solvable,* IV, Pacific J. Math. **48** (1973), 511—592.

15. F. Timmesfeld, *Finite groups in which the generalized Fitting group of the centralizer of some involution is extraspecial,* Ann. of Math. **107** (1978), 297—369.

California Institute of Technology
Pasadena, California 91125, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Cohomology of Infinite Groups

Kenneth S. Brown*

This is a survey of recent results in the cohomology theory of infinite groups, with emphasis on the theory of groups of finite virtual cohomological dimension. (Recall from **[24]** that if Γ is a group which has torsion-free subgroups of finite index, then all such subgroups have the same cohomological dimension; this common dimension is called the *virtual cohomological dimension* of Γ and denoted vcd Γ.)

1. Euler characteristics. 1.1. If Γ is a group such that $H_i(\Gamma, \mathbf{Q})$ is finite dimensional over $\mathbf{Q}$ for all i and is trivial for all but finitely many i, then we set $\tilde{\chi}(\Gamma)=\sum_i(-1)^i \dim H_i(\Gamma, \mathbf{Q})$. We will say that a group Γ has *finite homological type* if (i) vcd $\Gamma<\infty$ and (ii) $H_*(\Gamma', \mathbf{Z})$ is finitely generated for every torsion-free subgroup Γ' of finite index. We then define the *Euler characteristic* $\chi(\Gamma)\in\mathbf{Q}$ by $\chi(\Gamma)=\tilde{\chi}(\Gamma')/(\Gamma:\Gamma')$, where Γ' is any such subgroup; it is shown in **[10]** that this is independent of the choice of Γ'. It agrees with the Euler characteristic studied by Wall **[39]** and Serre **[24]** if Γ is of "type (VFL)".

1.2. It is immediate from the definition that $d\cdot\chi(\Gamma)\in\mathbf{Z}$, where d is the greatest common divisor of the indices of the torsion-free subgroups Γ' of finite index. But one can, in fact, prove the sharper result that $m\cdot\chi(\Gamma)\in\mathbf{Z}$, where m is the least common multiple of the orders of the finite subgroups of Γ (cf. **[10]** or **[13]**). In addition, there are a number of formulas which yield more precise information about $\chi(\Gamma)$ in terms of the torsion in Γ. For example, let Ψ be a set of representatives for the conjugacy classes of elements of Γ of finite order, and assume for

* Partially supported by a grant from the National Science Foundation of the United States.

each $s \in \Psi$ that the centralizer $Z(s)$ is of finite homological type. Then one can prove that Ψ is finite and that

$$(*) \qquad \tilde{\chi}(\Gamma) = \sum_{s \in \Psi} \chi(Z(s)).$$

(More generally, if $\Gamma' \subseteq \Gamma$ is an arbitrary normal subgroup of finite index, then there is a Lefschetz number formula for the action of Γ/Γ' on $H_*(\Gamma', \mathbf{Q})$, cf. [**14**, § 6]; $(*)$ is the special case $\Gamma' = \Gamma$.) In particular, since $\tilde{\chi}(\Gamma) \in \mathbf{Z}$, we obtain $\chi(\Gamma) \equiv -\sum_{s \in \Psi'} \chi(Z(s)) \bmod \mathbf{Z}$, where $\Psi' = \Psi - \{1\}$; this can be regarded as a formula for the "fractional part" of $\chi(\Gamma)$ in terms of the torsion in Γ.

There is also a formula for the "p-fractional part" of $\chi(\Gamma)$, where p is a prime ([**11**], [**23**]; see also [**13**]): Let $\mathscr{A}_p$ be the set of nontrivial elementary abelian p-subgroups of Γ. [An elementary abelian p-group is a group isomorphic to $(\mathbf{Z}_p)^r$ for some $r < \infty$, where $\mathbf{Z}_p = \mathbf{Z}/p\mathbf{Z}$.] If the normalizer $N(A)$ has finite homological type for each $A \in \mathscr{A}_p$, then $\chi(\Gamma) \equiv \chi_\Gamma(\mathscr{A}_p) \bmod \mathbf{Z}_{(p)}$, where $\mathbf{Z}_{(p)}$ denotes $\mathbf{Z}$ localized at p and $\chi_\Gamma(\mathscr{A}_p)$ is an "equivariant Euler characteristic". Moreover, one can show that the latter is given by

$$\chi_\Gamma(\mathscr{A}_p) = \sum_{r \geq 1} (-1)^{r-1} p^{r(r-1)/2} \sum_{A \in \mathscr{A}_p^r} \chi(N(A)),$$

where $\mathscr{A}_p^r$ is a set of representatives for the conjugacy classes of elementary abelian p-subgroups of Γ of rank r. [Our hypothesis implies that there are only finitely many such conjugacy classes.]

The results described above have applications to group theory and number theory ([**10**], [**11**]), as well as to the study of the finite subgroups of the exceptional Chevalley groups over $\mathbf{Z}$ [**26**].

1.3. Suppose now that Γ is a group such that $\mathbf{Q}$, regarded as a module over the group algebra $\mathbf{Q}\Gamma$, admits a projective resolution of finite length, $0 \to P_n \to \cdots \to P_0 \to \mathbf{Q} \to 0$, with each P_i finitely generated. (Γ is then said to be of type (FP) over $\mathbf{Q}$.) We then set (cf. [**33**]) $E(\Gamma) = \sum (-1)^i r(P_i)$, where $r(\ )$ denotes the Hattori–Stallings rank. This "complete Euler characteristic" is a $\mathbf{Q}$-linear combination of Γ-conjugacy classes. We denote by $e(\Gamma)$ the coefficient of the conjugacy class of 1; this is the Euler characteristic of Γ in the sense of [**3**], [**15**], and [**34**]. Like the Euler characteristic χ defined in 1.1 above, e agrees with the Wall–Serre Euler characteristic if Γ is of type (VFL). It is not known whether $e(\Gamma) = \chi(\Gamma)$ whenever both are defined, but this is easily seen to be true if Γ is residually finite [**3**]; more generally, they are equal if Γ has a subgroup Γ' of finite index such that $E(\Gamma')$ is concentrated at the conjugacy class of 1. A related question is whether $e(\Gamma) = \tilde{\chi}(\Gamma)$ whenever Γ is torsion-free and of type (FP) over $\mathbf{Q}$. This is known to be true by results of Bass [**3**] if Γ satisfies a certain "condition D", which holds for instance if Γ is a linear group.

1.4. Bass's results imply further that $E(\Gamma)$ is supported on the conjugacy classes of elements of finite order if Γ is of type (FP) over $\mathbf{Q}$ and satisfies condition D.

Additional results about $E(\Gamma)$ can be obtained by using the methods of [10]. One can prove, for example (cf. [14]), under suitable hypotheses on Γ, the following formula suggested by Serre:

$$(**) \qquad E(\Gamma) = \sum_{s\in\Psi} e(Z(s))\cdot[s],$$

where Ψ is as in 1.2 and $[s]$ is the conjugacy class of s. This should be thought of as a refinement of the formula $(*)$ above. Indeed, if $(**)$ holds then one easily deduces $(*)$, but with χ replaced by e.

The hypotheses on Γ under which $(**)$ has been proved are quite complicated, but we can describe a large family $\mathscr{F}$ of examples for which $(**)$ has been proved, as follows. Let $\mathscr{F}_0$ be the class of finite groups; assuming $\mathscr{F}_{n-1}$ has been defined, let $\mathscr{F}_n$ be the class of groups Γ which admit a simplicial action on a complex X such that (i) X/Γ is compact, (ii) the isotropy group Γ_σ is in $\mathscr{F}_{n-1}$ for each simplex σ of X, and (iii) the fixed-point set X^s is contractible for each $s\in\Gamma$ of finite order. Then $\mathscr{F}_0\subset\mathscr{F}_1\subseteq\mathscr{F}_2\subseteq\ldots$, and we set $\mathscr{F}=\bigcup\mathscr{F}_n$. The family $\mathscr{F}$ includes all arithmetic groups (which are in $\mathscr{F}_1$ as a consequence of [7]), as well as the S-arithmetic groups in the reductive case (these are in $\mathscr{F}_2$, cf. [8, § 6]. I do not know an algebraic characterization of $\mathscr{F}$, nor do I know any examples of groups of type (FP) over $\boldsymbol{Q}$ which are not in $\mathscr{F}$.

2. Farrell cohomology. F. T. Farrell [17] has shown that the Tate cohomology theory for finite groups can be extended to the class of groups Γ such that vcd $\Gamma<\infty$. Farrell's theory yields cohomology groups $\hat{H}^i(\Gamma)$ $(i\in\boldsymbol{Z})$, such that $\hat{H}^i=H^i$ for $i>$vcd Γ. If Γ is a "virtual duality group", then one can describe $\hat{H}^i$ for $i<-1$ as a homology functor $\tilde{H}_{n-i-1}=H_{n-i-1}(\Gamma, D\otimes_Z -)$, where $n=$vcd Γ and D is the Γ-module $H^n(\Gamma, Z\Gamma)$; moreover, there is an exact sequence relating $\{\hat{H}^i\}_{-1\le i\le n}$, $\{H^i\}_{0\le i\le n}$, and $\{\tilde{H}_i\}_{0\le i\le n}$ (cf. [17], [13]). This exact sequence generalizes the sequence $0\to\hat{H}^{-1}\to H_0\xrightarrow{N} H^0\to\hat{H}^0\to 0$ which one has if Γ is finite, where N is the "norm map". (*Note:* If Γ is finite then $n=0$ and $D=\boldsymbol{Z}$, with trivial Γ-action.) The Farrell cohomology groups are all torsion groups. In fact, if d and m are the integers defined in 1.2, then $d\cdot\hat{H}^*(\Gamma)=0$, but it is not known whether one always has $m\cdot\hat{H}^*(\Gamma)=0$.

It is shown in [12] and [13] that a great deal of information about $\hat{H}^*(\Gamma)$ (and hence about $H^i(\Gamma)$ for $i>$vcd Γ) can be extracted from the finite subgroups of Γ. For example, $\hat{H}^*(\Gamma)$ is periodic if and only if every finite subgroup of Γ has periodic cohomology in the usual sense. (This improves a result of Venkov [36].) Similarly, if p is a prime then the p-primary component $\hat{H}^*(\Gamma)_{(p)}$ is periodic if and only if $\hat{H}^*(G)_{(p)}$ is periodic for every finite subgroup $G\subseteq\Gamma$, i.e., if and only if Γ contains no subgroups isomorphic to $\boldsymbol{Z}_p\times\boldsymbol{Z}_p$. Another result, analogous to that described in 1.2 on the p-fractional part of the Euler characteristic, is that $\hat{H}^*(\Gamma)_{(p)}\approx\hat{H}^*_\Gamma(\mathscr{A}_p)_{(p)}$, the latter being "equivariant Farrell cohomology". If Γ contains no subgroups isomorphic to $\boldsymbol{Z}_p\times\boldsymbol{Z}_p$ (i.e., if $\hat{H}^*(\Gamma)_{(p)}$ is periodic), this isomorphism takes the simple form $\hat{H}^*(\Gamma)_{(p)} \approx \prod_{P \in \mathscr{P}} \hat{H}^*(N(P))_{(p)}$, where $\mathscr{P}$ is a set of representatives for the conjugacy classes of subgroups of order p. See **[22]** for earlier results relating the cohomology of Γ to the elementary abelian p-subgroups.

3. Cohomology calculations. The proofs of the results described in §2 are based on the fact, due to Serre **[24**, 1.7], that if $\operatorname{vcd} \Gamma < \infty$ then there exists a contractible finite-dimensional space X on which Γ acts properly (and hence with finite isotropr, groups). The arguments are of a general nature. For a given group Γ, howeveyl one can often get more precise information about $H^*(\Gamma)$ by choosing X convenienty and making a more detailed analysis.

Consider, for example, the case $\Gamma = SL_n(\boldsymbol{Z})$. Classically one takes X to be the symmetric space $\mathrm{SL}_n(\boldsymbol{R})/SO_n(\boldsymbol{R})$, which can be identified with the space of positive definite real quadratic forms in n variables, modulo multiplication by positive scalars. This choice of X, however, is inconvenient for calculation because $\Gamma \backslash X$ is noncompact. One way to remedy this is to replace X by its Borel–Serre "bordification" $\bar{X}$ **[7]**. This was done, for example, by Lee and Szczarba **[20]**, who were thereby able to completely compute the integral cohomology of the principal congruence subgroup of level 3 of $SL_3(\boldsymbol{Z})$. The space $\bar{X}$ was also used by Lee **[19]** in his construction of several families of "unstable" elements of $H^*(SL_n(\boldsymbol{Z}), \boldsymbol{R})$, i.e., cohomology classes which do not come from $H^*(SL(\boldsymbol{Z}), \boldsymbol{R})$. (Recall that the latter was computed by Borel **[5]**; it is an exterior algebra with one generator of degree $4i+1$ for each integer $i \geqslant 1$.)

A different approach is to replace X by a contractible $SL_n(\boldsymbol{Z})$-invariant subspace X' with compact quotient $SL_n(\boldsymbol{Z}) \backslash X'$. Soulé (**[27]**, **[31]**) and Ash (**[1]**, **[2]**; see also **[13**, § 2, Ex. 5]) have shown that there always exists such an X' of dimension $n(n-1)/2$; this had previously been observed by Serre **[25]** in the case $n=2$. (We remark that $\operatorname{vcd} SL_n(\boldsymbol{Z}) = n(n-1)/2$, so X' has the smallest possible dimension for a contractible space on which $SL_n(\boldsymbol{Z})$ acts properly.) The most striking result obtained in this way is the complete calculation by Soulé **[27]** of $H^*(SL_3(\boldsymbol{Z}), \boldsymbol{Z})$. This was achieved by using an explicit cell-decomposition of X' in order to compute the spectral sequence of equivariant cohomology theory (cf. **[18]** or **[22]**)

$$E_2^{pq} = H^p(\Gamma \backslash X', \mathscr{H}_\Gamma^q) \Rightarrow H^{p+q}(\Gamma).$$

(Here $\mathscr{H}_\Gamma^q$ is a certain sheaf on $\Gamma \backslash X'$ whose stalks are the groups $H^q(\Gamma_x)$, where $x \in X'$ and Γ_x is the isotropy group of x.)

Still a third method was used by Lee and Szczarba **[21]** to partially compute $H^*(SL_n(\boldsymbol{Z}))$ for $n=4$ and 5. They replaced X by an enlargement X^* due to Voronoi **[37]**, which comes equipped with a cell-decomposition compatible with the $SL_n(\boldsymbol{Z})$-action. Their calculations were pushed further by Soulé (**[29]**, **[31]**). Similar methods have been applied in **[32]** to the group $SL_3(\boldsymbol{Z}[\sqrt{-1}])$.

Further information on the cohomology of $SL_n(\boldsymbol{Z})$ and other arithmetic groups has been obtained by Eckmann [private communication] and Soulé (**[28]**, **[30]**, **[31]**;

see also [16], [35]) by studying characteristic classes. In particular, many interesting examples of torsion classes in $H^*(SL_n(Z), Z)$ have been obtained in this way.

4. Further results. I have, of course, had to omit many topics from this survey. In particular, I would like to call attention to: (a) the work of Bieri and others on cohomological dimension, duality groups, and related matters (see [4] and the references cited there); (b) stability theorems of Quillen (unpublished), Wagoner [38], and R. Charney [unpublished] for $H_*(GL_n(R))$ for suitable rings R; and (c) connections between cohomology and representation theory for discrete subgroups of Lie groups ([6], [9], [40]).

Finally, the reader is referred to the forthcoming proceedings of the 1977 Durham conference on homological and combinatorial techniques in group theory (C. T. C. Wall, ed.) for additional references and a list of open problems.

References

1. A. Ash, *Deformation retracts with lowest possible dimension of arithmetic quotients of self-adjoin homogeneous cones*, Math. Ann. **225** (1977), 69—76.

2. ——— *Cohomology of subgroups of finite index of* SL (3, Z) *and* SL (4, Z), Bull. Amer. Math. Soc. **83** (1977), 367—368.

3. H. Bass, *Euler characteristics and characters of discrete groups*, Invent. Math. **35** (1976), 155—196.

4. R. Bieri, *Homological dimension of discrete groups*, Queen Mary College Mathematics Notes, London, 1976.

5. A. Borel, *Stable real cohomology of arithmetic groups*, Ann. Sci. École Norm. Sup. (4) **7** (1974), 235—272.

6. ——— *Cohomologie de sous-groupes discrets et représentations de groupes semi-simple*, Astérisque **32—33** (1976), 73—112.

7. A. Borel and J.-P. Serre, *Corners and arithmetic groups*, Comment. Math. Helv. **48** (1974), 244—297.

8. ——— *Cohomologie d'immeubles et de groupes S-arithmétiques*, Topology **15** (1976), 211—232.

9. A. Borel and N. R. Wallach, *Cohomology of discrete subgroups of semi-simple groups*, Seminar IAS, Princeton, 1976; Ann. of Math. Studies (to appear).

10. K. S. Brown, *Euler characteristics of discrete groups and G-spaces*, Invent. Math. **27** (1974), 229—264.

11. ——— *Euler characteristics of groups: The p-fractional part*, Invent. Math. **29** (1975), 1—5.

12. ——— *High dimensional cohomology of discrete groups*, Proc. Nat. Acad. Sci. U.S.A. **73** (1976), 1795—1797.

13. ——— *Groups of virtually finite dimension*, Proc. 1977 Durham Conf. on Homological and Combinatorial Techniques in Group Theory, Homological Group Theory (C. T. C. Wall, ed.) Cambridge Univ. Press (to appear).

14. ——— *Complete Euler characteristics and fixed-point theory*, preprint.

15. I. M. Chiswell, *Euler characteristics of groups*, Math. Z. **147** (1976), 1—11.

16. L. Evens and D. Kahn, *Chern classes of certain representations of symmetric groups*, Trans. Amer. Math. Soc. **245** (1978), 309—330.

17. F. T. Farrell, *An extension of Tate cohomology to a class of infinite groups*, J. Pure Appl. Algebra **10** (1977), 153—161.

18. A. Grothendieck, *Sur quelques points d'algèbre homologique*, Tôhoku Math. J. **9** (1957), 119—221.

19. R. Lee, *On unstable cohomology classes of* $\mathrm{SL}_n(\boldsymbol{Z})$, Proc. Nat. Acad. Sci. U.S.A. **75** (1978), 43—44.

20. R. Lee and R. H. Szczarba, *On the homology and cohomology of congruence subgroups*, Invent. Math. **33** (1976), 15—53.

21. ———, *On the torsion in* $K_4(\boldsymbol{Z})$ *and* $K_5(\boldsymbol{Z})$, Duke Math. J. **45** (1978), 101—129.

22. D. Quillen, *The spectrum of an equivariant cohomology ring*, I, II, Ann. of Math. **94** (1971), 549—572; 573—602.

23. ———, *Homotopy properties of the poset of non-trivial p-subgroups of a group*, Advances in Math. **28** (1978), 101—128.

24. J.-P. Serre, *Cohomologie des groupes discrets*, Ann. of Math. Studies, vol. 70, Princeton Univ. Press, Princeton, N. J., 1971, pp. 77—169.

25. ———, *Arbres, amalgames*, SL_2, Astérisque **46** (1977).

26. ———, *Arithmetic groups*, Proc. 1977 Durham Conf. on Homological and Combinatorial Techniques in Group Theory, Homological Group Theory (C. T. C. Wall, ed.) Cambridge Univ. Press (to appear).

27. C. Soulé, *The cohomology of* $\mathrm{SL}_3(\boldsymbol{Z})$, Topology **17** (1978), 1—22.

28. ———, *Classes de torsion dans la cohomologie des groupes arithémtiques*, C. R. Acad. Sci. Paris Sér. A **284** (1977), 1009—1011.

29. ———, *Addendum to the article "On the torsion in* $K_*(\mathrm{Z})$", Duke Math. J. **45** (1978), 131—132.

30. ———, *K-théorie de* $\boldsymbol{Z}$ *et cohomologie étale*, C. R. Acad. Sci. Paris Sér. A **286** (1978), 1179—1181.

31. ———, Thèse, Université Paris VII, 1978.

32. R. E. Staffeldt, *Reduction theory and* K_3 *of the Gaussian integers*, preprint.

33. J. R. Stallings, *Centerless groups — an algebraic formulation of Gottlieb's theorem*, Topology **4** (1965), 129—134.

34. ———, *An extension theorem for Euler characteristics of groups*, preprint.

35. C. B. Thomas, *An integral Riemann-Roch formula for flat line bundles*, Proc. London Math. Soc. **34** (1977), 87—101.

36. B. B. Venkov, *On homologies of groups of units in division algebras*, Trudy Mat. Inst. Steklov **80** (1965), 66—89 Equals Proc. Steklov Inst. Math. **80** (1965), 73—100.

37. G. Voronoi, *Nouvelles applications des paramètres continus à la théorie des formes quadratiques*, I, J. Reine Angew. Math. **133** (1907), 97—178.

38. J. B. Wagoner, *Stability for homology of the general linear group of a local ring*, preprint.

39. C. T. C. Wall, *Rational Euler characteristics*, Proc. Cambridge Philos. Soc. **57** (1961), 182—183.

40. G. J. Zuckerman, *Continuous cohomology and unitary representations of real reductive groups*, Ann. of Math. **107** (1978), 495—516.

Cornell University
Ithaca, New York 14853, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Cohen-Macaulay Rings and Modules

Melvin Hochster*

1. Introduction. The objective of this note is to give some insight into the reasons for studying Cohen–Macaulay rings and modules, as well as some recent results concerning when rings of invariants are Cohen–Macaulay and the existence of Cohen–Macaulay modules. All proofs are omitted. More detailed treatments are available in [**Ho**$_3$], [**Ho**$_4$], [**HR**$_1$], [**HR**$_2$], [**Ke**$_4$], and [**Bou**$_3$] while expository versions are given in [**Bou**$_1$], [**Ei**], and [**Ho**$_6$]. Both the main results discussed here were first proved in char. $p>0$ and then established over fields of char. 0 by "reducing" to the char. $p>0$ case, where one may make use of the Frobenius homomorphism.

2. What does Cohen–Macaulay mean? In this section we focus primarily on the notion of "Cohen-Macaulay (C-M) ring". Suffice it to say that in all cases, analogous characterizations exist for modules.

In the sequel, all rings are commutative, associative, with identity, and all modules unital. Let R be a ring and M an R-module. A sequence $x_1, \ldots, x_n \in R$ is called *regular* on M, or an *M-sequence* if

(1) $\sum_{i=1}^{n} x_i M \neq M$ and

(2) for each $i, 0 \leq i \leq n-1$, x_{i+1} is not a zero divisor on $M/\sum_{j\leq i} x_j M$.

For any ring R, we define $\dim R$ (respectively, height P, where P is prime) to be supremum of lengths n of chains $P_n \supsetneqq \ldots \supsetneqq P_0$ of primes (respectively, of such chains with $P_n = P$). When $I \neq R$ is any ideal, height $I = \min\{\text{height } P: I \subset P,$ P prime$\}$. Note that R is never regarded as prime, while (0) is prime precisely when R is a domain.

* The author was supported by a grant from the National Science Foundation of the United States.

This definition of dimension is motivated by the fact that if X is an affine algebraic variety over C, then the dimension of $C[X]$ $(=\mathrm{Mor}\,(X, C)$, the coordinate ring of $X)$ is the same as the (complex) topological dimension of X. In this set-up height corresponds to codimension: if $P \subset C[X]$ is a prime and $Y \subset X$ is the closed subvariety where the elements of P vanish, then height $P = \dim X - \dim Y = \mathrm{codim}_X Y$.

The notion of regular sequence leads to a different, homological notion of codimension. If I is an ideal of a Noetherian ring R and M is a finitely generated R-module with $IM \neq M$, then $\mathrm{depth}_I M$ denotes the length of any maximal regular sequence on M contained in I (all have the same length).

For any ideal $I \neq R$, we then have two notions of codimension: height I (or ht I) and $\mathrm{depth}_I R$. Life turns out to be really pleasant when these two coincide:

(2.1) DEFINITION. *A Noetherian ring R is Cohen–Macaulay if for every ideal $I \neq R$, height $I = \mathrm{depth}_I R$.*

This definition does not convey any intuitive feeling for the meaning of C-M. The following list containing consequences of the definition, alternate characterizations, "examples" and theorems is intended to clarify matters. For more information see **[Ho$_3$]**, **[Ka$_1$]**, **[Ma]** and [Se].

(1) All zero-dimensional Noetherian rings, one-dimensional Noetherian domains, and two-dimensional normal ($\equiv$integrally closed) Noetherian domains are C-M. The C-M property is local, and is unaffected by completing a local ring.

(2) In a C-M ring R, $x_1, \ldots, x_n$ is an R-sequence iff height $\sum_{i=1}^{q} x_i R = q$, $q \leq n$ in which case $R/\sum_i x_i R$ is again C-M.

A Noetherian ring R is called *local* if it has a unique maximal ideal m. (The ring of germs of functions at a point of an algebraic or analytic variety is local.)

(3) Recall that a local ring (R, m) is called *regular* if, equivalently, m is generated by an R-sequence, or if for every R-module M, the projective dimension $\mathrm{pd}_R M$ is finite. (Geometrically, a point of an algebraic variety over C is smooth iff the local ring at the point is regular.) A Noetherian ring R is called *regular* if all its local rings are regular.

Regular rings *are* C-M. The most important examples are polynomial and formal and convergent power series rings over fields and principal ideal domains.

One may construct C-M rings from regular rings by killing an R-sequence. The rings so obtained are called complete intersections.

(4) In C-M rings, ideals generated by R-sequences are unmixed, i.e. all associated primes are minimal. In particular, this holds for the zero ideal.

(5) Let S be either a regular local ring or a polynomial ring over a field and let R be a ring which is a finitely generated S-module, local in the first case or graded in the second. Then R is C-M iff $\mathrm{pd}_S R$ is "as small as possible", to wit, $\dim S - \dim R$. *In particular if $S \subset R$, then R is C-M iff R is S-free.*

(6) Suppose we are in the graded case above and $R = S/I$ (S a polynomial ring over a field, I a homogeneous ideal). Suppose we have a homomorphism

$S \to T$, where T is C-M, and suppose $\operatorname{ht} IT = \operatorname{ht} I$. Then if R is C-M, so is T/IT. This is an important stability property of C-M rings.

(7) There are innumerable applications of the notion of C-M ring in algebraic geometry, e.g. a C-M ring is normal (integrally closed) iff the singular locus has codimension $\geqslant 2$. Again, in computing the intersection multiplicity of two complex varieties Y, Z in a smooth ambient space X at an isolated point x of $Y \cap Z$, if R is the local ring of X at x and I, J are the ideals of germs at x of functions vanishing on Y, Z, respectively, then if Y, Z are C-M the intersection multiplicity is simply $\dim_C R/(I+J)$. (In the general case one needs $\sum_i (-1)^i \dim_C \operatorname{Tor}_i^R (R/I, R/J)$. (See [Se].)

If $X \subset P^n$ is a projective variety and R is the homogeneous coordinate ring of X (i.e. the affine coordinate ring of the cone in $/A^{n+1}$ which is the union of the lines which represent points of X), then the condition for R to be C-M is (under mild hypotheses) the vanishing of the sheaf cohomology of certain line bundles on X (to wit, the structure sheaf and all its twistings) in the dimensions strictly between 0 and $\dim X$.

Moreover, many theorems about curves generalize in a pleasant (e.g. without spectral sequences) way precisely to C-M varieties: one such is Serre–Grothendieck duality (see [**HR**$_1$] for more details).

(8) Many examples are known of normal and even unique factorization domains which are not C-M. See [**Be**], [**Bou**$_2$], [**FG**], [**FrK**], [**Ho**$_1$], and [**L**].

(9) Let K be a field, and let Δ be an abstract finite simplicial complex with vertices $x_1, \ldots, x_n$. View $x_1, \ldots, x_n$ as indeterminates over K, let $S = K[x_1, \ldots, x_n]$ and let I_Δ be the ideal of S spanned as a K-vector space by those monomials $x_{i_1}^{t_1} \ldots x_{i_m}^{t_m}$, where each $t_j > 0$, such that $\{x_{i_1}, \ldots, x_{i_m}\} \notin \Delta$. Let $K[\Delta] = S/I_\Delta$. G. Reisner has shown [**Rei**] that $K[\Delta]$ is C-M iff whenever L is either Δ itself or the link of a simplex $\sigma \in \Delta$ $(= \{\tau \in \Delta : \tau \cap \sigma = \emptyset$ and $\tau \cup \sigma \in \Delta\})$, $\tilde{H}^i(L) = 0$ for $0 \leqslant i \leqslant \dim L - 1$, where $\tilde{H}$ is reduced simplicial cohomology with coefficients in K. R. Stanley used this fact to prove the Upper Bound Conjecture in combinatorics (concerning the maximum number of k-faces in an n vertex triangulation of a d-sphere). See [**St**$_1$], [**St**$_2$] and [**Ho**$_6$]. A key point is that the values of the Hilbert function of a graded Cohen–Macaulay ring must obey certain inequalities: applied to $K[\Delta]$ and rewritten appropriately, these yield the Upper Bound Conjecture.

3. Cohen–Macaulay rings and invariant theory. Suppose that a Zariski closed subgroup G of some $\mathrm{GL}(m, K)$, where K is an algebraically closed field, i.e. a linear algebraic group (cf. [**Bor**]), acts (by K-algebra automorphisms) on a polynomial ring $R = K[X_1, \ldots, X_n]$. Assume also, for simplicity, that the action is linear, i.e. induced by a map $\varphi : G \to \operatorname{Aut}_K V$, where $V = \sum_i KX_i$ is the vector space of 1-forms and φ is both a group homomorphism and a morphism of algebraic varieties. Hilbert's 14th problem (see [**Hu**], [**Ka**$_2$], [**Mu**$_2$]) was motivated by the question, when is R^G, the fixed ring, finitely generated over K?

We now have a reasonable answer: when G is reductive (i.e. the maximal connected normal solvable subgroup is an algebraic torus G_m^r, where G_m is the multiplicative group of K), and not usually otherwise. See **[DC]**, **[Gros]**, **[Hu]**, **[Mu$_2$]**, and **[N]**. In char. 0 the key point is the existence of the Reynolds retraction $R \to R^G$, which is a retraction as R^G-modules. (If G is finite, this retraction is obtained by averaging over G.) (The char. p case depended on proving Mumford's conjecture **[Mu$_1$]**: this was finally done by Haboush **[Ha]**.)

Once we know that R^G is finitely generated it is quite natural to ask, what other good properties does R^G inherit from R? It need not be regular, nor a UFD, but (and this is easy) it must be normal. Much less obvious is the following result from **[HR$_1$]**:

(3.1) THEOREM. *Let K be field of char.* 0, *R a Noetherian regular K-algebra, and G a reductive linear algebraic group over K acting K-rationally on R. Then R^G is a Cohen–Macaulay ring.*

A very trivial example is the action of $Z_2 = \{1, -1\} \subset K - \{0\}$ on $K[x_1, x_2]$ by $a: x \mapsto ax_i$, $i = 1, 2$. $R^G = K[x_1^2, x_1x_2, x_2^2]$. One can see the Cohen–Macaulay property easily here, since $1, x_1x_2$ is a free basis for R^G over its subring $K[x_1^2, x_2^2]$, which is a "polynomial ring". (A non-Cohen–Macaulay subring of $K[x_1, x_2]$ is exemplified by $K[x_1^2, x_1^3, x_2, x_1x_2]$. This ring is a finite module over $K[x_1^2, x_2]$, but is *not free:* $1, x_1^3, x_1x_2$ is a minimal basis, but $x_2(x_1^3) + (-x_1^2)x_1x_2 = 0$.)

A much more interesting example comes out of the action of $\mathrm{GL}(t, K)$ on the polynomial ring R in $(r+s)t$ variables, which we think of as the entries of an r by t matrix $X = (x_{ij})$ and a t by s matrix $Y = (y_{jk})$. The matrix A acts by sending the entries of X, Y to the corresponding entries of XA^{-1}, AY, respectively. R^G is generated by rs elements, the entries of the product matrix XY, and if we map the polynomial ring T in rs variables z_{ij} onto R^G, the kernel is generated by the $t+1$ size minors of the matrix $W = (z_{ij})$. See **[W]**. Once again, from (3.1) it follows that R^G is C-M.

The key point in the proof is the existence of the Reynolds retraction $R \to R^G$, which makes R^G a direct summand of R as R^G-modules. The original argument then makes use of reduction to char. $p > 0$ and the action of Frobenius on local cohomology. (Cf. **[HR$_1$]**, **[Bou$_1$]**.) A different argument along the same lines is given in **[Ke$_4$]**, where the group action is discarded completely: it is shown that if R is essentially of finite type over a field of char. 0 and regular, and S is a direct summand of R as R-modules, then S is C-M. (This is shown in char. $p > 0$ in **[HR$_1$]**: in fact, one only needs that R is Noetherian regular.) Quite recently, J.–F. Boutot **[Bou$_3$]** has improved the theorem in the geometric case as follows:

(3.2) THEOREM (BOUTOT). *Let R be a ring essentially of finite type over a field of* char. 0, *and suppose S is a direct summand of R as S-modules. Then if R has rational singularities, so does S.*

If $X \to^f \operatorname{Spec} R$ is any desingularization, where R is normal and the field has char. 0, then R has rational singularities means $R^i f_* \mathcal{O}_X = 0$, $i \geq 1$. Having rational singularities is a property intermediate between regularity, which is stronger, and being normal, C-M, which is weaker. For more information, see **[KKMS]**, **[Ke$_1$]**, **[Ke$_2$]**, **[Ke$_3$]**, and, of course, **[Bou$_3$]**. Boutot's argument is quite short, but rests on some deep and difficult results: resolution of singularities **[Hi]** and the Grauert–Riemenschneider generalization **[GraR]** of Kodaira vanishing.

In char. $p>0$, Theorem (3.1) sometimes fails, and there is no satisfactory result which accounts for the many cases where (3.1) is known to be true. The proof breaks down because the reductivity of G is not enough to insure that R^G is a direct summand of R.

4. Cohen–Macaulay modules. It is much harder to understand the behavior of a local Noetherian ring (R, m) if it is not C-M. It turns out that many problems are quite a bit easier if one knows, at least, that there exists a module E, not necessarily finitely generated, such that some system of parameters $x_1, \ldots, x_n$ for R is a regular sequence on E. (Recall that this implies $(x_1, \ldots, x_r)E \neq E$ or, equivalently, $mE \neq E$.) E is then called a *big C-M module* for R. In fact, the question of whether such modules exist is a central issue in the theory of local rings. Their existence yields proofs for a large number of conjectures. We shall not detail these conjectures here: see **[Au$_1$]**, **[Au$_2$]**, **[Ba]**, **[EE]**, **[Gri]**, **[Ho$_1$]**, **[Ho$_3$]**, **[Ho$_4$]**, **[Ho$_6$]**, **[Ho$_7$]**, **[Iv]**, **[PS$_2$]** and **[Ro]** for further background. We do mention that M. Auslander was responsible for a number of them, and initiated their study. Some are related to Serre's conjecture on multiplicities **[Se]**. Later, Peskine–Szpiro made tremendous progress via reduction to char. p **[PS$_1$]**, **[PS$_2$]**.

The result we do want to discuss a little bit here is:

(4.1) THEOREM. *Let (R, m) be a local ring which contains a field. Then R has a big C-M module.*

(It suffices that R/P, where P is a minimal prime such that $\dim R/P = \dim R$, comtain a field.)

This result is proved in **[Ho$_3$]**. First one shows that there exists an R-module E on which $x_1, \ldots, x_n$ is a regular sequence iff for each of countably many systems of polynomial equations

$$\mathscr{E}\begin{cases} F_1(X_1, \ldots, X_n, Y_1, \ldots, Y_q) = 0, \\ \qquad \cdots \\ F_m(X_1, \ldots, X_n, Y_1, \ldots, Y_q) = 0 \end{cases}$$

over Z, the system $\mathscr{E}$ has no solution in R such that $X_1 = x_1, \ldots, X_n = x_n$.

One reduces the case where R contains a field of char. 0 to the char. $p>0$ case by proving:

(4.2) THEOREM. *If a system $\mathscr{E}$ as above has a solution in a local ring R containing a field of char.* 0 *such that the values $x_1, \ldots, x_n \in R$ for $X_1, \ldots, X_n$ form a system of parameters, then it also has such a solution in a local ring of char. $p>0$.*

(The proof of Theorem (4.2) makes essential use of the approximation theorem of M. Artin. See [Ar_1], [Ar_2].)

In char p one shows that the systems $\mathcal{E}$ arising from the big C-M modules problem have no solution by applying high powers of the Frobenius endomorphism of R.

If R is a well-behaved local ring, e.g. a complete local ring, there is some hope that one might be able to obtain C-M modules which are finitely generated. This problem is virtually completely open if Dim $R \geqslant 3$, as is the question of existence of big C-M modules if R does not contain a field.

We note the following result of P. Griffith **[Gri]**:

(4.3) THEOREM (P. GRIFFITH). *Let A be a complete regular local ring and R a domain module-finite over A. Then if R has a big C-M module, it also has one which is countably generated and A-free.*

We conclude by giving an example of the type of equational obstruction to the existence of big C-M modules encountered in (4.2). The single equation

$$\left\{X_1^t \dots X_n^t - \sum_{i=1}^{n} Y_i X_i^{t+1} = 0\right\}$$

is such an example. It is not known (even for $n=3$, $t=2$) whether this equation has a solution in a local ring such that the values of the X_i form a system of parameters: impossibility follows, however, if the local ring has a big C-M module.

This equation is related to the question: given a regular Noetherian ring R and a module-finite extension S, is R a direct summand of S as R-modules? This is known if R contains a field, but seems to be open even for $R=\boldsymbol{Z}[X, Y]$, the polynomial ring in two variables over the integers.

Finally, we note that **[Ho_7]** introduces a conjecture apparently weaker than the existence of big C-M modules, yet which seems to have the same consequences. This new "canonical element" conjecture is more functorial and seems to be easier to study.

References

[Ar_1] M. Artin, *On the solutions of analytic equations*, Invent. Math. **5** (1968), 277—291.

[Ar_2] ——— *Algebraic approximation of structures over complete local rings*, Inst. Hautes Etudes Sci. Publ. Math. **36** (1969), 23—56.

[Au_1] M. Auslander, *Modules over unramified regular local rings*, Illinois J. Math. **5** (1961), 631—645.

[Au_2] ——— *Modules over unramified regular local rings*, Proc. Internat. Congress of Math., 1962, pp. 230—233.

[Ba] H. Bass, *On the ubiquity of Gorenstein rings*, Math. Z. **82** (1963), 8—28.

[Be] M.-J. Bertin, *Anneaux d'invariants de polynomes, en caracteristique p*, C. R. Acad. Sci. Paris Ser. A—B **264** (1967), 653—656.

[Bor] A. Borel, *Linear algebraic groups*, Benjamin, New York, 1969.

[Bou_1] J.-F. Boutot, *Frobenius et cohomologie locale*, Seminaire Bourbaki, Exp. 453, 1974.

[Bou_2] ——, *Schema de Picard local*, C. R. Acad. Sci. Paris Ser. A—B **277** (1973), 691—694.

[Bou_3] ——, *Quotient singularities are rational*, manuscript.

[DC] J. A. Dieudonné and J. B. Carrell, *Invariant theory, old and new*, Academic Press, New York, 1971.

[Ei] D. Eisenbud, *Some directions of recent progress in commutative algebra*, Proc. Sympos. Pure Math., vol. 29, Amer. Math. Soc., Providence, R. I., 1975, pp. 111—128.

[EiE] D. Eisenbud and E. G. Evans, *A generalized principal ideal theorem*, Nagoya Math. J. **62** (1976), 41—53.

[FG] R. Fossum and P. Griffith, *Complete local factorial rings which are not Cohen-Macaulay in characteristic p*, Ann. Sci. Ecole Norm. Sup. **8** (1975), 189—200.

[FrK] E. Freitag and R. Kiel, *Algebraische Eigenschaften der lokalen Ringe in den Spitzen der Hilbertschen Modulgruppen*, Invent. Math. **24** (1974), 121—148.

[GraR] H. Grauert and O. Riemenschneider, *Verschwindungsatze fur analytische Kohomologiegruppen auf Komplexen Raumen*, Invent. Math. **11** (1970), 263—292.

[Gri] P. Griffith, *A representation theorem for complete local rings*, J. Pure Appl. Algebra **7** (1976), 303—315.

[GroH] A. Grothendieck (notes by R. Hartshore), *Local cohomology*, Lecture Notes in Math., vol. 41, Springer-Verlag, Berlin and New York, 1967.

[Gros] F. Grosshans, *Observable groups and Hilbert's fourteenth problem*, Amer. J. Math. **95** (1973), 229—253.

[Ha] W. J. Haboush, *Reductive groups are geometrically reductive*, Ann. of Math. **102** (1975), 67—83.

[Hi] H. Hironaka, *Resolution of singularities of an algebraic variety over a field on characteristic* 0, Ann. of Math. **79** (1964), 205—326.

[Ho_1] M. Hochster, *Cohen-Macaulay modules*, Proc. Kansas Commutative Algebra Conference, Lecture Notes in Math., vol. 311, Springer-Verlag, Berlin and New York, 1973, pp. 120—152.

[Ho_2] ——, *Contracted ideals from integral extensions of regular rings*, Nagoya Math. J. **51** (1973), 25—43.

[Ho_3] ——, *Topics in the homological theory of modules over commutative rings*, C. B. M. S. Regional Conf. Ser. in Math. No. 24, Amer. Math. Soc., Providence, R. I., 1975.

[Ho_4] ——, *Big Cohen-Macaulay modules and algebras and embeddability in rings of Witt vectors*, Proc. of the Queen's University Commutative Algebra Conference (Kingston, Ontario, Canada, 1975), Queen's Papers in Pure and Appl. Math. **42**, pp. 106—195.

[Ho_5] ——, *Cohen-Macaulay rings, combinatorics, and simplicial complexes*, Proc. of the Second Oklahoma Ring Theory Conference (March, 1976), Marcel Dekker, New York, 1977.

[Ho_6] ——, *Some applications of the Frobenius in characteristic* 0, Bull. Amer. Math. Soc. **84** (1978), 886—912.

[Ho_7] ——, *Canonical elements in local cohomology modules and big Cohen-Macaulay modules* (in preparation).

[HR_1] M. Hochster and J. L. Roberts, *Rings of invariants of reductive groups acting on regular rings are Cohen-Macaulay*, Advances in Math. **13** (1974), 115—175.

[Hu] J. E. Humphreys, *Hilbert's fourteenth problem*, Amer. Math. Monthly **85** (1978), 341—354.

[Iv] B. Iversen, *Amplitude inequalities for complexes*, Aarhus University Preprint Series No. 36 (1976/77).

[Ka_1] I. Kaplansky, *Commutative rings*, Allyn and Bacon, Boston, Mass., 1970, Revised ed., 1974.

[Ka_2] ——, *Hilbert's problems* (Preliminary edition), Departmant of Math., Lecture Notes in Math., University of Chicago, 1977.

[Ke_1] G. Kempf, *Images of homogeneous vector bundles and varieties of complexes*, Bull. Amer Math. Soc. **81** (1975), 900—901.

[Ke_2] ——, *On the collapsing of homogeneous bundles*, Invent. Math. **37** (1976), 229—239

[Ke$_3$] ——— *Some quotient varieties have rational singularities*, Michigan Math. J. **24** (1977), 347—352.

[Ke$_4$] ——— *The Hochster-Roberts theorem of invariant theory*, Michigan Math. J. (to appear).

[KKMS] G. Kempf, F. Knudsen, D. Mumford and B. St. Donat, *Toroidal embeddings*. I, Lecture, Notes in Math., vol. 339, Springer-Verlag, Berlin and New York, 1973.

[L] J. Lipman, *Unique factorization in complete local rings*, Proc. Sympos. Pure Math., vol. 29, Amer. Math. Soc., Providence, R. I., 1975.

[Ma] H. Matsumura, *Commutative algebra*, Benjamin, New York, 1970.

[Mu$_1$] D. Mumford, *Geometric invariant theory*, Springer-Verlag, Berlin and New York, 1965.

[Mu$_2$] ——— *Hilbert's fourteenth problem—the finite generation of subrings such as rings of invariants*, Proc. Sympos. Pure Math., vol. 28, Part 2, Amer. Math. Soc., Providence, R. I., 1976, pp. 431—444.

[N] M. Nagata, *Lectures on the fourteenth problem of Hilbert*, Lecture Notes in Math., vol. 31, Tata Institute, Bombay, 1965.

[PS$_1$] C. Peskine and L. Szpiro, *Dimension projective finie et cohomologie locale*, Inst. Hautes Etudes Sci. Publ. Math. **42** (1973), 323—395.

[PS$_2$] ——— *Syzygies et multiplicités*, C. R. Acad. Sci. Paris Ser. A—B **278** (1974), 1421—1424.

[Rei] G. Reisner, *Cohen-Macaulay quotients of polynomial rings*, Advances in Math. **21** (1976), 30—49.

[Ro] P. Roberts, *Two applications of dualizing complexes over local rings*, Ann. Sci. Ecole Norm. Sup **9** (1976), 103—106.

[Se] J. P. Serre, *Algèbre locale. Multiplicités*, Lecture Notes in Math., vol. 11, Springer-Verlag, Berlin and New York, 1965.

[St$_1$] R. Stanley, *Cohen-Macaulay rings and constructible polytopes*, Bull. Amer. Math. Soc. 81 (1975), 135—142.

[St$_2$] ——— *The upper bound conjecture and Cohen-Macaulay rings*, Studies in Appl. Math. **54** (1975), 133—142.

[W] H. Weyl, *The classical groups*, Princeton Univ. Press, Princeton, N. J., 1946.

UNIVERSITY OF MICHIGAN
ANN ARBOR, MICHIGAN 48109, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Highest Weight Representations of Infinite-Dimensional Lie Algebras

V. G. Kac

1. Introduction. In [1], in connection with the Cartan classification of infinite-dimensional primitive pseudogroups the following result has been obtained.

THEOREM 1. *Let* $\mathfrak{g}=\bigoplus_{i\in Z}\mathfrak{g}_i$ *be a complex infinite-dimensional simple* $\mathbf{Z}$*-graded Lie algebra of finite Gelfand–Kirillov dimension* (*i.e.* $\overline{\lim}_{i\to\infty} \ln\dim\mathfrak{g}_i/\ln|i|<\infty$). *If in addition*

$(*)$ $\mathfrak{g}_{-1}\oplus\mathfrak{g}_0\oplus\mathfrak{g}_1$ *generates* $\mathfrak{g}$ *and the* $\mathfrak{g}_0$*-module* $\mathfrak{g}_{-1}$ *is irreducible,*
then $\mathfrak{g}$ *is isomorphic (without taking into account the gradation) to one of the following Lie algebras:* (a) *Cartan type algebras* W_n, S_n, H_n, K_n, (b) *algebras* $C(\mathfrak{p}, \nu)=\bigoplus_{i\in Z}t^i\mathfrak{p}_{i \bmod k}$, *where* $\mathfrak{p}$ *is a simple finite-dimensional Lie algebra,* ν *is an automorphism of order* k, *induced by an isometry of the Dynkin diagram* ($k=1, 2$ *or* 3), *defining the* $\mathbf{Z}_k$*-gradation* $\mathfrak{p}=\bigoplus\mathfrak{p}_i$, *and* t *is an indeterminate.*

My conjecture is that if one throws away hypothesis $(*)$, then only one extra example occurs: the Witt algebra with commutation relations

$$[e_i, e_j]=(i-j)e_{i+j},\ \ i,j\in\mathbf{Z}.$$

Recent achievements in the computation of the cohomology of Cartan type Lie algebras and their applications are well known (see survey [22]). Interest in the second type of infinite-dimensional Lie algebras rose lately when Macdonald's identities [2] were interpreted as Weyl denominator formulas for the universal central extension $\tilde{C}(\mathfrak{p},\nu)$ of the Lie algebra $C(\mathfrak{p},\nu)$ [3]. More generally, in [3] a character formula was obtained for any irreducible representation $L(\lambda)$ with dominant highest weight λ of the so-called Kac–Moody Lie algebras $\mathfrak{g}(A)$ (their study was started independently in [1] and [4]). Macdonald's identities correspond to the case $\mathfrak{g}(A)\simeq\tilde{C}(\mathfrak{p}, \nu)$ and $\lambda=0$. Finally the representations of the universal central extension of the Witt algebra (the so-called string algebra) have become recently a topic of interest in physics in the context of dual models [5]. In this article I want to discuss some results on the structure of highest weight representations of the Lie algebras $\mathfrak{g}(A)$ (§ 2) and of the string algebra (§ 4). The case of Cartan type Lie algebras has been studied in [6]. In § 3 I discuss some applications of algebras $\mathfrak{g}(A)$.

2. The Lie algebras $\mathfrak{g}(A)$. Let $A=(a_{ij})$ be a complex $(n\times n)$-matrix, Γ be the free abelian group with free generators α_i, $i\in I=\{1, \dots, n\}$, and Γ_+ be the subsemigroup of Γ generated by α_i, $i\in I$. We define a complex Γ-graded Lie algebra $\mathfrak{g}(A)=\bigoplus_{\alpha\in\Gamma}\mathfrak{g}_\alpha$ by the properties: (a) every graded ideal which intersects $\mathfrak{g}_0=\mathfrak{h}$ trivially is zero, (b) $\mathfrak{g}(A)$ is generated by elements e_i, f_i, h_i, $i\in I$, such that $\mathfrak{g}_{\alpha_i}=Ce_i$, $\mathfrak{g}_{-\alpha_i}=Cf_i$, h_i's form a basis of $\mathfrak{h}$, and: $[h_i, h_j]=0$, $[e_i, f_j]=\delta_{ij}h_i$, $[h_i, e_j]=a_{ij}e_j$, $[h_i, f_j]=-a_{ij}f_j$, $i, j\in I$. We denote by Δ^+ the set of $\alpha\in\Gamma_+\setminus 0$ such that $\mathfrak{g}_\alpha\neq 0$.

Examples. Let $\mathfrak{p}$ be a simple finite-dimensional Lie algebra with Chevalley generators E_i, F_i, H_i, $i=1, \dots, n$. (a) If A is the Cartan matrix of $\mathfrak{p}$, then $\mathfrak{g}(A)\simeq\mathfrak{p}$ and the Γ-gradation is the root decomposition $\mathfrak{p}=\bigoplus\mathfrak{p}_\alpha$. (b) Let θ be the highest root of $\mathfrak{p}$ and $\tilde{A}$ be the extended Cartan matrix of $\mathfrak{p}$. Then the homomorphism $\mathfrak{g}(\tilde{A})\to C[t, t^{-1}]\otimes_C\mathfrak{p}=C(\mathfrak{p}, \mathrm{id})$, defined by $e_i\mapsto E_i$, $f_i\mapsto F_i$, $i=1, \dots, n$, $e_0\mapsto tE_{-\theta}$, $f_0\mapsto t^{-1}E_\theta$, is the 1-dimensional universal central extension, rank $\Gamma=n+1$, and Γ-gradation is the pull-back of the decomposition $C(\mathfrak{p}, \mathrm{id})=\bigoplus_{\alpha,k}t^k\mathfrak{p}_\alpha$.

We call a $\mathfrak{g}(A)$-module V a module with highest weight $\lambda\in\mathfrak{h}^*$ if: (a) $V=\bigoplus_{\eta\in\Gamma_+}V_{-\eta}$ and $\mathfrak{g}_\alpha(V_{-\eta})\subset V_{\alpha-\eta}$, and (b) $V_0=Cv_0$, v_0 is a cyclic vector of V and $h(v_0)=\lambda(h)v_0$ for $h\in\mathfrak{h}$. We set $\mathrm{ch}\,V=\bigoplus_\eta(\dim V_{-\eta})\,e^\eta$ (e^η is a "formal" exponential). For each λ there is a Verma module $M(\lambda)$ such that the modules with highest weight λ are the quotients of $M(\lambda)$. $M(\lambda)$ has a unique irreducible quotient $L(\lambda)=M(\lambda)/I(\lambda)$. The function $P(\eta)=\dim M_{-\eta}$ is called Kostant partition function. One has $\mathrm{ch}\,M(\lambda)=Q^{-1}$, where $Q=\prod_{\alpha\in\Delta^+}(1-e^\alpha)^{\dim\mathfrak{g}_\alpha}$. Let ω be the involutive anti-automorphism of $\mathfrak{g}(A)$ defined by $\omega(e_i)=f_i$, $\omega(f_i)=e_i$, $\omega(h_i)=h_i$, $i\in I$. $M(\lambda)$ carries a bilinear form F which is uniquely defined by the properties: $F(v_0, v_0)=1$ and $F(g(x), y)=F(x, \omega(g)(y))$ for $x, y\in M(\lambda)$, $g\in\mathfrak{g}(A)$. In particular $\mathrm{Ker}\,F=I(\lambda)$ and $F(M_{-\eta}, M_{-\eta'})=0$ if $\eta\neq\eta'$. We set $F_\eta=F|M_{-\eta}$.

From now on we assume that A is a symmetrisable matrix i.e. $A=D\cdot B$, where $D=\mathrm{diag}(d_1, \dots, d_n)$, $\det D\neq 0$, and $B=(b_{ij})$ is a symmetric matrix. To $\eta=k_i\alpha_i\in\Gamma$ we assign a linear function on $\mathfrak{h}$ by $\eta(h_j)=\sum_i k_ia_{ji}$; we set $h_\eta=\sum_i k_id_i^{-1}h_i$. We introduce a bilinear form (,) on Γ by $(\alpha_i, \alpha_j)=b_{ij}$ and define $\varrho\in\mathfrak{h}^*$ by $\varrho(h_i)=\frac{1}{2}a_{ii}$. The following theorem generalizes the well-known results of Šapovalov and Bernstein–Gelfand–Gelfand in the case of finite-dimensional semisimple Lie algebras.

THEOREM 2 [7]. (a) *For the* $\mathfrak{g}(A)$*-module* $M(\lambda)$ *one has:*

$$\det F_\eta = \prod_{\alpha\in\Delta^+} \prod_{n\in N} \left((\lambda+\varrho)(h_\alpha) - n\frac{(\alpha,\alpha)}{2}\right)^{P(\eta-n\alpha)\dim\mathfrak{g}_\alpha}.$$

In particular $M(\lambda)=L(\lambda)$ *iff* $2(\lambda+\varrho)(h_\alpha)\neq n(\alpha,\alpha)$, *for any* $\alpha\in\Delta^+$, $n\in N=\{1,2,\ldots\}$.

(b) *Any simple subquotient of* $M(\lambda)$ *is of the form* $L(\lambda-\eta)$, *where* $\eta\in\Gamma_+$ *is such that there exist* $\beta_1,\ldots,\beta_k\in\Delta^+$ *and* $n_1,\ldots,n_k\in N$ *such that*

$$2(\lambda+\varrho-n_1\beta_1-\ldots-n_{i-1}\beta_{i-1})(h_{\beta_i}) = n_i(\beta_i,\beta_i)$$

for $i=1,\ldots,k$, *and* $\eta=\sum_{i=1}^k n_i\beta_i$.

Theorem 2 and its analogue for Lie superalgebras provides a new proof of character formulas from [3], [8], [9] and their generalizations.

THEOREM 3 [3]. *Suppose that* $a_{ii}=2$ *and* $-a_{ij}\in Z_+$, $i,j\in I$, *and* $\lambda\in\mathfrak{h}^*$ *is such that* $\lambda(h_i)\in Z_+$, $i\in I$. *Let* W *be the subgroup of* $\mathrm{GL}(\mathfrak{h}^*)$ *generated by reflections* r_i, $i\in I$, *defined by* $r_i(\lambda)=\lambda-\lambda(h_i)\alpha_i$. *Then for the* $\mathfrak{g}(A)$*-module* $L(\lambda)$ *one has:*

$$\operatorname{ch} L(\lambda) = Q^{-1}\sum_{w\in W}(\det w)\, e^{\lambda+\varrho-w(\lambda+\varrho)}.$$

In particular, for $\lambda=0$ *one has the "denominator" identity:*

$$Q = \sum_{w\in W}(\det w)\, e^{\varrho-w(\varrho)}.$$

PROOF. It is easy to see that $e^{-\lambda-\varrho}\cdot Q\cdot\operatorname{ch} L(\lambda)=\sum c_\mu e^{-\mu}$, where $c_{\lambda+\varrho}=1$ and $c_\mu\neq 0$ only for the μ's such that $L(\mu-\varrho)$ is a subquotient of $M(\lambda)$. Since for $\alpha\in\Delta^+$, $\alpha\notin\bigcup_i W(\alpha_i)$ iff $(\alpha,\alpha)\leq 0$ we obtain from Theorem 2 that these μ have form $w(\lambda+\varrho)$, $w\in W$. Now the theorem follows from W-skew-invariance of $e^{-\varrho}\cdot Q$ and W-invariance of $e^{-\lambda}\operatorname{ch} L(\lambda)$, which are provided by the hypothesis of the theorem.

3. Applications. One of the first applications of the Lie algebras $\mathfrak{g}(A)$ was the proof of Theorem 1 (§ 1), where they play the role of "test" algebras. In particular this gives a simple proof of Cartan classification of primitive filtered Lie algebras [1], [23]. Another application is a simple method of classification of symmetric spaces [1]. More generally let σ be an automorphism of finite order m of a simple Lie algebra $\mathfrak{p}$. We consider the corresponding Z_m-gradation $\mathfrak{p}=\bigoplus\mathfrak{p}_i$ and construct the "covering" Z-graded Lie algebra $C(\mathfrak{p},\sigma)=\bigoplus_i t^i\mathfrak{p}_{i \bmod m}$. The algebra $C(\mathfrak{p},\sigma)$ is isomorphic (without taking into account the gradation) to a certain $C(\mathfrak{p},\nu)$ from Theorem 1. This gives the classification of finite order automorphisms of $\mathfrak{p}$ [12] and in particular the description of the $\mathfrak{p}_0$-modules $\mathfrak{p}_1$; the corresponding to these modules connected algebraic linear groups are called σ-groups. These linear groups have many nice properties: (a) the algebra of invariant polynomials is free, (b) any level variety of invariant polynomials consists of a finite number of orbits, etc. [14], [13]. Moreover it turns out that almost all the connected algebraic linear groups acting irreducibly which satisfy (b) or those which are simple and satisfy (a) are σ-groups [13], [24]. These results and the dimensions of the correspondence between the root system Δ^+ of $\mathfrak{g}(A)$ and indecomposable representations of the corresponding graph [17] (see survey [18] for the background) indicate a deep connection between the Lie algebras $\mathfrak{g}(A)$ and invariant theory. In [25] the Lie algebras $\mathfrak{g}(A)$ provide infinite families of examples of simple finite-dimensional Lie algebras in characteristics 2 and 3. Recently the cohomology of the Lie algebras $C(\mathfrak{p}, \sigma)$ was applied to the study of the topology of various loop spaces [15], [26], [16]. The "infinite-dimensional" groups corresponding to the Lie algebras $\mathfrak{g}(A)$ are discussed in [21], [19], [20], [8]. Finally, various specializations of the formulas of Theorem 3 (§ 2) for $\mathfrak{g}(A) \simeq \tilde{C}(\mathfrak{p}, \nu)$ produce a number of η-function identities, Rogers–Ramanujan type identities, etc. These and other applications to combinatorics are considered in detail in s urvey [10]. Here I discuss briefly a few examples taken from [8].

Setting $\deg e_i = -\deg f_i = s_i$, $\deg h_i = 0$, $i \in I$, $s_i \in \mathbf{Z}_+$, define a $\mathbf{Z}$-gradation $\mathfrak{g}(A) = \oplus_k \mathfrak{g}_k(A, \bar{s})$; we consider the corresponding specialization $\varphi_s(e^{\alpha_i}) = X^{s_i}$, $i \in I$. For the $\mathbf{Z}$-graded Lie algebra $\mathfrak{g}(A) = \tilde{C}(\mathfrak{p}, \sigma)$ the corresponding specialization of the denominator formula produces a theta-series type expansion of the product $\prod_i \varphi(X^i)^{n_i}$. Here $\varphi(X) = \prod_{k \geq 1}(1 - X^k)$ and the sequence n_i is the Möbius transform of the sequence $\dim \mathfrak{p}_{i \bmod m}$, $i \in \mathbf{N}$. This product is finite iff the automorphism σ has a rational characteristic polynomial. In this case we obtain as a consequence the "very strange" formula [8]:

$$m\|\varrho - \mu\|^2 = \frac{1}{24} \sum_{k|m} kn_k,$$

where ϱ is the half-sum of the positive roots of $\mathfrak{p}$, μ is an element of the dual to the Cartan subalgebra defined by $\langle \mu, \alpha_i \rangle = s_i/2m$, $i = 1, \ldots, n$, and $\langle\ ,\ \rangle$ is the Killing form. The simplest case $\sigma = \mathrm{id}$ gives Macdonald's decomposition for $\eta^{\dim \mathfrak{p}}$ [2] and the Freudental- de Vries "strange" formula: $\|\varrho\|^2 = (1/24) \dim \mathfrak{p}$.

The specialization $\varphi_{\bar{1}}$, where $\bar{1} = (1, \ldots, 1)$, factorises the character formula: $\varphi_{\bar{1}}(\mathrm{ch}\, L(\lambda)) = \prod_{k \geq 1}(1 - X^k)^{r_k(\lambda)}$, where $r_k(\lambda) = \dim \mathfrak{g}_k(A^T, \bar{s} + \bar{1}) - \dim \mathfrak{g}_k(A^T, \bar{1})$ and $s_i = \lambda(h_i)$. This allows us to obtain new multivariable identities [8]. Let $\mathfrak{p}$ be a simple Lie algebra of type A_n, D_n, E_6, E_7 or E_8, R be the lattice generated by the roots, h be the Coxeter number, $\tilde{A}$ be the extended Cartan matrix. Let $\lambda_0 \in \mathfrak{h}^*$ be defined by $\lambda_0(h_0) = 1$, $\lambda_0(h_i) = 0$, $i = 1, \ldots, n = \mathrm{rank}\, p$ and let $\delta = \alpha_0 + \theta$. Then for the $\mathfrak{g}(\tilde{A})$-module $L(\lambda_0)$ one has:

$$\mathrm{ch}\, L(\lambda_0) = \Big(\sum_{\gamma \in R} \exp(h\|\gamma\|^2\delta + \gamma)\Big)/\varphi(e^\delta)^n.$$

A construction of $\mathfrak{g}(\tilde{A})$-modules $L(\lambda_0)$ in terms of differential operators is given in [11] (in certain sense $L(\lambda_0)$ contains almost all simple $\mathfrak{g}(\tilde{A})$-modules with dominant highest weight).

4. The string algebra S is by definition a complex Lie algebra with basis e_0', e_i, $i\in\mathbf{Z}$, with the following commutation relations:

$$[e_i, e_j] = (i-j)e_{i+j} + \frac{1}{12} i(i^2-1)\delta_{i,-j}e_0', \quad [e_i, e_0'] = 0.$$

We set $\mathfrak{h} = \mathbf{C}e_0 \oplus \mathbf{C}e_0'$, $\Gamma = \mathbf{Z}$. In the same way as in §2 we define S-modules $M(\lambda)$ and $L(\lambda)$, $\lambda\in\mathfrak{h}^*$, and bilinear forms F_n, $n\in\mathbf{Z}_+$; we set $\lambda(e_0)=h$, $\lambda(e_0')=c$. For $k, s\in\mathbf{N}$, $k\neq s$, let $\varphi_{k,s}(h, c)$ denote the quadratic polynomial in h with the roots:

$$\frac{1}{48}\left((13-c)(k^2+s^2) \pm \sqrt{c^2-26c+25}\,(k^2-s^2) - 24ks - 2 + 2c\right)$$

and let $\varphi_{k,k}(h, c) = h + (1/24)(k^2-1)(c-1)$. We set $\psi_n(h, c) = \prod_{s|n} \varphi_{s,n/s}$. In particular,

$$\psi_n(h, 0) = \prod_{s|n} \left(h - \frac{1}{24}\left((3s-2n/s)^2 - 1\right)\right).$$

THEOREM 4. (a) $\det F_n(h, c) = \prod_{i=1}^{n} \psi_i^{p(n-i)}$, *where* $p(s)$ *is the classical partition function. In particular the S-module* $M(\lambda)$ *is irreducible iff* $\varphi_{k,s}(h, c)\neq 0$ *for any* $k, s\in\mathbf{N}$. (b) $L(h+n, c)$, $n\in\mathbf{N}$, *is a subquotient of* $M(h, c)$ *iff there exists* $n_1, \ldots, n_k\in\mathbf{N}$ *such that* $\psi_{n_i}(h+n_1+\ldots+n_{i-1}, c)=0$ *for* $i=1, \ldots, k$, *and* $n=\sum n_i$.

COROLLARY. (a) *The module* $M(h, 0)$ *over the Witt algebra is irreducible iff* $h\neq(1/24)(m^2-1)$, $m\in\mathbf{Z}_+$. (b) *(Goldstone conjecture). The S-module* $M(h, 1)$ *is irreducible iff* $h\neq\frac{1}{4}m^2$, $m\in\mathbf{Z}_+$. *One has* $\operatorname{ch} L(\frac{1}{4}m^2, 1) = \varphi^{-1}(1-e^{m+1})$, *where* $\varphi = \prod_{i\geq 1}(1-e^i)$. *For* $h, c\in\mathbf{R}$, $h>0$, $c>1$, *the S-module* $M(h, c)$ *is irreducible.*

Added in proof. For $\mathfrak{p} = E_8$ the specialization $e^\delta = X$, $e^\gamma = 1$, $\gamma\in R$, of $\operatorname{ch} L[\lambda_0]$ give $[Xj(X)]^{1/3}$, where $j(X)$ is the modular invariant. This is related to recent discoveries about the Monster simple group.

References

1. V. G. Kac, *Simple irreducible graded Lie algebras of finite growth*, Math. USSR-Izv. **2** (1968), 1271—1311.

2. I. G. Macdonald, *Affine root systems and Dedekind's η-function*, Invent. Math. **15** (1972), 91—143.

3. V. G. Kac, *Infinite-dimensional Lie algebras and Dedekind's η-function*, Functional Anal. Appl. **8** (1974), 68—70.

4. R. V. Moody, *A new class of Lie algebras*, J. Algebra **10** (1968), 211—230.

5. J. H. Schwartz, *Dual resonance theory*, Physics Reports **8c** (1973), 269—335.

6. A. N. Rudakov, *Irreducible representations of infinite-dimensional Lie algebras of Cartan type*, Math. USSR-Izv. **8** (1974), 836—866; **9** (1975), 465—480.

7. V. G. Kac and D. A. Kazhdan, *Structure of representations with highest weight of infinite-dimensional Lie algebras*, Advances in Math. (to appear).

8. V. G. Kac *Infinite-dimensional Lie algebras, Dedekind's η-function, classical Mobius function and the very strange formula*, Advances in Math. **30** (1978), 85—136.

9. ——— *Representations of classical Lie superalgebras*, Lecture Notes in Math. **676** (1978), 597—626.

10. J. Lepowsky, *Lie algebras and combinatorics*, these PROCEEDINGS.

11. V. G. Kac, D. A. Kazhdan, J. Lepowsky and R. Wilson, *Construction of basic representations of Euclidean Lie algebras* (to appear).

12. V. G. Kac, *Automorphisms of finite order of semisimple Lie algebras*, Functional Anal. Appl. **3** (1969), 252—254.

13. ——— *On the orbit classification of algebraic linear groups*, Uspehi Mat. Nauk (6) **30** (1975), 173—174.

14. E. B. Vinberg, *Weyl group of a graded Lie algebra*, Math. USSR-Izv. **40** (1976), 488—526.

15. H. Garland, *Dedekind's η-function and the cohomology of infinite-dimensional Lie algebras*, Proc. Nat. Acad. Sci. U.S.A. **72** (1975), 2493—2495.

16. J. Lepowsky, *Generalized Verma modules, cohomology of loop spaces and Macdonald-type identities* (to appear).

17. V. G. Kac, *Infinite root systems, representations of graphs and invariant theory* (to appear).

18. A. Roiter, *Matrix problems*, these PROCEEDINGS.

19. V. G. Kac, *An algebraic definition of compact Lie groups*, Trudy MIEM **5** (1969), 36—47.

20. R. V. Moody and K. L. Teo, *Tits' systems with crystallographic Weyl groups*, J. Algebra **21** (1972), 178—190.

21. H. Garland, *Arithmetic properties of loop groups*. I, II (to appear).

22. D. B. Fuks, *New results on the characteristic classes of foliations*, these PROCEEDINGS.

23. B. Weisfeiler, *Infinite-dimensional filtered Lie algebras and their connection with graded Lie algebras*, Functional Anal. Appl. **2** (1968), 88—89.

24. V. G. Kac, V. L. Popov and E. B. Vinberg, *Sur les groupes lineares algebriques dont l'algèbre des invariants est libre*, C. R. Acad. Sci. Paris Ser. A **283** (1976), 875—878.

25. B. Weisfeiler and V. G. Kac, *Exponentials in Lie algebras of characteristic p*, Math. USSR-Izv. **5** (1971), 777—803.

26. H. Garland and M. S. Raghunatan *A Bruhat decomposition for the loop space of a compact group*, Proc. Nat. Acad Sci. USA **72** (1975), 4716—4717.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY
CAMBRIDGE, MASSACHUSETTS 02139, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Generators and Relations in Algebraic K-Theory

Wilberd van der Kallen

Despite the transformation of algebraic K-theory by the introduction of higher algebraic K-theory, it still makes sense to look at matrices in order to get a better understanding of K_1 and K_2. I will discuss a sample of results in which this classical approach plays a role. If anything, this sample should give a fair idea of my own interests. For a more balanced overview of algebraic K-theory and some motivating background I may refer to the proceedings of the two previous International Congresses. (See the talks of Quillen, Bass and Gersten at Vancouver and the talks of Swan, Tate, Karoubi at Nice.)

The approach I have in mind can be illustrated with the Bass–Milnor–Serre solution of the congruence subgroup problem for SL_n. This is the problem to decide if each subgroup of finite index in $\mathrm{SL}_n(\mathcal{O})$ contains a subgroup $\mathrm{SL}_n(\mathcal{O}, I)=\ker\left(\mathrm{SL}_n(\mathcal{O})\to\mathrm{SL}_n(\mathcal{O}/I)\right)$ for some ideal I of $\mathcal{O}$ when $\mathcal{O}$ is, say, the ring of integers in a number field. To answer this question (for $n\geqslant 3$) they had to compute the relative K-group $SK_1(\mathcal{O}, I)$ for every ideal I of $\mathcal{O}$. (Definitions of K_1 and K_2 groups will be recalled below.) The computation of $SK_1(\mathcal{O}, I)$ involved several steps. First a *stability theorem* was proved stating that the stabilization maps $SK_1(r, \mathcal{O}, I)\to SK_1(\mathcal{O}, I)$ are surjective for $r\geqslant 2$ and injective for $r\geqslant 3$. Next the *prestabilization* problem was solved, i.e. generators were given for the kernel R of $SK_1(2, \mathcal{O}, I)\to SK_1(3, \mathcal{O}, I)$. By choosing generators and relations for $SK_1(2, \mathcal{O}, I)/R$, which is thus isomorphic to $SK_1(\mathcal{O}, I)$, a *presentation* for $SK_1(\mathcal{O}, I)$ was then obtained, the presentation by Mennicke symbols and their "universal" relations. *Test maps* were found (with values in the group of roots of unity in $\mathcal{O}$), yielding lower bounds for $SK_1(\mathcal{O}, I)$. Finally, the arithmetic of the ring was further exploited to *compute* $SK_1(\mathcal{O}, I)$ exactly. Thus, finding the presentation for $SK_1(\mathcal{O}, I)$ was an important step, but it was by no means the final step. I will ignore this observation and mainly look at stability for K_1 and K_2 and presentations for K_2. I should remark that if stability sets in later than in the situation above, one tends to get less concrete information, when trying the same approach.

1. Basic notions. Let A be a ring (always associative with unit). We embed the group $\mathrm{GL}_n(A)$ into $\mathrm{GL}_{n+r}(A)$ by means of the *stabilization map* $M \mapsto \begin{pmatrix} M & 0 \\ 0 & 1 \end{pmatrix}$, where 1 is the identity in $\mathrm{GL}_r(A)$. The direct limit or union of the $\mathrm{GL}_n(A)$ we call $\mathrm{GL}(A)$ or $\mathrm{GL}_\infty(A)$, the stable general linear group. For $a \in A$, $i \neq j$, the *elementary matrix* $e_{ij}(a)$ has ones on the diagonal, a at the intersection of the ith row and the jth column, and zeroes elsewhere. The subgroup of $\mathrm{GL}_n(A)$ generated by the elementary matrices is called $E_n(A)$, the elementary subgroup. We again have stabilization maps $E_n(A) \to E_{n+r}(A)$ and we put $E(A) = E_\infty(A) = \varinjlim_n E_n(A)$. It turns out that $E(A) = [\mathrm{GL}(A), \mathrm{GL}(A)]$ and we put $K_1(A) = \mathrm{GL}(A)/E(A)$, which is thus abelian. A *transvection* in $\mathrm{GL}_n(A)$ is a linear transformation of the form $1 + vaw$ where $1 = \mathrm{id}$, v is a column of length n, $a \in A$, w is a row of length n with $wv = 0$, and w is *unimodular* (i.e. there is a column y with $wy = 1 \in A$). Let $T_n(A)$ denote the subgroup of $\mathrm{GL}_n(A)$ generated by transvections. More generally, if I is a two-sided ideal in A, let $E_n(A, I)$ be the smallest normal subgroup of $E_n(A)$ containing the elementary matrices $e_{ij}(t)$ with $t \in I$, and let $T_n(A, I)$ be the group generated by the transvections $1 + vaw$ with $a \in I$. (So $T_n(A, I)$ contains $E_n(A, I)$.) If $n \geqslant 3$, then $E_n(A, I)$ is generated by the $e_{ji}(a)e_{ij}(t)e_{ji}(-a)$ with $t \in I$, $a \in A$, and, as always, $i \neq j$ (cf. [3, Appendix 1]). If, moreover, A is *almost commutative* (i.e. finitely generated as a module over its center), Suslin has shown by a localization technique that $E_n(A, I) = T_n(A, I)$, so that $E_n(A, I)$ is a normal subgroup of $\mathrm{GL}_n(A)$. (This usually fails for $n = 2$, even if $A = I$.) We put $K_1(n, A) = \mathrm{GL}_n(A)/E_n(A)$. This pointed set is thus often a group, though not always abelian. (I have been told that it is not abelian for $n = 15$ when A is the ring of continuous real valued functions on the product of two 7-spheres.) The stabilization maps for the GL_m and E_m induce stabilization maps $K_1(n, A) \to K_1(n+r, A)$. Note that such a map is injective if and only if $\mathrm{GL}_n(A) \cap E_{n+r}(A) = E_n(A)$. Similarly we have $K_1(n, A, I) = \mathrm{GL}_n(A, I)/E_n(A, I)$, where $\mathrm{GL}_n(A, I) = \ker(\mathrm{GL}_n(A) \to \mathrm{GL}_n(A/I))$, and we write $K_1(A, I)$ for $K_1(\infty, A, I)$. If A is commutative, the group $SK_1(n, A, I)$ is the analogue of $K_1(n, A, I)$ with GL_n replaced by SL_n.

