SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Inventiones Mathematicae
Verlag: Springer
Jahr: 1987
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN356556735_0088
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN356556735_0088

Übergeordnetes Werk

Werk Id: PPN356556735
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN356556735
OPAC: http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=356556735

Terms and Conditions

The Goettingen State and University Library provides access to digitized documents strictly for noncommercial educational, research and private purposes and makes no warranty with regard to their use for other purposes. Some of our collections are protected by copyright. Publication and/or broadcast in any form (including electronic) requires prior written permission from the Goettingen State- and University Library.
Each copy of any part of this document must contain there Terms and Conditions. With the usage of the library's online system to access or download a digitized document you accept the Terms and Conditions.
Reproductions of material on the web site may not be made for or donated to other repositories, nor may be further reproduced without written permission from the Goettingen State- and University Library.

For reproduction requests and permissions, please contact us. If citing materials, please give proper attribution of the source.

Contact

Niedersächsische Staats- und Universitätsbibliothek Göttingen
Georg-August-Universität Göttingen
Platz der Göttinger Sieben 1
37073 Göttingen
Germany
Email: gdz@sub.uni-goettingen.de

Inventiones mathematicae

Volume 88 1987

Editores

M. Berger, Bures-sur-Yvette **A. Borel,** Princeton **J. Coates,** Cambridge
A. Connes, Bures-sur-Yvette **A. A. Kirillov,** Moskwa
B. Mazur, Cambridge, Mass. **J. W. Morgan,** New York
R. Remmert, Münster **J-P. Serre,** Paris **J. Sjöstrand,** Orsay
S.-T. Yau, La Jolla

Editores quondam

J. F. Adams (1974–1984) V. I. Arnold (1966–1978) H. Bauer (1966–1978)
B. J. Birch (1966–1978) E. Bombieri (1969–1979) P. Cartier (1966–1975)
P. Gabriel (1966–1967) A. Haefliger (1966–1968) L. Hörmander (1976–1984)
N. H. Kuiper (1966–1974) B. Malgrange (1966–1976) Y. Manin (1966–1978)
J. K. Moser (1974–1982) J-P. Serre (1967–1979) T.A. Springer (1976–1986)
J. Tits (1966–1975) E. Vesentini (1966–1973)



Springer International

Inventiones mathematicae

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review, or thesis); that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors, if any, as well as by the responsible authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher and that the manuscript will not be published elsewhere in any language without the consent of the copyright holders.

All articles published in this journal are protected by copyright, which covers the exclusive rights to reproduce and distribute the article (e.g., as offprints), as well as all translation rights. No material published in this journal may be reproduced photographically or stored on microfilm, in electronic data bases, video disks, etc., without first obtaining written permission from the publisher.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if not specifically identified, does not imply that these names are not protected by the relevant laws and regulations.

While the advice and information in this journal is believed to be true and accurate at the date of its going to press, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Special regulations for photocopies in the USA: Photocopies may be made for personal or in-house use beyond the limitations stipulated under Section 107 or 108 of U.S. Copyright Law, provided a fee is paid. This fee is US $ 0.20 per page, or a minimum of US $ 1.00 if an article contains fewer than five pages. All fees should be paid to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISSN 0020-9910, the volume, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

Urheberrecht. Voraussetzung für die Einreichung eines Manuskriptes an die Redaktion der Zeitschrift ist, daß die Arbeit noch nicht publiziert oder an anderer Stelle zur Publikation eingereicht wurde. Ferner wird vorausgesetzt, daß die Publikation von allen beteiligten Autoren einer Arbeitsgruppe genehmigt und daß die Arbeit, wenn sie zur Publikation angenommen wurde, nicht an anderer Stelle in gleicher Form, ohne die Genehmigung des Copyright-Inhabers, publiziert wird, entweder in derselben oder einer anderen Sprache.
Die Zeitschrift sowie alle in ihr enthaltenen einzelnen Beiträge und Abbildungen sind urheberrechtlich geschützt. Jede Verwertung, die nicht ausdrücklich vom Urheberrechtsgesetz zugelassen ist, bedarf der vorherigen schriftlichen Zustimmung des Verlags. Das gilt insbesondere für Vervielfältigungen, Bearbeitungen, Übersetzungen, Mikroverfilmungen und die Einspeicherung und Verarbeitung in elektronischen Systemen. Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden. Jeder Autor, der Deutscher ist oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebt oder Bürger Österreichs, der Schweiz oder eines Staates der Europäischen Gemeinschaft ist, kann unter bestimmten Voraussetzungen an der Ausschüttung der Bibliotheks- und Fotokopietantiemen teilnehmen. Nähere Einzelheiten können direkt von der Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestr. 49, D-8000 München 2, eingeholt werden.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in dieser Zeitschrift berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürfen.

Printers: Universitätsdruckerei H. Stürtz AG, Würzburg

© Springer-Verlag Berlin Heidelberg 1987

NIEDERSÄCHS.
STAATS. U. UNIV.-
BIBLIOTHEK
GÖTTINGEN

Index Volume 88 1987

Fasc. 1 1–224 issued on 13.03.1987
Fasc. 2 225–446 issued on 06.04.1987
Fasc. 3 447–668 issued on 04.05.1987

Adolphson, A., Sperber, S.: Newton polyhedra and the degree of the L-function associated to an exponential sum 555

Altmann, K.: Equisingular deformations of isolated 2-dimensional hypersurface singularities 619

Bănică, C., Putinar, M.: On complex vector bundles on projective threefolds 427

Bernstein, D., Katok, A.: Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonians 225

Bombieri, E., Schmidt, W.M.: On Thue's esquation 69

Bourgain, J., Milman, V.D.: New volume ratio properties for convex symmetric bodies in $\mathbb{R}^n$ 319

Boutot, J.-F.: Singularités rationnelles et quotients par les groupes réductifs 65

Bruce, J.W., Plessis, A.A. du, Wall, C.T.C.: Determinacy and unipotency 521

Buchweitz, R.-O., Greuel, G.-M., Schreyer, F.-O.: Cohen-Macaulay modules on hypersurface singularities II 165

Cox, D., Donagi, R., Tu, L.: Variational Torelli implies generic Torelli 439

Crew, R.: L-functions of p-adic characters and geometric Iwasawa theory 395

Donagi, R., see Cox, D., et al. 439

Du Cloux, F.: Jets de fonctions différentiables sur le dual d'un groupe de Lie nilpotent 375

Floyd, W.J., Plotnick, S.P.: Growth functions on Fuchsian groups and the Euler characteristic 1

Forman, R.: Functional determinants and geometry 447

Goldman, W.M., Millson, J.J.: Local rigidity of discrete groups acting on complex hyperbolic space 495

Greuel, G.-M., see Buchweitz, R.-O., et al. 165

Hain, R.M., Zucker, S.: Unipotent variations of mixed Hodge structure 83

Hardt, R., Lin, F.-H.: Regularity at infinity for area-minimizing hypersurfaces in hyperbolic space 217

Horie, K.: A note on basic Iwasawa λ-invariants of imaginary quadratic fields 31

Kalai, G.: Rigidity and the lower bound Theorem 1 125

Katok, A., see Bernstein, D. 225

Kitazume, M.: The Conway-Norton algebras for $\Omega^-(6,3)$, $\Omega(7,3)$, F'_{24}, and their full automorphism groups 277
Knörrer, H.: Cohen-Macaulay modules on hypersurface singularities I 153
Levitt, G.: 1-formes fermées singulières et groupe fondamental 635
Lichtenbaum, S.: The construction of weight-two arithmetic cohomology 183
Lin, F.-H., see Hardt, R. 217
Millson, J.J., see Goldman, W.M. 495
Milman, V.D., see Bourgain, J. 319
Nori, M.V.: On subgroups of $GL_n(\mathbb{F}_p)$ 257
Plessis, A.A. du, see Bruce, J.W., et al. 521
Plotnick, S.P., see Floyd, W.J. 1
Putinar, M., see Bănică, C. 427
Ratner, M.: Rigid reparametrizations and cohomology for horocycle flows 341
Reid, L.: A correction to a paper of H. King 423
Rotthaus, C.: On the approximation property of excellent rings 39
Rubin, K.: Local units, elliptic units, Heegner points and elliptic curves 405
Salvetti, M.: Topology of the complement of real hyperplanes in $\mathbb{C}^N$ . . 603
Schmidt, W.M., see Bombieri, E. 69
Schreyer, F.-O., see Buchweitz, R.-O., et al. 165
Sperber, S., see Adolphson, A. 555
Tu, L., see Cox, D., et al. 439
Wall, C.T.C., see Bruce, J.W., et al. 521
White, B.: Curvature estimates and compactness theorems in 3-manifolds for surfaces that are stationary for parametric elliptic functionals . . 243
Yamasaki, M.: L-groups of crystallographic groups 571
Zucker, S., see Hain, R.M. 83

Covered by Zentralblatt für Mathematik and Current Mathematical Publications

A20900E

Inventiones mathematicae

1-3 T J 2 A 25965

Vol. 88 Fasc. 1 1987

W.J. Floyd, S.P. Plotnick	Growth functions on Fuchsian groups and the Euler characteristic	1
K. Horie	A note on basic Iwasawa λ-invariants of imaginary quadratic fields	31
C. Rotthaus	On the approximation property of excellent rings	39
J.-F. Boutot	Singularités rationnelles et quotients par les groupes réductifs	65
E. Bombieri, W.M. Schmidt	On Thue's equation	69
R.M. Hain, S. Zucker	Unipotent variations of mixed Hodge structure	83
G. Kalai	Rigidity and the lower bound Theorem 1	125
H. Knörrer	Cohen-Macaulay modules on hypersurface singularities I	153
R.-O. Buchweitz, G.-M. Greuel, F.-O. Schreyer	Cohen-Macaulay modules on hypersurface singularities II	165
S. Lichtenbaum	The construction of weight-two arithmetic cohomology	183
R. Hardt, F.-H. Lin	Regularity at infinity for area-minimizing hypersurfaces in hyperbolic space	217

Covered by *Zentralblatt für Mathematik* and *Current Mathematical Publications*

Niedersächsische Staats- u. Universitätsbibliothek Göttingen

08. April 1987

Springer International

222 Invent. math. ISSN 0020-9910 INVMBH 88(1) 1–224 (1987) April 1987

Inventiones mathematicae

This journal is published at frequent intervals to bring out new contributions to mathematics. It is a policy of the journal to publish papers within four months of acceptance.

Copyright

Submission of a manuscript implies: that the work described has not been published before (except in the form of an abstract or as part of a published lecture, review, or thesis); that it is not under consideration for publication elsewhere; that its publication has been approved by all coauthors, if any, as well as by the responsible authorities at the institute where the work has been carried out; that, if and when the manuscript is accepted for publication, the authors agree to automatic transfer of the copyright to the publisher and that the manuscript will not be published elsewhere in any language without the consent of the copyright holders.

All articles published in this journal are protected by copyright, which covers the exclusive rights to reproduce and distribute the article (e.g., as offprints), as well as all translation rights. No material published in this journal may be reproduced photographically or stored on microfilm, in electronic data bases, video disks, etc., without first obtaining written permission from the publisher.

The use of general descriptive names, trade names, trademarks, etc., in this publication, even if not specifically identified, does not imply that these names are not protected by the relevant laws and regulations.

While the advice and information in this journal is believed to be true and accurate at the date of its going to press, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, express or implied, with respect to the material contained herein.

Special regulations for photocopies in the USA: Photocopies may be made for personal or in-house use beyond the limitations stipulated under Section 107 or 108 of U.S. Copyright Law, provided a fee is paid. This fee is US $ 0.20 per page, or a minimum of US $ 1.00 if an article contains fewer than five pages. All fees should be paid to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISSN 0020-9910, the volume, and the first and last page numbers of each article copied. The copyright owner's consent does not include copying for general distribution, promotion, new works, or resale. In these cases, specific written permission must first be obtained from the publisher.

Other regulations: Authors publishing in this journal can, under certain conditions, benefit from library and photocopy fees collected by VG WORT. Authors of German nationality and those resident in the Federal Republic of Germany or Berlin (West), as well as citizens of Austria, Switzerland and member countries of the European Community, may apply to Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, for detailed information.

Pour permettre une publication plus rapide, le journal «Inventiones mathematicae» paraît par fascicules séparés réunis ensuite en volumes. La rédaction s'efforce de publier les travaux reçus dans les quatre mois suivant leur acceptation.

Droits d'auteur

Ne peuvent être acceptés que des articles n'étant pas simultanément remis pour publication ailleurs et n'ayant jamais été publiés, ni dans leur pays d'origine, ni à l'étranger. L'auteur s'engage à ne pas soumettre son manuscrit ailleurs après qu'il a été accepté. Dès l'acceptation d'un manuscrit et sa publication par la maison d'édition, le droit d'impression en toutes langues et pour tous pays, y compris le droit de reproduction photographique, en microforme ou par tout autre moyen, est acquis par la maison d'édition.

Die Zeitschrift erscheint, um eine rasche Publikation zu ermöglichen, in einzelnen Heften, die zu Bändern vereinigt werden. Es ist das Ziel der Zeitschrift, Arbeiten innerhalb von 4 Monaten nach Annahme zu publizieren.

Urheberrecht

Voraussetzung für die Einreichung eines Manuskriptes an die Redaktion der Zeitschrift ist, daß die Arbeit noch nicht publiziert oder an anderer Stelle zur Publikation eingereicht wurde. Ferner wird vorausgesetzt, daß die Publikation von allen beteiligten Autoren einer Arbeitsgruppe genehmigt und daß die Arbeit, wenn sie zur Publikation angenommen wurde, nicht an anderer Stelle in gleicher Form, ohne die Genehmigung des Copyright-Inhabers, publiziert wird, entweder in derselben oder einer anderen Sprache.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten.

Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere von Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden.

Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder ähnlichem Wege bleiben vorbehalten.

Fotokopien für den persönlichen und sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopien hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte und benützte Kopie dient gewerblichen Zwecken gemäß § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, von der die einzelnen Zahlungsmodalitäten zu erfragen sind.

Jeder Autor, der Deutscher ist oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebt oder Bürger Österreichs, der Schweiz oder eines Staates der Europäischen Gemeinschaft ist, kann unter bestimmten Voraussetzungen an der Ausschüttung der Bibliotheks- und Fotokopietantiemen teilnehmen. Nähere Einzelheiten können direkt von der Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Anschrift s. oben, eingeholt werden.

Die Wiedergabe von Gebrauchsnamen, Handelsnamen, Warenbezeichnungen usw. in dieser Zeitschrift berechtigt auch ohne besondere Kennzeichnung nicht zu der Annahme, daß solche Namen im Sinne der Warenzeichen- und Markenschutz-Gesetzgebung als frei zu betrachten wären und daher von jedermann benutzt werden dürften.

Manuscripts in duplicate should be submitted to

Les manuscrits en deux exemplaires doivent être envoyés à

Manuskripte in doppelter Anfertigung erbeten an

Reinhold Remmert
Berliner Straße 7
D-4540 Lengerich

or to any of the editors
ou à un éditeur
oder an einen der Herausgeber

Invent. math. 88, 1–29 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Growth functions on Fuchsian groups and the Euler characteristic

William J. Floyd[1,*] and Steven P. Plotnick[2,**]

[1] Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA, and Department of Mathematics, VPI & SU, Blacksburg, VA 24061, USA

[2] Department of Mathematics, Columbia University, New York, NY 10027, USA

Let G be a finitely generated group, and Σ a finite generating set for G. Then Σ determines a norm on G, called the *word norm*, defined by letting $|g|$ be the minimal length of g as a word in Σ. Following Milnor [15], we define the growth series $g(z) = 1 + a_2 z + \ldots + a_n z^n \ldots$, where a_n is the number of elements of G of word norm exactly n. Much of the early work (e.g., [1, 16, 17, 23]) on growth series was concerned with asymptotic properties of the coefficients of $g(z)$. This work culminated in Gromov's theorem [12] that G has polynomial growth if and only if it is virtually nilpotent and in Grigorchuk's theorem [11] that there are finitely generated groups whose growth is neither polynomial nor exponential.

If G is a Coxeter group and Σ is the standard generating set for G, Bourbaki [3] showed that the growth series $g(z)$ is the series of a rational function $f(z)$ and Serre [20] showed that $f(1) = 1/\chi(G)$, where $\chi(G)$ is the rational Euler characteristic of G. Note that if G is finite then $g(1) = \text{order}(G) = 1/\chi(G)$, but $g(z)$ is not defined at one if G is infinite. If G is a compact hyperbolic or irreducible Euclidean Coxeter group and Σ is the standard generating set for G, Serre showed that $f(1/z) = \pm f(z)$, and so the poles of f are algebraic units. In contrast to the work on the asymptotic properties of the coefficients of $g(z)$, these results raised the possibility of a beautiful theory of the exact structure of growth functions on groups. However, a decade went by before there was much further work on this subject.

In two papers ([5] and [6]) Cannon studied growth functions for cocompact hyperbolic groups. If G is a cocompact hyperbolic group and Σ is any finite generating set for G, he showed that the coefficients a_n of the growth series $g(z)$ satisfy a linear recursion, and hence $g(z)$ is the power series of a rational function $f(z)$. He also computed some examples when G is a closed surface group or a compact hyperbolic Coxeter group. If G is a closed surface group and Σ is a geometric generating set (see below for a definition) for G coming from a one relator presentation, then f is reciprocal ($f(z) = f(1/z)$), $f(1) = 1/\chi(G)$, the

* Supported in part by NSF Grant DMS 84-01919

** Supported in part by NSF Grant DMS 85-03758

numerator of f is a product of cyclotomic polynomials, and the denominator of f is the product of a Salem polynomial and cyclotomic polynomials.

In 1983, Benson [2] showed that if G is a finite extension of $\mathbb{Z}^n$ and Σ is any finite generating set for G, then the growth series $g(z)$ is the series of a rational function $f(z)$, and f has a pole of order n at $z=1$. Since these groups have Euler characteristic $\chi(G)=0$, this fits in with the conjecture that $f(1)=1/\chi(G)$ when the power series $g(z)$ is the series of a rational function f and when $\chi(G)$ is defined.

Smythe [21] studied the relation between $f(1)$ and $\chi(G)$ by thinking of G as a filtered semigroup and studying RG-resolutions of R (where R is the integers or a field). While he gets conditions for which $f(1)=1/\chi(G)$, it seems to be very hard to work with these conditions.

There have been further calculations (e.g., [10] and [22]) to support the conjecture that $f(1)=1/\chi(G)$. However, a recent example due to Parry [18] shows that this conjecture is not true for all finite generating sets. Let $G=\mathbb{Z}*\mathbb{Z}$ with generators x and y, $\Sigma=\{x^{\pm 2}, x^{\pm 3}, y^{\pm 2}, y^{\pm 3}\}$, and $\Sigma'=\{g\in G\colon |g|_\Sigma<3\}$. Parry shows that the growth function of G with respect to the generating set Σ' is $f(z)=(1+24z-4z^2)/(1-32z+4z^2)$. Note that $f(1)=-7/9$ and $\chi(G)=-1$. As a result of this example, we decided to look at the case when Σ is a geometric generating set.

Let G be a discrete group of isometries of hyperbolic n-space $\mathbb{H}^n$ such that $\mathbb{H}^n/G$ has finite volume. Choose a basepoint x in $\mathbb{H}^n$ which is not fixed by any non-trivial element of G, and let $D=D_x=\{y\in\mathbb{H}^n\colon d(y,x)\leqq d(y,gx) \text{ for all } g\in G\}$. Then D is a finite sided polyhedron in $\mathbb{H}^n$ which is a fundamental region for the action of G on $\mathbb{H}^n$. D is called a Dirichlet region, and the number of sides of D depends on the choice of basepoint x. The set of face-pairing elements $\Sigma=\{g\in G\colon gD\cap D$ has dimension $n-1\}$ is a finite generating set for G. If G is finitely generated, we will call a finite generating set Σ for G *geometric* if there is a representation of G as a discrete group of isometries of hyperbolic space $\mathbb{H}^n$ so that $\mathbb{H}^n/G$ has finite volume and a basepoint $x\in\mathbb{H}^n$ so that Σ is the set of face pairing elements for D_x.

Let G be a finitely generated, discrete group of isometries of $\mathbb{H}^2$ with $\mathbb{H}^2/G$ of finite area and Σ a geometric generating set for G with respect to the Dirichlet region $D=D_x$. For convenience we also assume that Σ has no elements of order two. In § 1 we study the combinatorics of the tessellation of $\mathbb{H}^2$ by $\{gD\colon g\in G\}$. Using this we show: (1) $B(n)=\cup\{gD\colon |g|\leqq n\}$, the "combinatorial ball of radius n", is a topological ball, and (2) there is a square matrix $\mathbf{A}$ and a column vector $\mathbf{v}$ so that $g(z)$ is the series of the rational function $f(z)=1+(1\ldots1)(\mathbf{I}-z\mathbf{A})^{-1}z\mathbf{v}$. This is motivated by Cannon's proof that $g(z)$ is the series of a rational function, but we may have a different recursion matrix $\mathbf{A}$ and initial vector $\mathbf{v}$.

These facts are used in § 2, where we prove our main result. Loosely speaking, it may be stated as follows:

Theorem 1. *Given a Fuchsian group G, Dirichlet region D, generating set Σ, and growth function $f(z)$ as above, there is a row vector $\mathbf{w}$ (defined geometrically in terms of D) so that*

$$f(z)=1/\chi(G)+(1-z)(2\chi(G))^{-1}\mathbf{w}(\mathbf{I}-z\mathbf{A})^{-1}\mathbf{v}.$$

Thus $f(1)=1/\chi(G)$ if $\mathbf{I}-\mathbf{A}$ is invertible.

From this one would expect that $f(1)=1/\chi(G)$ for most but not all choices of geometric generating sets for Fuchsian groups. In § 3 we give some examples to illustrate this. We first compute the growth functions associated to tilings of $\mathbb{H}^2$ by (even-sided) regular polygons, with all angles $2\pi/p$, $p\geqq 3$. In these cases, it turns out that $\mathbf{I}-\mathbf{A}$ is invertible, so $f(1)=1/\chi$. We also give an example, on a genus two surface with a less symmetric Dirichlet region, where $\mathbf{I}-\mathbf{A}$ is not invertible but nevertheless $f(1)=1/\chi$. Finally, we give an example where $\mathbf{I}-\mathbf{A}$ is not invertible and $f(1)\neq 1/\chi$. Our example has genus 4 and 30 sides, with angles alternating between $2\pi/3$ and $2\pi/5$. We conclude, in § 4, with an indication of how our methods should extend to higher even-dimensional hyperbolic groups, and with some open problems.

§ 1. Combinatorics of planar tessellations

Let G be a finitely generated, discrete group of isometries of either the hyperbolic plane (in which case we assume $\mathbb{H}^2/G$ has finite area), the Euclidean plane (in which case we assume $\mathbb{E}^2/G$ is compact), or the round sphere S^2, and let Σ be a geometric generating set for G with respect to the Dirichlet region $D=D_x$. For each integer $n\geqq 0$ let $B(n)=\text{closure}(\cup\{gD\colon |g|_\Sigma\leqq n\})$, where in the first case we are using the Poincaré disc model for $\mathbb{H}^2$ and taking the closure in $\mathbb{R}^2$ (the closure is adding the parabolic points at infinity). One can think of $B(n)$ as the "combinatorial ball of radius n" around x.

Our goal in this section is to give an inductive construction of the group graph of G with generators Σ. As consequences of this construction, we will show:

(1) If G is a group of hyperbolic or Euclidean isometries, then $B(n)$ is a topological ball, and the cells gD with $|g|_\Sigma=n+1$ intersect $B(n)$ and each other in rather simple configurations.

(2) The growth series of G with respect to Σ is the series of a rational function.

We make one rather mild assumption on Σ, more out of convenience than necessity – namely, *we assume that Σ has no elements of order two.* Thus, our discussion is essentially disjoint from the case of Coxeter groups, where all elements of Σ have order two.

We first describe how the Dirichlet region determines a presentation for G with generators Σ. For simplicity, first assume that D is compact and G is torsion-free, so that G is the fundamental group of a closed surface. Let $V=\{\text{vertices } v \text{ of } D\}$, and put an equivalence relation $\sim$ on V by $v\sim w$ if there is a finite sequence $v=v_0$, $v_1,\ldots,v_n=w$ of vertices of D and elements $g_1,\ldots,g_n\in\Sigma$ with $g_jv_{j-1}=v_j$ for all $j\in\{1,\ldots,n\}$ (so each equivalence class corresponds to one vertex in the quotient surface). Each equivalence class of vertices is known as a *vertex cycle*. Let $\{v_1,\ldots,v_m\}$ consist of one vertex from each equivalence class. For each $i\in\{1,\ldots,m\}$, let c_i be the word in Σ determined by an edge-path in the 1-skeleton of the dual tessellation which traverses the link of v_i. If $a_0=v_i$, $a_1=g_1a_0$, $a_2=g_2a_1$, $\ldots$, $a_r=g_ra_{r-1}$, $a_0=g_{r+1}a_r$ is a minimal listing (i.e., $a_j\neq a_i$ if $i\neq j$) of the vertex cycle through v_i, then the corresponding word is $g_{r+1}g_r\ldots g_2g_1$. Note that choosing a different initial edge for the edge-path would give a cyclic permutation of c_i and changing the orientation of the edge-path would replace c_i by its inverse. Then G has presentation $\langle\Sigma|c_1,\ldots,c_m\rangle$. We give an example in Fig. 1.

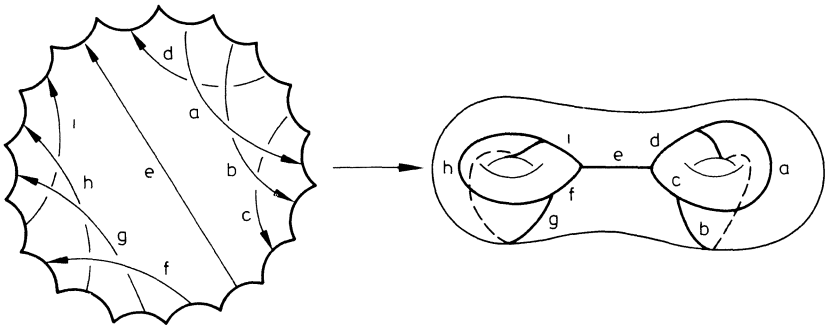


Fig. 1. $G \cong \langle a, b, c, d, e, f, g, h, i \mid ab^{-1}d,\ bc^{-1}a^{-1},\ ced^{-1},\ fie^{-1},\ gh^{-1}f^{-1},\ hi^{-1}g^{-1} \rangle$

In the general case, D may be non-compact and G may have torsion. Recall that Σ has no elements of order two. In particular, Σ contains no reflections, so each element of Σ pairs two different edges. Hence, D is an even-sided polygon in $\mathbb{H}^2$, $\mathbb{E}^2$, or S^2, possibly with some vertices at ∞ in the hyperbolic case. As before, we follow the edge pairings to compute the vertex cycles and words $c_1, \ldots, c_m$. We also assign an integer $n_i \in \{1, 2, \ldots, \infty\}$ to each cycle as follows: Given $i \in \{1, \ldots, m\}$, add the (interior) angles in D at each vertex of the cycle through v_i. This gives the total angle at the image of v_i, say $\bar{v}_i$, in the quotient $\mathbb{H}^2/G$, $\mathbb{E}^2/G$, or S^2/G. Since G is discrete, this angle must have the form $2\pi/n_i$ for some positive integer n_i. If $n_i = 1$, the quotient has a non-singular geometric structure near $\bar{v}_i$, and c_i has order 1 in G. If $1 < n_i < \infty$, $\bar{v}_i$ is a branch point of index n_i, and c_i is an elliptic element of order n_i. If $n_i = \infty$, the cycle through v_i consists of vertices at ∞, neighbourhoods of the vertices in the cycle are glued together to form a cusp in $\mathbb{H}^2/G$, and c_i is a parabolic element of infinite order. Then G has the presentation $\langle \Sigma \mid (c_1)^{n_1}, \ldots, (c_m)^{n_m} \rangle$, where infinitely long relators, of course, may be deleted.

We now describe the possibilities for the "spine" of the quotient; that is, the image of ∂D. In the torsion-free, cocompact case, the spine is a finite graph, and every vertex has valence at least three, as in Fig. 1. In the presence of torsion, there may be vertices of valence one or two. For example, the orientation-preserving subgroup of a triangle group can give valence one or two vertices, as shown in Fig. 2. Finally, if $\mathbb{H}^2/G$ is not compact, the spine might be disconnected, as shown in Fig. 3.

We now describe the properties that we need of the words $c_1, \ldots, c_m$ to inductively construct the group graph. Let $C = \{c_1, \ldots, c_m\}$. For each c_i, let $[c_i]$ consist of all cyclic permutations of c_i, and let $W = \cup \{c^{n_i} : c \in [c_i] \text{ and } i \in \{1, \ldots, m\}\}$. By convention, if $n^i = \infty$, c^{n_i} means the bi-infinite word $\ldots ccc \ldots$; if g is the first letter of c, we say c^∞ *begins* with g. Note that, since Σ has no element of order two, every element of W has length at least three.

Lemma 1 (compare Cannon [6, Lemma 3.2]). (i) *The elements of C are cyclically reduced.*

(ii) *For each $g \in \Sigma$, $g^{\pm 1}$ appears exactly twice in the elements of C.*

(iii) *Elements of $W \cup W^{-1}$ are determined by their first two letters.*

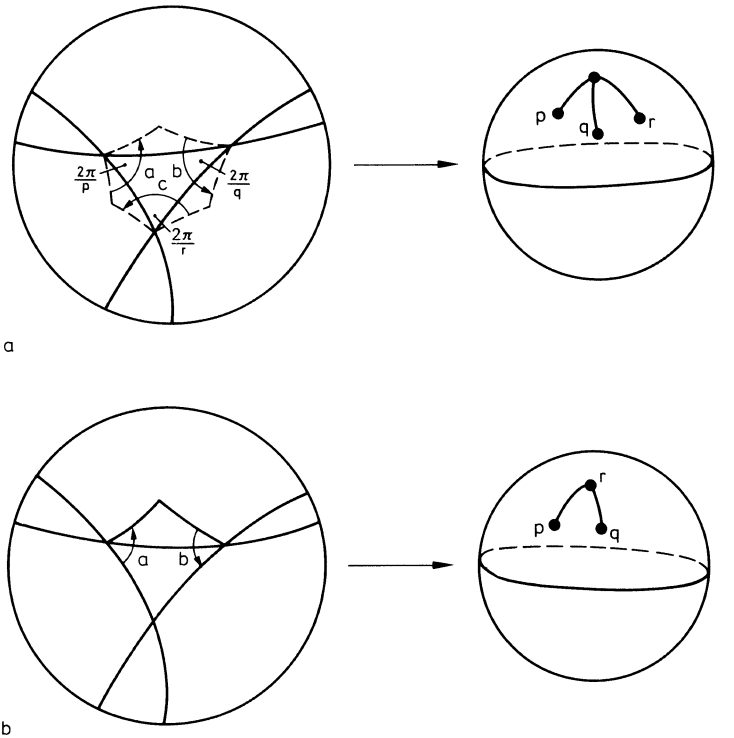


Fig. 2a, b. Two possible "spines" for a (p, p, r)-triangle group. **a** $G \cong \langle a, b, c \,|\, a^p, b^q, c^r, abc \rangle$. **b** $G \cong \langle a, b \,|\, a^p, b^q, (ab)^r \rangle$

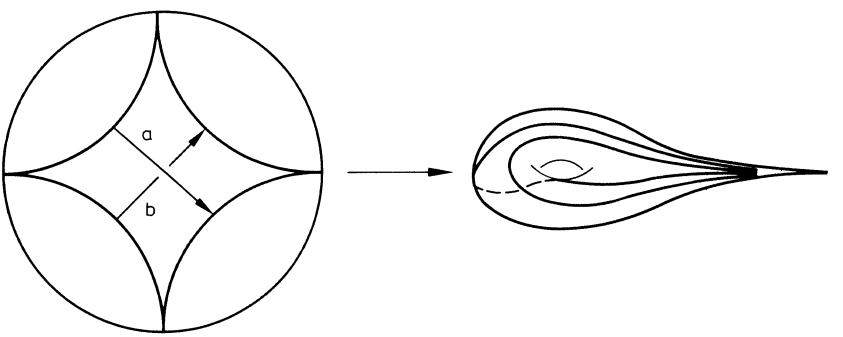


Fig. 3. $G \cong \langle a, b \,|\, (aba^{-1}b^{-1})^\infty \rangle$

(iv) *For each* $g \in \Sigma$, *exactly two of the elements of* $W \cup W^{-1}$ *begin with* g.

(v) *If* f, g *and* h *are distinct elements of* Σ, *then* fg^{-1}, fh *and* gh *do not all occur as initial segments of elements of* $W \cup W^{-1}$.

Proof. (i) The combination gg^{-1} could not occur for a geometric generating set.

(ii) Let e be the edge $D \cap gD$. Then $g^{\pm 1}$ occurs exactly once for each end or vertex of e.

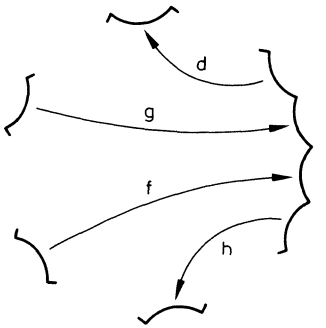


Fig. 4

(iii) Let $g \in \Sigma$, $w \in (W \cup W^{-1})$ such that w starts with g, and let e be the edge $D \cap gD$. Then w is the word determined by following one of the two vertices or ends of e. So there are only two possibilities, and these are detected by the second letter of w.

(iv) This follows from (ii).

(v) Let f, g and h be distinct elements of Σ, and suppose that fg^{-1} and fh are initial segments of elements of $W \cup W^{-1}$. Without loss of generality, we can assume that the edge-pairing elements near the edge $D \cap fD$ are as in Fig. 4. If gh occurs as an initial segment of an element of $W \cup W^{-1}$, then $d = h$ and D only has 3 sides, a contradiction. □

Our construction of the group graph follows an argument of Cannon in [6], and we adopt Cannon's notation from [6] and [7]. Given a graph L and a directed edge e in L, let $\bar{e}$ be the same edge but with the opposite direction. We say that a graph L is *colored by* Σ if each directed edge e in L has a label $c(e) \in \Sigma$ such that $c(\bar{e}) = (c(e))^{-1}$. We call $c(e)$ the *color* of e. If L is colored by Σ, then each edge-path α in L determines a word $c(\alpha)$ in Σ by concatenating the colors of the successive directed edges of the edge-path. A colored graph L is a *prepermutation* graph if for each vertex $v \in L$ and each color c in Σ there is at most one edge emanating from v with color c.

Now assume that G is infinite, i.e. either hyperbolic or Euclidean. Let $K \subset \mathbb{H}^2$, $\mathbb{E}^2$ be the 1-complex with vertices the elements $\{gx : g \in G\}$, and with two vertices gx, hx connected by a geodesic arc if gD and hD intersect in an edge (this occurs exactly when $hg^{-1} \in \Sigma$). Since Σ has no elements of order two, K is the 1-skeleton of the dual tessellation. For each vertex v in K and $g \in \Sigma$, there is exactly one directed edge e in K from v to gv, K is colored by Σ if we give e the color g, and K is exactly the group graph of G with generators Σ.

Using Lemma 1, we can now give an inductive construction of K. Our building blocks are graphs $\Lambda_\omega(v)$ which are colored by Σ, where $\omega \in W$ and v is a vertex. Given a symbol v and a finite word $\omega \in W$, $\Lambda_\omega(v)$ is a cyclic graph such that one vertex of $\Lambda_\omega(v)$ is labelled by v and the cyclic edge-loop, based at the vertex labelled v, which traverses each edge exactly one time determines the word ω or ω^{-1} (depending on the direction). If v is a vertex and $\omega = c^\infty \in W$, then $\Lambda_\omega(v)$ is an infinite graph homeomorphic to $(-\infty, \infty)$ by a homeomorphism taking the vertex labelled v to 0. In this case $\Lambda_\omega(v)$ is colored so that one of the two edge-paths which traverse each edge exactly one time determines the word c^∞ and the segment which starts at v

and has length $|c|$ determines the word c. We will construct graphs

$$\Gamma(0) \subset \Gamma(1) \subset \ldots \subset \Gamma(n) \subset \ldots \subset \bigcup_{n=0}^{\infty} \Gamma(n) = \Gamma$$

such that $\Gamma(0)$ is a base vertex v_0, and for all $n>0$ $\Gamma(n)$ is a connected, prepermutation graph which is obtained from $\Gamma(n-1)$ by adding some $\Lambda_\omega(v)$'s. We will abuse notation and also denote by $\Lambda_\omega(v)$ the image of $\Lambda_\omega(v)$ in $\Gamma(n)$, since each $\Lambda_\omega(v)$ will be embedded. However, it is possible that a Λ_ω in $\Gamma(n)$ can have two different base vertices v_1 and v_2; that is, the colored graphs $\Lambda_{w_1}(v_1)$ and $\Lambda_{w_2}(v_2)$ are identified in $\Gamma(n)$ even though $v_1 \neq v_2$. (Actually, in very special circumstances, Λ_ω may have three base vertices. See Condition 5 below.) We will put norms $\| \ \|_n$ on the $\Gamma(n)$'s so that if v is a vertex of a Λ_ω in $\Gamma(n)$, then $\|v\|_n = \min\{\|v\|_n$ plus the edge-path distance in Λ_ω from v to v: v is a base vertex for $\Lambda_\omega\}$. Since our norms will have $\|v\|_i = \|v\|_n$ if $i<n$ and $v \in \Gamma(i)$, we will usually denote $\|v\|_n$ by $\|v\|$. Let $V(n-1)$ denote the set of vertices of $\Gamma(n-1)$. For each $n>0$, we require that $\Gamma(n)$ satisfies the following five Conditions:

(1) Each $\Lambda_\omega(v)$ of $\Gamma(n)$ is based at a vertex $v \in V(n-1)$.

(2) An edge of $\Gamma(n)$ joining a vertex of norm $\leqq n$ to one of norm $<n$ belongs to exactly two of the $\Lambda_\omega(v)$'s in $\Gamma(n)$.

(3) An edge of $\Gamma(n)$ joining two vertices of norm $\geqq n$ belongs to exactly one $\Lambda_\omega(v)$ in $\Gamma(n)$.

(4) A vertex in $\Gamma(n)$ of norm n is joined by edges to at most two vertices of norm $n-1$. If it is joined to two vertices of norm $n-1$, then all three vertices belong to some finite $\Lambda_\omega(v)$ in $\Gamma(n)$.

(5) Except for the case of the orientation-preserving subgroup of a $(3,3,r)$-triangle group, with a 4-sided Dirichlet region as in Fig. 2.b, at most two vertices of a given $\Lambda_\omega(v)$ can have equal norm, and if $\Lambda_\omega(v)$ contains two vertices of minimal norm, then these vertices are connected by an edge. In this case we say $\Lambda_\omega(v)$ is based at both vertices. For the $(3,3,r)$-triangle group, some Λ_ω's are based at three vertices.

Let $\Gamma(0)$ be a vertex v_0. To construct $\Gamma(1)$, start with the union $\Gamma(0) \coprod \cup \{\Lambda_\omega(v_0)$: $\omega \in W\}$, identify all of the vertices labelled v_0, and call the resulting graph $\Gamma(1)'$. For each $g \in \Sigma$, by Lemma 1.iv) there are exactly two directed edges in $\Gamma(1)'$ which start at the vertex labelled v_0 and have color g, and these will belong to different Λ_ω's by Lemma 1.i). For each $g \in \Sigma$ identify this pair of directed edges, and call the resulting graph $\Gamma(1)$. Then $\Gamma(1)$ is connected, and $\Gamma(1)$ is a prepermutation graph by Lemma 1.iii). Furthermore, all relators of $\Gamma(1)$ are a consequence of finite elements of W, as in Cannon [6]. Put a norm $\| \ \|$ on the vertices of $\Gamma(1)$ by letting $\|v\|$ be the edge-path distance in $\Gamma(1)$ from the vertex labelled v_0 to v. Note that Conditions 1–5 hold for $\Gamma(1)$, and that each $\Lambda_w(v_0)$ is embedded in $\Gamma(1)$.

Now let $n>1$, and assume that $\Gamma(n-1)$ has been constructed satisfying Conditions 1–5. We will obtain $\Gamma(n)$ from $\Gamma(n-1)$ by adding $\Lambda_\omega(v)$'s for each vertex $v \in \Gamma(n-1)$ with $\|v\| = n-1$. We will first define graphs $\Gamma(v)$ for each $v \in V(n-1)$ with $\|v\| = n-1$, attach these graphs to $\Gamma(n-1)$, and then make whatever edge identifications are required for the result, $\Gamma(n)$, to be a prepermutation graph.

First assume that v is joined by an edge in $\Gamma(n-1)$ to a unique vertex $v' \in \Gamma(n-1)$ with norm $\|v'\| = n-2$. By Condition 2, the edge vv' joining v and v' belongs to exactly two $\Lambda_\omega(v)$'s, which we will denote by $\Lambda_{w'}(a)$ and $\Lambda_{w''}(b)$, in $\Gamma(n-1)$. By Condition 3, at least one of these, which we will assume is $\Lambda_{w''}(b)$, is not in $\Gamma(n-2)$. It follows that v' is a base vertex for $\Lambda_{w''}(b)$, and we inductively assume the local picture in $\Gamma(n-1)$ near vv' is that of Fig. 5a). Let w_1 be the cyclic permutation of w' which is determined by the edge-loop in $\Lambda_{w'}(a)$ which starts at v. If w' is an infinite word, this means that we read w' beginning at v. Let w_2 be the cyclic permutation of w'' which is determined by the edge-loop in $\Lambda_{w''}(v')$ which starts at v. Let $W(v) = W - \{w_1, w_2\}$. Take $\cup\{\Lambda_\omega(v) : \omega \in W(v)\}$, identify all of the vertices labelled v, and call the resulting graph $\Gamma(v)'$. For each $g \in \Sigma$, there will be at most two directed edges in $\Gamma(v)'$ which start at the vertex labelled v and have color g. Identify each pair of directed edges from this vertex which have the same color, call the resulting graph $\Gamma(v)$, and note that each $\Lambda_\omega(v)$ is embedded in $\Gamma(v)$.

Now assume that v is joined by edges in $\Gamma(n-1)$ to two distinct vertices x_1 and x_2 in $\Gamma(n-1)$ with $\|x_1\| = \|x_2\| = n-2$. Then by Condition 4 there is a finite word $w \in W$ and a vertex $x \in \Gamma(n-2)$ with x, x_1 and x_2 all in $\Lambda_w(x)$. By Conditions 2 and 3, there are words w', $w'' \in W$ such that $v \in \Lambda_{w'}(x_1)$, $v \in \Lambda_{w''}(x_2)$, and both $\Lambda_{w'}(x_1)$ and $\Lambda_{w''}(x_2)$ are in $\Gamma(n-1)$ but not in $\Gamma(n-2)$. Thus, we inductively assume the local picture in $\Gamma(n-1)$ near v is as in Fig. 5b. Note that there are no identifications among Λ_w, $\Lambda_{w'}$, and $\Lambda_{w''}$ other than the ones pictured. This is because otherwise either $d_2 = (c_1)^{-1}$, $d_1 = (c_2)^{-1}$, or $d_1 = d_2$. If $d_2 = (c_1)^{-1}$, then by Lemma 1.iii) the cyclic permutations of $w^{\pm 1}$ and $(w'')^{\pm 1}$ determined by the two edge loops which start with $x_2 v$ must be the same. But this implies that $\Lambda_w(x)$ and $\Lambda_{w''}(x_2)$ are the same graphs since $\Gamma(n-1)$ is a prepermutation graph, which gives a contradiction. Similar reasoning shows that $d_2 \neq (c_1)^{-1}$. If $d_1 = d_2$, then there are elements in $W \cup W^{-1}$ beginning $c_1(c_2)^{-1}$, $c_1 d_1$, and $c_2 d_2$, violating Lemma 1.v). Let w_1 be the cyclic permutation of w determined by the edge-loop in $\Lambda_w(x)$ which starts at v, let w_2 be the cyclic permutation of w' determined by the edge-loop in $\Lambda_{w'}(x_1)$ which starts at v, and let w_3 be the cyclic permutation of w'' determined by the edge-loop in $\Lambda_{w''}(x_2)$ which starts at v. Let $W(v) = W - \{w_1, w_2, w_3\}$. Let $\Gamma(v)'$ be the graph obtained from $\cup\{\Lambda_\omega(v) : w \in W(v)\}$ by identifying all of the vertices labelled v. For each $g \in \Sigma$, there will be at most two directed edges in $\Gamma(v)'$ which start at the base vertex labelled v and have color g. Identify each pair of directed edges from the base

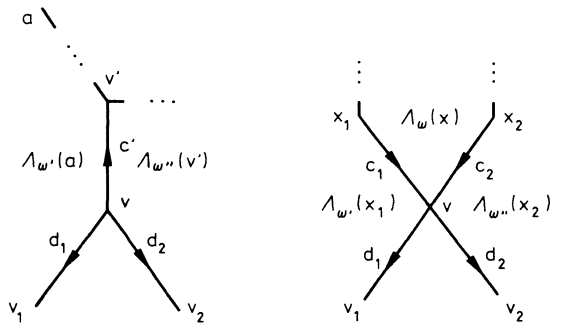


Fig. 5a, b a b

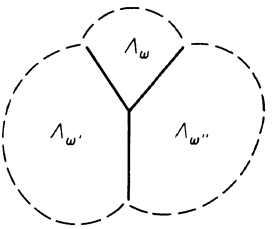


Fig. 6

vertex which have the same color, and call the resulting graph $\Gamma(v)$. As before, each $\Lambda_\omega(v)$ is embedded in $\Gamma(v)$.

For the future, note that the argument above ruling out $d_1 = d_2$ (see the proof of Lemma 1.v) merely says that one cannot find three Λ_ω's fitting nearly around a vertex, as in Fig. 6. Indeed, since K is the 1-skeleton of the dual tessellation, the link of v would be a copy of D, forcing D to be 3-sided. Similarly, if a future argument forces four Λ_ω's to fit nearly around a vertex, D would be 4-sided.

We will now attach the $\Gamma(v)$'s to $\Gamma(n-1)$ to obtain $\Gamma(n)$. Let $\Gamma(n)'$ be the graph obtained from $\Gamma(n-1) \coprod \cup \{\Gamma(v): v \in V(n-1) \text{ and } \|v\| = n-1\}$ by identifying each vertex v in $\Gamma(n-1)$ of norm $n-1$ with the vertex labelled v in $\Gamma(v)$. We will obtain $\Gamma(n)$ as a quotient of $\Gamma(n)'$ by further edge identifications, depending on the local structure near the vertices of norm $n-1$ in $\Gamma(n-1)$. Let v be a vertex in $\Gamma(n-1)$ with $\|v\| = n-1$, and let v_1 and v_2 be the vertices in $\Gamma(n-1)$ as shown in Fig. 5.a or 5.b (depending on which case v is in). Then each of $\|v_1\|$ and $\|v_2\|$ is either $n-1$ or n, depending on the lengths of w and w'. Identify the pair of directed edges in $\Gamma(n)'$ which start at v and have color d_1, and identify the pair of directed edges in $\Gamma(n)'$, which start at v and have color d_2. Do this for each vertex of norm $n-1$ in $\Gamma(n)'$.

We claim that no further identifications need be made near v, in order that $\Gamma(n)$ be a prepermutation graph, if $\|v_1\| = \|v_2\| = n$. For, suppose further identifications need be made, involving v and, say, v_2. Because of the way the Λ_ω are added to $\Gamma(n-1)$, this can *only* occur if v_2 is connected in $\Gamma(n-1)$ to another vertex, v_3, of norm $n-1$ or n, and we must have one of the following two possibilities (Fig. 7):

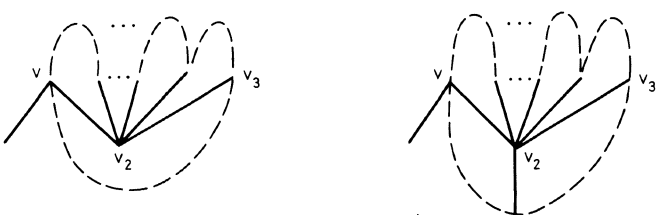


Fig. 7a, b

Note also that *all* of the Λ_ω pictured as lying above v_2 must already be present in $\Gamma(n-1)$. But then Condition 3 is violated if there are more than one of these. Hence, there is only one, but now the local picture near v_2 forces D to have two (a) or three (b) sides.

Note also that this argument shows that if some $\Lambda_\omega \in \Gamma(n) - \Gamma(n-1)$ is based at more than one vertex (of norm $n-1$), then there is an interval in Λ_ω connecting

these two vertices so that every vertex in the interval also has norm $n-1$. This is a step toward inductively verifying Condition 5.

Now suppose that $\|v_2\| = n-1$. (This occurs if either w'' has length 3 or w'' has length 4 but is based at two vertices.) Then in $\Gamma(n)'$ there is a copy of $\Gamma(v)$ attached to v and a copy of $\Gamma(v_2)$ attached to v_2. In $\Gamma(v)$ there is exactly one $\Lambda_\alpha(v)$ which has a directed edge which starts at v and has color d_2, and in $\Gamma(v_2)$ there is exactly one $\Lambda_\beta(v_2)$ which has a directed edge which starts at v_2 and has color $(d_2)^{-1}$. In the above we have identified the directed edge in $\Gamma(v)$ which starts at v and has color d_2 with the directed edge vv_2 in $\Gamma(n)'$ and we have identified the directed edge in $\Gamma(v_2)$ which starts at v_2 and has color $(d_2)^{-1}$ with the directed edge v_2v in $\Gamma(n)'$. In order that the resulting graph will be a prepermutation graph, we need to also identify the copies of $\Lambda_\alpha(v)$ and $\Lambda_\beta(v_2)$ in $\Gamma(n)'$ so that the colors match up. This is possible by Lemma 1.iv). If $\|v_1\| = n-1$ (This occurs if either $\Lambda_{w'}$ is a cycle of length $2r+1$, based at one vertex of norm $n-r-1$, or if $\Lambda_{w'}$ is a cycle of length $2r$, based at two vertices of norm $n-r-2$.) then do the same thing with the appropriate cycles in $\Gamma(v)$ and $\Gamma(v_1)$. Continue this for all vertices of norm $n-1$ in $\Gamma(n)'$, and call the resulting graph $\Gamma(n)$. By construction, $\Gamma(n)$ is a connected, prepermutation graph (this uses Lemma 1.v) if α and β are 3-cycles), and all relators in $\Gamma(n)$ are a consequence of finite elements of W. Let $V(n)$ be the set of vertices in $\Gamma(n)$. Since $\Gamma(n-1)$ and each $\Gamma(v)$ are embedded in $\Gamma(n)$, we will identify them with their images in $\Gamma(n)$. In the case where we have identified $\Lambda_\alpha(v)$ with $\Lambda_\beta(v_2)$, we will think of their image as a single $\Lambda_\omega(v)$, even though it has more than one choice of labelled vertex.

We now investigate the possibility that some Λ_ω could be based at more than two vertices. Suppose v, v_1, and v_2 are as in Fig. 5.a or 5.b, with $\|v\| = \|v_1\| = \|v_2\| = n-1$. If $\Lambda_\alpha(v)$ is identified with both $\Lambda_\beta(v_2)$ and $\Lambda_\psi(v_1)$, we find one of two possibilities (Fig. 8):

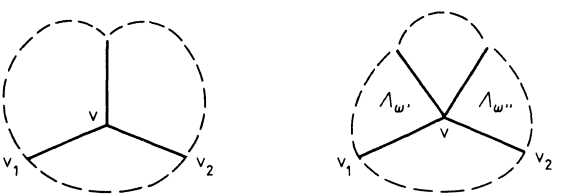


Fig. 8a, b a b

As usual, Fig. 8a is impossible. If Fig. 8b arises, D is 4-sided. Furthermore, if v has minimal norm among all vertices producing Λ_ω's based at three vertices, then $\Lambda_{w'}$ and $\Lambda_{w''}$ have length 3 or 4.

Since we will need to analyze the presentations coming from 4-sided fundamental domains, we list below the possibilities (Fig. 9).

Now, the situation where Λ_w is based at three vertices actually does arise. Consider Fig. 9a with $p=q=3$. Then $G \cong \langle a, b | a^3, b^3, (ab)^r \rangle$ is the $(3,3,r)$-triangle group from Fig. 2b. If $r=2$, G is the tetrahedral group, so assume $r \geqq 3$. Figure 10 shows a piece of the group graph for the Euclidean case, $r=3$, with vertices labelled by their norms. The hyperbolic cases ($r>3$) are similar.

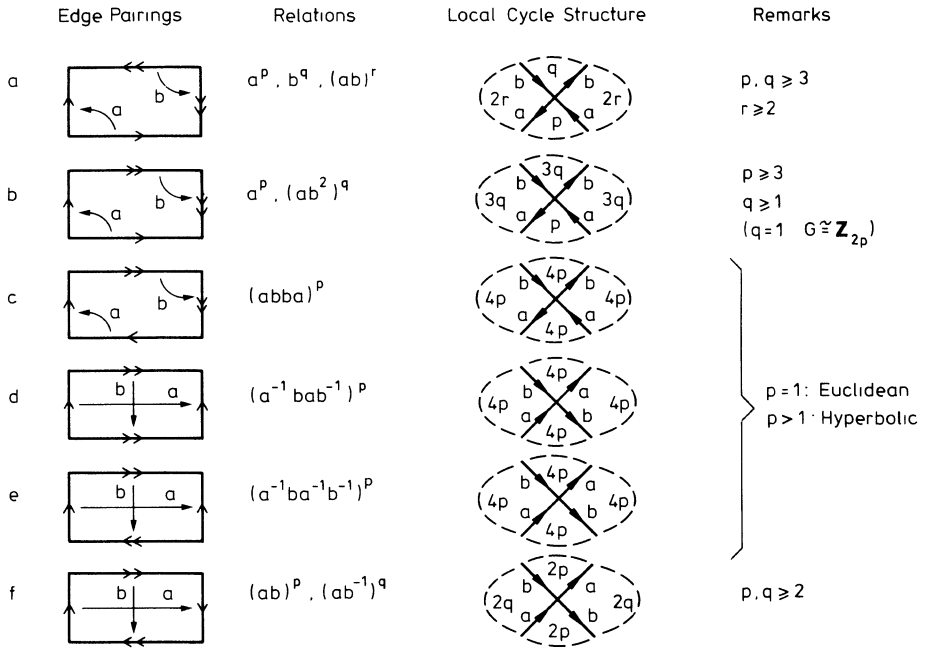


Fig. 9

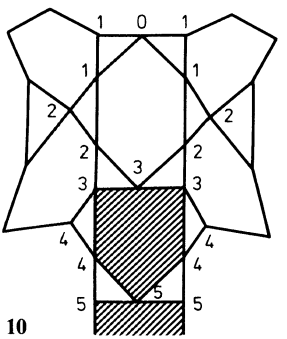


Fig. 10

We claim these are the only such examples. For instance, suppose both $\Lambda_{w'}$ and $\Lambda_{w''}$ (Fig. 8) have length 4, and that v, v_1, v_2 have minimal norm among such triples. Working backwards, we find (Fig. 11):

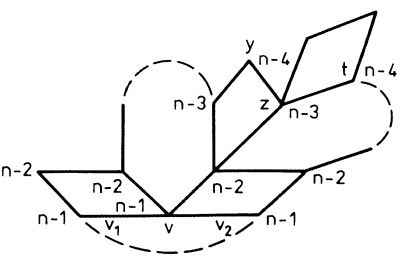


Fig. 11

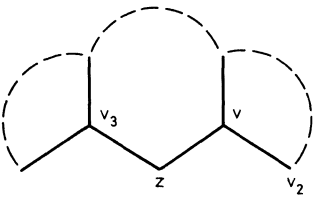


Fig. 12

But the vertices y, z, t violate Condition 4 for $\Gamma(n-3)$. A similar contradiction arises if one of $\Lambda_{w'}$, $\Lambda_{w''}$ has length 3 and the other has length 4. Refering to Fig. 9, we see that the only other possibility is Fig. 9b with $q=1$, but this is a finite group.

Extend the norm on $V(n-1)$ to $V(n)$ by defining the norm of a vertex $v \in V(n) - V(n-1)$ to be $\|v\| = \min\{\|v\|$ plus the edge-path distance in $\Gamma(v)$ from v to v: $v \in \Gamma(v)\}$.

It now remains to inductively verify Conditions 1–5 for $\Gamma(n)$. Observe that Condition 1 is obvious, and we have already given the arguments necessary to verify Condition 5 (modulo checking that the $(3, 3, r)$-triangle groups do not give rise to any other unusual behavior, but this is easy). The main work remaining consists in verifying Condition 4. So let z be a vertex of norm n. By a *predecessor* of z we mean a vertex v, of norm $n-1$, connected by an edge to z. We need to show z has at most two predecessors, as in Condition 4. Let v be a predecessor of z. There are two main cases.

Case I: Every predecessor of z has a unique predecessor.

Subcase A: z is either v_1 or v_2 (Fig. 5a).

Suppose $z = v_1$. Of course, z might have a predecessor belonging to $\Lambda_{w'}$ (Fig. 12). If z has another predecessor, it must have a predecessor that belongs to a cycle that shares either vz or $v_3 z$ with $\Lambda_{w'}$. If it shares vz it must also contain $v_2 v$, and then the local picture makes D 3-sided. If it shares $v_3 z$, then since v_3 has a unique predecessor, the situation is as shown in Fig. 12, and then the local picture at v_3 forces D to be 3-sided.

Subcase B: $z \neq v_1$, $z \neq v_2$.

(1) If both v_1 and v_2 are predecessors of z, we have (Fig. 13):

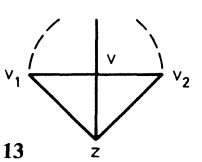


Fig. 13

Hence, D is 4-sided. But we see from Fig. 9 that the only infinite groups with 4-sided Dirichlet regions, and 3-cycles in their presentations, are the $(3, 3, r)$-triangle groups, in which case the two 3-cycles which emanate from a vertex are opposite, not adjacent.

(2) If (say) v_2 is a predecessor of z, but v_1 is not, there are various possibilities. If a cycle through v_1 passes through z, we have (Fig. 14):

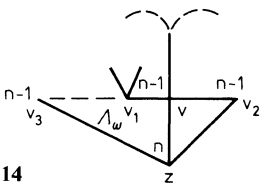


Fig. 14

Since $\Lambda_w \subset \Gamma(n) - \Gamma(n-1)$, the vertices of Λ_w have norm at least $n-1$. But if $\|v_1\| = n$, Λ_w violates Condition 5. Thus, $\|v_1\| = n-1$, Λ_w is a 4-cycle based at three vertices, and G must be the tetrahedral group. It is easy to draw the entire group graph in this case, and indeed this situation does occur. On the other hand, if a cycle through v_2 passes through z, we find a picture exactly as in Fig. 14, but with v replaced by v_2.

(3) If neither v_1 nor v_2 are predecessors of z, and a cycle through v_2, say, passes through z, we have (Fig. 15):

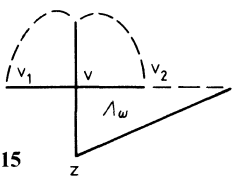


Fig. 15

But Condition 5 forces Λ_w to be a 4-cycle, based at three vertices ($\|v_2\| = n-1$), so we have a $(3, 3, r)$-triangle group. But this forces the cycles adjacent to Λ_w to have length 3, so v_1 is a predecessor of z after all.

Case II: Some predecessor, v, of z has two predecessors (Fig. 5b).

Subcase A: $\|v_1\| = \|v_2\| = n-1$.

(1) If both v_1 and v_2 are predecessors of z, D is forced to have 5 sides.

(2) If neither v_1 nor v_2 are predecessors of z, then the existence of another predecessor gives rise to Fig. 16:

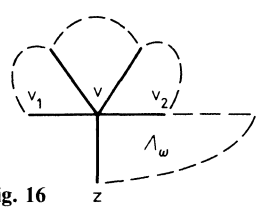


Fig. 16

But then Λ_w is based at three vertices, so D has 4 sides, which contradicts the local picture at v.

(3) If v_1 is a predecessor of z, but v_2 is not, we must have Fig. 17:

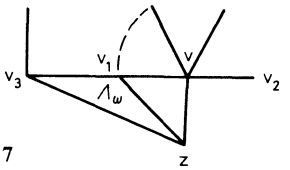


Fig. 17

This is because a cycle containing z and passing through v_2 would give a contradiction as in (2) above. If there was a vertex of Λ_w in the interior of v_1v_3, Condition 5 would force Λ_w to be based at three vertices, Λ_w would have length 4, and we would have the tetrahedral group. Thus, Λ_w is a 3-cycle, but now we have adjacent 3-cycles.

Subcase B: $\|v_1\| = n$, $v_1 \neq z$.

We have already seen that no Λ_w passing through v_1 can include z, and it is easy to see that at most one additional predecessor (v_2) of z can come via v_2.

Subcase C: $v_1 = z$.

We begin with (Fig. 18):

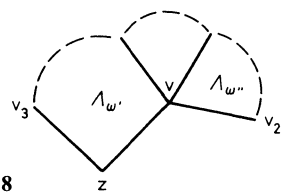


Fig. 18

If $\|v_3\| = n$, a relatively short argument shows that z has no predecessors coming via v_3. But previous arguments give that at most one predecessor (v_2) comes via v_2. Thus, assume $\|v_3\| = n-1$, and $\Lambda_{w'}$ has length 4 or 5. If another predecessor comes via v_3, G must be a $(3,3,r)$-triangle group, and the fact that $\Lambda_{w'}$ has length 4 or 5 forces G to be the tetrahedral group (but this does not arise). If $\|v_2\| = n$, we know (Condition 5) that no predecessor comes via v_2. Thus, assume $\|v_2\| = n-1$. Suppose a predecessor other than v_2 came via v_2. If v_2 is also a predecessor of z, we have a $(3,3,r)$-triangle group with two adjacent 3-cycles; whereas if v_2 is not a predecessor we again have a $(3,3,r)$-triangle group, but with two adjacent cycles of length greater than 3. We conclude that the situation is (Fig. 19):

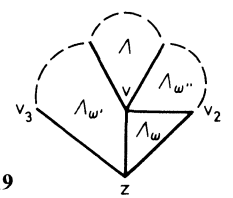


Fig. 19

Now $\Lambda_{w''}$ must have length 4, since otherwise there are adjacent 3-cycles. But now this forces (Fig. 9a) $\Lambda_{w'}$ to have length 4, and we have a $(p,3,2)$-triangle group, where Λ has length p. Letting z have minimal norm among all vertices with three predecessors, and working backwards, we find (Fig. 20):

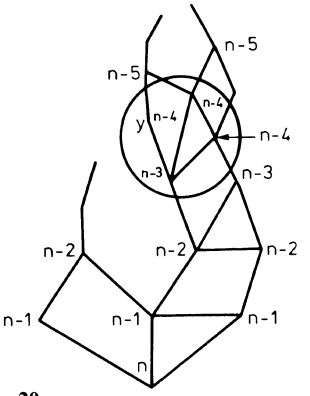


Fig. 20

The circled portion contradicts minimality! However, our labelling implicity used that $p \geqq 6$, since otherwise $\|y\| \neq n-4$. In fact, this is precisely how vertices with three predecessors arise in the graphs of the octahedral ($p=4$) and icosohedral ($p=5$) groups. It is instructive to draw out these graphs – as vertices arise with more than two predecessors, the cycles begin "wrapping" around the graph. Eventually, a cycle wraps completely around the graph. The graph is then complete, and tessellates a sphere.

This completes the inductive verification of Condition 4. We leave to the reader the relatively easy task of verifying Conditions 2 and 3. This now (finally!) completes the construction of Γ. As in [6], we find that $\Gamma = \bigcup_{n=0}^{\infty} \Gamma(n)$ is a connected prepermutation graph, each color emanates from every vertex of Γ, each finite word of W is a relator at every vertex, and every relator of Γ is a consequence of the finite words of W. Consequently, Γ is the graph of the group with presentation $\langle \Sigma | (c_1)^{n_1}, \ldots, (c_m)^{n_m} \rangle$. In fact, Γ is isomorphic to K by an isomorphism φ which takes v_0 to the base point x in K and is color-preserving. In addition, φ is norm-preserving between the norm $\| \ \|$ on the vertices of Γ and the norm $| \ |$ on G.

We now show that the $B(n)$'s are balls. Recall that G is a finitely generated, discrete, co-finite area group of hyperbolic or Euclidean isometries, Σ is a geometric generating set for G with respect to the Dirichlet region $D = D_x$, Σ has no elements of order two, and for each integer $n \geqq 0$ $B(n) = \text{closure}(\cup \{gD\colon |g| \leqq n\})$.

Lemma 2. *For every integer $n \geqq 0$, $B(n)$ is a ball.*

Proof. First exclude the $(3,3,r)$-triangle groups with 4-sided Dirichlet regions. The proof is by induction on n. The lemma is valid for $n=0$ since $B(0) = cl(D)$ is a ball. Let $n>0$, and assume that $B(n-1)$ is a ball. Let $\{b_1, \ldots, b_k\}$ be an enumeration of the elements in G of norm n. For each $i \in \{1, \ldots, k\}$, let $C_i = B(n-1) \cup cl(b_1 D \cup \ldots \cup b_i D)$. Let $C_0 = B(n-1)$. Then $B(n) = C_k$, so it suffices to show that each C_i is a ball.

Let $i \in \{1, \ldots, k\}$, and assume by induction that C_{i-1} is a ball. We will show that C_i is a ball by showing that $C_{i-1} \cap cl(b_i D)$ is a closed interval. By Condition 4, $b_i D$ intersects $B(n-1)$ in either one edge or in two edges which have a vertex in

common. By Condition 5, if $b_iD \cap b_jD$ is not empty for some $j \neq i$, then $b_iD \cap b_jD$ is an interval and $b_iD \cap b_jD \cap B(n-1)$ is a point in the boundary of $B(n-1)$. Thus $C_{i-1} \cap b_jD$ is an interval and hence C_i is a ball.

For the $(3, 3, r)$-triangle groups, it is easy to see that each C_i is a ball if one is careful in enumerating the elements of norm n. □

Note that the proof of Lemma 2 actually shows that, except for the $(3, 3, r)$-triangle groups, if $B(n)$ is built from $B(n-1)$ by adding the domains of norm n, one at a time, in *any* order, the result is a ball at each stage.

Actually, Brenner and Lyndon [4] have shown the following remarkable fact. Let Π be *any* cellulation of the plane (i.e. Π is a union of polygons covering the plane, and two polygons are either disjoint, meet in a vertex, or meet along an edge). Pick a base polygon $P_0 \in \Pi$. Define $P \in \Pi$ to have length 1 if $P \cap P_0$ is an edge. Assume inductively that polygons of length $n-1$ have been defined, let $B(n-1)$ be the union of all $P \in \Pi$ of length $\leqq n-1$, and define $P \in \Pi$ to have length n if $P \cap B(n-1)$ is a union of edges. Then, *assuming that every $P \in \Pi$ has at least six sides*, each $B(n)$ is a topological ball, and in fact, if we build $B(n)$ from $B(n-1)$ by adding length n polygons in any order, the result is a ball at each stage.

Brenner and Lyndon made no assumptions that a group acts transitively on the polygons, which is the case for us. Our approach, however, allows us to find a "recursive structure" on the elements of G, thereby showing the rationality of the growth function (see below). Note that the proof of Lemma 2 also establishes the following result.

Lemma 3. *Let $n \geqq 0$, and $g, h \in G$ with $|g| = |h| = n+1$. Then $B(n) \cap gD$ is an interval, and $gD \cap hD$ is either empty or an interval which intersects $B(n)$ in a point (again excluding the $(3, 3, r)$-triangle groups).*

Let $g(z)$ be the growth series of G with respect to the generating set Σ. We know from [5] that g is the power series of a rational function f. Cannon proves this by giving a finite labelling of G by "cone-types", and using this to show that the coefficients of g satisfy a linear recursion. Using the fact that Σ is a geometric generating set, we will describe another finite labelling of G (analogous to Cannon's method in [6]) that shows that the coefficients of g satisfy a linear recursion. This gives an alternate proof in this case that g is the power series of a rational function, and makes the proof of our main theorem much easier.

For each $\Lambda_\omega(v)$ in Γ, denote by $\Lambda_\omega(\varphi(v))$ the image $\varphi(\Lambda_\omega(v))$, as a subgraph of K which is colored by Σ and has a base vertex labelled by $\varphi(v)$. We will partition the elements of G into finitely many types, as follows. The identity element of G is in a type by itself, and each element $g \in \Sigma$ (these are the elements in G of norm one) is the only element in its type. Let $g \in G$ with $|g| \geqq 2$. Then gx is joined by either one or two edges in K to vertices of norm $|g|-1$. Each of these edges is contained in exactly two Λ_ω's in K. Let $\Lambda(g)$ denote the union of these two or three Λ_ω's, as a colored subgraph of K with certain vertices labelled as base vertices. (Each Λ_ω has one, two, or three choices of labelled vertex, and we will call each such choice a base vertex. However, we will forget the particular labelling of each base vertex.) Given two elements $g, g' \in G$ with $|g| \geqq 2$ and $|g'| \geqq 2$, then g and g' are in the same type if and only if there is a color-preserving isomorphism $\tau\colon \Lambda(g) \to \Lambda(g')$ such that τ takes the

base vertices of $\Lambda(g)$ to the base vertices of $\Lambda(g')$ and $\tau(gx)=g'x$. Let $t(g)$ denote the type of g, and let $\{t_1, \ldots, t_r\}$ be an enumeration of the non-identity types.

Let $j \in \{1, \ldots, r\}$, and choose an element $g \in G$ with $t(g)=t_j$. We define $n(j)$ ($=1$ or 2) to be the cardinality of $\{h \in G: |h|=|g|-1$ and gx and hx are connected by an edge in $K\}$, and for each $i \in \{1, \ldots, r\}$ we define $m(i, j)$ to be the cardinality of $\{h \in G: |h|=|g|+1$, gx and hx are connected by an edge in K, and $t(h)=t_i\}$. Then $n(j)$ is the number of predecessors of g, and $m(i, j)$ is the number of elements of type t_i which have g as a predecessor.

We claim that $n(j)$ and $m(i, j)$ depend only on i and j and not on the representative element g. This is clear if $|g|=1$, since in this case g is the only element with type t_j. So assume that $|g| \geqq 2$. Then $n(j)=1$ if gx is in the boundary of exactly three edges in $\Lambda(g)$ (see Fig. 5), and $n(j)=2$ if gx is in the boundary of four edges in $\Lambda(g)$. In particular, $n(j)$ depends only on the type of g. Let $N(g)=\{hx: h \in G, |h|=|g|+1$, and there is an edge in K joining gx and $hx\}$ ($N(g)$ is the next generation of g). Then $m(i, j)$ is the cardinality of $\{h \in N(g): t(h)=t_i\}$, so we need to show that the distribution by types of the elements of $N(g)$ depends only on the type of g. Let $\Lambda'(g)$ be the union of $\Lambda(g)$ and all edges in K whose boundaries contain gx (see Fig. 21). Since $\Lambda'(g)$ has exactly one edge of each color emanating from gx, $\Lambda'(g)$ is determined up to color-preserving isomorphism by the type of g. Let $h \in N(g)$, and let e be the edge in $\Lambda'(g)$ joining gx and hx. Then e is contained in two Λ_ω's, Λ_w and $\Lambda_{w'}$, in $\Lambda(h)$, and by Lemma 1.iv) these Λ_ω's are determined by the color of e. If the union of these two Λ_ω's is $\Lambda(h)$, then the Λ_ω's in $\Lambda(h)$ are determined by the color of e. Suppose $\Lambda(h) \neq \Lambda_w \cup \Lambda_{w'}$, and hx is not in $\Lambda(g)$. Then h has a second predecessor, and, unless G is a $(3, 3, r)$-triangle group, we can assume without loss of generality that it is g_2, so we have Fig. 22.

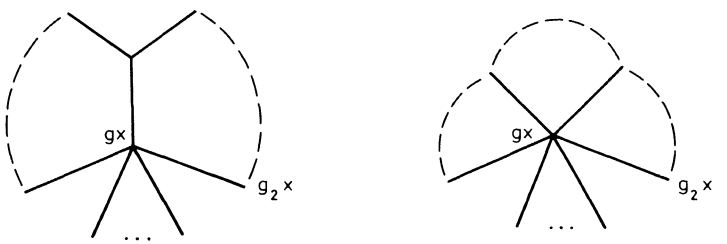


Fig. 21a, b a b

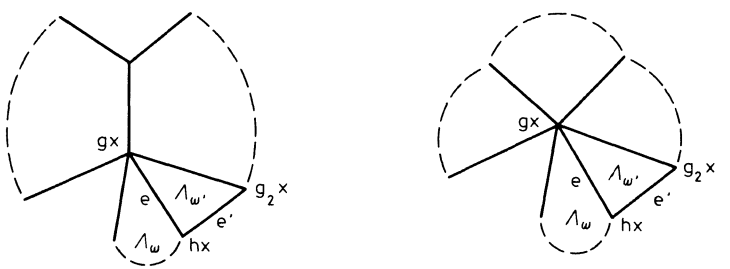


Fig. 22a, b a b

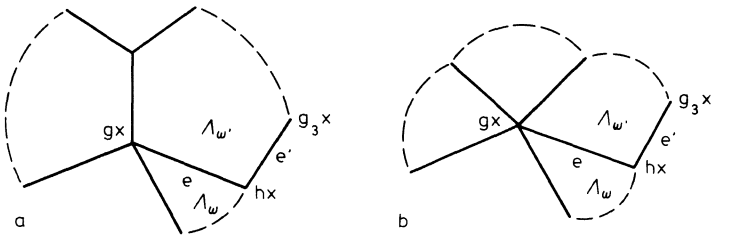


Fig. 23a, b

This occurs when $\Lambda_{w'}$ has length 3 and $|g_2|=|g|$. The third Λ_ω of $\Lambda(h)$ is the other Λ_ω attached to the edge e'. Since the color of e', the length of $\Lambda_{w'}$, and $|g_2|$ are determined by the type of g, in this case the Λ_ω's in $\Lambda(h)$ are determined by the color of e and the type of g.

The final case is that $\Lambda(h) \neq \Lambda_w \cup \Lambda_{w'}$ and $hx \in \Lambda(g)$. Without loss, we can assume that $h=g_2$, so we have Fig. 23. This occurs when $|g_3|=|g|$, and the third Λ_ω in $\Lambda(h)$ is determined by the color of e'. Since this information is determined by the type of g, in this case also the Λ_ω's in $\Lambda(h)$ are determined by the color of e and the type of g.

It remains to show that the base vertices of $\Lambda(h)$ are determined by the color of e and the type of g. It is easy to show that the base vertices of $\Lambda(h)$ are the obvious ones if G is not a $(3,3,r)$-triangle group or a $(3,4,r)$-triangle group. We leave it to the reader to check that the type of g determines the base vertices of $\Lambda(h)$ for the $(3,3,r)$- and $(3,4,r)$-triangle groups. Since the type of g determines the colors of the edges in $\Lambda'(g)$, each $m(i,j)$ depends only on the type t_j and not on the representative g of that type.

Let $\mathbf{A}$ be the $r \times r$ matrix with $\mathbf{A}_{ij}=m(i,j)/n(i)$, and let $\mathbf{v}$ be the $r \times 1$ matrix whose i^{th} entry is the cardinality of $\{g \in G\colon |g|=1 \text{ and } t(g)=t_i\}$. Then for each positive integer n, the matrix $\mathbf{A}^n\mathbf{v}$ has i^{th} entry the cardinality of $\{g \in G\colon |g|=n+1 \text{ and } t(g)=t_i\}$. Let $(1\ldots1)$ be the $1 \times r$ matrix with all entries 1. Then the nth coefficient a_n of g is $(1\ldots1)\mathbf{A}^{n-1}\mathbf{v}$.

Lemma 4. *With the notation as above, g is the Maclaurin series of the rational function $f(z)=1+(1\ldots1)(\mathbf{I}-z\mathbf{A})^{-1}z\mathbf{v}$.*

Proof. By the above,

$$\begin{aligned} g(z) &= 1+a_1z+\ldots+a_nz^n+\ldots \\ &= 1+(1\ldots1)z\mathbf{v}+\ldots+(1\ldots1)\mathbf{A}^{n-1}z^n\mathbf{v}+\ldots \\ &= 1+(1\ldots1)(\mathbf{I}+z\mathbf{A}+\ldots+z^{n-1}\mathbf{A}^{n-1}+\ldots)z\mathbf{v}. \end{aligned}$$

But this is the Maclaurin series of the rational function

$$f(z)=1+(1\ldots1)(\mathbf{I}-z\mathbf{A})^{-1}z\mathbf{v}. \quad \square$$

In computing examples, it is often possible to exploit symmetries in the Dirichlet region in order to reduce the number of types. For convenience, we will give a definition of the kind of structure we need. Let G be a finitely generated group, and Σ a finite generating set for G. A *recursive structure* for (G,Σ) is a

partition of the elements of G-{identity} into finitely many equivalence classes $\{t_1, \ldots, t_s\}$, called types, so that for each i, $j \in \{1, \ldots, s\}$ and $g \in G$ with type t_j, the cardinality, $n(j)$, of $\{h \in G: |h|_\Sigma = |g|_\Sigma - 1$ and g and h are connected by an edge in the graph of G with respect to $\Sigma\}$ and the cardinality, $m(i, j)$, of $\{h \in G: |h|_\Sigma = |g|_\Sigma + 1$, the type of h is t_i, and g and h are connected by an edge in the graph of G with respect to $\Sigma\}$ depend only on the type of g. The *recursion matrix* is the $s \times s$ matrix $\mathbf{A}$ with $\mathbf{A}_{ij} = m(i, j)/n(i)$ and the *initial vector* is the $s \times 1$ matrix $\mathbf{v}$ with i^{th} entry the cardinality of $\{g \in G: |g|_\Sigma = 1$ and the type of g is $t_i\}$. Note that a recursive structure is all that was used to prove Lemma 4.

§ 2. The Euler characteristic

Let G be a finitely generated, discrete group of isometries of $\mathbb{H}^2$ such that $\mathbb{H}^2/G$ has finite area, and Σ a finite generating set for G which is geometric with respect to the Dirichlet region $D = D_x$ and has no elements of order two. Let $\{t_1, \ldots, t_r\}$ be a recursive structure for (G, Σ), and for each $g \in G$-{identity} let $t(g)$ denote the type of g. Let $g \in G$-{identity}, and let $n = |g|$. Let $\sigma(g)$ be the number of edges of $g(D_x) \cap B(n-1)$, and let $\Theta(g)$ be the sum, over the vertices v in $g(D_x) \cap B(n-1)$, of the angle in $B(n-1)$ at the vertex v divided by the cardinality of $\{h \in G: |h| = n$ and $v \in h(D)\}$. So $\sigma(g)$ counts the number of sides with which $g(D)$ is attached to $B(n-1)$ and $\Theta(g)$ counts the portion of the angle sum of $\partial B(n-1)$ corresponding to $g(D_x) \cap B(n-1)$. The recursive structure is *geometric* if for each $g \in G$-{identity}, $\sigma(g)$ and $\Theta(g)$ depend only on the type of g. Note that the recursive structure defined in § 2 is clearly geometric. If the recursive structure is geometric, then the *attaching vector* is the $1 \times r$ matrix $\mathbf{w}$ whose j^{th} entry is $(\Theta(g)/\pi) - \sigma(g)$, for any $g \in G$ with type t_j.

Theorem 1. *Let G be a discrete group of isometries of $\mathbb{H}^2$ with $\mathbb{H}^2/G$ of finite area, Σ a geometric generating set for G with no elements of order two, and $\mathbf{A}$, $\mathbf{v}$, $\mathbf{w}$ determined as above by a geometric recursive structure for (G, Σ). Then the growth series for G with respect to Σ is the Maclaurin series of the rational function*

$$f(z) = 1/\chi + (1-z)(2\chi)^{-1}\mathbf{w}(\mathbf{I} - z\mathbf{A})^{-1}\mathbf{v},$$

where χ is the rational Euler characteristic of G (the orbifold Euler characteristic of $\mathbb{H}^2/G$).

Proof. At this point, it is somewhat more natural to work not with the growth series $g(z)$, but with the cumulative growth series $h(z) = \sum_{n=0}^{\infty} b_n z^n$, where $b_n = (1 + a_1 + \ldots + a_n)$ is the number of elements of G with length $\leqq n$. These are clearly related by $(1-z)h(z) = g(z)$.

For every $n \geqq 0$, let $\sigma(n)$ be the number of sides of $B(n)$ and let $\Theta(n)$ be the sum of the angles at the vertices of $B(n)$. Since $\mathbf{A}^n\mathbf{v}$ is the distribution by types of the elements of length $n+1$, $\mathbf{w}\mathbf{A}^n\mathbf{v} = (\Theta(n)/\pi) - \sigma(n)$. By the Gauss-Bonnet theorem and Lemma 2, the area of $B(n) = \pi(\sigma(n) - 2) - \Theta(n) = -\pi(2 + \mathbf{w}\mathbf{A}^n\mathbf{v})$. But also by Gauss-Bonnet, we have that the area of $D = \text{area}(\mathbb{H}^2/G) = -2\pi\chi$, hence the area of $B(n)$

$= -2\pi\chi b_n$, so that $b_n = 1/\chi + (1/2\chi)\mathbf{w}\mathbf{A}^n\mathbf{v}$. Thus,

$$h(z) = \sum_{n=0}^{\infty} (1/\chi + (1/2\chi)\mathbf{w}\mathbf{A}^n\mathbf{v})z^n$$
$$= 1/(1-z)\chi + (1/2\chi)\mathbf{w}(\mathbf{I} - z\mathbf{A})^{-1}\mathbf{v}.$$

Multiplying by $1-z$ now proves the theorem. □

An immediate consequence of Theorem 1 is that $f(1) = 1/\chi$ if $\mathbf{I} - \mathbf{A}$ is invertible. Since 1 is not an eigenvalue of A for a generic square matrix A, one might expect $f(1) = 1/\chi$ for most geometric generating sets for Fuchsian groups. As we will see in the next section, there are geometric generating sets for which $\mathbf{I} - \mathbf{A}$ is not invertible and $f(1) \neq 1/\chi(G)$.

§ 3. Examples

We want to describe some examples and computations of growth functions for Fuchsian groups with geometric generators Σ. We first consider growth functions associated to tessellations of $\mathbb{H}^2$ ($\mathbb{E}^2$, S^2) by *regular* n-gons, n even, $n \geqq 4$, with vertex angle $2\pi/p$, $p \geqq 3$. It is well-known that such regular polygons exist. Except for $n=4$, $p=3$ (spherical quadrilateral), $n=4$, $p=4$ (Euclidean square), and $n=6$, $p=3$ (Euclidean hexagon), they are hyperbolic. We will refer to them as $(n, 2\pi/p)$-gons.

Suppose an $(n, 2\pi/p)$-gon is the fundamental domain for a (torsion-free, closed) surface group. An Euler characteristic argument shows that if the surface is orientable, genus g, then $g = \frac{1}{2}\left(1 + \frac{n}{2} - \frac{n}{p}\right)$, so that $p|n$ and $\frac{n}{2} + \frac{n}{p}$ is odd. If the surface is non-orientable, genus g, then $g = 1 + \frac{n}{2} - \frac{n}{p}$, hence $p|n$.

Conversely, if $p|n$, an $(n, 2\pi/p)$-gon is a fundamental domain for a surface group. This is seen in the following way. If $g = \frac{1}{2}\left(1 + \frac{n}{2} - \frac{n}{p}\right)$ is an integer, let F_g be a closed, orientable surface of genus g. Now F_g has a spine with $2g$ edges and one vertex of valence $4g$. Perform Whitehead moves to create valence p vertices, one at a time, from the original vertex. When this has been done $\left(\frac{n}{p} - 1\right)$ times, we have a new spine with $\frac{n}{p}$ vertices, each of valence p. Equivalently, there are edge pairings on a polygon with $n = 2p(2g-1)/(p-2)$ sides so that each vertex cycle has length p. $\left(\text{If } \frac{1}{2}\left(1 + \frac{n}{2} - \frac{n}{p}\right) \text{ is not an integer, one can do the same construction on a non-orientable surface of genus } g = 1 + \frac{n}{2} - \frac{n}{p}.\right)$ Now let D be an $(n, 2\pi/p)$-gon in $\mathbb{H}^2$ (except for $(4, \pi/2)$ and $(6, 2\pi/3)$, which give tessellations of $\mathbb{E}^2$). Since D is regular, we can pick hyperbolic isometries to give the desired edge pairings, and now Poincaré's Theorem for Fundamental Polygons [19, 14] shows that the edge pairings generate a discrete, torsion-free group having D as a fundamental domain.

For fixed n, this gives only a finite collection of $(n, 2\pi/p)$-gons. We can do much better if we allow torsion in G. For example, suppose $(p, n) \geqq 3$, and pick q so that $q|p$, $q|n$, $q \geqq 3$. Then there is an $(n, 2\pi/q)$-gon which is the fundamental domain for a torsion-free group. Replace the $(n, 2\pi/q)$-gon by $D = (n, 2\pi/p)$-gon, keeping the same edge pairings. The angle sum for each vertex cycle is now $2\pi/(p/q)$, so Poincaré's Theorem again shows that the edge pairings generate a discrete group with D as fundamental domain.

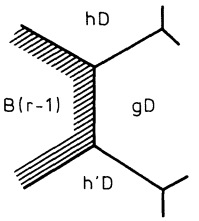
Of course, this construction has a lot of symmetry – every vertex has branching index p/q. When $(p, n) \leqq 2$, one could imagine that a less symmetric construction would produce an $(n, 2\pi/p)$-gon as fundamental domain. For example, let $n = 4$. If $4|p$, the edge pairings in Fig. 9c, d, or e, give a Fuchsian group G ($p > 4$), and $\mathbb{H}^2/G$ is a torus or Klein bottle with one vertex, branching index $p/4$. If 4 does not divide p but $2|p$, then Fig. 9a gives a $(p, p, p/2)$-triangle group. In the quotient S^2, two vertices have branching index p, one vertex has branching index $p/2$. If $3|p$, then Fig. 9b gives a discrete group with quotient the projective plane, one vertex having index p, the other having index $p/3$. (If $p = 3$, the group is $\mathbb{Z}_6 - S^2$ is tiled by six $(4, 2\pi/3)$-gons – but otherwise the group is Fuchsian.) But if p is odd and 3 does not divide p, the $(4, 2\pi/p)$-gon is *not* a fundamental domain for any discrete group (even allowing reflections).

Consider the general problem of tiling $\mathbb{H}^2$ by $(n, 2\pi/p)$-gons. If p is even, and we allow reflections, we simply use the Coxeter group based on the polygon $\langle a_1, \ldots, a_n | a_i^2, (a_i a_{i+1})^{p/2} \rangle$. The case $(p, n) = 1$ is clearly more complicated combinatorially, as the example above shows, and we do not know the exact conditions on p and n which allow the construction of a discrete group having the $(n, 2\pi/p)$-gon as a fundamental domain. *Nevertheless*, such tilings always exist. In fact, one can view the discussion which follows as a construction of the tilings. The type analysis shows how to inductively build the combinatorial ball $B(n)$, where length is taken in the sense of Brenner-Lyndon, as described earlier. It will be clear that each $B(n)$ has a non-singular, complete hyperbolic structure in its interior, and its boundary is a broken geodesic. Since each $B(n)$ is a topological ball, $\bigcup_{n=0}^{\infty} B(n)$ is a simply connected, complete, hyperbolic manifold, hence $\mathbb{H}^2$.

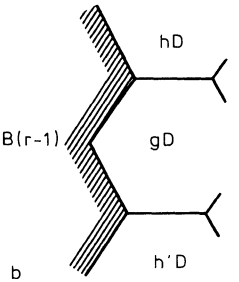
Thus, in what follows, we will not worry about the existence of a group G having the $(n, 2\pi/p)$-gon as a fundamental domain. The discussion is written as though G exists – if G does not exist, think of G as the set of polygons in the tiling, interpret length in the sense of Brenner-Lyndon, and let $\chi(G) = (-1/2\pi)\text{area}(D)$.

Given n ($n \geqq 4$) and p ($p \geqq 3$), we will find a geometric recursive structure for (G, Σ), where G is the discrete group generated by the edge pairing Σ on an $(n, 2\pi/p)$-gon, and explicitly calculate the initial vector $\mathbf{v}$, recursion matrix $\mathbf{A}$, and growth function f. We will see that the growth function depends only on n and p, and not on the choice of edge pairings. The argument is divided into the following four cases: (1) $p = 3$, (2) $p = 4$, (3) $p = 2k$, $k \in \mathbb{Z}$, $k > 2$, and (4) $p = 2k + 1$, $k \in \mathbb{Z}$, $k \geqq 2$.

Let $p = 3$. If $n = 4$ ($n = 5$), then the growth function is $1 + 4z + z^2(1 + 5z + 5z^2 + z^3)$. So assume $n \geqq 6$. Our recursive structure will have two types, t_1 and t_2. Let $g \in G$ with $|g|_\Sigma = r \geqq 1$. The type of g is t_1 if $g(D) \cap B(r-1)$ is a single edge e and both vertices of ∂e have angle $2\pi/3$ in $B(r-1)$. The type of g is t_2 if $g(D) \cap B(r-1)$ consists of two edges with a vertex in common, the interior vertex of gD



a



b

Fig. 24a, b

$\cap B(r-1)$ has angle $4\pi/3$ in $B(r-1)$, and the other two vertices each have angle $2\pi/3$ in $B(r-1)$. So $n(1)=1$ and $n(2)=2$. We need to show that every element in G-{identity} has one of these two types, and that this does give a recursive structure for (G, Σ). Suppose $r \geqq 1$, and $g \in G$ with $|g|=r$ and $t(g)=t_1$. By Lemmas 2 and 3, there exist h, $h' \in G$ so that $|h|=|h'|=r$ and the local picture near gD in $B(r)$ is that of Fig. 24a. Thus gx is connected by edges in K to $n-5$ vertices of type t_1 and to two vertices of type t_2. Similarly, if $|g|=r$ and $t(g)=t_2$, then there are elements h, h' in G so that the local picture near gD in $B(r)$ is that of Fig. 24b. In this case gx is connected by edges in K to $n-6$ vertices of type t_1 and to two vertices of type t_2. Since every element of norm one has type t_1, this shows inductively that every element of g has type t_1 or type t_2 and that $\{t_1, t_2\}$ gives a recursive structure with initial vector $\mathbf{v}=\begin{bmatrix} n \\ 0 \end{bmatrix}$ and recursion matrix $\mathbf{A}=\begin{bmatrix} n-5 & n-6 \\ 1 & 1 \end{bmatrix}$. By Lemma 4, the growth series g for G with respect to Σ is the Maclaurin series of the rational function $f(z)=\dfrac{z^2+4z+1}{z^2+(4-n)z+1}$. Notice that $n=6$ gives the hexagonal tiling of $\mathbb{E}^2$, and f has a double pole at $z=1$.

Now let $p=4$. An element $g \in G$ with $|g|=r \geqq 1$ has type t_1 if $gD \cap B(r-1)$ is a single edge with both vertices of angle $\pi/2$ in $B(r-1)$, and g has type t_2 if $gD \cap B(r-1)$ is two edges with a vertex in common, the two exterior vertices of $gD \cap B(r-1)$ have angle $\pi/2$ in $B(r-1)$, and the interior vertex has angle $3\pi/2$ in $B(r-1)$. An argument similar to the above shows that $\{t_1, t_2\}$ gives a recursive structure for (G, Σ) with initial vector $\mathbf{v}=\begin{bmatrix} n \\ 0 \end{bmatrix}$ and recursion matrix $\mathbf{A}=\begin{bmatrix} n-3 & n-4 \\ 1 & 1 \end{bmatrix}$. So the growth function $f(z)=\dfrac{z^2+2z+1}{z^2+(2-n)z+1}$. Notice that $n=4$ gives the tiling of $\mathbb{E}^2$ by squares, with the "usual" growth function $(1+z)^2/(1-z)^2$ for $\mathbb{Z}^2$.

Now suppose $p=2k$. In this case there are k types. For $i \in \{1, \ldots, k-1\}$, an element $g \in G$ with $|g|=r \geqq 1$ is of type t_i if $gD \cap B(r-1)$ is a single edge e, one vertex of ∂e has angle π/k in $B(r-1)$, and the other vertex has angle $(2i-1)\pi/k$ in $B(r-1)$. An element $g \in G$ is of type t_k if $gD \cap B(r-1)$ consists of two edges with a vertex in common, the exterior vertices of $gD \cap B(r-1)$ each have angle π/k in $B(r-1)$, and the interior vertex has angle $(2k-1)\pi/k$ in $B(r-1)$. Suppose $i \in \{1, \ldots, k-1\}$, and

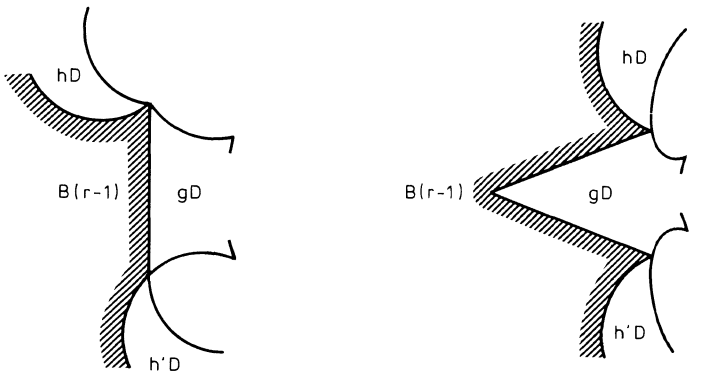


Fig. 25a, b

$g \in G$ with $t(g) = t_i$ and $|g| = r$. By Lemmas 2 and 3, there exist $h, h' \in G$ so that $|h| = |h'| = r$ and the local picture in $B(r)$ near gD is that of Fig. 25a. Thus gx is connected by edges in K to $n-1$ vertices of norm $r+1$, of which $n-3$ have type t_1, one has type t_2, and one has type t_{i+1}. If g has type t_k, then the local picture near gD in $B(r)$ is that of Fig. 25b. In this case gx is connected by edges in K to $n-2$ vertices of norm $r+1$, of which $n-4$ have type t_1 and two have type t_2. Since every element of norm one has type t_1, this shows that each element of $g \in G$-{identity} has type in $\{t_1, \ldots, t_k\}$. Hence $\{t_1, \ldots, t_k\}$ gives a recursive structure with initial vector

$$\mathbf{v} = \begin{bmatrix} n \\ 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{bmatrix}$$

and recursion matrix

$$\mathbf{A} = \begin{bmatrix} n-3 & n-3 & n-3 & n-3 & \ldots & n-3 & n-3 & n-4 \\ 2 & 1 & 1 & 1 & \ldots & 1 & 1 & 2 \\ 0 & 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & \ldots & 0 & 0 & 0 \\ \vdots & \vdots & & & \ddots & & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \ldots & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & \ldots & 0 & 1/2 & 0 \end{bmatrix},$$

and the growth function for G with respect to Σ is

$$f(z) = \frac{z^k + 2z^{k-1} + 2z^{k-2} + \ldots + 2z^2 + 2z + 1}{z^k + (2-n)z^{k-1} + (2-n)z^{k-2} + \ldots + (2-n)z^2 + (2-n)z + 1}.$$

The cases $n=4g$, $k=2g$ (orientable, genus g) and $n=2g$, $k=g$ (non-orientable, genus g), corresponding to (torsion-free) surface groups with one vertex, were first computed by Cannon [6].

When $p=2k+1$ there are $2k$ types. For $i\in\{1,\ldots,k\}$, an element $g\in G$ with $|g|=r\geqq 1$ has type $t(g)=t_i$ if $gD\cap B(r-1)$ is a single edge, with one vertex of angle $2\pi/(2k+1)$ in $B(r-1)$ and the other vertex of angle $(2i-1)2\pi/(2k+1)$. For $i\in\{k+1,\ldots,2k-1\}$, $t(g)=t_i$ if $gD\cap B(r-1)$ is a single edge, with one vertex of angle $2\pi/(2k+1)$ and the other of angle $4\pi(i-k)/(2k+1)$. Finally, g has type t_{2k} if $gD\cap B(r-1)$ is two edges with a vertex in common, the exterior vertices of $gD\cap B(r-1)$ each have angle $2\pi/(2k+1)$ in $B(r-1)$, and the interior vertex has angle $4k\pi/(2k+1)$ in $B(r-1)$. An argument like the above shows that this gives a recursive structure. The initial vector and recursion matrix are the same as the comparable size matrices for the case $p=2k$, except that in this case $A_{1k}=n-4$. The growth function

$$f(z)=\frac{z^{2k}+2z^{2k-1}+\ldots+2z^{k+1}+4z^k+2z^{k-1}+\ldots+2z+1}{z^{2k}+(2-n)z^{2k-1}+\ldots+(2-n)z^{k+1}+(4-n)z^k+(2-n)z^{k-1}+\ldots+(2-n)z+1}$$

In all of the above cases, the recursive structure is geometric and the recursion matrix $\mathbf{A}$ has $\mathbf{I}-\mathbf{A}$ invertible. By Theorem 1, we know (without explicitly checking) that $f(1)=1/\chi(G)$ in all of these examples. Note that the zeros of f are not all roots of unity, in contrast to the Coxeter groups [3] or the examples in [6].

We now give an example of a recursive structure for a surface group in which $\mathbf{I}-\mathbf{A}$ is not invertible. Let G be the fundamental group of a closed orientable surface of genus 2, and Σ a finite generating set for G coming from edge-pairing elements of a 16-gon whose image in the surface is shown in Fig. 26. By Poincaré's Theorem, we can assume that G is represented as a group of isometries of $\mathbb{H}^2$ and that Σ is a geometric generating set for G with respect to a Dirichlet region D for G in which each vertex has angle $2\pi/3$ or $\pi/2$. From Fig. 26, one sees that the angles alternate in the cycle $(2\pi/3\ 2\pi/3\ 2\pi/3\ \pi/2)$. Let $r\geqq 1$, and $g\in G$ with $|g|=r$. Then g is of type t_1 if $gD\cap B(r-1)$ is a single edge e, and both vertices of ∂e have angle $2\pi/3$ in $B(r-1)$. For $i\in\{2,3\}$, $t(g)=t_i$ if $gD\cap B(r-1)$ is a single edge e, one vertex of ∂e has angle $2\pi/3$ in $B(r-1)$, and the other vertex has angle $\pi(i-1)/2$ in $B(r-1)$. If $gD\cap B(r-1)$ is two edges with a vertex in common, then $t(g)\in\{t_4,t_5,t_6\}$. The type is t_4 if the interior vertex of $gD\cap B(r-1)$ has angle $4\pi/3$, one exterior vertex has angle $2\pi/3$, and the other exterior vertex has angle $\pi/2$. The type is t_5 if the interior vertex has angle $3\pi/2$ and each exterior vertex has angle $2\pi/3$, and $t(g)=t_6$ if the interior vertex has angle $4\pi/3$ and each exterior vertex has angle $2\pi/3$. An argument like the above shows that $\{t_1,\ldots,t_6\}$ gives a geometric recursive structure for (G,Σ) with initial vector

$$\mathbf{v}=\begin{bmatrix}8\\8\\0\\0\\0\\0\end{bmatrix}$$

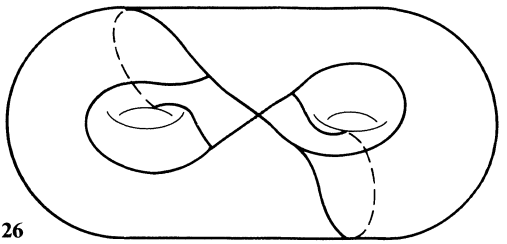


Fig. 26

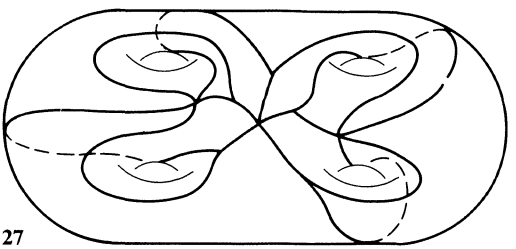


Fig. 27

and recursion matrix

$$\mathbf{A}=\begin{bmatrix} 6 & 6 & 5 & 6 & 4 & 6 \\ 5 & 6 & 6 & 5 & 6 & 4 \\ 1 & 0 & 0 & 0 & 0 & 2 \\ 1/2 & 1/2 & 1/2 & 1/2 & 1 & 0 \\ 0 & 1/2 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 & 0 \end{bmatrix}.$$

Thus the growth function

$$f(z)=\frac{z^4+5z^3+8z^2+5z+1}{z^4-11z^3-20z^2-11z+1}.$$

We note that even though $\mathbf{I}-\mathbf{A}$ is not invertible, we still find $f(1)=1/\chi$. In this case, $\mathbf{v}$ is in the image of $\mathbf{I}-\mathbf{A}$, say $(\mathbf{I}-\mathbf{A})\mathbf{v}_1=\mathbf{v}$. It is easy to rewrite the formula of Theorem 1 to give

$$f(z)=1/\chi+(1/2\chi)\left[(1-z)\mathbf{w}\mathbf{v}_1-(1-z)^2\mathbf{w}(\mathbf{I}-z\mathbf{A})^{-1}\mathbf{A}\mathbf{v}_1\right].$$

Since 1 has multiplicity 1 as an eigenvalue of A, we find $f(1)=1/\chi$.

Finally, we give an example in which $\mathbf{I}-\mathbf{A}$ is not invertible and $f(1)\neq 1/\chi$. Let G be the fundamental group of a closed orientable surface of genus 4, and Σ a finite generating set for G determined by the edge-pairing elements of a 30-gon whose image in the surface is shown in Fig. 27. By Poincaré's Theorem, we can represent G as a group of isometries of $\mathbb{H}^2$ and find a basepoint $x\in\mathbb{H}^2$ so that Σ is a geometric generating set with respect to the Dirichlet region $D=D_x$, where D has 30 sides of equal length, and the angles of D strictly alternate between $2\pi/3$ and

$2\pi/5$. We will give a geometric recursive structure $\{t_1, \ldots, t_5\}$ for (G, Σ). Let $g \in G$ with $|g| = r \geqq 1$. For $i \in \{1, 2, 3\}$, $t(g) = t_i$ if $gD \cap B(r-1)$ is a single edge e, one vertex of ∂e has angle $2\pi/3$ in $B(r-1)$, and the other vertex of ∂e has angle $2\pi i/5$ in $B(r-1)$. The type of g is t_4 if $gD \cap B(r-1)$ is two edges with a vertex in common, the interior vertex has angle $8\pi/5$ in $B(r-1)$ and each exterior vertex has angle $2\pi/3$ in $B(r-1)$. The type of g is t_5 if $gD \cap B(r-1)$ is two edges with a vertex in common, the interior vertex has angle $4\pi/3$ in $B(r-1)$, and each exterior vertex has angle $2\pi/5$ in $B(r-1)$. An argument like the above shows that this gives a geometric recursive structure with initial vector

$$\mathbf{v} = \begin{bmatrix} 30 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

and recursion matrix

$$\mathbf{A} = \begin{bmatrix} 26 & 26 & 25 & 24 & 26 \\ 1 & 1 & 1 & 2 & 0 \\ 1 & 0 & 0 & 0 & 2 \\ 0 & 1/2 & 0 & 0 & 0 \\ 0 & 0 & 1/2 & 0 & 0 \end{bmatrix}.$$

By Lemma 4, the growth function for G with respect to Σ is

$$f(z) = \frac{z^4 + 4z^3 + 7z^2 + 4z + 1}{z^4 - 26z^3 - 53z^2 - 26z + 1}.$$

Note that $\mathbf{I} - \mathbf{A}$ is not invertible, and $f(1) = -17/103 \neq -1/6 = 1/\chi$. So even for geometric generating sets for surface groups, it is not always true that $f(1) = 1/\chi$.

§ 4. Higher dimensions and open questions

We would like to explain briefly why Theorem 1 should have an analogue for even-dimensional co-finite volume hyperbolic groups. The two ingredients in the proof of Theorem 1 are:

(1) that $B(n)$ is a ball, and

(2) the Gauss-Bonnet formula, expressing area$(B(n))$ in terms of combinatorial data for $\partial B(n)$.

As far as (1) is concerned, nothing seems to be known in higher dimensions, except for Coxeter groups with the standard generating set, in which case $B(n)$ is a ball [8]. However, the generalized Gauss-Bonnet theorem does give an analogue for (2). For polyhedra in $\mathbb{H}^n$, this formula was known to Dehn [9], Klein [13, pp. 204–205], and certainly others.

To describe the formula, let B (a topological ball) be a polyhedron in $\mathbb{H}^n$, n even. Then ∂B is a union of lower-dimensional faces B^i_λ, $0 \leqq i \leqq n-1$, $1 \leqq \lambda \leqq m_i$, defined by the intersections of $n-i$ hyperplanes. Each face B^i_λ has an "inner angle" α^i_λ, $0 \leqq \alpha^i_\lambda \leqq 1$, defined as follows: Let $x \in B^i_\lambda$. The collection of unit tangent vectors to x which are normal to B^i_λ forms an $(n-i-1)$-dimensional sphere S^{n-i-1}, and α^i_λ is the fraction of S^{n-i-1} which points into B. This is independent of x. For example, the unique codimension zero face (B itself) has $\alpha^n = 1$, a codimension one face has $\alpha^{n-1}_\lambda = 1/2$, and a codimension two face has $\alpha^{n-2}_\lambda = \theta_\lambda/2\pi$, where θ_λ is the dihedral angle formed by the intersection of two hyperplanes. The Gauss-Bonnet formula states:

$$\text{volume}(B) = (-1)^{n/2} \frac{K_n}{2} \left[1 + \sum_{i=0}^{n-1} (-1)^i \sum_{\lambda=1}^{m} \alpha^i_\lambda \right],$$

where $K_n = \dfrac{2^n \left(\dfrac{n}{2} - 1\right)! \, \pi^{n/2}}{(n-1)!}$ is the volume of the unit n-sphere.

Now let G be a discrete group of isometries of $\mathbb{H}^n$, with finite volume, and let D be a Dirichlet region. By Cannon [5], the elements of G fall into finitely many "cone types", and we have a recursion matrix $\mathbf{A}$ and initial vector $\mathbf{v}$ which yield the growth function as in § 1. Each type is determined by local combinatorial data, in particular the way that a copy of D corresponding to an element of length n, and a fixed type, is attached to $B(n-1)$, hence each element of length n and a fixed type accounts for fixed angle contributions (of all codimensions $\geqq 1$) in $\partial B(n-1)$. Thus, as in § 2 there is an attaching vector $\mathbf{w}$ so that

$$\mathbf{w}\mathbf{A}^n\mathbf{v} = (-1)^{n/2} \frac{K_n}{2} \sum_{i=0}^{n-1} (-1)^i \sum_\lambda \alpha^i_\lambda = \text{vol}(B(n)) - (-1)^{n/2} \frac{K_n}{2}.$$

(Here we are assuming $B(n)$ is a ball.) But

$$\text{vol}(B(n)) = b_n \, \text{vol}(D) = b_n (-1)^{n/2} \frac{K_n}{2} \chi(G),$$

by the Gauss-Bonnet formula for $\mathbb{H}^n/G$. Hence

$$b_n = 1/\chi \left[1 + (-1)^{n/2} \frac{2}{K_n} \mathbf{w}\mathbf{A}^n\mathbf{v} \right].$$

As in Theorem 1, we find

$$f(z) = 1/\chi + (-1)^{n/2} \frac{2(1-z)}{K_n \chi} (\mathbf{w}, (\mathbf{I} - z\mathbf{A})^{-1}\mathbf{v}),$$

so one again expects (generically) that $f(1) = 1/\chi$.

In our view, the main open problems along the lines of this work are:

(1) Understand the "error term" $f(1) - 1/\chi$ when it is non-zero. In § 3 we gave one example – the fundamental group of a closed, orientable surface of genus 4,

and D is a 30-sided polygon whose angles alternate between $2\pi/3$ and $2\pi/5$. This is just one example of an infinite family. We have computed growth functions for all tilings of $\mathbb{H}^2$ by even-sided polygons whose angles alternate between $2\pi/p$ and $2\pi/q$. It turns out that the "error" is non-zero whenever both p and q are odd. We will describe these results elsewhere.

(2) Understand the symmetries of the growth functions. Notice that all of the examples in §3 give reciprocal functions: $f(z)=f(1/z)$. This symmetry in the coefficients of the numerator and denominator of $f(z)$ seems to be the rule. Actually, we do know of two classes of counterexamples – one class is the $(3, 3, r)$-triangle groups with 4-sided Dirichlet domains, which were already singled out in §1 as exceptional. In these cases, the denominators of $f(z)$ have symmetric coefficients, but not the numerators. Based on many computations, we are tempted to conjecture that for cocompact Fuchsian groups the denominators always have symmetric coefficients.

(3) Understand the poles of $f(z)$. This is closely related to (2), see Cannon [6]. How does the radius of convergence of $g(z)$ (= the smallest modulus of the poles of $f(z)$) change as the geometry of D changes? Can $f(z)$ have a pole at 1? While $f(z)$ often has only one pole in the open unit disk, there are examples for which $f(z)$ has more than one pole in the open disk. If $f(z)$ has only one pole in the open disk and the denominator has symmetric coefficients, then this pole is the reciprocal of a Salem number.

(4) Compute some higher dimensional examples. Some three-dimensional examples have been computed by Cannon and by Grayson (see [10]).

Note added in proof.
For cocompact planar groups with geometric generating sets, the authors have shown that $f(z)=f(1/2)$ except for the two classes of counterexamples mentioned above. The proof will appear in a paper titled Symmetries of planar growth functions.

References

1. Bass, H.: The degree of polynomial growth of finitely generated nilpotent groups. Proc. Lond. Math. Soc. **25**, 603–614 (1972)
2. Benson, M.: Growth Series of finite extension of $\mathbb{Z}^n$ are rational. Invent. math. **73**, 251–269 (1983)
3. Bourbaki, N.: Groupes et algèbres de Lie. Chapitres 4, 5 et 6. Paris: Hermann 1968
4. Brenner, J.L., Lyndon, R.C.: Infinite Eulerian tessellations. Discrete Math. **46**, 111–132 (1983)
5. Cannon, J.W.: The combinatorial structure of cocompact discrete hyperbolic groups. Geom. Dedicata **16**, 123–148 (1984)
6. Cannon, J.W.: The growth of the closed surface groups and the compact hyperbolic Coxeter groups. Preprint (1983)
7. Cannon, J.W.: Colored graphs. Preprint (1983)
8. Davis, M.: Groups generated by reflections and aspherical manifolds not covered by Euclidean space. Ann. Math. **117**, 293–324 (1983)
9. Dehn, M.: Die Eulersche Formel im Zusammenhang mit dem Inhalt in der Nicht-Euklidischen Geometrie. Math. Ann. **61**, 561–586 (1905)
10. Grayson, M.A.: Geometry and growth in three dimensions. Ph.D. thesis. Princeton University (1983)
11. Grigorchuk, R.I.: On Milnor's problem of group growth. Sov. Math., Dokl. **28**, 23–26 (1983)
12. Gromov, M.: Groups of polynomial growth and expanding maps. Publ. Math., Inst. Hautes Etud. Sci. **53**, 53–78 (1981)

13. Klein, F.: Vorlesungen über Nicht-Euklidische Geometrie. Die Grundlehren der Mathematischen Wissenschaften, Vol. 26. Berlin-Göttingen-Heidelberg: Springer 1928
14. Maskit, B.: On Poincaré's theorem for fundamental polygons. Adv. Math. **7**, 219–230 (1971)
15. Milnor, J.: Problem. Amer. Math. Monthly **75**, 685–686 (1968)
16. Milnor, J.: A note on curvature and fundamental groups. J. Differ. Geom. **2**, 1–7 (1968)
17. Milnor, J.: Growth of finitely generated solvable groups. J. Differ. Geom. **2**, 447–449 (1968)
18. Parry, W.: Counterexamples involving growth series and the Euler characteristic. Proc. Amer. Math. Soc. (to appear)
19. Poincaré, H.: Théorie des groupes fuchsiens. Acta Math. **1**, 1–62 (1882)
20. Serre, J-P.: Cohomologie des groups discrets, in Prospects in Mathematics. Ann. Math. Stud. No 70, Princeton, 1971
21. Smythe, N.: Growth functions and Euler series. Invent. math. **77**, 517–531 (1984)
22. Wagreich, P.: The growth function of a discrete group. Proc. Conference on Algebraic Varieties with Group Actions. Lecture Notes, Vol. 956, pp. 125–144 Berlin-Heidelberg-New York: Springer 1982
23. Wolf, J.A.: Growth of finitely generated groups and curvature of Riemannian manifolds. J. Differ. Geom. **2**, 421–446 (1968)

Oblatum 12-XI-1985 & VI-1986

Invent. math. 88, 31–38 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

A note on basic Iwasawa λ-invariants of imaginary quadratic fields

Kuniaki Horie*

Department of Mathematics, Tokyo Metropolitan University, Fukazawa, Setagaya-ku, Tokyo 158, Japan

Let p be a fixed prime number and let $\mathbb{Z}_p$ denote the p-adic integer ring. For each finite algebraic number field k, let h_k denote the class number of k; let λ_k and μ_k denote respectively the Iwasawa λ- and μ-invariants associated with the basic $\mathbb{Z}_p$-extension over k.

In this paper, we shall prove the following:

Theorem 1. *There exist infinitely many imaginary quadratic fields k such that p is not decomposed in k and does not divide h_k.*

Theorem 2. *There exist infinitely many imaginary quadratic fields k with $\lambda_k = \mu_k = 0$.*[1]

For the proof of these, it is sufficient only to show Theorem 1, because Iwasawa's criterion II in [12] says that Theorem 2 follows from Theorem 1 (moreover, when p is odd, the statements of the two theorems are equivalent. Cf. [1], [14]). In the case $p=2$, Theorem 1 is an immediate consequence of Gauss' genus theory. In the case $p>2$, our proof of Theorem 1 is based on the idea of Hartung [9] and on Eichler's trace formula combined with the p-adic Galois representation attached to the Jacobian variety $J=J_0(p)$ of the modular curve $X=X_0(p)$.

We shall also make some simple remarks on the value distribution of λ_k for imaginary quadratic fields k.

Throughout the paper, $\mathbb{C}$, $\mathbb{Q}_p$, $\mathbb{Q}$, $\mathbb{Z}$, and $\mathbb{N}$ denote the complex number field, the p-adic number field, the rational number field, the rational integer ring, and the set of positive integers, respectively. For any Galois extension K over $\mathbb{Q}$, we denote by $\mathrm{Gal}(K/\mathbb{Q})$ the Galois group of $K/\mathbb{Q}$. If K is a finite abelian extension over $\mathbb{Q}$, then we write f_K for the conductor of K. We shall use the Kronecker symbol $(-)$ and also the order notations O and o of Landau.

Now, let A be a commutative ring with unity. Then, for each $n\in\mathbb{N}$, $\mathrm{GL}_n(A)$

* Supported in part by a grant from the Yukawa Foundation

[1] As is well known, $\mu_K=0$ for every finite abelian extension K over the rational number field $\mathbb{Q}$ ([7])

denotes the general linear group of degree n over A and $\mathrm{SL}_n(A)$ the special linear group of degree n over A. For any square matrix M with coefficients in A, we let $\mathrm{tr}(M)$ denote the trace of M and $\det(M)$ the determinant of M.

Acknowledgements. The author would like to express his heartfelt gratitude to Professor Y. Yamamoto for very helpful suggestions, in particular, for calling the author's attension to the reference [19]. He is also thankful to his wife Mitsuko for her kind advice.

1. Proof of Theorem 1

By an argument in the introduction, we may assume that p is odd. Let $\Gamma_0(p)$ denote, as usual, the congruence modular group defined by

$$\Gamma_0(p)=\left\{\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in \mathrm{SL}_2(\mathbb{Z}) \mid c\equiv 0 \pmod p\right\}.$$

Let X be the modular curve associated with $\Gamma_0(p)$, g the genus of X, and J the Jacobian variety of X. Not only X but also J is defined over $\mathbb{Q}$. For each $m\in\mathbb{N}$, let J_m denote the finite subgroup of J consisting of all points in J of order dividing m, and $\mathbb{K}[m]$ the smallest algebraic number field over which the elements of J_m are rational, i.e., the field generated over $\mathbb{Q}$ by the affine coordinates of all points in J_m. Then $\mathbb{K}[m]$ is a finite Galois extension over $\mathbb{Q}$, with $\mathrm{Gal}(\mathbb{K}[m]/\mathbb{Q})$ acting faithfully on J_m. It is known that J_m is a free $\mathbb{Z}/m\mathbb{Z}$-module of rank $2g$. Hence the compositum

$$\mathbb{K}[p^\infty]=\bigcup_{n=1}^{\infty}\mathbb{K}[p^n]$$

becomes a Galois extension over $\mathbb{Q}$ and $\mathrm{Gal}(\mathbb{K}[p^\infty]/\mathbb{Q})$ acts continuously and faithfully on

$$G_p(J)=\bigcup_{n=1}^{\infty}J_{p^n},$$

which is a free $\mathbb{Q}_p/\mathbb{Z}_p$-module of rank $2g$. Thus, taking a p-adic coordinate system $\iota\colon G_p(J)\xrightarrow{\sim}(\mathbb{Q}_p/\mathbb{Z}_p)^{2g}$, we obtain a faithful p-adic representation

$$R_\infty\colon \mathrm{Gal}(\mathbb{K}[p^\infty]/\mathbb{Q})\to \mathrm{GL}_{2g}(\mathbb{Z}_p).$$

At the same time, the action of $\mathrm{Gal}(\mathbb{K}[p]/\mathbb{Q})$ on J_p, together with the restriction of ι to J_p, induces a faithful representation

$$R\colon \mathrm{Gal}(\mathbb{K}[p]/\mathbb{Q})\to \mathrm{GL}_{2g}(\mathbb{Z}/p\mathbb{Z}).$$

We then have, for every $\sigma\in\mathrm{Gal}(\mathbb{K}[p^\infty]/\mathbb{Q})$,

$$R(\sigma|\mathbb{K}[p])=R_\infty(\sigma) \bmod p \tag{1}$$

where $\sigma|\mathbb{K}[p]$ denotes the restriction of σ to $\mathbb{K}[p]$ and the right hand side the reduction (mod p) of $R_\infty(\sigma)\in\mathrm{GL}_{2g}(\mathbb{Z}_p)$.

Meanwhile, let $\mathfrak{S}=\mathfrak{S}_2(\Gamma_0(p))$ denote the space of cusp forms of weight 2 with respect to $\Gamma_0(p)$. It is known that $\mathfrak{S}$ is a $\mathbb{C}$-vektor space of dimension

g. At least on $\mathfrak{S}$, we understand the Hecke operator T_n, for each $n \in \mathbb{N}$, to be the linear transformation which maps each $\phi = \phi(s) \in \mathfrak{S}$ to

$$\phi \mid T_n = n \sum_{(a,d)} \sum_{b=1}^{d} \phi((as+b)d^{-1}) d^{-2} \in \mathfrak{S},$$

where (a, d) ranges over the pairs of positive integers such that $ad = n$ and $p \nmid a$. Let us fix a basis of $\mathfrak{S}$ over $\mathbb{C}$, so that each T_n is represented by a matrix of degree g with coefficients in $\mathbb{C}$. We denote this matrix again by T_n. The trace $\mathrm{tr}(T_n)$ is known to be an integer.

Now, let $\mathfrak{Q}$ be any finite prime of $\mathbb{K}[p^\infty]$ not lying above p. Since J has good reduction outside p, $\mathfrak{Q}$ is unramified for the extension $\mathbb{K}[p^\infty]/\mathbb{Q}$ (cf. [11, 17]). Hence there exists a unique Frobenius automorphism of $\mathfrak{Q}$ for $\mathbb{K}[p^\infty]/\mathbb{Q}$, which we denote by $\sigma_{\mathfrak{Q}}$. Let x be an indeterminate and q the prime number corresponding to the restriction of $\mathfrak{Q}$ on $\mathbb{Q}$. Then, by the congruence relations for modular correspondences,

$$\det(1_{2g} - R_\infty(\sigma_{\mathfrak{Q}})\, x) = \det(1_g - T_q\, x + q x^2)$$

(see [4] and [18]). In particular, comparing the coefficients of x on both sides, we have

$$\mathrm{tr}(R_\infty(\sigma_{\mathfrak{Q}})) = \mathrm{tr}(T_q). \tag{2}$$

We next recall Eichler's trace formula for T_q (cf. [5]):

$$\mathrm{tr}(T_q) = q - 1 - \sum_{(a,b)} \left(1 + \left\{\frac{(a^2-4q)\,b^{-2}}{p}\right\}\right) \frac{h((a^2-4q)\,b^{-2})}{w((a^2-4q)\,b^{-2})}. \tag{3}$$

Here the sum is taken over the pairs $(a, b) \in \mathbb{Z} \times \mathbb{N}$ such that $a^2 - 4q < 0, b^2 \mid (a^2-4q)$, and $(a^2-4q)\,b^{-2} \equiv 0$ or $1 \pmod 4$; further, for each negative integer $\Delta \equiv 0$ or $1 \pmod 4$,

$$\left\{\frac{\Delta}{p}\right\} = 1 \qquad \text{if } p^2 \mid \Delta \text{ and } \Delta p^{-2} \equiv 0 \text{ or } 1 \pmod 4,$$

$$= \left(\frac{\Delta}{p}\right) \qquad \text{otherwise,}$$

$h(\Delta)$ denoting the class number of the order of $\mathbb{Q}(\sqrt{\Delta})$ with discriminant Δ and $w(\Delta)$ the number of roots of unity in this order. Therefore we immediately obtain

$$\mathrm{tr}(T_q) \leqq q - 1. \tag{4}$$

Let us also recall Kronecker's class number relation (cf. [16] or [5]);

$$q = \sum_{(a,b)} \frac{h((a^2-4q)\,b^{-2})}{w((a^2-4q)\,b^{-2})},$$

where the sum is taken over the same pairs $(a, b) \in \mathbb{Z} \times \mathbb{N}$ as in (3). It then follows from this and (3) that

$$\mathrm{tr}(T_q) + q + 1 = \sum_{(a,b)} \left(1 - \left\{\frac{(a^2-4q)\,b^{-2}}{p}\right\}\right) \frac{h((a^2-4q)\,b^{-2})}{w((a^2-4q)\,b^{-2})}. \tag{5}$$

The following lemma will soon be useful.

Lemma. *Assume that* $p \geqq 5$, *and let* l *be the least positive quadratic non-residue* (mod p). *Then* l *is a prime number* $< \frac{p}{2}$ *such that*

$$\mathrm{tr}\,(T_l) + l + 1 \not\equiv 0 \pmod{p}.$$

Proof. It is obvious that l is a prime number. Since $p \geqq 5$, we can take $n \in \mathbb{N}$ such that $\frac{p}{2} < n^2 < p$. This implies that

$$l < \frac{p}{2}. \tag{6}$$

Next, substituting l for q in (5), we have

$$\mathrm{tr}\,(T_l) + l + 1 = \sum_{(a,b)} \left(1 - \left\{\frac{(a^2 - 4l)\,b^{-2}}{p}\right\}\right) \frac{h((a^2-4l)\,b^{-2})}{w((a^2-4l)\,b^{-2})},$$

where (a, b) ranges over the pairs of integers such that $a^2 < 4l$, $b > 0$, and $(a^2 - 4l)\,b^{-2}$ is an integer $\equiv 0$ or $1 \pmod 4$. It follows therefore that

$$\mathrm{tr}\,(T_l) + l + 1 \geqq \left(1 - \left(\frac{-l}{p}\right)\right) \frac{h(-4l)}{w(-4l)} + \left(1 - \left(\frac{1-l}{p}\right)\right) \frac{h(4-4l)}{w(4-4l)}.$$

We note here that $\left(\frac{l}{p}\right) = -1$ but $\left(\frac{l-1}{p}\right) = 1$. Hence the right hand side in the above inequality is always positive, so that

$$\mathrm{tr}\,(T_l) + l + 1 > 0.$$

On the other hand, by (4) and (6),

$$\mathrm{tr}\,(T_l) + l + 1 < p.$$

Since $\mathrm{tr}\,(T_l) \in \mathbb{Z}$, we then obtain

$$\mathrm{tr}\,(T_l) + l + 1 \not\equiv 0 \pmod{p}.$$

Thus the lemma is proved.

We are now ready to prove Theorem 1. Let us first consider the case $p \geqq 5$ (with assuming $g \geqq 1$). Take an arbitrary prime number r. Let l be again the least positive quadratic non-residue (mod p), let $\mathfrak{L}$ be a prime of $\mathbb{K}[p^\infty]$ lying above l, and let, for each odd prime v,

$$v^* = \left(\frac{-1}{v}\right) v.$$

Note that no prime number but p is ramified in $\mathbb{K}[p]$. The Tschebotareff density theorem then shows that there exists a prime number q, different from l and from p, with the following properties:

(i) For some prime $\mathfrak{Q}$ of $\mathbb{K}[p^\infty]$ lying above q,

$$\sigma_{\mathfrak{Q}} | \mathbb{K}[p] = \sigma_{\mathfrak{L}} | \mathbb{K}[p],$$

(ii) q remains prime in $\mathbb{Q}(\sqrt{-1})$, $\mathbb{Q}(\sqrt{-2})$, and $\mathbb{Q}(\sqrt{v^*})$ for every odd prime $v \leqq r$ different from p.[2]

The property (i) implies, by (1),

$$R_\infty(\sigma_{\mathfrak{L}}) \bmod p = R_\infty(\sigma_{\mathfrak{L}}) \bmod p.$$

Hence we obtain from (2) that

$$\operatorname{tr}(T_q) \equiv \operatorname{tr}(T_l) \pmod{p}. \tag{7}$$

Furthermore, $\mathbb{K}[p]$ is known to contain the cyclotomic field $\mathbb{K}_p$ of p-th roots of unity (cf. [20]). Therefore, restricting the equality in (i) to $\mathbb{K}_p$, we have

$$q \equiv l \pmod{p}.^3$$

Together with this and (7), Lemma shows that

$$\operatorname{tr}(T_q) + q + 1 \not\equiv 0 \pmod{p}.$$

Hence, by (5),

$$\frac{h((a_0^2 - 4q)\, b_0^{-2})}{w((a_0^2 - 4q)\, b_0^{-2})} \not\equiv 0 \pmod{p} \tag{8}$$

holds for some pair $(a_0, b_0) \in \mathbb{Z} \times \mathbb{N}$ such that $a_0^2 < 4q$, $b_0^2 \mid (a_0^2 - 4q)$, $(a_0^2 - 4q)\, b_0^{-2} \equiv 0$ or $1 \pmod 4$, and

$$\left\{\frac{(a_0^2 - 4q)\, b_0^{-2}}{p}\right\} \neq 1. \tag{9}$$

We now take an imaginary quadratic field

$$k = \mathbb{Q}(\sqrt{a_0^2 - 4q}).$$

Let u be the positive integer determined by

$$-f_k\, u^2 = (a_0^2 - 4q)\, b_0^{-2}. \tag{10}$$

Then

$$\frac{h((a_0^2 - 4q)\, b_0^{-2})}{w((a_0^2 - 4q)\, b_0^{-2})} = \frac{h(-f_k)}{w(-f_k)}\, u \prod_{v \mid u} \left(1 - \left(\frac{-f_k}{v}\right) v^{-1}\right),$$

with the product on the right taken over the prime divisors v of u (cf. [3]). Here, as $w(-f_k)$ divides 12, it follows from (8) that the class number $h_k = h(-f_k)$ is prime to p. We also have, by (9),

$$\left(\frac{-f_k}{p}\right) \neq 1.$$

In addition to the above, let us show that

$$f_k > r. \tag{11}$$

---

[2] Here we have used the fact that, among the fields in (i), (ii), any one is disjoint over $\mathbb{Q}$ from the compositum of the others

[3] In the case where $g = 0$ (i.e., $\mathbb{K}[p^\infty] = \mathbb{Q}$) and $p \geqq 5$, one can prove Theorem 1 similarly, adding this congruence to the property (i)

As (10) means, $4q$ is the norm of $a_0+b_0u\sqrt{-f_k}$ for $k/\mathbb{Q}$; so that q does not remain prime in k. Therefore, by (ii), it is impossible that $f_k=4$ or 8. There exists then an odd prime w dividing f_k. Furthermore, (10) implies $\left(\frac{q}{w}\right)\neq -1$. On the other hand, by (i), $\left(\frac{q}{p}\right)=\left(\frac{l}{p}\right)=-1$ and, by (ii), $\left(\frac{q}{v}\right)=-1$ for every odd prime $v\leqq r$ different from p. Consequently, it follows that $f_k\geqq w>r$.

After all, we obtain an imaginary quadratic field k such that p is not decomposed in k, h_k is prime to p, and the inequality (11) holds. Since r is an arbitrary prime number, we have proved Theorem 1 in the case $p\geqq 5$.

Next, suppose that $p=3$. For each $m\in\mathbb{N}$, let $\sigma_1(m)$ denote the sum of all positive integers dividing m. The following class number relation is due to Gierster [8] and Hurwitz [10]:

If m is a positive integer $\equiv 1 \pmod 3$, then

$$4\sum_{\substack{(a,b)\\ 3\mid a}}\frac{h((a^2-4m)b^{-2})}{w((a^2-4m)b^{-2})}=\sigma_1(m), \tag{12}$$

where the sum on the left is taken over the pairs $(a, b)\in\mathbb{Z}\times\mathbb{N}$ such that $a^2<4m$, $3\mid a$, $b^2\mid(a^2-4m)$, and $(a^2-4m)b^{-2}\equiv 0$ or $1 \pmod 4$.

Now, let r be any prime number. There exists then a prime number q such that $q\equiv 1 \pmod 3$, $\left(\frac{-1}{q}\right)=\left(\frac{-2}{q}\right)=-1$, and $\left(\frac{q}{v}\right)=-1$ for every odd prime $v\leqq r$ different from 3, q. In the above formula (12), we substitute q for m:

$$\sum_{\substack{(a,b)\\ 3\mid a}}\frac{4}{w((a^2-4q)b^{-2})}h((a^2-4q)b^{-2})=q+1.$$

Here, in each summand on the left, $w((a^2-4q)b^{-2})$ divides 4 since $(a^2-4q)b^{-2}\equiv 2 \pmod 3$. Hence we have

$$h((a_0^2-4q)b_0^{-2})\not\equiv 0 \pmod 3$$

for some pair $(a_0, b_0)\in\mathbb{Z}\times\mathbb{N}$ satisfying $3\mid a_0$, $a_0^2<4q$, $b_0^2\mid(a_0^2-4q)$, and $(a_0^2-4q)b_0^{-2}\equiv 0$ or $1 \pmod 4$. Let

$$k=\mathbb{Q}(\sqrt{a_0^2-4q})$$

so that

$$\left(\frac{-f_k}{3}\right)=\left(\frac{a_0^2-4q}{3}\right)=-1,\qquad w(-f_k)\mid 4.$$

Then, by an argument similar to that in the case $p\geqq 5$, we can conclude that $f_k>r$ as well as that 3 remains prime in k and does not divide h_k. Theorem 1 is therefore proved also when $p=3$.

Thus we have done with the proof of Theorem 1.

Remark 1. Let k be one of the imaginary quadratic fields provided by Theorem 1. Actually, it then follows from [12] that, for any $\mathbb{Z}_p$-extension K over k, the Iwasawa λ- and μ-invariants associated with K/k are both vanishing.

Remark 2. Not only when $p=3$ but also when $5 \leqq p \leqq 13$, using some class number relations due to Gierster [8] and Hurwitz [10], we can show that there exist infinitely many imaginary quadratic fields k in which p remains prime and for which p does not divide h_k.

2. Additional remarks

Let K be any imaginary biquadratic field and let

$$\lambda_K^- = \lambda_K - \lambda_{K^+},$$

where K^+ denotes the real quadratic subfield of K. Let k, k' be the imaginary quadratic subfields of K: $K=kk'$. Now, assume that $K^+ \neq \mathbb{Q}(\sqrt{2})$ if $p=2$. Then, by the analytic class number formula and by Iwasawa's class number formula (cf. [13]), we have

$$\lambda_K^- = \lambda_k + \lambda_{k'}.$$

In particular, $\lambda_K^- = 0$ if and only if $\lambda_k = \lambda_{k'} = 0$.

We thus obtain, from Theorem 2, the following result.

Corollary. *There exist infinitely many imaginary biquadratic fields K such that $\lambda_K^- = 0$.*

The rest of the paper owes much to conversations with J. Nakagawa who has kindly told us the review [9'] and his results related to this reference. Here we wish to thank him sincerely.

Let N be a fixed positive integer. For each $x>0$, let $\mathbf{q}(x)$ denote the number of imaginary quadratic fields with conductor $\leqq x$ and $\mathbf{q}^*(x)$ the number of imaginary quadratic fields k with $f_k \leqq x$ such that $\lambda_k \geqq N$. Then [6, 15] and a simple calculation in analytic number theory verify

$$\lim_{x\to\infty} \frac{\mathbf{q}^*(x)}{\mathbf{q}(x)} = 1$$

for the case $p=2$;

$$\mathbf{q}(x) - \mathbf{q}^*(x) = \frac{3x(\log\log x)^{N-1}}{(N-1)!\, 2^{N+2} \log x} + O\left(\frac{x(\log\log x)^{N-2}}{\log x}\right), \qquad x \to \infty,$$

while

$$\mathbf{q}(x) = 3\pi^{-2}x + o(x), \qquad x \to \infty.$$

It may be impossible, however, to expect a similar result for the case $p>2$. Indeed, by checking the proofs in [2], we can show that, if $p=3$, then

$$\limsup_{x\to\infty} \frac{\mathbf{q}^*(x)}{\mathbf{q}(x)} \leqq \frac{11}{16}.$$

On the other hand, in the special case where $N=1$ and $p>2$,

$$\liminf_{x\to\infty} \frac{\mathbf{q}^*(x)}{\mathbf{q}(x)} \geqq \frac{p}{2(p+1)},$$

because $\lambda_k \geqq 1$ for every imaginary quadratic field k in which p is decomposed (cf. [14]).

Finally we add another (essentially known) fact as follows: *if* $N=2$, *then* $\mathbf{q}^*(x) \to \infty$ *as* $x \to \infty$. *Namely, there exist infinitely many imaginary quadratic fields* k *such that* $\lambda_k \geqq 2$.

For the proof, we may assume $p>2$. It is known that, for every imaginary quadratic field k, λ_k is at least equal to the p-rank of the ideal class group of k. However, according to [21], there exist infinitely many imaginary quadratic fields whose ideal class groups are of p-rank $\geqq 2$.

References

1. Coates, J., Lichtenbaum, S.: On l-adic zeta functions. Ann. Math. **98**, 498–550 (1973)
2. Davenport, H., Heilbronn, H.: On the density of discriminants of cubic fields II. Proc. R. Soc. Lond. Ser. A **322**, 405–420 (1971)
3. Dirichlet, P.G.L., Dedekind, R.: Vorlesungen über Zahlentheorie (corrected reprint). New York: Chelsea 1968
4. Eichler, M.: Quaternäre quadratische Formen und die Riemannsche Vermutung für die Kongruenzzetafunktion. Arch. Math. **5**, 355–366 (1954)
5. Eichler, M.: Über die Darstellbarkeit von Modulformen durch Thetareihen. J. Reine Angew. Math. **195**, 156–171 (1956)
6. Ferrero, B.: The cyclotomic $\mathbb{Z}_2$-extension of imaginary quadratic fields. Am. J. Math. **102**, 447–459 (1980)
7. Ferrero, B., Washington, L.C.: The Iwasawa invariant μ_p vanishes for abelian number fields. Ann. Math. **109**, 377–395 (1979)
8. Gierster, J.: Über Relationen zwischen Klassenzahlen binärer quadratischer Formen von negativer Determinante (Erste Abhandlung, Zweite Abhandlung). Math. Ann. **21**, 1–50 (1883); **22**, 190–210 (1883)
9. Hartung, P.: Proof of the existence of infinitely many imaginary quadratic fields whose class number is not divisible by 3. J. Number Theory **6**, 276–278 (1974)

9′. Hartung, P. (Weinberger, P.J.): MR **50** #4528

10. Hurwitz, A.: Über Relationen zwischen Klassenzahlen binärer quadratischer Formen von negativer Determinante. Math. Ann. **25**, 157–196 (1885)
11. Igusa, J.: Kroneckerian model of fields of elliptic modular functions. Am. J. Math. **81**, 561–577 (1959)
12. Iwasawa, K.: A note on class numbers of algebraic number fields. Abh. Math. Sem. Univ. Hamburg **20**, 257–258 (1956)
13. Iwasawa, K.: On Γ-extensions of algebraic number fields. Bull. Am. Math. Soc. **65**, 183–226 (1959)
14. Iwasawa, K.: On $\mathbb{Z}_l$-extensions of algebraic number fields. Ann. Math. **98**, 246–326 (1973)
15. Kida, Y.: On cyclotomic $\mathbb{Z}_2$-extensions of imaginary quadratic fields. Tôhoku Math. J. **31**, 91–96 (1979)
16. Kronecker, L.: Über die Anzahl der verschiedenen Klassen quadratischer Formen von negativer Determinante. J. Reine Angew. Math. **57**, 248–255 (1860)
17. Serre, J.-P., Tate, J.: Good reduction of abelian varieties. Ann. Math. **88**, 492–517 (1968)
18. Shimura, G.: Correspondences modulaires et les fonctions ζ de courbes algébriques. J. Math. Soc. Japan **10**, 1–28 (1958)
19. Shimura, G.: Automorphic functions and number theory I, II (in Japanese). Sûgaku **11**, 193–205 (1960); **13**, 65–80 (1961)
20. Weil, A.: Courbes algébriques et variétés abéliennes. Paris: Hermann 1971
21. Yamamoto, Y.: On unramified Galois extensions of quadratic number fields. Osaka J. Math. **7**, 57–76 (1970)

Oblatum 3-I-1986 & 14-VII-1986

Invent. math. 88, 39-63 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

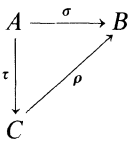
On the approximation property of excellent rings

Christel Rotthaus
Department of Mathematics, Michigan State University, East Lansing, MI 48824-1027, USA

In [3] Artin has introduced the following problem:

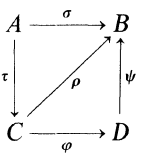
If $\sigma: A \to B$ is a regular morphism of noetherian rings factoring over an A-algebra C of finite type:

(*)



is then (*) embedded in a commutative diagram:

(**)



where D is a smooth finite type A-algebra?

A theorem, which would imply the existence of an embedding of the above type, yields the positive solution of the Bass-Quillen conjecture. So far this theorem is only known in special cases. There is the following result by Artin [3, 4].

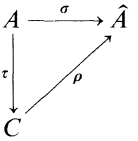
Theorem. *The power series ring $R[[\underline{X}]]$ in finitely many variables over an excellent discrete valuation ring R is a direct limit of smooth R-algebras.*

In this paper another special case of this smoothing problem is considered. More precisely we prove the following:

Let A be a semilocal ring whose generic formal fibres are geometrically regular and suppose that either $\mathbb{Q} \subseteq A$ or that certain separability conditions are satisfied. If $\hat{A}$ is the completion of A with respect to the Jacobson radical of

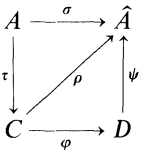
A, then every commutative diagram

(*)



where C is an A-algebra of finite type, can be embedded in a commutative diagram:

(**)



such that the following conditions are satisfied:

(i) D is a finitely generated A-algebra

(ii) D is smooth over A at every prime ideal of Spec D lying under a prime ideal of the generic fibres of $\hat{A}$, this means: for every $\mathfrak{p} \in \operatorname{Spec} \hat{A}$ such that $\mathfrak{p} \cap A$ is a minimal prime ideal in A, the A-algebra $D_{\psi^{-1}(\mathfrak{p})}$ is smooth over A.

To prove this theorem we have to make use of the above-mentioned theorem by Artin. As its application we show in §4 that excellent henselian rings containing the rationals satisfy the approximation property.

The result by Artin is discussed in several other papers [14, 19] but their proofs seem to be difficult to follow. The interested reader should also compare our result with Corollary (5.4), p. 108 of [20]. In his paper, Popescu is interested in proving the smoothing problem in case of rings containing the rationals. However, it is very difficult to understand his arguments.

§ 1. Preliminaries

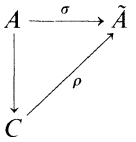
In the following we use the notations and definitions of EGA [10–12] and of [16]. Thus as in [11] by an excellent semilocal ring we mean a semilocal universally catenary ring whose formal fibres are geometrically regular. In particular we would like to mention the following definitions:

(1.1) *Definition.* Assume A is a semilocal noetherian ring and $\hat{A}$ is the completion of A with respect to the Jacobson radical of A. Then the generic formal fibres of A are geometrically regular if for every minimal prime ideal $\mathfrak{g}$ of A the ring $\hat{A} \otimes_A k(\mathfrak{g})$ is geometrically regular over the quotient field $k(\mathfrak{g})$ of $A/\mathfrak{g}$.

(1.2) *Definition.* Let $\sigma: A \to \tilde{A}$ be an injective morphism of noetherian rings.

(a) Assume there is a factorization of σ over an A-algebra C of finite type

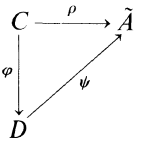
(1.2.1)



Then we say:

(a.1) ρ (or C resp.) factors over a finitely generated smooth A-algebra D if there exists a finitely generated smooth A-algebra D and A-algebra morphisms $\varphi\colon C\to D$ and $\psi\colon D\to\tilde{A}$ such that the diagram

(1.2.2)



commutes.

(a.2) ρ (or C resp.) factors over a generically smooth A-algebra D if there exist an A-algebra D of finite type and A-algebra morphisms $\varphi\colon B\to D$ and $\psi\colon D\to\tilde{A}$ such that

(α) The diagram (1.2.2) commutes; i.e. $\rho=\psi\cdot\varphi$

(β) For every prime ideal $\mathfrak{p}\in\operatorname{Spec}\tilde{A}$ such that $\sigma^{-1}(\mathfrak{p})$ is a minimal prime ideal in $\operatorname{Spec}A$ the A-algebra $D_{\psi^{-1}(\mathfrak{p})}$ is smooth over A.

(b) $\tilde{A}$ is a direct limit of smooth (or of generically smooth resp.) A-algebras if every finitely generated A-subalgebra B of $\tilde{A}$ factors over a smooth (or over a generically smooth resp.) A-algebra.

(1.3) *Remark.* (a) If $\sigma\colon A\to\tilde{A}$ is a direct limit of smooth A-algebras, then σ is a regular morphism. Similarly, if $\sigma\colon A\to\tilde{A}$ is a direct limit of generically smooth A-algebras, then $\sigma_s\colon S^{-1}A\to S^{-1}A\otimes_A\tilde{A}$ is a regular morphism, where S denotes the multiplicative set

$$S=A\setminus\bigcup_{\substack{\mathfrak{g}\in\operatorname{Spec}A\\ \mathfrak{g}\text{ minimal}}}\mathfrak{g}.$$

(b) With the notations of (1.2) (a.2) if D is a generically smooth A-algebra, then

$$ht(\psi(H_{D/A})\tilde{A}\cap A)\geqq 1$$

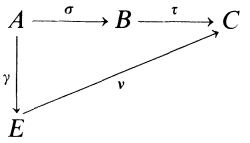
$H_{D/A}$ is the ideal defined in [9] and of the nonsmooth locus of D over A.

Further we mention some properties of smooth morphisms which are simple to prove and mostly well-known and which we will use later on in the proof of the theorem.

(1.4) **Proposition.** *Let $\sigma\colon A\to B$ and $\tau\colon B\to C$ be regular local morphisms of local noetherian rings. Assume that B is a direct limit of smooth A-algebras with respect to σ and that C is a direct limit of smooth B-algebras with respect to τ. Then C is a direct limit of smooth A-algebras with respect to $\tau\cdot\sigma$.*

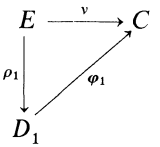
Proof. Assume there is given a factorization of $\tau \cdot \sigma$:

(1.4.1)



where ν is injective and $E = A[Y_1, \ldots, Y_N]/(g_1, \ldots, g_s)$ is a finitely generated A-algebra. Then we can find a smooth finitely generated B-algebra $D_1 = B[X_1, \ldots, X_n]/(f_1, \ldots, f_p)$, an A-algebra morphism $\rho_1 : E \to D_1$ and a B-algebra morphism $\varphi_1 : D_1 \to C$ such that the diagram

(1.4.2)



commutes. Let m be the maximal ideal in C and define $m_1 = \varphi_1^{-1}(m)$. By using [16] (20.E) we can assume that there is a $t \leqq p$ such that

(a) $(f_1, \ldots, f_p) B[X_1, \ldots, X_n]_{m_1} = (f_1, \ldots, f_t) B[X_1, \ldots, X_n]_{m_1}$.

(b) $\det \left(\dfrac{\partial f_i}{\partial X_j} \right)_{i,j=1,\ldots,t} \notin m_1$.

To begin with we lift ρ_1 to a morphism of the polynomial algebras:

$$\tilde{\rho}_1 : A[Y_1, \ldots, Y_N] \to B[X_1, \ldots, X_n]$$

such that $\tilde{\rho}_1 | A = \sigma$ and $\tilde{\rho}_1(Y_i) = h_i$, where h_i is a preimage of $\rho_1(\bar{Y}_i)$ in $B[X_1, \ldots, X_n]$. We consider the following relations in $B[X_1, \ldots, X_n]$:

(I) for all $j = 1, \ldots, s$

$$\tilde{\rho}_1(g_j(Y_1, \ldots, Y_N)) = g_j(h_1, \ldots, h_N) = \sum_{k=1}^{p} r_{jk} f_k.$$

(II) because $(f_1, \ldots, f_p) B[X_1, \ldots, X_n]_{m_1} = (f_1, \ldots, f_t) B[X_1, \ldots, X_n]_{m_1}$ for every $j \in \{t+1, \ldots, p\}$ there are elements $l_j \notin m_1$ such that

$$l_j f_j = \sum_{k=1}^{t} w_{jk} f_k.$$

Choose $\tilde{E} \subset B$ to be a finitely generated A-subalgebra containing all coefficients which occur in f_j, r_{jk}, h_i, l_j and w_{jk}. Then in particular the relations (I) and (II) are satisfied in $\tilde{E}[X_1, \ldots, X_n]$. By assumption there is a smooth finitely generated A-algebra D_2 such that the embedding $\tilde{E} \to B$ factors over D_2. Then the morphism

$$\tilde{\rho}_2 : A[Y_1, \ldots, Y_N] \to D_2[X_1, \ldots, X_n]$$
$$Y_i \to h_i$$

satisfies the condition

(I) for $j=1, \ldots, s$

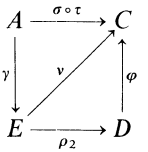
$$\tilde{\rho}_2(g_j(Y_1, \ldots, Y_N)) = \sum_{k=1}^{p} r_{jk} f_k$$

and induces therefore a mapping

$$\rho_2 \colon E \to D_2[X_1, \ldots, X_n]/(f_1, \ldots, f_p) = D.$$

This leads to a factorization:

(1.4.3)



where φ is defined by $\varphi(\bar{X}_i) = \varphi_1(\bar{X}_i)$ for every $i=1, \ldots, n$. If $\mathfrak{r} = \varphi^{-1}(m) \subset D$ then we get by means of (II):

$$(f_1, \ldots, f_p) D_2[X_1, \ldots, X_n]_{\mathfrak{r}} = (f_1, \ldots, f_t) D_2[X_1, \ldots, X_n]_{\mathfrak{r}}$$

and

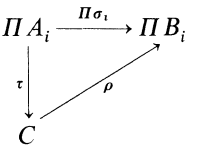
$$\det\left(\frac{\partial f_i}{\partial X_j}\right)_{i,j=1,\ldots,t} \notin \mathfrak{r}.$$

Therefore $D_{\mathfrak{r}}$ is smooth over D_2 and hence smooth over A. A localization of D in a multiplicative set generated by one suitable element of D results in a finitely generated smooth A-algebra over which ν factors.

(1.5) **Proposition.** *Assume there are given finitely many regular morphisms σ_i: $A_i \to B_i$ of noetherian rings such that for every $i=1, \ldots, n$, B_i is a direct limit of smooth A_i-algebras with respect to σ_i. Then $\prod_{i=1}^{n} B_i$ is a direct limit of smooth $\prod_{i=1}^{n} A_i$-algebras with respect to $\prod_{i=1}^{n} \sigma_i$.*

Proof. Assume there is a factorization of $\Pi \sigma_i$ over a ΠA_i-algebra C of finite type

(1.5.1)



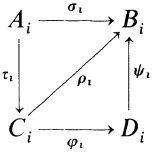
Assume further that ρ is injective, $C = (\Pi A_i)[X_1, \ldots, X_m]/I$, and

$$\rho(\bar{X}_j) = (x_{j1}, \ldots, x_{jn}) \in \prod_{i=1}^{n} B_i$$

where $\bar{X}_j$ is the image of X_j in C. For every $i=1, \ldots, n$ we consider the A_i-subalgebra $C_i = A_i[x_{1i}, \ldots, x_{mi}]$ of B_i. By assumption C_i can be embedded in a

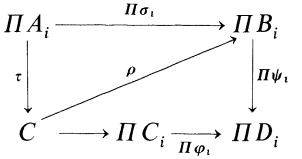
smooth finitely generated A_i-algebra D_i and we have a commutative diagram:

(1.5.2)



Then ΠD_i is a smooth finitely generated ΠA_i-algebra and the expected factorization is given by the following diagram:

(1.5.3)



For $i=1, \ldots, s$ let W_i be fields of characteristic p_i ($p_i=0$ or $p_i>0$) and P_i are prime fields contained in W_i and for $i=s+1, \ldots, n$ let W_i be a complete Cohen ring [10] EGA 0_{IV} (19.8.4) containing $\mathbb{Z}_{(p_i)}$ and define P_i to be $\mathbb{Z}_{(p_i)}$. Obviously every W_i is regular over P_i and by applying (1.4) and (1.5) we get by means of Artin's theorem.

(1.6) **Corollary.** *Let $A_i = W_i[[X_{i1}, \ldots, X_{im_i}]]$ be a power series ring in finitely many variables over W_i. Then $\prod_{i=1}^{n} A_i$ is a direct limit of smooth $\prod_{n=1}^{n} P_i$-algebras.*

Finally we would like to note some well-known facts about smooth morphisms and regular rings which are important for the following sections:

(1.7) **Proposition.** *Let k be a field and A an algebra essentially of finite type over k. Assume that the following conditions are satisfied:*

(i) *A is a regular ring.*

(ii) *For every maximal ideal m of A the residue class field A/m is separable over k.*

Then A is smooth over k.

Proof. By [13] Chap. II (4.8) A is smooth over k if and only if A_m is smooth over k for every maximal ideal m of A. For regular local rings the assertion follows with [13] Chap. II (6.8).

(1.8) **Proposition.** *Suppose $\varphi\colon A \to R$ is a local morphism of local regular rings. Let $x_1, \ldots, x_n$ be elements of the maximal ideal of A whose images $\varphi(x_1), \ldots, \varphi(x_n)$ are part of a regular system of parameters in R. Then $x_1, \ldots, x_n$ are part of a regular system of parameters of A.*

Proof. Let m_A and m_R denote the maximal ideals of A and R. If $n=1$, then $\varphi(x_1)$ is part of a regular system of parameters if and only if $\varphi(x_1) \in m_R \setminus m_R^2$. This implies $x_1 \in m_A \setminus m_A^2$ and x_1 is part of a regular system of parameters of A. The assertion follows by induction on n.

(1.9) **Proposition.** *Suppose R is a regular local ring, $\mathfrak{g} \subset R$ is a prime ideal and $R/\mathfrak{g}$ is regular. Then every system of generators $\{P_1, \ldots, P_m\}$ of the ideal $\mathfrak{g}$ contains a subset $\{P_{i_1}, \ldots, P_{i_s}\}$ with the following properties:*

(a) *The ideal $\mathfrak{g}$ is generated by $\{P_{i_1}, \ldots, P_{i_s}\}$.*

(b) *$\{P_{i_1}, \ldots, P_{i_s}\}$ is part of a regular system of parameters of R.*

Proof. Assume m is the maximal ideal of R and $k = R/m$ is the residue class field of R. By [16] (17.F) $\mathfrak{g}$ is generated by a part of a regular system of parameters of R. Therefore

$$\mathfrak{g}/m\mathfrak{g} \to m/m^2$$

is injective and the assertion follows by Nakayama's lemma.

(1.10) **Proposition.** *Let $\varphi: D \to R$ be a local morphism of regular local rings. Let E be a field contained in D and suppose that R is formally smooth over E* (w.r.t. *the maximal-adic topology on R*). *Then D is formally smooth over E* (w.r.t. *the maximal adic topology on D*). *In particular, if D is essentially of finite type over E, then D is smooth over E.*

Proof. Let k and K be the residue class fields of D and R respectively. By [1] Prop. 23, p. 93 and Prop. 17, p. 223, R is formally smooth over E iff $H_1(E, R, K) = 0$ and we have to show $H_1(E, D, k) = 0$.

Let P be a prime field contained in E. By considering $P \to D \to R$ we get the Zariski-Jacobi-sequence:

$$H_2(P, R, K) \to H_2(D, R, K) \to H_1(P, D, K).$$

Since D and R are regular local rings, they are formally smooth over their prime fields and by the above reference: $H_2(P, R, K) = 0$ and $H_1(P, D, K) = 0$. Hence $H_2(D, R, K) = 0$. Next consider $E \to D \to R$ and the Zariski-Jacobi-sequence:

$$H_2(D, R, K) \to H_1(E, D, K) \to H_1(E, R, K).$$

This shows $H_1(E, D, K) = 0$. Since K is flat over k by [1] Prop. 58, p. 59 also $H_1(E, D, k) = 0$ proving that D is formally smooth over E (w.r.t. the maximal-adic topology). If D is moreover essentially of finite type over E, then D is already smooth over E ([16], (29.E) Theorem 64).

§ 2. The reduced case

In this section we prove our main result by using Artin's theorem.

Assume in the following that A is a semilocal reduced ring whose generic formal fibres are geometrically regular and that either $\mathbb{Q} \subseteq A$ or A satisfies the following condition: for all minimal primes $\mathfrak{g}$ of A there is a perfect subfield P_k of $Q(A/\mathfrak{g})$ such that $Q(A/\mathfrak{g})$ has a separating transcendence basis over P_k such that $\prod P_k \subset A$. Let $m_1, \ldots, m_s$ be the maximal ideals of A, $J = \bigcap_{i=1}^{s} m_i$ be the Jacobson-radical of A and let $\hat{A}$ denote the J-adic completion of A. Then

(2.1) **Theorem.** *$\hat{A}$ is limit of generically smooth A-algebras.*

Proof. $\hat{A}$ can be presented by

$$\hat{A}=\prod_{i=1}^{s}\hat{A}_{m_i}$$

where $\hat{A}_{m_i}$ is the m_i-adic completion of A or A_{m_i}. For $i=1,\ldots,s$ let M_i be the total ring of quotients of A_{m_i}. Then by assumption the ring $A_{m_i}\otimes M_i$ is regular. Therefore $\hat{A}\otimes_A M$ is a regular ring where M denotes the total ring of quotients of A. This implies:

(i) If $\mathfrak{p}\in\operatorname{Spec}\hat{A}$ is a prime ideal whose contraction $\mathfrak{q}=\mathfrak{p}\cap A$ is a minimal prime ideal of A, then the local ring $\hat{A}_{\mathfrak{p}}$ is geometrically regular over the field $A_{\mathfrak{p}}$.

(ii) By Cohen's structure theorem each A_{m_i} is a homomorphic image of a power series ring $W_i[[X_{i1},\ldots,X_{im_i}]]$ where the W_i are fields. Hence

$$\hat{A}_{m_i}\cong W_i[[X_{i1},\ldots,X_{im_i}]]/\mathfrak{q}_i$$

with a reduced ideal $\mathfrak{q}_i=\mathfrak{q}_{i1}\cap\ldots\cap\mathfrak{q}_{in_i}$ in $W_i[[\underline{X}_i]]$. (A_{m_i} is reduced because of the regularity of the generic formal fibres. It follows immediately that $\hat{A}_{m_i}$ satisfies Serre's conditions (R_0) and (S_1).) Hence we have the following commutative diagram:

(ii.1)
$$\begin{array}{ccc}\mathbb{Z} & \xrightarrow{\nu} & \prod_{i=1}^{s} W_i[[\underline{X}_i]] \\ \downarrow{\scriptstyle\mu} & & \downarrow{\scriptstyle\gamma} \\ A & \xrightarrow{\sigma} & \hat{A}\end{array}$$

where γ is a surjective mapping and ν is the canonical morphism sending 1 to the unit element of $\prod_{i=1}^{s} W_i[[\underline{X}_i]]$. In the following we use the notation $R=\prod_{i=1}^{s} W_i[[\underline{X}_i]]$ and get:

(ii.2) $\hat{A}\cong R/\mathfrak{q}$ where

$$\mathfrak{q}=\bigcap_{i=1}^{s}\mathfrak{q}_i=\bigcap_{i=1}^{s}\bigcap_{j=1}^{n_i}\mathfrak{q}_{ij}$$

is a reduced ideal in R. Here the prime ideals $\mathfrak{q}_{ij}\subseteq W_i[[\underline{X}_i]]$ are canonically identified with prime ideals of R. Henceforth we identify $\hat{A}$ with $R/\mathfrak{q}$.

(iii) Let C be a (fixed) finitely generated A-subalgebra of $\hat{A}$, then

$$\rho\colon\ C=A[Y_1,\ldots,Y_N]/\mathfrak{r}\to\hat{A}$$

is the embedding morphism and $\mathfrak{r}$ is a reduced ideal of $A[\underline{Y}]=A[Y_1,\ldots,Y_N]$ which is generated by $f_1,\ldots,f_r\in A[Y_1,\ldots,Y_N]$.

In the following we propose to embed C into a generically smooth A-algebra D over which ρ factors. We start with the construction of a suitable

finitely generated algebra E which is contained in A and to which we have to refer later on.

(iv) Let $f_1, \ldots, f_r$ be the system of generators of the ideal $\mathfrak{r}$ given in (iii). Then for every $\nu = 1, \ldots, r$ we have

$$f_\nu = \sum_{|\mu| \leqq l} a_{\nu(\mu)} Y^{(\mu)} \in A[Y_1, \ldots, Y_N]$$

where (μ) is a multiple index and the coefficients $a_{\nu(\mu)}$ are elements of A.

(iv.1) Let M be the total ring of quotients of A. M is a finite product of fields $M = \Pi M_i$ with $M_i = k(\mathfrak{g}_i)$ and $\mathfrak{g}_i$ a minimal prime ideal of A. First we remark that if $\mathbb{Q} \nsubseteq A$ then M_i contains a maximal perfect subfield P_k with $P_k \subseteq A$. In particular by assumption M_i has a separating transcendence basis over P_k. In this case we define Z_0 to be the product of the P_k. If $\mathbb{Q} \subseteq A$ we simply take $Z_0 = \mathbb{Q}$. Then for every minimal prime $\mathfrak{g}_i$ of A $(Z_0)_{\mathfrak{g}_i \cap Z_0}$ is either $\mathbb{Q}$ or one of the above fields P_i. We may also suppose that the algebras W_j chosen in (ii) contain one of the P_i. We take elements t_i in A which are not contained in $\mathfrak{g}_i$ but are in all minimal primes of A different from $\mathfrak{g}_i$. Since every generating set of the field extension $M_i \supseteq P_i$ contains a separating transcendence basis there are, for all i subfields, E_i of M_i with the following properties:

(α) E_i is a P_i-algebra essentially of finite type and generated by (finitely many) elements $\bar{a}_{\nu(\mu)}$, $\bar{t}_i$ and $\bar{b}_{ij}$ of $A/\mathfrak{g}_i$.

(β) M_i is separable over E_i.

Let E be the Z_0-subalgebra of A generated by $a_{\nu(\mu)}$ and by the elements t_i, $(b_{ij} t_i)$ of A (for all i). Then by the canonical injection $A \to \Pi M_i$, E is mapped into ΠE_i and the total ring of quotients of E is exactly ΠE_i.

(iv.2) We choose representatives:

$$\{\tilde{a}_{\nu(\mu)}\}_{\substack{\nu = 1, \ldots, r \\ |\mu| \leqq l}} \subseteq R$$

and $\{\tilde{t}_i\}$, $\{\tilde{b}_{ij}\} \subseteq R$ such that $\gamma(\tilde{a}_{\nu(\mu)}) = a_{\nu(\mu)}$, $\gamma(\tilde{t}_i) = t_i$ and $\gamma(\tilde{b}_{ij}) = b_{ij}$. Further let

$$F_\nu = \sum_{|\mu| \leqq l} \tilde{a}_{\nu(\mu)} Y^{(\mu)} \in R[Y_1, \ldots, Y_N]$$

represent $f_\nu \in A[Y_1, \ldots, Y_N] \subseteq \hat{A}[Y_1, \ldots, Y_N]$ for every $\nu = 1, \ldots, r$. We define

$$E_0 = Z_0[\tilde{a}_{\nu(\mu)}, \tilde{t}_i, (\tilde{b}_{ij} \cdot \tilde{t}_i)] \subseteq R$$

to be the Z_0-subalgebra of R generated by $\tilde{a}_{\nu(\mu)}, \tilde{t}_i, (\tilde{b}_{ij} \cdot \tilde{t}_i)$. Then with $\gamma | E_0$ and $\mathfrak{g} = \operatorname{Ker} \gamma_0 = \mathfrak{q} \cap E_0$ we get $\gamma_0(E_0) \subseteq A$ and $\gamma_0(E_0) \cong E_0/\mathfrak{g} \cong E$.

(v) We construct in the following a smooth algebra D_0 and an ideal $\mathfrak{p} \subseteq D_0$ with the properties:

(v.1) There are algebra morphisms

$$\psi^*\colon D_0 \to R$$
$$\delta\colon E_0 \to D_0$$

such that $\kappa = \psi^* \cdot \delta\colon E_0 \to R$ is the canonical embedding of E_0 into R.

(v.2) $\delta(\mathfrak{g}) \subseteq \mathfrak{p}$ and $\psi^*(\mathfrak{p}) \subset \mathfrak{q}$.

(v.3) Consider the following commutative diagram:

$$\begin{array}{ccccc} E_0 & \xrightarrow{\delta} & D_0 & \xrightarrow{\psi^*} & R \\ \downarrow & & & & \downarrow{\scriptstyle \gamma} \\ A & & \longrightarrow & & \hat{A} \end{array}$$

Then the ring

$$(D_0/\mathfrak{p})_{(\gamma \cdot \psi^*)^{-1}(\mathfrak{n})}$$

is regular for every prime ideal $\mathfrak{n} \subseteq \hat{A}$ whose contraction $\mathfrak{n} \cap A$ is a minimal prime ideal in A. In particular in this case $(D_0/\mathfrak{p})_{(\gamma \cdot \psi^{*-1})(\mathfrak{n})}$ contains one of the fields E_i constructed in (iv). Since $\hat{A}_{\mathfrak{n}}$ is geometrically regular over the field $A_{(\mathfrak{n} \cap A)}$ and $A_{(\mathfrak{n} \cap A)}$ is separable over E_i, $(D_0/\mathfrak{p})_{(\gamma \cdot \psi^{*-1})(\mathfrak{n})}$ is already smooth over E_i by (1.10).

(v.4) The generically smooth A-algebra D over which the embedding ρ factors is given by $D = (D_0/\mathfrak{p}) \otimes_E A$.

Now we give the details of the construction:

(vi) To simplify the notation we assume that the ideal $\mathfrak{q}$, defined in (ii.2), is given by

$$\mathfrak{q} = \mathfrak{q}_1 \cap \ldots \cap \mathfrak{q}_t$$

where $\mathfrak{q}_j$ is (now) a minimal prime ideal of R containing $\mathfrak{q}$. Then we select systems of generators for the ideals $\mathfrak{q}_j$ in R:

$$\mathfrak{q}_j = \langle Q_{j1}, \ldots, Q_{jm_j} \rangle \quad \text{for every } j = 1, \ldots, t.$$

Furthermore we define the following subset Γ of $\operatorname{Spec}(R)$:

$$\Gamma = \left\{ \mathfrak{n} \in \operatorname{Spec}(R) \;\middle|\; \begin{array}{l} \mathfrak{n} \text{ is maximal with the property:} \\ \mathfrak{n} \supseteq \mathfrak{q} \text{ and } \varphi(\mathfrak{n}) \cap A \text{ is a minimal} \\ \text{prime ideal of } A \end{array} \right\} \subseteq V(\mathfrak{q}).$$

For every $\mathfrak{n} \in \Gamma$ we get:

(vi.1) The ring $(R/\mathfrak{q})_{\mathfrak{n}}$ is regular (by (i)). In particular this implies that $\mathfrak{n}$ contains exactly one of the minimal prime ideals $\mathfrak{q}_j$ over $\mathfrak{q}$.

(vi.2) If $\mathfrak{q}_{j_0} \subseteq \mathfrak{n}$ for one $j_0 \in \{1, \ldots, t\}$ then

$$\mathfrak{q}_{j_0 \mathfrak{n}} = \langle Q_{j_0 l_1}, \ldots, Q_{j_0 l_v} \rangle_{\mathfrak{n}}$$

and the $Q_{j0l_1}, \ldots, Q_{j0l_v}$ is part of a regular system of parameters of $R_{\mathfrak{n}}$ (by (1.10)). This defines a finite set Δ_j of ordered tuples of length $\leqq m_j$ for every $j \in \{1, \ldots, t\}$ by

(vi.3) $(l_1, \ldots, l_v) \in \Delta_j$ if and only if the following two conditions are satisfied:

(vi.3.1) $(l_1, \ldots, l_v) \leqq (1, \ldots, m_j)$ this means:

$$l_1 < l_2 < \ldots < l_v \quad \text{and} \quad \{l_1, \ldots, l_v\} \subseteq \{1, \ldots, m_j\}.$$

(vi.3.2) There is a $\mathfrak{n} \in \Gamma$ such that

(α) $\mathfrak{n} \supseteq \mathfrak{q}_j$

(β) $Q_{jl_1}, \ldots, Q_{jl_v}$ is part of a regular system of parameters of $R_{\mathfrak{n}}$

(γ) $\mathfrak{q}_{j\mathfrak{n}} = \langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle_{\mathfrak{n}}$.

(vii) For every $j \in \{1, \ldots, t\}$, $(l_1, \ldots, l_v) \in \Delta_j$ and $l \in \{1, \ldots, m_j\} \setminus \{l_1, \ldots, l_v\}$ we consider in R the ideals:

$$\mathfrak{a}_{j;(l_1, \ldots, l_v);l} := (\langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle : Q_{jl}).$$

Obviously if $\mathfrak{n} \in \Gamma$; $\mathfrak{q}_j \subseteq \mathfrak{n}$ and $\mathfrak{q}_{j\mathfrak{n}} = \langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle$ where $Q_{jl_1}, \ldots, Q_{jl_v}$ is part of a regular system of parameters of R, then for every $l \in \{1, \ldots, m_j\} \setminus \{l_1, \ldots, l_v\}$ the ideal $\mathfrak{a}_{j;(l_1, \ldots, l_v);l}$ is not contained in $\mathfrak{n}$. Furthermore we select a system of generators for every ideal $\mathfrak{a}_{j;(l_1, \ldots, l_v);l} \subseteq R$

$$\mathfrak{a}_{j;(l_1, \ldots, l_v);l} = \langle a^{\alpha}_{j;(l_1, \ldots, l_v);l} \mid \alpha = 1, \ldots, h \rangle.$$

Then we have to consider the following equations in R

$$\text{(VII)} \qquad a^{\alpha}_{j;(l_1, \ldots, l_v);l} Q_{jl} = \sum_{\beta=1}^{v} w^{\alpha,\beta}_{j;(l_1, \ldots, l_v);l} Q_{jl_\beta}$$

with elements

$$w^{\alpha,\beta}_{j;(l_1, \ldots, l_v);l} \in R.$$

(VII) is given for every $j \in \{1, \ldots, t\}$, $(l_1, \ldots, l_v) \in \Delta_j$, $l \in \{1, \ldots, m_j\} \setminus \{l_1, \ldots, l_v\}$, and $\alpha \in \{1, \ldots, h\}$.

(viii) The embedding

$$\rho\colon\ C = A[Y_1, \ldots, Y_N]/\mathfrak{r} \to \hat{A} = R/\mathfrak{q}$$

defined in (iii) provides additional relations. For every $i = 1, \ldots, N$ we choose representatives $u_i \in R$ of the elements $\rho(\bar{Y}_i) = t_i \in \hat{A}$. With the polynomials $F_v \in R[Y_1, \ldots, Y_N]$-defined in (iv) - we get:

$$F_v(u_1, \ldots, u_N) \in \mathfrak{q} \qquad \text{for every } v = 1, \ldots, r.$$

This implies

$$F_v(u_1, \ldots, u_N) \in \mathfrak{q}_j \qquad \text{for every } v = 1, \ldots, r;\ j = 1, \ldots, t$$

resulting in equations

$$\text{(VIII)} \qquad F_v(u_1, \ldots, u_N) = \sum_{k=1}^{m_j} g_{vjk} Q_{jk}$$

with elements $g_{vjk} \in R$.

(ix) Finally let $\{P_1, \ldots, P_q\}$ be a system of generators of the ideal $\mathfrak{g} \subseteq E_0$ defined in (iv). Because of $\mathfrak{g} \subseteq \mathfrak{p}$ there are for every $j = 1, \ldots, t$ and for every $\mu = 1, \ldots, q$ elements $b_{\mu jk} \in R$ such that

$$P_\mu = \sum_{k=l}^{m_j} b_{\mu jk} Q_{jk}. \tag{IX}$$

(x) We start with a subalgebra of R essentially of finite type over Z_0 containing all elements and relations considered above. This means take the subalgebra B_0 of R which is generated by the following algebras essentially of finite type over Z_0 and by the following elements:

(x.1) By assumption $R = \prod_{i=l}^{s} W_i[[\underline{X}_i]]$ where each W_j contains exactly one of the rings P_i and $Z_0 \subseteq R$. Let B_0 be the Z_0-subalgebra of R with:

(x.2) $E_0 \subseteq B_0$ compare with (iv)

(x.3) $\{Q_{j\gamma}\} \subseteq B_0$ (vi)

(x.4) $\{b_{\mu jk}\} \subseteq B_0$ (ix)

(x.5) $\{a^{\alpha}_{j;(l_1,\ldots,l_v);l}\} \subseteq B_0$ (vii)

(x.6) $\{w^{\alpha,\beta}_{j;(l_1,\ldots,l_v);l}\} \subseteq B_0$ (VII)

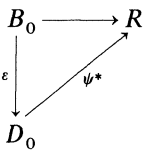
(x.7) $\{u_1, \ldots, u_N\} \subseteq B_0$ (viii)

(x.8) $\{g_{\nu jk}\} \subseteq B_0$. (VIII)

All indices mentioned above run through finite sets. Hence there is a finitely generated Z_0-algebra B_0 containing all elements in (x.2) through (x.8), and B_0 is essentially of finite type over Z_0. Moreover equations (VII) and (IX) are satisfied in B_0 and because of $E_0 \subseteq B_0$ relations (VIII) are also true in B_0.

(xi) By means of (1.7) (developed from Artin's theorem) we find a smooth finitely generated Z_0-algebra D_0 such that the embedding $B_0 \to R$ factors over D_0:

(xi.1)



In particular this implies that D_0 is a regular ring. In the following we identify B_0 with the subring $\varepsilon(B_0)$ of D_0. For $j = 1, \ldots, t$ we consider the ideals

$$\mathfrak{p}_j = \langle Q_{j1}, \ldots, Q_{jm_j} \rangle \subseteq \psi^{*-1}(\mathfrak{q}_j) \subseteq D_0.$$

We define

$$\mathfrak{p} = \bigcap_{j=l}^{t} \mathfrak{p}_j \subseteq \psi^{*-1}(\mathfrak{q}) \subseteq D_0$$

and

$$\tilde{D} = D_0/\mathfrak{p}.$$

We have to prove then that $\tilde{D}$, D_0 and $\mathfrak{p}$ satisfy the required properties of (v).

(xii) E_0 is a subalgebra of B_0, therefore by (xi.1) the embedding κ: $E_0 \to R$ factors over D_0. Relations (IX) are satisfied in B_0, hence also in D_0 and $\mathfrak{g}$ is mapped into $\mathfrak{p} \subseteq D_0$ by ε. By this we can assume that E is a subalgebra of $\tilde{D} = D_0/\mathfrak{p}$. From the construction of $\mathfrak{p}$ we also get $\psi^*(\mathfrak{p}) \subseteq \mathfrak{q}$. This means our requirements (v.1) and (v.2) are satisfied.

(xiii) **Statement** (v.3). *For every* $\mathfrak{n} \in \Gamma$ *the ring* $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ *is regular local.*

Proof of (xiii). For every $\mathfrak{n} \in \Gamma$ there is exactly one $j \in \{1, \ldots, t\}$ with $\mathfrak{q}_j \subseteq \mathfrak{n}$. Because $\mathfrak{p}_k$ contains a system of generators of the R-ideal $\mathfrak{q}_k$, there is also exactly one $j \in \{1, \ldots, t\}$ with $\mathfrak{p}_j \subseteq \rho^{-1}(\mathfrak{n})$. By assumption $(R/\mathfrak{q}_j)_{\mathfrak{n}}$ is a regular local ring; hence there is an element $(l_1, \ldots, l_v) \in \Delta_j$ with

(a) $\mathfrak{q}_j R_{\mathfrak{n}} = \langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle$.

(b) $\{Q_{jl_1}, \ldots, Q_{jl_v}\}$ is part of a regular system of parameters in $R_{\mathfrak{n}}$.

By construction D_0 contains the generators:

$$\{a^{\alpha}_{j;(l_1,\ldots,l_v);l}\}$$

of the R-ideals $\mathfrak{a}_{j(l_1,\ldots,l_v);l}$ for every $l \in \{1, \ldots, m_j\} \setminus \{l_1, \ldots, l_v\}$. Furthermore as noted in (vii) for every $l \in \{1, \ldots, m_j\} \setminus \{l_1, \ldots, l_v\}$

$$\mathfrak{a}_{j;(l_1,\ldots,l_v);l} \not\subseteq \mathfrak{n}.$$

This implies that for every l there is a $\alpha_l \in \{1, \ldots, h\}$ such that

$$a^{\alpha_1}_{j;(l_1,\ldots,l_v);l} \notin \mathfrak{n}.$$

In particular we get:

$$a^{\alpha_l}_{j;(l_1,\ldots,l_v);l} \in D_0 \setminus \psi^{*-1}(\mathfrak{n}).$$

On the other hand relations (VII) are satisfied in D_0, therefore

$$a^{\alpha_1}_{j;(l_1,\ldots,l_v);l} Q_{jl} \in \langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle D_0$$

resulting in:

$$\langle Q_{jl_1}, \ldots, Q_{jl_v} \rangle D_{0\psi^{*-1}(\mathfrak{n})} = \mathfrak{p}_j D_{0\psi^{*-1}(\mathfrak{n})}.$$

Furthermore by the induced morphism:

$$\psi^*\colon\ D_{0\psi^{*-1}(\mathfrak{n})} \to R_{\mathfrak{n}}.$$

$\{Q_{jl_1}, \ldots, Q_{jl_v}\}$ are mapped into the part of a regular system of parameters of $R_{\mathfrak{n}}$. By (1.8) $\{Q_{jl_1}, \ldots, Q_{jl_v}\}$ are already part of a regular system of parameters of $D_{0\psi^{*-1}(\mathfrak{n})}$ and the ring $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})} = (D_0/\mathfrak{p}_j)_{\psi^{*-1}(\mathfrak{n})}$ is regular.

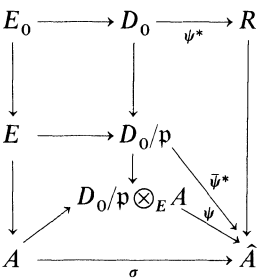
(xiv) **Statement.** *For every* $\mathfrak{n} \in \Gamma$: $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ *is smooth over* E.

Proof of (xiv). Since $\mathfrak{n}$ lies over a minimal prime ideal of A, the intersection of $\mathfrak{n}$ with E is also a minimal prime ideal of E and $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ contains one of the subfields E_i constructed in (iv). $\hat{A}_{\mathfrak{n}}$ is formally smooth over the field $A_{(\mathfrak{n}\cap A)}$ and $A_{(\mathfrak{n}\cap A)}$ is separable over E_i, therefore $\hat{A}_{\mathfrak{n}}$ is also formally smooth over E_i. By (1.10) $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ is formally smooth over E_i and since $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ is essentially of finite type over E_i, it is smooth over E_i, hence also smooth over E.

(xv) **Statement.** *$D=D_0/\mathfrak{p}\otimes_E A$ is a generically smooth A-algebra.*

Proof of (xv). Obviously D is an A-algebra essentially of finite type. We consider the following commutative diagram:

(xiv.1)



where ψ is the morphism induced by $\bar{\psi}^*$ and σ. We have to prove that $D_0/\mathfrak{p}\otimes_E A$ is a generically smooth A-algebra with respect to ψ. Let $\mathfrak{n}$ be an element of Γ. Then by (xiv) $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}$ is smooth over E and hence $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}\otimes_E A$ is smooth over E. Let $\bar{\mathfrak{n}}$ denote the image of $\mathfrak{n}$ in $\hat{A}$, then $[D_0/\mathfrak{p}\otimes_E A]_{\sigma^{-1}(\bar{\mathfrak{n}})}$ is a localization of $(D_0/\mathfrak{p})_{\psi^{*-1}(\mathfrak{n})}\otimes_E A$ and hence smooth over A or A_n respectively where $n=\bar{\mathfrak{n}}\cap A$.

(xvi) **Statement.** *The morphism $\rho\colon C\to\hat{A}$ factors over $\psi\colon D\to\hat{A}$.*

Proof of (xvi). In (viii) $u_1,\ldots,u_N$ are chosen to be representatives of $\rho(\bar{Y}_i)=t_i\in\hat{A}$. By construction $u_1,\ldots,u_N$ are elements of D_0 and let $\bar{u}_1,\ldots,\bar{u}_N$ be their images in $D_0/\mathfrak{p}$. We define an A-algebra morphism:

$$\begin{aligned}\eta\colon A[Y_1,\ldots,Y_N]&\to D=D_0/\mathfrak{p}\otimes_E A\\ Y_i&\to\quad \bar{u}_i\otimes 1.\end{aligned}$$

The composition $\psi\circ\eta$ maps Y_i into $\rho(\bar{Y}_i)=t_i$ and it is to show that $\mathfrak{r}=(f_1,\ldots,f_r)$ is contained in the kernel of η. It follows for every $v\in\{1,\ldots,r\}$:

$$\eta(f_v(Y_1,\ldots,Y_N))=\sum_{|\mu|\leq l}(\bar{u}^{(\mu)}\otimes a_{v(\mu)}).$$

By (iv) the $a_{v(\mu)}$ are contained in E and hence:

$$\eta(f_v(Y_1,\ldots,Y_N))=(\sum_{|\mu|\leq l}a_{v(\mu)}\bar{u}^{(\mu)})\otimes 1.$$

Again by (iv) and by the construction of D_0 we get that

$$\sum a_{\nu(\mu)} \bar{u}^{(\mu)}$$

is the residue class of $F_\nu(u_1, \ldots, u_N)$ in $D_0/\mathfrak{p}$. But in D_0 the relations (VIII) are satisfied. This implies that $F_\nu(u_1, \ldots, u_N)$ is contained in $\mathfrak{p} \subset D_0$. Therefore $\eta(f_\nu(Y_1, \ldots, Y_N)) = 0$ for every $\nu = 1, \ldots, r$ and ρ factors over D. D is an algebra essentially of finite type over A and may be replaced by a finite type A-algebra satisfying the above requirements.

§ 3. A lifting property for limits of smooth algebras

In § 2 we have shown that the completion $\hat{A}$ of a semilocal reduced ring A is a direct limit of generically smooth A-algebras, if A has geometrically regular generic formal fibres and satisfies certain separability conditions. For the application in the next section this result has to be extended to the non-reduced case.

To do so we need the following considerations:

(3.1) (see [10], EGA 0_{IV} (19.3.10.1)). Let A be a noetherian ring, $t \in A$ a nilpotent element and let D_0, D_1 and E be noetherian A-algebras. Suppose further that there are given a surjective A-algebra morphism

$$\psi_1 \colon D_1 \to E$$

and A-algebra morphisms

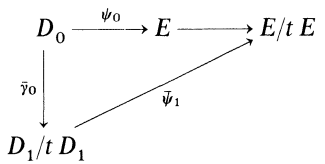
$$\psi_0 \colon D_0 \to E$$

and

$$\bar{\gamma}_0 \colon D_0 \to D_1/t D_1$$

such that the diagram

(3.1.1)



commutes where $\bar{\psi}_1$ is the morphism induced by ψ_1.

Lemma. *In the above situation, if D_0 is smooth over A then there is an A-algebra morphism*

$$\gamma_0 \colon D_0 \to D_1$$

such that

(i) $\psi_1 \circ \gamma_0 = \psi_0$.

(ii) $\bar{\gamma}_0$ *is the composed morphism:* $D_0 \xrightarrow{\gamma_0} D_1 \xrightarrow{\text{can}} D_1/t D_1$.

Proof. Consider the commutative diagram

$$\begin{array}{ccc} D_1 & \xrightarrow{\psi_1} & E \\ \downarrow & & \downarrow \\ D_1/t\,D_1 & \xrightarrow{\bar{\psi}_1} & E/t\,E \end{array}$$

where the vertical mappings are the canonical ones. Let $\mathfrak{r}$ be the kernel of ψ_1. Since ψ_1 is surjective E can be identified with $D_1/\mathfrak{r}$ and tE is the ideal $tD_1 + \mathfrak{r}/\mathfrak{r} \cong (t)/(t) \cap \mathfrak{r}$. Moreover the A-algebra $D_1/(t) \cap \mathfrak{r}$ is canonically isomorphic to the fibre product $(D_1/(t)) \times_{D_1/(t)+\mathfrak{r}} (D_1/\mathfrak{r})$. ([10] EGA 0_{IV} (19.3.10.2)).

With this property the commutative diagram

$$\begin{array}{ccc} D_0 & \xrightarrow{\psi_0} & E = D_1/\mathfrak{r} \\ {\scriptstyle \bar{\gamma}_0}\downarrow & & \downarrow \\ D_1/(t) & \xrightarrow[\bar{\psi}_1]{} & E/tE = D_1/\mathfrak{r} + (t) \end{array}$$

yields an A-algebra morphism

$$\omega\colon D_0 \to D_1/(t) \cap \mathfrak{r}$$

such that the morphisms $\bar{\gamma}_0\colon D_0 \to D_1/(t)$ and $\psi_0\colon D_0 \to E = D_1/\mathfrak{r}$ factor through ω. Since $(t) \cap \mathfrak{r}$ is nilpotent and D_0 is smooth over A, ω can be lifted to an A-algebra morphism $\gamma_0\colon D_0 \to D_1$ satisfying the required conditions.

(3.2) Let A be a noetherian ring, $\tilde{A}$ a noetherian A-algebra, D_0 a smooth A-algebra of finite type and let $t \in A$ be a nilpotent element. Suppose further that C_1 is a smooth finitely generated (A/tA)-algebra and that there are given A-algebra morphisms:

$$\psi_0\colon D_0 \to \tilde{A}$$

$$\bar{\varphi}_1\colon C_1 \to \tilde{A}/t\,\tilde{A}$$

and

$$\bar{\gamma}_0\colon D_0 \to C_1$$

such that the diagram:

$$\begin{array}{ccccc} D_0 & \xrightarrow{\psi_0} & \tilde{A} & \longrightarrow & \tilde{A}/t\,\tilde{A} \\ {\scriptstyle \bar{\gamma}_0}\downarrow & & & \nearrow_{\bar{\varphi}_1} & \\ C_1 & & & & \end{array}$$

commutes. Then:

Lemma. *There is a smooth finitely generated A-algebra D_1 and A-algebra morphisms*

$$\gamma_0\colon D_0 \to D_1 \quad \text{and} \quad \psi_1\colon D_1 \to \tilde{A}$$

such that the following conditions are satisfied:

(i) $D_1/t\,D_1 = C_1[Y_1, \ldots, Y_N]$ *is a polynomial ring over* C_1.

(ii) *The composed morphism* $D_0 \xrightarrow{\gamma_0} D_1 \xrightarrow{\text{can}} D_1/t\,D_1$ *is equal to the extended morphism*

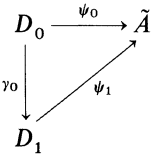
$$D_0 \xrightarrow{\bar{\gamma}_0} C_1 \xrightarrow{\text{can}} C_1[Y_1, \ldots, Y_N] = D_1/t\,D_1$$

and the morphism (induced by ψ_1*):*

$$\bar{\psi}_1 \colon D_1/t\,D_1 \to \tilde{A}/t\,\tilde{A}$$

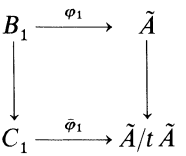
has the property: $\bar{\psi}_1 | C_1 = \bar{\varphi}_1$.

(iii) *The diagram:*



is commutative.

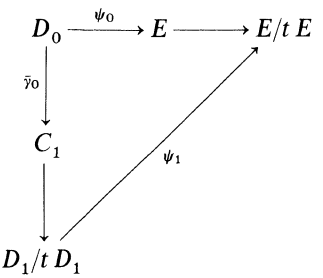
Proof. By [9] Theorem 6, p. 580, there is a finitely generated smooth A-algebra B_1 with $B_1/t\,B_1 \cong C_1$. Because t is nilpotent: $\bar{\psi}_1 \colon C_1 \to \tilde{A}/t\,\tilde{A}$ lifts to $\varphi_1 \colon B_1 \to \tilde{A}$ such that the diagram



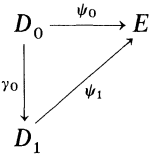
is commutative. The images of D_0 by ψ_0 and of B_1 by φ_1 are contained in a finitely generated A-subalgebra E of $\tilde{A}$. By adding variables $Y_1, \ldots, Y_N$ we get a surjective A-algebra morphism

$$\psi_1 \colon D_1 = B_1[Y_1, \ldots, Y_N] \to E$$

with $\psi_1/B_1 = \varphi_1$. Moreover the diagram:



is commutative and can be lifted by (3.1) to a commutative diagram



satisfying the required conditions.

(3.3) **Theorem.** *Assume that $\sigma\colon A\to\tilde{A}$ is a flat morphism of noetherian rings. Let $I\subseteq A$ be a nilpotent ideal and assume further that the induced morphism $\bar{\sigma}\colon A/I\to\tilde{A}/I\tilde{A}$ is a direct limit of smooth (A/I)-algebras. Then $\tilde{A}$ is a direct limit of smooth A-algebras with respect to σ.*

Proof. We assume that the statement of (3.3) is wrong. Then there is an ideal $I_0\subseteq I$ which is maximal with the property that the morphism

$$\bar{\sigma}\colon\ A/I_0\to\tilde{A}/I_0\tilde{A}$$

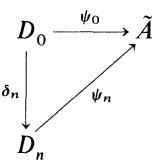
is not limit of smooth A/I_0-algebras. We replace A/I_0 by A; $\tilde{A}/I_0\tilde{A}$ by $\tilde{A}$ and $\bar{\sigma}$ by σ and may assume in the following that there is a nonzero element $t\in I$ with $t^2=0$ such that $\bar{\sigma}\colon A/tA\to\tilde{A}/t\tilde{A}$ is a direct limit of smooth A/tA-algebras.

Let C be a finitely generated A-subalgebra of $\tilde{A}$, $\rho\colon C\to\tilde{A}$ the embedding morphism. Then C is of the following form:

$$C=A[X_1,\ldots,X_n]/\mathfrak{q}.$$

We define $D_0=A[X_1,\ldots,X_n]$ and assume that $\psi_0\colon D_0\to\tilde{A}$ is the composition: $D_0\longrightarrow C\xrightarrow[\rho]{}\tilde{A}$.

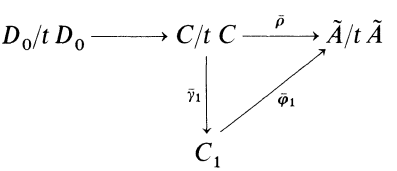
We intend to show that ρ factors over a smooth A-algebra D. Therefore we construct successively smooth A-algebras D_n and A-algebra morphisms $\delta_n\colon D_0\to D_n$ and $\psi_n\colon D_n\to\tilde{A}$ for certain $n\in\mathbb{N}$ such that the diagram



commutes. We are finished if for a $n_0\in\mathbb{N}$ $\mathfrak{q}=\operatorname{Ker}\psi_0=\operatorname{Ker}\delta_{n_0}$. Then $\rho\colon C\to A$ factors over $\psi_{n_0}\colon D_{n_0}\to A$ where D_{n_0} is a smooth A-algebra.

(i) By assumption $\tilde{A}/t\tilde{A}$ is limit of smooth A/tA-algebras. We take a smooth A/tA-algebra C_1 and A/tA-algebra morphisms $\bar{\gamma}_1\colon C/tC\to C_1$ and $\bar{\varphi}_1\colon C_1\to\tilde{A}/t\tilde{A}$ such that the diagram:

(3.3.1)



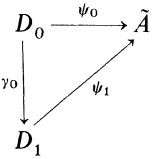
commutes.

By (3.2) there is a finitely generated smooth A-algebra D_1 and A-algebra morphisms

$$\gamma_0\colon D_0 \to D_1$$
$$\psi_1\colon D_1 \to \tilde{A}$$

such that the diagram

(3.3.2)



commutes and the morphisms γ_0 and ψ_1 are liftings and/or extensions of $\bar{\gamma}_0$ and $\bar{\psi}_1$ as described in (3.2). This implies in particular that $\gamma_0(\mathfrak{q})$ is contained in tD_1.

(iii) To continue we have to make use of the special choice of the element t and of the flatness σ. We take a system of generators $a_1, \ldots, a_r \in A$ of the ideal $\mathrm{Ann}(tA)$:

$$\mathrm{Ann}(tA) = (0:tA) = \langle a_1, \ldots, a_r \rangle A.$$

Because of the flatness of σ this implies ([16] (3.H) p. 23):

$$\mathrm{Ann}(t\tilde{A}) = (0:t\tilde{A}) = (0:tA)\tilde{A} = \langle a_1, \ldots, a_r \rangle \tilde{A}.$$

Assume further that $g_1, \ldots, g_s \in D_0$ is a system of generators of the ideal $\mathfrak{q} \subset D_0$. By (ii) we have for every $i = 1, \ldots, s$:

$$\gamma_0(g_i) = t\, l_i$$

for certain elements $l_i \in D_1$. Because of $t\psi_1(l_i) = 0$ in $\tilde{A}$ we get

$$\psi_1(l_i) \in \langle a_1, \ldots, a_r \rangle \tilde{A} \quad \text{for every } i = 1, \ldots, s.$$

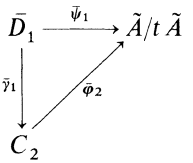
Again by using that $\tilde{A}/t\tilde{A}$ is a limit of smooth A/tA-algebras we find a smooth A/tA-algebra C_2 and A/tA-algebra morphisms:

$$\bar{\gamma}_1\colon \bar{D}_1 = D_1/tD_1 \longrightarrow C_2$$
$$\bar{\varphi}_2\colon C_2 \longrightarrow \tilde{A}/t\tilde{A}$$

satisfying the following properties:

(ii.1) The diagram

(3.3.3)



is commutative

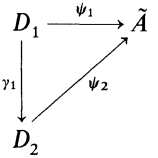
(ii.2) For every $i=1, \ldots, s$

$$\bar{\gamma}_1(\bar{l}_i) \in \langle \bar{a}_1, \ldots, \bar{a}_r \rangle C_2$$

where $\bar{l}_i$ denotes the image of l_i in $\bar{D}_1$ and $\bar{a}_j$ the image of a_j in A/tA or C_2 resp.

(iii) Again by (3.2) there is a finitely generated smooth A-algebra D_2 and A-algebra morphisms $\gamma_1: D_1 \to D_2$ and $\psi_2: D_2 \to \tilde{A}$ such that the diagram

(3.3.4)

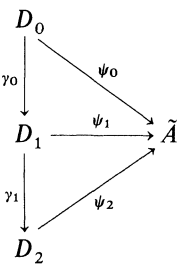


is commutative and may be considered as an extension and lifting of (3.3.3).

In particular this implies for every $i=1, \ldots, s$:

$$\gamma_1(l_i) \in \langle a_1, \ldots, a_r \rangle D_2 + t D_2 = \langle a_1, \ldots, a_r \rangle D_2.$$

(iv) We consider the composition of the diagrams (3.3.2) and (3.3.4):



Defining $\delta_2 = \gamma_1 \circ \gamma_0$ we get:

(iv.1) $\psi_0 = \psi_2 \circ \delta_2$

(iv.2) for every $i=1, \ldots, s$:

$$\delta_2(g_i) = \gamma_1 \circ \gamma_0(g_i) = \gamma_1(t\, l_i) = t\, \gamma_1(l_i) = 0$$

because $\gamma_1(l_i) \in \langle a_1, \ldots, a_r \rangle D_2 = \operatorname{Ann}(tA) D_2$.

Therefore $\mathfrak{q} \subset \operatorname{Ker} \delta_2$ and δ_2 factors over B. This proves the assertion.

(3.4) **Proposition.** *Let $\sigma: A \to \tilde{A}$ be a flat morphism of noetherian rings. Define S to be the multiplicative set in A:*

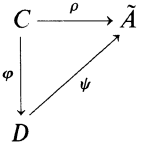
$$S = A \setminus \bigcup_{\mathfrak{g} \in \operatorname{Spec} A;\, \mathfrak{g} \min} \mathfrak{g}.$$

Assume further that the induced morphism $S^{-1}\sigma: S^{-1}A \to S^{-1}\tilde{A}$ is a limit of smooth $S^{-1}A$-algebras. Then σ is a limit of generically smooth A-algebras.

Proof. Suppose that C is a finitely generated A-subalgebra of $\tilde{A}$; $\rho: C \to \tilde{A}$ is the embedding morphism. Then we have to show that there is a finitely generated

A-algebra D and A-algebra morphisms $\varphi\colon C\to D$ and $\psi\colon D\to\tilde{A}$ such that the following conditions are satisfied:

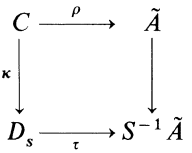
(a) The diagram



is commutative.

(b) For every $\mathfrak{p}\in\mathrm{Spec}(\tilde{A})$ with $\mathfrak{p}\cap S=\emptyset$ the A-algebra $D_{\psi^{-1}(\mathfrak{p})}$ is smooth over A.

By assumption there is a smooth $S^{-1}A$-algebra D_s and morphisms $\kappa\colon C\to D_s$ and $\tau\colon D_s\to S^{-1}\tilde{A}$ of A or $S^{-1}A$-algebras respectively such that the diagram



commutes.

Assume that D_s is of the form

$$D_s=S^{-1}A[Z_1,\ldots,Z_m]/J$$

and that $\tau(\bar{Z}_i)=z_i/t$, where $\bar{Z}_i$ is the image of Z_i in D_s and $z_i\in\tilde{A}$; $t\in S$.

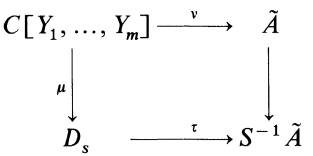
Then we define the following morphisms by extending κ or ρ resp.:

$$\mu\colon C[Y_1,\ldots,Y_m]\to D_s,\quad Y_i\to t\bar{Z}_i$$

and

$$\nu\colon C[Y_1,\ldots,Y_m]\to\tilde{A},\quad Y_i\to z_i$$

resulting in a commutative diagram:



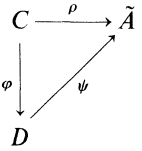
Defining

$$D=C[Y_1,\ldots,Y_m]/\mathrm{Ker}\,\mu\cap\mathrm{Ker}\,\nu$$

we get:

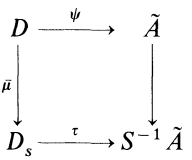
(a) ρ factors over D:

(a.1)



where φ is the canonical mapping and ψ is the morphism induced by v.

(b) Let $\bar{\mu}$: $D \to D_s$ be the morphism induced by μ then the diagram:



is commutative.

(c) By localization $S^{-1}\mu$: $S^{-1}D \to D_s$ is an isomorphism, because

(c.1) by construction of μ: $S^{-1}\mu$ is surjective

(c.2) $\operatorname{Ker} S^{-1}\mu = S^{-1}\operatorname{Ker}\mu \subseteq S^{-1}\operatorname{Ker} v = \operatorname{Ker} S^{-1} v$ hence: $S^{-1}(\operatorname{Ker}\mu \cap \operatorname{Ker} v) = S^{-1}\operatorname{Ker}\mu \cap S^{-1}\operatorname{Ker} v = \operatorname{Ker} S^{-1} v$ implying $S^{-1}\bar{\mu}$ is injective.

Therefore D is generically smooth over A with respect to ψ and (a.1) is the required factorization of ρ.

(3.3) and (3.4) imply by means of the result in § 2 the following theorem which is required in the next section. Let A be a semilocal noetherian ring whose generic formal fibres are geometrically regular. Suppose further that the separability conditions of § 2 are satisfied. Then:

(3.5) **Theorem.** *The completion of A with respect to the Jacobson radical is a direct limit of generically smooth A-algebras.*

§ 4. The approximation property of excellent rings over $\mathbb{Q}$

To prove the approximation property we use the following theorem by Elkik:

(4.1) **Theorem** ([9], Theorem 2, p. 560). *Let (A, J) be a noetherian henselian pair and h be an integer. Then there is a pair (n_0, r) of integers with the following property. Suppose that*

$$B = A[X_1, \ldots, X_N]/(f_1, \ldots, f_s)$$

is a finitely generated A-algebra; $n \geqq n_0$ is an integer and $(\underline{a}^0) = (a_1^0, \ldots, a_N^0) \in A^N$ such that

$$(*) \qquad \begin{aligned} f_j(a_1^0, \ldots, a_N^0) &\in J^n \quad \text{for every } j = 1, \ldots, s \\ H_B(\underline{a}^0) &\supseteq J^h. \end{aligned}$$

Then there is an $(\underline{a}) = (a_1, \ldots, a_N) \in A^N$ with $a_i^0 \equiv a_i \bmod J^{n-r}$ for $i = 1, \ldots, N$ and $f_j(a_1, \ldots, a_N) = 0$ for $j = 1, \ldots, s$.

The ideal $H_B \subset B$ is defined as in [9] by means of the minors of the Jacobian matrix

$$(\partial f_j/\partial X_i)_{\substack{j=1,\ldots,s \\ i=1,\ldots,N}}.$$

In particular $V(H_B) \subseteq \operatorname{Spec} B$ describes the nonsmooth locus of B over A (see also [6], (29.E), p. 219).

In the following discussion we assume that A is a semilocal excellent ring containing $\mathbb{Q}$. Let I be the Jacobson radical of A then $A^h = (A, I)^h$ denotes the henselisation of the pair (A, I) and $\hat{A} = (A, I)\hat{}$ the I-adic completion of A.

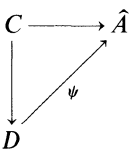
(4.2) **Theorem.** *A^h has the approximation property with respect to the I-adic completion. This means: Assume $(g_1, \ldots, g_s) \subseteq A^h[Y_1, \ldots, Y_N]$ is a system of polynomials over A^h having a solution $(\hat{y}_1, \ldots, \hat{y}_N) \in \hat{A}^N$ in $\hat{A}$, then there is a $(y_1, \ldots, y_N) \in (A^h)^N$ such that $g_i(y_1, \ldots, y_N) = 0$ for every $i = 1, \ldots, s$.*

Proof. We use induction on the dimension of A. There is nothing to prove for rings of dimension 0. Assume that the assertion is proved for rings of dimension smaller than k and let A^h be a ring of dimension k satisfying the above assumptions. Assume further that

$$C = A^h[Y_1, \ldots, Y_M]/(g_1, \ldots, g_s)$$

is an A^h-subalgebra of $\hat{A}$. We have to show that $(g_1, \ldots, g_s)$ has a solution in A^h.

By (3.5) the embedding $C \to \hat{A}$ factors over a finitely generated A^h-algebra D



such that the following condition is satisfied: For every $\mathfrak{p} \in \operatorname{Spec} \hat{A}$ contracting to a minimal prime ideal $\mathfrak{p} \cap A^h$ in A^h the A^h-algebra $D_{\psi^{-1}(\mathfrak{p})}$ is smooth over A^h. Suppose that D is presented by

$$D = A^h[X_1, \ldots, X_N]/(f_1, \ldots, f_r)$$

then the generic smoothness of D is equivalent to:

$$ht(\psi(H_D)\hat{A} \cap A^h) \geqq 1.$$

We take an element $t \in \psi(H_D)\hat{A} \cap I \subset A^h$ not contained in all minimal prime ideals of A^h and get for every $n \in \mathbb{N}$:

(a) $t^n \in I$

(b) $\dim(A^h/t^n A^h) < \dim A^h$.

We want to apply Elkik's theorem to the henselian pair $(A^h, (t))$ and to D (corresponding to B in (4.1)). Therefore we have to find an element $(\underline{a}^0)$

$=(a_1^0, \ldots, a_N^0) \in (A^h)^N$ satisfying the conditions (*) of (4.1). Here we have to make use of the induction hypothesis. First of all we define $h=1$ in (4.1) and take an integer $n \geqq \max(n_0, 2)$, where n_0 is given as in (4.1) for $h=1$. Then the ring $A^h/(t^n)$ satisfies the assumptions of (4.2). Furthermore $(A^h/(t^n), I/(t^n))$ is a henselian pair with $\dim A^h/(t^n) < k$. Therefore by induction hypothesis $A^h/(t^n)$ has the approximation property.

We consider the following system:

$$(**) \qquad \begin{aligned} & f_i(X_1, \ldots, X_N) = 0 \quad \text{for } i = 1, \ldots, r \\ & t \in H_D(X_1, \ldots, X_N). \end{aligned}$$

The second relation can be easily expressed by polynomial equations and we can consider (**) as a system of polynomials over A^h having a solution in $\hat{A}$. Taking residue classes of the coefficients in (**) we may as well consider (**) as equations over $A^h/(t^n)$ having a solution in $\hat{A}/(t^n)$. Then by the approximation property of $A^h/(t^n)$ there is an element $(\bar{a}) = (\bar{a}_1, \ldots, \bar{a}_N) \in A^h/(t^n)$ such that the following relations are satisfied in $A^h/(t^n)$:

$$\begin{aligned} & f_i(\bar{a}_1, \ldots, \bar{a}_N) = 0 \quad \text{for every } i = 1, \ldots, r \\ & t \in H_D(\bar{a}_1, \ldots, \bar{a}_N). \end{aligned}$$

Assume that $(\underline{a}^0) = (a_1^0, \ldots, a_N^0) \in (A^h)^N$ are representatives of $(\bar{a}_1, \ldots, \bar{a}_N)$. Because of $n \geqq 2$ we get:

$$(*) \qquad \begin{aligned} & f_i(a_1^0, \ldots, a_N^0) \in (t^n) \quad \text{for every } i = 1, \ldots, r \\ & t \in H_D(a_1^0, \ldots, a_N^0). \end{aligned}$$

Hence the assumptions in (4.1) are satisfied implying that $(f_1, \ldots, f_r)$ has a solution over A^h. Because C can be considered as a subalgebra of D the solution of the system $(g_1, \ldots, g_s)$ over A^h results.

References

1. André, M.: Homologie des algèbres commutatives. Berlin-Heidelberg-New York: Springer 1974
2. Artin, M.: Algebraic approximation of structures over complete local rings. Publ. Math. Inst. Hautes Etud. Sci. **36**, 23–58 (1969)
3. Artin, M.: Algebraic structure of power series rings. Contemp. Math. **13**, 223–227 (1982)
4. Artin, M., Rotthaus, C.: A structure theorem for power series rings. (preprint)
5. Artin, M., Denef, J.: Smoothing of a ring homomorphism along a section. Arithmetic and Geometry Vol. II, Birkhäuser, Boston 1983
6. Becker, J., Denef, J., Lipshitz, L., Dries, L. van den: Ultraproducts and approximation in local rings I. Invent. Math. **51**, 189–203 (1979)
7. Bourbaki, N.: Elements of mathematics: Commutative algebra. Paris: Hermann 1972
8. Denef, J., Lipshitz, L.: Ultraproducts and approximation in local rings II. Math. Ann. **253**, 1–28 (1980)
9. Elkik, R.: Solutions d'équations à coefficients dans un anneau hensélién. Ann Sci. Éc. Norm. Super. 4e sér. **6**, 553–604 (1973)
10. Grothendieck, A., Dieudonne, J.: Élements de géométrie algébrique. Publ. Math. Inst. Hautes Étud. Sci. **20** (1964)

11. Grothendieck, A., Dieudonne, J.: Élements de géométre algebrique. Publ. Math. Inst. Hautes Étud. Sci. **24** (1965)
12. Grothendieck, A., Dieudonne, J.: Élements de géométrie algebrique. Publ. Math. Inst. Hautes Étud. Sci. **32** (1967)
13. Iversen, B.: Generic local structure of the morphisms in commutative algebra. Lec. Notes Math. Vol. 310. Berlin-Heidelberg-New York: Springer 1973
14. Kurke, H., Mostowski, T., Pfister, G., Popescu, D., Roczen, M.: Die Approximationseigenschaft lokaler Ringe. Lec. Notes Math. Vol. 634. Berlin-Heidelberg-New York: Springer 1978
15. Lichtenbaum, S., Schlessinger, M.: The cotangent complex of a morphism. Trans. Am. Math. Soc. **128**, 41–70 (1967)
16. Matsumura, H.: Commutative algebra. Benjamin, New York, 1970
17. Pfister, G., Popescu, D.: Die strenge Approximationseigenschaft lokaler Ringe. Invent. Math. **30**, 145–174 (1975)
18. Pfister, G., Popescu, D.: On three-dimensional local rings with approximation property. Rev. Roum. Math. Pures Appl. **26**, 301–307 (1981)
19. Ploski, A.: Note on a theorem by M. Artin. Bull. Acad. Pol. Sci., Ser. Math. **22**, 1107–1109 (1974)
20. Popescu, D.: General Néron Desingularization. Nagoya Math. J. **100**, 97–126 (1985)
21. Raynaud, M.: Anneaux locaux henseliens. Lect. Notes Math., Vol. 169. Berlin-Heidelberg-New York: Springer 1970

Oblatum 19-XII-1985 & 19-VI-1986 & 15-IX-1986

Invent. math. 88, 65–68 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Singularités rationnelles et quotients par les groupes réductifs

Jean-François Boutot

Université Louis Pasteur, Département de Mathématiques, 7, rue René Descartes, F-67084 Strasbourg, France

Sauf mention expresse du contraire, nous emploierons le mot schéma pour désigner *un schéma de type fini sur un corps de caractéristique zéro* ou un localisé d'un tel schéma.

Rappelons tout d'abord quelques définitions.

Définition 1. *Un homomorphisme d'anneaux (commutatifs) $A \to A'$ est pur si, pour tout A-module M, l'application $M \to A' \otimes_A M$ est injective. Un morphisme de schémas affines $X' \to X$ est pur, si l'homomorphisme induit sur les anneaux de fonctions $\Gamma(X, \mathcal{O}_X) \to \Gamma(X', \mathcal{O}_{X'})$ est pur.*

Exemples. Pour que l'homomorphisme $A \to A'$ soit pur, il suffit que A soit facteur direct de A' comme A-module, ou que A' soit un A-module fidèlement plat.

Définition 2 (G. Kempf [13], chap. I, § 3). *Un schéma X est à singularités rationnelles s'il est normal et si, pour $f\colon Z \to X$ une résolution des singularités de X, l'une des conditions équivalentes suivantes est vérifiée*:

a) $R^i f_* \mathcal{O}_Z = 0$ *pour* $i > 0$.

b) X *est Cohen-Macaulay et* $f_* \omega_Z = \omega_X$

(où ω_Z est le faisceau dualisant sur Z et ω_X le faisceau dualisant sur X).

Remarque. D'après Hironaka [7], il existe toujours une résolution des singularités $f\colon Z \to X$ et les conditions a) et b) sont indépendantes de la résolution choisie. L'équivalence des deux conditions résulte du théorème de dualité pour le morphisme f.

Pour X normal, le faisceau ω_X peut s'interpréter comme faisceau des différentielles régulières. Le faisceau $f_* \omega_Z$ est indépendant de la résolution choisie, on le note parfois K_X, c'est le faisceau des différentielles absolument régulières. On a toujours une injection canonique $K_X \subset \omega_X$, la condition $f_* \omega_Z = \omega_X$ signifie donc que toute différentielle régulière est absolument régulière.

L'objet de cet article est de prouver le résultat suivant (annoncé dans [8]):

Théorème. *Soit $X' \to X$ un morphisme pur de schémas affines. Alors, si X' est à singularités rationnelles, il en est de même de X.*

La principale application de ce théorème est le

Corollaire. *Soient k un corps de caractéristique zéro et G un groupe algébrique réductif sur k agissant k-rationnellement sur un k-schéma affine $X=\mathrm{Spec}(R)$ à singularités rationnelles. Alors le quotient $X/G=\mathrm{Spec}(R^G)$ est à singularités rationnelles.*

Ceci résulte du théorème puisque l'opérateur de Reynolds $R\to R^G$ fait de R^G un facteur direct de R comme R^G-module.

Remarques. 1) Le corollaire améliore un résultat de M. Hochster et J.L. Roberts [9] qui montraient que X/G est Cohen-Macaulay lorsque X régulier. Leur démonstration procède par réduction à la caractéristique $p>0$ et ils montrent que, si $X'\to X$ est un morphisme pur de schémas affines de caractéristique $p>0$ et si X' est régulier, alors X est Cohen-Macaulay (voir aussi [2]).

2) Le corollaire étend également des résultats de E. Viehweg [14] et D. Burns [3] qui traitaient le cas d'un groupe G fini et de G. Kempf [12] qui traitait le cas d'un groupe G dont la composante neutre est semi-simple et d'un schéma X régulier et factoriel.

3) Ce résultat a été utilisé dans l'étude des singularités des adhérences des orbites de certaines représentations des groupes classiques: classes de conjugaison des groupes symplectique et orthogonal (H. Kraft et C. Procesi [10]) et classes d'équivalence de représentations du carquois A_n (S. Abeasis, A. Del Fra et H. Kraft [1]).

La démonstration utilise le théorème d'annulation de Grauert et Riemenschneider [5], variante du théorème de Kodaira, sous une forme duale déjà employée par Hartshorne et Ogus [6].

Proposition ([6], prop. 2.2.). *Soient X un schéma local normal, $f\colon Z\to X$ une résolution des singularités de X et Y la fibre fermée de f. Alors on a $H^i_Y(Z,\mathcal{O}_Z)=0$ pour $i<\dim X$.*

Rappelons-en brièvement la démonstration. Notons $d=\dim X$, $\mathfrak{m}$ l'idéal maximal de $\Gamma(X,\mathcal{O}_X)$ et, pour tout $\Gamma(X,\mathcal{O}_X)$-module M, $M^\wedge$ le complété de M pour la topologie $\mathfrak{m}$-adique. On a:

$$H^i_Y(Z,\mathcal{O}_Z)=\varinjlim \mathrm{Ext}^i(\mathcal{O}_Z/\mathfrak{m}^n\mathcal{O}_Z,\mathcal{O}_Z)$$

et, d'après le théorème de comparaison de Grothendieck,

$$H^j(Z,\omega_Z)^\wedge=\varprojlim H^j(Z,\omega_Z/\mathfrak{m}^n\omega_Z).$$

Mais $\mathrm{Ext}^i(\mathcal{O}_Z/\mathfrak{m}^n\mathcal{O}_Z,\mathcal{O}_Z)=\mathrm{Ext}^i(\omega_Z/\mathfrak{m}^n\omega_Z,\omega_Z)$ est dual de $H^{d-i}(Z,\omega_Z/\mathfrak{m}^n\omega_Z)$, donc $H^i_Y(Z,\mathcal{O}_Z)$ est dual de $H^{d-i}(Z,\omega_Z)^\wedge$. Enfin le théorème de Grauert-Riemenschneider assure que $H^j(Z,\omega_Z)=0$ pour $j>0$.

Démonstration du théorème. On montre facilement que X est normal (cf. [9], prop. 6.15). Par localisation et récurrence sur la dimension, on peut supposer de plus que X est local et à singularités rationnelles en dehors du point fermé. Enfin on peut supposer que X' est intègre (sinon $X'=\coprod X'_i$ et l'un des morphismes $X'_i\to X$ est pur (cf. *loc. cit.*, cor. 6.12)).

Soit $Z \to X$ une résolution des singularités de X. L'une des composantes irréductibles de $Z \times_X X'$ est birationnelle à X', soit Z' une résolution des singularités de cette composante. On obtient ainsi un diagramme commutatif

$$\begin{array}{ccc} X' & \longleftarrow & Z' \\ \downarrow & & \downarrow \\ X & \longleftarrow & Z \end{array}$$

où Z (resp. Z') est une résolution des singularités de X (resp. X').

Soient U l'ouvert complémentaire du point fermé dans X et U' (resp. V, V') l'image réciproque de U dans X' (resp. Z, Z'). Considérons, pour $i \geqq 1$, le diagramme commutatif d'applications canoniques:

$$\begin{array}{ccccc} H^i(Z', \mathcal{O}_{Z'}) & \longrightarrow & H^i(V', \mathcal{O}_{V'}) & \xleftarrow{h_{U'V'}} & H^i(U', \mathcal{O}_{U'}) \\ \uparrow{\scriptstyle h_{ZZ'}} & & \uparrow & & \uparrow{\scriptstyle h_{UU'}} \\ H^i(Z, \mathcal{O}_Z) & \xrightarrow[h_{ZV}]{} & H^i(V, \mathcal{O}_V) & \xleftarrow[h_{UV}]{} & H^i(U, \mathcal{O}_U) \end{array}$$

Il s'agit de montrer que $H^i(Z, \mathcal{O}_Z)=0$ sachant que $H^i(Z', \mathcal{O}_{Z'})=0$. En fait, nous allons montrer que l'application $h_{ZZ'}$ est injective.

La suite exacte de cohomologie locale pour $Y=Z-V$ et l'annulation de $H^i_Y(Z, \mathcal{O}_Z)$ pour $i<\dim X$ (proposition) montrent que h_{ZV} est toujours injectif. Comme V est une résolution des singularités de U et que U est à singularités rationnelles, la suite spectrale de Leray pour $V \to U$ montre que h_{UV} est bijectif; pour la même raison, $h_{U'V'}$ est bijectif. Enfin, en considérant un complexe $K^\bullet$ définissant la cohomologie $H^i(U, \mathcal{O}_U)$ et le complexe $K^\bullet \otimes_{\Gamma(X, \mathcal{O}_X)} \Gamma(X', \mathcal{O}_{X'})$ définissant $H^i(U', \mathcal{O}_{U'})$ et en utilisant la pureté du morphisme $X' \to X$, on montre facilement que $h_{UU'}$ est injectif (cf. [9], cor. 6.8).

Remarques. 1) Les conditions de finitude imposées à X et X' viennent de l'emploi du théorème de Grauert-Riemenschneider dans la démonstration, mais elles ne devraient pas être essentielles. En fait, il semble raisonnable de conjecturer que la proposition reste vraie lorsque X est le spectre d'un anneau local noethérien complet d'égale caractéristique zéro.

2) Il apparaît encore une fois que le théorème de Grauert-Riemenschneider est l'instrument idéal pour établir des énoncés généraux relatifs aux singularités rationnelles. On en trouvait déjà la preuve dans les travaux de G. Kempf [11] et de R. Elkik [4].

Bibliographie

1. Abeasis S., Del Fra A., Kraft H.: The geometry of representations of A_n. Math. Ann. **256**, 401–418 (1981)
2. Boutot J.-F.: Frobenius et cohomologie locale, in séminaire Bourbaki, exposé 453, nov. 1974. Lect. Notes Math. **514**. Berlin Heidelberg New York: Springer (1976)

3. Burns D.: On rational singularities in dimensions >2. Math. Ann. **211**, 237–244 (1974)
4. Elkik R.: Singularités rationnelles et déformations. Invent. math. **47**, 139–147 (1978)
5. Grauert H., Riemenschneider O.: Verschwindungssätze für analytische Kohomologiegruppen auf komplexen Räumen. Invent. math. **11**, 263–290 (1970)
6. Hartshorne R., Ogus A.: On the factoriality of local rings of small embedding codimension. Commun. Algebra **1**, 415–437 (1974)
7. Hironaka H.: Resolution of singularities of an algebraic variety over a field of characteristic zero. Ann. Math. **79**, 109–326 (1964)
8. Hochster M.: Cohen-Macaulay rings and modules. In: Proceedings of the International Congress of Mathematicians, Helsinki 1978, vol. 1, pp. 291–298 Helsinki: Academia Scientiarium Fennica 1980
9. Hochster M., Roberts J.L.: Rings of invariants of reductive groups acting on regular rings are Cohen-Macaulay. Adv. Math., **13**, 115–175 (1974)
10. Kraft H., Procesi C.: On the geometry of conjugacy classes in classical groups. Comment. Math. Helvetici **57**, 539–602 (1982)
11. Kempf G.: On the collapsing of homogeneous bundles. Invent. math. **37**, 229–239 (1976)
12. Kempf G.: Some quotient varieties have rational singularities, Michigan Math. J. **24**, 347–352 (1977)
13. Kempf G., Knudsen F., Mumford D., Saint-Donat B.: Toroidal embeddings I. Lect. Notes Math. **339**. Berlin Heidelberg New York: Springer 1973
14. Viehweg E.: Rational singularities of higher dimensional schemes. Proc. Am. Math. Soc. **63**, 6–8 (1977)

Oblatum 11-VII-1986

Invent. math. 88, 69–81 (1987)

Inventiones mathematicae

© Springer-Verlag 1987

On Thue's equation

E. Bombieri [1] and W.M. Schmidt [1,2]

[1] Institute for Advanced Study, Department of Mathematics, Princeton, NJ 08540, USA
[2] Department of Mathematics, University of Colorado, Boulder CO 80309, USA

I. Introduction

Let $F(x, y)=a_0 x^r+a_1 x^{r-1} y+\ldots+a_r y^r$ be a binary form with rational integral coefficients, of degree $r \geqq 3$.

In this paper we study the problem of providing upper bounds for the number of solutions of the Thue equation

$$F(x, y)=h, \tag{1.1}$$

and more precisely for the number of *primitive* solutions to (1.1), that is, solutions in coprime integers x, y.

The first important results on the number of solutions of Thue's equation were obtained by C.L. Siegel, in the case $r=3$ and in the binomial case $F(x, y)=ax^r-by^r$. In view of his results and comments, we should attribute to Siegel the problem to decide whether the number of primitive solutions to (1.1), for irreducible F of degree $r \geqq 3$, could be bounded by a function depending only on r and h, but otherwise independent of F.

In 1983, Siegel's question was answered in the affirmative by J.-H. Evertse [2]. As a special case of his results (he also treats equations in number fields), Evertse obtains the bound

$$7^{15((\binom{r}{3}+1)^2}+6 \times 7^{2\binom{r}{3}(t+1)}$$

for the number of primitive solutions of (1.1), where t is the number of prime factors of the constant term h. Evertse's method uses ideas of Siegel and Baker to reduce the problem to equations in units in a number field and then to equations of binomial type $ax^r-by^r=c$, again in number fields. Finally techniques of Padé approximations are used to deal with binomial equations.

We shall prove the following result.

Theorem. *Let $F(x, y)$ be an irreducible binary form of degree $r \geqq 3$, with rational integral coefficients. Then the number of primitive solutions of the equation*

$$|F(x, y)|=h \tag{1.2}$$

does not exceed

$$c_1 r^{1+t}$$

where c_1 *is an absolute constant and* t *is the number of distinct prime factors of* h. *When* $r > c_2$, *the number of primitive solutions (with* (x, y) *and* $(-x, -y)$ *regarded as the same) does not exceed*

$$215 r^{1+t}.$$

The arguments in this paper can be used to prove more general results. For example, if N_r is the corresponding bound in the special case $h=1$, one obtains $N_r r^t$ as a bound in the general case; this simple result appears to be new. Thus Evertse's theorem $N_3 \leqq 12$ (see [3]) yields the much better bound 12×3^t in the case $r=3$. Other improved bounds (for example independent of h) may be obtained on additional assumptions, such as F having a large discriminant in comparison with a power of h.

It is perhaps worthwhile to indicate at this point the new aspects of our approach. The general line of attack to the problem, inaugurated by Siegel and Mahler and followed by almost every other researcher, deals rather efficiently with solutions x, y to $F(x, y)=1$, as long as the maximum between $|x|$ and $|y|$ is larger than a certain power of the height of F. The reason for success in this case lies in the fact that if either x or y is large, then $\frac{x}{y}$ is a very good approximation to a root of the equation $F(\xi, 1)=0$ and one can use diophantine approximation methods quite profitably.

There remains the problem of dealing with small solutions. Our approach depends on the remark that if we transform $F(x, y)$ by the action of an element of $SL(2, \mathbf{Z})$ the problem of counting solutions remains unchanged, while the diophantine approximation properties of F can change very drastically. Transformations in $GL(2, \mathbf{Z})$ are also used to dispense with a technical hypothesis about the height of F.

The last part of the proof deals with the general case $|F(x, y)|=h$ and its reduction to equations of type $|F_i(x', y')|=1$. This is done by means of elementary p-adic analysis. In the case in which h is coprime with the discriminant of F our analysis is essentially equivalent to a simpler argument by Mahler [7]; needless to say, it was Mahler's paper which inspired the present treatment, although the final result also appears to be new. See however Hooley [4].

In the other direction, one may ask how large can N_r be. The example

$$F(x, y)=x^r+c(x-y)(2x-y)\ldots(rx-y)$$

shows that $N_r \geqq r$ for every r; hence our upper bound $c_1 r$ is best possible except for the determination of c_1.

We have not discussed here the very important results obtained by Baker and others on the problem of locating effectively the solutions of Thue's equation, having been content with our much more modest goal of counting solutions. The interested reader will find an ample bibliography on the subject of Thue's equation in Evertse [2].

We note also that there is a very different approach to the problem of counting the number of solutions of Thue's equation, invented by Silverman [9]; he relates the number of primitive solutions to the rank of the Mordell-Weil group on the jacobian. It would be of great interest to combine in an efficient way the old method based on diophantine approximation and Silverman's method.

We wish also to thank here Julia Mueller for the long conversations we had on this subject.

II. Construction and normalization of auxiliary forms

Let $F(x, y)=a_0 x^r+a_1 x^{r-1} y+\ldots+a_r y^r$ be an irreducible binary form with rational integral coefficients and let α_i, $i=1, \ldots, r$, be the set of roots of the associated equation $F(\xi, 1)=0$. We shall initially be concerned with counting the number of solutions of

$$|F(x, y)|=1 \tag{2.1}$$

in integers x, y. Let N_r be the maximum number of solutions, for all forms F as above, and with solutions (x, y), $(-x, -y)$ regarded as the same.

For $A=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ let us write F_A for the binary form

$$F_A(x, y)=F(ax+by, cx+dy).$$

We also write $\boldsymbol{x}$ for (x, y) and $A\boldsymbol{x}$ for $(ax+by, cx+dy)$.

If $D(F)=a_0^{2r-2}\prod_{i<j}(\alpha_i-\alpha_j)^2$ is the discriminant of F, then

$$D(F_A)=(\det A)^{r(r-1)} D(F). \tag{2.2}$$

Let $A_0=\begin{pmatrix} p & 0 \\ 0 & 1 \end{pmatrix}$, $A_j=\begin{pmatrix} 0 & -1 \\ p & j \end{pmatrix}$ $(j=1, \ldots, p)$ where p is a prime number. Then it is easily seen that

$$\mathbf{Z}^2=\bigcup_{j=0}^{p} A_j \mathbf{Z}^2.$$

Therefore the number of solutions of (2.1) does not exceed $n_0+n_1+\ldots+n_p$, where n_j is the number of solutions of $|F_{A_j}(x, y)|=1$. Note that $\det A_j=p$, so that $|D(F_{A_j})| \geqq p^{r(r-1)}$ by (2.2). Therefore

$$N_r \leqq (p+1) N_r(p), \tag{2.3}$$

where $N_r(p)$ is the maximum number of solutions of (2.1) for all forms F with

$$|D(F)| \geqq p^{r(r-1)}. \tag{2.4}$$

If $A \in SL(2, \mathbf{Z})$, then we say that F_A and $-F_A$ are *equivalent* to F. The number of solutions of (2.1) is the same for equivalent forms. It will be convenient to replace the form F by an equivalent form which is easy to work with. Suppose that (2.1) has at least one solution $\boldsymbol{x}^0$. There is $A \in SL(2, \mathbf{Z})$ such that $A^{-1}\boldsymbol{x}^0$

$=(1,0)$ and hence $(1,0)$ is a solution of the relation $|F_A(x,y)|=1$. Here F_A has leading coefficient ± 1, so that $\pm F_A$ has leading coefficient 1. We may therefore restrict ourselves to forms F which are *normalized* in the sense that their leading coefficient $a_0=1$.

Let

$$M(F)=\prod_{i=1}^{r}\max(1,|\alpha_i|)$$

be the Mahler height of the normalized form F. We say that a form F is *reduced* if it is normalized and it has smallest Mahler height among all normalized forms equivalent to F. Every form F with a solution to (2.1) is equivalent to a reduced form, although the reduced form need not be unique. In what follows we will suppose that our form F is reduced.

The discriminant of a form is invariant under equivalence, so that when dealing with $N_r(p)$ again we may suppose that F is reduced. It is known (Mahler [6]) that

$$|D(F)|\leqq r^r\, M(F)^{2r-2}, \tag{2.5}$$

so that (2.4) leads to

$$M(F)\geqq p^{r/2}\, r^{-r/(2r-2)}. \tag{2.6}$$

III. Large solutions

The *height* of an integer point (x,y) with coprime components is given by

$$H(x,y)=\max(|x|,|y|).$$

In this section we give an upper bound for the number of solutions (x,y) to a Thue equation $|F(x,y)|=1$, provided the height $H(x,y)$ is sufficiently large.

We begin with a refinement of an estimate of Lewis and Mahler [5].

Lemma 1. *For any (x,y) with $y\neq 0$ we have*

$$\min_{\alpha}\min\left(1,\left|\alpha-\frac{x}{y}\right|\right)\leqq\frac{(2r^{\frac{1}{2}}\,M(F))^r\,|F(x,y)|}{H(x,y)^r},$$

where the first minimum on the left is over the roots of $F(\xi,1)$.

Remark. It is easily seen that the proof in Lewis and Mahler allows to gain also a factor $|a_0|^2\sqrt{|D(F)|}$ where $D(F)$ is the discriminant of F. For simplicity, we will not make use here of this fact.

Proof. The proof is essentially in Lewis and Mahler [5], inequality (16), p. 341, with the following changes:

(a) in place of [5], (1), p. 335, one uses (2.5),

(b) in place of [5], (4), p. 336, one uses $M(g)\leqq M(f)$, which is obvious. Then [5], (8), p. 37 becomes

$$|f'(\xi)|\geqq\frac{|D(F)|^{\frac{1}{2}}}{n^{(n/2)-(\frac{1}{2})}\,M(F)^{n-2}}.$$

As a consequence, we obtain [5], (16), p. 341 with

$$\Lambda = \frac{|D(F)|^{\frac{1}{2}}}{2^{n-1}\, n^{(n/2)-(\frac{1}{2})}\, M(F)}.$$

Since $\sigma \leqq M(F)/|a_0| \leqq M(F)$, the lemma follows.