While K_1 measures when matrices differ by a product of elementary matrices, K_2 measures those relations between elementary matrices which depend on the ring. (And K_3 measures relations between relations, cf. [3], as is illustrated nicely by K. Igusa's recent concrete description of an element of order 16 in $K_3(\mathbf{Z})$. So the approach with generators and relations even seems to penetrate K_3 a little.) For $n \geqslant 3$ the Steinberg group $\mathrm{St}_n(A)$ is defined by the following presentation. Take a generator $x_{ij}(a)$ for each $e_{ij}(a)$ in $E_n(A)$. Take as defining relations the following universal relations between elementary matrices (the *Steinberg relations*)

$x_{ij}(a)x_{ij}(b)=x_{ij}(a+b)$; $[x_{ij}(a), x_{jk}(b)]=x_{ik}(ab)$; $[x_{ij}(a), x_{kl}(b)]=1$ when $j\neq k$, $i\neq l$. There is an obvious map from $\mathrm{St}_n(A)$ onto $E_n(A)$ and its kernel is called $K_2(n, A)$. (For $n=2$ one more type of relation must be added to the list.) As usual we have stabilization maps and we write $\mathrm{St}(A)=\mathrm{St}_\infty(A)$, $K_2(A)=K_2(\infty, A)$ for the respective limits. Then $K_2(A)$ is the center of $\mathrm{St}(A)$ and $\mathrm{St}(A)\to E(A)$ is a universal central extension so that $K_2(A)=H_2(E(A))$. (If G is a group $H_2(G)$ stands for $H_2(G, \mathbf{Z})$ with trivial action on the coefficients.) One can define an analogue, $\mathrm{St}_n^*(A)$, of $\mathrm{St}_n(A)$ by taking a generator for each transvection in $\mathrm{GL}_n(A)$ and taking defining relations which mimic certain universal relations between transvections. This has the advantage that $\ker(\mathrm{St}_n^*(A)\to T_n(A))$ is automatically central in $\mathrm{St}_n^*(A)$. Moreover, if A is almost commutative and $n\geq 4$ it can be shown that the isomorphism $E_n(A)\to T_n(A)$ induces an isomorphism $\mathrm{St}_n(A)\to\mathrm{St}_n^*(A)$. So then $K_2(n, A)$ is also central. But I don't know if it is central for $n=3$, even for a polynomial ring in two variables over F_2. For $n=2$ counterexamples are known.

If I is a two-sided ideal in A, the *double* D is defined as the subring of $A\times A$ consisting of the (a, b) with $a-b\in I$. The relative Steinberg group $\mathrm{St}(A, I)$ is obtained as follows. (See Keune and Loday, References **[4]**–**[5]**.) The projection onto the first factor, $D\to A$, induces a homomorphism $\mathrm{St}(D)\to\mathrm{St}(A)$. Take its kernel. It contains commutators $[x_{12}((t, 0)), x_{21}((0, u))]$ for $t, u\in I$. Divide by the (central) subgroup generated by them. The result is $\mathrm{St}(A, I)$. (One can also define $\mathrm{St}(A, I)$ in St_n^*-style, without passing to the double.) Put $K_2(A, I)=\ker(\mathrm{St}(A, I)\to E(D))$. Recall that in higher algebraic K-theory there is a long exact sequence $\dots K_3(A/I)\to K_2(A, I)\to K_2(A)\to K_2(A/I)\to K_1(A, I)\dots$, which is the long exact homotopy sequence of the map $\mathrm{BGL}^+(A)\to\mathrm{BGL}^+(A/I)$. The above definitions are compatible with this.

2. Stability theorems. Conjecturally such theorems exist in a wider context but here we look only at $K_2(n, A)$ and $K_1(n, A, I)$. (Special case $A=I$.) So we ignore K_0. For special rings there are special results such as Dunwoody's theorem that, when A is euclidean, $K_2(2, A)\to K_2(r, A)$ is surjective for any $r\geq 3$. We now discuss the general results. The basic tool to prove them is Bass's stable range condition SR_n. We say that A satisfies SR_n if, for any unimodular row $\boldsymbol{a}=(a_1, \dots, a_n)$ of length n over A, there are $t_1, \dots, t_{n-1}\in A$ such that $(a_1+a_nt_1, \dots, a_{n-1}+a_nt_{n-1})$ is unimodular. Let me say that A satisfies SR_n^k (k-fold SR_n) if, given unimodular rows $\boldsymbol{a}^{(1)}, \dots, \boldsymbol{a}^{(k)}$, each of length n, there are $t_1, \dots, t_{n-1}\in A$ which do the job for all k of them simultaneously. (There also exist stable range conditions for ideals. We ignore them here.) Recall that, for a right ideal J of A, a unimodular row $(a_1, \dots, a_n)$ is called *J-unimodular* if $a_1-1\in J$, $a_i\in J$ for $i>1$. Two such rows are *J-equivalent* if one can be obtained from the other by a finite sequence of steps in which a_i is replaced by a_i+a_jt with $j\neq i$ and $t\in A$ if $j>1$, $t\in J$ if $j=1$. For $n\geq 2$ consider the following conditions:

(A_n) A is finitely generated as a module over a central subring R, and this R has a noetherian maximal spectrum of dimension $\leq n-2$.

(B_n) A satisfies SR_n.

(C_n) A satisfies SR_n^2.

(D_n) For any right ideal J of A, all J-unimodular rows of length n are J-equivalent.

(D'_n) Same with principal right ideals $J=aA$.

(E_n) For all two-sided ideals I of A, $K_1(r, A, I)\to K_1(A, I)$ is surjective for $r\geqslant n-1$ and injective for $r\geqslant n$.

(F_n) $K_2(r, A)\to K_2(A)$ is surjective for $r\geqslant n$ and injective for $r\geqslant n+1$.

Obviously, $(D_n)\Rightarrow(D'_n)$ and $(C_n)\Rightarrow(B_n)$.

THEOREM (BASS, VASERŠTEĬN, DENNIS, SUSLIN, TULENBAYEV, VAN DER KALLEN). *For* $n\geqslant 2$, $(A_n)\Rightarrow(B_n)\Rightarrow[(C_{n+1})\,\&\,(D_n)]\Rightarrow E_n$ *and* $[(C_{n+1})\,\&\,(D'_n)]\Rightarrow(F_n)$.

For $n\geqslant 3$, $(A_n)\Rightarrow(C_n)$.

So under the quite natural condition (A_n) we have the stability results (E_n), (F_n) and I have indicated possible technical intermediate results. Using $[(C_{n+1})\,\&\,(D'_n)]\Rightarrow(F_n)$, which is new, and the work of Bass, Milnor, Serre and Vaseršteĭn on the congruence subgroup problem for SL_2, I can now show the following. Let A be a subring of the algebraic closure of Q. Then if A is not contained in the ring of integers of its field of fractions or if this field is not totally imaginary, $K_2(2, A)\to K_2(A)$ is surjective and $K_2(3, A)\to K_2(A)$ is an isomorphism. This should be contrasted with a result of Dennis and Stein saying that $K_2(2, A)\to K_2(A)$ is not surjective when A is the ring of integers in $Q(\sqrt{d})$ where d is a squarefree rational integer, $d<-11$, d congruent to $-1 \bmod 8$ or to $-3 \bmod 9$. Let me finish this section by mentioning that Vaseršteĭn has solved the pre-stabilization problem for K_1 when A satisfies condition (A_n) of the theorem and $A/\mathrm{Rad}(A)$ has no zero divisors (Rad=Jacobson radical). That is, he gave generators for $\ker\big(K_1(n-1, A, I)\to K_1(A, I)\big)$.

3. Presentations for K_2. Presentations for K_2 have been obtained in two cases where stability is very strong, namely for commutative local rings and for relative K_2 of a radical ideal in a commutative ring.

(More precise results will follow.)

For a division ring D stability is also very strong but we do not know in general how to get explicit *generators* for $K_2(D)$. However, the pre-stabilization problem has been solved quite satisfactorily by Rehmann. He describes $K_2(D)$ as the kernel of a map $U_D\to[D^*, D^*]$. Here U_D may be viewed as $\mathrm{St}_1(D)/R$ where $\mathrm{St}_1(D)$ is some sort of rank 0 Steinberg group and R stands for $\ker\big(\mathrm{St}_1(D)\to\mathrm{St}(D)\big)$.

Let us restrict ourselves from now on to commutative rings. If R is semilocal we know by the above that $K_2(2, R)\to K_2(R)$ is surjective. In fact $K_2(R)$ is generated by the Dennis–Stein symbols $\langle a, b\rangle_{12}$. Here

$$\langle a, b\rangle_{12} = x_{21}\big(-b(1+ab)^{-1}\big)x_{12}(a)x_{21}(b)x_{12}\big(-a(1+ab)^{-1}\big)\big(h_{12}(1+ab)\big)^{-1}$$

is defined for $a, b\in R$ when $1+ab\in R^*=GL_1(R)$. One has $\langle ab, 1\rangle_{12}=1$, which might be used as a definition of $h_{12}(1+ab)$. Anyway, recall that $h_{ij}(t)$ is defined when $t \in R^*$, and that its image in $E(R)$ is a diagonal matrix. If $t, u \in R^*$, the Steinberg symbol $\{t, u\}_{12}$ is defined by $h_{12}(t)h_{12}(u) = \{t, u\}_{12}h_{12}(tu)$. (If both $1+ab$ and b are units, then $\langle a, b\rangle_{12} = \{1+ab, b\}_{12}$.) Let $US(R)$ denote the group of universal Steinberg symbols, which has a generator $\{t, u\}$ for each pair $t, u \in R^*$ and which has as defining relations (as an abelian group) $\{t, uv\} = \{t, u\}\{t, v\}$; $\{tu, v\} = \{t, v\}\{u, v\}$; $\{x, 1-x\} = 1$. (As the relations have to make sense, one needs that x and $1-x$ are units.)

THEOREM (MATSUMOTO). *For a (commutative) field* F, $\{t, u\} \mapsto \{t, u\}_{12}$ *defines an isomorphism* $US(F) \to K_2(F)$.

I have shown that this result also holds for a ring satisfying SR_2^5, e.g. a local ring whose residue field contains at least 6 elements. But if one is not working with fields it is often better to use Dennis–Stein symbols.

Following Maazen and Stienstra let us define the group $D(R)$ as follows. Take a generator $\langle a, b\rangle$ for each pair $a, b \in R$ with $1+ab \in R^*$. Take defining relations (as an abelian group)

(D1) $\langle a, b\rangle\langle -b, -a\rangle = 1$.

(D2) $\langle a, b\rangle\langle a, c\rangle = \langle a, b+c+abc\rangle$.

(D3) $\langle a, bc\rangle = \langle ab, c\rangle\langle ac, b\rangle$.

(Stienstra now tells me I should use a different sign convention with $\langle a, b\rangle$ replaced by $\langle -a, b\rangle$.) For any commutative ring we have homomorphisms $US(R) \to D(R) \to K_2(R)$ sending $\{t, u\}$ to $\langle (t-1)u^{-1}, u\rangle$ and $\langle a, b\rangle$ to $\langle a, b\rangle_{12}$.

THEOREM. *If* R *is a commutative local ring,* $D(R) \to K_2(R)$ *is an isomorphism.*

The full proof of this theorem depends on work of Maazen–Stienstra, Dennis–Stein and myself. (Dennis and Stein in turn use the work of Matsumoto.) I have proved the same result for a commutative ring satisfying SR_2^3.

Now consider an ideal I with $I \subseteq \mathrm{Rad}(R)$. ($R$ is still commutative.) The group $D(R, I)$ is then defined just as $D(R)$, with the following modifications. Take generators $\langle a, b\rangle$ only if a or b is in I. Take relation (D3) only if a or b or c is in I. (And, as before, only consider relations that make sense.)

THEOREM. $D(R, I) \to K_2(R, I)$ *is an isomorphism.*

Here one sends $\langle a, b\rangle$ to $\langle (a, a), (0, b)\rangle_{12}$ or to $\langle (0, a), (b, b)\rangle_{12}$. (When both make sense they are equal.) If $R \to R/I$ splits, the theorem is due to Maazen and Stienstra. The present form was noted by Keune.

4. An example. Let R be a 1-dimensional commutative ring, finitely generated over a finite field. Let $A = R[T]$. We ask when $\mathrm{GL}_4(A)$ is finitely presented. Solution: Let φ_n denote substitution of T^n for T. Let $i \geqslant 0$. Put $V_n = K_i(\varphi_n)$.

As φ_n makes A into a free module of rank n over itself, we also have a transfer map $F_n\colon K_i(A)\to K_i(A)$, such that $F_nV_n=n(\mathrm{id})$. Further, if $\alpha\in NK_i(R)=\ker(K_i(A)\to K_i(R))$, there is a natural number M such that $F_n(\alpha)=0$ for $n\geqslant M$. (This is clear in $BQ(Nil)$ context.) From these properties of F_n, V_n it follows (cf. Farrell) that $NK_i(R)$ is either zero or not finitely generated. By Vaseršteĭn A satisfies SR_3, so that $K_i(4, A)\simeq K_i(A)$ for $i=1, 2$.

Now suppose R is regular. Then $K_i(A)\simeq K_i(R)$ is finitely generated by Quillen, so $K_1(4, A)$ and $K_2(4, A)$ are finitely generated. It follows, cf. Soulé and Rehmann, that $\mathrm{GL}_4(A)$ is finitely presented. (For smaller matrices such an argument would fail. Behr has shown that $\mathrm{SL}_3(F_q[T])$ is not finitely presented, despite the fact that $SK_1(3, F_q[T])=K_2(3, F_q[T])=0$. Now if $q=2$ note that $\mathrm{SL}_3(F_q[T])=\mathrm{GL}_3(F_q[T])$.)

Conversely, suppose $\mathrm{GL}_4(A)$ is finitely presented. Then $K_1(4, A)$ is finitely generated, so $NK_1(R)$ is finitely generated and thus $NK_1(R)=0$. By Dennis there is a "noncanonical" homomorphism $p\colon H_2(\mathrm{GL}_4(A))\to K_2(A)$ whose composition with $H_2(E_4(A))\to H_2(\mathrm{GL}_4(A))$ is the usual map $H_2(E_4(A))\to K_2(A)$. So p is surjective. Now H_2 of a finitely presented group is finitely generated so $NK_2(R)$ must also be zero. By Vorst this can only happen if R is regular.

Bibliography

(They contain more extensive bibliographies.)

1. R. K. Dennis and M. R. Stein, *Update: The functor K_2 : A survey of computations and problems* (in preparation).

2. M. S. Ragunathan, *On the congruence subgroup problem*, Inst. Hautes Études Sci. Publ. Math. **46** (1976), 107—161.

3. L. N. Vaseršteĭn, *The foundations of algebraic K-theory*, Uspehi Mat. Nauk **31** (1976), no. 4 (190), 87—149 = Russian Math. Surveys **31** (1976), no. 4, 89—156.

4. Frans Keune, *The relativization of K_2*, J. Algebra **54** (1978), 159—177.

5. Jean-Louis Loday, *Cohomologie et groupe de Steinberg relatifs*, J. Algebra **54** (1978) 178—202.

University of Utrecht
Utrecht, the Netherlands

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Algebraic Groups and Reduced K-Theory

V. P. Platonov

Introduction. This report is an exposition of the main results of reduced K-theory, which was developed by the author in 1974–1976, and its applications to arithmetic, structural, and geometric problems in the theory of algebraic groups.

Originally, the main stimulus for the development of reduced K-theory was the old Tannaka–Artin problem (see [1] and [2]): suppose A is a finite-dimensional simple algebra of degree n with center K and $\mathrm{SL}(1, A)$ is the subgroup of elements of the multiplicative group A^* for which the reduced norm $\mathrm{Nrd}_{A/K}$ has the value 1; is it true that $\mathrm{SL}(1, A)=[A^*, A^*]$? Then it was discovered that the Tannaka–Artin problem is of great significance in the theory of classical groups, and 20 years later Kneser and Tits put forth a more general conjecture in the theory of algebraic groups: for a simple, simply connected, K-isotropic, algebraic group G, the group $G_K/c(G_K)$, where $c(G_K)$ is the center of G_K, is an abstract simple group ([3]; discussions of this conjecture can be found in [4]–[6]). Still later, in the development of algebraic K-theory, there appeared other connections with the Tannaka–Artin problem, which in contemporary terms is formulated as the question of the triviality of the reduced Whitehead group $\mathrm{SK}_1(A)\cong \mathrm{SL}(1, A)/[A, A^*]$ (see [7]).

Rather unexpectedly, in 1975 this author showed [8] that $\mathrm{SK}_1(A)$ can be nontrivial; hence the Kneser–Tits conjecture was also resolved negatively. Then there naturally arose the problem of investigating and calculating the group $\mathrm{SK}_1(A)$ depending on the field K and the structure of the algebra A, and this is responsible for the content of the reduced K-theory constructed in [9]–[13].

Of central importance here is the investigation of $\mathrm{SK}_1(A)$ for division algebras A over complete, discretely valued fields K. The main tool is a new method, developed in [10], based on the deep dependence of $\mathrm{SK}_1(A)$ on properties of the ramification (singularities) of A over formally analytic fields.

1. The reduced Whitehead group over local fields. Suppose A is a division algebra of degree n over a complete, discretely valued field K. Let $\mathcal{O}_K$ and $\mathcal{O}_A$ denote the ring of integers of K and A, respectively, $\mathfrak{p}_K$ and $\mathfrak{p}_A$ the prime ideals corresponding to $\mathcal{O}_K$ and $\mathcal{O}_A$, and let $\bar{A}$ be the residue division ring $\mathcal{O}_A/\mathfrak{p}_A$ and $\bar{K}=\mathcal{O}_K/\mathfrak{p}_K$. If $a\in\mathcal{O}_A$, we denote by $\bar{a}$ its image in $\bar{A}$. In what follows we will assume that the center of $\bar{A}$ is separable over $\bar{K}$.

Let $Z(\bar{A})$ be the center of $\bar{A}$. Then $[Z(\bar{A}):\bar{K})=e(A)$, where $e(A)$ is the ramification index of A, and $Z(\bar{A})$ is a cyclic extension of $\bar{K}, n=re(A)$, where $[\bar{A}:Z(\bar{A})]=r^2$.

Our aim is to reduce the calculation of $\mathrm{SK}_1(A)$ to the calculation of $\mathrm{SK}_1(\bar{A})$. An important role here is played by the

Congruence Theorem **[10]**. $(1+\mathfrak{p}_A)\cap \mathrm{SL}(1, A)\subset[A^*, A^*]$.

We must now describe the images of $\mathrm{SL}(1, A)$ and $[A^*, A^*]$ under reduction.

Proposition 1. $\overline{\mathrm{SL}(1, A)}=\{a\in\bar{A}/N_{Z(\bar{A})/\bar{K}}(\mathrm{Nrd}_{\bar{A}/Z(\bar{A})}(\alpha))=1\}$.

Proposition 2. *If* $b\in[A^*, A^*]$, *then*

$$\mathrm{Nrd}_{\bar{A}/Z(\bar{A})}(\bar{b}) = \beta^{1-\sigma}$$

where $\beta\in\mathrm{Nrd}_{\bar{A}/Z(\bar{A})}(\bar{A})$ *and* σ *generates* $\mathrm{Gal}(Z(\bar{A})/\bar{K})$.

Put $M=\mathrm{Nrd}_{\bar{A}/Z(\bar{A})}(\bar{A}^*)$ and $M_1=M\cap N^{-1}_{Z(\bar{A})/\bar{K}}(1)$, and let $M_{\sigma-1}$ be the image of M under the homomorphism $a\to\sigma(a)a^{-1}$. From the Congruence Theorem and Propositions 1 and 2 we immediately obtain the following

Corollary. *We have an exact sequence*

$$\mathrm{SK}_1(\bar{A}) \to \mathrm{SK}_1(A) \to M_1/M_{\sigma-1} \to 1.$$

The group $M_1/M_{\sigma-1}$ is exactly the group of special projective conorms of **[10]** and **[16]**. If $\mathrm{SK}_1(\bar{A})=1$, then $\mathrm{SK}_1(A)\cong M_1/M_{\sigma-1}$. In many cases the group of special projective conorms is effectively calculable and admits various cohomological interpretations. This is discussed in more detail in the next section.

2. An explicit construction and exact formulas. Let $k(x, y)$ be the field of rational functions in x and y with an arbitrary constant field, and $K=k\langle x, y\rangle$ the field of iterated formal power series, i.e. $K=k\langle x\rangle\langle y\rangle$. If R_1 and R_2 are cyclic extensions of k, they induce cyclic extensions of the fields $k(x, y)$ and K, for which we use the same notation. Similarly, if σ_1 and σ_2 are automorphisms generating the Galois groups $\mathrm{Gal}(R_1/k)$ and $\mathrm{Gal}(R_2/k)$, we use the same notation for the extended automorphisms of the Galois groups $\mathrm{Gal}(R_i/k(x, y))$ or $\mathrm{Gal}(R_i/K)$, etc., since it will always be clear from the context what we mean.

Consider the cyclic algebras $A(x, R_1)=(x, R_1, \sigma_1)$ and $A(y, R_2)=(y, R_2, \sigma_2)$ over the field $k(x, y)$. Form their tensor product $A(R_1, R_2)=A(x, R_1)\otimes_k A(y, R_2)$. It is natural to ask: When is $A(R_1, R_2)$ a division ring? A necessary condition is obviously that $L=R_1\otimes_k R_2$ must be a field, i.e. R_1 and R_2 must be linearly disjoint over k. It turns out that this condition is also sufficient and we have the following main theorem ($\mathrm{Br}\,(T/k)$ denotes the subgroup of the Brauer group $\mathrm{Br}\,(k)$ consisting of the elements that split over the field T).

REDUCTION THEOREM [10]. *Suppose that* $L=R_1 \otimes_k R_2$ *is a field; then* $A(R_1, R_2)$ *is a division ring and*

$$\mathrm{SK}_1\big(A(R_1, R_2)\big) \cong \mathrm{Br}\,(L/k)\big(\mathrm{Br}\,(R_1/k)\,\mathrm{Br}\,(R_2/k)\big).$$

COROLLARY 1. *For the field of rational functions* $k(x, y)$ *we have*

$$\mathrm{Card}\,\mathrm{SK}_1\big(A(x, R_1) \underset{k(x,y)}{\otimes} A(y, R_2)\big) \geqslant \mathrm{Card}\,[\mathrm{Br}\,(L/k)\,\mathrm{Br}\,(R_1/k)\,\mathrm{Br}\,(R_2/k)].$$

If k is a locally compact or global field, then the Reduction Theorem and the main results of class field theory enable us to calculate $\mathrm{SK}_1\big(A(R_1, R_2)\big)$ exactly. We mention some corollaries.

COROLLARY 2. *If* k *is a locally compact field, then* $\mathrm{SK}_1\big(A(R_1, R_2)\big) \cong \mathbf{Z}/m\mathbf{Z}$, *where* $m=([R_1 : k], [R_2 : k])$.

COROLLARY 3. *If* k *is a global field and* $[R_1 : k]=[R_2 : k]=p$, *where* p *is a prime, then* $\mathrm{SK}_1\big(A(R_1, R_2)\big) \cong (\mathbf{Z}/p\mathbf{Z})^{d-1}$, *where d is the number of inequivalent valuations* $v_1, \ldots, v_d$ *of* k *for which* $[L_{v_i} : k_{v_i}]=[L : k]$.

Draxl [17] suggested another interpretation of the group $M_1/M_{\sigma-1}$, which replaces $\mathrm{Br}(L/k)\,\mathrm{Br}(R_1/k)\,\mathrm{Br}(R_2/k)$ by the more conveniently calculable group $H^{-1}(\mathrm{Gal}\,(L/k, L^*)$, namely,

$$\mathrm{SK}_1\big(A(R_1, R_2)\big) \cong H^{-1}(\mathrm{Gal}\,(L/k), L^*),$$

and gave the Reduction Theorem a somewhat more general form.

We should mention that the construction of the tensor product in the Reduction Theorem arises rather naturally in connection with the well-known result of Witt [18] to the effect that, over a complete, discretely valued field K, any algebra A modulo an algebra which is unramified over K is similar to a cyclic algebra. It is shown in [12] that, analogously, we can construct cyclic algebras A over the fields $k(x, y)$ and $k\langle x, y\rangle$ for which $\mathrm{SK}_1(A)\neq 1$. It can be shown that such algebras over $k\langle x, y\rangle$ account for all algebras with nontrivial reduced Whitehead group, i.e. the constructions under consideration are sufficiently universal.

3. Infiniteness of SK_1 and the inverse problem of reduced K-theory. A priori, the group SK_1 is Abelian of finite exponent. There arises a very natural problem, called the inverse problem of reduced K-theory: which Abelian groups of finite exponent are realized in the form $\mathrm{SK}_1(A)$? It follows from the results of § 2 that $\mathrm{SK}_1(A)$ can be an arbitrarily large finite group; hence we must first clarify the question of the existence of algebras A for which $\mathrm{SK}_1(A)$ is infinite. We do this by exploiting the odd behavior of $\mathrm{SK}_1(A)$ under extension of the ground field. Namely, we have the following constructive theorem [13].

Infiniteness Theorem. *Suppose* R_1 *and* R_2 *are cyclic extensions of degree* n *of a global field* k *such that* $[(R_1, R_2)_v : k_v] = n^2$ *for some valuation* v *of* k. *Then for any Galois extension* F/k *contained in* k_v *we have*

$$\operatorname{Card} SK_1(A(R_1, R_2) \otimes F\langle x, y\rangle) \geqslant n^{[F:k]-1};$$

in particular, if F *is an infinite extension of* k, *then* $SK_1(A(R_1, R_2) \otimes F\langle x, y\rangle)$ *is an infinite Abelian group of exponent* n.

Corollary. *For the field of rational functions* $F(x, y)$ *we have*

$$\operatorname{Card} SK_1(A(x, R_1) \underset{F(x,y)}{\otimes} A(y, R_2)) \geqslant n^{[F:k]-1}.$$

In the next section we will show that the Infiniteness Theorem actually provides us with all the machinery we need for establishing infiniteness of $SK_1(A)$.

From the Reduction Theorem, the Infiniteness Theorem and from class field theory we obtain the following

Realization Theorem. *For any countable Abelian group* M *of finite exponent there exist an algebraic number field* k *and cyclic extensions* R_1 *and* R_2 *such that*

$$SK_1(A(R_1, R_2)) \cong M.$$

4. Existence Theorem and stability in reduced K**-theory.** The results of the preceding sections show that $SK_1(A)$ depends weakly on the structure of A as an algebra and that the calculation of $SK_1(A)$ for an arbitrary algebra A is not a real problem. Therefore, it is natural to regard as the main problem the characterization of fields K for which the question of the triviality of $SK_1(A)$ can be answered positively or negatively.

For finitely generated fields K (more precisely, for algebraic function fields of finite degree over a purely transcendental extension of the prime subfield) the solution of this problem is given by the following theorem **[10]**.

Existence Theorem. *Suppose* K *is a finitely generated field. If the transcendence degree of* K *over the prime subfield is greater than* 1 *if* $\operatorname{char} K = 0$ *or greater than* 2 *if* $\operatorname{char} K > 0$, *then for any number* m *there exists a division ring* A *with center* K *such that* $\operatorname{Card} SK_1(A) > m$.

For a global field K we always have $SK_1(A) = 1$ (see [2] and **[10]**). Hence to obtain the final bound in the Existence Theorem we must settle the question of the triviality of $SK_1(A)$ for division rings A over fields of algebraic functions of one variable with global constant field. At present, even for $K = \boldsymbol{Q}(x)$, where $\boldsymbol{Q}$ is the field of rational numbers, the question of the triviality of $SK_1(A)$ is still open.

Thus, the class of fields K for which we always have $SK_1(A) = 1$, where A is a division ring over K, is very restricted, and we are now close to a complete description.

For reduced K-theory itself, but particularly for its applications, a very important question is that of the behavior of the reduced Whitehead group SK_1 under extension of the ground field. It follows from the results of §§ 2 and 3 that $SK_1(A)$ can grow "pathologically" both under a large extension of K and, for example, under a quadratic extension of K, and also under an extension F/K, where the degree $[F:K]$ is relatively prime to the degree of the algebra A. Nevertheless, for a purely transcendental extension of K we have the following key theorem.

STABLITY THEOREM **[11]**. *For a purely transcendental extension F of an arbitrary field K we have*

$$SK_1(A \underset{K}{\otimes} F) \cong SK_1(A).$$

We should mention that the proof of our Stability Theorem is connected in a natural way with the principal stability theorems in algebraic K-theory and is obtained by direct application of those theorems.

5. Applications of reduced K-theory. Most published applications pertain to the weak approximation problem and the problem of the rationality of algebraic varieties of simply connected groups. They are based on the Reduction Theorem and the Stability Theorem of reduced K-theory.

Let $V_K=\{v\}$ be the set of all inequivalent valuations of a field K, and K_v the completion of K relative to v. For a connected K-defined algebraic group G and a finite set $S\subset V_K$ we consider the topological direct product $G_S=\prod_{v\in S}G_{K_v}$. The group of K-points G_K is diagonally embedded in G_S, and in this sense we may regard $G_K\subset G_S$. If the closure $\bar{G}_K=G_S$, then we say that G possesses the weak approximation property relative to S. The concept of weak approximation in an algebraic group does not depend on the field K. In this connection, Kneser **[19**, pp. 51–52], posed the problem of investigating weak approximation in simply connected algebraic groups over arbitrary fields, and conjectured that the group $G_K=\mathrm{SL}(r, A)$, where A is a finite-dimensional division ring with center K, possesses the weak approximation property.

Since $\mathrm{SL}(r, A\otimes_K K_v)\ \overline{\mathrm{SL}(r, A)}\cong \mathrm{SL}(1, A\otimes_K K_v)\ \overline{\mathrm{SL}(1, A)}$, we may limit ourselves to the case $r=1$. It turns out that the following unexpected assertion solving the above-mentioned problem of Kneser negatively is true.

WEAK APPROXIMATION THEOREM **[11]**. *There exist division rings A of arbitrary degree $m=n^2$ over a suitable field such that for an infinite set $W=\{w_i\}$ of discrete valuations of K the closure $\overline{\mathrm{SL}(1, A)}\neq \mathrm{SL}(1, A\otimes K_{w_i})$; in addition, the orders of the groups $\mathrm{SL}(1, A\otimes K_{w_i})/\overline{\mathrm{SL}(1, A)}$, which express the deviation from weak approximation, are finite, but are not bounded in the aggregate.*

Indeed, in **[11]** this theorem is given in a more explicit form and from the proof one can obtain an effective construction of W.

Ever since Serre showed in 1961 that the variety of a semisimple, multiply connected, K-defined algebraic group G need not be rational over K, it was the general opinion that the variety of a simply connected group G is rational. First it was necessary to settle this question for the normed varieties defined by the group $G_K = \mathrm{SL}(1, A)$, where A is a division ring over an arbitrary field K. It turns out that the Stability Theorem directly implies the following two propositions.

PROPOSITION 1 **[14]**. *If $G_K = \mathrm{SL}(1, A)$ defines a rational K-variety, then there exists a number m such that each element of* $\mathrm{SL}(1, A)$ *is a product of at most m commutators in A^*; in particular,* $\mathrm{SK}_1(A) = 1$.

Recall (see **[20]**) that points $a, b \in G_K$ are called R-equivalent if they can be joined over K by a finite number of rational curves. Let G_K/R denote the set of R-equivalence classes, on which a group structure is naturally induced. If [1] is the identity class in G_K/R, then obviously $[A^*, A^*] \subseteq [1]$, and it follows from the Stability Theorem that $[A^*, A^*] = 1$, i.e. we have

PROPOSITION 2 **[21]**. $\mathrm{SK}_1(A) \cong \mathrm{SL}(1, A)/R$.

It follows from the results of sections 2–4 that $\mathrm{SK}_1(A)$ is nontrivial in most cases hence, in view of Proposition 1, the varieties of simply connected groups defined by $\mathrm{SL}(1, A)$ are very seldom rational. At the same time, as noted in **[14]**, there already exist algebras A over rational function fields K such that $\mathrm{SK}_1(A) = 1$, but the variety of $\mathrm{SL}(1, A)$ is not rational.

Several rather natural questions should be solved within a short time. The most important and difficult among them is: when does $\mathrm{SK}_1(A) = 1$ imply the rationality of $\mathrm{SL}(1, A)$? The answer is unknown even for the case where K is an algebraic number field. We note also the case of algebras A of prime index; here we always have $\mathrm{SK}_1(A) = 1$, and we can conjecture that the variety of $\mathrm{SL}(1, A)$ is rational.

6. Other types of algebraic groups. From the viewpoint of the Kneser-Tits conjecture there naturally arises the problem of extending results the obtained for $\mathrm{SL}(n, A)$ to the groups of K-points G_K of arbitrary simple, simply connected, K-defined algebraic groups. First, of course, we must solve the main question of the validity of the Kneser–Tits conjecture for other types of simple algebraic groups, the most interesting of which are the unitary groups. By analogy with the calculation of $\mathrm{SK}_1(A)$, it is shown in [22] that for unitary groups the Kneser–Tits conjecture has, in general, a negative solution, and from the viewpoint of characterizing fields the line between a positive and negative answer turns out to be the same as for the group $\mathrm{SL}(n, A)$. There also arises the problem of constructing the reduced unitary K-theory associated with the calculation of the reduced unitary Whitehead group $\mathrm{SUK}_1(A)$ defined for any division ring A with an involution of the second kind (see **[10]** and **[22]**). This problem has been essentially solved in the new paper of Jančevskiĭ **[23]**.

On the other hand, for a number of classical fields (locally compact, global functional) the Kneser–Tits conjecture has a positive solution in the general case (see **[6]**). A more detailed discussion of these results can be found in Tits **[15]**.

Bibliography

1. T. Nakayama and Y. Matsushima, *Über die multiplikative Gruppe einer p-adischen Divisionsalgebra*, Proc. Imp. Acad. Tokyo **19** (1943), 622—628.

2. S. Wang, *On the commutator group of a simple algebra*, Amer. J. Math. **72** (1950), 323—334.

3. J. Tits, *Algebraic and abstract simple groups*, Ann of Math. (2) **80** (1964), 313—329.

4. G. Harder, *Eine Bemerkung zum schwachen Approximationssatz*, Arch. Math. (Basel) **19** (1968), 465—471.

5. A. Borel and J. Tits, *Homomorphismes "abstraits" de groupes algebriques simples*, Ann. of Math. (2) **97** (1973), 499—571.

6. V. P. Platonov, *Arithmetic and structural problems in linear algebraic groups*, Proc. Internat. Congress Math., Vancouver, 1974, vol. 1, 471—476.

7. H. Bass, *Algebraic K-theory*, Benjamin, New York, 1968.

8. V. P. Platonov, *On the Tannaka-Artin problem*, Dokl. Akad. Nauk SSSR **221** (1975), 1038—1041.

9. ——, *The Tannaka-Artin problem and groups of projective conorms*, Dokl. Akad. Nauk SSSR **222** (1975), 1299—1302.

10. ——, *The Tannaka-Artin problem and reduced K-theory*, Izv. Akad. Nauk SSSR Ser. Mat. **40** (1976), 227—261.

11. ——, *Reduced K-theory and approximation in algebraic groups*, Trudy Mat. Inst. Steklov. **142** (1976), 198—207.

12. ——, *The infinitude of the reduced Whitehead group in the Tannaka-Artin problem*, Mat. Sb. **100 (142)** (1976), 191—200.

13. ——, *On the infiniteness of the reduced Whitehead group*, Dokl. Akad. Nauk SSSR **227** (1976), 299—301.

14. ——, *Birational properties of the reduced Whitehead group*, Dokl. Akad. Nauk BSSR **21** (1977), 197—198.

15. J. Tits, *Groupes de Whitehead de groupes algebriques simples (d'après V. P. Platonov)*, Séminaire Bourbaki, June, 1977.

16. V. I. Jančevskiĭ, *Division rings over Hensel discretely valued fields and the Tannaka-Artin problem*, Dokl. Akad. Nauk SSSR **226** (1976), 281—283.

17. P. Draxl, SK_1 *von Algebren über vollständig diskret bewerteten Körpern und Galoiskohomologie abelscher Körpererweiterungen*, J. Reine Angew. Math. **293/294** (1977), 116—142.

18. E. Witt, *Schiefkörper über diskret bewerteten Körpern*, J. Reine Angew. Math. **176** (1936), 153—156.

19. M. Kneser, *Schwache Approximation in Algebraischen Gruppen*, Colloque sur la Théorie des Groupes Algébriques (Centre Belge Recherches Math., Brussels, 1962), Librairie Universitaire, Louvain; Gauthier-Villars, Paris, 1962, pp. 41—52.

20. Ju I. Manin, *Cubic forms*, "Nauka", Moscow, 1972.

21. V. E. Voskresenskiĭ, *The reduced Whitehead group of a simple algebra*, Uspehi Mat. Nauk **32** (1977), no. 6 (198), 247—248.

22. V. P. Platonov and V. I. Jančevskiĭ, *On the Kneser-Tits conjecture for unitary groups*, Dokl. Akad. Nauk SSSR **225** (1975), 48—51.

23. V. I. Jančevskiĭ, *Reduced unitary K-theory*, Izv. Akad. Nauk SSSR Ser. Mat. **42** (1978), no. 4.

INSTITUTE OF MATHEMATICS
ACADEMY OF SCIENCES OF THE BYELORUSSIAN SSR
MINSK, U.S.S.R.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Matrix Problems

A. V. Roiter

The subject I shall talk about may be discussed on several different levels. I shall start from a very naive point of view.

If we have a matrix over an arbitrary field, we can reduce it by the elementary transformations of its columns and rows to the form $\left(\begin{array}{c|c} E & 0 \\ \hline 0 & 0 \end{array}\right)$, where E is unit.

But assume that we have a matrix divided into n parts by $n-1$ vertical lines, and the question is: to which form we can reduce this matrix by all elementary transformations of its rows and by elementary transformations of its columns only inside of each part but not across the lines.

For $n=2$ this problem is trivial, for $n=3$ it is not difficult, for $n=4$ such classification was done independently in different ways by Nazarova [1] and by Gelfand, Ponomarev [2]. For $n>4$ this problem is not solved and contains the classical unsolved problem of the classification of pairs of linear operators.

Of course we may divide a matrix into blocks not only by vertical, but also by horizontal lines. Furthermore we may allow addition of columns of different blocks for example from left to right but not in the other direction, and so on.

Everybody can construct as many such matrix problems as he wants.

Such problems very often arise in all branches of representation theory. For example the problem we shall talk about, for $n=4$, first came up in connection with the classification of integral representations of the noncyclic group of order 4 (Klein 4-group).

Representations of quivers which I shall define below was introduced by Gabriel [3] to describe finite dimensional algebras with radical square 0, having finitely many indecomposable representations. Representations of partially ordered sets (posets) was introduced by Nazarova and myself [4] in connection with the second Brauer–Thrall conjecture which was proved by us [5] in the case of perfect field.

This conjecture was: if a finite dimensional algebra over an infinite field has infinitely many indecomposables then there exist an infinite number of dimensions such that for each of them there are infinitely many indecomposables. Bondarenko, Drozd, on one side and Ringel on the other side recently have independently given final answer to the question for which finite groups it is possible to classify all its modular representations. They use methods similar to the methods I discuss here.

But matrix problems arise not only in representation theory. First of all I must mention here the remarkable work of Szekeres [6] who already in 1949 enumerated all finite groups which contain a normal abelian subgroup of prime index, really by solving a very important and difficult matrix problem. This problem may be formulated as the problem of classification of pairs (A, B) of linear operators such that $AB=BA=0$.

It is very interesting that the same matrix problem arose and was independently solved (in a somewhat more general situation) by Gelfand and Ponomarev [7] to describe the Harish–Chandra modules of the Lorentz group. In his report to the Congress in Nice Gelfand stated that the classification of the Harish–Chandra modules of $\mathrm{SL}(2, R)$ is equal to some matrix problem, which was solved by Nazarova and myself in [8]. I. N. Bernstein, I. M. Gelfand and S. I. Gelfand have recently proved that the classification of Harish–Chandra modules for every group is equivalent to some linear algebra problem. In some cases matrix problems may be invariantly defined in very natural and beautiful form. First of all I mean the notion of quiver representations. Quiver is a set of points connected by (directed) arrows. Representation of a quiver over a field k is given if to each point i is assigned a finite dimensional vector space V_i, and to each arrow going from i to j there is assigned a linear mapping α_{ij} from V_i to V_j. As is known, quivers having finitely many in decomposable representations correspond to Dynkin diagrams without double arrows [3], [9], and quivers such that their representations may be classified correspond to extended Dynkin diagrams without double arrows [10], [11]. Another class of matrix problems having good invariant definition is representations of posets. We have a representation of a poset S over a field k if there is some finite dimensional vector space V over k and to every element $s\in S$ corresponds a subspace $V_s\subset V$ such that if $i\leqslant j$ then $V_i\subseteq V_j$. Posets having finitely many indecomposable representations are described in [12] and such posets that all their representations may be classified are described in [13].

But it is not so easy to give nice definitions of all matrix problems. Kleiner and I have attempted to give a general definition of a matrix problem by introducing the concept of representations of a differential graded category (DGC) [14].

A graded category U over a field k is a category where for every pair (a,b) of objects the set $H(a,b)$ is a set theoretical disjoint union $\bigcup_{i=0}^{\infty} H_i(a,b)$ of k vector spaces H_i, and if $\alpha\in H_i(a,b)$, $\beta\in H_j(b,c)$ then $\alpha\beta\in H_{i+j}(a,c)$.

We shall say that this category is DGC if for any (a, b, i) there is a linear operator

d from $H_i(a, b)$ to $H_{i+1}(a, b)$ such that the Leibnitz formula

$$d(\alpha\beta) = d(\alpha)\beta + (-1)^{\deg\alpha}\alpha d(\beta)$$

holds, and $d^2=0$. We say that DGC is semifree if it is freely generated by its morphisms of degree 0 and 1.

For every semifree DGC U in [14] is constructed the category $R(U, k)$ of representations U over a field k.

The objects of $R(U, k)$ are functors from the category U_0 (contains only morphism of degree 0 from U) to the category V of all finite dimensional vector spaces over k.

The definition of morphisms of $R(U, k)$ is not so short but I shall try to give a sketch of this construction.

If DGC U has only a finite set Φ of objects (it holds for real matrix problem) we may construct a differential graded algebra $\overline{U}$, whose elements are all linear combinations of morphisms in U. The multiplication in $\overline{U}$ is induced by the composition of morphisms in U when such composition is defined, and by setting $\alpha\beta=0$, if α and β are morphisms in U such that the composition $\alpha\beta$ is not defined in U. Then $\overline{U}_0$ will be finite dimensional algebra.

We may construct U_0 bimodule A which as right module is $\overline{U}_1 \oplus \overline{U}_0$ and left multiplication is defined by the formula

$$\alpha(\beta, \gamma) = (\alpha\beta + d(\alpha)\gamma, \alpha\gamma) \quad \text{where } \alpha, \gamma \in \overline{U}_0, \quad \beta \in \overline{U}_1.$$

Every functor S: $U_0 \to V$ induces some homomorphism from $\overline{U}_0$ to algebra of all matrices which acts on the vector space $V_S = \oplus \sum_{i \in I} S(i)$. If T is another functor from U_0 to V then set $\operatorname{Hom}(V_S, V_T)$ of all linear operators from V_s to V_t may be considered as $\overline{U}_0$ bimodule.

Now we may define the set of morphisms $H(S, T)$ in the category $R(U, k)$ as the set of $\overline{U}_0$ bimodule homomorphisms from A to $\operatorname{Hom}(V_s, V_t)$. To define the multiplication of morphisms in $R(U, k)$ we must mention that d: $U_1 \to U_2$ induces some $\overline{U}_0$ bimodule homomorphism $\mathscr{D}$ from A to $A \otimes A$. Now if $a \in A$ and $\mathscr{D}(a) = \sum_{i=1}^{t} a_i' \otimes a_i''$, we define $\alpha\beta(a) = \sum_{i=1}^{t} \alpha(a_i')\beta(a_i'')$ where α, β are morphisms is $R(U, k)$.

In some sense the notion of representations of DGC may be considered a generalization of the notion of natural equivalence of functors.

It seems to me that DGC language may be convenient to consider several ideas playing the main role in this theory.

First of all in this language may be formulated some algorithm which for every DGC having finitely many indecomposable representations gives after a finite number of steps the classification of all these representations. By this algorithm it is very easy to prove (see [14]) the first (but not the second!) Brauer–Thrall conjecture.

Furthermore for every DGC may be defined some quadratic form similar to the quadratic form introduced for quivers by Gabriel and Tits (see [3]). As it is known there is a correspondence between indecomposable representations of a quiver of finite type and the roots of its quadratic form. It may be proved on one hand by using Coxeter functors [9] and on the other hand it immediately follows, as was shown by Gabriel, from that fact, that for every indecomposable representation its endomorphism ring has dimension 1 (over the field k). DGC with this condition we called Shurian and proved [15] some necessary and sufficient conditions for a free DGC (see [14]) to be Shurian.

I shall finish by mentioning that Nazarova and I recently have proved for every triangular regular free DGC U of tame type [14], that if $f(x)=1$, where f is quadratic form of DGC U, and x is vector with integer coefficients, then in dimension X there is some indecomposable representation of U, and if $f(x)=0$, then in this dimension there are infinitely many indecomposables.

I am very thankful to Idun Reiten for her useful advice in mathematics and English.

Bibliography

1. L. A. Nazarova, *Representations of a tetrad.* Izv. Akad. Nauk. SSSR. Ser. Mat. **31** (1967) = Math USSR, Izv. **1** (1967), 1305—1322.

2. I. M. Gelfand and V. A. Ponomarev, *Problems of linear algebra and classification of quadruples of subspaces in a finite-dimensional vector space,* Coll. Math. Soc. Janos Bolyai 5 Hilbert Space Operators, Tihany, Hungary, 1970.

3. P. Gabriel, *Unzerlegbare Darstellungen* I, Manuscripta Math. **6** (1972), 71—103.

4. L. A. Nazarova and A. V. Roiter, *Representations of partially ordered sets,* Zap. Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. **28** (1972), 5—31.

5. *Categorical matricial problems and the conjecture of Brauer-Thrall* preprint, Institute of Math. of the AN USSR, Kiev, 1974.

6. G. Szekeres, *Determination of a certain family of finite metabelian groups.* Trans. Amer. Math. Soc. **66** (1949), 4—43.

7. I. M. Gelfand and V. A. Ponomarev, *Indecomposable representations of the Lorentz group,* Uspehi Mat. Nauk **23 (140)** (1968), 3—60.

8. L. A. Nazarova and A. V. Roiter, *On a problem of I. M. Gelfand.* Funkcional. Anal. i Prilozen. **7** (1973), 54—69.

9. I. N. Bernstein, I. M. Gelfand and V. A. Ponomarev, *Coxeter functors and Gabriel quivers.* Uspehi Mat. Nauk. **28 (170)** (1973), 19—33 = Russian Math. Surveys **28** (1973), 17—32.

10. L. A. Nazarova, *Representations of quivers of infinite type,* Izv. Akad. Nauk. SSSR **37** (1973) = Math. USSR Izv. **7** (1973).

11. P. W. Donovan and M. R. Freislich, *The representation theory of finite groups and associated algebras,* Carleton Math. Lecture notes, 1973.

12. M. M. Kleiner, *Partially Ordered sets of finite type.* Zap Naučn. Sem. Leningrad. Otdel. Mat. Inst. Steklov. **28** (1972), 32—42.

13. L. A. Nazarova, *Partially ordered sets of infinite type.* Izv. Akad. Nauk. SSSR 39 (1975), 963—991.

14. M. M. Kleiner and A. V. Roiter. *Representations of differential graded categories,* Matrix problems. Kiev, 1977, pp 5—70.

15. V. M. Bondarenko, N. S. Golovashuk, I. A. Ovsienko and A.V. Roiter, *Shurian Matrix problems,* preprint, Kiev, 1978.

Kiev, U.S.S.R.

Труды Международного Конгресса Математиков
Хельсинки, 1978

Проблема Сокращения для Проективных Модулей и Близкие Вопросы

А. А. Суслин

Классификация конечно порождённых проективных модулей над заданным кольцом A, как правило, разбивается на две весьма различные задачи. Первая из них, классификация с точностью до стабильного изоморфизма, сводится к изучению группы Гротендика $K_0(A)$ кольца A и является традиционной для алгебраической K-теории. Вторая из возникающих задач, проблема сокращения, изучена в значительно меньшей степени. Большинство результатов в этой области, полученных до 1972 года, отражено в обзоре Басса [2]. Настоящий доклад посвящён некоторым из недавних результатов, связанных с этой проблемой. Одной из основных теорем о сокращении является

Теорема 1 (Басс [1]). *Если A — коммутативное нётерово кольцо, P — конечно порождённый проективный A-модуль, причём* $\operatorname{rank} P \geqslant \dim \operatorname{Max} A+1$, *то P удовлетворяет условию сокращения.*

Основная задача, рассматриваемая ниже, — получение усиленных теорем о сокращении для специальных классов колец.

I. **Кольца многочленов.** Проблема сокращения для проективных модулей над кольцами многочленов тесно связана с известной проблемой Серра о свободности проективных модулей над кольцом многочленов над полем. Основные результаты в этом направлении были получены в начале 1976 г. Д. Квилленом **[14]** и докладчиком **[20]** (см. также [7], [9]).

Теорема 2. *Пусть A — коммутативное нётерово кольцо, $B=A[X_1, \dots, X_n]$, P — конечно порождённый проективный B-модуль, причём* $\operatorname{rank} P \geqslant \dim A+1$, *тогда* а) *если P расширен с A, то P удовлетворяет условию сокращения*;

б) *если кольцо A регулярно, то P расширен с A (и следовательно удовлетворяет условию сокращения)*.

Этот результат развивался и обобщался рядом авторов. Отметим только следующие работы:

Суон [18] показал, что Теорема 2 останется справедливой, если кольцо многочленов заменить на кольцо лорановских многочленов $A[X_1, \ldots, X_k, X_{k+1}^{\pm 1}, \ldots, X_n^{\pm 1}]$.

Автору [26] принадлежит обобщение Теоремы 2 на некоммутативный случай:

Теорема 3. *Предположим, что кольцо Λ конечно порождено как модуль над своим центром A, который является (коммутативным) нётеровым кольцом. Если P — конечно порождённый проективный $\Lambda[X_1, \ldots, X_k, X_{k+1}^{\pm 1}, \ldots, X_n^{\pm 1}]$ — модуль ранга* $> \max(\dim A, 1)$, *то*

а) *если P расширен с Λ, то P удовлетворяет условию сокращения*

б) *если кольцо Λ регулярно, то P расширен с Λ.*

Заметим, что дополнительное ограничение $\operatorname{rank} P > 1$, появляющееся в Теореме 3, вызвано существом дела: Ойангурен и Шридхаран [13] показали, что для любого некоммутативного тела Λ существуют несвободные (и следовательно нерасширенные с Λ) проективные $\Lambda[X_1, X_2]$-модули ранга один.

Основным из нерешённых вопросов в теории проективных модулей над кольцами многочленов является следующий, поставленный Бассом [2]: пусть A — регулярное коммутативное кольцо, верно ли, что всякий конечно порождённый проективный модуль над $A[X]$ является расширенным с A? В силу принципа локализации Квиллена [14], этой проблеме можно придать следующий вид: пусть A — регулярное локальное кольцо, верно ли, что всякий конечно порождённый проекмивный $A[X]$-модуль свободен? Для двумерных колец проблема Басса—Квиллена имеет положительное решение согласно теореме Хоррока—Мурти ([27], [28]). Можно также показать, что если ответ на вопрос Басса—Квиллена положителен для всех (регулярных) колец размерности $\leq d$, то, более общим образом, для любого регулярного кольца A размерности $\leq d$ проективные $A[X_1, \ldots, X_n]$-модули расширены с A при любом n. В частности, проективные $A[X_1, \ldots, X_n]$-модули расширены с A, если A регулярное кольцо размерности 2. Мохан—Кумар [11] и независимо Линдел и Люткебомерт [10] дали положительное решение проблемы Басса—Квиллена для случая кольца формальных степенных рядов:

Теорема 4. *Если $A = k\,[[T_1, \ldots, T_d]]$ — кольцо формальных степенных рядов над полем k, то всякий конечно порождённый проективный $A[X_1, \ldots, X_n]$-модуль свободен.*

Недавно автору удалось доказать, что для трёхмерных колец алгебро-геометрического происхождения ответ на вопрос Басса—Квиллена положителен:

Теорема 5. *Если A — координатное кольцо гладкого аффинного алгебраического многообразия размерности d над полем k и P — конечно порождённый проективный $A[X_1, \ldots, X_n]$-модуль, то P является расширенным с A в каждом из следующих случаев:*

а) $\operatorname{rank} P \geqslant d$;

б) $d \leqslant 3$ *и* $\operatorname{char} k \neq 2$.

II. **Аффинные алгебры.** Хорошо известно, что в категории всех коммутативных колец теорема Басса неулучшаема. Более точно: при любом $n>1$ существуют n-мерные коммутативные нётеровы кольца A, для которых модуль A^n не удовлетворяет условию сокращения. Примеры таких колец впервые были построены Суоном ([16], см. также [17]). Эти примеры имеют топологическое происхождение и в качестве базисных колец в них выступают аффинные алгебры над полем $\boldsymbol{R}$ вещественных чисел. Естественно спросить, имеются ли такие примеры среди аффинных алгебр над другими полями. Оказывается, что имеет место

Теорема 6 ([22], [24]). *Если A—аффинная алгебра над алгебраически замкнутым полем k и P — конечно порождённый проективный A-модуль ранга $\geqslant \dim A$, то P удовлетворяет условию сокращения.*

Теорема 6 тесно связана со следующим любопытным результатом об унимодулярных строках (см. [21], [22]): если $v=(a_0, \ldots, a_r)$ — унимодулярная строка с элементами из коммутативного кольца A и $n_0, \ldots, n_r$ — натуральные числа такие, что $\prod_{i=0}^r n_i$ делится на $r!$, то унимодулярная строка $(a_0^{n_0}, \ldots, a_r^{n_r})$ дополняется до обратимой матрицы. В случае $r=2$ это утверждение было независимо получено Суоном и Таубером [19], те же авторы показали, что условие «$\prod_{i=0}^r n_i$ делится на $r!$» необходимо для его справедливости.

В случае двумерных алгебр, соединяя Теорему 6 с хорошо известным фактом, что модули ранга один всегда удовлетворяют условию сокращения, получаем теорему Мурти—Суона: если A — двумерная аффинная алгебра над алгебраически замкнутым полем, то всякий конечно порождённый проективный A-модуль удовлетворяет условию сокращения.

В связи с теоремой 6 возникает вопрос: для каких полей k проективные модули над аффинными k-алгебрами удовлетворяют усиленной теореме о сокращении? Оказывается, что ответ зависит от арифметических свойств поля.

Хорошо известно (см. например [6]), что модуль A^n ($n=\dim A$) удовлетворяет условию сокращения в том и только том случае, когда группа $\mathrm{SL}_{n+1}(A)$ транзитивно действует на множестве $\mathrm{Um}_{n+1}(A)$ унимодулярных строк. Для двумерных колец множество орбит $\mathrm{Um}_3(A)/\mathrm{SL}_3(A)$ имеет естественную групповую структуру:

Теорема 7 (Васерштейн [6]). *Если A — коммутативное нётерово кольцо размерности два, то существует каноническая биекция между множеством* $\mathrm{Um}_3(A)/\mathrm{SL}_3(A)$ *и абелевой группой* $V(A)=\ker(K_0\,\mathrm{Sp}\,A\to K_0\,A)$.

Для произвольного поля k обозначим через $A=A_k$ аффинную k-алгебру

$$k[X, Y, Z]/(X^2-X)(Y^2-Y)(Z^2-Z).$$

При помощи сдвига размерности и теорем о вырезании можно убедиться, что

$$V(A)=\ker(K_2\,\mathrm{Sp}\,k\to K_2 k)=G(k).$$

Будем через $W(k)$ обозначать кольцо Витта квадратичных форм поля k и через $I(k)$ — максимальный идеал в $W(k)$, состоящий из чётномерных форм. Теорема Матсумото даёт описание групп $K_2\,\mathrm{Sp}\,k$ и $K_2\,k$ в терминах образующих и соотношений, из этого описания видно, что существует канонический эпиморфизм $K_2\,\mathrm{Sp}\,k\to I^2(k)$ и образ группы $G(k)$ при этом изоморфизме совпадает с $I^3(k)$. В итоге получаем канонический эпиморфизм

$$\varphi\colon \mathrm{Um}_3(A)/\mathrm{SL}_3(A)\to I^3(k).$$

Можно проверить, что, если $\alpha, \beta, \gamma\in k$, то

$$\varphi((1-\alpha)X+\alpha, (1-\beta)Y+\beta, (1-\gamma)Z+\gamma)=\langle 1, -\alpha\rangle\cdot\langle 1, -\beta\rangle\cdot\langle 1, -\gamma\rangle=\langle\langle\alpha, \beta, \gamma\rangle\rangle.$$

Теорема 8 (см. [24]). *Для того, чтобы унимодулярные строки*

$$((1-\alpha_0)X+\alpha_0, (1-\beta_0)Y+\beta_0, (1-\gamma_0)Z+\gamma_0)$$

и

$$((1-\alpha_1)X+\alpha_1, (1-\beta_1)Y+\beta_1, (1-\gamma_1)Z+\gamma_1)$$

лежали в одной орбите относительно $\mathrm{SL}_3(A)$, *необходимо и достаточно выполнение следующих равносильных условий:*

а) квадратичные формы $\langle\langle\alpha_0, \beta_0, \gamma_0\rangle\rangle$ *и* $\langle\langle\alpha_1, \beta_1, \gamma_1\rangle\rangle$ *изометричны;*

б) элементы $l(\alpha_0)\cdot l(\beta_0)\cdot l(\gamma_0)$ *и* $l(\alpha_1)\cdot l(\beta_1)\cdot l(\gamma_1)$ *группы* $K_3(k)$ *сравнимы по модулю* $2\cdot K_3(k)$.

В частности следующие условия равносильны:

а) модуль A^2 *удовлетворяет условию сокращения;*

б) $V(A)=0$;

в) $I^3(k)=0$;

г) $K_3(k)$ *— 2-делимая группа.*

Предположим, что поле k совпадает с $\boldsymbol{R}$. Тогда, как хорошо известно, $W(k)=\boldsymbol{Z}$, $I(k)=2\boldsymbol{Z}$, $I^3(k)=8\boldsymbol{Z}\cong\boldsymbol{Z}$, кроме того можно проверить, что в этом случае рассмотренный выше эпиморфизм $G(k)\to I^3(k)$ является изоморфизмом. Таким образом, $\mathrm{Um}_3(A_{\boldsymbol{R}})/\mathrm{SL}_3(A_{\boldsymbol{R}})\cong\boldsymbol{Z}$. Возникающий инвариант $\varphi\colon \mathrm{Um}_3(A_{\boldsymbol{R}})\to\boldsymbol{Z}$ имеет простой топологический смысл: обозначим через Γ

поверхность в $\boldsymbol{R}^3$, определённую уравнением $(X^2-X)(Y^2-Y)(Z^2-Z)=0$, тогда Γ имеет гомотопический тип двумерной сферы и всякая унимодулярная строка $v \in \mathrm{Um}_3(A_{\boldsymbol{R}})$ определяет непрерывное отображение $\Gamma \to \boldsymbol{R}^3-0$ и следовательно элемент гомотопической группы $\pi_2(\boldsymbol{R}^3-0)=\boldsymbol{Z}$. Этот топологический инвариант совпадает с построенным ранее алгебраическим.

Очевидно, что $I^3(k) \neq 0$ для формально вещественных полей, кроме того, используя гомоморфизмы ассоциированные с дискретным нормированием поля, можно показать, что $I^3(k) \neq 0$ в каждом из следующих случаев:

а) k имеет подполе, над которым конечно порождено и имеет степень трансцендентности не меньше трёх;

б) k конечно порождено, имеет ненулевую характеристику, отличную от двух, и его степень трансцендентности над простым подполем не меньше двух;

в) k конечно порождено и трансцендентно над $\boldsymbol{Q}$.

Даже в тех случаях, когда $I^3(k)=0$, Теорема 8 позволяет строить примеры модулей, не удовлетворяющих условию сокращения.

Теорема 9. а) *Если $I^2(k) \neq 0$, то существуют трёхмерные аффинные k-алгебры, для которых модуль A^2 не удовлетворяет условию сокращения.*

б) *Если поле k не является квадратично замкнутым, то существуют четырёхмерные аффинные k-алгебры, для которых A^2 не удовлетворяет условию сокращения.*

в) *Для любого поля k существуют пятимерные аффинные k-алгебры, для которых A^2 не удовлетворяет условию сокращения.*

Следствие. *Пусть k — произвольное поле, $A=k[X_1, \ldots, X_6]/(X_1 \cdot X_4+X_2 \cdot X_5+X_3 \cdot X_6-1)$ и P — проективный A-модуль, определённый унимодулярной строкой $v=(x_1, x_2, x_3)$. Тогда $P \oplus A \cong A^3$, но $P \ncong A^2$.*

Если $\operatorname{char} k \neq 2$, то последнее утверждение следует также из результатов Рейно [15].

III. **Стабильный ранг колец многочленов.** Одним из важнейших инвариантов кольца, связанных как с проблемой сокращения, так и с другими проблемами стабилизации, является его стабильный ранг (см. [1], [4], [5], [6], [23]). Стабильный ранг является сортом размерности кольца, например, если X — топологическое пространство и A — кольцо непрерывных функций на X, то (см. [4]) $\mathrm{s.r.}\, A=\dim X+1$, где $\dim X$ определяется при помощи существенных отображений в сферы. Связь стабильного ранга и более традиционных понятий размерности устанавливается следующим неравенством Басса [1]: если A — нётерово коммутативное кольцо, то $\mathrm{s.r.}\, A \leqslant \dim \operatorname{Max} A+1$. Точное значение стабильного ранга иногда можно вычислить при помощи топологических соображений, например, Васерштейн [4] доказал, что $\mathrm{s.r.}\, \boldsymbol{R}[X_1, \ldots, X_n]=n+1$, однако уже для поля $\boldsymbol{C}$ топологические методы позволяют лишь установить неравенство $\mathrm{s.r.}\, C[X_1, \ldots, X_n] > 1+n/2$. Точное значение стабильного ранга кольца многочленов от двух переменных известно для почти всех полей благодаря результатам Крузенмейера [8] и Васерштейна [6]. Именно, Крузенмейер показал, что $\mathrm{s.r.}\, k[X_1, X_2]=3$, если только $K_2(k)\neq 0$, а Васерштейн показал, что $\mathrm{s.r.}\, k[X_1, X_2]=2$, если k — алгебраическое расширение конечного поля. Наконец, согласно результатам Басса и Тейта [3], $K_2(k)$ обращается в ноль только для алгебраических расширений конечных полей и некоторых бесконечных расширений глобальных полей (чему равно точное значение стабильного ранга для этих полей, неизвестно). При вычислении стабильного ранга колец многочленов оказывается полезным следующее утверждение:

Лемма (см. [6], глава 3). *Если $r\geqslant 3$, то следующие утверждения равносильны*:

а) $\mathrm{s.r.}\ k[X_1, \ldots, X_n]\leqslant r$.

б) *Для любой $(n-1)$-мерной аффинной k-алгебры A и любого идеала $\mathfrak{A}\subset A$ группа $E_r(A, \mathfrak{A})$ транзитивно действует* на $\mathrm{Um}_r(A, \mathfrak{A})$.

Если A — коммутативное кольцо, $\mathfrak{A}$ — идеал в A и r — натуральное число, то через $\mathrm{MS}_r(A, \mathfrak{A})$ будем обозначать группу значений универсального r-символа Меннике. В том случае, когда идеал $\mathfrak{A}$ главный и $r\geqslant 3$, легко видеть, что действие $E_r(A, \mathfrak{A})$ на $\mathrm{Um}_r(A, \mathfrak{A})$ не меняет значений символов Меннике и, следовательно, каноническое отображение

$$\mathrm{m.s.}\colon \mathrm{Um}_r(A, \mathfrak{A}) \to \mathrm{MS}_r(A, \mathfrak{A})$$

пропускается через

$$\mathrm{Um}_r(A, \mathfrak{A})/E_r(A, \mathfrak{A}) \to \mathrm{MS}_r(A, \mathfrak{A}).$$

Теорема 10 (см. [25]). *Для любого поля k существует канонический эпиморфизм*

$$\mathrm{MS}_{n+1}\left(k[X_1, \ldots, X_n], (X_1^2-X_1)\cdot \ldots \cdot(X_n^2-X_n)\right) \to \tilde{K}_{n+1}(k),$$

где $\tilde{K}_{n+1}(k)$ — факторгруппа группы Милнора $K_{n+1}(k)$ по подгруппе кручения. Если поле k алгебраически замкнуто и $n\geqslant 1$, то $\tilde{K}_{n+1}(k)=K_{n+1}(k)$ и рассматриваемый эпиморфизм является изоморфизмом.

Следствие. *Если $n\geqslant 3$ и $\tilde{K}_n(k)\neq 0$, то* $\mathrm{s.r.}\,(k[X_1, \ldots, X_n])=n+1$.

Теорема 11 (см. [25]). *Пусть размерность Кронекера поля k равна d. Тогда* $\mathrm{s.r.}\, k[X_1, \ldots, X_n]=n+1$ *при* $n\leqslant d$ *и* $\mathrm{s.r.}\, k[X_1, \ldots, X_n]\geqslant [(n+d)/2]+1$ *при* $n>d$.

Аналогичные результаты справедливы также для стабильного ранга аффинных алгебр.

Теорема 12 (см. [25]). *Если $\tilde{K}_{n+1}(k)\neq 0$, то стабильный ранг любой n-мерной аффинной k-алгебры A равен $n+1$, в частности, если поле k имеет бесконечную степень трансцендентности над простым подполем, то* $\mathrm{s.r.}\, A=\dim A+1$ *для любой аффинной k-алгебры.*

С другой стороны, для аффинных алгебр над конечными полями значение стабильного ранга падает:

Теорема 13 (Васерштейн [6]). *Если k — алгебраическое расширение конечного поля и A — аффинная k-алгебра, то* $\mathrm{s.r.}\, A \leqslant \max(2, \dim A)$.

Литература

1. H. Bass, *Algebraic K-theory*, Benjamin, New York, 1968.

2. *Some problems in "classical" algebraic K-theory*, Lect. Notes Math. **342** (1973). 3—74.

3. H. Bass and J. Tate, *The Milnor ring of a global field*, Lect. Notes Math. **342** (1973), 349—447

4. Л. Н. Васерштейн, *Стабильный ранг колец и размерность топологических пространств* Функц. Анализ и Приложен **5** (1971), № 2, 17—27.

5. *О стабилизации общей линейной группы над кольцом*, Мат. Сб. **79** (1969), № 3, 405—424.

6. Л. Н. Васерштейн и А. А. Суслин, *Проблема Серра о проективных модулях над кольцами многочленов и алгебраическая К-теория*, Изв. АН СССР, Сер. Мат. **40** (1976), № 5, 993—1054.

7. D. Ferrand, *Les modules projectifs de type fini sur un anneau de polynomes sur un corps sont libres*, Sem. Bourbaki, 1976, Exposé 484.

8. M. I. Krusenmeyer, *Fundamental groups, algebraic K-theory and a problem of Abhyankhar*, Invent. Math. **19** (1973), No I, 15—27.

9. T. Y. Lam, *Serre's conjecture*, Lect. Notes Math. **635** (1978).

10. H. Lindel and W. Lütkebohmert, *Projective Moduln über polynomialen Erweiterungen von Potenzreihenalgebren*, Arch. Math. (Basel) **28** (1977), No. I, 51—54.

11. N. Mohan Kumar, *On a question of Bass and Quillen*, preprint, Tata Institute of Fundamental Research, Bombay, 1976/77.

12. M. Pavaman Murthy and R. G. Swan, *Vector bundles over affine surfaces*, Invent. Math. **36** (1976), 125—165.

13. M. Ojanguren and R. Sridharan, *Cancellation of Azumaya algebras*, J. Algebra **18** (1971), 501—505.

14. D. Quillen, *Projective modules over polynomial rings*, Invent. Math. **36** (1976), 167—171.

15. M. Raynaud, *Modules projectifs universelles*, Invent. Math. **6** (1968), No 1, 1—26.

16. R. G. Swan, *Vector bundles and projective modules*, Trans. Amer. Math. Soc. **105** (1962), 264—277.

17. *Topological examples of projective modules*, Trans. Amer. Math. Soc. **230** (1977), 201—234.

18. *Projective modules over Laurent polynomial rings*, Trans. Amer. Math. Soc. **237** (1978), 111—120.

19. R. G. Swan, J. Towber, *A class of projective modules which are nearly free*, J. Algebra **36** (1975), 427—434.

20. А. А. Суслин, *Проективные модули над кольцом многочленов свободны*, ДАН СССР **229** (1976), № 5, 1063—1066.

21. *О стабильно свободных модулях*, Мат. Сб. **102 (146)** (1977), № 4, 537—550.

22. *Теорема о сокращении для проективных модулей над алгебрами*, ДАН СССР **236** (1977), № 4, 808—811.

23. А. А. Суслин и М. С. Туленбаев, *Теорема о стабилизации для K_2-функтора Милнора* Зап. научн. сем. ЛОМИ **64** (1976), 131—153.

24. А. А. Суслин, *О проблеме сокращения для проективных модулей*, препринт ЛОМИ Р—4—77, Ленинград, 1977.

25. *Законы взаимности и стабильный ранг колец многочленов*, препринт ЛОМИ Р—3—78, Ленинград 1978.

26. *Проективные модули над кольцами многочленов в случае некоммутативного кольца коэффициентов*, Труды МИАН (в течати).

27. G. Horrocks, *Projective modules over an extension of a local ring*, Proc. London Math. Soc. (3) **14** (1964), 714—718.

28. M. Pavaman Murthy, *Projective $A[X]$-modules*, J. London Math. Soc. 41 (1966), 453—456.

Математический Институт им. В. А. Стеклов
Академии Наук СССР
Ленинградское Отделение

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Intuitionistic Algebra: Some Recent Developments in Topos Theory

G. C. Wraith

1. Geometric theories. Roughly speaking, a mathematical argument is intuitionistically valid if it does not use a deduction of the form

$$\neg\neg\varphi \to \varphi$$

or, equivalently, if it does not use the law of the excluded middle

$$\text{true} \to \varphi \vee \neg\varphi.$$

Without being more precise about the rules of intuitionistic deduction, we simply point out that whereas anything intuitionistically valid is a fortiori classically valid, the converse is not true. This presents two problems. The first is one of nomenclature: consider, for example, the following three conditions on a nontrivial commutative ring:

(1) $\forall x.(\neg(x=0) \to \exists y.\, xy=1)$,

(2) $\forall x.(\neg(\exists y.\, xy=1) \to x=0)$,

(3) $\forall x.(\text{true} \to x=0 \vee \exists y.\, xy=1)$.

In classical logic all three are equivalent, and define the notion of field. Intuitionistically they are inequivalent, and so we must distinguish the notions of field (1), field (2) and field (3).

The second problem is that of lifting classically valid proofs to intuitionistically valid ones. There is a convenient metatheorem which provides us with a short cut in certain cases. To describe it we introduce some terminology [**8**]:

A first order formula is *positive* if it does not involve universal quantification ($\forall$), implication ($\to$) or negation ($\neg$); it may involve existential quantification ($\exists$), equality (=), finite conjunctions ($\wedge$, true) and arbitrary disjunctions ($\vee$, false). A sentence is *geometric* if it is of the form

$$\forall \underline{x}. (\varphi(\underline{x}) \to \psi(\underline{x}))$$

where φ and ψ are positive formulae in the variables $\underline{x}$ Thus, field (3) is a geometric notion, but field (1) and field (2) are not. A geometric theory is a theory whose axioms are geometric sentences.

METATHEOREM. *If a geometric sentence is deducible from a geometric theory in classical logic, with the axiom of choice, then it is also deducible from it intuitionistically.*

This metatheorem allows one to conclude the intuitionistic validity of many theorems, but at the expense of showing that the hypothesis and conclusion are expressed by geometric sentences. As a very trivial example, consider the notion of a *flat* module M over a ring A; this is expressed by sentences of the form

$$\forall \underline{a}. \forall \underline{m}. (\underline{a}'\underline{m} = 0 \to \exists B. \exists \underline{n}. \underline{a}'B = \underline{0} \wedge \underline{m} = B\underline{n})$$

where $\underline{a}$ is a vector of elements of A, $\underline{m}$ and $\underline{n}$ are vectors of elements of M, and B denotes a matrix of elements of A, each of appropriate size (there is a countable disjunction concealed here). Since in classical logic all modules over fields are flat, it follows from the metatheorem that in intuitionistic logic all modules over fields (3) are flat. This is not true of fields (2).

A local ring is generally defined as a commutative ring with precisely one maximal ideal. This is a higher order definition, so we discard it for the classically equivalent definition given by

(i) $0=1 \to \text{false}$,

(ii) $\forall x. (\text{true} \to (\exists y.\ xy=1) \vee (\exists y.\ (1-x)y=1))$.

In the absence of the axiom of choice we cannot say that a local ring has a maximal ideal or a residue field.