Let us fix a root α_0 of $F(\xi, 1)$, let Y_0 be given and let (x_1, y_1), (x_2, y_2), ... be the sequence of solutions to $|F(x, y)| = 1$ with $y > 0$,

$$\left|\alpha_0 - \frac{x}{y}\right| = \min_{\alpha} \left|\alpha - \frac{x}{y}\right|,$$

and of height $\geqq Y_0$, ordered such that

$$Y_0 \leqq H(x_1, y_1) \leqq H(x_2, y_2) \leqq \ldots.$$

By the preceding lemma, and assuming that $Y_0 \geqq C^{1/r}$,

$$\frac{1}{y_n y_{n+1}} \leqq \left|\frac{x_n}{y_n} - \frac{x_{n+1}}{y_{n+1}}\right| \leqq \frac{2C}{H(x_n, y_n)^r} \qquad (n = 1, 2, \ldots),$$

where we have abbreviated

$$C = (2r^{\frac{1}{2}}\, M(F))^r. \tag{3.1}$$

We infer that

$$H(x_{n+1}, y_{n+1}) \geqq y_{n+1} \geqq H(x_n, y_n)^r/(2C y_n) \geqq H(x_n, y_n)^{r-1}/2C.$$

An easy induction now yields

Strong gap principle. We have for $n = 1, 2, \ldots$

$$H(x_n, y_n) \geqq \left\{(2C)^{-\frac{1}{r-2}} H(x_1, y_1)\right\}^{(r-1)^{n-1}}.$$

The first step in obtaining a bound for the number of large solutions consists in showing that $H(x_n, y_n)$ cannot be very much larger than $H(x_1, y_1)$. This is done as follows.

Let t, τ be positive numbers such that

$$t < \sqrt{2/r}, \qquad \sqrt{2 - rt^2} < \tau < t, \tag{3.2}$$

and put

$$\lambda = 2/(t - \tau), \tag{3.3}$$

$$A_1 = \frac{t^2}{2 - rt^2}\left(\log M(F) + \frac{r}{2}\right). \tag{3.4}$$

We will suppose that $\lambda < r$. We say that a rational number x/y is a *very good approximation* to α if

$$\left|\alpha - \frac{x}{y}\right| < (4e^{A_1} H(x, y))^{-\lambda}.$$

We have the

Thue-Siegel principle. If α is of degree r and x/y, x'/y' are two very good approximations to α then

$$\log(4e^{A_1})+\log H(x', y')\leqq\delta^{-1}\{\log(4e^{A_1})+\log H(x, y)\}$$

where

$$\delta=(rt^2+\tau^2-2)/(r-1). \tag{3.5}$$

Proof. If $|\alpha|\leqq 1$, this is a special case of the Thue-Siegel principle in [1], p. 179, with $\alpha_1=\alpha_2=\alpha$, $\beta_1=x/y$, $\beta_2=x'/y'$, $K=\mathbf{Q}$, v the usual absolute value, $\vartheta=1$, on using the estimate (15), p. 184 of [1] for A_1 and the value (14), p. 183 of [1] for τ, noting also that

(i) the hypothesis $K_{\bar{v}}=k_v$ is not used in the proof and therefore may be omitted;

(ii) $c(\vartheta t)\leqq\log 3$;

(iii) the chosen value for A_1 implies a fortiori $\left|\alpha-\frac{x}{y}\right|<\vartheta(t-\tau)$;

(iv) $h(\beta)=H(x, y)$ when $\beta=x/y$;

(v) the exponent in (5A), p. 179 of [1] should be $2\vartheta/(t-\tau)$, not $2\vartheta^{-1}/(t-\tau)$.

If $|\alpha|>1$ we note that the assumption that x/y is a very good approximation to α implies that

$$\left|\alpha^{-1}-\frac{y}{x}\right|\leqq|\alpha|^{-1}\,|y/x|(4e^{A_1}H(x, y))^{-\lambda}<(3e^{A_1}H(x, y))^{-\lambda}$$

and we may apply the same argument with α^{-1}, y/x, y'/x' in place of α, x/y, x'/y', noting that $F(y, x)$ has the same Mahler height as $F(x, y)$.

In order to combine the strong gap principle and the Thue Siegel principle we must choose Y_0 so large that the approximations $\frac{x_1}{y_1}, \ldots, \frac{x_n}{y_n}, \ldots$ are all very good approximations to α. To this end it suffices by the preceding lemma that

$$\frac{C}{H(x_i, y_i)^r}\leqq\frac{1}{(4e^{A_1}H(x_i, y_i))^{\lambda}}$$

for every i; thus we need

$$Y_0\geqq C^{\frac{1}{r-\lambda}}(4e^{A_1})^{\frac{\lambda}{r-\lambda}},$$

which we shall suppose.

Now we can apply the Thue-Siegel principle with the lower bound for $\log H(x_n, y_n)$ provided by the strong gap principle, and we find

$$\log(4e^{A_1})+(r-1)^{n-1}\left\{\log H(x_1, y_1)-\frac{1}{r-2}\log(2C)\right\}$$
$$\leqq\delta^{-1}\{\log(4e^{A_1})+\log H(x_1, y_1)\},$$

hence

$$(r-1)^{n-1}\leqq\delta^{-1}\,\frac{\log Y_0+\log(4e^{A_1})}{\log Y_0-(r-2)^{-1}\log(2C)},$$

provided $\log Y_0>(r-2)^{-1}\log 2C$, which is the case.

We choose

$$Y_0=(2C)^{\frac{1}{r-\lambda}}(4e^{A_1})^{\frac{\lambda}{r-\lambda}} \tag{3.6}$$

and get

$$(r-1)^{n-1}\leqq\delta^{-1}\frac{(r-\lambda)^{-1}\log(2C)+r(r-\lambda)^{-1}\log(4e^{A_1})}{(\lambda-2)(r-2)^{-1}(r-\lambda)^{-1}\log 2C+\lambda(r-\lambda)^{-1}\log(4e^{A_1})}$$
$$\leqq\delta^{-1}\frac{r-2}{\lambda-2}.$$

This proves that

$$n<2+\frac{\log(\delta^{-1}(\lambda-2)^{-1})}{\log(r-1)}.$$

We summarize our result as follows. Let C, t, τ, λ, A_1, δ, Y_0 be given with (3.1), (3.2), (3.3), (3.4), (3.5), (3.6) and with $\lambda<r$.

Lemma 2. *With the above choice of Y_0, the number of solutions (x, y) of the Thue equation $|F(x, y)|=1$ (where (x, y), $(-x, -y)$ are counted as one) with $H(x, y)\geqq Y_0$ does not exceed*

$$\left(2+\left[\frac{\log(\delta^{-1}(\lambda-2)^{-1})}{\log(r-1)}\right]\right)r.$$

If we choose

$$t=\sqrt{2/(r+a^2)},\qquad \tau=bt \tag{3.7}$$

with $0<a<b<1$, then

$$\delta^{-1}<r^2/(2(b^2-a^2)),$$

$$\lambda=\frac{2}{(1-b)t}>\frac{\sqrt{2r}}{1-b},$$

and we see that our upper bound, for fixed a, b, is $3r$ provided r is sufficiently large.

IV. Small solutions

We suppose here that F is reduced, with (2.6). When

$$F(\boldsymbol{x})=F(x, y)=(x-\alpha_1 y)\ldots(x-\alpha_r y),$$

put

$$L_i(\boldsymbol{x})=x-\alpha_i y\qquad(i=1, \ldots, r).$$

Given $\boldsymbol{x}=(x, y)$, $\boldsymbol{x}_0=(x_0, y_0)$, put $D(\boldsymbol{x}, \boldsymbol{x}_0)=xy_0-x_0y$.

Lemma 3. *Suppose $\boldsymbol{x}$, $\boldsymbol{x}_0$ are solutions of* (2.1). *Then for $1\leqq i, j\leqq r$,*

$$\frac{L_i(\boldsymbol{x}_0)}{L_i(\boldsymbol{x})}-\frac{L_j(\boldsymbol{x}_0)}{L_j(\boldsymbol{x})}=(\beta_i-\beta_j)D(\boldsymbol{x}, \boldsymbol{x}_0), \tag{4.1}$$

where $\beta_1, \ldots, \beta_r$ depend on $\boldsymbol{x}$ and are such that the form $G(v, w)=(v-\beta_1 w)\ldots(v-\beta_r w)$ is equivalent to F.

Proof. Pick $x' \in \mathbf{Z}^2$ with $D(x', x)=1$; this is possible since x with $|F(x)|=1$ has coprime components. Then x, x' is a basis of $\mathbf{Z}^2$ and we may write x_0 as $x_0 = ax + bx'$. In fact $D(x_0, x)=b$, so that

$$x_0 = ax + D(x_0, x)\, x'.$$

Therefore for each i in $1 \leqq i \leqq r$

$$\frac{L_i(x_0)}{L_i(x)} = a + D(x_0, x)\frac{L_i(x')}{L_i(x)} = a - D(x, x_0)\,\beta_i$$

with $\beta_i = -L_i(x')/L_i(x)$, and (4.1) follows. Set

$$G(v, w) = \pm F(vx + wx'),$$

with the sign chosen such that $G(1, 0) = \pm F(x) = 1$, so that G is normalized. Also, $G \sim F$ and

$$G(v, w) = \pm \prod_{i=1}^{r} L_i(vx + wx') = \pm \prod_{i=1}^{r} (vL_i(x) + wL_i(x'))$$

$$= \prod_{i=1}^{r} (v + w(L_i(x')/L_i(x))) = \prod_{i=1}^{r} (v - \beta_i w).$$

In the special case when $x_0 = (1, 0)$, we have $L_i(x_0) = L_j(x_0) = 1$, and (4.1) becomes

$$\frac{1}{L_i(x)} - \frac{1}{L_j(x)} = (\beta_j - \beta_i)\, y.$$

For every x with $|F(x)|=1$, so that $|L_1(x)| \ldots |L_r(x)| = 1$, we choose $j = j(x)$ with $|L_j(x)| \geqq 1$; then

$$\frac{1}{|L_i(x)|} \geqq |\beta_j - \beta_i|\,|y| - 1.$$

Also, $|\overline{L_j(x)}| \geqq 1$, thus

$$\frac{1}{|L_i(x)|} \geqq |\bar{\beta}_j - \beta_i|\,|y| - 1$$

and finally

$$\frac{1}{|L_i(x)|} \geqq |\mathrm{Re}(\beta_j) - \beta_i|\,|y| - 1.$$

We also choose an integer $m = m(x)$ with $|m - \mathrm{Re}(\beta_j)| \leqq \frac{1}{2}$ and we obtain

$$\frac{1}{|L_i(x)|} \geqq (|m - \beta_i| - \tfrac{1}{2})\,|y| - 1 \qquad (i = 1, \ldots, r). \tag{4.2}$$

For $1 \leqq i \leqq r$, let $\mathfrak{X}_i$ be the set of solutions of (2.1) with $1 \leqq y \leqq Y_0$ and $|L_i(x)| \leqq 1/(2y)$.

Lemma 4. *Suppose $x \neq \tilde{x}$ with $y \leqq \tilde{y}$ belong to $\mathfrak{X}_i$. Then*

$$\tilde{y}/y \geqq \tfrac{2}{7} \max(1, |\beta_i - m|) \tag{4.3}$$

where $\beta_i = \beta_i(x)$, $m = m(x)$.

Proof.

$$1 \leqq |D(\boldsymbol{x}, \tilde{\boldsymbol{x}})| = \begin{Vmatrix} x-\alpha_i y & y \\ \tilde{x}-\alpha_i \tilde{y} & \tilde{y} \end{Vmatrix} \leqq y\,|L_i(\tilde{\boldsymbol{x}})| + \tilde{y}\,|L_i(\boldsymbol{x})|$$
$$\leqq (y/2\tilde{y}) + \tilde{y}\,|L_i(\boldsymbol{x})| \leqq \tfrac{1}{2} + \tilde{y}\,|L_i(\boldsymbol{x})|,$$

so that $\tilde{y}\,|L_i(\boldsymbol{x})| \geqq \frac{1}{2}$. Combining this with (4.2) we get

$$\tilde{y} \geqq \tfrac{1}{2}(|m-\beta_i| - \tfrac{1}{2})\,y - \tfrac{1}{2},$$

so that

$$\tilde{y}/y \geqq \max(1, \tfrac{1}{2}\,|m-\beta_i| - \tfrac{3}{4}).$$

But for any real $\xi \geqq 0$

$$\max(1, \tfrac{1}{2}\xi - \tfrac{3}{4}) \geqq \tfrac{2}{7}\max(1, \xi),$$

and (4.3) follows.

Lemma 5. *Suppose* $\boldsymbol{x}$ *with* $y>0$ *and* $|L_i(\boldsymbol{x})| > 1/(2y)$ *has* $|F(\boldsymbol{x})| = 1$. *Then*

$$|m-\beta_i| \leqq \tfrac{7}{2}$$

where again $\beta_i = \beta_i(\boldsymbol{x})$, $m = m(\boldsymbol{x})$.

Proof. From (4.2),

$$|m-\beta_i| \leqq \frac{1}{2} + \frac{1}{|L_i(\boldsymbol{x})|\,y} + \frac{1}{y} \leqq \frac{1}{2} + 2 + 1 = \frac{7}{2}.$$

For each set $\mathfrak{X}_i$ which is not empty, let $\boldsymbol{x}(i)$ be the element with the largest value of y. Let $\mathfrak{X}$ be the set of solutions of $|F(\boldsymbol{x})| = 1$ with $1 \leqq y \leqq Y_0$, where the elements $\boldsymbol{x}(1), \ldots, \boldsymbol{x}(r)$, in so far as they exist, have been removed.

Lemma 6. *Suppose* (2.6) *holds with a prime* $p > (\frac{7}{2})^2$. *Then when* $\varepsilon > 0$ *and* $r > r_1(p, \varepsilon)$, *the set* $\mathfrak{X}$ *has cardinality*

$$|\mathfrak{X}| < \frac{r(1+\varepsilon)}{1-(2\log(\frac{7}{2}))/\log p}.$$

Proof. Let i in $1 \leqq i \leqq r$ be fixed for the moment. In the case when $\mathfrak{X}_i$ is not empty, order its elements as $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_v$ with $y_1 \leqq \ldots \leqq y_v$ (so that $\boldsymbol{x}(i) = \boldsymbol{x}_v$). Lemma 4 yields

$$\tfrac{2}{7}\max(1, |\beta_i(\boldsymbol{x}_k) - m(\boldsymbol{x}_k)|) \leqq y_{k+1}/y_k$$

for $k = 1, 2, \ldots, v-1$, so that

$$\prod_{\boldsymbol{x} \in \mathfrak{X} \cap \mathfrak{X}_i} (\tfrac{2}{7}\max(1, |\beta_i(\boldsymbol{x}) - m(\boldsymbol{x})|)) \leqq Y_0.$$

For $\boldsymbol{x}$ in $\mathfrak{X}$ but not in $\mathfrak{X}_i$ we have

$$\tfrac{2}{7}\max(1, |\beta_i(\boldsymbol{x}) - m(\boldsymbol{x})|) \leqq 1$$

by Lemma 5. We may conclude that

$$\prod_{\boldsymbol{x} \in \mathfrak{X}} (\tfrac{2}{7}\max(1, |\beta_i(\boldsymbol{x}) - m(\boldsymbol{x})|)) \leqq Y_0. \tag{4.4}$$

The form $G=\Pi(v-\beta_i w)$ of Lemma 4 was equivalent to F, and hence so is the form $\hat{G}=\Pi(v-(\beta_i-m)w)$. Since F was reduced, $M(\hat{G})\geqq M(F)$, so that

$$\prod_{i=1}^{r}\max(1,|\beta_i(\boldsymbol{x})-m(\boldsymbol{x})|)=M(\hat{G})\geqq M(F).$$

Taking the product of (4.4) for $i=1,\ldots,r$ we find that

$$((\tfrac{2}{7})^r M(F))^{|\mathfrak{X}|}\leqq Y_0^r.$$

Since (2.6) holds with $p>(\frac{7}{2})^2$, we have $M(F)>(\frac{7}{2})^r$ when $r>r_2(p)$, and then

$$|\mathfrak{X}|\leqq\frac{r\log Y_0}{\log M(F)-r\log(\frac{7}{2})}. \tag{4.5}$$

With A_1 given by (3.4) with (3.7), we have

$$A_1=(1/a^2)(\log M(F)+\tfrac{1}{2}r),$$

so that with (3.1) and (3.6) we obtain

$$\log Y_0=\frac{r}{r-\lambda}\left(\log M(F)+\log(2r^{\frac{1}{2}})+\frac{\log 2}{r}\right)+\frac{\lambda}{r-\lambda}\left(\log 4+\frac{1}{a^2}\left(\log M(F)+\frac{r}{2}\right)\right).$$

Since λ with (3.3), (3.7) is $O(r^{\frac{1}{2}})$, we see that

$$\log Y_0=(1+O(r^{-\frac{1}{2}}))\log M(F)+O(r^{\frac{1}{2}}),$$

with the constants implicit in O depending only on our choice of a, b in (3.7). Observing that $\log M(F)>r\log(\frac{7}{2})$ we may conclude that

$$\log Y_0<\left(1+\frac{\varepsilon}{2}\right)\log M(F)$$

when $r>r_3(\varepsilon)$, which in conjunction with (4.5) yields

$$|\mathfrak{X}|<r\left(1+\frac{\varepsilon}{2}\right)\frac{\log M(F)}{\log M(F)-r\log(\frac{7}{2})}.$$

Thus for $r>r_2(p,\varepsilon)$, by (2.6),

$$\begin{aligned}|\mathfrak{X}|&<r\left(1+\frac{\varepsilon}{2}\right)\frac{1}{1-(r\log(\frac{7}{2}))/\log M(F)}\\&<r\left(1+\frac{\varepsilon}{2}\right)\frac{1}{1-(\log(\frac{7}{2}))/(\frac{1}{2}\log p-((\log r)/(r-2)))}\\&<r(1+\varepsilon)\frac{1}{1-(\log(\frac{7}{2}))/(\frac{1}{2}\log p)}.\end{aligned}$$

V. The number of solutions of $|F(x)|=1$

The solutions of $|F(\boldsymbol{x})|=1$ with $0<y\leqq Y_0$ lie in $\mathfrak{X}$, or they are among possible $\boldsymbol{x}(1),\ldots,\boldsymbol{x}(r)$. Considering also $\boldsymbol{x}=(1,0)$, we see that the number of solutions with $|y|\leqq Y_0$ is $\leqq|\mathfrak{X}|+r+1$. (Recall that we regard $\boldsymbol{x}$, $-\boldsymbol{x}$ as the same solution.) Since it was shown in Sect. III that the number of large solutions is $\leqq 3r$ when

r is large, we get an upper bound

$$|\mathfrak{X}|+4r+1.$$

When $p>(\frac{7}{2})^2$ and r is large, Lemma 6 yields the bound

$$N_r(p)<r(1+2\varepsilon)\left(\frac{1}{1-(2\log(\frac{7}{2}))/\log p}+4\right);$$

recall that $N_r(p)$ was an upper bound for the number of solutions when (2.4) holds. Together with (2.3) this gives

$$N_r<r(1+2\varepsilon)\,C(p)$$

where

$$C(p)=(p+1)\left(\frac{1}{1-(2\log(\frac{7}{2}))/\log p}+4\right).$$

With $p=19$ we get $C(19)=214.16\ldots$. Therefore $N_r<215r$ when r is large.

VI. The general Thue equation

We have shown that $N_r<215r$ when r is large. Since N_r is bounded for any fixed value of r, either by Evertse's work or by the method of this paper, Theorem 1 is true when $h=1$. To deduce the general case it will therefore suffice to show that

$$N_{r,h}\leqq r^t\,N_r, \tag{6.1}$$

where $N_{r,h}$ is the maximum number of primitive solutions of (1.2) and t is the number of distinct prime factors of h.

The idea of the proof, by induction on t, was partly inspired by Mahler's paper [7].

By splitting x/y according to congruence classes modulo high powers of a prime p, we can transform $F(x, y)$ into new forms $F_i(x', y')$ (say one form per congruence class), and now F_i itself will be divisible by a power of p; since h is also so divisible, we can reduce the power of p appearing in h, and repeat this process until p does not divide h anymore. The only thing which requires some care is to show that the number of forms F_i needed in this process stays bounded independent of p. An explicit construction of the forms F_i is certainly possible but it is rather complicated to describe, and it is easier, and conceptually simpler, to show only that the forms F_i exist by working in the p-adic field $\mathbf{Q}_p$ and in a completion Ω_p of an algebraic closure of $\mathbf{Q}_p$. We denote by $|\ |_p$ the usual absolute value in $\mathbf{Q}_p$, together with an extension to Ω_p; $\mathbf{Z}_p$ is the ring of integers in $\mathbf{Q}_p$, R_p is the ring of elements $|\zeta|_p\leqq 1$ in Ω_p.

For every $g\in\mathbf{Q}_p[x]$ let us define the norm

$$\|g\|_p=\sup_{R_p}|g(\zeta)|_p$$

and let us also define $\mu(g)$ by

$$p^{-\mu(g)}=(\sup_{\mathbf{Z}_p}|g(x)|_p)/\|g\|_p\,.$$

For every $\bar{u}\in\mathbf{Z}_p/p^{\mu(g)}\,\mathbf{Z}_p\subseteq R_p/p^{\mu(g)}\,R_p$, let $D(\bar{u})$ and $\Delta(\bar{u})$ denote the disks in $\mathbf{Z}_p$ and R_p having image $\bar{u}$ by the corresponding canonical homomorphism.

Our simple but basic result is as follows.

Lemma 7. *Let $f \in \mathbf{Z}_p[x]$ have norm $\|f\|_p = 1$ and let $\mu > 0$. Then we can find a finite set of elements A_σ, B_σ in $\mathbf{Z}_p$, $A_\sigma \neq 0$, $\sigma = 1, \ldots, s$ such that the polynomials*

$$f_\sigma(z) = p^{-\mu} f(A_\sigma z + B_\sigma)$$

have the following properties

(a) *the set of points $x \in \mathbf{Z}_p$ such that $|f(x)|_p \leqq p^{-\mu}$ is the disjoint union, for $\sigma = 1, \ldots, s$, of the disks $A_\sigma \mathbf{Z}_p + B_\sigma$;*

(b) *$f_\sigma(z) \in \mathbf{Z}_p[z]$ for $\sigma = 1, \ldots, s$;*

(c) *$s \leqq \#(\text{roots of } f \text{ in } R_p)$.*

Proof. We work first in R_p. Let

$$U = \{\zeta \in R_p, |f(\zeta)|_p \leqq p^{-\mu}\};$$

then U is open and closed. Let $\Delta(a_i, b_i) = a_i R_p + b_i$, $a_i \neq 0$, $a_i, b_i \in R_p$ be a maximal disk in R_p such that $|f(\zeta)|_p \leqq p^{-\mu}$ there. Since $\|f\|_p = 1$, it follows that $\Delta(a_i, b_i)$ is properly contained in R_p and moreover f has a root in $\Delta(a_i, b_i)$. The set U is the disjoint union of the $\Delta(a_i, b_i)$, and the number of disks $\Delta(a_i, b_i)$ does not exceed the number of roots of f in R_p. Let $D(A_i, B_i) = A_i \mathbf{Z}_p + B_i$, $A_i, B_i \in \mathbf{Z}_p$, $A_i \neq 0$, be the disk $D(A_i, B_i) = \Delta(a_i, b_i) \cap \mathbf{Z}_p$ (if not empty). Let

$$f_i(z) = p^{-\mu} f(A_i z + B_i).$$

If $\mu(f_i) = 0$ then since $f_i(z) \in \mathbf{Z}_p$ for $z \in \mathbf{Z}_p$ we deduce that $f_i \in \mathbf{Z}_p[x]$, and we take f_i as one of the f_σ's for Lemma 7. If instead $\mu(f_i) > 0$, each disk $\Delta(\bar{u})$ contains at least one root of f_i. Thus we first refine $\Delta(a_i, b_i)$ by writing it as the union of the disks $a_i \Delta(\bar{u}) + b_i$, and then repeat the procedure. It is now clear that

(i) the number of disks at any given stage does not exceed the number of roots of f in R_p;

(ii) if $\mu(f_i) > 0$ at a certain stage, the number of disks at the next stage increases by at least $p^{\mu(f_i)} - 1 \geqq 1$.

By (i) and (ii) this process must end and we remain with a disjoint decomposition

$$U = \cup \Delta(a_\sigma, b_\sigma)$$

with the following properties:

(1) $|f(\zeta)|_p \leqq p^{-\mu}$ in each $\Delta(a_\sigma, b_\sigma)$;

(2) if $D(A_\sigma, B_\sigma) = \Delta(a_\sigma, b_\sigma) \cap \mathbf{Z}_p$ (if not empty) then (a) of Lemma 7 holds;

(3) the number of $\Delta(a_\sigma, b_\sigma)$'s does not exceed the number of roots of $f = 0$ in R_p, and hence (c) of Lemma 7 holds;

(4) the polynomials $f_\sigma(z) = p^{-\mu} f(A_\sigma z + B_\sigma)$ all have $\mu(f_\sigma) = 0$, hence (b) of Lemma 7 holds.

The proof of Lemma 7 is complete.

We are now in position to prove our Theorem. Let

$$|F(x, y)| = p_1^{b_1} \ldots p_t^{b_t} \tag{6.2}$$

be a Thue equation. We perform the construction of Lemma 7 for the polynomial

$$f(x)=F(x, 1)$$

with $p=p_1$, $\mu=b_1$, and obtain

$$f_\sigma(z)=p_1^{-b_1} f(A_\sigma z+B_\sigma),$$

where we may further suppose that $B_\sigma\in\mathbf{Z}$, $A_\sigma=p^{l_\sigma}$ for some $l_\sigma\geqq 0$. Let $F_\sigma(x', y')$ be the binary form

$$F_\sigma(x', y')=p_1^{-b_1} F(A_\sigma x'+B_\sigma y', y').$$

In the notation of Lemma 7, we have $F_\sigma(x', y')=(y')^r f_\sigma\left(\frac{x'}{y'}\right)$, hence Lemma 7, (b) shows that F_σ has again integral coefficients. By Lemma 7, (a) every solution (x, y) to (6.2) with $(y, p_1)=1$ can be written as $\frac{x}{y}=A_\sigma z+B_\sigma$ for some σ and $z\in\mathbf{Z}_{p_1}$. We have $A_\sigma=p^{l_\sigma}$, $B_\sigma\in\mathbf{Z}$ thus $z=\frac{x^*}{y^*}$ with $x^*\in\mathbf{Z}$ and $y^*=y$. This shows that

$$|F_\sigma(x^*, y^*)|=p_1^{-b_1}|F(x, y)|=p_2^{b_2}\ldots p_t^{b_t},$$

i.e. (x^*, y^*) is a primitive solution to $|F_\sigma|=h/p_1^{b_1}$. If instead $(y, p_1)>1$, we must have $(x, p_1)=1$ and we may perform the same construction with the polynomial $f(x)=F(1, x)$ and interchanging the role of x, y. It is now clear, by Lemma 7, (c), that the total number of binary forms needed to remove the factor $p_1^{b_1}$ is at most r.

The proof of (6.1) and of our theorem is completed by induction on t, once it has been established for $t=0$.

References

1. Bombieri, E., Mueller, J.: On effective measures of irrationality for $\sqrt[r]{\frac{a}{b}}$ and related numbers. J. Reine Angew. Math. **342**, 173–196 (1983)
2. Evertse, J.-H.: Upper bounds for the number of solutions of diophantine equations. Math. Centrum. Amsterdam, pp. 1–127 (1983)
3. Evertse, J.-H.: On the representation of integers by binary cubic forms of positive discriminant. Invent. Math. **73**, 117–138 (1983)
4. Hooley, C.: On the representations of numbers by binary cubic forms. Glasgow Math. J. **27**, 95–98 (1985)
5. Lewis, D., Mahler, K.: Representation of integers by binary forms. Acta Arith. **6**, 333–363 (1961)
6. Mahler, K.: An inequality for the discriminant of a polynomial. Michigan Math. J. **11**, 257–262 (1964)
7. Mahler, K.: On Thue's theorem. Austr. Nat. Univ. Res. Rep. 24 (1982). Math. Scand. **55**, 188–200 (1984)
8. Siegel, C.L.: Über einige Anwendungen diophantischer Approximationen. Abh. Preuss. Akad. Wiss. Phys. Math. Kl. (1929), N° 1. Gesammelte Abhandlungen, Bd. 1, pp. 209–266. Berlin Heidelberg New York: Springer-Verlag 1966
9. Silverman, J.H.: Representation of integers by binary forms and the rank of the Mordell-Weil group. Invent. Math. **74**, 218–292 (1983)

Oblatum 14-IV-1986

Invent. math. 88, 83–124 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Unipotent variations of mixed Hodge structure

Richard M. Hain[1,*] and Steven Zucker[2,**]

[1] Department of Mathematics, GN-50, University of Washington, Seattle, WA 98195, USA
[2] Department of Mathematics, The Johns Hopkins University, Baltimore, MD 21218, USA

Contents

1. Introduction and allegro: Statement of theorem and outline of proof 83
2. In the group ring 88
3. Mixed Hodge structures for path space cohomology 90
4. The canonical variations of mixed Hodge structure 95
5. Classifying spaces and higher Albanese manifolds 100
6. Behavior at infinity 110
7. Monodromy in a unipotent variation 117
8. Rigidity 121
References 124

§ 1. Introduction and allegro: Statement of theorem and outline of proof

Let X be a smooth algebraic variety over the complex numbers $\mathbb{C}$, or more generally, a Zariski-open subset of a compact Kähler manifold. Let $\mathbb{V}$ be a local system on X underlying a *graded-polarizable variation of mixed Hodge structure*, defined over some real field (or ring of integers) $\mathbb{F}$. We recall (see [23], [24]) that the latter consists of:

(1.1) i) a flat $\mathbb{F}$-structure on $\mathbb{V}$,

ii) an increasing (*weight*) filtration W of $\mathbb{V}$ by local subsystems $\{\mathbb{W}_k\}$ defined over $\mathbb{F}$,[3]

iii) a decreasing (*Hodge*) filtration F of $\mathscr{V} := \mathscr{O}_X(\mathbb{V})$ by sub-bundles (i.e., locally free sheaves) $\{\mathscr{F}^p\}$, such that

$$d\mathscr{F}^p \subset \Omega^1_X \otimes \mathscr{F}^{p-1} \quad (\textit{horizontality}),$$

* Supported in part by the National Science Foundation, through grants MCS-8108814(A04) and DMS-8401175.

** Supported in part by the National Science Foundation, through grants MCS 81-01650 and DMS-8501005

[3] When $\mathbb{F}$ is a ring of integers, one usually imposes the integral structure only on $\mathbb{V}$, though not on the weight filtration

iv) for each $k \in \mathbb{Z}$, the filtration induced by F on Gr_k^W defines a polarizable variation of Hodge structure of weight k.

Such arise, for example, from systems of cohomology $\mathbb{V} = R^q f_* \mathbb{C}$, where $f: Z \to X$ is a topologically locally constant family of varieties (see [23: § 5] for a more precise statement).

Let $x \in X$. As is well-known, giving $\mathbb{V}$ (as a local system) is equivalent to specifying its monodromy representation

$$\rho: \pi_1(X, x) \to \mathrm{Aut}(V_x), \tag{1.2}$$

defined over $\mathbb{F}$; here V_x is the fiber of $\mathscr{V}$ at x, filtered by W. Since the weight filtration is flat, the image of ρ consists of linear transformations that are compatible with W (though not necessarily strictly). In other words, the image of ρ lies in the subgroup of $\mathrm{Aut}(V_x)$ that one would denote $W_0 \mathrm{Aut}(V_x)$, whose Lie algebra, $W_0 \mathrm{End}(V_x)$, is given the mixed Hodge structure induced by that of V_x by linear algebra. For universal reasons, ρ extends to an algebra homomorphism of the group ring:

$$\bar{\rho}: \mathbb{C}\pi_1(X, x) \to W_0 \mathrm{End}(V_x). \tag{1.3}$$

It is not difficult to see that the following four statements are equivalent:

(1.4) i) the representation ρ is unipotent,

ii) $\bar{\rho}$ factors through a quotient $\mathbb{C}\pi_1(X, x)/J^{r+1}$, where J is the kernel of the usual augmentation of the group ring,

iii) $\bar{\rho}(J) \subset W_{-1} \mathrm{End}(V_x)$,

iv) the variations of Hodge structure of the pure quotients (1.1, iv) are constant.

When the monodromy satisfies the above conditions, we will refer to the variation of mixed Hodge structure as a *unipotent* one. It then follows that ρ factors through a quotient $\pi_1(X, x)/\Gamma^{(r+1)}$, where $\Gamma^{(\cdot)}$ denotes the descending central series of $\pi_1(X, x)$. A fundamental example of a unipotent variation is provided by the mixed Hodge structures of $\mathbb{C}\pi_1(X, x)/J^{r+1}$, as x ranges over X, as defined in [11] - cf. also [16] - which we call the *tautological* variation. It is our primary goal to give the classification of unipotent variations of mixed Hodge structure, which was conjectured by Deligne.

In order to state our result, we must first recall and generalize a notion from [23]. The variation of mixed Hodge structure is considered *good* if it satisfies the following conditions at infinity (cf. [23: (3.13); App.] when X is a curve): there exists a compactification $\bar{X}$ of X, for which $\Sigma := \bar{X} - X$ is a divisor with normal crossings, such that

(1.5) i) the Hodge filtration bundles $\mathscr{F}^p$ extend over $\bar{X}$ to sub-bundles $\tilde{\mathscr{F}}^p$ of the canonical extension $\tilde{\mathscr{V}}$ of $\mathscr{V}$, such that they induce the corresponding thing for each pure subquotient $Gr_k^W \mathbb{V}$,

ii) for the nilpotent logarithm N_j of a local monodromy transformation about a component Σ_j of Σ, the weight filtration M of N relative to W exists.[4]

[4] It would seem that one wants to insist that for any non-empty intersection $\bigcap_{j \in I} \Sigma_j$, the same be true for all linear combinations $N = \sum_{j \in I} \lambda_j N_j$ with each $\lambda_j > 0$, and that the relative weight filtration depend only on I (cf. [3], [23: (3.18)]). However, this follows from (1.5, ii) by [30: § 4]

These conditions are known to be satisfied by geometrically-defined variations of mixed Hodge structure, at least when X is a curve (see [23: (5.7)]). For a unipotent variation, (1.5, ii) is equivalent to the condition $NW_k \subset W_{k-2}$, since the triviality of $Gr^W N$ forces M to equal W (see [23: (2.14)]).

The main result in this paper is:

(1.6) **Theorem.** *Fix any $x \in X$. Then the monodromy representation functor defines an equivalence of categories*

$$\left\{\begin{array}{l}\text{unipotent variations of}\\ \text{mixed Hodge structure satisfying}\\ \text{condition (1.5) at infinity, with}\\ \text{index of unipotency } \leqq r, \text{ defined over } \mathbb{F}\end{array}\right\} \to \left\{\begin{array}{l}\text{mixed Hodge theoretic}\\ \text{representations of}\\ \mathbb{C}\pi_1(X,x)/J^{r+1},\\ \text{defined over } \mathbb{F}.\end{array}\right\}$$

In order to dispel any mystery about the content of the above theorem, we give here the highlights of the proof; we show:

(1.7) a) For any good unipotent variation, the monodromy representation $\bar{\rho}$ (1.4, ii) is a morphism of mixed Hodge structures.

b) The tautological variation satisfies the good conditions at infinity. Then, given a mixed Hodge theoretic unipotent representation (i.e., V a finite dimensional vector space with mixed Hodge structure, and $\mathbb{C}\pi_1(X,x)/J^{r+1} \to \mathrm{End}(V)$ an algebra homomorphism that is also a morphism of mixed Hodge structures), there is a natural construction of a unipotent variation of mixed Hodge structure with fiber V at x, satisfying (1.5).

c) For unipotent variations of mixed Hodge structure satisfying (1.5), there is a rigidity theorem, which asserts that the variation is uniquely determined by its monodromy together with the Hodge filtration at one point (see [23: (4.20)] for the case where X is a curve, where unipotence is not assumed; cf. also [20: (7.24)]).

In short, the variation of mixed Hodge structure can be reconstituted from its monodromy (where the latter is viewed as a mixed Hodge theoretic representation). The correspondence of morphisms in the two categories is reduced to (1.7) via the use of "Hom" functors. As an interpretation of the theorem, we can say that in a unipotent variation of mixed Hodge structure on X, the mixed Hodge structure at the point x must be compatible with that of

$$\mathbb{C}\pi_1(X,x)^\wedge = \varprojlim_r \mathbb{C}\pi_1(X,x)/J^{r+1}$$

(associated to the choice of x as the basepoint of X), and, as x moves, the mixed Hodge structure is constrained to move in certain directions. We remark that there are obvious counterexamples to (1.6) if one neglects to impose (1.5) (see [23: (4.21)]), as (1.7(c)) breaks down.

We describe now the means by which the statements (1.7) are verified.

Let PX denote the space of piecewise-smooth paths (parametrized by $I = [0,1]$) in X, equipped with the compact-open topology. It becomes a (C^∞) differentiable space in the sense of Chen (see [5: (1.2)]) via the following designation: if N is a smooth manifold, a mapping $h: N \to PX$ is differentiable if and only if the associated mapping $\tilde{h}: N \times I \to X$ is piecewise smooth. The tautological mapping:

(1.8) $$p\colon PX \to X \times X \qquad p(\gamma) = (\gamma(0), \gamma(1))$$

defines the so-called *free path fibration*. The fiber $P_{x,y}$ over $(x, y) \in X \times X$ is just the set of paths from x to y; for $y = x$, we have that $P_{x,x}$ is the space of loops in X based at x – which we shall continue to denote this way, and not by the customary $\Omega_x X$ of topology, as Ω has a different meaning in analytic and algebraic geometry.

The story begins with the canonical identification

(1.9) $$\mathbb{Z}\pi_1(X, x) \cong H_0(P_{x,x}; \mathbb{Z}),$$

and likewise, of course, for field coefficients. Thus, we have

(1.10) $$\mathrm{Hom}_{\mathbb{Z}}(\mathbb{Z}\pi_1(X, x), \mathbb{Z}) \cong H^0(P_{x,x}; \mathbb{Z}).$$

Now, on any differentiable space, one can define a notion of differential forms (see [5: p. 835]). According to a theorem of Chen, we can (essentially) compute the 0-th cohomology group of the loop space $P_{x,x}$ with $\mathbb{C}$ coefficients by means of certain subcomplexes of its de Rham complex, the complexes of *iterated integrals* $\int_x A^\cdot$ (see our (3.18)) constructed from suitable differential graded algebras $A^\cdot$ that compute $H^\cdot(X, \mathbb{C})$. More precisely, we have that

(1.11) $$H^0(\textstyle\int_x A^\cdot)^* \cong \mathbb{C}\pi_1(X, x)^\wedge$$

(see [5: p. 852]).

The algebraic, or "combinatorial", data carried by the iterated integrals is contained in the so-called *reduced bar construction* on $A^\cdot$, denoted $\bar{B}(A^\cdot)_x$ (see our (3.4)):

(1.12) $$H^0(\bar{B}(A^\cdot)_x)^* \cong \mathbb{C}\pi_1(X, x)^\wedge.$$

From its definition, there is an obvious increasing filtration $\mathfrak{B}$ on $\bar{B}(A^\cdot)_x$, which satisfies:

(1.13) i) $Gr_r^{\mathfrak{B}} \bar{B}(A^\cdot)_x \cong \otimes^r (A^+/dA^0)[r]$ as graded vector spaces, where A^+ denotes the subalgebra of $A^\cdot$ consisting of elements of (strictly) positive degree,

ii) If one ignores the differential,

$$\bar{B}(A^\cdot)_x \cong \bigoplus_{r>0} Gr_r^{\mathfrak{B}} \bar{B}(A^\cdot)_x,$$

iii) For each $r > 0$, there is an isomorphism

(1.14) $$H^0(\bar{B}_r(A^\cdot)_x) \cong [\mathbb{C}\pi_1(X, x)/J^{r+1}]^*,$$

where we are writing $\bar{B}_r(A^\cdot)_x$ for $\mathfrak{B}_r \bar{B}(A^\cdot)_x$ ([5: p. 852], cf. [21]).

Assume from now on that $A^\cdot$, with weight filtration W and Hodge filtration F, is a mixed Hodge complex for $H^\cdot(X, \mathbb{C})$ in the sense of [8: (8.15)], so that (W, F) induce the mixed Hodge structure of $H^\cdot(X, \mathbb{C})$. In view of (1.13, ii), and by the properties of the differential of $\bar{B}(A^\cdot)_x$, these filtrations induce one on $\bar{B}(A^\cdot)_x$ in the usual way, such that each $\bar{B}_r(A^\cdot)_x$ becomes a mixed Hodge complex with respect to the filtrations $(W * \mathfrak{B}, F)$ [11: (3.2.1)]; it is, moreover,

filtered by $\mathfrak{B}$ (see [10] or [23: §6] for a discussion of filtered mixed Hodge complexes). By (1.13, iii) and duality, this defines a tower of mixed Hodge structures for $\{\mathbb{C}\pi_1(X,x)/J^{r+1}\}$.

The mixed Hodge structures associated to π_1 were originally given in terms of the so-called *Malčev Lie algebra* (cf. [16: (9.2)]). However, one can realize both the Malčev Lie algebra and the associated Malčev group inside the completed group ring $\mathbb{C}\pi_1(X,x)^\wedge$. With this said, (1.7(a)) will be seen in §7 to follow from the formula for the monodromy (more generally, parallel transport) in terms of iterated integrals ([12], cf. [5: p. 856]).

Next, let V be a vector space defined over $\mathbb{Q}$, and W an increasing filtration of V over $\mathbb{Q}$, and L a full lattice in V. Given a collection of Hodge numbers and a graded-polarization for V, there are classifying spaces for graded-polarized mixed Hodge structures, constructed in [2] or [25]. For evident reasons, these fiber over the product of the classifying spaces for Hodge structures (with the given data) on the $Gr_k^W V$'s. The fiber $U(V)$ is locally homogeneous under the unipotent group $W_{-1}GL(V)_{\mathbb{C}}$. It follows that a unipotent variation of mixed Hodge structure is classified by a horizontal mapping

$$\varphi\colon X \to U(V); \tag{1.15}$$

locally, such φ is given by

$$\tilde{\varphi}\colon X \to W_{-1}GL(V)_{\mathbb{C}}. \tag{1.16}$$

In case $V = J/J^{r+1}$, the interesting direct summand of $\mathbb{C}\pi_1/J^{r+1}$, we will see that for the tautological variations, the mapping $\tilde{\varphi}$ factors through the adjoint representation of the *Malčev group* $\hat{G}_{r-1} \subset \mathbb{C}\pi_1/J^r$:

$$\hat{G}_{r-1} \hookrightarrow \mathrm{Aut}(J/J^{r+1}).$$

Somewhat more canonically, we observe an analogous thing for the variations of mixed Hodge structure for $\{H_0(P_{x,y})\}$, with x fixed and y varying, – we call this the *canonical variation* with basepoint x – that (1.16) this time factors through the right-regular representation of $\hat{G}_r$ on $\mathbb{C}\pi_1/J^{r+1}$. These assertions are direct consequences of the variation of the bar complex $\bar{B}(A^\cdot)_{x,y}$ that is used for computing $H^0(P_{x,y};\mathbb{C})$ as x and y vary, which is in turn a reflection of the natural way in which the latter defines a local system on $X \times X$ (see our §3).

The resulting factorization of φ for the canonical variations with basepoint $x \in X$ will be denoted suggestively

$$\alpha_x^r\colon X \to \mathrm{Alb}_x^r(X) \tag{1.17}$$

and called the r-th *higher Albanese mappings*; they form an inverse system. (We remark that for $x, y \in X$, there is a canonical system of isomorphisms

$$\mathrm{Alb}_y^r(X) \cong \mathrm{Alb}_x^r(X)$$

compatible with (1.17).) In case $r=1$, $J/J^2 \cong H_1(X,\mathbb{C})$, and α_x^1 is the familiar Albanese mapping. A similar, but not identical, construction of higher Al-

baneses was given by Paršin [17] and studied by Hwang-Ma [14]. A mixed Hodge theoretic unipotent representation ρ is seen to induce a mapping

$$C_\rho\colon \mathrm{Alb}_x^r(X)\to U(V_x);$$

the composite, $C_\rho\circ\alpha_x^r$, defines a unipotent variation. It satisfies the conditions (1.5) at infinity automatically, once we know that the canonical variation does. The latter is verified by the methods of asymptotic Hodge theory (§6; cf. [23: (5.7)]).

It then remains to see that if we start with a unipotent variation of mixed Hodge structure, and apply the preceding discussion to the monodromy representation (recall (1.7, a)), that we have reconstructed our given variation. If X is a curve, we may appeal directly to the rigidity theorem of [23: (4.20)]. Since algebraic varieties contain plenty of curves, that result implies rigidity whenever X is algebraic. On the other hand, a general Kähler manifold may have no curves at all. In §8, we will derive a rigidity theorem for Zariski-open subsets of compact Kähler manifolds by extending the construction in [23: §4] (for general graded-polarizable variations of mixed Hodge structure over curves) to unipotent variations over higher dimensional bases. Because we are working with a very special case, we can circumvent the difficulties one encounters in general in trying to do "mixed Hodge theory with degenerating coefficients" when $\dim X>1$.

We wish to express our gratitude to P. Deligne for conversations and correspondence on unipotent variations, which included an initiation to the problem.

§2. In the group ring

The material in this section is well-known. For a reference, see [19: A.3].

Let G be a group, R a commutative ring (with unit). We denote by RG the group algebra of G over R. Its augmentation

$$\varepsilon\colon RG\to R \tag{2.1}$$

is the homomorphism sending each $g\in G$ to $1\in R$. Put

$$J=\ker\varepsilon. \tag{2.2}$$

One observes that the homomorphism

$$\iota\colon R\to RG \qquad \iota(r)=r\cdot e, \tag{2.3}$$

where e is the identity element of G, defines an R-module splitting

$$RG\cong R\oplus J. \tag{2.4}$$

The R-algebra homomorphism

$$\Delta\colon RG\to RG\otimes_R RG, \tag{2.5}$$

determined by the formula $\Delta(g)=g\otimes g$, defines the coproduct of a Hopf algebra structure on RG. Let $RG^\wedge$ denote the J-adic completion of RG, and $\hat{J}\subset RG^\wedge$ the closed ideal generated by the image of J. We have, again,

$$RG^\wedge \cong R\oplus\hat{J}. \tag{2.6}$$

The coproduct (2.5) induces

$$\hat{\Delta}\colon RG^\wedge \to RG^\wedge \hat{\otimes}_R RG^\wedge, \tag{2.7}$$

imparting $RG^\wedge$ with the structure of a complete Hopf algebra.

The set of *grouplike* elements of $RG^\wedge$ is defined to be

$$\hat{G}=\{x\in RG^\wedge\colon \varepsilon(x)=1, \Delta(x)=x\hat{\otimes}x\}\subset 1+\hat{J}. \tag{2.8}$$

The primitive elements are, by definition,

$$\mathfrak{g}=\{x\in\hat{J}\colon \Delta(x)=1\hat{\otimes}x+x\hat{\otimes}1\}. \tag{2.9}$$

Assume henceforth that R is a field of characteristic zero. One defines, by the usual power series, the mutually inverse exponential and logarithm functions

$$\hat{J}\underset{\log}{\overset{\exp}{\rightleftarrows}}1+\hat{J}. \tag{2.10}$$

Then:

(2.11) **Proposition.** i) *$\hat{G}$ is a multiplicative subgroup of $RG^\wedge$,*

ii) *$\mathfrak{g}$ is a Lie algebra, with commutator as bracket operation,*

iii) *The restriction of* (2.10) *sets up a bijection*

$$\mathfrak{g}\underset{\log}{\overset{\exp}{\rightleftarrows}}\hat{G}.$$

(2.12) *Definition.* The natural homomorphism

$$\theta\colon G\to\hat{G}$$

is called the *Malčev completion* of G over R; $\mathfrak{g}$ is its *Malčev Lie algebra.*

One defines filtrations of $\hat{G}$ and $\mathfrak{g}$ respectively by

$$\hat{G}^r=\hat{G}\cap(1+\hat{J}^r),\qquad \mathfrak{g}^r=\mathfrak{g}\cap\hat{J}^r. \tag{2.13}$$

We then put

$$\hat{G}_r=\hat{G}/\hat{G}^{r+1},\qquad \mathfrak{g}_r=\mathfrak{g}/\mathfrak{g}^{r+1}. \tag{2.14}$$

Let

$$G=\Gamma^{(1)}\supset\Gamma^{(2)}\supset\dots \tag{2.15}$$

denote the lower (i.e., descending) central series of G, and put

$$G_{(r)}=G/\Gamma^{(r+1)}. \tag{2.16}$$

(2.17) **Proposition.** i) ([19: (A.2)], [31]) *$\{\hat{G}^r\}$ is the lower central series of $\hat{G}$.*

ii) ([19: (A.2)], [31]) *The canonical mapping $RG^\wedge \to RG_{(r)}{}^\wedge$ induces a filtered isomorphism*

$$\hat{G}_r = (G_{(r)})^\wedge.$$

iii) *Letting $D^l = \{g \in G\colon (g-1) \in J^l\} = \ker\{G \to \hat{G}_{l-1}\}$, we have that $\{D^l\}$ is a decreasing sequence of normal subgroups with $[D^l, D^m] \subset D^{l+m}$, $\theta(D^l) \subset \hat{G}^l$ and G/D^l is torsion-free for all l* (see [18: p. 85]); *moreover, $D^l = \sqrt{\Gamma^{(l)}}$* (see [18: p. 474]), *and thus $D^l/\Gamma^{(l)}$ is a torsion group.*

iv) *If G is finitely generated, then so is each $\Gamma^{(l)}$ and D^l* (see [18: p. 93]), *and*

$$(\Gamma^{(l)}/\Gamma^{(l+1)}) \otimes_{\mathbb{Z}} \mathbb{R} \cong (D^l/D^{l+1}) \otimes_{\mathbb{Z}} \mathbb{R} \cong \hat{G}^l/\hat{G}^{l+1} \cong \mathfrak{g}^l/\mathfrak{g}^{l+1}$$

(with $R = \mathbb{R}$ for the group ring).

(2.18) *Remarks.* i) For a torsion-free *nilpotent* group G, we have that $G \to \hat{G}$ is injective.

ii) If G is finitely-generated and nilpotent, $\hat{G}$ is a (finite-dimensional) nilpotent Lie group.

§ 3. Mixed Hodge structures for path space homology

We recall the free path fibration $p\colon PX \to X \times X$ of (1.8). It is a fibration in the sense of Hurewicz, but is locally trivial only in the homotopy category, in a sense that we now make explicit. Fix a point $(x_0, y_0) \in X \times X$. For any $(x, y) \in X \times X$ and choice of paths τ_x from x_0 to x, τ_y from y_0 to y, the formula

$$\gamma \longmapsto \tau_x \gamma \tau_y^{-1}$$

– our convention is that path multiplication is carried out from left to right; read the above as: τ_x followed by γ... – defines an embedding

$$\tau\colon P_{x,y} \hookrightarrow P_{x_0,y_0} \tag{3.1}$$

that is a homotopy equivalence. Moreover, if τ_x lies in a contractible neighborhood U of x_0, and τ_y lies within a contractible neighborhood V of y_0, then the homotopy class of (3.1) is independent of the choice of τ's. It follows that we may identify canonically

$$H_0(P_{x,y}; \mathbb{Z}) \cong H_0(P_{x_0,y_0}; \mathbb{Z}) \tag{3.2}$$

for all $x \in U$, $y \in V$. That is, we have a local system

$$\{H_0(P_{x,y}; \mathbb{Z}\}_{(x,y) \in X \times X}. \tag{3.3}$$

We recall that in the case of a loop space ($y = x$),

$$H_0(P_{x,x}; \mathbb{Z}) \cong \mathbb{Z}\pi_1(X, x)$$

is filtered by the powers of its augmentation ideal $J_{x,x}$. If we restrict the path fibration p to the diagonal (i.e., to loops only), we see immediately from (3.2) that $\{J^l_{x,x}\}_{x \in X}$ defines a locally constant sheaf for each $l \geqq 0$. We can extend these to all of $X \times X$ by parallel transport: $J^l_{x,y} \subset H_0(P_{x,y}; \mathbb{Z})$ is the image of $J^l_{x,x}$

under any identification $H_0(P_{x,y}) \cong H_0(P_{x,x})$ [or $H_0(P_{y,y})$] of type (3.2). Equivalently, it is the filtration induced by the canonical left $H_0(P_{x,x})$-module [or right $H_0(P_{y,y})$-module] structure of $H_0(P_{x,y})$, defined via path multiplication. We write J^l generically for $J^l_{x,y}$.

The duals of the truncations $H_0(P_{x,y};\mathbb{C})/J^{r+1}$ are computable by complexes of iterated integrals of length at most r, as was mentioned in (1.13, iii). These, in turn, are expressible in terms of the bar construction, as follows.

Let $A = \bigoplus_{p \geq 0} A^p$ be a differential graded algebra over the field $\mathbb{F}$, i.e., a complex of $\mathbb{F}$-vector spaces with associative multiplication, such that

$$d(ab) = (da)b + (\tau a)(db),$$

where τa denotes $(-1)^p a$ for $a \in A^p$. Let $M^\cdot$ be a complex of vector spaces over $\mathbb{F}$ that has both the structure of a right and left module over $A^\cdot$ (such that the Leibniz rule holds).

(3.4) *Definition* (cf. [4]). The *reduced circular bar construction* on $(A^\cdot, M^\cdot)$ is the complex $\bar{B}(A^\cdot; M^\cdot)$, whose underlying graded vector space is a quotient of

$$M^\cdot \otimes_{\mathbb{F}} [\otimes^\cdot (A^+[1])], \tag{3.5}$$

where A^+ denotes the elements in $A^\cdot$ of strictly positive degree, specified as follows. As differential,

$$d = (\tau \otimes d_B) + (d_M \otimes 1) + d_C,$$

where d_B is the bar differential: if one writes

$$[a_1|\ldots|a_r] \quad \text{for } a_1 \otimes \ldots \otimes a_r \in \otimes^r A^+,$$

$$\begin{aligned} d_B[a_1|\ldots|a_r] = &\sum_{1 \leq j \leq r} (-1)^j [\tau a_1|\ldots|\tau a_{j-1}|da_j|a_{j+1}|\ldots|a_r] \\ &+ \sum_{1 \leq j < r} (-1)^{j+1} [\tau a_1|\ldots|\tau a_{j-1}|\tau a_j \cdot a_{j+1}|a_{j+2}|\ldots], \end{aligned} \tag{3.6}$$

where τ is defined in the expected way on $M^\cdot$; d_M is the differential of $M^\cdot$; and

$$d_C(m \otimes [a_1|\ldots|a_r]) = -(\tau m) \cdot a_1 \otimes [a_2|\ldots|a_r] + (-1)^v a_r \cdot m \otimes [a_1|\ldots|a_{r-1}], \tag{3.7}$$

where $v = (\deg a_r + 1)(\deg m + \deg[a_1|\ldots|a_{r-1}])$. The relations are given by:

(3.8) i) $m \otimes [dg|a_1|\ldots|a_r] = m \otimes [ga_1|a_2|\ldots|a_r] - (m \cdot g) \otimes [a_1|\ldots|a_r]$

ii) $m \otimes [a_1|\ldots|a_{i-1}|dg|a_i\ldots] = m \otimes [a_1|\ldots|a_{i-1}|ga_i|a_{i+1}|\ldots]$
$\quad - m \otimes [a_1|\ldots|ga_{i-1}|a_i|\ldots] \quad (i > 1)$

iii) $m \otimes [a_1|\ldots|a_r|dg] = -m \otimes [a_1|\ldots|a_{r-1}|ga_r] + (g \cdot m) \otimes [a_1|\ldots|a_r]$,

iv) $m \otimes [dg] = (g \cdot m) \otimes 1 - (m \cdot g) \otimes 1$

whenever $m \in M$, $g \in A^0$ and $a_1, \ldots, a_r \in A^+$.

Let $\mathfrak{B} = \{\bar{B}_r(A^\cdot; M^\cdot)\}$ be the filtration of $\bar{B}(A^\cdot; M^\cdot)$ by increasing A^+-tensor length. One sees from (3.8) that it is well-defined; it is a filtration by subcomplexes, as follows from the definition of d. It is easy to see that

$$Gr_r^{\mathfrak{B}} \bar{B}(A^\cdot; M^\cdot) \cong M^\cdot \otimes \{\otimes^r (A^+/dA^0)\}[r]. \tag{3.9}$$

The associated spectral sequence (Eilenberg-Moore) begins

$$E_1^{-r,\cdot} \cong H^\cdot(M^\cdot) \otimes \{\otimes^r H^\cdot(A^+/dA^0)\}[r], \tag{3.10}$$

which identifies

$$(E_1, d_1) \cong \bar{B}(H^\cdot(A^\cdot); H^\cdot(M^\cdot)).$$

Let A' be another differential graded algebra, and M' a two-sided A'-module. Suppose that we are given homomorphisms

$$\varphi\colon A \to A', \quad \psi\colon M \to M' \tag{3.11}$$

compatible with the module structures. It is easy to check that these determine morphisms of complexes

$$\bar{B}(\varphi, \psi)\colon \bar{B}(A; M) \to \bar{B}(A'; M'), \tag{3.12}$$

$$\bar{B}_r(\varphi, \psi)\colon \bar{B}_r(A; M) \to \bar{B}_r(A'; M'). \tag{3.13}$$

(3.14) **Proposition.** *If φ and ψ are quasi-isomorphisms* (i.e., *induce isomorphisms on cohomology*), *then* (3.13) *is a $\mathfrak{B}$-filtered quasi-isomorphism for all r.*

Proof. In view of (3.9) and the Künneth theorem, we have quasi-isomorphisms

$$Gr_l^{\mathfrak{B}} B(A; M) \to Gr_l^{\mathfrak{B}} B(A'; M').$$

The assertion now follows because $\mathfrak{B}$ defines a finite filtration of each $\bar{B}_r$. □

Assume now that A and M underlie mixed Hodge complexes [8: (8.1.5)], at least over $\mathbb{R}$, i.e., have filtrations W and F inducing mixed Hodge structures on their cohomology. If we assume that $H^0(A^\cdot) = \mathbb{C}$ and that the weight filtration of $A^\cdot$ is bounded below (because of (3.8)), there are induced filtrations W and F on $\bar{B}(A; M)$, termwise via tensor algebra (see [7: (1.1.12)]). We then have:

(3.15) **Proposition** [11: (3.4.1)]. *If M and A underlie mixed Hodge complexes over $\mathbb{F}$, then $\bar{B}_r(A; M)$, with filtrations $(W * \mathfrak{B}, F)$, is a mixed Hodge complex over $\mathbb{F}$, filtered by $\mathfrak{B}$.*

Proof (outline). We have

$$\begin{aligned} Gr_k^{W*\mathfrak{B}} \bar{B}(A; M) &\cong \bigoplus_{i+j=k} Gr_i^W Gr_j^{\mathfrak{B}} \bar{B}(A; M) \\ &\cong \bigoplus_{i+j=k} Gr_i^W (M \otimes (\otimes^j \bar{A}^\cdot))[j], \end{aligned} \tag{3.16}$$

where $\bar{A}^\cdot = A^+/dA^0$. One must observe that $\bar{A}^\cdot$ is a mixed Hodge complex. Then each summand in (3.16) is seen to be a Hodge complex (see [8: (8.1.1)]) of weight k: i from the Gr_i^W plus j from the shift, yielding the first assertion. It is filtered by $\mathfrak{B}$ (in the sense of [10]) by (3.16) and [23: (6.8)]. □

We will make use of the following bifiltered version of (3.14), which is easy to verify.

(3.17) **Proposition.** *If the homomorphisms* (3.11) *are compatible with the two filtrations W and F, and are bifiltered quasi-isomorphisms, then* (3.13) *is a bifiltered quasi-isomorphism with respect to $(W * \mathfrak{B}, F)$.*

For any morphism of smooth manifolds $f: N \to X \times X$, which we write as (f_0, f_1), we make $E_N^\cdot$, the C^∞ de Rham complex of N, into a right $E_X^\cdot$ module via f_0^* (i.e., for $\alpha \in E_X^\cdot$, $m \in E_N^\cdot$, $m \cdot \alpha = m \wedge f_0^* \alpha$) and a left module via f_1^*. Let $f^*PX \to N$ be the pullback of $PX \to X \times X$ along f. Combining [5: p. 867] and [4], we have:

(3.18) **Theorem.** *If X is connected and $A^\cdot$ is a differential graded subalgebra of $E_X^\cdot$ such that*

$$H^1(A^\cdot) \to H^1(E_X^\cdot) \quad \textit{is injective,}$$

then $\bar{B}(A^\cdot; E_N^\cdot)$ is canonically isomorphic to the complex of iterated integrals on f^*PX constructed from $A^\cdot$ and $E_N^\cdot$.

(3.19) *Examples.* i) Take N to be a point $\{p\}$, and let $f(p) = (x, y)$; then $E_N^\cdot \cong \mathbb{C}$. We write $\bar{B}(A^\cdot)_{x,y}$ for the resulting bar construction $\bar{B}(A^\cdot; \mathbb{C})$, related to $P_{x,y}$. Note that $\bar{B}_0(A^\cdot; \mathbb{C}) \cong \mathbb{C}$ splits off from $\bar{B}_r(A^\cdot; \mathbb{C})$ $(r > 0)$ only when $y = x$, as $[dg] = g(y) - g(x)$.

ii) Take $N = X$, $f(y) = (x_0, y)$. Then $\bar{B}(A^\cdot; E_N^\cdot)$ is related to the space of paths emanating from x_0.

iii) Another obvious choice is $N = X \times X$, $f = 1_N$. This gives "universal" iterated integrals when $A^\cdot = E_X^\cdot$.

Let now $X = V - D$, where V is a compact Kähler manifold and D is a divisor with normal crossings. As is well-known, the following algebras $A^\cdot$ can be used to compute $H^\cdot(X, \mathbb{C})$ and its mixed Hodge structure.

(3.20) i) The C^∞ logarithmic de Rham complex: $E_V^\cdot(\log D)$,

ii) The holomorphic logarithmic Cech (double) complex for a Stein covering $\mathfrak{U}$ of X:

$$C^\cdot(\mathfrak{U}, \Omega_V^\cdot(\log D)).$$

These have the customary weight filtration (by number of poles) and Hodge filtration F (by holomorphic degree), and are seen to be bifiltered quasi-isomorphic through their inclusion in

iii) $C^\cdot(\mathfrak{U}, E_V^\cdot(\log D))$.

We can apply (3.15) to (1.14) and (3.19, i) to conclude:

(3.21) **Proposition.** i) *The filtrations $W * \mathfrak{B}$ and F of $\bar{B}_r(A^\cdot)_{x,y}$ induce a $\mathfrak{B}$-filtered mixed Hodge structure on $H^i(\bar{B}_r(A^\cdot)_{x,y})$ for all i.*

ii) *In case $i = 0$, these define by duality,[5] a mixed Hodge structure on $H_0(P_{x,y}, \mathbb{C})/J^{r+1}$; in fact, we have a compatible system of mixed Hodge structures for all J^l/J^m.*

(3.22) *Remarks.* i) It will simplify the notation later to refer to the mixed Hodge complexes of (3.21, i) as

$$\bar{B}(R\Gamma \mathbb{C}_X)_{x,y},$$

with the understanding that we must take appropriate multiplicative models

[5] See (1.14) when $y = x$, and §4 for the general case

for $R\Gamma\mathbb{C}_X$ (see [11: (5.6)]), esp. (3.20, ii), in order to see the filtrations, not to mention to make the constructions in Sects. 4–6.

ii) The above mixed Hodge structures are defined over $\mathbb{Q}$, which one can see on the level of mixed Hodge complexes via the filtered quasi-isomorphism [7: (3.1.8)]

$$(Rj_*\mathbb{Q}_X\otimes\mathbb{C},\tau)\approx(\Omega_V^\cdot(\log D),W),$$

where $j\colon X\to V$ denotes the inclusion, with the help of (3.14), as $\mathfrak{B}$ is clearly defined over $\mathbb{Q}$. In $H_0(P_{x,y};\mathbb{C})/J^{r+1}$, the resulting $\mathbb{Q}$-form is just $H_0(P_{x,y};\mathbb{Q})/J^{r+1}$. Moreover, we let $H_0(P_{x,y};\mathbb{Z})$ determine the integer lattice of a $\mathbb{Z}$-structure.

iii) From (3.9), we deduce that the mixed Hodge structure on J^l/J^{l+1} is independent of x and y. For example, $J/J^2\cong H_1(X,\mathbb{C})$; more generally, J^l/J^{l+1} is isomorphic to a quotient of $\otimes^l H_1(X,\mathbb{C})$.

iv) By using algebras $A^\cdot$, quasi-isomorphic to $E_X^\cdot$, with $A^0=\mathbb{C}$, one sees from (3.14) that the quasi-isomorphism class of $\bar{B}(R\Gamma\mathbb{C}_X)_{x,y}$ is actually independent of x and y.

When $y=x$, there is further structure in

$$\varprojlim_r H_0(P_{x,x};\mathbb{C})/J^{r+1}\cong\mathbb{C}\pi_1(X,x)\hat{},$$

namely that of a complete Hopf algebra. It is dual to the Hopf algebra structure induced from the following operations on $\bar{B}(A^\cdot)=\bar{B}(A^\cdot)_{x,x}$, when $A^\cdot=E_X^\cdot$ (graded-commutativity of $A^\cdot$ is needed for μ below):

$$\Delta\colon \bar{B}(A^\cdot)\to\bar{B}(A^\cdot)\otimes\bar{B}(A^\cdot),\qquad \mu\colon \bar{B}(A^\cdot)\otimes\bar{B}(A^\cdot)\to\bar{B}(A^\cdot),$$

where

$$\Delta([a_1|\ldots|a_r])=\sum_{i=0}^{r}([a_1|\ldots|a_i]\otimes[a_{i+1}|\ldots|a_r]). \tag{3.23}$$

$$\mu([a_1|\ldots|a_r]\otimes[a_{r+1}|\ldots|a_{r+s}])=\sum_\sigma\varepsilon(\sigma)[a_{\sigma(1)}|\ldots|a_{\sigma(r+s)}]; \tag{3.24}$$

here, σ ranges over all (r,s)-shuffles and $\varepsilon\colon S_{r+s}\to\{1,-1\}$ is the weighted sign representation of the symmetric group obtained by assigning weight $-1+\deg a_j$ to a_j. From the definitions, one sees immediately:

(3.25) **Proposition** [11: (3.3.4)]. *For the mixed Hodge structures on $\mathbb{C}\pi_1(X,x)/J^{r+1}$ defined above, the Hopf algebra operations are morphisms of mixed Hodge structures.*

(3.26) **Corollary.** *The Malčev Lie algebra $\mathfrak{g}_r$ of (2.14) is a mixed Hodge substructure of $\mathbb{C}\pi_1/J^{r+1}$, and the bracket operation is a morphism of mixed Hodge structures.*[6]

(3.27) *Remark.* One sees that the formula (3.23) defines, more generally, a mapping

[6] Since $\mathfrak{g}_r$ generates $\mathbb{C}\pi_1/J^{r+1}$, the mixed Hodge structure of $\mathfrak{g}_r$ determines the one on $\mathbb{C}\pi_1/J^{r+1}$

$$\bar{B}_r(A^\cdot)_{x,y} \to \mathfrak{B}_r(\bar{B}(A^\cdot)_{x,z} \otimes \bar{B}(A^\cdot)_{z,y}),$$

for any $x, y, z \in X$, that is a morphism of complexes (the evaluations at z from d_C cancel because of the signs in (3.7)). This shows that truncations of the pairing

$$H_0(P_{x,z}; \mathbb{C}) \otimes H_0(P_{z,y}; \mathbb{C}) \to H_0(P_{x,y}; \mathbb{C}),$$

induced by path multiplication, are morphisms of mixed Hodge structures.

§ 4. The canonical variations of mixed Hodge structure

In this section, we show that the mixed Hodge structures associated to $H^0(P_{x,y}X; \mathbb{C})$ that were defined in (3.21) comprise a unipotent variation of mixed Hodge structure on $X \times X$, i.e., satisfy the conditions of (1.1) and (1.4).

We begin with a somewhat more general set-up. Let S be a complex manifold, $f\colon Z \to S$ a smooth and proper Kähler morphism, and E a divisor in Z with relative normal crossings. The restriction of f to $Y = Z - E$ is denoted g, and we put $Y_s = g^{-1}(s)$ for $s \in S$. Let σ_0 and σ_1 be two cross-sections of g. We consider the family of mixed Hodge structures associated to

$$\{\mathfrak{B}_r H^0(P_{\sigma_0(s),\sigma_1(s)} Y_s; \mathbb{C}\}_{s \in S}, \tag{4.1}$$

which reduces to our original situation when $S = X \times X$, $Y = X \times S$, g is the obvious projection, and

$$\sigma_0(x, y) = (x; x, y), \qquad \sigma_1(x, y) = (y; x, y).$$

The collection of vector spaces (4.1) forms a holomorphic vector bundle $\mathscr{B}_r$ with flat connection d, defined via PY as follows. Let O be a contractible open subset of S. Over O, there are diffeomorphisms

$$Y_{\mathsf{O}} := g^{-1}(0) \cong Y_t \times \mathsf{O} \qquad (\text{any } t \in \mathsf{O}), \tag{4.2}$$

compatible with projections onto I, which may be taken to extend across E. The sections σ_0, σ_1 correspond to maps $g_0, g_1\colon \mathsf{O} \to Y_t$. Pulling back the local system (3.2) along $g_0 \times g_1\colon \mathsf{O} \to Y_t \times Y_t$ we obtain the local system

$$\{H^0(P_{\sigma_0(s),\sigma_1(s)} Y_s)\}_{s \in \mathsf{O}}. \tag{4.3}$$

We wish to realize this flat vector bundle de Rham theoretically (cf. [15: § 2]). As it will soon become clear why we must use C^∞ forms to do so, we proceed as follows. We begin by sheafifying the bar construction, defining

$$\bar{\mathscr{B}}_r^\infty(Y/S) = \bar{B}_r(g_* \mathscr{E}^\cdot_{Y/S}; \mathscr{E}^0_S)_{\sigma_0,\sigma_1}, \tag{4.4}$$

via the description in Definition (3.4), taking tensor products over the sheaf of C^∞ functions $\mathscr{E}^0_S$, and using σ_0 and σ_1 to define the left and right module structures. The trivializations (4.2) give us that $\mathscr{E}^\cdot_Y$ is locally a certain completed tensor product of $\mathscr{E}^\cdot_S$ and $\mathscr{E}^\cdot_{Y_t}$. In fact, the inclusion

$$\mathscr{E}^{\cdot}_0 \otimes \mathscr{E}^{\cdot}_{Y_t} \to \mathscr{E}^{\cdot}_{Y_0}$$

is readily seen to be a quasi-isomorphism. It follows that

$$\mathscr{B}_r^{\infty} = \mathscr{H}^0(\bar{\mathscr{B}}_r^{\infty}(Y/S)) \tag{4.5}$$

is a locally free C^{∞} sheaf on S, with fibers given by (4.1), and $\mathscr{B}_r^{\infty}$ is the C^{∞} bundle underlying $\mathscr{B}_r$. We can also write (see (3.14))

$$\mathscr{B}_r^{\infty} = \mathscr{H}^0(\bar{B}_r(f_* \mathscr{E}^{\cdot}_{Z/S}(\log E); \mathscr{E}^0_S)_{\sigma_0, \sigma_1}).^{7} \tag{4.6}$$

Next, put

$$\mathscr{K}_r^{\infty} = \bar{B}_r(g_* \mathscr{E}^{\cdot}_Y; \mathscr{E}^{\cdot}_S)_{\sigma_0, \sigma_1}, \tag{4.7}$$

in which tensor products[8] are taken over $\mathscr{E}^{\cdot}_S$. Let L denote the decreasing filtration of $\mathscr{K}_r^{\infty}$ induced by "filtration bête" (truncation) in $\mathscr{E}^{\cdot}_S$. We note that the isomorphism

$$Gr_L^p[\otimes^r(g_* \mathscr{E}^{\cdot}_Y)] \cong \mathscr{E}^p_S \otimes_{\mathscr{E}^0_S} [\otimes^r(g_* \mathscr{E}^{\cdot}_{Y/S})]$$

implies that the underlying graded sheaf in (4.7) coincides with that of the bar construction

$$\bar{B}_r^{\infty}(g_* \mathscr{E}^{\cdot}_{Y/S}; \mathscr{E}^{\cdot}_S)_{\sigma_0, \sigma_1}$$

over $\mathscr{E}^0_S$. As their differentials become equal after passing to Gr_L, we see that for all p we have

$$Gr_L^p \mathscr{K}_r^{\infty} \cong \mathscr{E}^p_S \otimes_{\mathscr{E}^0_S} \bar{\mathscr{B}}_r^{\infty}(Y/S)[-p]. \tag{4.8}$$

The exact sequence

$$0 \to Gr_L^1 \mathscr{K}_r^{\infty} \to (L^0/L^2)\mathscr{K}_r^{\infty} \to Gr_L^0 \mathscr{K}_r^{\infty} \to 0 \tag{4.9}$$

determines on cohomology

$$\mathscr{H}^0(Gr_L^0 \mathscr{K}_r^{\infty}) \cong \mathscr{B}_r^{\infty} \xrightarrow{\delta} \mathscr{E}^1_S \otimes \mathscr{B}_r^{\infty} \cong \mathscr{H}^1(Gr_L^1 \mathscr{K}_r^{\infty}). \tag{4.10}$$

(4.11) **Proposition.** *δ defines the flat connection on $\mathscr{B}_r^{\infty}$.*

Proof. Since δ satisfies the Leibniz rule, it is clear that δ defines some flat connection on $\mathscr{B}_r^{\infty}$ (cf. [15]). We want to verify that this connection agrees with the one discussed previously. Now, any element of $\mathscr{H}^0(\mathscr{K}_r^{\infty})$ determines a section of $\mathscr{B}_r^{\infty}$ that is flat with respect to δ. It is enough to see then that these are flat in the other sense, and that they span the fibers of $\mathscr{B}_r^{\infty}$. As the question is local on S, we may place ourselves in the situation of (4.2). The result follows immediately by a direct calculation. □

The filtration W of $\mathscr{E}^{\cdot}_Z(\log E)$ induces filtrations of both $\mathscr{K}_r^{\infty}$ and $\bar{\mathscr{B}}_r^{\infty}(Y/S)$, as does the bar filtration $\mathfrak{B}$. By filtering the exact sequence (4.9), we see that

---

[7] Lest there be any confusion, the notation $\mathscr{E}^{\cdot}_{Z/S}(\log E)$ here indicates forms with only holomorphic logarithmic poles (no $d\bar{z}/\bar{z}$'s), but with all of $f^* \mathscr{E}^{\cdot}_S$ killed

[8] NB – This is in the graded sense: $\omega_1 \otimes \varphi \cdot \omega_2 = \omega_1 \cdot (\tau\varphi) \otimes \omega_2$. One then sees quite readily that d is well-defined

they induce flat sub-bundles of $\mathscr{B}_r^\infty$. Moreover, it is clear that these are defined over $\mathbb{Q}$. The same then holds for $W*\mathfrak{B}$, which induces the weight filtration of the fibers. This gives (1.1, ii).

Next, we use the Hodge filtration F of $\mathscr{E}_Z^\cdot(\log E)$ to define F on $\mathscr{K}_r^\infty$ (and $\bar{\mathscr{B}}_r^\infty(Y/S)$) by taking the total Hodge filtration of

$$\mathscr{E}_S^\cdot \otimes (\otimes^\cdot f_* \mathscr{E}_{Z/S}^\cdot(\log E)).$$

Because σ_0 and σ_1 are holomorphic, this is compatible with the differential in $\mathscr{K}_r^\infty$ (see (3.7)). The F-filtered analogue of (4.8) is then

$$Gr_L^i F^p \mathscr{K}_r^\infty \cong \bigoplus_k (\mathscr{E}_S^{k,i-k} \otimes F^{p-k} \bar{\mathscr{B}}_r^\infty(Y/S))[-i]. \tag{4.12}$$

Clearly,

$$\mathscr{E}\mathscr{F}^p := F^p \bar{\mathscr{B}}_r^\infty(Y/S) \tag{4.13}$$

is the q-th Hodge filtration C^∞ sub-bundle, and the filtered version of (4.10) shows that

$$d(\mathscr{E}\mathscr{F}^p) = \delta(\mathscr{E}\mathscr{F}^p) \subset (\mathscr{E}_S^{1,0} \otimes \mathscr{E}\mathscr{F}^{p-1}) \otimes (\mathscr{E}_S^{0,1} \otimes \mathscr{E}\mathscr{F}^p). \tag{4.14}$$

We next verify that $\mathscr{E}\mathscr{F}^p$ actually defines a holomorphic sub-bundle.

(4.15) **Lemma.** *Let F be a C^∞ sub-bundle of the holomorphic vector bundle E on a complex manifold. A necessary and sufficient condition for F to be a holomorphic sub-bundle is: if σ is a C^∞ section of F, then the values of the $(0,1)$-form $\bar{\partial}\sigma$ are also in F.*

Proof. Consider F as a C^∞ submanifold of the complex manifold E. The condition on F is equivalent to the assertion that its tangent bundle is closed under the almost complex structure of E, which is in turn equivalent to that F is a complex submanifold of E. Details are left to the reader. □

It now follows from (4.14) that the bundles $\mathscr{E}\mathscr{F}^p$ satisfy the criterion of (4.15). Let $\mathscr{F}^p$ denote the subsheaf of holomorphic sections. The following is elementary:

(4.16) **Lemma.** *$d\mathscr{F}^p \subset \Omega_S^1 \otimes \mathscr{F}^{p-1}$ if and only if $d(\mathscr{E}\mathscr{F}^p) \subset \mathscr{E}_S^1 \otimes (\mathscr{E}\mathscr{F}^{p-1})$.* □

Again appealing to (4.14), we conclude:

(4.17) **Proposition.** *The Hodge filtration sub-bundles satisfy* (1.1, iii). □

Finally, we have from (3.16), that

$$Gr_k^{W*\mathfrak{B}} \bar{\mathscr{B}}_r^\infty(Y/S) \cong \bigoplus_{i+j=k} Gr_i^W(\otimes^j f_* \mathscr{E}_{Z/S}^\cdot(\log E))[j], \tag{4.18}$$

from which we see that the C^∞ bundles

$$\mathscr{H}^i(Gr_k^{W*\mathfrak{B}} \bar{\mathscr{B}}_r^\infty(Y/S)), \tag{4.19}$$

together with their filtration $\mathscr{E}\mathscr{F}$, are the direct sum of variations of Hodge structure induced by tensor algebra from geometric ones, coming from the

relative cohomology of Z and the intersections of the components of E. It follows that $Gr_k^{W*\mathfrak{B}}\bar{\mathscr{B}}_r^\infty$, being a sub-bundle of (4.19) when $i=0$, is polarizable. This finishes the verification of:

(4.20) **Proposition.** i) *The filtrations* $W*\mathfrak{B}$ *and* $\mathscr{F}$ *of* $\mathscr{B}_r^\infty = \mathscr{H}^0(\bar{\mathscr{B}}_r^\infty(Y/S))$ *determine a graded-polarizable variation of mixed Hodge structure, defined over* $\mathbb{Q}$,

ii) *By duality, these determine a graded-polarizable variation of mixed Hodge structure with fibers*

$$H_0(P_{\sigma_0(s),\sigma_1(s)}Y_s;\mathbb{C})/J^{r+1}. \quad \square$$

(4.21) *Definition.* i) The r-th *canonical variation* of mixed Hodge structure on $X\times X$ is the variation of mixed Hodge structure of (4.20, ii), with fiber

$$H_0(P_{x,y}X;\mathbb{C})/J^{r+1}$$

at $(x,y)\in X\times X$.

ii) The r-th canonical variation of mixed Hodge structure on X with basepoint x is the restriction of (i) to $\{x\}\times X\cong X$.

iii) The r-th *tautological variation* on X is the pull-back of (i) to X along the diagonal embedding, i.e., the one with fiber $\mathbb{C}\pi_1(X,x)/J^{r+1}$ as x varies over X.

An important feature of the canonical variations is:

(4.22) **Proposition.** *The canonical variations are unipotent variations of mixed Hodge structure.*

Proof. We only need check, in this case, that $Gr_k^{W*\mathfrak{B}}\mathscr{B}_r$ underlies a constant Hodge structure. This follows at once from the fact that the variations of Hodge structure (4.19) are trivial, coming from the cohomology of fixed varieties: V and the intersections of components of D ($X=V-D$, as in (3.20)). $\square$

We close this section with a reformulation of (4.4)–(4.20), whose underlying principle will greatly simplify the discussion to come in §6. Of course, we would have preferred to mimic [15] more closely by working with the holomorphic de Rham complex. It seems impossible to get $\mathscr{B}_r$ in this manner; the best one can do is to replace $\mathscr{B}_r^\infty$ in (4.6) by

$$\mathscr{H}^0(\bar{B}_r(Rf_*\Omega^\bullet_{Z/S}(\log E);\mathscr{C}^0(\mathscr{O}_S))_{\sigma_0,\sigma_1}),$$

which is a locally free $\mathscr{C}^0(\mathscr{O}_S)$-module with "fibers" given by (4.1); here and elsewhere, $\mathscr{C}^i$ denotes the sheaf of Cech i-cochains (with values as indicated). However, a holomorphic construction is available if we restrict ourselves to sufficiently small open subsets of S (cf. (4.2)).

Let $T\subset S$ be an open set such that over T, Z admits a finite Leray covering $\mathfrak{U}$ with the property that

(4.23) i) $f(U)=T$ for every U in the nerve of $\mathfrak{U}$,

ii) $\sigma_0(T)$ and $\sigma_1(T)$ are each contained in only one element of $\mathfrak{U}$.

From (4.23, i), it follows that $f_* \mathscr{C}^\cdot(\Omega^\cdot_{Z/S}(\log E))|_T$ is quasi-isomorphic to the sheaf

$$C^\cdot(\mathfrak{U}, \Omega^\cdot_{Z/S}(\log E)) \tag{4.24}$$

that associates to each open set $V \subset T$

$$C^\cdot(\mathfrak{U} \cap V, \Omega^\cdot_{Z/S}(\log E));$$

from (ii), we get that $\mathscr{O}_T$ is a bimodule over (4.24). Likewise, $\Omega^\cdot_T$ is a bimodule over

$$C^\cdot(\mathfrak{U}, \Omega^\cdot_Z(\log E)).$$

We may then form

$$\mathscr{K}_r = \bar{B}_r(C^\cdot(\mathfrak{U}, \Omega^\cdot_Z(\log E)); \Omega^\cdot_T)_{\sigma_0, \sigma_1}, \tag{4.25}$$

as in (4.7), with $\Omega^\cdot_T$ as coefficient ring, and likewise define L to be the filtration induced by truncation in $\Omega^\cdot_T$. Then (cf. (4.8))

$$Gr^p_L \mathscr{K}_r \cong \Omega^p_T \otimes \bar{B}'_r[-p], \tag{4.26}$$

where

$$\bar{B}'_r = \bar{B}_r(C^\cdot(\mathfrak{U}, \Omega^\cdot_{Z/S}(\log E)); \mathscr{O}_T)_{\sigma_0, \sigma_1}, \tag{4.27}$$

and we have a connection δ', analogous to (4.10):

$$\delta' \colon \mathscr{H}^0(\bar{B}'_r) \to \Omega^1_T \otimes \mathscr{H}^0(\bar{B}'_r). \tag{4.28}$$

(4.29) **Proposition.** *As holomorphic vector bundles with connection, there is an isomorphism*

$$(\mathscr{H}^0(\bar{B}'_r), \delta') \cong (\mathscr{B}_r, d)|_T.$$

Proof. We consider the inclusions

$$\mathscr{K}_r \xhookrightarrow{i} \mathscr{K}_r{}^\wedge \xhookleftarrow{j} \mathscr{K}_r{}^\infty|_T,$$

where

$$\mathscr{K}_r{}^\wedge = \bar{B}_r(C^\cdot(\mathfrak{U}, \mathscr{E}^\cdot_Z(\log E)); \mathscr{E}^\cdot_T)_{\sigma_0, \sigma_1}$$

(with $\mathscr{E}^\cdot_T$ as coefficient ring). Defining L on $\mathscr{K}_r$ by truncation in $\mathscr{E}^\cdot_T$, we have that

i) j is a filtered quasi-isomorphism with respect to L,

ii) given (i), i induces a commutative diagram

$$\begin{array}{ccc} \mathscr{H}^0(\bar{B}'_r) & \xrightarrow{\delta'} & \Omega^1_T \otimes \mathscr{H}^0(\bar{B}'_r) \\ \downarrow & & \downarrow \\ \mathscr{B}_r^\infty|_T & \xrightarrow{\delta} & \mathscr{E}^1_T \otimes \mathscr{B}_r^\infty|_T. \end{array}$$

Since $\mathscr{H}^0(\bar{B}'_r)$ and $\mathscr{B}_r^\infty|_T$ have the same fibers, the desired assertion is a direct consequence of (ii). □

From here, we obtain rather easily:

(4.30) **Corollary.** *The filtrations* $W*\mathfrak{B}$ *and* F *of* $\bar{B}'_r$ *induce the weight and Hodge filtrations resp. of* $\mathscr{B}_r|_T$.

§ 5. Classifying spaces and higher Albanese manifolds

In this Section, we show how the Malčev groups $\hat{G}_r$ (2.14), associated to $G = \pi_1(X)$, can be used to construct a sort of classifying space for certain mixed Hodge structures with fixed pure graded quotients. This was known to Deligne from a somewhat different point of view [9].

First, we recall the "standard" theory of classifying spaces for graded-polarized mixed Hodge structures over $\mathbb{Z}$, beginning with the pure case (following [20: § 3]).

Let V be a complex vector space, containing a distinguished lattice Λ (so that $V = \Lambda \otimes_{\mathbb{Z}} \mathbb{C}$), and Q a $(-1)^k$-symmetric bilinear form on V that is defined over $\mathbb{Q}$. The set $D(V)$ of filtrations on V that define a Hodge structure of weight k, polarized by Q, with specified Hodge numbers, is a homogeneous space for the real points of the (semi-simple) isometry group $H = \mathbf{O}(Q)$[9]. If F is any one such filtration, the (compact) stabilizer of F is denoted F^0H; the reason for this notation is that the Lie algebra $\mathfrak{h}$ of H inherits a filtration from F, and F^0H is the subgroup corresponding to the Lie subalgebra $F^0\mathfrak{h}$. Then

$$D(V) \cong H_{\mathbb{R}}/F^0H_{\mathbb{R}} \tag{5.1}$$

has a natural structure of a homogeneous complex manifold[10]. If $\tilde{F} = hF$ ($h \in H_{\mathbb{R}}$) is another Hodge filtration in $D(V)$, then right multiplication by h in $H_{\mathbb{R}}$ sets up an isomorphism

$$H_{\mathbb{R}}/\tilde{F}^0H_{\mathbb{R}} \to H_{\mathbb{R}}/F^0H_{\mathbb{R}}. \tag{5.2}$$

Since $H_{\mathbb{Z}}$ acts properly and continuously on $D(V)$, $H_{\mathbb{Z}} \backslash D(V)$ is an analytic space with only quotient singularities; it parametrizes the isomorphism classes of Hodge structures over $\mathbb{Z}$ with the given polarization and Hodge numbers.

(5.3) *Remark.* Hodge structures of the above type defined only over some larger sub-ring $\mathbb{A} \subset \mathbb{R}$ are parametrized by $H_{\mathbb{A}} \backslash D(V)$, but the latter is usually not such a nice topological space.

The tangent bundle $TD(V)$ of $D(V)$ is the homogeneous vector bundle with fiber $\mathfrak{h}/F^0\mathfrak{h}$, associated to the adjoint representation of F^0H. The so-called *horizontal sub-bundle* $T_hD(V)$ is associated to the F^0H-invariant subspace

$$Gr_F^{-1}\mathfrak{h} = F^{-1}\mathfrak{h}/F^0\mathfrak{h}. \tag{5.4}$$

A variation of Hodge structure on S, defined over $\mathbb{A}$, of the specified type, is given up to isomorphism by a mapping

[9] Our convention is that H acts via *left* multiplication

[10] Canonically, the set of flagged frames for Hodge filtrations of V is principal homogeneous under $H_{\mathbb{R}}$, and $D(V)$ is the tautological quotient that forgets the frame

(5.5) $$\varphi\colon S \to H_{\mathbb{A}} \backslash D(V)$$

locally liftable to a holomorphic mapping into $D(V)$, such that (we pass to liftings, if necessary, to make sense of the following)

(5.6) $$\varphi_* TS \subset T_h D(V).$$

One says that φ is *horizontal* when (5.6) is satisfied.

There is a parallel theory for graded-polarized mixed Hodge structures defined over $\mathbb{A}$ [2, 25]. Let W denote the weight filtration of V, defined over the quotient field of $\mathbb{A}$. Roughly, a mixed Hodge structure consists of a finite number of polarized Hodge structures on the $Gr_k^W V$'s, together with some iterated extension data. So, given Hodge numbers and polarizing forms for each $Gr_k^W V$, we write H_k and D_k for the corresponding objects in (5.1). If $D(V)$ now denotes the set of filtrations of V that induce Hodge filtrations of the specified types, there is a tautological surjection

(5.7) $$\pi\colon D(V) \to \prod_k D_k.$$

To obtain a description of $D(V)$ analogous to (5.1), one considers the canonical mapping

$$\rho\colon W_0 \operatorname{Aut}(V) \to \prod_k \operatorname{Aut}(Gr_k^W V),$$

and puts

(5.8) $$\tilde{\mathsf{O}} = \rho^{-1}\left(\prod_k H_k\right).$$

Then

(5.9) $$H = \tilde{\mathsf{O}}_{\mathbb{R}} \cdot W_{-1} \operatorname{Aut}(V)_{\mathbb{C}}$$

acts transitively on $D(V)$. The subgroup $W_{-1}\operatorname{Aut}(V)_{\mathbb{C}}$, which is normal in H, preserves the fibers of π. (One can use a splitting of the filtration W to write

$$H \cong \left(\prod_k H_{k,\mathbb{R}}\right) \ltimes W_{-1} \operatorname{Aut}(V)_{\mathbb{C}}.)$$

Let F denote any one filtration in $D(V)$. Then $D(V) \cong H/F^0 H$, and $H_{\mathbb{A}} \backslash D(V)$ is the desired classifying space (it is again a nice space when $\mathbb{A} = \mathbb{Z}$). The tangent bundle and horizontal sub-bundle of $D(V)$ are homogeneous, and are defined as before; in the present setting, (5.5) and (5.6) describe graded-polarized variations of mixed Hodge structure.

For emphasis, we state an immediate consequence of the preceding discussion:

(5.10) **Proposition.** i) *The set $UD(V)$ of filtrations on V that induce fixed Hodge structures on the $Gr_k^W V$'s is canonically isomorphic to*

$$W_{-1} \operatorname{Aut}(V)_{\mathbb{C}} / F^0 W_{-1} \operatorname{Aut}(V)_{\mathbb{C}}.$$

ii) *The classifying space for mixed Hodge structures on V, defined over $\mathbb{A}$, that induce the given Hodge structures on the $Gr_k^W V$'s is*

$$UD(V)/\mathbb{A} := W_{-1}\operatorname{Aut}(V)_{\mathbb{A}} \backslash UD(V).$$

iii) *A unipotent variation of mixed Hodge structure on S is classified by a horizontal locally liftable holomorphic mapping of the form*

$$\varphi\colon S \to UD(V)/\mathbb{A}.$$

(5.11) *Remarks.* i) The graded-polarization is actually irrelevant for (5.10).

ii) As in (5.2), the complex structure on $UD(V)$ is canonical,

iii) $W_{-1}\operatorname{Aut}(V)_{\mathbb{Z}}$ acts *freely* on $D(V)$, hence also on $UD(V)$.

iv) The structure of $UD(V)$ as a homogeneous space admits the following interpretation. The group $W_{-1}\operatorname{Aut}(V)_{\mathbb{C}}$ acts simply transitively on the set of all frames for V that are flagged for W and lift given Hodge frames for the $Gr_k^W V$'s, and $F^0 H$ does likewise on the frames for the given filtration F.

v) Note that $UD(V)/\mathbb{C}$ is just a point. This is a reflection of the fact that there are no moduli for a *pair* of filtrations (W, F) on a vector space V: up to $\operatorname{Aut}(V)$, they are completely determined by the numbers

$$d_k^p := \dim Gr_F^p Gr_k^W.$$

We now take for V

$$V_r = \mathbb{C}\pi_1(X)/J^{r+1}, \tag{5.12}$$

and, of course, $\mathbb{Z}\pi_1(X)/J^{r+1}$ for Λ. We select the Hodge filtration F_x of $H_0(P_{x,x};\mathbb{C})/J^{r+1}$ from (3.21) for F. The right regular representation R of $\mathbb{C}\pi_1(X)$ sets up an embedding

$$V_r \hookrightarrow \operatorname{End}(V_r), \tag{5.13}$$

which is a morphism of mixed Hodge structures by (3.25). This in turn embeds the Malčev Lie algebra $\mathfrak{g}_r$ of (2.14) in $\operatorname{End}(V_r)$ as a mixed Hodge substructure, and the Malčev group $\hat{G}_r$ in $GL(V_r)$. By strictness,

$$\mathfrak{g}_{r,\mathbb{C}} \cap \iota^{-1}(F^0 \operatorname{End}(V_r)) = F^0 \mathfrak{g}_{r,\mathbb{C}},$$

which implies that (5.13) induces a mapping

$$\hat{G}_{r,\mathbb{Z}} \backslash \hat{G}_{r,\mathbb{C}} / F^0 \hat{G}_{r,\mathbb{C}} \to UD(V_r)/\mathbb{Z}, \tag{5.14}$$

which is a covering onto its image.

(5.15) *Definition.* The r-th *Albanese manifold* of X, with basepoint x, is the complex manifold $\operatorname{Alb}_x^r(X)$, given by the left-hand side of (5.14).

From the functoriality of the mixed Hodge structures associated to π_1 with respect to pointed mappings, we see that $f\colon (X,x) \to (Z,z)$ induces a tower of mappings

$$f_*\colon \operatorname{Alb}_x^r(X) \to \operatorname{Alb}_z^r(Z).$$

It appears, a priori, that the complex structure of $\mathrm{Alb}_x^r(X)$ depends on the point x. We will see soon (5.20, iii) that it does not.

The mapping (5.14) is an example of the basic:

(5.16) *Construction.* For any mixed Hodge structure V, defined over $\mathbb{A}$, an algebra homomorphism of mixed Hodge structures

$$\mathbb{C}\pi_1(X,x)/J^{r+1} \to \mathrm{End}(V) \quad \text{(defined over } \mathbb{A}),$$

which we also call a *mixed Hodge theoretic* (*unipotent*) *representation* of $\pi_1(X,x)$, determines (as in (5.14)) a holomorphic mapping

$$\mathrm{Alb}_x^r(X) \to UD(V)/\mathbb{A},$$

whose differential respects the horizontal sub-bundles.

We wish to construct a "universal" holomorphic and horizontal mapping

$$\alpha_x^r \colon X \to \mathrm{Alb}_x^r(X), \tag{5.17}$$

such that $\alpha_x^r(x)$ is the basepoint of $\mathrm{Alb}_x^r(X)$, namely the double coset of $1 \in \hat{G}_r$. Recall from (4.22) that the canonical variations are unipotent, hence are classified by horizontal mappings

$$X \times X \to UD(V_r)/\mathbb{Z}. \tag{5.18}$$

The following will be shown by direct calculation:

(5.19) **Proposition.** *The classifying mapping* (5.18) *for the* r-th *canonical variation factors through the isomorphism*

$$\mathrm{Alb}_x^r(X) \times \mathrm{Alb}_x^r(X) \xrightarrow{\sim} \mathrm{Alb}_{(x,x)}^r(X \times X)$$

via the representation ρ *of*

$$\pi_1(X \times X, (x,x)) \cong \pi_1(X,x) \times \pi_1(X,x)$$

on V_r *given by*

$$\rho(\gamma_0, \gamma_1) = L(\gamma_0^{-1})R(\gamma_1);$$

here L *denotes the left regular representation.*

This fact has some obvious consequences:

(5.20) **Corollary.** i) *The classifying mapping for the* r-th *canonical variation with basepoint* x *factors through* (5.17).

ii) *The classifying mapping for the* r-th *tautological variation* (4.21, iii) *factors through* $\mathrm{Alb}_x^r(X)$,[11] *as in* (5.16) *with* $V = V_r$, *via the adjoint representation of* $\hat{G}_r$.

iii) *For* $x, y \in X$, *there is a canonical isomorphism*

$$\mathrm{Alb}_x^r(X) \cong \mathrm{Alb}_y^r(X) \quad \text{(cf. (5.11, ii))}.$$

Of course, we will take the classifying mappings for the canonical variations with basepoint x to be the "universal" mapping referred to in (5.17).

[11] in fact, through $\mathrm{Alb}_x^{r-1}(X)$

(5.21) **Proposition-Construction.** *Every mixed Hodge theoretic unipotent representation of* $\pi_1(X,x)$, *defined over* $\mathbb{A}$, *determines a unipotent variation of mixed Hodge structure defined over* $\mathbb{A}$.

Proof. Composing α_x^r, which is horizontal by (5.20, i), with the mapping in (5.16) gives a holomorphic horizontal mapping

$$X \to UD(V)/\mathbb{A}.$$

It is locally liftable because α_x^r is. □

(5.22) *Remark.* i) Before beginning the proof of (5.19), we observe that it is obvious from (3.1) that ρ is the correct formula for the monodromy in the canonical variation. What the proposition asserts is that this discrete data in a sense "spans" the classifying mapping.

ii) The assertion of (5.19) is stronger than what can be deduced from (3.25) by general principles of Hodge theory.

As the essential elements of the proof of (5.19) all occur in the direct verification of (5.20, i), we will content ourselves here with the proof of the latter, slightly simpler assertion.

Let $F_{x,y}$ denote the Hodge filtration at (x,y) in the canonical variation, transported to V_r. Then we can write

$$F_{x,y} = h_y F_{x,x}$$

for some $h \in \mathrm{Aut}(V_r)$. What we want to show is that for all $y \in X$, h_y can be taken to be of the form $R(g_y)$, where g_y is in the Malčev group (lifting to the covering in (5.14) is not a problem here). Said another way, we want to see that there exists a linear transformation, expressing a flat frame in terms of what we shall call a *unipotent Hodge frame* (see (5.11, iv)), that lies in image of the right regular representation of the Malčev group. It is very helpful to observe that because questions of holomorphy are not at stake, *it is enough to find merely a* C^∞ *frame* with the above properties. Also, the existence of such frames will be determined for the *tower* of canonical variations, and not for each one individually.

We compute explicitly. Let $\gamma(t)$ be a parametrized path from x to y, and $\omega_r(t)$ the connection matrix with respect to some C^∞ frame – or better, the connection *transformation* with respect to its $\mathbb{C}$-span. The *transport* $T(t) \in \mathrm{Aut}(V_r)$, expressing a flat frame with respect to the given frame, is then the solution of the differential equation

$$T'(t) = T(t)\omega_r(t), \qquad T(0) = I. \tag{5.23}$$

We show:

(5.24) **Proposition.** *There exists a unipotent Hodge frame such that* $T(t) \in R(\hat{G}_r)$ *for all* t.

(5.25) **Lemma.** $T(t) \in R(\hat{G}_r)$ *for all* t *if and only if* $\omega_r(t) \in R(\mathfrak{g}_r)$ *for all* t.

Proof. For any closed subgroup G of $GL(V)$, with Lie algebra $\mathfrak{g} \subset \mathrm{End}(V)$, the tangent space to G at $M \in G$ is $M\mathfrak{g}$. Since $R(\mathfrak{g}_r)$ is the Lie algebra of $R(\hat{G}_r)$, the assertion of the lemma follows from (5.23). □

Thus, Proposition (5.24) is a consequence of

(5.26) **Proposition.** *There is a system of C^∞ unipotent Hodge frames for the V_r's such that $\omega_r(t) \in R(\mathfrak{g}_r)$ for all t.*

Since it is for the duals of the V_r's for which we have constructed mixed Hodge complexes (see (3.21)), we are obliged at this point to treat the dual connection in order to prove (5.26). To this end, we state an elementary fact:

(5.27) **Lemma.** *Let A be a complete Hopf algebra that is the algebraic dual of a Hopf algebra B.*

i) *Suppose that $M: A \to A$ is a linear mapping that is dual to $M^*: B \to B$. Then M is given by multiplication on the right by the element $\lambda \in A$ if and only if M^* factors as $M^* = \psi \circ \Delta$, where*

$$\psi: B \otimes B \to B$$

is given by $\psi(b_1 \otimes b_2) = \langle \lambda, b_2 \rangle b_1$;

ii) *An element $\lambda \in A$ is primitive if and only if*

$$\langle \lambda, b_1 \cdot b_2 \rangle = \langle 1, b_1 \rangle \langle \lambda, b_2 \rangle + \langle 1, b_2 \rangle \langle \lambda, b_1 \rangle \quad \text{for all } b_1, b_2 \in B.$$

Proof. These assertions are a direct consequence of the dual relationship between the coproduct in A and the multiplication in B. □

Thus, to prove (5.26), it suffices to show that with respect to a system of unipotent Hodge frames for the bundles $\mathscr{B}_r^\infty$, the values of connection transformation are the restriction of elements of

$$\mathrm{End}(H^0(\bar{B}(A^\bullet)_{x,y})$$

that satisfy the conditions of (5.27).

Let $(M^\bullet, W, F)$ be a bifiltered differential graded algebra and $\rho: M^\bullet \to A^\bullet$ a filtration preserving homomorphism such that

i) $M^0 = \mathbb{C}$,

ii) ρ is a bifiltered quasi-isomorphism with respect to (W, F).

The existence of such is a consequence of [16: (6.10)]. The induced maps

$$B(\rho): \bar{B}_r(M^\bullet) \to \bar{B}_r(A^\bullet)_{x,y} \tag{5.28}$$

are then also bifiltered quasi-isomorphisms, by (3.17). Thus, (5.28) determines a set of "false constant sections" of $\mathscr{B}_r^\infty$. We say "false" here because these are not flat sections of the bundle; indeed, they are not even holomorphic in general[12]. We can see this by computing δb, where b is a cocycle in $\bar{B}_r(M^\bullet)$, viewed as a section of $\bar{\mathscr{B}}_r^\infty(X \times X/X)$. From (4.10) and (3.7), we see that δ is induced by

[12] If $\tilde{A}^\bullet$ is a quasi-isomorphic DGA, it may define a different set of false constants, for the quasi-isomorphism of bar complexes makes use of exact 1-cochains, and hence involves the varying formulas (3.8)

$$[a_1|\dots|a_r]\mapsto\begin{cases}\rho(a_r)\otimes[a_1|\dots|a_{r-1}] & \text{if } \deg a_r=1,\\ 0 & \text{otherwise.}\end{cases} \tag{5.29}$$

Thus, if $\xi\in TX$, the connection transformation in the direction ξ is induced by

$$[a_1|\dots|a_r]\mapsto\begin{cases}\langle\rho(a_r),\xi\rangle[a_1|\dots|a_{r-1}] & \text{if } \deg a_r=1,\\ 0 & \text{otherwise.}\end{cases} \tag{5.30}$$

Recalling (3.23), we have that the coproduct in $\bar{B}(M^\cdot)$ is induced by

$$\Delta[a_1|\dots|a_r]=\sum_{i=0}^{r}([a_1|\dots|a_i]\otimes[a_{i+1}|\dots|a_r]).$$

With the help of (5.30), we see that we get (5.27, i), with λ determined by

$$\langle\lambda,[a_1|\dots|a_r]\rangle=\begin{cases}\langle\rho(a_1),\xi\rangle & \text{if } r=1 \text{ and } \deg a_1=1,\\ 0 & \text{otherwise,}\end{cases} \tag{5.31}$$

extended linearly.[13]

Turning to (5.27, ii), we see that it suffices to verify that $\langle\lambda,b_1\cdot b_2\rangle=0$ whenever b_1 and b_2 are annihilated by the unit element of A. But this is obvious from (3.24) and (5.31). □

Recall that the usual *Albanese manifold* of X is defined by

$$\mathrm{Alb}(X)=H_1(X,\mathbb{Z})\backslash(F^1H^1(X))^*, \tag{5.32}$$

and the *Albanese mapping* with basepoint x is

$$\alpha_x\colon X\to\mathrm{Alb}(X), \tag{5.33}$$

defined via

$$\omega\xmapsto{\alpha_x(y)}\int_\gamma\omega \quad (\text{any } \gamma\in P_{x,y}X)$$

for all $\omega\in H^0(\Omega^1_{\bar{X}}(\log D))$. Our goal is to identify (5.33) with α^1_x of (5.17).

One interprets (5.33) as a classifying mapping for a variation of mixed Hodge structure as follows (cf. [29: p. 116]). The exact sequence of complexes

$$0\to\mathbb{C}_y\to\mathrm{cone}(\Omega^\cdot_X\to\mathbb{C}_x\oplus\mathbb{C}_y)\to\mathrm{cone}(\Omega^\cdot_X\to\mathbb{C}_y)\to 0$$

gives rise to the extension of mixed Hodge structures

$$0\to H^0(\{y\},\mathbb{C})\to H^1(X,\{x,y\},\mathbb{C})\to H^1(X,\{x\},\mathbb{C})\to 0 \tag{5.34}$$

for $y\neq x$, and to the split extension when $y=x$. Its dual is

$$0\to H_1(X,\mathbb{C})\to H_1(X,\{x,y\};\mathbb{C})\to\mathbb{C}\to 0. \tag{5.35}$$

[13] It is readily verified that λ annihilates coboundaries

This latter extension is separated in the sense of [29: p. 110], in that it is of the form

$$0 \to W_{-1} H \to H \xrightarrow{p} Gr_0^W H \to 0. \tag{5.36}$$

The set of all extensions of mixed Hodge structures as in (5.36), with W_{-1} and Gr_0^W fixed, is naturally parametrized by the image of

$$\mathrm{Hom}(Gr_0^W H, W_{-1} H) \hookrightarrow W_{-1}\,\mathrm{End}(H) \xrightarrow{\exp} W_{-1}\,\mathrm{Aut}(H) \to UD(H)/\mathbb{Z}.$$

To represent the extension class of (5.36), one uses splittings

$$Gr_0^W H \underset{s'}{\overset{s}{\rightleftarrows}} H,$$

where s is compatible with the $\mathbb{Z}$-structure, s' with the Hodge filtrations; one then takes

$$(s'-s) \in \mathrm{Hom}(Gr_0^W H, W_{-1} H).$$

We can write

$$s'-s = r\,s' = -r'\,s,$$

where $r = 1 - sp$ and $r' = 1 - s'p$ retract H onto $W_{-1} H$, compatible with the $\mathbb{Z}$-structure and Hodge filtrations respectively. For

$$\omega \in (W_{-1} H)^* = H^*/W_0 H^*,$$

we get, following [29: p. 125],

$$\langle (s'-s)(h), \omega \rangle = -\langle r'\,s(h), \omega \rangle = \langle s(h), s'^*(\omega) \rangle \tag{5.37}$$

where s'^*: $H^*/W_0 H^* \to H^*$ is the section that is dual to r',[14] hence compatible with F. In the case of (5.35), we have an *isomorphism* (see (5.34))

$$F^1 H^1(X) \cong F^1 H^1(X, \{x, y\}),$$

i.e. $s'^*|_{F^1}$ is uniquely determined. For $h = [\gamma] \in Gr_0^W H$, $s(h)$ is the homology class of a path γ from x to y, and we obtain for (5.37)

$$\langle (s'-s)\,h, \omega \rangle = \int_\gamma \omega.$$

Thus, α_x classifies $H_1(X, \{x, y\}; \mathbb{C})$.

Next, we note that there is a canonical isomorphism

$$H_0(P_{x,y} X)/J^2 \to H_1(X, \{x, y\}). \tag{5.38}$$

(5.39) **Proposition.** *The mapping* (5.38) *is an isomorphism of mixed Hodge structures.*

[14] The description of a dual filtration, as a translate of the dual reference filtration, is given by the *inverse* transpose, whence the change of sign in (5.37)

Proof. Passing to the duals, we must compare the mixed Hodge complexes

i) $E_{\bar{X}}^{\bullet}(\log D) \xrightarrow{\varepsilon_x \oplus \varepsilon_y} (\mathbb{C} \oplus \mathbb{C})$

ii) $(\mathbb{C} \oplus E_{\bar{X}}^{+}(\log D)[-1])/[dg \sim (g(y) - g(x))$ whenever $g \in E_{\bar{X}}^{0}]$.

Of course, (i) is bifiltered quasi-isomorphic to

i)′ $\{\varphi \in E_{\bar{X}}^{\bullet}(\log D) : \varphi(x) = \varphi(y) = 0$ if $\varphi \in E_{\bar{X}}^{0}\}$,

and i)′ maps into ii), respecting the filtrations. □

(5.40) **Corollary.** $\alpha_x = \alpha_x^1$. □

At the negligible cost of changing H^0, we may replace (i) above by

$$E_{\bar{X}}^{\bullet}(\log D) \xrightarrow{\varepsilon_y - \varepsilon_x} \mathbb{C}.$$

This suggests the (correct) mixed Hodge theoretic interpretation of addition:

$$\alpha_x(y_1, \ldots, y_l) := \sum_{j=1}^{l} \alpha_x(y_j) \tag{5.41}$$

is the extension class corresponding to the mixed Hodge complex

$$E_{\bar{X}}^{\bullet}(\log D) \xrightarrow{e} \mathbb{C}, \qquad e = \sum_{j=1}^{l} (\varepsilon_{y_j} - \varepsilon_x).$$

(5.42) *Problem.* Show that the mapping $\alpha_x : X^l \to \mathrm{Alb}(X)$ defined by (5.41) cannot be lifted to give a compatible system of mappings

$$\alpha_x^r : X^l \to \mathrm{Alb}_x^r(X).$$

The higher Albaneses of a variety do not, in general, admit the structure of a smooth algebraic variety (or of a Zariski open subset of a compact Kähler manifold.) This is a consequence of a theorem of Morgan [16: (9.4)] that places restrictions on the fundamental groups of smooth varieties. In particular, very few nilpotent groups occur.

A finitely generated group G is *rationally nilpotent* if its Malčev Lie algebra $\mathfrak{g}$ is nilpotent.

(5.43) **Theorem.** *Suppose that X is a smooth algebraic variety (or a Zariski open in a compact Kähler manifold) whose fundamental group is not rationally nilpotent. If $r \geqq 4$ and $\mathrm{Alb}^r(X) \neq \mathrm{Alb}^3(X)$, then $\mathrm{Alb}^r(X)$ does not have the homotopy type of any Zariski open subset of a compact Kähler manifold. In particular, infinitely many of the higher Albanese manifolds of X are not algebraic varieties.*

Before giving the proof, we remark that the fundamental group of a smooth curve is not rationally nilpotent except in the case that it is abelian.

Proof. If G is a group, then there is a natural graded Lie algebra surjection

$$\mathbf{L}(H_1(G;\mathbb{Q}))\to Gr\,\mathfrak{g},$$

where $\mathfrak{g}$ is the Malčev Lie algebra of G and $\mathbf{L}(V)$ denotes the free Lie algebra generated by V [19: A.2, 2.13]. Thus

$$Gr\,\mathfrak{g}\cong\mathbf{L}(H_1(G;\mathbb{Q}))/\mathbf{R},$$

where $\mathbf{R}$ is a graded ideal.

Morgan's theorem [16: (9.4)] asserts that if G is the fundamental group of a Zariski open subset of a compact Kähler manifold, then $\mathbf{R}$ is generated by elements $\{R_j\}$ of degree 2, 3 and 4.

To deduce the theorem from this, one first uses (2.17) to show that

$$\mathfrak{g}(\mathrm{Alb}^r(X))=\mathfrak{g}(X)/\mathfrak{g}(X)^{r+1}.$$

Thus

$$Gr\,\mathfrak{g}(\mathrm{Alb}^r(X))\cong\mathbf{L}(H_1(X;\mathbb{Q}))/\mathbf{R}+\mathbf{I}^{r+1}, \tag{5.44}$$

where

$$\mathbf{L}(V)=\mathbf{I}\supseteq\mathbf{I}^2\supseteq\ldots$$

denotes the lower central series of $\mathbf{L}(V)$.

Since $\pi_1(X)$ is not rationally nilpotent, there exists $t\geqq r$ such that

$$\mathrm{Alb}^r(X)=\mathrm{Alb}^t(X)\neq\mathrm{Alb}^{t+1}(X).$$

Consequently, the projection

$$\mathfrak{g}(\mathrm{Alb}^{t+1}(X))\to\mathfrak{g}(\mathrm{Alb}^t(X))$$

is not an isomorphism, which, by (5.44), implies that

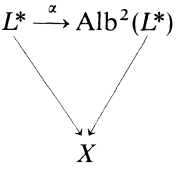
$$\mathbf{R}\not\supset\mathbf{I}^{t+1}.$$

Since $t\geqq r\geqq 4$, it follows that $\mathbf{R}+\mathbf{I}^{t+1}$ cannot be generated by elements of degree 2, 3 and 4. Morgan's theorem then implies that $\mathrm{Alb}^r(X)$ does not have the homotopy type of a Zariski open subset of a compact Kähler manifold. □

(5.45) *Remarks and Examples.* Examples of non-classical Albaneses that are algebraic can be constructed. For example, suppose that $L\to X$ is a line bundle over an abelian variety X with $c_1(L)\neq 0$ in $H^2(X)$. Consider the associated $\mathbb{C}^*$ bundle p: $L^*\to X$. The nontriviality of c_1 ensures that $\pi_1(L^*)$ is nonabelian and that $H_1(L^*)/\mathrm{torsion}\cong H_1(X)$, so that p: $\mathrm{Alb}^1(L^*)\to\mathrm{Alb}^1(X)=X$ is an isomorphism. The Albanese mapping

$$\alpha\colon\ L^*\to\mathrm{Alb}^2(L^*)$$

is a biholomorphism. To see this note that the diagram



commutes and that each is a principal $\mathbb{C}^*$ bundle over X. Since α induces an isomorphism on π_1, α is a biholomorphism. All of the higher Albaneses of L^* are algebraic since

$$\mathrm{Alb}^r(L^*) = \mathrm{Alb}^2(L^*)$$

whenever $r \geqq 2$.

Not all second Albaneses admit an algebraic (or analytic) structure such that the projection $\mathrm{Alb}^2(X) \to \mathrm{Alb}^1(X)$ is algebraic. For example, let $X = \mathbf{P}^1 - \{0, 1, \infty\}$. Then $\mathrm{Alb}^1(X) \cong \mathbb{C}^* \times \mathbb{C}^*$ and $\mathrm{Alb}^2(X)$ is a $\mathbb{C}^*$ bundle over $\mathbb{C}^* \times \mathbb{C}^*$ with $c_1 = 1$. Using a more precise statement of Morgan's theorem, one can show that $\mathrm{Alb}^2(X)$ is not biholomorphic to a smooth variety such that the map $\mathrm{Alb}^2(X) \to \mathbb{C}^* \times \mathbb{C}^*$ is a morphism. Note, however, that $\mathrm{Alb}^2(X)$ does have the homotopy type of a smooth algebraic variety: viz, a $\mathbb{C}^*$ bundle over an elliptic curve.

§ 6. Limit behavior

In this Section, we will show that the canonical variations of mixed Hodge structure, and also the more general path space variations (4.3) in the case where S is a curve, are good, in the sense that they satisfy the conditions of (1.5) (cf. [23: (3.13), App.]). It will follow that all variations constructed by the procedure of (5.21) are likewise good.

The proof comes down to incorporating the construction presented in [23: § 5] into the relative bar construction.

Let $f\colon Z \to S$ and $E \subset Z$ be as in § 4. We suppose that f admits a compactification

$$\bar{f}\colon \bar{Z} \to \bar{S}, \tag{6.1}$$

such that

i) $\bar{Z}$ and $\bar{S}$ are compact Kähler manifolds,
ii) $\Sigma = \bar{S} - S$ is a divisor with normal crossings,
iii) with $\bar{Z}_\Sigma$ denoting $\bar{f}^*\Sigma$, $\bar{Z}_\Sigma \cup \bar{E}$ is a divisor with normal crossings,
iv) σ_0 and σ_1 extend to sections $\bar{\sigma}_0$ and $\bar{\sigma}_1$ of $\bar{f}$.

For the canonical variation, in which we have

$$X = \bar{X} - D, \quad S = X \times X, \quad Z = \bar{X} \times S,$$

we have the natural choices for compactifications:

$$\bar{S} = \bar{X} \times \bar{X}, \quad \bar{Z} = \bar{X} \times \bar{S}, \quad \text{etc.} \tag{6.2}$$

In this case, the restriction of $\bar{f}$ over every smooth complex analytic curve containing a point of Σ and meeting $X \times X$ still satisfies (6.1, ii-iv) - a phenomenon that does not occur in general. We are therefore able to give a unified treatment of the questions of limit behavior by assuming that $\bar{S}$ *in* (6.1) *is a curve, which we do from now on,* and hence Σ is a finite set. Also, the questions are local around Σ, so *we may assume, without loss, that* $\bar{S}$ *is a disc* T, and S is the punctured disc T^*.

Having made these reductions, we can now define

$$\tilde{\mathscr{B}}_r = \mathscr{H}^0[\bar{B}_r(\bar{f}_* C^\cdot(\mathfrak{U}, \Omega^\cdot_{\bar{Z}/S}(\log(\bar{E} \cup \bar{Z}_\Sigma)); \mathscr{O}_S)_{\bar{\sigma}_0, \bar{\sigma}_1}], \tag{6.3}$$

where $\mathfrak{U}$ is a covering of $\bar{f}^{-1}(T)$ that obeys (4.23) (extended across Σ). The first task is to prove:

(6.4) **Proposition.** *Let* g *be as in* §4. *If the local monodromy of* $R^1 g_* \mathbb{C}$ *is unipotent,* $\tilde{\mathscr{B}}_r$ *is the canonical extension to* T *of the bundle* $\mathscr{B}_r|_{T^*}$.

Proof (cf. [23: (5.3)]): We begin by showing that $\tilde{\mathscr{B}}_r$ is a free $\mathscr{O}_T$-module. Let Z_∞ be the pullback of Z to the universal covering of T^*, $k\colon Z_\infty \to Z$, $E_\infty = k^{-1}(E)$, $Y_\infty = Z_\infty - E_\infty$, $i\colon \bar{Z}_0 \to \bar{Z}$. Then there is a quasi-isomorphism

$$\Omega^\cdot_{Z/S}(\log(\bar{E} \cup \bar{Z}_0)) \otimes \mathscr{O}_{Z_0} \approx i^{-1} k_* \Omega^\cdot_{Z_\infty}(\log E_\infty) \tag{6.5}$$

on $\bar{Z}_0$ (cf. [22: (2.16)]), from which one obtains

$$H^\cdot(\bar{Z}_0, \Omega^\cdot_{Z/S}(\log(\bar{E} \cup \bar{Z}_0)) \otimes \mathscr{O}_{Z_0}) \cong H^\cdot(Y_\infty, \mathbb{C}) \cong H^\cdot(Y_s, \mathbb{C}) \tag{6.6}$$

for any $s \in S$. It follows from (3.22, iv) that (6.5) sets up a $\mathfrak{B}$-filtered quasi-isomorphism

$$\bar{B}_r(R\Gamma \Omega^\cdot_{Z/\bar{S}}(\log(\bar{E} \cup \bar{Z}_0)) \otimes \mathscr{O}_{\bar{Z}_0})_{\bar{\sigma}_0(0), \bar{\sigma}_1(0)} \approx \bar{B}_r(R\Gamma \mathbb{C}_{Y_s})_{\sigma_0(s), \sigma_1(s)}. \tag{6.7}$$

Since the two sides of (6.7) then have isomorphic H^0's, we see that the dimension of

$$H^0(\bar{B}_r(R\Gamma \Omega^\cdot_{Z/\bar{S}}(\log(\bar{E} \cup \bar{Z}_0)) \otimes \mathscr{O}_{Z_s})_{\bar{\sigma}_0(s), \bar{\sigma}_1(s)}) \tag{6.8}$$

is independent of s. Thus, $\tilde{\mathscr{B}}_r$ is free, and its fibers are given by (6.8).

By extending (4.25) to $\bar{S}$ as

$$\tilde{\mathscr{K}}_r = \bar{B}_r(\bar{f}_* C^\cdot(\mathfrak{U}, \Omega^\cdot_{Z/S}(\log \bar{E} \cup \bar{Z}_\Sigma)); \Omega^\cdot_S(\log \Sigma))_{\bar{\sigma}_0, \bar{\sigma}_1} \tag{6.9}$$

and arguing as in (4.26)-(4.28), we obtain a logarithmic connection

$$\tilde{\delta}\colon \tilde{\mathscr{B}}_r \to \Omega^1_S(\log \Sigma) \otimes \tilde{\mathscr{B}}_r, \tag{6.10}$$

which restricts to δ over S. This shows that $\tilde{\mathscr{B}}_r$ is meromorphically commensurable with the canonical extension of $\mathscr{B}_r$. To prove that the two are equal, we must check the eigenvalues of the residues of $\tilde{\delta}$ on Σ. Since the connection is compatible with the filtration $\mathfrak{B}$, the eigenvalues are among those for

$$\bigoplus_k \mathscr{H}^0(Gr_k^{\mathfrak{B}} \mathscr{B}_r) \cong \otimes^\cdot(R^1 g_* \mathbb{C} \otimes_{\mathbb{C}} \mathscr{O}_S).$$

Under the hypothesis on g, we see that all such eigenvalues are 0. This shows that $\tilde{\mathscr{B}}_r$ is the canonical extension. □

(6.11) *Remark.* As usual, the monodromy hypothesis of (6.4) is satisfied whenever $\bar{Z}_\Sigma$ is reduced.

It is not hard to see that (6.5) is actually a filtered quasi-isomorphism with respect to $W(\bar{E})$, the weight filtration with respect to $\bar{E}$ only (cf. [23: (5.4)]). It follows that (6.7) is a bifiltered quasi-isomorphism with respect to $(W(\bar{E})*\mathfrak{B}, F)$. Thus $W(\bar{E})*\mathfrak{B}$ induces a filtration of $\tilde{\mathscr{B}}_r$ by canonical extensions of flat sub-bundles (compare the discussion following the proof of (4.22)).

(6.12) **Lemma.** *Under the assumptions of* (6.4), $Gr_k^{W(\bar{E})*\mathfrak{B}}\tilde{\mathscr{B}}_r$ *is the canonical extension of* $Gr_k^{W*\mathfrak{B}}\mathscr{B}_r$.

Let $\tilde{\mathscr{F}}$ denote the filtration of $\tilde{\mathscr{B}}_r$ induced by the Hodge filtration F of $\Omega^\bullet_Z(\log(\bar{E}\cup\bar{Z}_\Sigma))$.

(6.13) **Proposition.** i) $\tilde{\mathscr{F}}$ *is a filtration by sub-bundles, hence is given by the extended Hodge bundles. In fact,*

$$\tilde{\mathscr{F}}^p \cong \mathscr{H}^0[F^p \bar{B}_r(\bar{f}_* C^\bullet(\mathfrak{U}, \Omega^\bullet_{Z/S}(\log(\bar{E}\cup\bar{Z}_\Sigma); \mathscr{O}_S))_{\bar{\sigma}_0,\bar{\sigma}_1}];$$

ii) $\tilde{\mathscr{F}}$ *induces the extended Hodge bundles of* $Gr_k^{W(E)}\tilde{\mathscr{B}}_r$ *for all* k.

Proof. The techniques for proving such things appear already in [27: §2] and [23: (5.7)]. For (i) it is enough to show that the spectral sequence for F on

$$\bar{B}_r(R\bar{f}_*\,\Omega^\bullet_{Z/S}(\log(\bar{E}\cup\bar{Z}_0))\otimes\mathscr{O}_{Z_0})_{\bar{\sigma}_0(0),\bar{\sigma}_1(0)}$$

degenerates at E_1, for this implies that the sheaves $\mathscr{H}^0[F^p\bar{B}_r\ldots]$ above are free, and inject into $\tilde{\mathscr{B}}_r$ with free quotient. While this can be achieved via the argument of [7: (3.2.13)] (replace W there by $\mathfrak{B}$), it is more useful (for (ii)) to show that F induces the Hodge filtration of a (filtered) mixed Hodge structure.

By [23: (5.5), (5.6)], the complex

$$\Omega^\bullet_{Z/S}(\log(\bar{E}\cup\bar{Z}_0))\otimes\mathscr{O}_{Z_0} \tag{6.14}$$

has an $(F, W(\bar{E}))$-bifiltered resolution $A^\bullet$ that possesses another filtration M, such that $(A^\bullet, M, F)$ is a cohomological mixed Hodge complex filtered by $W(\bar{E})$; on $\mathbf{H}^\bullet(A^\bullet)$, this is the "limit mixed Hodge structure" for the cohomology of Y_s as $s\to 0$. Unfortunately, $A^\bullet$ does not have a multiplicative structure compatible with that of (6.14), so it is useless for bar constructions. However, the existence of a tri-filtered differential graded algebra $\tilde{A}^\bullet$ of sheaves, and a morphism of complexes $\tilde{A}^\bullet\to A^\bullet$ that is a filtered quasi-isomorphism with respect to all three filtrations, is demonstrated in [11: §7]. We now claim

(6.15) **Proposition.** *With the filtrations* $M*\mathfrak{B}$ *and* F,

$$B_r = \bar{B}_r(R\Gamma\tilde{A}^\bullet)$$

becomes a mixed Hodge complex, filtered by $W(\bar{E})*\mathfrak{B}$.

Proof of (6.15). Applying [23: (6.8)], we wish to verify an isomorphism of Hodge complexes of weight k:

$$Gr_k^{M*\mathfrak{B}}(W(\bar{E})*\mathfrak{B})_l B_r \cong \bigoplus_{j\leqq l} Gr_k^{M*\mathfrak{B}} Gr_j^{W(\bar{E})*\mathfrak{B}} B_r. \tag{6.16}$$

According to [28: (A.1)], there is a functorial decomposition of the successive quotients of a finite convolved filtration:

$$Gr_k^{W*L} V \cong \bigoplus_{i+j=k} Gr_i^W Gr_j^L V, \tag{6.17}$$

valid in any Abelian category, e.g. complexes, as should have been made clear in [28]. It is important to recall that one deduces (6.17) inductively from a functorial splitting - we assume, without loss of generality, that $W_{-1}=0$ -

$$\begin{aligned} Gr_k^{W*L} V &\cong Gr_k^{W*L} W_0 \oplus Gr_k^{W*L}(V/W_0) \\ &\cong Gr_k^L Gr_0^W V \oplus Gr_k^{\overline{W}*\bar{L}}(V/W_0), \end{aligned} \tag{6.18}$$

where $\overline{W}$ and $\bar{L}$ are the filtrations induced by W and L on V/W_0. We use (6.18) to calculate

$$Gr_k^{M*\mathfrak{B}}(W(\bar{E})*\mathfrak{B})_l B_r = (W(\bar{E})*\mathfrak{B})_l Gr_k^{M*\mathfrak{B}} B_r \hookrightarrow Gr_k^{M*\mathfrak{B}} B_r$$

to be

$$\begin{aligned} &Gr_k^M \mathfrak{B}_0 (W(\bar{E})*\mathfrak{B})_l B_r \oplus Gr_k^{M*\mathfrak{B}}[(\overline{W(E)}*\overline{\mathfrak{B}})_l (B_r/\mathfrak{B}_0)] \\ &\quad \subset Gr_k^M \mathfrak{B}_0 B_r \oplus Gr_k^{M*\mathfrak{B}}(B_r/\mathfrak{B}_0). \end{aligned} \tag{6.19}$$

With the care necessary in dealing with multiple filtrations, we will show that

$$(M*\mathfrak{B})_k[(\overline{W(E)}*\overline{\mathfrak{B}})_l(B_r/\mathfrak{B}_0)] = (\overline{M}*\overline{\mathfrak{B}})_k(\overline{W(E)}*\overline{\mathfrak{B}})_l(B_r/\mathfrak{B}_0), \tag{6.20}$$

Granting this, we deduce by induction that

$$Gr_k^{M*\mathfrak{B}}(W(E)*\mathfrak{B})_l B_r = \bigoplus_i Gr_{k-i}^M W(E)_{l-i} Gr_i^{\mathfrak{B}} B_r, \tag{6.21}$$

and thus also

$$Gr_k^{M*\mathfrak{B}} Gr_l^{W(E)*\mathfrak{B}} B_r = \bigoplus_i Gr_{k-i}^M Gr_{l-i}^{W(E)}(Gr_i^{\mathfrak{B}} B_r). \tag{6.22}$$

In view of (3.9), the desired assertion (6.16) reduces to the analogous assertion for $R\Gamma\tilde{A}^\bullet$ (see [11: (7.5.1)]) and its tensor powers (see Appendix, (6.32)).

Thus, we must verify (6.20), which is equivalent to showing that

$$(M*\mathfrak{B})_k \cap (W(E)*\mathfrak{B})_l + \mathfrak{B}_0 \supset [(M*\mathfrak{B})_k + \mathfrak{B}_0] \cap (W(E)*\mathfrak{B})_l. \tag{6.23}$$

It is helpful to realize that (6.23) does not hold in complete generality; there are counterexamples for filtrations on two-dimensional vector spaces! Also, the question is fortunately about vector spaces, and not really about complexes. We may therefore exploit the fact that $\mathfrak{B}$ has an obvious splitting (as a graded vector space) by tensor length, compatible with both M and $W(E)$, which

follows if we utilize a model of $R\Gamma\tilde{A}^{\bullet}$ that contains no non-zero exact 1-forms, and thus the relations (3.8) in the bar complex are vacuous. It is an easy exercise to see that this gives (6.23), and hence (6.22), which imparts to B_r the structure of a filtered mixed Hodge complex, as desired. □

From here, it is a standard argument (see [23: (5.7)]) to deduce (6.13), and in fact all of (1.5). We may replace $\tilde{A}^{\bullet}$ by (6.14) in discussing the filtrations F and $W(E)*\mathfrak{B}$. Thus, on

$$\tilde{\mathscr{B}}_r(0) = H^0[\bar{B}_r(R\Gamma\Omega^{\bullet}_{Z/S}(\log(\bar{E}\cup\bar{Z}_0))\otimes\mathscr{O}_{Z_0})_{\bar{\sigma}_0(0),\bar{\sigma}_1(0)}], \tag{6.24}$$

the spectral sequence for F degenerates at E_1, while that for $W(E)*\mathfrak{B}$ degenerates at E_2. From the assertion about F, one deduces that the right-hand member of (6.13, i) is a free $\mathscr{O}_S$-module that defines a sub-bundle of $\tilde{\mathscr{B}}_r$, necessarily $\tilde{\mathscr{F}}^p$ (cf., [27: p. 130]). To obtain (6.13, ii), we note that F induces on

$$Gr_l^{W(E)*\mathfrak{B}} \cong \bigoplus_j Gr_{l-j}^{W(E)}(\otimes^j[R\Gamma\Omega^{\bullet}_{Z/S}(\log(\bar{E}\cup\bar{Z}_0)\otimes\mathscr{O}_{Z_0}])[j] \tag{6.25}$$

the Hodge filtration of this limit complex. The filtration induced on

$$Gr_l^{W(E)*\mathfrak{B}}\tilde{\mathscr{B}}_r(0)$$

through the spectral sequence for $W(E)*\mathfrak{B}$ is the limit Hodge filtration for $Gr_l^{W(E)*\mathfrak{B}}\mathscr{B}_r$; from (6.15), it follows that this coincides with the filtration induced by $\tilde{\mathscr{F}}^p(0)$, as desired.

We complete our main task by showing

(6.26) **Proposition.** *On $\tilde{\mathscr{B}}_r(0)$, $M*\mathfrak{B}$ induces the weight filtration of the local monodromy relative to $W(E)*\mathfrak{B}$.*

Proof. Again, the argument (due to El Zein) goes as in [23: (5.7)]. On

$$_{W(E)*\mathfrak{B}}E_1^{-l,l+m} \cong \bigoplus_j \mathbf{H}^{m+j}(\bar{Z}_0, Gr_{l-j}^{W(E)}\otimes^j R\Gamma\Omega^{\bullet}_{Z/S}(\log(\bar{E}\cup\bar{Z}_0))\otimes\mathscr{O}_{Z_0}), \tag{6.27}$$

M (more accurately, Dec M) induces the monodromy weight filtration centered at $m+l$ (one uses elementary properties of the tensor product to reduce, in effect, to the case $j=1$). Thus the differential d_1 of (6.27) is compatible with these monodromy weights, and in fact is strictly compatible, since M is the weight filtration of a mixed Hodge structure. One deduces by linear algebra that M induces the monodromy weight filtration, again centered at $m+l$, on the cohomology

$$_{W(E)*\mathfrak{B}}E_2^{-l,l+m} \cong Gr_l^{W(E)*\mathfrak{B}}H^m(B_r).$$

of (6.27). This shows that M induces the relative weight filtration we have been seeking. □

(6.28) *Remark.* In the case of a canonical variation, we get from (6.3) and (6.4) that

i) $\tilde{\mathscr{B}}_r \cong \mathscr{H}^0\bar{B}_r(R\bar{f}_*\Omega^{\bullet}_{\bar{X}\times\bar{S}/\bar{S}}(\log[(D\times\bar{S})\cup(\bar{X}\times\Sigma)]);\mathscr{O}_{\bar{S}})_{\bar{\sigma}_0,\bar{\sigma}_1}$,

where $\bar{S}=\bar{X}\times\bar{X}$ here. It is not hard to see that $\tilde{\mathscr{B}}_r$ can be realized in this case from a somewhat smaller complex, viz.

ii) $\tilde{\mathscr{B}}_r\cong\mathscr{H}^0\bar{B}_r(R\bar{f}_*\Omega^\cdot_{\bar{X}\times\bar{S}/\bar{S}}(\log(D\times\bar{S});\mathcal{O}_{\bar{S}})_{\bar{\sigma}_0,\bar{\sigma}_1}$.

The quickest way to see this is to take successive quotients with respect to $\mathfrak{B}$; in both cases, we get the canonical (=trivial here) extensions of cohomology bundles associated to the product $X\times S$. Using (ii), one could bypass the use of $\tilde{A}^\cdot$ in deriving the limit mixed Hodge theory (6.15) in this case, but one could not get at the connection (4.29).

It remains to check that the construction (5.21) always yields a (unipotent) variation of mixed Hodge structure that is good at infinity, i.e., satisfies (1.5, i, ii). Let

$$\theta\colon\ \mathbb{C}\pi_1(X,x)/J^{r+1}\to\mathrm{End}(V) \tag{6.29}$$

be a mixed Hodge theoretic representation, with respect to the Hodge filtration F_x on the left-hand member of (6.29), and F_V of V. On any open subset of X, on which the canonical variation with basepoint x can be expressed by $F_y=h(y)F_x$, where $h(y)$ has values in the Malčev group (see §5), the Hodge filtration for the constructed variation is given by

$$\Phi(y)=\theta(h(y))F_V. \tag{6.30}$$

We work in a distinguished polydisc $(\Delta^*)^n$ about a point of D, and let $\bar{N}_1,\dots,\bar{N}_n\in\mathrm{End}(V)$ denote the local monodromy logarithms. These are in $W_{-2}\mathrm{End}(V)$ since they are the images under θ of those for $\mathbb{C}\pi_1(X)/J^{r+1}$; call the latter $N_1,\dots,N_n$. This gives (1.5, ii) immediately.

As in [20: (4.7)], we consider the twisted classifying mapping (i.e., the one adapted to the canonical extension; see [23: p. 510]):

$$\Psi(\mathbf{t})=\exp(-\sum_j z_j\bar{N}_j)\,\Phi(\mathbf{t}). \tag{6.31}$$

For $\mathbf{t}$ in a contractible subset U of $(\Delta^*)^n$, we can write

$$\begin{aligned}\Psi(\mathbf{t})&=\exp(-\textstyle\sum z_j\theta(N_j))\,\theta(h(\mathbf{t}))F_V\\&=\theta[\exp(-\textstyle\sum z_jN_j)h(\mathbf{t})]F_V.\end{aligned}$$

Since the canonical variation is good, we can choose $h(\mathbf{t})$ so that

$$\lim_{U\ni\mathbf{t}\to\mathbf{0}}\exp(-\textstyle\sum z_jN_j)h(\mathbf{t})$$

exists. As we can cover $(\Delta^*)^n$ by a finite number of such U's, it follows then that

$$\lim_{\mathbf{t}\to\mathbf{0}}\Psi(\mathbf{t})\quad\textit{exists}.$$

In other words, the Hodge filtration extends inside $\tilde{V}$, as desired for (1.5, i).

(6.32) *Remark.* The arguments of this section apply equally well to proving that, when the fibers of $f\colon Z\to S$ are simply connected and the local mono-

dromy of $R^\cdot f_* \mathbb{C}$ is unipotent, the canonical extension of the holomorphic vector bundle

$$\{H^k(P_{\sigma_0(s),\sigma_1(s)} Z_s; \mathbb{C})\}_{s\in S} \otimes \mathcal{O}_S$$

can be realized as

$$\mathscr{H}^k[\bar{B}\bar{f}_* C^\cdot(\mathfrak{U}, \Omega^\cdot_{Z/S}(\log \bar{E} \cup \bar{Z}_S); \mathcal{O}_S)_{\sigma_0,\sigma_1}].$$

This is used in [11: 7.5] to prove the existence of a limit mixed Hodge structure on the homotopy groups $\pi_k(Z_s, \sigma(s))$ in the case when $\sigma_0 = \sigma_1$.

Appendix to § 6

In the course of the proof of (6.16), we needed the following:

(6.33) **Lemma.** *Let $A_1^\cdot$ and $A_2^\cdot$ be complexes of vector spaces, each equipped with a pair of increasing filtrations, denoted M and W, that are bounded from below. Assume that*

$$Gr_k^M W_l A_m^\cdot \cong \bigoplus_{j\leqq l} Gr_k^M Gr_j^W A_m^\cdot \quad (m=1,2).$$

Give $A^\cdot = A_1^\cdot \otimes A_2^\cdot$ the usual filtrations M and W defined on a tensor product. Then

$$Gr_k^M W_l A^\cdot = \bigoplus_{j\leqq l} Gr_k^M Gr_j^W A^\cdot.$$

Proof. We characterize the above isomorphism in terms of splittings. Let

$$A^i = \oplus A^i(k)$$

be any splitting of the filtration M, and let

$$A^i(k) = \bigoplus_{k,j} A^i(k,j)$$

be the lifting of the decomposition of W on $Gr_k^M A^\cdot$. This gives

$$M_k A^i = \bigoplus_{\substack{r\leqq k \\ j}} A^i(r,j), \qquad W_l A^i = \bigoplus_{\substack{r \\ j\leqq l}} A^i(r,j). \tag{6.34}$$

and

$$Gr_k^M Gr_j^W A^i \cong A^i(k,j) \tag{6.35}$$

By construction,

$$dA^i(k,l) \subset A^{i+1}(k,l) \oplus \bigoplus_{\substack{r\leqq k-1 \\ j\leqq l}} A^{i+1}(r,j). \tag{6.36}$$

Conversely, the existence of a simultaneous splitting (6.34) of M and W on the graded vector space $A^\cdot$ for which (6.36) holds, implies that

$$Gr_k^M W_l A^\cdot \cong \bigoplus_{j\leqq l} Gr_k^M Gr_j^W A^\cdot$$

as complexes. With this said, the lemma is now easy to prove. Let $\oplus A_1^i(k,j)$ and $\oplus A_2^i(k,j)$ be the above splittings for $A_1^\cdot$ and $A_2^\cdot$. For $A^\cdot = A_1^\cdot \otimes A_2^\cdot$, we put

$$A^i(k,j) = \bigoplus_{\substack{p+q=i \\ a+b=k \\ m+n=j}} [A_1^p(a,m) \otimes A_2^q(b,n)]. \tag{6.37}$$

This clearly determines a splitting of M and W on $A^\cdot$, so it remains to check (6.36), which is obvious from the definition of d on the tensor product and the fact that (6.36) holds for $A_1^\cdot$ and $A_2^\cdot$. □

§ 7. Monodromy in a unipotent variation

Let $\mathbb{V}$ be a local system underlying a good unipotent variation of mixed Hodge structure on $X = \bar{X} - D$ (where $\bar{X}$ is a compact Kähler manifold and D is a divisor with normal crossings), defined over a real field (or ring) $\mathbb{F}$. For any $x \in X$, the monodromy representation on the fiber V_x of $\mathbb{V}$ at x,

$$\rho\colon \pi_1(X,x) \to \mathrm{Aut}(V_x),$$

induces

$$\bar{\rho}\colon \mathbb{C}\pi_1(X,x) \to \mathrm{End}(V_x),$$

which descends (see (1.4)) to define

$$\mathbb{C}\pi_1(X,x)/J^{r+1} \xrightarrow{\bar{\rho}} \mathrm{End}(V_x) \tag{7.1}$$

for some $r \geqq 0$. Both sides of (7.1) are endowed with mixed Hodge structures, defined over $\mathbb{F}$: the left-hand side's is given by (3.21), and $\mathrm{End}(V_x)$ inherits one from V_x. We prove in this section:

(7.2) **Theorem.** *The mapping $\bar{\rho}$ of* (7.1) *is a morphism of mixed Hodge structures.*

We begin by rewriting (7.1) as

$$\bar{\rho} \in (\mathbb{C}\pi_1(X,x)/J^{r+1})^* \otimes \mathrm{End}(V_x),$$

and interpreting this, via (1.14), as:

$$\bar{\rho} \in H^0(\bar{B}_r(R\Gamma\mathbb{C}_X)_{x,x} \otimes \mathrm{End}(V_x)). \tag{7.3}$$

The assertion of (7.2) becomes less mysterious, and indeed transparent, once we find a natural 0-cocycle that represents (7.3).

We must first go through some generalities on vector bundles and connections. Let Z be a complex manifold. Recall that a holomorphic vector bundle with fiber V is given (up to isomorphism) by an element

$\xi \in H^1(Z, \mathcal{O}(\mathrm{Aut}(V)))$. Explicitly, if the bundle is trivialized over the elements of an open covering $\mathfrak{U} = \{O_\alpha\}$ of Z, then g is represented by the Cech cochain

$$(\xi_{\alpha\beta}) \in C^1(\mathfrak{U}, \mathcal{O}_Z(\mathrm{Aut}(V))) \tag{7.4}$$

of transition functions, which satisfies the cocycle condition

$$\delta\xi = 0\colon\ \xi_{\alpha\beta}\xi_{\beta\gamma} = \xi_{\alpha\gamma} \quad \text{on } O_\alpha \cap O_\beta \cap O_\gamma. \tag{7.5}$$

We define $\varphi \in C^1(\mathfrak{U}, \mathcal{O}_Z(\mathrm{End}(V)))$ by

$$\varphi_{\alpha\beta} = \xi_{\alpha\beta} - I. \tag{7.6}$$

One easily checks that (7.5) is equivalent to:

$$\varphi_{\alpha\beta} - \varphi_{\alpha\gamma} + \varphi_{\beta\gamma} + \varphi_{\alpha\beta}\varphi_{\beta\gamma} = 0; \tag{7.7}$$

in other words,

$$\delta\varphi + \varphi \cup \varphi = 0.^{15} \tag{7.8}$$

Assume that a connection ∇ is given on the bundle. The connection can be described explicitly by a compatible system of connection matrices ω_α for the covering $\mathfrak{U}$:

$$\nabla v = dv - v \cdot \omega_\alpha \quad \text{in } O_\alpha, \tag{7.9}$$

$$\xi_{\alpha\beta}\omega_\beta = \omega_\alpha \xi_{\alpha\beta} + d\xi_{\alpha\beta} \quad \text{in } O_\alpha \cap O_\beta. \tag{7.10}$$

This defines a cochain

$$\omega \in C^0(\mathfrak{U}, \Omega^1_Z \otimes_{\mathbb{C}} \mathrm{End}(V)), \tag{7.11}$$

such that

$$\delta\omega + (\varphi \cup \omega) + (\omega \cup \varphi) - d\varphi = 0 \quad \text{in } C^1(\mathfrak{U}, \Omega^1_Z \otimes_{\mathbb{C}} \mathrm{End}(V)). \tag{7.12}$$

The connection is flat (integrable) if and only if

$$d\omega + \omega \wedge \omega = 0 \quad \text{in } C^0(\mathfrak{U}, \Omega^2_Z \otimes \mathrm{End}(V)) \tag{7.13}$$

(zero curvature). The following combines (7.8), (7.12) and (7.13):

(7.14) **Proposition.** *Let*

$$\varphi \in C^1(\mathfrak{U}, \mathcal{O}_Z \otimes \mathrm{End}(V)), \quad \omega \in C^0(\mathfrak{U}, \Omega^1_Z \otimes \mathrm{End}(V)),$$

and assume that $I + \varphi$ takes values in $\mathrm{Aut}(V)$. *The degree one cochain $\tau = \varphi \oplus \omega$ in the double complex $C^\cdot(\mathfrak{U}, \Omega^\cdot_Z)$ comes from a flat holomorphic vector bundle with fiber V if and only if*

$$D\tau + \tau \cup \tau = 0.$$

[15] As usual, the cup product entails multiplication in the Cech sense, and ordinary composition in $\mathrm{End}(V)$

(Here, D is the total differential in the double complex; sign conventions are as in [1: §14].)

Let $h \in C^0(\mathfrak{U}, \mathcal{O}_Z(\mathrm{Aut}(V)))$ define a change of trivialization; this gives rise to a new cocycle

$$\xi' \in C^1(\mathfrak{U}, \mathcal{O}_Z(\mathrm{Aut}(V))),$$

which satisfies the relation

$$h_\alpha \xi'_{\alpha\beta} = \xi_{\alpha\beta} h_\beta \quad \text{in } O_\alpha \cap O_\beta. \tag{7.15}$$

We get $\tau' = \varphi' \oplus \omega'$ as before. Set

$$\psi = h - I \in C^0(\mathfrak{U}, \mathcal{O}_Z \otimes \mathrm{End}(V)). \tag{7.16}$$

We leave it to the reader to verify:

(7.17) **Proposition.** $\tau' - \tau = D\psi + \tau \cup \psi - \psi \cup \tau'$.

(7.18) *Remark.* Separating (7.17) into components, we are asserting:

i) $\varphi' - \varphi = \delta\psi + \varphi \cup \psi - \psi \cup \varphi'$,
ii) $\omega' - \omega = d\psi + \omega \cup \psi - \psi \cup \omega'$.

For reasons to become apparent, one calls, given a differential graded algebra $A^\cdot$, an element τ of $A^1 \otimes \mathrm{End}(V)$ a *twisting cochain* if $d\tau + \tau \cdot \tau = 0$. Also, for any $\tau \in A^1 \otimes \mathrm{End}(V)$ one defines its *formal transport*

$$T(\tau) \in \bar{B}_\infty(A^\cdot)^0 \otimes \mathrm{End}(V),$$

where the sub-∞ indicates that we now allow expressions of infinite length in (3.5), by the formula

$$T(\tau) = 1 + \sum_{n \geq 1} [\tau]_n, \tag{7.19}$$

where

$$[\tau]_n = \underbrace{[\tau|\ldots|\tau]}_{n \text{ times}} \tag{7.20}$$

(7.21) **Proposition.** i) *The formal transport $T(\tau)$ is a cocycle in $\bar{B}_\infty(A^\cdot) \otimes \mathrm{End}(V)$ if and only if τ is a twisting cochain.*

ii) *If τ and τ' are related by* (7.17)*, then $T(\tau)$ and $T(\tau')$ are cohomologous.*

Proof. By (3.6), we have

$$\begin{aligned} dT(\tau) &= -\sum_{n \geq 1} \Big[\sum_{i+j=n-1} ([\tau]_i | [d\tau] | [\tau]_j) + \sum_{i+j=n-2} ([\tau]_i | \tau \cdot \tau | [\tau]_j) \Big] \\ &= -\sum_{m \geq 0} \sum_{i+j=n} ([\tau]_i | (d\tau + \tau \cdot \tau) | [\tau]_j), \end{aligned}$$

from which we obtain (i). For (ii), let

$$S(\tau, \tau') = \sum_{i,j \geq 0} ([\tau]_i | \psi | [\tau']_j) \in \bar{B}_\infty(A^\cdot)^{-1} \otimes \mathrm{End}(V),$$

with ψ as in (7.16). One checks directly that

$$T(\tau) - T(\tau') = dS(\tau, \tau').$$

(7.22) **Corollary.** *The twisting cochains* $\tau \in \bar{B}_\infty(C^\bullet(\mathfrak{U}, \Omega_Z^\bullet) \otimes \mathrm{End}(V))$ *that define a given flat holomorphic vector bundle on* Z *determine a single cohomology class*

$$\langle T(\tau) \rangle \in H^0(\bar{B}_\infty(R\Gamma \mathbb{C}_Z)) \otimes \mathrm{End}(V).$$

(7.23) *Remark.* One readily checks that the entire discussion from (7.4) to (7.22) goes through verbatim in the C^∞ category. We will use this variant without further comment.

Assume next that the flat bundle has unipotent monodromy. Then the sum in (7.19) is actually a finite sum, and therefore

$$\langle T(\tau) \rangle \in H^0(\bar{B}_r(R\Gamma \mathbb{C}_Z)) \otimes \mathrm{End}(V) \tag{7.24}$$

for some $r \in \mathbb{N}$.

(7.25) **Proposition.** *If* τ *is a twisting cochain for a flat vector bundle with unipotent monodromy, then* (7.24) *gives the monodromy representation of* $\mathbb{C}\pi_1(Z)/J^{r+1}$, *and more generally the parallel transport for* $H_0(P_{z_0, z_1} Z, \mathbb{C})/J^{r+1}$.

Proof. A flat bundle with unipotent monodromy admits a filtration by flat subbundles, such that the successive quotients are (flatly) trivial[16]. It follows that the bundle is trivial as a C^∞ vector bundle. We compute with a global C^∞ frame, with respect to which τ equals

$$\omega \in H^0(Z, \mathscr{E}_Z^1 \otimes \mathrm{End}(V)),$$

the global connection form. That $T(\tau)$ defines parallel transport follows from the classical formula for parallel translation (see [12]).

(7.26) *Remark.* Without appealing to a C^∞ trivialization of the bundle, one can prove Proposition (7.25) by pulling back to its unipotent frame bundle $U\Phi \to Z$. There, the bundle admits a tautological holomorphic global frame, while $\pi_1(Z) \to \pi_1(U\Phi)$ is an isomorphism.

Let W and F now denote the weight and Hodge filtrations respectively at the basepoint in X. One can identify local frames of $\tilde{\mathscr{V}}$ adapted to its filtrations with local V-valued functions adapted to W and F. (It may help at this point to recall (5.11, v).) Also, by working on $\bar{X}$, we can keep track of the singularities of the connection.

(7.27) **Proposition.** *Let* $\tilde{\mathscr{V}}$ *be the canonical extension to* $\bar{X}$ *of the bundle on* X *underlying a good unipotent variation of mixed Hodge structure. If* τ *is computed with respect to local unipotent Hodge frames of* $\tilde{\mathscr{V}}$ *for the covering* $\bar{\mathfrak{U}}$ *of* $\bar{X}$, *we have*

$$\tau \in F^0 W_{-1} C^\bullet(\bar{\mathfrak{U}}, \Omega_{\bar{X}}^\bullet(\log D) \otimes \mathrm{End}\, V)^1.$$

[16] e.g., the weight filtration in a unipotent variation of mixed Hodge structure

Proof. Using the notation from (7.14), we wish to verify

i) $\varphi \in F^0 W_{-1} C^1(\bar{\mathfrak{U}}, \mathcal{O}_{\bar{X}} \otimes \mathrm{End}(V))$,

ii) $\omega \in F^0 W_{-1} C^0(\bar{\mathfrak{U}}, \Omega^1_{\bar{X}}(\log D) \otimes \mathrm{End}(V))$.

The first statement follows directly, since φ gives a change of basis between unipotent Hodge frames. In (ii), that $\omega \in F^0$ is an immediate consequence of horizontality (5.6). To see that $\omega \in W_{-1}$, we note that the connection transformations always take values in $W_{-1}\mathrm{End}(V)$; near D, where ω picks up a logarithmic pole, its residues are actually in $W_{-2}\mathrm{End}(V)$, as follows from (1.5, ii). □

(7.28) **Corollary.** *Under the hypotheses of* (7.27),

$$\langle T(\tau) \rangle \in F^0 W_0 [H^0(\bar{B}_r(R\Gamma \mathbb{C}_X)) \otimes \mathrm{End}(V)].$$

In other words, the parallel transport

$$H_0(P_{x,y} X, \mathbb{C}) \to \mathrm{Hom}(V_x, V_y)$$

is a morphism of mixed Hodge structures.

(7.29) *Remark.* i) By a similar argument, one can reprove the implication b) ⇒ a) of [13: (6.4)], which asserts that for a unipotent flat holomorphic bundle on a smooth variety, with *holomorphically* trivial canonical extension, the monodromy is trivial on

$$(F * J)^1 = F^0 \cap J + F^{-1} \cap J^2 + \ldots .$$

ii) It was tempting to try to prove (7.28) along the lines of (7.26). This does not work, because one must also compactify $U\Phi$ in the vertical direction in order to discuss mixed Hodge theory, and the tautological frames then have essential singularities at infinity.

§ 8. Rigidity

Let $X = \bar{X} - D$, where $\bar{X}$ is a compact Kähler manifold and D is a divisor with normal crossings in $\bar{X}$. We complete our program by establishing a "rigidity theorem" for good unipotent variations of mixed Hodge structure on X; that is, the variation of mixed Hodge structure is uniquely determined by the underlying W-filtered local system and the Hodge filtration at one point. By a standard argument (cf. [23: (4.20)], [20: (7.24)]), it is enough to prove:

(8.1) **Proposition.** *Let* $\mathbb{V}$ *underlie a good unipotent variation of mixed Hodge structure on* X. *Then for any* $x \in X$, $H^0(X, \mathbb{V})$ *is a mixed Hodge substructure of the fiber* V_x, *via the evaluation mapping at* x.

The case where X is a curve (but $\mathbb{V}$ need not be unipotent) is carried out in [23: § 4]. As was remarked earlier, at the end of Sect. 1, we can use that case to get the desired rigidity result without further work, by merely restricting to curves, *if* we know that any two points of X lie on an algebraic curve in X, e.g. if X is an algebraic variety. However, there even exist complex tori that have

no analytic subvarieties other than points, so a direct argument is needed for the general case.

As in [23], we prove (8.1) by constructing a cohomological mixed Hodge complex for $\mathbb{V}$, compatible with restriction mappings to the points of X. The fact that in a unipotent variation, the relative weight filtration M is given by W itself (recall (1.5)) allows for a rather simple construction.

Recall that $H^*(X, \mathbb{V})$ can be computed as the hypercohomology of the complex of sheaves on $\bar{X}$

$$\Omega^*_{\bar{X}}(\log D) \otimes \tilde{\mathscr{V}}, \tag{8.2}$$

for the latter is quasi-isomorphic to $Rj_* \mathbb{V}$, where j denotes the inclusion mapping of X in $\bar{X}$ (see [6: p. 103]). We define filtrations F (decreasing) and $\mathfrak{M}$ (increasing) on (8.2) as follows. For F, we take the usual thing (cf. [23: (4.10)], [26: (1.12)]):

$$F^p(\Omega^i_{\bar{X}}(\log D) \otimes \tilde{\mathscr{V}}) = \Omega^i_{\bar{X}}(\log D) \otimes \tilde{\mathscr{F}}^{p-i}. \tag{8.3}$$

To define $\mathfrak{M}$, we adapt [23: (4.9)] by putting

$$\mathfrak{M}_k = \{\varphi \in \Omega^*_{\bar{X}}(\log D) \otimes \tilde{\mathscr{W}}_k\text{: all } l\text{-fold residues of } \varphi \text{ lie in } \tilde{\mathscr{W}}_{k-l}\}, \tag{8.4}$$

where $\tilde{\mathscr{W}}_k$ is the canonical extension bundle associated to the local system $\mathbb{W}_k$ on X. To explain the condition in (8.4), one lets $\tilde{D}^{(l)}$ denote the disjoint union of the l-fold intersections of the components of D, and

$$a^{(l)}\colon \tilde{D}^{(l)} \to \bar{X}$$

the tautological mapping. Then

$$Gr_l^W \Omega^*_{\bar{X}}(\log D) \cong a^{(l)}_* \Omega^*_{\tilde{D}^{(l)}}[-l]$$

via l-fold residues. This gives a residue mapping

$$W_l \Omega^*_{\bar{X}}(\log D) \otimes \tilde{\mathscr{W}}_k \to a^{(l)}_*(\Omega^*_{\tilde{D}^{(l)}} \otimes a^{(l)*} \tilde{\mathscr{W}}_k)[-l]. \tag{8.5}$$

Thus, a vector bundle-valued form is in $\mathfrak{M}_k$ if and only if the coefficients of its terms with l logarithmic singularities lie in $\tilde{\mathscr{W}}_{k-l}$.

Clearly, $\mathfrak{M}_k$ is closed under d; there is a corresponding filtration of $Rj_* \mathbb{V}$, showing that the filtration induced on cohomology has the same field of definition as $\mathbb{V}$. Also, there is a bifiltered morphism onto V_x for any $x \in X$.

(8.6) **Proposition.** *If $(\mathbb{V}, W, \mathscr{F})$ is a good unipotent variation of mixed Hodge structure, defined over $\mathbb{F}$, then*

$$(\Omega^*_{\bar{X}}(\log D) \otimes \tilde{\mathscr{V}}, \mathfrak{M}, F)$$

is a cohomological mixed Hodge complex, defined over $\mathbb{F}$ and filtered by W.

Proof. We compute $Gr_k^{\mathfrak{M}}$. Consider the diagram

$$\begin{array}{ccc} \mathfrak{M}_k \subset \Omega^*_{\bar{X}}(\log D) \otimes \tilde{\mathscr{W}}_k & \xrightarrow{\psi} & a^{(1)}_*(\Omega^*_{\tilde{D}^{(1)}}(\log D_1) \otimes a^{(1)*} \tilde{\mathscr{W}}_k)[-1] \\ \downarrow \pi & & \\ \Omega^*_{\bar{X}}(\log D) \otimes_{\mathscr{O}_{\bar{X}}} (\tilde{\mathscr{W}}_k / \tilde{\mathscr{W}}_{k-1}) \cong \Omega^*_{\bar{X}}(\log D) \otimes_{\mathbb{C}} Gr_k^W \mathbb{V}, & & \end{array} \tag{8.7}$$

where $D_1 = a^{(1)*}(a^{(2)}_* \tilde{D}^{(2)})$ and ψ is the residue mapping. We will apply the following elementary lemma:

(8.8) **Lemma.** *In an abelian category, let* $\varphi_1 \colon A \to B_1$ *and* $\varphi_2 \colon A \to B_2$ *be morphisms, and put*

$$\varphi = (\varphi_1, \varphi_2) \colon A \to B_1 \oplus B_2 .$$

Assume that $\operatorname{im}\varphi = \operatorname{im}\varphi_1 \oplus \operatorname{im}\varphi_2$. *Then* φ *sets up an isomorphism*

$$\frac{A}{(\ker\varphi_1)\cap(\ker\varphi_2)} \xrightarrow{\sim} (\operatorname{im}\varphi_1)\oplus(\operatorname{im}\varphi_2).$$

We take $A = \mathfrak{M}_k$, φ_1 the restriction of π to $\mathfrak{M}_k$, and φ_2 the restriction of ψ followed by the quotient modulo $\psi(\mathfrak{M}_{k-1})$. It is not hard to see that, in fact,

$$\psi(\mathfrak{M}_i) = a^{(1)}_*(\Omega^\bullet_{\tilde{D}^{(1)}}(\log D_1) \otimes \mathfrak{M}^{(1)}_{i-1})[-1], \tag{8.9}$$

where $\mathfrak{M}^{(1)}$ is the filtration analogous to (8.4) for $a^{(1)*}\tilde{\mathcal{V}}$ on $\tilde{D}^{(1)}$. Since $a^{(1)}$ is a finite mapping onto its image, we get

$$\operatorname{im}\varphi_2 = a^{(1)}_*(\Omega^\bullet_{\tilde{D}^{(1)}}(\log D_1) \otimes Gr^{\mathfrak{M}^{(1)}}_{k-1}(a^{(1)*}\tilde{\mathcal{V}}))[-1].$$

Of course, φ_1 maps onto $\Omega^\bullet_{\bar{X}} \otimes Gr^W_k \mathbb{V}$, and

$$\ker\varphi_2 \supset \ker\varphi_1 = \mathfrak{M}_{k-1}.$$

The hypothesis about the independence of the images is also quite obvious, so (8.8) yields

$$Gr^{\mathfrak{M}}_k \cong (\Omega^\bullet_{\bar{X}} \otimes Gr^W_k \mathbb{V}) \oplus (a^{(1)}_*(\Omega^\bullet_{\tilde{D}^{(1)}}(\log D_1) \otimes Gr^{\mathfrak{M}^{(1)}}_{k-1}))[-1].$$

This gives, by induction,

$$Gr^{\mathfrak{M}}_k \cong \bigoplus_{l \geq 0} (a^{(l)}_* \Omega^\bullet_{\tilde{D}^{(l)}} \otimes Gr^W_{k-l} \mathbb{V})[-l] \tag{8.10}$$

in which $\tilde{D}^{(0)} = \bar{X}$, etc., and we extend $Gr^W_{k-l}\mathbb{V}$ across D as a constant sheaf. It is clear now that $Gr^{\mathfrak{M}}_k$ is a cohomological Hodge complex of weight k, since each summand is one. Moreover, the filtration induced by W on (8.10) is just given by truncating the sum, so

$$Gr^{\mathfrak{M}}_k W_l \cong \bigoplus_{j \leq l} Gr^{\mathfrak{M}}_k Gr^W_j \mathbb{V};$$

by [23: (6.8)], the mixed Hodge complex is filtered by W. □

Having finished the last step of the proof of Theorem (1.6), we sum up briefly what we have shown:

i) For any good unipotent variation of mixed Hodge structure, the monodromy representation (7.1) is a morphism of mixed Hodge structures.

ii) When one applies Construction (5.21) to (7.1), one recovers the original variation.

References

1. Bott, R., Tu, L.: Differential Forms in Algebraic Topology. GTM, vol. 82. Berlin-Heidelberg-New York: Springer 1982
2. Carlson, J.: Mixed Hodge structure over the integers. (Preprint 1979)
3. Cattani, E., Kaplan, A.: Polarized mixed Hodge structure and the monodromy of a variation of Hodge structure. Invent. Math. **67**, 101–115 (1982)
4. Chen, K.-T.: Reduced bar constructions on de Rham complexes. In: Heller, A., Tierney, M. (eds) Algebra, topology and category theory, pp. 19–32. New York: Academic Press 1976
5. Chen, K.-T.: Iterated path integrals. Bull. Am. Math. Soc., New Ser. **83**, 831–879 (1977)
6. Deligne, P.: Equations différentielles a points singuliers réguliers. Lect. Notes Math. vol. 163. Berlin-Heidelberg-New York: Springer 1970
7. Deligne, P.: Théorie de Hodge, II. Publ. Math., Inst. Hautes Etud. Sci. **40**, 5–57 (1971)
8. Deligne, P.: Théorie de Hodge, III. Publ. Math., Inst. Hautes Etud. Sci. **44**, 5–77 (1974)
9. Deligne, P.: Letter to Wojtkowiak, October 25, 1983
10. El Zein, F.: Complexe de Hodge mixte filtré. C. R. Acad. Sci. Paris **295**, 669–672 (1982)
11. Hain, R.: The de Rham homotopy theory of complex algebraic varieties. J. K-Theory (to appear)
12. Hain, R.: The geometry of the mixed Hodge structure on the fundamental group. Algebraic geometry, Bowdoin, Am. Math. Soc.: Proc. Symp. Pure Math. (to appear)
13. Hain, R.: On a generalization of Hilbert's 21st problem. Ann. Sci. Ec. Norm. Super., IV. Ser. in press (1987)
14. Hwang-Ma, S.-Y.: Periods of iterated integrals of holomorphic forms on a compact Riemann surface. Trans. Am. Math. Soc. **264**, 295–300 (1981)
15. Katz, N., Oda, T.: On the differentiation of de Rham cohomology classes with respect to parameters. J. Math. Kyoto Univ. **8**-2, 199–213 (1968)
16. Morgan, J.: The algebraic topology of smooth algebraic varieties. Publ. Math., Inst. Hautes Etud. Sci. **48**, 137–204 (1978)
17. Parsin, A.: A generalization of the Jacobian variety. Am. Math. Soc. Transl. **84**, 187–196 (1969)
18. Passman, D.: The algebraic structure of group rings. New York: John Wiley 1977
19. Quillen, D.: Rational homotopy theory. Ann. Math. **90**, 205–295 (1969)
20. Schmid, W.: Variation of Hodge structure: the singularities of the period mapping. Invent. Math. **22**, 211–319 (1973)
21. Stallings, J.: Quotients of the powers of the augmentation ideal in a group ring. In: Neuwirth, L. (ed.), Knots, groups and 3-manifolds, Papers dedicated to the memory of R.H. Fox. Annals of Math. Studies **84**, Princeton 1975
22. Steenbrink, J.: Limits of Hodge structures. Invent. Math. **31**, 229–257 (1976)
23. Steenbrink, J., Zucker, S.: Variation of mixed Hodge structure, I. Invent. Math. **80**, 489–542 (1985)
24. Usui, S.: Variation of mixed Hodge structures arising from family of logarithmic deformations. Ann. Sci. Ec. Norm Super., IV. Ser. **16**, 91–107 (1983)
25. Usui, S.: Variation of mixed Hodge structure arising from family of logarithmic deformations II: Classifying space. Duke Math. J. **51**, 851–875 (1984); supplement by Saito, Shimizu and Usui **52**, 529–534 (1985)
26. Zucker, S.: Hodge theory with degenerating coefficients: L_2 cohomology in the Poincaré metric. Ann. Math. **109**, 415–476 (1979)
27. Zucker, S.: Degeneration of Hodge bundles (after Steenbrink). In: Topics in transcendental algebraic geometry. Ann. Math. Studies **106**, 121–141 (1984)
28. Zucker, S.: Variation of mixed Hodge structure, II. Invent. Math. **80**, 543–565 (1985)
29. Carlson, J.: Extensions of mixed Hodge structures. In: Algebraic geometry, Angers, 1979, pp. 107–127. Sijthoff & Noordhoff, Alphen a.d. Rijn, the Netherlands 1980
30. Kashiwara, M.: Variation of mixed Hodge structure. (Preprint 1986)
31. Quillen, D.: On the associated graded ring of a group ring. J. Algebra **10**, 411–418 (1968)
32. Hain, R., Zucker, S.: A guide to unipotent variations of mixed Hodge structure. Proc. Conf. on Hodge Theory, Sant Cugat, 1985 (to appear)

Oblatum 6-I-86 7-VIII-1986

Invent. math. 88, 125–151 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Rigidity and the lower bound Theorem 1

Gil Kalai

Institute of Mathematics, Hebrew University, Jerusalem, Israel and
Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Summary. For an arbitrary triangulated $(d-1)$-manifold without boundary C with f_0 vertices and f_1 edges, define $\gamma(C)=f_1-df_0+\binom{d+1}{2}$. Barnette proved that $\gamma(C)\geqq 0$. We use the rigidity theory of frameworks and, in particular, results related to Cauchy's rigidity theorem for polytopes, to give another proof for this result. We prove that for $d\geqq 4$, if $\gamma(C)=0$ then C is a triangulated sphere and is isomorphic to the boundary complex of a stacked polytope. Other results: (a) We prove a lower bound, conjectured by Björner, for the number of k-faces of a triangulated $(d-1)$-manifold with specified numbers of interior vertices and boundary vertices. (b) If C is a simply connected triangulated d-manifold, $d\geqq 4$, and $\gamma(\mathrm{lk}(v, C))=0$ for every vertex v of C, then $\gamma(C)=0$. ($\mathrm{lk}(v, C)$ is the link of v in C.) (c) Let C be a triangulated d-manifold, $d\geqq 3$. Then $\mathrm{skel}_1(\Delta_{d+2})$ can be embedded in $\mathrm{skel}_1(C)$ iff $\gamma(C)>0$. (Δ_d is the d-dimensional simplex.) (d) If P is a 2-simplicial d-polytope then $f_1(P)\geqq df_0(P)-\binom{d+1}{2}$. Related problems concerning pseudomanifolds, manifolds with boundary and polyhedral manifolds are discussed.

1. The lower bound theorem

Barnette's *lower bound theorem* (LBT) ([9, 10]) asserts that if P is a simplicial d-polytope with n vertices, then $f_k(P)$, the number of k-dimensional faces of P, satisfies the inequality $f_k(P)\geqq\varphi_k(n, d)$, where

$$\varphi_k(n,d)=\begin{cases}\binom{d}{k}n-\binom{d+1}{k+1}k & \text{for } 1\leqq k\leqq d-2\\ (d-1)n-(d+1)(d-2) & \text{for } k=d-1.\end{cases} \tag{1.1}$$

Barnette's theorem settled an old conjecture in the theory of convex polytopes. (See [31, pp. 183–188] for the history of this conjecture.)

The main purpose of this paper is to show the connection between the lower bound theorem and the *rigidity theory of frameworks*. The basic idea is quite simple. Let P be a simplicial d-polytope, $d \geqq 3$, with n vertices. The inequality $f_1(P) \geqq \varphi_1(n,d)$ follows from the fact that P is *rigid*. This means that every small perturbation of the vertices of P, which does not change the length of the edges of P, is induced by an affine rigid motion of $\mathbb{R}^d$. The crucial result is Cauchy's rigidity theorem ([22]) which gives the rigidity of simplicial 3-polytopes. The result for higher dimensions follows by a simple inductive argument. (See [66, 60, p. 119]). We use rigidity theory to prove several extensions of the lower bound theorem and to study the cases of equality.

Barnette's inequality $f_k(P) \geqq \varphi_k(n,d)$ is sharp, and equality holds for every $1 \leqq k < d$ if P belongs to the family of *stacked* polytopes defined as follows: A d-simplex is stacked, and each simplicial d-polytope obtained from a stacked d-polytope with one fewer vertex by adding a pyramid over some facet is stacked. Alternatively, a simplicial d-polytope P is stacked if P is the union of simplices $S_1, S_2, \ldots, S_t$ such that each $(d-2)$-face of any of these simplices is a face of P.

Let P be a simplicial d-polytope. The set $\mathscr{B}(P)$ of proper faces of P forms a triangulation of the boundary of P. Thus, $\mathscr{B}(P)$ can be regarded as an abstract triangulation of S^{d-1}, the $(d-1)$-dimensional sphere. $\mathscr{B}(P)$ is called the *boundary complex* of P, [31, *Sect.* 3.2]. Define a *stacked* $(d-1)$*-sphere* to be a triangulated $(d-1)$-sphere which is isomorphic to the boundary complex of a stacked d-polytope.

A few years before Barnette proved the LBT, Walkup ([63]) settled the cases $d \leqq 5$. Walkup considered arbitrary triangulated $(d-1)$-manifolds and proved the case $d=4$ of the following theorem.

Theorem 1.1. *Let C be triangulated $(d-1)$-manifold, $d \geqq 4$, with n vertices, then:*

(i) $f_k(C) \geqq \varphi_k(n,d)$ *for* $1 \leqq k \leqq d-1$,

(ii) *If* $f_k(C) = \varphi_k(n,d)$ *for some* k, $1 \leqq k < d$, *then C is a stacked* $(d-1)$*-sphere.*

Note that the situation for $d=3$ is quite simple. A triangulated 2-manifold C with n vertices has $3n-3\chi(C)$ edges and $2n-2\chi(C)$ triangles, where $\chi(C)$ is the Euler characteristics of C. For every 2-manifold M, $\chi(M) \leqq 2$ and $\chi(M) = 2$ iff M is a 2-sphere. Thus, $f_i(C) = \varphi_i(n,3)$ for $i=1$ or $i=2$ iff C is a triangulated 2-sphere.

Our first purpose is to prove Theorem 1.1 for every $d \geqq 4$. Major portions of this result have been proved before by other methods: Part (i) and the special case $k=d-1$ of part (ii) were proved by Barnette (see [10, p. 354], [11]). Part (ii) for the special case of simplicial d-polytopes was proved by Billera and Lee in [15]. Their proof relies on the (necessity part of the) "g-theorem" - the complete characterization of f-vectors of simplicial polytopes, which was conjectured by McMullen ([47, 48]), and was proved by Stanley (necessity, [55]) and Billera and Lee (sufficiency, [15])). However, it was not known before that $f_k(C) = \varphi_k(n,d)$ occurs only if C is a triangulated sphere, (this was conjectured by Walkup [63, p. 77]). Nor was it known whether equality may holds for non-polytopal spheres.

A well-known and easy reduction due to McMullen, Perles and Walkup (see Sect. 5) reduces Theorem 1.1 to the case $k=1$.

We recall some definitions on rigidity of graph embeddings (frameworks). (See [5, 51, 28, 26, 33]). Given a graph $G=\langle V,E\rangle$, a *d-embedding* of G is a map $\psi: V\to\mathbb{R}^d$. A d-embedding ψ is *rigid* if any small perturbation φ of ψ which keeps the distances fixed between the images of adjacent vertices in G, keeps the distances fixed between every pair of vertices of G (and thus extends to an isometry of $\mathbb{R}^d$). A graph G is generically d-rigid if "almost all" embeddings of G into $\mathbb{R}^d$ are rigid. Such a graph having n vertices must have at least $dn-\binom{d+1}{2}$ edges. (Detailed definitions are given in Sect. 3.)

The inquality $f_1(C)\geqq\varphi_1(n,d)=dn-\binom{d+1}{2}$ for a triangulated $(d-1)$-manifolds C, $d\geqq 4$, with n vertices, follows from

Theorem 1.2. *The graph (1-skeleton) of every triangulated $(d-1)$-manifold, $d\geqq 4$, is generically d-rigid.*

The proof is given in Sect. 6. Using some basic results on rigidity we reduce Theorem 1.2 to the generic 3-rigidity of graphs of triangulated 2-spheres which was proved by Gluck [28] (see Sect. 4). (Compare Gromov [67, Ch. 2.4.10].)

For a triangulated $(d-1)$-manifold C define $\gamma(C)=f_1(C)-dn+\binom{d+1}{2}$. (The same definition applies to simplicial d-polytopes.) For $d\geqq 4$, $\gamma(C)$ is, by Theorem 1.2, the dimension of the space of *stresses* of a generic d-embedding of the graph of C.

In Sect. 7 we study those triangulated manifolds C for which $\gamma(C)=0$. We prove that if $\gamma(C)=0$ then $\gamma(\mathrm{lk}(v,C))=0$ for every vertex v of C. ($\mathrm{lk}(v,C)$ is the link of v in C, see Sect. 2.) Using this result, we reduce Theorem 1.1 (ii) to the known case $k=d-1$. A direct proof of Theorem 1.1 (ii) is given in Sect. 9.

In Sect. 8 we determine the class of triangulated d-manifolds C, $d\geqq 4$, which satisfy the condition: $\mathrm{lk}(v,C)$ is a stacked $(d-1)$-sphere for every vertex v of C. This condition implies a severe restriction on C, and, in particular, if C is simply-connected, then C itself must be a stacked d-sphere. We also derive a useful combinatorial characterization of stacked spheres among all triangulated manifolds.

Klee proved in [42] that the inequality $f_{d-1}(C)\geqq\varphi_{d-1}(n,d)$ holds for an arbitrary $(d-1)$-pseudomanifold C. Other cases of Theorem 1.1 are still open for this general setting. In Sect. 10 we show how the assertion of Theorem 1.1 for arbitrary $(d-1)$-pseudomanifold reduces to the old standing conjecture:

Conjecture G [28, 25]. *The graph of every triangulated 2-manifold is generically 3-rigid.*

In Sect. 11 we prove a sharp lower bound, conjectured by Björner [17], for the number of k-faces of a triangulated manifold with boundary, when the numbers of interior vertices and boundary vertices are specified.

Theorem 1.3. *Let C be a triangulated $(d-1)$-manifold $d\geqq 3$ with non-empty boundary. If C has n_i vertices in the interior and n_b vertices on the boundary then*

$f_k(C) \geqq \varphi_k^{\mathbf{b}}(n_i, n_b, d)$, *where*

$$\varphi_k^{\mathbf{b}}(n_i, n_b, d) = \begin{cases} \binom{d-1}{k} n_b + \binom{d}{k} n_i - \binom{d}{k+1} k & \text{for } 1 \leqq k \leqq d-2 \\ n_b + (d-1) n_i - (d-1) & \text{for } k = d-1. \end{cases} \tag{1.2}$$

Equality occurs only for a special type of triangulated balls. Theorem 1.3 for the special case when C is the dual of an unbounded simple polyhedra was proved (using the "g-theorem",) by Billera and Lee [16].

In Sect. 12 we discuss an extension of the LBT to arbitrary polytopes and polyhedral manifolds. For a polyhedral complex C, let $f_2^k(C)$ denotes the number of 2-faces of C which are k-gons. The following theorem, which was conjectured in [35, p. 67], extends the lower bound theorem to arbitrary d-polytopes.

Theorem 1.4. *If P is a d-polytope with n vertices then*

$$f_1(P) + \sum_{k \geqq 3} (k-3) \tilde{f}_2^k(P) \geqq dn - \binom{d+1}{2} \tag{1.3}$$

The analogous statement for arbitrary polyhedral $(d-1)$-manifolds (even polyhedral $(d-1)$-spheres,) is still open.

Theorem 1.4 follows from a recent theorem of Whiteley ([66], Sect. 4) on infinitesimal rigidity of certain embedded graphs associated with d-polytopes. (See Sect. 4.) Previously, it was proved for rational d-polytopes (namely, d-polytopes whose vertices have rational coordinates,) using some deep results from algebraic geometry ([58, Ch. 4, 46, 59]). In the second part of this paper ([38]) we study the class of d-polytopes which satisfy (1.3) as an equality.

Grünbaum proved ([31, p. 200],) that the graph of every d-polytope contains a refinement of the complete graph on $d+1$ vertices. Barnette extended this result ([11]) to arbitrary polyhedral $(d-1)$-manifolds. In Sect. 13 we prove

Theorem 1.5. *The graph of a triangulated $(d-1)$-manifold C, $d \geqq 4$, contains a refinement of the complete graph on $d+2$ vertices* iff *C is not a stacked $(d-1)$-sphere.*

In Sect. 14 we present a few open problems which were raised during this research. In particular, we briefly consider the LBT in the context of McMullen-Walkup "generalized lower bound conjecture" and discuss related problems on f-vectors of triangulated manifolds.

The basic reference (and source of inspiration) for convex polytope theory is Grünbaum's book [31]. We try to follow the definitions and notations of [31]. Other books on the subject are [48] and [21].

I would like to thank Richard Stanley for many valuable discussions on f-vectors of polytopes, and for introducing to me the recent exotic applications of algebraic geometry. I am thankful to Margaret Bayer, Louis Billera, Anders Björner, Robert Connelly, Henri Crapo, Micha Perles, and Walter Whiteley for helpful discussions during the various stages of this work. I would like to thank Lou Billera also for the warm hospitality during my visit at Cornell in

summer 1984. Being familiar to some of Perles' unpublished work was, as usual, a great advantage. This research was supported by the Weizmann postdoctoral fellowship. My work at Cornell was supported in part by NSF grant DMS 8403225.

2. Preliminaries

We shall use the following definitions and notation on simplicial complexes: Let C be a finite abstract simplicial complex on the vertex set V. Thus, C is a collection of subsets of V (called the *faces* of C,) and if $T\in C$ and $S\subseteq T$ then $S\in C$. For $S\in C$ the *dimension* of S is $\dim S=|S|-1$. $f_k(C)$ denotes the number of k-dimensional faces (briefly k-faces) of C. The *f-vector* of C is the vector $f(C)=(1,f_0(C),f_1(C),\ldots)$. The k-th dimensional *skeleton* of C, $\mathrm{skel}_k(C)$ is defined by

$$\mathrm{skel}_k(C)=\{S\in C:\dim S\leqq k\}.$$

$V(C)$ denotes the set of vertices (0-faces) of C. (Thus, $V(C)\subseteq V$.) 1-faces of C are called *edges* and $\mathrm{skel}_1(C)$ is called the *graph* of C and is denoted by $G(C)$.

For a face $S\in C$ the *link* of S in C, $\mathrm{lk}(S,C)$, is defined by:

$$\mathrm{lk}(S,C)=\{T\setminus S: T\in C, T\supset S\}.$$

($\mathrm{lk}(S,C)$ is also called the quotient complex of C by S.) Let V be a set of vertices and A be a family of subsets of V. $\bar{A}$ denotes the simplicial complex spanned by A. (I.e., $\bar{A}=\{S\subset V: S\subset T \text{ for some } T\in A\}$.) For a face $S\in C$, the *star* of S in C is defined by $st(S,C)=\overline{\{T\in C: T\supset S\}}$. The *antistar* of S in C is defined by $ast(S,C)=\{T\in C: T\cap S=\emptyset\}$.

Let C and D be simplicial complexes with $V=V(C)$, $U=V(D)$ and $V\cap U=\emptyset$. $C*D$, the *join* of C and D is defined by:

$$C*D=\{T\in V\cup U: T\cap V\in C, T\cap U\in D\}.$$

Note that $st(S,C)=\bar{S}*\mathrm{lk}(S,C)$.

A simplicial complex C is *pure* if all its maximal faces have the same size. Maximal faces of a pure simplicial complex are called *facets*. Two facets S, T of a pure simplicial complex are *adjacent* if they intersect in a maximal proper face of each. A pure simplicial complex C is *strongly connected* if for every two facets S and T of C, there is a sequence of facets $S=S_0,S_1,\ldots,S_m=T$, such that S_i and S_{i+1} are adjacent, $0\leqq i<m$.

A *d-pseudomanifold* is a strongly connected d-dimensional simplicial complex, such that every $(d-1)$-face is contained in exactly two facets. A *d-pseudomanifold with boundary* is a strongly connected d-dimensional simplicial complex, such that every $(d-1)$-face is contained in at most two facets. For a d-pseudomanifold with boundary C, the boundary of C, ∂C, is the $(d-1)$-dimensional pure simplicial complex whose facets are those $(d-1)$-faces of C which are included in a unique facet of C.

Let C be a pure simplicial complex and let F be a facet of C. The *stellar subdivision* of C at the facet F is defined by $C[F]=(C\setminus F)\cup\{R\cup\{u\}: R\subset F, R\neq F\}$. Here, u is a new vertex.

C is a *triangulated manifold* if $|C|$ is a manifold. ($|C|$ is the topological space associated with C. See, [53, Ch. 3]). It is usually more convenient to consider the larger class of homology manifolds. A pure d-dimensional complex C is a *homology manifold* if for every $\phi \neq S \in C$, $|S| = k$, the link of S in C has the same homology groups as a $(d-k)$-dimensional sphere. A homology d-manifold which has the same homology groups as a d-sphere is called a *homology d-sphere.*

3. Rigidity of frameworks

Let $G = \langle V, E \rangle$ be a graph with vertex set $V = V(G)$ and edge set $E = E(G)$. A *d-embedding* of G into $\mathbb{R}^d$ is a map $\varphi: V \to \mathbb{R}^d$. A *framework* $\mathcal{F}$ is a pair $\mathcal{F} = (G, \varphi)$ where $G = \langle V, E \rangle$ is a graph and φ is a d-embedding of G.

Two d-embeddings φ and ψ of a graph G are *isometric* if for every two vertices $a, b \in V$, $d(\varphi(a), \varphi(b)) = d(\psi(a), \psi(b))$. ($d(x, y)$ denotes the Euclidian distance between x and y.) Equivalently, φ and ψ are isometric if there is an affine rigid motion T of $\mathbb{R}^d$ such that $\varphi = T(\psi)$. Two d-embeddings φ and ψ of a graph G are *G-isometric* if for every two adjacent vertices $a, b \in V$, $d(\varphi(a), \varphi(b)) = d(\psi(a), \psi(b))$. (The vertices a and b are called *adjacent* if $\{a, b\} \in E(G)$.)

For two d-embeddings φ and ψ of G define their distance $d(\varphi, \psi) = \max_{a \in V} d(\varphi(a), \psi(a))$.

Definition 3.1. A *d-embedding* φ of a graph G is *rigid* if there is an $\varepsilon > 0$ such that every embedding ψ of G which is G-isometric to φ and satisfies $d(\varphi, \psi) < \varepsilon$ is isometric to φ. φ is *flexible* if it is not rigid.

Definition 3.2. A graph G is *generically d-rigid* if the set of rigid d-embeddings of G is an open dense set in the set of all embeddings. (The set of all embeddings is a topological vector space of dimension $|V| \times d$.)

Remarks. (1) We will freely use these definitions for an arbitrary simplicial (or more general) complex C and they will apply to the graph of C. (2) When we consider rigidity of d-polytopes or embedded manifolds this will be (unless stated otherwise) w.r.t. the given embedding.

A systematic study of rigidity of frameworks may be found in [5, 6, 28, 51]. We shall need the following basic facts:

0. If $H = \langle V, E' \rangle$ is generically d-rigid and $G = \langle V, E \rangle$ where $E \supset E'$ then G is generically d-rigid (obvious).

1. If G is not generically d-rigid then the set of rigid d-embeddings of G has empty interior. (In this case G is *generically d-flexible.*)

2. *If* G is a generically d-rigid graph with n vertices and e edges then $e \geqq dn - \binom{d+1}{2}$.

3. Let $G = \langle V(G), E(G) \rangle$ be a graph and let u be a vertex not in $V(G)$. Define $G^*\{u\} = \langle V', E' \rangle$ where $V' = V(G) \cup \{u\}$ and $E' = E(G) \cup \{\{u, v\} : v \in V(G)\}$. $G^*\{u\}$ is called a *cone* over G.

Cone Lemma (Whiteley, [65]). *G is generically d-rigid* iff *$G^*\{u\}$ is generically $(d+1)$-rigid.*

4. **Replacement Lemma.** *Let $G=\langle V,E\rangle$ be a graph and let U be a subset of V. If the restriction of G to U is generically d-rigid and $G\cup K(U)$ is generically d-rigid then G is generically d-rigid. ($K(U)$ denotes the complete graph on U.)* (The proof is easy.)

Given a fixed set V of vertices, the set of edges of minimal (w.r.t. inclusion) generically d-rigid graphs on V, is the set of bases of a matroid $\mathscr{R}_d^n$ of rank $\binom{n}{2}-\binom{n-d}{2}\left(=dn-\binom{d+1}{2}\text{ for } n\geqq d\right)$.

Definition 3.3. A graph $G=\langle V,E\rangle$ is (generically) *d-acyclic*[1] if the set of its edges is independent in $\mathscr{R}_d^n$.

For the reader who is not familiar with matroid theory terminology (a good reference is Welsh [64]), here is an equivalent definition: Let φ be a d-embedding of a graph G. An edge $\{a,b\}$, not in $E(G)$ depends on G (w.r.t. φ), if for every embedding ψ which is G-isometric and close enough to φ, $d(\psi(a),\psi(b))=d(\varphi(a),\varphi(b))$. G is d-acyclic if for a generic d-embedding of G no edge E of G depends on $G'=\langle V(G),E(G)\setminus E\rangle$.

An important variant of rigidity is the notion of *infinitesimal rigidity*. The definition given below follows Connelly [26]. For the geometric motivation behind the definition and a full treatment of the relations between the different notions of this section see [26] and [51].

Let φ be a d-embedding of a graph G. An *infinitesimal flex* of φ is a d-embedding ψ of G such that for every two adjacent vertices a and b of G, $(\varphi(a)-\varphi(b))\cdot(\psi(a)-\psi(b))=0$. (Here, $\cdot$ is the usual scalar product.) An infinitesimal flex ψ of φ is *trivial* of for every two vertices a,b of G, $(\varphi(a)-\varphi(b))\cdot(\psi(a)-\psi(b))=0$. A d-embedding φ of G is *infinitesimally rigid* if every infinitesimal flex of φ is trivial.

Infinitesimally rigid frameworks are rigid, and the generic behavior w.r.t. rigidity and infinitesimal rigidity coincide. If a graph G is infinitesimally rigid w.r.t. one d-embedding then it is generically d-rigid. (In particular, $|E(G)|\geqq d|V(G)|-\binom{d+1}{2}$.)

Given a d-embedding φ of a graph G, a *stress* of G w.r.t. φ is a function $w: E(G)\to R$ such that for every vertex $v\in V$

$$\sum\{w(\{v,u\})(\varphi(v)-\varphi(u)): \{v,u\}\in E(G)\}=0.$$

For a graph $G=\langle V,E\rangle$, $a_d(G)$ will denote the rank of G in $\mathscr{R}_d^n$. Alternatively, $a_d(G)$ is the number of edges of a maximal d-acyclic subgraph of G. (All maximal d-acyclic subgraphs of G have the same number of edges.) Define $b_d(G)=|E(G)|-a_d(G)$. $b_d(G)$ is the dimension of the space of stresses of G w.r.t. a generic d-embedding. In particular, G is d-acyclic if a generic d-embedding of G has no non-zero stress.

[1] This definition is slightly different from the definition in [37] which relies on a different matroid

4. Theorems of Cauchy, Steinitz, Alexandrov, Gluck and Whiteley

We make an essential use on the following theorem of Gluck [28].

Theorem G. *A triangulated 2-sphere is generically 2-rigid.*

Let us give a quick survey of Gluck's proof. Theorem G follows from the fundamental theorems of Cauchy and Steinitz. Cauchy's rigidity theorem ([22]) asserts that if P and Q are two convex 3-polytopes and $\varphi: V(P) \to V(Q)$ is a combinatorial isomorphism, which induces an isometry between every face of P and its image in Q, then P and Q are isometric. Steinitz's theorem (see [61, 31, p. 235, 14]) asserts that every polyhedral 2-sphere is combinatorially isomorphic to the boundary complex of a 3-polytope.

Cauchy's theorem implies that every simplicial 3-polytope P is rigid. Since the set of embeddings of P which actually realize P as a convex polytope is an open subset of the set of all embeddings, the graph of P is generically 3-rigid. By Steinitz's theorem every triangulated 2-sphere is isomorphic to the boundary complex of a simplicial 3-polytope and is therefore generically 3-rigid.

A *d-polytopal framework* is an embedded graph obtained from the graph of a d-polytope P by triangulating the 2-faces of P in an arbitrary way.

Alexandrov ([1]) extended Cauchy's arguments and proved that every 3-polytopal framework is infinitesimally rigid. (Note that Alexandrov's theorem combined with Steinitz's theorem give an even more direct proof of Theorem G. This is the variant in [28].)

Whiteley ([66]) have recently found a significant generalization of Alexandrov's theorem to higher dimensions

Theorem W. *A d-polytopal framework, $d \geqq 3$, is infinitesimally rigid.*

The basic connection between rigidity and the LBT can be seen at this point. Note that in a d-polytopal framework $\mathscr{F}(P)$, based on a d-polytope P, there are $(k-3)$ additional edges for each k-gonal 2-face. Thus, $\mathscr{F}(P)$ has exactly $f_1(C) + \sum_{k \geq 3} (k-3) f_2^k(C)$ edges. Theorem 1.4 follows from Theorem W and the basic inequality $e \geqq dn - \binom{d+1}{2}$ for the number e of edges of an infinitesimally rigid d-embedded graph with n vertices. In particular, this gives the essential case $k = 1$ of the lower bound inequalities for simplicial polytopes.

Remark. Gluck's proof of the generic 3-rigidity of triangulated 2-spheres is unusual. Convexity is not involved in the assertion of the theorem but is very much present in the proof. Steinitz's theorem is a sort of a low dimensional miracle, and Cauchy's theorem gives a much stronger rigidity property than needed. Recently, Tay and Whiteley ([62]) found a direct proof for Gluck's theorem which does not depend on Cauchy's or Steinitz's theorems. Graver's approach ([30]) may also supply a direct proof for Gluck's theorem.

5. The MPW-reduction

The result of this section were found (independently) by McMullen, Perles and Walkup (see [10, 49]). Recall that $\varphi_k(n, d)$ is the number of k-faces in a stacked

d-polytope with n vertices and is given by formula (1.1). For a pure $(d-1)$-dimensional simplicial complex C with n vertices define $\gamma(C)=f_1(C)-\varphi_1(n,d)$. Thus, for $d\geqq 3$, $\gamma(C)=f_1(C)-dn+\binom{d+1}{2}$ and for $d=2$, $\gamma(C)=f_1(C)-n$. Define also

$$\gamma_k(C)=f_k(C)-\varphi_k(n,d),$$

and

$$\gamma^k(C)=\sum\{\gamma(\mathrm{lk}(S,C)): S\in C, |S|=k\}.$$

Thus, $\gamma_1(C)=\gamma^0(C)=\gamma(C)$.

Proposition 5.1. *Let C be a $(d-1)$-dimensional simplicial complex, and let k, d be integers, $1\leqq k\leqq d-1$. There are* positive *constants $w_i(k,d)$, $0\leqq i\leqq k-1$, such that*

$$\gamma_k(C)=\sum_{i=0}^{k-1} w_i(k,d)\,\gamma^i(C). \tag{5.1}$$

Proof. First note that

$$(k+1)f_k(C)=\sum_{i=1}^{n} f_{k-1}(\mathrm{lk}(v,C)). \tag{5.2}$$

Put $\varphi_k(n,d)=a_k(d)\,n+b_k(d)$. (Thus, $a_k(d)=\binom{d}{k}$ for $1\leqq k\leqq d-2$ and $a_{d-1}(d)=d-1$.) Easy calculation gives

$$2\left(dn-\binom{d+1}{2}\right)a_{k-1}(d-1)+n\,b_{k-1}(d-1)=(k+1)\,\varphi_k(n,d). \tag{5.3}$$

Let C be a pure $(d-1)$-dimensional simplicial complex, $d\geqq 3$, with n *vertices* $v_1,\ldots,v_n$. Assume that the degree of v_i in $G(C)$ is n_i (i.e., $f_0(\mathrm{lk}(v_i,C))=n_i$). Note that $\sum_{i=0}^{n} n_i=2f_1(C)=2\left(dn-\binom{d+1}{2}+\gamma(C)\right)$. Therefore

$$\begin{aligned}\sum_{1=1}^{n}\varphi_{k-1}(n_i,d-1)&=a_{k-1}(d-1)\sum_{i=1}^{n} n_i+n\,b_{k-1}(d-1)\\&=a_{k-1}(d-1)\,2\left(dn-\binom{d+1}{2}\right)+2a_{k-1}(d-1)\,\gamma(C)+nb_{k-1}(d-1).\\&=(k+1)\,\varphi_k(n,d)+2a_{k-1}(d-1)\,\gamma(C).\end{aligned} \tag{5.4}$$

From (5.2) and (5.4) we get

$$(1+k)\,\gamma_k(C)=2a_{k-1}(d-1)\,\gamma(C)+\sum_{i=1}^{n}\gamma_{k-1}(\mathrm{lk}(v_i,C)). \tag{5.5}$$

Repeated applications of formula (5.5) give (5.1). The value of $w_i(k,d)$ is

$$w_i(k,d)=\begin{cases}2(a_{k-i-1}(d-i-1))/(k+1)\binom{k}{i} & 0\leqq i\leqq k-2,\\ 2/(k+1)\,k & i=k-1\end{cases} \tag{5.6}$$

Corollary (the MPW-reduction). *Let $d \geqq 2$ be an integer. Let C be a $(d-1)$-dimensional simplicial complex with n vertices, such that $\gamma(\mathrm{lk}(S, C)) \geqq 0$ for every $S \in C$, $|S| < k$. Then* (i) $f_k(C) \geqq \varphi_k(n, d)$. (ii) *If $f_k(C) = \varphi_k(n, d)$ then $\gamma(C) = 0$.*

Remark. Note that if C is a $(d-1)$-pseudomanifold then $\gamma^{d-2}(C) = 0$.

6. The lower bound inequalities for triangulated manifolds

For $d \geqq 3$ define a class $\mathscr{C}_d$ of $(d-1)$-pseudomanifolds inductively as follows: $\mathscr{C}_3$ is the class of triangulated 2-spheres. For $d \geqq 4$, a $(d-1)$-pseudomanifold C belongs to $\mathscr{C}_d$ if for every vertex v of C, $\mathrm{lk}(v, C) \in \mathscr{C}_{d-1}$. Note that every homology 2-sphere is a triangulated 2-sphere. Therefore for $d \geqq 4$, $\mathscr{C}_d$ includes all homology $(d-1)$-manifolds (and, in particular, all triangulated $(d-1)$-manifolds). $\mathscr{C}_4$ is exactly the class of homology 3-manifolds.

Theorem 6.1. *If $C \in \mathscr{C}_d$ then C is generically d-rigid.*

Lemma 6.2. *Let C be a strongly connected d-dimensional simplicial complex. Then C is generically d-rigid.*

Proof. (Compare [36].) If every two vertices of C are adjacent then C is clearly generically d-rigid. Otherwise, since C is strongly connected, there are two non-adjacent vertices u, v of C, and two adjacent d-faces S and T, such that $u \in S$ and $v \in T$. Let $\bar{C}$ be the simplicial complex obtained from C by adding to C all d-faces of $\overline{S \cup T}$. The affect of the operation $C \to \bar{C}$ on $G(C)$ is just adding one new edge $\{u, v\}$. The graph induced by $G(C)$ on the vertices of $S \cup T$ is a complete graph on $d+2$ vertices minus an edge ("$\{u, v\}$"). This graph is clearly generically d-rigid and by the Replacement Lemma (Sect. 3) if $\bar{C}$ is generically d-rigid so is C. Repeated application of this operation will terminate with a complex $\hat{C}$ whose graph is complete. $\hat{C}$ is clearly generically d-rigid.

Proof of Theorem 6.1. By induction on d. For $d = 3$, $\mathscr{C}_d$ is the class of triangulated 2-spheres which are generically 3-rigid by Gluck's theorem, we assume the truth of the theorem for $d-1$ and prove it for d. Let $C \in \mathscr{C}_d$. For a vertex $v \in C$, the neighborhood $N(v)$ of v is defined by $N(v) = \{v\} \cup \{u \in V(C): \{u, v\} \in C\}$. For a vertex $v \in C$, $\mathrm{lk}(v, C) \in \mathscr{C}_{d-1}$ and by the induction hypothesis $\mathrm{lk}(v, C)$ is generically $(d-1)$-rigid. By the cone lemma (Sect. 3), $st(v, C) = \{v\} * \mathrm{lk}(v, C)$ is generically d-rigid. Let $K_d(N(v))$ denote the complete d-dimensional complex on $N(v)$. By the replacement lemma (Sect. 3), C is generically d-rigid iff $C \cup K_d(N(v))$ is generically d-rigid. Repeated application of this argument shows that C is generically d-rigid iff $\bar{C} = \bigcup \{K_d(N(v)): v \in V(C)\}$ is generically d-rigid. But $\bar{C}$ easily seen to be a strongly connected d-dimensional complex, hence generically d-rigid.

Theorem 6.1 and the MPW reduction give:

Theorem 6.2. *If $C \in \mathscr{C}_d$ and C has n vertices then $f_k(C) \geqq \varphi_k(n, d)$ for all $d \geqq k \geqq 1$.*

Remarks. 1. The inductive argument in the proof of Theorem 6.1 seems to be quite old. It is hinted in [60, foornote p. 119] and perhaps goes back to the

works of Alexandrov and Pogorelov. Whiteley's proof of Theorem W uses similar (but more delicate) inductive argument.

2. Theorem 6.1 strengthened the fact that the graph of a triangulated $(d-1)$-manifold is d-connected. Barnette [11] proved that the graph of every polyhedral $(d-1)$-manifold is d-connected, thus extending a result of Balinski [7] which asserts that the graph of every d-polytope is d-connected.

Let G be a graph with n vertices and e edges, $n \geqq d$. Recall that $b_d(G)$ is the dimension of the space of stresses of G w.r.t. a generic d-embedding. $b_d(G) \geqq e - dn + \binom{d+1}{2}$ and equality holds iff G is generically d-rigid. Theorem 6.1 thus implies that for $C \in \mathscr{C}_d$, $\gamma(C)$ is the dimension of the space of stresses of a generic d-embedding of $G(C)$.

Theorem 6.1 implies also an upper bound for the number of edges of subgraphs of graphs of triangulated manifolds.

Theorem 6.3. *Let $C \in \mathscr{C}_d$ and let H be a subgraph of $G(C)$. Then $f_1(H) \leqq d f_0(H) - \binom{d+1}{2} + \gamma(C)$.*

Proof. Let H be a subgraph of $G(C)$. (We may assume that H has at least d vertices.) Denote $\gamma(H) = f_1(H) - d f_0(H) + \binom{d+1}{2}$. Note that if H is a subgraph of G then $b_d(H) \leqq b_d(G)$. Therefore,

$$\gamma(C) = b_d(G(C)) \geqq b_d(H) \geqq \gamma(H).$$

We conclude this section by showing that the proof of Theorem 6.1 applies in a slightly more general situation. (We use this fact in Sects. 9 and 11.) Let C be a strongly connected $(d-1)$-dimensional simplicial complex and let T be a tree in $G(C)$. It is easy to see that $\bigcup \{K_d(N(v))\colon v \text{ a vertex of } T\}$ is a strongly connected d-dimensional simplicial complex. Therefore, the proof of Theorem 6.1 gives.

Proposition 6.4. *Let C be a strongly connected $(d-1)$-dimensional simplicial complex. Let T be a tree in $G(C)$ which satisfy:* (i) *Every vertex u of C is adjacent to some vertex of T,* (ii) $\mathrm{lk}(v, C)$ *is generically $(d-1)$-rigid for every vertex v of T. Then C is generically d-rigid.*

7. The extremal cases in the lower bound theorem

Recall that a stacked $(d-1)$-sphere is a triangulated $(d-1)$-sphere which is isomorphic to the boundary complex of a stacked d-polytope. As easily seen, C is a stacked $(d-1)$-sphere iff C can be obtained from the boundary complex of a d-simplex by repeated applications of stellar subdivisions of facets.

Theorem 7.1. *Let d, k be fixed integers $d > 3$, $d > k \geqq 1$. Let C be a simplicial complex in $\mathscr{C}_d$ with n vertices and $\varphi_k(n, d)$ k-faces. Then C is a stacked $(d-1)$-sphere.*

Proof. The MPW reduction shows that for $C \in \mathscr{C}_d$, if $f_0(C)=n$ and $f_k(C)=\varphi_k(n,d)$ for some $1<k<d$, then $f_1(C)=\varphi_1(n,d)$, i.e., $\gamma(C)=0$. Define: $\mathscr{C}_d^0=\{C \in \mathscr{C}_d : \gamma(C)=0\}$. By Theorem 6.1 every $C \in \mathscr{C}_d$ is generically d-rigid. Therefore, for $C \in \mathscr{C}_d$, $C \in \mathscr{C}_d^0$ iff C is d-acyclic. (See Sect. 3.)

Lemma 7.2. *If $C \in \mathscr{C}_d^0$, $d \geqq 4$, then for every vertex $v \in C$, $\mathrm{lk}(v, C) \in \mathscr{C}_{d-1}^0$.*

Proof. Assume to the contrary that $C \in \mathscr{C}_d^0$, v is a vertex of C, and $\mathrm{lk}(v, C) \notin \mathscr{C}_{d-1}^0$. Thus, $\mathrm{lk}(v, C)$ is not $(d-1)$-acyclic and from the Cone Lemma (Sect. 3) it follows that $st(v, C)=v * \mathrm{lk}(v, C)$ is not d-acyclic. Since $C \supset st(v, C)$, C is not d-acyclic as well. A contradiction.

Proof of Theorem 7.1 (end). The case $d=4$ of Theorem 7.1 was proved already by Walkup ([63, Th. 1]). (Barnette's result mentioned below also covers this case.) Assume now that for $d \geqq 5$, if $C \in \mathscr{C}_{d-1}^0$ then C is a stacked $(d-1)$-sphere. Let $C \in \mathscr{C}_d^0$, $d \geqq 5$. Recall that $\gamma^k(C)=\sum\{\gamma(\mathrm{lk}(S, C)) : S \in C, |S|=k\}$. (See Sect. 5.) Lemma 7.2 implies that for every $S \in C$, $\gamma(\mathrm{lk}(S, C))=0$. Therefore, for every $k \geqq 1$, $\gamma^k(C)=0$. By Proposition 5.1, $f_{d-1}(C)=\varphi_{d-1}(n,d)$. By Lemma 7.2 for every vertex $v \in C$, $\mathrm{lk}(v, C) \in \mathscr{C}_{d-1}^0$. By the induction hypothesis $\mathrm{lk}(v, C)$ is a stacked sphere, and therefore C is a triangulated $(d-1)$-manifold. Barnette proved ([9, 11]) that if a triangulated $(d-1)$-manifold C with n vertices satisfies $f_{d-1}(C)=\varphi_{d-1}(n,d)$ then C is a stacked $(d-1)$-sphere. This completes the proof of Theorem 7.1.

A direct proof of Theorem 7.1 is given in Sect. 9. We use there a characterization of stacked spheres which is proved in the next section.

The proof of Lemma 7.2 gives more:

Theorem 7.3. *Let C be a generically d-rigid pure $(d-1)$-dimensional simplicial complex. Then for every vertex v of C, $\gamma(\mathrm{lk}(v, C)) \leqq \gamma(C)$.*

Proof. Define $G_1=G(\mathrm{lk}(v, C))$, $G_2=G(st(v, C))$ $(=G_1 * \{v\})$. Let H be a maximal $(d-1)$-acyclic subgraph of G_1. By the cone Lemma, $H * \{v\}$ is a maximum d-acyclic subgraph of G_2. Therefore

$$\gamma(\mathrm{lk}(v, C)) \leqq b_{d-1}(G_1)=b_d(G_2) \leqq b_d(G(C))=\gamma(C).$$

8. Triangulated manifolds with stacked links

In this section we study triangulated manifolds C such that $\mathrm{lk}(v, C)$ is a stacked sphere for every vertex v of C. For manifolds of dimensions greater than 3 this condition implies a severe topological restriction. We also derive a characterization of stacked spheres among pseudomanifolds in $\mathscr{C}_d$ which is used in the next sections.

Consider the following two operations on triangulated manifolds. Let C and D be pure simplicial complexes with disjoint sets of vertices, S be a facet of C and T be a facet of D. Let ψ be a bijection between $V(S)$ and $V(T)$. The *connected sum* $C \#_\psi D$ of C and D is the simplicial complex obtained by identifying the vertices of S with the vertices of T by ψ and deleting the facet S

$(=T)$. Connected sums of two triangulated manifolds is a triangulated manifold. Note that if $E=C\#_{\psi}D$ then for $v\in S, \mathrm{lk}(v,E)$ is a connected sum of $\mathrm{lk}(v,C)$ and $\mathrm{lk}(\psi(v),D)$. All other links are unchanged.

Let C be a pure $(d-1)$-dimensional simplicial complex, S and T be two disjoint facets of C, and ψ be a bijection from $V(S)$ to $V(T)$. Assume further that no vertex of S is adjacent to a vertex of T and that no vertex in C is adjacent to both a vertex v in S and to its image $\psi(v)$ in T. Let C^{ψ} be the simplicial complex obtained from C by identifying the vertices of S to the vertices of T via ψ and deleting the facet $S(=T)$. We say that C^{ψ} is obtained by *forming a handle* over C. Note that $\mathrm{lk}(v,C^{\psi})=\mathrm{lk}(v,C)$ unless $v\in S$ $(=T)$, and then $\mathrm{lk}(v,C^{\psi})=\mathrm{lk}(v,C)\#\mathrm{lk}(\psi(v),C)$.

Note also that

$$\gamma(C\#_{\psi}D)=\gamma(C)+\gamma(D), \tag{8.1}$$

$$\gamma(C^{\psi})=\gamma(C)+\binom{d+1}{2}\ (d=\dim C-1). \tag{8.2}$$

Walkup defined the class $\mathscr{H}^d(k)$ of $(d-1)$-dimensional simplicial complexes as follows: $\mathscr{H}^d(0)$ is the class of stacked $(d-1)$-spheres. $C\in\mathscr{H}^d(k)$ if $C=D^{\psi}$ for some $D\in\mathscr{H}^d(k-1)$. Define $\mathscr{H}^d=\bigcup\{\mathscr{H}^d(k): k\geqq 0\}$. Note that a connected sum of two complexes in $\mathscr{H}^d$ is in $\mathscr{H}^d$. In fact, $\mathscr{H}^d$ is exactly the class of simplicial complexes obtained from boundary complexes of d-simplices by successively applying the operations $C\#_{\psi}D$ and C^{ψ}. For $d\geqq 4$, if $C\in\mathscr{H}^d(k)$ then rank $H_1(C)=k$ and $\gamma(C)=k\binom{d+1}{2}$. From the description of links of vertices of $C\#_{\psi}D$ and C^{ψ}, it follows that if $C\in\mathscr{H}^d$, then $\mathrm{lk}(v,C)$ is a stacked $(d-2)$-sphere for every vertex v of C.

The notion of a *missing face* (see [3],) will play an important role from now on.

Definition 8.1. *Let C be a simplicial complex on the vertex set V. A subset S of V is a missing face of C, if $S\notin C$ but for every proper subset R of S, $R\in C$. A k-missing face is a missing face with $k+1$ vertices.*

Theorem 8.2. *Let C be a $(d-1)$-pseudomanifold, $d\geqq 4$. If for every vertex $v\in C, \mathrm{lk}(v,C)\in\mathscr{H}^{d-1}(0)$ and C has no $(d-2)$-missing faces, then $C\in\mathscr{H}^d$.*

Lemma 8.3. *Let P be a stacked d-polytope.* (i) *P has no k-missing faces for $1<k<d-1$.* (ii) *If P is not a d-simplex then P has a missing $(d-1)$-face.*

Proof. Let P and Q be two simplicial d-polytopes such that Q is obtained from P by adding a pyramid over a facet T of P. (The boundary complex of Q is obtained from the boundary complex of P by a stellar subdivision of T.) It is easy to see that every missing face of P is a missing face of Q and, in addition, Q has one new $(d-1)$-missing face T and $f_0(P)-d$ new 1-missing faces of the form $\{u,v\}$ where u is the new vertex of Q and $v\notin T$. Lemma 8.3 follows by induction from the definition of stacked polytopes.

Proof of Theorem 8.2. Let $v\in C$ and let S be a $(d-2)$-missing face in $\mathrm{lk}(v,C)$. (Unless C is a simplex there is a vertex v in C whose degree is more than d

and therefore $\mathrm{lk}(v, C)$ has a $(d-2)$-missing face.) Since C has no $(d-2)$-missing faces, S must be a face of C and therefore $T=S\cup\{v\}$ is a $(d-1)$-missing face of C. Cut C along ∂T and patch with two $(d-1)$-simplices. (As was shown by Walkup, [63, Lemma 4.2], this operation can always be performed.) The resulting complex is a (possibly not connected) triangulated $(d-1)$-manifold $\bar{C}$. If $\bar{C}$ is connected then C is obtained from $\bar{C}$ by forming a handle. If $\bar{C}$ is not connected it has two connected components and C is their connected sum. Theorem 8.2 follows by double induction on $\gamma(C)$ and $f_0(C)$.

Corollary 8.4. *Let C be a $(d-1)$-pseudomanifold, $d\geqq 5$. If for every vertex $v\in C$, $\mathrm{lk}(v, C)$ is a stacked $(d-2)$-sphere, then $C\in\mathscr{H}^d$.*

Proof. It is enough to show that C does not have $(d-2)$-missing faces. Indeed, if S is a $(d-2)$-missing face of C and v is a vertex of S then $S\setminus\{v\}$ is a $(d-3)$-missing face of $\mathrm{lk}(v, C)$. This is impossible by Lemma 8.2(i) since $\mathrm{lk}(v, C)$ is a stacked $(d-2)$-sphere and $(d-3)>1$.

Remark 8.5. Perles proved (see [4]) that if P is a neighborly 4-polytope then every link of a vertex of P is stacked. Thus, the class of triangulated 3-manifolds with stacked 2-spheres as the only links of vertices, is much larger than $\mathscr{H}^4$. Having only stacked spheres as links impose a severe topological restriction on d-manifolds for $d\geqq 4$. *Problem:* Which 3-manifolds admit a triangulation with only stacked 2-spheres as links of vertices? (Compare [23].)

We derive now from Theorem 8.2 a useful characterization of stacked spheres. Recall that a cycle M in a graph G is *chordless* if M is an induced subgraph of G. (Thus, M is a subgraph of G with a set of vertices $V(M)=\{v_1,\ldots,v_m\}$, $m\geqq 3$ and edges $\{v_1,v_2\},\ldots,\{v_{m-1},v_m\},\{v_m,v_1\}$ and the only edges of G with endpoints in $V(M)$ are edges of M.) A graph is *chordal* if it does not contain a chordless m-cycles for $m\geqq 4$.

Theorem 8.5. *Let $C\in\mathscr{C}_d$, $d\geqq 3$. The following are equivalent:*

(i) *C is a stacked $(d-1)$-sphere,*
(ii) *$G(C)$ is chordal and C has no k-missing faces for $1<k<d-1$.*

Proof. (i)$\rightarrow$(ii). Let C be a stacked $(d-1)$-sphere, $d\geqq 3$. By Lemma 8.2, C has no k-missing faces for $1<k<d-1$. It is left to show that $G(C)$ is chordal. Let P and Q be two simplicial d-polytopes such that Q is obtained from P by adding a pyramid over a facet T of P. $G(Q)$ is obtained from $G(P)$ by adding a new vertex u and connecting it to all vertices of T. From this description it is clear that if $G(P)$ is chordal then so is $G(Q)$. Therefore, graphs of stacked $(d-1)$-spheres are chordal.

(ii)$\rightarrow$(i). The proof will proceed by induction on d. For $d=3$ we have to prove that every triangulated 2-sphere C with a chordal graph, is a stacked 2-sphere. Assume to the contrary, that C is a counterexample with a minimal number of vertices. If C has a 2-missing face then C is the connected sum of two smaller triangulated 2-spheres C_1 and C_2. $G(C_1)$ and $G(C_2)$ are chordal and by the minimality of C, C_1 and C_2 are stacked and therefore so is C. Thus, C does not have a 2-missing face. Let v be a vertex of degree 4 or 5 in C. (Such a vertex always exists unless C is the boundary of a 3-simplex.) If v

has 4 neighbors they form a 4-cycle (with the edges of $\mathrm{lk}(v, C)$) and this 4-cycle must have a diagonal. Since C has no 2-missing faces C is a stacked 2-sphere with 5 vertices. If v has 5 neighbors then by the same argument C is a stacked 2-sphere with 6 vertices. A contradiction.

Let $d \geqq 4$, and assume that the implication (ii)$\rightarrow$(i) holds for every $d' < d$. Let C be a member of $\mathscr{C}_d$ with a chordal graph and no k-missing faces for $1 < k < d-1$. First we show that C is simply-connected. Otherwise, let M be a minimal cycle in $G(C)$ which is not null-homotopic in C. M must be chordless and if M is a triangle it must be a 2-missing face. Let v be a vertex of C. If S is a k-missing face of $\mathrm{lk}(v, C)$, $1 < k < d-2$ then either S itself or $S \cup \{v\}$ is a missing face of C. This is impossible by the assumption on C. If M is a chordless cycle in $\mathrm{lk}(v, C)$ then since C has no 2-missing faces, M is chordless in C as well. Thus, by the induction hypothesis, $\mathrm{lk}(v, C)$ is a stacked $(d-2)$-sphere for every vertex v of C. Since C does not have $(d-2)$-missing faces and is simply-connected, by Theorem 8.2, C is a stacked $(d-1)$-sphere.

Both conditions of Theorem 8.5(ii) are necessary. The graph of every 2-neighborly d-polytope is chordal. The d-cross polytope has k-missing faces only for $k=1$. The implication (ii)$\rightarrow$(i) does not hold for arbitrary $(d-1)$-pseudomanifolds as shown by the 3-neighborly 3-pseudomanifolds of Altshuler ([2]).

9. Direct proof of Theorem 7.1

Lemma 9.1. *If $C \in \mathscr{C}_d^0$, S is a missing face of C then either* $\dim S = 1$ *or* $\dim S = d-1$.

Proof. The lemma says nothing for $d=3$. Let $C \in \mathscr{C}_d^0$, $d \geqq 4$. Let us first show that C has no 2-missing faces. Assume to the contrary that T is a 2-missing face of C. Let v be a vertex of T and let $E = T \setminus \{v\}$. E is an edge of C, the vertices of E are adjacent to v and are therefore vertices of $st(v, C)$. But E itself does not belong to $\mathrm{lk}(v, C)(E \cup \{v\} \notin C)$, and therefore E does not belong to $st(v, C)$. Since $st(v, C)$ is generically d-rigid, E depends on $st(v, C)$ w.r.t. a generic d-embedding. However, $E \notin st(v, C)$ and therefore $st(v, C) \cup E$ is not d-acyclic. Since $st(v, C) \cup E \subset C$, C is not d-acyclic. A contradiction.

If T is a k-missing face of C, $2 < k < d-1$ then for every subset S of T of size $k-2$, $T \setminus S$ is a 2-missing face of $\mathrm{lk}(S, C)$. By Lemma 7.2, $\mathrm{lk}(S, C) \in \mathscr{C}_{d-k+2}^0$. But $d-k+2 \geqq 4$ and therefore $\mathrm{lk}(S, C)$ does not have a 2-missing face. A contradiction.

Lemma 9.2. *If $C \in \mathscr{C}_d^0$ then $G(C)$ is chordal.*

Proof. Assume to the contrary that $C \in \mathscr{C}_d^0$ and M is a chordless m-gon in C, $m \geqq 4$. Let $E = \{v_1, v_2\}$ be an edge in M. Let U be the set of vertices of M which are not in E, and let H be the induced subgraph of M on U. (H is a path.) Let W be the set of vertices of C which are adjacent to some vertex of U. Clearly $v_1, v_2 \in W$. Define a simplicial complex D on W by $D = \bigcup \{st(u, C) : u \in U\}$. Since M is chordless $E \notin D$. By Proposition 6.4, D is generally d-rigid. But the vertices of

E belongs to D, therefore $D \cup E$ is not d-acyclic and since $D \cup E \subseteq C$. C is not d-acyclic. A contradiction.

Direct proof of Theorem 7.1 (end). Let $C \in \mathscr{C}_d^0$, by Lemmas 9.1 and 9.2, C has no k-missing faces for $1 < k < d-1$ and no chordless m-gons for $m \geqq 4$. By Theorem 8.5, C is a stacked $(d-1)$-sphere.

Remark. Most of the work is needed just for the case $d=4$. If one assumes the assertion of Theorem 7.1 for $d=4$, then the general case follows easily by induction, from Lemma 7.2 and Theorem 8.2.

Corollary 8.4 and Theorem 7.1 imply:

Theorem 9.3. *If C is a simply-connected triangulated $(d-1)$-manifold, $d \geqq 5$, and for every vertex $v \in C$, $\gamma(\mathrm{lk}(v, C)) = 0$ then $\gamma(C) = 0$.*

Second proof of Theorem 9.3 (hint). In order to show that $\gamma(C)=0$ it is enough to prove that for every edge $E \in C$, a generic d-embedding ρ of $C \setminus E$ has a non-trivial infinitesimal flex v. (See Sect. 3). Let $E = \{v_1, v_2\}$ be an edge of C. Since $\mathrm{lk}(v_1, C)$ is acyclic, $st(v_1, C) \setminus E$ has a non-trivial infinitesimal flex. Choose such a flex v_0. We will extend this infinitesimal flex to an infinitesimal flex of $C \setminus E$. Let $r = d(v_0(v_1), v_0(v_2))$.

Let $\{w, u\}$ be an edge in C, and let ξ be an infinitesimal flex of $st(w, C) \setminus E$. Consider the restriction of ξ to $D_0 = st(\{u, w\}, C) \setminus E$ and extend it to an infinitesimal flex $\bar{\xi}$ of $D_1 = st(u, C) \setminus E$. This can always be done (here we use the fact that $d \geqq 5$). The extension is unique unless $v_1, v_2 \in D_1$, but either v_1 or v_2 are not in D_0. In this case extend ξ under the condition that $d(\bar{\xi}(v_1), \bar{\xi}(v_2)) = r$.

Apply this operation to extend v_0 to stars of all the vertices in C. It can be shown that if an infinitesimal flex is defined on $st(v, C)$ using this procedure via a path l from v_1 to v, then it depends only on the homotopy class of the path l. Therefore, if C is simply-connected one gets a well-defined non-trivial infinitesimal flex on $C \setminus E$.

Third proof of Theorem 9.3 for boundary complexes of simplicial polytopes. Let $\delta(C) = f_2(C) - (d-1) f_1(C) + \binom{d}{2} f_1(C) - \binom{d+1}{3}$ $(= h_3(C) - h_2(C)$, see Sect. 14). It is easy to check that

$$\sum \{\gamma(\mathrm{lk}(v, C)) : v \in V(C)\} = 3\,\delta(C) + (d-1)\,\gamma(C). \tag{9.1}$$

It is plausible that $\delta(C) \geqq 0$ holds for every simply-connected triangulated $(d-1)$-manifold C, $d \geqq 5$. This is known only when $d=5$ and when $d>5$ and C is the boundary complex of a d-polytope. Clearly if $\delta(C) \geqq 0$ and the left hand side of Eq. (9.1) is equal to zero then: $\gamma(C) = \delta(C) = 0$.

For a triangulated 4-manifold C, the Dehn-Sommerville equations assert that $\delta(C) = 10(\chi(C) - 2)$ where $\chi(C)$ is the Euler characteristic of C. In particular, if C is simply-connected then $\delta(C) = 10\, b_2 \geqq 0$ where $b_2 = \mathrm{rank}\, H_2(C) \geqq 0$ is the second Betti-number of C.

The inequality $\delta(P) \geqq 0$ for a simplicial d-polytope P, $d \geqq 5$, is a special case of the "generalized lower bound inequalities" [49, 55] (see Sect. 14). (In fact, the "g-theorem" in its full strength implies that if $\gamma(P) = 0$ then $\delta(P) = 0$. This implies also, by (9.1), Lemma 7.2 for polytopes.)

10. The lower bound conjecture for pseudomanifolds

The lower bound conjecture for pseudomanifolds. (a) If C is a $(d-1)$-pseudomanifold with n vertices, then $f_k(C) \geqq \varphi_k(n,d)$ for $1 \leqq k \leqq d-1$. (b) If equality holds for some k, $d>k>1$, then C is a stacked sphere.

The case $k=d-1$ of part (a) of this conjecture was proved by Klee [42]. The remaining cases are still open.

Definition 10.1. A $(d-1)$-pseudomanifold is *normal* if every face $S \in \mathrm{skel}_{d-3} C$ has a connected link.

Note that the class $\mathscr{C}_d$ of $(d-1)$-pseudomanifolds defined in Sect. 6 is the class of normal $(d-1)$-pseudomanifolds whose singular part has codimension greater than 2. (If $C \in \mathscr{C}_d$ and S is a face of C of size $d-3$, then $\mathrm{lk}(S, C)$ is a triangulated 2-sphere.)

The class of normal pseudomanifolds is closed under taking links of faces. Therefore, the LBT for normal pseudomanifolds reduces by the MPW-reduction to the case $k=1$. As in the proof of Theorem 6.1 the generic d-rigidity of normal $(d-1)$-pseudomanifolds follows from the generic 3-rigidity of normal 2-pseudomanifolds, which are just triangulated 2-manifolds. Part (a) of the LBC for normal pseudomanifolds would thus follow from the following old standing conjecture:

Conjecture G [28, 25]. *The graph of every triangulated 2-manifold is generically 3-rigid*[2].

Remark. Connelly gave in [24] an example of a flexible embedding of a triangulated 2-sphere, and thus refuted the old conjecture (going back to Euler,) that *every* triangulated 2-manifold embedded in R^3 is rigid.

Conjecture G would also imply part (b) of the LBC for normal pseudomanifolds as follows: It is enough to show it for normal 3-pseudomanifolds and then to proceed as in Sect. 7. Conjecture G implies that a 3-pseudomanifold C is generically 4-rigid. Thus if $\gamma(C)=0$ then C must be 4-acyclic, and every link of a vertex of C must be 3-acyclic hence a triangulated 2-sphere.

In order to reduce the LBC for arbitrary pseudomanifolds to the normal case, and also to extend Theorem 1.1 to arbitrary pseudomanifolds with singular set of codimension greater than two, we need the following normalization process [57, p. 83] (compare [29, p. 151, 17]).

Let C be a $(d-1)$-pseudomanifold. Choose a non-empty face S of C of smallest possible dimension k, $k<d-2$ with a non-connected link. "Pull apart" C at S to get a new complex $N_S(C)$ as follows: Create a copy F_i of F for each component K_i of $\mathrm{lk}(F, C)$ so that the link of F_i in the new complex $N_S(C)$ is K_i. Repeated applications of this operation will terminate with a normal $(d-1)$-pseudomanifold $N(C)$.

Direct computation gives:

$$(f_k(C)-\varphi_k(n,d)) > (f_k(N_S(C))-\varphi_k(n,d)) \quad \text{for every } 1 \leqq k < d. \tag{10.1}$$

[2] Whiteley and Graver have recently proved (independently) that all triangulations of the torus are generically 2-rigid. Connelly proved (private communication) that every triangulated 2-manifold admits a generically 3-rigid subdivision

It is likely but unknown that if $N_S(C)$ is generically d-rigid so is C.

Remarks 1. Altshuler constructed in [2] 3-pseudomanifolds such that *none* of their 2-dimensional links are spheres.

2. Note that the lower bound inequalities need not hold for a strongly connected $(d-1)$-dimensional complex, in which every $(d-2)$-face is included in *at least* two facets. A counterexample is two tetrahedra identified along an edge.

11. Manifolds with boundary

In this section we prove a lower bound, conjectured by Björner [17], for the number of k-faces of a triangulated $(d-1)$-manifold *with* boundary as a function of the numbers of interior vertices and boundary vertices. The problem was originated in the study of polytope pairs, see [40, 41, 17, 16]. We first need a few definitions.

A *d-tree* ([34]) is defined inductively as follows: A complete simplicial complex on $d+1$ vertices is a d-tree. If C is a d-tree on the vertex set V, $u \notin V(C)$, and S is any $(d-1)$-face of C, then the simplicial complex obtained from C by adding u to the vertex set V and adding the new facet $S \cup \{u\}$, is a d-tree. A *simple d-tree* ([63]) is a d-tree in which every $(d-1)$-face is included in at most two facets. (I.e., it is a pseudomanifold with boundary.) A simple d-tree is actually a triangulated d-ball. In fact, given a stacked d-polytope P, $d \geqq 3$, P can be divided uniquely into d-simplices $S_1, \ldots, S_m$, such that every $(d-2)$-face of any of these simplices is a face of P. The sets of vertices of these simplices form the set of facets of a simple d-tree. This gives a $1-1$ correspondence between simple d-trees and stacked d-polytopes, $d \geqq 3$.