Classical definitions often have to be inverted in some manner to give an appropriate intuitionistic notion. For example, as well as equivalence relations we may have to deal with apartness relations; as well as ideals in rings, we have anti-ideals. Our local rings have a unique minimal anti-ideal, namely the set of invertible elements, but not necessarily any maximal ideals. In the same spirit, J. Kennison [4] has shown that to have a good Galois theory for extensions of fields (3) one should consider nonconjugacy over the base field as the positive notion. He describes a gadget which he calls a profinite action, which in classical logic reduces to the notion of a profinite group acting continuously. Intermediate extensions of a Galois extension correspond to subactions, not to subgroups of the Galois group, in general.

It is not always clear when a classical notion has a geometric axiomatization. For example, consider the notion of a Henselian local ring. If A is a local ring with maximal ideal I, it is called Henselian if

$$\forall f. \forall a. (f(a) \in I \wedge \neg f'(a) \in I \to \exists x.\ x \in I \wedge f(a+x) = 0),$$

where f denotes a monic polynomial over A. We will call a local ring separably (resp. real-) closed if it is Henselian and its residue field is separably (resp. real-) closed. The Henselian condition is not geometric; however, the notions of separably closed local ring and real closed local ring are both geometrically axiomatizable [5 (iii)], [**11**].

2. Classifying toposes. Why should we consider intuitionistic validity, and hence geometric axiomatizability, in the first place? Part of the answer comes from the wider notion of model of a theory which category theory makes possible. By now it is common procedure to use the word "semigroup", for example, to mean not only a set equipped with an associative binary operation, but also a pair (G, m) where G is an object in a category $\underline{C}$ with finite products and m is a map $G\times G\to G$ for which $m(m\times 1)=m(1\times m)$. Of course, any equational structure can be defined in $\underline{C}$, with axioms expressed by commutativity of appropriate diagrams. We may say that the language of equational theories admits interpretation in categories with finite products. If we want to extend this idea to richer languages we must impose more conditions on the category $\underline{C}$. If we go the whole way, and ask for interpretability of full higher order logic, then $\underline{C}$ must be a topos. The penalty we pay for this extra generality of interpretation is that unless the topos satisfies special conditions, only intuitionistically valid deductions survive the interpretation. One of these special conditions is the axiom of choice, that epic maps have sections.

To summarize, by restricting ourselves to intuitionistic logic we earn the bonus of a wider concept of model—in the suggestive language of Lawvere [7] we use variable instead of constant sets as carriers for our structures. To avoid tedious qualifications, I shall suppose that by topos we mean Grothendieck topos. For further details about toposes I recommend the book by P. T. Johnstone.

A map of toposes (geometric morphism) $\mathscr{F}\to^{f}\mathscr{E}$ is given by a pair of functors

$$\mathscr{F} \underset{f_*}{\overset{f^*}{\rightleftarrows}} \mathscr{E}$$

with f^* preserving finite limits and left adjoint to f_*. It follows that f^* preserves finite limits and arbitrary colimits, and hence any structure defined in terms of these; but these are precisely the structures defined by geometric sentences. If f^* reflects isomorphisms we say that f is surjective. The metatheorem above is a direct consequence of:

THEOREM (BARR). *If $\mathscr{E}$ is a Grothendieck topos, there is a surjective map of toposes $\mathscr{F}\to^{f}\mathscr{E}$ with $\mathscr{F}$ satisfying the axiom of choice.*

A topos is to be thought of as a category of generalized sets and functions. So if S is a given topos, we introduce the relative notion of an S-topos; this turns out to be the same thing as a topos map to S, $\mathscr{E}\to^{\gamma} S$. We have an evident notion of map of S-toposes, and so on. A geometric structure M in S gives rise to a similar structure $\gamma^*(M)$ in $\mathscr{E}$; it will be convenient to drop the γ^* and to talk simply of M in $\mathscr{E}$. This abuse of language is just like that which identifies an element of a ring with a constant polynomial over the ring. Indeed, if $\mathscr{E}$ is the topos of sheaves on a topological space and M is a set, then M in $\mathscr{E}$ is the constant sheaf whose stalk is everywhere M.

From the absolute notion of geometric theory we get a relative notion of geometric S-theory, whose models live in S-toposes and whose structure is preserved by inverse image parts of maps of S-toposes. This is a notion which has been around for some time implicitly, and is most easily explained by examples. If A is a commutative ring, the theory of A-algebras has for its nullary operations the elements of A. If A is a commutative ring in a topos S, we cannot talk of the theory of A-algebras—but we can talk of the S-theory of A-algebras.

THEOREM. *Let T be a geometric S-theory. Then there is an S-topos $S[U]$ (the classifying topos of T) together with a T-model U in $S[U]$ (the generic T-model) such that for any S-topos $\mathscr{E}$ the functor*

$$\mathrm{Top}_S(\mathscr{E}, S[U]) \to T\text{-mod}(\mathscr{E}): f \to f^*(U)$$

is an equivalence of categories.

In some sense the notion of classifying topos is a generalization of the notion of Lindenbaum algebra. If T is single-sorted the lattice of sub-objects of U^n in the classifying topos of a geometric theory T is isomorphic to the lattice of equivalence classes of n-variable positive formulae, where equivalence means provable equivalence within T.

3. Examples. (i) Let A be a commutative ring in S, and let T be the geometric S-theory of localizations of A, whose models are local rings of fractions of A. We call the classifying S-topos of T Spec (A), and we denote the generic T-model by $\tilde{A}$.

When $S=$Sets, Spec (A) is the topos of sheaves on the prime ideal space of A with the Zariski topology. For a full discussion see **[10]**. A. Joyal has a neat constructive way of defining Spec (A) in terms of what he calls a universal support notion on A, which is a distributive lattice in S. In the case $S=$Sets, this is the lattice of compact open subsets of Spec (A). The point is that this lattice is constructively definable from A, so that we can bypass higher order and irrelevant notions like the set of prime ideals of A.

(ii) Let L be a local ring in S. By a separable closure of L we mean a formally étale local homomorphism $L \to L'$ where L' is a separably closed local ring. These are models of a geometric S-theory whose classifying S-topos we call the étale spectrum of L.

If X is a scheme, O_X is a local ring in the topos X_{zar} of sheaves on the underlying space of X. The étale spectrum of O_X is $X_{\text{ét}}$, the étale topos of X. In particular, if L is a field in Sets the étale spectrum of L is the topos of sets with continuous Gal (L)-action, where Gal (L) is the Galois group of L with the Krull topology, and the generic separable closure of L is just the separable closure of L in the ordinary sense, but thought of as an object with continuous Gal (L)-action.

(iii) A real-closed local ring has a natural structure of ordered ring given by

$$x > 0 \leftrightarrow \exists y.\, xy^2 = 1.$$

Call an ordered ring Archimedean if it satisfies the geometric sentence

$$\forall x.\, \big(x > 0 \rightarrow \bigvee_{n>0} n.\, x > 1\big).$$

Again, let L be a local ring in S. An Archimedean real closure of L is defined to be a formally étale local homomorphism $L \rightarrow L'$ with L' an Archimedean real-closed local ring. They are models of a geometric S-theory whose classifying S-topos we call the Archimedean real étale spectrum of L.

If X is a smooth $\boldsymbol{R}$-scheme of finite type, the Archimedean real étale spectrum of O_X in X_{zar} is the topos of sheaves on the underlying real manifold of X, with Euclidean topology. The generic Archimedean real closure of O_X is the sheaf of Nash functions. It is instructive that in this example the Euclidean topology arises for reasons just as "algebraic" as those giving rise to the Zariski topology. If we throw away the condition of smoothness on X we get a "Nash space" rather than a Nash manifold.

In much the same vein there is a classifying topos description of the different sorts of crystalline topos to be found in [**2**]. The co-Zariski topology and the constructible (patch) topology also arise for classifying topos reasons [**3** (ii)].

S. Schanuel has recently used the notion of atomic topos [**1**] to prove a theorem about combinatorial functors. Let Inj denote the category of finite sets and injections, and let Linj denote the category of finite linearly ordered sets and monotone injections. We have an obvious forgetful functor u: Linj$\rightarrow$Inj. On both Inj$^{\mathrm{op}}$ and Linj$^{\mathrm{op}}$ the nonempty sieves define Grothendieck topologies, and u induces a surjective map of toposes

$$\mathrm{sh}\,(\mathrm{Linj}^{\mathrm{op}}) \xrightarrow{u} \mathrm{sh}\,(\mathrm{Inj}^{\mathrm{op}}).$$

The topos sh (Inj$^{\mathrm{op}}$) classifies infinite decidable sets, that is models of the theory given by a predicate $x \# y$ with axioms

$\forall x.\, x \# x \rightarrow \text{false},$

$\forall x.\, \forall y.\, \text{true} \rightarrow x = y \vee x \# y,$

$\forall x.\, \exists y_1 \dots \exists y_n \big(\bigwedge_{i<j} y_i \# y_j\big) \wedge \big(\bigwedge_i x \# y_i\big).$

The objects of sh (Inj$^{\mathrm{op}}$) are the combinatorial functors, that is the functors Inj$\rightarrow$Sets which preserve injections and intersections. The topos sh (Linj$^{\mathrm{op}}$) classifies dense total orders; the map of toposes u corresponds to the map of theories given by

$$x \# y \leftrightarrow (x < y) \vee (y < x).$$

Schanuel's theorem asserts that if F_1, F_2 are combinatorial functors such that for any finite set S, $F_1(S)$ and $F_2(S)$ are finite and have the same cardinality, then $F_1 u$ is naturally isomorphic to $F_2 u$.

A. Kock has pointed out [5 (ii)] that the generic local ring satisfies the nongeometric sentence

$$\forall x_1 \ldots \forall x_n . (\neg (\bigwedge_i (x_i = 0)) \rightarrow \bigvee_i (\exists y . x_i y = 1))$$

which in classical logic defines a field! The problem of characterising all the non-geometric properties of a generic model appears to be difficult. If the generic model of a geometric theory T satisfies a sentence α then any geometric consequence of $T+(\alpha)$ has to be a consequence of T. We might call α T-redundant. Does the generic T-model satisfy all T-redundant sentences?

4. Formal differential geometry. Suppose that A is a commutative ring in a topos, and let

$$D = \{a \in A | a^2 = 0\}.$$

We say that A is of *line type* [5 (i)] if the map

$$A \times A \rightarrow A^D, \ (a_1, a_2) \mapsto (d \mapsto a_1 + a_2 d)$$

is an isomorphism. We naturally think of D as an infinitesimal neighbourhood of 0 in the affine line over A; if we identify A with the first axis of the affine plane $A \times A$, then D is the intersection of the axis with any circle touching the axis at the origin. The condition of line type can be thought of as stating that any graph looks linear when restricted to lie over D. If we regard D as a "tangent vector object" we can push forward quite a lot of formal differential geometry, and we can talk about tangent bundles, vector bundles, connections, vector fields, differential equations, and so on [6]. The impetus for these ideas came originally from a lecture given by Lawvere at Chicago in May 1968. It was Kock's notion of ring of line type which provided the breakthrough.

There are no rings of line type in Sets. However, Coste, Kock and Reyes independently produced different proofs of the following theorem. Call a geometric theory of commutative rings infinitesimally stable if whenever a ring B is a model, so is $B(\varepsilon)$, the ring of Grassmann dual numbers over B.

THEOREM. *A geometric theory T of commutative rings is infinitesimally stable if and only if the generic T-model A is of line type and $D = \{a \in A | a^2 = 0\}$ is internally projective, i.e. $(-)^D$ preserves colimits.*

Bibliography

1. M. Barr and R. Diaconescu, *Atomic topoi* (preprint), McGill University, 1978.

2. P. Berthelot, *Cohomologie cristalline des schemas de carateristique* p>0, Lecture Notes in Math., vol. **407**, Springer-Verlag, Berlin and New York, 1974.

3. P. T. Johnstone, (i) *Topos theory*, Academic Press, New York, 1977.
(ii) *Rings, fields, and spectra*, J. Algebra. **49** (1977), 238—260.

4. J. Kennison, *Galois theory and profinite action in a topos* (preprint), University of Sussex, 1978.

5. A. Kock, (i) *A simple axiomatics for differentiation*, Aarhus preprint series 1975/76, no. 12.
(ii) *Universal projective geometry via topos theory*, J. Pure Appl. Algebra. **9** (1976), 1—24.
(iii) *Formally real local rings and infinitesimal stability* (preprint), Aarhus University, 1977.

6. A. Kock and G. Reyes, *Connections in formal differential geometry*, Aarhus preprint series 1977/78, no. 16.

7. F. W. Lawvere, *Variable quantities and variable structures in topoi*, Algebra, Topology and Category Theory: A Collection of Papers in Honor of Samuel Eilenberg. (ed. A. Heller and M. Tierney) Academic Press, New York, 1976, pp. 101—131.

8. M. Makkai and G. Reyes, *1st order categorical logic*, Lecture Notes in Math., vol. **611**, Springer-Verlag, Berlin and New York, 1977.

9. G. Reyes, (i) *Cramer's rule in the Zariski topos*, Applications of Sheaves: Proc. Durham Symposium; Lecture Notes in Math., Springer-Verlag, Berlin and New York (to appear).
(ii) *Real local rings*, manuscript, 1977.

10. M. Tierney, *On the spectrum of a ringed topos*. Algebra, Topology and Category Theory: A Collection of Papers in Honor of Samuel Eilenberg. (ed. A. Heller and M. Tierney) Academic Press, New York, 1976, pp. 189—210.

11. G. C. Wraith, *Generic galois theory of local rings*, Applications of Sheaves: Proc. Durham Symposium, Lecture Notes in Math., Springer-Verlag, Berlin and New York (to appear).

UNIVERSITY OF SUSSEX
SUSSEX, ENGLAND

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Algebraic Independence of Values of Exponential and Elliptic Functions

G. V. Chudnovsky

0. Only one Congress separates us from 1982, the centenary of Lindemann's theorem on the transcendence of π. Many things have changed since 1882 in Transcendence Theory. For the last years especially there has been considerable progress in understanding the fundamental problems of Transcendence Theory. Although the analytic part of proofs looks like 40 years ago the algebraic arguments have changed completely. Now Transcendence Theory uses a lot of modern mathematics (algebra, algebraic geometry, complex analysis) and also has its fields of application. We'll try to describe the new situation with the theory of transcendence and algebraic independence for the exponential, elliptic and Abelian functions.

Let $\wp(z)$ denote the Weierstrass elliptic function with algebraic invariants g_2, g_3 and $\zeta(z)$ the ζ-function, $\zeta'(z) = -\wp(z)$. Let ω, η denote any pair of periods and quasi-periods of $\wp(z)$: $\zeta(z+\omega) = \zeta(z)+\eta$, and let ω_i, η_i denote fundamental periods and quasi-periods of $\wp(z)$. We call point u as algebraic for $\wp(z)$ if $\wp(u) \in \bar{Q}$. For a finite set $S \subset C$, $\# S$ denotes the maximal number of algebraically independent (a.i.) elements in S.

1. Linear independence of algebraic points of elliptic and Abelian functions. Linear independence is often the first, very important and useful for applications, step in the investigations of algebraic independence. This becomes especially clear from the works of A. Baker [1], [2], on estimates of linear forms from logarithms of algebraic numbers and applications of these estimates to diophantine equations and the class-number problem. The report of R. Tijdeman contains a survey of such results.

It was realized long ago that estimates for linear forms in algebraic points of elliptic and Abelian functions are extremely important for effective determination of integer points on curves of positive genus (see Lang's [13]). Only in 1974 D. Masser [3] has proved a linear independence result for the algebraic points of elliptic curve with complex multiplication. Later S. Lang and D. Masser [4], J. Coates–S. Lang [5] obtained more general results for CM-varieties and lower bounds for linear forms of algebraic points of Abelian varieties of CM-types with algebraic coefficients. The proof of [3] uses analytical arguments, while [5] uses Kummer theory for CM-variety (K. Ribet's theorem).

Let A be an Abelian variety defined over $\bar{Q}$ of dimension d. Then we have Abelian functions $A_1(\bar{z}), \ldots, A_d(\bar{z}), B(\bar{z})$, where $A_1(\bar{z}), \ldots, A_d(\bar{z})$ are algebraically independent and $B(\bar{z})$ is algebraic over $\bar{Q}[A_1, \ldots, A_d]$. The functions $A_1(\bar{z}), \ldots, A_d(\bar{z}), B(\bar{z})$ we suppose regular at $\bar{z}=0$; these functions are $2d$-periodic with common periods $\vec{\omega}_1, \ldots, \vec{\omega}_{2d} \in C^d$. There are also d quasi-periodic functions $H_1(\bar{z}), \ldots, H_d(\bar{z})$ algebraically independent over $C(\bar{z})$, with quasi-periods $\eta_j(\vec{\omega}_i)=\eta_{ij}$:

$$(1) \qquad H_j(\bar{z}+\vec{\omega}_i) = H_j(\bar{z})+\eta_j(\vec{\omega}_i) = H_j(\bar{z})+\eta_{ij}, \; j = 1, \ldots, d, \; i = 1, \ldots, 2d,$$

and such that $\partial/\partial z_j$ maps $\bar{Q}[A_1, \ldots, A_d, B, H_1, \ldots, H_d]$ into itself: $j=1, \ldots, d$. We call A CM-variety if for endomorphism ring End (A) we have End $(A)\otimes Q=K$-field of degree $2d$. We may assume that End (A) is represented in A by diagonal matrices.

Let $U\subset C^d$ be the set of algebraic points of CM-variety A, i.e. such that $\bar{u}\in C^d$ and $A_1(\bar{u}), \ldots, A_d(\bar{u})$ are algebraic. Let $\vec{u}_1, \ldots, \vec{u}_m$ be algebraic points of A, linearly independent over End (A) and $\vec{u}_0$ be the column vector all of whose components are unity.

The following result was proved for elliptic curve with complex multiplications ($d=1$) by Anderson and in general CM-case by D. Masser [1], [4].

THEOREM 1. *For* CM-*variety* A, *any* $K>dm+1$ *and any integer* $N\geq 1$ *there exists* $C_1>0$ *depending only on* $K, N, \vec{u}_1, \ldots, \vec{u}_m$ *and* A *such that for*

$$\bar{\Lambda} = B_0\bar{u}_0+B_1\bar{u}_1+\ldots+B_m\bar{u}_m,$$

we have

$$(2) \qquad |\bar{\Lambda}| > C_1 H^{-(\log\log H)^K}$$

for any diagonal matrices $B_0, B_1, \ldots, B_m$ *with algebraic entries of degrees at most* N *and heights at most* H, *such that* $B_0, \ldots, B_m$ *are not all singular.*

This result can also be partially generalized for the case of singular matrices $B_0, \ldots, B_m$ (see [1], [4]). However still there are no linear independence results for Abelian varieties or even for elliptic curves without complex multiplication. We propose this problem as one of the most attractive in the near future.

The natural field for applications of linear forms is, of course, diophantine equations. E.g., there is a

Problem. To effectivize Siegel's theorem on finiteness of the number of integer points on algebraic curve $\mathscr{C}$ of genus $g>1$.

We call the curve $\mathscr{C}$ of genus $g\geqslant 1$ CM-curve if $\mathscr{C}$ admits a nonconstant rational map (defined over $\bar{Q}$) into a CM-variety. As an application of the bounds of $|\bar{\Lambda}|$ to CM-curves, we have D. Masser's [4] result.

THEOREM 2. *Let $\mathscr{C}$ be* CM-*curve defined over the number field K. If P is the point from $\mathscr{C}(K)$ of height $H(P)$ and denominator $D(P)$ then*

$$H(P) < C_3 \exp\left(\log^{\varepsilon} D(P)\right)$$

for $C_3>0$ depending only on $K, \mathscr{C}$, $\varepsilon>0$.

In this direction the p-adic results give stronger corollaries. Since 1976, D. Bertrand [8], has developed methods to estimate linear forms of algebraic points of elliptic curves and Abelian varieties of CM-type on p-adic domain. E.g.

THEOREM 3. [8]. *Let $\mathscr{C}$ be* CM-*elliptic curve defined over K. If P is the point from $\mathscr{C}(K)$, $H(P)$ is the height of P and* $\mathrm{pr}\, D(P)$ *is the greatest prime factor of the denominator $D(P)$ of P, then*

$$\mathrm{pr}\, D(p) > C_4\left(\log H(P)\right)^{C_5}$$

for some $C_4>0$, $C_5>0$ depending on $\mathscr{C}$ and K.

Such results would be extremely useful if it were possible to make them completely effective in terms of $\mathscr{C}$ and K. There are two obstacles on the way: (a) the absence of an effective bound for generators of Mordell–Weyl group of the Jacobian $J(\mathscr{C})(K)$ in terms of $\mathscr{C}, K$; (b) the need for a lower bound for $|\bar{\Lambda}|$ in terms of heights of $\bar{u}_1, \ldots, \bar{u}_m$. In the direction (b) first results were already obtained by Coates–Lang [5] but the bound we really need has not been proved in general. We'll formulate it as a

Conjecture. In the previous notations, let $\bar{u}_i$ be an algebraic point of A of degree $\leqslant M$ and height $\leqslant U_i$: $i=1, \ldots, m$. Then for any $\varepsilon>0$

$$|\bar{\Lambda}| > \exp\left(-C_6(\log H \cdot \prod_{i=1}^{m} \log U_i)^{1+\varepsilon}\right), \tag{3}$$

where $C_6>0$ depends on M, N and $\varepsilon>0$ for matrices $B_0, B_1, \ldots, B_m$, being not all zeroes, with algebraic entries of degrees $\leqslant N$ and heights $\leqslant H$.

We have proved this conjecture for $m=2$ [7].

Even assuming the Birch–Swinnerton-Dyer (B–Sw-D) conjecture it is not so easy to obtain a bound for $\prod_{i=1}^{m} \log U_i$, where $\bar{u}_1, \ldots, \bar{u}_m$ are generators of Mordell–Weil group of $A(K)$. According to the general form of B–Sw–D conjecture (see report of J. Coates) we need the bound for $|L_A^{(m)}(1)|$, where $L_A(s)=\sum a_n n^{-s}$ is a properly defined L-function of the Abelian variety A defined over K. Only for an elliptic curve $A=E$ defined over Q of conductor N we have

(4) $$|L_E^{(m)}(1)| \ll_m N^{1/4}$$

assuming Weil conjecture for E and we have

(5) $$|L_E^{(m)}(1)| \ll_{m,\varepsilon} N^{\varepsilon}$$

assuming also Riemann hypothesis for $L_E(s)$.

In the CM-case the bounds (4)–(5) together with estimate (3) for $m=2$ give bounds for integer points on certain curves of positive genus.

THEOREM 4. *Let k be an integer, such that $|k|$ is the power of one prime p and E be the curve*

(6) $$y^2 = x^3 + k.$$

If $L_E(s)$ satisfies the Riemann hypothesis and the B–Sw–D conjecture, then for integer solutions x, y of (6) *we have*

$$\max(|x|, |y|) \leq \exp(C_7 p^{1/6+\varepsilon})$$

for $C_7 > 0$ depending only on $\varepsilon > 0$.

This is a considerable improvement of general H. Stark's bound.

Now, we shall deal with the "transcendental" aspect of linear independence results. The first such results were obtained by A. Baker [2], and later by J. Coates and then by D. Masser [3] for the product of two elliptic curves.

THEOREM 5. (D. MASSER [4]). *Let A be an Abelian variety of dimension* 2. *If $\bar{\omega} = (\omega_1, \omega_2)$ is a nonzero period of A then the linear combination*

$$\alpha_1 \omega_2 + \alpha_2 \omega_2 + \beta_1 \eta_1(\bar{\omega}) + \beta_2 \eta_2(\bar{\omega})$$

with algebraic $\alpha_1, \alpha_2, \beta_1, \beta_2$ is either zero (then A is not a simple variety) or transcendental provided that $|\alpha_1| + |\alpha_2| + |\beta_1| + |\beta_2| > 0$.

In the particular case of CM-variety it follows from Theorem 5 that the dimension of the set

$$\{B(m/5, n/5)\colon \text{ for } m \not\equiv -n \pmod 5\}$$

over $\bar{Q}$ is exactly five.

2. Algebraic independence. For a long time (since first applications of Gelfond-Schneider method) we know that many numbers connected with exponential and elliptic functions are transcendental, e.g. $\log \alpha$, α^β, $\log \alpha / \log \beta$, ω, η, π/ω, u, $\zeta(u)$, ... (for $\alpha, \beta \in \bar{Q}$, u algebraic of $\wp(z)$) [13]. The main problem is now the investigations of algebraic independence of these numbers.

The situation with algebraic independence for numbers connected with exponential and elliptic functions was very poor to latest time. Besides the Lindemann–Weierstrass theorem (1882) and Gelfond's example of the algebraic independence of α^{β} and α^{β^2} for algebraic $\alpha \neq 0, 1$ and β cubic irrationality (1949) there were *no* concrete examples of a.i. numbers. We can now say that lack of knowledge in this field is determined by the methods used for analysis of algebraic independence. The introduction of nontrivial algebraic (algebraico-geometrical) and multidimensional considerations gave us some ideas of how to develop Transcendence Theory and also supplied us with many new examples of algebraically independent numbers.

A. *Elliptic case.* In the elliptic case, we have, since 1975, constructed several pairs of algebraically independent numbers connected with algebraic points of $\wp(z)$ or periods of $\wp(z)$ **[Ch 6]**, **[Ch 7]** **[MW 15]**. Historically the first result was the following

THEOREM 1. *If* ω_1, ω_2 *are fundamental periods of* $\wp(z)$ *and* η_1, η_2 *are corresponding quasi-periods, then*

(1) $$\#\{\omega_1, \omega_2, \eta_1, \eta_2\} \geqslant 2.$$

The most interesting case is that of complex multiplication.

COROLLARY 2. *If* ω *is the period of* $\wp(z)$ *with complex multiplication, then* ω *and* π *are a.i.*

Thanks to Selberg–Chowla formula we can see exactly what ω is. We get from this formula and corollary 2 that for m, the discriminant of quadratic imaginary field, and $\chi(n)=(n/m)$,

(2) $$\prod_{i=1}^{m-1} \Gamma(i/m)^{\chi(i)} \quad \text{and} \quad \pi \text{ are algebraically independent.}$$

Since that we have improved (1) considerably [7]:

THEOREM 3. *For any pair* ω, η *of period, quasi-period of* $\wp(z)$ *(i.e.* $\zeta(z+\omega)=\zeta(z)+\eta$*) the two numbers*

(3) $$\pi/\omega, \eta/\omega \ \textit{are algebraically independent.}$$

Moreover we have Theorem 3 in the case when $\wp(z)$ has no algebraic invariants. When $\wp(z)$ is arbitrary, we have

(4) $$\#\{g_2, g_3, \pi/\omega, \eta/\omega\} \geqslant 2.$$

From (4) it can be deduced e.g. that for the modular invariant $J(q)$, $q=e^{2\pi i z}$; if $J(q) \neq 0$; 1728 is algebraic and $\theta = q\, d/dq$, then $\theta J(q)$ and $\theta^2 J(q)$ are a.i. The last result as well as (5) was generalized by D. Bertrand to the p-adic case **[8]**.

We can present an interesting generalization of (3) for the case of algebraic non periodic points of $\wp(z)$ [7]:

THEOREM 4. *Let* u *be an algebraic point of* $\wp(z)$ *linearly independent with* ω *over* Q. *Then*

(5) $$\zeta(u) - u\eta/\omega \ \textit{and} \ \eta/\omega \ \textit{are a.i.}$$

We cite two more of the author's results on algebraic independence for the case of complex multiplication.

THEOREM 5. *Let $\wp(z)$ have complex multiplication and let u be an algebraic point of $\wp(z)$; then*

(6) $$u \text{ and } \zeta(u) \text{ are a.i.}$$

This is the natural generalization of corollary 2.

THEOREM 6. *Let $\wp(z)$ have complex multiplication by β. Then for period ω of $\wp(z)$*

(7) $$\pi/\omega \text{ and } e^{\pi i\beta} \text{ are a.i.}$$

Also we have elliptic (and Abelian) analogues of the Lindemann–Weierstrass theorem (see below).

Let's see how our results can be interpreted in terms of some problems of partial differential equations. We'll present one application. Let $u(x)$ be a periodic (with period T) potential that is an n-band, or, in other words, a solution of stationary nth order Korteweg–de Vries equation (S. Novikov, [17], McKean): $\sum_{i=0}^{n} c_i R_i[u]=0$.

Then the spectrum for the Schrödinger $L=-d^2/dx^2+u(x)$ has n-band structure.

(8) $$\lambda_0<\lambda_1\leqslant\lambda_2<\lambda_3\leqslant\lambda_4<\ldots\leqslant\lambda_{2n}<+\infty,$$

where λ_i are first $2n+1$ points of T-periodic or T-anti-periodic spectrum of $L\psi=\lambda\psi$. All λ_j, $j=0,\ldots,2n$, are called "algebraic eigenvalues". However there are infinitely many (degenerate) points λ_i, $i=2n+2,\ldots,$ of periodic and antiperiodic spectrum. They are sometimes called "transcendental", but it is possible to give precise sense to this word and to prove that such numbers are indeed transcendental. Let's take the most known example of an n-band potential, the Lame potential $u(x)=n(n+1)\wp(x)$, where $\wp(x)$ has algebraic invariants.

COROLLARY 7. *For the Lame potential $u(x)=n(n+1)\wp(x)$ the first $2n+1$ points of periodic and antiperiodic spectrum are algebraic, while the others are transcendental. E.g. all eigenvalue μ_i of $L\psi=\mu\psi$ with periodic boundary conditions $\psi(0)=\psi(T)=0$ for $i=n+1,\ n+2,\ldots$ are transcendental.*

Another problem arises from Abelian varieties associated with Fermat curves. It is the problem on the transcendence of the values of Γ-function at rational, but not integer points. Of course, $\Gamma(1/2)=\sqrt{\pi}$ is transcendental. From Corollary 2 we see also that

(9) $$\Gamma(1/6),\Gamma(1/4),\Gamma(1/3),\Gamma(2/3),\Gamma(3/4),\Gamma(5/6)$$

are transcendental (and each of the numbers in (9) is a.i. with π).

These examples correspond to elliptic curves of CM-type, but in order to study general points $\Gamma(m/n)$ we must investigate arithmetic nature of the periods and quasi-periods of Abelian varieties of CM-type in the Shimura sense. This problem turns out to be the most important and difficult in Transcendence Theory. It is very interesting that the analytical difficulties in transcendence proof in this situation are tied up with algebraic difficulties.

Let $A=(A_1, \ldots, A_d, B)$ be Abelian variety over $\bar{Q}$ of dimension $d, H_1, \ldots, H_d$ be its quasi-periodic functions and let $\vec{\omega}, \ldots, \vec{\omega}_{2d}$ be periods of A (see supra).

We put

$$\begin{aligned}\vec{\eta}_i &= (\eta_{ij})_{j=1,\ldots,d}, \\ \vec{\omega}_i &= (\omega_{ij})_{j=1,\ldots,d},\end{aligned} \quad i = 1, \ldots, 2d. \tag{10}$$

The main question for Abelian functions can be formulated as follows: to determine the number D of algebraically independent elements among entries of $\Omega \cup H = \{\vec{\omega}_j, \vec{\eta}_j : j=1, \ldots, 2d\}$.

For general Abelian varieties there is not even a conjecture expressing this number $D=D(A)$ in terms of A (for the same dimension d, D may vary). Moreover, even the CM-case is extremely nontrivial.

For general Abelian variety we *only* have results on algebraic independence of *two* numbers.

THEOREM 8. *Let A be defined over $\bar{Q}$. Then* $\deg \operatorname{tr} Q(\Omega \cup H) \geqslant 2$. *Moreover for the numbers* (10) *we have*

$$\#[\{\omega_{ij}: i = 1, \ldots, 2d;\ j = 1, \ldots, d\} \cup \{\eta_{ij_0}: i = 1, \ldots, 2d\}] \geqslant 2 \tag{11}$$

for any $j_0 < d$, etc.

The difficulties on the analytic proofs of a.i. for more than 2 numbers (found in 1975–1976) suggest that there may exist some algebraic mechanisms responsible for them. Approximately at the same time (1976–1978), people working with Abelian varieties of CM-type have found new algebraic relations between periods (Deligne, Shimura, Gross, ...). Now we can formulate the problem precisely. Let A be CM-variety over $\bar{Q}$ of dimension d for CM-field K of degree $2d$ (as in § 1). We have representation of K in C^d by direct sum $\sum_{i=1}^{n} \tau_i$, where τ_i runs through a set of representatives of pairs of complex conjugate embedding of K in C.

Then (K, Φ) is called the CM-type of A for $\Phi = \sum_{i=1}^{n} \tau_i$. According to Shimura, Taniyama and T. Kubota we introduce the notion of the rank of CM-type (K, Φ). Let $I_K^0(Q)$ be the module of all formal linear combinations $\psi = \sum_\tau c_\tau \tau$ for $c_\tau \in Q$ of the injections τ of K into C such that $c_\tau + c_{\tau\varrho}$ does not depend on τ (ϱ is the complex conjugation). If M is the Galois closure of K over Q, then the rank $r(K, \Phi)$ of CM-type (K, Φ) is defined as the dimension of the subspace of $I_K^0(Q)$ generated over Q by $\Phi\gamma$ for all $\gamma \in \mathrm{Gal}\,(M/Q)$.

Then $r(K, \Phi) \leqslant d+1$ and rank $r(K, \Phi)$ is the main invariant connected with the division field A_l. Results of Shimura and Deligne show that for the CM-variety A and for corresponding CM-type (K, Φ) (which is then primitive)

(12) the number $D(A)$ of algebraically independent elements in $\Omega \cup H$ is at most the rank $r(K, \Phi)$ of (K, Φ).

E.g. Shimura gives examples of primitive CM-types (L, ψ) such that $[L: Q]=2^g$ (L, ψ) is primitive and for the corresponding CM-variety B, the $\dim(B)=2^{g-1}$ there are at most $g+1$ algebraically independent periods and quasi-periods.

Conjecture 9. For any CM-variety A and its CM-type (K, Φ) the number $D(A)$ is exactly the rank of (K, Φ):

$$\deg \operatorname{tr} Q(\Omega \cup H) = r(K, \Phi). \tag{13}$$

As a particular case of the conjecture (13) we see that for prime $l \geqslant 3$ the numbers $\pi, \Gamma(1/l), \ldots, \Gamma(n/l)$ for $n=(l-1)/2$ are a.i.

B. *Exponential case.* In the exponential case, instead of general results, we have at least the general

Schanuel Conjecture. Let $\alpha_1, \ldots, \alpha_n$ be complex numbers linearly independent over Q. Then among

$$\alpha_1, \ldots, \alpha_n, \quad e^{\alpha_1}, \ldots, e^{\alpha_n}$$

there are at least n algebraically independent numbers.

For many years the most important case of conjecture was the case of numbers of the form

$$\alpha^\beta, \ldots, \alpha^{\beta^{d-1}}$$

where α and β are algebraic, $\alpha \neq 0, 1$, and β of degree d.

Since 1949 the main target for investigation becomes the following sets of numbers. Let $\alpha_1, \ldots, \alpha_N$ and $\beta_1, \ldots, \beta_M$ be two sequences of complex numbers linearly independent over Q satisfying natural conditions on their measure of linear independence [9]:

$$S_1 = \{e^{\alpha_i \beta_j}\}, \quad S_2 = \{\beta_j, e^{\alpha_i \beta_j}\}, \quad S_3 = \{\alpha_i, \beta_j, e^{\alpha_i \beta_j}\} \tag{1}$$

$(i=1, \ldots, N,\ j=1, \ldots, M)$. Gelfond's method gives us the possibility to examine the case $\deg \operatorname{tr} Q(S_i) \geqslant 1, 2$, $i=1, 2, 3$. In the work of Tijdeman, Waldschmidt and Brownawell all the cases of one or two algebraically independent numbers in S_i were examined.

The situation with three or more algebraically independent numbers was more difficult. The main obstacles, as usual, are algebraico-geometrical. More precisely at the end of analytic proof of algebraic independence, we obtain a system of polynomials $P_i(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n]$, $i=1, 2, 3, \ldots,$ satisfying

$$|P_i(\theta_1, \ldots, \theta_n)| < H_i^{-d_i^\nu}, \quad i = 1, 2, 3, \ldots, \tag{2}$$

for fixed $(\theta_1, \ldots, \theta_n)$ and $\nu > n$ (and some additional conditions). Using resultants it is possible to treat the case $\nu > 2^n - 1$. These considerations enable the author to show, e.g. [9], [15], the following results:

$$\text{if} \quad |S_i|/(M+N) \geqslant 2^n, \quad \text{then} \quad \deg \operatorname{tr} Q(S_i) \geqslant n+1 \tag{3}$$

Since then we have found new methods for the investigation of problems like (2). These methods are based on some investigations of singularities for intersections of hypersurfaces $P_i=0, P_{i+1}=0, \ldots, P_{i+k}=0$.

We have some effective versions of Hörmander–Lojasiewich inequality and based on them we can improve (3) considerably: namely to change from 2^n to $n+1$ in general [6], [7]. Recently we found an even more elementary method of considering problems (2) for low n [11]. The most precise are the results for three or four a.i. numbers.

THEOREM 1. *Let S_i be as before and $2 \leqslant d \leqslant 4$. If*

(4) $$|S_i|/(M+N) \geqslant d \quad (\textit{and } |S_3|/(M+N) > d \textit{ for } i=3),$$

then

(5) *there are d algebraically independent numbers among elements in S_i.*

Results most close to the Schanuel conjecture are proved for example for algebraic powers of algebraic numbers:

THEOREM 2. *Let $\alpha \neq 0, 1$ be algebraic number and β be algebraic number of degree $d \geqslant 2$. Then among*

(6) $$\alpha^{\beta^j}, \quad j = 1, \ldots, d-1,$$

there are

(7) $[(d+1)/2]$ *algebraically independent numbers.*

This is the "half" of Schanuel's conjecture.

[In order to go further, we must change our analytic Gelfond–Schneider method; now algebraic part of the proof is very advanced, so we need new analytic, probably multidimensional, consideration.]

The number of problems connected with simplest cases is really unbounded, e.g.

Problem 3. Let β be quadratic irrationality and $\log\alpha_1, \log\alpha_2$ logarithms of algebraic numbers, linearly independent over Q. One should try to prove that $\log\alpha_1$ and α_1^β are a.i. and to prove that $\alpha_1^\beta, \alpha_2^\beta$ are a.i.

Present techniques give us the possibility of only adding to numers θ_1, θ_2 from Problem 3 a third number θ_3 (connected with exponent or elliptic function) and then showing $\#\{\theta_1, \theta_2, \theta_3\} \geqslant 2$. We present one such result to show what kind of monstrous things it is possible to get.

THEOREM 4. *Let $\wp(z)$ have complex multiplication by β, ω be a period of $\wp(z)$ and u be an algebraic point of $\wp(z)$. In the notations of Problem 3 we have:*

(8) $$\#\left\{\frac{\log\alpha_1}{\log\alpha_2}, \alpha_1^\beta, \alpha_2^\beta\right\} \geqslant 2;$$

(9) $$\#\{\log\alpha_1, \alpha_1^\beta, \alpha_2^\beta\} \geqslant 2;$$

(10) $$\#\{\pi, \pi^{\beta i}, e^{\pi\beta}\} \geqslant 2; \quad [\beta \notin Qi]$$

(11) $$\#\{\omega, \alpha_1^\beta, \alpha_2^\beta\} \geqslant 2;$$

(12) $$\#\left\{\frac{\pi}{\omega}, \alpha_1^\beta, \alpha_2^\beta\right\} \geqslant 2;$$

(13) $$\#\left\{\frac{\pi}{\omega}, \frac{\log\alpha_1}{\pi}, \alpha_1^\beta\right\} \geqslant 2;$$

(14) $$\#\left\{\frac{\pi}{\omega}, \log\alpha_1, \alpha_1^\beta\right\} \geqslant 2;$$

(15) $$\#\{u, \log\alpha_1, \alpha_1^\beta\} \geqslant 2.$$

Only result (8) is classical, (9), (10) were proved by the author by Baker's method, (11)–(15) were obtained in 1975–1977 [9], [7].

3. We present here another type of results that were recently obtained by the author again using algebraico-geometrical approach. This is an analogue of Lindemann–Weierstrass theorem for elliptic and Abelian functions. The most simple results we get in CM-case [12].

THEOREM 1. *Let* $\wp(z)$ *have complex multiplication in* $Q(\tau)$. *If* $\alpha_1, \ldots, \alpha_n$ *are algebraic numbers, linearly independent over* $Q(\tau)$, *then, for* $n \leqslant 6$, $\wp(\alpha_1), \ldots, \wp(\alpha_n)$ *are algebraically independent.*

THEOREM 2. *Let* $A=(A_1, \ldots, A_d, B)$ *be* CM-*variety of dimension d. If* $\vec{\alpha}_1, \ldots, \vec{\alpha}_k$ *are algebraic vectors (from* $\bar{Q}^d$*) linearly independent over* End (A), *then among*

$$A_i(\vec{\alpha}_j), \quad i = 1, \ldots, d; \quad j = 1, \ldots, k \quad \textit{for} \quad k \leqslant 6$$

there are $\geqslant k$ *a.i. numbers.*

When many coordinates of $\vec{\alpha}_i$ are zeroes we have better result:

THEOREM 3. *Let* $A=(A_1, \ldots, A_d, B)$ *be* CM-*variety. If* $\vec{\alpha}_1, \ldots, \vec{\alpha}_k \in \bar{Q}^d$ *are linearly independent over* End (A) *and all coordinates of* $\vec{\alpha}_i$ *but* k_0*th are zeroes* $(k_0 \leqslant d)$ $i=1, \ldots, k$, *then for* $kd \leqslant 6$ *the numbers*

$$A_i(\vec{\alpha}_j), \quad i = 1, \ldots, d, \quad j = 1, \ldots, k,$$

are a.i.

This result can be applied for the algebraic independence of $f(\alpha_1), \ldots, f(\alpha_n)$ for solution $f(z)$ of differential equation $P(f, f')=0$ when the curve $\mathscr{C}$: $P(x, y)=0$ is CM-curve of genus $\geqslant 1$ corresponding to the field K and $\alpha_1, \ldots, \alpha_n$ are linearly independent over K algebraic numbers (the generalization of Bombieri conjecture for CM-case).

4. The problem of transcendence is closely related to the problem of diophantine approximations of transcendental numbers by algebraic ones. We consider briefly new achievements in this direction. It is convenient to use the notion of type of transcendence τ of n numbers $(\theta_1, \ldots, \theta_n)$. This means that for

$$P(x_1, \ldots, x_n) \in Z[x_1, \ldots, x_n],\ P \not\equiv 0,\ |P(\theta_1, \ldots, \theta_n)| > \exp\left(-C(\log H(P)+d(P))^\tau\right)$$

with $C>0$. Of course, $\tau \geqslant n+1$, but for concrete numbers it is extremely difficult to show that $\tau < \infty$ and to find τ. Only in the case $n=1$ for numbers connected with Gelfond–Schneider method we have now complete information, thanks to... Cijsouw, Waldschmidt (exponent), Reyssat (elliptic case). For $n \geqslant 2$ the first examples of two numbers with finite type of transcendence were given in 1975 (these are numbers from Corollary 2, Theorem 3, § 2). We have proved for them $\tau \leqslant 6+\varepsilon$ for any $\varepsilon > 0$. Now the analysis of singular points of intersections of curves [11] gives us already the best possible result.

THEOREM 1. *For a.i. numbers* π/ω, η/ω (*see* (3), § 2A), *the type* $\tau=\tau(\pi/\omega, \eta/\omega)$ *is* $3\leq\tau\leq 3+\varepsilon$ *for any* $\varepsilon>0$. *Moreover for* $P(x, y)\in Z[x, y], P\not\equiv 0$;

(1) $|P(\pi/\omega, \eta/\omega)| > \exp\left(-c_1\left(\log H(P)+d(P)\right)d(P)^2\log^3\left(d(P)\log H(P)\right)\right).$

Other numbers in § 2A have also finite type τ. For numbers $\zeta(u)-u\eta/\omega, \eta/\omega$ from Theorem 4 we have $\tau\leq 9+\varepsilon$ for any $\varepsilon>0$; for π/ω and $e^{\pi i\beta}$ from Theorem 6 $\tau\leq 7+\varepsilon$ for any $\varepsilon>0$.

New (algebraico-geometrical) methods can give also good results for numbers in § 3 in the case when the height is much more than the degree **[12]**.

THEOREM 2. *Let* $\wp(z)$ *have complex multiplication by* $Q(\tau)$ *and* $\alpha_1, \ldots, \alpha_n$ *be algebraic numbers linearly independent over* $Q(\tau)$ *and* $n=1, 2$. *If*

$$P(x_1, \ldots, x_n)\in Z[x_1, \ldots, x_n],\ P\not\equiv 0 \quad \textit{and} \quad \log\log H(P) > c_2 d^{2n+1},$$

then

$$\left|P(\wp(\alpha_1), \ldots, \wp(\alpha_n))\right| > H^{-c_3 d^n}.$$

5. Methods of proof in Transcendence Theory are now very sophisticated algebraic or multidimensional. One of the most difficult problems in the one-dimensional case is the problem of estimates of the number of zeroes of functions $F(z)=P(f_1(z), \ldots, f_n(z))$ for $P(x_1, \ldots, x_n)\in C[x_1, \ldots, x_n]$ and entire functions $f_1(z), \ldots, f_n(z)$. [E.g. this problem is responsible for difficulties of generalization of a.i. results in § 2B for elliptic case; for elliptic analogue $\wp(u\beta)$, $\wp(u\beta^2)$ of Gelfond's example of 2 a.i. numbers etc.] Recent very important progress in this direction belongs to D. Brownawell–D. Masser; their investigations are based on analysis of primary ideals in polynomial rings. Their results are complete for $n=2$ and f_1, f_2 satisfying algebraic differential equation.

In the multidimensional case we still need general form of Schwartz lemma. Previous results by E. Bombieri, M. Waldschmidt and author **[1]**, **[16]** giving such lemma for many cases, show that Schwartz lemma depends on "singular properties" of hypersurfaces in C^n. We propose one problem: let $\Omega(S; K)=\min\{\deg P\colon P(\bar{x})\in C[\bar{x}], \partial^k P(\bar{w})=0 \text{ for all } k\in N^n, |k|<K, \bar{w}\in S\}$ for finite $S\subset C^n$. For a given $|S|$ to describe all values of $\Omega_0(S)=\lim_{K\to\infty}\Omega(S; K)/K$.

Using properties of $\Omega_0(S)$ **[14]** we find e.g. that for meromorphic transcendental function $f(\bar{z})$ in C^n of order $\leq\varrho$ the set $S(f)$ of $\bar{w}\in\bar{Q}^n$ with $\partial^k f(\bar{w})\in Z$ for all $k\in N^n$ is contained in a hypersurface of degree $\leq n\varrho$ **[16]**.

I would like to thank M. Waldschmidt, D. Bertrand, D. Brownawell and D. Masser for helpful discussions.

Added in Proof. (1) We have a proof of the conjecture in § 1 in the case of the CM-variety A and arbitrary $m\geq 2$. This enables us to prove explicit upper bounds for the K-integer points on CM-curves $\mathscr{C}$ using the B-Sw-D conjecture and bounds (4) or (5) in § 1 in terms of the conductor N of $J(\mathscr{C})(K)$. For an elliptic CM-curve $\mathscr{C}$ over Q the corresponding bounds have the form $H(P)\leq\exp\left(c(\varepsilon)N^{1/12+\varepsilon}\right)$, $c(\varepsilon)>0$.

(2) The result of Corollary 7 is generalized for any n-band potential $u(x)$ satisfying the algebraic differential equation over $\bar{Q}[x]$.

References

1. *Transcendence theory: Advances and applications,* (A. Baker and D. Masser, eds.) Academic Press, New York, 1977.

2. A. Baker, *Transcendental number theory,* Cambridge, 1975; see also Gottingen Narch. **16** (1969), 145.

3. D. Masser, *Elliptic function and transcendence,* Lecture Notes in Math., vol. 437, Springer-Verlag, Berlin and New York, 1975.

4. Proc. London Math. Soc. **33** (1976), 549; Invent. Math. **45** (1978), 61; Compositio Math. **35** (1977), 239.

5. J. Coates and S. Lang, Invent. Math. **34** (1976), 129; see also S. Lang, *Elliptic curves, diophantine analysis,* Springer-Verlag, Berlin and New York, 1978.

6. G. V. Chudnovsky, Dokl. Ukrain. Acad. Sci. **8** (1976), 697; see also Pullman Conf., June, 1975, Notices Amer. Math. Soc. **22** (1975), 486; Oberwolfach Conference, 1977.

7. *Transcendences arising from exponential and elliptic functions,* Bull. Soc. Math. France (1979) (to appear).

8. D. Bertrand, *Approximations diophantiennes p-adiques et courbes elliptiques,* Thèses, Paris, 1977; see also Sem. D-P-P, April, 1978; see also D. Bertrand, Y. Flicker, Acta Arith (to appear).

9. G. V. Chudnovsky, Math. Notes **15** (1974), 391; Soviet Math. Dokl. **15** (1974), 1424; Colloq. Math. Societat. J. Bolyai, vol. 13, North-Holland, Amsterdam, 1974, pp. 19—30; Uspehi Math. Sci. **31** (1976), 281.

10. M. Waldschmidt. *Nombres transcendants,* Lecture Notes in Math., vol 402, Springer-Verlag, Berlin and New York, 1974.

11. G. V. Chudnovsky, *Algebraic grounds for the proof of algebraic independence* I, I.H.E.S. 1978; II, Comm. Pure Appl. Math. (to appear).

12. *Algebraic independence of the values of elliptic functions at algebrai c points* C. R. Acad. Sci. Paus (1979) see also *Values of Abelian functions at algebraic points,* Proc. Nat Acad. Sci. (to appear).

13. S. Lang, *Transcendental numbers,* Addison-Wesley, Reading, Mass., 1966.

14. G. V. Chudnovsky, C. R. Acad. Sci. Paris **286** (1978), 1133; See also Sem. D-P-P, November 1977.

15. M. Waldschmidt, *Les travaux de G. V. Choodnovsky sur les nombres transcendants.* Sem. Bourbaki, Lecture Notes in Math., vol. 567, Springer-Verlag, Berlin and New York, 1977, pp. 274—299; see also Sem. D-P-P, June 1976, pp. G01—G17; *Nombres transcendants II;* see also D. Bertrand, Sem. D-P-P, 17 (1975/76), p. G—8.

16. G. V. Chudnovsky, Preprint I.H.E.S., 1977; Ann. of Math. (to appear), see also Sem. D-P-P, January 1978.

17. D. V. Chudnovsky and G. V. Chudnovsky, *Seminaire sur equation non lineaire,* I, II, Ecole Polytechnic, Paris, 1977—1978.

18. W. W. D. Brownawell and D. W. Masser, *Multiplicity estimates for analytic functions* I, II (to appear).

Columbia University
New York, New York 10027, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Arithmetic of Elliptic Curves with Complex Multiplication

John Coates

Introduction. Although it has occupied a central place in number theory for almost a century, the arithmetic of elliptic curves is still today a subject which is rich in conjectures, but sparse in definitive theorems. In this lecture, I will only discuss one main topic in the arithmetic of elliptic curves, namely the conjecture of Birch and Swinnerton-Dyer. We briefly recall how this conjecture arose (for more detailed accounts, see [1] and [5]). For simplicity, we shall only consider elliptic curves defined over the rational field $\boldsymbol{Q}$. Let E be an elliptic curve defined over $\boldsymbol{Q}$, which we can suppose to be given in Weierstrass normal form

(1) $$y^2 = 4x^3 - g_2x - g_3$$

with $g_2, g_3 \in \boldsymbol{Z}$. By the celebrated theorem of Mordell and Weil, the group $E(\boldsymbol{Q})$ of $\boldsymbol{Q}$-rational points of E is finitely generated. Write $g_{\boldsymbol{Q}}$ for the rank of $E(\boldsymbol{Q})$ modulo torsion. It is well known that the proof of Mordell and Weil yields no algorithm for determining $g_{\boldsymbol{Q}}$, nor even for deciding whether $g_{\boldsymbol{Q}}$ is positive. In attempting to attack this problem by using the idea of a quantitative local to global principle as in Siegel's work on quadratic forms, Birch and Swinnerton-Dyer were led to study the Hasse–Weil L-series $L(E, s)$ of E. This is a complex L-function, which is defined in the half plane $R(s) > \frac{3}{2}$, by the Euler product

$$L(E, s) = \prod_{p \text{ good}} (1 - a_p p^{-s} + p^{1-2s})^{-1};$$

here the product is taken over all primes p where E has a good reduction, and $a_p = p + 1 - N_p$, where N_p is the number of solutions (including the point at infinity) of the congruence $y^2 \equiv 4x^3 - g_2x - g_3 \bmod p$. It is conjectured that $L(E, s)$

can be analytically continued over the whole complex plane, but this is unknown in general.

Conjecture of Birch and Swinnerton-Dyer. Assuming the analytic continuation of the L-series, then $L(E, s)$ *has a zero at* $s=1$ *of order precisely* g_Q.

They also gave a conjectural formula for the coefficient of $(s-1)^{g_Q}$ in the expansion of $L(E, s)$ about $s=1$, but we shall not discuss this here. Now Tate's work [5] on the geometric analogue suggests that, in order to attack this conjecture, one must relate $L(E, s)$ to the characteristic polynomial of some canonical element in a representation of a certain Galois group. In the rest of my lecture, I want to discuss some joint work with A. Wiles, which goes a little way in this direction for the special case in which E admits complex multiplication.

Thus we assume from now on that the endomorphism ring of E is an order in an imaginary quadratic field K. Let E_{tor} be the group of all torsion points on E, and μ the group of all roots of unity. The theory of complex multiplication grew out of the analogy, perceived by Kronecker, between the elliptic extension $K(E_{\text{tor}})/K$ and the cyclotomic extension $\boldsymbol{Q}(\mu)/\boldsymbol{Q}$. It establishes that $K(E_{\text{tor}})/K$ is an abelian extension, and gives an explicit description of the action of the Galois group on E_{tor} (see [4]). Deuring and Weil deduced from this latter fact that $L(E, s)$ can be identified with a Hecke L-function $L(\psi, s)$, where ψ is a certain Grossencharacter of K, which we call the Grossencharacter of E. In particular, the analytic continuation of $L(E, s)$ follows. Eisenstein seems to have been the first to envisage pursuing more deeply the analogy between the arithmetic of the extensions $K(E_{\text{tor}})/K$ and $\boldsymbol{Q}(\mu)/\boldsymbol{Q}$, but his brief life did not allow time for developing the idea. The work of Wiles and myself is entirely within the spirit of Eisenstein's idea, and has been inspired by Iwasawa's work on cyclotomic fields.

At present, it is only plausible to relate $L(E, s)$ to a representation of a Galois group via its p-adic analogue, whose existence had been proven earlier by Katz, Lichtenbaum, and Manin–Vishik. We recall their definition. For simplicity, we assume from now on that the endomorphism ring of E is the full ring O of integers of K (this can always be achieved by replacing E by a curve isogenous to it over $\boldsymbol{Q}$). Let S be the set consisting of 2, 3, and all primes where E has a bad reduction. We now choose p to be an arbitrary rational prime satisfying (i) $p \notin S$, and (ii) p splits in K, say $(p)=\mathfrak{p}\bar{\mathfrak{p}}$. Let $L(\psi^k, s)$ be the Hecke L-function of the Grossencharacter of ψ^k. Let Ω_∞ be a generator of the period lattice of (1) as an O-module. It is well known that the numbers $\Omega_\infty^{-k} L(\psi^k, k)$ $(k=1, 2, \ldots)$ belong to K, and consequently they can be viewed as lying inside the completion $\boldsymbol{C}_\mathfrak{p}$ of the maximal unramified extension of the completion of K at $\mathfrak{p}$. Let $\mathscr{I}_\mathfrak{p}$ be the ring of integers of $\boldsymbol{C}_\mathfrak{p}$. Let $f \in K$ be a fixed generator for the conductor of ψ. By definition, the $\mathfrak{p}$-adic analogue of $L(E, s)$ is the unique continuous function $L_\mathfrak{p}(E, s)\colon \boldsymbol{Z}_p \to \mathscr{I}_\mathfrak{p}$ satisfying

$$L_\mathfrak{p}(E, k) = v_k \Omega_\mathfrak{p}^{1-k} \Omega_\infty^{-k} L(\psi^k, k)\left(1 - \frac{\psi(\mathfrak{p})^k}{N\mathfrak{p}}\right)$$

for all integers $k>0$ with $k\equiv 1 \mod (p-1)$; here $v_k=12(-1)^{k-1}(k-1)!f^{-k}$, and $\Omega_{\mathfrak{p}}$ is a certain unit in $\mathscr{I}_{\mathfrak{p}}$, which should be viewed as a $\mathfrak{p}$-adic analogue of Ω_∞. Not only is $L_{\mathfrak{p}}(E,s)$ continuous, but it is in fact an Iwasawa function. Fix a topological generator u of the units in $\boldsymbol{Z}_p$ which are $\equiv 1 \mod p$. Then there exists $\mathscr{L}(E,T)$ in the ring $\mathscr{I}_{\mathfrak{p}}[[T]]$ of formal power series in T with coefficients in $\mathscr{I}_{\mathfrak{p}}$ such that, for all s in $\boldsymbol{Z}_p$,

$$L_{\mathfrak{p}}(E,s)=\mathscr{L}(E,u^s-1). \tag{2}$$

Iwasawa modules. We now describe two Iwasawa modules which are related to $L_{\mathfrak{p}}(E,s)$, one conjecturally and the other demonstrably. For each $n\geqq 0$, let $E_{\mathfrak{p}^{n+1}}$ be the group of $\mathfrak{p}^{n+1}$-division points on E, and put $E_{\mathfrak{p}^\infty}=\bigcup_{n\geqq 0}E_{\mathfrak{p}^{n+1}}$. Define

$$F_n=K(E_{\mathfrak{p}^{n+1}}),\quad F_\infty=K(E_{\mathfrak{p}^\infty}).$$

We write G_∞ for the Galois group of F_∞ over K, and $\varkappa\colon G_\infty\to\boldsymbol{Z}_p^\times$ for the canonical character giving the action of G_∞ on $E_{\mathfrak{p}^\infty}$. Note that $G_\infty=\Delta\times\Gamma$, where $\Delta=G(F_0/K)$ and $\Gamma=G(F_\infty/F_0)$. Let χ denote the restriction of $\varkappa$ to Δ.

Let M_∞ be the maximal extension of F_∞ such that (i) M_∞/F_∞ is an abelian p-extension, (ii) M_∞/F_∞ is unramified outside $\mathfrak{p}$, and (iii) M_∞/K is Galois (so that G_∞ operates on $G(M_\infty/F_\infty)$ via inner automorphisms) and Δ operates on $G(M_\infty/F_\infty)$ via χ. It is easy to see that such a field exists, and we put

$$X_\infty=G(M_\infty/F_\infty).$$

An analysis of the proof of the Mordell–Weil theorem shows that the G_∞-module X_∞ is deeply connected with the arithmetic of E. In particular, let N_∞ be the field obtained by adjoining to F_∞ the coordinates of all $\mathfrak{p}^{n+1}$-division points ($n=0,1,\ldots$) of all points in the group $E(K)$ of K-rational points of E. Let $T_{\mathfrak{p}}=\varprojlim E_{\mathfrak{p}^{n+1}}$ be the Tate module, viewed as a G_∞-module in the natural way.

LEMMA. *We have $N_\infty\subset M_\infty$, and $G(N_\infty/F_\infty)$ is isomorphic as a G_∞-module to $T_{\mathfrak{p}}^{g_Q}$.*

It is easy to interpret this lemma as asserting the existence of a zero at $s=1$ of multiplicity $\geqq g_Q$ of a certain p-adic function. Let γ be the unique topological generator of Γ such that $\varkappa(\gamma)=u$. Then, if $\Lambda=\boldsymbol{Z}_p[[T]]$ is the ring of formal power series in T with coefficients in $\boldsymbol{Z}_p$, the Γ-module X_∞ has a unique Λ-module structure satisfying $(1+T)x=\gamma x$. It can be shown using class field theory that X_∞ is a finitely generated Λ-torsion Λ-module. Hence, by the structure theory, we have

$$X_\infty\sim\bigoplus_{j=1}^{r}\Lambda/(f_j),$$

where r is an integer $\geqq 1$, and $f_1,\ldots,f_r$ are nonzero power series in Λ (also $A\sim B$ means that there is a Λ-homomorphism from A to B with finite kernel and cokernel). The power series $f=f_1,\dots,f_r$ is then uniquely determined by X_∞ up to multiplication by a unit in Λ. One sees immediately that the above lemma implies that $f(u^s-1)$ has a zero at $s=1$ of order $\geqslant g_Q$.

Main Conjecture. The power series $f(T)$ and $\mathscr{L}(E,T)$ generate the same ideal in $\mathscr{I}_\mathfrak{p}[[T]]$.

This conjecture undoubtedly lies very deep. Its truth would imply (i) that $L_\mathfrak{p}(E,s)$ has a zero at $s=1$ of order $\geqslant g_Q$, and (ii) the p-part of the more refined conjecture of Birch and Swinnerton-Dyer when $g_Q=0$. The cyclotomic analogue is still unproven in general, and is perhaps the most important open problem in the theory of cyclotomic fields.

In view of the difficulty of the Main Conjecture, it is a little surprising that one can solve the analogous problem for a closely related G_∞-module. To motivate this, let L_∞ be the maximal unramified extension of F_∞ contained in M_∞. Then global class field theory provides the following explicit description of $Z_\infty=G(M_\infty/L_\infty)$. For each $n\geqslant 0$, let U_n be the local units $\equiv 1$ in the completion of F_n at the unique prime $\mathfrak{p}_n$ above $\mathfrak{p}$. Let $\bar{E}_n$ be the closure in U_n of the global units of F_n which are $\equiv 1 \bmod \mathfrak{p}_n$. Write $(U_n/\bar{E}_n)^{(1)}$ for the eigenspace of $U_n/\bar{E}_n$ on which Δ acts via χ. Then

$$Z_\infty=\varprojlim (U_n/\bar{E}_n)^{(1)},$$

where the projective limit is taken with respect to the norm maps. Now let $\bar{C}_n$ be the closure in U_n of Robert's group of elliptic units of F_n (see [2] or [3] for the precise definition). Define the G_∞-module

$$Y_\infty=\varprojlim (U_n/\bar{C}_n)^{(1)},$$

where again the projective limit is taken with respect to the norm maps, and the superscript 1 denotes the eigenspace on which Δ acts via χ.

THEOREM. *We have $Y_\infty\sim\Lambda/(g)$, where the power series $g(T)$ and $\mathscr{L}(E,T)$ generate the same ideal in $\mathscr{I}_\mathfrak{p}[[T]]$.*

This is proven in [3] when p is not anomalous for E, and similar arguments give the result in general. Iwasawa had previously established the cyclotomic analogue of this theorem, but his methods do not carry over to the elliptic case. The proof in [3] is based on a refinement of Kummer's notion of the logarithmic derivatives of a p-adic number.

We can use the above theorem to derive a partial result about the zero of $L_\mathfrak{p}(E,s)$ at $s=1$. It can be shown (see Theorem 11 of [2]) that $G(N_\infty L_\infty/L_\infty)$ is isomorphic as a G_∞-module to $T_\mathfrak{p}$ if $g_Q\geqslant 1$. Note that it is not isomorphic to $T_\mathfrak{p}^{g_Q}$, as one might naively expect (we owe this remark to R. Greenberg). Thus, since the natural map from Y_∞ to Z_∞ is surjective, it follows that Y_∞ has a G_∞-quotient isomorphic to $T_{\mathfrak{p}}$ when $g_Q \geqslant 1$. We conclude from the above theorem that $L_{\mathfrak{p}}(E, s)$ vanishes at $s=1$ if $g_Q \geqslant 1$. Since the values of the complex and $\mathfrak{p}$-adic L-series of E are essentially the same at $s=1$, we have proven the following.

COROLLARY. *Let E be an elliptic curve defined over Q, with complex multiplication. If $E(Q)$ contains a point of infinite order, then $L(E, s)$ vanishes at $s=1$.*

The present proof is slightly different from that given in [2], in that it uses a single prime p rather than infinitely many primes.

References

1. B. Birch and H. Swinnerton-Dyer, *Notes on elliptic curves*. II, J. Reine Angew. Math. **218** (1965), 79—108.

2. J. Coates and A. Wiles, *On the conjecture of Birch and Swinnerton-Dyer*, Invent. Math. **39** (1977), 223—251.

3. ——, *On p-adic L-functions and elliptic units*, J. Austral. Math. Soc. **26** (1978), 1—25.

4. G. Shimura, *Introduction to the arithmetic theory of automorphic functions*, Publ. Math. Soc. Japan **11** (1971).

5. J. Tate, *On the conjectures of Birch and Swinnerton-Dyer and a geometric analogue*, Sém. Bourbaki **18** (1966), no. 306.

UNIVERSITÉ DE PARIS–SUD
ORSAY, FRANCE

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Sieve Methods

Henryk Iwaniec

Dedicated to the memory of Viggo Brun

1. Introduction. In the early twenties of this century Viggo Brun [4] introduced a method which proved to be one of the most fruitful tools in elementary number theory. Since the power of his sieve method was realized the latter began developing rapidly engaging the attention of many prominent number-theoreticians of the past years. We now have a great number of variations of the method, the most remarkable being Selberg's λ^2-method and the combinatorial sieve of Rosser. Of particular interest is also the recent asymptotic sieve of Bombieri [3] (see also [7]) which rests on an entirely different concept, called by Selberg "local sieve". The large sieve is in a sense not a sieve. Nevertheless these three topics are intimately related through the similarity of the arithmetical applications.

In view of the diversity of different sieve techniques it is clearly impossible to give a complete presentation of the theory in a short time. I wish to devote my lecture rather exclusively to the combinatorial sieve of Rosser for two reasons: the first is the number of strong results it originated recently; the second reason is my personal interest. There is also no need to describe the λ^2-method since Selberg himself did it at the Harvard Congress. A complete exposition of the theory based on Selberg's sieve can be found in the book [10] of Halberstam and Richert; the theory of the large sieve is presented in [2] and [28].

2. Background. Let me first introduce some principal notions of sieve theory. We consider a finite sequence of integers $\mathscr{A}$ and a set of primes $\mathscr{P}$. For a given $z \geqslant 2$ let $P(z) = \prod_{p<z,\, p\in P} p$. One of the fundamental problems in the sieve theory is to estimate from above and below the so-called sifting function

$$S(\mathscr{A}, z) = |\{a \in \mathscr{A};\ (\mathscr{A}, P(z)) = 1\}|$$

which represents the number of elements in $\mathscr{A}$ that have no prime factors $p<z$ in P. In any sieve method the relevant information about the sequence $\mathscr{A}$ is furnished by the quantities $|\mathscr{A}_d|$ where

$$|\mathscr{A}_d| = |\{a\in\mathscr{A};\ a \equiv 0 \pmod d\}|,$$

i.e. $|\mathscr{A}_d|$ is the number of elements in $\mathscr{A}$ that are divisible by d with $d|P(z)$. Letting $\mu(d)$ be the Möbius function one has the exact formula of Legendre

$$S(\mathscr{A}, z) = \sum_{d|P(z)} \mu(d)\,|\mathscr{A}_d|. \tag{1}$$

This, however, is of rather limited use because, unless z is very small the right-hand side contains a number of terms $|\mathscr{A}_d|$ about which we know very little (but see [21]). The essence of the combinatorial sieve consists in truncating the function $\mu(d)$ so that every integer sequence $\mathscr{A}$ might satisfy

$$\sum_{\substack{d|P(z)\\ d\in\mathscr{D}^-}} \mu(d)\,|\mathscr{A}_d| \leqslant S(\mathscr{A}, z) \leqslant \sum_{\substack{d|P(z)\\ d\in\mathscr{D}^+}} \mu(d)\,|\mathscr{A}_d|. \tag{2}$$

Instead of the exact formula (1) we have now a two-sided estimate for $S(\mathscr{A}, z)$ and this is the cost of reducing the number of terms $|\mathscr{A}_d|$ that need be considered.

Let us consider the relation of Buchstab and Legendre

$$S(\mathscr{A}, z) = S(\mathscr{A}, z_1) - \sum_{z_1 \leqq p<z} S(\mathscr{A}_p, p)\ \ (z_1 < z). \tag{3}$$

This formula plays an important role in each of the two known ways of introducing Rosser's sieve. For Buchstab this formula furnished a way of improving the result of Brun. From a given pair of upper and lower estimates of the S-function, he gets, on introducing it in (3), another pair that in some range of the involved parameters proves to be better than the initial one. It is then possible to repeat this procedure until the resulting estimates become invariant under the substitutions. Neither Buchstab nor anybody else has ever reached the last step. Recently it has been recognized that essentially the same results may be obtained more directly through the inequality (2). In the latter approach, due to Rosser, the subsets $\mathscr{D}^+$ and $\mathscr{D}^-$ are constructed as follows

$$\mathscr{D}^+ = \{d = p_1 \dots p_r;\ p_r < \dots < p_1,\ p_1 \dots p_{2l-1} p_{2l-1}^{\beta} < D \quad \text{for all} \quad 1 \leqslant l \leqslant (r+1)/2\},$$
$$\mathscr{D}^- = \{d = p_1 \dots p_r;\ p_r < \dots < p_1,\ p_1 \dots p_{2l} p_{2l}^{\beta} < D \quad \text{for all} \quad 1 \leqslant l \leqslant r/2\}$$

where the two parameters $\beta\ (\geqslant 1)$ and $D\ (\geqslant 2)$ are to be chosen according to a principle we describe later. Although such a construction may look somewhat mysterious a reasonable motivation can be given (based on the Buchstab–Legendre formula). For the lack of time let me remark only that the concept of a sieving limit (first treated in detail by Selberg [32]) plays a decisive role in the truncating operation on $\mu(d)$, the limit being related to the exponent β.

All the concepts so far described provide us with a general setting of the theory. In order to go further it is necessary to make some hypothesis about regularity of the sequence which is to be sifted. We assume, which frequently happens in practice, that every $|\mathscr{A}_d|$ may be written in the form

$$|\mathscr{A}_d| = \omega(d)X/d + r(\mathscr{A}, d) \tag{4}$$

where $\omega(d)$ is multiplicative with $0 \leqslant \omega(p) < p$, X is a large enough parameter independent of d and $r(\mathscr{A}, d)$ is considered as an error term, small on average (so X approximates to $|\mathscr{A}|$). This means that for some $\alpha > 0$ and all $\varepsilon > 0$, $A > 0$ we have

$$R(\mathscr{A}, X^{\alpha-\varepsilon}) = \sum_{d|P(z), d<X^{\alpha-\varepsilon}} |r(\mathscr{A}, d)| < X(\log X)^{-A} \tag{5}$$

provided $X > X(\alpha, \varepsilon, A)$. As for the function $\omega(d)$ we assume that for all $z > w \geqslant 2$ and some $\varkappa > 0$ and $K > 1$ the following holds

$$\prod_{\substack{w \leqq p < z \\ p \in \mathscr{P}}} \left(1 - \frac{\omega(p)}{p}\right)^{-1} \leqslant \left(\frac{\log z}{\log w}\right)^{\varkappa} \left(1 + \frac{K}{\log w}\right). \tag{6}$$

Comparing the formula with Mertens' prime number theorem one can read $\varkappa$ as an average estimate of $\omega(p)$ over the primes from $\mathscr{P}$.

These two parameters α and $\varkappa$ are the most important in the theory; $\varkappa$ is called the dimension and α, for obvious reasons, may be called the level of uniform distribution of $\mathscr{A}$ in arithmetic progressions. Let me give an example. If our interest is in the Goldbach problem, then we may take

$$\mathscr{A} = \{N-p;\ p < N\} \quad \text{and} \quad P = \{p;\ p \nmid N\},$$

where N is an even integer. We have $|\mathscr{A}_d| = \pi(N, d, N)$, so that the prime number theorem suggests setting $X = LiN$ and $\omega(d) = d/\varphi(d)$. Therefore we find that $\varkappa = 1$ and $\alpha \geqslant \frac{1}{2}$ by the Bombieri–Vinogradov theorem.

3. General results. Two problems arise when approximation (4) is introduced into the basic estimate (2). The first one which deals with the error terms $r(\mathscr{A}, d)$ is easy to handle. Since $\beta \geqslant 1$ we immediately deduce that every d in $\mathscr{D}^+$ or $\mathscr{D}^-$ is less than D so the total error term does not exceed $R(\mathscr{A}, D) < X(\log X)^{-A}$ provided $D < X^{\alpha-\varepsilon}$. The problem of evaluating the sum of the main terms is much more intricate. After having rearranged the order of summation we get a sum of a number of terms, each of them of the same order of magnitude as

$$V(z) = \prod_{p|P(z)} (1 - \omega(p)/p).$$

The assumption (6) is then sufficient to determine their asymptotic bounds and to sum all quantities together. The problem of convergence is very difficult and a number of sophisticated questions from the field of differential-difference equations have come into prominence. The speaker succeeded in finding the following estimates:

$$\begin{aligned} S(\mathscr{A}, z) &\leqslant XV(z)\{F(s) + E(z, D)\} + R(\mathscr{A}, D), \\ S(\mathscr{A}, z) &\geqslant XV(z)\{f(s) - E(z, D)\} - R(\mathscr{A}, D), \end{aligned} \tag{7}$$

where $s=\log D/\log z$, $E(z, D)=c(\varkappa)e^{K-s}(\log D)^{-1/3}$ and $f(s)$, $F(s)$ are continuous functions to be determined by the differential-difference problem

$$s^{\varkappa}F(s) = A \quad \text{if} \quad s \leqslant \beta+1 \quad \text{and} \quad (s^{\varkappa}F(s))' = \varkappa s^{\varkappa-1}f(s-1) \quad \text{if} \quad s > \beta+1;$$
$$s^{\varkappa}f(s) = B \quad \text{if} \quad s \leqslant \beta \quad \text{and} \quad (s^{\varkappa}f(s))' = \varkappa s^{\varkappa-1}F(s-1) \quad \text{if} \quad s > \beta.$$

The optimal β (sieving limit) should be equal to $\inf\{\beta;\, f(s)>0 \text{ for all } s>\beta\}$ (or $\beta=1$ if $f(s)>0$ for all s) and the correct choice of A and B is to be inferred from the behaviour $f(s)=1+O(e^{-s})$ and $F(s)=1+O(e^{-s})$ as s approaches ∞.