A d-tree on n-vertices has $\psi_k(n,d) = \binom{d}{k} n - \binom{d+1}{k+1} k$ k-faces ([34]). A simple result of Beineke and Pippert [19] and Björner [17], asserts that every strongly connected d-dimensional simplicial complex C with n vertices has at least $\psi_k(n,d)$ k-faces. This bound applies, in particular, to $(d-1)$-pseudomanifolds with boundary. Beineke and Pippert showed that equality holds only for k-trees. (The earliest result of this type was proved by Klee [40].)

Define a stacked $(d-1)$-ball to be a triangulated $(d-1)$-ball C which is obtained from a simple $(d-1)$-tree by repeated stellar subdivisions of facets. Equivalently, C is a stacked $(d-1)$-ball if C is the antistar of a vertex of a stacked $(d-1)$-sphere.

Let C be a stacked $(d-1)$-ball with n vertices, n_b of them on the boundary and n_i in the interior. (n_b is always at least d.) Thus, C is obtained from a simple $(d-1)$-tree with n_b vertices by n_i applications of stellar subdivisions of facets. Let $\varphi_k^{\mathbf{b}}(n_i, n_b, d)$ be the number of k-faces of C. As easily seen this number depends only on n_i, n_b and d, and is given by formula (1.2):

Theorem 11.1. *Let C be a triangulated $(d-1)$-manifold, $d \geqq 4$, with non-empty boundary. If C has n_i vertices in the interior and n_b vertices in the boundary then*

(i) $f_k(C) \geqq \varphi_k^{\mathbf{b}}(n_i, n_b, d)$, *for every* k, $1 \leqq k \leqq d-1$.

(ii) *If* $f_k(C) = \varphi_k^{\mathbf{b}}(n_i, n_b, d)$ *for some* k, $1 \leqq k \leqq d-1$ *then* C *is a stacked* $(d-1)$*-ball.*

Proof. Let u be a vertex not in C and $D = C \cup (\{u\} * \partial C)$. D is a $(d-1)$-pseudomanifold (without boundary).

Claim 11.2. D *is generically* d*-rigid.*

Proof. Note that for every vertex $v \in D$, different from u, $\mathrm{lk}(v, C)$ is a homology $(d-2)$-sphere, and is generically $(d-1)$-rigid by Theorem 6.1. Choose any tree T in $G(D)$ which contains all vertices of D except u. The conditions of Proposition 6.4 hold and therefore C is generically d-rigid.

Proof of Theorem 11.1 (continued). Put $n = f_0(D)$ $(= n_i + n_b + 1)$. Recall that for $k \geqq 1$, $\gamma_k(D) = f_k(D) - \varphi_k(n, d)$, (Sect. 5.) Put $\gamma_0(D) = 0$. A simple inspection shows that

$$f_k(C) - \varphi_k^{\mathbf{b}}(n_i, n_b, d) = \gamma_k(D) - \gamma_{k-1}(\mathrm{lk}(u, D)). \tag{11.1}$$

Proposition 11.3. *Let* D *be a generically* d*-rigid* $(d-1)$*-pseudomanifold. Then for every vertex* v *of* D, $\gamma_k(D) \geqq \gamma_{k-1}(\mathrm{lk}(v, D))$. *If equality holds then* $\gamma(D) = 0$.

Proof. Recall that $\gamma^i(D) = \sum \{\gamma(\mathrm{lk}(S, D)) : S \in D, |S| = i\}$. Proposition 5.1 asserts that $\gamma_k(D) = \sum_{i=0}^{k-1} w_i(k, d)\, \gamma^i(D)$. The coefficients $w_i(k, d)$ are given by formula (5.6).

We need the following two inequalities:

$$\gamma^i(D) \geqq \gamma^{i-1}(\mathrm{lk}(v, D)) + \gamma^i(\mathrm{lk}(v, D)). \tag{11.2}$$

$$w_i(k, d) + w_{i+1}(k, d) > w_i(k-1, d-1) \quad \text{for every } 1 \leqq i \leqq k-1. \tag{11.3}$$

To prove (11.2) divide the set of $(i-1)$-faces of D into three parts. (a) Those faces S which contain the vertex v, (b) Those faces S which do not contain v but $S \cup \{v\} \in D$ and (c) the remaining $(i-1)$-faces of D. Note that the sum of $\gamma(\mathrm{lk}(S, C)$ over faces in the first family is exactly $\gamma^{i-1}(\mathrm{lk}(v, C)$. If S belongs to the second family and $T = S \cup \{v\}$ then by Theorem 7.3, $\gamma(\mathrm{lk}(S, D)) \geqq \gamma(\mathrm{lk}(T, D)$. But $\mathrm{lk}(T, D) = \mathrm{lk}(S, \mathrm{lk}(v, D))$, and therefore the sum of $\mathrm{lk}(S, C)$ over all faces in the second family is at least $\gamma^i(\mathrm{lk}(v, D))$.

To prove (11.3) use formula (5.6) and note that always $a_k(d) > a_{k-1}(d-1)$ and $1/(k+1)\binom{k}{i-1} + 1/(k+1)\binom{k}{i} = 1/k\binom{k-1}{i-1}$.)

By Proposition 4.1, (11.2) and (11.3),

$$\begin{aligned}
\gamma_k(D) &= \sum_{i=0}^{k-1} w_i(k, d)\, \gamma^i(D) \geqq w_0(k, d)\, \gamma(\mathrm{lk}(v, D)) + \sum_{i=1}^{k-1} w_i(k, d)(\gamma^{i-1}(\mathrm{lk}(v, D)) + \gamma^i(\mathrm{lk}(v, D)) \\
&\geqq \sum_{i=0}^{k-2} (w_i(k, d) + w_{i+1}(k, d))\, \gamma^i(\mathrm{lk}(v, D)) \geqq \sum_{i=0}^{k-2} w_i(k-1, d-1)\, \gamma^i(\mathrm{lk}(v, D)) \\
&= \gamma_{k-1}(\mathrm{lk}(v, D)).
\end{aligned}$$

This gives the required inequality. Since (11.2) is a strict inequality if $\gamma_k(D) = \gamma_{k-1}(\mathrm{lk}(v, D))$ then $\gamma(D) = 0$.

Back to the proof of Theorem 11.1. By Claim 11.2, D is generically d-rigid. Formula (11.1) and Proposition 11.3 give part (i) and show that in case of equality $\gamma(D)=0$. In order to prove part (ii) we need

Claim 11.4. If $\gamma(D)=0$ then D is a stacked $(d-1)$-sphere.

Proof. Let $E=\mathrm{lk}(u,D)$. If E is not connected, apply the normalization procedure described in Sect. 10 to the vertex u. The proof of Claim 11.2 apply for the resulting complex $\hat{D}$ and by formula (10.1), $\gamma(D)>\gamma(\hat{D})\geqq 0$. A contradiction. If E is connected then for $d\geqq 5$, $D\in\mathscr{C}_d$ and by Theorem 7.1, D is a stacked $(d-1)$-sphere. For $d=4$, $\mathrm{lk}(u,D)$ may be any triangulated 2-manifold. However, since $\gamma(D)=0$, $G(D)$ is 4-acyclic and $\mathrm{lk}(u,D)$ must be 3-acyclic. Therefore, $\mathrm{lk}(u,D)$ is a triangulated 2-sphere, $D\in\mathscr{C}_4$ and by Theorem 7.1, D is a stacked 3-sphere.

Proof of Theorem 11.1(ii) (end). By Claim 11.4, D is a stacked $(d-1)$-sphere, hence C is a stacked $(d-1)$-ball.

Remark 11.5. Björner conjectured in [17] that Theorem 11.1(i) holds for every $(d-1)$-pseudomanifold with boundary. Björner proved the case $d=3$ of this conjecture, and showed that the conjecture imply the lower bound inequalities for pseudomanifolds without boundary. It can be shown that the assertion of Theorem 11.1 for arbitrary pseudomanifolds with boundary would also follow from the generic 3-rigidity of all triangulated 2-manifolds (Conjecture G). Our proof can be applied to all normal $(d-1)$-pseudomanifolds with boundary with singular part of codimension 3 or more.

12. A lower bound conjecture for polyhedral manifolds

For a polyhedral complex C, $f_2^k(C)$ is the number of 2-faces of C which are k-gons. For a polyhedral $(d-1)$-dimensional complex C define:

$$\gamma(C)=f_1(P)+\sum_{k\geqq 3}(k-3)f_2^k(P)-dn+\binom{d+1}{2} \tag{12.1}$$

For a d-polytope P, with boundary complex $\mathscr{B}(P)$, $\gamma(P)$ stands for $\gamma(\mathscr{B}(P))$.

Conjecture 12.1. *If P is a polyhedral $(d-1)$-manifold then $\gamma(C)\geqq 0$.*

Perhaps the ultimate generality for conjecture 12.1 (and a convenient context to study this conjecture,) is for "graph manifolds" which are defined in [11]. (See also [12].)

As we already mentioned in Sect. 5, Whiteley's theorem implies the truth of Conjecture 12.1 for boundary complexes of d-polytopes (Theorem 1.4). Previously, it was proved for rational polytopes as a consequence of some deep results in algebraic-geometry. In fact, for such a polytope P, $\gamma(P)$ is the dimension of the second primitive intersection homology group ([29]) of the toric variety associated with P. (See [46], [58, Ch. 4], [59].) However, as was shown by Perles [31, pp. 92–95], there are polytopes which are *not* combinatorially equivalent to rational polytopes.

One difficulty in dealing with Conjecture 12.1 is the fact that the generic d-rigidity of d-polytopal frameworks is not a local property as in the simplicial case (for $d \geqq 4$). In the case of a simplicial d-polytope, $d \geqq 4$, (or a triangulated $(d-1)$-manifold,) the graph induced on a neighborhood of any vertex is already generically d-rigid. This is not the case for d-polytopal frameworks. As an example, let P be a pyramid over the octahedron Q, and consider the neighborhood of any vertex of Q.

For a d-polytopal framework $\mathscr{F}$ based on a d-polytope P it is only for the (highly non-generic) embeddings which realize P as a convex polytope that it is possible to prove "local" infinitesimal rigidity at any vertex [66, p. 456]. This in turn, implies the infinitesimal rigidity and hence the generic rigidity of $\mathscr{F}$. We do not know how to find such a pleasant embedding for arbitrary polyhedral $(d-1)$-manifolds (or even polyhedral $(d-1)$-spheres).

We mention now two corollaries of Theorem 1.4. A polyhedral complex P is *k-simplicial* if every j-face S of P, $j \leqq k$ is a simplex. Theorem 1.4 and the MPW-reduction imply:

Theorem 12.2. *Let P be a k-simplicial d-polytope with n vertices then $f_i(P) \geqq \varphi_i(n,d)$ for $1 \leqq i \leqq k$.*

Let us check now what does Theorem 1.4 says for simple polytopes. If P is a simple polytope with n vertices then $f_1(P) = \frac{dn}{2}$ and $\sum k f_2^k = f_1(P)(d-1)$. The inequality $\gamma(P) \geqq 0$ reduces in this case to:

$$f_2(P) - (d-2) f_1(P) \leqq \binom{d+1}{2}.$$

A posteriori this follows, of course, from Billera, Lee and Stanley's complete characterization of f-vectors of simplicial polytopes.

A d-polytope P is *elementary* if $\gamma(P) = 0$. In [38] we study the function $\gamma(P)$ for polytopes and especially the class of elementary polytopes. We prove there that quotients and faces of elementary polytopes are elementary and that for every face S of an elementary polytope P either S or $\mathrm{lk}(S,P)$ is a simplex. We prove also that the class of elementary polytopes is self-dual. The starting point for the proof is the fairly simple identity: For a 4-polytope P, $\gamma(P) = \gamma(P^*)$. It would be interesting to find a natural isomorphism between the spaces of stresses of the polytopal frameworks based on a 4-polytope P and its dual P^*.

13. Topological subgraphs of triangulated manifolds

In this section we diverge from lower bound theorems. We prove using some of our previous results a property of graphs of traingulated manifolds of a different nature.

A graph H is embeddable in a graph G if G contains some subgraph homeomorphic to H. Grünbaum proved ([31, p. 200]) that K_{d+1}, the complete graph on $d+1$ vertices, is embeddable in the graph of every d-polytope. Barnette proved in [11] that K_{d+1} is embeddable in the graph of every

polyhedral $(d-1)$-manifold. (These results are immediate in the simplicial case.) For a graph G with no vertices of degree 2, TG stands for any graph homeomorphic to G.

Theorem 13.1. Let $d \geqq 4$ be a fixed integer. K_{d+2} is embeddable in the graph of a triangulated $(d-1)$-manifold C iff C is not a stacked $(d-1)$-sphere.

Proof. It is well-known and easy that if C is isomorphic to a stacked d-polytope then K_{d+2} is not embeddable in $G(C)$. In fact, $G(C)$ does not contain K_{d+2} even as a minor.

Let K_d^- denotes a K_d minus an edge. The two vertices of a $TK_d^-(d>4)$ of degree $d-2$ are called *special.*

Lemma 13.2. *Every two non-adjacent vertices of a simplicial* 3*-polytope serve as special vertices of a* TK_5^-; *every two non-adjacent vertices of a stacked* d*-polytope* $(d>3)$ *serve as the special vertices of some* TK_{d+2}^-.

Proof. The first part follows from the 3-connectivity of C, the second part can easily be checked directly.

Proof of Theorem 13.1 (end). Let C be a triangulated $(d-1)$-manifold, and assume that C does not contain a TK_{d+2}. We can assume that C has no vertices of degree d (otherwise we delete them successively). We apply induction on d. Let v be a vertex of C and u, w be a pair of non-adjacent vertices in $\mathrm{lk}(v, C)$. By Lemma 13.2, (and the induction hypothesis if $d>4$,) u and w are the two special vertices of some TK_{d+1}^- in $\mathrm{lk}(v, C)$. Therefore u and w are not adjacent in C nor they are connected in a path that avoids $st(v, C)$. This directly implies that C has no 2-missing faces and no chordless m-gons for $m \geqq 4$. For $d>4$ the induction hypothesis implies that C does not contain missing k-faces for $2<k<d-1$ as well. By Theorem 8.5, C is isomorphic to a stacked sphere.

Remarks. (1) For triangulated 2-manifolds the situation is this. K_5 is not embeddable in any triangulated 2-sphere (stacked or not) by (the easy part of) Kuratowski's Theorem. It is plausible but unknown that K_5 is embeddable in every triangulated 2-manifold which is not a sphere. This will follow from an oldstanding conjecture of Dirac [27] which asserts that K_5 is embeddable in every graph with n vertices and more than $3n-6$ edges. Assuming the the truth of Dirac's conjecture it can be shown that Theorem 13.1 holds for arbitrary $(d-1)$-pseudomanifolds. (While our proof applies only for pseudomanifolds in $\mathscr{C}_d$.)

(2) Grünbaum proved ([31, p. 200]) that for every d-polytope P, $\mathrm{skel}_i(\Delta_d)$ is embeddable in $\mathrm{skel}_i(P)$. *Problem:* For which simplicial d-polytopes is $\mathrm{skel}_i(\Delta_{d+1})$ embeddable in $\mathrm{skel}_i(P)$? By van Kampen-Flores theorem ([31, Ch. 11]) this may never occur if $i \geqq [d+1/2]-1$.

14. Concluding remarks and open problems

14.1. $\gamma(M)$ *and the topology of* M. For a manifold M, (of dimension at least 2,) define $\gamma(M)=\min\{\gamma(C)\colon C \text{ is a triangulation of } M\}$. For every manifold

M, which admits some finite triangulation, $\gamma(M)$ is a non-negative integer, and we have proved that $\gamma(M)=0$ only if M is a sphere. If M is two dimensional $\gamma(M)=3(2-\chi(M))$, where $\chi(M)$ is the Euler characteristics of M.

Walkup proved [63] that (i). For a 3-manifold M which is not a sphere, $\gamma(M)\geqq 10$, and $\gamma(M)=10$ iff M is $S^1\times S^2$ or the corresponding non-orientable "handle". (He also showed that the only triangulations C of these manifolds which satisfy $\gamma(C)=10$ are in $\mathscr{H}^4(1)$.) (ii) For all other 3-manifolds M, $\gamma(M)\geqq 17$ and $\gamma(M)=17$ only when M is the three dimensional projective space.

In [39] we show that for every fixed non-negative integers $d, c, d\geqq 2$, there are only finitely many d-manifolds M for which $\gamma(M)<c$.

We would like to understand how the topology of M affects the invariant $\gamma(M)$. Let $b_i(M)$ denotes the i-th (reduced) Betti number of M. (Thus, $b_i(M)=\operatorname{rank}\tilde{H}_i(M,\mathbb{Z})$.)

Conjecture 14.1. *For a $(d-1)$-manifold M, $d\geqq 4$, $\gamma(M)\geqq b_1(M)\binom{d+1}{2}$.*

If $C\in\mathscr{H}^d(k)$ then $\gamma(C)=b_1(C)\binom{d+1}{2}$. (Are these the only cases of equality?) Walkup proved ([63]) that for every 4-manifold M, (and even every 4-pseudomanifolds in $\mathscr{C}_5$,) $\gamma(M)\geqq\frac{15}{2}(2-\chi(M))$ and equality holds iff $C\in\mathscr{H}^5\left(1-\frac{\chi(M)}{2}\right)$.

The problem of finding $\gamma(M)$ for a $(d-1)$-manifold M resembles the well-known problem of finding $\alpha(M)$ the minimal number of vertices in a triangulation of M (see [50]). Let i, d be fixed integers, $d\geqq 3$, $0<i<\left[\frac{d-1}{2}\right]$. It can be shown quite easily that $\alpha(M)\geqq C(i,d)\,b_i(M)^{\frac{1}{i+1}}$, where $C(i,d)$ is a positive constant depending on i and d (compare [18]). We conjecture that similarly (for $i>0$,) $\gamma(M)\geqq D(i,d)\,b_i(M)^{\frac{1}{i}}$, where $D(i,d)$ is another positive constant depending on i and d.

We would like to know the exact values of $\gamma(S^1\times S^1\times S^1)$, $\gamma(S^2\times S^2)$ and $\gamma(\mathbb{C}P^2)$. Kühnel's 3-neighborly complex projective plane with 9 vertices ([43, 44]) shows that $\gamma(\mathbb{C}P^2)\leqq 6$.

14.2. *The generalized lower bound conjecture*

Let d be a fixed integer, $d\geqq 1$. For a vector of non-negative integers $f=(f_{-1},f_0,f_1,\ldots,f_{d-1}),f_{-1}=1$ define $h[f]=(h_0,h_1,\ldots,h_d)$ where

$$h_k=\sum_{i=0}^{k}(-1)^i\binom{d-k+i}{i}f_{k-i-1}.$$

If f is the f-vector of a simplicial d-polytope or a $(d-1)$-dimensional complex C, $h[f]$ is called the *h-vector* of C. h-vectors of simplicial polytopes were introduced by McMullen and Walkup [49]. This concept plays a crucial part in the combinatorial theory of simplicial polytopes and in several other

areas of combinatorics ([48, 54, 55]). (The original notation was $g_i^{(d)}$ for h_i and $g_i^{(d+1)}$ for $h_{i+1}-h_i$.)

A simplicial d-polytope P is *k-stacked* if P can be triangulated without introducing new j-faces for $j \leqq d-k-1$. A triangulated $(d-1)$-sphere C is k-stacked if it is the boundary of a triangulated d-ball B with the same $(d-k-1)$-skeleton.

McMullen and Walkup suggested in [49] the following far reaching generalization of the lower bound conjecture.

The generalized lower bound conjecture. (i) (The generalized lower bound inequalities.) If P is a simplicial d-polytope and $0 \leqq k \leqq \left[\frac{d}{2}\right]-1$ then $h_{k+1}(P)-h_k(P) \geqq 0$. (ii) If $h_{k+1}(P)-h_k(P)=0$ then P is a k-stacked polytope.

The generalized lower bound inequalities were proved by Stanley [55] as part of his proof of the necessity part of the "g-theorem". Part (ii) is still open.

Note that $\gamma(P)=h_2(P)-h_1(P)$. The Dehn-Sommerville equations (see [48, 56]) assert that $h_i=h_{d-i}$, $0 \leqq i \leqq d$. In particular, if $d=2k+1$ then $h_{k+1}-h_k=0$ for every simplicial d-polytope.

It is widely believed that the assertions of the GLBC and the "g-theorem" are true for arbitrary triangulated spheres. (See [32, 56].) In part (ii), "a k-stacked polytope" should be replaced by "a k-stacked sphere"[3].

For a triangulated $(d-1)$-manifold C define

$$\hat{h}_k(C)=h_k(C)-\binom{d}{k}\sum_{i=0}^{k-1}(-1)^i b_{k-i-1}(C).$$

Schenzel proved ([52], see also [57, pp. 84–85]) that every triangulated $(d-1)$-manifold with boundary C satisfies $\hat{h}_k(C) \geqq 0$, for every $k \geqq 0$.

Conjecture 14.2. Let C be a triangulated $(d-1)$-manifold (without boundary). Then for every k, $0 \leqq k \leqq \left[\frac{d}{2}\right]-1$, $\hat{h}_{k+1}(C)-\hat{h}_k(C) \geqq \binom{d}{k-1} b_k(C)$.

Note that Conjecture 14.1 is a special case of conjecture 14.2. The Dehn-Sommerville equations assert that $\hat{h}_d(C)=0$ and $\hat{h}_i(C)=\hat{h}_{d-i}(C)$, $1 \leqq i<d$.

Many of the results of this paper have obvious analogs in the context of the generalized lower bound inequalities. Proving them seems hard. Only the third proof of Theorem 9.3 and the proof hinted there for Lemma 7.2 extend directly.

14.3. *Flexible weak embeddings.* Let C be a pure simplicial complex. An embedding of the vertices of C into $\mathbb{R}^d$ is a *weak embedding* of C if the images of the vertices of every facet of C are affinely independent. If C is the boundary complex of a stacked d-polytope then every weak embedding of C into $\mathbb{R}^d$ is rigid. Bricard constructed in 1897 ([20]) a flexible weak embedding of the octahedron into $\mathbb{R}^3$.

---

[3] The conjectured equality cases for spheres do not imply the conjecture for polytopes. One consequence from the GLBC would be that for $1 \leqq k \leqq \left[\frac{d}{2}\right]-1$ every d-polytope whose boundary complex is k-stacked (as a sphere) is a k-stacked polytope. We doubt if this is true for $k \geqq \left[\frac{d}{2}\right]$

Conjecture 14.3. *Every non-stacked 3-polytope have a flexible weak embedding.*

14.4. *Rigidity of spaces and separations properties.* All topological spaces mentioned here admit finite triangulations. A topological space X is d-rigid if every triangulation of X is generically d-rigid. A simple sufficient condition for d-rigidity follows from the generic d-rigidity of strongly connected d-dimensional simplicial complexes.

Theorem 14.4. *Let X be a topological space. If for every $Y \subset X$ which separates X, $\dim Y \geqq d-1$ then X is d-rigid.*

Conjecture 14.5. *Let $d \geqq 3$. Let X be a topological space. If for every $Y \subset X$ which separates X, $H_{d-2}(Y) \neq 0$ then X is d-rigid.*

14.5. *Rigidity of tight manifolds.* A triangulated 2-manifold M enbedded in R^3 is *tight* (see [45, 8]) if $M \cap H$ is connected for every half space H of $\mathbb{R}^3$. (This property is known as Banchoff's two piece property and is weaker then tightness in more general contexts.) M is *strictly tight* if it is tight and no two adjacent facets of M are in the same plane (in particular if the vertices of C are in general position).

Strictly tight embeddings of a triangulated 2-sphere C are just realizations of C as the boundary of simplicial polytopes. All these embeddings are rigid by Cauchy's theorem. Connelly proved in [26] that all tight embeddings of a 2-sphere, i.e., embeddings as convex surfaces, are rigid.

Conjecture 14.6. *A tight embedding of a triangulated 2-manifold in $\mathbb{R}^3$ is rigid.*

Conjecture 14.6 implies that triangulated 2-manifolds which admits strictly tight embeddings are generically 3-rigid. Yet, it is hard to suggest this approach for proving Conjecture *G* (Sect. 10) for orientable 2-manifolds. It is not even known whether every orientable triangulated 2-manifold can be geometrically embedded in $\mathbb{R}^3$ (See [32, 13]).

References

1. Alexandrov, A.D.: Convex polyhedra. Moscow, 1950 (Russian); German transl., Konvexe Polyeder. Berlin: Akademie-Verlag 1958
2. Altshuler, A.: 3-pseudomanifolds with preassigned links. Trans. Am. Math. Soc. **241**, 213–237 (1978)
3. Altshuler, A., Perles, M.A.: Quotient polytopes of cyclic polytopes, Part I: Structure and characterization. Isr. J. Math. **36**, 97–125 (1980)
4. Altshuler, A., Steinberg, L.: Neighborly 4-polytopes with 9 vertices. J. Comb. Theory, Ser. A **15**, 270–287 (1973)
5. Asimow, L., Roth, B.: The rigidity of graphs I. Trans. Am. Math. Soc. **245**, 279–289 (1978)
6. Asimow, L., Roth, B.: The rigidity of graphs II. J. Math. Anal. Appl. **68**, 171–190 (1979)
7. Balinski, M.: On the graph structure of convex polyhedra in n-space. Pac. J. Math. **11**, 431–434 (1961)
8. Banchoff, T.F.: Tightly embedded 2-dimensional polyhedral manifolds. Am. J. Math. **87**, 462–472 (1975)
9. Barnette, D.: The minimum number of vertices of a simple polytope. Isr. J. Math. **10**, 121–125 (1971)
10. Barnette, D.: A proof of the lower bound conjecture for convex polytopes. Pac. J. Math. **46**, 349–354 (1973)

11. Barnette, D.: Graph theorems for manifolds. Isr. J. Math. **16**, 62–72 (1973)
12. Barnette, D.: Generalized combinatorial cells and facet splitting. Pac. J. Math. **46**, 349–354 (1973)
13. Barnette, D.: Polyhedral maps on 2-manifolds. In: Convexity and related combinatorial geometry. Kay, D.C., Breen, M. (eds.), pp. 7–19. New York: Marcel Dekker Inc. 1982
14. Barnette, D., Grünbaum, B.: On Steinitz's theorem concerning convex polytopes and on some properties of planar graphs. The many facets of graph theory. Lecture Notes in Math., vol. 110, pp. 27–40. Berlin-Heidelberg-New York: Springer 1969
15. Billera, L.J., Lee, C.W.: A proof of the sufficiency of McMullen's conditions for f-vectors of simplicial polytopes. J. Comb. Theory, Ser. A **31**, 237–255 (1981)
16. Billera, L.J., Lee, C.W.: The number of faces of polytopes pairs and unbounded polyhedra. Eur. J. Comb. **2**, 307–322 (1981)
17. Björner, A.: The minimum number of faces of a simple polyhedron. Eur. J. Comb. **1**, 27–31 (1980)
18. Björner, A., Kalai, G.: An extended Euler-Poincaré formula. (Preprint)
19. Beineke, L.W., Pippert, R.E.: Properties and characterizations of k-trees. Mathematica **18**, 141–151 (1971)
20. Bricard, R.: Mémoire sur la théorie de l'octa dre articulé. J. Math. Pures Appl., IX. Ser. 3, 113–148 (1897)
21. Brönsted, A.: An introduction to convex polytopes. Graduate text in Mathematics, vol. 90. Berlin-Heidelberg-New York: Springer 1983
22. Cauchy, L.: Sur les polygones et les polyèdres, Second Memoire, I. Ecole Polytechnique 9 (1813), 87 (=Oeuvres complètes d'Augustin Cauchy, 2nd sér., Tome 1, 1905, pp. 26–38)
23. Cooper, D., Thurston, W.P.: Triangulating 3-manifolds using 5 vertex link types. Preprint
24. Connelly, R.: A counterexample to the rigidity conjecture for polyhedra. Publ. Math., Inst. Hautes Étud. Sci. **47**, 333–338 (1978)
25. Connelly, R.: Conjectures and open problems in rigidity. Proc. International Congress of Math., Helsinki, 1978
26. Connelly, R.: The rigidity of certain cabled frameworks and the second order rigidity of arbitrarily triangulated convex surfaces. Adv. Math. **37**, 272–299 (1980)
27. Dirac, G.A.: Homomorphism theorems for graphs. Math. Ann. **153**, 69–80 (1964)
28. Gluck, H.: Almost all simply connected closed surfaces are rigid. In: Geometric topology. Lecture Notes in Math., vol. 438, pp. 225–239. Berlin-Heidelberg-New York: Springer 1975
29. Goresky, M., MacPherson, R.: Intersection homology theory. Topology 19, 135–162 (1980)
30. Graver, J.: A combinatorial approach to infinitesimal rigidity. (Preprint)
31. Grünbaum, B.: Convex polytopes. New York: Wiley Interscience 1967
32. Grünbaum, B.: Polytopes, graphs and complexes. Bull. Am. Math. Soc. **76**, 1131–1201 (1970)
33. Grünbaum, B., Shephard, G.C.: Lecture on lost mathematics. Preprint Syracuse University, 1978
34. Harary, F., Palmer, E.M.: On acyclic simplicial complexes. Mathematika **15**, 115–122 (1968)
35. Kalai, G.: Combinatorial problems in convexity and the combinatorics of simplicial complexes. Ph.D. Thesis, Jerusalem, 1983
36. Kalai, G.: Weakly saturated graphs are rigid. Ann. Discrete Math. **20**, 189–190 (1984)
37. Kalai, G.: Heperconnectivity of graphs. Graphs Comb. **1**, 65–80 (1985)
38. Kalai, G.: Rigidity and the lower bound theorem II: Elementary polytopes. (Preprint)
39. Kalai, G.: Rigidity and the lower bound theorem III: Triangulated manifolds with "few edges".(Preprint)
40. Klee, V.: A comparison of primal and dual method of linear programming Num. Math. **9**, 227–235 (1966)
41. Klee, V.: Polytope pairs and their relations to linear programming. Acta Math. **113**, 1–25 (1974)
42. Klee, V.: A d-pseudomanifold with f_0 vertices has at least $df_0-(d-1)(d+2)$ d-simplices. Houston J. Math. **1**, 81–86 (1975)
43. Kühnel, W., Banchoff, T.F.: The 9 vertex complex projective plane. The Math. Intell. **5**, issue 3, 11–22 (1983)
44. Kühnel, W., Lassmann, G.: The unique 3-neighborly 4-manifold with few vertices. J. Comb. Theory, Ser. A **35**, 173–184 (1983)
45. Kuiper, N.H.: Tight embeddings and maps. Submanifolds of geometric type tree in E^N, The Chern Symposium. Singer, I.M. (ed.), pp. 97–145, Berlin Heidelberg New York: Springer 1980

46. MacPherson, R.: Intersection homology. Herman Weyl lectures. (in press) (1987)
47. McMullen, P.: The numbers of faces of simplicial polytopes. Isr. J. Math. **9**, 559–570 (1971)
48. McMullen, P., Shephard, G.C.: Convex polytopes and the upper bound conjecture. London Math. Soc. Lecture Notes Series, vol. 3. London/New York: Cambridge Univ. Press 1971
49. McMullen, P., Walkup, D.W.: A generalized lower bound conjecture for simplicial polytopes. Mathematika **18**, 264–273 (1971)
50. Ringel, G.: Map color theorem. Berlin-Heidelberg-New York: Springer 1974
51. Roth, B.: Rigid and flexible frameworks. Am. Math. Mon. **88**, 6–21 (1981)
52. Schenzel, P.: On the number of faces of simplicial complexes and the purity of Frobenius. Math. Z. **178**, 125–142 (1981)
53. Spanier, E.H.: Algebraic topology. New York: McGraw Hill 1966
54. Stanley R.: The upper bound conjecture and Cohen-Macaulay rings. Stud. Appl. Math. **54**, 135–142 (1975)
55. Stanley, R.: The number of faces of simplicial convex polytopes. Adv. Math. **35**, 236–238 (1980)
56. Stanley, R.: The number of faces of simplicial polytopes and spheres. Discrete geometry and convexity, Goodman, J.E., Lutwak, E., Malkevitch, J., Pollack, R. (eds.), pp. 212–223. New York: Ann. NY Acad Sci 1985
57. Stanley, R.: Interactions between commutative algebra and combinatorics. Boston: Birkhäuser 1983
58. Stanley, R.: Enumerative combinatorics. Vol. I, Wadsworth, Monterey, 1986
59. Stanley, R.: Generalized h-vectors, intersection cohomology of toric varieties, and related results. Proc. Japan-USA workshop on commutative algebra and combinatorics (to appear)
60. Stoker, J.J.: Geometric problems concerning polyhedra in the large. Commun. Pure Appl. Math. **21**, 119–168 (1968)
61. Steinitz, E., Rademacher, H.: Vorlesungen über die Theorie der Polyeder. Berlin-Göttingen: Springer 1934
62. Tay, T.S., Whiteley, W.: Generating all isostatic frameworks. Struct. Topology **11**, 21–70 (1985)
63. Walkup, D.: The lower bound conjecture for 3- and 4-manifolds. Acta Math. **125**, 75–107 (1970)
64. Welsh, D.: Matroid theory. London: Academic Press 1976
65. Whiteley, W.: Cones, infinity and 1-story buildings. Struct. Topology **8**, 53–70 (1983)
66. Whiteley, W.: Infinitesimally rigid polyhedra I. Statics of Frameworks. Trans. Am. Math. Soc. **285**, 431–465 (1984)
67. Gromov, M.: Partial differential relations. Berlin Heidelberg New York: Springer 1986

Oblatum 26-VI-1985 & 3-IX-1985 & 3-III-1986

Note added in proof

The basic relation between the LBT and rigidity is observed independently by M. Gromov in [67, Ch. 2.4.10]. Moreover, Gromov presents a purly combinatorial "substitute" for rigidity. Using Gromov's "rigidity" concept combined with the results of Sections 7–11, it is possible to prove Theorems 1.1 and 11.1 for arbitrary pseudomanifolds.

Invent. math. 88, 153–164 (1987)

Inventiones mathematicae

© Springer-Verlag 1987

Cohen-Macaulay modules on hypersurface singularities I

Horst Knörrer

Mathematisches Institut, Universität Bonn, Wegelerstr. 10, D-5300 Bonn 1, Federal Republic of Germany

Egbert Brieskorn gewidmet

The main goal of this paper and the subsequent paper [8] by R.-O. Buchweitz, G.M. Greuel and F.-O. Schreyer is to prove the following characterization of the simple hypersurface singularities:

Theorem. *Let P be a regular analytic k-algebra,* i.e. *a formal or convergent (if k is a complete non trivial valuated field) power series ring $k\langle x_0 \ldots, x_n\rangle$, where k is an algebraically closed field of characteristic different from two. Let f be a non-zero element in the maximal ideal of P, and denote by $R:=P/(f)$ the associated hypersurface ring. Then the following statements are equivalent*

a) *There are only finitely many isomorphism classes of indecomposable maximal Cohen-Macaulay modules*[1] *over R,*

b) *R is the local ring of a simple hypersurface singularity in the sense of* [1] *resp.* [16].

In this paper we will prove the implication "(b)⇒(a)" of the theorem. It is well known that an n-dimensional simple hypersurface singularity is isomorphic to a singularity with an equation $g(x, y)+z_1^2+\ldots+z_{n-1}^2=0$ where $g(x, y)=0$ defines a simple plane curve singularity (cf. [8] 0.2, 0.3). Therefore we will study in general the relation between maximal Cohen-Macaulay modules of stably equivalent singularities, i.e. of singularities whose equations only differ by sums of squares in independent variables. In particular we show that there is a canonical bijection between the sets of isomorphism classes of maximal Cohen-Macaulay modules over the singularities given by the equations $f(x_0, \ldots, x_n)=0$ and $f(x_0, \ldots, x_n)+y^2+z^2=0$ (theorem 3.1).

This construction also yields a description of the Auslander-Reiten quivers of maximal Cohen-Macaulay modules over simple hypersurface singularities. (For the theory of Auslander-Reiten quivers and almost split sequences we refer to [5], [3], or the survey article [12].) The Auslander-Reiten quivers of the two- (resp. one-) dimensional simple hypersurface singularities have been computed by M. Auslander and I. Reiten [3], [6], resp. E. Dietrich, A. Wiede-

[1] i.e. R-modules M with depth $M=\dim R$

mann [9] for $\operatorname{char} k=0$ and K. Kiyek, G. Steinke [16] for $\operatorname{char} k \neq 0$. In the case of curves and $\operatorname{char} k=0$ these quivers coincide with certain graphs associated to representations of finite reflection groups in $GL(2,k)$ (as described in [13], 3.7), and in chapter 4 will give a conceptual proof for this coincidence of graphs.

This isomorphism of graphs had been observed by J.-L. Verdier and – together with the paper [14] – was the starting point for this work. R.-O. Buchweitz pointed out – among many other things – how to use matrix factorizations in this context. I also want to thank M. Auslander, H. Esnault, Chr. Riedtmann, E. Viehweg and A. Wiedemann for many interesting and fruitful discussions.

1. Matrix factorizations

For later use wer recall some facts about the structure of maximal Cohen-Macaulay modules[2] over $R=P/(f)$. These results are due to D. Eisenbud [10], chapter 6.

Definition. A matrix factorization of f is a pair $(\varphi: F \to G, \psi: G \to F)$ of morphisms between free P-modules such that $\varphi \circ \psi = f \cdot \mathrm{id}_G$, $\psi \circ \varphi = f \cdot \mathrm{id}_F$. A morphism between two matrix factorizations $(\varphi: F \to G, \psi: G \to F)$ and $(\varphi': F' \to G', \psi': G' \to F')$ for f is a pair of morphisms $(\alpha: F \to F', \beta: G \to G')$ which makes the following diagram commute

$$\begin{array}{ccccc} F & \xrightarrow{\varphi} & G & \xrightarrow{\psi} & F \\ \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \alpha} \\ F' & \xrightarrow{\varphi'} & G' & \xrightarrow{\psi'} & F' \end{array}$$

The matrix factorizations of f from an additive category which we denote by $\mathcal{MF}(f)$. We will call a sequence in $\mathcal{MF}(f)$ exact if the induced sequences of free P-modules are exact.

For a matrix factorization (φ, ψ) of f we put $\operatorname{cok}(\varphi, \psi) := \operatorname{coker} \varphi$, this is a module over $R=P/(f)$. If $(\alpha, \beta): (\varphi, \psi) \to (\varphi', \psi')$ is a morphism of matrix factorizations then β induces a morphism $\operatorname{cok}(\alpha, \beta): \operatorname{coker} \varphi \to \operatorname{coker} \varphi'$. So cok is a functor from $\mathcal{MF}(f)$ to the category of R-modules. One easily sees that it maps exact sequences to exact sequences. Furthermore it is clear that the trivial matrix factorization $(\times 1: P \to P, \times f: P \to P)$ is mapped to zero. So let $\mathcal{J}$ be the ideal in $\mathcal{MF}(f)$ (in the sense of [18] 2.2) generated by the morphisms that factor over $(1, f)$.

Theorem 1.1 ([10] chapter 6). *The functor* cok *induces an equivalence between the category* $\mathcal{MF}(f)/\mathcal{J}$ *and the category* MCM(R) *of maximal Cohen-Macaulay modules over* R

Corollary 1.2. *An exact sequence* $0 \to X_1 \to X_2 \to X_3 \to 0$ *in* $\mathcal{MF}(f)$ *is almost split if and only if the induced sequence* $0 \to \operatorname{cok} X_1 \to \operatorname{cok} X_2 \to \operatorname{cok} X_3 \to 0$ is almost split in MCM(R).

[2] In the sequel we abbreviate "maximal Cohen-Macaulay module" by "MCM"

Let T: $\mathcal{MF}(f) \to \mathcal{MF}(f)$ be the involutive transformation $(\varphi, \psi) \mapsto (\psi, \varphi)$. Then cok TX is isomorphic to the first syzygy module of cok X.

Lemma 1.3. *Let $X \in \mathcal{MF}(f)$ such that $TX \cong X$. Then X is isomorphic to a matrix factorization (φ_0, φ_0) where $\varphi_0 : P^r \to P^r$ is a morphism with $\varphi_0^2 = f \cdot \mathrm{id}$.*

Proof. We may assume that $X = (\varphi\colon P^r \to P^r,\ \psi\colon P^r \to P^r)$ is indecomposable. By assumption there is an isomorphism $(\alpha, \beta)\colon (\varphi, \psi) \to (\psi, \varphi)$. Then $(\beta \circ \alpha,\ \alpha \circ \beta)$ is an automorphism of X. As X is indecomposable it follows from [20] 2.19 and Theorem 1.1 that the quotient of the endomorphism ring $\mathrm{End}(X)$ of X by its radical I is isomorphic to k. So we may assume that

$$(\beta \circ \alpha, \alpha \circ \beta) = (\mathrm{id}, \mathrm{id}) + (\rho_1, \rho_2) \qquad \text{with } (\rho_1, \rho_2) \in I.$$

This implies that $\alpha \circ \rho_1 = \rho_2 \circ \alpha$, $\beta \circ \rho_2 = \rho_1 \circ \beta$. Now choose (convergent) series for $(1+x)^{-1/2}$ and put

$$\alpha' := \alpha_0 (\mathrm{id} + \rho_1)^{-1/2} = (\mathrm{id} + \rho_2)^{-1/2} \circ \alpha$$

$$\beta' := \beta_0 (\mathrm{id} + \rho_2)^{-1/2} = (\mathrm{id} + \rho_1)^{-1/2} \circ \beta.$$

Then (α', β') is a morphism from X to TX and $\alpha' \circ \beta' = \beta' \circ \alpha' = \mathrm{id}$. Now choose an automorphism γ of P^r such that $\gamma^2 = \alpha'$.

Then $\varphi_0 := \gamma \circ \psi \circ \gamma = \gamma^{-1} \circ \varphi \circ \gamma^{-1}$ has the desired properties. □

2. Double branched covers

We now study the relation between maximal Cohen-Macaulay modules over $R = P/(f)$ and over $R_1 := P_1/(f + y^2)$, where $P_1 := P\langle y \rangle$. The group $\mathbb{Z}/2\mathbb{Z}$ acts on R_1 by σ: $(x, y) \mapsto (x, -y)$. We denote the twisted group ring by $R_1[\sigma]$, i.e. $R_1[\sigma] = R_1 \oplus R_1 \cdot \sigma$ with the multiplication

$$(r_1 + r_2 \cdot \sigma) \cdot (r_1' + r_2' \cdot \sigma) = r_1 r_1' + r_2 \sigma(r_2') + (r_1 r_2' + r_2 \sigma(r_1')) \cdot \sigma.$$

Modules over $R_1[\sigma]$ correspond to R_1-modules M with an action of the involution σ such that $\sigma(r \cdot m) = \sigma(r) \cdot \sigma(m)$ for all $r \in R_1$, $m \in M$. We call an $R_1[\sigma]$ module maximal Cohen-Macaulay if it is an MCM over R_1, and denote the category of such modules by $\mathrm{MCM}_\sigma(R_1)$.

R_1 itself admits two σ-actions, namely $\varphi \mapsto \varphi \circ \sigma$ and $\varphi \mapsto -\varphi \circ \sigma$. We denote the corresponding $R_1[\sigma]$-modules by R_1^+ (or sometimes just R_1) and R_1^-. Then $R_1[\sigma] = R_1^+ \oplus R_1^-$ as $R_1[\sigma]$-module, and R_1^+ and R_1^- are the only indecomposable projective MCM's over $R_1[\sigma]$. For $M \in \mathrm{MCM}_\sigma(R_1)$ we denote by M^σ resp. M^a the set of σ-invariant (resp. σ-antiinvariant) elements of M. They are both MCM's over $R_1^\sigma \cong P$, so M^σ and M^a are free P-modules of equal rank.

If $M \in \mathrm{MCM}_\sigma(R_1)$ then multiplication by y resp. $-y$ gives a pair of P-linear maps

$$\varphi\colon M^\sigma \to M^a, \qquad \psi\colon M^a \to M^\sigma.$$

Obviously (φ, ψ) is a matrix factorization of f; we denote it by $\mathcal{A}(M)$. $\mathcal{A}$ defines a functor from $\mathrm{MCM}_\sigma(R_1)$ to $\mathcal{MF}(f)$, and we have

Proposition 2.1. $\mathscr{A}: \mathrm{MCM}_\sigma(R_1) \to \mathscr{MF}(f)$ *is an equivalence of categories.*

Proof. If $(\varphi: F \to G, \psi: G \to F)$ is a matrix factorization of f we put

$$M := F \oplus G \text{ as } P\text{-module}$$

A structure of an $R_1[\sigma]$-module on M is defined by

$$\begin{aligned} y \cdot (v_1, v_2) &= (-\psi(v_2), \varphi(v_2)) \\ \sigma \cdot (v_1, v_2) &= (v_1, -v_2) \end{aligned} \qquad \text{for } (v_1, v_2) \in F \oplus G.$$

M is an MCM over $R_1[\sigma]$ since it is free as a P-module. This construction describes the functor inverse to $\mathscr{A}$. □

2.2. Remarks. (i) $\mathscr{A}(R_1^+) = (1, f)$, and $\mathscr{A}(M \otimes R_1^-) \cong T \circ \mathscr{A}(M)$ for all $M \in \mathrm{MCM}_\sigma(R_1)$

(ii) The functor $\mathrm{cok} \circ \mathscr{A}: \mathrm{MCM}_\sigma(R_1) \to \mathrm{MCM}(R)$ is explicitly described by $M \mapsto M^a/y\, M^\sigma = M/R_1 \cdot M^\sigma$. In particular we have

$$M/y \cdot M \cong \mathrm{cok} \circ \mathscr{A}(M) \oplus \mathrm{cok} \circ \mathscr{A}(M \otimes R^-)$$

for all $M \in \mathrm{MCM}_\sigma(R_1)$.

If R has an isolated singularity then almost split sequences exist in the categories $\mathrm{MCM}(R)$, $\mathrm{MCM}_\sigma(R_1)$ (cf. [4]).

So from (1.1), (1.2) and (2.1) we get

Corollary 2.3. *Suppose that R has an isolated singularity. Then the Auslander-Reiten quiver of* $\mathrm{MCM}(R)$ *is isomorphic to the full subgraph of the Auslander-Reiter quiver of* $\mathrm{MCM}_\sigma(R_1)$ *obtained by deleting the vertex corresponding to* R_1^+.

To study the relation between MCM's over $R_1[\sigma]$ and R_1 we consider the following functors:

The forget-functor $F: \mathrm{MCM}_\sigma(R_1) \to \mathrm{MCM}(R_1)$, its adjoint

$$\begin{aligned} \mathrm{ad}: \mathrm{MCM}(R_1) &\to \mathrm{MCM}_\sigma(R_1), \\ M &\mapsto R_1[\sigma] \otimes_{R_1} M \end{aligned}$$

the restriction $\mathrm{rest}: \mathrm{MCM}(R) \to \mathrm{MCM}(R)$,

$$M \mapsto M/y\,M.$$

On the level of matrix factorizations we have the functors

Rest: $\mathscr{MF}(f + y^2) \to \mathscr{MF}(f)$ which sends a matrix factorization

$$(\Phi: F_1 \to F_2, \psi: F_2 \to F_1)$$

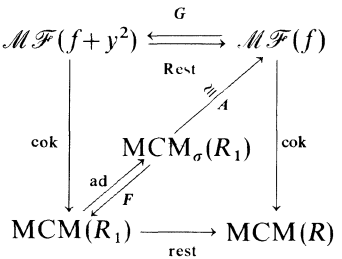
to the matrix factorization of f defined by the induced maps between F_1/yF_1 and F_2/yF_2.

G: $\mathscr{MF}(f) \to \mathscr{MF}(f + y^2)$ mapping a matrix factorization $(\varphi: F_1 \to F_2$, $\psi: F_2 \to F_1)$ of f to the matrix factorization (Φ, Φ) of $f + y^2$ where Φ is the map

$$\Phi = \begin{pmatrix} y \cdot \mathrm{id} & \psi \\ \varphi & -y \cdot \mathrm{id} \end{pmatrix} : (F_1 \oplus F_2) \otimes_P P_1 \to (F_1 \oplus F_2) \otimes_P P_1.$$

To a morphism (α, β) in $\mathcal{MF}(f)$ the functor G associates the morphism $\left(\begin{pmatrix}\alpha & 0\\ 0 & \beta\end{pmatrix}, \begin{pmatrix}\alpha & 0\\ 0 & \beta\end{pmatrix}\right)$.

In total we get the following – non commutative – diagram



Proposition 2.4. *The following functors are pairwise equivalent*:

i) rest ∘ cok *and* cok ∘ Rest	*from* $\mathcal{MF}(f+y^2)$	*to* $\mathrm{MCM}(R)$
ii) cok ∘ $\mathcal{A}$ ∘ ad *and* rest	*from* $\mathrm{MCM}(R_1)$	*to* $\mathrm{MCM}(R)$
iii) cok ∘ G *and* $F \circ \mathcal{A}^{-1}$	*from* $\mathcal{MF}(f)$	*to* $\mathrm{MCM}(R_1)$

Proof. The equivalence (i) is obvious, and (ii) follows from remark 2.2.ii. To prove the third part of the proposition let $(\varphi\colon P^r \to P^r,\ \psi\colon P^r \to P^r)$ be a matrix factorization of f. By construction

$$F \circ \mathcal{A}^{-1}(\varphi, \psi) = P^r \oplus P^r \text{ as } P\text{-module},$$

and the R_1-module structure on it is given by

$$y \cdot (v_1, v_2) = (-\psi(v_2), \varphi(v_1))$$

(cf. the proof of proposition 2.1). On the other hand $\mathrm{cok} \circ G(\varphi, \psi)$ is the quotient of $R_1^r \oplus R_1^r$ by the R_1-submodule generated by the elements $(y \cdot v, \psi(v))$ and $(\varphi(v), -y \cdot v)$ with $v \in R_1^r$. So if we split R_1 as $R_1 = P \oplus y \cdot P$, then $\mathrm{cok} \circ G(\varphi, \psi)$ can be described as follows:

Let $I \subset P^r \oplus P^r \oplus P^r \oplus P^r$ be the P-submodule generated by the elements $(0, v, \psi(v), 0)$, $(-fv, 0, 0, \psi(v))$, $(\varphi(v), 0, 0, -v)$, $0, \varphi(v), fv, 0)$ with $v \in P^r$. Then $\mathrm{cok} \circ G(\varphi, \psi) \cong P^r \oplus P^r \oplus P^r \oplus P^r / I$ as a P-module, and the R_1-module structure on it is given by $y \cdot (v_1, \ldots, v_4) = (-fv_2, v_1, -fv_4, v_3)$.

One now computes directly that

$$(v_1, v_2, v_3, v_4) \to (v_3 - \psi(v_2), v_1 + \varphi(v_4))$$

defines an isomorphism between the R_1-modules $\mathrm{cok} \circ G(\varphi, \psi)$ and $F \circ \mathcal{A}^{-1}(\varphi, \psi)$ as described above. This construction yields the desired equivalence of functors. □

The involution σ acts on $\mathcal{MF}(f+y^2)$ in the obvious way, and clearly $\mathrm{cok} \circ \sigma(Y) \cong \sigma^*(\mathrm{cok}(Y))$ for all $Y \in \mathcal{MF}(f+y^2)$. Furthermore the involutive transformation T: $(\varphi, \psi) \to (\psi, \varphi)$ described in chapter 1 acts both on $\mathcal{MF}(f)$ and

$\mathcal{MF}(f+y^2)$ and is clearly compatible with the functor Rest. Also $G \circ T$ is equivalent to $T \circ G = G$, the equivalence being given by

$$\begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}.$$

Lemma 2.5. i) $\mathrm{ad} \circ F$ *is equivalent to the functor* $M \mapsto M \oplus (M \otimes R_1^-)$.

$F \circ \mathrm{ad}$ *is equivalent to the functor* $M \mapsto M \oplus \sigma^* M$

ii) $\mathrm{Rest} \circ G$ *is equivalent to* $\mathrm{id} \oplus T$ *on* $\mathcal{MF}(f)$

$G \circ \mathrm{Rest}(Y) \cong Y \oplus \sigma(Y)$ *for all non-trivial*[3] *indecomposable matrix factorizations* Y *of* $f+y^2$.

Proof. i) As R_1-module $\mathrm{ad} \circ F(M)$ is equal to $M \oplus \sigma^*(M)$, and the action of σ on it is given by $(v_1, v_2) \mapsto (v_2, v_1)$. Then the map

$$\begin{aligned} M \oplus (M \otimes R_1^-) &\to M \oplus \sigma^*(M) \\ (x, y) &\mapsto (x+y, \sigma(x) - \sigma(y)) \end{aligned}$$

defines an isomorphism between the $R_1[\sigma]$-modules $M \oplus (M \otimes R_1^-)$ and $\mathrm{ad} \circ F(M)$. This construction is clearly functorial and gives the desired equivalence of functors.

The equivalence between $F \circ \mathrm{ad}$ and $\mathrm{id} \oplus \sigma^*$ is obtained by splitting $R_1[\sigma] = R_1 \oplus R_1 \cdot \sigma$ as R_1-module.

ii) The first claim is easily checked directly, and the second claim follows from (i) above Theorem 1.1 and Proposition 2.4. □

Lemma 2.6. i) *An element* $Y \in \mathcal{MF}(d+y^2)$ *is isomorphic to a matrix factorization in the image of* G *if and only if* $\sigma Y \cong Y$.

ii) $TY \cong \sigma(Y)$ *for all non-trivial indecomposable* $Y \in \mathcal{MF}(f+y^2)$.

Proof. i) One easily sees that $\sigma \circ G(X) \cong G(X)$ for all $X \in \mathcal{MF}(f)$. Conversely let $Y \in \mathcal{MF}(f+y^2)$ such that $\sigma Y \cong Y$. By (2.5.i) we may assume that Y is indecomposable. We want to show that $M := \mathrm{cok}(Y)$ admits the structure of an $R_1[\sigma]$-module (i.e. that M is the image of F), the claim then follows from (2.4.iii).

By assumption there is an isomorphism $\Psi: M \to \sigma^*(M)$. It induces an isomorphism $\Psi^*: \sigma^*(M) \to M$. As M is indecomposable we may – as in the proof of Lemma 1.3 – assume that

$$\Psi^* \circ \Psi = \mathrm{id} + \rho \text{ with } \rho \text{ in the radical of } \mathrm{End}(M)$$

Now choose again a (convergent) power series of $(1+x)^{-1/2}$ and put

$$\tilde{\Psi} := \Psi \circ (\mathrm{id} + \rho)^{-1/2}$$

Then $\tilde{\Psi}^* \circ \Psi = \mathrm{id}$, i.e. $\tilde{\Psi}$ defines the structure of an $R_1[\sigma]$-module on M.

ii) $M := \mathrm{cok}(Y)$ is a free module over P – say of rank r. If we choose a P-basis for this module then multiplication by y is described by an $(r \times r)$-matrix A with coefficients in P such that $A^2 = -f \cdot \mathrm{id}$. One sees that $(A + y \cdot \mathrm{id}, A - y \cdot \mathrm{id})$ is also a matrix factorisation for M. So Y and $(A + y \cdot \mathrm{id}, A - y \cdot \mathrm{id})$ are isomorphic up to trivial summands, and (2.6.ii) is obvious.

[3] i.e. not isomorphic to $(1, f+y^2)$ or $(f+y^2, 1)$

Proposition 2.7. i) *Let $X \in \mathcal{MF}(f)$ be indecomposable and non-trivial. Then $G(X)$ is decomposable if and only if $TX \cong X$. In this case $G(X)$ is isomorphic to a matrix factorization of the form $Y \oplus TY$ with indecomposable Y such that $Y \ncong TY$.*

ii) *Let $X \in \mathcal{MF}(f+y^2)$ be indecomposable. Then* Rest(Y) *is decomposable if and only if $TY \cong Y$* (i.e. *if Y is in the image of G*). *In this case* Rest(Y) *is of the form $X \oplus TX$ with indecomposable X such that $X \ncong TX$.*

Proof. First we show that for a non-trivial indecomposable $X \in \mathcal{MF}(f)$ the matrix factorization $G(X)$ decomposes if and only if $TX \cong X$: If $TX \cong X$ we may assume by Lemma 1.3 that $X = (\varphi, \varphi)$ with $\varphi^2 = f \cdot \mathrm{id}$. Then $G(X)$ is isomorphic to

$$(\varphi + \mathrm{iy} \cdot \mathrm{id}, \varphi - \mathrm{iy} \cdot \mathrm{id}) \oplus (\varphi - \mathrm{iy} \cdot \mathrm{id}, \varphi + \mathrm{iy} \cdot \mathrm{id}).$$

Conversely if GX has a non-trivial summand Y then Rest Y is a summand of Rest $\circ\, G(X) \cong X \oplus TX$ (Lemma 2.5.ii). So Rest(Y) is isomorphic to X or TX. On the other hand $T \circ$ Rest $(Y) \cong$ Rest $(TY) \cong$ Rest $(\sigma Y) \cong$ Rest(Y), hence $X \cong TX$.

Next we remark that for a non-trivial indecomposable $Y \in \mathcal{MF}(f+y^2)$ the matrix factorization Rest(Y) decomposes if and only if $TY \cong Y$: If $Y \cong TY$ then Y is the image of G by (2.6), so Rest(Y) decomposes by (2.5.ii). The converse is shown as above.

To complete the proof of (i) let $X \in \mathcal{MF}(f)$ be a non-trivial matrix factorization such that $TX \cong X$. We write $GX = Y \oplus TY$ as above. Now Rest $\circ\, G(X) \cong X \cong X \oplus TX$ has only two indecomposable summands, so Y is indecomposable. If Y were isomorphic to TY then Rest(Y) would decompose by what we showed above, and Rest $\circ\, G(X) =$ Rest $Y \oplus$ Rest(TY) would then have at least four summands.

The proof of ii) is completed in the same way. □

Corollary 2.8. *There are only finitely many isomorphism classes of indecomposable maximal Cohen-Macaulay modules over R if and only if this is the case for R_1.*

Proof. By Proposition 2.1 and Theorem 1.1 the number of indecomposable objects in MCM(R) and MCM$_\sigma(R_1)$ only differs by one. But by Lemma 2.5.i and the Krull-Schmidt-theorem MCM$_\sigma(R_1)$ has finitely many isomorphism classes of indecomposable objects if and only if this is true for MCM(R_1). □

Proposition 2.7 allows a more detailed control over the relation between MCM's over R and R_1, this will be used in Chapter 3. Concerning morphism, we note the following two observations:

Corollary 2.9. *Let Φ: $M \to N$ be a morphism in* MCM(R_1) *such that* rest(Φ): rest$(M) \to$ rest(N) *factors over a projective* (=*free*) *module in* MCM(R). *Then Φ itself factors over a projective module.*

Proof. By (2.4.ii) and Theorem 1.1 the morphism $\mathcal{A} \circ \mathrm{ad}(\Phi)$ factors over projectives, hence by (2.5.i) also the map

$$\Phi \oplus \sigma \circ \Phi \circ \sigma : M \oplus \sigma^*(M) \to N \oplus \sigma^*(N)$$

factors over a projective module. Now the claim is immediate. □

Corollary 2.10. *If* $0 \to M_1 \to M_2 \to M_3 \to 0$ *is an almost split sequence in* $\mathrm{MCM}_\sigma(R_1)$ *(resp.* $\mathrm{MCM}(R_1)$*) then the induced sequence*

$$0 \to F(M_1) \to F(M_2) \to F(M_3) \to 0$$

(resp. $0 \to \mathrm{ad}(M_1) \to \mathrm{ad}(M_2) \to \mathrm{ad}(M_3) \to 0$*) is a direct sum of almost split sequences in* $\mathrm{MCM}(R_1)$ *(resp.* $\mathrm{MCM}_\sigma(R_1)$*).*

Proof. This follows from (2.5.i) as in [19] Theorem 3.8. □

3. Periodicity

We want to compare maximal Cohen-Macaulay modules over R and over the ring $R_2 := P_2/(f+y^2+z^2)$, where $P_2 := P\langle y, z\rangle = P_1\langle z\rangle$. For this purpose we write $u = z+iy$, $v = z-\mathrm{iy}$ (so that $y^2+z^2 = uv$) and introduce the functor

$$H: \mathcal{MF}(f) \to \mathcal{MF}(f+uv)$$

that associates to $(\varphi, \psi) \in \mathcal{MF}(f)$ the matrix factorization

$$\left(\begin{pmatrix} u & \psi \\ \varphi & -v \end{pmatrix}, \begin{pmatrix} v & \psi \\ \varphi & -u \end{pmatrix}\right)$$

of $f+uv$, and to a morphism (α, β) in $\mathcal{MF}(f)$ the morphism

$$\left(\begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}, \begin{pmatrix} \alpha & 0 \\ 0 & \beta \end{pmatrix}\right) \text{ in } \mathcal{MF}(f+uv).$$

To formulate the properties of this functor let $\underline{\mathrm{MCM}}(R)$ be the quotient of the category $\mathrm{MCM}(R)$ by the ideal that is generated by the morphism factoring over projective (=free) modules. By Theorem 1.1 this category is also the quotient of $\mathcal{MF}(f)$ by the ideal of morphism factoring over projectives. So H induces a functor from $\underline{\mathrm{MCM}}(R)$ to the corresponding category $\underline{\mathrm{MCM}}(R_2)$, and we have

Theorem 3.1. *H induces an equivalence between the categories $\underline{MCM}(R)$ and $\underline{\mathrm{MCM}}(R_2)$.*

This immediately implies

Corollary 3.2. *The stable Auslander-Reiten quivers of* $\mathrm{MCM}(R)$ *and* $\mathrm{MCM}(R_2)$ *are isomorphic.*

The rest of this chapter is devoted to the proof of Theorem 3.1. For convenience we denote by G_1 resp. Rest_1 the functors between $\mathcal{MF}(f)$ and $\mathcal{MF}(f+y^2)$ described in Chapter 2, and by G_2 resp. Rest_2 the corresponding functors between $\mathcal{MF}(f+y^2)$ and $\mathcal{MF}(f+y^2+z^2)$. Furthermore put

$$\rho := \mathrm{Rest}_1 \circ \mathrm{Rest}_2 : \mathcal{MF}(f+y^2+z^2) \to \mathcal{MF}(f).$$

One easily checks that

$$G_2 \circ G_1 \simeq H \oplus T \circ H \tag{3.3}$$

$$\rho \circ H \simeq \mathrm{id} \oplus T \tag{3.4}$$

$$T \circ H \simeq H \circ T \tag{3.5}$$

Lemma 3.6. *H induces a bijection between the sets of isomorphism classes of non-trivial indecomposable matrix factorizations of f and $f+y^2+z^2$.*

Proof. Let $X \in \mathcal{MF}(f)$ be non-trivial and indecomposable. $G_2 \circ G_1(X)$ splits into precisely two indecomposable summands by Proposition 2.7.i, so $H(X)$ is indecomposable by (3.3). If X' is another matrix factorization of f with $H(X) \cong H(X')$ then by (3.4) we have $X' \cong X$ or $X' = TX$. Assume that $X' \ncong X$. By iterated application of proposition 2.7.i we see first that $G_1(X)$ is indecomposable and then that $G_2 \circ G_1(X)$ splits into two non-isomorphic summands. So by (3.3) we have $H(X) \ncong T \circ H(X)$. On the other hand by (3.5) we have $T \circ H(X) \cong H \circ T(X) \cong H(X') \cong H(X)$, which is a contradiction.

Finally we have to show that each non-trivial indecomposable object in $\mathcal{MF}(f+y^2+z^2)$ is a direct summand of an element in the image of H. But this follows directly from (3.3) and Lemma 2.5. □

For the proof of Theorem 3.1 we still have to show that for any two matrix factorizations $X = (\varphi\colon F_1 \to F_2,\ \psi\colon F_2 \to F_1)$, $X' = (\varphi'\colon F_1 \to F'_2,\ \psi'\colon F'_2 \to F'_1)$ of f the functor H induces an isomorphism between the quotient of $\mathrm{Hom}(X, X')$ by the subspace of morphisms factoring over projective (= sum of trivial) matrix factorizations and the corresponding quotient of $\mathrm{Hom}(H(X), H(X'))$. That this map is injective follows directly from (3.4). The key step in proving its surjectivity is

Lemma 3.7. *Let* $(A, B) = \left(\begin{pmatrix} \alpha_1 & \alpha_2 \\ \alpha_3 & \alpha_4 \end{pmatrix}, \begin{pmatrix} \beta_1 & \beta_2 \\ \beta_3 & \beta_4 \end{pmatrix}\right)$ *be a morphism from $H(X)$ to $H(X')$. Then*

$$\rho(A, B) =: \left(\begin{pmatrix} \bar{\alpha}_1 & \bar{\alpha}_2 \\ \bar{\alpha}_3 & \bar{\alpha}_4 \end{pmatrix}, \begin{pmatrix} \bar{\beta}_1 & \bar{\beta}_2 \\ \bar{\beta}_3 & \bar{\beta}_4 \end{pmatrix}\right)$$

is a morphism in $\mathcal{MF}(f)$ from

$$\rho \circ H(X) = \left(\begin{pmatrix} 0 & \psi \\ \varphi & 0 \end{pmatrix}, \begin{pmatrix} 0 & \psi \\ \varphi & 0 \end{pmatrix}\right) \text{ to}$$

$$\rho \circ H(X') = \left(\begin{pmatrix} 0 & \psi' \\ \varphi' & 0 \end{pmatrix}, \begin{pmatrix} 0 & \psi' \\ \varphi' & 0 \end{pmatrix}\right)$$

With this notation we have

i) *The morphisms $(\bar{\alpha}_3, \bar{\beta}_2)\colon X \to TX'$ and $(\bar{\alpha}_2, \bar{\beta}_3)\colon TX \to X'$ factor over projectives.*

ii) *The morphisms $(\bar{\alpha}_1, \bar{\beta}_4)$ and $(\bar{\beta}_1, \bar{\alpha}_4)$ from X to X' only differ by a morphism that factors over projectives.*

Proof. i) The fact that (A, B) is a morphism of matrix factorizations implies that

$$\varphi' \circ \alpha_1 - v\alpha_3 = u\beta_3 + \beta_4 \circ \varphi$$

$$u\alpha_2 + \psi' \circ \alpha_4 = \beta_1 \circ \psi - v\beta_2$$

Expanding α_1, β_1, α_4, β_4 in powers of u and v we see that $\bar{\beta}_2$ and $\bar{\beta}_3$ can be written in the form

$$\begin{aligned}\bar{\beta}_2 &= \psi' \circ \gamma_2 + \delta_2 \circ \psi \\ \bar{\beta}_3 &= \varphi' \circ \gamma_3 + \delta_3 \circ \varphi\end{aligned} \quad \ldots \text{ with certain maps } \gamma_i, \delta_i$$

Clearly $\psi' \circ \gamma_2$ induces the zero map from $\operatorname{cok} X = \operatorname{coker}(F_1 \xrightarrow{\varphi} F_2)$ to $\operatorname{cok} TX' = \operatorname{coker}(F_2' \xrightarrow{\psi'} F_1')$. On the other hand the morphism $\delta_2 \circ \psi$: $\operatorname{coker} \varphi \to \operatorname{coker} \psi'$ factors over the free R-module $F_1 \otimes_p R$, since $\psi \circ \varphi = 0$ over R. This shows that the map induced by $\bar{\beta}_2$ from $\operatorname{cok} X$ to $\operatorname{cok} TX'$ factors over projectives, and similarly one sees that the map induced by $\bar{\beta}_3$ from $\operatorname{cok}(TX)$ to $\operatorname{cok} X'$ factors over projectives. Using Theorem 1.1 part i of the lemma now follows.

ii) Again from the fact that (A, B) is a morphism of matrix factorizations we see that

$$\varphi' \circ \alpha_2 - v\alpha_4 = \beta_3 \circ \psi - v\beta_4 .$$

Expanding in powers of u and v we see that $\bar{\alpha}_4 - \bar{\beta}_4$ has the form $\varphi' \circ \gamma + \delta \circ \psi$, and the claim follows as above. □

Lemma 3.7 shows that any morphism (A, B) from $H(X)$ to $H(X')$ can be altered by an element in $H(\operatorname{Hom}(X, X'))$ in such a way that its restriction $\rho(A, B)$ factors over projectives. Applying Corollary 2.9 twice now gives the desired result.

4. Simple hypersurface singularities

The two-dimensional simple singularities are just the rational double points, and by [2] there are only finitely many isomorphism classes of MCM's over their local rings. By iterated application of the Corollary 2.8 we get the implication $(b) \Rightarrow (a)$ of the theorem stated in the introduction.

The Auslander-Reiten quivers of the one- and two-dimensional simple singularities are known – for dimension two they are determined in [3] and [6], for dimension one in [9] and [16]. Applying the results of Chapter 3 this suffices to determine the Auslander-Reiten quivers for simple hypersurface singularities of arbitrary dimension.

Auslander's computation of the Auslander-Reiten quiver for two-dimensional simple singularities in $\operatorname{char} k = 0$ given in [3] is based on the fact that these singularities are of the form k^2/Γ, where Γ is a finite subgroup of $\mathrm{SL}(2, k)$. He used the McKay correspondence [17] to identify them as Dynkin quivers of type A, D, E. We want to give here a similar description of the Auslander-Reiten quivers of the simple plane curve singularities in $\operatorname{char} k = 0$:

Let G be a finite subgroup of $GL(2, k)$ which is generated by reflections, and denote by ε: $G \to \{\pm 1\}$ the linear character $g \mapsto \det g$. The kernel Γ of ε is a subgroup of index 2 in G; we denote a generator of the group $G/\Gamma = \mathbb{Z}/2\mathbb{Z}$ by σ. The singularity of $X := k^2/\Gamma$ at the origin is a simple surface singularity. σ acts as an involution on X, and by Chevalley's theorem [7] V.5.3 the quotient

$X/\langle\sigma\rangle$ is smooth. The branch locus $(Y,0)\subset(X,0)$ of the projection $X\to X/\langle\sigma\rangle$ is isomorphic to a simple plane curve singularity – and every simple plane curve singularity can be obtained in this way.

Let S be the local ring $\mathcal{O}_{k^2,0}$. Put $R_1:=\mathcal{O}_{X,0}=S^\Gamma$. To determine the Auslander-Reiten quivers of indecomposable MCM's over $R:=\mathcal{O}_{Y,0}$ it suffices by Corollary 2.3 to compute the Auslander-Reiten quiver of indecomposable MCM's over $R_1[\sigma]$. If $\rho\colon G\to GL(E)$ is a representation of G over k we put

$$M_\rho:=(S\otimes E)^\Gamma.$$

This is in a natural way an MCM over $R_1[\sigma]$ (cf. [15], [3], [11]). $\rho\mapsto M_\rho$ is an additive functor from the category of $k[G]$-modules to $\mathrm{MCM}_\sigma(R_1)$. By c we denote the canonical representation $c\colon G\to GL(2,k)$.

Theorem 4.1. i) *For each M in $\mathrm{MCM}_\sigma(R_1)$ there is a unique representation ρ of G such that $M\cong M_\rho$.*

ii) *Let ρ,ρ' be irreducible representations of G. Then there exists an irreducible morphism $M_\rho\to M_{\rho'}$ if and only if ρ is a direct summand of $\rho'\otimes c$.*

Remark 4.2. This shows that the Auslander-Reiten quiver of $\mathcal{O}_{Y,0}$ is isomorphic to the graph computed in [13], 3.7, with the vertex corresponding to the trivial representation deleted.

Proof of Theorem 4.1. i) For a representation $\rho\colon G\to GL(E)$ we have $S\oplus E\cong(M_\rho\otimes_{R_1}S)^{\wedge\wedge}$ (cf. [15], [3], [11]), hence E as $k[G]$-module can be recovered from M. So $\rho\mapsto M_\rho$ is an injection from the set of indecomposable $k[G]$-modules to the set of indecomposable MCM's over $R_1[\sigma]$.

To prove that it is surjective let M be an indecomposable MCM over $R_1[\sigma]$. By [3] there is a representation $\rho'\colon\Gamma\to GL(E')$ such that $M\cong(S\otimes E')^\Gamma$ as R_1-module. Let $\rho''\colon\Gamma\to GL(E'')$ be the representation $\rho'':=\rho'\circ\sigma$, where the generator σ of G/Γ acts on Γ by conjugation. Obviously $\sigma^*(M)=(S\oplus E'')^\Gamma$ as R_1-module. As $M\cong\sigma^*(M)$ it follows from [3], [11] or [15] that the representations ρ' and $\rho''=\rho'\circ\sigma$ of Γ are isomorphic. Hence there is a representation ρ of G such that $\rho|\Gamma=\rho'$. By Lemma 2.5i we have $M\cong M_\rho$ or $M\cong M_{\varepsilon\cdot\rho}$.

ii) Consider the exact sequence of $R_1[\sigma]$-modules

$$0\to R_1^-\to M_c\to\mathfrak{m}_{R_1^+}\to 0 \tag{4.3}$$

representing the unique non-trivial extension of $\mathfrak{m}_{R_1^+}$ by R_1^- (cf. [3], observe that $R_1^-\cong\omega_{X,0}$ as $R_1[\sigma]$-module). By [3] one obtains for each non-trivial indecomposable M in $\mathrm{MCM}(R_1)$ the almost split sequence in $\mathrm{MCM}(R_1)$ ending with M by tensoring M with the sequence (4.3) and taking reflexive hulls. By Corollary 2.10 this statement also holds in the category $\mathrm{MCM}_\sigma(R_1)$. Now if $M=M_\rho$ for some irreducible representation ρ of G then one easily checks that $(M_c\otimes M_\rho)^{\wedge\wedge}=M_{c\otimes\rho}$ (cf. [3]). So the MCM's over $R_1[\sigma]$ admitting an irreducible morphism to M_ρ are just the modules $M_{\rho'}$, where ρ' is an irreducible summand of ρ' is an irreducible summand of $\rho\otimes c$. □

Remark 4.4. Part i) of Theorem 4.1 can also be proven using ii) and the completeness result [21] Proposition 1.

References

1. Arnol'd, V.I.: Critical points of smooth functions. Proc. Int. Congr. Math., Vancouver 1974, vol. 1, 19–39
2. Artin, M., Verdier, J.-L.: Reflexive modules over rational double points. Math. Ann. **270**, 79–82 (1985)
3. Auslander, M.: Rational singularities and almost split sequences. Trans. Am. Math. Soc. **293**, 511–532 (1986)
4. Auslander, M.: Isolated singularities and existence of almost split sequences. In: Representation theory I, Groups and orders. Lect. Notes Math., vol. 1178, pp. 194–242. Berlin – Heidelberg – New York – Tokyo: Springer 1986
5. Auslander, M., Reiten, I.: Representation theory of Artin algebras III. Commun. Algebra **3**, 239–294 (1974); IV ibid. **5**, 443–518 (1977); V ibid. **5**, 519–554 (1977); VI ibid. **6**, 257–300 (1978)
6. Auslander, M., Reiten, I.: Almost split sequences for rational double points. Trans. Am. Math. Soc. (to appear)
7. Bourbaki, N.: Groupes et algèbres de Lie 4, 5, 6. Paris: Hermann 1968
8. Buchweitz, R.-O., Greuel, G.-M., Schreyer, F.-O.: Cohen-Macaulay modules on hypersurface singularities II. Invent. Math. **88**, 165–182 (1987)
9. Dietrich, E., Wiedemann, A.: The Auslander-Reiten quiver of a simple curve singularity. Trans. Am. Math. Soc. **294**, 455–475 (1986)
10. Eisenbud, D.: Homological algebra on a complete intersection, with an application to group representations. Trans. Am. Math. Soc. **260**, 35–64 (1980)
11. Esnault, H.: Reflexive modules on quotient singularities. J. Reine Angew. Math. **362**, 63–71 (1985)
12. Gabriel, P.: Auslander-Reiten sequences and representation-finite algebras. In: Representation theory I, Lecture Notes in Math., Vol. 831 pp. 1–71. Berlin – Heidelberg – New York: Springer 1980
13. Gonzalez-Sprinberg, G., Verdier, J.-L.: Construction géométrique de la correspondance de McKay. Ann. Sci. Ec. Norm. Super., IV. Ser. **16**, 409–449 (1983)
14. Greuel, G.-M., Knörrer, H.: Einfache Kurvensingularitäten und torsionsfreie Moduln. Math. Ann. **270**, 417–425 (1985)
15. Herzog, J.: Ringe mit nur endlich vielen Isomorphieklassen unzerlegbarer Cohen-Macaulay-Moduln. Math. Ann. **233**, 21–34 (1978)
16. Kiyek, K., Steinke, G.: Einfache Kurvensingularitäten in beliebiger Charakteristik. Arch. Math. **45**, 565–573 (1985)
17. McKay, J.: Graphs, singularities and finite groups. Proc. Symp. Pure Math. **37**, 183–186 (1980)
18. Popescu, N.: Abelian categories with applications to rings and modules. Academic Press. London & New York 1973
19. Reiten, I., Riedtmann, Chr.: Skew group algebras in the representation theory of Artin algebras. J. Algebra **92**, 224–282 (1985)
20. Swan, R.: Algebraic K-Theory. Lect. Notes in Math., Vol. 76. Berlin – Heidelberg – New York: Springer 1968
21. Wiedemann, A.: Orders with loops in their Auslander-Reiten graph. Commun. Algebra **9**, 641–656 (1981)

Oblatum 30-V-1986

Invent. math. 88, 165–182 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Cohen-Macaulay modules on hypersurface singularities II

R.-O. Buchweitz[1,*], G.-M. Greuel[2], and F.-O. Schreyer[2]

1 Institut für Mathematik, Technische Universität Hannover, Welfengarten 1, D-3000 Hannover, Federal Republic of Germany

2 Fachbereich Mathematik, Universität Kaiserslautern, Erwin-Schrödinger-Str., D-6750 Kaiserslautern, Federal Republic of Germany

Egbert Brieskorn gewidmet

Contents

0. Introduction and results 165
1. Periodic complexes and matrix factorizations 169
2. Construction of matrix factorizations and MCM's 173
3. Proof of the main results 175
4. MCM's on A_∞ and D_∞ 179
5. References 182

0. Introduction and results

0.1. H. Knörrer [Knö] showed the following property of the local ring R of an isolated simple hypersurface singularity: There are only finitely many isomorphism classes of indecomposable maximal Cohen-Macaulay R-modules.

Recall that a module M – always assumed to be finitely generated and unitary – over a local Cohen-Macaulay ring R is said to be *maximal Cohen-Macaulay*, if depth $M = \dim R$ (M is an MCM for short). M is called *indecomposable* if it is not a nontrivial direct sum. Let us call R of *finite* (resp. *countable*) CM-*representation type*, if there are only finitely (resp. countably) many isomorphism classes of indecomposable MCM's over R. If R is not of finite CM-representation type we say that it is of infinite CM-representation type.

The main object of this paper is to prove the converse of Knörrer's result, namely that a non-simple hypersurface singularity is of infinite CM-representation type (cf. Theorem A and B). This characterization of the simple hypersurface singularities has been proved before by Artin and Verdier [A-V], Auslander [Aus1] and Esnault [Esn] in Dimension 2 and by Greuel and Knörrer [G-K] in Dimension 1.

In addition we classify the hypersurface singularities which are of countable CM-representation type (Theorem B). Our method yields also an application to

* Supported by a "Heisenberg-Stipendium", Bu-398/3-1 of the DFG

vector bundles on projective hypersurfaces with "no cohomology in the middle" (Theorem C).

In our proof we explicitly construct infinitely many non isomorphic MCM's. The construction uses the fact that MCM's over a hypersurface ring R have periodic resolutions and that every MCM arises from a matrix factorization of the defining function. This was first proved by D. Eisenbud [Eis] and we recall the main facts in § 1.

0.2. Let P denote a regular noetherian local ring with maximal ideal $\mathfrak{m}$ and algebraically closed residue field $k=P/\mathfrak{m}$. Let $f\in\mathfrak{m}$, $f\neq 0$, $R=P/(f)$ and $\operatorname{mult} R=\operatorname{mult}(f)=\sup\{v\mid f\in\mathfrak{m}^v\}$.

We show in Proposition 3.1, that R is of finite CM-representation type if either

a) $\dim R\geqq 1$ and $\operatorname{mult} R\geqq 4$ or

b) $\dim R\geqq 2$ and $\operatorname{mult} R\geqq 3$.

The modules which we construct are parametrized in Case a) by the points of $\mathbb{P}_k(\mathfrak{m}/\mathfrak{m}^2)$ and in Case b) by the points of a cubic in $\mathbb{P}_k(\mathfrak{m}/\mathfrak{m}^2)$. In order to arrive at the simple hypersurface singularities we specialize to analytic hypersurfaces. Suppose that P is a regular analytic k-algebra, i.e. isomorphic to a formal or convergent (if k is a complete non trivial valuated field) power series ring $k\langle z_0,\ldots,z_n\rangle$. $R=P/(f)$, $f\in\mathfrak{m}$, is then called an *analytic hypersurface ring* and $(X,0)$ denotes the analytic hypersurface defined by R.

Arnold introduced the notion of a *simple* hypersurface singularity over $k=\mathbb{C}$ (or $\mathbb{R}$) for a singularity which is of "finite deformation type" and he classified them (cf. [Arn]). The classification yields the well-known singularities of type A_k, D_k, E_6, E_7, E_8. Barth, Peters and van de Ven [B-P-V] obtained for reduced plane curve singularities the same list from a different definition which has the advantage that it carries over immediately to arbitrary algebraically closed fields: A 1-dimensional reduced hypersurface ring $R=k\langle z_0,z_1\rangle/(f)$ is called *simple* if $\operatorname{mult}(f)\leqq 3$ and if after any finitely many successive quadratic transformations, the reduced total transform of f still has multiplicity $\leqq 3$ in each point. We then also call f or $(X,0)$ simple.

For the following theorem we assume that R is an n-dimensional analytic hypersurface ring, $n\geqq 1$, k algebraically closed and $\operatorname{char}(k)\neq 2$.

Theorem A. *R is of finite* CM*-representation type if and only if R is isomorphic to $k\langle z_0,\ldots,z_n\rangle/(f)$ such that*

$$f(z_0,\ldots,z_n)=g(z_0,z_1)+z_2^2+\ldots+z_n^2$$

where $g\in k\langle z_0,z_1\rangle$ is simple.

Remarks. 1) We need the assumption that R is an analytic hypersurface ring in order to apply the Weierstraß preparation theorem and that $\operatorname{char}(k)\neq 2$ in order to split off $z_2^2+\ldots+z_n^2$. The assumption $\operatorname{char}(k)\neq 2$ is superfluous if $\dim R=1$ (cf. 3.5).

2) Equations of the simple singularities are given in [Arn], [B-P-V] and 0.3 for $k=\mathbb{C}$ and in [K-S] for an arbitrary algebraically closed field. As was

noticed in [K-S], if $n=\dim R=2$ and $\mathrm{char}(k)\neq 2$, the singularities occurring in the theorem are exactly the rational double points in the sense of [A-V].

3) As mentioned above, the theorem was proved before in Dimension 1 by [G-K] for $\mathrm{char}(k)=0$ and then extended to arbitrary characteristic by [K-S] (cf. also 3.5). In Dimension 2 it is due to [A-V], with different proofs also given in [Aus1] and [Esn]. Their result implies that a two dimensional normal analytic hypersurface singularity is of finite CM-representation type iff it is a rational double point, even in characteristic 2.

0.3. In the following let $k=\mathbb{C}$, $f\in\mathbb{C}\{z_0,\ldots,z_n\}$ (the convergent power series ring), $R=\mathbb{C}\{z_0,\ldots,z_n\}/(f)$ and $(X,0)$ the complex space germ defined by R. In this case, the singularities exhibited in Theorem B are exactly the *simple singularities* in the sense of Arnold, mentioned above:

$$
\begin{aligned}
&A_k\colon\ f=z_0^{k+1}+z_1^2+\ldots+z_n^2, && k\geqq 1,\\
&D_k\colon\ f=z_0^{k-1}+z_0z_1^2+z_2^2+\ldots+z_n^2, && k\geqq 4,\\
&E_6\colon\ f=z_0^3+z_1^4+z_2^2+\ldots+z_n^2,\\
&E_7\colon\ f=z_0^3+z_0z_1^3+z_2^2+\ldots+z_n^2,\\
&E_8\colon\ f=z_0^3+z_1^5+z_2^2+\ldots+z_n^2.
\end{aligned}
$$

Notice that there are two natural limit cases of A_k resp. D_k as k tends to infinity, namely

$$
\begin{aligned}
&A_\infty\colon\ f=z_1^2+\ldots+z_n^2,\\
&D_\infty\colon\ f=z_0z_1^2+z_2^2+\ldots+z_n^2,
\end{aligned}
$$

considered as an element of $\mathbb{C}\{z_0,\ldots,z_n\}$.

Besides the simple ones, these two (non-isolated) singularities play a special role with respect to MCM's (cf. Theorem B).

We mention also that A_∞ and D_∞ are basic for topological considerations in the deformation theory of non-isolated singularities (cf. [Sie]).

Let us call a singularity (i.e. a complex space germ) $(X,0)$ of *finite* (resp. *countable*) *deformation type*, if there exist a finite (resp. countable) list of singularities with the following property: For any (flat) deformation $(Y,0)\to(T,0)$ of $(X,0)$ over a complex space germ $(T,0)$, there exists a representative $Y\to T$ such that each fibre Y_t has only singularities which are isomorphic to one of the members of the list. A hypersurface singularity of finite deformation type is called *simple*. This is (a slight modification of) Arnold's definition and the classification leads to the singularities A_k, D_k $(k\neq\infty)$ and E_6, E_7, E_8.

Similarily we define an $R=\mathcal{O}_{X,0}$-module M to be of *finite* (resp. *countable*) *deformation type* if there exists a finite (resp. countable) list of R-modules such that for each deformation $\mathcal{M}$ of M over $(T,0)$ (i.e. $\mathcal{M}$ is a flat $\mathcal{O}_{X\times T,0}$-module and $M=\mathcal{M}\otimes\mathcal{O}_{X\times\{0\},0}$) there exists a representative T of $(T,0)$ and a sheaf on $(X,0)\times T$ inducing $\mathcal{M}$ on $(X\times T,0)$, (which we also call $\mathcal{M}$) such that for each $t\in T$ the $\mathcal{O}_{(X,0)\times\{t\}}=\mathcal{O}_{X,0}$-module $\mathcal{M}\otimes\mathcal{O}_{(X,0)\times\{t\}}$ is isomorphic to a module of the list.

Theorem B. *Let* $R=\mathbb{C}\{z_0,\ldots,z_n\}/(f)$, $f\in m-\{0\}$, $n\geqq 1$. *Then the statements* (i)–(iii) *of* (1) *are equivalent and so are the statements* (i)–(iv) *of* (2).

(1) (i) *f defines a simple hypersurface singularity.*
(ii) *R is of finite* CM*-representation type.*
(iii) *Each* MCM *over R is of finite deformation type.*

(2) (i) *f defines an A_∞ or D_∞ singularity.*
(ii) *R is of countable but not finite deformation type.*
(iii) *R if of countable but not finite* CM*-representation type.*
(iv) *Each* MCM *over R is of countable deformation type and there exists an* MCM *over R which is not of finite deformation type.*

Remarks. 1) See [Dur] for further characterizations of simple singularities.

2) Since the singularities A_∞ and D_∞ are not isolated, it follows already from a general theorem of Auslander [Aus2] that they are of infinite CM-representation type. The point is, that these are the only hypersurface singularities which are of *countable* CM-representation type. We give a complete classification of all MCM's over A_∞ and D_∞ in §4.

3) For the simple hypersurface singularities all indecomposable MCM's are explicitly known and their Auslander-Reiten quivers are related to the corresponding Dynkin diagram of type A_k, D_k or E_k respectively (cf. [Knö] for further references). For A_∞ and D_∞ the Auslander-Reiten quiver of those MCM's which are locally free on a punctured neighbourhood of the singularity are again related to the Dynkin diagrams A_∞ and D_∞ respectively (cf. [Sch]).

0.4. The modules $M^{\pm}(\underline{x},\underline{y})$ which we construct are all of a fixed rank 2^{n-y} where $n=\dim R$ and of a fixed minimal number of generators equal to 2^n (cf. 2.5). There is an estimate, due to W. Bruns – and obtained as a consequence of the results in [Bru] – to the effect that any MCM on a hypersurface singularity $R=P/(f)$ without free summands has rank at least $\frac{1}{2}(\operatorname{codim}_R \operatorname{Sing}(R)-1)$ ($\operatorname{Sing}(R)$ denotes the singular locus of R). The actual minimal bound is not known.

We offer the following:

Conjecture A. *Any maximal Cohen-Macaulay module M without free summands on an irreducible hypersurface singularity R satisfies:*

$$\operatorname{rank} M\geqq 2^e$$

where $e=[\frac{1}{2}(\operatorname{codim}_R \operatorname{Sing}(R)-2)]$ *and* $[q]$ *denotes the largest integer* $\leqq q$.

Remark that the guess above is stable under taking general hypersurface sections, so that it would be enough to establish it for isolated hypersurface singularities.

0.5. We give an application to projective geometry:

Instead of the local case one may consider the graded case. For this replace P by the graded polynomial ring $S=k[z_0,\ldots,z_n]$ over an algebraically closed field k and assume that $f\in S$ is homogenous of positive degree.

Let $X = \mathrm{Proj}_k(S/(f)) \subset \mathbb{P}_k^n$ denote the projective hypersurface defined by $f = 0$.

If $n > 1$, by a theorem of Grothendieck-Serre (cf. for example [K-L]), graded maximal Cohen-Macaulay modules M over $R = P/(f)$ correspond via sheafification biuniquely to coherent sheaves of $\mathcal{O}_X$-modules $\mathcal{M}$ with "*no cohomology in the middle*" i.e.

$$H^i(X, \mathcal{M}(v)) = 0 \quad \text{for } 0 < i < n-1 = \dim X \text{ and } v \in \mathbb{Z}.$$

Shifting the grading of M by v corresponds to twisting the sheaf $\tilde{M}$ by $\mathcal{O}_{\mathbb{P}_k^n}(v)$. If X is smooth, these sheaves are vector bundles on X.

We obtain:

Theorem C. *Let X be a projective hypersurface in $\mathbb{P}_k^n$, $n > 1$, k algebraically closed. X is a linear or quadratic hypersurface if and only if there are (up to twist) only finitely many isomorphism classes of indecomposable locally free sheaves $\mathcal{M}$ of $\mathcal{O}_X$-modules with vanishing cohomology in the middle:*

$$H^i(X, \mathcal{M}(v)) = 0, \quad v \in \mathbb{Z},\ 0 < i < n-1.$$

In the graded case Conjecture A implies

Conjecture B. *Let $X \subset \mathbb{P}_k^n$ be a smooth hypersurface over an algebraically closed field k. Let $\mathcal{M}$ be a vector bundle on X which satisfies*

$$H^i(X, \mathcal{M}(v)) = 0, \quad v \in \mathbb{Z},\ 0 < i < n-1 = \dim X.$$

If

$$\operatorname{rank} \mathcal{M} < 2^e$$

with $e = \left[\frac{n-2}{2}\right]$, *then $\mathcal{M}$ is a direct sum of line bundles*

$$\mathcal{M} \cong \bigoplus_i \mathcal{O}_X(a_i).$$

Remarks. 1) For a linear hypersurface, Conjecture B is true without assumption on the rank by a theorem of Horrocks (cf. [O-S-S]). Since over a regular local ring every MCM is free, this follows also from the Grothendieck-Serre theorem mentioned above.

2) For a quadratic hypersurface, Conjecture B is also true. This follows from the classification of MCM's over the A_1-singularities in Dimension 1 and 2 and Knörrer's periodicity theorem (cf. also 2.5(4)).

3) The conjecture is at least sharp in all degrees. Look at the Fermat hypersurface X of degree d, defined by $f = z_0^d + \ldots + z_n^d$. It follows from our construction in § 3 and sheafification that there are vector bundles on X of rank 2^e with no cohomology in the middle and which do not contain a line bundle as a direct summand.

1. Periodic complexes and matrix factorizations

1.1. We recall the main results about periodic resolutions, maximal Cohen-Macaulay modules and matrix factorizations as obtained by D. Eisenbud in [Eis, §§ 4–6]. These are essential for our construction in the next section.

A (right-bounded) complex

$$\mathbb{F}.: \ldots \to F_k \to F_{k-1} \to \ldots \to F_a \to 0, \quad a \in \mathbb{Z},$$

is called *periodic* (*of period* 2) if there is a map of complexes t: $\mathbb{F}. \to \mathbb{F}.$ of degree -2 such that t: $F_{i+2} \to F_i$ is an isomorphism for all $i \geqq a$. Identifying F_{i+2} with F_i by means of t and setting $G = F_a = F_{a+2} = \ldots$ and $F = F_{a+1} = F_{a+3} = \ldots$, we may assume that a periodic complex $\mathbb{F}.$ has the form

$$\mathbb{F}.: \ldots \xrightarrow{\Psi} F \xrightarrow{\Phi} G \xrightarrow{\Psi} F \xrightarrow{\Phi} G \longrightarrow 0.$$

1.2. Special periodic complexes arise from matrix factorizations. Given a ring P and an element $f \in P$, a *matrix factorization of* f is an ordered pair of morphisms of free P-modules, $(\Phi: F \to G;\ \Psi: G \to F)$, such that

$$\Phi \circ \Psi = f.\,\mathrm{id}_G \quad \text{and} \quad \Psi \circ \Phi = f.\,\mathrm{id}_F.$$

If P is local with maximal ideal $\mathfrak{m}$, the matrix factorization (Φ, Ψ) is called *reduced* iff

$$\operatorname{Im} \Phi \subseteq \mathfrak{m}.\,G \quad \text{and} \quad \operatorname{Im} \Psi \subseteq \mathfrak{m}.\,F.$$

Two matrix factorizations $(\Phi_i: F_i \to G_i;\ \Psi_i: G_i \to F_i)$, $i = 1, 2$, of the same element $f \in P$ are *equivalent* iff there are P-linear isomorphisms α: $F_1 \to F_2$, β: $G_1 \to G_2$ such that

$$\beta \Phi_1 = \Phi_2 \alpha \quad \text{and} \quad \alpha \Psi_1 = \Psi_2 \beta.$$

If we now set $R = P/(f)$, write $\bar{}$ for reduction modulo f and denote $\varphi = \bar{\Phi}$, $\psi = \bar{\Psi}$, then a matrix factorization (Φ, Ψ) yields a periodic complex of R-modules

$$\mathbb{F}(\Phi, \Psi): \ldots \longrightarrow \bar{F} \xrightarrow{\varphi} \bar{G} \xrightarrow{\psi} \bar{F} \xrightarrow{\varphi} \bar{G} \longrightarrow 0.$$

If f is not a zero-divisor on P, it follows (cf. [Eis, 5.1]) that $\mathbb{F}(\Phi, \Psi)$ is exact and hence an R-free resolution of coker φ. Furthermore, one has $\operatorname{rank} F = \operatorname{rank} G$, (cf. [Eis, 5.4]). Let us also remark (cf. [Eis, 5.5]), that for f a non-zero divisor on P and (Φ, Ψ) a matrix factorization of f, Ψ is already uniquely determined by Φ.

1.3. Assume now that R is any local Cohen-Macaulay ring and that

$$\ldots \xrightarrow{\varphi} G \xrightarrow{\psi} F \xrightarrow{\varphi} G \longrightarrow M \longrightarrow 0$$

is a periodic R-free resolution of the R-module M (with F and G of finite rank). Then M is a *maximal Cohen-Macaulay R-module* (MCM for short) i.e. $\operatorname{depth} M = \dim R$.

This follows from the characterization of depth by Ext's and the obvious fact that $M \cong \ker \varphi$. If $R = P/(f)$ for a non-zero-divisor f in P and if the periodic resolution comes from a matrix-factorization (Φ, Ψ) of f as above, then $\mathbb{F}(\Phi, \Psi)$ is a minimal resolution of $M = \operatorname{coker} \Phi$ iff (Φ, Ψ) is reduced. In this case, M has no R-free summand (cf. [Eis, 0.1; 6.1]).

1.4. From now on let P denote a regular local ring with maximal ideal $\mathfrak{m}$, f a non-zero element of $\mathfrak{m}$ and R the hypersurface ring $P/(f)$.

The main result on maximal Cohen-Macaulay R-modules is that every such module has a periodic free resolution and that all such resolutions arise from matrix factorizations of f. More precisely

Theorem ([Eis, 6.1, 6.3]). 1) *Every maximal Cohen-Macaulay R-module M has a periodic free resolution of the form $\mathbb{F}(\Phi, \Psi)$ for some matrix factorization (Φ, Ψ) of f. This resolution can be chosen minimal - and hence (Φ, Ψ) reduced - iff M has no free summand.*

2) The associations $(\Phi, \Psi) \mapsto \mathbb{F}(\Phi, \Psi)$ *and* $(\Phi, \Psi) \mapsto \operatorname{Coker} \Phi$ *induce bijections between the sets of*

(i) *equivalence classes of reduced matrix factorizations of f over P,*

(ii) *isomorphism classes of periodic minimal free resolutions over R,*

(iii) *isomorphism classes of maximal Cohen-Macaulay R-modules without free summands (the 0-module included).*

1.5. Note that the proof of this theorem (as given in [Eis]) is quite elementary. We indicate how the circle closes by repeating the argument which associates to a maximal Cohen-Macaulay R-module M a matrix factorization of f:

By the Auslander-Buchsbaum-Serre theorem one has

$$\operatorname{proj\,dim}_P M = \dim P - \operatorname{depth} M = 1.$$

Hence there is a P-free resolution of M of length 1,

$$0 \longrightarrow F \xrightarrow{\Phi} G \longrightarrow M \longrightarrow 0,$$

where M is considered as a P-module.

Since f annihilates M, $f \cdot G \subseteq \operatorname{Im} \Phi$ and hence there exists a P-linear map Ψ: $G \to F$ with

$$\Phi \circ \Psi = f \cdot \mathrm{id}_G.$$

Applying Φ from the right and noting that Φ is injective yields then $\Psi \circ \Phi = f \cdot \mathrm{id}_F$, so that (Φ, Ψ) is a matrix factorization of f with $\operatorname{Coker} \Phi = M$.

1.6. Now assume that M is a maximal Cohen-Macaulay module over $R = P/(f)$ *without free summand*. Then, by the theorem above, there exists a reduced matrix factorization (Φ, Ψ) of f such that $M = \operatorname{Coker} \Phi$.

Keeping the notations from (1.5), Φ defines a natural map $F \underset{P}{\otimes} G^* \to P$, given by $x \otimes \lambda \to \lambda(\Phi(x))$ for $x \in F$, $\lambda \in G^*$. The image of this map is an ideal denoted by $I(\Phi)$. Since (Φ, Ψ) is reduced, it follows that $I(\Phi) \subseteq \mathfrak{m}$. (Choosing bases for the free P-modules F and G, $I(\Phi)$ is just the ideal generated by the entries of the matrix of Φ with respect to these bases.)

In the same way define $I(\Psi) = \operatorname{Im}(G \otimes F^* \to P)$. Then we have:

Lemma. *With the notations and hypotheses as above, $I(\Phi)$ and $I(\Psi)$ only depend on the isomorphism class of $M = \operatorname{Coker} \Phi$ as an R-module. More precisely:*

(i) *If M is minimally generated by μ elements then $I(\Phi)=F_{\mu-1}(M)$,*

(ii) $I(\Psi)\cdot\det\Phi=F_1(M)\cdot f$,

(iii) $f\in I(\Phi)\cdot I(\Psi)\subseteq(I(\Phi)+I(\Psi))^2\subseteq\mathfrak{m}^2$.

Here $F_a(-)$ denotes the a-th Fitting ideal of a P-module, cf. [Kap, App. 4-3(b)].

Proof. (i) is essentially the definition of $F_{\mu-1}(M)$, remarking that $\operatorname{Im}\Phi\subseteq\mathfrak{m}G$ just means $\mu=rk\,G$.

(ii) (cf. also [Eis, 5.5]): Denote by Φ^{ad} the adjoint morphism of Φ,

$$\Phi^{\mathrm{ad}}=\bigwedge^{rk\,G-1}(\operatorname{Hom}(\Phi,P)).$$

Then one has:

$$\Phi\circ\Phi^{\mathrm{ad}}=\det\Phi\cdot\mathrm{id}_G,$$

(where $\bigwedge^{rk\,G-1}G$ is identified with G) and

$$I(\Phi^{\mathrm{ad}})=F_1(M)$$

by the definition of the first Fitting ideal of M. From the properties of a matrix factorization one gets

$$\det\Phi\cdot(\Phi\circ\Psi)=\det\Phi\cdot f\cdot\mathrm{id}_G=f\cdot\Phi\circ\Phi^{\mathrm{ad}},$$

hence, as Φ is injective,

$$\det\Phi\cdot\Psi=f\cdot\Phi^{\mathrm{ad}},$$

whence (ii).

(iii) is a direct consequence of the definitions. □

Remark. Although not explicitely required, statement (iii) above shows $f\in\mathfrak{m}^2$. The reason is of course that if $f\in\mathfrak{m}-\mathfrak{m}^2$, the quotient $R=P/(f)$ is again local regular and all maximal Cohen-Macaulay modules over R are actually free (for which (iii) above yields a new proof).

1.7. To summarize the above, *we have associated to every reduced matrix factorization (Φ,Ψ) an ideal - namely $I(\Phi)+I(\Psi)$ - in P, whose square contains f.*

The key to our results is the following

Proposition. *The association $(\Phi,\Psi)\to I(\Phi)+I(\Psi)$ from (reduced) matrix factorizations of f to ideals of P yields a surjection from the set of isomorphism classes of maximal Cohen-Macaulay modules over R (without free summands) onto the set of those (proper) ideals of P whose square contains f.*

The proof of this is the content of the next section, see especially (2.4).

As an immediate consequence we get

Corollary. *Let P be a regular local ring, $f\in\mathfrak{m}$ a non-zero element in its maximal ideal and $R=P/(f)$.*

If there are infinitely many different ideals $I \subseteq P$ for which $f \in I^2$, then there are infinitely many isomorphism classes of indecomposable maximal Cohen-Macaulay R-modules.

This will be exactly the criterion to prove our main results.

Proof of the Corollary. If $M = \bigoplus M_i$ is a decomposition of M into indecomposable summands, a matrix factorization (Φ, Ψ) for M may be obtained by taking the direct sum of matrix factorizations (Φ_i, Ψ_i) for the M_i. *W.* l.o.g. we may assume that no M_i is free, hence that M has no free summand. Then, in particular, $I(\Phi)+I(\Psi)=\sum_i (I(\Phi_i)+I(\Psi_i))$. If there are only finitely many isomorphism classes of indecomposable MCM's over R, there are also only finitely many distinct ideals of the form $I(\Phi)+I(\Psi)$ for reduced matrix factorizations (Φ, Ψ) of f. This, together with the proposition, establishes the corollary.

1.8. The constructions and statements above hold similarly in the graded instead of the local case. For this, just replace in 1.4–1.7 the regular local ring P by a graded polynomial ring $S=k[z_0, \ldots, z_n]$ over a field k and assume f to be a homogenous element of $S_+ = (z_0, \ldots, z_n)S$ instead of being in $\mathfrak{m}$.

We leave it to the reader to define graded matrix factorizations and graded resolutions "periodic up to a shift of degrees" and to reformulate (1.4) up to (1.7) in this case.

As mentioned in the introduction, we shall apply the graded case to construct vector bundles on projective hypersurfaces.

2. Construction of matrix factorizations and MCM's

2.1. Let P be again an arbitrary commutative ring and $f \neq 0$ an element of P.

Suppose f is given in the form

$$f = \sum_{i=1}^{r} x_i y_i$$

with $x_i, y_i \in P$ for $i = 1, \ldots, r$. To such a presentation of f we will associate a matrix factorization of f – and hence a maximal Cohen-Macaulay module $M(\underline{x}, \underline{y})$ over $R = P/(f)$ in case P is local regular.

2.2. Let $\mathbb{K}(\underline{x}, P)$ denote the Koszul-complex associated to the sequence $(x_1, \ldots, x_r)$ of elements in P. Then the underlying graded P-module is the exterior algebra $\Lambda^\cdot P^r$, with P^r a free P-module on the basis $e_1, \ldots, e_r$. The Koszul-differential will be denoted by δ_{-1} and is explicitly given by

$$\delta_{-1}(\omega) = \left(\sum_{i=1}^{r} x_i e_i\right) \lrcorner \omega$$

for any $\omega \in \Lambda^\cdot P^r$ where $\lrcorner$ denotes the contraction operator (or "inner product") on $\Lambda^\cdot P^r$ – (cf. [ALG, §9] for terminology).

As f is by assumption contained in the ideal generated by the sequence $(x_1, \ldots, x_r)$ it annihilates the homology of $\mathbb{K}(\underline{x}, P)$.

But even better, multiplication with f on $\mathbb{K}(\underline{x}, P)$ is zero-homotopic. A homotopy is given by the following endomorphism of degree $+1$ on $\mathbb{K}(\underline{x}, P)$,

$$\delta_{+1}(\omega) = \left(\sum_{i=1}^{r} y_i e_i\right) \wedge \omega$$

for any $\omega \in \Lambda^{\cdot} P^r$.

Indeed,

$$\begin{aligned}\delta_{-1} \circ \delta_{+1}(\omega) &= (\Sigma\, x_i e_i^*) \lrcorner (\Sigma\, y_i e_i \wedge \omega) \\ &= ((\Sigma\, x_i e_i^*) \lrcorner (\Sigma\, y_i e_i)) \wedge \omega - (\Sigma\, y_i e_i) \wedge ((\Sigma\, x_i e_i^*) \lrcorner \omega) \\ &= f\omega - \delta_{+1} \circ \delta_{-1}(\omega).\end{aligned}$$

Since furthermore $\delta_{\pm 1}^2 = 0$, we have statement (1) of

2.3. **Proposition.** *With the notations as above, set* $\delta = \delta_{+1} + \delta_{-1}$.

(1) δ *is an endomorphism of the P-module* $\Lambda^{\cdot} P^r$ *satisfying* $\delta^2 = f \cdot \mathrm{id}_{\Lambda^{\cdot} P^r}$. *In particular,* (δ, δ) *is a matrix factorization of* f.

(2) *If* P *is regular local,* $f \neq 0$ *and* $x_i, y_i \in \mathfrak{m}$ *for all* i, *then:*

(i) (δ, δ) *is reduced and* $M(\underline{x}, \underline{y}) = \operatorname{coker} \delta$ *is a maximal Cohen-Macaulay module over* $R = P/(f)$ *without free summand.*

(ii) $I(\delta) = (x_1, \ldots, x_r, y_1, \ldots, y_r)$.

Remark. It is left as an exercise to the reader to show that $M(\underline{x}, \underline{y})$ is actually free, if one of the x_i or y_i is a unit.

Proof. Assertion (2.i) follows from (1) by (1.2), (1.3), (ii) follows immediately from the explicit description of $\delta_{\pm 1}$.

2.4. *Proof of Proposition 1.7.* Proposition 1.7 is now an easy consequence of Proposition 2.3. Let $I \subset \mathfrak{m}$ be any ideal with $f \in I^2$. Choose a set of generators $\{x_1, \ldots, x_r\}$ of I. Since $f \in I^2$ there exists a presentation $f = \sum_{i=1}^{r} x_i y_i$ with all $y_i \in I$. Take the reduced matrix factorization $(\Phi, \Psi) = (\delta, \delta)$ of f from 2.3. Then $I(\Phi) + I(\Psi) = I(\delta)$ and by 2.3 (2.ii)

$$I(\delta) = (x_1, \ldots, x_r, y_1, \ldots, y_r) = I,$$

which finishes the proof.

2.5. *Remarks.* 1) As $\delta_{\pm 1}$ are graded maps of degree ± 1, δ maps $\Lambda^{\mathrm{even}} P^r = \bigoplus_{i \equiv 0(2)} \Lambda^i P^r$ into $\Lambda^{\mathrm{odd}} P^r = \bigoplus_{i \equiv y(2)} \Lambda^i P^r$ and conversely.

In particular, denoting by δ^+ (resp. δ^-) the restriction of δ to $\Lambda^{\mathrm{even}} P^r$ (resp. $\Lambda^{\mathrm{odd}} P^r$), the matrix factorization (δ, δ) of f decomposes as a direct sum

$$(\delta, \delta) = (\delta^+, \delta^-) \oplus (\delta^-, \delta^+).$$

Accordingly $M(\underline{x}, \underline{y})$ is a direct sum $M(\underline{x}, \underline{y}) = M^+(\underline{x}, \underline{y}) \oplus M^-(\underline{x}, \underline{y})$ with $M^{\pm}(\underline{x}, \underline{y}) = \operatorname{coker} \delta^{\pm}$.

2) Starting from the presentation (2.1) of f, one might of course choose another partition of $\{x_1, \ldots, x_r, y_1, \ldots, y_r\}$ into two subsets of r elements such that every index $i=1, \ldots, r$ occurs exactly once in each subset. But it can be seen that the module M obtained from this choice is then isomorphic to $M(\underline{x}, \underline{y})$ - whereas the two summands M^+ and M^- might get interchanged. Remark also that M^+ and M^- are isomorphic to their own duals if $r \equiv 1 \bmod 2$ and isomorphic to the dual of the other if $r \equiv 0 \bmod 2$.

3) If f is irreducible in the UFD P, the rank of $M(\underline{x}, \underline{y})$ as an R-module is 2^{r-1}. Namely by [Eis, 5.6] one has $\det \Phi = f^{\operatorname{rank}_R \operatorname{coker} \Phi} \cdot u$ (u a unit in P) for every matrix factorization (Φ, Ψ) of f and by 2.3(1)

$$\det \delta \cdot \det \delta = \det(f \operatorname{id}_{\Lambda \cdot P^r}) = f^{2r},$$

whence $\det \delta = \pm f^{2^{r-1}}$.

It follows easily from this that $\operatorname{rank} M^{\pm}(\underline{x}, \underline{y}) = 2^{r-2}$ and the reader is encouraged to interpret this result for $r=1$!

4) Assume $P = k\langle x_1, \ldots, x_r, y_1, \ldots, y_r \rangle$ and $f = \sum_{i=1}^{r} x_i y_i$. Then $R = P/f$ is a singularity of type A_1 of odd dimension.

If $\operatorname{char} k \neq 2$ it can be deduced from the proof of H. Knörrer's periodicity theorem, [Knö], that indeed the two modules $M^{\pm}(\underline{x}, \underline{y})$ are non-isomorphic and are the only indecomposable non-free MCM's over R. A singularity of type A_1 of even dimension has exactly one indecomposable non-free MCM. This follows again from the periodicity theorem and the obvious result for $R = k\langle x \rangle / x^2$.

3. Proof of the main results

3.1. **Proposition.** *Let P be a regular local ring with algebraically closed residue class field $k = P/\mathfrak{m}$, $f \in \mathfrak{m}$ a non zero element and $R = P/(f)$. Suppose*

(a) $\operatorname{mult} R \geqq 4$ *and* $\dim R \geqq 1$, *or*

(b) $\operatorname{mult} R \geqq 3$ *and* $\dim R \geqq 2$.

Then R has infinite CM*-representation type.*

Proof. By assumption $f \in \mathfrak{m}^3$. Consider the residue $f_3 \in \mathfrak{m}^3/\mathfrak{m}^4$ of f has a homogenous polynomial on $\mathfrak{m}/\mathfrak{m}^2$ and let $C(f_3)$ be the variety defined by $f_3 = 0$ in $\mathbb{P}(\mathfrak{m}/\mathfrak{m}^2) \cong \mathbb{P}_k^n$; $n = \dim R$. For every point $\lambda \in C(f)$ we will construct an ideal $I_\lambda \subset P$ with $f \in I_\lambda^2$ as follows:

Choose generators $z_0, \ldots, z_n$ of $\mathfrak{m}$ such that in terms of the residue classes $\overline{z_i} \in \mathfrak{m}/\mathfrak{m}^2$, λ is defined by $\overline{z_1} = \ldots = \overline{z_n} = 0$. Then $f \in (z_1, \ldots, z_n) + \mathfrak{m}^4$, since $\lambda \in C(f)$.

We define

$$I_\lambda := (z_1, \ldots, z_n) + \mathfrak{m}^2 = (z_1, \ldots, z_n, z_0^2).$$

Then $I_\lambda \neq I_\mu$ for $\lambda \neq \mu$ and

$$I_\lambda^2 = (z_i z_j,\ 1 \leqq i \leqq j \leqq n,\ z_i \cdot z_0^2,\ 1 \leqq i \leqq n,\ z_0^4)$$

contains f:

$$f \in [(z_1, \ldots, z_n) + \mathfrak{m}^4] \cap \mathfrak{m}^3 = I_\lambda^2 \cap \mathfrak{m}^3 \subset I_\lambda^2.$$

By our assumption we have

(a) $C(f)=\mathbb{P}_k^n$ and $n\geqq 1$, or

(b) $C(f)$ is a cubic hypersurface in $\mathbb{P}_k^n$ and $n\geqq 2$.

In both cases $\dim C(f)\geqq 1$. So $C(f)$ contains infinitely many points, and there are hence infinitely many different ideals I with $f\in I^2$. Now Corollary 1.7 implies the result.

3.2. *Proof of Theorem C.* We replace P by the polynomial ring $S=k[z_0,\ldots,z_n]$, k an algebraically closed field, and f by a homogeneous polynomials of positive degree, finally $\mathfrak{m}$ by $S_+=(z_0,\ldots,z_n)S$, the irrelevant ideal, and set $R=S/(f)$.

As mentioned in 1.8, the assertions of §1 are still true in the graded category. Furthermore, from a presentation

$$f=\sum_{i=1}^{r} x_i y_i$$

by homogeneous elements $x_i, y_i\in S$, the construction of §2 gives a graded maximal Cohen-Macaulay module over R. Hence, if we choose in the proof of Proposition 3.1 correspondingly homogeneous generators $\{z_1,\ldots,z_n\}$ of the homogeneous ideal S_+, we obtain for $n\geqq 2$ and degree $f\geqq 3$ infinitely many graded MCM's over R which are pairwise not isomorphic, even if we neglect the grading.

By Remark 2.3 and the fact that the ideals I_λ constructed will contain S_+^2, it follows that the associated modules are actually locally free off $R_+=S_+/(f)$, the irrelevant ideal of R. Sheafification now yields Theorem C.