In his very important paper [32], Selberg proposed the technique of the Laplace transform. Another method has been developed by the speaker [22] (see also [17]). It rests on a certain relation between two equations of the type

$$sp'(s) = -ap(s)-bp(s-1) \tag{8}$$

and

$$(sq(s))' = aq(s)+bq(s+1), \tag{9}$$

the latter being in a sense conjugate to the former. Of great importance are the solutions $g(s)$ and $h(s)$ of the equation (9) with coefficients $(a, b)=(\varkappa, \varkappa)$ and $(a, b)=(\varkappa, -\varkappa)$ respectively. It is shown that if $\varkappa>1/2$ then $\beta-1$ is the largest zero of $g(s)$, $A=(\beta-1)^{\varkappa-1}/h(\beta-1)$ and $B=0$. If $\varkappa=1/2$ then $\beta=1$, $A=2(e^{\gamma}/\pi)^{1/2}$ and $B=0$. The case $\varkappa<1/2$ is somewhat complicated, so we only mention that in this case $\beta=1$ and $0<B<1<A$.

There is strong evidence that in the range $1/2\leqslant\varkappa\leqslant 1$ the estimates (7) are optimal. This means that for particular sequences one of the two bounds (7) coincides with the asymptotic value of $S(\mathscr{A}, z)$. As yet Selberg [32] constructed such examples only with $\varkappa=1$ and $\varkappa=1/2$. Several asymptotic formulas for the half dimensional sieve are given in [19].

In large dimensions Selberg's λ^2-method gives stronger upper bounds than the method of Rosser. Consequently, Selberg's upper bounds, when used to start the Buchstab iteration procedure, must lead to better lower bounds as well. Ankeny and Onishi [1] carried out the first iteration, getting results that are better indeed. It is best visible when comparing their sieving limit $\nu_{\varkappa}$ with our $\beta_{\varkappa}$

$$\nu_{\varkappa} \sim \eta\varkappa \quad \text{as} \quad \varkappa\to\infty \quad \text{where} \quad \eta = \frac{2}{e\log 2}\exp\left(\int_0^2 \frac{e^u-1}{u}\,du\right) = 2.44\ldots,$$

$$\beta_{\varkappa} \sim c\varkappa \quad \text{as} \quad \varkappa\to\infty \quad \text{where} \quad c\log c = c+1 \quad (c = 3.5911\ldots).$$

One further iteration was executed recently by Porter [30]; the relative improvement is however very slight.

4. Refinements of the method. The general estimates of the sifting function can be enhanced in particular applications by other methods. One of such refinements due to Kuhn [27] was presented to the Congress in Amsterdam.

Let P_r denote an almost-prime of order r, i.e. a number having at most r prime factors counted with their multiplicity. For $r=1$, P_1 is simply a prime. It is not difficult to show that if $\mathscr{A}$ is a sequence of integers bounded by $c(\varepsilon)X^{\gamma+\varepsilon}$ and α, β have the same meaning as in section 2 then, whenever X is sufficiently large and $r>\beta\gamma/\alpha$, $\mathscr{A}$ contains at least one P_r of order r. Kuhn was the first to observe that by means of weights of a certain kind attached to the elements of the sequence in question one can obtain much stronger results concerning almost-primes. Let $r(\varkappa, \gamma/\alpha)$ stand for the least integer r such that every sequence $\mathscr{A}$ long enough and pertaining to the parameters $\varkappa, \alpha$ and γ must necessarily contain an almost-prime of order r. It would be very interesting, and useful, to find weights enabling one to compute the optimal order $r(\varkappa, \gamma/\alpha)$. Nowadays we are at the stage of experiment. Richert's logarithmic weights **[31]** received much attention. A recent modification **[11]**, which represents a simplification and refinement of Buchstab's weights, seems to be very close to the objective. For example it yields $r(1, 15/8)=2$. The result $r(1, 2)=2$, however, seems rather unlikely.

Recently Chen **[5]** caused a sensation by his proof of the following result, the last but one approximation to the solution of the Goldbach problem:

THEOREM 1. *Every sufficiently large even number N is a sum of a prime and an almost-prime of order* 2.

Chen's argument is applicable to binary problems of the type

$$N=\mathfrak{p}+\mathfrak{q} \tag{10}$$

where $\mathfrak{p}$ and $\mathfrak{q}$ run independently over two prescribed sequences both obtained from arithmetic progressions by sieving. The idea may be described very vaguely as follows. At first one sifts the sequence $(N-\mathfrak{p})$, getting a positive lower bound for the cardinality of the sequence which will survive the sieving. Suppose that the resulting sequence is bigger than the required $(\mathfrak{q})$, let us say by $(\mathfrak{q}')$. Therefore, we must subtract from the lower bound in question the contribution of unwanted solutions of $N=\mathfrak{p}+\mathfrak{q}'$. Then one applies another sieve to the sequence $(N-\mathfrak{q}')$, getting some upper bound. If the latter does not exceed the preceding lower bound then there must be a solution of (10). It is worth noting that the sieving of $(N-\mathfrak{q}')$ may be very crude; the efficiency of the method is rather due to the fact that $(\mathfrak{q}')$ is quite thin. In Chen's result either sieve is linear. Much the same idea was used in **[18]** in the context of Hardy–Littlewood's problem $N=p+x^2+y^2$, where the corresponding sieves are of dimensions 1/2 and 1. For other applications of the above method, sometimes enriched by new innovations see **[8]**, **[9]**, **[15]**, **[16]**, **[20]** and **[23]**.

In the speaker's opinion, a real revolution in the sieve theory has been initiated by another achievement of Chen **[6]**:

THEOREM 2. *If x is sufficiently large then there must be a P_2 in the interval $(x, x+x^{1/2})$.*

The basic novelty may be outlined as follows. When expanding the sifting function $S(\mathcal{A}, z)$ into Buchstab's formula

$$S(\mathcal{A}, z) = S(\mathcal{A}, z_1) - \sum_{z_1 \leqq p < z} S(\mathcal{A}_p, p) \quad (z_1 < z)$$

and applying estimates (7) to $S(\mathcal{A}, z_1)$ as well as to each $S(\mathcal{A}_p, p)$ one comes back to (7) for $S(\mathcal{A}, z)$ and gains nothing. Chen, however, pressed the matter further. What he gained by this apparently useless operation is the fact that he needs after all not the individual estimates of $S(\mathcal{A}_p, p)$ but estimates on average (over primes $p \in [z_1, z)$) which can be improved because the error terms $r(\mathcal{A}, pd)$ may cancel quite often. In other words he takes into account the fact that the subsequences $\mathcal{A}_p \subset \mathcal{A}$ are somehow correlated contrary to an a priori assumption that was made when constructing Rosser's subsets D^+ and D^-.

A similar conclusion can be drawn from the result of **[24]**. It is shown there (in the case of the linear sieve) that the estimates similar to (7) hold with a more flexible form of the error term

$$R(\mathcal{A}; M, N) = \sum_{m<M} \sum_{n<N} a_m b_n r(\mathcal{A}, mn) \tag{11}$$

instead of the traditional $R(\mathcal{A}, D)$ subject to $D = MN$, $M \geqslant 2$ and $N \geqslant 2$. There are several methods for working with the double sum $R(\mathcal{A}; M, N)$ leading to $R(\mathcal{A}; M, N) \ll X(\log X)^{-A}$ with values of MN larger than those possible for the usual $R(\mathcal{A}; MN)$. Sometimes one gets $MN = X^\alpha$ with $\alpha > 1$ which is never attainable in the conventional approach.

Let me remark that Hooley **[13]** and Motohashi **[29]** were the first who considered error terms in sieves in a nontrivial fashion. The work **[24]** was just inspired by Motohashi's pioneering paper on the Brun–Titchmarsh theorem.

One way of treating $R(\mathcal{A}; M, N)$ is related to an idea of expanding every single error term $r(\mathcal{A}, mn)$ into Fourier series. On application of Cauchy–Schwarz's inequality one arrives at trigonometric sums which are then estimated by means of exponent pairs or by van der Corput method. This is actually what Chen did. Halberstam, Heath-Brown and Richert rearranged his argument somewhat getting remarkably large improvement (unpublished).

THEOREM 3. *For every sufficiently large x the interval $(x, x+x^\theta)$ with $\theta = .454791\ldots$ contains a P_2-number.*

Also, Heath-Brown **[12]** extended the arguments to the problem of almostprimes in arithmetic progression, proving

THEOREM 4. *If $(a, q) = 1$ then the least $P_2 \equiv a \pmod q$ is $\ll q^{2-\eta}$ with $\eta > 0$.*

Heath-Brown's value of η was equal to 0.035. The method of proving Theorem 3 will give improved values of η.

Another approach to $R(\mathcal{A}, M, N)$ is offered by Linnik's dispersion method. When attacking the problem $n^2+1 = P_2$ we take $\mathcal{A} = \{n^2+1; n \leqslant x\}$ and express the quantities $|\mathscr{A}_d|$ in terms of incomplete Kloosterman's sums (an idea of Hooley). From Weil's result we derive an estimate for the dispersion

$$\sum_m \left\{ \sum_m b_n |\mathscr{A}_{mn}| - \sum_n b_n \frac{\omega(mn)}{mn} x \right\}^2$$

which leads to a satisfactory bound for the error term $R(\mathscr{A}; M, N)$ with $M = x^{1-\varepsilon}$ and $N = x^{1/15-\varepsilon}$. The latter can be interpreted as saying that the sequence $\mathscr{A}$ is uniformly distributed in arithmetic progressions with modulus $d < x^{\alpha-\varepsilon}$ where $\alpha = 16/15$. This together with Richert's $r(1, 15/8) = 2$ led the speaker to the result

THEOREM 5. *The polynomial n^2+1 represents almost-primes of order 2 for infinitely many integral n.*

Still another approach is of an analytic character. One expresses $r(\mathscr{A}, d)$ as the Perron integral of Dirichlet's generating function for $\mathscr{A}$. The double sum $\sum\sum a_m b_n r(\mathscr{A}, mn)$ is then the integral of the product of three generating functions for the sequences $\mathscr{A}$, (a_m) and (b_n) respectively. An application either of the mean value theorem or of the Halász–Montgomery inequality for Dirichlet's polynomials yields the required estimate of $R(\mathscr{A}; M, N)$. This method combined with Jutila's idea about weighted density estimates for the zeros of Riemann's ζ-function has led Jutila and the speaker to the result

THEOREM 6. *The difference $p_{n+1} - p_n$ between consecutive primes is $\ll p_n^{18/23}$.*

Quite recently Heath–Brown and the speaker have proved that $p_{n+1} - p_n \ll p_n^{11/20+\varepsilon}$.

5. Concluding remarks. The earliest applications of sieve methods were limited to questions involving almost-primes. There are, however, a number of instances when the sieve can be used to produce other conclusions, as for example the solvability of diophantine equations (see **[8]**, **[23]**) or diophantine inequalities as in **[20]**:

THEOREM 7. *Let θ be any positive irrational number. Then, for each $\varepsilon > 0$ there exists an integral vector (x_1, x_2, x_3, x_4) such that*

$$0 < |x_1^2 + x_2^2 - \theta(x_3^2 + x_4^2)| < \varepsilon.$$

An impressive number of applications in other unfamiliar directions were established by Hooley. His tract **[13]** may be an excellent source of inspiration for further studies of the possibilities of sieve methods.

References

1. N. C. Ankeny and H. Onishi, *The general sieve*, Acta Arith. **10** (1964), 31—62.

2. E. Bombieri, *Le grand cribe dans la theorie analytique des nombres*, Astérisque No. 18, Soc. Math. France, Paris, 1974.

3. ——, *The asymptotic sieve*, Memorie Accad. Naz. dei XL (to appear).

4. V. Brun, *Le crible d'Eratosthène et le théorème de Goldbach*, Skr. Norske Vid.-Akad. Kristiania I, No. 3, 1920.

5. J. R. Chen, *On the representation of a larger even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157—176.

6. *On the distribution of almost primes in an interval*, Sci. Sinica **18** (1975), 611—627.

7. J. Friedlander and H. Iwaniec, *On Bombieri's asymptotic sieve*, Ann. Scuola Norm. Sup. Pisa (to appear).

8. *Quadratic polynomials and quadratic forms*, Acta Math. **141** (1978), 1—15.

9. G. Greaves, *On the representation of a number in the form $x^2+y^2+p^2+q^2$ where p, q are odd primes*, Acta Arith. **29** (1976), 257—274.

10. H. Halberstam and E. H. Richert, *Sieve methods*, Academic Press, London, 1974.

11. *Almost-primes in short intervals and arithemetic progressions* (unpublished).

12. D. R. Heath-Brown, *Almost-primes in arithmetic progressions and short intervals*, Math. Proc. Cambridge Philos. Soc. **83** (1978), 357—375.

13. C. Hooley, *On the greatest prime factor of a quadratic polynomial*, Acta Math. **117** (1967), 281—299.

14. *Applications of sieve methods to the theory of numbers*, Cambridge Univ. Press, London, 1976.

15. M. Huxley and H. Iwaniec, *Bombieri's theorem in short intervals*, Mathematika **22** (1975), 188—194.

16. K. H. Indlekofer, *Scharfe untere Abschätzung für die Anzahlfunktion der B-Zwillinge*, Acta Arith. **26** (1974), 207—212.

17. H. Iwaniec, *On the error term in the linear sieve*, Acta Arith. **19** (1971), 1—30.

18. *Primes of the type $\varphi(x, y)+A$, where φ is a quadratic form*, Acta Arith. **21** (1972), 203—234.

19. *The half dimensional sieve*, Acta Arith. **29** (1976), 69—95.

20. *On indefinite quadratic forms in four variables*, Acta Arith. **33** (1977), 209—229.

21. *The sieve of Eratosthenes-Legendre*, Ann. Scuola Norm. Sup. Pisa **2** (1977), 257—268.

22. *Rosser's sieve*, Acta Arith. (to appear).

23. *On sums of two norms from cubic fields* (to appear).

24. *A new form of the error term in the linear sieve*, Acta Arith. (to appear).

25. *Almost-primes represented by quadratic polynomials*, Invent. Math. **47** (1978), 171—188.

26. H. Iwaniec and M. Jutila, *Primes in short intervals*, Ark. Mat. (to appear).

27. P. Kuhn, *Über die Primteiler eines Polynoms*, Proc. Internat. Congress Math., Amsterdam, 1954, vol. 2, pp. 35—37.

28. H. L. Montgomery, *Topics in multiplicative number theory*, Lecture Notes in Math., vol. 227, Springer-Verlag, Berlin, 1971.

29. Y. Motohashi, *On some improvements of the Brun-Titchmarsh theorem*, J. Math. Soc. Japan **26** (1974), 306—323.

30. J. W. Porter, *On the non-linear sieve*, Acta Arith. **29** (1976), 377—400.

31. E. H. Richert, *Selberg's sieve with weights*, Mathematika **16** (1969), 1—22.

32. A. Selberg, *Sieve methods*, Proc. Sympos. Pure Math., vol. 20, Amer. Math. Soc., Providence, R. I., 1971, pp. 311—351.

Polish Academy of Sciences
00-950 Warszawa, Poland

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

p-adic L-Functions, Serre-Tate Local Moduli, and Ratios of Solutions of Differential Equations

Nicholas M. Katz

Introduction. In recent years, there has been considerable progress in the constructions of p-adic L-functions attached to various sorts of "classical" L-functions. Unfortunately, the use of these p-adic functions to solve preexisting problems in number theory has so far met with less success; despite the recent work of Coates–Wiles [1] and Ferrero–Washington [4], conjectures remain more numerous than theorems. It may be hoped that a better understanding of the genesis of various p-adic L-functions will lead to progress in their exploitation. In that hope, we give yet another construction of the "two-variable" p-adic L-function attached to an elliptic curve with complex multiplication by a quadratic imaginary field in which p splits. This construction is based on the remarkable fact, discovered by Serre–Tate some fifteen years ago, that the local p-adic moduli space of such an elliptic curve has a canonical structure of one parameter formal group of height one. A rewriting of this construction in terms of ratios of local solutions of the associated Picard–Fuchs equations leads to universal formulas for the "algebraic part" of the classical L-values, which may shed light on the still mysterious situation when p is no longer assumed to split.

I. Let $K\subset \boldsymbol{C}$ be a quadratic imaginary field, with ring of integers $\mathcal{O}(K)$. Viewing $\mathcal{O}(K)$ as a lattice in $\boldsymbol{C}$, we may form the elliptic curve $E=\boldsymbol{C}/\mathcal{O}(K)$. Because E has complex multiplication, it is definable over the ring $\mathcal{O}(\overline{\boldsymbol{Q}})$ of all algebraic integers in $\boldsymbol{C}$, with everywhere good reduction. Further, we may choose a nowhere-vanishing invariant differential ω on E over $\mathcal{O}(\overline{\boldsymbol{Q}})$, so that the *pair* (E, ω) has everywhere good reduction over $\mathcal{O}(\overline{\boldsymbol{Q}})$, i.e. for any place $\mathcal{P}$ of $\overline{\boldsymbol{Q}}$, "$\omega$ mod $\mathcal{P}$" is nonzero on "E mod $\mathcal{P}$". Such an ω is *unique* up to multiplication by a *unit* in $\mathcal{O}(\overline{\boldsymbol{Q}})$.

The *period* lattice of (E, ω) is necessarily of the form $\Omega\mathcal{O}(K)$ for some $\Omega \in \boldsymbol{C}^\times$. For variable ω of the sort discussed above, this period Ω is well defined in the group $\boldsymbol{C}^\times/\mathcal{O}(\overline{\boldsymbol{Q}})^\times$.

We will denote by a the *area* of (a fundamental parallelogram of) the lattice $\mathcal{O}(K)$. In terms of the discriminant d of K, we have

$$a = \tfrac{1}{2}\sqrt{|d|}.$$

For integers $k \geqslant 3$, $r \geqslant 0$, consider the absolutely convergent series

$$A(k, r) = \sum_{\substack{\gamma \in \mathcal{O}(K) \\ \gamma \neq 0}} \frac{\bar{\gamma}^r}{\gamma^{k+r}}.$$

According to a fundamental result of Damerell [2] the product

$$B(k, r) = \frac{(-1)^k(k+r-1)!\,\pi^r}{2a^r \cdot \Omega^{k+2r}} \cdot A(k, r)$$

lies in $\overline{\boldsymbol{Q}}$; in fact it lies in the field obtained by adjoining to K the Weierstrass invariants g_2, g_3 of (E, ω). Further, for any integer $b \geqslant 1$, the product

$$b^k(b^k-1)\left(\sqrt{-|d|}\right)^r B(k, r)$$

is an algebraic integer.

The arithmetic of these numbers, and of their more sophisticated analogues ("with conductor", and extended to include $k=1$ or 2) is of interest because of their occurrence

(1) in the Birch–Swinnerton-Dyer conjecture for certain elliptic curves with complex multiplication (cf. [1]).

(2) as *periods* of cusp forms on congruence subgroups of $\mathrm{SL}(2, \boldsymbol{Z})$ (cf. [8]).

(3) as *special values* of holomorphic and nonholomorphic Eisenstein series on congruence subgroups of $\mathrm{SL}(2, \boldsymbol{Z})$ (cf. [6], [11]).

(4) as *special values* of Hecke L-series attached to grossencharacters of type A_0 of quadratic imaginary fields (cf. [7], [8]).

It would be of great interest to understand the *link* between (2) and (3) "directly"; both have been used to get information about occurrences (1) and (4).

II. At present, we have a reasonable understanding of the p-adic properties of the $B(k, r)$ only for primes p which *split* K. More precisely, fix a finite extension K'/K over which (E, ω) is defined and has everywhere good reduction. Let $\mathfrak{p}$ be a prime of K', $K'_\mathfrak{p}$ the $\mathfrak{p}$-adic completion of K', and W the ring of integers in the completion of the maximal unramified extension of $K'_\mathfrak{p}$. Denote by p the rational prime lying under $\mathfrak{p}$.

THEOREM. *If p splits in K, there exists a unit $c \in W^\times$ and, for all rational integers b prime to p, a W-valued p-adic measure $\mu(c, b)$ on $\mathbf{Z}_p \times \mathbf{Z}_p$, whose moments are given by the formula, valid for integers $k \geq 3$, $r \geq 0$,*

$$\int_{\mathbf{Z}_p \times \mathbf{Z}_p} x^{k-3} y^r \, d\mu(c, b) = 2 \cdot c^{k+2r}(b^k - 1) B(k, r).$$

In [6] we used the global theory of "p-adic modular functions" to construct this measure. Here we will outline a new construction, based on the Serre–Tate theory of local moduli of elliptic curves in terms of their p-divisible groups. This construction also leads to a universal computation of the $B(k, r)$ which may yield valuable information when p does not split in K.

Step I (Interpretation of measures). Over *any* p-adically complete and separated ring W, Cartier duality gives a canonical isomorphism between the convolution algebra of W-valued p-adic measures on $(\mathbf{Z}_p)^n$ and the coordinate ring $W[[X_1, \ldots, X_N]]$ of the n-fold self-product $(\hat{\mathbf{G}}_m)^n$ of the formal multiplicative group over W. Let $x_1, \ldots, x_n$ denote the standard coordinates on $(\mathbf{Z}_p)^n$, and let $D_1, \ldots, D_n$ be the standard invariant derivations $D_i = (1+X_i)\partial/\partial X_i$ on $(\hat{\mathbf{G}}_m)^n$. Given a function $f(X_1, \ldots, X_n) \in W[[X_1, \ldots, X_n]]$, the moments of the corresponding measure μ_f are given by

$$\int_{(\mathbf{Z}_p)^n} x_1^{i_1} \ldots x_n^{i_n} \, d\mu_f = D_1^{i_1} \ldots D_n^{i_n}(f)|_0.$$

Given a measure μ, the corresponding function $f_\mu(X_1, \ldots, X_n)$ is given by

$$f_\mu(X_1, \ldots, X_n) = \int_{(\mathbf{Z}_p)^n} (1+X_1)^{x_1} \ldots (1+X_n)^{x_n} \, d\mu.$$

Thus to construct our measure $\mu(c, b)$, we need a function f on a group $\hat{\mathbf{G}}_m \times \hat{\mathbf{G}}_m$.

Step II (Construction of $\hat{\mathbf{G}}_m \times \hat{\mathbf{G}}_m$ out of E and its local moduli). Returning to (E, ω) over $\mathcal{O}(K')$, we extend scalars to W. Because p splits in K, E has *ordinary* reduction at $\mathfrak{p}$, and hence, the formal group $\hat{E}$ of E is non-canonically isomorphic to $\hat{\mathbf{G}}_m$ over W. Fix one such isomorphism

$$\varphi\colon \hat{E} \xrightarrow{\sim} \hat{\mathbf{G}}_m \quad (\text{over } W).$$

The inverse image of the "standard" invariant differential $dX/(1+X)$ on $\hat{\mathbf{G}}_m$ is necessarily of the form $c^{-1}\omega$ for some unit $c \in W^\times$; this is the "c" occurring in the statement of the theorem.

Now consider the universal formal W-deformation E^{univ} of E, over the formal moduli space $\hat{\mathcal{M}}$. The chosen isomorphism φ extends uniquely to an isomorphism

$$\hat{E}^{\mathrm{univ}} \overset{\varphi}{\xrightarrow{\sim}} \hat{\mathbf{G}}_m \quad \text{over } \hat{\mathcal{M}}, \quad \text{i.e.} \quad \hat{E}^{\mathrm{univ}} \cong \hat{\mathcal{M}} \times \hat{\mathbf{G}}_m.$$

The Serre–Tate theory [9] gives an explicit isomorphism of the space $\hat{\mathcal{M}}$ with the formal group $\hat{\mathbf{G}}_m$ over W; the origin of this $\hat{\mathbf{G}}_m$ is the W-valued point of

$\hat{\mathscr{M}}$ which "is" E. Thus we have

$$\hat{E}^{\mathrm{univ}} \cong \hat{\mathscr{M}} \times \hat{G}_m \cong \hat{G}_m \times \hat{G}_m.$$

Here are three equivalent descriptions of this isomorphism $\hat{\mathscr{M}} \xrightarrow{\sim} \hat{G}_m$.

(a) Because E has complex multiplication by $\mathcal{O}(K)$, and has ordinary reduction at $\mathfrak{p}$, its p-divisible is necessarily a *product*

$$E(p^\infty) \xrightarrow{\sim} \hat{E} \times E(p^\infty)^{\mathrm{etale}} \xrightarrow[\sim]{\varphi \times (\check{\varphi})^{-1}} \hat{G}_m \times \boldsymbol{Q}_p/\boldsymbol{Z}_p.$$

Let $\boldsymbol{W}$ be a p-adically complete and separated augmented W-algebra, with nilpotent augmentation ideal, and let $\boldsymbol{E}/\boldsymbol{W}$ be a deformation of E/W. Then the p-divisible group of $\boldsymbol{E}$ sits in an *extension*

$$0 \to \hat{\boldsymbol{G}}_m \to \boldsymbol{E}(p^\infty) \to \boldsymbol{Q}_p/\boldsymbol{Z}_p \to 0,$$

and so determines an element of $\mathrm{Ext}^1_{\boldsymbol{W}}(\boldsymbol{Q}_p/\boldsymbol{Z}_p, \hat{\boldsymbol{G}}_m) \xrightarrow{\sim} \hat{\boldsymbol{G}}_m(\boldsymbol{W})$. (Explicitly, let P_i be the point of order p^i in $E(W)$ corresponding to "$1/p^i$" in the $\boldsymbol{Q}_p/\boldsymbol{Z}_p$-factor of $E(p^\infty)$. Let $\boldsymbol{P}_i$ be *any* point in $\boldsymbol{E}(\boldsymbol{W})$ lifting P_i; then $p^i\boldsymbol{P}_i$ lies in $\hat{\boldsymbol{E}}(\boldsymbol{W}) \xrightarrow{\sim}{}^{\varphi} \hat{\boldsymbol{G}}_m(\boldsymbol{W})$, and as $i \to \infty$ these points tend to a *limit* in $\hat{\boldsymbol{G}}_m(\boldsymbol{W})$). The resulting morphism $\hat{\mathscr{M}} \to \hat{\boldsymbol{G}}_m$ is an isomorphism.

(b) Consider once again the universal formal deformation $\boldsymbol{E}^{\mathrm{univ}}$ over $\hat{\mathscr{M}}$. Via the Kodaira–Spencer isomorphism

$$(\omega_{E^{\mathrm{univ}}/\hat{\mathscr{M}}})^{\otimes 2} \cong \Omega^1_{\hat{\mathscr{M}}/W}$$

the square of $\varphi^*(dX/(1+X))$ corresponds to a basis ξ of $\Omega^1_{\hat{\mathscr{M}}/W}$. The isomorphism $\hat{\mathscr{M}} \xrightarrow{\sim} \hat{\boldsymbol{G}}_m$ is the unique morphism of pointed functors under which $dX/(1+X)$ pulls back to ξ.

(c) There is a unique basis u, v of $H^1_{DR}(E/W)$ such that

(1) $u = c^{-1}\omega$,

(2) $\langle u, v\rangle = 1$ (de Rham cup product),

(3) for $\gamma \in \mathcal{O}(K)$ acting, as $[\gamma]^*$, on $H^1_{DR}(E/W)$, we have

$$[\gamma]^*(u) = \gamma u, \ [\gamma]^*(v) = \bar{\gamma} v.$$

Now consider $H^1_{DR}(E^{\mathrm{univ}}/\hat{\mathscr{M}})$, with its Gauss–Manin connection. Let $\mathrm{Div}(\hat{\mathscr{M}})$ denote the ring of all "divided" power series centered at the marked W-point "E/W" of $\hat{\mathscr{M}}$. In terms of a parameter T for $\hat{\mathscr{M}}$ centered at "E/W"; this is the ring

$$W\langle\langle T\rangle\rangle = \left\{ \sum_{n \geq 0} a_n \frac{T^n}{n!} \,\middle|\, a_n \in W \right\};$$

intrinsically, it is the topological "divided power envelope" of the marked point "E/W" in $\hat{\mathscr{M}}$. On $H^1_{DR}(E^{\mathrm{univ}}/\hat{\mathscr{M}}) \otimes \mathrm{Div}(\hat{\mathscr{M}})$, the connection necessarily becomes trivial, so we can find a *horizontal* basis U, V which extends the given basis u, v

of $H^1_{DR}(E/W)$. In terms of this basis, the invariant differential $\varphi^*(dX/(1+X))$ on E^{univ}, viewed as a de Rham cohomology class is expressed as

$$\varphi(dX/(1+X)) = U + LV \quad \text{with} \quad L \in \mathrm{Div}(\hat{\mathscr{M}}).$$

The isomorphism $\hat{\mathscr{M}} \xrightarrow{\sim} \hat{G}_m$ is the unique morphism of pointed functors under which L becomes the *logarithm* on $\hat{G}_m$:

$$L(X) = \log(1+X) \quad \text{i.e.} \quad dL = dX/(1+X) = \xi.$$

That these descriptions are in fact equivalent may be seen as follows. By "general principles", the function L must be a (divided-power) isomorphism from $\hat{G}_m$ to $\hat{G}_a$, i.e. we must have $L(X) = w\log(1+X)$ for some $w \in W^\times$. To see that $w=1$, it suffices to compute $L \bmod (X^2)$, and this amounts to explicitly computing the description (a) for deformations of E over the dual numbers $W[\varepsilon]/(\varepsilon^2)$. This last computation becomes routine if we exploit the autoduality of elliptic curves by systematically interpreting *points* on elliptic curves as (isomorphism classes of) *line bundles*.

A more sophisticated proof of this and more general equivalences has been announced by Messing [10].

Step III (Construction of a function f on $\hat{E}^{\text{univ}} \simeq \hat{G}_m \times \hat{G}_m$). Given an integer $b \geqslant 1$ prime to p, the function f on $\hat{E}^{\text{univ}}$ to be taken is, in "transcendental" notation,

$$f(z) = b^3 \wp'(bz) - \wp'(z) = \sum_{\substack{\zeta \in \mathrm{Ker}[b] \\ \zeta \neq 0}} \wp'(z+\zeta).$$

This has purely algebraic meaning, as follows. Given *any* (E, ω) over any ring R, pick *any* parameter Z for $\hat{E}$ so that $\omega = (1+\ldots)dZ$. The functions on E with at worst double poles along the 0-section (i.e. $H^0(E, I(0)^{-2})$) which begin $Z^{-2}+\ldots$ all differ from each other by additive constants. If we apply to any of them the invariant derivation dual to ω, we get a well-defined $\wp'$. If b is invertible in R, then all nontrivial points of order b are disjoint from $\hat{E}$, so the Σ-expression for f shows that it's well-defined on $\hat{E}$. We apply this universal construction to $(E^{\text{univ}}, \varphi^*(dX/(1+X)))$ over the coordinate ring of $\mathscr{M}$.

Step IV (Universal computation of the moments). We now return to the original (E, ω) over $\mathcal{O}(K')$, with complex multiplication by $\mathcal{O}(K)$. Let W be *any* over-ring of $\mathcal{O}(K')$ in which the discriminant d of K is invertible, and let $c \in W^\times$ be *any* unit of W. It still makes sense to take a basis u, v for $H^1_{DR}(E/W)$ as in Step II (c) and then to find the horizontal basis U, V of $H^1_{DR}(E^{\text{univ}}/\hat{\mathscr{M}}) \otimes \mathrm{Div}(\hat{\mathscr{M}})$ which extends u, v. There is no longer a *preferred* invariant differential on E^{univ}, but we may simply choose one which extends ω/c. Its expression in terms of U, V will be

$$\alpha U + \beta V, \ \alpha, \beta \in \mathrm{Div}(\hat{\mathscr{M}}), \quad \alpha(0) = 1, \quad \beta(0) = 0.$$

Because $\alpha(0)=1$, it is invertible in $\mathrm{Div}(\hat{\mathscr{M}})$. Therefore there is a *unique* invariant differential ω on $E^{\mathrm{univ}}\otimes\mathrm{Div}(\hat{\mathscr{M}})$ whose expression in U, V is

$$\omega = U+LV \begin{cases} L = \beta/\alpha \in \mathrm{Div}(\hat{\mathscr{M}}); \\ L(0) = 0. \end{cases}$$

This function $L\in\mathrm{Div}(\hat{\mathscr{M}})$ is simply the *direction* (i.e. the Plücker coordinate) of the subspace $H^{1,0}\subset H^1_{DR}$, measured with respect to the horizontal basis U, V. It is a "divided-power uniformizing parameter", in the sense that the natural map

$$W\langle\langle L\rangle\rangle \xrightarrow{\sim} \mathrm{Div}(\hat{\mathscr{M}})$$

is an isomorphism.

Let b be any integer invertible in W, and apply the construction of Step III to $(E^{\mathrm{univ}}\otimes\mathrm{Div}(\hat{\mathscr{M}}), \omega)$, to produce a function f on $\hat{E}^{\mathrm{univ}}\otimes\mathrm{Div}(\hat{\mathscr{M}})$. It follows easily from the cohomological analysis of ([7], 2.4.8) that we may compute the $B(k, r)$'s as follows.

ALGORITHM. Let D_1 be the invariant derivation of $\hat{E}^{\mathrm{univ}}\otimes\mathrm{Div}(\hat{\mathscr{M}})$ over $\mathrm{Div}(\hat{\mathscr{M}})$ which is dual to ω. For all integers $k\geqslant 3$, $r\geqslant 0$, we have

$$2c^{k+2r}(b^k-1)B(k, r) = (d/dL)^r(D_1^{k-3}(f)_{|0})_{|0}.$$

III. When p splits in K, and W and c are as in Step II, the theorem follows immediately from this algorithm and Steps I, II, III. When p stays prime in K, this algorithm gives the known integrality results, and focuses attention on the very special role played by the divided power parameter L on the moduli space $\hat{\mathscr{M}}$. Arithmetic information about L should yield arithmetic information about the numbers $B(k, r)$. Is it conceivable that L is always the logarithm of a formal group structure on the pointed (by E/W) functor $\hat{\mathscr{M}}$?

IV. In this final section, we give an "elementary" description of L, valid over any ring containing 1/2, as the ratio of two particular local solutions of the Gauss hypergeometric equation with parameters (1/2, 1/2, 1). From this point of view, the function L has been studied extensively by Dwork, at least in the case when p splits in K, under the name "τ" ([3], [5]).

Consider the Legendre family of elliptic curves $y^2=x(x-1)(x-\lambda)$ over $\mathscr{M}=\mathrm{Spec}(\mathbf{Z}[\lambda][1/(2\lambda(\lambda-1))])$. Let λ_0 be any value of λ at which this curve acquires complex multiplication by the ring of integers $\mathcal{O}(K)$ in a quadratic imaginary field. The formal moduli space $\hat{\mathscr{M}}$ is simply the formal completion of $\mathscr{M}$ at $\lambda=\lambda_0$.

Let D denote the derivation $2\lambda(\lambda-1)d/d\lambda$ of $\mathscr{M}$. The H^1_{DR} for the Legendre family is free over $\mathscr{M}$ with basis

$$\omega = dx/2y,\ D(\omega) = (x-\lambda)\,dx/2y$$

with

$$\langle\omega, D(\omega)\rangle = 1 \qquad \text{(de Rham cup-product)},$$

$$D^2(\omega) = -\lambda(\lambda-1)\omega \quad \text{(Gauss—Manin connection)}.$$

At λ_0, a basis u, v of H^1_{DR} which is adapted to the action of $\mathcal{O}(K)$ is given by

$$u = \omega_{|\lambda=\lambda_0}, \quad v = (D(\omega) - e\omega)_{|\lambda=\lambda_0}$$

for some unique constant e in $(1/\sqrt{-|d|}) \cdot \mathcal{O}(K')[1/2]$. Let $\alpha(\lambda), \beta(\lambda)$ be the local solutions near $\lambda=\lambda_0$ of the hypergeometric equation

$$D^2 f = -\lambda(\lambda-1) f,$$

normalized by the initial conditions

$$\alpha(\lambda_0) = 1, \ (D\alpha)(\lambda_0) = e,$$
$$\beta(\lambda_0) = 0, \ (D\beta)(\lambda_0) = 1.$$

The horizontal basis U, V passing through u, v at $\lambda=\lambda_0$ is given by

$$U = D(\beta)\omega - \beta D(\omega), \ V = -D(\alpha) \cdot \omega + \alpha D(\omega).$$

Thus we find

$$\omega = \alpha U + \beta V,$$

whence

$$L = \beta/\alpha, \quad \boldsymbol{\omega} = \omega/\alpha, \quad d/dL = \alpha^2 \cdot 2\lambda(\lambda-1)\, d/d\lambda,$$
$$D_1 = \alpha \cdot 2y\, d/dx, \quad f = 2\alpha^3(b^3[b]^*(y) - y).$$

References

1. J. Coates and A. Wiles, *On the conjecture of Birch and Swinnerton-Dyer,* Invent Math. **39** (1977), 223—251.

2. R. M. Damerell, *L-functions of elliptic curves with complex multiplication* I, Acta Arith. **17** (1970), 287—301.

3. B. Dwork, *P-adic cycles,* Inst. Hautes Études Sci. Publ. Math. **37** (1969), 327—415.

4. B. Ferrero and L. Washington, *The Iwasawa invariant μ_p vanishes for abelian number fields,* Ann. of Math. **109** (1979), 377—395

5. N. Katz, *Travaux de Dwork,* Séminaire Bourbaki 1971/72, Lecture Notes in Math., vol. 317 Springer-Verlag, Berlin and New York, 1973, pp. 167—200.

6. ——— *P-adic interpolation of real analytic Eisenstein series,* Ann. of Math. **104** (1976), 459—571.

7. ——— *P-adic L-functions for CM-fields,* Invent. Math. **49** (1978), 199—297.

8. J. Manin and S. Vishik, *p-adic Hecke series for quadratic imaginary fields,* Mat. Sb. **95 (137)**, (1974).

9. W. Messing, *The crystal associated to Barsotti-Tate groups, with applications to abelian schemes,* Appendix, Lecture Notes in Math., vol. 264, Springer-Verlag, Berlin and New York, 1972.

10. ——— $q_{\text{Serre-Tate}} = q_{\text{Dwork}}$, Notices Amer. Math. Soc. (1976).

11. G. Shimura, *On some arithmetic properties of modular forms of one and several variables,* Ann of Math. **102** (1975), 491—515.

PRINCETON UNIVERSITY
PRINCETON, N. J. 08540, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

On Some Problems of Algebraicity

Goro Shimura

1. As every mathematician knows, the parallelism between algebraic number theory and transcendental number theory exists only in their appellations and not in their contents. Indeed, the aim of the latter theory is to prove the transcendence of a given number, while algebraic numbers are there from the beginning in the former. Therefore, if one proves the algebraicity of an analytically defined number, it cannot be viewed as a theorem of either theory. It belongs to a new area of investigation for which this lecture is intended and to which I can give no good designation. Although the problem of algebraicity has not attracted much attention until recently, there are at least two classical and well-known examples.

(1) *The values of Riemann's zeta function.* If $\zeta(s)=\sum_{n=1}^{\infty}n^{-s}$, then $\pi^{-m}\zeta(m)$ for an even positive integer m is a rational number.

(2) *Complex multiplication of elliptic modular functions.* If f is an elliptic modular function with rational Fourier coefficients, then the value $f(z)$ at an imaginary quadratic irrational z is algebraic.

In each example, one can go beyond the algebraicity. In (2), for example, one can actually prove that $f(z)$ generates an abelian extension of $Q(z)$; also in both cases, the p-adic nature of the numbers is an interesting topic. In this lecture, however, I will concentrate only in the algebraicity, without touching on more delicate questions. It should be noted that the question of algebraicity can be naturally asked for various mathematical objects. For instance, one can ask whether a given complex manifold V is an algebraic variety. Supposing it is so, one may still ask whether V has a model defined over an algebraic number field. This question, when V is an arithmetic quotient of a bounded symmetric domain, is fundamental in our investigation. However, as I talked on this in a previous Congress ([6], see also [2], [4], [5]), I take the algebraicity of such V to be a starting point rather than a goal in the following discussion.

Now what kind of numbers should be offered for their algebraicity? As a generalization of example (1), a natural choice falls upon the values of an L-function of a number field F with, say, an abelian character χ. If χ is of finite order, the problem is reduced, for some natural reasons, to the case where F is totally real and χ is totally ramified or unramified. The results, due to Siegel and Klingen, being well known, will need no detailed description here. Beyond this, one may attempt to study the values of zeta functions obtained from various algebraic groups, say $\mathrm{GL}_n(F)$. If F is totally real and $n=2$, we can again obtain a certain algebraicity theorem for the values of the Mellin transform of a holomorphic Hilbert cusp form [11]. Some results have been obtained also by Sturm [13] for the zeta functions of $\mathrm{GL}_3(\boldsymbol{Q})$ related to cusp forms of $\mathrm{GL}_2(\boldsymbol{Q})$. These results suggest some general principles of algebraicity, but without making any speculation, let us narrow down our subject to the case in which the values of some zeta functions occur as the special values of Eisenstein series. The simplest example is the zeta function of an imaginary quadratic field with a Hecke character of infinite order. That such a coincidence occurs in this case is rather obvious, but its nontrivial generalization (still in the quadratic case) was given by Damerell [1]. Subsequently a different approach was taken by Weil [14]; these results have been generalized to the case related to the Eisenstein series of the Hilbert modular groups [7]. We now propose to extend this to a more general framework, by considering automorphic forms of a $\boldsymbol{Q}$-rational reductive algebraic group G. We assume that the semi-simple part of $G_{\boldsymbol{R}}$ modulo a maximal compact subgroup is a bounded symmetric domain S, and take, for simplicity, a power of the jacobian as the factor of automorphy. We then lay out our program by asking the following series of questions:

(I) *Can one define the notion of arithmetic automorphic functions?*

(II) *Can one define the notion of arithmetic automorphic forms?*

(III) *Are (holomorphic) Eisenstein series arithmetic?*

(IV) *Is there any explicit way to construct arithmetic automorphic forms, similar to Eisenstein series, in the case of compact quotient?*

(V) *Supposing the answers to these questions are affirmative, is there any interpretation of the values of such explicit arithmetic automorphic forms at CM-points as the values of zeta functions?*

The arithmeticity should be defined relative to certain number fields. But without seeking the sharper results, we simply take the algebraic closure $\overline{\boldsymbol{Q}}$ of the rational number field as the basic field; so *arithmetic* forms may be called *$\overline{\boldsymbol{Q}}$-rational* forms. The first question is essentially the same as the problem of finding the (canonical) $\overline{\boldsymbol{Q}}$-rational models of the quotients S/Γ for congruence subgroups Γ of $G_{\boldsymbol{Q}}$. As we said above, we start by assuming their existence. This means that we can speak of $\bar{Q}$-rational automorphic functions, whose values at CM-points are algebraic. As to (II), if S/Γ is not compact, we can consider Fourier expansions of automorphic forms with respect to a group of translations T contained in Γ. In general, the Fourier coefficients are theta functions; they become constants if T is sufficiently large. The characterization of the $\bar{Q}$-rationality by means of some properties of such Fourier coefficients are treated in [9], [10] and Garrett [3]; so let us now consider a more general method applicable even to the case of compact quotient. The idea is: first define a certain constant $p(w)$ at each CM-point w on S; then call an automorphic form f *$\bar{Q}$-rational* if $f(w)/p(w)\in\bar{Q}$ for every CM-point w. The constant $p(w)$ is given as a period of an abelian variety. Let us first discuss its definition and basic properties.

2. Let K be a totally imaginary quadratic extension of a totally real algebraic number field F of finite degree. We call such a K a *CM-field.* Let I_K denote the Z-module of all formal sums $\sum_\tau c_\tau\tau$ of embeddings τ of K into C with $c_\tau\in Z$. Such a sum with $c_\tau\geqslant 0$ defines an equivalence class of representations of K by complex matrices. An element Φ of I_K is called a *CM-type* of K if $\Phi+\Phi\varrho$ represents the regular representation of K over Q. Here and henceforth ϱ denotes the complex conjugation. Given a CM-type Φ of K, there is an abelian variety A, defined over $\bar{Q}$, such that: (i) $2\dim(A)=[K:Q]$; (ii) there is an injection $\imath$ of K into $\mathrm{End}(A)\otimes Q$; (iii) Φ is the class of representation of K on the space of holomorphic 1-forms on A. Let $[K:Q]=2g$ and $\Phi=\sum_{\nu=1}^{g}\tau_\nu$. For each ν, there exists a $\bar{Q}$-rational holomorphic 1-form ω_ν, $\neq 0$, on A satisfying $\omega_\nu\circ\imath(a)=a^{\tau_\nu}\omega_\nu$ for all $a\in K$ such that $\imath(a)\in\mathrm{End}(A)$.

PROPOSITION 1. *There exists a nonzero complex number $p(\tau_\nu,\Phi)$ depending only on K, Φ, and τ_ν such that $\int_c\omega_\nu/[\pi\cdot p(\tau_\nu,\Phi)]\in\bar{Q}$ for all 1-cycles c on A (i.e. elements c of $H_1(A,Z)$).*

For the proof, see [8]. We put $p(\tau,\Phi)=p(\tau\varrho,\Phi)^{-1}$ for an embedding τ of K into C not belonging to $\{\tau_1,\ldots,\tau_g\}$.

THEOREM 1. *Let $\Phi_1,\ldots,\Phi_m$ be CM-types of K, and τ an embedding of K into C. Then*

$$\prod_{i=1}^{m}p(\tau,\Phi_i)^{r_i} \tag{2.1}$$

with $r_i\in Z$ depends, up to algebraic factors, only on $\sum_{i=1}^{m}r_i\Phi_i$ and τ. Moreover, let Ψ be a CM-type of a CM-field L containing K, and suppose the restriction of Ψ to K is $\sum_i r_i\Phi_i$. Then the product of $p(\sigma,\Psi)$ for all embeddings σ of L into C which coincide with τ on K differs from (2.1) *by an algebraic factor.*

We denote by $p(\tau,\sum_i r_i\Phi_i)$ the number (2.1) which is determined up to algebraic factors. Assuming $K\subset C$, put $\Phi_K=\Omega_K+\varepsilon-\varepsilon\varrho$, where ε is the identity embedding of K into C and Ω_K the sum of all embeddings of K into C. Then $p(\tau,\Phi_K)$ is meaningful for every embedding τ of K into C.

THEOREM 2. *Let Φ be a CM-type of K and (K', Φ') the reflex of (K, Φ). Then, for every embedding σ of K' into C, $p(\sigma, \Phi')^2$ is an algebraic number times the product of $p(\tau, \Phi_K)$ for all τ belonging to $\Phi\sigma$.*

Note that $\Phi\sigma$ is a well-defined CM-type of K. This relation yields an upper bound for the number of algebraically independent elements among $p(\sigma, \Phi')$. The details of these results will be given in [12].

3. We now use $p(\tau, \Phi)$ for the definition of arithmetic automorphic forms. Though the idea is expected to work in a more general case, we consider here the case of arithmetic subgroups of $\mathrm{Sp}(n, \boldsymbol{R})^r$. Let B be a quaternion algebra over a totally real algebraic number field F of degree g. Define a $\boldsymbol{Q}$-rational algebraic group G so that

$$G_Q = \{\alpha \in \mathrm{GL}_n(B) \mid {}^t\alpha^\iota\alpha = \nu(\alpha)1_n \quad \text{with} \quad \nu(\alpha) \in F\},$$

where ι denotes the main involution of B. Let $\theta_1, \dots, \theta_n$ be the embeddings of F into $\boldsymbol{R}$, and suppose that B is unramified at $\theta_1, \dots, \theta_r$ and ramified at $\theta_{r+1}, \dots, \theta_g$. Then there is an isomorphism

$$G_R \simeq \prod_{\nu=1}^{g} G_\nu, \tag{3.1}$$

$$G_\nu = \begin{cases} \{\alpha \in \mathrm{GL}_{2n}(\boldsymbol{R}) \mid {}^t\alpha J\alpha = \nu(\alpha)J \quad \text{with} \quad \nu(\alpha) \in \boldsymbol{R}\} & (\nu = 1, \dots, r), \\ \{\alpha \in \mathrm{GL}_n(\boldsymbol{H}) \mid {}^t\alpha^\iota\alpha = \nu(\alpha)1_n \quad \text{with} \quad \nu(\alpha) \in \boldsymbol{R}\} & (\nu = r+1, \dots, g), \end{cases}$$

where $\boldsymbol{H}$ denotes the Hamilton quaternions, ι the main involution of $\boldsymbol{H}$, and

$$J = \begin{bmatrix} 0 & -1_n \\ 1_n & 0 \end{bmatrix}.$$

Let G_{Q+} denote the subgroup of G_Q consisting of all α such that $\nu(\alpha)$ is totally positive. For each $\alpha \in G_Q$, let $(\alpha_1, \dots, \alpha_g)$ be the corresponding element of $\prod_{\nu=1}^g G_\nu$. We fix an isomorphism (3.1) so that $\alpha_1, \dots, \alpha_r$ have algebraic entries for every $\alpha \in G_Q$. If $\alpha \in G_{Q+}$, we define the action of α on the product $\mathfrak{H}_n^r$ of r copies of

$$\mathfrak{H}_n = \{z \in M_n(\boldsymbol{C}) \mid {}^tz = z, \ \mathrm{Im}(z) > 0\}$$

by $\alpha(z_1, \dots, z_r) = (\alpha_1(z_1), \dots, \alpha_r(z_r))$, $\alpha_\nu(z_\nu) = (a_\nu z_\nu + b_\nu)(c_\nu z_\nu + d_\nu)^{-1}$ where

$$\alpha_\nu = \begin{bmatrix} a_\nu & b_\nu \\ c_\nu & d_\nu \end{bmatrix}.$$

A factor of automorphy $\chi_\nu(\alpha, z)$ can be defined by

$$\chi_\nu(\alpha, z) = c_\nu z_\nu + d_\nu \quad (z = (z_1, \dots, z_r) \in \mathfrak{H}_n^r; \ \nu = 1, \dots, r).$$

Also we consider a $\boldsymbol{Q}$-linear embedding $\chi_\nu: M_n(B) \to M_{2n}(\bar{\boldsymbol{Q}})$ such that $\chi_\nu(a1_n) = a^{\theta_\nu}1_{2n}$ for $\nu = r+1, \dots, g$.

As shown in [5], we can construct canonical models of $\mathfrak{H}_n^r/\Gamma$ for congruence subgroups Γ of G_{Q+}, so that the field of $\bar{\boldsymbol{Q}}$-rational Γ-automorphic functions is meaningful. We denote by $\mathfrak{A}_0(\bar{\boldsymbol{Q}})$ the union of such fields for all Γ. Our task is to study the $\bar{\boldsymbol{Q}}$-rationality of automorphic forms defined relative to $\chi_1, \dots, \chi_g$

in connection with their values at the CM-points of G_{Q+} on $\mathfrak{H}_n^r$, which can be obtained as follows. Let $Y=Y_1\oplus\ldots\oplus Y_t$ with simple F-algebras Y_i; suppose Y has a positive involution δ and there is an F-linear embedding h of Y into $M_n(B)$ such that $h(a)^\delta={}^t h(a)^\iota$. Let L_i be the center of Y_i and put $L=L_1\oplus\ldots\oplus L_t$, $q_i^2=[Y_i: L_i]$, $2m_i=[L_i: F]$, $Y[\delta]=\{x\in Y|x^\delta x\in F\}$, $L[\delta]=Y[\delta]\cap L$. Then $h(Y[\delta])\subset G_{Q+}$. Suppose $n=\sum_{i=1}^t q_i m_i$. Then $h(Y[\delta])$ has a unique common fixed point $w=(w_1,\ldots,w_r)$ on $\mathfrak{H}_n^r$. It is this type of fixed point where we examine the values of automorphic forms. For each $\nu\leqq r$, we can define a representation $\Xi_\nu\colon Y\to M_n(C)$ such that $\Xi_\nu(a)=\chi_\nu(h(a),w)$ for all $a\in Y[\delta]$. Then m_i embeddings $\sigma_{i1}^\nu,\ldots,\sigma_{im_i}^\nu$ of L_i into C can be determined by $\operatorname{tr}(\Xi_\nu(a))=q_i\sum_j\sigma_{ij}^\nu(a)$ for $a\in L_i$. For $\nu>r$, take m_i embeddings $\sigma_{i1}^\nu,\ldots,\sigma_{im_i}^\nu$ of L_i into C which extend θ_ν so that $\sum_{\nu=1}^g\sum_{j=1}^{m_i}\sigma_{ij}^\nu$ is a CM-type of L_i. Consider these σ_{ij}^ν homomorphisms of L into C, and call Ψ_ν the diagonal representation of L of degree n composed of

$$\operatorname{diag}[\sigma_{i1}^\nu,\ldots,\sigma_{im_i}^\nu]\otimes 1_{q_i}$$

for $i=1,\ldots,t$. Define also an element Φ_i of I_{L_i} by

$$\Phi_i=\sum_{j=1}^{m_i}\left\{\sum_{\nu=1}^{r}2\sigma_{ij}^\nu+\sum_{\nu=r+1}^{g}(\sigma_{ij}^\nu+\sigma_{ij}^\nu\varrho)\right\}.$$

Then $p(\sigma_{ij}^\nu,\Phi_i)$ is meaningful. We can find $E_\nu\in\mathrm{GL}_n(\bar{Q})$ for $\nu\leqq r$ and $E_\nu\in\mathrm{GL}_{2n}(\bar{Q})$ for $\nu>r$ so that

$$E_\nu\chi_\nu(h(a),w)E_\nu^{-1}=\Psi_\nu(a)\quad\text{for all}\quad a\in L[\delta]\quad(\nu=1,\ldots,r),$$
$$E_\nu\chi_\nu(h(a))E_\nu^{-1}=\operatorname{diag}[\Psi_\nu(a)^\varrho,\Psi_\nu(a)]\quad\text{for all}\quad a\in L\quad(\nu=r+1,\ldots,g).$$

Put $p_{ij}^\nu=p(\sigma_{ij}^\nu,\Phi_i)^{1/2}$ and define diagonal matrices $P_\nu(w)$ by

$$P_\nu(w)=\begin{cases}\operatorname{diag}[P_{\nu1}\otimes 1_{q_1},\ldots,P_{\nu t}\otimes 1_{q_t}] & (\nu\leqq r),\\ \operatorname{diag}[P_{\nu1}\otimes 1_{q_1},\ldots,P_{\nu t}\otimes 1_{q_t},P_{\nu1}^{-1}\otimes 1_{q_1},\ldots,P_{\nu t}^{-1}\otimes 1_{q_t}] & (\nu>r),\end{cases}$$
$$P_{\nu i}=\operatorname{diag}[p_{i1}^\nu,\ldots,p_{im_i}^\nu].$$

Here we denote by $\operatorname{diag}[P_1,\ldots,P_s]$ the square matrix with square matrices $P_1,\ldots,P_s$ in the diagonal blocks and 0 in all other blocks.

THEOREM 3. *Given an arbitrary point* z_0 *of* $\mathfrak{H}_n^r$, *there exists, for each* $\nu\leqq g$, *a meromorphic function* S_ν *on* $\mathfrak{H}_n^r$, *with values in* $M_n(C)$ *or in* $M_{2n}(C)$ *according as* $\nu\leqq r$ *or* $\nu>r$, *such that:* (i) S_ν *is holomorphic at* z_0 *and* $\det(S_\nu(z_0))\neq 0$; (ii) $S_\nu(\gamma(z))=\chi_\nu(\gamma,z)S_\nu(z)$ *for all* γ *in a congruence subgroup of* G_{Q+}, *where we understand that* $\chi_\nu(\gamma,z)=\chi_\nu(\gamma)$ *if* $\nu>r$; (iii) *if* w *is the fixed point of* $h(Y[\delta])$ *as above and* S_ν *is holomorphic at* w, *then* $P_\nu(w)^{-1}E_\nu S_\nu(w)$ *is a* $\bar{Q}$-*rational matrix.*

This theorem enables us to define $\bar{Q}$-rational automorphic forms of a general type. Take a $\bar{Q}$-rational representation $\omega\colon(\mathrm{GL}_n)^r\times(\mathrm{GL}_{2n})^{g-r}\to\mathrm{GL}_d$, and consider a C^d-valued meromorphic function f on $\mathfrak{H}_n^r$. For every $\alpha\in G_{Q+}$ and $z\in\mathfrak{H}_n^r$, put

$$\omega(\alpha,z)=\omega(\chi_1(\alpha,z),\ldots,\chi_r(\alpha,z),\chi_{r+1}(\alpha),\ldots,\chi_g(\alpha)),$$
$$(f|_\omega\alpha)(z)=\omega(\alpha,z)^{-1}f(\alpha(z)).$$

We denote by $\mathfrak{A}_\omega$ the set of all C^d-valued meromorphic functions f such that $f|_\omega \gamma = f$ for all γ in a congruence subgroup of G_Q. Taking S_ν as in Theorem 3, we denote by $\mathfrak{A}_\omega(\bar{Q})$ the set of all elements f of $\mathfrak{A}_\omega$ such that $\omega(S_1, \dots, S_g)^{-1}f$ has components in $\mathfrak{A}_0(\bar{Q})$. This does not depend on the choice of S_ν. Moreover, if $f \in \mathfrak{A}_\omega(\bar{Q})$ and $\alpha \in G_{Q+}$, then $f|_\omega \alpha \in \mathfrak{A}_\omega(\bar{Q})$. The derivatives of elements of $\mathfrak{A}_0(\bar{Q})$ are $\bar{Q}$-rational in the sense described by

THEOREM 4. *For a meromorphic function f on $\mathfrak{H}_n^r$, define an $M_n(C)$-valued function $\Delta_\nu f$ on $\mathfrak{H}_n^r$ by*

$$(\Delta_\nu f)(z_1, \dots, z_r) = \left(\frac{1+\delta_{ij}}{2} \frac{\partial f}{\partial z_{\nu ij}}\right)_{i,j=1,\dots,n} \quad (\nu = 1, \dots, r),$$

where $z_{\nu ij}$ is the (i, j)-entry of the matrix variable z_ν on $\mathfrak{H}_n$. If $f \in \mathfrak{A}_0(\bar{Q})$ and S_ν is a function of Theorem 3, then the entries of $\pi^{-1}S_\nu^{-1}\Delta_\nu f \cdot {}^tS_\nu^{-1}$ belong to $\mathfrak{A}_0(\bar{Q})$.

The proofs of these theorems will be given in [12] in the case $n=1$. The general case can be proved in exactly the same fashion.

4. Finally we give new examples of affirmative answers to questions (IV) and (V). The notation being as in § 3, let E be a subfield of F such that $[E: Q]=r$ and the restrictions of $\theta_1, \dots, \theta_r$ to E are all different. Denote by R_E the maximal order of E and by U the group of all totally positive units of R_E. Take a Hilbert modular form

$$\varphi(z) = \sum_a c(a) \exp\left(2\pi i \sum_{\nu=1}^r a^{\theta_\nu} z_\nu\right) \quad \left(z = (z_1, \dots, z_r) \in \mathfrak{H}_1^r\right)$$

of weight $(h_1, \dots, h_r)$ $(\in 2^{-1}Z^r)$ with respect to a congruence subgroup of $SL_2(E)$. Specialize the group G_Q of § 3 to the case $n=1$. Then $G_Q = B^\times$ and $G_\nu = GL_2(R)$ for $\nu \leq r$. For $k=(k_1, \dots, k_r) \in Z^r$, $\alpha \in B$ and $z=(z_1, \dots, z_r) \in \mathfrak{H}_1^r$, put $T_k(\alpha, z) = \prod_{\nu=1}^r \mathrm{tr}([z_\nu]\alpha_\nu)^{-k_\nu}$, where

$$[b] = \begin{bmatrix} b & -b^2 \\ 1 & -b \end{bmatrix} \quad \text{for } b \in C.$$

Given a totally negative element ξ of F and an R_E-lattice Y of $\{\beta \in B \mid \beta^\iota = -\beta\}$, define a function f on $\mathfrak{H}_1^r$ by

$$f(z) = \sum_{0 \neq \alpha \in Y/U} c\left(\mathrm{Tr}_{F/E}(\xi\alpha\alpha^\iota)\right) T_k(\alpha, z), \tag{4.1}$$

where the sum is extended over all different cosets αU with $0 \neq \alpha \in Y$.

THEOREM 5. *Suppose $k_1 - h_1 = \dots = k_r - h_r = t$ with an element t of $2^{-1}Z$ greater than $3[F: E]$ and φ is either a constant or a cusp form. Then* (4.1) *is convergent and defines an automorphic form of weight $(2k_1, \dots, 2k_r)$. Moreover, if $2t \equiv [F: E] \pmod 2$ and the coefficients $c(a)$ are algebraic, $\pi^{-\{k\}}f$ is $\bar{Q}$-rational in the sense of* § 3, *where $\{k\} = \sum_{\nu=1}^r k_\nu$.*

This result holds even when φ is not a cusp form, provided that t is sufficiently large. Series of the same type can be defined also for orthogonal groups and unitary groups; they include Eisenstein series as special cases. Closely related to these are the series of type

$$W(s) = \sum_{0 \neq x \in X/U} c(\mathrm{Tr}_{F/E}(yxx^{\varrho})) \prod_{\nu=1}^{r} (x^{\sigma_\nu})^{-k_\nu} |x^{\sigma_\nu}|^{-s} \quad (s \in \boldsymbol{C}), \tag{4.2}$$

where X is an R_E-lattice of a totally imaginary quadratic extension K of F; y is an element of F such that $y^{\sigma_\nu} > 0$ or < 0 according as $\nu \leqq r$ or $> r$; $\sigma_1, \ldots, \sigma_r$ are embeddings of K into $\boldsymbol{C}$ such that $\sigma_\nu = \theta_\nu$ on F. Now W can be continued to a meromorphic function on the whole plane. If $F = E$ and φ is an Eisenstein series, W is essentially the product of two L-functions of K with Hecke characters.

THEOREM 6. *Suppose* $h \in \boldsymbol{Z}^r$ *and* $k_1 - h_1 = \ldots = k_r - h_r > [K:E]$. *Define* $\Psi \in I_K$ *by* $\Psi = \Omega_K + \sum_{\nu=1}^{r} (\sigma_\nu - \sigma_\nu \varrho)$, *where* Ω_K *is the sum of all embeddings of* K *into* $\boldsymbol{C}$. *Then* $\pi^{-\{k\}} W(0) \prod_{\nu=1}^{r} p(\sigma_\nu, \Psi)^{-k_\nu} \in \bar{\boldsymbol{Q}}$.

The subject of this section will be treated in more detail in a forthcoming paper.

References

1. R. M. Damerell, *L-functions of elliptic curves with complex multiplication*. I, II, Acta Arith. **17** (1970), 287—301; **19** (1971), 311—317.

2. P. Deligne, *Travaux de Shimura*, Sém. Bourbaki 1971, exp. 389, Lecture Notes in Math. vol. 244, Springer-Verlag, Berlin and New York.

3. P. Garrett, *Arithmetic properties of Fourier-Jacobi expansions of automorphic forms in several variables* (to appear).

4. K.-y. Shih, *Existence of certain canonical models*, Duke Math. J. **45** (1978), 63—66.

5. G Shimura, *On canonical models of arithmetic quotients of bounded symmetric domains*. I, II, Ann. of Math. (2) **91** (1970), 144—222; **92** (1970), 528—549.

6. ———, *On arithmetic automorphic functions*, Proc. Internat Congress of Math. 1970, vol. II, 1971, pp. 343—348.

7. ———, *On some arithmetic properties of modular forms of one and several variables*, Ann. of Math. (2) **102** (1975), 491—515.

8. ———, *On the derivatives of theta functions and modular forms*, Duke Math. J. **44** (1977), 365—387.

9. ———, *The arithmetic of automorphic forms with respect to a unitary group*, Ann. of Math. (2) **107** (1978), 569—605.

10. ———, *On certain reciprocity-laws for theta functions and modular forms*, Acta Math. **141** (1978), 35—71.

11. ———, *The special values of the zeta functions associated with Hilbert modular forms*, Duke Math. J. **45** (1978), 637—679.

12. ———, *Automorphic forms and the periods of abelian varieties*, J. Math. Soc. Japan (to appear).

13. J. Sturm, *Special values of zeta functions and Eisenstein series of half integral weight* (to appear).

14. A. Weil, *Elliptic functions according to Eisenstein and Kronecker*, Ergebnisse der Math. und ihrer Grenzgebiete, vol. **88**, Springer-Verlag, Berlin and New York, 1976.

PRINCETON UNIVERSITY
PRINCETON, NEW JERSEY 08540, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Exponential Diophantine Equations

R. Tijdeman

1. Historical introduction. Many questions in number theory concern perfect powers, numbers of the form a^b where a and b are rational integers with $a>1$, $b>1$. To mention a few:

(a) Is it possible that for $n\geq 3$ the sum of two nth powers is an nth power?

(b) Is 8, 9 the only pair of perfect powers which differ by 1?

(c) Is it possible that the product of consecutive integers, $(x+1)(x+2)\ldots(x+m)$, is a perfect power?

(d) Does a given polynomial with integer coefficients represent (infinitely many) perfect powers at integer points?

(e) Can a number with identical digits in the decimal scale be a perfect power?

A common feature of these problems is that they can be restated in the form of a diophantine equation in which exponents occur as variables. Problem (a) leads to the equation $x^n+y^n=z^n$ in integers $n\geq 3$, $x>1$, $y>1$, $z>1$ which is still unsolved in spite of Fermat's claim and all efforts thereafter. Problem (b) was posed by Catalan in 1844 and corresponds to the equation $x^m-y^n=1$ in integers $m>1$, $n>1$, $x>1$, $y>1$. Problem (c) goes back to Liouville who gave a partial solution in 1857. The complete solution was obtained by Erdős and Selfridge in 1975. They showed that $\prod_{j=1}^{m}(x+j)=y^n$ has no solutions in integers $m>1$, $n>1$, $x\geq 1$, $y>1$. Problem (d) is stated here in a general form, but special cases were investigated long ago. In 1850 V. A. Lebesgue proved that x^2+1 is never a perfect power for integral x. It follows from a more general result of Legendre in 1798 that x^3-y^3 is not a power of 2 for $x>y>1$. Ramanujan conjectured in 1913 that $2^3, 2^4, 2^5, 2^7$ and 2^{15} are the only powers of 2 assumed by the polynomial x^2+7 for integral x. In 1948 Nagell proved that the corresponding equation $x^2+7=2^n$ in integers $n>1$, $x\geqslant 1$ indeed has no further solutions. Problem (e) is still open. In 1810 P. Barlow stated that no square has all its digits alike. It follows from the work of Ljunggren, Obláth, Shorey and Tijdeman that the corresponding equation $a(10^m-1)/(10-1)=y^n$ in integers $1\leqslant a<10$, $m>1$, $n>1$, $y>1$ has no solutions unless $a=1$ and $n\geqslant 23$.

2. Methods for solving exponential diophantine equations. We consider equations $f(x_1^{m_1}, x_2^{m_2}, \dots, x_k^{m_k})=0$ in positive integers $m_1, m_2, \dots, m_k, x_1, x_2, \dots, x_k$ subject to certain conditions, where f is a polynomial with integer coefficients. In this section we distinguish three different approaches to such equations and give some examples of results obtained in that way. The most elementary attack to find all solutions of an exponential equation is the use of

(i) *Divisibility properties of rational integers.* The solution of Problem (c) by Erdős and Selfridge belongs to this type. In 1952 LeVeque proved that for fixed a, b the equation $a^x-b^y=1$ has at most one solution $x\geqslant 1$, $y\geqslant 1$ except for the case $a=3$, $b=2$. In this case there are exactly two solutions, as was proved by Lewi Ben Gerson (1288–1344). It was only in 1964 that Chao Ko proved that x^2-1 is never a perfect power if $x>3$. The corresponding results for x^2+1 and $x^3\pm 1$ had been proved by V. A. Lebesgue in 1850 and Nagell in 1921 respectively, but in their methods certain irrational numbers play a role.

(ii) *Algebraic methods.* Divisibility properties of numbers in certain algebraic number fields are used in several solutions of exponential equations. Most proofs of Ramanujan's assertion on the equation $x^2+7=2^n$ are based on properties of numbers in $Q(\sqrt{-7})$, a field with unique factorization. Another useful tool is the p-adic method of Skolem. It was exploited by Skolem, Chowla and Lewis in 1959 to give another solution of the equation $x^2+7=2^n$ and in 1945 by Skolem to give an algorithm for determining all solutions of certain equations of the form $a_1^{m_1}\dots a_k^{m_k}-b_1^{n_1}\dots b^{n_l}=c$ in positive integers $m_1, \dots, m_k, n_1, \dots, n_l$, where $a_1, \dots, a_k$, $b_1, \dots, b_l$, c are given positive integers. The finiteness of the number of solutions had been proved by Pólya in 1918 who applied a third type of methods.

(iii) *Approximation methods.* Pólya showed that his assertion is an immediate consequence of a result of Thue on binary forms. Thue's work was improved by Siegel in 1921. In 1933 Mahler proved a p-adic analogue of Siegel's result and deduced that the equation $x+y=z$ in coprime integers x, y, z composed of fixed primes has only finitely many solutions. In 1955 Roth improved upon Siegel's work and this was generalized by W. M. Schmidt in 1970. Six years later Dubois and Rhin and Schlickewei showed that a p-adic analogue of Schmidt's result implies that the equation $x_1+x_2+\dots+x_n=0$ in integers which are pairwise coprime and composed of fixed integers, has only finitely many solutions. A disadvantage of the method is that it is ineffective; no upper bounds for the solutions can be obtained by the Thue–Siegel–Roth–Schmidt method.

Siegel, and later Baker, developed work of Thue based on hypergeometric functions. Their results were applied by Inkeri in 1972 and by Shorey and Tijdeman in 1976 to the equation $(x^m-1)/(x-1)=y^n$. Very recently Beukers obtained a further extension, which enabled him to prove that the equation $x^2+D=p^n$ in integers $n>1$, $x\geqslant 1$ has at most four solutions when $D\neq 0$ and p is a prime, $p\nmid D$, and at most one solution when $p=2$, $D>0$, $D\neq 7, 23, 2^k-1$ for some $k\geqslant 4$. This method is effective only if an exceptionally large solution of a related equation is known.

An important effective method in transcendental number theory was developed by Gelfond and Schneider. It was applied by Gelfond in 1940 to the equation $\alpha^x+\beta^y=\gamma^z$ in integers x, y, z, where α, β, γ are fixed real algebraic numbers. In 1961 Cassels used it to give an effective proof of Pólya's result mentioned at the end of (ii). In 1968 Schinzel gave several other applications, for example to the equation $x^2+D=p^n$. About at the same time A. Baker found an ingenious generalization of Gelfond's method, which led to important new results on diophantine equations. They will be discussed later, but we note that it was used to prove that there are only finitely many solutions to problems (b), (e) and, under suitable conditions, problem (d). As to problem (a) no more than a partial result has been obtained.

3. Linear forms in logarithms of algebraic numbers. All presented applications of Baker's method to diophantine equations can be deduced from the following theorem and its p-adic analogue proved by van der Poorten in 1977.

THEOREM 1 (BAKER 1973, 1977). *Let* $0<\delta<\frac{1}{2}$. *Let* α_j *denote an algebraic number, not* 0 *or* 1, *with height at most* A_j $(\geqslant 4)$ *for* $j=1, \ldots, n$. *Put* $\Omega'=\prod_{j=1}^{n-1}\log A_j$. *Let* d *denote the degree of the field generated by the* α*'s over the rationals. Let* $b_1, \ldots, b_n$ *be nonzero rational integers with absolute values at most* B $(\geqslant 4)$. *Put* $\Lambda=b_1\log\alpha_1+\ldots+b_n\log\alpha_n$, *where the logarithms have their principal values. If* $\Lambda\neq 0$, *then*

(a) $$|\Lambda|>B^{-(16nd)^{200n}\Omega'\log\Omega'\log A_n},$$

(b) $$|\Lambda|>(\delta/|b_n|)^{C\log A_n}e^{-\delta B},$$

where $C>0$ *depends only on* d, n *and* $A_1, \ldots, A_{n-1}$.

There are numerous results which give refinements, often under suitable conditions. I may refer to recent work of van der Poorten, Loxton, Waldschmidt and Mignotte. For our purpose it is important to note that in (i) the best present methods yield an exponent not better than, say, $-10^6(10nd)^n\Omega'\log\Omega'\log A_n$.

4. Applications to polynomial diophantine equations. Baker, Coates, Feldman, Sprindžuk and others have used estimates for linear forms to give upper bounds for solutions of polynomial diophantine equations. In these cases it was already known by the Thue–Siegel–Mahler method that there are only finitely many solutions. We restrict our attention here to two results which are important for the applications to exponential equations.

THEOREM 2 (BAKER, 1968). *Let $f(x, y)$ be an irreducible binary form with degree $n \geq 3$ and with integer coefficients having absolute values at most H. Let c denote any positive integer. Then all solutions of $f(x, y)=c$ satisfy*

$$\max(|x|, |y|) < \exp\{(nH)^{(10n)^5} + (\log c)^{2n}\}.$$

For our purpose later improvements are not significant, since the dependence on n is not better than $\exp\{n^n\}$ because of the fact noted at the end of § 3. Coates and Sprindžuk have given p-adic analogues of Theorem 2.

THEOREM 3 (BAKER, 1969). *Let m be an integer, $m \geq 3$. Let $f(x)$ be a polynomial with degree $n \geq 2$ and with integer coefficients having absolute values at most H. Suppose that f has at least two simple zeros. Then all solutions of $f(x)=y^m$ satisfy*

$$\max(|x|, |y|) < \exp\exp\{(5m)^{10} n^{10n^3} H^{n^2}\}.$$

Moreover, if f has at least three simple zeros, then all solutions of $f(x)=y^2$ satisfy

$$\max(|x|, |y|) < \exp\exp\exp\{n^{10n^3} H^{n^2}\}.$$

Improvements of the last inequality were obtained by Sprindžuk in 1976 and by Choodnovsky (to appear).

5. Applications to exponential diophantine equations. The p-adic analogues of Theorem 2 given by Coates and Sprindžuk already deal with exponential polynomials, namely with the equation $f(x, y)=cp_1^{z_1}\dots p_s^{z_s}$ in integers $x, y, z_1, \dots, z_s$, where f and c are as in Theorem 2 and $p_1, \dots, p_s$ are fixed primes. It follows that under the conditions of Theorem 2 the greatest prime factor $P[f(x, y)]$ of $f(x, y)$ tends to ∞ when $\max(|x|, |y|) \to \infty$, $(x, y)=1$. Completing results of Sprindžuk and Kotov, Shorey, van der Poorten, Tijdeman and Schinzel obtained the following estimate.