If the degree of f is 1 or 2, then there are (up to a shift of the grading) only finitely many indecomposable graded MCM's on R (cf. 2.5(4)).

Proof of Theorem A

3.3. Let $P=k\langle z_0,\ldots,z_n\rangle$ denote now a formal or convergent power series ring over an algebraically closed field k of arbitrary characteristic.

If $\mathrm{char}(k)\neq 2$ and

$$f(z_0,\ldots,z_n)=g(z_0,z_1)+z_2^2+\ldots+z_n^2,$$

where g defines a simple plane curve singularity, (0.2), then $R=P/(f)$ has finite CM-representation type by [Knö]. In order to prove the converse, we establish the existence of infinitely many ideals I with $f\in I^2$ for every f which does not decompose into a simple plane g and a sum of squares as above. Then we can apply Corollary 1.7. We proceed by induction on $n=\dim R$.

3.4. **Lemma.** *$f\in k\langle z_0,z_1\rangle$ defines a simple curve singularity if and only if for all $x,y\in\mathfrak{m}$ the following conditions are simultaneously satisfied:*

(a) $f\notin(x,y)^4$,

(b) $f\notin(x^3,x^2y^2,xy^4,y^6)$,

(c) $f\notin(x^2)$.

Proof. Note that $f \notin (x^2)$ for all $x \in \mathfrak{m}$ iff f defines a reduced curve singularity. Then applying the method to classify (reduced) simple curve singularities as in [B-P-V, §II.8, pp. 61–65] – with modifications as in [K-S] for fields of arbitrary characteristic – gives the lemma. We leave the details to the reader.

3.5. **Case $n=1$.** If f does not define a simple plane curve singularity, then by the lemma there exist $x, y \in \mathfrak{m}$ with either

(a) $f \in (x, y)^4$, or
(b) $f \in (x^3, x^2y^2, xy^4, y^6)$, or
(c) $f \in (x^2)$.

We consider in these cases correspondingly the ideals

(a) $I_\lambda = (\lambda_0 x + \lambda_1 y) + (x, y)^2$, $\lambda = (\lambda_0 : \lambda_1) \in \mathbb{P}^1_k$,
(b) $I_\lambda = (x + \lambda y^2, y^3)$, $\lambda \in k$,
(c) $I_\lambda = (x, y^\lambda)$, $\lambda \in \mathbb{N}$.

Then $f \in I_\lambda^2$. This is obvious in Case (a) and (c) and in Case (b) it follows from:

(b_1) $y^6 = (y^3)^2$
(b_2) $xy^4 = y(x + \lambda y^2)y^3 - \lambda y^6$
(b_3) $x^2y^2 = y^2(x + \lambda y^2)^2 - 2\lambda x y^4 - \lambda^2 y^6$

and

(b_4) $x^3 = (x + \lambda y^2)^3 - 3\lambda x^2 y^2 - 3\lambda^2 x y^4 - \lambda^3 y^6$.

For $\lambda \neq \mu$, one has $I_\lambda \neq I_\mu$ and 1.7 gives infinitely many isomorphism classes of indecomposable MCM's over $R = P/(f)$. □

3.6. **Case $n>1$.** We now assume $\mathrm{char}(k) \neq 2$ and proceed by induction on n. If $\mathrm{mult}\, R \geqq 3$ then there exist infinitely many ideals I with $f \in I^2$ by the proof of 3.1. If $\mathrm{mult}\, R = 2$ then we may apply the Weierstraß preparation theorem and Tschirnhausen transformation ($\mathrm{char}(k) \neq 2$) to obtain

$$f(z_0, \ldots, z_n) = g(z_0, \ldots, z_{n-1}) + z_n^2$$

with $g \in k\langle z_0, \ldots, z_{n-1}\rangle$. By the induction hypothesis there are infinitely many pairwise distinct ideals $\tilde{I}_\lambda \subset k\langle z_0, \ldots, z_{n-1}\rangle$ with $g \in \tilde{I}_\lambda^2$. (Notice that this is trivially true if $g = 0$.)

The ideals

$$I_\lambda = \tilde{I}_\lambda \cdot k\langle z_0, \ldots, z_n\rangle + (z_n) \subset k\langle z_0, \ldots, z_n\rangle$$

are pairwise distinct and $f \in I_\lambda^2$, hence 1.7 gives the result. □

Proof of Theorem B

3.7. From now on let

$$P = \mathbb{C}\{z_0, \ldots, z_n\}$$

denote the convergent power series ring over the complex numbers. As for the simple singularities in Lemma 3.4 we need a similar characterization of the curve singularities A_∞ and D_∞.

Lemma. *Let* $f \in \mathbb{C}\{z_0, z_1\}$ *satisfy* (a), (b) *but not* (c) *of* 3.4. *Then there is a coordinate system* $x, y \in \mathfrak{m}$ *such that*

$$A_\infty\colon f = x^2, \quad \text{or}$$
$$D_\infty\colon f = x^2 y.$$

The converse is also true.

Proof. $\mathbb{C}\{z_0, z_1\}$ is a unique factorization domain, hence we may write

$$f = u f_1^{e_1} \dots f_r^{e_r},$$

with u a unit, $f_i \in \mathfrak{m}$ the irreducible factors of f and e_i their multiplicities.

Since (c) is not satisfied, $e_i \geqq 2$ for at least one i, say $e_1 \geqq 2$. By (a) $\mathrm{mult}(f) < 4$, hence $r \leqq 2$ and $\mathrm{mult}(f_i) = 1$ for all i. Set $x = \sqrt{u} f_1$. Then in case $r = 1$, $f = x^2$. If $r = 2$, set $y = f_2$. The residue classes of x, y in $\mathfrak{m}/\mathfrak{m}^2$ are linearly independent since $f \notin (x^3) + x^2 \mathfrak{m}^2$ by (b). So $(x, y) = \mathfrak{m}$ and $f = x^2 y$. Since the converse is obvious, the lemma follows. □

3.8. By the lemma we see that the construction in 3.5 and 3.6 gives countably many isomorphism classes of MCM's for A_∞ and D_∞. The complete classification of all indecomposable MCM's on A_∞ and D_∞ in the curve case is given in §4. There are only countably many of such modules and by [Knö] this holds in all dimensions. Moreover, in both cases the module presented by the matrix $(x) \oplus (x)$ deforms into each of the modules presented by $\begin{pmatrix} x & y^k \\ 0 & x \end{pmatrix}$ for $k \geqq 1$, hence it is not of finite deformation type. Again by [Knö] the same holds in higher dimensions.

For the other non simple singularities we constructed in 3.5 and 3.6 families of modules parametrized by $\mathbb{C}$ or $\mathbb{P}^1_{\mathbb{C}}$ and it is easy to see that these families are flat over the parameter-space. Hence none of these modules is of countable deformation type and in particular the singularities are not of countable CM-representation type.

Putting everything together, we have almost completed the proof of Theorem B. Using Theorem A, the equivalences in (1) as well as the equivalence of (i), (iii) and (iv) in (2) follow immediately.

Now A_∞ deforms into the singularities of type A_k and D_∞ into the singularities of type A_k or D_k, hence these two singularities are of countable but not finite deformation type.

The converse is essentially contained in Arnold's proof of the classification of simple singularities. A direct argument goes as follows: Let f define a singularity which is neither simple nor A_∞ or D_∞. Since the deformation theory of $f \in \mathbb{C}\{z_0, \dots, z_n\}$ and $f + z_{n+1}^2 \in \mathbb{C}\{z_0, \dots, z_{n+1}\}$ is essentially the same, we may assume that $\mathrm{mult}(f) \geqq 3$. If n, the dimension of the singularity, is $\geqq 2$ or if $n = 1$ and $\mathrm{mult}(f) \geqq 4$, the initial part of f defines a projective variety which has moduli, hence the singularity defined by f itself is not of countable deformation type. In case $n = 1$, $\mathrm{mult}(f) = 3$, we may assume by Lemma 3.4

that, in suitable coordinates x, y,

$$f(x, y) = x^3 + a x^2 y^2 + b x y^4 + c y^6$$

with $a, b, c \in \mathbb{C}\{x, y\}$. Replacing x^3 by $x(x+\lambda y^2)(x+\lambda^2 y^2)$, $\lambda \in \mathbb{C}$, gives a family of nonisomorphic singularities (blowing up two times exhibits exactly four points on a rational component of the exceptional divisor, whose cross-ratio varies with λ). This completes the proof of Theorem B.

4. MCM's on A_∞ and D_∞

4.1. Let P be a two-dimensional noetherian regular local ring, $\mathfrak{m} \subset P$ the maximal ideal and x, y generators of $\mathfrak{m}$. We call a one-dimensional hypersurface ring R

$$A_\infty, \quad \text{if } R \cong P/(x^2), \text{ or}$$

$$D_\infty, \quad \text{if } R \cong P/(x^2 y).$$

Proposition. *Let M be an indecomposable* MCM *over A_∞. Then M has a presentation*

$$0 \to P^n \xrightarrow{\varphi} P^n \to M \to 0$$

with $n=1$ or $n=2$ and φ one of the following matrices

$$n=1: \quad (x), (x^2)$$

$$n=2: \quad \begin{pmatrix} x & y^k \\ 0 & -x \end{pmatrix}, \qquad k = 1, 2, 3, \ldots.$$

Proof. Consider the short exact sequence

$$0 \to \ker(x) \to M \to xM \to 0,$$

where M is any MCM over A_∞. Both, $\ker(x)$ and xM are torsion free $L = P/(x)$-modules. Since L is a discrete valuation ring both modules are free L-modules, say

$$xM \cong L^m, \qquad \ker(x) \cong L^n,$$

(xM may be 0).

By a suitable choice of bases, the natural inclusion $xM \to \ker(x)$ is given by a diagonal matrix

$$\left.\begin{pmatrix} y^{k_1} & & \\ & \ddots & \\ & & y^{k_m} \\ \hline & 0 & \end{pmatrix}\right\} n$$

$$\underbrace{\qquad\qquad}_{m}$$

with $0 \leqq k_1 \leqq \ldots \leqq k_m$. M is an extension of xM by $\ker(x)$ and we obtain a presentation

$$0 \to P^{n+m} \xrightarrow{\varphi} P^{n+m} \to M \to 0$$

with

$$\varphi = \left(\begin{array}{c|c} \begin{matrix} x & & \\ & \ddots & \\ & & x \end{matrix} & \begin{matrix} y^{k_1} & & \\ & \ddots & \\ & & y^{k_m} \end{matrix} \\ \hline 0 & \begin{matrix} -x & & \\ & \ddots & \\ & & -x \end{matrix} \end{array}\right) \begin{matrix} \} n \\ \} m \end{matrix}$$

$$\underbrace{\qquad}_{n} \qquad \underbrace{\qquad}_{m}$$

which shows that M is a direct sum of modules of the desired type. □

4.2. **Proposition.** *Let M be an indecomposable* MCM *over D_∞. Then M has a representation*

$$0 \to P^n \xrightarrow{\varphi} P^n \to M \to 0$$

with $n=1$ or $n=2$ and φ one of the following matrices

$$n=1: \quad (x), (y), (x^2), (xy), (x^2 y)$$

$$n=2: \quad \begin{pmatrix} x & y^k \\ 0 & -x \end{pmatrix}, \begin{pmatrix} xy & y^{k+1} \\ 0 & -xy \end{pmatrix},$$

$$\begin{pmatrix} xy & y^k \\ 0 & -x \end{pmatrix}, \begin{pmatrix} x & y^{k+1} \\ 0 & -xy \end{pmatrix}, \qquad k=1,2,3,\ldots.$$

Proof. If M is any MCM over D_∞, it yields a short exact sequence

$$0 \to \ker(y) \to M \to yM \to 0$$

which exhibits M as an extension of the torsion free (hence MCM) $A_\infty = P/(x^2)$-module yM by the free $P/(y)$-module $\ker(y)$. Using the classification of MCM's over A_∞ we obtain a presentation

$$0 \to P^n \xrightarrow{\varphi} P^n \to M \to 0$$

of M, where

$$\varphi = \left(\begin{array}{c|c} \begin{matrix} y & & \\ & \ddots & \\ & & y \end{matrix} & E \\ \hline 0 & \begin{matrix} x & 0 & 0 & y^{k_1} \\ 0 & x & y^{k_2} & 0 \\ & \ddots & & \\ 0 & & -x & 0 \\ & & 0 & -x \end{matrix} \end{array}\right) \begin{matrix} \} q \\ \} p \end{matrix}$$

$$\underbrace{\qquad}_{q} \qquad \underbrace{\qquad}_{p}$$

with $0 \leqq k_1 \leqq \ldots \leqq k_r \leqq \infty$, where we put $y^\infty := 0$, $y^0 := 1$ and use the presentation $\operatorname{coker}\begin{pmatrix} x & 1 \\ 0 & -x \end{pmatrix}$ of $\operatorname{coker}(x^2)$. The order of the generators of yM is chosen in such a way that we can apply the Gaußalgorithm to E later on. q is the rank of $\ker(y)$ as $P/(y)$-module, $p=2r$ (resp. $p=2r+1$) in case the pairity of the summands of yM isomorphic to (x) is even (resp. odd).

Hence

$$\begin{aligned} \operatorname{Ext}^1_R(yM, \ker y) &\cong \operatorname{Ext}^1_P(yM, \ker y) \\ &\cong (P/(x,y))^{p\cdot q} \quad (\text{as } k = P/\mathfrak{m}\text{-vectorspaces}) \\ &\cong k^{p\cdot q}. \end{aligned}$$

We now use the action of $\operatorname{Aut}(\ker(y)) \times \operatorname{Aut}(yM)$ on $\operatorname{Ext}^1_R(yM, \ker y)$ via pullback and pushout to simplify E:

Claim. After suitable row and column operations on φ the entries of E are either 1 or 0 and no row and column of E contains more than one nonzero entry.

Indeed: The orbits of the action of $Gl(q,k) \subset \operatorname{Aut}(\ker(y))$ correspond to a point in the union of Grassmannians $\bigcup_{0 \leqq t \leqq \min(p,q)} G(t,p)$; where $t = \operatorname{rank} E$, E regarded as a $q \times p$ matrix.

Hence we may assume that

$$E = \begin{pmatrix} 0 \ldots 1 & * \ldots & * & 0 & * \ldots \\ 0 \ldots 0 & 0 \ldots & * & 0 & * \ldots \\ \vdots \quad \vdots & \vdots & \vdots & \vdots & \vdots \\ & & * & 0 & * \ldots \\ & & 0 & 1 & * \ldots \\ & & 0 & 0 & 0 \ldots \\ & & \vdots & \vdots & \vdots \end{pmatrix}$$

where $*$ indicates an arbitrary entry in k.

Now use the action of $\operatorname{Aut}(yM)$, i.e. row operations on E. For the given order of generators of yM, the claim is easily checked, using $k_1 \leqq k_2 \leqq \ldots \leqq k_r$ and an induction on the cells of the Grassmannians.

Passing now to a minimal presentation, we obtain a direct sum decomposition of M into modules of the desired type. E.g.:

$$\operatorname{coker}\left(\begin{array}{cc|cc} y & 0 & 1 & 0 \\ 0 & y & 0 & 1 \\ \hline 0 & 0 & x & y^k \\ 0 & 0 & 0 & -x \end{array}\right) \cong \operatorname{coker}\begin{pmatrix} xy & y^{k+1} \\ 0 & -xy \end{pmatrix}.$$

References

[ALG] Bourbaki, N.: Algèbre, Chap. X. Paris: Hermann 1980

[Arn] Arnol'd, V.I.: Normal forms for functions near degenerate critical points, the Weyl groups of A_k, D_k, E_k and Lagrangian singularities. Funkts. Anal. **6**, 254–272 (1972)

[A-V] Artin, M., Verdier, J.L.: Reflexive modules over rational double points. Math. Ann. **270**, 79–82 (1985)

[Aus1] Auslander, M.: Rational singularities and almost split sequences. Trans. Am. Math. Soc. **293**, 511–531 (1986)

[Aus2] Auslander, M.: Isolated singularities and almost split sequences. Ottawa Proceedings 1984, SLN

[B-P-V] Barth, W., Peters, C., van de Ven, A.: Compact complex surfaces. Erg. d. Math. Band 4, Berlin-Heidelberg-New York-Tokio: Springer 1984

[Bru] Bruns, W.: The Eisenbud-Evans generalized principal ideal theorem and determinantal ideals. Proc. Am. Math. Soc. **83**, 19–26 (1981)

[Dur] Durfee, A.: Fifteen characterizations of rational double points and simple critical points. L'Enseignement Math. **25**, 131–163 (1979)

[Eis] Eisenbud, D.: Homological Algebra on a complete intersection, with an application to group representations. Trans. Am. Math. Soc. **260**, 35–64 (1980)

[Esn] Esnault, H.: Reflexive modules on quotient singularities. J. Reine Angew. Math. (in press) (1987)

[G-K] Greuel, G.-M., Knörrer, H.: Einfache Kurvensingularitäten und torsionsfreie Moduln. Math. Ann. **270**, 417–425 (1985)

[Kap] Kaplansky, I.: Commutative Rings, Rev. Edition, Univ. of Chicago Press, Chicago, 1974

[K-S] Kiyek, K., Steinke, G.: Einfache Kurvensingularitäten in beliebiger Charakteristik. Arch. Math. (in press) (1987)

[K-L] Kleiman, S.L., Landolfi, J.: Geometry and deformation of special Schubert varieties. In: Oort, F. (ed.), Algebraic geometry, Oslo 1970, pp. 97–124. Wolters-Noordhoff Publ. Groningen 1972

[Knö] Knörrer, H.: Cohen-Macaulay modules on hypersurface singularities I. Invent. Math. **88**, 153–164 (1987)

[O-S-S] Okonek, C., Schneider, M., Spindler, H.: Vectorbundles on complex projective spaces. Boston-Basel-Stuttgart: Birkhäuser 1980

[Sch] Schreyer, F.-O.: Finite and countable CM-representation type. To appear in Proc. Symp. on Singularities, Representations of algebra and vector bundles

[Sie] Siersma, D.: Isolated line singularities. In: Singularities. Proc. Symp. Pure Math. **40**, 485–496 (1983)

Oblatum 30-V-1986

Invent. math. 88, 183–215 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

The construction of weight-two arithmetic cohomology

Stephen Lichtenbaum*

Department of Mathematics, Cornell University, Ithaca, NY 14853, USA

Introduction

Let X be a regular noetherian scheme. In a letter to Soulé ([Be]) Beilinson conjectured for each non-negative integer n the existence of complexes $\Gamma(n, X)$ of Zariski sheaves on X whose hypercohomology should be in the same relation to algebraic K-theory as classical cohomology is to topological K-theory, and proposed certain properties that these complexes should satisfy. This investigation was continued in [L], where the conjectured complexes were moved to the étale site. Further conjectures were also made in [L] relating the hypercohomology of these complexes in arithmetic situations to values of zeta-functions and to duality theorems, thus justifying the name "arithmetic cohomology groups of X". These complexes were conjectured to satisfy six "Axioms" closely related to the properties proposed by Beilinson for his complexes of Zariski sheaves. Since this paper is almost exclusively concerned with the ^{5}tale site, we will refer to the ^{5}tale complex as $\Gamma(n, X)$, and use $\Gamma_B(n, X)$ when talking about the Zariski site.

The complex $\Gamma(0, X)$ is just the sheaf $\mathbb{Z}$ in degree 0, and the complex $\Gamma(1, X)$ is the sheaf G_m in degree 1. In this paper we construct a two-term complex which is an excellent candidate for $\Gamma(2, X)$, in that it essentially satisfies the aforementioned "Axioms". (We use the word "essentially" because at some points, in order to use Gersten's conjecture, we are forced to assume that X is of finite type over a field; at other points we work only up to 2-torsion or up to torsion involving primes dividing the residue field characteristics. There seems good reason to expect our complex $\Gamma(2, X)$ to in fact exactly satisfy all the axioms.)

We now sketch the construction of $\Gamma(2, X)$ and recall the axioms. Let A be a regular noetherian ring, let $W = \operatorname{Spec} A[T]$, $Z = \operatorname{Spec} A[T]/T(T-1)$. Let $B = \{b_1, b_2, \ldots, b_n\}$ be a finite sequence of "exceptional units" of A, i.e., b_i and $1-b_i$ are both units for all i. Let $Y_B = \operatorname{Spec} A[T] \Big/ \prod_{i=1}^{n} (T-b_i)$. Then there is an exact sequence

$$K_3(A) \to K_2(W - Y_B, Z) \xrightarrow{\varphi_{A,B}} K_1'(Y_B) \to K_2(A).$$

* Partially supported by a grant from the National Science Foundation

Let $C_{2,1}(A)=\varinjlim_B K_2(W-Y_B, Z)$, $C_{2,2}(A)=\varinjlim K_1'(Y_B)$, and $\varphi_A=\varinjlim \varphi_{A,B}$. Let $\tilde{\Gamma}(2, A)$ be the two-term complex $(C_{2,1} \xrightarrow{\varphi_A} C_{2,2}(A))$, with $C_{2,1}(A)$ in degree 1 and $C_{2,2}(A)$ in degree 2.

Now let X be a regular noetherian scheme. If $U=\operatorname{Spec} A$ is an open affine étale over X, the functors $U \to C_{2,1}(A)$ for $i=1,2$ evidently determine a two-term complex of presheaves on X for the étale topology and we define $\Gamma(2, X)$ to be the associated complex of sheaves.

We now recall the axioms:

Axiom 1. $\Gamma(2, X)$ is acyclic outside of $[1, 2]$.

This is clearly satisfied.

Axiom 2 (Hilbert's Theorem 90). Let α_* be the functor which associates to every étale sheaf on X its restriction to the Zariski site. Then the Zariski sheaf $R^3\alpha_*\Gamma(2, X)=0$. In particular if F is a field, then $H^3(F, \Gamma(2, F))=0$.

If X is a scheme of finite type over a field F of characteristic $p \geqq 0$, this is proven (Corollary 9.7) up to p-torsion and 2-torsion. If $X=\operatorname{Spec} F$, it is true up to p-torsion (Proposition 4.3).

Axiom 3 (Kummer theory). Let n be a positive integer prime to all residue field characteristics. Then there exists a triangle in the derived category

$$\begin{array}{ccc} & \mu_n \otimes \mu_n & \\ \swarrow & & \nwarrow \\ \Gamma(2, X) & \xrightarrow{n} & \Gamma(2, X). \end{array}$$

This is proven if X is of finite type over a field (Corollary 8.3).

Axiom 4 (Products). There is a product map

$$\Gamma(1, X) \overset{L}{\otimes} \Gamma(1, X) \to \Gamma(2, X).$$

This is true for all X (*Proposition* 2.6).

Axiom 5 (Connection with K-theory). The cohomology sheaves $\mathscr{H}^1(X, \Gamma(2, X))$ and $\mathscr{H}^2(X, \Gamma(2, X))$ are isomorphic to the étale sheaves $gr_\gamma^2 K_3(X)$ and $gr_\gamma^2 K_2(X)$, respectively, up to 2-torsion.

This is proven if X is of finite type over a field (Corollary 8.4).

Axiom 6. If F is a field, $H^2(F, \Gamma(2))$ is canonically isomorphic to $K_2(F)$.

We prove (Proposition 4.3) that there is a natural injection of $K_2(F)$ into $H^2(F, \Gamma(2))$, with p-torsion cokernel, if p is the characteristic of F.

In addition to the construction of $\Gamma(2, X)$ and the proof that it satisfies the axioms, this paper contains other results of independent interest. A key point in the proof that $\Gamma(2, X)$ satisfies the axioms is the construction, following Bloch, of a map from $\operatorname{Tor}_1(F^*, F^*)$ to $K_3(F)_{\text{ind}}$ (where F is a field), and the proof that this map is injective. The injectivity is shown by reducing to the

case when $F=\mathbb{C}$, and then constructing an explicit "dilogarithm map" from $K_3(\mathbb{C})_{\text{ind}}$ to $\mathbb{C}^*$ which is injective on the image of $\text{Tor}_1(\mathbb{C}^*, \mathbb{C}^*)$. (See Theorem 7.1 and Corollary 7.2.)

Also, if F is a field of characteristic $p \geqq 0$, we obtain (mod p-torsion) the following exact sequence which completely describes the failure of $K_2(F)$ to satisfy Galois descent. (Let F_s be a separable closure of F and $G=\text{Gal}(F_s/F)$.)

$$0 \to \bigoplus_{\ell \neq p} (K_3(F_s)_{\text{ind}})^G \otimes Q_\ell/Z_\ell) \to \bigoplus_{\ell \neq p} H^1(F, Q_\ell/Z_\ell(2)) \to K_2(F) \to K_2(F_s)^G \to \bigoplus_{\ell \neq p} H^2(F, Q_\ell/Z_\ell(2)) \to 0, \tag{0.1}$$

where the direct sums are taken over all primes $\ell \neq p$. (The description of the kernel of the map from $K_2(F)$ to $K_2(F_s)^G$ is due to Suslin and made use of in this paper; it is the computation of the cokernel which seems to be new.)

Finally, our construction yields *two* natural candidates for Beilinson's conjectured complexes of Zariski sheaves: $\Gamma(2)$ restricted (as presheaf) to the Zariski site and $\tau_{\leqq 2} R\alpha_* \Gamma(2)$, where α_* is the natural map from étale sheaves to Zariski sheaves. It is natural to ask whether these are in fact the same; this amounts to asking whether $K_3(F)_{\text{ind}}$ satisfies Galois descent, i.e., whether the natural map from $K_3(F)_{\text{ind}}$ to $(K_3(F_s)_{\text{ind}})$ is an isomorphism. I am not sure what to expect.

The interested reader should also refer to Bloch's new preprint [B12], where a proposal is made for a definition of $\Gamma(n, X)$ for all $n \geqq 0$ and all schemes X of finite type over a field. It would be interesting to verify that Bloch's definition agrees with ours for $n=2$ (in the derived category, of course).

I would like to take this opportunity to thank S. Bloch, K. Dennis, B. Mazur, D. Ramakrishnan and C. Soulé for many helpful conversations.

§ 1. The necessary foundations

We begin with some notation and terminology. If $\mathscr{A}$ is an exact category, let $BQ(\mathscr{A})$ denote the classifying space of the Quillen category of $\mathscr{A}$. If X is a scheme, let $\mathscr{M}(X)$ denote the category of coherent sheaves of O_X-modules and let $\mathscr{P}(X)$ denote the category of locally free sheaves of O_X-modules of finite rank. Let $BQ(X)$ denote $BQ\mathscr{P}(X)$, and $BQ'(X)$ denote $BQ\mathscr{M}(X)$. If A is a commutative ring, let $BQ(A)=BQ(\text{Spec}\, A)$ and $BQ'(A)=BQ'(\text{Spec}\, A)$. We have then $K_n(X)=\pi_{n+1}(BQ(X))$, and $K'_n(X)=\pi_{n+1}(BQ'(X))$. If (Y, i) is a closed subscheme of X, let $F_{X,Y}$ be the fiber of the map $BQ(X) \to BQ(Y)$ induced by i^*. Then the relative K-groups are defined by $K_n(X, Y)=\pi_{n+1}(F_{X,Y})$. If X is regular and quasi-projective over an affine scheme, then the inclusion map induces a homotopy equivalence of $BQ'(X)$ with $BQ(X)$.

We next give a slight modification of Grayson's elucidation [Gr] of Waldhausen's definition of products in K-theory [W]. Let $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$ be exact categories together with a biexact functor from $\mathscr{A} \times \mathscr{B}$ to $\mathscr{C}$. Then Waldhausen defines a map of topological spaces $BQ\mathscr{A} \wedge BQ\mathscr{B} \to BQQ\mathscr{C}$, where $BQQ\mathscr{C}$ is Waldhausen's "double Q construction". (We recall that $\pi_i(BQQ\mathscr{C})$ is

naturally isomorphic to $K_{i-2}(\mathscr{C})$.) This map of topological spaces is functorial in the exact categories $\mathscr{A}$, $\mathscr{B}$ and $\mathscr{C}$ with respect to exact functors (in the sense that the obvious diagrams of topological spaces *commute*, not merely commute up to homotopy). This map induces natural pairings $K_n(\mathscr{A}) \times K_m(\mathscr{B}) \to K_{m+n}(\mathscr{C})$. In particular, this applies to the case when X is a scheme, $\mathscr{A}=\mathscr{P}(X)$, $\mathscr{B}=\mathscr{C}$ is either $\mathscr{P}(X)$ or $\mathscr{M}(X)$, and the map from $\mathscr{A}\otimes\mathscr{B}$ to $\mathscr{C}$ is given by tensor product.

This also extends to the relative case. Let A be a regular noetherian ring, and let $i\colon Z \to X$ be a closed immersion of A-schemes of finite type. Then i^* induces a map from $BQ(X)$ to $BQ(Z)$. Let $\widetilde{BQ}(X) = BQ(X) \underset{BQ(Z)}{\times} (BQ(Z)^I)$, so we have a fibration $\widetilde{BQ}(X)$ to $BQ(Z)$. Let $F_{X,Z}$ be the fiber of this map, and recall that the relative K-groups are defined by the formula $K_i(X,Z) = \pi_{i+1}(F_{X,Z})$ for $i \geqq 0$.

Similarly i^* induces a map from $BQQ(X)$ to $BQQ(Z)$. Let $\widetilde{BQQ}(X)$ be defined as above, and let $FF_{X,Z}$ be the fiber of the map from $\widetilde{BQQ}(X)$ to $\widetilde{BQQ}(Z)$. Then we have a commutative diagram of spaces

$$\begin{array}{ccc} BQ(A) \wedge \widetilde{BQ}(X) & \longrightarrow & \widetilde{BQQ}(X) \\ \downarrow & & \downarrow \\ BQ(A) \wedge \widetilde{BQ}(Z) & \longrightarrow & \widetilde{BQQ}(Z) \end{array}$$

which induces a map $BQ(A) \wedge F_{X,Z} \to FF_{X,Z}$. We also have ([W], p. 196) a commutative diagram

$$\begin{array}{ccc} BQ(X) & \xrightarrow{\varphi_X} & \Omega BQQ(X) \\ \downarrow & & \downarrow \\ BQ(Z) & \xrightarrow{\varphi_Z} & \Omega BQQ(Z) \end{array}$$

which induces (after taking path-spaces) a map ψ from $F_{X,Z}$ to $\Omega FF_{X,Z}$. Since φ_X and φ_Z are homotopy equivalences, by the five-lemma applied to the exact sequences of the fibrations we see that ψ induces isomorphisms from $\pi_i F_{X,Z}$ to $\pi_{i+2}\Omega FF_{X,Z} = \pi_{i+1}(FF_{X,Z})$. Hence we may identify $K_i(X,Z)$ with $\pi_{i+2}(FF_{X,Z})$. Also, every map defined is natural in X and Z. We then obtain a long exact sequence coming from the fibration $F_{X,Z} \to BQ(X) \to BQ(Z)$.

$$(*) \qquad K_{i+1}(Z) \to K_i(X,Z) \to K_i(X) \to K_i(Z) \to \ldots$$

We also have maps ψ from $K_j(A) \otimes K_i(X,Z) \to K_{i+j}(X,Z)$, $K_j(A) \otimes K_i(X) \to K_{i+j}(X)$ and $K_j(A) \otimes K_i(Z) \to K_{i+j}(Z)$, and it follows from Theorem 2.5 of Grayson [Gr] that these maps ψ are compatible with any of the maps φ in the exact sequence $(*)$, where the indices are chosen suitably.

Proposition 1.1. i) *Let A be a regular noetherian ring. Let X be a regular noetherian scheme over A, and let $Y, Z \subset X$ be closed subschemes such that $Y \cap Z = \emptyset$. Then there is a long exact localization sequence:*

$$\ldots \to K_n(X, Z) \to K_n(X - Y, Z) \to K'_{n-1}(Y) \to K_{n-1}(X, Z) \to \ldots$$

with compatible $K_*(A)$-actions in the sense of the preceding paragraph. This sequence is functorial in Y and Z in the obvious sense.

ii) *Now assume that X, Y and Z are faithfully flat over A with the property that every coherent sheaf on X, Y or Z is the quotient of a locally free sheaf. Then, if A' is a regular ring over A, $X' = X \underset{A}{\times} A'$, $Y' = Y \underset{A}{\times} A'$, and $Z' = Z \underset{A}{\times} A'$, there are natural maps from the sequence for (X, Y, Z) to the sequence for (X', Y', Z'), making the obvious diagram commute.*

Proof. We follow Bloch [Bl1] in considering the diagram of fibrations

$$\begin{array}{ccccc} F_{X-Y,Z} & \longrightarrow & \widetilde{BQ}(X-Y) & \longrightarrow & BQ(Z) \\ \uparrow & & \uparrow & & \| \\ F_{X,Z} & \longrightarrow & BQ(X) & \longrightarrow & BQ(Z). \end{array}$$

Replacing $\widetilde{BQ}(X)$ by $\widetilde{\widetilde{BQ}}(X) = \widetilde{BQ}(X) \underset{\widetilde{BQ}(X-Y)}{\times} \widetilde{BQ}(X-Y)^I$, we obtain the diagram of fibrations:

$$\begin{array}{ccccc} F_{X-Y,Z} & \longrightarrow & \widetilde{BQ}(X-Y) & \longrightarrow & BQ(Z) \\ \uparrow & & \uparrow & & \| \\ \tilde{F}_{X,Z} & \longrightarrow & \widetilde{\widetilde{BQ}}(X) & \longrightarrow & BQ(Z) \\ & & \uparrow & & \uparrow \\ & & H_{X,Y} & \longrightarrow & * \end{array}$$

where $H_{X,Y}$ is defined to be the fiber of the map from $\widetilde{\widetilde{BQ}}(X)$ to $\widetilde{BQ}(X-Y)$ and $\tilde{F}_{X,Z} = \tilde{F}_{X,Z} \underset{\widetilde{BQ}(X-Y)}{\times} \widetilde{BQ}(X-Y)^I$. This diagram in turn induces the fibration:

$$H_{X,Y} \to \tilde{F}_{X,Z} \to F_{X-Y,Z}.$$

Here $\tilde{F}_{X,Z}$ is naturally homotopic to $F_{X,Z}$, and $H_{X,Y}$ is naturally homotopic to the homotopy fiber of $BQ(X) \to BQ(X-Y)$, which is in turn [Q1, p. 128] naturally homotopic to $BQ'(Y)$.

We then repeat the same construction with FF's replacing F's and QQ's replacing Q's, define $HH_{X,Y}$ to be the fiber of the map from $\widetilde{\widetilde{BQQ}}(X)$ to $\widetilde{BQQ}(X-Y)$, and obtain the fibration:

$$HH_{X,Y} \to \widetilde{FF}_{X,Z} \to FF_{X-Y,Z}.$$

As above, we have a natural homotopy equivalence from $H_{X,Y}$ to $\Omega HH_{X,Y}$. As above, we get compatible maps $BQ_A \wedge H_{X,Y} \to HH_{X,Y}$, $BQ_A \wedge \tilde{F}_{X,Z} \to \widetilde{FF}_{X,Z}$ and $BQ_A \wedge F_{X-Y,Z} \to FF_{X-Y,Z}$, and again we use Theorem 2.5 of [Gr] to finish the

demonstration of the first part of the proposition once we observe that Quillen's identification up to homotopy of $H_{X,Y}$ with $BQ'(Y)$ is compatible with the action of $K_*(A)$.

We now prove ii). The functoriality of BQ assures us of compatible natural maps from $BQ(X)$, $BQ(X-Y)$ and $BQ(Z)$ to $BQ(X')$, $BQ(X'-Y')$ and $BQ(Z')$ respectively. These immediately induce a map from the fibration $H_{X,Y} \to F_{X,Z} \to F_{X-Y,Z}$ to the fibration $H_{X',Y'} \to F_{X',Z'} \to F_{X'-Y',Z'}$, and hence natural maps from one homotopy sequence to the other.

We claim that, in addition, these maps are compatible with Quillen's identification of $H_{X,Y}$ with $BQ'(Y)$. It suffices to prove that they are compatible with the natural map from $BQ'(Y)$ to $BQ'(X)$ induced by i_*, where $i: Y \to X$. We have to show that the following diagram of spaces commutes:

$$\begin{array}{ccc} BQ'(Y) & \longrightarrow & BQ'(X) \\ \downarrow & & \downarrow \\ BQ'(Y') & \longrightarrow & BQ'(X'). \end{array}$$

Recall Quillen's definition of the vertical maps [Q1, p. 125]. We have inside $\mathscr{M}_Y$ the exact subcategory $P(Y, Y')$ consisting of those coherent sheaves F on Y such that $\operatorname{Tor}_i^{O_Y}(O'_Y, F)=0$ for all $i>0$. If Y'/Y is of finite Tor-dimension and every sheaf on Y is the quotient of a locally free sheaf, then Quillen shows that $BQP(Y, Y') \to BQ'(Y)$ is a homotopy equivalence, and hence the map from $BQP(Y, Y') \to BQ'(Y')$ induced from $F \to F \underset{O_Y}{\otimes} O_{Y'}$ induces a map $BQ'(Y)$ to $BQ'(Y')$.

The only non-obvious step in proving the naturality is to show that $F \to F \underset{O_Y}{\otimes} O_X$ sends $P(Y, Y')$ to $P(X, X')$. Since this question is local on X, we may assume that X (and hence Y, X' and Y') are affine. Let $X=\operatorname{Spec} B$, $Y=\operatorname{Spec} C$, $X'=\operatorname{Spec} B'$, $Y'=\operatorname{Spec} C'$. Let F correspond to the C-module M. Suppose $\operatorname{Tor}_i^C(C', M)=0$. Since C is flat over A and

$$C' = C \underset{A}{\otimes} A', \quad \operatorname{Tor}_i^C(C', M) \simeq \operatorname{Tor}_i^A(A', M) \underset{A'}{\otimes} C'.$$

Since C' is faithfully flat over A', $\operatorname{Tor}_i^A(A', M)=0$, and hence

$$\operatorname{Tor}_i^B(B', M) \simeq \operatorname{Tor}_i^A(A', M) \underset{A'}{\otimes} B' = 0.$$

Lemma 1.2. *Let A be a regular noetherian ring. Let $X=\operatorname{Spec} A[T]$, and $Z=\operatorname{Spec} A[T]/T(T-1)$. Then $K_n(X, Z)$ is naturally isomorphic to $K_{n+1}(A)$.*

Proof. The basic exact sequence of relative K-theory yields

$$\ldots \to K_n(X, Z) \to K_n(X) \to K_n(Z) \to K_{n-1}(X, Z) \to \ldots$$

By Quillen [Q1, Corollary to Theorem 8], we know that $K_n(X)$ is naturally isomorphic to $K_n(A)$. Since Z is isomorphic to $A \times A$, $K_n(Z)$ is isomorphic to $K_n(A) \otimes K_n(A)$, and it is easily seen that the map from $K_n(X)$ to $K_n(Z)$ induces the diagonal map Δ from $K_n(A)$ to $K_n(A) \otimes K_n(A)$. The lemma now follows immediately.

1.3. Now let $B=\{b_1, b_2, \ldots, b_n\}$ be a finite sequence of elements of A, and assume that the b_i's are "*exceptional units*", i.e., that b_i and $1-b_i$ are both units, for all i. Let $Y_B=\operatorname{Spec} A[T]\Big/\prod_{i=1}^{n}(T-b_i)$, and X and Z be as in Lemma 1.2. Then Proposition 1.1 and Lemma 1.2 combine to show (since $Y_B\cap Z=\emptyset$), that we have

Lemma 1.4. *There exists a long exact sequence*

$$\ldots \to K_{n+1}(A)\to K_n(X-Y_B, Z)\xrightarrow{\varphi_{n,B}} K'_{n-1}(Y_B)\xrightarrow{\psi_{n,B}} K_n(A)\to \ldots$$

This sequence is a covariant functor of A, is natural in B in the obvious sense, and is compatible with the $K_*(A)$-action.

Definition 1.5. Let $C_{n,1}(A)=\varinjlim_B K_n(X-Y_B, Z)$, $C_{n,2}(A)=\varinjlim_B K'_{n-1}(Y_B)$, $\varphi_n\colon C_{n,1}(A)\to C_{n,2}(A)=\varinjlim_B \varphi_{n,B}$, and $\psi_n\colon C_{n,2}(A)\to K_n(A)=\varinjlim_B \psi_{n,B}$.

Corollary 1.6. *There exists a long exact sequence of $K_*(A)$-modules:*

$$\ldots \to K_{n+1}(A)\to C_{n,1}(A)\to C_{n,2}(A)\to K_n(A)\to \ldots$$

This sequence is a covariant functor of A.

Lemma 1.7. *Let A be local and assume that the residue field of A has more than two elements or A is a field. Then φ_1 is surjective.*

Proof. In this case $C_{1,2}(A)=\varinjlim_B K'_0(Y_B)$. By functoriality, there is a commutative diagram

$$\begin{array}{ccc} \coprod_{a_i} K_0(A[T]/T-a_i) & & \\ \downarrow & \searrow & \\ C_{1,2}(A) & \longrightarrow & K_1(A)\simeq A^*. \end{array}$$

Now $K_0(A[T]/T-a_i)\simeq K_0(A)\simeq Z$, and we claim that the above diagonal map takes the generator $[A]$ to $(a_i-1)/a_i\in A^*$.

By functoriality and the fact that $K_1(A)$ injects into $K_1(K)$, where K is the quotient field of A, we may assume that A is a field. Let $\alpha=a_i$ and let $Y=\operatorname{Spec} A[T]/(T-\alpha)$. Then consider the commutative diagram with exact columns and rows:

$$\begin{array}{ccccc} K_1(X) & \longrightarrow & K_1(X-Y) & \longrightarrow & K'_0(Y) \\ \downarrow & & \downarrow & & \\ K_1(Z) & \longrightarrow & K_1(Z) & & \\ \downarrow & & \downarrow & & \\ K_0(X, Z) & \longrightarrow & K_0(X-Y, Z). & & \end{array}$$

To compute the image of $[A]$ in $K_0(X, Z) \simeq K_1(A)$ we lift it back to $K_1(X-Y)$, obtaining for example $(T-\alpha)$ in $(A[T]_{T-\alpha})^*$. This maps by specialization at 0 and 1 to $(-\alpha, 1-\alpha) \in K_1(Z) = A^* \otimes A^*$, and the map $(\beta, \gamma) \to \gamma/\beta$ identifies $K_1(Z)/\Delta K_1(A)$ with $K_1(A)$, in this case yielding $(\alpha-1)/\alpha$.

Since the map $\alpha \to (\alpha-1)/\alpha$ permutes the exceptional units, it suffices to show the exceptional units generate A^*. In a local ring, a unit is exceptional is it is not congruent to 1 modulo the maximal ideal, so if there are more than two elements in the residue field the exceptional units generate.

Let $K_n(A)_{\text{dec}}$ denote the image in $K_n(A)$ of $K_1(A) \otimes K_{n-1}(A)$ under the product map in K-theory, and let $K_n(A)_{\text{ind}} = K_n(A)/K_n(A)_{\text{dec}}$.

Proposition 1.8. a) *If A is local with residue field containing more than two elements, then the image of ψ_n contains $K_n(A)_{\text{dec}}$.*

b) *If F is a field, then the image of ψ_n is equal to $K_n(F)_{\text{dec}}$.*

Proof. a) We use the compatibility of our long-exact localization sequence with the action of $K_*(A)$. Let $\beta \in K_{n-1}(A)$, $\gamma \in K'_0(A[T]/T-\alpha))$, and $B = \{\alpha\}$.

Let $\tilde{\gamma}$ be the image of γ in $C_{n,2}(A)$. Then $\psi_{n,B}(\beta\gamma) = \beta\psi_{0,B}(\gamma)$ and hence $\psi_n(\beta\tilde{\gamma}) = \beta\psi_0(\tilde{\gamma})$. Since the $\psi_0(\tilde{\gamma})$'s generate $K_1(A)$, the image of ψ_n contains $K_n(A)_{\text{dec}}$.

b) If F is a field, then $C_{n,2}(F)$ is naturally isomorphic to $\coprod_{x \in F-\{0,1\}} K_{n-1}(F)_x$. Let $\gamma \in K_{n-1}(F)_x$. Then $\psi_n(\gamma) = \psi_n(\gamma, 1_x)$, $(\gamma \in K_{n-1}(F), 1_x \in K_0(F)_x) = \gamma\psi_1(1_x)$, which shows that $\psi_n(\gamma) \in K_n(F)_{\text{dec}}$. So the image of ψ_n is contained in $K_n(F)_{\text{dec}}$, so is equal to it by a).

§ 2. The construction of $\Gamma(1, X)$, $\Gamma(2, X)$ and the product map

Let A be a regular noetherian ring.

Definition 2.1. Let $\Gamma(i, A)$ be the complex of abelian groups

$$C_{i,1}(A) \xrightarrow{\varphi_{i,A}} C_{i,2}(A), \quad \text{for } i = 1 \text{ or } 2.$$

Remark 2.2. We define the degree of $C_{2,i}(A)$ to be i, and the degree of $C_{1,i}(A)$ to be $i-1$. The $K_*(A)$-structure on the $C_{i,j}(A)$ induces a map of complexes of degree 1 from $\tilde{\Gamma}(1, A) \otimes K_1(A)$ to $\tilde{\Gamma}(2, A)$, which is of course functorial in A.

Definition 2.3. Let X be a regular noetherian scheme. If $U = \operatorname{Spec} A$ is an open affine lying over X, the functors $U \to C_{i,j}(A)$ for $1 \leqq i, j \leqq 2$ evidently determine presheaves on X and hence associated sheaves of abelian groups $C_{i,j}(X)$ with respect to any of the usual topologies on X (Zariski, étale, flat), and we define $\Gamma_{\text{Zar}}(i, x)$ (resp. $\Gamma(i, x)$) to be the complex of sheaves on X for the Zariski (resp. étale) topologies on

$$X\colon C_{i,1} \xrightarrow{\varphi_{i,x}} C_{i,2}(X).$$

Proposition 2.4. *Then natural map from $C_{1,2}(X)$ to $K_1(X)$ induces an isomorphism in the derived category of sheaves on X for the étale topology from $\Gamma(1, X)$ to $G_m[-1]$. The same is true in the Zariski topology if we assume that all residue fields of X have more than two elements.*

Proof. This only amounts to the statement that if A is a regular local ring whose residue field has more than two elements, the sequence

$$0 \to C_{1,1}(A) \to C_{1,2}(A) \to K_1(A) \to 0$$

is exact, since $K_1(A)$ is naturally isomorphic to A^*. But this follows immediately from Corollary 1.6, Lemma 1.7, and Proposition 1.8, since $K_2(A) = K_2(A)_{\text{dec}}$ [D-S, p. 287].

Proposition 2.5. *The map of complexes referred to in Remark 2.2 induces a map in the derived category of étale sheaves on X from $\Gamma(1, X) \overset{L}{\otimes} \Gamma(1, X)$ to $\Gamma(2, X)$, where $\overset{L}{\otimes}$ denotes the usual tensor product in the derived category. If every residue field of X has more than two elements, the same is true in the Zariski topology.*

Proof. It suffices to show that if A is regular local, $C_{1,1}(A)$ and $C_{1,2}(A)$ are flat, i.e., torsion-free. But it is an easy exercise to check that for $R = A[T] \Big/ \prod_{i=1}^{n} (T - a_i)$, a_i distinct, the natural map from $\oplus K'_0(A[T]/(T-a_i))$ to $K'_0(R)$ induces an isomorphism of Z^n with $K'_0(R)$. This implies that $C_{1,2}(A)$ is the direct limit of free groups, hence torsion-free.

Remark 2.6. For future reference, we write this out more explicitly. We have:

$$(**) \qquad \begin{array}{ccccccccc} 0 \to & \mathrm{Tor}_1(A^*, A^*) & \longrightarrow & C_{1,1}(A) \otimes A^* & \longrightarrow & C_{1,2}(A) \otimes A^* & \longrightarrow & A^* \otimes A^* & \to 0 \\ & \downarrow \rho_A & & \downarrow & & \downarrow & & \downarrow \psi_A & \\ 0 \to & \tilde{K}_3(A) & \longrightarrow & C_{2,1}(A) & \xrightarrow{\varphi_A} & C_{2,2}(A) & \longrightarrow & K_2(A) & \to 0. \end{array}$$

Here ψ_A is the natural surjection, and $\tilde{K}_3(A)$ is defined to be $\mathrm{Ker}(\varphi_A)$. It follows from Proposition 1.8 that if A is a field F, the natural map from $K_3(F)$ to $\tilde{K}_3(F)$ induces an isomorphism of $K_3(F)_{\text{ind}}$ with $\tilde{K}_3(F)$.

§ 3. The verification of the axioms for fields

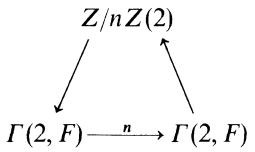
In this section we begin the demonstration that the complex $\Gamma(2, X)$ of étale sheaves defined in the previous section satisfy in large part the "axioms" listed in our paper [L]. We have already seen § 2 that Axioms 1 and 4 are satisfied, and we will demonstrate here that Axioms 2, 3, 5 and 6 are satisfied for fields (up to p-torsion, where p is the characteristic).

We begin by remarking that we already know Axiom 5 for fields, since we have seen Remark 2.6 that we have a natural exact sequence

$$(*) \qquad 0 \to K_3(F)_{\text{ind}} \to C_{2,1}(F) \xrightarrow{\varphi_F} C_{2,2}(F) \to K_2(F) \to 0$$

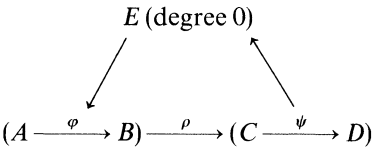
and Soulé has shown in [So] that $K_3(F)_{\text{ind}}$ is naturally isomorphic to $gr^2_\gamma K_3(F)$ and $K_2(F)$ is naturally isomorphic to $gr^2_\gamma K_2(F)$. (See § 8 for more information on the γ-filtration.)

Axiom 3 asserts the existence of a triangle in the derived category:



where n is any positive integer prime to the characteristic of F. We leave the following exercise in derived categories to the reader:

Exercise 3.1. Let $\mathscr{A}$ be an abelian category, $A \xrightarrow{\varphi} B$ and $C \xrightarrow{\psi} D$ complexes of objects of $\mathscr{A}$. Assume A and C are in degree 1, B and D in degree 2. Let $\rho: (A \xrightarrow{\varphi} B) \to (C \xrightarrow{\psi} D)$, and assume that ρ induces an isomorphism $\operatorname{Coker} \varphi \xrightarrow{\rho} \operatorname{Coker} \psi$, and a surjection $\operatorname{Ker} \varphi \xrightarrow{\rho} \operatorname{Ker} \psi$, with kernel E. Then there is a triangle in the derived category of $\mathscr{A}$:



We need at this point an easy generalization of results of Suslin from the case of algebraically closed fields to that of separably closed fields:

Lemma 3.2. *Let F be a separably closed field, of characteristic $p \geqq 0$, $\bar{F}$ its algebraic closure. Then:*

a) *The kernel and cokernel of the natural map $K_n(F) \to K_n(\bar{F})$ are p-torsion, and hence uniquely divisible by n prime to p.*

b) *The Milnor K-theory groups $K_{r,M}(F)$ are uniquely n-divisible for $r \geqq 2$ and n prime to p.*

c) *The image of $K_{3,M}(F)$ in $K_3(F)$ is uniquely divisible by n prime to p.*

Proof. (See also [D-W].) a) The standard formula "cores o res = multiplication by degree" shows that the kernel is p-torsion. Quillen's formula ([Q1, p. 130]) that "res o cores = multiplication by degree" for finite inseparable extensions shows that the cokernel is p-torsion.

b) The groups $K_{r,M}(\bar{F})$ are Q-vector spaces ([Ba-T], p. 357). The existence of a norm map on Milnor K-theory shows that $K_{r,M}(F)$ has no n-torsion for n prime to p. Since $K_{r,M}(F)$ is a quotient of $(F^*)^{\otimes r}$, F^* being n-divisible implies that $K_{r,M}(F)$ is n-divisible.

c) By Suslin [Su3, p. 3], $\operatorname{Ker}(K_{3,M}(F) \to K_3(F))$ is annihilated by 2, hence zero if p is not 2, by b). If p is 2, then the kernel in any case is uniquely n-divisible for n odd. In either case, c) immediately follows from b).

Proposition 3.3. *Axiom 3 is true for $\Gamma(2, F)$,* i.e., *for n prime to p, there is a triangle in the derived category of étale sheaves on F:*

$$\begin{array}{ccc} & Z/nZ(2) & \\ \swarrow & & \nwarrow \\ \Gamma(2,F) & \xrightarrow{n} & \Gamma(2,F) \end{array}$$

Proof. Let F_s be a separable closure of F. By Exercise 3.1 and the exact sequence (*), it is enough to show that $K_2(F_s)$ is uniquely divisible by n and $K_3(F)_{\text{ind}}$ is divisible by n with kernel $Z/nZ(2)$. The first statement is Lemma 3.2b, since $K_2 = K_{2,M}$. The second statement follows immediately from the corresponding theorem for $K_3(F)$ and algebraically closed fields ([Su1]), Lemma 3.2a, and Lemma 3.2c.

§ 4. Verification of axioms 2 and 6

We now begin the proof of Axioms 2 and 6. For the remainder of this section, let F be a field of characteristic $p \geqq 0$, and let n be a positive integer prime to p. Let F_s denote the separable closure of F and $G = \mathrm{Gal}(F_s/F)$.

We have seen in § 1 that there is a natural exact sequence

$$0 \to C_{1,1}(F) \xrightarrow{\psi_F} C_{1,2}(F) \xrightarrow{\rho_F} F^* \to 0.$$

We have also seen that $C_{1,2}(F)$ may be identified with $\coprod_{x \in F, x \neq 0,1} Z_x$, and that the map ρ_F may be described by saying that $\rho_F(1_x) = \dfrac{x-1}{x}$. It is then easy to see that $C_{1,1}(F)$ may be identified with the multiplicative group $J(F)$ of rational functions which have F-rational zeros and poles and are equal to 1 at 0 and 1. The map ψ_F is just the map taking any such function to the finite part of its divisor.

In view of the preceding paragraph, the complex $C_{1,1}(F) \xrightarrow{\psi_F} C_{1,2}(F)$ is equal to $F^*[-1]$ in the derived category, where we regard $C_{1,1}(F)$ as being in degree 0 and $C_{1,2}(F)$ in degree 1. Since $C_{1,1}(F_s)$ and $C_{1,2}(F_s)$ are torsion-free,

$$C_{1,1}(F_s) \underset{Z}{\otimes} F_s^* \xrightarrow{\psi_F \otimes 1} C_{1,2}(F_s) \underset{Z}{\otimes} F_s^*,$$

regarded as a complex in degrees 1 and 2, represents the derived tensor product complex $G_m[-1] \overset{L}{\otimes} G_m[-1]$.

Lemma 4.1. *There is a distinguished triangle in the derived category of étale sheaves of abelian groups on* $\mathrm{Spec}\, F$

$$\begin{array}{ccc} & \mu_n \otimes \mu_n & \\ {\scriptstyle r}\swarrow & & \nwarrow \\ G_m[-1] \overset{L}{\otimes} G_m[-1] & \xrightarrow{n} & G_m[-1] \overset{L}{\otimes} G_m[-1] \end{array}$$

such that the map r identifies $\zeta \otimes \omega$ *with* $(1-t+\zeta t)^n \underset{n}{\otimes} \omega$ *in*

$$J(F_s) \otimes F_s^* = C_{1,1}(F_s) \underset{Z}{\otimes} F_s^*.$$

Proof. This is an elementary computation, given Exercise 3.1. There is a natural map i from $F^* \otimes F^* = \mathrm{Coker}(\psi_F \otimes 1)$ to $\mathrm{Coker}(\psi_{F_s} \otimes 1)^G$, and from the hypercohomology spectral sequence (or from elementary diagram chasing) we obtain a natural map ρ from $\mathrm{Coker}(\psi_{F_s} \otimes 1)^G$ to $H^2(F, G_m[-1] \overset{L}{\underset{Z}{\otimes}} G_m[-1])$. Let Γ denote $G_m[-1] \overset{L}{\underset{Z}{\otimes}} G_m[-1]$, and ∂ denote the coboundary map from $H^2(F, \Gamma)$ to $H^2(F, \mu_n \otimes \mu_n)$.

Lemma 4.2. *The composite map* $\partial \circ \rho \circ i$: $F^* \otimes F^* \to H^2(F, \mu_n \otimes \mu_n)$ *is the negative of Tate's Galois symbol* h_F ([T], p. 262).

Proof. Let $A = C_{1,1}(F_s^*) \underset{Z}{\otimes} F_s^*$ and $B = C_{1,2}(F_s^*) \otimes F_s^*$. Since multiplication by n is surjective on A and B, it easily follows that we have the following commutative diagram with exact rows and columns:

$$\begin{array}{ccccccc}
H^0(F, A) & \longrightarrow & H^0(F, B) & \longrightarrow & H^2(F, \Gamma) & \longrightarrow & H^1(F, A) \\
\downarrow{\scriptstyle n} & & \downarrow{\scriptstyle n} & & \downarrow{\scriptstyle n} & & \downarrow{\scriptstyle n} \\
H^0(F, A) & \longrightarrow & H^0(F, B) & \xrightarrow{\rho} & H^2(F, \Gamma) & \longrightarrow & H^1(F, A) \\
\downarrow & & \downarrow{\scriptstyle \lambda} & & \downarrow{\scriptstyle \partial} & & \downarrow \\
H^1(F, A_n) & \longrightarrow & H^1(F, B_n) & \xrightarrow{\mu} & H^2(F, \mu_n \otimes \mu_n) & \longrightarrow & H_2(F, A_n)
\end{array}$$

so it suffices to compute $\mu \circ \lambda \circ i$.

Let $a \otimes b$ be in $F^* \otimes F^*$. Then $i(a \otimes b)$ is represented by $(b)_{\frac{1}{1-a}}$ in $H^0(F, B) = \coprod_{x \neq 0, 1} ((F_s)^*)^G$. Continuing with λ, $\lambda \circ i(a \otimes b)$ is represented by the class of the cocycle $\sigma \to (\sigma(\sqrt[n]{b})/\sqrt[n]{b})_{\frac{1}{1-a}}$ in $H^1(F, B_n)$.

Let $Q \subseteq B$ be $(\psi_{F_s} \otimes 1)(A)$, and let $K = \mathrm{Ker}(\psi_{F_s} \otimes 1) = \mathrm{Tor}_1(F_s^*, F_s^*)$. Then $Q_n = B_n$ and we may compute μ by looking at the exact sequence $0 \to K_n \to A_n \to Q_n \to 0$ and computing the coboundary map, remembering that $\mu_n \otimes \mu_n$ is inside A_n via the identification $\zeta \otimes \omega \to (1-t+\zeta t)^n \otimes \omega$.

We next observe that the element $\dfrac{1-t+at}{(1-t+t\sqrt[n]{a})^n} \otimes c$ maps to $c_{\frac{1}{1-a}} - c^n{}_{\frac{1}{1-\sqrt[n]{a}}}$ in $\coprod F_s^* = C_{1,2}(F_s) \underset{Z}{\otimes} F_s^*$, so if c is an n-th root of unity ω, $\dfrac{1-t+at}{(1-t+t\sqrt[n]{a})^n} \otimes \omega \to \omega_{\frac{1}{1-a}}$. From this it follows that a representative cocycle for $\mu \circ \lambda \circ i(a \otimes b)$ is given by f where

$$f(\sigma,\tau)=\sigma\left(\frac{1-t+at}{(1-t+t\sqrt[n]{a})^n}\right)\otimes\sigma\left(\frac{\tau(\sqrt[n]{b})}{\sqrt[n]{b}}\right)$$
$$-\frac{1-t+at}{(1-t+t\sqrt[n]{a})^n}\otimes\frac{\sigma\tau(\sqrt[n]{b})}{\sqrt[n]{b}}$$
$$+\frac{1-t+at}{(1-t+t\sqrt[n]{a})^n}\otimes\frac{\sigma(\sqrt[n]{b})}{\sqrt[n]{b}},$$

which simplifies to $\left(\text{using } \frac{(1-t+\alpha t)^n}{(1-t+\beta t)^n}\otimes\omega=\left(1-t+\frac{\alpha t}{\beta}\right)^n\otimes\omega\right)$

$$(1-t+\rho_\sigma^{-1}t)\otimes\sigma\zeta_\tau,$$

where $\rho_\sigma=\sigma(\sqrt[n]{a})/\sqrt[n]{a}$, $\zeta_\tau=\tau(\sqrt[n]{b})/\sqrt[n]{b}$. Under our identification, this corresponds to $(\rho_\sigma\otimes\sigma\zeta_\tau)^{-1}$, and since Tate's map sends $a\otimes b$ to $\rho_\sigma\otimes\sigma\zeta_\tau$, we are done.

We now begin the proof of Axiom 6 by constructing a map θ_F from $K_2(F)$ to $H^2(F,\Gamma(2,F_s))$. We will then prove that θ_F is injective, with p-torsion cokernel.

The spectral sequence of hypercohomology yields the exact sequence $0\to\mathrm{Coker}(\varphi_{F_s}^G)\to H^2(F,\Gamma(2,F_s))\to\mathrm{Ker}(H^1(G,\varphi_{F_s}))$. There is an evident map from $K_2(F)$ to $\mathrm{Coker}(\varphi_{F_s}^G)$ and hence to $H^2(F,\Gamma(2,F_s))$, and we call this map θ_F. In the previous section we have defined a map $\rho\circ i$ from $F^*\otimes F^*$ to $H^2(F,G_m[-1]\overset{L}{\otimes}G_m[-1])$, and it is evident that the diagram

$$\begin{array}{ccc} F^*\otimes F^* & \xrightarrow{\rho\circ i} & H^2(F,G_m[-1]\overset{L}{\otimes}G_m[-1]) \\ \downarrow & & \downarrow \\ K_2(F) & \longrightarrow & H^2(F,\Gamma(2,F_s)) \end{array}$$

commutes. (See Remark 2.6 and the comments preceding Lemma 4.1.) We can adjoin to this the commutative diagram

$$\begin{array}{ccc} H^2(F,G_m[-1]\overset{L}{\otimes}G_m[-1]) & \xrightarrow{\delta} & H^2(F,\mu_n\otimes\mu_n) \\ \downarrow & & \downarrow \\ H^2(F,\Gamma(2,F_s)) & \longrightarrow & H^2(F,\mu_n\otimes\mu_n) \end{array}$$

which comes (eventually) from the commutative diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & \mu_n\otimes\mu_n & \longrightarrow & \mathrm{Tor}_1(F_s^*,F_s^*) \\ & & \downarrow\beta & & \downarrow\alpha \\ 0 & \longrightarrow & \mu_n\otimes\mu_n & \longrightarrow & K_3(F_s)_{\mathrm{ind}}, \end{array}$$

where α is defined via diagram (**) of § 2, and $\mu_n\otimes\mu_n$ is identified with the n-torsion subgroup of $\mathrm{Tor}_1(F_s^*,F_s^*)$ and of $K_3(F_s)_{\mathrm{ind}}$ by the two horizontal maps. Since we will see in § 7 that α is injective, it follows that β is an isomorphism.

Let $\tilde{K}_2(F)=H^2(F,\Gamma(2,F_s))$. It follows from Proposition 3.3 and Lemma 3.5 that we have the commutative diagram:

$$\begin{array}{ccc} K_2(F)/nK_2(F) & \xrightarrow{\theta_F^{(n)}} & \tilde{K}_2(F)/n\tilde{K}_2(F) \\ \downarrow{\scriptstyle h_F} & & \downarrow{\scriptstyle \tilde{h}_F} \\ H^2(F,\mu_n\otimes\mu_n) & \xrightarrow{\tilde{\beta}} & H^2(F,\mu_n\otimes\mu_n) \end{array}$$

where $\tilde{\beta}$ is an isomorphism, and h_F is, up to sign, the map defined by Tate in [T]. The Mercuriev-Suslin theorem [M-S] asserts that h_F is an isomorphism, hence $\theta_F^{(n)}$ is injective and $\tilde{h}_F$ is surjective.

We now prove Axiom 2 (up to p-torsion).

Proposition 4.3. *$H^3(F,\Gamma(2,F_s))$ is a p-torsion group. In particular, if F has characteristic zero, then $H^3(F,\Gamma(2,F_s))=0$.*

Proof. It follows immediately from the hypercohomology spectral sequence that $H^3(F,C)$ is torsion for any complex C such that $C_n=0$ for $n\geqq 3$. It follows from the Kummer sequence (Axiom 3) that we have the exact sequence

$$0\to\tilde{K}_2(F)/n\tilde{K}_2(F)\xrightarrow{\tilde{h}_F}H^2(F,\mu_n\otimes\mu_n)\to H^3(F,\Gamma(2,F_s)_n\to 0.$$

Since we have just seen that $\tilde{h}_F$ is surjective, we obtain $H^3(F,\Gamma(2,F_s))_n=0$.

Corollary 4.4. *$\theta_F^{(n)}$ is an isomorphism between $K_2(F)/nK_2(F)$ and $\tilde{K}_2(F)/n\tilde{K}_2(F)$.*

Theorem 4.5. (Axiom 6). *θ_F is injective, with p-torsion cokernel.*

Proof. We first observe that θ_F is clearly an isomorphism if F is separably closed. By considering the commutative diagram

$$\begin{array}{ccc} K_2(F) & \xrightarrow{\theta_F} & \tilde{K}_2(F) \\ \downarrow{\scriptstyle \gamma} & & \downarrow \\ K_2(F_s) & \xrightarrow{\theta_{F_s}} & \tilde{K}_2(F_s), \end{array}$$

we see, since $\mathrm{Ker}(\gamma)$ is torsion, that $\mathrm{Ker}(\theta_F)$ is torsion. Let $A=C_{1,2}(F_s)$ and $B=C_{2,2}(F_s)$, so $\Gamma(2,F_s)$ is the complex $A\to B$. Now θ_F factors into $\theta_F^{(1)}\colon K_2(F)\to\mathrm{Coker}(A^G\to B^G)$ and $\theta_F^{(2)}\colon\mathrm{Coker}(A^G\to B^G)\to\tilde{K}_2(F)$. It follows from the standard properties of the transfer on K_2 that $K_2(F)\to K_2(F_s)^G$ has torsion cokernel, and it is easy to see that $\mathrm{Coker}(A^G\to B^G)$ maps to $K_2(F_s)^G$ with torsion kernel and cokernel, so $\theta_F^{(1)}$ has torsion cokernel. The basic exact sequence

$$H^0(F,A)\xrightarrow{\varphi_{F_s}^G}H^0(F,B)\to H^2(F,\Gamma(2,F_s))\to H^1(F,A)$$

shows that $\mathrm{Coker}\,\theta_F^{(2)}$ is contained in the torsion group $H^1(F,A)$ and so is torsion. Hence $\mathrm{coker}\,\theta_F$ is torsion.

We next claim that if ℓ is a prime different from p, then the map from $K_2(F)_\ell$ to $\tilde{K}_2(F)_\ell$ is surjective. We first treat the case when F contains an ℓ-th root of unity. In that case we have a natural map α from $H^1(F,\mu_\ell\otimes\mu_\ell)$

$= F^*/F^{*\ell} \otimes \mu_\ell$ to $K_2(F)_\ell$ given by $a \otimes \zeta \to (a, \zeta)$. We may then consider the diagram

$$\begin{array}{ccccc} H^1(F, \mu_\ell \otimes \mu_\ell) & \xrightarrow{\alpha} & K_2(F)_\ell & & \\ \downarrow{\scriptstyle \tilde{\beta}} & & \downarrow{\scriptstyle \theta_{F,\ell}} & & \\ H^1(F, \mu_\ell \otimes \mu_\ell) & \xrightarrow{\gamma} & \tilde{K}_2(F)_\ell & \longrightarrow & 0 \end{array}$$

where α is as above, γ is induced from the Kummer sequence for $\Gamma(2, F_s)$ and $\tilde{\beta}$ is induced by the isomorphism β, and so is an isomorphism. In order to show $\theta_{F,\ell}$ is surjective, it suffices to show that the diagram commutes.

We first compute $\theta_{F,\ell} \circ \alpha$. We start with $a \otimes \zeta$ in $F^*/F^{*\ell} \otimes \mu_\ell = H^1(F, \mu_\ell \otimes \mu_\ell)$. This maps to (a, ζ) in $K_2(F)_\ell$, which equals $-(\zeta, a)$, and so is represented in $\mathrm{Coker}(\varphi_{F_s})^G$ by $-\zeta_{\frac{1}{1-a}}$.

We next compute $\gamma \circ \tilde{\beta}$. We have the commutative diagram:

$$\begin{array}{ccccccc} H^0(F, A_\ell) & \longrightarrow & H^0(F, B_\ell) & \xrightarrow{\varepsilon} & H^1(F, \mu_\ell \otimes \mu_\ell) & \longrightarrow & H^1(F, A_\ell) \\ \downarrow & & \downarrow & & \downarrow & & \\ H^0(F, A) & \longrightarrow & H^0(F, B) & \longrightarrow & H^2(F, \Gamma(2, F_s)) = \tilde{K}_2(F) & & \end{array}$$

with exact rows. The coboundary map ε may be computed using our explicit realization of $G_m \overset{L}{\otimes} G_m$ and the evident commutative diagram, and we find that $\varepsilon(\zeta_{\frac{1}{1-a}}) = -(a \otimes \zeta)$. In particular, ε is surjective. We then see immediately that a representative for $(\gamma \circ \tilde{\beta})(a \otimes \zeta)$ in $\mathrm{Coker}(\varphi_{F_s})^G$ is also given by $-\zeta_{\frac{1}{1-a}}$.

Next let F be arbitrary.

We know from the existence of a norm map on K_2 and the fact that $[F(\mu_\ell): F]$ is prime to ℓ that we may identify $K_2(F)_\ell$ with $K_2(F(\mu_\ell))_\ell^{\Delta}$, where $\Delta = G(F(\mu_\ell)/F)$. Similarly, the existence of a norm map on hypercohomology permits us to identify $\tilde{K}_2(F)_\ell$ with $\tilde{K}_2(F(\mu_\ell))_\ell^{\Delta}$. Since the order of Δ is prime to ℓ, starting with $0 \to K \to K_2(F(\mu_\ell))_\ell \to \tilde{K}_2(F(\mu_\ell))_\ell \to 0$, we obtain

$$K_2(F(\mu_\ell))_\ell^{\Delta} \to \tilde{K}_2(F(\mu_\ell))_\ell^{\Delta} \to H^1(\Delta, K) = 0,$$

which gives us the surjectivity in the general case.

We are now in a position to axiomatize our situation.

Lemma 4.6. *Let θ: $A \to B$ be a map of abelian groups such that*

1) *θ has torsion cokernel.*

2) *θ induces an isomorphism $A/\ell A \to B/\ell B$ for all primes ℓ except possibly for a distinguished prime p.*

3) *θ induces a surjection $A_\ell \to B_\ell$ for all $\ell \neq p$. Then the cokernel of θ is a p-torsion group, and the kernel of θ is ℓ-divisible for $\ell \neq p$.*

Proof. Let $K = \mathrm{Ker}\,\theta$, $C = \mathrm{Coker}\,\theta$, $Q = \mathrm{Im}\,\theta$, so we have the exact sequence $0 \to K \to A \to Q \to 0$, $0 \to Q \to B \to C \to 0$, giving rise to:

$$0 \to K_\ell \to A_\ell \to Q_\ell \to K/\ell K \to A/\ell A \to Q/\ell Q \to 0$$
$$0 \to Q_\ell \to B_\ell \to C_\ell \to Q/\ell Q \to B/\ell B \to C/\ell C \to 0.$$

First 2) $\Rightarrow C/\ell C = 0$, $0 \to K_\ell \to A_\ell \to Q_\ell \to K/\ell K \to 0$, and $0 \to Q_\ell \to B_\ell \to C_\ell \to 0$.

Next 3) $\Rightarrow Q_\ell \xrightarrow{\sim} B_\ell$, $C_\ell = 0$, and $K/\ell K = 0$.

Finally 1) $\Rightarrow C$ is p-torsion.

We now need another lemma about K_2:

Lemma 4.7. *Let F be a field of characteristic $p \geqq 0$ and finitely-generated over the prime field. Let ℓ be any prime. Then $K_2(F)(\ell)$ contains no non-zero ℓ-divisible subgroups.*

Proof. First let $\ell \neq p$. Suslin shows [Su2, Thms. 3.9, 3.10 and Corollary 2.8] that $K_2(F)(\ell) \simeq H^2(F, Z_\ell(2))(\ell)$, and Tate shows [T, p. 240] that $H^2(F, Z_\ell(2))$ contains no non-zero ℓ-divisible subgroups. If $\ell = p$, $K_2(F)(p) = 0$ by Suslin [Su2, Thm. 1.10], so we are done.

Now assume F is finitely-generated and look at the map θ_F: $K_2(F) \to \tilde{K}_2(F)$. By Lemma 4.4 $\operatorname{Ker}\theta_F$ is ℓ-divisible for $\ell \neq p$, and $\operatorname{Coker}\theta_F$ is p-torsion. Then $\operatorname{Ker}\theta_F(\ell)$ is ℓ-divisible, so zero by Lemma 4.7. Hence $\operatorname{Ker}\theta_F$ is p-torsion. But, by the result of Suslin quoted above, $K_2(F)(p) = 0$, so θ_F is injective.

For F arbitrary, we write F as the limit of finitely-generated fields, and use the fact that $K_2(F)$ and $\tilde{K}_2(F)$ commute with direct limits to conclude that θ_F is injective with p-torsion cokernel.

§ 5. A new version of the Bloch-Wigner dilogarithm function (see [Bl1, p. 59ff])

We begin with the value group of this function. Let $\overset{0}{\coprod_{x \in \mathbb{C}}} \mathbb{C}^*$ be the kernel of the augmentation map $\coprod_{x \in \mathbb{C}} \mathbb{C}^* \to \mathbb{C}^*$, and let $S \subset \overset{0}{\coprod_{x \in \mathbb{C}}} \mathbb{C}^*$ be the subgroup generated by all elements of the form $\{\langle a \rangle_{x+1} - \langle a \rangle_x - \langle a \rangle_1 + \langle a \rangle_0\}$ for $a \in \mathbb{C}^*$, $x \in \mathbb{C}$. Then it is immediate that we have the exact sequence

$$0 \to \mathbb{C}^* \xrightarrow{b_1} \overset{0}{\coprod_{x \in \mathbb{C}}} \mathbb{C}^*/S \xrightarrow{b_2} \coprod_{x \in \mathbb{C}^* - \{1\}} \mathbb{C}^* \to 0 \tag{5.1}$$

where $b_1(a) =$ the class of $\langle a \rangle_1$ - the class of $\langle a \rangle_0$ and

$$b_2(\langle a \rangle)_x = \langle a \rangle_{\exp(2\pi i x)} \quad \text{if } x \notin Z$$
$$= 0 \quad \text{if } x \in Z.$$

Now we define a dilogarithm function $\tilde{\varepsilon}$: $\mathbb{C} - \{0, 1\} \to \coprod_{x \in \mathbb{C}} \mathbb{C}^*/S$ by the following formula:

$$\tilde{\varepsilon}(a) = \langle 1-a \rangle_{\frac{1}{2\pi i}\log a} + \left\langle \exp -\frac{1}{2\pi i} \int_1^{1-a} \log(1-t)\frac{dt}{t} \right\rangle_1 \tag{5.2}$$
$$- \left\langle \exp -\frac{1}{2\pi i} \int_1^{1-a} \log(1-t)\frac{dt}{t} \right\rangle_0 + \langle 1-a \rangle_0.$$

This definition needs a few words of explanation. In order to define $\log a$ we must choose a path γ from 1 to a. This then determines a path $1-\gamma$ from 0 to $1-a$. The integral $\int_1^{1-a} \log(1-t)\frac{dt}{t}$ is evaluated by choosing a real number ε, evaluating $\int_{1-\varepsilon}^{1-a} \log\frac{(1-t)dt}{t}$ along the path consisting of $[1-\varepsilon, 0]$ followed by $1-\gamma$, and letting ε go to zero. On the first part of this path we take the principal value of $\log(1-t)$ and on the second part the value determined by the path γ.

Lemma 5.3. *The expression $\tilde{\varepsilon}(a)$ is well-defined in $\coprod_{x\in\mathbb{C}} \mathbb{C}^*/S$, independent of the choice of γ.*

Proof. (After Bloch, [Bl1].) If we change γ by inserting a small clockwise circle C around 0, thus obtaining a new path γ', then

$$\frac{1}{2\pi i}\log a \text{ (evaluated via } \gamma') - \frac{1}{2\pi i}\log a \text{ (evaluated via } \gamma) = 1.$$

Also,

$$\int_{1-\gamma'} \log_{\gamma'}(1-t)\frac{dt}{t} = \int_{1-\gamma} \log_{\gamma}(1-t)\frac{dt}{t} + \int_{1\,(\text{via } 1-\gamma)}^{1-a} 2\pi i\frac{dt}{t}$$

plus terms which go to zero as C get small. (We may assume without loss of generality that γ passes as near zero as we like, so $1-\gamma$ gets close to 1. The second integral is then taken along the part of $1-\gamma$ from 1 to $1-a$.) Hence if we let $x=\frac{1}{2\pi i}\log a$ (evaluated via γ), the change in $\tilde{\varepsilon}(a)$ is equal to $\langle 1-a\rangle_{x+1} - \langle 1-a\rangle_x - \langle 1-a\rangle_1 + \langle 1-a\rangle_0$ since $\exp\left(-\frac{1}{2\pi i}\int_\gamma 2\pi i\frac{dt}{t}\right) = (1-a)^{-1}$. Since the change in $\tilde{\varepsilon}(a)$ lies in S, $\tilde{\varepsilon}$ is well-defined. We must also consider the effect of changing γ by inserting a small circle around 1, but, again as in Bloch [Bl1], one sees that each of the individual terms remains unchanged.

Remark 6.4. It is immediate that $b_2(\varepsilon(a)) = \langle 1-a\rangle_a$. For later use we make the conventions that $\tilde{\varepsilon}(1) = \left\langle \exp\left(\frac{-\pi i}{12}\right)\right\rangle_1$ and that $\langle 0\rangle_a = 0$. Also in $\coprod_{x\in\mathbb{C}^*-\{1\}} \mathbb{C}^*$, $\langle a\rangle_1 = 0$. (Note

$$\exp -\frac{1}{2\pi i}\int_1^0 \log(1-t)\frac{dt}{t} = \exp\left(\frac{1}{2\pi i}\sum_{n=1}^{\infty}\frac{1}{n^2}\right) = \exp\left(\frac{-\pi i}{12}\right).)$$

Let $(1+I)$ denote the set of functions f in $\mathbb{C}(t)^*$ such that $f(0)=f(\infty)=1$. Explicitly, such a function f is represented by $f(t)=\Pi(t-\lambda_i)^{d_i}$, where $d_i=\pm 1$, $\Pi\lambda_i^{d_i}=1$ and $\Sigma d_i=0$. (The λ_i's are not necessarily distinct.) We will now use the dilogarithm to define a map β from $(1+I)\otimes_{\mathbb{Z}} \mathbb{C}(t)^*$ to $\coprod_{x\in\mathbb{C}}{}^0\, \mathbb{C}^*/S$ such that the following diagram commutes:

$$(5.4)\qquad \begin{array}{ccc} (1+T)\underset{Z}{\otimes}\mathbb{C}(t)^* & \xrightarrow{T} & \coprod_{x\in\mathbb{C}^*-\{1\}}\mathbb{C}^* \\ \downarrow{\scriptstyle\beta} & & \downarrow \\ \coprod_{x\in\mathbb{C}}\mathbb{C}^*/S & \xrightarrow{b_2} & \coprod_{x\in\mathbb{C}^*-\{1\}}\mathbb{C}^* \end{array}$$

where T is given by the tame symbols

$$T=\{T_x\},\qquad T_x(f\otimes g)=(-1)^{v_x(f)v_x(g)}\frac{g^{v_x(f)}}{f^{v_x(g)}}(x)\in\mathbb{C}^*.$$

We define β by describing it on elements of the forms

1) $\Pi(t-\lambda_i)^{d_i}\otimes c$ $(c\in\mathbb{C}^*)$
2) $\Pi(t-\lambda_i)^{d_i}\otimes t$
3) $\Pi(t-\lambda_i)^{d_i}\otimes(t-\mu)$, $\mu\neq 0$.

We begin with 1). We choose paths γ_i from 1 to λ_i such that $\Pi\lambda_i^{d_i}=1$. We then define

$$\beta(\Pi(t-\lambda_i)^{d_i}\otimes c)=\Sigma\langle c^{d_i}\rangle_{\frac{1}{2\pi i}\log\lambda_i},$$

using the paths γ_i to define $\log\lambda_i$. It is easy to verify that changing the paths γ_i changes $\Sigma\langle c^{d_i}\rangle_{\frac{1}{2\pi i}\log\lambda_i}$ by an element of S. It is also immediate that

$$b_2(\beta(\Pi(t-\lambda_i)^{d_i}\otimes c))=\sum_{\lambda_i\neq 1}\langle c^{d_i}\rangle_{\lambda_i}=T(\Pi(t-\lambda_i)^{d_i}\otimes c).$$

We now define $\beta(\Pi(t-\lambda_i)^{d_i}\otimes t)$ to be

$$\Sigma d_i(\tilde{\varepsilon}(\lambda_i)+\tilde{\varepsilon}(\lambda_i^{-1})-\langle 1-\lambda_i^{-1}\rangle_{\frac{1}{2\pi i}\log\lambda_i}-\langle 1-\lambda_i^{-1}\rangle_{\frac{1}{2\pi i}\log\lambda_i^{-1}}$$
$$+\langle -1\rangle_{\frac{1}{2\pi i}\log\lambda_i}+2\langle 1-\lambda_i^{-1}\rangle_0-\langle -1\rangle_0),$$

where we use the paths γ_i to define $\log\lambda_i$ and γ_i^{-1} to define $\log\lambda_i^{-1}$. It is again easy to see that changing the paths γ_i changes this expression by an element of S, and

$$b_2\beta\Pi(t-\lambda_i)^{d_i}\otimes t=\sum_{\lambda_i\neq 1}d_i\langle\lambda_i\rangle_{\lambda_i}=T(\Pi(t-\lambda_i)^{d_i}\otimes t).$$

Finally, we define $\beta(\Pi(t-\lambda_i)^{d_i}\otimes(t-\mu))$ if $\mu\neq 0$, to be

$$-\Sigma d_i(\tilde{\varepsilon}(\mu/\lambda_i)-\tilde{\varepsilon}(\lambda_i)-\tilde{\varepsilon}(\lambda_i^{-1})-\langle 1-\mu/\lambda_i\rangle_{\frac{1}{2\pi i}\log\lambda_i}$$
$$-\langle 1-\mu/\lambda_i\rangle_{\frac{1}{2\pi i}\log\mu/\lambda_i}+\langle 1-\mu/\lambda_i\rangle_{\frac{1}{2\pi i}}+\langle 1-\mu_i/\lambda_i\rangle_0$$
$$+\langle 1-\lambda_i^{-1}\rangle_{\frac{1}{2\pi i}\log\lambda_i}-2\langle 1-\lambda_i^{-1}\rangle_0+\langle 1-\lambda_i^{-1}\rangle_{\frac{1}{2\pi i}\log\lambda_i^{-1}}$$
$$-\langle -1\rangle_{\frac{1}{2\pi i}\log\lambda_i}+\langle -1\rangle_0)$$

where, in addition to the previous path conventions, the path defining $\log \mu$ should be the product of the paths defining $\log \lambda_i$ and $\log \mu/\lambda_i$. Again the definition of β is seen to be independent of the choice of paths subject to the given conditions, and

$$\begin{aligned} b_2\beta(\Pi(t-\lambda_i)^{d_i}\otimes(t-\mu)) &= \Sigma d_i\langle\lambda_i-\mu\rangle_{\lambda_i}-\langle 1-\mu/\lambda_i\rangle_{\mu}+\langle\lambda_i\rangle_1) \\ &= (\text{since } \Sigma d_i=0 \text{ and } \Pi\lambda_i^{d_i}=1)\Sigma d_i\langle\lambda_i-\mu\rangle_{\lambda_i}-\langle\mu-\lambda_i\rangle_{\mu_i}) \\ &= T(\Pi(t-\lambda_i)^{d_i}\otimes(t-\mu)). \end{aligned}$$

The careful reader will verify that these formulas hold even when $\mu=\lambda_i$ for some i or when some $\lambda_i=1$, keeping in mind our conventions.

§ 6. The rigidity of the dilogarithm

In this section we shall show that Bloch's rigidity result proven in [Bl1] for the Bloch-Wigner dilogarithm remains valid in the situation of § 5. We begin with our notation.

Let K be the field of meromorphic functions defined in a neighborhood of $(0, 1)$. Let $\tilde{I}\subset K(t)$ be the set of all functions $g(t)$ such that $g(0)=g(\infty)=0$ in K. Let $A\subset K$ be the set of functions f holomorphic on $(0, 1)$ and such that

1) either $\lim_{z\to 0} f(z)$ exists or $\lim_{z\to 0}\exp 2\pi i f(z)=0$
2) either $\lim_{z\to 1} f(z)$ exists or $\lim_{z\to 1}\exp 2\pi i f(z)=0$.

Let A^* be the set of functions holomorphic and non-zero on $(0, 1)$ whose limits exist at the endpoints. Note that neither A nor A^* are groups, but the group of constant integer-valued functions Z acts on A by addition.

Let ψ be the map from A to A^* which sends $f(z)$ to $\exp 2\pi i f(z)$. Then ψ induces an identification of A^* with $Z\backslash A$. (A branch of the logarithm is fixed once it is defined at any point on the interval.) This gives rise to the exact sequence

$$0\to K^*\to \coprod_{y\in A}{}^{0} K^*/\mathscr{S} \xrightarrow{b_2} \coprod_{\omega\in A}{}^{0} K^*\to 0. \tag{6.1}$$

Here $\mathscr{S}$, analogously to the set S of § 1, is defined to be the subgroup generated by $\{\langle f\rangle_{y+1}-\langle f\rangle_y-\langle f\rangle_1+\langle f\rangle_0\}$ for $f\in K^*$, $y\in A$, and the map b_2 sends $\langle f\rangle_y$ to $\langle f\rangle_{\exp 2\pi i y}$.

Let C_2 be the set of functions $\delta\prod_j (t-\lambda_j(z))^{d_j}$, where the product is finite, the λ_j's are in A^*, $\delta\in A^*$, $d_j\in Z$. Let $C_1\subset C_2$ be the set of such functions with $\delta\equiv 1$, $\Sigma d_j=0$ and $\Pi\lambda_j(z)\equiv 1$.

We define a "dilogarithm map" β from $C_1\underset{Z}{\otimes} C_2$ *to* $\coprod_{y\in A} K^*/\mathscr{S}$ in an exactly analogous fashion to the definition given in § 1. For example, we begin by defining $\tilde{\varepsilon}(f)$ for $f\in A^*$ to be

$$\langle 1-f\rangle_{\frac{1}{2\pi i}\log f}+\left\langle \exp-\frac{1}{2\pi i}\int_1^{1-f(z)}\log(1-t)\frac{dt}{d}\right\rangle_1$$

$$-\left\langle \exp-\frac{1}{2\pi i}\int_1^{1-f(z)}\log(1-t)\frac{dt}{d}\right\rangle_0-\langle 1-f\rangle_0,$$

with the same conventions governing the choice of branches as in § 5. This map β has the following properties:

1) There exists a commutative diagram with exact rows:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & B & \longrightarrow & C_1\underset{Z}{\otimes} C_2 & \xrightarrow{T} & \coprod_{\omega\in A^*}^{0} K^* & & \\ & & \downarrow{\scriptstyle\varphi} & & \downarrow{\scriptstyle\beta} & & \downarrow{\scriptstyle(\mathrm{id})} & & \\ 0 & \longrightarrow & K^* & \longrightarrow & \coprod_{y\in A}^{0} K^*/\mathscr{S} & \xrightarrow{b_2} & \coprod_{\omega\in A^*}^{0} K^* & \longrightarrow & 0 \end{array} \tag{6.2}$$

where T is the obvious "tame symbol" map, B is the kernel of T, and φ is induced by β. The proof of commutativity of the right-hand square is directly modeled on the corresponding proof in § 5, and will be omitted.

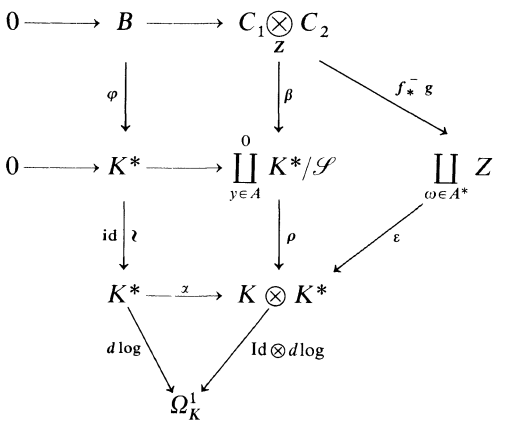
2) Let α be any point in $[0,1]$. Evaluation at α induces a map from $\coprod_{y\in A}^{0} K^*/\mathscr{S}$ to $\coprod_{x\in\mathbb{C}}^{0} K^*/\bar{\mathscr{S}}$, where $\bar{\mathscr{S}}$ denotes the image of $\mathscr{S}$ in $\coprod_{x\in\mathbb{C}}^{0} K^*$, i.e., $\langle f\rangle_y\to\langle f\rangle_{y(\alpha)}$. If α is an endpoint, and y is not defined at α, we map $\langle f\rangle_y$ to $\langle f\rangle_0$. We consider now the subgroup $(\coprod_{x\in\mathbb{C}}^{0} K^*/\bar{\mathscr{S}})_\alpha$ generated by $\langle f\rangle_x$ such that $f(\alpha)$ is defined and non-zero. Then the diagram

$$\begin{array}{ccccc} C_1\underset{Z}{\otimes} C_2 & & \longrightarrow & & (1+T)\underset{Z}{\otimes}\mathbb{C}(t)^* \\ \downarrow{\scriptstyle\beta} & & & & \downarrow{\scriptstyle\beta} \\ \coprod_{y\in A}^{0} K^*/\mathscr{S} & \longrightarrow & \coprod_{x\in\mathbb{C}}^{0} K^*/\bar{\mathscr{S}} & \longrightarrow & \coprod_{x\in\mathbb{C}}^{0}\mathbb{C}^*/S \end{array} \tag{6.3}$$

commutes, in the sense that the image of $C_1\underset{Z}{\otimes} C_2$ in $\coprod_{x\in\mathbb{C}}^{0} K^*/\bar{\mathscr{S}}$ lies in $\left(\coprod_{x\in\mathbb{C}}^{0} K^*/\bar{\mathscr{S}}\right)_\alpha$ which then goes by evaluation at α to $\coprod_{x\in\mathbb{C}}^{0}\mathbb{C}^*/S$. The verification is straightforward. However if we are considering a generator $\Pi(t-\lambda_i(z))^{d_i}\otimes(t-\mu(z))$ of $C_1\otimes C_2$, special attention has to be paid to the α's where either $\lambda_i(\alpha)=1$, $\mu_i((\alpha)=0$, or $\lambda_i(\alpha)=\mu_i(\alpha)$.

Now we repeat another argument of Bloch's. Let Ω^1_K denote the K-module of meromorphic 1-forms on (0, 1), and consider the commutative diagram:

(6.4)



Here $\rho(\langle f \rangle_y) = y \otimes f$ is well-defined, and $\alpha(f) = 1 \otimes f$. As in Bloch [Bl1, p. 63]. $\mathrm{Id} \otimes d\log)(g \otimes f) \overset{\text{def}}{=} \frac{g\, df}{f}$. Also $\varepsilon(\langle 1 \rangle_\omega \overset{\text{def}}{=} \rho(\tilde{\varepsilon}(\omega))$, and the map from $C_1 \underset{Z}{\otimes} C_2$ is as defined by Bloch, namely

$$(f_*^- g)(\Pi(t-\lambda_i(z))^{d_i} \otimes \delta(z)) = 0$$
$$(f_*^- g)(\Pi(t-\lambda_i(z))^{d_i}) \otimes t) = \Sigma \langle d_i \rangle_{\lambda_i} + \langle d_i \rangle_{\lambda_i^{-1}}$$
$$(f_*^- g)(\Pi(t-\lambda_i(z))^{d_i} \otimes (t-\mu))(\mu \not\equiv 0)$$
$$= \Sigma \langle d_i \rangle_{\lambda_i} + \Sigma \langle d_i \rangle_{\lambda_i^{-1}} - \Sigma \langle d_i \rangle_{\mu\lambda_i^{-1}}.$$

The proof is routine. For the commutativity of the triangle, it should be observed that with our conventions in picking branches of the logarithm, we have $\Sigma d_i \log \lambda_i \equiv 0$, $\log \lambda_i + \log \mu/\lambda_i \equiv \log \mu$, and $\log \lambda_i + \log \lambda_i^{-1} \equiv 0$.

We now have, as in [Bl1],

$$(\mathrm{Id} \otimes d\log)(\varepsilon(\langle 1 \rangle_\omega)$$
$$= (\mathrm{Id} \otimes d\log)) \left(\frac{1}{2\pi i} \log \omega \otimes (1-\omega) + 1 \otimes \exp - \frac{1}{2\pi i} \int_1^{1-\omega(z)} \log(1-t) \frac{dt}{t} \right)$$
$$= \frac{1}{2\pi i} \log \omega(z) \frac{-\omega'(z)}{1-\omega(z)} - \frac{1}{2\pi i} \frac{\log(\omega(z))(-\omega'(z))}{1-\omega(z)} = 0.$$

Hence $(\mathrm{Id} \otimes d\log) \circ \rho \circ \beta = 0$, hence $d\log(\varphi(b)) = 0$ for any $b \in B$, hence $\varphi(b) \in \mathbb{C}^* \subset K^*$ for any $b \in B$.

Now let f_z be given by $\prod_{i \in I} (t - \lambda_i(z))^{d_i}$, where the $\lambda_i(z)$ are holomorphic in a neighborhood of $(0, 1)$ and non-zero. I is finite, $\Pi \lambda_i(z)^{d_i} \equiv 1$, and $\Sigma d_i = 0$. Assume that $1 - f_z$ does not have any multiple roots except at 0 and 1. Then, restricting to a smaller neighborhood if necessary, we may write $1 - f_z = \delta(z) \prod_{j \in J} (t - \mu_j(z))^{e_j}$, where the μ_j are holomorphic. (Roots are analytic functions of the coefficients.)

Lemma 6.5. *Let f_z be as above. Then $f_z \otimes 1 - f_z \in B$, hence $j(f_z \otimes 1 - f_z)$ is constant.*

Proof. By the standard tame symbol argument, $T(f_z \otimes 1 - f_z) = 0$, and $B = \mathrm{Ker}(T)$.

Proposition 6.6. *If $f \in 1+I$, $\beta(f \otimes 1 - f) = 0$, so β induces a function from $(1+I) \otimes \mathbb{C}(t)^*/\mathscr{R}$ to $\coprod_{x \in \mathbb{C}} \mathbb{C}^*/S$, where $\mathscr{R}$ is generated by $\{(f \otimes 1 - f)\}$.*

Proof. Given $f, g \in I$, one can find an analytic family f_z with $f_0 = f$, $f_1 = g$ and f_z as above. (One can easily find such an analytic family in an open neighborhood U of $(0, 1)$ with only finitely many z's such that $1 - f_z$ has multiple roots. So there is an analytic curve C joining 0 and 1 where $1 - f_z$ has no multiple roots. By choosing an analytic isomorphism of C with $(0, 1)$ we obtain what we want.) Since we know by Lemma 6.5 that $\beta(f_z \otimes 1 - f_z)$ is constant, it remains to show that this constant is zero. For this it suffices to choose any $f \in I$, and we pick $f = 1 - \dfrac{(t-\lambda_1)(t-\lambda_2)}{(t-\lambda_1\lambda_2)(t-1)} = \dfrac{t(\lambda_1+\lambda_2-1-\lambda_1\lambda_2)}{(t-\lambda_1\lambda_2)(t-1)}$. If we apply our definition of β to f, we see immediately that all the terms formally cancel.

§ 7. $\mathrm{Tor}_1(F^*, F^*) \to K_3(F)_{\mathrm{ind}}$ is injective

Using the results on the dilogarithm of the preceding two sections, we will proceed to show:

Theorem 7.1. *Let F be a field. Then the natural map ρ_F from $\mathrm{Tor}_1(F^*, F^*) \to K_3(F)_{\mathrm{ind}}$ defined in Remark 2.6 is injective.*

Proof. We first will prove the theorem for $F = \mathbb{C}$, then for any field of characteristic zero, and then in general. We first recall from Remark 2.6 that, for any field F, we have the exact sequence

$$0 \to K_3(F)_{\mathrm{ind}} \to C_{2,1}(F) \xrightarrow{\varphi_F} C_{2,2}(F) \to K_2(F) \to 0,$$

and from Proposition 1.8 that $C_{2,2}(F)$ is isomorphic to $\coprod_{x \in F - \{0,1\}} F_x^*$. With the notation as in § 1, $C_{2,1}(F)$ is $\varinjlim K_2(X - Y_B, Z)$. Let F now be algebraically closed. Then this in turn is immediately seen to be $K_2(S, J)$, where S is the semi-local ring contained in $F(t)$ consisting of all rational functions defined at 0 and 1, and J is the Jacobson radical of S. The exact sequence

$$0 \to K_3(F)_{\mathrm{ind}} \to K_2(S, J) \xrightarrow{\varphi_F} \coprod_{x \in F - \{0,1\}} F^* \to K_2(F) \to 0$$

may be identified with the analogous sequence in [Bl1], showing that the map φ_F is given by the tame symbol.

Let $R \subset F(t)$ be the semi-local ring consisting of all rational functions vanishing at 0 and ∞, and I its Jacobson radical. The map $f \to f\left(\dfrac{1}{1-x}\right)$ is

clearly an isomorphism of the pair (S, J) with the pair (R, I), and hence induces an isomorphism $\mu: K_2(S, J) \to K_2(R, I)$. In [Bl1], Bloch describes a natural map from $K_2(R, I)$ to $(1+I) \underset{Z}{\otimes} \mathbb{C}(t)^*/\mathscr{R}$, where $\mathscr{R}$ is the subgroup generated by elements of the form $(f \otimes (1-f))$, $f \in 1+I$, as in Proposition 6.6.

We now go back and rewrite the commutative diagram (**) of Remark 2.6 in the case when $A = \mathbb{C}$:

$$\begin{array}{ccccccccc}
0 \to \mathrm{Tor}_1(\mathbb{C}^*, \mathbb{C}^*) & \longrightarrow & (1+J) \underset{Z}{\otimes} \mathbb{C}^* & \longrightarrow & \coprod_{x \in \mathbb{C} - \{0,1\}} \mathbb{C}^* & \longrightarrow & \mathbb{C}^* \otimes \mathbb{C}^* \to 0 \\
\downarrow \rho_{\mathbb{C}} & & \downarrow & & \downarrow & & \downarrow \\
0 \to K_3(\mathbb{C})_{\mathrm{ind}} & \longrightarrow & K_2(S, J) & \xrightarrow{\varphi_{\mathbb{C}}} & \coprod_{x \in \mathbb{C} - \{0,1\}} \mathbb{C}^* & \longrightarrow & K_2(\mathbb{C}) \to 0.
\end{array}$$

We recall from Lemma 4.1 that the above map from $\mathrm{Tor}_1(\mathbb{C}^*, \mathbb{C}^*)$ to $(1+J) \otimes \mathbb{C}^*$, when restricted to $\mathrm{Tor}_1(\mathbb{C}^*, \mathbb{C}^*)_n$, has as image the set of elements $(1-t+\zeta t)^n \otimes \omega$ as ζ, ω run through the group of n-th roots of unity. So to show that $\rho_{\mathbb{C}}$ is injective it suffices, after making the obvious identifications, to show that the composite map

$$(1+I) \otimes \mathbb{C}^* \to K_2(R, I) \to ((1+I) \otimes \mathbb{C}(t)^*)/\mathscr{R} \xrightarrow{\beta} \coprod_{x \in \mathbb{C}} \mathbb{C}^*/S$$

takes $\eta = (t-\zeta)^n/(t-1)^n \otimes \omega$ to an element of exact order n if ζ and ω are both primitive n-th roots of unity, since the element $(1-t+\zeta t^n) \otimes \omega$ in $(1+J) \otimes \mathbb{C}$ corresponds to the element $\dfrac{(t-\zeta)^n}{(t-1)^1} \otimes \omega$ in $(1+I) \otimes \mathbb{C}^*$.

We interrupt the proof of Theorem 7.1 to state the following

Corollary 7.2. *The map β induces a "dilogarithm map", (also called β), from $K_3(\mathbb{C})_{\mathrm{ind}}$ to $\mathbb{C}^*$ such that β is injective when restricted to $\mathrm{Tor}_1(\mathbb{C}^*, \mathbb{C}^*)$, viewed as a subgroup of $K_3(\mathbb{C})_{\mathrm{ind}}$ by means of $\rho_{\mathbb{C}}$.*

Proof. This follows immediately by considering the following commutative diagram, where the top and bottom rows are exact:

$$\begin{array}{ccccccc}
0 \longrightarrow K_3(\mathbb{C})_{\mathrm{ind}} & \longrightarrow & K_2(R, I) & \longrightarrow & \coprod_{x \in \mathbb{C} - \{0,1\}} \mathbb{C}^* & & \\
 & & \downarrow & & \downarrow \lambda\,(\mathrm{id}) & & \\
 & & ((1+I) \otimes \mathbb{C}(t)^*)/\mathscr{R} & \longrightarrow & \coprod_{x \in \mathbb{C} - \{0,1\}} \mathbb{C}^* & & \\
 & & \downarrow \beta & & \downarrow \lambda\,(\mathrm{id}) & & \\
0 \longrightarrow \mathbb{C}^* & \xrightarrow{b_1} & \coprod_{x \in \mathbb{C}}^{0} \mathbb{C}^*/S & \xrightarrow{b_2} & \coprod_{x \in \mathbb{C}^* - \{1\}} \mathbb{C}^* & &
\end{array}$$

Remark 7.3. This map presumably coincides (when $n=1$) with the map θ' defined by Karoubi ([Ka], Theorem 4.21) from $K_{2n+1}(\mathbb{C})$ to $\mathbb{C}^*$ by using the K-theory of Banach algebras. It would be interesting to see a proof of this fact.

Now the map from $(1+I)\otimes\mathbb{C}^*$ to $((1+I)\otimes\mathbb{C}(t)^*)/\mathscr{R}$ is easily believed to be (and may be shown to be - see the Appendix) the obvious one. If we now apply the dilogarithm map β to η we obtain $\sum_{i=1}^{n}\langle\omega\rangle_{\alpha_i/n}$ where the α_i are integers all congruent modulo n, $\sum_{i=1}^{n}\alpha_i=0$, and α_i is prime to n. Writing $\alpha_i/n=\alpha_1/n+c_i$, $c_i\in Z$, and observing that

$$\langle\omega\rangle_{\beta+k}=\langle\omega\rangle_{\beta}+k(\langle\omega\rangle_1-\langle\omega\rangle_0)\in\coprod_{x\in\mathbb{C}}\mathbb{C}^*/S,$$

we see that

$$\sum_{i=1}^{n}\langle\omega\rangle_{\alpha_i/n}=n\langle\omega\rangle_{\alpha_1/n}+\sum_{i=1}^{n}c_i(\langle\omega\rangle_1-\langle\omega\rangle_0).$$

Since $\sum_{i=1}^{n}d_i=0$, $\sum_{i=1}^{n}c_i=-\alpha_1$, and we obtain

$$\sum_{i=1}^{n}\langle\omega\rangle_{\alpha_i/n}=-\alpha_1(\langle\omega\rangle_1-\langle\omega\rangle_0)$$

which is equal to $\omega^{-\alpha_1}$ in $\mathbb{C}^*\subseteq\coprod_{x\in\mathbb{C}}^{0}\mathbb{C}^*/S$ (via b_1). Since α_1 is prime to n, $\omega^{-\alpha_1}$ is of exact order n in $\mathbb{C}^*$, which was to be shown.

Suslin has proven [Su1] that if $F_1\subseteq F_2$ are any two algebraically closed fields, the map from $K_3(F_1)_{\text{tor}}$ to $K_3(F_2)_{\text{tor}}$ is bijective. Since $K_3^M(F)$ is a Q-vector space if F is algebraically closed, $K_3(F)_{\text{tor}}$ is naturally isomorphic to $(K_3(F)_{\text{ind}})_{\text{tor}}$. Hence the natural map from $(K_3(F_1)_{\text{ind}})_{\text{tor}}$ to $(K_3(F_2)_{\text{ind}})_{\text{tor}}$ is bijective. It then follows immediately from the fact that Theorem 7.1 is true for $\mathbb{C}$ that it is true for $\bar{Q}$, then for any algebraically closed field of characteristic zero, then for any field of characteristic zero.

We now consider the case when F has characteristic p. We may assume F is algebraically closed. In fact, as above, we may assume that $F=\bar{F}_p$, where F_p is the prime field. Let $\mathscr{O}$ be a Henselian local ring with residue field F and quotient field E contained in $\bar{Q}$. We have the following commutative diagram:

$$\begin{array}{ccccc}
\text{Tor}_1(E^*,E^*) & \xleftarrow{\psi_1} & \text{Tor}_1(\mathscr{O}^*,\mathscr{O}^*) & \xrightarrow{\psi_2} & \text{Tor}_1(F^*,F^*) \\
\downarrow{\scriptstyle\varphi_1} & & \downarrow{\scriptstyle\varphi_2} & & \downarrow{\scriptstyle\varphi_3} \\
K_3(E)_{\text{ind}} & \xleftarrow{\rho_1} & \tilde{K}_3(\mathscr{O}) & \xrightarrow{\rho_2} & K_3(F)_{\text{ind}}
\end{array}$$

We have just seen that φ_1 is injective. Since ψ_1 is injective it follows that φ_2 is injective.

We next claim that the kernel of the map from $K_3(\mathscr{O})$ to $\tilde{K}_3(\mathscr{O})$ is torsion. For this purpose, we consider the commutative diagram with exact row and column:

$$\begin{array}{ccccc} & & & & K_3^M(E) \\ & & & & \downarrow \\ K_3(F) & \longrightarrow & K_3(o) & \longrightarrow & K_3(E) \\ & & \downarrow & & \downarrow \\ & & \tilde{K}_3(o) & \longrightarrow & K_3(E)_{\text{ind}} \end{array}$$

Since $K_3(F)$ and $K_3^M(E)$ are torsion (see [Q2] and [Ba-T] respectively), the claim follows immediately. But Gabber shows in [Ga] that if $p \nmid n$, $K_i(o, Z/n)$ is isomorphic to $K_i(F, Z/n)$ under the natural map. This, since $K_j(F)$ is divisible by n for $j>0$, immediately implies that the natural map from $K_3(o)$ to $K_3(F)$ is an isomorphism on torsion prime to p. Also, $K_3(F)_{\text{ind}} = K_3(F)$ for $F = F_p$. It immediately follows that $\rho_2\varphi_2$ is injective on torsion prime to p, and hence, since ψ_2 is surjective, φ_3 is injective.

§ 8. The proof of Axioms 3 and 5

We begin by reviewing Soulé's theory of weights and the γ-filtration on K-theory, with special reference to K_3 of local rings. The following is extracted from [So], pp. 1–29.

Let A be either a field or a regular local ring containing an infinite field. Then there is a filtration (the *γ-filtration*) on $K_3(A)$ such that $0 = F_\gamma^4(A) \subseteq F_\gamma^3(A) \subseteq F_\gamma^2(A) = K_3(A)$ and $F_\gamma^3(A) = K_3(A)_{\text{dec}}$, up to 2-torsion. There are functorial Adams operations ψ^k on $K_i(A)$ for all $k \in Z - \{0\}$ and all $i \geqq 0$, such that ψ^k is compatible with the γ-filtration and the product structure on K-theory, ψ^k acts by multiplication by k on $K_1(A)$ and ψ^k acts by multiplication by k^i on $gr_\gamma^i K_3(A) = F_\gamma^i(A)/F_\gamma^{i+1}(A)$.

Let $K_m(A)^{(i)}$ be the subgroup of $K_m(A)$ on which all the Adams operations ψ^k act by multiplication by k^i. Then $K_3(A) = K_3(A)^{(2)} \oplus K_3(A)^{(3)}$, up to 2-torsion. It is an immediate consequence of the above results that, up to 2-torsion, $K_3(A)_{\text{dec}} = K_3(A)^{(3)}$ and the natural map of $K_3(A)^{(2)}$ to $K_3(A)_{\text{ind}}$ induces an isomorphism.

Proposition 8.1. *Let A be a regular local ring which is the localization of an algebra of finite type over a field. Then the natural map from $K_3(A)^{(2)}$ to $\tilde{K}_3(A)$ is an isomorphism up to 2-torsion, and hence $\tilde{K}_3(A)$ is naturally isomorphic to $gr_\gamma^2 K_3(A)$, up to 2-torsion.*

Proof. We always work modulo 2-torsion. Let F be the quotient field of A. Since $\tilde{K}_3(A)$ is a quotient of $K_3(A)/K_3(A)_{\text{dec}}$, we have the commutative diagram

$$\begin{array}{ccc} K_3(A)^{(2)} & \xrightarrow{\varphi} & K_3(F)^{(2)} \\ \downarrow{\scriptstyle \rho} & & \downarrow{\scriptstyle \rho_F} \\ \tilde{K}_3(A) & \longrightarrow & \tilde{K}_3(F) \end{array}$$

Since A is the localization of an algebra of finite type over a field, A satisfies the Gersten conjecture [Q1] and hence φ is injective. Since we have already seen (Remark 2.6) that $\tilde{K}_3(F)$ is isomorphic to $K_3(F)_{\text{ind}}$, which in turn is isomorphic to $K_3(F)^{(2)}$, ρ_F is an isomorphism, and it follows that ρ is bijective.

Lemma 8.2. *Let A be the strict Henselization of the localization of a regular algebra of finite type over a field k of characteristic $p \geqq 0$. Let n be an odd positive integer not divisible by p. Then*

a) *The natural map $\mathrm{Tor}_1(A^*, A^*) \to \tilde{K}_3(A)$ induces an isomorphism of $\mathrm{Tor}_1(\mu_n, \mu_n)$ with $\tilde{K}_3(A)_n$.*

b) *$\tilde{K}_3(A)$ is n-divisible.*

c) *$K_2(A)$ is uniquely n-divisible.*

Proof. a) Consider the commutative diagram

$$\begin{array}{ccc} \mathrm{Tor}_1(A^*, A^*)_n & \xrightarrow{\psi} & \tilde{K}_3(A)_n \\ \downarrow & & \downarrow \tilde{\varphi} \\ \mathrm{Tor}_1(k^*, k^*)_n & \longrightarrow & \tilde{K}_3(k)_n \end{array}$$

Clearly $\tilde{\varphi}$ is surjective. Consider the commutative diagram

$$\begin{array}{ccccccccc} 0 & \longrightarrow & K_4(A)/nK_4(A) & \longrightarrow & K_4(A, Z/nZ) & \longrightarrow & K_3(A)_n & \longrightarrow & 0 \\ & & \downarrow & & \downarrow \theta & & \downarrow \varphi & & \\ 0 & \longrightarrow & K_4(k)/nK_4(k) & \longrightarrow & K_4(k, Z/nZ) & \longrightarrow & K_3(k)_n & \longrightarrow & 0 \end{array}$$

$K_4(k)$ is n-divisible for k separably closed by Suslin's theorem that $K_4(k)$ is n-divisible for k algebraically closed ([Su1]) and Lemma 3.2. We know that θ is an isomorphism by [Ga]. Hence φ is an isomorphism.

We know that, up to 2-torsion, we have $K_3(A) = K_3(A)^{(2)} \oplus K_3(A)^{(2)}$ and this is compatible with the natural map from A to k. It follows then that φ induces an isomorphism $K_3(A)_n^{(2)} \to K_3(k)_n^{(2)}$. By using Proposition 8.1 and taking direct limits, the natural maps from $K_3(A)^{(2)}$ to $\tilde{K}_3(A)$ and $K_3(k)^{(2)}$ to $\tilde{K}_3(k)$ are isomorphisms modulo 2-torsion, hence $\tilde{\varphi}$ induces an isomorphism from $\tilde{K}_3(A)_n$ to $\tilde{K}_3(k)_n$, and $\tilde{\psi}$ is an isomorphism.

b) Suslin [Su1] shows that $K_3(k, Z/nZ)$ is zero for k algebraically closed and hence for k separably closed by using Lemma 3.2. By the aforementioned result of Gabber, $K_3(A, Z/nZ) = 0$, hence $K_3(A)$ is n-divisible, hence $\tilde{K}_3(A)$ is n-divisible.

c) Since $K_3(A, Z/nZ) = 0$, $K_2(A)_n = 0$. We also have the commutative diagram

$$\begin{array}{ccccc} 0 & \longrightarrow & K_2(A)/nK_2(A) & \longrightarrow & K_2(A, Z/nZ) \\ & & \downarrow \psi & & \downarrow \varphi \\ 0 & \longrightarrow & K_2(k)/nK_2(k) & \longrightarrow & K_2(k, Z/nZ) \end{array}$$

Gabber's result implies φ is injective, hence ψ is injective, hence $K_2(A)/nK_2(A) = 0$.

Corollary 8.3. *Let X be a regular scheme of finite type over a field of characteristic $p \geqq 0$, and let n be an odd integer prime to p. Then Axiom 3 is true for X;* i.e., *there is a triangle in the derived category of étale sheaves of abelian groups on X:*

$$\begin{array}{ccc} & Z/nZ(2) & \\ \swarrow & & \nwarrow \\ \Gamma(2, X) & \xrightarrow{n} & \Gamma(2, X), \end{array}$$

Proof. This is an immediate consequence of Remark 2.6, Exercise 3.1, and Lemma 8.2.

Corollary 8.4. *Let X be as in Corollary 8.3. Then Axiom 5 is true for X;* i.e., *there exists a sequence which is exact up to 2-torsion:*

$$0 \to gr_2^{\gamma} K_3(X) \to C_{2,2}(X) \to C_{2,1}(X) \to K_2(X) \to 0.$$

Proof. This is an immediate consequence of Remark 2.6 and Lemma 8.1.