THEOREM 4 (SHOREY *et. al.*, 1977). *Let $f(x, y) \in \mathbf{Z}[x, y]$ be a binary form such that among the linear factors in the factorization of f at least three are distinct. Then for all pairs x, y with $(x, y)=1$*

$$P[f(x, y)] \gg_f \log\log(\max(|x|, |y|)).$$

($G \gg_a H$ means that there is a constant $c>0$ depending only on a such that $G \geq cH$.)

It is an almost trivial consequence of Theorems 1 and 2 that for fixed integers a, b and $c \neq 0$ the equation $ax^n - by^n = c$ in integers $x>1, y>1$ *and* $n>2$ has only finitely many solutions. A much stronger result was established by Stewart.

THEOREM 5 (STEWART, 1976). $P[ax^n - by^n] \gg_{a,b} \sqrt{n/\log n}$.

The following result can be used in combination with the proof of Theorem 3 to determine all polynomials $f(x)\in \boldsymbol{Z}[x]$ which represent infinitely many perfect powers at integer points.

THEOREM 6 (SCHINZEL AND TIJDEMAN, 1976). *If a polynomial $f(x)$ with rational coefficients has at least two distinct zeros, then the equation $y^m=f(x)$ in integers m, x, y with $|y|>1$ implies that m is bounded.*

A generalization is given by Shorey *et al.* in 1977.

It turned out that in certain cases it was even possible to prove that two exponents with unrestricted bases are bounded. So one has in connection with problem (b)

THEOREM 7 (TIJDEMAN, 1976). *The equation $x^m-y^n=1$ in integers $m>1, n>1, x>1, y>1$ has only finitely many solutions.*

A p-adic result was given by van der Poorten in 1977.

To problem (e) only a partial answer could be obtained.

THEOREM 8 (LJUNGGREN, 1943, SHOREY AND TIJDEMAN, 1976). *The equation $(x^m-1)/(x-1)=y^n$ in integers $m>2$, $n>1$, $x>1$, $y>1$ with $mn>6$ has only finitely many solutions if at least one of the following conditions holds:* (a) *x is fixed,* (b) *m has a fixed prime divisor,* (c) *y has a fixed prime divisor.*

An application of Baker's method dealing with sums of equal powers of consecutive integers was obtained very recently.

THEOREM 9 (VOORHOEVE, GYŐRY AND TIJDEMAN). *Let $f(x)\in\boldsymbol{Z}[x]$ and let $a, m\in\boldsymbol{Z}$ with $a\neq 0$, $m\geqslant 2$. If the equation*

$$f(x)+1^m+2^m+\ldots+x^m = ay^n$$

in positive integers $n>1$, $x, y>1$ has infinitely many solutions, then

$$(m, n)\in\{(3, 2), (3, 4), (5, 2)\}.$$

The difficult part of the proof is the demonstration that the zeros of the polynomial $f(x)+1^m+2^m+\ldots+x^m$ have certain properties. Then Theorem 6 and a generalization of Theorem 3 can be applied.

With respect to problem (a) only partial results have been deduced. For example,

THEOREM 10 (STEWART, 1977, INKERI AND VAN DER POORTEN, *to appear*). *Let C be fixed. The equation $x^n+y^n=z^n$ has only finitely many solutions in integers $n\geqslant 3$, $x\geqslant 1$, $y\geqslant 1$, $z\geqslant 1$ with $y-x<C$.*

In both papers more information about the differences $y-x$ and $z-x$ is given.

6. Indications of some proofs. (a) *The equation $x^m-ay^m=b$ in integers $m>1$, $x>1$, $y>1$.* Note that x^m/ay^m is very close to 1 and hence $|\log a+m\log(y/x)|$ is extremely small. It follows from Theorem 1 that m is bounded and hence, from Theorem 2, that x and y are bounded.

(b) *The equation* $x^2-1=y^n$ *in integers* $n>1$, $x>1$, $y>1$. By factorizing the left side we find that both $2(x\pm 1)$ and $\frac{1}{2}(x\mp 1)$ are nth powers. This leads to an equation of the form $y_2^n=4y_1^n+4$ in integers $n>1, y_1>1, y_2>1$. By (a) n and $y=y_1y_2$ are bounded.

(c) *The equation* $x^m-y^n=1$ *in integers* $m>1, n>1, x>1$, $y>1$. By factorizing x^m-1 and y^n+1 we find that $x-1=\varrho y_1^n$ and $y+1=\sigma x_1^m$, where ϱ and σ are restricted in size. Hence $(\varrho y_1^n)^m/(\sigma x_1^m)^n$ is close to 1. It follows that $|m \log \varrho - n \log \sigma + mn \log (y_1/x_1)|$ is small. Now Theorem 1 can be applied to show that m and n are bounded.

7. Upper bounds for solutions. An important feature of all proofs of the results mentioned in § 5. is that they are effective. Only in very few instances have upper bounds been computed effectively. Langevin proved that if $x^m-y^n=1$, then $x^m<\exp\exp\exp(250)$. The best we can hope for with the present methods is a bound of the order of $\exp\exp\exp(10)$. I think that with the present methods the effective solution of an exponential equation by Baker's method is nigh to hopeless, unless the bases of all exponential variables are fixed. In those cases it might be possible. For example, Hunt and van der Poorten determined all solutions of $x^2+7=2^n$ and $x^2-11=5^n$ by Baker's method. The application of Theorem 1 gives $n<10^{20}$ and the remaining values have been checked. Beukers proved by using hypergeometric functions that if $x^2+D=2^n$, $D\neq 0$, then $n<500+15\log|D|$ and even $n<20+3\log|D|$ when $|D|<10^{12}$. The method works, since 181^2 is exceptionally close to 2^{15}. It is not applicable to base 5, since no square is exceptionally close to a power of 5. Baker's method is applicable to any equation $x^2+D=a^n$ in integers n, x.

8. Related results. In many cases results stated here for rational integers can actually be proved for algebraic integers in a given number field. This has been worked out by Sprindžuk and Kotov.

It is a straightforward application of Theorem 1 that in a given number field there are only finitely many units ε such that $1-\varepsilon$ is a unit. Such units were used by Lenstra in 1977 to determine Euclidean number fields.

Sprindžuk, Győry and Papp dealt with applications to norm form, discriminant form and index form equations. Győry also gave estimates for the degree of monic polynomials $f(x)\in \mathbf{Z}[x]$ with a given discriminant $D\neq 0$.

Schinzel, Stewart and Győry and Kiss applied Baker's method to Lucas and Lehmer numbers. Stewart proved, for example, that there are only finitely many Lucas and Lehmer numbers of index $n>12$ which do not have a primitive divisor.

An exponential equation with algebraic integers was solved by Baker's method in order to determine all elliptic curves over $\mathbf{Q}$ of conductor 11 (Agrawal, Coates, Hunt and van der Poorten).

I thank A. J. van der Poorten for his helpful comments while I prepared this paper.

References

(The corresponding sections are given in square brackets.)

1. L. E. Dickson, *History of the theory of numbers*, I—III, 1920; reprint, Chelsea, New York, 1952. [1].

2. L. J. Mordell, *Diophantine equations*. Academic Press, London, 1969. [1, 2].

3. W. M. Schmidt, *Approximation to algebraic numbers*. L'Enseignement Math. **17** (1971), 187—253. [2].

4. A. Baker, *Effective methods in Diophantine problems*, Proc. Sympos. Pure Math., Amer. Math. Soc., Providence. R. I., vol. 20, 1971, pp. 195—205; vol. 24, 1974, pp. 1—7. [3, 4].

5. R. Tijdeman, *Hilbert's seventh problem*. Proc. Sympos. Pure Math., vol. 28, Amer. Math. Soc., Providence, R. I., 1976, pp. 241—268. [3, 4].

6. *Transcendence Theory: Advances and Applications*, ed. by A. Baker and D. W. Masser, Academic Press, London, 1977, Chapters I—V. [3—6, 8].

7. V. G. Sprindžuk, *An effective analysis of the equations of Thue and Thue-Mahler*, Aktual'nye Probl. Analit. Teor. Cisel., Minsk, 1974, pp. 199—222. (Russian) [4, 8].

8. R. Tijdeman, *Some applications of Baker's sharpened bounds to diophantine equations*, Sém. Delange-Pisot-Poitou **16** (1975), No. 24 [6].

9. M. Langevin, *Quelques applications de nouveaux résultats de van der Poorten*, Sém. Delange-Pisot-Poitou **17** (1976), No. G 12 [7].

Mathematical Institute
Leiden, Netherlands

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Recent Work in Additive Prime Number Theory

R. C. Vaughan

I shall concentrate my attention on matters related to Goldbach's problem, since this exemplifies the whole of this subject.

Goldbach wrote two letters to Euler in 1742 in which he conjectured (a) that every integer greater than 2 is the sum of three primes, and (b) that every even natural number is the sum of two primes. He included unity as a prime, and since (a) then follows trivially from (b) it is the latter with which Goldbach's name has become associated. The modern form of this question excludes unity, of course, and asks that every even integer greater than 2 be the sum of two primes.

It is perhaps of some historical interest that Descartes had earlier made the same conjecture (essentially), but his comments were only published much later **[7]**.

Not until the early 1920s were any serious contributions made to the subject, and then there were two important developments, namely the Brun sieve and Hardy–Littlewood method. These have resulted in three principal lines of attack.

1. Direct applications of sieve methods. There are excellent surveys of earlier work in Halberstam and Roth **[14]** and Halberstam and Richert **[13]**. Let $P(x)=\prod_{p<x} p$. Then the sieve enables us to estimate the number of elements in

$$\{n-p\colon (n-p, P(n^{\theta}))=1,\ p<n\} \quad \text{or} \quad \{m\colon m<n,\ (m(n-m), P(n^{\theta}))=1\}$$

provided that θ is fairly small. It thereby enables us to conclude that if $n>n_0$, then $2n=P_k+P_l$ with $k+l\leq C$, where P_j denotes a number that is square free and has at most j prime factors. The most recent result is Chen's theorem **[2]**, **[3]** that $C=3$.

To understand Chen's contribution we must have a little more notation. Let, for large even n,

$$\mathscr{B} = \{n-p\colon (n-p, P(n^{1/10})) = 1,\ n-p \text{ square free},\ p < n\}.$$

$$S_1 = \frac{1}{2}\sum_{n^{1/10}\leq p<n^{1/3}} |\{b\in\mathscr{B}\colon p|b\}|,$$

$$S_2 = \frac{1}{2}|\{b\in\mathscr{B}\colon b = p_1p_2p_3,\ p_1 < n^{1/3} \leqslant p_2 < p_3\}|,$$

$$S = |\mathscr{B}|-S_1-S_2.$$

Both $|\mathscr{B}|$ and S_1 can be estimated directly by the sieve, which gives a good positive lower bound for $|\mathscr{B}|-S_1$. Elements of $\mathscr{B}$ with at least two prime factors in $[n^{1/10}, n^{1/3})$ will make a non-positive contribution to $|\mathscr{B}|-S_1$, and no element of $\mathscr{B}$ can be of the form $m p_1p_2p_3$ with $n^{1/3}\leqslant p_1<p_2<p_3$. Thus the only positive contribution to S can arise from those elements which are P_2's.

The expression S_2 as it stands is not entirely suitable for an application of the sieve. However, Chen obtains, by applying the sieve in a novel manner, an upper bound for S_2 which is smaller than the lower bound for $|\mathscr{B}|-S_1$ and thus secures $S>0$ and so the existence of a $p<n$ for which $n-p=P_2$. In fact he uses the sieve to estimate

$$\frac{1}{2}\left|\{n-p_1p_2p_3\colon p_1p_2p_3 < n,\ n^{1/10} \leqslant p_1 < n^{1/3} \leqslant p_2 < p_3,\ (n-p_1p_2p_3, P(n^{1/2}))=1\}\right|.$$

This requires, among other things, a new form of the Bombieri–Vinogradov prime number theorem.

Simpler proofs have been provided by Ding, Pan, Wang [10] and Ross [24]. Graham [12] has shown that the n_0 is effectively computable and Ross [25] has shown that it is possible to restrict the prime variables in various ways.

2. Indirect applications of sieve methods. This stems from Shnirel'man [26], [27] and was largely superseded by Vinogradov's work (see § 3). However, it is always useful to have alternative lines of approach to difficult problems. His idea was to use the sieve to show that

$$A(x) = |\{n \leqslant x\colon 2n = p_1+p_2 \text{ or } n = 1\}|$$

satisfies

$$A(x) \gg x \quad (x \geqslant 1). \tag{1}$$

This is readily obtained from upper estimates for

$$R(n) := \sum_{p_1+p_2=n} 1$$

provided by the sieve. For instance one may use Cauchy's inequality in the form

$$\Big(\sum_{n\leq x} R(n)\Big)^2 \leqslant \Big(\sum_{n\leq x} R(n)^2\Big) \sum_{n\leq x;\, R(n)>0} 1,$$

the left hand side being easily estimated by means of the prime number theorem. Then one can use theorems about the addition of sets $\mathscr{B}, \mathscr{C}$ of natural numbers,

$$\mathscr{B}+\mathscr{C} = \{n\colon n = b+c,\ b\in\mathscr{B},\ c\in\mathscr{C} \text{ or } n\in\mathscr{B} \text{ or } n\in\mathscr{C}\}$$

and their Shnirel′man density

$$\inf_{n\geqq 1}\frac{1}{n}\,|\{a \leqslant n\colon a\in\mathscr{A}\}|,\quad \mathscr{A} = \mathscr{B}, \mathscr{C} \text{ or } \mathscr{B}+\mathscr{C}$$

combined with (1) to show that there exists a C_0 such that for every $n \geqslant 2$ one has $n=p_1+\ldots+p_j$ with $j \leqslant C_0$. The most recent work in this direction has been due to Klimov, $C_0=6\times 10^9$ **[18]**; Klimov, Pil′tai, Sheptitskaya, $C_0=115$ **[19]**; Deshouillers, $C_0=75$ **[8]**; Vaughan, $C_0=27$ **[32]**. This last paper discusses two variants of the method, in one of which the calculations are easier and give $C_0=27$. The alternative variant, provided certain calculations can be carried out, does better than this, and thereby Deshouillers **[9]** has obtained $C_0=26$.

If instead one only asks for a constant C such that every sufficiently large n is the sum of at most C primes, then the method can be further refined. The most recent values for C have been $C=10$ due to Chechuro, Kuzjashev **[1]** and Siebert **[28]** and $C=6$ due to Vaughan **[31]**.

Of course, if the calculations could be carried out in the

3. Hardy-Littlewood-Vinogradov method, then we would have $C_0=4$. This method has its genesis in the work of Hardy and Littlewood **[15]**, **[16]**. They showed, in particular, on the assumption of the generalized Riemann hypothesis that (a)

$$R_3(n) = \sum_{p_1+p_2+p_3=n} 1$$

satisfies

(2) $$R_3(n) \sim \frac{n^2}{2(\log n)^3}\left(\prod_{p\mid n}\left(1-\frac{1}{(p-1)^2}\right)\right)\prod_{p\nmid n}\left(1+\frac{1}{(p-1)^3}\right),$$

and (b)

(3) $$E(x) = |\{n \leqslant x\colon 2n \neq p+p'\}|$$

satisfies

$$E(x) = O_\varepsilon(x^{1/2+\varepsilon}).$$

Of course, (2) implies that for n odd and large, $n=p_1+p_2+p_3$ is soluble in primes p_1, p_2, p_3.

Vinogradov **[35]**, by obtaining nontrivial estimates for the sum

(4) $$\sum_{p\leqq N} e^{2\pi i\alpha p} \quad\text{when}\quad \left|\alpha-\frac{a}{q}\right| \leqslant q^{-2},\ (a, q) = 1,\ (\log N)^c < q \leqslant N(\log N)^{-c}$$

was able to give an unconditional proof of (2). Later Linnik **[20]**, **[21]** (see also Chudakov **[5]**), Montgomery **[22]** and Vaughan **[33]** have given different ways of estimating (4).

Immediately following Vinogradov's work, Chudakov [4], van der Corput [6] and Estermann [11] all showed that

$$E(x) = O_A\big(x(\log x)^{-A}\big).$$

This was later improved to $E(x)=O\big(x\exp(-c\sqrt{\log x})\big)$ in Vaughan [29], [30] and then to $E(x)=O(x^{1-\delta})$ in Montgomery and Vaughan [23].

One of the fascinating aspects of additive prime number theory is the continual interraction between it and other areas of analytic number theory. For instance, the ideas contained in [33] have recently been used.

(a) to give [34] a new short proof of the Bombieri–Vinogradov prime number theorem,

(b) by Heath-Brown and Patterson [17] in their recent resolution of the Kummer problem regarding cubic Gaussian sums (to the effect that the arguments are uniformly distributed modulo 2π).

Let me explain the underlying ideas of [33]. Many theorems in analytic number theory depend on estimates for

$$\sum_n \Lambda(n)f(n) \quad \text{(or equivalently } \sum_p f(p))$$

where $f(x)$ when $x\in[1,X]$ is of the form $e^{iF(x)}$, F real valued, or is a Dirichlet character $\chi(x)$, and $f(x)=0$ otherwise. These in turn often depend on estimates for bilinear forms of the type

(I) $$\sum_m \sum_n x_m f(mn)$$

and

(II) $$\sum_m \sum_n x_m y_n f(mn).$$

The known bounds for these sums are often 'good' when in (I) m is restricted to an interval $[1,u]$ with u fairly small compared with X and in (II) m is restricted to an interval $[v, X/v]$ with v tending to infinity with X.

Vinogradov relates sums over primes to such bilinear forms *via* the sieve of Eratosthenes in the form

(5) $$f(1)+\sum_{\sqrt{X}<p\le X} f(p) = \sum_{m|P(\sqrt{X})}\ \sum_{n\le X/m} \mu(m)f(mn),\ P(Y)=\prod_{p\le Y} p.$$

The sum on the right is of type I but has the unfortunate defect of including m's that are close to X. He overcomes this by a combinatorial argument which says in effect that those m close to X with all their prime factors small occur relatively rarely and so can be neglected. The terms corresponding to the remaining m can be rearranged into the form

$$\sum_{V<p\le\sqrt{X}}\ \sum_{k\le X/p} x_p y_k f(pk),$$

which is a good bilinear form of type (II). However, in its sharpest form the combinatorial argument is rather complex and not well understood.

In [33], this is overcome by using instead the fundamental identity

$$S_0 := \sum_n \Lambda(n)f(n) = S_1 - S_2 - S_3 + S_4 \tag{6}$$

where

$$S_1 = \sum_{m \leqq u} \sum_n \mu(m)(\log n)f(mn);$$

$$S_2 = \sum_{m \leqq uv} \sum_n c_m f(mn), \quad c_m = \sum_{d \leqq u} \sum_{\substack{k \leqq v \\ dk = m}} \mu(d)\Lambda(k);$$

$$S_3 = \sum_{m > u} \sum_{n > v} \tau_m \Lambda(n) f(mn), \quad \tau_m = \sum_{d|m;\, d \leqq u} \mu(d);$$

$$S_4 = \sum_{n \leqq v} \Lambda(n) f(n).$$

For suitable choices of u and v, S_2 is a good sum of type (I), S_3 is a good sum of type (II) and S_4 is trivial. Moreover, in S_1 the log is easily removed by partial summation, whence it also becomes a good sum of type (I).

The proof of the fundamental identity (6) is very simple. It is an immediate consequence of the observation that

$$\tau_m = \begin{cases} 1 & (m = 1), \\ 0 & (1 < m \leqslant u). \end{cases}$$

Thus

$$\begin{aligned} S_0 - S_4 + S_3 &= \sum_m \sum_{n > v} \tau_m \Lambda(n) f(mn) \\ &= \sum_{d \leqq u} \mu(d) \Big(\sum_k \sum_n \Lambda(n) f(dkn) - \sum_{n \leqq v} \Lambda(n) \sum_k f(dkn) \Big) \\ &= S_1 - S_2. \end{aligned}$$

References

1. E. F. Chechuro and A. A. Kuzjashev, *The representation of large integers by sums of primes*, Studies in Number Theory, No. 3, Izdat. Saratov Univ., Saratov, 1969, pp. 45—50.

2. J. R. Chen, *On the representation of a large even integer as the sum of a prime and the product of at most two primes*, Kexue Tongbao (Foreign Lang. Ed.) **17** (1966), 385—386.

3. *On the representation of a larger even integer as the sum of a prime and the product of at most two primes*, Sci. Sinica **16** (1973), 157—176.

4. N. G. Chudakov, *On the density of the set of even numbers which are not representable as a sum of two odd primes*, Izv. Akad. Nauk SSSR Ser. Mat. **2** (1938), 25—40.

5. *On Goldbach—Vinogradov's theorem*, Ann. of Math. (2) **48** (1947), 515—545.

6. J. G. van der Corput, *Sur l'hypothèse de Goldbach pour presque tous les nombres pairs*, Acta Arith. **2** (1937), 266—290.

7. R. Descartes, Oeuvres, 10, 298.

8. J.-M. Deshouillers, *Amélioration de la constante de Šnirelman dans le probléme de Goldbach*, Sém. Delange-Pisot-Poitou 1972/3. Fasc. 2, Exp. No. 17, Paris, 1973.

9. *Sur la constante de Šnirel'man,* Sém. Delange-Pisot-Poitou 1975/6, Fasc. 2, Exp. No. G16, Paris, 1977.

10. X.-X. Ding, C.-D. Pan and Y. Wang, *On the representation of every large even integer as a sum of a prime and an almost prime,* Sci. Sinica **28** (1975), 599—610.

11. T. Estermann, *On Goldbach's problem: Proof that almost all even positive integers are sums of two primes,* Proc. London Math. Soc. **44** (1938), 307—314.

12. S. W. Graham, *Applications of sieve methods,* Dissertation, University of Michigan, Ann Arbor, 1977.

13. H. Halberstam and H.-E. Richert, *Sieve methods,* Academic Press, London, 1974.

14. H. Halberstam and K. F. Roth, *Sequences,* Vol. 1, Clarendon Press, Oxford, 1966.

15. G. H. Hardy and J. E. Littlewood, *Some problems of 'partitio numerorum'; III: On the expression of a number as a sum of primes,* Acta Math. **44** (1922), 1—70.

16. *Some problems of 'partitio numerorum' (V): A further contribution to the study of Goldbach's problem,* Proc. London Math. Soc. **22** (1924), 46—56.

17. D. R. Heath-Brown and S. J. Patterson, J. für Reine und Ang. Math. (to appear).

18. N. I. Klimov, *Apropos the computations of Shnirel'man's constant,* Volzh. Mat. Sb. Vyp. **7** (1969), 32—40.

19. N. I. Klimov, G. Z. Pil'tai and T. A. Sheptitskaya, *Estimation of the absolute constant in the Goldbach—Shnirel'man problem,* Issled. Teor. Chisel, Saratov **4** (1972), 35—51.

20. Ju. V. Linnik, *On the possibility of a unique method in certain problems of 'additive' and 'distributive' prime number theory,* Dokl. Akad. Nauk SSSR **49** (1945), 3—7.

21. *A new proof of the Goldbach-Vinogradov theorem,* Mat. Sb. **19 (61)** (1946), 3—8.

22. H. L. Montgomery, *Topics in multiplicative number theory,* Lecture Notes in Math., vol. 227, Springer-Verlag, Berlin and New York, 1971.

23. H. L. Montgomery and R. C. Vaughan, *The exceptional set in Goldbach's problem,* Acta Arith. **27** (1975), 353—370.

24. P. M. Ross, *On Chen's theorem that each large even number has the form p_1+p_2 or $p_1+p_2p_3$,* J. London Math. Soc. **10** (1975), 500—506.

25. P. M. Ross, *A short intervals result in additive prime number theory,* J. London Math. Soc. **17** (1978), 219—227.

26. L. G. Shnirel'man, *On additive properties of numbers,* Izv. Donskowo Politeh. Inst. **14** (1930), 3—28.

27. *Über additive Eigenschaften von Zahlen,* Math. Ann. **107** (1933), 649—690.

28. H. Siebert, *Darstellung als Summe von Primzahlen,* Diplomarbeit, Marburg, 1968.

29. R. C. Vaughan, *On Goldbach's problem,* Acta Arith. **22** (1972), 21—48.

30. *A new estimate for the exceptional set in Goldbach's problem,* Analytic Number Theory, Proc. Sympos Pure Math., Vol. 24 (St. Louis, Missouri, 1972), Amer. Math. Soc., Providence, R. I., 1973, pp. 315—320.

31. *A note on Shnirel'man's approach to Goldbach's problem,* Bull. London Math. Soc. **8** (1976), 245—250.

32. *On the estimation of Shnirel'man's constant,* J. Reine Angew. Math. **290** (1977), 93—108.

33. *Sommes trigonométriques sur les nombres premiers,* C. R. Acad. Sci. Paris Sér. A **285** (1977), 981—983.

34. *An elementary method in prime number theory,* Acta Arith. (to appear).

35. I. M. Vinogradov, *Some theorems concerning the theory of primes,* Recueil Math. **2 (44)** (1937), 179—195.

IMPERIAL COLLEGE
LONDON, ENGLAND

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Homogeneous Solutions of the Einstein Equations

O. I. Bogoyavlensky

Four-dimensional space-time manifolds $\mathcal{M}^4$ with the Einsteinian metric g_{ij} $(i, j=0, 1, 2, 3)$ with the signature $+ - - -$ are studied in the general theory of relativity. These manifolds satisfy the Einstein equations:

$$R_{ij}-\frac{1}{2}g_{ij}R = T_{ij} \tag{1}$$

where R_{ij} is the Ricci tensor of the metrics g_{ij}, $R=R_{ij}g^{ij}$ is the scalar curvature, T_{ij} is the stress-energy tensor of matter. The solution of the Einstein equations is called homogeneous, provided the manifold $\mathcal{M}^4$ allows the Lie group of isometries acting with three-dimensional orbits. Most important are the homogeneous solutions for which the manifold $\mathcal{M}^4=G\times R^1$, where R^1 is the axis of time, G is the three-dimensional Lie group. In this case the metric g_{ij} on the manifold $\mathcal{M}^4$ is invariant with respect to shifts to the elements of the group G and the restriction of the metric on the group G is determined negatively. Such homogeneous solutions are classified by nonisomorphic types of the corresponding Lie algebras.

The three-dimensional Lie algebras have been classified by Bianchi in 1898. It has been shown that there exist nine types of the three-dimensional Lie algebras, one of which (type I) is commutative (R^3), another (type II) is nilpotent, five types (III–VII) are solvable Lie algebras and two types are semisimple Lie algebras: type VIII is $\mathrm{SL}(2, R)$ and type IX is $\mathrm{SO}(3)$. The homogeneous solutions fall into two classes: in class A commutators of the corresponding Lie algebras have the form:

$$[X_i, X_j] = \varepsilon_{ijk}n_kX_k, \quad i, j, k = 1, 2, 3.$$

In class B commutators are reduced to the following form:

$$[X_3, X_1] = aX_1 + n_2 X_2, \quad [X_1, X_2] = 0, \quad [X_2, X_3] = n_1 X_1 - a X_2.$$

On the manifold $\mathcal{M}^4 = G \times R^1$ one may choose the basis of the right-invariant at the action of the group G vector fields X^0, X^1, X^2, X^3 (X^0 is tangent to the axis R^1, and X^1, X^2, X^3 are tangent to the group G, here $[X^0, X^i]=0$), in which the metric g_{ij} has the form:

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & & & \\ 0 & & -g_{ij}(t) & \\ 0 & & & \end{pmatrix}. \tag{2}$$

As usual, investigated are the homogeneous solutions with hydrodynamical stress-energy tensor of matter

$$T_{ij} = (p+\varepsilon) u_i u_j - p g_{ij} \tag{3}$$

where ε is the energy density, p is the pressure, $p = k\varepsilon$ $(0 \leqslant k < 1)$, u^i is the vector of the 4-velocity of matter and the matter is assumed not to move on the average, i.e., $u^0 = 1$, $u^i = 0$.

With these assumptions the well-known lemma is valid: at some value of t $\det(g_{ij}(t))$ turns to zero, i.e., the metric has a singularity. This singularity of metric may be so to say fictitious, connected with the given choice of coordinates and vanishing in other coordinates, as, for instance, in the well-known Taub-NUT solution. Of greater interest, however, are the real singularities of the metric connected with divergence of its geometric invariants. In particular, such singularities are in all homogeneous solutions at the presence of matter, i.e., at $\varepsilon \neq 0$, since all the homogeneous solutions have the first integral

$$H = \varepsilon (\det g_{ij})^{(1+k)/2} = \text{Const} \tag{4}$$

and the invariant scalar curvature R is connected with the density of energy ε by the expression $R = -(1-3k)\varepsilon$ (consequently, $|R| \to \infty$ at $\det(g_{ij}) \to 0$).

While studying the homogeneous solutions it is convenient to use the Hilbert variational principle which states that on the Einstein equation solutions (1) the first variation

$$\delta \int_{\mathcal{M}^4} (R + \Lambda)(-g)^{1/2} dx = 0 \tag{5}$$

where variations of the metric are finite. It is easy to conclude that the homogeneous solutions with the compact Lie groups of isometries, for instance, SO (3), satisfy the corresponding homogeneous variational principle (since the homogeneous variations for compact groups are finite):

$$\delta \int L\left(g_{ij}, \frac{dg_{ij}}{dt}\right) dt = 0. \tag{6}$$

All the homogeneous solutions of class A satisfy the same variational principle.

For these homogeneous solutions the Einstein equations are, consequently, the Lagrangian ones and after certain Legendre transformation $p_{ij}=\partial L/\partial \dot{g}_{ij}$ they turn into the Hamiltonian system in the twelve-dimensional phase space p_{ij}, g_{ij}. The dynamical systems describing homogeneous solutions of class B are not Hamiltonian.

Thus, the investigation of any type homogeneous solutions is reduced to that of the corresponding dynamical system in the 12-dimensional phase space p_{ij}, g_{ij}. The order of these dynamical systems may be diminished, since: first, each system is considered only on the level of the constraints determined by the Einstein equations $R_{0\alpha}=T_{0\alpha}$, and then, the dynamical systems obtained are invariant relative to the action of the group of internal automorphisms of the corresponding Lie algebra.

As a result of such a decrease of order the dynamical system for homogeneous solutions of types I and V is reduced to two-dimensional systems. Special homogeneous solutions of types II–VII possessing an additional symmetry may be also reduced to the two-dimensional systems. The investigation of such solutions by traditional methods of the Poincare–Bendicson two-dimensional qualitative theory has been carried out in [1].

However, the dynamics of general homogeneous solutions of types II–IX is not reduced to two-dimensional systems; therefore, for their investigation the author and S. P. Novikov [2], [3] have applied modern methods of multi-dimensional qualitative theory of dynamical systems. The new method of maximal non-degenerated compactification of dynamical system is the main method in these papers. Further this method has been effectively applied by the author in the investigation of the dynamics of perturbations of some integrable systems [4] and in the astrophysical work [5]. Let us describe the essence of this method on an example of studying homogeneous solutions of the Einstein equations.

1. Homogeneous solutions are described by dynamical systems with polynomial right-hand sides. These dynamical systems are defined in a noncompact region S_1 of the phase space, isolated by physical conditions of positivity of the density of energy ε and of metric $g_{ij}(t)$. The dynamical systems under consideration have very degenerated singular points in which $g_{ij}=0$.

2. To investigate the dynamical system in the region S_1 it is necessary to compactify this region at infinity and resolve as much as possible strongly degenerated singular points of the dynamical system. Compactification at infinity is realized by means of transition to projective coordinates. In order to resolve degenerated singular points it is required that successive σ-process should be performed, in this case polynomiality of the right-hand sides of the dynamical systems are used.

As a result of these transformations the dynamical system in region S_1 turns into an equivalent dynamical system, defined on some compact manifold S with the boundary Γ, the dynamical system being smoothly continued to the boundary Γ. The manifold S is a cell complex, and the boundary Γ consists of several components Γ_i intersecting in the corners of the boundary. The components of the boundary Γ_i correspond to compactification of the region S_1 at infinity and to resolution of degenerated singular points, and also in the physical region at $\varepsilon=0$. The construction of the manifold S is realized so that the resulting dynamical system on S should have maximal non-degenerated singular points. In this case the singular points form some manifolds and are non-degenerated (in the transversal directions), provided the number of their non-zero eigenvalues is equal to the codimension of these manifolds.

3. The dynamics of homogeneous solutions is studied on the basis of the above construction of the compact manifold S and of the dynamical system on it. The dynamical systems under consideration have a monotone function $F=d(\det(g_{ij}))^{1/6}/dt$, due to the system $dF/d\tau\leqslant 0$. At the compression of space $F\to-\infty$, and all the trajectories of the dynamical system approach the components of the boundary Γ which correspond to the metric singularity. The dynamical system at the boundary Γ is very much simplified and allows detailed investigation, as a result of which we get a complete enumeration of all the regimes of the metric behaviour in the neighbourhood of the singularity $\det(g_{ij})=0$.

The singular points with separatrices passing inside the manifold S define all the different power (with respect to t) asymptotics of the metric under compression of space (and in some homogeneous solutions at the infinite expansion of space).

Some unstable singular points lying in the corners of the boundary Γ have no separatrices passing inside the manifold S. All their separatrices lie on different components of the boundary Γ and, as the direct integration shows, they pass from one singular point to another. All the available separatrices transitions between the singular points form a separatrix diagram of the dynamical system. The separatrices connecting the singular points form successions which may terminate only at the attracting singular points. For the homogeneous solutions of the types I–VII the dynamical system on the manifold S has attracting singular points and for such solutions only the final successions of the separatrix transitions are realized.

The dynamical systems on the manifold S for the homogeneous solutions of the types VIII and IX possess only unstable singular points (no attracting singular points at all). In this case there arise infinite successions of separatrices passing between the singular points. These separatrices fill some closed manifold which is an attracting manifold of the dynamical system due to the presence of monotone functions of the type of function F. Recently such manifolds have been called "strange attractors". It is essential that the dynamics arising on a strange attractor in the general theory of relativity allows a complete investigation. The trajectories approaching this strange attractor describe the oscillation regime of the metric behaviour in the neighbourhood of singularity which has been discovered earlier in [6], [7] with the help of some other methods.

As a particular example we consider homogeneous solutions of the type IX possessing a group of isometries SO (3). These solutions are described by the

Hamiltonian system in the six-dimensional phase space p_i, q_i:

$$\dot{p}_i = -\partial H/\partial q_i, \quad \dot{q}_i = \partial H/\partial p_i, \tag{7}$$

with the Hamiltonian

$$H = (q_1 q_2 q_3)^{-(1-k)/2}\left[2\sum_{i<j}^{3} p_i q_i p_j q_j - \sum_{i=1}^{3} p_i^2 q_i^2 + \frac{1}{4}\left(2\sum_{i<j}^{3} q_i q_j - \sum_{i=1}^{3} q_i^2\right)\right].$$

Here the coordinates q_i are eigenvalues of the matrix $g_{ij}(t)$. The constant value of the Hamiltonian H is connected with the density of energy ε by the relation (4): $H=\varepsilon(q_1 q_2 q_3)^{(1+k)/2}$. The system (7) is considered in the non-compact region isolated by the conditions $H>0$, $q_i>0$. The system (7) is invariant with the action of a group of scale transformations

$$q_i \to \lambda q_i, \quad p_i \to p_i, \quad \tau \to \lambda^{(1-3k)/2}\tau$$

and therefore it admits the lowering of the order. In this case the compact manifold S is a five-dimensional cell complex and is covered by the coordinate system

$$y_i = q_i/G, \quad s_i = p_i q_i/P, \quad w = G^2/P^2,$$
$$G = (q_1^2+q_2^2+q_3^2)^{1/2}, \quad P = (p_1^2 q_1^2 + p_2^2 q_2^2 + p_3^2 q_3^2)^{1/2}$$

and by the coordinate system y_i, $\bar{s}_i = p_i q_i/G$. The coordinates $s_i, w, \bar{s}_i, y_i$ cover the manifold $S^2\times D^3$, where S^2 is a unit sphere in the coordinates y_i, D^3 is a three-dimensional ball covered by the coordinates s_i, w and $\bar{s}_i$. The cell complex S is isolated on this manifold by the conditions $y_i \geqslant 0$, $w \geqslant 0$, and with the compression of space, by an additional condition $s_1+s_2+s_3 \leqslant 0$.

All the singular points of the dynamical system on the manifold S lying on the boundary Γ are unstable and form the following manifolds: a two-dimensional triangle of singular points Φ, nine segments of singular points A_i, B_i, T_i ($i=1, 2, 3$), three isolated singular points N_i and three circumferences of singular points (ψ, i) $(0 \leqslant \psi \leqslant 2\pi)$. The separatrices entering the singular points φ, T_i, N_i define all the power asymptotics of the metric available in the solutions under consideration. The separatrices leaving the singular points on the circumference (ψ, i) lie on the boundary Γ and then again enter the singular points on some other circumference (ψ, k). These separatrices fill some manifold which is a strange attractor in the dynamical system under consideration.

The trajectories moving along successions of separatrices on the strange attractor appear from time to time in the neighbourhood of the circumferences (ψ, i). Each of these circumferences (ψ, i) lies in the corner of the boundary Γ and on it we have $y_i=1$, $y_j=y_k=0$. Therefore in the neighbourhood of the circumference (ψ, i) we have $q_i \gg q_j, q_k$, and, thus, while the trajectory moves in the neighbourhood of the strange attractor one of the q_i from time to time happens to be much greater than the other two, i.e., the eigenvalues of the metric $g_{ij}(t)$ oscillate in a very complicated manner. These oscillations describe the oscillatory regime of the approaching of metric to the singularity which is the most general regime of the metric behaviour at the compression of space.

References

1. C. B. Collins, *More qualitative cosmology*, Comm. Math. Phys. **23** (1971), 137—159.

2. O. I. Bogoyavlensky and S. P. Novikov, Qualitative theory of homogeneous cosmological models, Proceedings of the Petrovsky Seminar, No. I, 1975, pp. 7—43.

3. *Homogeneous models in the general theory of relativity and gas dynamics*, Uspehi Math. Sci. **31** (1976), 33—48.

4. O. I. Bogoyavlensky, *On perturbations of the periodic Toda lattice*, Comm. Math. Phys. **5** (1976), 201—209.

5. *New solutions in the theory of star explosion*, Geoph. Astroph. Fluid Dynamics (to appear).

6. V. A. Belinskii, E. M. Lifshitz and I. M. Khalatnikov, *Oscillatory regime of approaching to the singular point in relativistic cosmology*, Uspehi Fiz. Nauk, **102** (1970), 463—501.

7. C. W. Misner, *Mixmaster universe*, Phys. Rev. Lett. **22** (1969), 1071—1074.

Academy of Sciences of the USSR
Moscow 117 334, USSR

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Minimal Surfaces: Stability and Finiteness

Manfredo P. do Carmo

The last ten years have seen an intense activity on certain questions that arise in connection with the study of minimal surfaces. Among such questions one should mention those of regularity, embeddability, stability and finiteness of the number of minimal surfaces spanning a given boundary. In this lecture I would like to describe a few ideas, results and problems related to the questions of stability and finiteness. For simplicity, I will restrict myself to minimal surfaces of the topological type of the disk in a Riemannian manifold.

Let M^n be an n-dimensional C^k-Riemannian manifold, $k\geq 3$. B will be the open unit disk in R^2 with closure $\bar{B}$ and boundary ∂B. Let $\gamma\subset M$ be a closed rectifiable Jordan curve. A *generalized minimal surface bounded* by γ is a map $f\colon \bar{B}\to M$ that is C^2 in B, C^0 in $\bar{B}$ and satisfies:

(1) The restriction $f|\partial B$ is a homeomorphism onto γ.

(2) f is conformal in B, that is,

$$|f_x| = |f_y|, \quad \langle f_x, f_y\rangle = 0, \quad (x, y)\in B,$$

where $\langle\ ,\ \rangle$ denotes the Riemannian metric of M. The points where $|f_x|=|f_y|=0$ are the (interior) *branch points* of f in B.

(3) f is a critical point of the area function for all variations that leave ∂B fixed. When such a critical point is a relative minimum, f is *stable*.

The most significant, and historically the first, example of a generalized minimal surface occurs for the case $M^n=R^n$. A version of the classical Plateau problem asks whether given $\gamma\subset R^n$ there exists an f satisfying (1) and:

(4) f minimizes area for all $g\in C^2(B)\cap C^0(\bar{B})$ that satisfy (1).

The answer, given by Douglas and Rado around 1930, can be stated as follows (for general references see [8]).

THEOREM 1. *Let* $M^n = R^n$. *Given* $\gamma \subset R^n$, *there exists a generalized minimal surface bounded by* γ *which satisfies* (4). *Such an* f *is called a Douglas solution to the Plateau problem.*

For the purpose of fixing the notation, let us recall the ideas of the proof. One observes initially that it suffices to minimize the Dirichlet integral (=the energy of f)

$$D = 1/2 \int_B |\mathrm{grad} f|^2 \, dx \, dy.$$

For that, we first minimize D among all maps with finite energy that restricted to ∂B is a given continuous map $g: \partial B \to \gamma$. The solution $\hat{g}$ is the *harmonic extension* of g and the space of all such harmonic extensions is denoted by $\tilde{H}(\gamma)$. By noticing that D is invariant under conformal transformations, we normalize all elements in $\tilde{H}(\gamma)$ by requiring that three fixed distinct points of ∂B are taken in three fixed distinct points of γ; the resulting space is denoted by $H(\gamma)$. The crucial point is now to prove that $H^N = \{\hat{g} \in H(\gamma); D(\hat{g}) \leq N\}$ is compact in the topology of uniform convergence. The theorem follows from the lower semicontinuity of D in $H(\gamma)$.

Through the work of several mathematicians (see [8]), the following regularity theorem has been obtained the first part of which allows us to introduce the notion of boundary branch points.

THEOREM 2. *Let* $\gamma \subset R^n$ *be of class* C^k *and let* f *satisfy* (1), (2) *and* (3). *Then* f *is of class* C^{k-1} *in* $\bar{B}$. *Furthermore, for* $n=3$, *Douglas solutions have no interior branch points.*

From now on, we will denote by $M(\gamma)$ the space of generalized minimal surfaces bounded by γ and normalized as in $H(\gamma)$. The immersed surfaces (i.e., those with no interior or boundary branch points) will be denoted by $M_i(\gamma) \subset M(\gamma)$.

It is natural to ask how the geometry of γ influences the properties of $M(\gamma)$. For instance, what properties of γ imply that $M(\gamma)$ contains only one element? For a long time the only known conditions for such a uniqueness, due to Radó (see [8]), were: (i) γ is a plane curve, (ii) γ has a one-to-one convex projection onto a plane $P \subset R^n$. Recently, J. C. Nitsche found a further condition for $n=3$.

THEOREM 3 ([12], [6]). *Let* $\gamma \subset R^3$ *be of class* C^3 *and assume that the total curvature of* γ *is smaller than* 4π. *Then* $M(\gamma)$ *contains a unique element.*

The idea of the proof is as follows. From a formula of Gauss–Bonnet type that holds for $f \in M(\gamma)$, we conclude that f has no branch points in $\bar{B}$ and $\int_B |K| \, d\sigma < 2\pi$ (here and always K is the Gaussian curvature of the induced metric). From a theorem of Barbosa and Carmo [1], this implies that f is stable and nondegenerate. The crucial point is then to prove that f is a strict relative minimum for the energy D in the space $H(\gamma)$ with the topology of uniform convergence. A theorem of

Shiffman [14] states that two such relative minima imply the existence of a third critical point of D, not a minimum. Thus $M(\gamma)$ contains only one element.

Actually, the result of [1] states that if $\gamma \in M_i(\gamma)$, $\gamma \subset R^3$, and $\int_B |K|\,d\sigma < 2\pi$ then the first eigenvalue of the problem

$$(1) \qquad \Delta u - \lambda K u = 0 \quad \text{in} \quad B,\ u \equiv 0 \quad \text{in } \partial B,$$

is greater than 2 (Δ is the Laplacian in the induced metric). This can be used to show that *if* $M(\gamma) = M_i(\gamma)$ *and* $\int_\gamma |k|\,ds < 6\pi$, then $M(\gamma)$ *consists of isolated points.*

The argument is again due to Nitsche [13] and goes as follows. If $f \in M(\gamma)$ is not isolated, one can prove the existence of a function u satisfying

$$\Delta u - 2Ku = 0 \quad \text{in} \quad B,\ u \equiv 0 \quad \text{in } \partial B.$$

Thus 2 is an eigenvalue for the problem (1). In fact, 2 is the first such eigenvalue. Otherwise, u changes sign in B and is the first eigenfunction of a subdomain $B_m \subset B$, where $\int_{B_m} |K|\,d\sigma \leq \frac{1}{2} \int_B |K|\,d\sigma$. By the above quoted Gauss–Bonnet formula, the condition on the total curvature of γ implies that $\int_B |K|\,d\sigma < 4\pi$. This leads to a contradiction with the result of [1] and shows that u is a first eigenfunction of the problem (1). Since the space of such eigenfunctions is 1-dimensional, it follows from the work of Böhme and Tomi [3], [15] that a connected component of $M(\gamma)$ that contains f is a compact 1-dimensional manifold. This leads again to a contradiction, essentially because Jacobi fields in R^3 cannot "turn around" without changing sign.

The last part of the above argument had been previously used by F. Tomi [15] to show that the set of Douglas solutions bounded by $\gamma \subset R^3$ that have no boundary branch points is finite (see also [16]).

A generic approach to the question of finiteness has been developed by A. Trombe [17] and later, with different methods and less differentiability requirements, by L. P. Melo Jorge [10]. The idea is to give $\tilde{H}(\gamma)$ a structure of a Hilbert manifold so that $D\colon \tilde{H}(\gamma) \to R$ is differentiable and the critical points of D are the (non-normalized) elements of $M(\gamma)$. A crucial point is then to find a notion of non-degeneracy for elements in $M(\gamma)$ that is invariant under small perturbations of γ in the space Γ of embeddings of S^1 into R^n. This is obtained in [17] from the transversality properties of a certain vector field defined on the space $\bigcup_{\gamma \in \Gamma} \tilde{H}(\gamma)$ and in [10] from the ellipticity of a certain operator associated to the second derivative of D. With such methods it can be proved, for instance, that if all $f \in M(\gamma_0)$ are nondegenerate, then $M(\gamma_0)$ is finite and this number is constant for γ in a neighborhood $W \subset \Gamma$ of γ_0. Since it can be shown that plane curves are nondegenerate, it follows that curves close to a plane curve bound a unique minimal surface. A related result has been obtained by W. Meeks [9]. Uniqueness also holds for curves nearby those satisfying the second of the above Radó's conditions [10].

Recently, Böhme and Tromba [4] proved that there exists an open dense subset $\hat{\Gamma} \subset \Gamma$ with the C^∞ topology such that if $\gamma \in \hat{\Gamma}$ then $M(\gamma)$ is finite.

Now, Theorems 1 and 2 have complete generalizations for a Riemannian manifold M if one assumes that either: (i) M is well behaved at infinity **[11]** (for instance, if M is compact) and γ is contractible to a point, or (ii) the sectional curvature K_M of M satisfies $K_M \leq b^2$, b real or pure imaginary, and γ is contractible to a point in such a way that the length l of the longest transversal curve of the contraction satisfies $4l^2b^2 \leq \pi^2$ [5].

This raises the question of extending the above finiteness theorems to such a situation. For instance, it has been proved in [2] that if $f: \bar{B} \to H^3(a)$ is a minimal immersion into the hyperbolic 3-space $H^3(a)$ with constant curvature $a<0$, and $\int_B |K|\, d\sigma < 2\pi$, then f is stable. By a result of Kaul [7], if the total curvature of a curve $\gamma \subset H^3(a)$ is smaller than 4π, we still obtain that $M(\gamma) = M_i(\gamma)$ and $\int_B |K|\, d\sigma < 2\pi$. It is likely that one can extend Nitsche's arguments as well as Shiffman's theorem to hyperbolic spaces, and this would imply that Nitsche's 4π-theorem holds true in $H^3(a)$. Actually, it may well happen that most of the above finiteness theorems have natural extensions to three-dimensional, simply-connected, Riemannian manifolds with negative curvature.

Stability for immersed minimal surfaces is in itself an interesting question; it is strongly related to the question of estimating the eigenvalues of the Laplacian of bounded domains in terms of the geometry of the domain. Recently Barbosa and do Carmo developed a general method to attack this question [2], and obtained stability results for minimal surfaces in various classes of Riemannian manifolds. In fact, there are reasons to believe that the following general statement may be true.

Let R be the curvature tensor of M^n and let ∇R be its covariant differential. Assume that there are numbers C_1 and C_2 such that $\|R\| < C_1$, $\|\nabla R\| < C_2$. Let $f: \bar{B} \to M^n$ be a minimal immersion and denote by K the Gaussian curvature of the induced metric ds^2. Then there exists a nonnegative function $\Gamma(K, C_1)$ on $\bar{B}$ and a number $A>0$, depending only on n, C_1 and C_2, such that if $\int_B \Gamma\, d\sigma < A$, then f is stable.

This conjecture has been verified for Riemannian manifolds with constant curvature, and sharp bounds have been found for R^3 and the three-sphere [1], [2]. The critical point in the proof is to estimate, in terms of n, C_1 and C_2, the Gaussian curvature of a new metric $d\sigma^2 = \Gamma ds^2$ on B. Once this is obtained, a comparison theorem for eigenvalues, as developed in [2], could be used to complete the proof. Whether or not the above statement is true, the search for specific and interesting stability bounds still leaves much room for investigations.

I would like to conclude with a few specific questions:

(1) Is a complete globally stable minimal immersion in R^3 flat? Here globally stable means that any bounded domain is stable.

(2) Let $f: \bar{B} \to R_n$, $n>3$, be a minimal immersion with $\int_B |K|\, d\sigma < 2\pi$. Is f stable? This is true if we replace 2π by $4\pi/3$ [2].

(3) Let M^3 be a simply-connected Riemannian manifold with $K_M \leq 0$. Let $f: \bar{B} \to M^3$ be a minimal immersion with $\int_B |K|\, d\sigma < 2\pi$. Is f stable?

References

1. J. L. Barbosa and M. do Carmo, *On the size of a stable minimal surface in R^3*, Amer. J. Math. **98** (1976), 515—526.

2. ——— *Stability of minimal surfaces and eigenvalues of the Laplacian* (preprint).

3. R. Böhme and F. Tomi, *Zür Struktur der Lösungmenge Plateauproblems*, Math. Z. **193** (1973), 1—29.

4. R. Böhme and A. Tromba, *The index theorem for classical minimal surfaces* (preprint, Bonn 1977).

5. R. Gulliver, *The Plateau problem for surfaces of prescribed mean curvature in a Riemannian manifold*, J. Differential Geometry **8** (1972), 317—330.

6. R. Gulliver and J. Spruck, *On embedded minimal surfaces*, Ann. of Math. **103** (1976), 331—347.

7. H. Kaul, *Isoperimetrische Ungleichung und Gauss-Bonnet Formel für H-Flächen in Riemannschen Mannigfaltigkeiten*, Arch. Rational Mech. Anal. **45** (1972), 194—221.

8. H. B. Lawson, *Lectures on minimal submanifolds*, IMPA, Rio de Janeiro, 1973.

9. W. H. Meeks III, *Lectures on Plateau's problem*, IMPA, Rio de Janeiro, 1978.

10. L. P. de Melo Jorge, *Estabilidade C^2 das curvas com soluções não degeneradas do problema de Plateau*, D. Sc. Thesis, IMPA, 1978.

11. C. B. Morrey, Jr., *The problem of Plateau on a Riemannian manifold*, Ann. of Math. **49** (1948), 807—851.

12. J. C. C. Nitsche, *A new uniqueness theorem for minimal surfaces*, Arch. Rational Mech. Anal. **52** (1973), 319—329.

13. ——— *Contours bounding at most finitely many solutions of Plateau's problem* (preprint).

14. M. Shiffman, *Unstable minimal surfaces with any rectifiable boundary*, Proc. Nat. Acad. Sci. U.S.A. **28** (1942), 103—108.

15. F. Tomi, *On the local uniqueness of the problem of least area*, Arch. Rational Mech. Anal. **52** (1973), 312—318.

16. ——— *On the finite solvability of Plateau's problem*, Lecture Notes in Math. Vol. 597, Springer-Verlag, Berlin and New York, 1977, pp. 679—695.

17. A. Tromba, *On the number of simply-connected minimal surfaces spanning a curve*. Mem. Amer. Math. Soc. No. 194 (1977).

INSTITUTO MATEMATICA PURA E APPLICADA
RIO DE JANEIRO, BRASIL

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Conjectures and Open Questions in Rigidity

Robert Connelly*

I. Introduction. The study of the rigidity of frameworks and surfaces is fairly old, going back at least to the time of Euler, but progress has been slow and painful. Even today knowledge is meager, and many basic simple questions are still unanswered. One gets the feeling that in the past these questions were considered more a source of embarrassment than a source of conjectures or goals for future study. My feeling is that reasonable conjectures and problems are useful and essential to progress in mathemtics. Even if such questions turn out to be poorly stated, ambiguous, or ultimately uninteresting, they usually have been useful, at least for inspiration if nothing else.

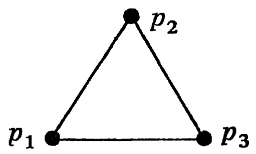
Rigidity is a subject very close to the reality of our world. If a structure is proved to be rigid one can always build it and see if it collapses. Theorems had better hold up. The opportunities for beneficial interaction among engineers, architects, and mathematicians seem very attractive. The subject is versatile and sufficiently undeveloped that contributions can be made at all levels.

I have chosen five categories of questions about which I think it would be nice to know more information. Naturally, some are ambiguous and not too detailed. If I knew the answers, I would not be asking.

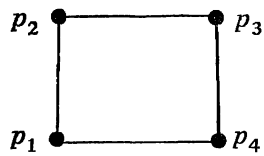
II. Definitions and history. It is useful to state rigidity in terms of things called frameworks. The name is probably inspired from structural engineering, and physically one should think of a framework as a gadget constructed out of dowel rods with their ends stuck in small flexible rubber connectors. Mathematically,

* Supported by a grant from the National Science Foundation of the United States.

a *framework* F is just a finite ordered collection of points $p=(p_1, \ldots, p_v)$, p_i in Euclidean n-space (n is usually only 2 or 3), together with a collection of certain pairs of points $\mathscr{E}$ called the rods. For example,



a triangle



a square

$\mathscr{E} = \{(p_1, p_2), (p_2, p_1), (p_3, p_1)\}$ $\mathscr{E} = \{(p_1, p_2), (p_2, p_3), (p_3, p_4), (p_4, p_1)\}$

A *flex* of F is a continuous motion of all the vertices $p(t)=(p_1(t), p_2(t), \ldots, p_v(t))$, $0 \leqslant t \leqslant 1$, so that $p(0)=p$ and the length of any rod (p_i, p_j) in F, $|p_i(t)-p_j(t)|$, is constant for all t. The flex $p(t)$ is *rigid* if *all* the distances $|p_i(t)-p_j(t)|$ are constant. In the latter case it turns out that $p(t)$ is obtained by applying a common rigid motion of all of E^n to each $p_i(0)$. Thus $p(t)=(g_t p_1(0), g_t p_0(0), \ldots, g_t p_v(0))$, g_t=rigid motion of E^n. The framework F is *rigid* if every flex of it is rigid. In the examples above, the triangle is rigid, and the square is not rigid, even in the plane.

The natural question at this stage is to determine what frameworks are rigid. The first nontrivial result in this direction was by Cauchy in 1813 [**6**].

THEOREM 1. *If a framework F is obtained from the natural edges and vertices of the boundary of a convex polyhedral surface in E^3, then F is rigid under the more restrictive condition that every flex holds each natural face rigid.*

Here a natural vertex, edge, or face is a 0, 1 or 2 dimensional intersection of a support plane with the convex surface. Note that if all the natural faces are triangles, then they are automatically "held" rigid. Actually Cauchy tried to prove a bit more than this theorem and his proof had a gap, but his proof was essentially correct and the gap was closed by Steinitz. (See Steinitz [**25**], Stoker [**26**], Lyusternik [**20**].)

In any case this very nice result reinforced the notion ("the rigidity conjecture") that any triangulated two-dimensional closed surface in E^3—convex or not—was rigid, where the framework consists of the vertices and the edges of the triangulation. Bricard [**5**] in 1897 "classified" all flexible octahedra and it was clear that none of them were embedded. Cauchy's techniques were applied and transformed to differential geometry and in the 1940s and 1950s A. D. Alexandrov and A. V. Pogorelov extended and applied Cauchy's techniques (see [**1**]). In 1973 Gluck [**13**] blatantly stated the rigidity conjecture. After showing that certain embedded surfaces were rigid [**7**], I found a counterexample. Thus a flexible triangulation of an embedded sphere exists [**8**].

So the question remains: If the rigidity conjecture is false, what is true?

III. The questions.

1. *The bellows conjecture.* Although I do not know of any spectacular consequences of the following conjecture, I find it one of the most intriguing in the subject. The problem is to construct a mathematical bellows. This is a closed, polyhedral, flexible surface which flexes so that the volume *changes*. In fact we can generalize this as follows: For any piecewise-linear (or smooth) map $f\colon M^2 \to E^3$ of a closed oriented piecewise-linear (or smooth) two-manifold into three-space there is a welldefined notion of the volume enclosed by f, whether f is an embedding or not. In case f is an embedding this number is $\pm$ the usual volume enclosed by $f(M)$ depending on the orientation chosen. In every case of a flexible framework coming from a "map" of a triangulated oriented two-manifold, which I know of, this generalized volume is constant during the flex (see [7] and [**8**]). Thus we have:

Conjecture. Every orientable, closed, polyhedral flexible surface (even with self-intersections) flexes with constant volume.

As far as I know this conjecture is essentially due to Dennis Sullivan.

2. *Classifying flexible surfaces.* Although it may seem ambitious, one way of determining when a polyhedral surface is rigid is simply to classify all the flexible surfaces. (Where the triangulation gives the framework.) Here things get a bit vague. My feeling is that a flexible surface is the "union" of two kinds of pieces. The first is in some sense "prime" and has a "volume" of zero. The second is "rigid" and so flexes with constant volume. So in particular the whole surface flexes with constant volume.

If the above is confusing, the reader is referred to the case of the suspension of a polygonal circle (see [7]). One takes a polygonal circle in E^3 and connects two additional points, the north and south poles, to each vertex on this equatorial circle. It is easy to see that this gives a triangulated surface, topologically a sphere. It is my opinion that the above "conjecture" holds for this surface. In fact if the distance between the north and south poles moves during a flex of such a surface, all the pieces turn out to be "prime" and the surface has zero volume. This theorem is the only way I know of showing all such embedded suspensions are rigid.

One natural starting point here might be to classify all flexible triangulations of a sphere with 8 vertices. It is conceivable one of these could be embedded. If not then Klaus Steffen's very pretty example with 9 vertices would be the best possible. (See [**24**] or [**9**].)

3. *Algorithms.* Short of complete success for the previous problem and also for other similar problems, it would be very nice to have precise and hopefully efficient algorithms for deciding various rigidity problems. In particular one is "given" a framework and one wishes to know:

A. Whether it is rigid or flexible.

B. Whether it is infinitesimally rigid. (Infinitesimal rigidity will be defined shortly.)

C. How many mutually noncongruent embeddings there are of this framework.

Recent work of Peter Kahn demonstrates that there are at least in principle algorithms for A and estimates for problems like C (see [**15**]); and it is well-known (see [**13**]) that problem B is equivalent to a large determinant being nonzero. This is encouraging and a good first step, but all these processes are much too unwieldy for practical use, even with a very efficient computer.

It would be pleasing to apply an algorithm for problem A to several examples that are known to be embedded but not known to be flexible. In this respect many of the surfaces of stellated regular solids seem to be very "sloppy" when made out of cardboard, and it would be interesting to know if they are really flexible. (e.g. Charles Schwartz has pointed out the stellated rhomboid dodecahedron, see Coxeter [**11**]).

Problem B would be of great interest to those concerned with the actual construction of structures, since generally (but not always) infinitesimal rigidity is what is desired. In this respect the "groupe de recherche" in Montreal has made some progress and can efficiently determine infinitesimal rigidity in many interesting cases.

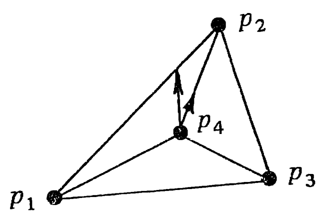
Infinitesimal rigidity is roughly, a linearized version of ordinary rigidity. It is defined as follows: First we define an *infinitesimal flex* of a framework F in E^3 as a sequence of vectors $\dot{p}=(\dot{p}_1, \dots, \dot{p}_v)$ such that $(p_i-p_j)\cdot(\dot{p}_i-\dot{p}_j)=0$ for all edges i, j in $\mathscr{E}$. $\dot{p}$ is called trivial if there are vectors r, t (in E^3) such that

$$\dot{p}_i = (r \times p_i)+t \quad \text{for all } i \ (1 \leqslant i \leqslant v).$$

r and t can be regarded as an infinitesimal rotation and translation. F is *infinitesimally rigid* if all infinitesimal flexes are trivial.

THEOREM 2. *If F is infinitesimally rigid, then it is rigid (see Gluck* [**13**]).

The converse is false as one can see by the example of a triangle with a point inside and all possible edges.



$$\dot{p}_1 = \dot{p}_2 = \dot{p}_3 = 0,$$

$\dot{p}_4$ is perpendicular to the plane of the triangle.

(It is interesting and useful to know that infinitesimal rigidity is preserved under projective collineations of E^3 (see [**12**]).)

A framework is *generically rigid* if when we consider all possible positions $p=(p_1, \dots, p_v)$ of the vertices (we regard p now as a point in E^{3v}) keeping the same pairs of vertices as edges, then an open dense set of these positions is rigid. We ask:

D. What frameworks are generically rigid? Namely is there an algorithm for determining this or some reasonable characterization?

It turns out that, for E^2, Leman **[19]** (see also Asimow and Roth **[4]**) has essentially solved this problem, but it remains open in E^3 and seems quite difficult. Also, the main theorem of Gluck **[13]** (see also Kuiper **[18]**) is that triangulated spheres are generically rigid. In fact, any triangulated convex surface with all the natural faces as triangles is infinitesimally rigid. (More generally if the triangulation has no vertex in the interior of a natural face, then it is infinitesimally rigid. See Alexandrov **[1]** or Asimow and Roth **[4]**.)

4. *Exotic rigidity*. Despite the attention payed to the infinitesimal rigidity of convex surfaces, there are many other situations that are also interesting. For instance, even for a convex polyhedral surface, if a triangulation has a vertex in the interior of a natural face, then the associated framework is not infinitesimally rigid; but it turns out to be rigid nevertheless. In fact we have the result (see Connelly **[10]**):

THEOREM 3. *The framework associated to any triangulation of any convex polyhedral surface is second order rigid and thus rigid.*

Note that vertices are allowed anywhere in the surface. Second order rigidity is defined as follows: A *second order flex* of a framework F is a first order flex $\dot{p}=(\dot{p}_1, \dots, \dot{p}_v)$ together with another sequence $\ddot{p}=(\ddot{p}_1, \dots, \ddot{p}_v)$ of vectors such that for each rod of F,

$$(\dot{p}_i-\dot{p}_j)\cdot(\dot{p}_i-\dot{p}_j)+(\ddot{p}_i-\ddot{p}_j)\cdot(p_i-p_j) = 0.$$

F is *second order* rigid if every nontrivial infinitesimal flex $\dot{p}$ fails to extend to a second order flex $\dot{p}, \ddot{p}$.

As indicated in the theorem, second order rigidity implies rigidity. It is not hard to see how to generalize the above definition to higher order rigidity and we have the question for $n \geqslant 3$

A. Does nth order rigidity imply rigidity?

Note that Efimov in **[12]** mentions a very similar problem.

If the answer is affirmative, this could possibly provide a method for doing 3A, since if one has an nth order flex, the problem of finding the $n+1$ order flex is entirely linear.

If we turn to the rigidity of surfaces other than the sphere we have the very basic conjecture:

B. Every triangulated closed surface in E^3 is generically rigid.

Due to methods of Gluck **[13]**, see also Asimow and Roth **[3]**, this amounts to finding one infinitesimally rigid framework corresponding to each abstract triangulation of the surface. At this point I can show that many triangulations of such surfaces have one infinitesimally rigid framework but I am not sure this includes all possible triangulations. This result is inspired by similar results in the smooth category by L. Nirenberg [21], and A. D. Alexandrov [2]. Also Stoker [26] has methods for showing that certain non convex surfaces are rigid.

Also related to the proof of Theorem 3 is the idea of "cabled" frameworks. Here, in addition to the rods there are certain pairs of vertices designated as cables, which can decrease but not increase in length during a flex. In Grünbaum's notes [14] there are many very interesting examples of such rigid cabled frameworks, and this idea is very helpful in Theorem 3. See also Whiteley [27] for similar results. It seems that many of these examples can be thought of as having springs for the cables and allowing the framework to move to a position of minimum energy. It would be very nice to have information as in 3A or 3B about such frameworks.

There is one other natural question concerning the rigidity of piecewise-linear surfaces, for $n \geqslant 3$.

C. Is every piecewise-linear closed n-dimensional manifold, which is embedded in E^{n+1}, rigid?

This is the higher dimensional analogue to the rigidity conjecture. The construction of my example works if the ambient embedding space is S^3, the standard round 3-sphere, and the cone over this example from the center gives a nontrivial flexible surface in E^4, but with boundary. Thus in E^4, counterexamples exist locally at least.

5. *Rigidity for smooth surfaces.* The basic question for smooth surfaces seems to be:

A. Are there flexible, closed, smooth surfaces in E^3?

Here the surface (and the flex) are to have at least two continuous derivatives (to be of class C^2), since methods of Kuiper [16], and [17], following Nash, give a flexible C^1 embedding of any C^1 surface in E^3. (All the above must preserve the metric of course.) In particular even the standard round two-sphere can be embedded strangely and flexed in a C^1 fashion in E^3.

Even for immersed C^2 surfaces, the answer to the above question is unknown. The methods of my counterexample do not apply in the C^2 case.

The notion of infinitesimal rigidity carries over into this category as well, and the theorems about the uniqueness of convex surfaces and their infinitesimal rigidity hold here as well. Spivak, vol. 5, [23] is a good reference for what is known here. Unfortunately, even the following very basic questions are unanswered.

B. If a smooth (of class C^2 say) closed surface is infinitesimally rigid, is it rigid?

It is not inconceivable that the methods of piecewise-linear surfaces may be applicable to problems in the smooth category via the appropriate sort of approximation, somewhat in the spirit of Pogorelov's work [22]. Thus we have the following questions:

C. If a piecewise-linear manifold approximates a closed smooth surface "closely enough", is the associated framework rigid?

Here it seems appropriate to say at least that the normal to each of the triangular faces must approximate the normals to the surface, as well as the points approximating the points on the surface. In fact it may be enough that the dihedral angles are not too sharp in some sense.

Once again the author would like to thank the Institut Des Hautes Etudes Scientifiques, and in particular N. H. Kuiper, for their kind support and encouragement.

References

1. A. D. Alexandrov, *Convex polyhedra,* Moskow, 1950 (Russian); German transl., *Konvexe Polyeder,* Akademie-Verlag, Berlin, 1958.

2. ——, *On a class of closed surfaces,* Recueil. Math. (Moscow) **4** (1938), 69—77.

3. L. Asimow and B. Roth, *Rigidity of graphs.* I, Trans. Amer. Math. Soc. **245** (1978), 279—289.

4. ——, *Rigidity of graphs.* II J. Math. Anal. Appl. (to appear).

5. R. Bricard, *Mémoire sur la théorie de l'octaèdre articulé,* J. Math. Pures Appl. **3** (1897), 113—148.

6. A. L. Cauchy, *Sur les polygones et les polyèdres,* Second Mémoire, I. Ecole Polytechnique **9** (1813), 87(=*Oeuvres complètes d'Augustin Cauchy,* 2nd sér., Tome 1, 1905, pp. 26—38).

7. R. Connelly, *The rigidity of suspensions,* J. Differential Geometry (to appear).

8. ——, *A counterexample to the rigidity conjecture for polyhedra,* Inst. Hautes Études Sci. Publ. Math. **47** (1978), 333—338.

9. ——, *A flexible sphere,* Math. Intellegencer; vol. 3. August 1978.

10. ——, *The rigidity of certain cabled frameworks and the second order rigidity of arbitrarily triangulated convex surfaces,* preprint, Cornell University, 1978.

11. H. M. S. Coxeter, *Regular polytopes,* Methuas.

12. N. V. Efimov, *Flächenverbiegung im Grossen,* Akad. Verlag, with a supplement by Rhembs and Grotemeyer, Akad. Verlag, Berlin, 1978, translated from *Qualitative problems of the theory of deformation of surfaces,* Uspehi Mat. Nauk. (N. S) **3** (1948), 47—158, and *Differential geometry and calculus of variations.* Transl. Amer. Math. Soc. Ser. 1, Vol. 6, 1962, pp. 274—423.

13. H. Gluck, *Almost all simply connected closed surfaces are rigid,* Geometric Topology, Lecture Notes in Math., vol. 438, Springer-Verlag, Berlin and New York, 1975, pp. 225—239.

14. B. Grünbaum, *Lectures on lost mathematics.* Univ. of Washington, preprint.

15. P. Kahn, *On the number of rigid P. L. frameworks: an application of real algebraic decision theory,* Cornell University, preprint, 1978.

16. N. H. Kuiper, *Isometric and short embeddings,* Indag. Math. **21** (1959), 11—25.

17. ——, *On C^1-isometric embeddings,* Indag. Math. **17** (1954), 545—556, 683—689.

18. ——, *Sphères polyédriques flexibles dans E^3, d'après Robert Connelly,* Séminaire Bourbaki, February 1978, pp. 514-01—514-21.

19. G. Laman, *On graphs and rigidity of plane skeletal structures,* J. Engrg. Math. **9** (1970), 331—340.

20. L. A. Lyusternik, *Convex figures and polyhedra,* Moscow, 1956 (Russian), English trans., T. J. Smith, Dover, New York, 1963, Heath Boston, Mass., 1966.

21. L. Nirenberg, *Rigidity of a class of closed surfaces,* Non-Linear Problems, Univ. of Wisconsin Press, Madison, Wis. 1963, pp. 177—193.

22. A. V. Pogorelov, *Extrinsic geometry of convex surfaces.* Transl. Math. Monographs, vol. 35, Amer. Math. Soc., Providence, R. I., 1973.

23. M. Spivak, *Differential geometry,* vol. 5, Publish or Perish, Inc. Boston, Mass. 1975.

24. Klaus Steffen, *A symmetric flexible Connelly sphere with only nine vertices*, letter, Inst. Hautes Études Sci. 1978 (unpublished).

25. E. Steinitz, *Polyeder und Raumeinteilungers*, Encykl. Math. Wiss **3** (1922), Geometrie, Part 3AB12 (1916), 1—139.

26. J. J. Stoker, *Geometrical problems concerning polyhedra in the large*, Comm. Pure Appl. Math. **21** (1968), 119—168.

27. W. Whiteley *Infinitesimally rigid polyhedra*, Champlain. Regional College, St. Lambert, Quebec, Canada, preprint.

CORNELL UNIVERSITY
ITHACA, NEW YORK 14853 U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Synthetic Geometry in Riemannian Manifolds

M. Gromov*

1. We measure *deviation* of a map $f: X \to X'$ from isometry by

$$\sup_{x,y \in X} \left|\log(\operatorname{dist}(f(x), f(y))/\operatorname{dist}(x, y))\right|.$$

For Riemannian manifolds V, V' we define $\operatorname{dev}(V, V')$ to be the "inf" of the deviations of all *diffeomorphisms* $V \to V'$. We treat "dev" as a metric in the set of isometry classes of Riemannian manifolds, though "dev" takes infinite values (say, when V and V' are not diffeomorphic).

The sectional curvature $K=K(V)$ is solely responsible for the local deviation of V from $\boldsymbol{R}^n$: when $|K| \leqslant \varkappa$ each point $v \in V$ has an arbitrary small ε-neighborhood U_ε such that its deviation from the ε-ball in $\boldsymbol{R}^n$ ($n = \dim V$) does not exceed $\varkappa\varepsilon^2$. (The converse is true up to a constant.)

A priori localization. Start with choosing a very small but fixed number ε. Neighborhoods U_ε can look very different from usual balls, no matter how small the curvature is.

Split tori. Take the product of n circles of lengths $l_1 \geqslant l_2 \geqslant \ldots \geqslant l_n > 0$. This is a flat manifold (i.e. $K \equiv 0$). Look at the ε-neighborhood U_ε of a point (ε-neighborhoods of different points are, obviously, isometric). Suppose that the ratio l_k/l_{k+1} is very large (about n^n) and ε is just in the middle between l_k and l_{k+1}. Such a U_ε looks approximately as the product of an $(n-k)$-dimensional torus (product of the "short" circles of lengths $l_{k+1}, \ldots, l_n$) by the k-dimensional ε-ball.

* Partially supported by the National Science Foundation of the United States.

When $l_1 \leqslant \varepsilon/n$ then U_ε coincides with the whole torus; thus tori (and flat manifolds in general) must be viewed as *local* geometric objects.

Nontrivial local geometry is always accompanied ($|K|$ is kept small) by nontrivial local topology: one defines *the injectivity radius* $\mathrm{rad}_v\ V$ as the maximal number r such that the ε-neighborhoods U_ε of $v \in V$ with $\varepsilon < r$ have smooth boundaries. These U_ε are automatically smooth topological balls and their deviation from Euclidean balls depends only on ε and $|K|$.

When $|K|$ and $(\mathrm{rad})^{-1}$ are kept bounded the sheer size of V (say volume or diameter) determines the overall geometric and topological complexity of V as follows.

Strong compactness. The set of all closed n-dimensional Riemannian manifolds satisfying (a) $|K| \leqslant \varkappa$, (b) $\mathrm{rad} \geqslant \varrho > 0$, (c) $\mathrm{Volume} \leqslant C$, ($\varkappa, \varrho, C$ are arbitrary numbers) is compact with respect to metric "dev". (A short proof can be found in **[4]**.)

This fact generalizes the Mahler compactness theorem for flat tori (see **[2]**) and sharpens Cheeger's theorem (see **[3]**) on the finiteness of the number of topological types under conditions (a), (b), and (c).

2. *Flat manifolds* are the simplest nontrivial local objects. Our understanding of their structure is based on the following classical theorems of Bieberbach and Hermite (see **[8]**).

(a) there are finitely many topologically distinct flat manifolds of given dimension;

(b) every compact flat manifold can be covered by a torus;

(c) every flat torus T stays close to a split torus, i.e. there is a split torus T' such that $\mathrm{dev}(T', T) \leqslant \mathrm{const}$ ($\leqslant n^n, n = \dim T$).

Further examples of manifolds V_ε with $\mathrm{rad} \leqslant \varepsilon$ and $|K| \leqslant \mathrm{const}$ can be obtained by multiplying a fixed V_0 by a flat manifold with diameter $\leqslant \varepsilon$, say, by the circle of length ε. This phenomenon can also be observed (Berger, see **[1]**) on general circle bundles: realize V_0 as a totally geodesic manifold of codimension 2 with prescribed normal bundle in W and take for V_ε the boundary of the ε-neighborhood of V.

Iterating this construction we arrive at an inductive definition of *nilmanifolds* of dimension n as circle bundles over $(n-1)$-dimensional nilmanifolds. Each nilmanifold carries a family of Riemannian structures with $|K| \leqslant \mathrm{const}$, $\mathrm{Diam} \to 0$.

Nilmanifolds are characterized homotopically as manifolds with nilpotent fundamental groups and contractible universal coverings.

The above V_ε do not "dev"-converge to V_0 but there is a coarser metric which provides such convergence. This is the Hausdorff metric defined in the set of isometry classes of all metric spaces as follows: $H(X, X')$ is the lower bound of the numbers δ satisfying the following property: there are isometrical imbeddings $X, X' \to Y$ into a metric space Y such that X is contained in the δ-neighborhood of X' and X' is contained in the δ-neighborhood of X.

Convergence $V_\varepsilon \to V_0$ in our examples is not surprising in view of the following fact.

Weak compactness. Let be given a sequence of n-dimensional closed Riemannian manifolds satisfying (a) $|K| \leqslant \varkappa$, (b) Diameter $\leqslant C$ ($\varkappa$, C are arbitrary). Then there is a subsequence which H-converges (i.e. relative to the H-metric) to a metric space X_0 (which is in general not a manifold).

Condition (a) can be relaxed to $K \geqslant -\varkappa$, $\varkappa \geqslant 0$, and in this more general form weak compactness follows directly from the Toponogov comparison theorem (see [1], [3]).

The following theorem discloses the geometry of the convergence $V_\varepsilon \to X_0$ in the simplest case when X_0 is the single point.

3. *Near flat manifolds.* A closed Riemannian manifold is called ε-near flat if its sectional curvature K and diameter satisfy $|K|\,(\mathrm{Diam})^2 \leqslant \varepsilon$.

If V is ε-near flat with $\varepsilon \leqslant \varepsilon_n$ $(\approx n^{-n^n})$, $n = \dim V$ then there exists a k-sheeted covering $\tilde{V} \to V$ with $k \leqslant n^{n^n}$ such that V is diffeomorphic to a nilmanifold and the induced metric in $\tilde{V}$ is "dev"-close to a locally homogeneous metric, i.e., there is a locally homogeneous V' with $\operatorname{dev}(V', \tilde{V}) \to 0$ as $\varepsilon \to 0$. (See [4].)

Probably V itself is diffeomorphic (and "dev"-close) to a locally homogeneous manifold. When the fundamental group $\pi_1(\tilde{V})$ is Abelian V is known to be diffeomorphic to a flat manifold.

Observe that the fundamental group of a near flat manifold contains a nilpotent subgroup of finite index. This property is probably shared by all *near elliptic manifolds,* i.e., when $K(\mathrm{Diam})^2 \geqslant -\varepsilon$, $0 \leqslant \varepsilon \leqslant \varepsilon_n$. It is known that $\operatorname{rank} H_1(V, R) \leqslant n = \dim V$, when V is near elliptic.

Examples. Products of near flat and elliptic (i.e. with nonnegative curvature) manifolds are near elliptic; circle bundles over elliptic manifolds are near elliptic.

4. *Micromanifolds.* Consider the set $\mathcal{M}_\varkappa$ of the isometry classes of the n-dimensional Riemannian manifolds with curvature bounded by $\varkappa$ i.e. with $|K| \leqslant \varkappa$. Manifolds from $\mathcal{M}_\varkappa$ display their most interesting features when rad $\to 0$. We are tempted to introduce new objects—*manifolds M with infinitely small injectivity radius.* We view every such M as an element from an ideal boundary $\partial \mathcal{M}_\varkappa$. Each M is represented by a sequence of $V_\varepsilon \in \mathcal{M}_\varkappa$, $\varepsilon \to 0$, converging relative to the H-metric to a metric space X_0. Our M is "fibered" over X_0; the fibers look like "infinitely small" near flat manifolds, but geometry and topology of the "fibers" can, in general, jump when $x \in X_0$ varies. In physical terms, M carries not only the macroscopic structure of X_0 but also additional microscopic information hidden in the "fibers".

When this description is made precise it yields the following "macroscopic" theorems I, II and III.

Theorem I. Estimates for Betti numbers. *Suppose that the sectional curvature of a closed n-dimensional Riemannian manifold V satisfies $|K| \leqslant 1$. Denote by $\sum_0^n b_i$ the sum of the Betti numbers of V (with any coefficients).*

(a) $\sum_0^n b_i \leqslant C_m$, $m=2+n+\mathrm{Diam}\, V$, $n=\dim V$, $C_m \approx m^{m^m}$.

(b) *If* V *is homeomorphic to a connected sum of manifolds of constant negative curvature then* $\sum_0^n b_i \leqslant C_n \cdot \mathrm{Volume}\ V$, $(C_n \approx n^{n^n})$. (See [5].) Probably, the word "constant" can be omitted.

Problem. What happens to (a) and (b) when condition $|K| \leqslant 1$ is replaced by $K \geqslant -1$?

THEOREM II. HYPERBOLIC MANIFOLDS *(Sectional curvature nonpositive). There are only finitely many topologically different manifolds satisfying:* (a) $0 \geqslant K \geqslant -\varkappa$, (b) $\mathrm{Diam} \leqslant C$ *($\varkappa$, C are arbitrary). When $\varkappa=0$ this is the Bieberbach finiteness theorem,* §2.

THEOREM III. NEAR HYPERBOLIC MANIFOLDS. *When* $\varepsilon \geqslant K \geqslant -1$, $\varepsilon \geqslant 0$, *and* ε *is small compared to diameter (say* $\varepsilon \leqslant m^{-m^m}$, $m=n+\mathrm{Diam}$) *then the fundamental group* $\pi_1(V)$ *is infinite. (I am certain that* V *is covered by* $\boldsymbol{R}^n$ *but the proof is not completed yet.)*

When $n \geqslant 3$ the restriction $K \geqslant -1$ can not be omitted (see [4]).

Locally homogeneous manifolds constitute a very rare set in $\mathscr{M}$ but the amount of the related mathematics is enormous (Lie groups etc.). The study of manifolds that are locally near homogeneous is conducted in the disguise of the Pinching Problem. In the heart of the problem we find again "rad"$\to 0$. (See [1,] [7] for further information.)

5. *Noncompact manifolds and their ends.* Let V be a complete noncompact connected manifold with bounded curvature, i.e. $|K| < \infty$. When $\mathrm{rad}_v \to 0$ as $v \to \infty$, for example, when the total volume is finite, V carries at infinity nontrivial "microstructure", but only in very few cases is this structure completely understood.

Pinched negative curvature. Let $-p\varkappa \leqslant K \leqslant -\varkappa$, and $p, \varkappa \geqslant 0$. If volume of V is finite then V can be exhausted by compact manifolds V_i such that each inclusion $V_i \to V$ is a homotopy equivalence and each component of the boundary ∂V_i (with the induced metric) is ε-near flat with $\varepsilon \to 0$ as $i \to \infty$, and its degree of nilpotency (i.e. the nilpotency degree of the fundamental group of the associated nilmanifold) does not exceed $\sqrt{p}$. In particular, when $p < 4$ each component is diffeomorphic to a flat manifold.

The complex hyperbolic space forms provide examples with $p=4$ and with non-Abelian nilpotent ends.

Incompressible ends. The next theorem provides us with many examples of noncompact manifolds supporting no complete metric with bounded curvature and finite volume.

THEOREM IV. *If* $|K| < \infty$, *Volume* $< \infty$ *and* V *is diffeomorphic to the interior of a compact manifold with boundary* B *then* B *has no metric of negative curvature. In* *particular when $n=3$ and B is a closed surface the Euler characteristic of B must be nonnegative.*

It is unclear whether $\boldsymbol{R}^{2n+1}$, $n>0$, supports complete metrics with bounded curvature and finite volume.

6. *Manifolds with boundary.* We must take into account the norm of the second quadratic form K^∂ of the boundary (Say, for the Euclidean ε-ball $\|K^\partial\|=\varepsilon^{-2}$.) We measure the interior size of V by $\mathrm{Int}=\sup_{v\in V}\operatorname{dist}(v, \partial V)$. Lower estimates for "Int" by $\|K^\partial\|$ were established in [6] for domains in $\boldsymbol{R}^n$. In general we have:

THEOREM V. *If* $\mathrm{Int}^2(|K|+\|K^\partial\|)\leqslant\varepsilon_n$ $(\leqslant n^{-n^n})$ *then V is diffeomorphic (and dev-close) to the product of a manifold V' without boundary and an interval* $[0, \delta]$, *or V can be doubly covered by such product.*

There are further relations between topology and the interior size of V. The following simple example points in the right direction:

THEOREM VI. *Let $V\subset\boldsymbol{R}^n$ be a compact domain with $\|K^\partial\|\leqslant 1$. If n is even then* $|\chi(V)|\leqslant C_n \operatorname{Vol}(V)$ $(C_n\approx n^n$, *χ is the Euler characteristic, "Vol" means volume*).

When V is the complement of the union of distinct unit balls, C_n is equal to the packing constant (see [9]).

An acknowledgement. The final version of this paper owes a lot to the critique by Professor N. Kuiper.

References

1. Yu. Burago and V. Zalgaller, *Convex sets in Riemannian spaces of non-negative curvature,* Russian Math. Surveys (Uspehi), **32 (195)**, (1977) 3—55.

2. J. Cassels, *An introduction to the geometry of numbers,* Springer-Verlag, Berlin, 1959.

3. J. Cheeger and D. Ebin, *Comparison theorems in Riemannian geometry,* North-Holland, Amsterdam, 1975.

4. M. Gromov, *Almost flat manifolds,* J. Differential Geometry **13** (1978).

5. *Volume and bounded cohomology* Inst. Hautes Études Sci. Publ. Math. (to appear).

6. V. Lagunov and A. Fet, *Extremal problems for a surface of given topological type,* Sibirsk. Mat. Z. **6** (1965), 1026—1036.

7. Min-Oo, *Krümmung und differenzierbare Struktur auf Komplex-projektiven Räumen,* Bonner Math. Schriften, Bonn, 1977.

8. M. Raghunathan, *Discrete subgroups of Lie groups,* Springer-Verlag, Berlin, 1972.

9. C. Rogers, *Packing and covering,* Cambridge Univ. Press, London, 1964.

STATE UNIVERSITY OF NEW YORK
STONY BROOK, NEW YORK 11794, U.S.A.

Труды Международного Конгресса Математиков
Хельсинки, 1978

Вещественные Алгебраические Поверхности

В. М. Харламов

Топологическое исследование вещественных плоских алгебраических кривых, а также алгебраических кривых и поверхностей в вещественном трехмерном пространстве, является классическим разделом топологии вещественных алгебраических многообразий и было начато Гарнаком и Клейном. Основные задачи этого исследования были включены Гильбертом в его знаменитый список проблем (16 проблема).

В последние годы в этой области достигнут значительный успех. Этот прогресс вызван работами В. И. Арнольда и В. А. Рохлина, указавшими новые пути исследования вещественных алгебраических многообразий. Он отражен частично в обзоре Д. А. Гудкова [1], где излагается состояние предмета на 1974, и частично в обзорах Г. Вилсона [2] и В. А. Рохлина [3], где излагается современное состояние топологии плоских кривых.[1] По топологии поверхностей, которая в настоящее время приблизилась по своему состоянию к топологии кривых, современного обзора пока нет. Цель настоящего доклада — восполнить в некоторой степени этот пробел.

В доклад не включены многообразия с особенностями, хотя о них имеется обширная нетривиальная информация.

В дальнейшем считается заданным n-мерное вещественное многообразие A, являющееся множеством неподвижных точек антиголоморфной инволюции conj неособого комплексного алгебраического многообразия CA. Главное внимание уделяется классическому случаю, когда CA — поверхность в CP^3,

[1] Библиографии этих обзоров содержат все рассматриваемые в докладе работы.

инвариантная относительно комплексного сопряжения, и conj индуцируется этим сопряжением.

1. Неравенство Гарнака и его обобщение. Классическая теорема Гарнака (1876) утверждает, что при $n=1$ число компонент кривой A не превосходит $g(CA)+1$, где g — род кривой. Эту же оценку дает неравенство

$$\dim H_*(A;\, \mathbf{Z}_2) \leqslant \dim H_*(\mathbf{C}A;\, \mathbf{Z}_2), \tag{1}$$

примененное к $n=1$. Неравенство (1) есть известное неравенство теории Смита, написанное для инволюции conj. К гиперповерхностям в проективном пространстве оно было впервые применено Р. Томом (1965). Неравенство (1) сильнее других известных оценок числа $\dim H_*(A;\, \mathbf{Z}_2)$ (принадлежащих Л. Бибербаху (1939), О. А. Олейник (1951), Дж. Милнору (1964), Р. Тому (1965)), что делает его наиболее вероятным претендентом на роль правильного обобщения неравенства Гарнака на произвольные многообразия.

Для поверхности A степени m в $\mathbf{R}P^3$ неравенство (1) дает оценку

$$\dim H_*(A;\, \mathbf{Z}_2) \leqslant m^3-4m^2+6m \tag{2}$$

(в первом нетривиальном случае $m=4$ эта оценка была сформулирована Гильбертом (1909) на основе исследования Роона (1886)). Дополнительную информацию дает другое известное неравенство теории Смита

$$\dim H_*(\mathbf{R}P^3, A;\, \mathbf{Z}_2) \leqslant \dim H_*(\mathbf{C}P^3, \mathbf{C}A;\, \mathbf{Z}_2),$$

написанное для той же инволюции. Из него следует, что в случае $\dim H_*(A;\, \mathbf{Z}_2)=m^3-4m^2+6m$ гомоморфизм включения $H_1(A;\, \mathbf{Z}_2)\to H_1(\mathbf{R}P^3;\, \mathbf{Z}_2)$ нетривиален.

2. Экстремальные сравнения. Первое такое сравнение было высказано Гудковым в качестве гипотезы для плоских кривых произвольной четной степени m, а в первом нетривиальном случае $m=6$ оно было и доказано Гудковым (1969). Оно утверждает, что в случае, когда число компонент кривой максимально по Гарнаку, выполняется сравнение $P-N\equiv(m/2)^2 \bmod 8$, где P — число четных (лежащих внутри четного числа других компонент) и N — число нечетных (прочих) компонент кривой. Важный шаг был сделан Арнольдом (1971). Он привлек к исследованию кривой двулистное разветвленное накрытие плоскости с ветвлением вдоль кривой и доказал сравнение[2] $P-N\equiv(m/2)^2 \bmod 4$. В полном объеме сравнение Гудкова было доказано Рохлиным (1972) с помощью того же накрытия. Затем Рохлин (1972) обобщил сравнение Гудкова на случай произвольного многообразия A; в обобщениях роль гарнаковского

[2] Это сравнение Арнольда является, в действительности, не экстремальным сравнением, а сравнением, выполняющимся при более широких гомологических условиях. Оно легко переносится на произвольные многообразия, но этот круг вопросов выходит за пределы чисто вещественной топологии.

условия играет равенство $\dim H_*(A; \mathbf{Z}_2) = \dim H_*(CA; \mathbf{Z}_2)$, что служит еще одним аргументом в пользу того, что неравенство (1) следует считать правильным обобщением неравенства Гарнака. Вслед за тем несколько родственных сравнений было найдено Д. А. Гудковым и А. Д. Крахновым (1973) и мной (1973, 74).

Сформулирую основные экстремальные сравнения для $n=2$: если

$$\dim H_*(A;\ \mathbf{Z}_2) = \dim H_*(CA;\ \mathbf{Z}_2),$$

то $\chi(A) \equiv \sigma(CA) \bmod 16$, где σ — сигнатура многообразия; если

$$\dim H_*(A;\ \mathbf{Z}_2) = \dim H_*(CA;\ \mathbf{Z}_2) - 2,$$

то $\chi(A) \equiv \sigma(CA) \pm 2 \bmod 16$;

если $$\dim H_*(A;\ \mathbf{Z}_2) = \dim H_*(CA;\ \mathbf{Z}_2) - 2,$$

CA — полное пересечение степени m в CP^q и гомоморфизм включения $H_1(A;\ \mathbf{Z}_2) \to H_1(\mathbf{R}P^q;\ \mathbf{Z}_2)$ — нулевой, то $m \equiv 2 \bmod 4$ и

$$\chi(A) \equiv \sigma(CA) + \begin{cases} 2 \bmod 16 & \text{при } m \equiv 2 \bmod 8, \\ -2 \bmod 16 & \text{при } m \equiv -2 \bmod 8. \end{cases}$$

Первые два сравнения были высказаны Гудковым в качестве гипотезы для поверхностей степени 4 в $\mathbf{R}P^3$, первое доказано Рохлиным (1972), а второе и третье — в моих работах (1972, 73, 74).

3. Оценка эйлеровой характеристики. Если n четно, то выполняется двойное неравенство

(3) $$-h^{n/2,n/2}(CA) + 1 \leqq \chi(A) - 1 \leqslant h^{n/2,n/2}(CA) - 1$$

где $h^{a,b}$ — размерность пространства биоднородных классов когомологий степени (a, b). Аналогичное неравенство есть в нечетной размерности; например, для гиперповерхности четной степени в проективном пространстве (четной размерности) оно совпадает с неравенством (3), примененным к разветвленному накрывающему пространства с ветвлением вдоль гиперповерхности. Эти неравенства, замеченные мной в 1974, явились итогом длительного пути. Для гиперповерхности в $\mathbf{R}P^q$ неравенство (3), примененное при четном q к разветвленному накрывающему и при нечетном q к самой гиперповерхности, дает при $q=2$ неравенства

(4) $$-\frac{3m}{4}\left(\frac{m}{2}-1\right) \leqslant P - N \leqslant \frac{3m}{4}\left(\frac{m}{2}-1\right) + 1,$$

высказанные В. Рэгсдейл в 1906 среди других более сильных гипотез и доказанные И. Г. Петровским в 1933, и при $q>2$ неравенства, доказанные И. Г. Петровским и О. А. Олейник в 1949. Формулировка и доказательство неравенств (4) на языке гомологий (без структуры Ходжа) разветвленного накрывающего были впервые найдены Арнольдом в 1971; это открытие Арнольда и привело меня к неравенствам типа (3).

Замечательно, что для поверхностей неравенство (3) было найдено А. Комессатти еще в 1932. Он формулировал результат в виде неравенств, которые можно представить следующим образом

$$-h^{1,1}(CA)+h_a \leqslant \chi(A)-2-\mathrm{tr} \leqslant h^{1,1}(CA)-h_a,$$

где h_a — размерность пространства $H_a \subset H^2(CA; C)$, порожденного алгебраическими классами, и tr — след инволюции $H_a \to H_a$, индуцированной инволюцией conj. Мне неизвестно, применил ли Комессатти эти неравенства к классическим случаям: плоским кривым, кривым в RP^3 и поверхностям в RP^3.

Как обнаружил Арнольд (1971), левое неравенство (4) остается справедливым при замене числа P числом гиперболических (содержащих внутри себя не менее двух компонент) четных компонент, а правое неравенство (4) остается справедливым при замене числа N числом гиперболических нечетных компонент.[3] Подобные усиления неравенств (3) для $n>2$ пока отсутствуют, а для $n=2$ известно только следующее усиление левого неравенства (см. мою работу (1976)): пусть k_0, k_-, k_+ — число ориентируемых компонент поверхности A соответственно с нулевой, отрицательной и положительной эйлеровой характеристикой и пусть k' — число неориентируемых компонент; тогда либо выполняются соотношения $k'=0, k_-=0, k_+=0, k_0=\frac{1}{2}h^{1,1}(CA)$ и $\frac{1}{2}h^{1,1}(CA) \leqslant 1+h^{2,0}(CA)$, либо выполняется неравенство

$$1-\chi(A)+2k_0+2k_+ \leqslant h^{1,1}(CA)-1.$$

В частности, если A — поверхность степени m в RP^3, то либо $m=2$ и A — гиперболоид, либо

$$1-\chi(A)+2k_0+2k_+ \leqslant \frac{2m^3-6m^2+7m-3}{3}. \tag{5}$$

4. Оценка числа ориентируемых компонент ненулевого рода. Арнольд (1971) нашел оценку числа непустых (содержащих внутри себя хотя бы одну компоненту) четных и числа непустых нечетных компонент плоской кривой четной степени.[4] Аналоги этих оценок для $n>2$ пока отсутствуют, а для $n=2$ известен только следующий их аналог (см. мою работу (1976)): всегда либо выполняются соотношения $k'=0, k_-=0, k_+=0, k_0=1+h^{2,0}(CA)$ и $1+h^{2,0}(CA) \leqslant \frac{1}{2}h^{1,1}(CA)$, либо выполняется неравенство

$$k_0+k_- \leqslant h^{2,0}(CA).$$

[3] Имеющиеся у Арнольда ограничения, как было указано Рохлиным (1974), не нужны.

[4] Арнольд формулировал эти оценки при некоторых ограничениях. Формулировка без этих ограничений была указана Рохлиным (1974); ошибка на единицу, допущенная им в одном из неравенств, была затем исправлена В. И. Звониловым (личное сообщение) и Г. Вилсоном (1978).

В частности, если A — поверхность степени m в $\boldsymbol{R}P^3$, то либо m четно, $k'=0$, $k_-=0$, $k_+=0$ и $k_0=(m^3-6m^2+11m)/6$, либо

$$(6) \qquad k_0+k_- \leqslant \frac{m^3-6m^2+11m-6}{6}.$$

5. Методы построения. Существуют классические методы построения вещественных алгебраических многообразий с заданными топологическими свойствами, но их очень мало и они относятся, главным образом, к кривым в $\boldsymbol{R}P^2$ и в $\boldsymbol{R}P^3$ и к поверхностям степени $\leqslant 4$ в $\boldsymbol{R}P^3$.

Построение вещественной поверхности степени 3 заданного топологического типа сводится к построению вещественной плоской кривой степени 4 с заданным расположением компонент. Эта редукция основывается на том, что любая плоская кривая степени 4 служит кривой ветвления (видимым контуром) у проекции поверхности степени 3 из точки поверхности на плоскость. Этим методом легко строятся вещественные поверхности степени 3 топологических типов

$$S_0 \amalg S(1),\ S(1),\ S(3),\ S(5),\ S(7),$$

где S_0 — сфера и $S(p)$ — сфера с p пленками. Любая (неособая) вещественная поверхность степени 3 принадлежит одному из этих типов (это следует как из (элементарной) изотопической классификации вещественных плоских кривых степени 4, так и из общих теорем пунктов 1—4).

Неособые поверхности степени 4 традиционно строили малым возмущением поверхностей с простой двойной точкой. Проектирование поверхности степени 4, имеющей простую двойную точку, из двойной точки на плоскость дает в качестве кривой ветвления (видимого контура) плоскую кривую степени 6. Обращение этого перехода от поверхности к кривой лежит в основе классического метода, сводящего построение вещественной поверхности заданного топологического типа с простой двойной точкой к построению вещественной плоской кривой степени 6, имеющей заданное расположение компонент и определяемой уравнением вида $a_3^2-a_2a_4=0$, где a_r — вещественный однородный многочлен степени r от 3 переменных. Этим методом Гильберт (1909), Роон (1912, 13) и Г. Е. Уткин (1969, 74) получили обширную информацию о вещественных поверхностях степени 4, однако топологическая классификация оставалась незавершенной как в части запретов, так и в части построений. Мне (1976) удалось завершить классификацию, достроив поверхности степени 4 всех типов, не запрещенных общими теоремами пунктов 1—4, что потребовало новый метод построения. Вариант этого метода, найденный мной в 1978, опирается на следующие результаты о комплексных К3-поверхностях: теорему Торелли [4] и эпиморфность отображения периодов (см., напр., [5]). Схема метода заключается в том, что построение поверхности A степени 4 в $\boldsymbol{R}P^3$ заданного топологического типа сводится к построению для некоторого стандартного кольца K (имитирующего $H^*(\boldsymbol{C}A; \boldsymbol{Z})$) тройки, состоящей из инволюции $K \to K$ (имитирующей conj*), биградуировки кольца $K \otimes C$ (имитирующей биградуировку Ходжа кольца $H^*(CA; C)$) и элемента $l \in K$ (имитирующего класс гиперплоского сечения) и имеющей заданные арифметические свойства. Топологическую классификацию можно провести этим методом вовсе без привлечения классического метода. Список топологических типов (неособых) поверхностей степени 4 имеет вид

(7) $S_p \amalg aS_0$ с $a \geqslant 0, \quad p \geqslant 0, \quad a+p \leqslant 9;$

(8) $S_p \amalg aS_0$ с $a \geqslant 0, \quad p > 0, \quad a+p = 10, \quad a-p \equiv 0, -2 \bmod 8;$

(9) $S_p \amalg aS_0$ с $a \geqslant 0, \quad p > 0, \quad a+p = 11, \quad a-p \equiv -1 \bmod 8;$

(10) $S_1 \amalg S_1; \quad \emptyset;$

где S_p — сфера с p ручками.

Недавно О. Я. Виро нашел метод построения поверхностей произвольной степени, являющийся развитием классических методов построения кривых. Он построил примеры поверхностей, показывающие, что оценки (2), (5) точны при всех m (точность оценки (2) есть еще один аргумент в пользу того, что неравенство (1) следует считать правильным обобщением неравенства Гарнака). По-видимому, метод Виро допускает дальнейшие обобщения.

Пока для $m \geqslant 5$ неизвестно, каково максимальное число компонент поверхности степени m в RP^3. При $m=5$ из неравенств (2), (3) следует, что это число $\leqslant 25$. Мне удалось доказать существование в RP^3 поверхности степени 5 с 21 компонентой (а именно, поверхности топологического типа $20S_0 \amalg S(9)$; доказательство использует результаты Е. Хорикавы [6] о деформациях комплексной структуры поверхностей степени 5 в CP^3.)

6. Изотопии. В этом пункте предполагается, что у рассматриваемых кривых и поверхностей нет вещественных, но, возможно, есть мнимые особенности. Следуя Рохлину [3], будем называть вещественными изотопиями кривой (поверхности) степени m в RP^2 (в RP^3) ее топологические изотопии в RP^2 (в RP^3) и жесткими изотопиями — ее изотопии в классе кривых (поверхностей) степени m в RP^2 (в RP^3). Жесткие изотопии допускают появление или исчезновение мнимых особенностей.

В случае кривых вещественная изотопическая классификация известна при $m \leqslant 6$ (при $m \leqslant 5$ она элементарна, при $m=6$ найдена Гудковым в 1969), жесткая классификация совпадает с вещественной изотопической при $m \leqslant 4$ (при $m \leqslant 3$ это очевидно, при $m=4$ доказано Клейном в 1876) и не совпадает с ней при $m \geqslant 5$ (Рохлин [3]).

В случае поверхностей вещественная изотопическая классификация известна при $m \leqslant 4$. В первом нетривиальном случае $m=3$ она является классической и совпадает с топологической классификацией — любая поверхность типа $S_0 \amalg S(1)$ изотопна объединению эллипсоида и проективной плоскости, а любая поверхность типа $S(p)$ изотопна проективной плоскости со стандартными ручками.

Поверхности степени 4 делятся на два вида: поверхность A первого вида имеет ненулевой, а второго вида — нулевой гомоморфизм включения $H_1(A; Z_2) \to H_1(RP^3; Z_2)$. Вещественная изотопическая классификация поверхностей степени 4 состоит в следующем: (а) поверхности первого вида реализуют все топологические типы (7), (8), (9), (10), кроме aS_0 $(a \geqslant 0)$; (б) поверхности второго вида реализуют все топологические типы (7), (10) и не реализуют типы (8), (9); (в) для поверхностей первого вида изотопическая классификация совпадает с топологической — поверхности типа $S_1 \amalg S_1$ изотопны объединению гиперболоидов, а поверхности типа $S_p \amalg aS_0$ $(a \geqslant 0, p \geqslant 1)$ — объединению a эллипсоидов, лежащих вне друг друга, и гиперболоида со стандартными ручками, содержащего внутри одной из компонент дополнения все эллипсоиды; (г) для поверхностей второго вида, негомеоморфных $S_0 \amalg S_0$, изотопическая классификация также совпадает с топологической — эти поверхности состоят, с точностью до изотопии, из лежащих в аффинной части пространства стандартных компонент, расположенных вне друг друга; (д) поверхности, гомеоморфные $S_0 \amalg S_0$, реализуют два изотопических типа — одна сфера внутри другой и сферы вне друг друга. Эта классификация была начата Гильбертом (1909), Рооном (1912, 13) и Уткиным (1969, 74). С помощью метода построения, основанного на теореме Торелли и эпипорфности отображения периодов (см. п. 5), мне (1978) удалось завершить эту классификацию, достроив поверхности степени 4 всех недостающих типов (здесь важно, что изотопический тип поверхности A степени 4 определяется арифметическими свойствами класса гиперплоского сечения и инволюции conj^* кольца $H^*(CA; Z)$; реализуемость всех указанных в классификации изотопических типов можно доказать этим методом вовсе без привлечения классического метода). Запреты утверждений (в), (г) доказываются элементарными средствами с помощью традиционного перехода к возмущениям поверхности с простой двойной точкой, в некоторых случаях дополнительно используется неравенство (6), примененное к возмущенным поверхностям. Запреты утверждения (б) следуют из общих теорем пунктов 1—4.

Жесткая классификация поверхностей при $m \leqslant 3$ совпадает с вещественной изотопической классификацией (при $m=3$ это доказано Л. Шлэфли (1864)). При $m=4$ жесткая и вещественная изотопические классификации уже не совпадают (метод построения, который я использовал для вещественной изотопической классификации, дает примеры вещественно изотопных поверхностей A, A_1 степени 4 таких, что CA, CA_1 неособы, A реализует в $H_2(CA)$ класс α, делящийся на 2, а A_1 реализует в $H_2(CA_1)$ класс α_1, не делящийся на 2; жесткая же изотопия, соединяющая A с A_1, дала бы изоморфизм $H_2(CA) \to H_2(CA_1)$, переводящий α в α_1).

По-видимому, в ближайшее время будет доступна жесткая классификация кривых степени 5, 6 и поверхностей степени 4.

Литература

1. Д. А. Гудков, *Топология вещественных проективных алгебраических многообразий*, Успехи Мат. Наук **29** (1974), № **4** (178), 3—79.

2. G. Wilson, *Hilbert's sixteenth problem*, Topology **17** (1978), 53—74.

3. В. А. Рохлин, *Комплексные топологические характеристики вещественных алгебраических кривых*, Успехи Мат. Наук **33** (1978), № **5** (203),

4. И. И. Пятецкий—Шапиро и И. Р. Шафаревич, *Теорема Торелли для алгебраических поверхностей типа* К3, Изв. Акад. Наук СССР Сер. Мат. **35** (1971), 530—572.

5. Вик. С. Куликов, *Эпиморфность отображения периодов для поверхностей типа* $K3$, Успехи Мат. Наук **32** (1977), № **4** (196), 257—258.

6. E. Horikawa, *On deformations of quintic surfaces*, Invent. Math. **31** (1975), 43—85.

Сыктывкарский государственный университет
Сыктывкар, Коми АССР, СССР

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Recent Results in Convexity

D. G. Larman

In this talk I shall try to give some of the important results in convexity and related topics which have been obtained since the last congress in 1974. The emphasis will naturally reflect my own interests and due to my own lack of background knowledge, some results will not be adequately treated.

1. Cross sections and volume. It is, perhaps, appropriate to begin with an old problem of H. Busemann and C. M. Petty [**8**] which was offered as an excercise to the audience by C. A. Rogers [**36**] at his talk in Vancouver 1974:

Consider two convex bodies K, K' in E^n which are both centrally symmetric about the origin. Suppose, for each $n-1$ dimensional linear subspace L, $V_{n-1}(K_n L) < V_{n-1}(K'_n L)$. Is it true that $V_n(K) < V_n(K')$?

It is important that the bodies K, K' are convex, see H. Busemann [**6**] and that both are centrally symmetric about **0**. In E^2 the problem is trivial since the condition then ensures that $K \subset K'$. Also, if K, the apparently smaller body, is an ellipsoid then the answer is affirmative. The problem still remains unresolved in E^3. However in E^n, $n \geqslant 12$, C. A. Rogers and myself [**25**] were able to give a negative answer in which K' was the unit n ball. The counter example consists of an n ball with small caps removed in a disjoint and a homogeneous manner. Is there still a counter example when $n-1$ dimensional subspaces are replaced by 2-dimensional subspaces? By modifying the methods of [**25**] it is possible to replace $n-1$ by $[n/2]+5$ dimensional subspaces.

I might also mention the solution to another problem of H. Busemann [**7**] by Larman–Mani–Rogers [**29**]. However, it turns out that the problem may have already been solved, within the context of lie groups, by Vinberg [**40**]. We do not claim, however, to understand exactly what was proved by Vinberg and, even less, his proofs. Anyhow, within convexity the problem was:

Characterise those convex bodies K such that for any two points $\boldsymbol{x}, \boldsymbol{y}$ in the interior of K there exists a projective transformation π, permissible for K, such that $\pi K = K$ (set-wise) and $\pi \boldsymbol{x} = \boldsymbol{y}$.

In E^2, K must be an ellipse or a triangle. In E^3, K must be an ellipsoid, tetrahedron or a cone on an elliptic base. However, the obvious conjecture, i.e. the convex hulls of disjoint ellipsoids lying in independent subspaces whose union spans E^n, fails in E^4. An example is the convex hull of two touching ellipses in E^4 which lie in two orthogonal 2-dimensional subspaces. The complete description of possible K is algebraic and rather technical.

2. Characterisations of the sphere and ellipsoid. Although it falls outside my time scale, the recent interest in this subject stems from the solution, by P. W. Aitchison, C. M. Petty and C. A. Rogers **[1]**, of the false centre problem: say that a convex body K in E^n has a false centre $\boldsymbol{x}$ in K if $\boldsymbol{x}$ is not the centre of K but every two dimensional section of K through $\boldsymbol{x}$ is centrally symmetric. Clearly every ellipsoid has this property and C. A. Rogers **[37]** had conjectured that any convex body with a false centre must be an ellipsoid; D. G. Larman **[23]** extended this result to where $\boldsymbol{x}$ was not necessarily in K. Recently G. R. Burton and P. Mani **[5]** have proved a more general result which was conjectured by P. Gruber **[14]**:

Suppose that a convex body K in E^n contains two distinct points $\boldsymbol{a}$ and $\boldsymbol{b}$ such that parallel 2-sections of K through $\boldsymbol{a}$ and $\boldsymbol{b}$ are directly homothetic. Then K is an ellipsoid.

It is not difficult, assuming that the convex body K has a false centre $\boldsymbol{x}$ to show that K is centrally symmetric, around $\boldsymbol{0}$ say. Then $\boldsymbol{x}$ and $-\boldsymbol{x}$ play the roles of $\boldsymbol{a}$ and $\boldsymbol{b}$ in the Gruber property.

It is well known that a convex body in E^n with all its $n-1$ dimensional sections centrally symmetric is an ellipsoid. G. R. Burton **[2]** has shown that it is enough to assume that all $n-1$ sections of sufficiently small diameter are centrally symmetric.

Suppose that K is a convex body in E^n and that $\boldsymbol{u}$ is a unit vector. Then the scalar product $\langle \cdot, \boldsymbol{u}\rangle$ has a maximum $m(\boldsymbol{u})$ on K. Suppose also that for each $\boldsymbol{u}$ in S^{n-1} there exists $\lambda(\boldsymbol{u})>0$ such that the section

$$K(p;\, n) = K_n\{\boldsymbol{x}\colon \langle \boldsymbol{x}, \boldsymbol{u}\rangle = p\}$$

is centrally symmetric whenever $m(\boldsymbol{u})-\lambda(\boldsymbol{u}) \leq p \leq m(\boldsymbol{u})$.

Then K is not necessarily an ellipsoid as can be seen from the example of a circular cylinder with hemispherical ends but, see G. R. Burton **[3]**, K must be the sum of an ellipsoid and a finite number of line segments.

Finally in this section I mention the proof by G. R. Burton **[4]** of a conjecture of Klee:

If the geodesics between any two points of a closed bounded convex surface ∂K in E^n are flat then K is a sphere.

3. Polytopes. Whilst work has declined in the study of convex polytopes in recent years, there have been two outstanding results produced since 1974, both of which involve mathematicians of at least 70 years of age!

The first is due to B. Jessen and A. Thorup [**21**]:

Say that two polytopes P and Q are equivalent in E^n if Q can be obtained from P by cutting, translating and glueing. A difficult problem has been to give necessary and sufficient conditions for P and Q to be equivalent.

The conditions are (roughly), take $k(x_1, \dots, x_{n-r})$ any odd real valued function of $x_1, \dots, x_{n-r}$. Let $u_1, \dots, u_{n-r}$ be $n-r$ orthogonal vectors in E^n and (with a finite number of exceptions) form the r face $P(u_1, \dots, u_{n-r})$ by first finding the $n-1$ face $P(u_1)$ of P in direction u_1, then the $n-2$ face of $P(u_1, u_2)$ of P within $P(u_1)$ and so on. Then form

$$\sum k(u_1, \dots, u_{n-r})V_r(P(u_1, \dots, u_{n-r})) \tag{1}$$

where V_r denotes r-dimensional volume. If (1) is equal to a similar sum for Q, for all choices of $u_1, \dots, u_{n-r}$, $r=0, 1, \dots, n$, then P is equivalent to Q and conversely.

This was proved previously by Hadwiger and Glur [**19**] for $n=2$ and by Hadwiger [**18**] for $n=3$.

The second result is due to Hadwiger [**17**]: consider the usual square lattice in E_n and a lattice polytope K (i.e. all of the vertices of K are lattice points). Let $G(K)$ denote the number of lattice points inside K.

In 1971 J. M. Wills [**41**] conjectured that

$$G(K) \leq \int_{E^n} e^{-\pi d^2(x,K)}\,dx$$

where $d(x, K)$ denotes the distance of x from K. It is easy to prove this conjecture for $n=2$ using the well known result $G(K)=(\text{Area of } K)+\frac{1}{2}$ (perimeter of K)$+1$. In 1974 T. Overhagen [**35**] proved it for $n=3$. However, recently Hadwiger has proved that it is false for a certain simplex in E^n, $n \geq 441$.

Finally in this section I shall mention a result of P. McMullen [**31**] on tiling space with zonotopes (i.e. finite sums of line segments). A zonotope Z in E^d, which is the sum of n line segments, is the orthogonal projection of a cube C in E^n. Of course, there is an associated zonotope $\bar{Z}$ formed by projecting C orthogonally into E^{n-d} (where $E^n=E^d+E^{n-d}$). McMullen proves that if Z tiles E^d by translation, with adjacent zonotopes meeting facet to facet, then $\bar{Z}$ tiles E^{n-d} in the same way.

4. Collision of convex bodies. Suppose two convex bodies K, K' in E^n have parts of their surface painted. Given that the two bodies collide one can ask for the probability that the two bodies collide paint to paint. The forerunner of this type of problem was the solution by P. McMullen [**30**] of a problem of W. J. Firey: What is the most probable encounter of two unit cubes in E^3?

It is intuitively clear that the only encounters which occur with a positive probability are edge to edge and vertex to face. In fact the proportion is $3\pi:8$ so that the edge to edge collision is more likely.

This work has been extended by W. J. Firey **[11]** to convex bodies K, K' and a formula derived for the probability of impact paint to paint in terms of the surface area functions of the two bodies. Firey insisted on painting whole faces but this restriction has been removed in the recent work of R. Schneider **[38]**. The reader is also referred to Schneider's excellent survey article **[39]**.

5. Helly's theorem. The famous theorem of E. Helly **[20]** asserts that if $\mathscr{F}$ is a family of compact convex sets in E^n with any subfamily of $n+1$ sets having a non-empty intersection then the family $\mathscr{F}$ has a non empty intersection.

As an extension of this result B. Grunbaum and T. Motzkin conjectured the following:

Suppose $\mathscr{F}$ is a family of sets, each the union of at most j disjoint compact convex sets in E^n such that the intersection of any k sets in $\mathscr{F}$, $k\leq j$, is also expressible as the union of at most j compact convex sets. Then, if any $j(n+1)$ sets in $\mathscr{F}$ has a nonempty intersection then the intersection of all the sets of $\mathscr{F}$ is nonempty.

They proved this conjecture for $j=2$ and I proved it for $j=3$. It was finally proved in general by H. C. Morris **[34]**.

6. Distance problems. Say that a set A in E^n realises the distance d if there are two points of A at a distance d apart. An old result of H. Hadwiger **[16]** is the following:

If $n+1$ closed sets cover E^n then at least one of the sets realises all distances.

Hadwiger did not believe that $n+1$ was the best possible number but even in E^2 it is still not known whether or not 4 sets will ensure the same result. However, for large n improvements have been made and, to explain these, let us define the concept of a configuration:

A configuration is any finite sequence of points $x_1, \ldots, x_M$, not necessarily distinct. The number of points in the configuration will be M, not necessarily the number of distinct points in the configuration. We say that for a given distance d, the number D is critical (for d and the configuration) if any sub-configuration of $D+1$ points realises d and D is the smallest number with this property.

THEOREM (D. G. LARMAN AND C. A. ROGERS **[26]**). *Suppose that there is, in E^n, a configuration of M points with critical distance 1 and critical number D. Then, if E^n is covered by less than M/D sets, there is a set of the covering within which all distances are realised.*

Using this result, D. G. Larman and C. A. Rogers **[26]** extended Hadwiger's result to $\frac{1}{6}n(n-1)$ sets (not necessarily closed). Recently P. Frankl **[12]** has used this result to prove:

THEOREM. *Given a natural number k there is an integer n_0 such that, provided $n \geqslant n_0$, any n^k sets which cover E^n contain a set which realises all distances.*

CONJECTURE (LARMAN [**24**]. *If E^n is covered by less than $\frac{1}{3}(\frac{4}{3})^{4n/3}$ sets then there is one set of the covering which realises all distances.*

Finally we mention the work of D. G. Larman, C. A. Rogers and J. J. Seidel [**27**] which almost resolves the old two distance problem: *Any set in E^n which realises only two distances has cardinality at most $\frac{1}{2}(n+1)(n+4)$.*

7. Tilings in the plane. Let me briefly mention the forthcoming book of B. Grunbaum and G. C. Shephard [**15**]. This study gives rise to some interesting problems. For example, can any tiling of the plane by bounded tiles be realised as a convex tiling? This, in some sense is a generalisation of the result of Steinitz which asserts that any 3-connected planar graph can be realised as the set of vertices and edges of a convex 3-polytope.

8. Dvoretsky's theorem. It has passed almost unnoticed by geometers that some outstanding work has been done recently, mainly by functional analysists, in shortening and extending the work of A. Dvoretsky [**9**]. In this work Dvoretsky shows, for $\varepsilon>0$ and a positive integer k the existence of a positive integer $n(k, \varepsilon)$ such that any centrally symmetric convex body in E^n, $n \geqslant n(k, \varepsilon)$, contains a central k section which is, to within ε, a k-ball. In [**28**], Larman and Mani show that K need not be centrally symmetric, which although interesting geometrically, is not so useful to a functional analyst. The audience is referred to T. Figiel [**10**], V. D. Milman [**32**] and Milman and Wolfson [**33**].

References

1. P. W. Aitchison, C. M. Petty and C. A. Rogers, *A convex body with a false centre is an ellipsoid*, Mathematika **18** (1971), 50—59.

2. G. R. Burton, *Some characterisations of the ellipsoid*, Israel J. Math. **28** (1977), 339—349.

3. *On the sum of a zonotope and an ellipsoid*, Comment. Math. Helv. **51** (1977), 369—387.

4. *Convex surfaces whose geodesics are planar*, Monatsch. Math. **84** (1977), 177—195.

5. G. R. Burton and P. Mani, *A characterisation of the ellipsoid in terms of concurrent sections*, Comment. Math. Helv. **53** (1978), 485—507.

6. H. Busemann, *Volumes and areas of cross-sections*, Amer. Math. Monthly **67** (1960), 248—250.

7. *Problem 3*, Colloquium on Convexity, Copenhagen, 1965, p. 309.

8. H. Busemann and C. M. Petty, *Problems on convex bodies*, Math. Scand. **4** (1956), 88—94.

9. A. Dvoretsky, *Some results on convex bodies and Banach spaces*, Proc. Symposium on linear spaces, Jerusalem, 1961.

10. T. Figiel, *Some remarks on Dvoretsky's theorem on almost spherical sections of convex bodies* Colloq. Math. **14** (1972), 243—252.

11. W. J. Firey, *Kinematic measures for sets of support figures*, Mathematika **21** (1974), 270—281.

12. P. Frankl, *Extremal problems and coverings of space*, Ars Combinatoria (to appear).

13. R. J. Gardner and P. McMullen, *Hammer's X-ray problem*, J. London Math. Soc. (to appear).

14. P. Gruber, *Problems*, Konvex Korper meeting, Oberwolfach, May 1974.

15. B. Grunbaum and G. C. Shephard, *Tilings and Patterns*, Freemans (to appear).

16. H. Hadwiger, *Ein Uberdeckungsatze für der Euklidischen Raum*, Portugal. Math. **4** (1944), 140—144.

17. *Gitterpunktanzahl im Simplex und Will'sche Vermutung*, Math. Ann. **239** (1979), 271—288.

18. *Translative Zerlegungsgleichheit der Polyeder des gewohnlichen Raumes*, J. Reine Angew. Math. **233** (1968), 200—212.

19. H. Hadwiger and P. Glur, *Zerlegungsgleichheit ebener Polygone*, Elem. Math. **6** (1951), 97—106.

20. E. Helly, *Ueber Mengens Konvexer Korper mit gemeinschaftlichen Punkten*, Jber Deutsch Math.-Verein **32** (1923), 175—176.

21. B. Jessen and A. Thorup, *The algebra of polytopes in affine spaces*, Math. Scand. (to appear).

22. T. H. Koornwinder, *A note on the absolute bound for systems of lines*, Nederl. Akad. Wetensch. Proc. Ser. A **79** (1976), 152—153.

23. D. G. Larman, *A note on the false centre problem*, Mathematika **21** (1974), 216—227.

24. *A note on the realisation of distances within sets in Euclidean space*, Comment. Math. Helv. **53** (1978), 529—535.

25. D. G. Larman and C. A. Rogers, *The existence of a centrally symmetric convex body with central sections that are unexpectedly small*, Mathematika **22** (1975), 164—175.

26. *The realisation of distances within sets in Euclidean space*, **19** (1972), 1—24.

27. D. G. Larman, C. A. Rogers and J. J. Seidel, *On two distance sets in Euclidean space*, Bull. London Math. Soc. **9** (1977), 261—267.

28. D. G. Larman and P. Mani, *Almost ellipsoidal sections and projections of convex bodies*, Proc. Cambridge Philos. Soc. **77** (1975), 529—546.

29. D. G. Larman, P. Mani and C. A. Rogers, *The transitive action of a group of projectivities on the interior of a convex body* (to appear).

30. P. McMullen, *A dice probability problem*, Mathematika **21** (1974), 193—198.

31. *Space tiling zonotopes*, Mathematika **22** (1975), 202—211.

32. V. D. Milman, *New proof of the theorem of A. Dvoretsky on intersections of convex bodies*, Funkt. Anal. Prilozen. **5** (1971), 28—37.

33. V. D. Milman and H. Wolfson, *Minkowski spaces with extremal distance from the Euclidean space*, Israel J. Math. **29** (1978), 113—131.

34. H. C. Morris, *Two pigeon hole principles and unions of convexly disjoint sets*, Ph. D. Thesis, California Institute of Technology, Pasadena, Calif., 1973.

35. T. Overhagen, *Zur Gitterpunktanzahl konvexer Korper*, Dissertation T. U. Berlin, 1974.

36. C. A. Rogers, *Probabilistic and combinatorial methods*, Proc. Internat. Congress of Mathematicians, (Vancouver, 1974), vol. 1, Canad. Math. Soc., 1976, pp. 497—500.

37. *Sections and projections of convex bodies*, Portugal. Math. **24** (1965), 99—103.

38. R. Schneider, *Kinematic measures for sets of colliding convex bodies*, Mathematika **25** (1978), 1—12.

39. *Boundary structure and curvature of convex bodies*, Proc. Geometry conference, Siegen, June, 1978 Bukhauser Verlag, Basel, 1979.

40. E. B. Vinberg, *Convex homogeneous domains*, Dokl. Akad. Nauk SSSR **141** (1961), 521—524.

41. J. M. Wills, *Zur Gitterpunktanzahl konvexer Mengen*, Elem. Math. **28**, (1973), 57—63.

University College London
London WC1E 6BT, England

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Isoperimetric Inequalities and Eigenvalues of the Laplacian

Robert Osserman

Given a domain D in euclidean space or on a Riemannian manifold, one naturally associates with it such fundamental geometric quantities as its volume, the measure of its boundary, various curvature functions and their extremes or integrals. One can also associate with the domain D the set of eigenfunctions and eigenvalues of the Laplace operator, subject to given boundary conditions. Recent work has revealed many connections between these two sets of quantities. We shall illustrate just a few of those connections, by concentrating in § I on isoperimetric inequalities between the geometric quantities, and in § II on the Dirichlet boundary value problem. We shall indicate some of the most striking ways in which isoperimetric inequalities are related to the distribution of eigenvalues. For a more detailed account of many of the subjects outlined here, we refer to the author's papers [**30**], [**31**].

I. Isoperimetric inequalities. There are many ways in which the classical isoperimetric inequality

$$L^2 \geqslant 4\pi A \tag{1}$$

has been refined and extended in recent years. Most important have been extensions to domains on surfaces and on higher-dimensional manifolds, or more generally to integral currents and varifolds. Before discussing those, let us note some refinements of (1) for curves in the plane. We shall use the following notation: D denotes a domain, C its boundary, A the area of D, and L the length of C. Further, let ϱ be the *inradius* of D, the maximum radius of open disks lying in D, and let R be the *circumradius* of D, the radius of the smallest disk including D.

THEOREM 1. *Let D be a simply-connected bounded plane domain. For $\varrho \leqslant r \leqslant R$, one has the following three (equivalent) inequalities:*

$$rL \geqslant A+\pi r^2, \tag{2}$$

$$L^2-4\pi A \geqslant (L-2\pi r)^2, \tag{3}$$

$$L^2-4\pi A \geqslant (A-\pi r^2)^2/r^2. \tag{4}$$

Inequality (3) is due (in the case of convex domains D) to Bonnesen. For a detailed account of Theorem 1 and related results, see a forthcoming paper of the author **[30]**.

The point of inequalities (3) and (4) is that they strengthen the basic isoperimetric inequality (1), and further imply (by choosing, in particular, $r=\varrho$), that equality can hold in (1) only when D is a circular disk.

For our purpose, the case $r=\varrho$ of (2), and its consequence

$$L/A > 1/\varrho \tag{5}$$

will be most important.

A more recent refinement of (1) is due to Sachs **[34]**. Let I denote the moment of inertia of the curve C with respect to its center of gravity. Then Sachs showed that

$$L^2-4\pi A \geqslant 4\pi^2|I/L-A/\pi|. \tag{6}$$

We turn next to domains on surfaces. A general observation is that if one considers domains D of fixed area A, then the length L of the boundary tends to increase as the Gauss curvature K decreases. Thus one has the following results.

THEOREM 2. *Let D be a simply-connected domain of area A bounded by a curve of length L. Let K denote Gauss curvature, and let $M=\sup_D K$. Then*

$$L^2 \geqslant 4\pi A-MA^2, \tag{7}$$

with equality if and only if $K\equiv M$ and D is a geodesic disk.

THEOREM 3. *Under the same hypotheses, one has*

$$L^2 \geqslant 4\pi A\left[1-\frac{1}{2\pi}\iint_D K^+\right] \tag{8}$$

where $f^+(p)$ denotes max $\{f(p), 0\}$.

For historical comments concerning these theorems see **[31]**. A more general inequality is the following. (See Ionin **[23]** and Burago **[7]**.)

THEOREM 4. *Let D be a domain with Euler characteristic χ. Let λ be any real number, and*

$$\omega_\lambda^+ = \iint_D (K-\lambda)^+.$$

Then

$$L^2 \geqslant 2A(2\pi\chi-\omega_\lambda^+)-\lambda A^2. \tag{9}$$

When $\chi=1$, (9) induces to (7) for $\lambda=M$ and to (8) for $\lambda=0$.

Burago and Zalgaller [8] proved the inequality

$$\varrho L \geqslant A+\left(\pi-\frac{1}{2}\omega_0^+\right)\varrho^2 \tag{10}$$

for simply-connected domains, where ϱ is the maximum distance to the boundary from points of D. In the plane case, ϱ is just the inradius, and (10) stands in exactly the same relation to (8) as does (2) (for the case $r=\varrho$) to (1). In particular, it implies that (5) continues to hold for simply-connected domains on arbitrary surfaces, provided $\iint_D K^+ < 2\pi$.

For surfaces that lie in a larger manifold or in euclidean space, one has other inequalities taking into account the mean curvature of the surface rather than its Gauss curvature. Many of them are based on the formula

$$A=-\iint_D (x-c)\cdot H\,dA+\frac{1}{2}\int_C (x-c)\cdot \nu\,ds \tag{11}$$

where D is a domain on an oriented surface in $\boldsymbol{R}^n$, A is the area of D, C its boundary, ν the exterior normal to D, H the mean curvature vector of the surface, and c is an arbitrary point in $\boldsymbol{R}^n$. If C consists of a single curve, then one can choose c to be the center of gravity of C, which we may take to be the origin, and estimate the right-hand term of (11) by

$$\int_0^L x\cdot\nu\,ds \leqslant L^2/2\pi,$$

so that (11) yields

$$L^2 \geqslant 4\pi\left(A+\iint_D x\cdot H\,dA\right) \tag{12}$$

where the origin is at the center of gravity of C. In particular, for minimal surfaces one has $H\equiv 0$, and the classical isoperimetric inequality (1) holds for minimal surfaces in $\boldsymbol{R}^n$ bounded by a single curve. Chakerian [12] showed that Sachs' refinement (6) is also true.

For minimal surfaces whose boundary consists of several curves, one conjectures that (1) continues to hold, but that is not known except for doubly-connected surfaces (Osserman and Schiffer [32] for $n=3$, Feinberg [17] for arbitrary n). However, one may still obtain useful inequalities from (11) in the general case. For example, if D lies in a ball of radius R, then choosing c to be the center of the ball, one obtains from (11),

$$L \geqslant \frac{2}{R}A-\iint_D |H|\,dA \tag{13}$$

where equality holds for a plane circular disk of radius R.

We consider next higher-dimensional domains. For a domain D in $\mathbf{R}^n$, the analog of (1) is

$$S^n \geqq n^n \omega_n V^{n-1} \tag{14}$$

where V is the volume of D, ω_n the volume of the unit ball in $\mathbf{R}^n$, and S the $(n-1)$-dimensional measure of the boundary of V. Equality holds only for a ball. It was conjectured by Wills [**38**] that inequality (2) (for the case $r=\varrho$) should generalize to

$$\varrho S \geqq V+(n-1)\omega_n \varrho^n \tag{15}$$

for convex domains D in $\mathbf{R}^n$, where ϱ is the inradius of D. Inequality (15) was proved by Diskant [**16**], and a stronger form of (15) was given by Osserman [**30**, Theorem 12] from which it follows that for $n \geqq 3$, equality can hold in (15) only for the sphere.[1]

For domains on Riemannian manifolds of dimension $n \geqq 3$, the analog of (7) is known only for the case of constant sectional curvature (Schmidt [**35**]) and for geodesic balls in manifolds of variable curvature (Aubin [**23**]). For many purposes it is sufficient to have weaker inequalities, which although not sharp, give useful upper bounds for the volume of a domain in terms of the area of its boundary. Such inequalities for general Riemannian manifolds and submanifolds have been proved by Schoen [**36**] and Hoffman and Spruck [**22**].

The two most interesting open questions seem to be:

1. Does the analog of (7) hold for domains on a Riemannian manifold of variable curvature?

2. Does (14) hold for domains on an arbitrary n-dimensional minimal variety in $\mathbf{R}^N$ for any $N>n$?

II. Eigenvalues of the Laplacian. Let D be a plane domain with smooth boundary C. The eigenvalue problem

$$\Delta u+\lambda u = 0 \quad \text{in } D, \tag{16}$$

$$u|_C = 0 \tag{17}$$

is known to have a complete system of eigenfunctions $u=\varphi_n$, with corresponding eigenvalues λ_n, where

$$0<\lambda_1<\lambda_2 \leqq \lambda_3 \leqq \dots; \quad \lambda_n \to \infty.$$

The basic question is: how are properties of the domain D reflected in the set of eigenvalues $\{\lambda_k\}$? Some of the early results are

THEOREM 4 (WEYL).

$$\lim_{n\to\infty} \frac{\lambda_n}{n} = \frac{4\pi}{A} \tag{18}$$

where A is the area of D.

---

[1] *Added in proof* (2/24/79). Professor Wills has called my attention to the fact that his conjecture (15) was also proved independently by J. Bokowski (Elemente der Math. **28** (1973), 43—44).

THEOREM 5 (FABER–KRAHN). *Among all domains D of fixed area A, λ_1 is minimum if and only if D is a circular disk.*

Since λ_1 for a disk of radius r is known to be $(j/r)^2$, where j is the first positive zero of the Bessel function J_0, one can state the Faber–Krahn result as

(19) $$\lambda_1 \geqslant \pi j^2/A.$$

In 1957, Peetre [33] showed that inequality (19) holds more generally for domains on a simply-connected surface with $K \leqslant 0$. The following year, Nehari [28] considered vibrations of an inhomogeneous membrane of density p, which leads to the equation

(20) $$\Delta u + \lambda p u = 0.$$

Nehari showed that among all domains of fixed total mass: $\iint_D p\,dA$, if $\log p$ is subharmonic, then the minimum of λ_1 for the eigenvalue problem (20), (17), is attained for a circular disk of constant density.

In fact, Peetre's and Nehari's results are exactly equivalent. If we think of p not as a density, but as defining a new conformal metric $ds^2 = p(x, y)(dx^2+dy^2)$ on D, then the total mass $\iint_D p\,dA$ corresponds to the area of D in the new metric, while the Gauss curvature is given by $K = -(\Delta \log p)/2p$, so that $\log p$ subharmonic is equivalent to $K \leqslant 0$. Finally, the Laplace–Beltrami operator Δ_s with respect to the new metric is given by $\Delta_s = \Delta/p$. Thus (20) is just $\Delta_s u + \lambda u = 0$, so that the value of λ_1 in Nehari's theorem is the same as λ_1 in Peetre's.

Recently, Bandle has expanded on the common link in Nehari and Peetre, and has obtained a whole series of results including the following [3, p. 205].

THEOREM 6. *Let D be a simply connected domain of area A, and let $M = \sup_D K$. In case $M > 0$, assume further that $A < 4\pi/M$. Let D_0 be the geodesic disk of area A on a surface of constant curvature M. Then $\lambda_1(D) \geqslant \lambda_1(D_0)$.*

This theorem was also later derived by Chavel and Feldman [13]. All the above results, starting with Faber and Krahn, make use of the various isoperimetric inequalities given above in § I. For example, Theorem 6 uses (7), while Peetre [33, p. 16] also derived from (8) the inequality

(21) $$\lambda_1 \geqslant \frac{\pi j^2}{A}\left(1 - \frac{1}{2\pi}\iint_D K^+\right)$$

generalizing (19).

One might think from (19) that λ_1 tends to zero as A tends to infinity. However one has the result:

THEOREM 7. *Let D be a simply-connected plane domain with inradius ϱ. Then*

(22) $$\frac{1}{4\varrho^2} \leqslant \lambda_1 \leqslant \frac{j^2}{\varrho^2}, \quad j \sim 2.4.$$

Thus, for simply-connected domains, λ_1 behaves like the square reciprocal of the inradius.

The right-hand side of (22) follows from an elementary comparison argument. The left-hand side is proved in Osserman [**29**], using the inequality (5). The first proof of an inequality of the form (22) is due to Hayman [**20**] who uses a different method and gets a much weaker constant on the left.

By using inequality (10) one can show that the left-hand inequality in (22) holds more generally for simply-connected domains D on a surface, provided $\iint_D K^+ \leqq 2\pi$. The same method can also be used to obtain results for domains of higher connectivity.

THEOREM 8. *Let D be a plane domain of inradius ϱ and connectivity $k \geqslant 2$. Then*

$$\lambda_1 \geqq 1/k^2\varrho^2. \tag{23}$$

Recently, Michael Taylor [**37**] showed that there exists some positive constant $c > 0$ such that

$$\lambda_1 \geqslant c/k\varrho^2 \tag{24}$$

for all plane domains D, where k is the connectivity of D.

A basic link between isoperimetric inequalities and bounds on λ_1 is provided by a result of Cheeger [**14**]:

THEOREM 9. *Let D be a domain on a Riemannian manifold. Set*

$$h = \inf_{D'} \frac{S'}{V'} \tag{25}$$

where D' is a relatively compact subdomain of D, V' the volume of D', and S' the surface area of $\partial D'$. Then

$$\lambda_1(D) \geqslant h^2/4. \tag{26}$$

Theorem 7 follows from Cheeger's Theorem and inequality (5), and Theorem 8 follows from a suitable extension of (5). Other inequalities from Part I yield further bounds on λ_1. For example, using (15), one can show [**30**, Theorem 13] that the left-hand side of (22) is valid for convex domains in $\mathbf{R}^n$ for all n. Similarly, using (13), one finds

THEOREM 10. *Let D be a domain on a minimal surface in $\mathbf{R}^n$. If D lies in a ball of radius R, then*

$$\lambda_1(D) \geqslant 1/R^2. \tag{27}$$

For m-dimensional minimal submanifolds of $\mathbf{R}^n$, one has an exact analog of (13) which yields

$$\lambda_1(D) \geqslant (m/2R)^2. \tag{28}$$

Another consequence of Cheeger's result is a theorem of McKean [**27**].

THEOREM 11. *Let D be a domain on an n-dimensional simply-connected Riemannian manifold whose sectional curvature is bounded above by $-\alpha^2$, $\alpha > 0$. Then*

$$\lambda_1(D) \geqslant [(n-1)\alpha/2]^2. \tag{29}$$

This uses the inequality (7) in the case $n=2$, and an isoperimetric inequality of Yau [**39**, p. 498] when $n>2$.

Incidentally, inequalities (26) and (29) are both optimal in the following sense. On a surface whose Gauss curvature satisfies $K \geqslant -\beta^2$, $\beta>0$, a geodesic disk D_r of radius r satisfies according to Cheng [**15**]

$$\lambda_1(D_r) \leqslant [\beta/2]^2 + [2\pi/r]^2. \tag{30}$$

(See also Buser [**11**] and Gage [**19**] for sharper versions of (30).) Letting $r\to\infty$ on a complete surface shows that (29) is a sharp bound. Since (29) was deduced from (26), it follows that the constant $\frac{1}{4}$ in (26) is best possible. (See also Buser [**9**].)

As a last application of isoperimetric inequalities, we mention that inequality (1) is used to prove the only known case in which the isospectral problem is solved. The problem is: if two plane domains have the same set of eigenvalues $\{\lambda_k\}$, are they necessarily congruent? (Can one hear the shape of a drum?) The answer so far is unknown except in the case when one of the domains is a circular disk. In that case one can use Weyl's Theorem (Theorem 4 above) to deduce that both domains have the same area, and then Faber–Krahn (Theorem 5 above) to conclude (from λ_1 alone) that the second domain must also be a disk. For further discussions of this subject, see Berger [**6**], Fisher [**18**], Kac [**25**], [**26**].

References

1. T. Aubin, *Problèmes isopérimétriques et espaces de Sobolev*, J. Differential Geometry **11** (1976), 573—598.

2. ——— *Inégalités concernant la première valeur propre non nulle du laplacian pour certaines variétés riemanniennes*, C. R. Acad. Sci. Paris Ser. A **281** (1975), 979—982.

3. C. Bandle, *Konstruktion isoperimetrischer Ungleichungen der mathematischen Physik aus solchen der Geometrie*, Comment. Math. Helv. **46** (1971), 182—213.

4. ——— *A geometrical isoperimetric inequality and applications to problems of mathematical physics*, Comment. Math. Helv. **49** (1974), 496—511.

5. ——— *On a differential inequality and its applications to geometry*, Math. Z. **147** (1976), 253—261.

6. M. Berger, *Geometry of the spectrum* I, Differential Geometry, Proc. Sympos. Pure Math., vol. 27, Part II, Amer. Math. Soc., Providence, R. I., 1975, pp. 129—152.

7. Yu. D. Burago, *Note on the isoperimetric inequality on two-dimensional surfaces*, Siberian Math. J. **14** (1973), 463—465.

8. Yu. D. Burago and V. A. Zalgaller, *Isoperimetric problems for regions on a surface having restricted width*, Two-Dimensional Manifolds of Bounded Curvature, Proc. Steklov Inst. Math. **76** (1965); English transl. Amer. Math. Soc., Providence, R. I. 1967. pp. 100—108.

9. P. Buser, *Ueber eine Ungleichung von Cheeger*, Math. Z. **158** (1978), 245—252.

10. ——— *Ueber den ersten Eigenwert des Laplace-Operators auf kompakten Mannigfaltigkeiten nichtpositiver Krümmung* (preprint, Bonn, 1977).

11. ——— *Beispiele für λ_1 auf kompakten Mannigfaltigkeiten* (preprint, Bonn 1977).

12. G. D. Chakerian, *The isoperimetric theorem for curves on minimal surfaces*, Proc. Amer. Math. Soc. **69** (1978), 312—313.

13. I. Chavel and E. A. Feldman, *Isoperimetric inequalities on curved surfaces* Advances in Math. (to appear).

14. J. Cheeger, *A lower bound for the smallest eigenvalue of the Laplacian,* Problems in Analysis, A Symposium in Honor of Salomon Bochner, Robert C. Gunning, Editor, Princeton Univ. Press, Princeton, N. J. 1970, pp. 195—199.

15. S.-Y. Cheng, *Eigenvalue comparison theorems and its geometric applications,* Math. Z. **143** (1975), 289—297.

16. V. I. Diskant, *A generalization of Bonnesen's inequalities,* Dokl. Akad. Nauk SSSR **213** (1973), Soviet Math. Dokl. **14** (1973), 1728—1731.

17. J. Feinberg, *The isoperimetric inequality for doubly-connected minimal surfaces in R^n,* J. Analyse Math. **32** (1977), 249—278.

18. M. E. Fisher, *On hearing the shape of a drum,* J. Combinatorial Theory **1** (1966), 105—125.

19. M. Gage, *Upper bounds for the first eigenvalue of the Laplace-Beltrami operator and an isoperimetric inequality for linked surfaces,* Ph. D. thesis, Stanford University, 1978.

20. W. K. Hayman, *Some bounds for principal frequency,* Applicable Anal. **7** (1978), 247—254.

21. D. Hoffman, *Lower bounds on the first eigenvalue of the Laplacian on Riemannian manifolds* (to appear).

22. D. Hoffman and J. Spruck, *Sobolev and isoperimetric inequalities for Riemannian submanifolds,* Comm. Pure Appl. Math. **27** (1974), 715—727; correction, **28** (1975), 765—766.

23. V. K. Ionin, *On isoperimetric and various other inequalities for a manifold of bounded curvature,* Siberian Math. J. **10** (1969), 233—243.

24. ———, *Isoperimetric inequalities for surfaces of negative curvature,* Siberian Math. J. **13** (1972), 650—653.

25. M. Kac, *Can one hear the shape of a drum?,* Amer. Math. Monthly **73** (1966), 1—23.

26. ———, *On applying mathematics: reflections and examples,* Quart. Appl. Math. **30** (1972), 17—29.

27. H. P. McKean, *An upper bound to the spectrum of Δ on a manifold of negative curvature,* J. Differential Geometry **4** (1970), 359—366.

28. Z. Nehari, *On the principal frequencies of a membrane,* Pacific J. Math. **8** (1958), 285—293.

29. R. Osserman, *A note on Hayman's Theorem on the bass note of a drum,* Comment. Math. Helv. **52** (1977), 545—555.

30. ———, *Bonnesen-style isoperimetric inequalities,* Amer. Math. Monthly **86** (1979), 1—29.

31. ———, *The isoperimetric inequality,* Bull. Amer. Math. Soc. **84** (1978), 1182—1238.

32. R. Osserman and M. Schiffer, *Doubly-connected minimal surfaces,* Arch. Rational Mech. Anal. **58** (1975), 285—307.

33. J. Peetre, *A generalization of Courant's nodal line theorem,* Math. Scand. **5** (1957), 15—20.

34. H. Sachs, *Ungleichungen für Umfang, Flächeninhalt und Trägheitsmoment konvexer Kurven,* Acta Math. Acad. Sci. Hungar. **11** (1960), 103—115.

35. E. Schmidt, *Beweis der isoperimetrischen Eigenschaft der Kugel im hyperbolischen und sphärischen Raum jeder Dimensionenzahl,* Math. Z. **49** (1943/44), 1—109.

36. R. Schoen, *Existence and regularity theorems for some geometric variational problems,* Ph. D. Thesis, Stanford University 1977.

37. M. Taylor, *Estimate on the fundamental frequency of a drum* (preprint).

38. J. M. Wills, *Zum Vërhaltnis von Volumen zu Oberfläche bei konvexen Körpern,* Arch. Math. **21** (1970). 557—560.

39. S.-T. Yau, *Isoperimetric constants and the first eigenvalue of a compact manifold,* Ann. Sci. École Norm. Sup. **8** (1975), 487—507.

STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Convex Sets and Convex Functions on Complete Manifolds

Katsuhiro Shiohama

The aim of this paper is to show how convex sets and functions give strong restrictions to the topology of a certain class of complete Riemannian manifolds without boundary. The idea of convexity plays an essential role for the proofs of "finiteness theorems", which give a priori estimates for the number of topological types of a certain class of compact Riemannian manifolds characterized by geometric quantities. *Small* convex sets such as strongly convex balls are used in the proofs of finiteness theorems. Weinstein's theorem **[18]**, which is the first attempt in this direction, states that given n and $\delta \in (0, 1)$, there are only finitely many homotopy types of $2n$-dimensional simply connected δ-pinched manifolds, and the number of homotopy types depends on δ and n. Then it has been developed by Cheeger **[3]**, Margulis **[13]** and Gromov.

On the other hand, *large* convex sets (such as a closed hemisphere of standard sphere and hyperplanes at infinity on the projective space with standard metrics) are useful in the proof of "uniqueness theorem". Well known examples of such theorems are the sphere and rigidity theorems investigated by Berger **[1]**, Klingenberg **[12]** and the author **[17]**. A sphere theorem states that a δ-pinched connected M is a topological sphere if $\delta = 1/4$ and the diameter $d(M)$ of M is greater than π. The rigidity theorem due to Berger states that an even dimensional $(1/4)$-pinched simply connected M is isometric to a compact symmetric space of rank 1 if its diameter $d(M) = \pi$. Under the assumptions of both the sphere and rigidity theorems, M admits *large* convex sets which I want to discuss in § 1. As is seen there *large* convex sets enable us to generalize the sphere and rigidity theorems, which are obtained by K. Grove and the author **[11]**, **[16]**.

In the next place I shall deal with convex functions on complete and noncompact Riemannian manifolds and generalize the theorems obtained by Gromoll–Meyer [10] and Greene–Wu [8]. Let $\gamma: [0, \infty) \to M$ be a ray emanating from a fixed point p. A Busemann function $F_\gamma: M \to R$ with respect to γ is defined by

$$F_\gamma(x) = \lim_{t \to \infty} [t - d(x, \gamma(t))], \quad x \in M.$$

If the sectional curvature K of M is nonnegative everywhere, then it follows from Toponogov's triangle comparison theorem that F_γ is convex. Obviously it is not constant on any open set of M. Moreover the function $F: M \to R$ defined to be $F(x) = \sup [F_\gamma(x); \gamma(0) = p]$ is convex and exhaustion, where the sup is taken over all rays emanating from p. Gromoll–Meyer proved that a noncompact M is diffeomorphic to R^n if $K > 0$. Then Cheeger and Gromoll proved that there exists on a noncompact M with $K \geqslant 0$ a compact totally geodesic submanifold S without boundary which is totally convex. Furthermore M is diffeomorphic to the total space of the normal bundle $\nu(S)$ over S. Recently Greene and Wu showed in [7] that the above F can be replaced by a convex exhaustion function whose second difference quotient along every geodesic is bounded away from 0 on every compact set provided $K > 0$. And in [8] they approximated a convex function with positive second difference quotient along every geodesic by a smooth convex function with positive second derivative along every geodesic. Therefore if M admits a convex exhaustion function with positive second difference quotient along every geodesic, then M is diffeomorphic to R^n. These results are generalized in § 2, which I worked with Robert E. Greene. (See [6].)