§ 9. "Hilbert's Theorem 90"

We prove in this section that our Axiom 2 (the appropriate generalization of Hilbert's Theorem 90) is true up to p-torsion and 2-torsion. We also show that the Mercuriev-Suslin theorem remains valid for regular local rings A which are localizations of algebras of finite type over a field k. For the rest of this section, let A be such a ring, and let $p \geqq 0$ be the characteristic of k.

Theorem 9.1. i) *The natural map from $K_2(A)/nK_2(A)$ to $H^2_{\text{ét}}(A, \mu_n \otimes \mu_n)$ is an isomorphism for n prime to p.*

ii) *$H^3_{\text{ét}}(A, \Gamma(2, A))$ has no odd torsion prime to p.*

Proof. i) Let E be the quotient field of A. Since the Gersten conjecture is true for A, we have the exact sequence:

$$\begin{aligned} 0 \to K_2(A) \to K_2(E) \xrightarrow{\varphi} & \coprod_{x \text{ codim } 1} K_1(\kappa(x)) \\ \to & \coprod_{y \text{ codim } 2} K_0(\kappa(y)) \to 0. \end{aligned}$$

We claim that this sequence remains exact after tensoring with Z/nZ. Let $Q = \text{Im}(\varphi)$. Since $\coprod_y K_0(\kappa(y))$ is free, it suffices to show that $\varphi_n : K_2(E)_n \to Q_n$ is surjective.

We have the following diagram:

$$\begin{array}{ccccc} H^1(E, \mu_n \otimes \mu_n) & \xrightarrow{\rho} & \coprod_{x \text{ codim } 1} H^0(\kappa(x), \mu_n) & \longrightarrow & 0 \\ \downarrow \psi & & \downarrow & & \\ K_2(E)_n & \xrightarrow{\varphi_n} & Q_n & & \end{array}$$

where we know that ρ is surjective by Bloch-Ogus ([Bl-O], Theorem 4.2) and ψ is Suslin's map [Su2]. This diagram commutes (evidently if $\mu_n \subset E$ and by reduction to that case in general). Hence φ_n is surjective. Let

$$S_1 = K_2(A)/nK_2(A), \qquad S_2 = K_2(E)/nK_2(E),$$
$$S_3 = \coprod_y K_1(\kappa(x))/nK_1(\kappa(x)), \qquad S_4 = \coprod_y K_0(\kappa(y))/nK_0\kappa(y),$$
$$T_1 = H^2_{\text{ét}}(A, \mu_n \otimes \mu_n), \qquad T_2 = H^2_{\text{ét}}(E, \mu_n \otimes \mu_n),$$
$$T_3 = \coprod_x H^1(\kappa(x), \mu_n), \qquad T_4 = \coprod_y H^0(\kappa(y), Z/nZ).$$

We now have the map of exact sequences:

$$\begin{array}{ccccccccc} 0 & \longrightarrow & S_1 & \longrightarrow & S_2 & \longrightarrow & S_3 & \longrightarrow & S_4 & \longrightarrow & 0 \\ & & \downarrow{\scriptstyle \theta'} & & \downarrow{\scriptstyle \theta} & & \downarrow & & \downarrow & & \\ 0 & \longrightarrow & T_1 & \longrightarrow & T_2 & \longrightarrow & T_3 & \longrightarrow & T_4 & \longrightarrow & 0 \end{array}$$

where the bottom row is again exact by Bloch-Ogus and the last three vertical maps are isomorphisms. (The last two are clear and θ is the Mercuriev-Suslin map.) So θ' is also an isomorphism.

ii) Let n be odd and prime to p. Then we have the commutative diagram with the bottom row exact (Corollary 8.3):

$$\begin{array}{ccccccc} & & K_2(A)/nK_2(A) & \xrightarrow{\theta'} & H^2_{\text{ét}}(A, \mu_n \otimes \mu_n) & & \\ & & \downarrow & & \downarrow & & \\ 0 \to \tilde{K}_2(A)/n\tilde{K}_2(A) & \longrightarrow & H^2_{\text{ét}}(A, \mu_n \otimes \mu_n) & \longrightarrow & H^3_{\text{ét}}(A, \Gamma(2))_n \to 0 & & \end{array}$$

where $\tilde{K}_2(A) = H^2_{\text{ét}}(A, \Gamma(2))$ and the commutativity is checked by reducing the case of a field. Since θ' is an isomorphism by i) it follows that $H^3_{\text{ét}}(A, \Gamma(2, A))_n = 0$.

Theorem 9.2. *$H^3_{\text{ét}}(A, \Gamma(2, A))$ is torsion.*

Proof. Let L be any residue field of A, and let j: Spec $L \to$ Spec A be the natural map. Let F be an étale sheaf on Spec L. Let α_* denote the natural restriction map from étale sheaves to Zariski sheaves. We claim that $R^q(\alpha_*)_A(j_*F)$ is torsion for $q > 0$. We clearly have $(\alpha_*)_A(j_*)_{\text{ét}} = (j_*)_{\text{Zar}}(\alpha_*)_F$. Since $R^q((j_*)_{\text{ét}})$, $R^q((j_*)_{\text{Zar}})$, and $R^q((\alpha_*)_F)$ are all zero modulo torsion for $q > 0$, our claim follows from the composite functor spectral sequences.

Since any finite étale extension of A is regular and semi-local, it also satisfies the Gersten conjecture, so we have the exact sequence of étale *presheaves* on A:

$$0 \to \underline{K}_2 \to j_* K_2(E) \xrightarrow{\psi} \coprod_{\operatorname{codim} x = 1} j_* K_1(\kappa(x)) \to \coprod_{\operatorname{codim} x = 2} j_* K_0(\kappa(x)) \to 0,$$

where E is the quotient field of A.

By using the basic properties of the transfer map on K_2, we see that the direct image by j_* of the sheaf associated to the presheaf $\underline{\underline{K}}_2$ on E is isomorphic, up to torsion, to the sheaf associated to the presheaf $j_* K_2$ on A. So, by a slight abuse of notation, the above exact sequence of presheaves becomes a sequence of sheaves which is exact up to torsion. Let $Q = \mathrm{Im}(\psi)$, so that we have

$$0 \to \alpha_* \underline{\underline{K}}_2 \to \alpha_* \coprod j_* K_2 \to \alpha_* Q \to R^1 \alpha_* \underline{\underline{K}}_2. \tag{9.3}$$

Again using the basic properties of the transfer map on K-theory, we see that the natural map from $(\underline{\underline{K}}_2)_{\mathrm{Zar}}$ to $\alpha_*(\underline{\underline{K}}_2)_{\text{ét}}$ is an isomorphism up to torsion. Similarly, we may identify $\alpha_*(\coprod j_*(K_i)_{\text{ét}})$ with $\coprod j_*(K_i)_{\mathrm{Zar}}$ for $i = 0, 1, 2$, modulo torsion.

This time using the Gersten conjecture for Zariski sheaves, we find:

$$0 \to (\underline{\underline{K}}_2)_{\mathrm{Zar}} \to \coprod j_*(K_2)_{\mathrm{Zar}} \to \coprod j_*(K_1)_{\mathrm{Zar}} \to \coprod j_*(K_0)_{\mathrm{Zar}} \to 0 \tag{9.4}$$

is exact.

So we have the exact sequence (modulo torsion)

$$0 \to \alpha_* \underline{\underline{K}}_2 \to \alpha_*(\coprod j_* K_2) \to \alpha_*(\coprod j_* K_1) \to \alpha_*(\coprod j_* K_0) \to 0, \tag{9.5}$$

and so

$$0 \to \alpha_* Q \to \alpha_*(\coprod j_* K_1) \to \alpha_*(\coprod j_* K_0). \tag{9.6}$$

Comparing (9.3), (9.5), and (9.6) we see that $R^1 \alpha_* \underline{\underline{K}}_2$ is torsion.

Now using the spectral sequence

$$H^p_{\mathrm{Zar}}(A, R^q \alpha_* \underline{\underline{K}}_2) \Rightarrow H^{p+q}_{\text{ét}}(A, \underline{\underline{K}}_2)$$

together with the fact that $H^p_{\mathrm{Zar}}(A, \cdot) = 0$ for $p > 0$ since A is local, we find that $H^1_{\text{ét}}(A, \underline{\underline{K}}_2) \simeq H^0_{\mathrm{Zar}}(A, R^1 \alpha_* \underline{\underline{K}}_2)$ and hence is torsion.

Now look at the exact sequence of étale sheaves

$$0 \to \underline{\underline{K}}_3 \to \coprod j_* K_3 \to \coprod j_* K_2 \to \cdots.$$

We take weight 2 components of this sequence, remembering that j_* increases weight by codimension, and conclude that $\underline{\underline{K}}_3^{(2)} \simeq j_* \underline{\underline{K}}_3^{(2)}$ up to torsion, where here j: Spec $E \to$ Spec A. Hence $R^q \alpha \underline{\underline{K}}_3^{(2)}$ is torsion for $q > 0$, and, as above, $H^q_{\text{ét}}(A, \underline{\underline{K}}_3^{(2)})$ is torsion for $q > 0$. Since the homology sheaves of $\Gamma(2, A)$ are, up to 2-torsion, $\underline{\underline{K}}_3^{(2)}$ and $\underline{\underline{K}}_2$, the spectral sequence of hypercohomology yields the exact sequence (modulo 2-torsion):

$$H^2_{\text{ét}}(A, \underline{\underline{K}}_3^{(2)}) \to H^3_{\text{ét}}(A, \Gamma(2, A)) \to H^1_{\text{ét}}(A, \underline{\underline{K}}_2)$$

and hence $H^3_{\text{ét}}(A, \Gamma(2, A))$ is torsion.

Corollary 9.7. *Let X be a scheme of finite type over a field k of characteristic $p \geqq 0$. Then $R^3 \alpha_* \Gamma(2, X)$ is zero up to 2-torsion and p-torsion.*

Proof. $R^3\alpha_*\Gamma(2, X)$ is the Zariski sheaf associated to the presheaf $U \to H^3_{\text{ét}}(U, \Gamma(2, U))$. The stalk of this sheaf at the local ring A is $H^3_{\text{ét}}(A, \Gamma(2, A))$, which is zero up to 2-torsion and p-torsion by Theorems 9.1 and 9.2.

§ 10. Galois descent for K_2

Let F be a field of characteristic $p \geqq 0$. We show in this section how our previous results enable us to describe up to p-torsion the extent to which $K_2(F)$ fails to satisfy Galois descent. Let $G = \text{Gal}(F_s/F)$.

Theorem 10.1. *The following sequence is exact (up to p-torsion).*

$$0 \to \bigoplus_{\ell \neq p} (\tilde{K}_3(F_s)^G \otimes Q_\ell/Z_\ell) \to \bigoplus_{\ell \neq p} H^1(F, Q_\ell/Z_\ell(2)) \to K_2(F)$$
$$\to K_2(F_x)^G \to \bigoplus_{\ell \neq p} H^2(F, Q_\ell/Z_\ell(2)) \to 0.$$

Proof. We have the hypercohomology spectral sequence

$$H^p(F, \mathscr{H}^q(\Gamma(2, F_s)) \Rightarrow H^{p+q}(F, \Gamma(2, F_s)),$$

where $\mathscr{H}^q$ of a complex of sheaves denotes the q-th cohomology sheaf of the complex.

Since $\mathscr{H}^q(\Gamma(2, F_s)) = 0$ for $q \neq 1, 2$, $\mathscr{H}^1(\Gamma(2, F_s)) = \tilde{K}_3(F_s)$, and $\mathscr{H}^2(\Gamma(2, F_s)) = K_2(F_s)$, we have the exact sequence

$$0 \to H^1(F, \tilde{K}_3(F_s)) \to H^2(F, \Gamma(2, F_s)) \to K_2(F_s)^G \to H^2(F, \tilde{K}_3(F_s))$$
$$\to H^3(F, \Gamma(2)) = 0 \quad \text{(up to } p\text{-torsion)}.$$

We first observe that $H^2(F, \Gamma(2, F_s) = K_2(F)$ up to p-torsion. We also have the exact sequence $0 \to Z/\ell^n Z(2) \to \tilde{K}_3(F) \xrightarrow{\ell^n} \tilde{K}_3(F) \to 0$ which gives rise to the exact sequences:

(1) $$0 \to H^1(F, \tilde{K}_3(F)) \otimes Q_\ell/Z_\ell \to H^2(F, Q_\ell/Z_\ell)(2))$$
$$\to H^2(F, \tilde{K}_3(F))(\ell) \to 0.$$

(2) $$0 \to \tilde{K_3}(F)^G \otimes Q_\ell/Z_\ell \to H^1(F, Q_\ell/Z_\ell(2)) \to H^1(F, \tilde{K}_3(F)(\ell) \to 0.$$

Since $H^1(F, \tilde{K}_3(F)) \otimes Q_\ell/Z_\ell$ is the tensor product of a torsion group with a divisible one, it must be zero and the Theorem easily follows.

§ 11. The relationship to Beilinson's complexes

The construction that we have described in this paper yields two possible definitions of weight-two Beilinson complexes $\Gamma_B(2, X)$ of Zariski sheaves, namely $\Gamma_{\text{Zar}}(2, X)$ and $\tau_{\leqq 2} R\alpha_* \Gamma(2, X)$, where α_* is the restriction functor from étale sheaves to Zariski sheaves. We discuss the extent to which they satisfy

Beilinson's conjectures. We use the numbering of Beilinson's conjectures given in our paper [L].

Beilinson's conjecture 1* (acyclicity outside of $[1, 2]$ is trivially true both for $\Gamma_{\mathrm{Zar}}(2, X)$ and $\tau_{\leqq 2} R\alpha_* \Gamma(2, X)$.

Beilinson's conjecture 3* asserts that

$$\Gamma_B(2, X) \overset{L}{\otimes} Z/nZ = \tau_{\leqq 2} R\alpha_* Z/nZ(2),$$

if n is invertible on X. As outlined in [L], this is true for $\tau_{\leqq 2} R\alpha_* \Gamma(2, X)$ if X is of finite type over a field and n is odd, in view of Corollary 8.3.

Beilinson's conjecture 5* asserts that $H^{4-i}_{\mathrm{Zar}}(X, \Gamma_B(2))$ is isomorphic to $gr^2_\gamma K_i(X)$ up to torsion, or possibly up to 2-torsion. In view of Remark 2.6, this conjecture is true for $\Gamma_{\mathrm{Zar}}(2, X)$, (even up to 2-torsion) if $X = \operatorname{Spec} F$.

Beilinson's conjecture 6* for fields asserts that $H^2_{\mathrm{Zar}}(F, \Gamma(2, F)) = K_2(F)$. This is true for $\Gamma_{\mathrm{Zar}}(2, F)$ and also for $\tau_{\leqq 2} R\alpha_* \Gamma(2, F)$, at least up to p-torsion.

In view of the oscillation between our two choices it is natural to ask if they are equal in the derived category. This would, by looking at $H^1(F, \Gamma(2))$, essentially imply that if F is a field, the natural map from $K_3(F)_{\mathrm{ind}}$ to $(K_3(F_s)_{\mathrm{ind}})^{G_F}$ is an isomorphism. Whether this is true for general F seems to be a very interesting question.

Appendix

Let R be a commutative ring. I an ideal of R. Recall that the "double" D of R with respect to I is defined to be the fibered product $R \underset{R/I}{\times} R$.

Lemma A1. *Let R be a semi-local ring, I an ideal of R contained in the Jacobson radical of R. Then D is a semi-local ring, and $\operatorname{Ker}(\pi_1)$ is contained in the Jacobson radical of D.*

Proof. Let $\pi_1 : D \to R$, $\operatorname{Ker}(\pi_1) = J$. Let m be a maximal ideal of D. We claim that $\pi_1(m) \neq R$; i.e. that $J \subset m$. Suppose $\pi_1(m) = R$. Then $m \ni (1, z)$ for some $z \in R$, $z \equiv 1 \pmod{I}$. Since $z \in 1 + I$, and I is contained in the Jacobson radical of R, z has an inverse $y \in R$, and y is clearly in $1 + I$. Hence $(1, z)(1, y) = 1_D$ and $m = D$ which is a contradiction. So J is contained in the Jacobson radical of D. Now the maximal ideals of D correspond bijectively to the maximal ideals of $R = D/J$, hence if R is semi-local, D is semi-local.

Lemma A2. *Let φ be a natural pairing from $K_1(R) \times K_1(R, I)$ for all commutative rings R and ideals I. Assume that if R is semi-local (so that $K_1(R)) = R^*$) and $I = 0$, then φ is given by the symbol map $a \times b \to (a, b)$. Then if R is semi-local, and I is contained in the Jacobson radical of R (so that $K_1(R, I)) = (1 + I)$, then φ is given by the map $a \times b \to \langle -a, (1-b)/a \rangle$, where $\langle x, y \rangle$ is a generator in the Keune presentation of $K_2(R, I)$* ([Ke], p. 175]), *$a \in R^*$.*

Proof. Recall from [Ke] that a set of generators for $K_2(R, I)$ is given by elements $\langle \alpha, \beta \rangle$ with $\alpha \in R$, $\beta \in R$ or $\alpha \in I$, $\beta \in R$. (The condition that $1 + \alpha\beta$ be a unit is automatic in this case.) Furthermore, it is well-known (see for instance [D-S]) that if $\alpha \in R^*$, $\beta \in I$, the natural map from $K_2(R, I)$ to $K_2(R)$ takes $\langle \alpha, \beta \rangle$ to $(-\alpha, 1 + \alpha\beta)$.

Let D be the "double" of R with respect to I. Let $J = \ker(\pi_1 : D \to R)$, and consider the commutative diagram:

$$\begin{array}{ccc} K_1(D) \times K_1(D, J) & \xrightarrow{\varphi_J} & K_2(D, J) \\ \downarrow \quad \downarrow & & \downarrow \wr \\ K_1(D) \times K_1(D) & \xrightarrow{\varphi} & K_2(D) \end{array}$$

Since the map π_1 from D to $R=D/J$ has a section given by the diagonal map Δ, the long exact sequence of relative K-theory splits into short exact sequences

$$0 \to K_j(D, J) \xrightarrow{i_j} K_j(D) \xrightarrow{\rho_j} K_j(R) \to 0.$$

In particular, $i=i_2$ is injective, and ρ_1 is surjective.

By assumption, $\varphi(a, b)=(a, b)=i(\langle -a,\ (1-b)/a\rangle)$ for $a \in K_1(D)=D^*$, $b \in K_1(D, J)=1+J$. So $i\varphi_J(a, b)=i(\langle -a,\ (1-b)/a\rangle)$, which implies $\varphi_J(a, b)=\langle -a, (1-b)/a\rangle$. So the lemma is true for the pair (D, J).

Now consider the commutative diagram:

$$\begin{array}{ccc} K_1(D) \times K_1(D, J) & \longrightarrow & K_2(D, J) \\ \downarrow \rho_1 \quad \downarrow \rho & & \downarrow \\ K_1(R) \times K_1(R, I) & \longrightarrow & K_2(R, I) \end{array}$$

Since π_1 is surjective and maps J isomorphically onto I, ρ is an isomorphism by Lemma 6.3 of [Mi]. We have seen that ρ_1 is surjective, so the truth of the lemma for R follows from its truth for D.

Lemma A3. *Let R be a semi-local domain and I an ideal contained in the Jacobson radical of I. Let F be the quotient field of R. In* [Bl1] *Bloch defines a map ψ from $K_2(R, I)$ to $(1+I) \underset{Z}{\otimes} F^*/\mathscr{R}$, where $\mathscr{R}$ is generated by the set of elements $(1-x) \otimes x$, $x \in I$. This map ψ is given by $\psi(\langle \alpha, \beta\rangle)=(1+\alpha\beta) \otimes \beta^{-1}=(1+\alpha\beta) \otimes -\alpha$. Bloch shows that the Keune relations between the generators are sent to elements of $\mathscr{R}$, so ψ is well-defined.*

Proposition A4. *Let φ be the pairing defined in* § 1 *from $K_1(R) \times K_1(R, I) \to K_2(R, I)$. Then $\psi\varphi(a \times b)=b \otimes a$.*

Proof. By definition, φ agrees with Waldhausen's product on absolute K-theory, hence (see [We]) with Loday's product [Lo]. Since Loday's product is given by the symbol pairing $a \times b \to (a, b)$ for $a, b \in R^*$, R semi-local, φ satisfies the hypotheses of Lemma 2. Hence $\psi\varphi(a \times b)=\psi\langle -a,\ (1-b)/a\rangle = b \otimes a$.

References

[Ba-T] Bass, H., Tate, J.: The Milnor ring of a global field. Algebraic K-Theory II. Lect. Notes in Math., vol. 342, pp. 348–446. Berlin-Heidelberg-New York: Springer 1973

[Be] Beilinson, A.: Letter to C. Soulé, 11/1/82

[Bl1] Bloch, S.: Higher regulators, algebraic K-theory, and zeta-functions of elliptic curves. Lecture Notes, U.C. Irvine (1977)

[Bl2] Bloch, S.: Algebraic cycles and higher K-theory. Adv. Math. **61**, 267–304 (1986)

[Bl-O] Bloch, S., Ogus, A.: Gersten's conjecture and the homology of schemes. Ann. Sci. Ec. Norm. Super., IV. Ser. **7**, 181–202 (1974)

[D-W] Deninger, C., Wingberg, K.: Artin-Verdier duality for n-dimensional local fields involving higher algebraic K-sheaves. Preprint

[D-S] Dennis, R.K., Stein, M.R.: K_2 of radical ideals and semi-local rings revisited. Algebraic K-Theory II. Lect. Notes in Math., vol. 342, pp. 281–303. Berlin-Heidelberg-New York: Springer 1973

[Ga] Gabber, O.: K-theory of Henselian local rings and Henselian pairs. Preprint

[Gr] Grayson, D.: Products in K-theory and intersecting algebraic cycles. Invent. Math. **47**, 71–84 (1978)

[Ka] Karoubi, M.: Homologie de groups discrets associés **S** des algebres d'opérateurs. Preprint

[Ke] Keune, F.: The relativization of K_2. J. Algebra **54**, 159–177 (1978)

[L] Lichtenbaum, S.: Values of zeta-functions at non-negative integers. Number Theory, Noordwijkerhout 1983. Lect. Notes in Math., vol. 1068, pp. 127–138. Berlin-Heidelberg-New York: Springer 1984

[Lo] Loday, J.L.: K-théorie algébrique et représentations des groups. Ann. Sci. Ec. Norm. Super., IV. Ser. **9**, 309–377 (1976)

[M-S] Mercuriev, A.S., Suslin, A.A.: K-cohomology of Severi-Brauer varieties and the norm residue homomorphism. Mathematics of the USSR-Izvestiya (Translation) **21**, 307–340 (1983)

[Mi] Milnor, J.: Introduction to Algebraic K-theory. Annals of Math. Studies No. 72. Princeton, N.J.: Princeton University Press 1971

[Q1] Quillen, D.: Higher algebraic K-theory I. Algebraic K-theory I. Lect. Notes in Math., vol. 341, pp. 85–147. Berlin-Heidelberg-New York: Springer 1973

[Q2] Quillen, D.: On the cohomology and K-theory of the general linear groups over a finite field. Ann. Math. **96**, 552–586 (1972)

[So] Soulé, C.: Opérations en K-théorie algébrique. Gn. J. Math. **37**, 488–550 (1985)

[Su1] Suslin, A.A.: On the K-theory of algebraically closed fields. Invents. Math. **73**, 241–245 (1983)

[Su2] Suslin, A.A.: Torsion in K_2 of fields. Preprint, LOMI E-2-82

[Su3] Suslin, A.A.: Homology of GL_n, characteristic classes and Milnor K-theory. Lect. Notes Math., 1046, pp. 357–375, Berlin Heidelberg New York: Springer 1984

[T] Tate, J.T.: Relations between K_2 and Galois cohomology. Invent. Math. **36**, 257–274 (1976)

[W] Waldhausen, F.: Algebraic K-theory of generalized free products. Ann. Math. **108**, 135–256 (1978)

[We] Weibel, C.A.: A survey of products in Alg. K-Theory. Algebraic K-Theory Evanston 1980. Lect. Notes in Math., vol. 854, pp. 494–517. Berlin-Heidelberg-New York: Springer 1981

Oblatum 31-XII-1985

Invent. math. 88, 217–224 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Regularity at infinity for area-minimizing hypersurfaces in hyperbolic space

Robert Hardt[1]* and Fang-Hua Lin[2]**

1 School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
2 Courant Institute of Mathematical Sciences, New York University, New York, NY 10012, USA

In the upper half space model, $n+1$ dimensional hyperbolic space is given as the set

$$\mathbb{H} = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : y > 0\}$$

equipped with the hyperbolic metric $y^{-2}(dx^2 + dy^2)$. A standard compactification of $\mathbb{H}$ involves adding the boundary $(\mathbb{R}^n \times \{0\}) \cup \{*\}$ so that $\bar{\mathbb{H}}$ is simply the one point compactification of the Euclidean closed half-space $\mathbb{R}^n \times [0, \infty)$. Suppose $0 \leqq \alpha \leqq 1$ and Γ is a compact $n-1$ dimensional $\mathscr{C}^{1,\alpha}$ smooth submanifold of $\mathbb{R}^n \times \{0\}$. In [$A_1$], M. Anderson proved that *there exists an n dimensional hyperbolic-area minimizing locally rectifiable current T in* $\mathbb{H}$ *whose support has* Γ *as its asymptotic limit.* (See also [A_2].) By the interior regularity theory of geometric measure theory, the support M of any such hyperbolic-area minimizing T is a relatively closed subset of $\mathbb{H}$ which is a real analytic submanifold away from a relatively closed singular set of Hausdorff dimension $n-7$. Anderson's construction gives M for which, in the Euclidean topology, $\bar{M} \sim M = \Gamma$. The question of the behavior of M near Γ was raised in [A_1] and [LR]. Here we prove the "boundary regularity at infinity" result that, *for any such hyperbolic-area minimizing T, the set* $M \cup \Gamma$, *in the ordinary Euclidean metric, is, near* Γ, *a finite union of* $\mathscr{C}^{1,\alpha}$ *submanifolds with boundary* Γ*; these have disjoint analytic interiors and meet* $\mathbb{R}^n \times \{0\}$ *orthogonally at* Γ. (For the particular T, constructed by Anderson only one submanifold will occur.) It follows that, for $n \leqq 6$, $M \cup \Gamma$ has finite genus, and, for $n \geqq 7$, any interior singularities of M must remain in a bounded region of hyperbolic space. Near points of Γ (in the Euclidean topology), $M \cup \Gamma$ may thus be described as the graph of a function. This function is the solution of an interesting partial differential equation that becomes degenerate along the part of the boundary corresponding to Γ. The second author has recently established [L] a boundary higher-regularity result for this equation, which implies, in particular, that $M \cup \Gamma$ is $\mathscr{C}^{k,\alpha}$ if Γ is $\mathscr{C}^{k,\alpha}$ for $k = 2, 3, \ldots, \infty$. Finally, in case Γ bounds a star-

* Research partially supported by the National Science Foundation
** Research partially supported by the Mathematical Sciences Research Institute

shaped domain in $\mathbb{R}^n \times \{0\}$, we deduce in § 4 the uniqueness (up to a constant factor) of T, among stationary currents having asymptotic limit Γ.

§ 1. Preliminaries

For convenience, we adopt the following convention:

All unspecified statements about topology and metric will refer implicitly to the usual Euclidean metric on $\mathbb{R}^n \times \mathbb{R}$, and the word "hyperbolic" will be stated explicitly when appropriate.

Throughout, we assume that

$\Gamma = \tilde{\Gamma} \times \{0\}$ is a fixed compact submanifold of $\mathbb{R}^n \times \{0\}$ and that T is an n dimensional hyperbolic-area minimizing locally rectifiable current in $\mathbb{H}$ with $\Gamma = \bar{M} \sim M$ where $M = \operatorname{spt} T$.

We make some preliminary observations on the location of M near Γ. These are all based on the following:

(1.1) *If $x \in \mathbb{R}^n$ and $0 < r < d(x) = \operatorname{dist}(x, \tilde{\Gamma})$, then $M \cap \mathbb{B}_r(x, 0) = \emptyset$.*

This follows from a standard argument (see e.g. [A_1, Lemma 5]) using first variation because the sets $\mathbb{H} \cap \partial \mathbb{B}_s(x, 0)$, for $0 < s < r$, are totally geodesic hyperbolic hyperplanes that foliate $\mathbb{H} \cap \mathbb{B}_r(x, 0)$ and serve as barriers. Let v_Γ be a unit normal vectorfield for $\tilde{\Gamma}$ in $\mathbb{R}^n$ and, for $a \in \tilde{\Gamma}$ and $r > 0$,

$$\delta(a, r) = \min \{d(a + r v_\Gamma(a)), d(a - r v_\Gamma(a))\}.$$

We observe that

(i) $r^{-1} \delta(a, r) \to 1$ as $r \to 0$ if Γ is differentiable at $(a, 0)$.

(ii) $\sup_{a \in \Gamma} [1 - r^{-1} \delta(a, r)] \to 0$ as $r \to 0$ if Γ is $\mathscr{C}^1 = \mathscr{C}^{1,0}$.

(iii) $\sup_{a \in \Gamma} [1 - r^{-1} \delta(a, r)] \leqq c_\Gamma r^\alpha$ if Γ is $\mathscr{C}^{1,\alpha}$ for $0 < \alpha < 1$.

(iv) $r^{-1} \delta(a, r) = 1$ for all positive $r < 1/\|\text{max. principle curv.}\|_{L^\infty(\Gamma)}$ if Γ is $\mathscr{C}^{1,1}$.

We will first consider the case when Γ is $\mathscr{C}^1$. The $\mathscr{C}^{1,\alpha}$ cases, with $0 < \alpha \leqq 1$, will be treated in § 3. Choose a positive number ρ_Γ so that $r^{-1} \delta(a, r) > \frac{1}{2}$ for all $(a, 0) \in \Gamma$ and $0 < r < 2\rho_\Gamma$. By (1.1),

$M \cap \{y < \rho_\Gamma\}$ is contained in the set

$$W = [\mathbb{R}^n \times (0, \rho_\Gamma)] \sim \bigcup_{\mathbf{d(x)} > \mathbf{2}\rho_\Gamma} \mathbb{B}_{2\rho_\Gamma}(x, 0) \sim \bigcup_{\mathbf{0} < \mathbf{d(x)} \leqq \mathbf{2}\rho_\Gamma} \mathbb{B}_{\mathbf{d(x)}}(x, 0). \tag{1.2}$$

Clearly W contains the product $\tilde{\Gamma} \times (0, \rho_\Gamma)$. We will show that W is also asymptotic to $\tilde{\Gamma} \times (0, \rho_\Gamma)$ as $y \to 0$. To see this, suppose that (x_i, y_i) is a sequence of points in W with $y_i \to 0$. Since every point $x \in \mathbb{R}^n \sim \tilde{\Gamma}$ is included in the above definition of W, $d(x_i)$ must approach 0. We now wish to show that $d(x_i)/y_i$ also approaches 0. For this, we may assume that $y_i < d(x_i)^{\frac{3}{4}}$. Choose a point $(a_i, 0) \in \Gamma$

with $d(x_i)=|x_i-a_i|$; hence, $x_i=a_i\pm d(x_i)\nu_\Gamma(a_i)$. Letting $r_i=d(x_i)+d(x_i)^{-1}y_i^2$, we see that the three points, $(a_i,0)$, (x_i,y_i), and $(a_i\pm r_i\nu_\Gamma(a_i),0)$, are vertices of a right triangle. Letting δ_i be the distance between the latter two vertices, we conclude that, as $i\to\infty$.

$$d(x_i)/y_i=[r_i^2-\delta_i^2]^{\frac{1}{2}}/\delta_i=[(r_i/\delta_i)^2-1]^{\frac{1}{2}}\to 0$$

because $r_i\leqq d(x_i)+d(x_i)^{\frac{1}{2}}\to 0$, $\delta(a_i,r_i)\leqq d(a_i\pm r_i\nu_\Gamma(a_i))\leqq\delta_i\leqq r_i$, and $r_i/\delta(a_i,r_i)\to 1$ by (ii). It now follows that

(1.3) *at each edge point* $(a,0)\in\Gamma$, *the tangent cone of the containing set* W *equals the vertical half-hyperplane* $\mathrm{Tan}(\tilde{\Gamma},a)\times[0,\infty)$.

§ 2. $\mathscr{C}^1-\mathscr{C}^1$ regularity at ∞

We will sometimes identify $\mathbb{R}^n$ with $\mathbb{R}^{n-1}\times\mathbb{R}$ and use the projection

$$p\colon(\mathbb{R}^{n-1}\times\mathbb{R})\times\mathbb{R}\to\mathbb{R}^{n-1}\times\mathbb{R},\qquad p((w,z),y)=(w,y).$$

2.1. Lemma. (Interior regularity.) *For any nonnegative integer* κ *and positive number* ε, *there exists a positive* δ *so that if* S *is an* n *dimensional hyperbolic-area minimizing rectifiable current in* $\mathbb{H}$ *with*

$$\mathrm{spt}\,S\subset\mathbb{B}_2^{n-1}(0)\times[-\delta,\delta]\times[\tfrac{1}{2},4],\qquad \mathrm{spt}\,\partial S\subset p^{-1}\partial(\mathbb{B}_2^{n-1}(0)\times[\tfrac{1}{2},4]),$$

and $p_*S=\kappa[\![\mathbb{B}_2^{n-1}(0)\times[\tfrac{1}{4},4]]\!]$, *then* $S\llcorner p^{-1}(\mathbb{B}_1^{n-1}(0)\times[\tfrac{1}{2},2])$ *is the sum of* κ *oriented graphs (which are pairwise either disjoint or identical) of real analytic functions* $z=u_j(w,y)$ *with* $\|u_j\|_{\mathscr{C}^{1,1}}\leqq\varepsilon$.

(*If* $\kappa=0$, *then* $S\llcorner p^{-1}(\mathbb{B}_1^{n-1}(0)\times[1,2])=0$.)

Proof. Here $\kappa[\![\mathbb{B}_2^{n-1}(0)\times[\tfrac{1}{4},4]]\!]$ denotes κ times the current corresponding to oriented integration, with respect to the hyperbolic metric, over the set $\mathbb{B}_2^{n-1}(0)\times[\tfrac{1}{4},4]$. Hyperbolic area may be described by the parametric integrand [F, 5.1.1] $\Psi((x,y),\xi)=y^{-n}|\xi|$. On the region $\mathbb{R}^n\times(\frac{1}{4},4)$, this integrand is elliptic [F, 5.1.2] with ellipticity bound 4^n. As in the proof of [F, 5.3.18] one may find a decomposition $S=\bigcup_{j=-\infty}^{\infty}S^{(j)}$ where each nonzero $S^{(j)}$ satisfies the hypothesis with $\kappa\in\{-1,0,+1\}$. Then one deduces, as in [F, 5.1.1], a mass estimate for each $S^{(j)}$ from the Ψ-minimality and the small height.

To prove the lemma one may now argue by contradiction, and consider, for any positive ε, a sequence of positive numbers δ_i approaching 0 and a corresponding sequence $S_i=\bigcup_{j=-\infty}^{\infty}S_i^{(j)}$ as above satisfying the hypothesis but not the conclusion of 2.2. By the lower density bound [F, 5.1.6], the support of each $S_i^{(j)}$ with zero projection does not meet $p^{-1}(\mathbb{B}_1^{n-1}(0)\times[\tfrac{1}{2},2])$ for i large. Using the mass bounds, the BV compactness theorem, the convergence of supports, and the regularity theory [F, 5.3.14], one finds that, for i large, the support of each $S_i^{(j)}$ with nonzero projection, is, on the smaller cylinder

$p^{-1}(\mathbb{B}^n_{3/2}{}^{-1}(0)\times[\frac{1}{3},3])$, an oriented graph of a $\mathscr{C}^{1,\frac{1}{2}}$ function $z=u_{i,j}(w,y)$. Moreover, a simple cutting and pasting argument shows that, for each such i, these all have parallel orientation. Thus, for each such i, there are precisely κ currents $S_i^{(j)}$ with $p_\# S_i^{(j)}\neq 0$, and each has $p_\# S_i^{(j)}=+[\![\mathbb{B}^{n-1}_2(0)\times[\frac{1}{4},4]]\!]$. Finally each $u_{i,j}$ is analytic by [M, 5.8], and

$$\lim_{i\to\infty}\|u_{i,j}\|_{\mathscr{C}^{1,1}(\mathbb{B}^{n-1}_1(0)\times[\frac{1}{2},2])}=0$$

by interior estimates [GT, 15.2] because $\lim_{i\to\infty}\delta_i=0$. This now contradicts, for i sufficiently large, the choice of S_i. □

2.2. Theorem. *If Γ, T, and M are as in § 1, and Γ is $\mathscr{C}^1$, then there exists a positive $\rho<\rho_\Gamma$ so that $(M\cup\Gamma)\cap\{y<\rho\}$ is a finite union of $\mathscr{C}^1$ submanifolds with boundary; these have disjoint analytic interiors and meet $\mathbb{R}^n\times\{0\}$ orthogonally at Γ.*

Proof. We will first verify the statement:
for some positive ρ, the set $M\cap\{y<\rho\}$ has no interior singularities and

$$v_{\mathbf{M}}(x_i,y_i)\to(v_\Gamma(a),0)\qquad \textit{whenever } (x_i,y_i)\in M\to(a,0)\in\Gamma.$$

If the statement were false, then there would exist a sequence (x_i,y_i) in M approaching a point $(a,0)$ in Γ so that

either $(x_i,y_i)\in\mathrm{Sing}(M)$ for all i

or $\lim_{i\to\infty}|v_{\mathbf{M}}(x_i,y_i)-(v_\Gamma(a_i),0)|>0$

for any choice of orienting normal field $v_{\mathbf{M}}$. Since a rotation or translation of $\mathbb{R}^n$ induces a hyperbolic isometry of $\mathbb{H}$, we may assume, for convenience, that $a=0$ and that $v_\Gamma(a)=(0,1)\in\mathbb{R}^{n-1}\times\mathbb{R}$.

By the $\mathscr{C}^1$ smoothness of Γ, we may choose, for each positive ρ, a positive σ so that p projects $\Gamma\cap[\mathbb{B}^n_{2\sigma}{}^{-1}(0)\times(-\sigma\rho,+\sigma\rho)\times\{0\}]$ $\mathscr{C}^1$-diffeomorphically onto $\mathbb{B}^{n-1}_{2\sigma}(0)\times\{0\}$. It then follows from the discussion of the set W in § 1 that

$$M\cap[\mathbb{B}^{n-1}_\sigma(0)\times\{\pm\sigma\rho\}\times(0,\tau)]=\emptyset$$

for some sufficiently small positive τ. By the constancy theorem [F, 4.1.7] (which holds in any metric),

$$p_\#(T\llcorner[\mathbb{B}^{n-1}_\sigma(0)\times(-\sigma\rho,+\sigma\rho)\times(0,\tau)])=\kappa[\![\mathbb{B}^{n-1}_\sigma\times(0,\tau)]\!]$$

for some integer κ. Replacing T by $-T$ if necessary, we may assume that κ is nonnegative.

Next we wish to scale T appropriately. To do this note that, for each positive r, the homothety map that sends (x,y) to (rx,ry) induces a hyperbolic isometry of $\mathbb{H}$. Thus, for each i, the expression

$$\Phi_i(x,y)=y_i^{-1}(x-x_i,y),$$

defines a hyperbolic isometry, and the current

$$S_i=\Phi_{i\#}(T\llcorner[\mathbb{R}^{n-1}\times(-\sigma\rho,+\sigma\rho)\times(0,\infty)])\llcorner p^{-1}(\mathbb{B}^{n-1}_2(0)\times[\tfrac{1}{4},4])$$

is hyperbolic-area minimizing. Moreover, for any positive δ, we easily verify, using (1.2) and (1.3), that S_i satisfies the hypotheses of Lemma 1.1 for all i sufficiently large. We conclude that, for such i, (x_i, y_i) is a regular point of M because $((0,0),1)=\Phi_i(x_i,y_i)$ is, by 2.1, a regular point of $\operatorname{spt} S_i$. Moreover, letting $u_{i,j}$ denote the functions provided by the conclusion of 2.1 and letting δ_i correspond in 2.1 to a sequence of positive ε_i approaching 0, we may, by (1.2) and (1.3), pass to a subsequence to have

$$\lim_{i\to\infty} \|u_{i,j}\|_{\mathscr{C}^{1,1}}=0.$$

In particular

$$\lim_{i\to\infty} |v_{\mathbf{M}}(x_i,y_i)-(v_{\mathbf{\Gamma}}(a),0)|=\lim_{i\to\infty} |v_{\mathbf{M}}(x_i,y_i)-((0,1),0)|=0$$

because $\lim_{i\to\infty} \nabla u_{i,j}(0)=0$. This now contradicts the choice of (x_i, y_i) and a, and completes the proof of the statement.

Defining $v_{\mathbf{M}}(a,0)=(v_{\mathbf{\Gamma}}(a),0)$ for $(a,0)\in\Gamma$, we infer from the statement, the continuity of $v_{\mathbf{\Gamma}}$ on Γ, and the continuity of $v_{\mathbf{M}}$ on $M\cap\{y<\rho\}$, that $v_{\mathbf{M}}$, so extended, is continuous on $\Gamma\cup[M\cap\{y<\rho\}]$.

Let Γ_k be a component of Γ, and consider a component M_j of $M\cap\{y<\rho\}$ with $\Gamma_k\subset\bar{M}_j$. Then M_j is, by interior regularity, an analytic submanifold, and $\Gamma_k\cup M_j$ is $\mathscr{C}^1$ regular at Γ_k. The M_j are disjoint by the maximum principle. By orienting each $n-1$ plane $\operatorname{Tan}(\Gamma_k,a)$ continuously in $a\in\Gamma_k$, we may define a continuous family of currents $[\![\operatorname{Tan}(\Gamma_k,a)\times(0,\rho)]\!]$. Arguing as above, there is an integer κ_k, independent of $a\in\Gamma_k$, so that the orthogonal projection onto $\operatorname{Tan}(\Gamma_k,a)\times\mathbb{R}$ locally projects $T\llcorner\{y<\rho\}$ to a piece of $\kappa_k[\![\operatorname{Tan}(\Gamma_k,a)\times(0,\rho)]\!]$. Again by minimality and cutting and pasting, we see that the currents $T\llcorner M_j$ all must induce the same orientations under these projections. Thus there are at most κ_k such M_j. In particular, the number of components of $M\cap\{y<\rho\}$ is finite. □

§ 3. $\mathscr{C}^{1,\alpha}$–$\mathscr{C}^{1,\alpha}$ regularity at ∞

If Γ is $\mathscr{C}^1$, then, we may, as in 2.2, change coordinates to view each component of $M\cap\{y<\rho\}$ locally, near $(0,0)\in\Gamma$, as lying in the graph of a function $z=u(w,y)$ that is $\mathscr{C}^1$ on a region $\mathbb{B}_\rho^{n-1}(0)\times[0,\rho)$. For any compact $K\subset\mathbb{B}_\rho^{n-1}(0)\times(0,\rho)$, the hyperbolic area of the graph of $u|K$ is

$$\int_{\mathbf{K}} (1+u_w^2+u_y^2)^{\frac{1}{2}} y^{-n}\,dw\,dy.$$

The minimality of T leads to the Euler-Lagrange equation

$$\sum_{j=1}^{n-1} [(1+u_w^2+u_y^2)^{-\frac{1}{2}}u_{\mathbf{w}_j}]_{\mathbf{w}_j}+[(1+u_w^2+u_y^2)^{-\frac{1}{2}}u_{\mathbf{y}}]_{\mathbf{y}}-\binom{n}{y}[(1+u_w^2+u_y^2)^{-\frac{1}{2}}u_{\mathbf{y}}]=0.$$

We find it more convenient and natural to view M using polar graphs over regions in a hemisphere centered in $\mathbb{R}^n\times\{0\}$.

Let $\mathbb{S}^n_+ = \{(x, y) \in \mathbb{R}^n \times \mathbb{R} : |x|^2 + y^2 = 1,\ y > 0\}$. For a region $\Omega \subset \mathbb{S}^n_+$ and a positive $\mathscr{C}^2$ function u on Ω, the polar graph of u is the set

$$\operatorname{graph}(u) = \{u(\omega)\omega : \omega \in \Omega\}.$$

We want to calculate the hyperbolic area of $\operatorname{graph}(u|K)$ for any compact $K \subset \mathbb{S}^n_+$. For this purpose, we represent points in $\mathbb{S}^n_+ \sim \{(0, 0, \ldots, 1)\}$ using pairs $(\theta, \eta) \in \mathbb{S}^{n-1} \times (0, \frac{1}{2}\pi]$ where $\theta = x/|x|$ and $\eta = \sin^{-1} y$ (i.e., $\eta = \frac{1}{2}\pi - \varphi$). In these coordinates, $\mathbb{S}^n_+$ has the standard metric $d\eta^2 + \cos^2 \eta d\theta^2$. The hyperbolic area is

$$\int_K [(u^2 + |\nabla u|^2) u^{2(n-1)}]^{\frac{1}{2}} (u \cdot \sin \eta)^{-n} \cos \eta \, d\theta \, d\eta$$
$$= \int_K (1 + u^{-2} |\nabla u|^2)^{\frac{1}{2}} (\sin \eta)^{-n} \cos \eta \, d\theta \, d\eta$$

where $|\nabla u|^2 = |\nabla_\theta u|^2 + |\nabla_\eta u|^2$. Note that, as expected, the graphs of $u|K$ and $\lambda u|K$, for $\lambda > 0$, have the same hyperbolic area.

3.1. Theorem. *If $0 < \alpha \leqq 1$ and Γ is $\mathscr{C}^{1,\alpha}$, then each of the manifolds in the conclusion of Theorem 2.1 are $\mathscr{C}^{1,\alpha}$ regular at their boundaries.*

Proof. We first consider the case $\alpha = 1$. Suppose that N is one of the components of $M \cap \{y < \rho\}$ and $(a, 0) \in \Gamma \cap \bar{N}$. After applying hyperbolic isometries induced by a suitable translation and rotation of $\mathbb{R}^n \times \{0\}$, we may assume that $a = (0, \ldots, 0, 1)$ and the $\nu_\Gamma(a) = (0, \ldots, 0, 1, 0)$. By Theorem 2.1, there is a positive δ so that $N \cup \Gamma$ is, near a, the polar graph of a positive $\mathscr{C}^1$ function u on a region $\Omega \subset \mathbb{S}^n_+$ defined by $|\theta - a| < \delta$ and $0 \leqq \eta < \delta$. Here $u(a, 0) = 1$ and $\nabla u(a, 0) = 0$.

The function $v(\theta, \eta) = \log(u(\theta, \eta))$ is also $\mathscr{C}^1$ on Ω. It locally minimizes the integral

$$\int (1 + |\nabla v|^2)^{\frac{1}{2}} (\sin \eta)^{-n} \cos \eta \, d\theta \, d\eta,$$

and has $v(a, 0) = 0$ and $\nabla v(a, 0) = 0$. Since Γ is $\mathscr{C}^{1,1}$ there is, as in § 1(iv), a positive r, depending only on Γ, so that the two closed balls in $\mathbb{R}^n \times \mathbb{R}$,

$$\bar{\mathbb{B}}_r(0, \ldots, 0, 1 - r, 0) \quad \text{and} \quad \bar{\mathbb{B}}_r(0, \ldots, 0, 1 + r, 0),$$

which are tangent to Γ at a, do not intersect M. Since these have quadratic contact with the unit sphere $\mathbb{S}^n$,

$$|v(\theta, \eta)| \leqq c(\Gamma)(|\theta - a|^2 + \eta^2) \quad \text{for all } (\theta, \eta) \in \Omega,$$

where $c(\Gamma)$ depends only on Γ. Letting $\gamma_\theta(\lambda)$ be the shortest constant speed geodesic on $\mathbb{S}^n$ going from $\gamma_\theta(0) = a$ to $\gamma_\theta(1) = \theta$, it follows that, for $0 < \lambda < 1$, the scaled function v_λ, defined by $v_\lambda(\theta, \eta) = \lambda^{-1} v(\gamma_\theta(\lambda), \lambda\eta)$, satisfies

$$|v_\lambda(\theta, \eta)| \leqq c(\Gamma) \lambda (|\theta - a|^2 + \eta^2).$$

Moreover, since v_λ locally minimizes an integral of the type

$$\int (1 + |\nabla v|^2)^{\frac{1}{2}} \lambda^n (\sin \lambda\eta)^{-n} \cos \eta \, d\theta \, d\eta,$$

we infer from the interior gradient estimates [GT, 15.2] that

$$|\nabla v_\lambda(\theta,\eta)| \leqq c(n,\Gamma)\lambda \quad \text{and} \quad |\nabla^2 v_\lambda(\theta,\eta)| \leqq c(n,\Gamma)\lambda$$

for $0<\lambda<\frac{1}{2}\delta$, $|\theta-a| \leqq 1$, and $\frac{1}{2} \leqq \eta \leqq 1$. For such λ, θ, η, this means that

$$|\nabla v(\lambda\theta,\lambda\eta)| \leqq c(n,\Gamma)\lambda \quad \text{and} \quad |\nabla^2 v(\lambda\theta,\lambda\eta)| \leqq c(n,\Gamma),$$

hence,

$$|\nabla v(\theta,\eta)| \leqq C(n,\Gamma)(|\theta-a|+\eta) \quad \text{and} \quad |\nabla^2 v(\theta,\eta)| \leqq C(n,\Gamma)$$

for $|\theta-a| \leqq \frac{1}{2}\delta$, and $\frac{1}{2}|\theta-a| \leqq \eta \leqq \frac{1}{2}\delta$. Since these estimates are independent of a, the function v, and hence u, is $\mathscr{C}^{1,1}$ in a neighborhood of $(a,0)$ in Ω, and $\Gamma \cup N$ is $\mathscr{C}^{1,1}$ near $(a,0)$.

To treat the $\mathscr{C}^{1,\alpha}$ case with $0<\alpha<1$, we again choose u and v as above. Now by § 1 the two closed balls in $\mathbb{R}^n \times \mathbb{R}$,

$$\mathbb{B}_{b(a,r)}(0,\ldots,0,1-r,0) \quad \text{and} \quad \mathbb{B}_{b(a,r)}(0,\ldots,0,1+r,0),$$

which are no longer tangent to Γ at a, do not intersect M. Taking the envelopes of these balls over all small positive r, we obtain functions $v_\pm$ on Ω which have $v_- \leqq v \leqq v_+$ and which satisfy, by § 1(iii) and the argument in § 1, estimates

$$\sup_{|\theta-a|+|\eta|<r} |v_\pm(\theta,\eta)| \leqq c(\Gamma) r^{1+\alpha}.$$

Following the $\mathscr{C}^{1,1}$ argument above, we now infer that

$$|\nabla v(\theta,\eta)| \leqq C(n,\Gamma)(|\theta-a|+\eta)^\alpha$$

and

$$|\nabla^2 v(\theta,\eta)| \leqq C(n,\Gamma)(|\theta-a|+\eta)^{\alpha-1}$$

for $|\theta-a| \leqq \frac{1}{2}\delta$, and $0 \leqq \eta \leqq \frac{1}{2}\delta$. Now, v, and hence u, is $\mathscr{C}^{1,\alpha}$ in a neighborhood of $(a,0)$ in Ω, and $\Gamma \cup N$ is $\mathscr{C}^{1,\alpha}$ near $(a,0)$. □

§ 4. Star-shaped domains at ∞

4.1. Theorem. *Suppose Ω is a bounded $\mathscr{C}^1$ star-shaped domain in $\mathbb{R}^n$ and $\Gamma = \partial\Omega \times \{0\}$. There exists a unique, up to multiplicity, n dimensional hyperbolic-stationary locally rectifiable current T in $\mathbb{H}$ such that $\Gamma = \bar{M} \sim M$ where $M = \operatorname{spt} T$. Moreover, T is hyperbolic-area minimizing, and M is the polar graph of a function defined on an open hemisphere centered in $\Omega \times \{0\}$.*

Proof. As noted in the introduction, M. Anderson showed, in [A_1, Theorem 3], the existence of a hyperbolic-area minimizing T with $\Gamma = \bar{M} \sim M$ where $M = \operatorname{spt} T$. We will first verify that this particular M is a completely regular polar graph.

To see this, we first apply a suitable translation to get Ω to be star-shaped about the origin. We may choose positive numbers r and s so that

$$\mathbb{B}^{n-1}_r(0) \subset \Omega \subset \bar{\Omega} \subset \mathbb{B}^{n-1}_s(0).$$

Then, by (1.1), $M \subset \mathbb{B}_s(0) \sim \bar{\mathbb{B}}_r(0)$; hence,

$$\left(\frac{r}{s}\right) M \cap M = \emptyset = M \cap \left(\frac{s}{r}\right) M.$$

We claim that $(\lambda M) \cap \mathrm{M} = \emptyset$ for all positive $\lambda \neq 1$. Otherwise, we could choose λ with either $\frac{r}{s} < \lambda < 1$ or $1 < \lambda < \frac{s}{r}$ so that the sets of regular points, $\mathrm{Reg}(M)$ and $\mathrm{Reg}(\lambda M)$, intersect transversally. Near $\mathbb{R}^n \times \{0\}$ the sets M and λM are disjoint because Ω is star-shaped about the origin. One may then, by minimality, replace a piece of T by the piece of $\mu_{\lambda\#} T$ that is cut off by T and still have a minimizing current. But then the singular set of the resulting current would include the intersection $\mathrm{Reg}(M) \cap \mathrm{Reg}(\lambda M)$, whose $n-1$ dimensionality contradicts the $n-7$ dimensional bound on the singular set. From this claim, it follows that M is a polar graph of a function u on $\mathbb{S}^n_+$. Moreover, by the small size of $\mathrm{Sing}(M)$, all the singularities of the function u are, as in the proof of [HS, 2.1], removable, and u is analytic on $\mathbb{S}^n_+$.

Next we assume that S is any n dimensional hyperbolic-stationary locally rectifiable current in $\mathbb{H}$ with $\overline{\mathrm{spt}\, S} \sim \mathrm{spt}\, S = \Gamma$. By the argument in § 1, we again find that

$$\mathrm{spt}\, S \subset \mathbb{B}_s(0) \sim \bar{\mathbb{B}}_r(0).$$

Let $\lambda_+ = \inf\{\lambda : \lambda M \cap \mathrm{spr}\, S \neq \emptyset\}$. If λ_+ were strictly greater than one, then we could (because Ω is star-shaped about the origin) find a point $b \in \lambda M \cap \mathrm{spt}\, S$. By the regularity and connectedness of M and the argument of [H, 4.4], we would find that $\lambda M \subset \mathrm{spt}\, S$, contradicting that $\overline{\lambda M} \sim \lambda M \subset \lambda \Gamma$ and $\lambda \Gamma \cap \Gamma = \emptyset$. Thus $\lambda_+ \leqq 1$. Similarly, $\sup\{\lambda : \lambda M \cap \mathrm{spt}\, S \neq \emptyset\} \geqq 1$. Thus $\mathrm{spt}\, S \subset M$, and S is, by the constancy theorem [F, 4.1.4], a multiple of T. □

References

[A_1] Anderson, M.: Complete minimal varieties in hyperbolic space. Invent. math. **69**, 477–494 (1982)

[A_2] Anderson, M.: Complete minimal hypersurfaces in hyperbolic n-manifolds. Comment. Math. Helv. **83**, 264–290 (1983)

[F] Federer, H.: Geometric measure theory. Berlin, Heidelberg, New York: Springer 1969

[GT] Gilbarg, D., Trudinger, N.: Elliptic partial differential equations of second order. Berlin, Heidelberg, New York: Springer 1977

[H] Hardt, R.: On boundary regularity for integral currents or flat chains modulo two minimizing the integral of an elliptic integrand. Comment. P.D.E. **2**, 1163–1232 (1977)

[HS] Hardt, R., Simon, L.: Area minimizing hypersurfaces with isolated singularities. J. Reine Angew. Math. **362**, 102–129 (1985)

[L] Lin, F.-H.: On the Dirichlet problem for minimal graphs in hyperbolic space. Preprint, MSRI

[LR] Levitt, G., Rosenberg, H.: Symmetry of constant mean curvature hypersurfaces in hyperbolic space. Duke Math. J. **52**, 53–59 (1985)

[M] Morrey, C.B., Jr.: Multiple integrals in the calculus of variations. Berlin, Heidelberg, New York: Springer 1966

Oblatum 14-VII-1986

Invent. math. 88, 225–241 (1987)

Inventiones mathematicae

© Springer-Verlag 1987

Birkhoff periodic orbits for small perturbations of completely integrable Hamiltonian systems with convex Hamiltonians

David Bernstein[1] and Anatole Katok[2]*

[1] Department of Mathematics, Northeastern University, Boston, MA 02115, USA
[2] Department of Mathematics, California Institute of Technology, Pasadena, CA 91125, USA

Dedicated to the memory of C.C. Conley

1. Introduction

By far the most remarkable property of small perturbations of completely integrable Hamiltonian systems is the preservation of invariant tori corresponding to irrational frequency vectors which are not too well approximable by rationals. This fact and its various ramifications was discovered by A.N. Kolmogorov, V.I. Arnold and J. Moser and became commonly known as KAM theory. This theory leaves open the basic question, namely, what happens to the rest of the invariant tori of the unperturbed completely integrable system? It is relatively easy to show that generically most of those tori, both rational and irrational, disappear. Thus, a more precise formulation of the above question should be like this. Are there special sufficiently simple motions in the perturbed system which are similar to the periodic and quasi-periodic motions on the destroyed tori, and therefore can be viewed as "traces" or "ghosts" of those tori?

The study of small perturbations of non-degenerate completely integrable systems with two degrees of freedom can be reduced to the consideration of area-preserving twist maps of the annulus or the cylinder [15]. For such maps S. Aubry [4, 5] and J. Mather [17] (cf. also [13]) established the existence of special invariant sets which are projected injectively to the circle and carry motions with any given admissible rotation number. Furthermore, the map preserves the cyclic order of points on any of those invariant sets. For any irrational rotation number such set is either an invariant circle, or, if the invariant circle for the given rotation number does not exist, it is a Cantor set with the motion described by A. Denjoy [11]. Moreover, the Cantor sets are always accompanied by order-preserving orbits doubly asymptotic to them [6, 14, 18]. As might be expected, for any rational rotation number one has a collection of at least two order-preserving (Birkhoff) periodic orbits together with

* Partially supported by NSF Grant DMS 8414400

homo- or heteroclinic orbits [14, 18]. All those objects possess certain continuity properties with respect to the rotation number [18].

Translating the above-mentioned results to the case of small perturbations of non-degenerate completely integrable Hamiltonian systems one obtains a more than satisfactory solution to the problem of vanishing tori for systems with two degrees of freedom. Namely, there are always traces of those tori present in the form of either Denjoy type minimal sets together with double asymptotic orbits (for any admissible irrational rotation number), or of at least two Birkhoff periodic orbits accompanied by homo- or heteroclinic orbits (for any admissible rational rotation number). Furthermore, the Denjoy minimal sets appear as limits of Birkhoff periodic orbits.

The Aubry-Mather approach is based on two key ingredients: the variational principle for finding desired motions and the regularity of the projection of any order-preserving orbit to the circle. Let us point out that the variational principle for finding order-preserving periodic orbits for twist maps can be substituted by certain topological arguments [7, 12]. The regularity of the projection allows us to take limits with respect to rotation number, thus, producing invariant circles or Denjoy type Cantor sets without any use of invariant measures or variational methods.

On the other hand, solutions representing global minima in various variational problems, associated to a twist map and posed without assuming preservation of order, turn out to be order preserving [6].

The earliest result concerning the preservation of some periodic orbits for Hamiltonian systems with more than two degrees of freedom is the Birkhoff-Lewis theorem [8, 9], (cf. also [2]) whose accurate proof was given by Moser [19]. Applying the method from [9] and [19] to our situation, one can find periodic orbits for a perturbation of a completely integrable system in a neighborhood of an invariant torus of the unperturbed system filled by periodic orbits of period, say, T, if the size of the perturbation is so small, that the orbits of the perturbed system stay sufficiently close to the original orbits during time T. We do not see how that method can be made to work uniformly in T for perturbations of fixed size.

Conley and Zehnder [10] discovered a remarkable global method for finding periodic orbits for sympletic maps and Hamiltonian systems. In this paper we use a version of their main trick. The result of the [10] most relevant for our discussion is Theorem 3 which represents a global generalization of the Birkhoff-Lewis theorem. This theorem depends only on a sort of a boundary condition rather than on closeness of perturbed and unperturbed systems. However, it can not be directly applied to the finding of very long orbits for a perturbed system which stay near the original torus.

Any attempt to carry out the Aubry-Mather approach to the case of more than two degrees of freedom faces the obvious problem that the arguments based on the preservation of order are no longer available. On the other hand, the variational arguments can be used, at least under some extra assumptions on the unperturbed system. In the present paper we make the first modest, but we believe non-trivial, step in that direction. Our main result for systems with n degrees of freedom is the existence of at least n distinct periodic orbits with any admissible rational frequency vector near the corresponding torus of the

unperturbed system under the extra assumption that the Hamiltonian of the unperturbed system has convex energy surfaces in the action-angle variables.

The reduction to the discrete-time case described in Sect. 7 allows obtaining this result from the corresponding result for sympletic maps (Theorem A, Sect. 1). This latter theorem follows immediately from Proposition 2 (Sect. 4) which provides certain estimates for specific critical points of the Lagrangian introduced in Sect. 2, and from Proposition 4 (Sect. 5) which establishes the existence of required critical points.

In Sect. 6 we show that at least one of the orbits described in Theorem A satisfies certain regularity conditions (Theorem B) which allows one to take limits in frequency vectors (Theorem C). Unfortunately, the structure of the limit objects corresponding to irrational frequencies is not completely clear at that stage.

One of the crucial ingredients of our method is a global topological trick which is very similar to the one used by C.C. Conley and E. Zehnder in [10]. This trick allows us to show the existence of solutions of the variational problem with sufficiently low values of our Lagrangian. Then the estimates of the Lagrangian show that these solutions correspond to periodic orbits with desired properties.

We work with discrete time symplectic maps and in the last section show how the results about continuous time Hamiltonian systems are derived from those for symplectic maps.

2. Preliminaries and formulation of main result

Let us consider the space $M=\mathbb{T}^n\times\mathbb{R}^n=\{(\varphi_1 \dots \varphi_n, r_1 \dots r_n),\ \varphi_i\in\mathbb{R}/\mathbb{Z},\ r_i\in\mathbb{R}\}$ with the natural sympletic 2-form

$$\Omega=\sum_{i=1}^{n} d\varphi_i\wedge dr_i$$

and let $f_0\colon \mathbb{T}^n\times U\to\mathbb{T}^n\times U$ be an integrable symplectic diffeomorphism, i.e. an Ω-preserving diffeomorphism of the form

$$f_0(\varphi, r)=(\varphi+\alpha(r), r), \qquad \varphi\in\mathbb{T}^n,\ r\in U.$$

Here $U\subset\mathbb{R}^n$ is a set diffeomorphic to an open n-disc.

Let, furthermore $F_0\colon \mathbb{R}^n\times U\to\mathbb{R}^n\times U$ be a lift of f_0 to the universal cover, so that for $x\in\mathbb{R}^n$, $r\in U$

$$F_0(x, r)=(x+a(r), r). \tag{1}$$

Throughout this paper we will assume the following non-degeneracy condition

(i) *$a\colon U\to\mathbb{R}^n$ is a regular injective map.*

Then the map F_0 can be represented via a generating function $H_0(x, x')$ so that if $F_0(x, r)=(x', r')$ then

$$r=\frac{\partial H_0}{\partial x}, \qquad r'=-\frac{\partial H_0}{\partial x'} \tag{2}$$

It follows immediately from (1) and (2) that the function H_0 actually depends only on the difference $x'-x$, $H_0(x, x')=h(x'-x)$. Let $b: a(U)\to U$ be the map inverse to a. Then from (2)

$$dh(\delta)=b(\delta)\,d\delta.$$

Condition (i) is a discrete-time equivalent of the standard non-degeneracy condition in KAM theory, cf. e.g. [1] or [3], Apprendix 8. Next we will introduce an additional non-trivial restriction which will also be assumed in all subsequent considerations, usually without any separate mentioning.

(ii) *h is a strictly convex function on $a(U)$, i.e. the Hessian of h at every point $\delta\in a(U)$ is a positive definite quadratic form.*

In Sect. 7 we will interpret conditions (i) and (ii) in terms of continuous time Hamiltonian systems.

Suppose that a lift F of f can be represented by a generating function $H(x, x')$ so that $F(x, r)=(x', r')$ if and only if

$$\frac{\partial H}{\partial x}(x, x')=r \qquad \frac{\partial H}{\partial x'}(x, x')=-r'.$$

Suppose in addition that the perturbation f preserves the r-component of the center of masses on each torus $\mathbb{T}^n\times\{r_0\}$ for $r_0\in U$, or, equivalently, that for any $m\in\mathbb{Z}^n$

$$H(x+m, x'+m)=H(x, x'). \tag{3}$$

So we can write $H(x, x')=h(x'-x)+P(x, x')$ where P satisfies (3). In order for $H(x, x')$ to exist, it suffices to assume that f is C^1-close to f_0. Then H is C^2-close to H_0. Conversely, if $H(x, x')$ is a small C^2-perturbation of H_0 then it defines a small C^1-perturbation of the map f_0. However, our results depend only on the smallness of the C^1 size of the perturbation $P(x, x')$ of the generating function. So it suffices to assume that f is C^0 close to f_0 provided that f can be represented by a generating function.

Let $r_0\in U$ and $s_0=a(r_0)$. We will study orbits of the map f which stay sufficiently close to the torus $\mathbb{T}^n_{r_0}=\mathbb{T}^n\times\{r_0\}$.

Our first main result establishes the existence of such orbits if the vector s_0 has rational coordinates.

Let $(\varphi, r)\in\mathbb{T}^n\times\mathbb{R}^n$ be a periodic orbit of the map f with the prime period q and let $(x, r)\in\mathbb{R}^n\times\mathbb{R}^n$ be a lift of the point (φ, r). Then there exists a vector $w\in\mathbb{Z}^n$ such that $F^q(x, r)=(x+w, r)$. We will call the vector $\frac{w}{q}$ the *rotation vector* of the point (φ, r). The rotation vector depends on the choice of the lift F but it is uniquely defined modulo $\mathbb{Z}^n$.

Theorem A. *Let f be a perturbation of an integrable symplectic map f_0 satisfying* (i) *and* (ii). *Let $w=(w_1 \dots w_n)\in\mathbb{Z}^n$, q be a positive integer such that $w_1 \dots w_n, q$ are relatively prime and the vector $\frac{w}{q}$ belongs to $a(U)$. Let furthermore $r_{w,q}=a^{-1}\left(\frac{w}{q}\right)$. There exists a constant Δ depending on f_0 but not on w and q such that for any $\delta<\Delta$ if the map f is defined by the generating function $H=h+P$*

where the C^1 norm of the perturbation part P of H is equal to δ, then the map f has at least $n+1$ different periodic orbits with rotation vector $\frac{w}{q}$ which lie completely inside the $C\delta^{\frac{1}{2}}$ neighborhood of the torus $\mathbb{T}^n \times \{r_{w,q}\}$ and at least one of those orbits lies inside the $C\delta^{\frac{1}{2}}$ neighborhood of that torus. Here C depends only on the unperturbed map f_0.

This theorem follows immediately from Proposition 2 which is proved in Sect. 4 and Proposition 4 proved in Sect. 5.

We conclude this section with the description of a reduction of our problem to another one which is defined globally in $\mathbb{T}^n \times \mathbb{R}^n$ and which coincides with our problem in a neighborhood of $\mathbb{T}^n_{r_0}$. We use the word "problem" instead of "map" because we are going to modify the generating function H and we do not care whether the perturbed generating function defines a map.

Let V be a neighborhood of δ_0 such that the function h in V is sufficiently close to its second Taylor polynomial T_2. By the convexity assumption (ii), T_2 is a convex second degree polynomial which can be defined in $\mathbb{R}^n$. Thus, it is easy to see that one can construct a C^2 small function $\bar{q}$ on $\mathbb{R}^n$ which coincides with $h - T_2$ on V, and vanishes outside of a compact set. Therefore, $\bar{h} \stackrel{\text{def}}{=} T_2 + \bar{q}$ is a strictly convex function which coincides with h in V and with T_2 outside of a compact set. Similarly we modify the generating function $H(x, x') = h(x' - x) + P(x, x')$ of the diffeomorphism f into a function $\bar{H}(x, x') = \bar{h}(x' - x) + \bar{P}(x, x')$ where the function $\bar{P}$ is uniformly C^1 small, coincides with P for $x' - x \in V$ and vanishes when $x' - x$ lies outside of a certain compact set. Naturally, we can make $\bar{H}$ satisfy periodicity condition (3).

Obviously, if $x' - x \in V$ then setting

$$r = \frac{\partial \bar{H}}{\partial x}, \qquad r' = -\frac{\partial \bar{H}}{\partial x'}$$

we have $F(x, r) = (x', r')$ because locally $\bar{H} = H$. Furthermore if $x_i \in \mathbb{R}^n$ is a finite or infinite sequence of vectors such that

$$\frac{\partial \bar{H}}{\partial x}(x_i, x_{i+1}) = -\frac{\partial \bar{H}}{\partial x'}(x_{i-1}, x_i) \stackrel{\text{def}}{=} r_i$$

and

$$x_{i+1} - x_i \in V \quad \text{for all } i \tag{4}$$

then the sequence (x_i, r_i) is an orbit or an orbit segment for F.

3. Periodic states with given rotation vector

Let us fix $w = (w_1 \dots w_n) \in \mathbb{Z}^n$ and $q \in \mathbb{Z}_+$ such that $w_1, \dots, w_n, q$ are relatively prime. Let us consider the space $\Psi_{w,q}$ of all double-infinite sequences $x = (\dots, x_{-1}, x_0, x_1, \dots)$ of vectors from $\mathbb{R}^n$ satisfying the following periodicity condition

$$x_{i+q} = x_i + w \quad \text{for all } i \in \mathbb{Z}. \tag{5}$$

Every sequence $\{x_i\}$ is uniquely determined by a q-tuple of vectors $(x_1, \ldots, x_q)$, so $\Psi_{w,q}$ can be naturally identified with $(\mathbb{R}^n)^q$. We introduce two kinds of identifications in $\Psi_{w,q}$. First, for any $m \in \mathbb{Z}^n$ we identify $\{x_i\}$ with $\{x_i + m\}$. Secondly we identify every sequence $\{x_i\}$ with its shift $\{y_i\}$, where $y_i = x_{i+1}$. In other words, our identifications are generated by the translations $T_m: \{x_i\} \mapsto \{x_i + m\}$ and the shift $S: \{x_i\} \mapsto \{x_{i-1}\}$. The quotient space of $\Psi_{w,q}$ corresponding to the first identification will be denoted by $\Phi^*_{w,q}$, and the result of both identifications will be denoted by $\Phi_{w,q}$.

A convenient coordinate system in $\Psi_{w,q}$ is given by the parameters $(v, t) = (v, t_1, \ldots, t_{q-1})$ where

$$v = \frac{x_0 + \ldots + x_{q-1}}{q}, \qquad t_i = x_i - x_{i-1} - \frac{w}{q}, \qquad i = 1, \ldots, q-1.$$

In terms of these coordinates we have

$$T_m(v, t) = (v + m, t)$$
$$S(v, t_1, \ldots, t_{q-1}) = \left(v + \frac{w}{q}, t_2, \ldots, t_{q-1}, -t_1 - t_2 - \ldots - t_{q-1}\right). \tag{6}$$

Thus, the space $\Phi^*_{w,q}$ is diffeomorphic to $T^m \times \mathbb{R}^{n(q-1)}$ and it represents a q-fold covering of $\Phi_{w,q}$. By (6) the latter space is an $\mathbb{R}^{n(q-1)}$ bundle over the torus $\mathbb{T}^n$ and thus it is homotopically equivalent to $\mathbb{T}^n$.

We define the function $L_{w,q}$ on $\Psi_{w,q}$ by

$$L_{w,q}(x) = \sum_{i=1}^{q} \bar{H}(x_i, x_{i+1}). \tag{7}$$

Obviously, the function $L_{w,q}$ is both T_m and S invariant, so it defines a function on $\Phi_{w,q}$ which we denote by the same symbol $L_{w,q}$ and will sometimes call the Lagrangian.

A point $x \in \Phi_{w,q}$ is called *an equilibrium state* if x is a critical point of $L_{w,q}$. Such a state must satisfy the following conditions:

$$0 = \frac{\partial L_{w,q}}{\partial x_i} = \frac{\partial \bar{H}(x_i, x_{i+1})}{\partial x} + \frac{\partial \bar{H}(x_{i-1}, x_i)}{\partial x'}.$$

If an equilibrium $\{x_i\}$ satisfies additional conditions: $x_{i+1} - x_i \in V$ for all i then, as we noted at the end of Sect. 2, it corresponds to an orbit of F which by the periodicity condition (5) can be projected onto a periodic orbit of f with rotation vector $\frac{w}{q}$.

4. Critical points of the Lagrangian $L_{w,q}$

Proposition 1. *$L_{w,q}$ is a proper function on $\Phi_{w,q}$ and is bounded from below.*

Proof. Since $\bar{H}(x, x')$ is bounded from below and goes to $+\infty$ as $|x' - x| \to \infty$, the statement follows from (7). □

Thus we have the following

Corollary 1. *The function $L_{w,q}$ reaches its absolute minimum l_0 at some equilibrium state x^0.*

In our next statement we will show that if the rotation vector $\frac{w}{q}$ is sufficiently close to $a(r_0)$, $\left(\text{in particular, if } a(r_0)=\frac{w}{q}\right)$, then this equilibrium state satisfies the condition (4) and determines a periodic orbit of f with rotation vector $\frac{w}{q}$ which lies near the torus $\mathbb{T}^n_{r_0}$. Moreover, under a somewhat stronger assumption on the size of the perturbation we can prove (4) for every critical point x of $L_{w,q}$ with the critical value less than l_0+C, where C does not depend on w or q.

We begin with some general estimates. We will use the letter C with various indices to denote various unspecified constants which may depend on the map f_0 but not on w or q. Let $x=\{x_i\}$ be an arbitrary state, set $a_i=x_i-x_{i-1}$ and let $\delta_k (k=0, 1)$ be the C^k-norm of the perturbation $\bar{P}$ of the generating function. We will assume throughout the rest of the paper that $\delta_k<\mathscr{D}$, $k=0, 1$ where the constant $\mathscr{D}$ is chosen once and for all for a given map f_0.

Lemma 1. *If x is a critical point of $L_{w,q}$ then $|a_{i+1}-a_i|<C\delta_1$ for every i. Besides, if x is a point of absolute minimum for $L_{w,q}$ then $|a_{i+1}-a_i|<C\delta_0^{\frac{1}{2}}$.*

Proof. Let $a_i=a+\tau$, $a_{i+1}=a-\tau$, and let $y=\frac{1}{2}(x_{i+1}+x_{i-1})$ so that $x_i=y+\tau$. Consider a family of states $x(\sigma)$, $\sigma\in\mathbb{R}^n$ defined by

$$x_j(\sigma)=\begin{cases} x_j & \text{for } j\not\equiv i \pmod q \\ \frac{1}{2}(x_{j+1}+x_{j-1})+\sigma & \text{for } j\equiv i \pmod q. \end{cases}$$

We have

$$\begin{aligned} &L_{w,q}(x(\sigma))-L_{w,q}(x) \\ &\quad=\bar{H}(x_{i-1}, y+\sigma)+\bar{H}(y+\sigma, x_{i+1})-\bar{H}(x_{i-1}, y+\tau)-\bar{H}(y+\tau, x_{i+1}) \\ &\quad=\bar{P}(x_{i-1}, y+\sigma)+\bar{P}(y+\sigma, x_{i+1})-\bar{P}(x_{i-1}, y+\tau)-\bar{P}(y+\tau, x_{i+1}) \\ &\qquad+\bar{h}(a+\sigma)+\bar{h}(a-\sigma)-\bar{h}(a+\tau)-\bar{h}(a-\tau). \end{aligned} \tag{8}$$

The sum of the first four terms is bounded by $4\delta_0$. If x is a point of absolute minimum then for $\sigma=0$

$$L_{w,q}(x(0))-L_{w,q}(x)\geqq 0$$

so we have

$$\bar{h}(a+\tau)+\bar{h}(a-\tau)-2\bar{h}(a)\leqq 4\delta_0.$$

On the other hand, by the convexity of $\bar{h}$

$$\bar{h}(a+\tau)+\bar{h}(a-\tau)-2\bar{h}(a)\geqq C_1|\tau|^2.$$

Therefore, $C_1|\tau|^2\leqq 4\delta_0$ and $|a_{i+1}-a_i|<C_2\delta_0^{\frac{1}{2}}$. This proves the second statement of the lemma.

If x is a critical point then $\left(\frac{\partial}{\partial\sigma} L_{w,q}(x(\sigma))|_{\sigma=\tau}\right)=0$. The differential of the sum of the first four terms in (8) at $\sigma=\tau$ is bounded by $2\delta_1$. The last two terms do not depend on σ and the differntial of $\bar{h}(a+\sigma)+\bar{h}(a-\sigma)$ at $\sigma=\tau$ equals $d\bar{h}(a+\tau)-d\bar{h}(a-\tau)$. Let $Q(s)$ be the matrix of second derivatives of $\bar{h}$ at s. By our construction of the function $\bar{h}$ (cf. Sect. 2) $\langle Q(s)\tau,\tau\rangle \geqq C_3|\tau|^2$. On the other hand by the Mean Value theorem $\langle d\bar{h}(a+\tau)-d\bar{h}(a-\tau),\tau\rangle=\langle Q(a+\lambda\tau)2\tau,\tau\rangle$ for some $\lambda\in(-1,1)$. Therefore, $|d\bar{h}(a+\tau)-d\bar{h}(a-\tau)|\geqq 2C_3|\tau|$.

Combining all terms in (8) we have $C_3|\tau|<\delta_1$ or $|a_{i+1}-a_i|<C_4\delta_1$. □

Lemma 2. *Suppose x is a point of absolute minimum for $L_{w,q}$. Then for any i, j*

$$|a_i-a_j|<C\min(\delta_1^{\frac{1}{2}},\delta_0^{\frac{1}{2}}).$$

Proof. Let $\beta_0=\max_l|a_l-a_{l+1}|$, β be any number greater than β_0, $\tau=|a_i-a_j|$ and suppose that k is an integer such that $k\beta<\tau/5$. We can assume that $i<j\leqq i+\frac{q}{2}$. Then the sets $\{i,\ldots,i+k\}$ and $\{j,\ldots,j+k\}$ are disjoint mod q so $i_1=i+k<j$ and $j_1=j+k<i+q$. Next we perform a sort of surgery and construct for every $m\in\mathbb{Z}^n$ a state $y(m)$ such that

$$y_l(m)=\begin{cases} x_i+\dfrac{l-i}{k}\,[x_{i_1}-m-x_i] & i\leqq l\leqq i_1\\ x_l-m & i_1\leqq l\leqq j\\ x_j-m+\dfrac{l-j}{k}\,[x_{j_1}-(x_j-m)] & j\leqq l\leqq j_1\\ x_l & j_1\leqq l\leqq i+q\end{cases}$$

and extend it for all l by periodicity.

We will try to choose m to make the vectors $(y_{i_1}-y_i)$ and $(y_{j_1}-y_j)$ as close to each other as possible. For that purpose choose as m any of the integer lattice vectors nearest to $\frac{1}{2}(x_{i_1}-x_i-x_{j_1}+x_j)$. Let us estimate the difference $L_{w,q}(x)-L_{w,q}(y(m))$ from below. This difference is equal to

$$\begin{aligned}&\sum_{l=1}^{k}[\bar{H}(x_{i+l},x_{i+l-1})+\bar{H}(x_{j+l},x_{j+l-1})\\&\quad-\bar{H}(y_{i+l},y_{i+l-1})-\bar{H}(y_{j+l},y_{j+l-1})]\\&=\sum_{l=1}^{k}[\bar{P}(x_{i+l},x_{i+l-1})+\bar{P}(x_{j+l},x_{j+l-1})\\&\quad-\bar{P}(y_{i+l},y_{i+l-1})-\bar{P}(y_{j+l},y_{j+l-1})]\\&\quad+\sum_{l=1}^{k}\left[\bar{h}(a_{i+l})+\bar{h}(a_{j+l})-\bar{h}\left(\frac{y_{i_1}-y_i}{k}\right)-\bar{h}\left(\frac{y_{j_1}-y_j}{k}\right)\right].\end{aligned}$$

The absolute value for the first sum is obviously bounded by $4k\delta_0$. Let us rewrite each term of the second sum. Let $a_{i+l}=b+\rho_l$ where b is the average of $\{a_{i+1},\ldots,a_{i+k}\}$ and $\sum\rho_l=0$. Similarly let $a_{j+l}=c+\sigma_l$, $\sum\sigma_l=0$. Obviously

$|\rho_l|<k\beta$, $|\sigma_l|<k\beta$. In particular, $\tau=|a_i-a_j|=|b-c+\rho_0-\sigma_0|$ and $|\tau-|b-c||\leqq 2k\beta\leqq\frac{2}{5}\tau$ so $|b-c|>\frac{\tau}{2}$. Also

$$\frac{y_{i_1}-y_i}{k}=\frac{b+c}{2}+\zeta \quad \text{and} \quad \frac{y_{j_1}-y_j}{k}=\frac{b+c}{2}-\zeta \quad \text{where } |\zeta|<\frac{d}{k}$$

(d depends only on the dimension n of $\mathbb{T}^n$). Furthermore

$$\sum_{l=1}^{k}\bar{h}(a_{i+l})=k\bar{h}(b)+\langle d\bar{h}(b),\sum\rho_l\rangle+O(k(k\beta)^2)=k\bar{h}(b)+O(k^3\beta^2).$$

Similarly

$$\sum_{l=1}^{k}\bar{h}(a_{j+l})=k\bar{h}(c)+O(k^3\beta^2),$$

$$\bar{h}\left(\frac{b+c}{2}+\zeta\right)+\bar{h}\left(\frac{b+c}{2}-\zeta\right)=2\bar{h}\left(\frac{b+c}{2}\right)+O(k^{-2}).$$

By the convexity of $\bar{h}$

$$\bar{h}(b)+\bar{h}(c)-2\bar{h}\left(\frac{b+c}{2}\right)>C_1\left|\frac{b-c}{2}\right|^2>C_5\tau^2.$$

Combining all these inequalities we have

$$L_{w,q}(x)-L_{w,q}(y)>C_5k\tau^2-4k\delta_0-C_6k^3\beta^2-C_7k^{-1}. \tag{9}$$

Since x is a point of absolute minimum

$$\tau^2<C_8\delta_0+C_9k^2\beta^2+C_{10}k^{-2}. \tag{10}$$

Now we are ready to complete the proof. By Lemma 1 we can choose $\beta=C_4\delta_1$ or $C_2\delta_0^{\frac{1}{2}}$. Take $k\approx\beta^{-\frac{1}{2}}$. If $k\beta>\tau/5$ then $\tau<C_{11}\beta^{\frac{1}{2}}$ and for $\beta=C_4\delta_1$ (resp. $C_2\delta_0^{\frac{1}{2}}$) $\tau<C_{12}\delta_1^{\frac{1}{2}}$ (resp. $\tau<C_{13}\delta_0^{\frac{1}{4}}$). If however $k\beta<\tau/5$ then we can apply the surgery described above and from (10) obtain the estimate $\tau^2<C_{14}\beta$ and $\tau<C_{15}\beta^{\frac{1}{2}}$ and we repeat the preceding argument. □

Lemma 3. *For every $E>0$ there is $C>0$ such that if x is a critical point of $L_{w,q}$ with $L_{w,q}(x)<E+\inf(L_{w,q})$ then for any i,j $|a_i-a_j|<C\delta_1^{\frac{1}{3}}$.*

Proof. By Lemma 1 we can choose $\beta=C_4\delta_1$. As before $\tau=|a_i-a_j|$. Take $k\approx\beta^{-\frac{2}{3}}$. If $k\beta>\tau/5$ then $\tau<C_{16}\delta_1^{\frac{1}{3}}$ and we are done. Otherwise we can apply the surgery from the proof of Lemma 2 and from (9) obtain the estimate:

$$\tau^2<(C_{17}E+C_{18})\beta^{\frac{2}{3}}.$$

Since $\beta=C_4\delta_1$ we have $\tau<C_{19}\delta_1^{\frac{1}{3}}$. □

Proposition 2. *Suppose $\frac{w}{q}\in V$ and $r_{w,q}=a^{-1}\left(\frac{w}{q}\right)$. Let δ be the C^1-norm of the perturbation $\bar{P}$ of the generating function. Then*

a) *any point in $\Phi_{w,q}$ of absolute minimum for $L_{w,q}$ determines a periodic orbit with rotation vector $\frac{w}{q}$ which lies completely inside the ε-neighborhood of the torus $\mathbb{T}^n \times \{r_{w,q}\}$, where $\varepsilon = C\delta^{\frac{1}{2}}$.*

b) *for any $E>0$ there is $\Delta>0$, independent on w and q, such that if $\delta<\Delta$ then any critical point $x \in \Phi_{w,q}$ of $L_{w,q}$ with $L_{w,q}(x)<E+\inf(L_{w,q})$ determines a periodic orbit with rotation vector $\frac{w}{q}$ which lies completely inside the ε-neighborhood of the torus $T^n \times \{r_{w,q}\}$ where $\varepsilon = C\delta^{\frac{1}{3}}$.*

Proof. Let x be a critical point of $L_{w,q}$ satisfying conditions of a) or b). Let

$$r_i = \frac{\partial \bar{H}}{\partial x}(x_i, x_{i+1}) = d\bar{h}(a_{i+1}) + \frac{\partial \bar{P}}{\partial x}(x_i, x_{i+1}).$$

By Lemmas 2 and 3 for any i, j $|a_i - a_j| < \varepsilon$. Since $w = \sum_1^q a_i$ we have $\left|a_i - \frac{w}{q}\right| < \varepsilon$. $\frac{\partial \bar{P}}{\partial x} = O(\delta)$ by definition of δ. But $\delta = O(\varepsilon)$ so $\frac{\partial \bar{P}}{\partial x} = O(\varepsilon)$. On the other hand $d\bar{h}\left(\frac{w}{q}\right) = r_{w,q}$ and $d\bar{h}(a_{i+1}) - d\bar{h}\left(\frac{w}{q}\right) = O\left(a_{i+1} - \frac{w}{q}\right) = O(\varepsilon)$. Therefore $|r_i - r_{w,q}| = O(\varepsilon)$. □

5. Existence of $(n+1)$ critical points

Let us recall that we denoted the quotient space of $\Psi_{w,q}$ modulo the group of integral translations by $\Phi^*_{w,q}$. The function $L_{w,q}$ is naturally defined on $\Phi^*_{w,q}$. Let as before $l_0 = \inf(L_{w,q})$ and let

$$M^*_t = \{x \in \Phi^*_{w,q} \mid L_{w,q}(x) \leqq t\},$$
$$M_t = \{x \in \Phi_{w,q} \mid L_{w,q}(x) \leqq t\}.$$

Proposition 3. *For a given $\bar{h}$ and δ_0 there exists a constant E such that for any perturbation $\bar{P}$ uniformly bounded by δ_0 there exists a continuous map $\Gamma: \mathbb{T}^n \to M^*_{l_0+E}$, such that for any w and q the composition of Γ with the inclusion $i_1: M^*_{l_0+E} \to \Phi^*_{w,q}$ is a homotopy equivalence between $\mathbb{T}^n$ and $\Phi^*_{w,q}$.*

Proof. We begin with defining some preliminary tools. Let

$$\varphi(t) = \begin{cases} 1, & \text{if } t - kq \in [0,1] \text{ for some } k \in \mathbb{Z} \\ 0, & \text{otherwise} \end{cases},$$

$$a_i(t) = \int_{i-t}^{i} \varphi(t)\, dt, \qquad i \in \mathbb{Z}.$$

It can be easily seen that

(a) $a_i(t+q) = a_i(t) + 1$

(b) $a_{i+q}(t) = a_i(t)$

(c) $a_i(t)=a_{i+1}(t)\in\mathbb{Z}$ for all but at most three values of $i\in[1, q]$ and for these exceptional i's $|a_i(t)-a_{i+1}(t)|\leqq 1$.

Let furthermore $A(t)=\{a_i(t),\ i\in\mathbb{Z}\}$ be the sequence of numbers $a_i(t)$. For any vector $z=(t_1, \ldots, t_n)\in\mathbb{R}^n$ let $V(z)=\{v_i(z),\ i\in\mathbb{Z}\}\overset{\text{def}}{=}\sum_{j=1}^{n} A(t_j)\,e_j$ where $e_1, \ldots, e_n$ is the standard basis of $\mathbb{R}^n$. So $V(z)$ is a family of double-infinite sequences of vectors depending on z as a parameter. Conditions (a), (b), (c) immediately imply that

(a′) $v_i(z+qm)=v_i(z)+m,\ m\in\mathbb{Z}^n$

(b′) $v_{i+q}(z)=v_i(z)$

(c′) $v_i(z)=v_{i+1}(z)\in Z^n$ for all but at most $3n$ values of $i\in\{1, \ldots, q\}$ and for those exceptional i each coordinate of $v_{i+1}(z)-v_i(z)$ does not exceed one in absolute value.

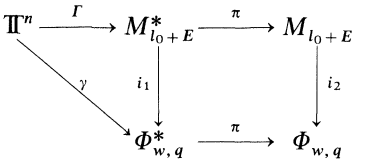
Now we are ready to define the map Γ. Let x be a point of absolute minimum of $L_{w,q}$ in $\Psi_{w,q}$. Let $x(z)$ be the following n-parameter family of states: $x(z)=x+V(z)$. By property (b′) $x(z)\in\Psi_{w,q}$. Denote the map $z\mapsto x(z)$ by $X:\mathbb{R}^n\to\Psi_{w,q}$. Furthermore by property (a′) the map $X:\mathbb{R}^n\to\Psi_{w,q}$ can be projected to a map $\Gamma:\mathbb{T}^n=\mathbb{R}^n/q\mathbb{Z}^n\to\Phi^*_{w,q}$. First we show that Γ is homotopic to the standard embedding $\Gamma_1: z\mapsto\{w_i(z)\}$ where $w_i(z)=\frac{z}{q}+i\frac{w}{q}$ and therefore effects a homotopy equivalence between $\mathbb{T}^n$ and $\Phi^*_{w,q}$. This homotopy is given by the explicit formula Γ_ε: $z\mapsto\{w_i(\varepsilon, z)\}$ where $w_i(\varepsilon, z)=(1-\varepsilon)\,x_i+\varepsilon\left(i\frac{w}{q}\right)+(1-\varepsilon)\,v_i(z)+\varepsilon\frac{z}{q}$. Next we show that $\Gamma(\mathbb{T}^n)\subset M^*_{l_0+E}$. To that end we estimate the difference $L_{w,q}(x(z))-L_{w,q}(x)=L_{w,q}(x(z))-l_0$. By property (c′) $\bar H(x_i(z), x_{i+1}(z))$ coincides with $\bar H(x_i, x_{i+1})$ for all but $3n$ values of $i\in\{1, \ldots, q\}$. Let i be one of the exceptional values. Then all coordinates of $v_i(z)-v_{i+1}(z)$ are less than or equal to one in absolute value. Since by Lemma 2 for a given δ_0 the differences $|x_{i+1}-x_i|$ are uniformly bounded, $|x_{i+1}(z)-x_i(z)|=|(x_{i+1}-x_i)+(v_{i+1}(z)-v_i(z))|$ are also uniformly bounded and

$$|\bar H(x_i(z), x_{i+1}(z))-\bar H(x_i, x_{i+1})|\leqq|\bar H(x_i(z), x_{i+1}(z))|+|\bar H(x_i, x_{i+1})|\leqq C.$$

Therefore $|L_{w,q}(x(z))-L_{w,q}(x)|\leqq 3nC$. □

Proposition 4. *For given $\bar h$ and δ_0 there exists a constant E independent of w, q such that for any perturbation $\bar P$ uniformly bounded by δ_0 there are at least $n+1$ different critical points of $L_{w,q}$ in $\Phi_{p,w}$ with critical values less than l_0+E.*

Proof. We will show that the Lusternik-Shnirelman category (for definition see, e.g. [16], Sect. 3 or [20], Chap. 5), of the set M_{l_0+E} is at least $n+1$. This implies the existence of $n+1$ critical points ([16], Sect. 3). The category of M is greater than k if there are $\omega_1 \ldots \omega_k\in H^1(M, \mathbb{Z})$ such that $\omega_1\cup\omega_2\cup\ldots\cup\omega_k\neq 0$ (cf. [20], Theorem 5.14). We will derive the existence of such cohomology classes from the following commutative diagram.



Here i_i, i_2 are inclusions, π is the projection, and Γ is the map constructed in Proposition 3. Let $\omega_1 \dots \omega_n$ be generators of $H^1(\Phi_{w,q}; \mathbb{Z})$. Then $\omega = \omega_1 \cup \dots \cup \omega_n$ is a generator of $H^n(\Phi_{w,q}; \mathbb{Z})$ and $\pi^*\omega = q\omega_0$ where ω_0 is a generator of $H^n(\Phi^*_{w,q}; \mathbb{Z})$. Here $\cup$ means cup-product in cohomology groups. Since γ is a homotopy equivalence we have $0 \neq \gamma^*(q\omega_0) = \gamma^*\pi^*\omega = \Gamma^*\pi^* i_2^*(\cup \omega_i) = \Gamma^*\pi^*(\cup(i_2^*\omega_i))$. So if we denote $i_2^*\omega_i$ by $\psi_i \in H^1(M_{l_0+E}, \mathbb{Z})$ and their cup product by ψ we have $\Gamma^*\pi^*\psi \neq 0$, thus $\psi \neq 0$. And the category of M_{l_0+E} is greater than n. □

6. Weak regularity of minimal orbits

Let ω: $[0, 1) \to \mathbb{R}_+$, $\omega(0) = 0$, be a modulus of continuity i.e. a non-decreasing continuous non-negative function which is strictly positive outside 0 and let $(\varphi_i, r_i) = f^i(\varphi_0, r_0) \in \mathbb{T}^n \times \mathbb{R}^n$ be an orbit of f. We say that this orbit is *ω-regular* if for any $i, j \in \mathbb{Z}$

$$|r_i - r_j| \leqq \omega(\mathrm{dist}(\varphi_i, \varphi_j)).$$

In particular, ω-regularity implies that two different points on the orbit do not have the same φ coordinates. If $\omega(t) = Lt$ for some constant L we call ω-regular orbits *Lipschitz regular*. For $n = 1$ all orbits corresponding to the absolute minimum of the Lagrangian are Lipschitz regular [13] with a fixed constant L which depends only on the map. As we mentioned in the introduction, regularity plays the key role in the whole theory of twist maps. We are not able to prove any ω-regularity for the minimal orbits for $n > 1$. However, a slightly weaker property does hold in that case.

Proposition 5. *Under the assumptions of Theorem A let $(\varphi_i, r_i) = f^i(\varphi_0, r_0)$, $(\varphi_q, r_q) = (\varphi_0, r_0)$ be any periodic orbit of f corresponding to a state which minimizes the functional $L_{w,q}$. Assume in addition that the second derivatives of H are bounded. Then there exist ε_0 and C depending only on f such that if i, j, k are different* mod q *and* $\max(\mathrm{dist}(\varphi_i, \varphi_j), \mathrm{dist}(\varphi_j, \varphi_k)) < \varepsilon_0$ *then*

$$|r_i - r_k| < C(\mathrm{dist}(\varphi_i, \varphi_j) + \mathrm{dist}(\varphi_i, \varphi_k))^{\frac{1}{2}}.$$

Proof. As usual we pass to the universal cover and consider the state $x = \{x_n\}_{n \in \mathbb{Z}} \in \Psi_{w,q}$ generating the given periodic orbit. We will assume that $i = 0$ and $0 < j < k < q$. The case $0 < k < j < q$ is considered similarly. From the conditions of the proposition we have for some $m_1, m_2 \in \mathbb{Z}^n$

$$|x_j + m_1 - x_k| < \varepsilon_0, \qquad |x_k + m_2 - x_q| < \varepsilon_0.$$

The idea of the proof is to rearrange the pieces $x_0, \dots, x_{j-1}$; $x_j, \dots, x_{k-1}$ and $x_k, \dots, x_{q-1}$ of the state x into a new state y and then estimate both the value

of the Lagrangian and its derivatives at y. Thus, we define

$$y_l = \begin{cases} x_l, & 0 \leqq l < j \\ x_{l+k-j} - m_1, & j \leqq l < j+q-k \\ x_{l+k-q} + m_2, & j+q-k \leqq l < q \end{cases}$$

and then extend y_l to other l according to (5) in order to obtain a state $y \in \Psi_{w,q}$.

Let us estimate $L_{w,q}(y)$ from above. We have

$$\begin{aligned} L_{w,q}(y) - l_0 &= L_{w,q}(y) - L_{w,q}(x) \\ &= H(x_{j-1}, x_k - m_1) - H(x_{j-1}, x_j) \\ &\quad + H(x_{k-1}, x_q - m_2) - H(x_{k-1}, x_k) \\ &\quad + H(x_{q-1}, x_j + m_1 + m_2) - H(x_{q-1}, x_q) \\ &< C_1(|x_j + m_1 - x_k| + |x_k + m_2 - x_q| + |x_q - m_1 - m_2 - x_j|) \\ &= C_1(\operatorname{dist}(\varphi_j, \varphi_k) + \operatorname{dist}(\varphi_k, \varphi_q) + \operatorname{dist}(\varphi_q, \varphi_j)) \\ &< C_2(\operatorname{dist}(\varphi_0, \varphi_j) + \operatorname{dist}(\varphi_0, \varphi_k)). \end{aligned} \tag{11}$$

On the other hand

$$\frac{\partial L_{w,q}}{\partial x_0}(y) = \frac{\partial \bar{H}}{\partial x'}(x_{k-1} - m_2 - m_1, x_0) + \frac{\partial \bar{H}}{\partial x}(x_0, x_1)$$

and

$$\frac{\partial \bar{H}}{\partial x'}(x_{k-1}, x_k) = -r_k, \quad \frac{\partial \bar{H}}{\partial x}(x_0, x_1) = r_0.$$

Since the second derivatives of H are bounded, we can assume that the second derivatives of $\bar{H}$ are bounded too. Thus, we have

$$\left| \frac{\partial L_{w,q}}{\partial x_0}(y) - (r_0 - r_k) \right| < C_3 |x_k - x_0 - m_1 - m_2| = C_3 \operatorname{dist}(\varphi_0, \varphi_k). \tag{12}$$

Using the fact that x is a point of absolute minimum for $L_{w,q}$ and boundness of the second derivatives of $\bar{H}$ we obtain from (12) that for all positive t and for some $C_4 > 0$

$$L_{w,q}(y) - (|r_k - r_0| - C_3 \operatorname{dist}(\varphi_0, \varphi_k))\, t + C_4 t^2 \geqq L_{w,q}(x)$$

and from (11)

$$C_4 t^2 - (|r_k - r_0| - C_3 \operatorname{dist}(\varphi_0, \varphi_k))\, t + C_2(\operatorname{dist}(\varphi_0, \varphi_j) + \operatorname{dist}(\varphi_0, \varphi_k)) \geqq 0.$$

Consequently

$$(|r_k - r_0| - C_3 \operatorname{dist}(\varphi_0, \varphi_k))^2 < C_5(\operatorname{dist}(\varphi_0, \varphi_j) + \operatorname{dist}(\varphi_0, \varphi_k))$$

which implies the statement of the proposition. □

Weak regularity does not imply regularity for two reasons. First, it is possible that the orbit in consideration contains two points whose φ-coor.

dinates are very close or even equal, while all other points have φ-coordinates far away from the two. In this case the assertion of Proposition 5 is vacuous. Second, in the situation of Proposition 5 one may have $\operatorname{dist}(\varphi_i, \varphi_k) \ll \operatorname{dist}(\varphi_i, \varphi_j)$. In this case the estimate for $|r_i - r_k|$ becomes very weak, although the corresponding estimate for $|r_i - r_j|$ which can be obtained by re-naming indices is good. The last observation is refined in the following statement.

Theorem B. *Under the assumptions of Proposition 5 for every point (φ_i, r_i) of the orbit the inequality*

$$|r_i - r_j| < c(\operatorname{dist}(\varphi_i, \varphi_j))^{\frac{1}{2}}$$

holds for all except probably one $j \in \{0, \ldots, q-1\}$. The constant c depends only on f, but not on w and q.

Proof. Let k, $i < k < i+q$ be such that

$$\operatorname{dist}(\varphi_i, \varphi_k) = \min_{i<j<i+q} \operatorname{dist}(\varphi_i, \varphi_j).$$

Then for every $j \neq k$, $i < j < i+q$ one has from Proposition 5

$$|r_i - r_j| < c(\operatorname{dist}(\varphi_i, \varphi_j) + \operatorname{dist}(\varphi_i, \varphi_k))^{\frac{1}{2}} < c'(\operatorname{dist}(\varphi_i, \varphi_j))^{\frac{1}{2}}. \quad \square$$

Theorem B allows us to make the first modest step toward the extension of results by Aubry and Mather to higher dimensions. Namely we will show that limits of minimal periodic orbits are regular.

Theorem C. *Under the assumptions of Proposition 5 let $\frac{w^{(n)}}{q^{(n)}}$, $n = 1, 2 \ldots$ be a sequence of rotation vectors and $(\varphi^{(n)}, r^{(n)})$ be a sequence of points whose f-orbits correspond to absolute minima of the functionals $L_{w^{(n)}, q^{(n)}}$. Suppose that the sequence $(\varphi^{(n)}, r^{(n)})$ converges to a point (φ, r) which is not an isolated point of its orbit. Then the orbit of (φ, r) is ω-regular where $\omega(t) = ct^{\frac{1}{2}}$.*

Proof. Let us note first that (φ, r) is not a periodic point and all points of its orbit are not isolated. Consider any two points on the orbit of (φ, r), say $(\varphi_i, r_i) = f^i(\varphi, r)$ and $(\varphi_j, r_j) = f^j(\varphi, r)$. Let us show first that $\varphi_i \neq \varphi_j$. For otherwise one can find k such that $\operatorname{dist}(\varphi_i, \varphi_k)$ is very small compared to $|r_i - r_j|^2$, then approximate the piece of orbit containing all three points by a piece of orbit of $(\varphi^{(n)}, r^{(n)})$ with very high precision and use Proposition 5 for the approximation.

Similarly if $\varphi_i \neq \varphi_j$, let us find k such that $\operatorname{dist}(\varphi_i, \varphi_k) \ll \operatorname{dist}(\varphi_i, \varphi_j)$. Assume $i < j < k$. Other cases can be treated similarly. Approximate the piece of orbit from i to k by a piece of orbit of $(\varphi^{(n)}, r^{(n)})$ and again apply Proposition 5 for the approximation. $\square$

At the current stage we are not able to make a comprehensive description of orbits which appear as limits defined in the last theorem. We hope that if $\frac{w^{(n)}}{q^{(n)}} \to \alpha$ then the limit orbits have rotation vector α and are fairly similar to quasi-periodic orbits with that rotation vector.

Let us note that in the situation described in Theorem C even if (φ, r) happens to be a periodic orbit, it satisfies the same regularity condition if the approximating orbits pass near that orbit several times.

7. Continuous time Hamiltonian systems

Let us consider a completely integrable Hamiltonian system with n degrees of freedom. Let $(I, \varphi)=(I_1 \dots I_n, \varphi_1 \dots \varphi_n)$, $I \subset U$ be an open set in $\mathbb{R}^n$, $\varphi \in \mathbb{T}^n$ be the action-angle coordinates for that system ([3], Sect. 50) so that the Hamiltonian which we will denote by H_0 depends only on I and the time evolution leaves every torus $\{I^{(0)}\} \times \mathbb{T}^n$ invariant. Let us fix $I^{(0)} \in U$ and assume that

$$\frac{\partial H_0}{\partial I_1}(I^{(0)}) \neq 0. \tag{13}$$

Then in the neighborhood of the invariant torus $\{I^{(0)}\} \times \mathbb{T}^n$ one can define the Poincaré map T on the hypersurface $\varphi_1 = 0$. This map has the following form

$$T(I, \varphi_2 \dots \varphi_n) = \left(I, \varphi_2 + \frac{\partial H_0}{\partial I_2} \Big| \frac{\partial H_0}{\partial I_1}, \dots, \varphi_n + \frac{\partial H_0}{\partial I_n} \Big| \frac{\partial H_0}{\partial I_1}\right). \tag{14}$$

The restriction of the map T to the $2n-2$-dimensional invariant manifold $N = \{H_0(I) = H_0(I^{(0)}),\ \varphi_1 = 0\}$ is a symplectic map with respect to the induced symplectic form $\sum_{i=2}^{n} dI_i \wedge d\varphi_i$.

By the implicit function theorem, on the hypersurface $H_0(I) = H_0(I^{(0)})$ one can locally express

$$I_1 = S(I_2 \dots I_n). \tag{15}$$

We will make the following assumption:

(iii) *the hypersurface in I-space $H_0(I) = H_0(I^{(0)})$ is strictly differentiably convex at the point $I^{(0)}$.*

This condition is satisfied for example, if the function H_0 is strictly convex in I at $I^{(0)}$, i.e. if the Hessian $\left.\frac{\partial^2 H_0}{\partial I^2}\right|_{I = I^{(0)}}$ is a positive definite quadratic form. Condition (iii) implies that S is a strictly convex function of $(I_2 \dots I_n)$ near $(I_2^{(0)}, \dots, I_n^{(0)})$.

Proposition 6. *The generating function for the lift of the map $T: N \to N$ to the universal cover has the form $\mathscr{H}(x, x') = h(x' - x)$ where h is the Legendre transformation of the function S.*

Proof. One has from (15) for $i = 2, \dots, n$

$$\frac{\partial S}{\partial I_i} = -\frac{\partial H_0}{\partial I_i} \cdot \left(\frac{\partial H_0}{\partial I_1}\right)^{-1}$$

so that from (2) if $\delta_i = \frac{\partial S}{\partial I_i}$ then $\frac{\partial h}{\partial \delta_i} = I_i$. □

By Proposition 6 if assumption (iii) is satisfied, h is a strictly convex function because it is a Legendre transform of a strictly convex function. Thus condition (iii) for H_0 implies that conditions (i) and (ii) from Sect. 2 are locally satisfied for the Poincaré map T.

Consider now a C^2 small Hamiltonian perturbation $H(I, \varphi)$ of the Hamiltonian H_0. The first-return map on the manifold $N_H = \{\varphi_1 = 0,\ H(I, \varphi) = H_0(I^{(0)})\}$ is still symplectic. Since manifolds N_H and N are close and the projection along I_1 direction is a symplectic map, the new map can be viewed as a C^1 small symplectic perturbation of the map T. Thus, this map is determined by a generating function which is C^2 (and hence C^1) close to h so that all results from the previous sections can be applied to this case. The properties of the Poincaré map can be in the obvious way extended to the continuous time systems. We leave exact formulations of the results for Hamiltonian systems corresponding to Theorems A, B, C to the reader.

Remark. The C^2 closeness of the perturbed Hamiltonian to the one for the integrable system is only sufficient but not necessary for application of our results because we only need C^1 closeness for the generating function (compare with discussion in Sect. 2). However since the relationship between the Hamiltonian and the generating function of a Poincaré map is rather complicated for the non-integrable case, we do not try to interpret the weaker condition in terms of the Hamiltonian.

Acknowledgements. This paper was written during the visit of the first author to California Institute of Technology in December 1984. We would like to thank Caltech's Mathematics Department for the support for that visit. We also feel a pleasant obligation to thank the referee for his thoughtful remarks.

References

1. Arnold, V.I.: Small denominators II. Proof of a theorem of A.N. Kolmogorov on the preservation of conditionally-periodic motions under a small perturbation of the Hamiltonian. Russ. Math. Surv. **18**, 9–36 (1963)
2. Arnold, V.I.: Sur une propriété topologique des applications globalement canoniques de la mécanique classique. C.R. Acad. Sci. Paris **261**, 3719–3722 (1965)
3. Arnold, V.I.: Mathematical methods of classical mechanics. Berlin-Heidelberg-New York: Springer 1978
4. Aubry, S.: Theory of the devil's staircase. Seminar on the Riemann Problem and Complete Integrability. Lect. Notes Math. **925**. Berlin-Heidelberg-New York: Springer 1978–79
5. Aubry, S., La Daeron, P.Y.: The discrete Frenkel-Kontorova model and its extensions. Physica 8D, 381–422 (1983)
6. Aubry, S., La Daeron, P.Y., André, G.: Classical ground-states of a one-dimensional model for incommensurate structures. (Preprint, 1982)
7. Bernstein, D.: Birkhoff periodic orbits twist maps with graph intersection property. Ergodic Theory Dyn. Syst. **5**, 531–532 (1985)
8. Birkhoff, G.D.: Une généralisation à n dimension du dernier théorème géométrique de Poincaré. C.R. Acad. Sci. Paris **192**, 196–198 (1931)
9. Birkhoff, G.D., Lewis, D.C.: On the periodic motions near a given periodic motion of a dynamical system. Ann. Mat. **12**, 117–133 (1933)
10. Conley, C.C., Zehnder, E.: The Birkhoff-Lewis fixed point theorem and a conjecture of V.I. Arnold. Invent. Math. **73**, 33–49 (1983)

11. Denjoy, A.: Sur les courbes définies par les équations différentielles à la surface du tore. J. Math. Pure Appl. (Ser. 9), **11**, 333–375 (1932)
12. Hall, G.R.: A topological version of a theorem of Aubry-Mather on twist maps. Ergodic Theory Dyn. Syst. **4**, 585–603 (1984)
13. Katok, A.: Some remarks on Birkhoff and Mather twist map theorems. Ergodic Theory Dyn. Syst. **2**, 185–194 (1982)
14. Katok, A.: More about Birkhoff periodic orbits and Mather sets of twist maps. (Preprint with continuation, 1982)
15. Katok, A.: Periodic and quasi-periodic orbits for twist maps. Dynamical Systems and Chaos, Proceedings, Sitges, Lect. Notes Phys. **179** (Springer, 1982)
16. Ljusternik, L.: The topology of calculus of variations in the large. AMS, Translation of Math Monograph v. 16, 1966
17. Mather, J.N.: Existence of quasi-periodic orbits for twist homeomorphisms of the annulus. Topology **21**, 457–467 (1982)
18. Mather, J.N.: A criterion for the non-existence of invariant circles. Publ. Math. IHES, **63**, 153–204 (1982)
19. Moser, J.: Proof of a generalized form of a fixed point theorem due to G.D. Birkhoff. Geometry and Topology, Rio de Janeiro. Lect. Notes Math. **597**, 464–494 (1976)
20. Schwartz, J.T.: Nonlinear functional analysis, 1963–64. Courant Inst. of Math. Science 1965

Oblatum 14-IV-1985

Invent. math. 88, 243–256 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Curvature estimates and compactness theorems in 3-manifolds for surfaces that are stationary for parametric elliptic functionals

B. White

Department of Mathematics, Stanford University, Stanford, CA 94305, USA

Let N be a three dimensional riemannian manifold and Φ a function that assigns a real number $\Phi(x, v)$ to each $x \in N$ and unit vector v in the tangent space to N at x. Then Φ defines a functional on surfaces as follows:

$$(*) \qquad \Phi(M) = \int_{x \in M} \Phi(x, \nu(x)) dx$$

where $\nu(x)$ is the unit normal to M at x and the integration is with respect to surface area on M. Note if Φ is even in that $\Phi(x, v) \equiv \Phi(x, -v)$, then $(*)$ makes sense even if M is non-orientable. We say that M is Φ-stationary if it satisfies the Euler-Lagrange equations for Φ. For example, if $\Phi(x, v) \equiv 1$, then $\Phi(M)$ is the area of M, and M is Φ-stationary if and only if M is a minimal surface. In this paper we prove curvature estimates and compactness theorems for regular Φ-stationary surfaces provided Φ is elliptic, i.e. if for some $\lambda > 0$, $(\Phi(x, v/|v|) - \lambda)|v|$ is a convex function of $v \in \mathrm{Tan}_x N$. For example we prove

Theorem 0. *If Φ is an even elliptic integrand, Φ and $D_2\Phi$ are $C^{2,\alpha}$, N is a compact manifold with strictly Φ-convex boundary, and M is an embedded Φ-stationary surface with $\partial M \subset \partial N$, then the principal curvatures of M are bounded by a constant depending only on Φ, N, the area and genus of M, $\|\partial M\|_{2,\alpha}$, and the "embeddedness" of ∂M,* i.e.

$$\max \{\mathrm{dist}_{\partial M}(x, y)/\mathrm{dist}_N(x, y) \colon x, y \in \partial M, x \neq y\}$$

(Φ-convexity of ∂N means that the Φ analogue of the mean curvature of ∂N points into N).

In the case of minimal surfaces, this theorem was proved by Michael Anderson [A] with similar methods. (In [A] ∂M is only assumed to be C^1, but there is an error: the conclusion is clearly false unless ∂M is at least C^2.) However, his arguments rely heavily on monotonicity and hence do not apply to general integrands Φ. Also, whereas his proofs use rather sophisticated

geometric measure theory (such as the Allard regularity theorem), the proofs in this paper are fairly elementary.

The methods are also similar to some of those used by Choi and Schoen [CS] to prove compactness theorems for embedded minimal surface in compact 3-manifolds of positive ricci curvature.

Regular Φ-stationary surfaces must have quasiconformal gauss maps, and there are interior estimates due to Schoen and Simon [SS2] for simply connected embedded surfaces in small balls with quasiconformal gauss maps. But even in the case of simply connected M, our results do not follow from [SS2]; for even if M is simply connected, its intersection with small balls need not be.