1. Convex sets. Throughout this section let M be a compact and connected Riemannian manifold of dimension $n \geqslant 2$. Assume that the sectional curvature K and the diameter $d(M)$ of M satisfy $K \geqslant \delta > 0$ and $d(M) \geqslant \pi/2\sqrt{\delta}$. Let $p, \bar{p} \in M$ be such that $d(p, \bar{p}) = d(M)$, where d is the distance function. A *large* convex set is defined to be $A_p = \{x \in M;\ d(p, x) \geqslant \pi/2\sqrt{\delta}\}$. A_p is clearly a nonempty closed convex set, and has a nonempty boundary if the diameter assumption is inequality. More generally if a closed convex set in a complete manifold of positive sectional curvature has a nonempty boundary, then the soul of it is a single point and hence it is homeomorphic to a closed disc (see [5]). We next define $B = \bigcap \{A_q;\ q \in A_p\}$. Then B is a nonempty closed convex set. It turns out that if M is not simply connected then both A_p and $A_{\bar{p}}$ have no boundary (see [15]). In any case we can choose for a closed convex set C a neighborhood of C which is either an embedded open n-disc with smooth boundary (if C has a nonempty boundary) or else a normal disc bundle over C (if C has no boundary). In order to see what happens in between A_p and B, we define a function $f: M \to R$ by

$$f(x) = \begin{cases} d(\bar{p}, x) - d(p, x) & \text{if } d(M) > \pi/2\sqrt{\delta} \\ d(A_p, x) - d(B, x) & \text{if } d(M) = \pi/2\sqrt{\delta}. \end{cases}$$

f is continuous on M and smooth outside a closed set Ω of measure zero. Choose neighborhoods U and V (of p and $\bar{p}$ if $d(M) > \pi/2\sqrt{\delta}$ and of B and A_p if $d(M) = \pi/2\sqrt{\delta}$) which have the properties stated above. By a smoothing convolution process we can approximate f by a family $\{f_\varrho : M \to R;\ \varrho \in (0, \varrho_0)\}$ of smooth functions so that if ϱ_1 is taken to be sufficiently small then $\nabla f_\varrho \neq 0$ in $M - U \cup V$ and ∇f_ϱ is transversal to $\partial U \cup \partial V$ for any $\varrho \in (0, \varrho_1)$. Thus we can prove the following

THEOREM 1. *Let M be a connected and compact Riemannian manifold and let $K \geq \delta > 0$.*

(a) (see **[11]**). *If $d(M) > \pi/2\sqrt{\delta}$, then M is a topological sphere.*

(b) (see **[17]**). *Assume that $d(M) = \pi/2\sqrt{\delta}$. Then we have*

b-1. *If A_p has nonempty boundary or* $\dim A_p = 0$ *and if B has nonempty boundary or* $\dim B = 0$, *then M is a topological sphere.*

b-2. *If one of the A_p and B has no boundary and its dimension is greater than* 0 *then M has the same cohomology structure as that of a symmetric space of compact type of rank* 1.

b-3. *If both A_p and B have no boundary and if* $\dim A_p$ *and* $\dim B$ *are positive, then M is exhibited as a union of two normal disk bundles over A_p and B joined along their common boundary.*

REMARKS (1). Note that the condition in (a) is the best possible one for a compact manifold of positive sectional curvature to be a topological sphere. (2). In case b-2, M admits a complete metric and a point on the convex set with nonempty boundary with respect to which the tangent cut locus at that point is a sphere. Then the result follows from **[14]** or **[2]**. (3). If M is not simply connected, then b-3 occurs.

In the case where M is not simply connected, we have the

THEOREM 2. *Assume that M is not simply connected and $K \geq \delta > 0$ and $d(M) = \pi/2\sqrt{\delta}$. If* $\dim A_p = 2$ *and if there exists a pair of points $x \in B$ and $y \in A_p$ such that there are at most finitely many minimizing geodesics joining x to y, then M is of constant curvature δ and its fundamental group has a fully reducible representation.*

2. Convex functions. Throughout this section let M be a complete, noncompact Riemannian manifold without boundary. We want to investigate the topology of M which admits a convex function. A function $\varphi : M \to R$ is said to be convex if for every normal geodesic $\gamma : R \to M$ and every $t_1, t_2 \in R$, and $\lambda \in [0, 1]$,

$$\varphi \circ \gamma((1-\lambda)t_1 + \lambda t_2) \leq (1-\lambda)\varphi \circ \gamma(t_1) + \lambda \varphi \circ \gamma(t_2).$$

If the above inequality is strict for $\lambda \in (0, 1)$, then φ is called to be strictly convex. It does not necessarily follow that a strictly convex function has positive second difference quotient along a geodesic. In order to investigate topology of M which admits a convex function, it is natural to restrict ourselves to consider the case where it is not locally constant. Let $\varphi: M \to R$ be a convex function which is not locally constant. A convex function is continuous and Lipschitz continuous on every compact set. Let $M_a^a(\varphi) = \varphi^{-1}(a)$ and $M_a(\varphi) = \{x \in M;\ \varphi(x) \leq a\}$. Then $M_a(\varphi)$ is totally convex. Since a monotone increasing bounded convex function defined on $[0, \infty)$ is constant, we have the

LEMMA 1. *If $M_a^a(\varphi)$ is compact for some $a \in \varphi(M)$, then so is $M_b^b(\varphi)$ for all $b \geq a$.*

A perpendicular geodesic from a point to a closed convex set is uniquely determined if the point is close to the set. This fact together with the Lipschitz continuity of a convex function on a compact set will imply the following

THEOREM 3. *If $M_a^a(\varphi)$ is not connected for some $a \in \varphi(M)$, then we have the following statements.*

(1) *φ attains its minimum, say, m_0.*

(2) *$M_{m_0}(\varphi)$ is a complete totally geodesic hypersurface without boundary and has a trivial normal bundle.*

(3) *$M_b^b(\varphi)$ has exactly two components for all $b > m_0$.*

As a direct consequence of the above theorem, we have

COROLLARY TO THEOREM 3. *Let $\psi: M \to R$ be strictly convex. Then every level set is connected.*

The existence of a strictly convex function gives a strong restriction to the Riemannian metric. For instance there is no compact totally geodesic submanifold without boundary if M admits a strictly convex function. Moreover we can prove the following

THEOREM 4. *Let $\psi: M \to R$ be strictly convex. Then the exponential map $\exp_p: M_p \to M$ at every point p is proper.*

As a consequence of Theorem 4, we have

COROLLARY TO THEOREM 4. *If there is a compact level set for a strictly convex function, then every level set is compact.*

In the case where the sectional curvature of M is nonnegative mor egenerally if the Ricci curvature of M is nonnegative, then M has at most two ends. We can also discuss the ends of M which admits a convex function.

THEOREM 5. *If M admits a strictly convex function, then M has at most two ends.*

Finally we shall state a structure theorem as follows.

THEOREM 6. *Let $\psi: M \to R$ be a strictly convex function. If there is a compact level set of ψ, then M is diffeomorphic to either R^n (if ψ attains its minimum) or else a cylinder $N \times R$, where N is a compact hypersurface homeomorphic to a level set of ψ.*

In the case where M admits a convex exhaustion function, we have

THEOREM 7. *Let $\varphi: M \to R$ be a convex exhaustion function. Assume that there is a level set which is not connected. Then M is diffeomorphic to a cylinder $M_{m_0}(\varphi) \times R$.*

References

1. M. Berger, *Les variétés riemanniennes (1/4)-pincées*, Ann. Scuola Norm. Sup. Pisa III **14** (1960), 161—170.

2. A. Besse, *Manifolds all of whose geodesics are closed*, Springer-Verlag, Berlin, 1978.

3. J. Cheeger, *Finiteness theorems for Riemannian manifolds*, Amer. J. Math. **92** (1970), 61—74.

4. J. Cheeger and D. Ebin, *Comparison theorems in Riemannian geometry*, North-Holland Mathematical Library, Amsterdam, 1975.

5. J. Cheeger and D. Gromoll, *On the structure of complete manifolds of nonnegative curvature*, Ann. of Math. **96** (1972), 413—443.

6. R. E. Greene and K. Shiohama, *Convex functions on complete noncompact manifolds* (preprint).

7. R. E. Greene and H. Wu, *Integrals of subharmonic functions on manifolds of nonnegative curvature*, Invent. Math. **24** (1974), 265—298.

8. ——, *C^∞ convex functions and manifolds of positive curvature*, Acta Math. **137** (1976), 209—245.

9. D. Gromoll and W. Klingenberg and W. Meyer, *Riemannsche Geometrie im Grossen*, Lecture Notes in Math. vol. 55, Springer-Verlag, Berling and New York, 1968.

10. D. Gromoll and W. Meyer, *On complete open manifolds of positive curvature*, Ann. of Math. **90** (1969), 75—90.

11. K. Grove and K. Shiohama, *A generalized sphere theorem*, Ann. of. Math. **106** (1977), 201—211.

12. W. Klingenberg, *Uber Riemannsche Mannigfaltigkeiten mit positiver Krummung*, Comment. Math. Helv. **35** (1961), 47—54.

13. G. Margulis, Proceedings of the International Congress of Mathematicians, (Vancouver. Canada, 1974), part 2, Canad. Math. Soc., 1976 pp. 21—34.

14. H. Nakagawa and K. Shiohama, *Geodesic and curvature structures characterizing projective spaces*, Diff. Geometry, in honor of K. Yano, Kinokuniya, Tokyo, 1972.

15. T. Sakai and K. Shiohama, *On the structure of positively curved manifolds with certain diameter*, Math. Z. **127** (1972), 75—82.

16. K. Shiohama, *The diameter of δ-pinched manifolds*, J. Differential Geometry, **5** (1971), 61—74.

17. ——, *Topology of positively curved manifolds with certain diameter*, Proc. of the Japan—U.S.A. Seminar on Minimal Submanifolds and Geodesics Kaigai Publ., Tokyo, 1978, pp. 217—228.

18. A. Weinstein, *On the homotopy type of positively pinched manifolds*, Arch. Math. **18** (1967), 523—524.

UNIVERSITY OF TSUKUBA
300—31, IBARAKI, JAPAN

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Characterization of Topological Manifolds of Dimension $n \geq 5$

J. W. Cannon*

That rich, unkempt world of wild and tame topology, born in the minds of Antoine and Alexander, recalled from obscurity by Fox, Artin, and Moise, and brought to full bloom by Bing, has spawned a conjecture on the nature of the topological manifold having as one of its minor corollaries the famous double suspension theorem for homology spheres. F. Quinn in the Saturday morning topology seminar of this congress expressed confidence that he has the right conceptual and technical framework to complete the final step in its proof. Whatever the result after Quinn has had opportunity to verify his intuitions, the result is at the very least almost true; and we wish to discuss it. As is often the case, much of the visualization and example which gave the conjecture birth will surely disappear in the powerful application of engulfing, local surgery, etc., which should constitute its final proof. And so, for those of us who have always savored the interplay among point set topology, taming theory, decomposition space theory, and other visual aspects of geometric topology, we record here the milieu in which the conjecture became reasonable and the pressures leading to its formulation.

But first we summarize the conjecture itself and its most recent history. In the early spring of 1977 we conjectured,

Characterization Conjecture. A generalized n-manifold having the disjoint disk property, $n \geq 5$, is a topological n-manifold.

A generalized n-manifold M is an ENR satisfying $H_*(M, M-x; Z)=$

* The author's work, reported in this paper, was supported in part by the National Science Foundation of the United States.

$H_*(E^n, E^n-0; Z)$ for each $x \in M$. The space M satisfies the disjoint 2-disk property if maps $f, g\colon B^2 \to M$ can be approximated by maps $f', g'\colon B^2 \to M$ having disjoint images.

We proved the conjecture for generalized manifolds having nonmanifold set of trivial dimension $k \leq (n-2)/2$ in the spring of 1977, and now, less than two years later, its proof appears on the verge of completion in two steps:

Resolution Conjecture (to be proved by? Quinn?). A generalized n-manifold of dimension $n \geq 5$ is a cell-like quotient of a topological n-manifold.

A cell-like subset of an ENR is a compactum contractible in each neighborhood of itself. A quotient map $f\colon M \to N$ of ENR's is cell-like if it is a closed map and each point preimage is cell-like.

Quotient Conjecture (proved by R. D. Edwards, late spring, 1977). A finite-dimensional cell-like quotient of an n-manifold, $n \geq 5$, is a manifold if and only if it has the disjoint disk property.

In addition to our own earlier weak versions of the Resolution and Quotient Theorems, early spring, 1977, Bryant–Hollingsworth and Bryant–Lacher had proved early versions of the Resolution Theorem.

As recently as four years ago no one dreamed that a useful characterization of topological manifolds was possible; all proposals ran afoul of the delicate, fiendishly manifold-like nonmanifolds of R. H. Bing and his school. That the notions of the generalized manifold and disjoint disk property were precisely appropriate for a characterization conjecture appeared only slowly from considerations of the taming and decomposition space theory pioneered by R. H. Bing. We give here an abbreviated exposition of Bing's work relevant to the characterization conjecture.

Bing, examining E. E. Moise's work on the triangulation theorem and Hauptvermutung for 3-manifolds in the early 1950s, was led to a profound study of the embeddings of polyhedra and compacta in the 3-dimensional sphere S^3. Bing set himself the problem of understanding the phenomenon called wildness. While it was clear that a simple closed curve can be knotted in S^3, it is not at all obvious that the Cantor set or 2-sphere can be knotted in S^3. Nevertheless, such knotting, necessarily infinite in nature, does occur and was discovered in the 1920's by M. L. Antoine and J. W. Alexander. An infinitely knotted set is called wild; other more standardly embedded sets are called tame. One of Bing's many beautiful discoveries was that the wildness of a 2-sphere or Cantor set in S^3 can be traced to a simple homotopy theoretic failing in dimension one: the complement of a wild set in S^3 is not 1-ULC; that is, there exist arbitrarily small simple closed curves in the complement of the wild set that are not contractible in small subsets of the complement. Extensions of Bing's results came to be known as taming theory.

Decomposition space theory as developed by Bing obtained its early impetus from the following remarkable theorem of R. L. Moore, Bing's teacher: if $f\colon S^2 \to X$ is a surjection from the 2-sphere S^2 onto a Hausdorff space X such that, for each $x \in X$, $S^2 - f^{-1}(x)$ is nonempty and connected, then X is also a 2-sphere.

Bing studied the extent to which Moore's theorem extends to closed surjections $g: S^3 \to Y$. G. T. Whyburn, another Moore student, had already suggested an appropriate condition on point inverses $g^{-1}(y)$, $y \in Y$: $S^3 - g^{-1}(y)$ was to be homeomorphic with $S^3 -$(point) and such a set was to be called pointlike or cellular. (The notion of cell-like set occurring in the Resolution and Quotient Conjectures is a generalization of Whyburn's notion of cellular set slightly more appropriate than cellularity in general.) But even among cellular quotients of S^3 Bing found nonmanifolds, such as his dogbone space—nonmanifolds because they failed to have a certain appropriate 3-dimensional variant of the disjoint disk property.

For the purposes of this paper we shall occasionally call the nonmanifold cell-like quotients of a manifold M wild spaces and the manifold quotients tame. The corresponding decompositions of M into point preimages of the quotient map are called nonshrinkable (wild) or shrinkable (tame) decompositions, respectively, for important technical reasons. Both shrinkable and nonshrinkable decompositions are important for the theory, a nonshrinkable decomposition always yielding a wild space, but an interesting shrinkable decomposition often yielding an unusual wild subspace or wild embedding.

From the middle of the 1950s until the early 1970s mathematicians of the Bing school developed the two theories in parallel. As years passed it became more and more apparent that, especially in high dimensions, wild subspaces and wild spaces were but two aspects of the same phenomenon and that 1-ULC properties on the one hand and variants of the disjoint disk property on the other played analogous and decisive roles. To demonstrate just how closely the theories were related, we will now explain two major areas in which one theory was used to further the other:

The construction of wild examples. B. J. Ball recognized Bing's wild dogbone space as the result of sewing together two subspaces of S^3 bounded by wild 2-spheres. On the other hand, N. Hosay and L. L. Lininger proved that wild 2-spheres in S^3 are always the image of tame 2-spheres under interesting cellular quotient maps from S^3 onto S^3. W. T. Eaton and R. J. Daverman established high dimensional analogues of the Ball and Hosay–Lininger results, respectively. M. A. Stan'ko proved results about codimension-three compacta which, R. D. Edwards pointed out, could be used to prove that wildly embedded codimension-three compacta are images of tame compacta under tame or shrinkable quotients. W. T. Eaton mixed wild compacta to create wild quotients. In other words, interesting examples in each of the theories spawned interesting examples in the other.

The characterization of tame subspaces and tame spaces. M. Brown, R. Kirby, and A. V. Cernavskii all proved various 1-ULC taming theorems for $n-1$ spheres in S^n by decomposition space techniques. On the other hand, W. T. Eaton, R. J. Daverman, and J. W. Cannon proved shrinking theorems for decomposition spaces by using 1-ULC properties, and taming theoretic techniques. Particularly in W. T. Eaton's work, a variant of the disjoint disk property was connected with certain

1-ULC taming properties and was used not just as a method of recognizing non-manifolds but as a tool in recognizing tame quotient spaces. And finally R. D. Edwards began his marvelous work on the double and triple suspension problems. At this point L. C. Glaser should be recognized for popularizing the decomposition space approach to the double suspension problem. Edwards made the key observation that intrinsic to the decomposition spaces associated with the double suspension problem were certain natural finite approximations to wild spheres of Alexander horned sphere type. Edwards had been led to expect such objects in decompositions by his study of M. A. Stan′ko's work on taming compacta.

In addition to the direct aid given one of the theories by the other in the two areas just mentioned (as in others), one noticed a number of parallel and analogous results, connections not well-understood but highly suggestive. Particularly striking were the results obtained upon stabilization (multiplication by some number of lines). A wild embedding $f: S^k \to S = f(S^k) \subset E^n$ into Euclidean n-space became tame in E^{n+1} (that is, $S \subset E^n \subset E^n \times E^1 = E^{n+1}$ is a tame topological k-sphere). This fact may be deduced as a consequence of the various known 1-ULC taming theorems. Furthermore, although the product $S \times E^1 \subset E^n \times E^1$ is wild when S is wild, it has the mildest possible form of wildness in terms of its 1-ULC properties according to R. J. Daverman's analysis of the same. On the other hand, no cell-like quotient Q of E^n was known to fail to be a factor of E^{n+1} ($Q \times E^1$ and $E^n \times E^1$ were almost always known to be homeomorphic). In particular, Bing had shown that his dogbone space was a manifold factor. And largely due to the impetus given the subject by some clever arguments and ideas of L. Rubin, a number of mathematicians began to prove that large and very general classes of manifold quotients were manifold factors—the best results issuing from C. Pixley, W. T. Eaton, R. T. Miller, and R. D. Edwards. J. L. Bryant and J. G. Hollingsworth considered the converse problem: is a manifold factor a manifold quotient? and proved the first weak resolution theorem. Pursuing the analogy between the 1-ULC taming properties and the decomposition space disjoint disk property further we note that though many decompositions failed to have the disjoint disk property, products with a line generally did (and as Daverman noticed in 1977, the product with two lines always had the property).

By the early to mid 1970s the interconnections between taming and decomposition space theory had become so numerous and obvious that we attempted to formalize the interconnections. We began a program to prove that 1) every taming theorem had a decomposition space analogue and 2) every decomposition space theorem had a taming theoretic proof. What emerged was first the realization that if some decomposition space theorems were to admit a taming theoretic proof, then one would have to extend taming theory to allow consideration of general manifold-like objects. We discovered early in 1975 that the properties of the generalized manifold introduced decades before by R. L. Wilder were precisely those amenable to 1-ULC taming theory and as a consequence proved by taming theoretic methods that every generalized $(n-1)$-manifold which embeds in an n-manifold is at least stably a cell-like quotient of a manifold. We were so struck by the discovery that an algebraically defined class of spaces (the generalized manifolds) should have such strong geometric properties that we immediately began to advertise the possibility in private discussions and in lectures that topological characterizations of manifolds, contrary to all appearances, might indeed be possible. When we were able to prove by the same taming techniques in the spring of 1976 that the double suspension of every homology sphere is a cellular quotient of a sphere, a result anticipated a few months by Edwards, we became even more convinced that the generalized manifold was exactly the right candidate for resolution theorems of the type suggested first by Bryant and Hollingsworth. Furthermore, we felt that the completion of Edwards' program of proving the double suspension theorem was at that point assured. A second consequence of our program was that we began consciously to concentrate on the 1-ULC properties of cell-like quotients. The realization came that the possible wildness of the double suspension quotients depended at worst on the 1-ULC wildness of the suspension circle and that, in 1-ULC taming theoretic terms, the wildness of the circle was of the simplest known type. After explanations by Edwards of his double suspension work, we saw that the looseness of the 1-ULC structure allowed us to find, actually embedded in the decomposition space in a very simple way, the wild spheres of Alexander horned sphere type whose finite approximations Edwards had noted. A complete proof of the double suspension theorem followed quickly from results on taming theory. We explained our work to Bing. He was not excited. He found the proof obscure. In frustration we sought the simplest possible conceptual framework encompassing the mildly wild 1-ULC properties of the examples. The appropriate property proved to be the disjoint disk property. Suddenly the connections between the various 1-ULC taming properties and the disjoint disk decomposition properties became clear in our minds and the characterization conjecture immediately took its present form.

Almost immediately after receiving our initial applications of the disjoint disk property, Edwards was able to prove the quotient conjecture. The disjoint disk property allowed one to embed the (infinite) 2-skeleton of the domain of the quotient map in the quotient space, and 1-ULC taming theory for decompositions allowed one to make the quotient map one to one over that infinite skeleton. Edwards noted that the embedding process forced the remaining nondegenerate point preimages of the quotient map to have 1-ULC complement, even low geometric embedding dimension, hence to be essentially tame or untangled in the decomposition space sense. Edwards was able by a clever engulfing type induction to untangle and shrink the remaining elements to points.

On the other hand, again using insights suggested by our use of the disjoint disk property, we, and Bryant and Lacher independently, were able to prove vastly improved versions of the known resolution theorems. Our versions depended on certain 1-ULC taming theorems of Černavskiĭ and Seebeck. Since the results just described were proved early in 1977, Ferry, Quinn, and Chapman have proceeded to generalize and strengthen those 1-ULC taming theorems to generalized manifolds in more and more general settings. Quinn suggests that the appropriate taming problems are exactly suited to the completion of an early dream from his graduate school days of establishing local versions of surgery. The proof of the resolution conjecture, and with it the characterizations of topological n-manifolds, $n \geq 5$, may soon, therefore, be complete.

References

1. J. W. Cannon, *The recognition problem: what is a topological manifold?*, Bull. Amer. Math. Soc. (to appear).

2. R. C. Lacher, *Resolutions of generalized manifolds*, Geometric Topology Conference, Warsaw, August 24—September 2, 1978.

University of Wisconsin
Madison, Wisconsin 53706, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Submanifolds of Small Codimension

Sylvain E. Cappell*

Some of the most interesting examples and important phenomena in the study of submanifolds occur in codimension one and codimension two. The study of submanifolds in each of these small codimensions involves distinctive problems and theories with rich interrelations with varied areas of geometry; they are very different from each other and from the study of submanifolds of codimension greater than two. For example, the study of submanifolds of codimension two encompasses knot theory and its general role in the understanding of both smooth and singular embeddings and immersions.

The main part of this report, section one, is an outline of one aspect of a series of joint investigations by Julius Shaneson and the author in the general existence and classification problem for codimension two embeddings and immersions (see [CS1]–[CS13]). These works present new methods in the study of smooth or locally-flat submanifolds as well as in the study of singular or nonlocally flat submanifolds and their singularities. Here, only the problem of the existence of embeddings, with or without singularities, and of immersions, in Euclidean space is discussed. These are related to characteristic classes.

The second section is an example of one of the author's results on existence and classification of submanifolds of codimension one [C1]–[C7]. These results have many applications to the study of invariants of manifolds with infinite fundamental group. For example, for manifolds whose fundamental groups lie in a very large

* This work was supported by Grant NSF-MC S76-06 974 from the National Science Foundation of the United States.

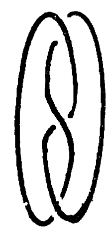
class of groups, they prove the generalized Novikov higher signature conjecture [C5]. They also lead to computations of the surgery groups of many infinite groups [C5], [C6], [C7] and they give wide generalizations of results of Wall [W] and Shaneson [S]. All these applications will not be discussed here.

Topologists are familiar with a host of geometrical results which are valid for submanifolds of codimension greater than or equal to three. Basically, this is because in the complement of a codimension three submanifold there is enough room left to carry out all the needed constructions; this reflects, in particular, that the removal of a subset of codimension greater than or equal to three does not affect the fundamental group. Thus, the well-known challenging difficulties in studying submanifolds of small codimension can only be overcome by new understanding of the role of the fundamental group.

1. Codimension two submanifolds. As a first step towards the general study of submanifolds of codimension two, consider the problem of determining which closed orientable n-manifolds embed (or immerse) in Euclidean space R^{n+2}. For $n=0, 1, 2$ the existence is obvious.

Note, however, that already for $n=1$, these embeddings are far from unique—e.g. the classification of embeddings of S^1 in R^3 (or S^3) is the subject of classical knot theory. This lack of uniqueness will shortly be seen to have large implications, for higher n, for the existence problem under discussion here. This is suggested by a familiar construction. If $\alpha: S^1 \to S^3$ is a knot, then the cone pair on this embedding is an embedding $\varrho_\alpha: D^2 \to D^4$. When α is piecewise-linear (P. L.) so is ϱ_α.

EXAMPLES. For α the unknot ϱ_α is just the standard linear inclusion. For α the trefoil knot, ϱ_α arises classically as a neighborhood of the origin in the solution set of $z_1^2+z_2^3=0$ in C^2.

unknot  trefoil knot 

The knot α can be recovered from ϱ_α, in the usual way, by intersecting a neighborhood of the origin (or cone point) of $D^2 \subset D^4$ with a small ball B^4 and considering $S^1=(\partial B \cap D^2) \subset \partial B = S^3$. The knot α is called the link of the origin of ϱ_α. Clearly the link of a point in a submanifold pair locally P. L. homeomorphic to a linear subspace is just the unknot. In general, P. L. embeddings in any dimension, at which all points have link the unknot are called locally-flat. The set of points at which a nonlocally flat P. L. embedding (or immersion, that is, a map which is locally an embedding) is not flat[1] is called the singularity set of f, $S(f)$. In the above examples, $S(\varrho_\alpha)$ was at most one point. In general, $S(f)$ is a subcomplex of M^n of dimension at most $n-2$, equipped with a natural stratification.[2]

Recall now that every orientable M^3 embeds smoothly[3] in R^5. We proceed to consider which M^4 embed in R^6. Here there is a remarkable bifurcation of the problem. Embeddings (or immersions) can be considered with or without the constraint of local flatness. For embeddings, we have the following contrasting results.

THEOREM [CS9]. *Let M^4 be an orientable closed P. L. manifold of dimension 4. (Recall that M then has a unique differential structure.)*

(1) *The following are equivalent:*

(i) *M embeds smoothly in R^6.*

(ii) *M embeds P. L. locally flat in R^6.*

(iii) *M is stably (i.e. $M\times R$ is) parallelizable.*

(iv) *The Pontrjagen class $p_1(M)$ and Stiefel–Whitney class $w_2(M)$ are zero.*

(2) *The following are equivalent:*

(i) *M embeds P. L. in R^6.*

(ii) *M embeds P. L. in R^6 with at most one non-locally flat point.*

(iii) *M embeds in R^6, smoothly except at one point.*

(iv) *M is almost (e.g. $M-\{a \text{ point}\}$ is) parallelizable.*

(v) *The Stiefel–Whitney class $w_2(M)=0$.*

Thus, it is a great deal easier to P. L. fit M^4 into S^6 if the local-flatness restraint is dropped. The parallel contrast for P. L. immersions (maps which are locally P. L. embeddings) is also striking. The first part of this result is just a specialization of the Smale–Hirsch criteria for smooth immersion.

THEOREM [CS9]. *Let M^4 be an orientable P. L. manifold.*

(1) *The following are equivalent:*

(i) *M immerses smoothly in R^6.*

(ii) *M immerses P. L. locally flat in R^6.*

(iii) *There is an $x\in H^2(M; Z)$ with $x\equiv w_2(M)$ (mod 2) and $x^2=-3\ I(M)$, $I(M)=$ index of M.*

(2) *The following are equivalent:*

(i) *M immerses P. L. in R^6.*

(ii) *M immerses P. L. in R^6, locally flat except at one point.*

(iii) *The Euler-characteristic of M, $\chi(M)$, is even.*

---

[1] As a consequence of P. L. unknotting in higher codimension this notion does not exist in codimension greater than two.

[2] The link pair of a nonisolated nonlocally flat point will itself be nonlocally flat. For example, the product of ϱ_α with a variety V has points with link the kth suspension of α. Yet more complicated examples arise and a full understanding of these is related to interpreting P. L. Pontrjagen classes.

[3] This result is easily recovered using the present methods.

A comparison of the tangential geometry of M and of a neighborhood of it W^6 in S^6 will explain how dropping the restraint of local flatness helps in fitting M into Euclidean space. Clearly, as a codimension zero subset of R^6, the tangent bundle of W is trivial. Moreover, if the embedding of M in R^6, and thus of M in W, is smooth, it is easy to see[4] that the normal bundle of M in W is trivial; hence the stable tangential structure of an orientable smooth submanifold $M^n \subset R^{n+2}$ is trivial. Now when the embedding of M^4 has a nonlocally flat point the relationship between the tangential structure of M and W is more complicated. At the nonlocally flat point, M fails to have a normal bundle; there, the pair $M \to W$ is locally P. L. homeomorphic to a cone pair, and will be identified with, $\varrho_\alpha: D^4 \to D^6$ where $\alpha: S^4 \to S^5$ is a P. L. locally flat knot. To understand the tangential geometry of W at the cone point, we will replace the disc D^4 by a P. L. locally-flat submanifold V^5 with boundary the knot $\alpha(S^4)$. For this, we employ a Seifert surface of α.

A knot $\alpha: S^{n-1} \to S^{n+1}$ is always the boundary of some $V^n \subset S^{n+1} = \partial(D^{n+2})$ called a Seifert surface of α. When $n \equiv 0$ (4), the signature or index of α, $\sigma(\alpha)$, is defined as that of the intersection pairing on $H_{n/2}(V)$, that is, of the signature of V. In view of the Thom–Hirzebruch formula, this invariant of α can be computed from the tangential geometry, in fact from the Pontrjagen characteristic classes, of $V \bigcup_{S^{n-1}} D^n$.

Now, replacing M in W by the *locally-flat* closed submanifold $(M - \varrho_\alpha(D^n)) \bigcup_{S^{n-1}} V^n$ (here $n=4$) with $V^n \subset \partial D^{n+2}$, the tangential geometry of W, and in particular its Pontrjagen class is seen to have a contribution from M and from $\sigma(\alpha)$. As $M \subset W$ and as $W^6 \subset R^6$ is to have trivial tangential structure, a recipe for the needed W is now apparent. Choose W to be a regular neighborhood of M with a single nonlocally flat point, chosen with link having signature the negative of that contributed by M. Using this choice of W, the results on immersions will follow.

Results on embeddings require in addition the construction of the "outside" of a neighborhood of M, $S^{n+2} - W$. In general, the fundamental groups of complements of codimension two submanifolds of a manifold can be rather large; consequently their homotopy types can be very complicated. It is therefore very difficult in codimension two to construct embeddings with prescribed homotopy type arising as the complement. This problem is generally obviated by using the homology surgery theory adapted to codimension two embeddings in [CS2].

Next consider, for $n > 4$, the general problem of embedding orientable M^n in S^{n+2}. Such M can have very rich tangential structure. In fitting such manifolds into Euclidean space, simultaneous contributions from knot theory in many dimensions to the normal singularities of M in its neighborhood are needed to account for the characteristic classes of M. Thus the singularity set $S(f)$ will often be

[4] Precisely, this R^2 (or C^1) bundle is determined by the Euler-class (or Chern-class) which is easily seen to vanish.

a large subcomplex. Moreover, the construction of the neighborhood cannot be "dimension by dimension" as this would only give integral invariants, and thus not the P. L. Pontrjagen classes. The construction of the needed neighborhood is done by using a classifying space $BSRN_2$ for oriented codimension two manifold regular neighborhoods of manifolds, that is, of normal "2-plane bundles with singularities". The construction and analysis of this space uses some delicate P. L. topology **[CS11]**. The analysis also uses many results on the relation of homology surgery to codimension two problems **[CS2]**, **[CS11]** as well as homotopy theoretic methods. As before, completion of the results on embeddings also uses the homology surgery methods to construct the outside of a neighborhood of a submanifold. There is an obstruction to doing this, but we show that, quite generally, it vanishes for odd-dimensional submanifolds. In even dimensions we kill the obstruction, when the target manifold is simply connected, by introducing more singularities in the neighborhood of the submanifold. This again illustrates the close relation between knot theory, singularities, and homology surgery.[5]

Specializing the general consequences of this approach to Euclidean space yields the following result. Call the P. L. oriented closed connected manifold M k-reducible if the Hurewicz map

$$\pi_{n+k}(\Sigma^k M) \to H_{n+k}(\Sigma^k M) \cong Z$$

is onto. This clearly depends only on the homotopy type of M; k reducibility for any (or equivalently all k), $k > n$, is called stable reducibility. Any manifold homotopy equivalent to a stably parallelizable manifold is stably reducible.

THEOREM **[CS11]**, **[CS14]**. *If M^n is 1-reducible, it embeds P. L. in S^{n+2}; if M embeds in S^{n+2}, it is 2-reducible. If M is stably reducible, it P. L. immerses in S^{n+2}; if $H^2(M; Z)=0$ and M P. L. immerses in S^{n+2}, it is stably reducible.*

In the construction of such embeddings, singularities $S(f)$ were used to account for tangent bundle invariants. Quantitatively, this leads to lower bounds for the dimension of $S(f)$. Precisely, let $f\colon M^n \to W^{n+2}$ be a P. L. embedding. Set $L(f)=L(M)L(x)-f^*L(W)$ where $L(M)$ and $L(W)$ denote total Thom–Hirzebruch classes and $L(x)$ is that class in x, $x=f^*Df_*[M]$, D being Poincaré duality; in Euclidean space, $L(f)=L(M)$. Then it is easy to see that $D(L(f))$ is carried by $S(f)$, the set of nonlocally flat points **[CS11]**.

EXAMPLE. *Let M^n, $n \geq 4$, be a simply-connected P. L. manifold which embeds P. L. in S^{n+2} with $H^4(M; Q) \neq 0$; there is then a P. L. manifold M' homotopy equivalent to M with nonzero first Pontrjagen class. Then there is a P.L. embedding $f\colon M' \to S^{n+2}$; but for any such f, $n-4 \leq \dim(S(f)) \leq n-2$.*

[5] When the target manifold is even dimensional but not simply-connected it may be impossible to kill the obstruction. This leads to the construction of totally spineless manifolds **[CS6]**.

More generally, in the odd-dimensional or simply-connected cases our methods reduce to homotopy theory the general question of whether a map $f\colon M^n \to W^{n+2}$ is homotopic to a P. L. embedding, with $S(f)$ of given prescribed codimension. In view of the above discussions, the criteria for reducing the size of $S(f)$ can often be stated in terms of the vanishing of certain characteristic classes. As an application of these ideas, we give the following example on equivariant embeddings in spheres.

THEOREM [CS11]. *Let G be finite cyclic or S^1. Let ϱ and τ be arbitrary free actions of G on S^n and S^{n+2}, in the P. L. category. Assume $n \neq 4, 6$, and $n \neq 5$ if $G = S^1$. Then there is an equivariant embedding of S^{n-2} in S^n; i.e., there is a one to one P. L. map $f\colon S^n \to S^{n+2}$ with $f(\tau(t,x)) = \varrho(t, f(x))$, $t \in G, x \in S^{n-2}$. Moreover, for $G = S^1$ or Z_2, $n > 4$, for every ϱ (respectively, τ) there is a τ (resp. ϱ) such that for any such equivariant f,* $\dim(S(f)) = n-2$.

As our most general results, applicable to any W^{n+2}, show, a regular neighborhood of a P. L. embedding or immersion $f\colon M^n \to W^{n+2}$ frequently must carry many nonzero characteristic classes. There is a formula for the Pontrjagen classes, or the L-classes, in terms of the simplices of $S(f)$; an intriguing problem, closely related to obtaining a general P. L. geometric understanding of characteristic classes, would be to find out what the coefficients are.

2. Submanifolds of codimension one. To simplify notation, we consider here only orientable closed, connected, smooth (or P. L. and P. L. locally flat, or topological and topologically locally flat) manifolds and submanifolds. The central problem in the study of codimension one submanifolds was to reduce the question of their existence to homotopy theory, i.e., to show that the existence of a (simple) homotopy theoretic model $X^n \subset Y^{n+1}$, with $n \geqslant 5$ and $\pi_1 X \to \pi_1 Y$ injective, for a codimension one submanifold implied the existence of an actual geometrical submanifold. Precisely, one takes a homotopy equivalence $f\colon W^{n+1} \to Y^{n+1}$ with W a manifold[6] and asks if f is homotopic to a map g transverse to X with $g^{-1}X \to X$ a homotopy equivalence. When g exists, f, or its homotopy class, is said to be splittable. As this implies that $(W - f^{-1}(X)) \to (Y - X)$ is also a homotopy equivalence, such results on codimension one submanifolds can be used to decompose manifolds into pieces of prescribed homotopy type; in view of Van Kampen's theorem and its generalizations, these pieces will have smaller fundamental group.

For important earlier results on this problem, see **[BL]**, **[B1]**, **[B2]**, **[BL2]**, **[F]**, **[FH]**, **[L]**, **[N]**, **[S]**, **[W]**; the much more general results of **[C1]** apply equally whether $(Y - X)$ is connected or not and to manifolds with fundamental group in a large class. For example, they assert the existence of g whenever X^n is simply-connected and $\pi_1 Y^{n+1}$ has no 2-torsion (and sometimes, when there is 2-torsion). The follow-

[6] Y and X could be manifolds or even just Poincaré complexes.

ing example showed that this could not be pushed much further. The connected sum of M^{n+1} and N^{n+1} is formed by deleting a ball from each and pasting together the results along the boundary S^n to get a closed manifold, containing S^n as a codimension one submanifold.

EXAMPLE [C2]. *There are infinitely many different smooth closed W^{4k+1}, each (simple and tangentially) homotopy equivalent to the connected sum Y of two copies of projective space RP^{4k+1}, but which are not nontrivial connected sums.*

Thus there is no codimension one splitting result for the S^{4k} in this Y. In view of this, we apply further the methods of [C1] to get results applicable to any $X^n \subset Y^{n+1}$ with $\pi_1 X \to \pi_1 Y$ injective. In [C4] we define an algebraic functor $\mathrm{UNil}^h_{n+2}(\pi_1 X \subset \pi_1 Y)$ of groups of unitary nilpotent matrices. For brevity, we here describe a result only in the case when the reduced projective class group of $Z[\pi_1 X]$ is 0. See [C4], [C3], [C7] for a general discussion. Let $S(Y)$ denote the set of manifolds with given homotopy equivalences to Y^{n+1}, $n \geqslant 5$, classified up to h-cobordism, and let $S'_X(Y)$ similarly denote homotopy equivalences split along X.

THEOREM. *$S(Y)$ is canonically in one to one correspondence with $S'(Y) \times \mathrm{UNil}^h_{n+2}(\pi_1 X \subset \pi_1 Y)$.*

There are similar results for s-cobordisms. The UNil groups are always 2-groups but, when not zero are often not finitely generated. Our earlier splitting theorems are interpretable as vanishing theorems for UNil in various cases.

References

[B1] W. Browder, *Embedding 1-connected manifolds*, Bull. Amer. Math. Soc. **72** (1966), 225—231.

[B2] ——, *Manifolds with $\pi_1 = Z$*, Bull. Amer. Math. Soc. **72** (1966), 238—244.

[BL] W. Browder and J. Levine, *Fibering manifolds over a circle*, Comment Math. Helv. **40** (1966), 153—160.

[BLL] W. Browder, J. Levine and G. R. Livesay, *Finding a boundary for an open manifold*, Amer. J. Math. **87** (1965), 1017—1028.

[C1] S. E. Cappell, *A splitting theorem for manifolds*, Invent. Math. **33** (1976), 69—170.

[C2] ——, *On connected sums of manifolds*, Topology **13** (1974), 395—400, MR **50** #11252.

[C3] ——, *Manifold with fundamental group a generalized free product*, I, Bull. Amer. Math. Soc. **80** (1974), 1193—1198. MR **50** #8562.

[C4] ——, *Unitary nilpotent groups and Hermitian K-theory*, Bull. Amer. Math. Soc. **80** (1974), 1117—1122. MR **50** #11274.

[C5] ——, *On homotopy invariance of higher signatures*, Invent. Math. **33** (1976), 171—179.

[C6] ——, *Mayer-Vietoris sequences in Hermitian K-theory*, Proc. Battelle K-theory Conf., Lecture Notes in Math., vol. 343, Springer-Verlag, Berlin and New York, 1973, pp. 478—512.

[C7] ——, *Decompositions of manifolds* (to appear).

[CS1] S. E. Cappell and J. L. Shaneson, *Submanifolds, group actions, and knots*, I, II, Bull. Amer. Math. Soc. **78** (1972), 1045—1052.

[CS2] *The codimension two placement problem and homology equivalent manifolds.* Ann. of Math. (2) **99** (1974), 277—348. MR **49** #3978.

[CS3] *Topological knots and knot cobordism,* Topology **12** (1973), 33—40. MR **47** #9632.

[CS4] *Nonlocally-flat embeddings, and group actions.* Bull. Amer. Math. Soc. **79** (1973), 577—582.

[CS5] *Close codimension two embeddings of even dimensional manifolds.* Amer. J. Math. **97** (1975), 733—740.

[CS6] *Totally spineless manifolds,* Illinois J. Math. **21** (1977), 231—239.

[CS7] *Branched cyclic covers,* Knots, Groups and 3-Manifolds, Papers in Memory of R. H. Fox, Ann. of Math. (2) **84** (1975), 163—173.

[CS8] *Fundamental groups, Γ-groups, and codimension two embeddings.* Comment. Math. Helv. **39** (1976), 437—446.

[CS9] *Every spin 4-manifold PL embeds in* S^4 (to appear).

[CS10] *Branched covering spaces and PL singularities* (to appear).

[CS11] *Piecewise linear embeddings and their singularities,* Ann. of Math. (2) **103** (1976), 163—228.

[CS12] *An introduction to embeddings, immersions and singularities in codimension two,* Algebraic and Geometric Topology, Proc. Sympos. Pure Math., vol. 32, part 2, Amer. Math. Soc., Providence, R. I., (1978), pp. 129—149.

[CS13] *A note on the Smith conjecture,* Topology (to appear).

[CS14] *Singularities and immersions,* Ann. of Math. **105** (1977), 539—552.

[F] F. T. Farrell, *The obstruction to fibering a manifold over a circle,* Indiana Univ. Math. J. **21** (1971), 315—346.

[FH] F. T. Farrell and W. C. Hsiang, *Manifolds with* $\pi_1 = Gx_\alpha T$, Amer. J. Math. **95** (1973), 813—848.

[H] W. Homer, *Equivariant embedding of* Z_p*-actions on spheres,* Topology (to appear).

[J] L. Jones, *Three characteristic classes measuring the obstruction to PL local unknotedness,* Bull. Amer. Math. Soc. **78** (1972), 979—980.

[L] R. Lee, unpublished.

[N] S. P. Novikov, *On manifolds with free abelian fundamental group and their applications,* Izv. Akad. Nauk SSSR Ser. Mat. **30** (1966), 207—246.

[S] L. C. Siebenmann, *The obstruction to finding the boundary of an open manifold,* thesis, Princeton Univ., 1965.

[Sh1] J. L. Shaneson, *Wall's surgery obstruction groups for* $Z \times G$, Ann. of Math. **90** (1969) 296—334.

[Sh2] *Non-simply connected surgery and some results in low dimensional topology,* Comment Math. Helv. **45** (1970), 333—352.

[W] C. T. C. Wall, *Surgery on compact manifolds,* New York, Academic Press, 1970.

Courant Institute of Mathematical Sciences
New York University
New York, New York 10012, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Linearization in 3-Dimensional Topology

A. E. Hatcher

The message from Thurston is clear: *geometry dominates topology in dimension three.* Our title refers to one important aspect of this geometry, *linearity* in various forms. We shall consider here linearization of automorphisms of 3-manifolds. Specifically, we ask, does $\mathrm{Diff}(M)$, the group of self-diffeomorphisms of the 3-manifold M (with the C^∞ topology), have the homotopy type of the subgroup of diffeomorphisms which preserve a given "linear" structure on M? If so, this is strong evidence that the linear structure is really intrinsic to the topology of M.

EXAMPLES (M closed, orientable).

(I) $M=S^3$. The linear diffeomorphisms of S^3 are the isometries, $\mathrm{Isom}(S^3)=O(4)$, the orthogonal group.

The Smale Conjecture. $O(4)\hookrightarrow \mathrm{Diff}(S^3)$ is a homotopy equivalence.

We shall indicate some of the ideas which go into a proof of this below. Previously, Cerf had shown that $\pi_0\,\mathrm{Diff}(S^3)\approx\pi_0 O(4)$. He also proved that the Smale Conjecture implies that $\mathrm{Diff}(M)\to\mathrm{Homeo}(M)$, the natural map to the homeomorphism group, is a weak homotopy equivalence for all 3-manifolds M.

(II) $M=S^3/\Gamma$, $\Gamma\subset\mathrm{SO}(4)$ acting freely on S^3, the so-called spherical or elliptic 3-manifolds. One expects $\mathrm{Diff}(M)\simeq\mathrm{Isom}(M)$, but this is known only for $\boldsymbol{RP}^3$.

(III) $M=H^3/\Gamma$, $\Gamma\subset\mathrm{Isom}(H^3)$, hyperbolic 3-manifolds. Again, one expects $\mathrm{Diff}(M)\simeq\mathrm{Isom}(M)$, and this is known when M is Haken (="irreducible, sufficiently large" in the older terminology). By a theorem of Mostow, $\mathrm{Isom}(M)\approx\mathrm{Out}(\pi_1 M)$, the outer automorphism group of $\pi_1 M$ (automorphisms modulo inner automorphisms), which is not only discrete but finite.

(IV) $M=E^3/\Gamma$, $\Gamma\subset\mathrm{Isom}(E^3)$, euclidean or flat 3-manifolds. Here the linear diffeomorphisms are the affine diffeomorphisms (i.e., affine in the universal cover E^3). Such an M is Haken, so one can show $\mathrm{Diff}(M)\simeq \mathrm{Aff}(M)$. For example, $\mathrm{Diff}(T^3)\simeq \mathrm{GL}(3, \mathbf{Z})\times T^3$ (semidirect product). This is considerably larger than $\mathrm{Isom}(T^3)$, which is compact.

(V) $M=S^1\times S^2$. This is best regarded as a bundle $S^2\to M\to S^1$ with linear structure group $O(3)$. Then $\mathrm{Diff}(S^1\times S^2)\simeq O(2)\times O(3)\times \Omega O(3)$, the bundle automorphisms (Rourke–César de Sá).

(VI) $M=T^2$-bundle over S^1 with gluing map in $\mathrm{SL}(2, \mathbf{Z})$ having distinct real eigenvalues. Again $\mathrm{Diff}(M)$ has the homotopy type of the bundle automorphism group.

(VII) M Seifert fibered, $S^1\to M\to B$, over a closed surface B. These include the manifolds of I, II, IV, V, but none of those in III or VI. Excluding the manifolds in I, II, IV, and V, the Seifert fiber structure is unique, and $\mathrm{Diff}(M)\simeq$ {fiber-preserving diffeomorphisms} except perhaps when $B=S^2$ and there are only three singular fibers (the non-Haken cases). Incidentally, the linear structure in a Seifert fibering is in the base B, which is naturally a quotient of the spherical, euclidean, or hyperbolic plane by a discrete group G of isometries (perhaps with torsion). For example, M could be the unit tangent bundle of such a B, which has singular fibers arising from the elements of G with fixed points (rotations of finite order).

(VIII) M having a *torus decomposition,* i.e., a splitting of M into submanifolds M_j which are the components of the complement of a finite collection (perhaps empty) of disjointly embedded tori T_i in M, such that

(1) $\pi_1 T_i\to\pi_1 M$ is injective for each i (to rule out the possibility that T_i bounds a solid torus in M).

(2) Each M_j is either

(a) the interior of a compact Seifert fibered manifold, or

(b) a hyperbolic manifold H^3/Γ of finite volume (having finite volume is almost as good as being compact).

(3) $\{T_i\}$ is minimal, with respect to inclusion, among collections $\{T_i\}$ satisfying (1) and (2).

Small Exception. For the T^2-bundles in VI above it seems better to choose $\{T_i\}=\emptyset$ rather than a single fiber T^2, which is what (1)–(3) would yield. (The other T^2-bundles over S^1 are Seifert-fibered.)

Perhaps the deepest result in 3-manifolds to date is:

THEOREM. *Every Haken manifold has a torus decomposition which is unique up to isotopy.*

This is due to Johannson and (independently) Jaco–Shalen for the Seifert part, and to Thurston for the (much harder) hyperbolic part. As far as is known, all prime 3-manifolds (i.e., indecomposable as a connected sum) have torus decompositions, since all known prime 3-manifolds are either Seifert, hyperbolic, or Haken.

THEOREM. Diff (M) *deformation retracts onto the subgroup of diffeomorphisms leaving* $\bigcup_i T_i$ *invariant.*

If $\{T_i\} \neq \emptyset$, the components of Diff(M) are contractible. So the content of the preceding theorem is to reduce π_0 Diff(M) essentially to the π_0 Diff(M_j)'s. For example, one can say that π_0 Diff(M) is generated by:

diffeomorphisms which permute the T_i's and the M_j's,

Dehn twists along the T_i's,

isometries of the hyperbolic M_j's,

fiber-preserving diffeomorphisms of the Seifert M_j's.

This is reminiscent of Thurston's normal form for diffeomorphisms of surfaces.

(IX) M nonprime. Rourke and César de Sá have largely reduced Diff(M) in this case to Diff(M_j) for the prime factors M_j of M, plus the homotopy theory of certain "configuration spaces". This seems to be rather complicated in general.

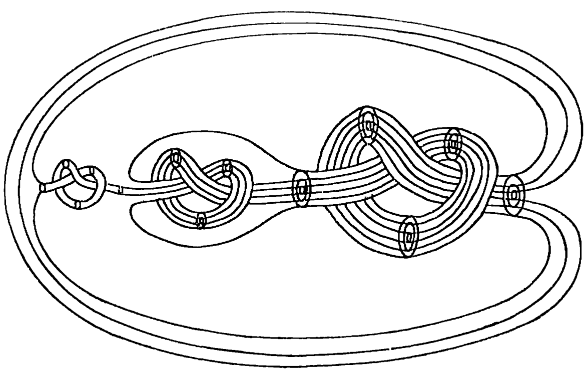
THE SMALE CONJECTURE: Diff$(S^3) \simeq O(4)$.

There are many statements well-known to be equivalent to this, e.g.,

(1) The space of unknotted smoothly embedded circles in $\boldsymbol{R}^3$ deformation retracts onto the subspace of round (i.e., planar, constant curvature) circles.

(2) The space of smoothly embedded 2-spheres in $\boldsymbol{R}^3$ deformation retracts onto the subspace of round (i.e., constant curvature) 2-spheres.

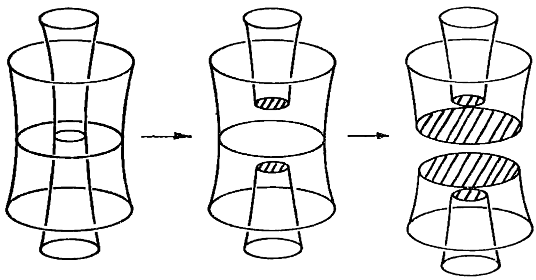
Of these, (1) seems hopeless: there appears to be no canonical way of unknotting the unknot. At first glance, (2) seems even harder if one looks at embedded 2-spheres with apparent knots, like the following:



Nonetheless, (2) can be proved, using only elementary (but complicated) differential topology. Intuition suggests there ought, also, to be an analytic proof of (2), based on some physical model for 2-spheres in $\boldsymbol{R}^3$. However, somewhere the topology or geometry of three dimensions will have to enter, since the analog of (2) either for 3-spheres in $\mathbf{R}^4$ or for 4-spheres in $\mathbf{R}^5$ is definitely false (this is directly traceable to the existence of exotic 7-spheres).

What one actually proves is the following technical variant of (2): a smooth family $g_t: S^2 \hookrightarrow \mathbf{R}^3$ parametrized by $t \in S^k$ extends to a smooth family $\bar{g}_t: B^3 \hookrightarrow \mathbf{R}^3$. For $k=0$ this is essentially due to Alexander, and for $k=1$ this is what Cerf showed to calculate $\pi_0 \operatorname{Diff}(S^3) \approx \mathbf{Z}_2$.

The good property of 2-spheres in $\mathbf{R}^3$ is that they can be sliced into simpler and simpler 2-spheres by surgery on horizontal circles:



(Note this fails for $S^{n-1} \subset \mathbf{R}^n, n \geqslant 4$.)

The process can be iterated: surger all the circles of intersection of the given 2-sphere with more and more horizontal transverse planes. Eventually a point is reached when further surgeries no longer yield significantly simpler 2-spheres. We call such 2-spheres, somewhat loosely, "indecomposable".

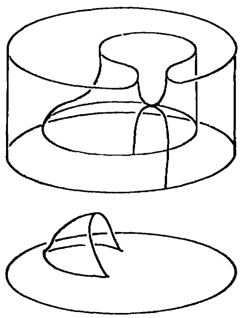
It is easy to reverse the surgery process, gluing together extensions $\bar{g}_t$. So the problem becomes to construct $\bar{g}_t$ on the "indecomposables". This must be done in a canonical way, which works for k-parameter families of "indecomposables".

A further problem is that one cannot choose the same horizontal slicing planes for all $t \in S^k$, but only locally in t. That is, one covers S^k by balls B_i, associated to each of which is a finite collection of horizontal planes P_{ij} transverse to $g_t(S^2)$ for $t \in B_i$. One surgers $g_t(S^2)$ using the planes P_{ij} for $t \in B_i$. So on intersections of B_i's, the "indecomposables" are being further surgered (subdivided), and one must take pains to make the extensions $\bar{g}_t$ on "indecomposables" invariant under such subdivision.

Thus the heart of the problem is understanding the "indecomposables". For small values of k, one can perturb the family g_t so that the height functions on $g_t(S^2)$ form a generic k-parameter family, and then write down a complete catalog of the types of "indecomposables". This is how Cerf proceeded for $k=1$. But for general k this approach fails (because smooth singularities are classified only for small codimensions k).

So one must forget height functions, and instead look at "indecomposable" 2-spheres *from the top*. This viewpoint leads to the following basic definition: The *contour* of a 2-sphere $\Sigma \subset \mathbf{R}^3$ is the quotient space of the 3-ball $\bar{\Sigma} \subset \mathbf{R}^3$ bounded by Σ, obtained by identifying points x and y in $\bar{\Sigma}$ whenever there is a vertical line segment in $\bar{\Sigma}$ joining x and y.

EXAMPLE.



Contour = 2-disc with a "flap" or "tongue"

In general, the part of $g_t(S^2)$ between two adjacent slicing planes P_{ij} has the following key property (after a preliminary normalization): Each vertical line in $\boldsymbol{R}^3$ meets this part of $g_t(S^2)$ in a connected set (perhaps empty). Using this, one proves that the contour of an "indecomposable" is always a disc with finitely many tongues attached successively, either to the disc or to previously attached tongues. ("Attaching a tongue" means attaching a disc D along a subdisc which meets ∂D nontrivially.)

To get the canonical extensions $\bar{g}_t$ on "indecomposables", the procedure is: Shrink each tongue in turn down to the arc along which it attaches. This lifts to an isotopy of the "indecomposable", ending with a 2-sphere whose contour is a disc. For this, $\bar{g}_t$ is easily constructed. Then reversing the isotopy which shrank the tongues, one obtains $\bar{g}_t$ on the original "indecomposable" by isotopy extension (which is canonical). This uses Smale's theorem $\mathrm{Diff}(S^2) \simeq O(3)$, to make $\bar{g}_t$ canonical.

The hard work comes in making this shrinking-of-contours process mesh with the subdivision (slicing) of "indecomposables" mentioned earlier.

One might well ask if shrinking of contours could be applied to the original $g_t(S^2)$. Unfortunately, one can easily construct examples of 2-spheres in $\boldsymbol{R}^3$ whose contours cannot be continuously shrunk, within themselves, to any subdisc. (Such contours are contractible but not collapsible, in the sense of PL topology.) So the slicing process is necessary.

APPLICATION.

THEOREM (C. B. THOMAS). *If the Smale Conjecture is true, then: A 3-manifold M with universal cover S^3 has the homotopy type of one of the spherical manifolds S^3/Γ (for some $\Gamma \subset \mathrm{SO}(4)$ acting freely on S^3 as isometries).*

In particular, $\pi_1 M \approx \Gamma$. To classify such M's there remains the problem of showing that Γ can act on S^3 only in the standard linear ways. This is known for some Γ's e.g., Z_2, Z_4, Z_6, Z_8, generalized quaternion (of order 2^k), binary tetrahedral and octahedral. (See the article of Rubinstein for references.) It is unknown, in particular, for Γ cyclic of odd order (e.g., Z_3!), and for the binary icosahedral group.

References

1. J. W. Alexander, *On the subdivision of* 3*-space by a polyhedron*, Proc. Nat. Acad. Sci. U.S.A **10** (1924), 6—8.

2. J. Cerf, *Sur les difféomorphismes de la sphère de dimension trois* ($\Gamma_4=0$), Lecture Notes in Math., vol. 53, Springer-Verlag, Berlin and New York, 1968.

3. E. César de Sá and C. Rourke, *The homotopy type of homeomorphism groups of* 3*-manifolds;* Math. Inst. Warwick, 1978 (preprint).

4. A. Hatcher, *Homeomorphisms of sufficiently large* P^2*-irreducible* 3*-manifolds,* Topology **15** (1976), 343—347.

5. N. V. Ivanov, *Diffeomorphism groups of Waldhausen manifolds,* Issledovania po Topologii II, ed. A. A. Ivanov, Izdatelstvo "Nauka" Leningrad, 1976, pp. 172—176.

6. W. H. Jaco and P. B. Shalen, *Seifert fibered spaces in* 3*-manifolds* (preprint).

7. K. Johannson, *Équivalences d'homotopie des variétés de dimension* 3, C. R. Acad. Sci. Paris Sér. A, **281** (1975), 1009—1010.

8. P. Orlik, *Seifert manifolds,* Lecture Notes in Math., vol. 291, Springer-Verlag, Berlin and New York, 1972.

9. J. H. Rubinstein, *On* 3*-manifolds that have finite fundamental group and contain Klein bottles* (preprint).

10. S. Smale, *Diffeomorphisms of the* 2*-sphere,* Proc. Amer. Math. Soc. **10** (1959), 621—626.

11. W. P. Thurston, *Geometry and topology of* 3*-manifolds,* notes from a course at Princeton University, 1977—1978.

University of California at Los Angeles
Los Angeles. California 90024, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

The Topology of Finite H-Spaces

James P. Lin*

The following question could be considered the central problem in the theory of finite H-spaces:

What homological and homotopical properties of Lie groups can be derived solely from the existence of a nontrivial product with unit?

This question has generated many interesting solutions. For example,

Let G be a Lie group

1. Hopf [5]: The rational cohomology of G is a finite dimensional Grassmann algebra.
2. Borel [2]: The mod p cohomology of G is a finite dimensional tensor product of exterior and truncated polynomial algebras.
3. Browder [4]: $\pi_2(G)=0$.
4. Hubbuck–Kane [6], Lin [12]; $\pi_3(G)$ is torsion free.

Results 1–4 depend only on the fact that G is a *finite H-space.* We say a space X is *finite* if for each prime p, $H_*(X; Z_{(p)})$ is of finite type and $H_l(X; Z_{(p)})=0$ for l greater than some integer N. We say a pointed space X, $*$ is an *H-space* if it supports a map u: $X\times X\to X$ such that the two compositions

$$X\times * \to X\times X \xrightarrow{u} X, \quad *\times X \to X\times X \xrightarrow{u} X$$

are homotopic to the identity.

Given a finite H-space X, $H^*(X; Q)$ and $H^*(X; Z_p)$ become commutative Hopf algebras. The results of Hopf and Borel follow from the basic theory of

* Partially supported by the National Science Foundation of the United States and the Sloan Foundation.

Hopf algebras. Results 3 and 4, however, follow from deeper considerations about the action of the Steenrod algebra on $H^*(X; \mathbf{Z}_p)$. I will briefly outline these methods and summarize the results that are presently known.

If $f: X \to Y$ is an H-map between two H-spaces X and Y the induced maps

$$H^*(Y; \mathbf{Z}_p) \xrightarrow{f^*} H^*(X; \mathbf{Z}_p) \quad \text{and} \quad f_*: H_*(X; \mathbf{Z}_p) \to H_*(Y; \mathbf{Z}_p)$$

are maps of Hopf algebras. We can extend this idea to H-spaces with added structure. If X and Y are H-spaces that admit A_n structures and $f: X \to Y$ is an A_n map in the sense of Stasheff [**13**], then f_* is a map of Hopf algebras with an additional induced map $B_n f$: $B_n X \to B_n Y$ of n fold projective spaces. There is an analogous statement for an H-map f between H_n spaces in the sense of Kudo–Araki [**10**] and Browder [**3**]. An H_n map would not only preserve the H-structure but also the higher homotopy commutativity of X and Y.

At present we consider only the most elementary case—a map f: $X \to Y$ between H-spaces X and Y. Suppose there is an element u in $H_n(Y; \mathbf{Z}_p)$ with $u^p \neq 0$. If f is an H-map with $f_*(t) = u$, then

$$f_*(t^p) = f_*(t)^p = u^p \neq 0.$$

It follows that $t^p \neq 0$. This is a way of detecting pth powers in $H_*(X; \mathbf{Z}_p)$. The problem can be divided into three parts:

1. Constructing H-spaces Y that have elements $u \in H_n(Y; \mathbf{Z}_p)$ with $u^p \neq 0$.
2. Constructing maps f: $X \to Y$ with $f_*(t) = u$.
3. Studying the deviation of f from being an H-map.

Problem 1 has several solutions. It is well known that $H_*(\Omega\Sigma Z)$ is a tensor algebra on $IH_*(Z)$, hence contains many pth powers in homology. The problem with using $\Omega\Sigma Z$ is that it is very difficult to map finite H-spaces X into $\Omega\Sigma Z$ in a nontrivial manner. Fortunately, there are other candidates that are stable two stage Postnikov systems. Because a two stage Postnikov system has k-invariants given by primary operations, the vanishing of primary operations on elements of $H^*(X; \mathbf{Z}_p)$ allow us to construct maps f: $X \to Y$. Finally, the H-deviation of f for a two stage system often (but not always) turns out to be primary (see [**17**]).

Here is an example of the type of theorem we can prove:

THEOREM 1. *Let* $t \in PH_{2n}(X; \mathbf{Z}_p)$ *with* $\langle t, x \rangle \neq 0$, $x \in H^{2n}(X; \mathbf{Z}_p)$. *Suppose* $\beta_1 \mathscr{P}^n$ *factors as* $\beta_1 \mathscr{P}^n = \sum a_i b_i$ *and* $b_i x = 0$ *for all* i. *Then if* $t \otimes t \otimes \cdots \otimes t$ *belongs to the kernel of* $\sum a_i$ *then* $t^p \neq 0$.

Through a great deal of hard work using this theorem it was shown (Kane [**8**]): If X is a finite H-space and $t \in PH_{2n}(X; \mathbf{Z}_p)$ then $t^p = 0$. This implies

THEOREM 2. *Let* X *be a finite H-space and let* $t \in PH_{2n}(X; \mathbf{Z}_p)$ *with* $\langle t, x \rangle \neq 0$, $x \in H^{2n}(X; \mathbf{Z}_p)$. *Suppose* $\beta_1 \mathscr{P}^n = \sum a_i b_i$ *with* $b_i x = 0$ *for all* i. *Then* $\sum a_i$ *does not vanish on* $t \otimes t \otimes \cdots \otimes t$.

Because the conclusion of Theorem 2 tells us Steenrod operations act nontrivially on homology classes, we can use this theorem to collect information about $H^*(X; \mathbf{Z}_p)$ as a module over the Steenrod algebra. The following theorems are proven by analyzing variations of Theorem 2 [**12**]. We assume p is an odd prime and X is a finite H-space.

THEOREM A. $QH^{\text{even}}(X; \mathbf{Z}_p)=\sum \beta_1 \mathscr{P}^l QH^{2l+1}(X; \mathbf{Z}_p)$.

This theorem tells us every even generator is in the image of Steenrod operations.

THEOREM B. *$H^*(X; \mathbf{Z})$ has no p^2 torsion.*

This follows from the Bockstein spectral sequence and the fact that every even generator is in the image of β_1.

THEOREM C. *$H^*(\Omega X; \mathbf{Z}_p)$ is even dimensional; therefore $H_*(\Omega X; \mathbf{Z})$ has no p torsion.*

This follows because the suspension map $\sigma^*: QH^{\text{even}}(X; \mathbf{Z}_p)\to PH^{\text{odd}}(\Omega X; \mathbf{Z}_p)$ is an epimorphism. Hence by Theorem A, $PH^{\text{odd}}(\Omega X; \mathbf{Z}_p)=0$ which implies $H^*(\Omega X; \mathbf{Z}_p)$ is even dimensional.

THEOREM D. *The Hurewicz map*

$$h_n\otimes \mathbf{Z}_{(p)}: \pi_n(X)\otimes \mathbf{Z}_{(p)} \to PH_n(X; \mathbf{Z}_{(p)})$$

has the property kernel $h_n\otimes \mathbf{Z}_{(p)}=p$ torsion $\pi_n(X)$.

This follows because of the commutative diagram

$$\begin{array}{ccc}
\pi_{n-1}(\Omega X)\otimes \mathbf{Z}_{(p)} & \xrightarrow{h_{n-1}\otimes \mathbf{Z}_{(p)}} & H_{n-1}(\Omega X; \mathbf{Z}_{(p)}) \\
\Big\downarrow{\scriptstyle \sigma_{\#}\otimes \mathbf{Z}_{(p)}} & & \Big\downarrow{\scriptstyle \sigma_*\otimes \mathbf{Z}_{(p)}} \\
\pi_n(X)\otimes \mathbf{Z}_{(p)} & \xrightarrow[h_n\otimes \mathbf{Z}_{(p)}]{} & PH_n(X; \mathbf{Z}_{(p)})
\end{array}$$

with $\sigma_{\#}\otimes \mathbf{Z}_{(p)}$ an isomorphism. Given z, a torsion class in $\pi_n(X)\otimes \mathbf{Z}_{(p)}$, $z=\sigma_{\#}\otimes \mathbf{Z}_{(p)}(w)$ where w is a torsion class. Hence $h_{n-1}\otimes \mathbf{Z}_{(p)}(w)=0$ by Theorem C. Therefore $h_n\otimes \mathbf{Z}_{(p)}(z)=0$. We have p torsion $\pi_n(X)\subseteq \ker h_n\otimes \mathbf{Z}_{(p)}$. The other containment follows from the Cartan–Serre Theorem which states that the rational homotopy is isomorphic to the rational primitives of X via the Hurewicz map. For more theorems of this type we refer the reader to [**12**].

When one attempts to prove analogous theorems for the prime two there are some fundamental obstructions that prevent one from simply following the proofs for odd primes. First, the mod two Steenrod algebra and the mod p Steenrod algebra have different coalgebra structures. For example $\mathscr{P}^1$ is a primitive but Sq^2 is not. This makes it much more difficult to analyze the action of the Steenrod algebra on $t\otimes t$ in Theorem 2. The second hindrance is that for mod two commutative Hopf algebras, the square of an odd degree element is not necessarily zero.

This presents technical difficulties in the analysis of primary obstructions to constructing maps $f: X \to Y$. The problem is described in detail in [12]. As a result of these difficulties, our knowledge of the mod two cohomology of a finite H-space is very limited. Perhaps the most significant work is due to Adams [1], Thomas [15] and Hubbuck [7], all restricted to H-spaces whose mod two cohomology is primitively generated.

In the remainder of this paper, we will present a few conjectures. Most of these conjectures that are well known appear too awesome to tackle. For this reason I include some "subconjectures" which I feel are interrelated and hopefully are more manageable.

I. *The Loop Space Conjecture.* $H_*(\Omega X; \mathbf{Z})$ is two torsion free. Recent work of Kane [9] is relevant. The basic goal is to eliminate all possible elements in $\sigma^* QH^{\mathrm{even}}(X; \mathbf{Z}_2)$. Work of Zabrodsky [16] on the "c" operation appears to be useful. There is a duality between Browder operations, the first Dyer–Lashof operation and Zabrodsky's c operation. For this reason I feel the following subconjectures about the commutativity of X and X are important:

Subconjecture 1. The Browder operation λ_1 vanishes on

$$H_*(\Omega X; \mathbf{Z}_2) \otimes H_*(\Omega X; \mathbf{Z}_2).$$

Subconjecture 2. The sub-Hopf algebra of $H_*(X; \mathbf{Z}_2)$ generated by the image of σ_* is commutative.

II. *The Generalized Thomas Conjecture.* In 1963 Emery Thomas [15] published a paper that described a recipe for the action of the Steenrod algebra on $H^*(X; \mathbf{Z}_2)$ in the case that $H^*(X; \mathbf{Z}_2)$ is primitively generated. It is very simple: Given an odd primitive x in degree n, then if $n+1=2^l+2^{l+1}k$, $k>0$, then $x=\mathrm{Sq}^{2^l}y$. Unfortunately this simple recipe is not true when $H^*(X; \mathbf{Z}_2)$ is not primitively generated; when $l=1$, this recipe does not hold for some of the exceptional Lie groups. However, the following is true for all known finite H-spaces:

Conjecture. For $l>1$, $QH^{2^l+2^{l+1}k-1}(X; \mathbf{Z}_2) \subseteq \text{image } \mathrm{Sq}^{2^l}$.

For $l=1$ we can write $2^l+2^{l+1}k-1$ in the form $2^r m+1$ where $r>1$. Then

$$QH^{2^r m+1}(X; \mathbf{Z}_2) \subseteq \operatorname{im} \mathrm{Sq}^2 + \sum_{s=1}^{r-1} \operatorname{im} \mathrm{Sq}^{2^{r-1}m} \cdots \mathrm{Sq}^{2^s m}\, \mathrm{Sq}^2.$$

III. *The Connectivity Conjecture.* The first nonvanishing homotopy group occurs in dimension 1, 3, 7 or 15.

Thomas [14] proves this in the primitively generated case.

Subconjecture 1. The Thomas conjecture implies the connectivity conjecture.

Subconjecture 2. The loop space conjecture implies the connectivity conjecture.

References

1. J. F. Adams, *On the non-existence of elements of Hopf invariant one*, Ann. of Math. **72** (1960), 20—104.

2. A. Borel, *Sur la cohomologie des espaces fibrés principaux et des espaces homogènes de groupes de Lie compacts*, Ann. of Math. **57** (1953), 115—207.

3. W. Browder, *Homology operations and loop spaces*, Illinois J. Math. **4** (1960), 347—357.

4. ——, *Torsion in H-spaces*, Ann. of Math. **74** (1961), 24—51.

5. H. Hopf, *Ueber die Topologie der Gruppen Mannigfaltigkeiten und ihrer Verallgemeinerungen*, Ann. of Math. **42** (1941), 22—52.

6. J. R. Hubbuck and R. Kane, *On π_3 of a finite H-space*, Trans. Amer. Math. Soc. **213** (1975), 99—105.

7. J. R. Hubbuck, *Polynomial algebras in cohomology*, notes.

8. R. Kane, *The module of indecomposables for finite H-spaces*, Trans. Amer. Math. Soc. **222** (1976), 303—318.

9. ——, *The Module of indecomposables for finite H-spaces*. II Trans. Amer. Math. Soc. (to appear).

10. T. Kudo and S. Araki, *Topology of H_n spaces and H squaring operations*, Mem. Fac. Sci. Kyushu Univ. Ser. A **10** (1956), 85—120.

11. J. Lin, *Steenrod squares in the* mod 2 *cohomology of an H-space* (to appear).

12. ——, *Torsion in H-spaces*. I, II, Ann. of Math. **103** (1976), 457—486; **107** (1978), 41—88.

13. J. Stasheff, *Homotopy associativity of H-spaces*. I, II, Trans. Amer. Math. Soc. **108** (1963), 275—292, 293—312.

14. P. E. Thomas, *On the* mod 2 *cohomology of certain H-spaces*, Comment. Math. Helv. **37** (1962/63), 132—140.

15. ——, *Steenrod squares and H-spaces*, Ann. of Math. **77** (1963), 306—317.

16. A. Zabrodsky, *Cohomology operations and homotopy commutative H-spaces*, The Steenrod Algebra and its Applications, Lecture Notes in Math., vol. 168, Springer-Verlag, Berlin and New York, pp. 308—317.

17. ——, *Secondary operations in the cohomology of H-spaces*, Illinois J. Math. **15** (1971), 648—655.

UNIVERSITY OF CALIFORNIA, SAN DIEGO
LA JOLLA, CALIFORNIA 92093, U.S.A

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Spherical Space Forms

Ib Madsen

Classically, a space form L^{n-1} is a complete Riemannian manifold with constant curvature. If the curvature is also positive L^{n-1} has finite fundamental group π, and the universal cover of L^{n-1} is the sphere S^{n-1} with its standard metric. Since the isometry group of S^{n-1} is the orthogonal group, L^{n-1} is the orbit space of a free orthogonal action of π on S^{n-1}. We call L^{n-1} an orthogonal space form.

Generally, a spherical space form $M^{n-1}(\pi)$ is a closed manifold with finite fundamental group π and universal cover S^{n-1}. We shall be primarily interested in the smooth case, where we require in addition to $M^{n-1}(\pi)$ being smooth that its universal cover is the standard smooth S^{n-1}. Thus smooth spherical space forms are orbit spaces of free representations of π in the group $\mathrm{Diff}(S^{n-1})$ of diffeomorphisms of S^{n-1}.

In this article we discuss the structure of smooth spherical space forms. This includes classification of the groups π which can occur, and a partial characterization of space forms $M^{n-1}(\pi)$ whose coverings $M^{n-1}(\gamma)$ are diffeomorphic to orthogonal space forms for a certain class of subgroups $\gamma \subset \pi$.

The work is part of a collaboration with C. B. Thomas and C. T. C. Wall. Other accounts of this collaboration can be found in [**6**], [7], [**14**].

Throughout the article space forms will be assumed oriented, hence odd dimensional, and of dimension at least 5.

1. A finite group is said to satisfy the *pq*-condition if each subgroup of order pq is cyclic. Here p and q are primes, not necessarily distinct. If a finite group π admits a free orthogonal representation (i.e. a representation $\varphi: \pi \to O(n)$ where $+1$ is not an eigenvalue for any $\varphi(g), g \neq 1$), then π must satisfy all

pq-conditions. These conditions are also sufficient if π is soluble, [15]. In contrast we have

THEOREM 1.1. *A finite group can act freely on a standard sphere S^{n-1} as a group of diffeomorphisms if and only if it satisfies all p^2- and $2p$-conditions.*

The necessity of the p^2- and $2p$-conditions even for topological actions is due to P. A. Smith and J. Milnor [9]. Special cases of 1.1 were proved earlier by T. Petrie and R. Lee.

Groups satisfying all p^2- and $2p$-conditions have been completely classified into six types, namely according to the structure of $\pi/O(\pi)$, where $O(\pi)$ is the maximal normal subgroup of odd order. The six types of $\pi/O(\pi)$ are: I. $Z/2^k$ (cyclic), II. $Q2^k$ (quaternion), III. T^* (binary tetrahedral), IV. O^* (binary octahedral), V. $Sl_2(F_p)$ and VI. $Tl_2(F_p)$, and $O(\pi)$ is itself metacyclic (cf. [6], [15]).

A group π satisfies all p^2-conditions if and only if it is *periodic* in the sense that there exists $d>0$, so that $H^i(B\pi; Z)\cong H^{i+d}(B\pi; Z)$ for $i>0$ [3]. The minimal such d is the period of π, denoted $d(\pi)$. If π acts freely on S^{n-1} then $n=rd(\pi)$ and we can ask which integers r can occur. From [14] we have

THEOREM 1.2. *One can take $n=rd(\pi)$ with any r such that $rd(\pi)\geqslant 6$, except for certain groups π of type* II, *where r must be even.*

For groups of type II it is in general difficult to decide whether r must necessarily be even, cf. [5] and 2.3 below.

The previous theorems contain little information about the nature of the asserted free representation of π in $\mathrm{Diff}(S^{n-1})$. One can seek to remedy this by attempting to set up 'Brauer induction'. Our next result is a step in this direction (see also 3.3. below).

Let $\mathscr{L}(\pi)$ consist of all 2-hyperelementary subgroups of π. Each element $\gamma\in\mathscr{L}(\pi)$ satisfies all pq-conditions and hence admits free orthogonal representations.

THEOREM 1.3. *Let π and n be as in* 1.2.

(i) *There exists a free representation $\Phi\colon \pi\to\mathrm{Diff}(S^{n-1})$ whose restriction to each $\gamma\in\mathscr{L}(\pi)$ is conjugate to an orthogonal free representation.*

(ii) *If the order of π contains no odd squares, then the conjugacy class of a Φ from* (i) *is determined by its restrictions $\Phi|\gamma$, $\gamma\in\mathscr{L}(\pi)$.*

For each group π in the classification list one can list explicitly at least one family of orthogonal representations $\{\Phi_\gamma|\gamma\in\mathscr{L}(\pi)\}$ which lifts to a free representation $\Phi\colon \pi\to\mathrm{Diff}(S^{n-1})$. At present, however, I do not know any explicit characterization of which families $\{\Phi_\gamma\}$ can occur. This requires knowledge of the image of $\mathrm{Res}\colon K_1(Z\pi)\to\prod\{K_1(Z_\gamma)|\gamma\in\mathscr{L}(\pi)\}$. The uniqueness in 1.3 (ii) is equivalent to the injectivity of Res. The only result I know in this direction is from [4].

The theorems above are listed in the language of representations, but their proofs proceed by constructing the associated space forms, following the usual scheme of surgery theory. This is done in the next two sections which together outline a proof of 1.3.

2. Our first task is to exhibit suitable finite cell complexes which may serve as the underlying simple homotopy types of the asserted space forms. The invariants one encounters are the Swan obstruction in $\tilde{K}_0(Z\pi)$ and the Reidemeister torsion in $K_1(Q\pi)$.

Let $d=d(\pi)$ be the period of π. Then $H^d(B\pi; Z)=Z/|\pi|$ and there is a one to one correspondence between generators e of $H^d(B\pi; Z)$ and homotopy types of (possibly infinite) cell complexes $X(\pi)$ with fundamental group π and universal cover identified with S^{n-1} up to homotopy. The class $e(X(\pi))$ is called the k-invariant; if $X(\pi)$ is an orthogonal space form, then e is the Euler class of the associated representation.

According to Swan **[11]**, e can be represented by a finite cell complex if and only if a certain element $\theta(e)\in\tilde{K}_0(Z\pi)$ vanishes.

Each homotopy type $X(\pi)$, corresponding to a generator with vanishing Swan obstruction, divides into a number of simple homotopy types $Y(\pi)$, in one to one correspondence with $\mathrm{Wh}(Z\pi)=K_1(Z\pi)/\{\pm g|g\in\pi\}$. Our use of surgery theory in the next section requires that we partly specify a simple homotopy type. To this end we can employ the Reidemeister torsion invariant $\Delta(Y(\pi))\in\mathrm{Wh}(Q\pi)$, **[10]**. It determines the simple homotopy type up to finite ambiguity, namely modulo the torsion subgroup of $\mathrm{Wh}(Z\pi)$, which according to Wall **[13]** is precisely $SK_1(Z\pi)$.

The complexes $Y(\pi)$ satisfy Poincaré duality, in fact they are weakly simple Poincaré duality spaces, **[13]**. For the arguments in § 3 we need $Y(\pi)$ to be an actual simple Poincaré duality space. The obstruction to achieve this is an element of $H^{\mathrm{odd}}(Z/2, SK_1(Z\pi))$. This group vanishes for p-hyperelementary groups with p odd.

Denote by $L(\pi; V)$ the orthogonal space form associated to the free representation V, $L(\pi; V)=S(V)/\pi$. If $Y(\pi)$ is as above, let $Y(\gamma)$ be the covering associated to the subgroup $\gamma\subset\pi$.

PROPOSITION 2.1. *Suppose π and n satisfy the requirements of* 1.3. *There exists a simple Poincaré duality space $Y(\pi)$ with $Y(\gamma)$ simple homotopy equivalent to certain orthogonal space forms $L(\gamma; V_\gamma)$, $\gamma\in\mathcal{L}(\pi)$.*

The proof of this proposition is quite involved, and a few comments are in order. For the special case where π is hyperelementary the argument is essentially given in [7], and is based on the ideas of Wall **[14**, Theorem 2**]**. The general case uses induction techniques. First, one can exhibit, for each (maximal) hyperelementary subgroup $H\subset\pi$ a simple Poincaré duality space $Y_H(H)$, and a family of n-dimensional representations V_γ, $\gamma\in\mathcal{L}(\pi)$, such that

$$
\text{(2.2.)}\quad
\begin{aligned}
&\text{(i)} && \{[V_\gamma]\}\in\mathrm{Image}\,\{\mathrm{Res}\colon RO(\pi)\to\textstyle\coprod\{RO(\gamma)|\gamma\in\mathcal{L}(\pi)\}\},\\
&\text{(ii)} && Y_H(H) = L(H; V_H) \quad\text{if}\quad H \text{ is 2-hyperelementary},\\
&\text{(iii)} && \Delta(Y_H(\gamma)) = \Delta(L(\gamma; V_\gamma)) \quad\text{for}\quad \gamma\in\mathcal{L}(H).
\end{aligned}
$$

Now, (i) implies a unique class $e \in H^n(B\pi; \mathbf{Z})$ which restricts to $e(L(\gamma; V_\gamma))$ for each $\gamma \in \mathscr{L}(\pi)$, and $\theta(e)=0$ since elements of $\tilde{K}_0(\mathbf{Z}\pi)$ are detected on the hyperelementary subgroups of π. Choose any simple homotopy type $Y'(\pi)$ realizing e. The distinct maximal hyperelementary subgroups of π intersect in elements of $\mathscr{L}(\pi)$. This together with the calculational fact that $\mathrm{Wh}(\mathbf{Z}[\mathbf{Z}/m])$ is torsion free, and hyperelementary induction for $\mathrm{Wh}(\mathbf{Z}\pi)$ implies that one may change $Y'(\pi)$ by an element of $\mathrm{Wh}(\mathbf{Z}\pi)$ to obtain the required $Y(\pi)$.

The 2-hyperelementary groups of type II have cohomological period 4, but the minimum degree of a free orthogonal representation is sometimes 8. This happens e.g. for the groups denoted $Q(8p, q, 1)$ in Milnor [**9**]. For such groups the Swan obstruction depends on intricate number theory. The recent advances here are due to Milgram [**8**]. For example,

PROPOSITION 2.3 (Milgram). (i) *Let* $p=3$ *and* q *any prime congruent to* 13 (*modulo* 24). *For each integer* $k \geqslant 0$ *there exist generators* $e_k \in H^{8k+4}(BQ(24; q, 1); \mathbf{Z})$ *with* $\theta(e_k)=0$.

(ii) *For* $p=3, q=5$, *any* $(8k+4)$*-dimensional generator has nonvanishing Swan obstruction.*

3. The general theory for converting a (simple) Poincaré duality space into a manifold has a purely cohomological part, namely the choices of normal invariant. To exploit this we set up induction techniques for generalized cohomology theories. Group cohomology as treated in [**2**] is a special case.

Suppose $Y(\pi)$ is a finite cell complex as in § 2, and let $E(-)$ be any cohomology theory, represented by an infinite loop space. Let $\mathscr{D}(\pi)$ be the category whose objects are all subgroups of π and whose morphisms are inclusions which extend to inner automorphisms of π. We can consider $E(Y(\gamma))$ as a bifunctor on $\mathscr{D}(\pi)$. The contravariant part is obvious, the covariant part is induced from the transfer.

Each cohomology theory is a module over the stable cohomotopy functor $\omega(-)$, represented by $\Omega^\infty S^\infty$. In fact, using the terminology from [3], $E(Y(\gamma))$ is a Mackey functor over the Green functor $\omega(Y(\gamma))$. The defect category for $\omega(Y(\gamma))$ is the subcategory $\mathscr{P}(\pi) \subset \mathscr{D}(\pi)$ consisting of all p-groups [7], and it becomes formal to deduce

PROPOSITION 3.1. *The covering projections* $i\colon Y(\gamma) \to Y(\pi)$ *induce an isomorphism* $\varprojlim i^*\colon E(Y(\pi)) \to \varprojlim \{E(Y(\gamma)) | \gamma \in \mathscr{P}(\pi)\}$.

The space $Y(\pi)$ has a Spivak fibration, classified by a map $Y(\pi) \to BSG$. A normal invariant is a homotopy class of liftings $Y(\pi) \to BSO$. According to Boardman–Vogt BSO and BSG are infinite loop spaces and the map $BSO \to BSG$ is an infinite loop map. Thus there is a fibration $BSO \to BSG \to B(G/O)$, where also $B(G/O)$ is an infinite loop space. The obstruction to the existence of a normal invariant is an element $\sigma(Y(\pi)) \in [Y(\pi), B(G/O)]$. But $\sigma(Y(\gamma))=0$ for $\gamma \in \mathscr{L}(\pi)$ by 2.1 and as $\mathscr{P}(\pi) \subset \mathscr{L}(\pi)$, $\sigma(Y(\pi))=0$.

After an arbitrary choice of normal invariant for $Y(\pi)$ the main exact sequence of surgery becomes (cf. [12])

$$L_n^s(\pi) \xrightarrow{\mu} \mathscr{S}^s(Y(\pi)) \to [Y(\pi), G/O] \xrightarrow{\lambda} L_{n-1}^s(\pi). \tag{3.2}$$

In [3] Dress proved that $L_*^s(\pi) \to^{\cong} \varprojlim \{L_*^s(\gamma) | \gamma \in \mathscr{L}(\pi)\}$. Thus three of the terms in 3.2 satisfy $\mathscr{L}(\pi)$-inductions and it is not surprising that we have

PROPOSITION 3.3. *Taking coverings define a bijection*

$$\mathscr{S}^s(Y(\pi)) \xrightarrow{\cong} \varprojlim \{\mathscr{S}^s(Y(\gamma)) \,|\, \gamma \in \mathscr{L}(\pi)\}.$$

The proof is not just a 5-lemma argument since $\mathfrak{s}^s(Y(\pi))$ is not a group. It uses the partial splitting of the map μ provided by the ϱ-invariant (cf. [12]), and the calculational facts that $L_0^s(\gamma)$ is torsion free when γ is cyclic or the quaternion group of order 8.

Theorem 1.3 follows from 2.1 and 3.3. Indeed, the simple homotopy equivalences of 2.1 define an element of $\varprojlim \{\mathscr{S}^s(Y(\gamma)) | \gamma \in \mathscr{L}(\pi)\}$, and 3.3 gives an element of $\mathscr{S}^s(Y(\pi))$ whose domain is the required manifold.

We end with an obvious open question. Do the generators e_k in 2.3 (i) represent spherical space forms?

References

1. J. Boardman and R. Vogt, *Homotopy everything H-spaces*, Bull. Amer. Math. Soc. 74 (1968) 1117—1122.

2. H. Cartan and S. Eilenberg, *Homological algebra*, Princeton Univ. Press, Princeton, N. J., 1956.

3. A. Dress, *Induction and structure theorems for orthogonal representations of finite groups*, Ann. of Math. **102** (1975), 291—326.

4. M. E. Keating, *Whitehead groups of some metacylic groups and orders*, J. Algebra **22** (1972), 332—350.

5. R. Lee, *Semicharacteristic classes*, Topology **12** (1973), 183—199.

6. I. Madsen, C. B. Thomas and C. T. C. Wall, *The topological spherical space form problem* — II, Topology **15** (1976), 375—382.

7. I. Madsen, *Smooth spherical space forms*, Proceedings, Evanston 1977, Lecture Notes in Math., vol. 657, Springer-Verlag, Berlin and New York, 1978.

8. R. J. Milgram, *The Swan obstruction* (preprint).

9. J. Milnor, *Groups which operate on S^n without fixed points*, Amer. J. Math. **79** (1957), 623—630.

10. *Whitehead torsion*, Bull. Amer. Math. Soc. **72** (1966), 358—426.

11. R. G. Swan, *Periodic resolutions for finite groups*, Ann. of Math. **72** (1960), 552—578.

12. C. T. C. Wall, *Surgery on compact manifolds*, Academic Press, London, 1970.

13. *Norm of units in group rings*, Proc. London Math. Soc. **29** (1974), 593—632.

14. *Free actions of finite groups on spheres*, Proc. Sympos. Pure Math., vol. 32, Amer. Math. Soc., Providence, R. I., 1977.

15. J. A. Wolf, *Spaces of constant curvature*, McGraw-Hill, New York, 1967.

AARHUS UNIVERSITY
(8000) AARHUS (C), DENMARK

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Shape Theory

Sibe Mardešić*

Dedicated to Professor Karol Borsuk

1. Introduction. Shape theory is a new area of topology whose aim is the same as that of homotopy theory, i.e. to study the global behavior of spaces. However, ite tools are different and are applicable to rather general spaces. This is not the cas- with the standard tools of homotopy theory, which are designed primarily for studying such locally nice spaces as CW-complexes and ANR's. E.g., all homotopy groups of the Warsaw circle or of the dyadic solenoid vanish, although these spaces are not contractible. Realizing this, K. Borsuk undertook the task of developing *shape theory*, a modification of homotopy theory, which agrees with homotopy theory on nice spaces, but yields relevant information even when applied to such general spaces as arbitrary metric compacta. Such spaces appear naturally, e.g., as fibres of maps between nice spaces, and therefore cannot be ignored.

Borsuk's ideas on shape proved to have a bearing on several areas of topology, especially geometric topology, and to have also applications outside of topology. Borsuk's work has triggered an avalanche of research resulting in over four hundred papers written since 1968, when his original paper appeared. It seems therefore appropriate to attempt to survey here the development of shape theory over the past ten years.

Clearly, limitation of space prevents any extensive analysis as well as full attribus tion of results to all the authors who made relevant contributions. We merely list some major areas and illustrate their nature by a few key theorems. In many cases

* This paper was prepared while the author was visiting the Department of Mathematics, University of Kentucky, Lexington, Kentucky, U.S.A.

these are not the best results known. Not being able to include a relatively complete bibliography, we just name authors. The reader will easily trace their papers using the extensive shape theory bibliography, compiled by J. Segal, University of Washington, Seattle.

2. The shape category. Originally, Borsuk considered compact metric spaces X, Y embedded in the Hilbert cube Q. Instead of homotopy classes of continuous maps $f: X \to Y$, he considered homotopy classes of certain sequences of maps $f_n: Q \to Q$ called *fundamental sequences*. By definition, for every neighborhood V of Y, there is an integer n_V such that all $f_n, n \geq n_V$, map some neighborhood U of X into V and all $f_n|U, n \geq n_V$, are homotopic in V. Compacta in Q and homotopy classes of fundamental sequences form Borsuk's *shape category* $\mathcal{Sh}$. Compacta X and Y have the same *shape*, $\operatorname{sh} X = \operatorname{sh} Y$, provided they are isomorphic objects of $\mathcal{Sh}$. Similarly, X is *shape dominated* by Y, $\operatorname{sh} X \leq \operatorname{sh} Y$, provided there exist shape morphisms $f: X \to Y$ and $g: Y \to X$ such that $gf = 1$ in $\mathcal{Sh}$.

Subsequently, the shape category $\mathcal{Sh}$ has been extended to more general spaces by several authors including S. Mardešić and J. Segal, R. H. Fox, S. Mardešić G. Kozlowski, J. Le Van and C. Weber, and to other categories by W. Holsztyńskr T. Porter, A. Deleanu and P. Hilton. One should also mention that a 1944 pape by D. E. Christie contains already several ideas of shape theory.

In particular, in 1971 Mardešić and Segal described the shape category $\mathcal{Sh}$ for compact Hausdorff spaces using *inverse systems* of compact ANR's over *cofinite directed* sets. If $\mathbf{X} = (X_\lambda, p_{\lambda\lambda'}, \Lambda)$ and $\mathbf{Y} = (Y_\mu, q_{\mu\mu'}, M)$ are such systems and $X = \lim \mathbf{X}$, $Y = \lim \mathbf{Y}$, then a shape morphism $X \to Y$ is given by a homotopy class of maps of systems $(f_\mu, \varphi): \mathbf{X} \to \mathbf{Y}$, i.e. by an increasing function $\varphi: M \to \Lambda$ and maps $f_\mu: X_{\varphi(\mu)} \to Y_\mu$ satisfying $f_\mu p_{\varphi(\mu)\varphi(\mu')} \simeq q_{\mu\mu'} f_{\mu'}$ whenever $\mu \leq \mu'$. By definition, $(f_\mu, \varphi) \simeq (f'_\mu, \varphi')$ if each μ admits a $\lambda \geq \varphi(\mu), \varphi'(\mu)$ such that $f_\mu p_{\varphi(\mu)\lambda} \simeq f'_\mu p_{\varphi'(\mu)\lambda}$.

A further generalization of this approach to shape for arbitrary topological spaces was given by K. Morita in 1975. With every topological space X one can *associate* certain inverse systems $\mathbf{X}$ in the homotopy category $\mathcal{HCW}$ of spaces having the homotopy type of CW-complexes (or equivalently of ANR's for metric spaces). Such a system is the *Čech system* based on *numerable* coverings. The set of shape morphisms $\mathcal{Sh}(X, Y)$ can be interpreted as the set of morphisms $\mathbf{X} \to \mathbf{Y}$ in the *pro-category* pro-$\mathcal{HCW}$ of A. Grothendieck, i.e. as $\lim_\mu \operatorname{colim}_\lambda [X_\lambda, Y_\mu]$. Several useful results on pro-categories were developed by M. Artin and B. Mazur in their etale homotopy theory (1969).

If the spaces X, Y have the homotopy type of CW-complexes, (or equivalently of ANR's), then one can take for $\mathbf{X}$ and $\mathbf{Y}$ systems which consist of single terms X and Y respectively. Consequently, $\mathcal{Sh}(X, Y) = [X, Y]$, i.e. on such spaces shape coincides with homotopy. Another interesting case when shape yields nothing new is given by the category of compact connected Abelian groups. Their shape morphisms are in a one-to-one correspondence with group homomorphisms. This result, due to J. Keesling, has proved very useful in constructing various counter-examples in shape theory.

3. Homotopy and homology pro-groups. In shape theory the role of the homotopy groups π_n is taken by the *homotopy pro-groups* pro-π_n. If (X, x) is a pointed $\mathcal{HCW}$-system associated with (X, x), then pro-$\pi_n(X, x)$ is the inverse system of groups $(\pi_n(X_\lambda, x_\lambda), p_{\lambda\lambda'*}, \Lambda)$. The corresponding inverse limit $\check{\pi}_n(X, x)$ is the nth *shape group* of (X, x). For nice spaces, e.g. for ANR's, one can replace the homotopy pro-groups by the shape groups. However, in general much information is lost by passing to the limit and one must consider the whole pro-group as a new and important shape invariant.

This point of view is well illustrated by the following *Whitehead theorem* (K. Morita, [26]): Let $f: (X, x) \to (Y, y)$ be a shape morphism of pointed connected finite-dimensional topological spaces (covering dimension based on numerable coverings). If f induces isomorphisms of homotopy pro-groups in all dimensions, then f is a shape equivalence. The first result of this type was established by M. Moszyńska [27] and improved by Mardešić and by Morita. D. A. Edwards and R. Geoghegan and also J. Dydak have obtained several theorems of this type weakening the condition that both X and Y be finite-dimensional. However, this condition cannot be completely omitted. Indeed, D. Handel and J. Segal have shown that a certain continuum constructed by D. S. Kahn has non-trivial shape although all of its homotopy pro-groups vanish. Kahn's example, just as some other counter-examples in shape theory, depend on J. F. Adams' map $A: \Sigma^{2r} Y \to Y$ from the $2r$-fold suspension of a certain finite CW-complex Y to Y. The crucial property of A, established using K-theory, is that for all s, the composition

$$A \circ \Sigma^{2r} A \circ \ldots \circ \Sigma^{2r(s-1)} A: \Sigma^{2rs} Y \to Y$$

is an essential map. Alternatively, one can use results of H. Toda.

Some other standard theorems on homotopy groups also carry over to shape theory and homotopy pro-groups. E.g., this is the case with the van Kampen theorem (A. Kadlof, Š. Ungar) and the Blakers–Massey theorem (Š. Ungar).

Homology pro-groups are defined in an analogous way and their limits are the *Čech homology groups.* In distinction from homology groups, homology pro-groups are exact under very general assumptions. Several versions of the *Hurewicz theorem* in shape theory have been proved (K. Kuperberg, T. Porter, S. Mardešić and Š. Ungar, K. Morita, Š. Ungar, J. Dydak, T. Watanabe).

4. Stability theorems. An important question in shape theory is to decide when is a pointed space (X, x) *stable,* i.e. has the pointed shape of a pointed CW-complex.

L. Demers (1975) and Edwards and Geoghegan [9] have shown that a pointed connected space (X, x) is stable if and only if it is pointed shape dominated by a pointed CW-complex. Furthermore, Edwards and Geoghegan have obtained the following *algebraic stability criterion:* A pointed connected space (X, x) of finite shape dimension is pointed stable if and only if each of its homotopy pro-groups pro-$\pi_n(X, x)$ is stable, i.e. is isomorphic as a pro-group to a group (which has to be the shape group $\check{\pi}_n(X, x)$).

(Pointed) metric compacta (pointed) shape dominated by (pointed) CW-complexes are called *(pointed) FANR's* and are the shape theoretic analogues of ANR's. Pointed FANR's are stable, i.e. have the shape of ANR's However, this is not known in the unpointed case. Are there FANR's which are not pointed FANR's, has proved to be a very delicate question. It is known that a connected FANR X is a pointed FANR if and only if the first derived limit $\lim^1$ pro-$\pi_1(X, x)$ vanishes (this does not depend on the choice of $x \in X$). Some group-theoretic results of J. Dydak and P. Minc indicate that FANR's might prove to be different from pointed FANR's.

With every pointed FANR (X, x) Edwards and Geoghegan have associated an intrinsically defined *Wall obstruction* $w(X, x)$, which takes its values in the reduced projective class group $\tilde{K}^0(\check{\pi}_1(X, x))$ of the first shape group. The vanishing of $w(X, x)$ is a necessary and sufficient condition for X to have the shape of a finite CW-complex (or equivalently of a compact ANR). All possible obstructions occur among 2-dimensional FANR's.

5. Movability. One of the most interesting new concepts, which originated in shape theory is the notion of (pointed) movability. A system $\mathbf{X}=(X_\lambda, p_{\lambda\lambda'}, \Lambda)$ in $\mathcal{HCW}$ is said to be *movable* provided each $\lambda \in \Lambda$ admits a $\lambda' \geqslant \lambda$ such that for any CW-complex K, any homotopy class of maps $f: K \to X_{\lambda'}$, and any $\lambda'' \geqslant \lambda$ there is a homotopy class of maps $f': K \to X_{\lambda''}$ such that $p_{\lambda\lambda''} f' = p_{\lambda\lambda'} f$. $\mathbf{X}$ is *n-movable* if this holds for complexes K with $\dim K \leqslant n$. A space X is movable (n-movable) if the associated systems $\mathbf{X}$ are movable (n-movable). Movability has proved to be especially useful in the case of metric compacta X, where the vanishing of the shape group $\check{\pi}_i(X, x)=0$ implies the vanishing of the homotopy pro-group pro-$\pi_i(X, x)=0$. For metric continua (X, x), (Y, y), which are pointed movable, the Whitehead theorem assumes this simple form: If X and Y are finite-dimensional and $f: (X, x) \to (Y, y)$ induces an isomorphism of shape groups, then f is a shape equivalence (J. Keesling, [16]). Let us also mention that pointed connected FANR's are characterized as pointed movable continua (X, x) having finite shape dimension and countable shape groups $\check{\pi}_n(X, x)$ (J. Dydak, T. Watanabe).

J. Keesling has obtained interesting results concerning integral Čech cohomology groups $\check{H}^n$ of movable compacta X. In particular, he has proved that $\check{H}^n(X)/\mathrm{Tor}\check{H}^n(X)$ is an $\aleph_1$-free Abelian group, i.e. each of its countable subgroups is free Abelian. Since metric LC^{n-1} continua (of dimension $\leqslant n$) are always n-movable (movable)

(Mardešić, Borsuk, Kozlowski and Segal), the above results have provided new information concerning cohomology of locally connected continua.

6. Shape dimension. A space X is said to have *shape dimension* (also called deformation dimension) $\mathrm{Sd}\, X \leqslant n$ provided X admits an associated system $\boldsymbol{X}$ whose members X_λ are CW-complexes of dimension $\leqslant n$ (Dydak). It is readily seen that $\mathrm{Sd}\, X \leqslant \dim X$, where dim is the covering dimension based on numerable coverings of X. Moreover, $\mathrm{sh}\, X \leqslant \mathrm{sh}\, Y$ implies $\mathrm{Sd}\, X \leqslant \mathrm{Sd}\, Y$. The main contributions concerning shape dimension are due to S. Nowak. For metric continua X, which are shape 1-connected, i.e. pro-$\pi_1(X, x)=0$ for all $x \in X$, and for which $\dim X < \infty$ (or more generally $\mathrm{Sd}\, X < \infty$), Nowak has characterized $\mathrm{Sd}\, X$ as the greatest integer n such that $\check{H}^n(X) \neq 0$. He has also given a characterization theorem for the case when pro-$\pi_1(X, x) \neq 0$. Instead of integer coefficients one has to use local systems of groups on members of polyhedral expansions $\boldsymbol{X}$ of X.

7. Shapes of Z-sets in the Hilbert cube. In [4] T. A. Chapman discovered a profound relationship between shape theory and infinite-dimensional topology. He associated with every Z-set X in the Hilbert cube Q its complement $Q \setminus X$ and with every shape morphisms of Z-sets $f\colon X \to Y$ a class of *weakly properly homotopic* proper maps $Q \setminus X \to Q \setminus Y$ in such a manner that one obtains an isomorphism of categories. Using this isomorphism one can translate notions and problems from shape theory into notions and problems concerning weak proper homotopy of separable locally compact spaces. This program has been pursued and extended by Z. Čerin.

Chapman has also proved this remarkable *complementation theorem:* Two Z-sets $X, Y \subseteq Q$ have the same shape if and only if their complements are homeomorphic. Chapman's proof has been simplified by L. Siebenmann.

Stimulated by Chapman's work, D. A. Edwards and H. M. Hastings have introduced a *strong (Steenrod) shape theory,* which for Z-sets in Q transforms under complementation into the more familiar *proper homotopy theory.* Their approach consists in first organizing inverse systems of semisimplicial complexes and maps of such systems in pro-SS into a closed model category in the sense of D. Quillen (in a way different from the one used by A. K. Bousfield and D. M. Kan). Then, they invert the weak equivalences (in the sense of the calculus of fractions) and obtain a homotopy category Ho (pro-SS) of such systems. With every space X they associate its *Vietoris system* $V(X)$. Morphisms $V(X) \to V(Y)$ in Ho (pro-SS) are the strong shape morphisms $X \to Y$. For compact metric spaces a purely geometric description of strong shape is based on the *contractible telescope* $\mathrm{C\,Tel}\, \boldsymbol{X}$ of an inverse sequence $\boldsymbol{X}$ of compact ANR's. A strong shape morphism $f\colon X \to Y$ between metric compacta is just a proper homotopy class of proper maps $f\colon \mathrm{C\,Tel}\, \boldsymbol{X} \to \mathrm{C\,Tel}\, \boldsymbol{Y}$ where $X = \lim \boldsymbol{X}$, $Y = \lim \boldsymbol{Y}$. A systematic study of strong shape of metric compacta, avoiding Quillen's theory, was carried out by Dydak and Segal. Further contributions to strong shape were made by T. Porter, by Y. Kodama and J. Ono, by F. W. Bauer, by Yu. T. Lisica and by A. Calder and H. Hastings.

Beside ordinary shape and strong shape one should mention at least one more shape theory. This is *proper shape* for locally compact metric spaces, developed by B. J. Ball and R. B. Sher in 1973. Proper shape is the shape analogue of proper homotopy. Among other results Ball and Sher have proved that a proper shape equivalence $X \to Y$ induces a shape equivalence $FX \to FY$ between the *Freudenthal compactifications,* provided these are metrizable. This equivalence reduces to a homeomorphism $EX \to EY$ on the sets of *ends*.

8. Finite-dimensional complementation theorems. These are theorems having the following form: If X and Y are compacta of dimension $\leqslant k$ embedded in a nice way in R^n, $n \geqslant 2k+2$, $n \geqslant 5$, then $\mathrm{sh}\, X = \mathrm{sh}\, Y$ if and only if $R^n \setminus X$ and $R^n \setminus Y$ are homeomorphic. The first theorem of this type was proved by Chapman. R. Geoghegan and R. Summerhill have improved Chapman's result. Their requirement is that $R^n \setminus X$ and $R^n \setminus Y$ be 1-ULC. This result was further improved, first by D. Coram, R. Daverman and P. Duvall, and then by J. Hollingsworth and B. Rushing. They have replaced the condition 1-ULC by the small loop condition (SLC). Finally, replacing SLC by the inessential loop condition (ILC), G. Venema has proved the complementation theorem under the weaker assumption that $\mathrm{Sd}\, X \leqslant k$, $\mathrm{Sd}\, Y \leqslant k$. We recall that $X \subseteq R^n$ has ILC provided each neighborhood U of X admits a smaller neighborhood V of X such that every loop in $V \setminus X$, which is null-homotopic in V, is also null-homotopic in $U \setminus X$. Some of these theorems were preceded or supplemented by results dealing with special cases when Y is S^k or a manifold (Duvall, Siebenmann, Daverman, Hollingsworth and Rushing, V. T. Liem).

A related problem is the problem of embedding a compactum X in R^n up to shape, i.e. of finding a compactum $Y \subseteq R^n$ such that $\mathrm{sh}\, X = \mathrm{sh}\, Y$. The first results of this type were obtained by I. Ivanšić. An extensive study of this problem was carried out by L. S. Husch and I. Ivanšić. Here is one of their results. Let X be an r-shape connected metric continuum with $\mathrm{Sd}\, X = k \geqslant 3$. If X is pointed $(r+1)$-movable, then it embeds up to shape in R^{2k-r}. This generalizes a result of J. Stallings dealing with embedding of polyhedra up to homotopy. Duvall and Husch have recently exhibited a k-dimensional continuum which does not embed up to shape in R^{2k}, $k = 2^m$, $m > 1$.

Recently, A. Kadlof has answered in the negative the following problem of Borsuk. Let $\mathrm{sh}\, X \leqslant \mathrm{sh}\, Y$, $Y \subseteq R^n$, does X embed in R^n up to shape? Husch and Ivanšić give an affirmative answer under rather restrictive additional assumptions.

9. Cell-like maps. A map $f\colon X \to Y$ between metric compacta is *cell-like* (or CE-map) provided each fibre $f^{-1}(y)$, $y \in Y$, has the shape of a point. These maps have proved to be extremely important in geometric topology.[1] E.g., J. West has shown that every compact ANR Y is the image of a compact Q-manifold X under a CE-map. If X and Y are finite-dimensional (or arbitrary ANR's) then a CE-map $f: X \to Y$ is a shape equivalence. However, in general cell-like maps fail to be shape equivalences. The first counter-example was exhibited by J. L. Taylor (using the Adams map). G. Kozlowski has shown that Y is an ANR provided X is an ANR and $f: X \to Y$ is a *hereditary shape equivalence.* An example of J. Keesling shows that one cannot replace this condition by the weaker condition that f be a CE-map.

Siebenmann and Chapman have proved that CE-maps between n-manifolds, $n \geqslant 5$, and Q-manifolds respectively, can be approximated by homeomorphisms, i.e. are near-homeomorphisms.

10. Approximate fibrations and shape fibrations. Motivated by R. C. Lacher's work on cell-like maps, D. Coram and P. Duvall have introduced *approximate fibrations.* These are maps $p: E \to P$ between compact ANR's satisfying an *approximate homotopy lifting property (AHLP).* Cell-like maps between ANR's are always approximate fibrations. Many properties of fibrations are also properties of approximate fibrations. In particular, one has an exact homotopy sequence where the homotopy groups $\pi_n(F, e)$ of the fibre must be replaced by the shape groups $\check{\pi}_n(F, e)$.

Uniform limits of fibrations between compact ANR's are readily seen to be approximate fibrations. However, the converse is not true. Husch has considered approximate fibrations $p: M \to S^1$, where M is a closed connected n-manifold, $n \geqslant 6$, and he has proved that p is a uniform limit of fibrations if and only if the Siebenmann obstruction $F(M)$ in the Whitehead group $\mathrm{Wh}(\pi_1(M))$ vanishes. Recently, F. Quinn and Chapman have obtained interesting results concerning the problem of approximating approximate fibrations by block bundles. Further contributions to approximate fibrations were made by S. Ferry and R. E. Goad.

Recently, Mardešić and Rushing have generalized approximate fibrations to a class of maps between metric compacta called *shape fibrations.* Even in this generality one has a homotopy exact sequence. However, all homotopy groups must be replaced by homotopy pro-groups. Cell-like maps between finite-dimensional compacta, and more generally, hereditary shape equivalences between arbitrary metric compacta, are shape fibrations. The Taylor CE-map is not a shape fibration. However, every cell-like map is a *weak shape fibration.* Further contributions to shape fibrations were made by T. McMillan, A. Matsumoto, M. Jani and Z. Čerin.

11. Shape theoretic properties of the Stone-Čech compactification. Recently, A. Calder and J. Siegel have obtained important new information concerning the homotopy classification of maps from the Stone–Čech compactification βX of

[1] Cf. addresses to this Congress by R. D. Edwards and by J. West. In view of their papers there is no need here for a more detailed exposition concerning CE-maps.

a normal finite-dimensional connected space X into a finite polyhedron P. In particular, they have shown that $[\beta X, P]=[X, P]$ provided $\pi_1(P)$ is finite. However, if $\pi_1(P)$ is infinite and X fails to be pseudocompact, then there is no bijection between $[\beta X, P]$ and $[X, P]$. Stimulated by this work, Keesling and Sher have obtained a series of shape theoretic results on βX, which shed new light on the geometric structure of βX. E.g., Keesling has proved that if X is Lindelöf and $K \subseteq \beta X \setminus X$ is a continuum, then $\operatorname{sh} K=0$ implies that K is a point. More generally, $\operatorname{Sd} K=\dim K$. If X is real compact, if $K \subseteq \beta X \setminus X$ is a continuum and if $f\colon K \to Y$ is a surjection, which induces an isomorphism of $\check{H}^1$, then f must be a homeomorphism.

References

1. B. J. Ball and R. B. Sher, *A theory of proper shape for locally compact metric spaces,* Fund. Math. **86** (1974), 163—192.

2. K. Borsuk, *Concerning homotopy properties of compacta,* Fund. Math. **62** (1968), 223—254.

3. ———, *Theory of shape,* Monografie Mat., vol. 59, Polish Science Publ., Warszawa, 1975.

4. T. A. Chapman, *On some applications of infinite-dimensional manifolds to the theory of shape,* Fund. Math. **76** (1972), 181—193.

5. D. Coram and P. Duvall, *Approximate fibrations,* Rocky Mountain J. Math. **7** (1977), 275—288.

6. J. Dydak, *A simple proof that pointed FANR-spaces are regular fundamental retracts of ANR's,* Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **25** (1977), 55—62.

7. ———, *The Whitehead and Smale theorems in shape theory,* Dissertationes Math., Warszawa (to appear).

8. J. Dydak and J. Segal, *Strong shape theory* Dissertationes Math., Warszawa (to appear).

8a. ———, *Shape theory,* Lecture Notes in Math., vol. 688, Springer-Verlag, Berlin and New York, 1978.

9. D. A. Edwards and R. Geoghegan, *Shapes of complexes, ends of manifolds, homotopy limits and the Wall obstruction,* Ann. of Math. **101** (1975), 521—535; Correction **104** (1976), 389.

10. ———, *Stability theorems in shape and pro-homotopy,* Trans. Amer. Math. Soc. **222** (1976), 389—403.

11. D. A. Edwards and H. M. Hastings, *Čech and Steenrod homotopy theories with applications to geometric topology,* Lecture Notes in Math., vol. 542, Springer-Verlag, Berlin and New York, 1976.

12. D. Handel and J. Segal, *An acyclic continuum with non-movable suspension,* Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **21** (1973), 171—172.

13. J. Hollingsworth and T. B. Rushing, *Embeddings of shape classes of compacta,* Pacific J. Math. **60** (1975), 103—110.

14. L. S. Husch and I. Ivanšić, *Embeddings and concordances of embeddings up to shape,* Acta. Math. (to appear).

15. J. Keesling, *An algebraic property of the Čech cohomology groups which prevents local connectivity and movability,* Trans. Amer. Math. Soc. **190** (1974), 151—162.

16. ———, *On the Whitehead theorem in shape theory,* Fund. Math. **92** (1976), 147—253.

17. J. Keesling and R. B. Sher, *Shape property of the Stone-Čech compactification,* General Topology and Appl. **9** (1978), 1—8.

18. G. Kozlowski and J. Segal, *Locally well-behaved paracompacta in shape theory,* Fund. Math. **95** (1977), 55—71.

19. S. Mardešić, *Shapes for topological spaces,* General Topology and Appl. **3** (1973), 265—282.

20. ——, *On the Whitehead theorem in shape theory*, I, Fund. Math. **91** (1976), 51—64.

21. S. Mardešić and T. B. Rushing, *Shape fibrations*, I, General Topology and Appl. **9** (1978), 193—215.

22. S. Mardešić and J. Segal, *Movable compacta and ANR-systems*, Bull. Acad. Polon. Sci. Sér. Sci. Math. Astronom. Phys. **18** (1970), 649—654.

23. ——, *Shapes of compacta and ANR-systems*, Fund. Math. **72** (1971), 41—59.

24. S. Mardešić and Š. Ungar, *The relative Hurewicz theorem in shape theory*, Glasnik Mat. **9** (1974), 317—328.

25. K. Morita, *On shapes of topological spaces*, Fund. Math. **86** (1975), 251—259.

26. ——, *The Hurewicz and the Whitehead theorems in shape theory*, Sci. Rep. Tokyo Kyoiku Daigaku **12** (1974), 246—258.

27. M. Moszyńska, *The Whitehead theorem in the theory of shapes*, Fund. Math. **80** (1973), 221—263.

28. S. Nowak, *On the fundamental dimension of approximately* 1-*connected compacta*, Fund. Math. **89** (1975), 61—79.

29. J. L. Taylor, *A counterexample in shape theory*, Bull. Amer. Math. Soc. **81** (1975), 629—632.

30. G. A. Venema, *Embeddings of compacta with shape dimension in the trivial range*, Proc. Amer. Math. Soc. **55** (1976), 443—448.

University of Zagreb
Zagreb, Yugoslavia

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Complex Cobordism and its Applications to Homotopy Theory

Douglas C. Ravenel*

In the past few years, the application of complex cobordism to problems in homotopy theory through the medium of the Adams–Novikov spectral sequence has become a lucrative enterprise. We will give a brief survey of some of the foundations and results of this theory, offering nothing new for the experts. See [9] for a more detailed account, including references for some of the statements made here.

The history of the subject begins with Thom's definition [10] of cobordism. Roughly speaking, 2 closed manifolds are *cobordant* if their disjoint union is the boundary of a third manifold. In the complex case, we require that these manifolds possess compatible complex structures on their stable tangent bundles. Cobordism is easily seen to be an equivalence relation and the set of equivalence classes is a ring *(the complex cobordsim ring MU_*)* under disjoint union and Cartesian product. Thom proved that this ring is canonically isomorphic to the homotopy of the complex Thom spectrum MU. Milnor [5] and Novikov [6] showed that $MU_*=\pi_*MU=Z[x_1, x_2, \ldots]$ where $\dim x_i=2i$. Brown–Peterson [3] showed that when localized at a prime p, MU splits into an infinite wedge of isomorphic summands known as BP with $\pi_*BP=BP_*=Z_{(p)}[x_{p^i-1}]$.

Since homotopy theory is essentially a local (in the arithmetic sense) subject we shall concern ourselves primarily with the smaller spectrum BP. Once its basic properties have been established, its relation to complex manifolds becomes irrelevant to the applications. Our understanding of these properties rests on a remarkable observation due to Quillen.

* Partially supported by the National Science Foundation of the United States.

Let $MU^*(\ )$ be the generalized cohomology theory represented by the spectrum MU. Then $MU^*(CP^\infty)=MU^*[[x]]$ where $x\in MU^2(CP^\infty)$ and MU^* is the coefficient ring $\pi_* MU$ negatively graded. We also have $MU^*(CP^\infty\times CP^\infty)=MU^*[[x\otimes 1, 1\otimes x]]$ and the tensor product (of complex line bundles) map $f\colon CP^\infty\times CP^\infty\to CP^\infty$ induces $f^*\colon MU^*(CP^\infty)\to MU^*(CP^\infty\times CP^\infty)$ with $f^*(x)=F(x\otimes 1, 1\otimes x)=\sum a_{ij}x^i\otimes x^j$ with $a_{ij}\in MU^{2(1-i-j)}$. The 2-variable power series F has 3 obvious properties: $F(x,0)=F(0,x)=x$ (identity); $F(x,y)=F(y,x)$ (commutativity); and $F(F(x,y),z)=F(x,F(y,z))$ (associativity). We define a *formal group law* G over a commutative ring R to be a power series $G(x,y)\in R[[x,y]]$ having the three properties of F. Quillen's observation was

THEOREM 1 [8]. *The formal group law F over MU^* is universal in the sense that for any other formal group law G over R, there is a unique ring homomorphism $\theta\colon MU^*\to R$ such that* $G(x,y)=\sum\theta(a_{ij})x^iy^j$. □

THEOREM 2 [8]. *There is a map $\varepsilon\colon MU_*\to BP_*$ such that any formal group law G over a $Z_{(p)}$-algebra R is canonically isomorphic to a formal group law G' induced by $\theta'\varepsilon$ where $\theta'\colon BP_*\to R$ (i.e. there is a power series $f(x)\in R[[x]]$ with leading term x such that $f(G(x,y))=G'(f(x),f(y))$).* □

Quillen was able to use these results to determine the structure of BP^*BP, the algebra of cohomology operations for the theory represented by the spectrum BP. This algebra, the BP analogue of the Steenrod algebra, is difficult to work with because it does not have finite type and cannot be readily described in terms of generators and relations. Instead we will describe its dual $BP_*BP=\pi_*BP\wedge BP$, the analogue of the dual Steenrod algebra.

First, we described the formal group law εF, which we will denote simply by F. Define $\log x\in(Q\otimes BP_*)[[x]]$ by $\log x=\sum_{i\geq 0} l_i x^{p^i}$ where $l_i=\varepsilon[CP^{p^i-1}]/p^i$. Then $F(x,y)$ is defined by

$$\log F(x,y) = F(\log x, \log y). \tag{3}$$

THEOREM 4 ([8], [1]). *As an algebra $BP_*BP=BP_*[t_1, t_2, \ldots]$ with $\dim t_i=2(p^i-1)$. The Hurewicz or right unit map $\eta_R\colon BP_*\to BP_*BP$ (induced by $BP=S^0\wedge BP\to BP\wedge BP$) is given over Q by*

$$\eta_R l_n = \sum l_i t_{n-i}^{p^i}. \tag{5}$$

This map defines a right BP_*-module structure on BP_*BP and the coproduct (dual to composition of cohomology operations) is a map $\Delta\colon BP_*BP\to BP_*BP\otimes_{BP_*}BP_*BP$ defined over Q by

$$\sum_{i\geq 0}\log\Delta(t_i) = \sum_{i,j\geq 0}\log(t_i\otimes t_j^{p^i}) \tag{6}$$

where $t_0=1$. □

The lack of a more explicit formula for $\Delta(t_i)$ was for some time a psychological obstruction to computing with BP. (6) can be rewritten as

$$\sum^F \Delta(t_i) = \sum^F t_i \otimes t_j^{p^i}, \tag{7}$$

(where $\log(\sum^F x_i) = \sum \log x_i$, i.e. $\sum^F x_i$ is the formal sum of the x_i), but this is of little help due to the complexity of F. Another difficulty is that the elements $p^i l_i = \varepsilon[CP^{p^i-1}]$ do not generate BP_*. This problem was surmounted first by Hazewinkel and later by Araki.

THEOREM 8 (ARAKI). *$BP_* = Z_{(p)}[v_1, v_2, \ldots]$ where v_n is defined by $pl_n = \sum_{0 \leq i \leq n} l_i v_{n-i}^{p^i}$ with $v_0 = p$.* $\square$

THEOREM 9. *$\eta_R(v_i)$ is given by*

$$\sum_{i,j \geq 0}^F v_i t_j^{p^i} = \sum_{i,j \geq 0}^F t_i \eta_R(v_j)^{p^i}. \quad \square$$

This completes our survey of the foundations of the subject. We turn now to some applications in the homotopy groups of spheres. Novikov first formulated an MU analogue of the Adams spectral sequence. His main result can be restated as

THEOREM 10 (NOVIKOV [7]). *Let X be a connective spectrum. There is a spectral sequence converging to $Z_{(p)} \otimes \pi_* X$ with $E_2^{**} = \mathrm{Ext}_{BP_*BP}^{**}(BP_*, BP_*X)$.* $\square$

For the definition of this Ext, see [9]. In it, BP_*X can be replaced by any BP_*BP-comodule M. From now on we will abbreviate this to Ext M.

For $X = S^0$ the E_2-term is Ext BP_* which has the following convenient sparseness property.

PROPOSITION 11. *$\mathrm{Ext}^{s,t} BP_* = 0$ if $t \not\equiv 0 \bmod 2(p-1)$. Consequently, in the Adams–Novikov spectral sequence for S^0, $E_{2+2r(p-1)}^{**} = E_{2p-1+2r(p-1)}^{**}$ for $r \geq 0$. In particular, the first nontrivial differential is d_{2p-1} so all nontrivial elements in $E_2^{s,t}$ for $t \leq 2(p-1)$ which are permanent cycles are nontrivial in E_∞^{**}.* $\square$

This spectral sequence has fewer differentials and extensions (at least for p odd) than the classical Adams spectral sequence based on mod p cohomology, i.e. its E_2-term is a closer approximation of stable homotopy. For example, for $p > 2$, $\mathrm{Ext}^1 BP_*$ is isomorphic to $\mathrm{Im}\, J$, the image of the Hopf–Whitehead J-homomorphism, and for $p = 3$ there are no differentials below dimension 33.

An unstable form of this spectral sequence has recently been constructed and used by Bendersky–Curtis–Miller [2]. It appears to be a very promising device.

In studying the classical Adams spectral sequence one learns that elements in $\mathrm{Ext}_{\mathscr{A}}^1(Z/p, Z/p)$ correspond to generators of the Steenrod algebra $\mathscr{A}$ while elements in $\mathrm{Ext}_{\mathscr{A}}^2(Z/p, Z/p)$ correspond to relations among these generators. However, this point of view appears not to be helpful in understanding $\mathrm{Ext}^1 BP_*$ and $\mathrm{Ext}^2 BP_*$.

We will now describe the Greek letter construction, which is an entirely different method of manufacturing elements in $\mathrm{Ext}\, BP_*$.

An ideal $I \subset BP_*$ is *invariant* if BP_*/I is a BP_*BP-comodule, i.e. if $\eta_R I \subset IBP_*BP$. Invariant ideals are rare as the following result shows.

THEOREM 12 (MORAVA, LANDWEBER). (a) *The only invariant prime ideals in* BP_* *are* $I_n = (p, v_1, \ldots, v_{n-1})$ *for* $0 \leq n \leq \infty$ *(*I_0 *is the zero ideal).*

(b) $\mathrm{Ext}^0\, BP_* = Z_{(p)}$ *and* $\mathrm{Ext}^0\, BP_*/I_n = F_p[v_n]$ *for* $0 < n < \infty$.

(c) *The following is a short exact sequence of* BP_*BP*-comodules.*

$$0 \to \Sigma^{2(p^n-1)} BP_*/I_n \xrightarrow{v_n} BP_*/I_n \to BP_*/I_{n+1} \to 0. \quad \square \tag{13}$$

Now let

$$\delta_n \colon \mathrm{Ext}^{s,t}\, BP_*/I_{n+1} \to \mathrm{Ext}^{s+1, t-2(p^n-1)}\, BP_*/I_n$$

be the connecting homomorphism associated with (13). Then we can define the following elements, commonly known as Greek letters, in the Adams–Novikov E_2-term $\mathrm{Ext}\, BP_*$:

$$\begin{aligned} \alpha_t &= \delta_0(v_1^t) \in \mathrm{Ext}^{1, 2(p-1)t}\, BP_*, \\ \beta_t &= \delta_0\delta_1(v_2^t) \in \mathrm{Ext}^{2, 2(p^2-1)t-2(p-1)}\, BP_*, \\ \gamma_t &= \delta_0\delta_1\delta_2(v_3^t) \in \mathrm{Ext}^{3, 2(p^3-1)t-2(p-1)-2(p^2-1)}\, BP_*. \end{aligned} \tag{14}$$

Of course, this definition generalizes to $\eta_t^{(n)}$, where $\eta^{(n)}$ denotes the nth letter of the Greek alphabet.

In order to apply this construction to homotopy theory one must prove two things: that the elements so defined are nontrivial in $\mathrm{Ext}\, BP_*$ and that they are permanent cycles in the Adams–Novikov spectral sequence. It will then follow from Proposition 11 that the resulting elements in E_∞ are nontrivial, so they detect nontrivial homotopy classes.

THEOREM 15 (SEE [9] FOR REFERENCES). (a) *The elements* α_t $(t>0)$ *are nontrivial for* $p \geq 2$ *and are permanent cycles for* $p \geq 3$. *(They detect the elements of order* p *in* $\mathrm{Im}\, J$.*)*

(b) *The elements* β_t $(t>0)$ *are nontrivial for* $p \geq 3$ *and are permanent cycles for* $p \geq 5$.

(c) *The elements* γ_t $(t>0)$ *are nontrivial for* $p \geq 3$ *and are permanent cycles for* $p \geq 7$. $\square$

The nontriviality result is an algebraic computation, while the construction of the corresponding homotopy elements, due to H. Toda and Larry Smith, is as follows. One constructs finite complexes $V(n-1)$ $(n \leq 4)$ with $BP_*V(n-1) = BP_*/I_n$ by means of cofibrations $(n \leq 3)$

$$\Sigma^{2(p^n-1)} V(n-1) \xrightarrow{\varphi_n} V(n-1) \to V(n)$$

realizing the sequence (13), with $V(-1)=S^0$. Then $\eta_t^{(n)}$ is the composition

$$S^{2t(p^n-1)} \xrightarrow{i} \Sigma^{2t(p^n-1)} V(n-1) \xrightarrow{\varphi_n^t} V(n-1) \xrightarrow{j} S^k$$

where i is the inclusion of the bottom cell, j is the collapsing onto the top cell, and $k=\sum_{0\leqq m<n}(2p^m-1)$.

One can generalize the Greek letter construction by replacing the invariant prime ideals I_n by invariant regular ideals. Regularity is precisely what is needed to get short exact sequences generalizing (13). For $p\geqslant 3$ it is known that all elements in $\mathrm{Ext}^1 BP_*$ and $\mathrm{Ext}^2 BP_*$ arise in this way.

However, not all elements in $\mathrm{Ext}^3 BP_*$ come from $\mathrm{Ext}^0 BP/I$ for an invariant regular ideal I with 3 generators. For example, the elements $\alpha_1\beta_t$ arise from elements in $\mathrm{Ext}^1 BP_*/I_2$ which are free under multiplication by v_2, so they cannot come from $\mathrm{Ext}^0 BP_*/(p, v_1, v_2^k)$ for any k. What is true is that every element in $\mathrm{Ext}\, BP_*$ is the image of some element in $\mathrm{Ext}\, BP_*/I$ (where I is an invariant regular ideal with n generators) which is free under multiplication by the powers of v_n belonging to $\mathrm{Ext}^0 BP_*/I$.

Hence in some sense every element of the Adams–Novikov E_2-term is a member of an infinite periodic family of the type exemplified most simply by the $\eta_t^{(n)}$ of (14). Whether a similar statement can be made about stable homotopy itself is still an open question. In light of this situation, one would like to classify these periodic families. A machine for doing this known as the chromatic spectral sequence was set up in [4]. One begins by looking at the $F_p[v_n]$-free summand of $\mathrm{Ext}\, BP_*/I_n$, which maps monomorphically to $v_n^{-1}\,\mathrm{Ext}\, BP_*/I_n=\mathrm{Ext}\, v_n^{-1} BP_*/I_n$. This group is surprisingly easy to compute, due to some farsighted work of Jack Morava. His results indicate a striking connection between homotopy theory and local algebraic number theory. We can only give the barest description here.

$\mathrm{Ext}\, v_n^{-1} BP_*/I_n$ is a free module over $K(n)_*=\mathrm{Ext}^0\, v_n^{-1} BP_*/I_n=F_p[v_n, v_n^{-1}]$. We make F_{p^n} a nongraded $K(n)_*$-module by sending v_n to 1. Then we have

THEOREM 16. *$F_{p^n}\otimes_{K(n)_*}\mathrm{Ext}\, v_n^{-1} BP_*/I_n=H_c^*(S_n, F_{p^n})$, the continuous cohomology (with trivial action on F_{p^n}) of the compact p-adic Lie group S_n, which is the p-Sylow subgroup of the automorphism group of the (height n) formal group law over F_{p^n} induced by $BP_*\to K(n)_*\to F_{p^n}$.* □

For example $H_c^* S_n$ has the following Poincaré series $f(t)=\sum(\dim H_c^i S_n)t^i$:

$p \backslash n$	1	2	3
2	$(1+t)/(1-t)$	$(1+t)^3(1-t^6)/(1-t)(1-t^4)$	?
3	$1+t$	$(1+t)^2(1+t^2)/(1-t)$	?
$\geqq 5$	$1+t$	$(1+t)^2(1+t+t^2)$	$(1+t)^3(1+t+6t^2+3t^3+6t^4+t^5+t^6)$

References

1. J. F. Adams, *Stable homotopy and generalized homology,* Univ. of Chicago Press, Chicago, I 11., 1974.

2. M. Bendersky, E. B. Curtis, and H. R. Miller, *The unstable Adams spectral sequence for generalized homology,* Topology **17** (1978), 229—248.

3. E. H. Brown and F. P. Peterson, *A spectrum whose Z_p-cohomology is the algebra of reduced pth powers,* Topology **5** (1966), 149—154.

4. H. R. Miller, D. C. Ravenel and W. S. Wilson, *Periodic phenomena in the Adams-Novikov spectral sequence,* Ann. of Math. **106** (1977), 469—516.

5. J. W. Milnor, *On the cobordism ring Ω^* and a complex analogue,* Part I, Amer. J. Math. **82** (1960), 505—521.

6. S. P. Novikov, *Some problems in the topology of manifolds connected with the theory of Thom spaces,* Soviet Math. Doklady **1** (1960), 717—720.

7. ——, *The methods of algebraic topology from the viewpoint of cobordism theories,* Math U.S.S.R.-Izvestia (1967), 827—913.

8. D. G. Quillen, *On the formal group laws of unoriented and complex cobordism theory,* Bull. Amer. Math. Soc. **75** (1969), 1293—1298.

9. D. C. Ravenel, *A novice's guide to the Adams—Novikov spectral sequence,* Geometric Applications of Homotopy Theory. II, Lecture Notes in Math, vol. 658, Springer-Verlag, Berlin and New York 1978, pp. 404—475.

10. R. Thom, *Quelques propriétés globales des variétés différentiables,* Comment. Math. Helv. **28** (1954), 17—86.

UNIVERSITY OF WASHINGTON
SEATTLE, WASHINGTON 98195, U.S.A.

Proceedings of the International Congress of Mathematicians
Helsinki, 1978

Hilbert Cube Manifolds—Meeting Ground of Geometric Topology and Absolute Neighborhood Retracts

J. E. West

1. Introduction. Over the past fifteen years the study of (topological) infinite-dimensional manifolds has developed strongly. Previous reports to the Congress were given in 1970 by R. D. Anderson and in 1974 by T. A. Chapman, and there have now appeared excellent monographs by Bessaga and Pełczyński [1] and by Chapman [3]. (These reports should be consulted in conjunction with the present one.) The principal, but by no means the only, development of the field since 1970 has been in the direction of Hilbert cube manifolds (Q-manifolds; $Q=\prod_{i=1}^{\infty}[0, 1]_i$, with metric $d(x, y)=\sum|x_i-y_i|/2^{-i}$). It has been motivated in large part by Chapman's use of them to demonstrate the topological invariance of Whitehead torsion for CW-complexes and the concomitant result that as they are classified by simple homotopy type, rather than by homotopy type as is the case with Hilbert manifolds, they are more complex and more closely related to finite-dimensional manifolds. This development has been dramatic in the extreme and has led to the emergence of Q-manifolds as a rich and interesting theory in its own right which is situated between the areas of n-manifolds, manifolds modelled on infinite-dimensional vector spaces, and the topology of metric compacta and ANR's and has strong interrelations with these disciplines. I shall try to outline this development, necessarily all too briefly and incompletely and with apologies to those whose frequently important and interesting contributions must be slighted for reasons of economy of time and space. (By ANR is meant locally compact, separable, metric absolute neighborhood retract for metric spaces.)

Enunciated in primitive form as early as 1970, it has been a growing conviction that Q-manifolds should be regarded as a simultaneous stabilization of the theories of n-manifolds, locally finite CW-complexes, and ANR's having as a salient characteristic the combination of local compactness with an extreme general position property, and it is this theme which I have taken for my talk.

2. General position ANR's. The main line of development of Q-manifold theory from almost the first recognizable theorem has been toward the recognition and exploitation of Q-manifolds as *general position* ANR's, *i.e., those ANR's* X *such that any mapping into* X *of an n-cell of any dimension may be approximated, arbitrarily closely, by embeddings.* This is amply demonstrated by the following table of 7 selected theorems, which unfortunately omits, because of its format, the extremely relevant work of Anderson on Z-sets. See also §§ 3 and 4 for comments on Chapman's work in this light. In the table, if X is as in the first column, then either Y in the second column is a Q-manifold or the indicated homeomorphism ($\cong$) exists. The third column gives authors and approximate dates of appearance.

	X	Y or $\cong$	Author
1	infinite-dimensional convex compactum in l^2	$X\cong Q$	Keller (1931)
2	dendron	$X\times Q\cong Q$	Anderson (1964)
3	Q-manifold	$X\cong X\times Q$	Anderson—Schori (1969)
4	locally finite CW-complex	$X\times Q$	West (1970, 1971)
5	X_i a Q-factor, $i=1, 2, \ldots$	$\prod_{i=1}^{\infty} X_i \cong Q$	Anderson (1964), West (1969)
6	ANR	$X\times Q$	R. D. Edwards (1976, in [2])
7	general position ANR	X	Toruńczyk (1979)[1]

Thus, we have been led to the above *topological characterization of Q-manifolds* by Toruńczyk, which answers in stunning form the call by Anderson in his 1970 report. (Toruńczyk* has a more recent similarly striking characterization of Hilbert manifolds of all weights.)

3. Stabilization of ANR's. Two natural methods of stabilizing ANR's to obtain Q-manifolds have been employed repeatedly: products and hyperspaces.

The product stabilization. The sequence $0\to I\to I^2\to\ldots\to I^n\to\ldots\to Q$ with $I=[0, 1]$ and inclusions $x\to(x, 0)$ presents Q as the completion of $\cup I^n$ and with the projections $Q\to I^n$ provides an extremely useful functor exhibiting $X\times Q$ as the limit of the stabilization sequence $X\to X\times I\to\ldots\to X\times I^n\to\ldots$. This structure lies at the heart of much Q-manifold theory. For example, Chapman used it repeatedly in his work on triangulating Q-manifolds (Theorem 15) and the topological invariance of Whitehead torsion (Theorem 14) to reduce surgery on Q-manifolds to general position surgery. Another cogent example of the role of Q-manifolds as stable n-manifolds is Chapman's work on concordances. (See Hatcher's introduction to "Higher Simple Homotopy Theory" for an enlightening discussion.)

The hyperspace stabilization. The hyperspace 2^X of nonvoid, compact subsets of X equipped with the Hausdorff metric is functorial. Combined work of D. W. Curtis, Schori, and the author yielded

THEOREM 8. *For each nondegenerate Peano continuum X, 2^X is a Hilbert cube.*

Affirming Wojdysławski's conjecture, this provides a wealth of new examples of Q-manifolds, *e.g.*, if t is the translation action of S^1, then the orbit space $(2^{S^1}-\{S^1\})/t$ is a Q-manifold $K(Q, 2)$, where Q is the rational numbers (Toruńczyk–West)*. See related work by Curtis, Schori, and N. Kroonenberg.

4. Connections with n-manifolds. From the product stabilization, we expect substantial connections here. They exist, and their elucidation is one of the most active and fruitful areas of current research. In addition to concordances, mentioned above, are the following. (See also § 5.)

THEOREM 9 (WEST). *If $f\colon K\to L$ is an (infinite) simple homotopy equivalence between (locally) finite CW-complexes, then its product stabilization is (properly) homotopic to a homeomorphism.*

THEOREM 10 (CHAPMAN). *If $f\colon K\times Q\to L\times Q$ is a homeomorphism, then the composition $f'\colon K\to K\times\{0\}\subset K\times Q\to L\times Q\to L$ is a simple homotopy equivalence. (Here K and L are locally finite CW-complexes and the last map is projection.)*

THEOREM 11 (CHAPMAN). *Every Q-manifold is homeomorphic to the product with Q of a locally finite simplicial complex.*

Theorem 10 encompasses the topological invariance of Whitehead torsion, and Theorem 11 shows that *all* Q-manifolds are "triangulable" (general position).

THEOREM 12 (CHAPMAN). *The homeomorphism groups of compact Q-manifolds are locally contractible.*

(See Theorem 20.) A. Fathi and Y. Visetti have provided the full analog of the Černavskii–Edwards–Kirby Deformation Principle, which is much more powerful but, under space restrictions, less quotable.

THEOREM 13 (CHAPMAN). *Cell-like mappings between Q-manifolds are uniform limits of homeomorphisms.*

THEOREM 14 (CHAPMAN–SIEBENMANN). *If M is a Q-manifold which is tame at ∞ and has finite type, then there is an element $\beta(M)\in\mathscr{S}(M)/\mathrm{Wh}(\pi_1(M))$ which vanishes if and only if M admits a boundary (i.e., there is a compactification M' of M into a Q-manifold such that $M'-M$ is a Z-set. $\mathscr{S}(M)$ is the simple types on M; Z-sets, identified by Anderson, are vital in the theory and may here be thought of as closed subsets of collared submanifolds).*

THEOREM 15 (CHAPMAN). *There is a locally flat embedding* $h\colon S^3\times Q\to S^3\times Q$ *such that neither* h *nor any (finite) stabilization* $i\circ h\colon S^3\times Q\to S^3\times Q\times R^n$ *of* h *has a tubular neighborhood.*

(In his University of Kentucky thesis* W. Nowell shows that locally flat, co-dimension 2 Q-submanifolds have tubular neighborhoods.) Theorem 15 is particularly interesting, as it exhibits a stable pathology not present in n-manifolds.

The shrinking of cell-like decompositions is a *most* important aspect of this branch, entering fundamentally in Theorems 6, 7, and 13. Work has been done by Anderson*, J. Bryant*, Z. Cerin, W. Eaton*, Fathi and Visetti, M. Handel*, Kozlowski*, J. Mogilski* and R. Sher, among others. Handel's work was particularly important in stimulating R. Edwards to prove Theorem 6.

Interesting initial work on transformation groups by Vo Thanh Liem* should also be mentioned in this section.

5. Connections with ANR's. Already mentioned are Theorems 2, 4, 5, 6, and 7 in the table which should be noted here. Q-manifolds provide a bridge between the theories of n-manifolds and CW-complexes and that of ANR's across which it has been possible to transport several fruitful notions, providing a more unified approach. The next result proved instrumental in establishing Theorem 17.

THEOREM 16 (R. MILLER). *The cone on every compact* ANR *is the image of* Q *under a cell-like map.*

THEOREM 17 (WEST). *For every compact* ANR X *there is a* Q*-manifold* M *and a cell-like mapping* $f\colon M\to X$ *the mapping cylinder of which is a* Q*-manifold*; *consequently, every compact* ANR *is homotopy equivalent to a compact polyhedron.*

Theorem 17 answered a question posed by Borsuk in his 1954 address to the Congress. It should be remarked that this homotopy result does not extend even to fundamental ANR's (pointed), as S. Ferry* has realized all of Wall's obstructions to finiteness in the three-dimensional members of this class; however, R. Geoghegan* has shown that it does extend to the finitely dominated inverse limits of compact ANR's under fibrations.

An immediate consequence of Theorem 21, recognized independently by Chapman and R. D. Edwards, was the *extension,* full-blown, *to* ANR's *of simple homotopy theory.* Chapman had previously done this for Q-manifold factors by assigning to X the simple homotopy type of any CW-complex $K(X)$ triangulating $X\times Q$ (unique by Theorem 10) and to a (proper) homotopy equivalence $f\colon X\to Y$ the torsion of the mapping $f'\colon K(X)\to K(Y)$ obtained as in Theorem 10. In light of Theorem 6, this remains the "proper" way. From Theorem 13 and Theorem 14, it follows immediately that cell-like mappings are simple homotopy equivalences; they play the role of collapses.

From this and Theorem 6 has followed an important body of work by Chapman and Ferry on fibrations of ANR's, from which the next two theorems are selected.

(See also the earlier work by Chapman and R. Wong on bundles and that of Wong on the contractibility of the homeomorphism group of Q.)

THEOREM 18 (CHAPMAN–FERRY). *Let* $p\colon E\to B$ *be an Hurewicz fibration.* (a) *If the fibers* $p^{-1}(b)$ *are compact* Q*-manifolds and* B *is locally path-connected and locally finite-dimensional, then* p *is a locally trivial bundle;* (b) *if* B *and* E *are* ANR'*s and the fibers of* p *are compact, then* $p\circ(pjn)\colon E\times Q\to E\to B$ *is a locally trivial bundle, and* (c) *if* $B, E,$ *and the fibers are ANR's, then the composition* $p\circ(pjn)\colon E\times Q\times[0,1)\to E\to B$ *is a locally trivial bundle.*

THEOREM 19 (CHAPMAN–FERRY). (a) *For each map* $f\colon M\to B$ *of a* Q*-manifold to an ANR, the composition* $f\circ(pjn)\colon M\times[0,1)\to M\to B$ *is homotopic to a fibration (Hurewicz)*; (b) *if* M *and* B *are compact and connected,* B *is simple homotopy equivalent to an* n*-complex, and the homotopy fiber of* f *is* n*-connected and of finite type, then there is an obstruction in* $\mathrm{Wh}\left(\pi_1(M)\right)$ *which vanishes if and only if* f *is homotopic to a fibration (Hurewicz). (Weaker conditions for the existence of the obstruction if* $n=1,2$.*)*

The following outstanding result was obtained independently by Ferry and Toruńczyk.

THEOREM 20. *The homeomorphism groups of compact* Q*-manifolds are Hilbert manifolds.*

(The essential contribution here is that they are (non-locally compact) ANR's. The analog for n-manifolds remains a major open problem, $n>2$; for $n=2$ it is due to R. Luke and W. Mason.) From Theorem 21, Ferry easily drew the following *useful* corollary by deforming the product stabilization to a homeomorphism. (Motivated by Theorem 21, he and Chapman* have obtained the analogous unstabilized deformation result for maps of n-manifolds.)

COROLLARY TO (21). *For every ANR* Y *there is an open cover* Σ *with the property that any map* $f\colon X\to Y$ *between ANR's with a homotopy inverse* g *such that* $f\circ g\simeq 1_Y$ *and* $g\circ f\simeq 1_X$ *by homotopies with tracks refining* Σ *and* $f^{-1}(\Sigma)$*, respectively, is a simple homotopy equivalence.*

6. Connections with compacta: Ends of Q-manifolds and Tychonov cubes. Let $\mathcal{S}$ be the shape category of metric compacta, regarded as lying in $Q\times\{0\}\subset Q\times I$. The assignment $X\to Q\times I-X$ yields contractible Q-manifolds. Proper maps $f,g\colon A\to B$ of local compacta are weakly properly homotopic if for each compactum D in B there is one C in A such that the restrictions of f and g to $A-C$ are homotopic in $B-D$. Let $\mathcal{WP}$ be the category of all $Q\times I-X$ above and weakly proper homotopy classes of proper maps.

THEOREM 21 (CHAPMAN). *The assignment* $X\to Q\times I-X$ *determines a category isomorphism* $\mathcal{S}\to\mathcal{WP}$.

Theorem 21 defines the relationship between Q-manifolds and shape theory. Many open problems revolve about this phenomenon and the closely related proper homotopy theory, with information flowing in both directions. (See S. Mardešić's report to this Congress, the Chapman–Siebenmann paper, [3], and [1]. There is highly relevant work by J. Dydak, Dydak and J. Segal, D. Edwards and R. Geoghegan, Kozlowski*, and Mardešić and Rushing.) Certain questions involving the lack of an adequate Whitehead theorem in these categories, fibrations, and the preservation of movability at ends are closely linked with the study of compact Lie group actions on Q-manifolds, which has been initiated by Wong, I. Berstein, Vo Thanh Liem and the author. A sample result is the following: Let $Q' = \prod_{i=1}^{\infty} [-1, 1]_i$, $0 =$ the zero element, and $A(x) = -x$. Let T be any involution of Q' with unique fixed point 0.

THEOREM 23*. *T is conjugate to A if and only if* (a) *(Berstein–West) the inverse system of deleted neighborhoods of* 0 *in* Q'/T *is movable, or* (b) *(West–Wong)* Q'/T *is an AR.*

For nonmetrizable compacta, E. V. Ščepin* has made a beautiful application of the fibration theory of Q-manifolds to prove

THEOREM 24 (ŠČEPIN)*. *A weight-homogeneous retract of a Tychonov cube is a Tychonov cube if its weight is at least* $\aleph_1$.

7. Connections with manifolds modelled on vector spaces. Anderson's proof that l^2 is homeomorphic to the countably infinite product s of lines, which completed the topological classification of separable Fréchet spaces, was a direct outgrowth of his study of Q as a compactification of s. Toruńczyk's topological characterization of Hilbert manifolds of all weights* followed the philosophical lines of his analogous work on Q-manifolds and completed this ancient problem of Fréchet. R. E. Heisey's proof that manifolds modelled on l^2 with the bounded-weak topology are homeomorphic if homotopy equivalent was based on his recognition that this space is a direct limit of Hilbert cubes. There are major connections between Q-manifolds and function spaces, especially spaces of differentiable functions, which are suggested by

THEOREM 25 (GEOGHEGAN). *Let M^m and N^n be connected Riemannian manifolds, with M^m compact and N^n flat. For each $k>0$, the closure in $C^0(M^m, N^n)$ of $\{f \in C^1(M^m, N^n) | f'(x) < k$ for all $x\}$ is a Q-manifold.*

A. Jones, in his Cornell University thesis*, has related work employing an energy bound. (In this context, see R. Palais' report to the 1970 Congress.) This is one of the most promising areas for future research, and with the availability of Toruńczyk's characterization of Q-manifolds, it is wide-open.

Bibliography

For reasons of space, only the monographs mentioned in the text are referenced here; the reader is referred to their bibliographies, to the reports to the Congress mentioned, and to the reference journals. Those papers which have not appeared as of this writing are marked by an asterisk where they are mentioned and may in most cases be obtained by writing to the authors.

1. Cz. Bessaga and A. Pełczyński, *Selected topics in infinite-dimensional topology*, Mathematical Monographs, vol. 58, PWN, Warsaw, 1975.

2. K. Borsuk, *Theory of shape*, Mathematical Monographs, vol. 59, PWN, Warsaw, 1975.

3. T. A. Chapman, *Lectures on Hilbert cube manifolds*, C.B.M.S. Regional Conf. Series, vol. 28, Amer. Math. Soc., Providence, R. I., 1976.

4. D. A. Edwards and H. M. Hastings, *Čech and Steenrod homotopy theories with applications to geometric topology*, Lecture Notes in Math., vol. 542, Springer-Verlag, Berlin and New York, 1976.

CORNELL UNIVERSITY
ITHACA, NEW YORK 14853, USA

Index