Sometimes one can prove curvature estimates without assuming a bound on the genus. If M is a Φ-minimizing current, one can prove such interior estimates ([ASS], using [S]). In case Φ is the area integrand, interior estimates can be proved for oriented hypersurfaces of dimension <7 [SS1], and interior and boundary estimates for oriented area minimizing hypersurfaces of dimension <7 ([S], [HS]).

1. Preliminaries

1. N is either a smooth compact riemanian 3-manifold (with or without boundary) or R^3. All measurements such as length and area are with respect to the riemannian metric on N (even if Φ is the area integrand for some other riemannian metric on N).

2. If Γ is an immersed curve or surface in N (or N itself) and $x, y \in \Gamma$, then $\mathrm{dist}_\Gamma(x, y)$ is the geodesic distance along Γ from x to y.

3. If M is an immersed surface in N, $B_M(x)$ denotes the second fundamental form of M at x. In particular, $|B_M(x)|^2$ is the sum of the squares of the principal curvatures of M at x. $K_M(x)$ denotes the gauss curvature of M at x (so if $M \subset R^3$, then $K_M(x)$ is the product of the principal curvatures of M at x).

Observe that if M' is obtained from M by dilation by a factor μ, then

$$|B_{M'}(x')| = \mu^{-1} |B_M(x)|$$
$$K_{M'}(x') = \mu^{-2} K_M(x)$$

so

$$\int_{M'} |B_{M'}|^2 = \int_M |B_M|^2$$
$$\int_{M'} K_{M'} = \int_M K_M.$$

These facts will be used repeatedly.

4. If Γ is an embedded curve in N, we let $\|\Gamma\|_{2,\alpha}$ be the $C^{2,\alpha}$ norm of Γ as parametrized by arc-length, and

$$\|\Gamma\|^*_{2,\alpha} = \|\Gamma\|_{2,\alpha} + \max \{\mathrm{dist}_\Gamma(x, y)/\mathrm{dist}_N(x, y) \colon x, y \in \Gamma (x \neq y)\}.$$

(If x and y are in different connected components of Γ, we define $\mathrm{dist}_\Gamma(x, y)$ to be 1.)

5. If Φ is an even integrand (i.e. $\Phi(x, y) \equiv \Phi(x, -v)$), ∂N is said to be strictly Φ-convex provided the first variation vectorfield of ∂N with respect to Φ is everywhere nonzero and points into N. This implies by the maximum principle (cf. the proof of Lemma 1) that if $M \subset N$ is Φ-stationary, then $(M \tilde{} \partial M) \cap \partial N$ is empty.

6. The generalized Chern-Osserman theorem.

Theorem. *Let M be a complete connected oriented surface immersed in R^3. If*

$$\int_M |B_M|^2 < \infty$$

then

$$\int_M K_M$$

is an integer multiple of 4π.

Osserman [O] proved this for minimal surfaces; the general theorem is proved in [W2]. The same argument shows that if M has a straight line L as boundary, then the theorem remains true if we replace 4π by 2π. (Note that the image of L under the gauss map of M lies in a great circle in S^2.)

Note also that if M is non-orientable, then we can apply the theorem to the double cover $M \tilde{}$ of M to deduce that

$$\int_M K = (1/2) \int_{M \tilde{}} K = \begin{cases} 2\pi m & \text{if } \partial M \text{ is empty} \\ \pi m & \text{if not.} \end{cases}$$

2. Local results

Lemma 1. *For every $A > 0$ there is an*

$$\varepsilon = \varepsilon(A, \lambda, \|\Phi\|_{1,\alpha}, \|D_2 \Phi\|_{1,\alpha}, \alpha) > 0$$

such that if M is a C^2 Φ-stationary immersed surface and if

(1) $$|B_M(x)| \leqq A \quad \text{when } \operatorname{dist}_M(x, x_0) < \varepsilon$$

$$\|\partial M \cap \{x \in M : \operatorname{dist}_M(x, x_0) < \varepsilon\}\|^*_{2,\alpha} \leqq A$$

then $\{x \in M : \operatorname{dist}_M(x, x_0) < \varepsilon\}$ is the graph over a subset of $\operatorname{Tan}_M(x_0)$ of a function u such that

$$\|u|(B_{\varepsilon/4}(x_0) \cap \operatorname{Tan}_M(x_0))\|_{2,\alpha} \leqq \varepsilon^{-1}.$$

Remark. In general, the same proof shows that for any integer $k \geqq 0$

$$\|u|(B_{\varepsilon/4}(x_0) \cap \operatorname{Tan}_M(x_0))\|_{2+k,\alpha} \leqq [\varepsilon(A, \lambda, \|\Phi\|_{k,\alpha}, \|D_2 \Phi\|_{k,\alpha}, \alpha)]^{-1}.$$

Proof. For simplicity, consider the case

$$\partial M \cap \{x \in M : \operatorname{dist}_M(x, x_0) < \varepsilon\} = \emptyset$$

and let $x_0 = 0 \in R^3$ and $\mathrm{Tan}_M(x_0) = R^2 \times (0)$. From the definition of B_M one has

$$|\nu(x) - \nu(x_0)| \leqq A \operatorname{dist}_M(x, x_0) \tag{2}$$

(where $\nu(x)$ is the unit normal to M at x), which implies that if $A\varepsilon \leqq 1$, then $\{x: \operatorname{dist}_M(x, x_0) < \varepsilon\}$ is the graph of a function u with

$$\sup |Du| \leqq CA\varepsilon. \tag{3}$$

Thus

$$\begin{aligned} |x - x_0| &\leqq \operatorname{dist}_M((x, u(x)), (x_0, u(x_0))) \\ &\leqq (1 + (CA\varepsilon)^2)^{1/2} |x - x_0| \end{aligned}$$

which implies that the domain of u includes a ball of radius $\varepsilon/2$ about x_0 (provided $A\varepsilon$ is sufficiently small). Note also that (2) and (3) imply that Du is lipschitz with lipschitz constant $\approx A$.

Now let $F = F(x, y, p)$ $(x \in R^2,\ y \in R,\ p \in R^2)$ be the nonparametric functional associated with Φ (see appendix), so that (for $U \subset M$ near x_0)

$$\int_U \Phi(z, \nu(z)) dz = \int_{\Pi(U)} F(x, u(x), Du(x)) dx.$$

Then u satisfies the Euler-Lagrange equation

$$a_{ij}(x) D_{ij} u(x) + b_i(x) D_i u(x) = f(x) \qquad (1 \leqq i, j \leqq 2) \tag{4}$$

where

$$a_{ij}(x) = \frac{\partial^2 F}{\partial p_i \partial p_j}(x, u(x), Du(x))$$

$$b_i(x) = \frac{\partial^2 F}{\partial y \partial p_i}(x, u(x), Du(x))$$

$$f(x) = \frac{\partial F}{\partial y}(x, u(x), Du(x)) - \frac{\partial^2 F}{\partial x_i \partial p_i}(x, u(x), Du(x)).$$

Because Φ is elliptic and Du is bounded, one can check that (4) is uniformly elliptic, with bounds depending only on λ and $\|D_2 \Phi\|_1$. Furthermore

$$\begin{aligned} |a_{ij}(x) - a_{ij}(y)| &\leqq \|D_2 \Phi\|_{1,\alpha} (|u(x) - u(y)|^\alpha + |Du(x) - Du(y)|^\alpha) \\ &\leqq C \|D_2 \Phi\|_{1,a} A^\alpha |x - y|^\alpha \end{aligned}$$

(by (2) and (3)).

In the same way one sees that

$$\|a_{ij}\|_{0,\alpha},\ \|b_i\|_{0,\alpha},\ \|f\|_{0,\alpha} \leqq C(A, \|\Phi\|_{1,\alpha}, \|D_2 \Phi\|_{1,\alpha}).$$

But now the conclusion follows immediately from the Schauder estimates for the linear equation (4)

$$\|u | B_{\varepsilon/4}\|_{2,\alpha} \leqq C(\|u | B_{\varepsilon/2}\|_0 + \|f | B_{\varepsilon/2}\|_{0,\alpha})$$

([GT, Corollaries 6.3, 6.7]). □

Lemma 2. *If M is an immersed Φ-stationary surface, then*

$$|B_M(x)|^2 \leqq -c_1 K_M(x) + c_2$$

where $c_1 > 0$ and c_2 depend only on λ, $\|\Phi\|_1$, and $\|D_2\Phi\|_1$ (and the sectional curvatures of N if N is not flat).

Proof. Without loss of generality, $x = 0$ and M is the graph of a function u with $Du(0) = 0$. By a rotation, we may assume $D^2u(0)$ is diagonal, its entries then being the principal curvatures κ_1 and κ_2 of M at 0. Then M satisfies the Euler-Lagrange equation ((4) in the proof of Lemma 1):

$$a_{ij}(x) D_{ij} u(x) + b_i(x) D_i u(x) = f(x)$$

which at 0 is

$$a_{11}\kappa_1 + a_{22}\kappa_2 = f(0).$$

Therefore

$$(a_{11}\kappa_1)^2 + (a_{22}\kappa_2)^2 = -2a_{11}a_{22}\kappa_1\kappa_2 + f(0)^2.$$

But $\lambda \leqq a_{ii} \leqq \|D_{22}\Phi\|_0$ and $|f(0)| \leqq \|\Phi\|_1 + \|D_2\Phi\|_1$, so

$$(\kappa_1)^2 + (\kappa_2)^2 \leqq -c_1\kappa_1\kappa_2 + c_2'. \tag{1}$$

Thus

$$|B|^2 \leqq -c_1 K_M + c_2'.$$

This proves the theorem if N is R^3 (or if N is flat). If N is not flat, one uses normal coordinates at $x \in N$ and derives (1) exactly as above. But then K_M is just the sum of $\kappa_1\kappa_2$ and the sectional curvature $R = R_N(e_1, e_2, e_1, e_2)$, so

$$\begin{aligned} |B|^2 &\leqq -c_1(K_M + R) + c_2' \\ &= -c_1 K_M + (c_2' - c_1 R) \\ &= -c_1 K_M + c_2. \quad \square \end{aligned}$$

Remark. Note that if

$$|B_M|^2 \leqq -c_1 K_M(x) + c_2$$

and if M' is obtained from M by dilation by μ, then

$$|B_{M'}(x')|^2 \leqq -c_1 K_{M'}(x') + \mu^{-2} c_2.$$

Theorem. *For every A and $\varepsilon > 0$, there exists*

$$c = c(\varepsilon, A, \lambda, \|\Phi\|_{1,\alpha}, \|D_2\Phi\|_{1,\alpha}, \alpha)$$

with the following property. Suppose M is an immersed Φ-stationary surface such that

$$\int_M |B|^2 \leqq A.$$

If

$$\int_M |K| \leqq 2\pi - \varepsilon \tag{1}$$

then

$$|B(x)| \inf\{1, \operatorname{dist}_M(x, \partial M)\} \leqq C. \tag{2}$$

Furthermore, if M has as a portion of its boundary an embedded curve Γ with

$$\|\Gamma\|^*_{2,\alpha} \leqq A$$

and if

$$\int_M |K| \leqq \pi - \varepsilon \tag{1'}$$

then

$$|B(x)| \inf\{1, \operatorname{dist}_M(x, \partial M \sim \Gamma)\} \leqq C. \tag{2'}$$

Corollary. *If M is a Φ-stationary surface properly immersed in an open set U, $\partial M \cap U$ is an embedded curve with $\|\partial M \cap U\|^*_{2,\alpha} \leqq A$, and if M satisfies* (1') *(or* (1) *if $\partial M \cap U$ is empty) then*

$$|B(x)| \inf\{1, \operatorname{dist}(x, \partial U)\} \leqq C.$$

Remark. If M is orientable, 2π may be replaced by 4π in (1) and π by 2π in (1').

Proof. Suppose (2) does not hold. Then there exist sequences M_n and Φ_n satisfying the hypotheses of the theorem and such that for some x_n in M_n

$$|B_{M_n}(x_n)| \cdot (\inf\{1, \operatorname{dist}_{M_n}(x_n, \partial M_n)\} - (1/n)) \to \infty. \tag{3}$$

Note we may assume that

$$\operatorname{dist}_{M_n}(y, x_n) < 1 \quad \text{for } y \in M_n$$

(otherwise replace M_n by the set of such y in M_n). Now let

$$f_n(x) = |B_{M_n}(x)| \cdot (\operatorname{dist}_{M_n}(x, \partial M_n) - (1/n)).$$

Then f_n is negative outside of a compact subset of M_n, so it attains a maximum at some point z_n. Note that $f_n(x_n) \to \infty$ by (3), so

$$f_n(z_n) \to \infty. \tag{4}$$

Since the diameter of M_n is $\leqq 2$, this implies

$$S_n = |B_{M_n}(z_n)| \to \infty.$$

Now translate M_n and Φ_n by $-z_n$ and dilate by S_n to get M'_n and Φ'_n with M'_n stationary for Φ'_n. Then

$$|B_{M'_n}(0)| = 1$$

and (writing dist for $\operatorname{dist}_{M'_n}$)

$$\begin{aligned}|B_{M_n'}(x)|\cdot(\operatorname{dist}(0,\partial M_n')-\operatorname{dist}(0,x)-S_n/n)&\\ \leqq|B_{M_n'}(x)|\cdot(\operatorname{dist}(x,\partial M_n')-S_n/n)&\\ \leqq|B_{M_n'}(0)|\cdot(\operatorname{dist}(0,\partial M_n')-S_n/n)&\end{aligned}$$

(since $f_n \leqq f_n(z_n)$). But $f_n(z_n)=(\operatorname{dist}(0,\partial M_n')-S_n/n)$, so

$$|B_{M_n'}(x)| \leqq f_n(z_n)/(f_n(z_n)-\operatorname{dist}(0,x)).$$

Since $f_n(z_n)\to\infty$, this means $|B_{M_n'}(x)|$ is uniformly bounded on compact sets as $n\to\infty$.

(To be more precise, let ψ_n be the map from a subset of R^2 onto $M_n'\subset R^3$ determined by geodesic polar coordinates for M_n' at 0. Then $|B_{M_n'}(\psi_n(x))|$ is uniformly bounded on compact subsets of R^2 as $n\to\infty$.)

Then it follows from Lemma 1 that (a subsequence of) the M_n' (or, more precisely, the ψ_n) converge uniformly in C^2 on compact sets to some surface M'. Note that

$$|B_{M'}(x)|^2 \leqq -c_1 K_{M'}(x) \tag{5}$$

by the remark after Lemma 2. Since

$$f_n(z_n)=\operatorname{dist}(0,\partial M_n')-S_n/n\to\infty,$$

M' is complete. Also

$$\int_{M'}|B_{M'}|^2=\lim_{r\to\infty}\int_{M'(0,r)}|B_{M'}|^2$$

(where $M'(0,r)$ is the geodesic ball in M' of radius r centered at 0)

$$\begin{aligned}&=\lim_{r\to\infty}\lim_{n\to\infty}\int_{M_n'}|B_{M_n'(0,r)}|^2\\ &\leqq\lim_{n\to\infty}\int_{M_n'}|B_{M_n'}|^2\\ &=\lim_{n\to\infty}\int_{M_n}|B_{M_n}|^2\end{aligned}$$

$$\therefore\int_{M'}|B_{M'}|^2\leqq A. \tag{6}$$

Similarly

$$\int_{M'}|K_{M'}|\leqq 2\pi-\varepsilon. \tag{7}$$

By the generalized Chern-Osserman theorem (8 in Sect. 1), (6) implies that

$$\int_{M'}K_{M'}$$

is an integer multiple of 2π (indeed of 4π if M' is orientable). Thus by (7)

$$\int_{M'}K_{M'}=0.$$

Therefore by (5)

$$\int_{M'} |B_{M'}|^2 = 0$$

so $B_{M'} \equiv 0$. But

$$|B_{M'}(0)| = \lim |B_{M'_n}(0)| = 1,$$

a contradiction. This proves (2).

The proof of (2′) is essentially the same, except that M' may have a straight line L as boundary. In that case $\int_{M'} K_{M'}$ is an integer multiple of π (or of 2π if M' is orientable). □

Theorem 2. *(Removal of singularities.) Let U be the interior of the unit ball in R^3 and let*

$$U^+ = \{(x, y, z) \in U : z \geqq 0\}.$$

Let Φ be an even elliptic integrand such that Φ and $D_2\Phi$ are $C^{2,\alpha}$, and let M be a Φ-stationary surface properly embedded in $U \sim \{0\}$ such that $0 \in \bar{M}$ and $\int_M |B_M|^2 < \infty$. If

(1) $$\partial M \cap (U \sim \{0\}) = \emptyset$$

or if

(2) $M \subset U^+$,
$(\partial M \cap U) \cup \{0\}$ *is a $C^{2,\alpha}$ embedded curve in ∂U^+, and*
$(\partial U^+) \cap U$ *is strictly Φ-convex*

then $M \cup \{0\}$ is a $C^{2,\alpha}$ embedded Φ-stationary surface.

Proof. Suppose (1) holds. (The proof for (2) is similar.) We may assume

$$\int_M |B|^2 < \pi/2.$$

(Otherwise replace M by its intersection with a small ball centered at 0 and dilate.) Then by the corollary to Theorem 1,

(3) $$|B_M(x)| \leqq C|x|^{-1}$$

for x near 0.

Now let R_i be any sequence decreasing to 0, and let M_i be the image of M under dilation by R_i^{-1}. Note that M_i satisfies the same bound (3), so by Lemma 1, a subsequence $M_{i'}$ will converge nicely in $R^3 \sim \{0\}$ to a complete surface M'. As before (in the proof of Theorem 1), if E is a compact subset of $R^3 \sim \{0\}$

$$\begin{aligned} \int_{E \cap M'} |B_{M'}|^2 &= \lim \int_{E \cap M_{i'}} |B_{M_{i'}}|^2 \\ &= \lim \int_{(R_i E) \cap M} |B_M|^2 \\ &= 0. \end{aligned}$$

So $B_{M'} \equiv 0$. It follows that

$$B_M(x) = o(|x|^{-1}). \tag{4}$$

Since $|B_{M'}| \equiv 0$, M' is a union of planes. Note that this immediately implies that, for small enough r, the function $x \to |x|$ on $M \cap B_r$ has no critical points except for strict local minima. This means (by Morse theory) that each connected component of $M \cap B_r$ is either a disk with exactly one critical point, or a punctured disk with no critical points. Let $\tilde{M}$ be a connected component that is a punctured disk with no critical point and let $\tilde{M}_{i'}$ be the component of $M_{i'}$ corresponding to $\tilde{M}$. Then $\tilde{M}_{i'}$ converges to a plane through (0), say $R^2 \times \{0\}$.

Since $\tilde{M}_{i'}$ is embedded,

$$\tilde{M}_{i'} \cap \{(x, y, z): x^2 + y^2 = 1\}$$

is the graph of a function w_i with

$$\|w_i\|_{2,\beta} \to 0 \qquad (0 < \beta < \alpha)$$

(by (4) and Lemma 1).

Note that the dilated integrands $\Phi_{i'}$ converge nicely to the "frozen" integrand Φ_0:

$$\Phi_0(x, v) \equiv \Phi(0, v).$$

It follows by the implicit function theorem (see appendix) that for $|t| \leqq 1$ and sufficiently large i, there is a unique function

$$v_{i,t}: B_1(0) \cap R^2 \to R$$

such that

$$v_{i,t}(x) = w_i(x) + t \qquad \text{if } |x| = 1$$

and such that $v_{i,t}$ is a solution to the Euler-Lagrange equation for $\Phi_{i'}$. Furthermore, $v_{i,t}$ depends smoothly on t, so in fact the graphs of the $v_{i,t}$ foliate a region of $B^2 \times R$. By the maximum principle (applied to $\tilde{M}_{i'} \sim \{0\}$), the leaf such that $v_{i,t}(0) = 0$ lies on one side of $\tilde{M}_{i'}$. But then (since 0 is in the closure of $\tilde{M}'$), any sequence of dilations of $\tilde{M}$ must converge to the same limiting plane. Thus $\tilde{M} \cup \{0\}$ is, in a neighborhood of 0, the graph of a C^1 function v. Since v is a $C^{2,\alpha}$ solution to the Euler-Lagrange equation for Φ except at 0, in fact it must be $C^{2,\alpha}$. (See appendix.)

We have shown that $(M \cup \{0\}) \cap B_r$ is a union of smooth disks. By the maximum principle, only one disk can pass through 0. Hence $M \cup \{0\}$ is embedded. □

3. The compactness theorem

Theorem 3. *Let N be a compact 3-manifold with (possibly empty) boundary. Let M_i be a sequence of compact connected immersed Φ_i-stationary surfaces (with or without boundary), where*

$$\|\Phi_i - \Phi\|_{1,\alpha} + \|D_2\Phi_i - D_2\Phi\|_{1,\alpha} \to 0 \qquad (\text{as } i \to \infty).$$

Suppose that the area and the genus of M_i and the total curvature of ∂M_i are uniformly bounded, and that the ∂M_i converge as sets (i.e. *in the Hausdorff metric*) *to some set Γ. Then*

(1) *There is a finite set $S \subset N$ and a subsequence $M_{i'}$ that converges uniformly in $C^{2,\beta}$ $(\beta < \alpha)$ on compact subsets of $\Omega = N \sim (S \cup \Gamma)$ to a Φ-stationary surface M.*

(2) *If $\|\partial M_i\|^*_{2,\alpha}$ is bounded, then we can let $\Omega = N \sim S$.*

Now suppose that each M_i is embedded, Φ_i is even, and

$$\|\Phi_i - \Phi\|_{2,\alpha} + \|D_2 \Phi_i - D_2 \Phi\|_{2,\alpha} \to 0$$

Then

(3) *$(M \cup S) \sim \Gamma$ is a regular embedded surface.*

(4) *If ∂N is strictly convex, $\partial M_i \subset \partial N$ is not empty, and $\|\partial M_i\|^*_{2,\alpha}$ is uniformly bounded, then S is empty.*

Example. Let $\Phi_n = \Phi$ be the area integrand and let M_n be the set of (x, y, z) in the unit 3-ball such that

$$(x^2 + y^2)^{1/2} = n^{-1} \cosh(nz).$$

Then the hypotheses for (3) are satisfied and M is the unit disk. But the hypotheses for (4) are not satisfied (because $\|\partial M\|^*_{2,\alpha}$ is not bounded), and the curvatures of the M_n blow up near $S = \{0\}$.

Proof. Define measures μ_i on N by

$$\mu_i(U) = \int_{M_i \cap U} |B_{M_i}|^2$$

Then

$$\mu_i(N) = \int_{M_i} |B_{M_i}|^2$$

$$\leqq \int_{M_i} (-c_i K_{M_i}(x) + c_2) dA$$

(by Lemma 2)

$$= c_1 \Big(\int_{\partial M} k - 2\pi\chi(M_i) \Big) + c_2 \operatorname{Area}(M_i)$$

(by Gauss-Bonnet)

$$\leqq C. \tag{5}$$

Thus we may assume that the μ_i converge weakly to a Radon measure μ.

Let $S = \{x \in N\colon \mu(\{x\}) \geqq 1\}$. Then by (5), S has at most C elements. Let $x \in N \sim (S \cup \Gamma)$. Then there is a neighborhood $B_\varepsilon(x)$ with

$$\mu(B_\varepsilon(x)) < 1 \quad \text{and} \quad B_\varepsilon(x) \cap \Gamma = \emptyset.$$

Thus for sufficiently large i

$$\mu_i(B_i(x)) < 1 \quad \text{and} \quad B_\varepsilon(x) \cap \partial M_i = \emptyset.$$

But then by the corollary to Theorem 1,

$$\sup \{|B_{M_i}(y)|\colon y \in B_{\varepsilon/2}(x)\} \leqq C$$

(where C does not depend on i). Then by Lemma 1, there is a subsequence $M_{i'}$ so that $M_{i'} \cap B_{\varepsilon/2}(x)$ converges nicely. Since Ω can be covered by countably many such balls, this proves (1). Conclusion (2) is proved in the same way, using the boundary versions of the corollary to Theorem 1 and of Lemma 1.

To prove (3), note that two pieces of $M \sim (S \cup \Gamma)$ cannot cross each other since $M \sim (S \cup \Gamma)$ is a limit of embedded surfaces. Nor can two pieces touch each other (by the maximum principle). Thus $M \sim (S \cup \Gamma)$ is embedded. But then by Theorem 2, $M \sim \Gamma$ is embedded.

To prove (4), note that since the $M_{i'}$ converge nicely to M away from S, it follows that for large i, $M_{i'}$ is (away from S) an m-sheeted cover of M. (In other words, each small piece of $M \sim S$ has m pieces of $M_{i'}$ converging to it, for some $m \geqq 1$.) But because $\|\partial M_i\|^*_{2,\alpha}$ is uniformly bounded, m must be 1. Now suppose $x \in S \sim \partial M$. (The case $x \in S \cap \partial M$ is similar.) Without loss of generality let $x = 0$ and $\mathrm{Tan}_0 M = R^2 \times \{0\}$. Let $(\Phi_i)_r$, $(M_i)_r$, and M_r be obtained from Φ_i, M_i, and M by dilation by r^{-1}. Note that for each $\varepsilon > 0$ there is a $\delta > 0$ such that for $r < \delta$

$$M_r \cap (B^2 \times [-1, 1]) = \mathrm{graph}\,(w)$$

for some w: $B^2 \to R$ with $\|w\|_{2,\alpha} < \varepsilon$. Thus for large enough i

$$\partial((M_i)_r \cap (B^2 \times [-1, 1])) = \mathrm{graph}\,(w_i)$$

where

$$w_i \colon B^2 \to R, \qquad \|w_i\|_{2,\alpha} < \varepsilon.$$

(This is where we use $m = 1$: otherwise w_i would be an m-valued function.) Now by the implicit function theorem (see appendix for details), there is a foliation of $B^2 \times [-1, 1]$ by $(\Phi_i)_r$-stationary graphs of $C^{2,\alpha}$ functions $v_{i,t}$, one of which (say $v_{i,0}$) has boundary values w_i. But then the maximum principle implies that $M_i \cap (B^2 \times [-1, 1])$ must coincide with the graph of $v_{i,0}$. This means the curvatures of $M_{i'}$ are not blowing up near 0, which contradicts the assumption that $0 \in S$. □

Remark. Notice that Theorem 0 is an immediate consequence of (4).

Appendix

Proposition. *Let Φ be an even elliptic integrand where Φ and $D_2\Phi$ are $C^{2,\alpha}$. Let Φ_r be the integrand defined by*

$$\Phi_r(x, v) = \Phi(rx, v)$$

There is an $\varepsilon > 0$ such that if $r < \varepsilon$ and if

$$w \colon B_1 \subset R^2 \to R,$$
$$\|w\|_{2,\alpha} < \varepsilon$$

then for each $t \in [-1, 1]$, there is a $C^{2,\alpha}$ function v_t: $B_1 \to R$ whose graph is Φ_r-stationary and such that

(1) $$v_t(x) = w(x) + t \quad \text{if } x \in \partial B_1.$$

Furthermore, v_t depends in a C^1 way on t so that the graphs of the v_t foliate a region of R^3. If M is a C^1 properly immersed Φ_r-stationary surface in $B_{1/2}(0)$ with $\partial M \subset graph(v_t)$, then $M \subset graph(v_t)$.

Note. The proof actually gives an $\varepsilon > 0$ that works for all even integrands Ψ with

$$\|\Psi - \Phi\|_{2,\alpha} + \|D_2\Psi - D_2\Phi\|_{2,\alpha}$$

sufficiently small.

Proof. Let $C_0^{2,\alpha}(B) = \{u \in C^{2,\alpha}(B): u = 0 \text{ on } \partial B\}$. Let $F_r = F_r(x, z, p)$ $(x \in R^2,\ z \in R,\ p \in R^2)$ be the nonparametric functional associated with Φ_r:

$$\begin{aligned} F_r(x, z, p) &= \Phi_r((x, z), (-p/(1+|p|^2)^{1/2}, 1/(1+|p|^2)^{1/2}) \\ &= \Phi((rx, rz), (-p/(1+|p|^2)^{1/2}, 1/(1+|p|^2)^{1/2})) \\ &= F(rx, rz, p). \end{aligned}$$

Then the Euler-Lagrange equation for Φ_r is

$$\begin{aligned} H_r(u) &= \frac{\partial}{\partial x_i}\left[\left(\frac{\partial F_r}{\partial p_i}\right)(x, u(x), Du(x))\right] - \left(\frac{\partial F_r}{\partial z}\right)(x, u(x), Du(x)) \\ &= \frac{\partial}{\partial x_i}\left[\left(\frac{\partial F}{\partial p_i}\right)(rx, ru(x), Du(x))\right] - r\left(\frac{\partial F}{\partial z}\right)(rx, ru(x), Du(x)). \end{aligned}$$

Define

$$h: R \times R \times C^{2,\alpha}(B) \times C_0^{2,\alpha}(B) \to C^{0,\alpha}(B)$$

by

$$h(r, t, w, u) = H_r(u + w + t).$$

Then h is C^1 (see the appendix of [W1]) and

$$D_3 h(0, t, 0, 0) = L$$

where

$$Lu = D_i(a^{ij} D_j u)$$

for some positive definite metrix a^{ij}. By [GT, 6.6]

$$\|u\|_{2,\alpha} \leqq C\|Lu\|_{0,\alpha} \qquad (u \in C_0^{1,\alpha}(B)).$$

Thus L is a Banach space isomorphism, so by the implicit function theorem, there exists, for sufficiently small r and w, a function $u = U_{r,t,w}$ in $C_0^{2,\alpha}(B)$ such that

$$h(r, t, w, u) = 0.$$

Now let

$$v_t = U_{r,t,w} + w + t.$$

Since U depends in a C^1 way on t, the graphs of the v_t cover a region of R^3. In particular, they cover $B_{1/2}(0)$.

Now let M be a C^1 properly immersed Φ_r-stationary surface in $B_{1/2}(0)$, with $\partial M \subset \text{graph}(v_0)$. Let M_t be the portion of M above v_t:

$$M_t = \{(x, z) \in M: z \geqq v_t(x)\}.$$

If M_t is not empty for some $t>0$, let T be the largest such t. Then if $\tau<T$ is sufficiently close to T, M_τ must be the graph of a C^1 function u:

$$M_\tau=\{(x,u(x)):\ x\in\Omega\}.$$

This is because M is C^1; indeed, by choosing τ close enough to T, we can make

$$\sup\{|Du(x)-Dv_\tau(x)|:\ x\in\Omega\}$$

arbitrarily small. Now subtract the Euler-Lagrange equation for v_τ from the corresponding equation for u:

$$\begin{aligned}(2)\qquad 0&=H_r(u)-H_r(v_\tau)\\&=\frac{\partial}{\partial x_i}\left[\int_0^1\frac{d}{ds}\frac{\partial F}{\partial p_i}(x,v_\tau+s(u-v_\tau),Dv_\tau+sD(u-v_\tau))ds\right]\\&\quad-\int_0^1\frac{d}{ds}\frac{\partial}{\partial y}F(x,v_\tau+s(u-v_\tau),Dv_\tau+sD(u-v_\tau))ds\\&=D_i[A^{ij}(x)D_j(u-v_\tau)+B^i(x)D_i(u-v_\tau)]-B^i(x)D_i(u-v_\tau)-C(x)(u-v_\tau)\end{aligned}$$

where

$$\begin{aligned}A^{ij}(x)&=\int_0^1\frac{\partial^2}{\partial p_i\partial p_j}F_r(x,v_\tau+s(u-v_\tau),Dv_\tau+sD(u-v_\tau))ds\\&=\int_0^1\frac{\partial^2}{\partial p_i\partial p_j}F(rx,r(v_\tau+s(u-v_\tau)),Dv_\tau+sD(u-v_\tau))ds\\B^i(x)&=\int_0^1\frac{\partial^2}{\partial z\partial p_i}F_r(x,v_\tau+s(u-v_\tau),Dv_\tau+sD(u-v_\tau))ds\\&=r\int_0^1\frac{\partial^2}{\partial z\partial p_i}F(rx,r(v_\tau+s(u-v_\tau)),Dv_\tau+sD(u-v_\tau))ds\\C(x)&=r^2\int_0^1\frac{\partial^2}{\partial z^2}F(rx,rv_\tau+s(u-v_\tau)),Dv_\tau+sD(u-v_\tau))ds.\end{aligned}$$

Thus (given a bound on Du and Dv_τ)

$$\begin{aligned}(3)\qquad |A^{ij}(x)-a^{ij}|&\leqq c\|D_2\Phi\|_{1,\alpha}\langle r^\alpha+\|Du\|_0^\alpha+\|Dv_\tau\|_0^\alpha\rangle\\|B^i(x)|&\leqq r\|\Phi\|_2\\|C(x)|&\leqq r^2\|\Phi\|_2.\end{aligned}$$

By choosing $\varepsilon>0$ sufficiently small, we can make r and $\|Du\|_0$ as small as we wish; and by choosing τ close to T, we can make $\|Du-Dv_\tau\|_0$ as small as we wish. In particular, by (3) we can make the lowest eigenvalue of

$$-D_i(A^{ij}D_jw+B^iw)+B^iD_iw+Cw$$

on B be positive. But since $u=v_\tau$ on $\partial\Omega$, this forces $u\equiv v_\tau$ (by (2)), which contradicts the choice of T. Thus M lies on or below the graph of v_0. But the same argument also shows that M lies on or above the graph of v_0, so M must be contained the graph of v_0. □

Remark 1. Given a linear function $A: R^2 \to R$, we could replace (1) by

$$(1') \qquad v_t(x) = (w(x) + A(x) + t \quad \text{if } |x| = 1.$$

The proof is as above (except that ε now may also depend on A). This is useful in proving the boundary version of Theorem 2.

Remark 2. Exactly the same proof works for hypersurfaces in R^n. Also, note that it is not necessary to assume that Φ is even provided M is embedded and the orientation on ∂M induced from M is the same as that induced by the portion of graph(v_t) that it bounds.

References

[A] Anderson, M.: Curvature estimates and compactness theorems for minimal surfaces in 3-manifolds. Ann. Sci. Ec. Norm. Super., IV. Ser. **18**, 89–105 (1985)

[ASS] Almgren, F.J., Jr., Schoen, R., Simon, L.: Regularity and singularity estimates on hypersurfaces minimizing parametric elliptic variational integrals. Acta Math **139**, 217–265 (1977)

[CS] Choi, H., Schoen, R.: The space of minimal embeddings of a surface into a three dimensional manifold of positive ricci curvature. Invent. Math. **81**, 387–394 (1985)

[GT] Gilbarg, D., Trudinger, N.: Elliptic partial differential equations of second order (2nd edition). Berlin-Heidelberg-New York: Springer 1983

[HS] Hardt, R., Simon, L.: Boundary regularity and embedded solutions for the oriented plateau problem. Ann. Math. **110**, 439–486 (1979)

[O] Osserman, R.: Global properties of minimal surfaces in E^3 and E^n. Ann. Math. **80**, 340–364 (1964)

[S] Simon, L.: Remarks on curvature estimates for minimal hypersurfaces. Duke Math. J. **43**, 545–553 (1976)

[SS1] Schoen, R., Simon, L.: Regularity of stable minimal hypersurfaces. Commun. Pure Appl. Math. **34**, 741–797 (1981)

[SS2] Schoen, R., Simon, L.: Regularity of simply connected surfaces with quasiconformal gauss map, in Seminar on minimal submanifolds. Ann. Math. Stud. **103**, 127–145 (1983)

[W1] White, B.: The space of k-dimensional surfaces that are stationary with respect to a parametric elliptic functional. Indiana Univ. Math. J. (in press) (1987)

[W2] White, B.: Complete surfaces of finite total curvature. J. Differ. Geom. (in press) (1987)

Oblatum 1-IV-1986

Invent. math. 88, 257–275 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

On subgroups of $GL_n(\mathbf{F}_p)$

Madhav V. Nori*

Tata Institute of Fundamental Research, School of Mathematics, Homi Bhabha Road, 400005 Bombay, India

Dedicated to Boris Weisfeiler

The aim of this paper is to understand subgroups of $GL_n(\mathbf{F}_p)$, where $\mathbf{F}_p$ is the prime field with p elements. To explain our results we need a few simple definitions.

Let G be a subgroup of $GL_n(\mathbf{F}_p)$. Let $X=\{x\in G \mid x^p=1\}$. Denote by $\tilde{G}$ the (connected) algebraic subgroup of GL_n, defined over $\mathbf{F}_p$, generated by the one-parameter subgroups

$$t\mapsto x^t=\exp(t\log x)$$

for all $x\in X$. The (normal) subgroup of G generated by X is denoted by G^+. Our main result (see Theorem B, § 3) says that if the prime p is greater than some constant that depends only on n, then $G^+=\tilde{G}(\mathbf{F}_p)^+$. If $\tilde{G}$ is semi-simple and simply connected, then $\tilde{G}(\mathbf{F}_p)=\tilde{G}(\mathbf{F}_p)^+$ and therefore $G^+=\tilde{G}(\mathbf{F}_p)$ in this case. In other words, G^+ is realized as the group of rational points of the connected algebraic group $\tilde{G}$.

The algebraic groups $\tilde{G}$ thus obtained are not arbitrary - by their very construction they are generated by a finite number of unipotent one-parameter subgroups of $(GL_n)_{\mathbf{F}_p}$ of the type:

$$t\mapsto \exp(ty)$$

where $y\in M_n(\mathbf{F}_p)$ satisfies $y^p=0$. Algebraic subgroups with the above property are said to be *exponentially generated*. Theorem B also shows that $G\mapsto\tilde{G}$ sets up a one-to-one correspondence between subgroups G of $GL_n(\mathbf{F}_p)$ such that $G=G^+$ and exponentially generated algebraic subgroups of $(GL_n)_{\mathbf{F}_p}$.

With G, X and $\tilde{G}$ as above, the Lie algebra of $\tilde{G}$ can be obtained directly from G. It is simply the linear span of $\{\log x \mid x\in X\}$, if p is sufficiently large with respect to n. We denote this linear span by $\langle\log G\rangle$. The Lie subalgebras $\langle\log G\rangle$ of $M_n(\mathbf{F}_p)$ thus obtained are *nilpotently generated*, in the sense that they are spanned by their nilpotent elements. We show (Theorem A, § 2) that nil-

* Research supported by the NSF Grant #MCS-8108814 (A04) through the Institute for Advanced Study, Princeton, NJ

potently generated Lie subalgebras of $M_n(F)$ are in one-to-one correspondence with exponentially generated algebraic subgroups of $(GL_n)_F$ where F is any field of sufficiently large characteristic.

The above results do not say enough about subgroups G of $GL_n(\mathbf{F}_p)$, when the prime p is large, because $G \neq G^+$ in general. However, p does not divide the order of G/G^+. By "lifting to characteristic zero" and using crucially a result of C. Jordan on finite subgroups of $GL_n(\mathbf{C})$, we get a grip on G/G^+ (see Theorem C, § 3).

The above statements have applications to two different problems, and were in fact motivated by them. The first is the following:

Given a subgroup π of $GL_n(\mathbf{Z})$, to what extent is the closure of π in the profinite group $GL_n(\hat{\mathbf{Z}})$ determined by the Zariski closure of π?

A detailed solution to this problem is given in § 5. Part of the results are also obtained by Matthews, Vaserstein and Weisfeiler (see [7]) who used the classification of finite simple groups.

The second motivation was the following. It looked plausible that an extension of the work of J-P. Serre [13] to the Galois action on the points of finite order on Abelian varieties would require, among other things, an understanding of the subgroups of $A(\mathbf{F}_p)$ where A is the Mumford-Tate group of the Abelian variety in question (see [11]). And in fact the theorems of this paper have been utilized by Serre precisely for this purpose (see his "Résumé des cours de 1985–1986", Annuaire du Collège de France, 1986).

Let G be a subgroup of $GL(V)$ where V a vector space of dimension n over $\mathbf{F}_p$ and let $G^+ \subset GL(V)$ and $\langle \log G \rangle \subset \operatorname{End} V$ be as before. We show (see Theorem D) that if $p > c(n)$, then there is a natural isomorphism of $H^1(G^+, V)$ with $H^1(\langle \log G \rangle, V)$. From this we deduce easily (see Theorem E) that if $p > c_3(n)$ and if the action of G on V is semi-simple, then $H^1(G, V)$ vanishes. A proof of Theorem E, different from the one given here, can be read off from the results of § 9–12 of [8].

Yet another application to Abelian varieties is the following. Theorem E (see § 4), combined with the celebrated work of G. Faltings [3], shows that the axioms of K. Ribet (see [10]) are valid for any Abelian variety defined over a number field. (Both the application and the statement of Theorem E itself were suggested to me by J-P. Serre.)

In spite of the group-theoretic nature of the results of this paper, the proof uses only the most rudimentary properties of schemes and very little of the theory of algebraic groups and Lie algebras (except perhaps for 3.5–3.7). This is not surprising because many statements concerning a field F are shown to be true when char. F is large, simply because these statements are true in characteristic zero. A drawback of this approach is that the constants C thus obtained are usually ineffective. And even when the proofs can be made effective, these constants appear to be too large.

A rough plan of the paper is as follows. In § 1, the basic properties of the exponential and logarithm in positive characteristic are dealt with. It is shown here that $\langle \log G \rangle$ is a Lie subalgebra, if $p \geqq 2n-1$ and $G \subset GL_n(\mathbf{F}_p)$. In § 2, schemes are used systematically to show that nilpotently generated Lie sub-

algebras and exponentially generated algebraic subgroups over a field of sufficiently large characteristic behave as they would in characteristic zero. The initial part of § 3 is devoted to the proof of the statement "$G^+ = \tilde{G}(\mathbf{F}_p)^+$" using a counting argument and a stratification obtained in § 2 (I have been informed by N. Katz that a similar proof of a similar theorem is due to O. Gabber, who used the Lang-Weil estimates instead to do his counting). The latter part of § 3 (viz. 3.6, 3.7 and Theorem C) is devoted to the study of G/G^+. The cohomology vanishing statement (Theorem E) is proved in § 4, and the final section contains the application to subgroups of $GL_n(\mathbf{Z})$.

I would like to thank Gopal Prasad, M.S. Raghunathan and T.N. Venkataramana with whom I had many discussions while the original version [8] of this paper was being worked out. I would also like to thank J-P. Serre who informed me of the application to Ribet's axioms, showed me the precise estimate of 3.5, helped me with many of the references, and gave me several tips as to how the presentation might be improved.

§ 0. Notation

E and F always denote fields; $\bar{F}$ denotes an algebraic closure of F. G is always a subgroup of $GL_n(F)$.

$\mathbf{A}^n$ denotes affine n-space over $\operatorname{Spec}\mathbf{Z}$.

$\mathbf{G}_a$ and $\mathbf{G}_m$ are the standard additive and multiplicative group-scheme over $\operatorname{Spec}\mathbf{Z}$.

If X is a scheme over $\operatorname{Spec}R$ and S is a R-algebra, the set of R-morphisms from $\operatorname{Spec}S$ to X is denoted by $X(S)$. The fibre-product $X \times_{\operatorname{Spec}R} \operatorname{Spec}S$ is often abbreviated to $X \times S$ or to X .

The letter A always denotes a closed subgroup-scheme of $(GL_n)_R$ where R is a commutative ring.

§ 1. Basic properties of exp and log

1.1. Preliminaries. Let B be an associative $\mathbf{F}_p$-algebra. Put

$$N = \{z \in B \mid z^p = 0\}$$

and

$$U = \{z \in B \mid z^p = 1\}.$$

Define exp and log by

$$\exp z = \sum_{i=0}^{p-1} \frac{z^i}{i!} \quad \text{and} \quad \log z = -\sum_{i=1}^{p-1} \frac{(1-z)^i}{i}.$$

Then exp: $N \to U$ and log: $U \to N$ are inverses of each other.

Note that if α and β are commuting elements of N, then $\alpha+\beta$ belongs to N, but $\exp(\alpha+\beta)$ does not coincide with $(\exp\alpha)(\exp\beta)$ in general. If in addition it is assumed that $\alpha^i \beta^{p-i} = 0$ whenever $1 \leqq i \leqq p-1$, then $\exp(\alpha+\beta) = (\exp\alpha)(\exp\beta)$.

If $B=M_n(F)$ where F is a field of characteristic $p>0$, then the above N and U will be denoted by $N_n(F)$ and $U_n(F)$. If char. $F=0$, then $N_n(F)$ and $U_n(F)$ are the sets of nilpotent and unipotent elements of $M_n(F)$ respectively.

Throughout this section F is a field of characteristic p containing the prime field $\mathbf{F}$. Thus $\mathbf{F}=\mathbf{Q}$ if $p=0$ and $\mathbf{F}=\mathbf{F}_p$ if $p>0$.

Lemma 1.2. *If $z\in N_n(F)$, then $\operatorname{ad} z$ is nilpotent. If $p>0$, then $(\operatorname{ad} z)^p=0$. If $p=0$ or if $p\geqq 2n-1$, then*

$$\exp(\operatorname{ad} z)=\operatorname{Ad}(\exp z).$$

Proof. For all $x\in M_n(F)$, denote by L_x and R_x, the operators on $M_n(F)$ given by left and right multiplication by x respectively. Recall that $\operatorname{ad} x=L_x-R_x$ and if $x\in GL_n(F)$, then $\operatorname{Ad} x=L_xR_x^{-1}$. Since the above lemma is standard when $p=0$, we shall assume that $p>0$. If $z\in N_n(F)$, $(\operatorname{ad} z)^p=(L_z-R_z)^p=0$ because L_z and R_z commute. Furthermore, it is clear that $\exp L_z=L_{\exp z}$ and $\exp R_z=R_{\exp z}$, so the last assertion of the lemma is simply $\exp(L_z-R_z)=(\exp L_z)(\exp(-R_z))$. If $p\geqq 2n-1$, then $L_z^iR_z^{p-1}=0$ whenever $1\leqq i\leqq p-1$ and from the remarks in 1.1 above, the lemma follows.

Definition 1.3. Let S be a subset of $M_n(F)$. The subgroup of $GL_n(F)$ generated by $\{\exp s\,|\,s\in N_n(F)\cap S\}$ is denoted by $\langle\exp S\rangle$.

Let G be a subgroup of $GL_n(F)$. Then the subgroup of G generated by $G\cap U_n(F)$ is denoted by G^+ and the $\mathbf{F}$-linear span of $\{\log g\,|\,g\in G\cap U_n(F)\}$ is denoted by $\langle\log G\rangle$.

Lemma 1.4. *Let $W_1\subset W_2\subset F^n$ be $\mathbf{F}$-linear subspaces and let S be a subset of $N_n(F)$. Then A and B below are equivalent.*

(A) *$sW_2\subset W_1$ for all $s\in S$.*

(B) *W_1 and W_2 are stable under $\langle\exp S\rangle$ and the action of $\langle\exp S\rangle$ on W_2/W_1 is trivial.*

The above lemma is an immediate consequence of the fact that exp and log are polynomials (in this context) with coefficients in $\mathbf{F}$ with constant term equal to 1 and 0 respectively.

Lemma 1.5. *Let $W_1\subset W_2\subset M_n(F)$ be $\mathbf{F}$-linear subspaces and let $S\subset N_n(F)$. If $p\geqq 2n-1$ or if $p=0$, then A and B below are equivalent:*

(A) *$(\operatorname{ad} s)W_2\subset W_1$ for all $s\in S$.*

(B) *W_1 and W_2 are both stable under the adjoint action of $\langle\exp S\rangle$ and the induced action of $\langle\exp S\rangle$ on W_2/W_1 is trivial.*

Proof. Define $\bar{S}\subset\operatorname{End}_F(M_n(F))$ by $\bar{S}=\{\operatorname{ad} s\,|\,s\in S\}$. By Lemma 1.2, $\langle\exp\bar{S}\rangle=\{\operatorname{Ad} h\,|\,h\in\langle\exp S\rangle\}$. The result now follows from Lemma 1.4.

Lemma 1.6. *Let G be a subgroup of $GL_n(F)$. If $p\geqq 2n-1$, or if $p=0$, then $\langle\log G\rangle$ is a $\mathbf{F}$-Lie subalgebra of $M_n(F)$.*

Proof. Put $S=\{\log g\,|\,g\in G\cap U_n(F)\}$. Then $G^+=\langle\exp S\rangle$ and since $\langle\log G\rangle\subset M_n(F)$ is clearly stable under $\operatorname{Ad} g$ for all $g\in G^+$, from Lemma 1.5 we see that $[s,\langle\log G\rangle]\subset\langle\log G\rangle$ for all $s\in S$. The lemma follows from this because $\langle\log G\rangle$ is the $\mathbf{F}$-linear span of S.

Remark 1.7. Let A be an algebraic subgroup of GL_n defined over a field F. If $\operatorname{char} F>0$, $\langle \log A(F)\rangle$ is not a F-Lie subalgebra of $M_n(F)$ in general. However, if A is generated by its unipotents and if $\operatorname{char} F=0$, then $\langle \log A(F)\rangle = \operatorname{Lie} A$ (see 1.12), and this motivates the various definitions of § 2.

Remark 1.8. Let G be a subgroup of $GL_n(F)$ and let B be the $\bar{F}$-linear span of G^+ in $M_n(\bar{F})$, where $\bar{F}$ is an algebraic closure of F. To show that $\langle \log G\rangle$ is a $\mathbf{F}$-Lie subalgebra, it suffices to assume that the operators $L_z^i R_z^{p-i}$ vanish on B for all $z\in N_n(F)$ such that $\exp z\in G$, whenever $1\leqq i\leqq p-1$; and for this to hold it suffices to assume that $p\geqq 2n-m$ where m is the length of the Artinian B-module $\bar{F}^n$. This is therefore a mild strengthening of Lemma 1.6.

In the extreme case where $m=n$, G^+ is contained in the unipotent radical of a Borel subgroup. And here, if $p\geqq n$, the set $\{\log g \mid g\in G^+\}$ is itself a $\mathbf{F}$-Lie subalgebra of $M_n(F)$ as is seen easily from the Campbell-Hausdorff formula (see § 10, [8], for further details).

Lemma 1.9. *Let $S\subset N_n(F)$ and let L be the smallest* $\mathbf{F}$-Lie *subalgebra of $M_n(F)$ containing S. Then L is the* $\mathbf{F}$-*linear span of $L\cap N_n(F)$, if $p=0$ or if $p\geqq 2n-1$.*

Proof. Let $T=\{g(\exp s)g^{-1}\mid g\in\langle \exp S\rangle \text{ and } s\in S\}$. Let W be the $\mathbf{F}$-linear span of $\{\log t\mid t\in T\}$. Clearly W is the smallest $\mathbf{F}$-linear subspace of $M_n(F)$ containing S and stable under the adjoint action of $\langle \exp S\rangle$. From Lemma 1.5 it follows that W is the smallest $\mathbf{F}$-subspace of $M_n(F)$ containing S and stable under $\operatorname{ad} s$ for all $s\in S$. In other words, $W=L$, and this proves the lemma.

Definition 1.10. A F-Lie subalgebra L *of* $M_n(F)$ is *nilpotently generated* if L is the F-linear span of $L\cap N_n(F)$. For every $z\in L\cap N_n(F)$ we have $f_z\colon \mathbf{A}_F^1\to (GL_n)_F$ defined over F given by $f_z(t)=\exp(tz)$. The algebraic subgroup of $(GL_n)_F$ generated by the f_z for all $z\in L\cap N_n(F)$ will be denoted by $\exp L$. That $\exp L$ is connected follows from Proposition 2.2, p. 106, [1].

Remark. It is also natural to call a Lie subalgebra L of $M_n(F)$ nilpotently generated if L is generated as a Lie subalgebra by $L\cap N_n(F)$. In view of Lemma 1.9 however, if $p=0$ or if $p\geqq 2n-1$, this definition of "nilpotently generated" coincides with 1.10.

Proposition 1.11. *If $M=\operatorname{Lie}(\exp L)$ where L is a nilpotently generated F-Lie subalgebra of $M_n(F)$, then*

(A) *$M\supset L$ and $[M,M]=[L,L]$ if $p\geqq 2n-1$*

and

(B) *$M=L$ if $p=0$.*

Proof. From the construction of $\exp L$ we see that $N_n(F)\cap L$ is contained in M. Because L is nilpotently generated, it follows that $L\subset M$.

If $W_1\subset W_2\subset M_n(F)$ are F-linear subspaces such that $[L,W_2]\subset W_1$, then by Lemma 1.5 essentially, W_1 and W_2 are stable under the action of $\exp L$ and the action of $\exp L$ on W_2/W_1 is trivial. From this it follows that $[M,W_2]\subset W_1$.

Putting $W_1=[L,L]$ and $W_2=L$, we get $[M,L]\subset[L,L]$. Now putting $W_1=[L,L]$ and $W_2=M$, we get $[M,M]\subset[L,L]$. This proves part (A) of the lemma.

Put $A=(\exp L)_{ab}$ and assume that $\operatorname{char} F$ is zero. It is well known that $\operatorname{Lie} A = M/[M, M]$; see e.g. Proposition 7.8, p. 193, [1]. For $z \in N_n(F) \cap L$, the composites

$$\mathbf{A}_F^1 \xrightarrow{f_z} \exp L \longrightarrow A$$

generate the algebraic group A, and again because $\operatorname{char} F = 0$, their images on tangent-spaces generate $\operatorname{Lie} A$. In other words, $L + [M, M] = M$. Since $L \supset [M, M]$, we see that $L = M$. This proves part (B) of the Lemma.

Corollary 1.12. *With L and F as in 1.11, assume that* $\operatorname{char} F$ *is zero and let E be any field containing F. Let $z \in N_n(E)$. Then*

$$z \in L \otimes_F E \Leftrightarrow \exp z \in (\exp L)(E).$$

Proof. First note that $\exp(L \otimes_F E) = (\exp L)_E$ because the Lie algebras of these algebraic groups coincide from 1.11(B); see (7.1), p. 186, [1]. Thus it suffices to prove the result when $E = F$.

If $\exp z$ is a rational point of $\exp L$, then so is $\exp(kz)$ for all $k \in \mathbf{Z}$. With the morphism $f_z: \mathbf{A}_F^1 \to (GL_n)_F$ as in 1.10, the Zariski-closed subset $f_z^{-1}(\exp L)$ contains $\mathbf{Z}$ and therefore $f_z^{-1}(\exp L) = \mathbf{A}_F^1$, from which $z \in \operatorname{Lie}(\exp L)$. By 1.11(B), we see that $z \in L$.

If $z \in N_n(F) \cap L$, then $\exp z$ is a rational point of $\exp L$ by the very construction of $\exp L$. This completes the proof of 1.12.

It is hoped that 1.11(B) and 1.12 are also true whenever $p \geqq 2n-1$. However, we do not know a proof of this now. In the next section, by routine scheme-theoretic methods, it will be shown that 1.11(B) and 1.12 are valid whenever p is sufficiently large with respect to n.

§ 2. Families of Lie subalgebras

The main result of this section is Theorem A which sets up a one-to-one correspondence between nilpotently generated Lie subalgebras of $M_n(F)$ and exponentially generated algebraic subgroups of $(GL_n)_F$ if the characteristic of the field F is sufficiently large. In the process we also prove 1.11(B) and 1.12 under such assumptions on F.

A natural number n will remain fixed throughout this section. Put $C = \mathbf{Z}[1/(n-1)!]$. All commutative rings (resp. *schemes*) *considered in this section are C-algebras* (resp. *schemes*/Spec C).

2.1. (*Notation*). For any commutative ring R, put

$$N_n(R) = \{w \in M_n(R) \mid w^n = 0\}$$

and

$$U_n(R) = \{x \in M_n(R) \mid (x-1)^n = 0\}.$$

In view of our hypothesis on R, this is consistent with the notation introduced in 1.1. Again $\exp: N_n(R) \to U_n(R)$ and $\log: U_n(R) \to N_n(R)$ are inverses of each other.

Definition 2.2. A R-submodule L of $M_n(R)$ is a *strict Lie subalgebra* of $M_n(R)$ if

(a) $M_n(R)/L$ is a projective R-module

and

(b) $l_1, l_2 \in L \Rightarrow [l_1, l_2] = l_1 l_2 - l_2 l_1 \in L$.

If $M_n(R)/L$ is locally free of rank $n^2 - k$, then L is a strict *Lie subalgebra of dimension* k. In particular, L is then a projective R-module of rank k.

Definition 2.3. Let L be a strict Lie subalgebra of $M_n(R)$ and let A be a closed subgroup-scheme of $(GL_n)_R = GL_n \times \operatorname{Spec} R$. Then (L, A) is an *acceptable pair* if (a), (b) and (c) below are true:

(a) The projection $A \to \operatorname{Spec} R$ is a smooth morphism with all fibres connected.

(b) $\operatorname{Lie}(A/R) = L$.

(c) For all fields F, for all homomorphisms $R \to F$, and for all $w \in N_n(F)$, we have:

$$w \in L \otimes_R F \Leftrightarrow \exp w \in A(F).$$

L is *acceptable* if there exists A so that (L, A) is an acceptable pair.

Definition 2.4. A closed subgroup-scheme A of $(GL_n)_F$ is *exponentially generated* if there is a subset S of $N_n(F)$ such that A is generated by the one-parameter groups $f_w(t) = \exp tw$ for all $w \in S$. Here F is a field.

Theorem A. *There is a natural number* $c_1(n) \geqq 2n-2$ *so that whenever* F *is a field of characteristic* p *and* $p = 0$ *or* $p > c_1(n)$, *then* (1) *and* (2) *below are true:*

(1) *If* L *is a nilpotently generated Lie subalgebra of* $M_n(F)$, *then* $(L, \exp L)$ *is an acceptable pair, and in particular* $L = \operatorname{Lie}(\exp L)$.

(2) *If* $A \subset (GL_n)_F$ *is exponentially generated, then* A *is smooth and* $\operatorname{Lie} A = \log A(F)$ *and furthermore* $A = \exp(\operatorname{Lie} A)$.

Thus $L \mapsto \exp L$ sets up a one-to-one correspondence between nilpotently generated Lie subalgebras of $M_n(F)$ and exponentially generated subgroup-schemes of GL_n defined over F. The above theorem will be proved later in this section, after the proof of 2.12.

Notation. Let $z = (z_1, \ldots, z_k) \in N_n(R)^k = N_n(R) \times \ldots \times N_n(R)$ k times, and put $L(z) = \sum_{i=1}^{k} R z_i$.

Definition 2.5. $Y_{n,k}(R) = \{z \in N_n(R)^k \mid L(z)$ is a k-dimensional strict Lie subalgebra of $M_n(R)$ in the sense of Definition 2.2)$\}$.

Remark 2.6. The various functors from C-algebras R to sets given by $R \mapsto M_n(R)$, $N_n(R)$, $U_n(R)$, $Y_{n,k}(R)$ are all clearly representable by schemes M_n, N_n, U_n and $Y_{n,k}$, where M_n is affine n^2-dimensional space over Spec C, N_n is a closed subscheme of M_n, U_n is a closed subscheme of $(GL_n)_C$, and finally $Y_{n,k}$ is a locally closed subscheme of $(N_n)^k$. Also exp: $N_n \to U_n$ and log: $U_n \to N_n$ are inverses of each other.

The main step in proving Theorem A is to establish Proposition 2.7 below. It should be noted that $z \in Y_{n,k}(R)$ can also be identified with a morphism z: Spec $R \to Y_{n,k}$.

Proposition 2.7. *There are integral domains R^λ and $z^\lambda \in Y_{n,k}(R^\lambda)$ indexed by $\lambda \in \Lambda$, where Λ is a finite set, such that*

(a) *for all $\lambda \in \Lambda$, the strict Lie subalgebra $L(z^\lambda)$ of $M_n(R^\lambda)$ is acceptable or R^λ has positive characteristic,*

(b) *for all $\lambda \in \Lambda$, z^λ: $\operatorname{Spec} R^\lambda \to Y_{n,k}$ is a locally closed immersion, and the sets $z^\lambda(\operatorname{Spec} R^\lambda)$ are disjoint and their union is $Y_{n,k}$.*

For Proposition 2.7 we need the two lemmas below. The proof of Lemma 2.8 we shall postpone, while Lemma 2.9 is merely a restatement of 1.11(B) and 1.12.

Lemma 2.8. *If R is a Noetherian integral domain with quotient field K and L is a strict Lie subalgebra of $M_n(R)$ such that $L \times_R K$ is acceptable, then there exists $g \in R$, $g \neq 0$ such that $L_g \subset M_n(R_g)$ is acceptable.*

Lemma 2.9. *If L is a nilpotently generated Lie subalgebra of $M_n(F)$ where F is a field of characteristic zero, then the pair $(L, \exp L)$ is acceptable.*

Proof of Proposition 2.7. We shall define a strictly decreasing sequence of closed subsets $(Y_{n,k})_{\mathrm{red}} = V^0 \supset V^1 \supset \ldots \supset V^m = \emptyset$ as follows. Assume that V^j has been defined an $V^j \neq \emptyset$. Then let $\operatorname{Spec} R$ be and irreducible affine open in V^j. If $\operatorname{char} R > 0$, let V^{j+1} be the complement of $\operatorname{Spec} R$ in V^j. If $\operatorname{char} R = 0$, denote by z: $\operatorname{Spec} R \to Y_{n,k}$ the inclusion morphisms, and by 2.9 and 2.8 above $\exists g \in R$, $g \neq 0$, such that $L(z)_g \subset M_n(R_g)$ is acceptable. Put $V^{j+1} = V^j - (\operatorname{Spec} R_g)$.

Because $Y_{n,k}$ is a Noetherian scheme, $V^m = \emptyset$ for some natural number m, and then the locally closed subsets $V^j - V^{j+1}$ of $Y_{n,k}$ for $0 \leqq j \leqq m-1$ give the desired stratification of $Y_{n,k}$.

Lemma 2.10. *If $L \subset M_n(R)$ is an acceptable Lie subalgebra, then $L \otimes_R S \subset M_n(S)$ is also acceptable for all* (commutative) *R-algebras S.*

Lemma 2.10 is immediate from Definition 2.3.

Corollary 2.11. *For evrery pair of natural numbers (n, k), there is a constant $c_1(n, k) \geqq n-1$ with the following property. If F is a field of characteristic $> c_1(n, k)$ and if L is a k-dimensional nilpotently generated* Lie *subalgebra of $M_n(F)$, then L is acceptable.*

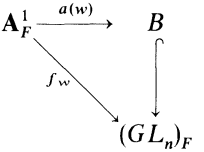
Proof. With the notation of Proposition 2.7, let $c_1(n, k)$ be any natural number such that $c_1(n, k) \geqq n-1$ and $c_1(n, k) \geqq \operatorname{char} R^\lambda$, for all $\lambda \in \Lambda$. Now given F and L as in the corollary, choose $z_1, \ldots, z_k \in N_n(F)$ so that the z_i form a basis for L. Then clearly $z = (z_1, \ldots, z_k) \in Y_{n,k}(F)$. By Proposition 2.7, there exist $\lambda \in \Lambda$ and s: $R^\lambda \to F$ such that $z = s(z^\lambda)$. Because $\operatorname{char} F > \operatorname{char} R^\lambda$ it follows that $\operatorname{char} R^\lambda = 0$ and therefore $L(z^\lambda)$ is acceptable. By Lemma 2.10 above, $L(z)$ is itself acceptable.

Lemma 2.12. *Let (L, B) be an acceptable pair over a field F, and assume that L is nilpotently generated. Then $B = \exp L$. Moreover, if $z_1, z_2, \ldots, z_k \in N_n(F) \cap L$ form a basis for L then B is generated by the one-parameter groups $t \to \exp(t z_i)$, for $1 \leqq i \leqq k$.*

An immediate consequence of 2.11 and 2.12 is the corollary below.

Corollary 2.13. *If L and F are as in 2.11, then the pair $(L, \exp L)$ is acceptable.*

Proof of 2.12. For every $w \in N_n(F) \cap L$ and for every $t \in E$ where $E \supset F$, $\exp(tw) \in B(E)$ by Definition 2.3 and therefore there is a commutative diagram



where $f_w(t) = \exp(t_w)$. It follows that $B \subset \exp L$.

Now if $z_i \in N_n(F) \cap L$, $1 \leqq i \leqq k$, form a basis for L, define $\phi: \mathbf{A}^k_F \to B$ by $\phi(t_1, \ldots, t_k) = (a(z_1)t_1) \ldots (a(z_k)t_k))$. Because ϕ clearly induces an isomorphism of tangent spaces at $0 \in \mathbf{A}^k(F) = F^k$ and because $\mathbf{A}^k_F$ and B are both smooth it follows that ϕ is étale at $0 \in \mathbf{A}^k(F)$. Consequently, there is a non-empty Zariski-open $V \subset B$ such that $V \subset \text{Image } \emptyset$. Because B is connected, V generates B as an algebraic group (see Proposition 1.3(a), p. 87, [1]) and this proves the second assertion of the lemma.

Proof of Theorem A. We take $c_1(n)$ to be the maximum of $2n-2$ and the $c_1(n, k)$ for $1 \leqq k \leqq n^2 - 1$, where the $c_1(n, k)$ are as in Corollary 2.11. Part (1) of the theorem then follows from (2.13).

Now let $A \subset (GL_n)_F$ be exponentially generated and let S be the subset of $N_n(F)$ consisting of those w for which the $f_w(t) = \exp(tw)$ are contained in A. Let L be the F-linear span of S. Clearly $L \subset \langle \log A(F) \rangle$. Moreover, because L is clearly invariant under $\text{Ad } A(F)$, by Lemma 1.5 we deduce that $[\langle \log A(F) \rangle, L] \subset L$. In particular, L is a nilpotent generated Lie subalgebra and by 2.13, the pair $(L, \exp L)$ is acceptable.

The one-parameter groups $\exp(tw)$, $w \in S$, generate A by Definition 2.4, and on the other hand generate $\exp L$ by the second assertion of Lemma 2.12 and therefore $A = \exp L$. Now part (2) of Theorem A follows from the fact that (L, A) is an acceptable pair. This completes the proof of Theorem A modulo Lemma 2.8 and we take up the proof of this lemma after the next remark.

Remark 2.14. We want to examine condition 2.3(c) more closely. Let (L, A) be an acceptable pair over R. The functor $S \mapsto L \otimes_R S$ defined for all commutative R-algebras S is clearly representable by a closed subscheme $\mathbf{L}$ of $(M_n)_R$. We put $P = \exp(\mathbf{L} \cap (N_n)_R)$ and $Q = A \cap (U_n)_R$. These are closed subschemes of $(GL_n)_R$ and condition (c) of Definition 2.3 says that $P(F) = Q(F)$ for all fields F. Thus (c) is equivalent to (d) below:

(d) $P_{\text{red}} = Q_{\text{red}}$.

Proof of Lemma 2.8. Assume that $(L \otimes_R K, H)$ is an acceptable pair and let A be the Zariski-closure of H in $(GL_n)_R$. Because H is smooth and geometrically irreducible, $\exists\, \alpha \in R$, $\alpha \neq 0$, so that the projection $A_\alpha \to \text{Spec } R_\alpha$ is a smooth morphism with fibres geometrically irreducible (see Theorem 9.7.7(i), p. 79, and Theorem 12.2.4(iii), p. 183 of EGA IV, Troisième Partie).

Because $\text{Lie}(A_\alpha/R_\alpha) \otimes_R K = L \otimes_R K$ there is a non-zero $\beta \in R$ such that $\text{Lie}(A_{\alpha\beta}/R_{\alpha\beta}) = L_{\alpha\beta}$.

Finally, if P and Q are the closed-subschemes associated to the pair (L_α, A_α), then $(P_{\text{red}})_\gamma = (Q_{\text{red}})_\gamma$ for some non-zero $\gamma \in R$. Putting $g = \alpha\beta\gamma$, the proof of Lemma 2.8 is now complete.

§ 3. Subgroups of $GL_n(\mathbf{F}_p)$

Let G be a subgroup of $GL_n(\mathbf{F}_p)$. The normal subgroup G^+ of G has been defined in 1.3. We denote by $\tilde{G}$ the algebraic subgroup generated by the one-parameter groups $t \mapsto x^t = \exp(t \log x)$ for all $x \in G$ such that $x^p = 1$.

Theorem B. *For every natural number n there is a constant $c_2(n) \geqq 2n-2$ such that statements* (1) *and* (2) *below are true for all primes $p > c_2(n)$:*

(1) *If G is a subgroup of $GL_n(\mathbf{F}_p)$, then $G^+ = \tilde{G}(\mathbf{F}_p)^+$ and $\langle \log G \rangle = \operatorname{Lie} \tilde{G}$.*

(2) *If $A \subset (GL_n)_{\mathbf{F}_p}$ is exponentially generated, then $A = \widetilde{A(\mathbf{F}_p)}$.*

Remark 3.1. Thus the above theorem says that $G \mapsto \tilde{G}$ sets up a one-to-one correspondence between the collection $\mathcal{G}$ of subgroups G of $GL_n(\mathbf{F}_p)$ such that $G = G^+$ and the collection $\mathcal{A}$ of exponentially generated algebraic subgroups of GL_n defined over $\mathbf{F}_p$. In view of Theorem A, $G \mapsto \langle \log G \rangle$ also sets a $1-1$ correspondence between $\mathcal{G}$ and the collection of nilpotently generated Lie subalgebras of $M_n(\mathbf{F}_p)$.

Remark 3.2. To compare $G \subset GL_n(\mathbf{F}_p)$ and $\tilde{G}(\mathbf{F}_p)$, we need to complement Theorem B with some information on G/G^+ and $\tilde{G}(\mathbf{F}_p)/\tilde{G}(\mathbf{F}_p)^+$, and this is done once Theorem B has been proved.

Proof of Theorem B. With $c_1(n)$ as in Theorem A, we shall always assume that p is a prime $> c_1(n)$.

If $A \subset (GL_n)_{\mathbf{F}_p}$ is exponentially generated and $L = \langle \log A(\mathbf{F}_p) \rangle$, by Theorem A we know that $A = \exp L$ and $(L, \exp L)$ is an acceptable pair. In particular, if $a^p = 1$ and $a \in A(\mathbf{F}_p)$, then $\log a \in L$ and therefore the group $t \mapsto a^t$ is contained in $\exp L$. On the other hand these one-parameter groups generate $\exp L$, so $\exp L = \widetilde{A(\mathbf{F}_p)}$, and this proves Theorem B, part (2).

We shall next show that $\operatorname{Lie} \tilde{G} = \langle \log G \rangle$ where G is a subgroup of $GL_n(\mathbf{F}_p)$. Let $k = \dim \langle \log G \rangle$ and choose $x_1, \ldots, x_k \in G$ such that $x_i^p = 1$ for all i and $\log x_1, \ldots, \log x_k$ form a basis for $\langle \log G \rangle$. By Theorem A, if $L = \langle \log G \rangle$, then $(L, \exp L)$ is an acceptable pair, and by Lemma 2.12, we see that $\exp L$ is generated by the one-parameter groups $t \mapsto x_i^t$ for $1 \leqq i \leqq k$. In particular, $\exp L \subset \tilde{G}$. But the reverse inclusion is clear, so we get $\exp L = \tilde{G}$, and $\operatorname{Lie} \tilde{G} = L$ follows from the acceptability of $(L, \exp L)$.

With the x_i as above, consider the subset T of G^+ given by $T = \{x_1^{a_1} \ldots x_k^{a_k} \mid a_1, a_2, \ldots, a_k \in \mathbf{F}_p\}$. We shall obtain a lower bound on the cardinality of T, which will then show that $[\tilde{G}(\mathbf{F}_p)^+ : G^+] < p$, if p is sufficiently large. From this inequality, it follows that $\tilde{G}(\mathbf{F}_p)^+ = G^+$, because if X denotes the coset-space $\tilde{G}(\mathbf{F}_p)^+/G^+$, then the left-action of every $u \in \tilde{G}(\mathbf{F}_p)^+$ of order p on X is necessarily trivial. And since such u generate $\tilde{G}(\mathbf{F}_p)^+$ it follows that the (transitive) action of $\tilde{G}(\mathbf{F}_p)^+$ on X is itself trivial, and therefore $G^+ = \tilde{G}(\mathbf{F}_p)^+$.

To estimate the cardinality of T, we use the morphism $\phi\colon \mathbf{A}^k_{\mathbf{F}_p} \to \tilde{G}$ in the proof of Lemma 2.12 given by

$$\phi(t_1, \ldots, t_k) = x_1^{t_1} x_2^{t_2} \ldots x_k^{t_k}.$$

Note first that $T = \phi(\mathbf{A}^k(\mathbf{F}_p)) = \phi(\mathbf{F}_p^k)$. Next note that ϕ is a separable morphism because it induces an isomorphism of the tangent-space of $0 \in \mathbf{A}^k(\mathbf{F}_p)$ with the tangent-space of $\tilde{G}$ at $1 \in \tilde{G}$ at $1 \in \tilde{G}(\mathbf{F}_p)$. Denote by $J \in \mathbf{F}_p[t_1, \ldots, t_k]$ the Jacobian of this morphism defined as follows. Let $X_1, \ldots, X_k$ be the basis of $(\mathrm{Lie}\,\tilde{G})^*$ dual to the basis $\log x_1, \ldots, \log x_k$ of $\mathrm{Lie}\,\tilde{G}$ and then extend the X_i to global right-invariant one-forms $\bar{X}_i$ on $\tilde{G}$ and finally define J by

$$\phi^*(\bar{X}_1 \wedge \bar{X}_2 \wedge \ldots \wedge \bar{X}_k) = J(dt_1 \wedge dt_2 \wedge \ldots \wedge dt_k).$$

Clearly $J(0) = 1$. Denote by m the degree of the polynomial J and by W the principal open subset of $\mathbf{A}^k_{\mathbf{F}_p}$ given by "$J \neq 0$". Now let $\# S$ denote the cardinality of a set S. By Lemma 3.3 below, $\# W(\mathbf{F}_p) \geqq p^k(1 - (m/p))$. Since $\phi | W$ is an étale morphism, if d denotes the degree of $\emptyset$, then all the fibres of $\phi | W$ have cardinality $\leqq d$ (from an application of Zariski's Main Theorem, or see Lemma 3.4 below), and therefore we get

$$\# \phi(W(\mathbf{F}_p)) \geqq d^{-1} p^k \left(1 - \frac{m}{p}\right).$$

The inclusion $G^+ \supset T \supset \phi(W(\mathbf{F}_p))$ combined with the estimate 3.5 on $\tilde{G}(\mathbf{F}_p)$ now give

$$[\tilde{G}(\mathbf{F}_p)\colon G^+] \leqq d \left(1 - \frac{m}{p}\right)^{-1} \left(1 + \frac{1}{p}\right)^k.$$

We shall soon obtain constants D and M that depend only on n such that $d \leqq D$ and $m \leqq M$. Since $0 \leqq k \leqq n^2 - 1$, this gives

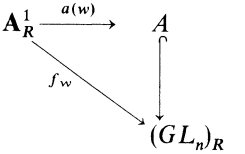
$$[\tilde{G}(\mathbf{F}_p)\colon G^+] \leqq D \left(1 - \frac{M}{p}\right)^{-1} \left(1 + \frac{1}{p}\right)^{n^2 - 1} = h(p)$$

and since $\lim_{p \to \infty} h(p) = D$, it follows that there exists a constant $c_2(n) \geqq c_1(n)$ such that

$$p > c_2(n) \Rightarrow h(p) < p.$$

As we have remarked earlier, this shows that $\tilde{G}(\mathbf{F}_p)^+ = G^+$ if $p > c_2(n)$.

To prove the existence of D and M as above one needs to define the morphism ϕ in greater generality, for which we need some preliminaries. Let R be an integral domain and let $z = (z_1, \ldots, z_k) \in Y_{n,k}(R)$ and assume that $(L(z), A)$ is an acceptable pair (see Definition 2.3 and 2.5). First note that $\forall\, w \in L(z) \cap N_n(R)$ we have:



where $f_w(t)=\exp tw$. To see this, by the definition of P and Q in 2.14, f_w factors through the inclusion of P in $(GL_n)_R$, and because $P_{\text{red}}=Q_{\text{red}}$ and because $\mathbf{A}^1_R$ is a reduced scheme, f_w factors through the inclusion of Q. Finally $Q\subset A$, from which we deduce the above commutative diagram.

Now define $\phi(z)\colon \mathbf{A}^k_R\to A$ by

$$\phi(z)(t_1,\ldots,t_k)=(a(z_1)t_1)(a(z_2)t_2)\ldots(a(z_k)t_k).$$

We let X_i be the basis of $\mathrm{Hom}_R(\mathrm{Lie}(A/R),R)$ dual to the basis z_i of $\mathrm{Lie}(A/R)$ and extend the X_i to global right-invariant sections $\bar{X}_i$ of $\Omega_{A/R}$ and define $J(z)\in R[t_1,\ldots,t_k]$ by

$$\phi(z)^*(\bar{X}_1\wedge\bar{X}_2\wedge\ldots\wedge\bar{X}_k)=J(z)(dt_1\wedge dt_2\wedge\ldots\wedge dt_k).$$

We put $m(z)=\text{degree}\,J(z)$ and denote by $d(z)$ the degree of the morphism $\phi(z)$. Note that $\phi(z)|W(z)$ is an étale morphism where $W(z)$ is the principal open subset of $\mathbf{A}^k_R$ where $J(z)$ is invertible.

We now put $R=\mathbf{F}_p$ and $w_i=\log x_i$ for $1\leqq i\leqq k$ where the x_i are the elements of G chosen earlier. Associated to $w=(w_1,\ldots,w_k)\in Y_{n,k}(\mathbf{F}_p)$ are the $\phi(w)$, $J(w)$, $W(w)$, $m(w)$ and $d(w)$; and these clearly coincide with ϕ, J, W, m and d used to estimate $\#T$.

By Proposition 2.7, $\exists\,\lambda\in\Lambda$ and $s\colon R^\lambda\to\mathbf{F}_p$ such that $w=s(z^\lambda)$. Because $J=J(w)=s(J(z^\lambda))$, it follows that $m=m(w)\leqq m(z^\lambda)$.

Also $\phi(z^\lambda)|W(z^\lambda)\colon W(z^\lambda)\to A$ is a morphism of schemes over $\mathrm{Spec}\,R^\lambda$ and its base-change by $s\colon \mathrm{Spec}\,\mathbf{F}_p\to\mathrm{Spec}\,R^\lambda$ is simply $\phi|W\colon W\to\tilde{G}$. From Lemma 3.4 below it follows that $d\leqq d(z^\lambda)$.

So we take $d(n,k)=\max\{d(z^\lambda)|\lambda\in\Lambda\}$ with the notation of notation of Proposition 2.7, and then put $D=\max\{d(n,k)|1\leqq k\leqq n^2-1\}$. Similarly defining M, the proof of Theorem B is now complete modulo Lemmas 3.3, 3.4 and 3.5.

Lemma 3.3. *If $h\in\mathbf{F}_p[t_1,\ldots,t_k]$ is a non-zero polynomial of degree m and if $S=\{a\in\mathbf{F}_p^k|h(a)=0\}$, then $\#S\leqq mp^{k-1}$.*

Lemma 3.3 is elementary. For a proof, see Lemma 3A, p. 147, [16].

Lemma 3.4. *If $f\colon X\to Y$ is an étale morphism where X and Y are reduced irreducible Noetherian schemes, then for all $y\in Y$ we have:*

$$\sum_{x\in f^{-1}(y)}[k(x)\colon k(y)]\leqq\text{degree}\,f,$$

where $k(z)$ denotes the residue field of a point z of a scheme Z.

In the above lemma, degree f is $[k(x)\colon k(y)]$ where x and y are the generic points of X and Y respectively. For a stronger version of this lemma, see Proposition 10.9, p. 24, [17].

Lemma 3.5. *Let A be a connected algebraic group of dimension k, defined over $\mathbf{F}_p$. Then*

$$(p-1)^k\leqq o(A(\mathbf{F}_p))\leqq(p+1)^k,$$

where $o(A(\mathbf{F}_p))$ *denotes the order of the finite group $A(\mathbf{F}_p)$.*

Proof. Put $r(A)=o(A(\mathbf{F}_p))p^{-k}$. If A is unipotent, then $r(A)=1$ and so the lemma holds here. If A is semi-simple, it is known that (see e.g. Lemma 11.16, p. 79, [14]) there are roots of unity $z_1, \ldots, z_l$ and natural numbers $m_1, \ldots, m_l$ where l is the rank of A, such that

$$o(A(\mathbf{F}_p))=\prod_{i=1}^{l}(1-z_i p^{-m_i})$$

thus proving the lemma in this case too. The above formula also holds when A is a torus with $k=l=\dim A$ and $m_i=1$ for all i and the z_i being the eigenvalues of the Frobenius in $\mathrm{Gal}(\bar{\mathbf{F}}_p/\mathbf{F}_p)$ acting on the group of characters of A.

Thanks to a result of S. Lang (see [6]), if W is a connected algebraic subgroup of A, then $A(\mathbf{F}_p)/W(\mathbf{F}_p)$ is in fact the group of rational points of A/W, and therefore the lemma follows from the three special cases of it discussed earlier.

Remark. The statement and proof of the above lemma were shown to me by J-P. Serre. However, in the interest of keeping the proof of Theorem B at an absolutely elementary level, it should be noted that by stratifying $Y_{n,k}$ as in Proposition 2.7 and using the Noether normalization lemma we get easily the inequality $o(A(\mathbf{F}_p))\leqq cp^k$ where c is a constant depending only on $n, k=\dim A$, and A is an exponentially generated algebraic subgroup of $(GL_n)_{\mathbf{F}_p}$, and $p>c_1(n)$. This inequality clearly suffices for the proof of Theorem B.

Remark 3.6. We now examine the group $\tilde{G}(\mathbf{F}_p)/\tilde{G}(\mathbf{F}_p)^+$ where $G\subset GL_n(\mathbf{F}_p)$ is a subgroup and $p\geqq n$.

Let A be the quotient of $\tilde{G}$ by its unipotent radical W. Because $H^1(\mathbf{F}_p, W)$ vanishes we get the exact sequence

3.6(i) $$1\to W(\mathbf{F}_p)\to \tilde{G}(\mathbf{F}_p)\to A(\mathbf{F}_p)\to 1.$$

Denoting by $A(\mathbf{F}_p)^+$ the subgroup of $A(\mathbf{F}_p)$ generated by its elements of order p, and noting

(a) $W(\mathbf{F}_p)$ is a p-group, and

(b) $u\in GL_n(\mathbf{F}_p)$ and $u^{p^k}=1\Rightarrow u^p=1$,

we deduce the exactness of (ii) below.

3.6(ii) $$1\to W(\mathbf{F}_p)\to \tilde{G}(\mathbf{F}_p)^+\to A(\mathbf{F}_p)^+\to 1.$$

3.6(iii) In particular, $\tilde{G}(\mathbf{F}_p)/\tilde{G}(\mathbf{F}_p)^+\to A(\mathbf{F}_p)/A(\mathbf{F}_p)^+$ is an isomorphism.

Next note that A is reductive, and being generated by its unipotents, is in fact semi-simple. Let $A_1\to A$ be the simply connected covering group of A. Then $A_1(\mathbf{F}_p)$ is generated by its elements of order p and its image in $A(\mathbf{F}_p)$ is precisely $A(\mathbf{F}_p)^+$. There is a unique factoring $A_1\to A_2\to A$ of $A_1\to A$ so that $A_2\to A$ is étale and $A_1\to A_2$ is purely inseparable. Clearly $A_1(\mathbf{F}_p)\to A_2(\mathbf{F}_p)$ is an isomorphism and if Z denotes the kernel of $A_2\to A$, we have the exact sequence below:

$$A_2(\mathbf{F}_p)\to A(\mathbf{F}_p)\to H^1(\mathbf{F}_p, Z)\to H^1(\mathbf{F}_p, A_2)$$

where the group on the extreme right vanishes by a theorem of S. Lang (see [6]) from which we deduce

3.6(iv) $$A(\mathbf{F}_p)/A(\mathbf{F}_p)^+ \cong H^1(\mathbf{F}_p, Z).$$

Now $H^1(\mathbf{F}_p, Z)$ is canonically isomorphic to $H_0(\mathrm{Gal}(\bar{\mathbf{F}}_p/\mathbf{F}_p), Z(\bar{\mathbf{F}}_p))$ and so its order divides that of $Z(\bar{\mathbf{F}}_p)$. Because the split form of A has only finitely many possibilities (the rank of A being bounded above by $n-1$), there is an upper bound on the order of $Z(\bar{\mathbf{F}}_p)$ which is independent of p.

In fact, as J-P. Serre remarked, 2^k is an upper bound for the order of the centre of all semi-simple algebraic groups of rank $\leqq k$, the bound being attained by $(SL_2)^k$. Thus we have

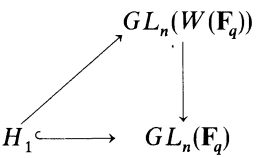
3.6(v) $\tilde{G}(\mathbf{F}_p)/\tilde{G}(\mathbf{F}_p)^+$ is a commutative group of order $\leqq 2^{n-1}$.

Theorem C. *If $G \subset GL_n(\mathbf{F}_p)$ and $\mathrm{char}\,\mathbf{F}_q = p \geqq n$, then G has a commutative subgroup H such that HG^+ is normal in G and $[G: HG^+] \leqq d(n)$, where $d(n)$ is a constant depending only on n (and not on q, p, G).*

The constant $d(n)$ is obtained from the following result of C. Jordan (see Theorem 8.29, p. 151, [9]).

Theorem (C. Jordan). *Every finite subgroup X of $GL_n(\mathbf{C})$ has a commutative normal subgroup Y such that $[X: Y] \leqq d(n)$, where $d(n)$ is a constant depending only on n.*

Proof of Theorem C. Let P be a p-Sylow subgroup of G and let $N_G(P)$ be its normaliser. Because $p \nmid [G:G^+]$, we see that $G = G^+ N_G(P)$, see e.g. p. 224, [28]. The orders of P and $N_G(P)/P$ being relatively prime, we get a subgroup H_1 of $N_G(P)$ such that $H_1 \to N_G(P)/P$ is an isomorphism. Thus $H_1 G^+ = G$ and $p \nmid o(H_1)$. Consequently, if $W(\mathbf{F}_q)$ denotes the Witt ring of $\mathbf{F}_q$, there is a lifting



and now embedding $W(\mathbf{F}_q)$ in $\mathbf{C}$, we may regard H_1 as a subgroup of $GL_n(\mathbf{C})$. Putting $X = H_1$ and $Y = H$ in Jordan's theorem completes the proof of Theorem C.

Remark 3.7. We would like to prove "$G^+ = \tilde{G}(\mathbf{F}_p)^+$" for subgroups $G \subset GL_n(\mathbf{F}_q)$ too where $q = p^k$ and p is a prime. Embedding $GL_n(\mathbf{F}_q)$ in $GL_{nk}(\mathbf{F}_p)$ in the usual way and applying Theorem B, we get this if $p > c_2(nk)$.

Combining the results of B. Weisfeiler [15] with the methods of § 9, 10, 11 of [8], it can be shown that $G^+ = \tilde{G}(\mathbf{F}_p)^+$ if p is greater than some constant that depends only on n and not on k. However [15] uses the classification of finite simple groups, and it is hoped that this form of Theorem B can be obtained by elementary methods similar to those given here.

§ 4. Semi-simplicity and a cohomology vanishing theorem

Theorem D. *With c_1 and c_2 as in Theorems A and B respectively, let p be a prime $>\max\{c_1(n+1), c_2(n+1)\}$ and let V be a n-dimensional vector-space over $\bar{\mathbf{F}}_p$. For every subgroup G of $GL(V)$ we have an isomorphism of abelian groups:*

$$H^1(G^+, V) \cong H^1(\langle \log G \rangle, V).$$

Theorem E. *There is a constant $c_3(n)$ so that if V is a vector space of dimension n over $\mathbf{F}_p$ and if $p > c_3(n)$, and if $G \subset GL(V)$ is a subgroup that acts semi-simply on V, then*

$$H^1(G, V) = 0.$$

Proof of Theorem D. Given $G \subset GL(V)$ and an element of $H^1(G^+, V)$, we get an exact sequence of $\mathbf{F}_p[G^+]$-modules:

$$0 \to V \to W \to \mathbf{F}_p \to 0.$$

Let $B = \{f \in \mathrm{End}\, W \mid f(V) \subset V\}$. Then $f \mapsto f|V$ gives a ring homomorphism $j: B \to \mathrm{End}\, V$. With $G \subset GL(V)$ as given, the action of G^+ on W gives a homomorphism $s: G^+ \to B^*$ such that $js(g) = g$ for all $g \in G^+$.

Applying Theorem B to $H = s(G^+) \subset GL(W)$ we see that $H = \tilde{H}(\mathbf{F}_p)^+$ and $\tilde{H} = \exp(\langle \log G \rangle)$ because $\dim W = n+1$, and $p > c_2(n+1)$. Clearly $\langle \log H \rangle \subset B$ and $j(\langle \log H \rangle) = \langle \log G \rangle$.

Next note that $j|\langle \log H \rangle: \langle \log H \rangle \to \langle \log G \rangle$ is an isomorphism. Because if $j(z) = 0$ and $z \in \langle \log H \rangle$, then $z^2 = 0$ and $\exp z \in H$, and furthermore $j(\exp z) = \exp j(z) = 1$, from which it follows that $\exp z = 1$, or equivalently we have $z = 0$.

Thus W may be regarded as a $\langle \log G \rangle$-module and the exact sequence $0 \to V \to W \to \mathbf{F}_p \to 0$ now gives an element of $H^1(\langle \log G \rangle, V)$. In a similar manner, an exact sequence $0 \to V \to W \to \mathbf{F}_p \to 0$ of modules over $\langle \log G \rangle$ gives rise to an exact sequence of G^+-modules, and this establishes a one-to-one correspondence between $H^1(G^+, V)$ and $H^1(\langle \log G \rangle, V)$. That this correspondence respects the group-structures on these cohomology groups is routine, and we omit its proof.

For the proof of Theorem E we need the two lemmas below.

Lemma 4.1. *With $Y_{n,k}$ and $L(z)$ as in 2.5, there are Zariski-open subsets W_1 and W_2 of $Y_{n,k}$ such* (a) *and* (b) *below are true for all fields F and for all $z \in Y_{n,k}(F)$:*

(a) $z \in W_1(F) \Leftrightarrow H^1(L(z), F^n) = 0$.

(b) $z \in W_2(F) \Leftrightarrow$ *for all fields $E \supset F$ and for all chains of $(L(z) \otimes_F E)$-invariant subspaces $0 = V_0 \subset V_1 \subset \ldots \subset V_m = E^n$, the action of $L(z) \otimes_F E$ on $\bigoplus_{i=1}^{m} V_i/V_{i-1}$ is faithful.*

Lemma 4.2. *With W_1 and W_2 as in* 4.1 *above, we have:*

$$(W_2)_{\mathbf{Q}} \subset (W_1)_{\mathbf{Q}}.$$

Proof of Lemma 4.1(a). Using the standard resolution for Lie algebra cohomology (see Theorem 7.1, Chapter XIII, [2]), we see that there is a complex $C^\cdot$ of finitely generated free $\mathcal{O}$-modules where $\mathcal{O}$ is the structure sheaf of $Y_{n,k}$ so that whenever F is a field and $z \in Y_{n,k}(F)$ we have:

$$H^i(C^\cdot \otimes_{\mathcal{O}} F) \cong H^i(L(z), F^n)$$

where $L(z) \subset M_n(F)$ acts on F^n in the standard manner. Because the dimension of $H^i(C^\cdot \otimes_{\mathcal{O}} k(y))$, where y is a closed point of $Y_{n,k}$ and $k(y)$ is the residue field, is an upper semi-continuous function of y, the lemma follows.

Proof of Lemma 4.1(b), Fix natural numbers n, k, f such that $1 \leqq f \leqq k \leqq n^2 - 1$ and define, for a commutative ring R, the set $Z_{n,k,f}(R)$ to be the collection of $(z, I, D_1, D_2, \ldots, D_{n-1})$ such that:

(i) $z \in X_{n,k}(R)$,

(ii) $I \subset L(z)$ is a R-submodule so that $L(z)/I$ is locally free of rank $(k-f)$, and $[L(z), I] \subset I$,

(iii) $0 = D_0 \subset D_1 \subset \ldots \subset D_{n-1} \subset D_n = R^n$ is a chain of R-submodules so that $ID_i \subset D_{i-1}$ and D_i/D_{i-1} is locally free of rank one, for $1 \leqq i \leqq n$.

Clearly the functor $R \mapsto Z_{n,k,f}(R)$ is representable by a scheme $Z_{n,k,f}$ which is proper over $Y_{n,k}$ and therefore the image X_f of $Z_{n,k,f} \to Y_{n,k}$ is a closed subset. Denote by W_2 the complement in $Y_{n,k}$ of the union of all the X_f, $1 \leqq f \leqq k$.

Let $z \in Y_{n,k}(F)$ and assume that $z \notin W_2(F)$. Then there is a field $E \supset F$ and a non-zero ideal $I \subset L(z) \otimes_F E$ and E-linear subspaces $0 = D_0 \subset \ldots \subset D_n = E^n$ such that $ID_i \subset D_{i-1}$ whenever $1 \leqq i \leqq n$. We define $W_0 = 0$ and $W_{j+1}/W_j = H^0(I, E^n/W_j)$. The subspaces W_j are invariant under $L(z) \otimes_F E$ because I is an ideal. Clearly $W_j \supset D_j$ for all j, and therefore $W_n = D_n$. The action of $L(z) \otimes_F E$ on $\bigoplus_j W_j/W_{j-1}$ is not faithful because the action of I here is trivial.

Conversely, if $0 = V_0 \subset \ldots \subset V_m = E^n$ is a chain of $(L(z) \otimes_F E)$-invariant subspaces such that the action of $L(z) \otimes_F E$ on $\bigoplus_j V_j/V_{j-1}$ is not faithful, let I be the kernel of this action. Put $f = \dim I$. Extend the chain $0 = V_0 \subset \ldots \subset V_m = E^n$ arbitrarily to a maximal chain of E-linear subspaces $0 = D_0 \subset \ldots \subset D_n = E^n$. Then $(z, I, D_1, \ldots, D_{n-1}) \in Z_{n,k,f}(E)$ and its image in $Y_{n,k}(E)$ is precisely z, and therefore $z \notin W_2(F)$. This proves the lemma.

Proof of Lemma 4.2. Let $z \in Y_{n,k}(F)$ and let $0 \subset V_0 \subset \ldots \subset V_m = F^n$ be a *maximal* chain of $L(z)$-invariant subspaces. If $z \in W_2(F)$, then the semi-simple action of $L(z)$ on $\bigoplus_{i=1}^{m} V_i/V_{i-1}$ is also faithful. Therefore $\exp L(z)$ is a semi-simple algebraic group. If char. F is zero, then $L(z) = \mathrm{Lie}(\exp L(z))$ is a semi-simple Lie algebra. Because all the finite-dimensional representation of $L(z)$ are semi-simple (see for example page LA 6.4, [12]), it follows that $H^1(L(z), F^n) = 0$. Thus $z \in W_1(F)$ and the lemma is proved.

Proof of Theorem E. Let $G \subset GL_n(\mathbf{F}_p)$ be a subgroup whose action on $\mathbf{F}_p^n$ is semi-simple. Because G^+ is a normal subgroup, its action on $\mathbf{F}_p^n$ is also semi-

simple. Also p does not divide $[G:G^+]$ if $p \geqq n$, from which we have

$$H^1(G, \mathbf{F}_p^n) = H^0(G/G^+, H^1(G^+, \mathbf{F}_p^n)).$$

Assuming that $p > \max\{c_1(n+1), c_2(n+1)\}$, from Theorem D, it suffices to prove that $H^1(\langle \log G \rangle, \mathbf{F}_p^n) = 0$. Now let k be the dimension of $\langle \log G \rangle$ and let $z = (z_1, \ldots, z_k) \in Y_{n,k}(\mathbf{F}_p)$ so that $L(z) = \langle \log G \rangle$. Because the action of $\langle \log G \rangle$ on $\mathbf{F}_p^n$ is also semi-simple from Lemma 1.4 and because $\mathbf{F}_p$ is perfect, the action of $\langle \log G \rangle \otimes E$ on E^n is semi-simple for every field E of characteristic p, and from this we see that $z \in W_2(\mathbf{F}_p)$.

From Lemma 4.2, however, there exists a constant $c_3(n,k)$ such that

$$w \in W_2(F) \quad \text{and} \quad \operatorname{char} F > c_3(n,k) \Rightarrow w \in W_1(F).$$

Therefore, assuming that $p > c_3(n,k)$ gives $z \in W_1(\mathbf{F}_p)$ and therefore $H^1(\langle \log G \rangle, \mathbf{F}_p^n) = 0$. Taking $c_3(n)$ to the maximum of $c_1(n+1)$, $c_2(n+1)$, $c_3(n,k)$, where $1 \leqq k \leqq n^2 - 1$, completes the proof of Theorem E.

§ 5. Applications to subgroups of $GL_n(\mathbf{Q})$

Let π be a finitely generated subgroup of $GL_n(\mathbf{Q})$. There is a finite set S of primes so that if $R = S^{-1}\mathbf{Z}$, then $\pi \subset GL_n(R)$. Two measures of "how large π is" are obtained by determining:

(A) the Zariski-closure A of π in $(GL_n)_R$,

and

(B) the closure $\bar{\pi}$ of π in the profinite group $GL_n(\hat{R})$ where

$$\hat{R} = \prod_{p \notin S} \mathbf{Z}_p.$$

The theorems below indicate that both these measures of the size of π are roughly equivalent. Assume for the moment that $A_{\mathbf{Q}}$ is connected and let Γ be a subgroup of finite index in π. Then the Zariski-closure of Γ is also A, but $\bar{\Gamma}$ is an open subgroup of $\bar{\pi}$. Thus we want to determine $\bar{\pi}$ only up to commensurability (two closed subgroups P and Q of $GL_n(\hat{R})$ are said to be commensurable with each other if $P \cap Q$ is an open subgroup of both P and Q).

The A above is the closed subscheme defined by the ideal I in $R[x_{ij}; \det(x_{ij})^{-1}]$ consisting of those f that vanish identically on π. Because $\pi \subset A(R)$, it follows that $\bar{\pi} \subset A(\hat{R})$. Also, if $\pi(p)$ denotes the image of $\pi \to GL_n(\mathbf{F}_p)$ for a prime $p \notin S$, then $\pi(p) \subset A(p)$.

Theorem 5.1. *For all but finitely many primes, we have*

$$A(\mathbf{F}_p)^+ \subset \pi(p) \subset A(\mathbf{F}_p).$$

Theorem 5.2. *If $A_{\mathbf{Q}}$ is connected, the closed subgroups $[\bar{\pi}, \bar{\pi}]$, $[A(\hat{R}), A(\hat{R})]$ and $[A, A](\hat{R})$ are all commensurable with each other.*

Theorem 5.3. *If $A(\mathbf{C})$ has finite fundamental group, then $\bar{\pi}$ is open in the closure of $A(R)$ in $GL_n(\hat{R})$.*

Theorem 5.4. *If $A(\mathbf{C})$ is connected and simply connected, then $\bar{\pi}$ is open in $A(\hat{R})$.*

Theorem 5.1 is deducible from Theorem B and C (see § 3 of this paper) and 5.2, 5.3, 5.4 are easily deducible from Theorem 5.1 (for detailed proofs of all these results, see [8]). The special case of Theorem 5.4, when A is also assumed to be semi-simple, was proved in [7]. While the semi-simple case is perhaps the most interesting, the general case in Theorem 5.4 does not follow from it.

The above results fail to be complete only when the group of characters, $\mathrm{Hom}(A_{\mathbf{Q}}, \mathbf{G}_m)$, is non-trivial (assuming that $A_{\mathbf{Q}}$ is connected). But in this case, any "open-ness" statement is hopelessly false. For example, if A is a torus, then $A(\hat{R})/A(\hat{R})^m$ is infinite for every $m \geqq 2$ while π/π^m is a finite group. And this shows that $\bar{\pi}$ is not open in $A(\hat{R})$.

A more meaningful question is the following:

Question 5.5. *Is the dimension $d(p)$ of the p-adic Lie group obtained by taking the closure of π in $GL_n(\mathbf{Z}_p)$ a constant function of p, where p is a prime not belonging to S?*

An affirmative answer to 5.5 in the case when A is a torus implies an affirmative answer to 5.5 in general. For this question, we may assume that $A_{\mathbf{Q}}$ is connected. Let B be the quotient of $A_{\mathbf{Q}}$ by $[A_{\mathbf{Q}}, A_{\mathbf{Q}}]$ and let $B = V \times T$ where V is a vector-group (isomorphic to $\mathbf{G}_a \oplus \mathbf{G}_a \oplus \ldots \oplus \mathbf{G}_a$) of dimension f and T is a torus. Let π_1 and π_2 be the image and kernel of the homomorphism

$$\pi' \to V(\mathbf{Q})$$

where π' is itself the image of $\pi \to B(\mathbf{Q})$. Because π_1 is finitely generated and Zariski-dense in V, it follows that π_1 is a free Abelian group of rank f.

In view of this and Theorem 5.2, it can be seen that

$$d(p) = e(p) + d + f$$

where $d = \dim [A_{\mathbf{Q}}, A_{\mathbf{Q}}]$ and $e(p)$ is the dimension of the closure of π_2 in $T(\mathbf{Q}_p)$. Thus, for 5.5, we may replace $\pi \subset A(\mathbf{Q})$ by $\pi_2 \subset T(\mathbf{Q})$.

If K is a number field with r_1 real and r_2 complex embeddings, if $G = R_{K/\mathbf{Q}}\mathbf{G}_m$ and if π is the group of units in the ring of integers, it is conjectured that $d(p) = r_1 + r_2 - 1$ for all primes p (see p. 41, [5]).

References

1. Borel, A.: Linear algebraic groups. New York: Benjamin 1969
2. Cartan, H., Eilenberg, S.: Homological algebra. Princeton Math. Ser. No. 19, 1956
3. Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern. Invent. Math. **73**, 349-366 (1983)
4. Grothendieck, A.: EGA IV troisième partie. Publ. Math., Inst. Hautes Etud. Sci. **28** (1966)

5. Iwasawa, K.: Lectures on p-adic L-functions. Ann. Math. Stud. vol. 74. Princeton Univ. Press 1972
6. Lang, S.: Algebraic groups over finite fields. Am. J. Math. **78**, 555–563 (1956)
7. Matthews, C.R., Vaserstein, L.N., Weisfeiler, B.: Congruence properties for Zariski-dense subgroups. Proc. London Math. Soc. **48**, 514–532 (1984)
8. Nori, M.V.: On subgroups of $SL_n(\mathbb{Z})$ and $SL_n(\mathbb{F}_p)$. Preprint 1983
9. Raghunathan, M.S.: Discrete subgroups of Lie groups. Ergebn. Math., vol. 68. Berlin-Heidelberg-New York: Springer 1972
10. Ribet, K.: Kummer theory on extensions of abelian varieties by tori. Duke Math. J. **46**, 743–761 (1979)
11. Serre, J-P.: Représentations l-adiques. Kyoto Internat. Symp. on Algebraic Number Theory, 1977, pp. 177–193
12. Serre, J-P.: Lie algebras and Lie groups. New York: Benjamin 1965
13. Serre, J-P.: Propriétés galoisiennes des points d'ordre fini des courbes elliptiques. Invent. Math. **15**, 259–331 (1972)
14. Steinberg, R.: Endomorphisms of linear algebraic groups. Mem. Am. Math. Soc. **80** (1968)
15. Weisfeiler, B.: Strong approximation for Zariski-dense subgroups of semi-simple algebraic groups. Ann. Math. **120**, 271–315 (1984)
16. Schmidt, W.M.: Equations over finite fields. An elementary approach. (Lecture Notes in Math., vol. 536). Berlin-Heidelberg-New York: Springer 1976
17. SGA 1: Lect. Notes in Math., vol. 224. Berlin-Heidelberg-New York: Springer 1971
18. Hall, M., Jr.: The theory of groups. New York: Macmillan 1959

Oblatum 4-IX-1985

Invent. math. 88, 277-318 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

The Conway-Norton algebras for $\Omega^-(6,3)$, $\Omega(7,3)$, F'_{24}, and their full automorphism groups

Dedicated to Professor Nagayoshi Iwahori on the occasion of his 60th birthday

Masaaki Kitazume

Department of Mathematics, Faculty of Science, University of Tokyo, Hongo, Tokyo, 113 Japan

Contents

0. Introduction 277
1. The Conway-Norten algebras 279
2. Tits' theorem 280
3. The algebra $(B(6), \tau(6))$ for $\Omega^-(6,3)$ 281
4. The algebra $(B(7), \tau(7))$ for $\Omega(7,3)$ 287
5. The algebra $(B(24), \tau(24))$ for F'_{24} 291
6. The automorphism group $\mathrm{Aut}(B(24), \tau(24))$ 302
7. The automorphism groups $\mathrm{Aut}(B(6), \tau(6))$ and $\mathrm{Aut}(B(7), \tau(7))$ 306
8. Appendix: $\tau(24)(B_1, B_1)$ and the group $C_{N(24)}(s)$ 311

0. Introduction

In [10], R. Griess has constructed the simple group F_1, the Friendly Giant or the Monster, as an automorphism group of a 196883-dimensional commutative (non-associative) algebra. In [17], J. Tits has proved the finiteness of the full automorphism group of this algebra (and consequently it is a group of Monster type). Such algebras related to some finite simple groups are studied by many finite group theorists.

In this paper, we are concerned with the Conway-Norton algebras described in S.D. Smith [15]. There exist five algebras which are called the Conway-Norton algebras. Each algebra is defined so that its automorphism group contains a certain group N related to some 3-transpositions group, whose simple component is A_6, A_7, $\Omega^-(6,3)$, $\Omega(7,3)$ or F'_{24}. The first two cases are handled in [15]. In this paper, we consider the latter three cases, and write $N=N(6)$, $N(7)$ or $N(24)$. Here the group $N=N(n)$ satisfies:

$$|N:N'|=2, \quad |Z(N')|=3, \quad N''=N'=C_N(Z(N')),$$

and

$$N'/Z(N') \simeq \begin{cases} \Omega^-(6, 3) & (n=6), \\ \Omega(7, 3) & (n=7), \\ F'_{24} & (n=24). \end{cases}$$

We denote by $(B, \tau)=(B(n), \tau(n))$ $(n=6, 7, 24)$ the corresponding Conway-Norton algebra, where $B=B(n)$ is a vector space over $\mathbb{C}$, and $\tau=\tau(24)$: $B(n) \times B(n) \to B(n)$ is a bilinear symmetric map, which is called the algebra product.

The main purpose of this paper is to determine the full automorphism group:

$$\mathrm{Aut}(B, \tau)=\{\alpha \in GL(B) \mid \alpha(\tau(x, y))=\tau(\alpha(x), \alpha(y)) \text{ for any } x, y \text{ of } B\}.$$

The main theorems in this paper are as follows:

Theorem 2. $\mathrm{Aut}(B(6), \tau(6))=C_*$,
Theorem 4. $\mathrm{Aut}(B(7), \tau(7))=N(7)$,
Theorem 5. $\mathrm{Aut}(B(24), \tau(24))=N(24)$.

Here the group C_*, which contains $N(6)$ with index 2, is defined in Sect. 3.

The proofs of these theorems will follow to a great extent the pattern of the work of Tits [17]. In this way, we obtain the structure of the centralizer of an involution, and then the finiteness of $\mathrm{Aut}(B, \tau)$ by Theorem B (in Sect. 2) due to Tits. Then we are able to use characterization theorems about finite simple groups, by which $N_{\mathrm{Aut}(B, \tau)}(Z(N'))$ is determined in each case. In order to prove that it coincides with $\mathrm{Aut}(B, \tau)$, we need the famous theorem (Theorem C in Sect. 7) for strongly embedded subgroups due to H. Bender [4].

In order to apply Tits' arguments, we first determine the structure constants of each algebra, and describe the action of the centralizer of an involution in N. In the case of $N(6)$, the representation space $B(6)$ is constructed by using the Coxeter-Todd lattice ([8]). Thanks to this complex lattice, the algebra $(B(6), \tau(6))$, and the actions of its automorphisms are presented quite explicitly (see Sect. 3). Using them, we are able to describe the algebras $(B(7), \tau(7))$, $(B(24), \tau(24))$ in Sects. 4, 5. The algebra $(B(6), \tau(6))$ appears as subalgebras of $B(7)$ and $B(24)$ respectively, and centralizers of some involutions in $N(7)$, $N(24)$ are groups related to $N(6)$. In addition, in the case of $B(24)$, the Mathieu group M_{24} and the Steiner system $S(24, 8, 5)$ play important roles (see Sect. 5).

Now we remark the similarity between $(B(24), \tau(24))$ and the algebra for F_1. In each case, the centralizer of an involution possesses just three irreducible components in the algebra. We denote them by B_*, B_2, B_1, the same symbols in both cases.

The space B_* is constructed from some lattice, that is, the Coxeter-Todd lattice in the $N(24)$-case (we may regard $B_*=B(6)$: see Sects. 3, 5), and the Leech lattice in the F_1-case ([10], [17]) by using the symmetric tensor.

The space B_1 is constructed as the tensor product of the representation space of such a lattice and the unique non-linear irreducible representation space of some extraspecial 2-group, which is the maximal normal 2-subgroup of the centralizer of an involution in the corresponding group.

The only exception is the space B_2. In the F_1-case, the basis of B_2 is labeled by the vectors (modulo ± 1 times) of minimal length in the Leech lattice. In the $N(24)$-case, such vectors in the Coxeter-Todd lattice do not appear in the description of B_2. (We use some vectors, $D(6)$-vectors, in $B_* = B(6)$: see Sect. 5.) But on another viewpoint, in each case, the space B_2 is labeled by the elements of an orbit in the extraspecial 2-group, mentioned above, by the action of some group of its automorphisms.

Is there any mathematical background of such similarity? The author cannot answer this question yet.

Finally we mention a remark, that is, Theorem 5 does not mean that $N(24)$ (nor F_{24}') has been constructed nor defined as the group $\mathrm{Aut}(B(24), \tau(24))$. We prove Theorem 5 by using the fact that $N(24)$ does act on $(B(24), \tau(24))$. This fact is guaranteed by the construction of the algebra in [15], and there we need the original construction of the group F_{24} by B. Fischer [9]. We can also describe the algebra without using such original ones (see Sect. 8). But then, in order to prove Theorem 5, we must construct some automorphism outside the centralizer of an involution. The author does not do it yet.

For groups, we use the following notation. The symbol $\Omega^-(6, 3)$ (resp. $\Omega(7, 3)$) denotes the simple component of the 6- (resp. 7-) dimensional orthogonal groups over the three elements field $\mathbb{F}_3$, and F_{24} denotes the largest 3-transpositions group of Fischer ($M(24)$ in [9]), and its commutator subgroup F_{24}' is the Fischer simple group. The symbol $U(6, 2)$ (resp. $SU(6, 2)$) denotes the general (resp. special) 6-dimensional unitary group over $\mathbb{F}_4$, whose simple component is denoted by $PSU(6, 2)$. Moreover we denote by $\Gamma U(6, 2)$ the group generated by $U(6, 2)$ and the field automorphism on $\mathbb{F}_4$.

For a group X, we denote by $\mathrm{Aut}(X)$ the automorphism group of X, and denote by $\mathrm{Out}(X)$ the outer-automorphism group of X.

For an integer m, and a prime integer p, we denote by p^m the elementary abelian group of order p^m, and denote by Z_m the cyclic group of order m. Moreover we denote by 2_+^{1+2m} the extraspecial 2-group of order 2^{1+2m} of plus-type, that is, the central product of m dihedral group D_8 of order 8 with amalgamated centers. We write $X \simeq A.B$ to mean the group X is an extension of the group A by the group B.

For subsets X, Y and elements a, b of a group, we set $a^b = bab^{-1}$, and $X^Y = \{x^y \mid x \in X, y \in Y\}$. We write a^Y instead of $\{a\}^Y$. We denote by $\langle \cdots \rangle$ a group generated by $\cdots$.

For vector spaces, we also denote by $\langle \cdots \rangle$ a space spanned by $\cdots$. Let (B, τ) be as above, and A be a subspace of B, and g be an automorphism of (B, τ). If A satisfies $\langle \tau(a, a') \mid a, a' \in A \rangle \subseteqq A$, then we denote by $\tau|_A$ the restriction of τ on $A \times A$. If A is stable by g, then we denote by $g|_A$ the restriction of g on A.

1. The Conway-Norton algebras

In this section, we give a brief description of the Conway-Norton algebras which appeared in [15].

(1.1) First, we consider a finite group N satisfying the following conditions:

(i) $|N: N'|=2$, $|Z(N')|=3$, and $C_N(Z(N'))=N'=N''$,

(ii) $N/Z(N')$ is generated by a conjugacy class $\bar{D}$ of 3-transpositions (i.e. if $\bar{x}, \bar{y}\in\bar{D}$, then $\bar{x}$ is an involution, and the order of $\bar{x}\bar{y}$ is 1, 2 or 3).

Then $N'/Z(N')$ is isomorphic to A_6, A_7, $\Omega^-(6, 3)$, $\Omega(7, 3)$ or F'_{24}. In this paper, we consider the latter three cases, and write $N=N(6)$, $N(7)$ and $N(24)$ when $N'/Z(N')\simeq\Omega^-(6, 3)$, $\Omega(7, 3)$ and F'_{24} respectively. We set $Z(N')=\langle z\rangle$.

Let $D=D(n)$ be the inverse image of $\bar{D}$ in $N=N(n)$, $n=6, 7, 24$. Then D is a conjugacy class of N, and N acts on D with rank 5. In fact, if we set $D_d=\{x\in D\,|\,|xd|=2\}$ and $A_d=\{x\in D\,|\,|xd|=3,\ x\neq zd,\ z^2d\}$ for each $d\in D$, then the $C_N(d)$-orbits are $\{d\}$, $\{zd, z^2d\}$, D_d, $z(D_d)\cup z^2(D_d)$ and A_d.

To define the Conway-Norton algebras, we consider the permutation module V with basis $\{v(d)\,|\,d\in D\}$ and the action $g(v(d))=v(d^g)$ for $g\in N$. Then V affords two N-irreducible and faithful representations. Let $B=B(n)$ be the corresponding submodule with the smaller dimension. Then we may write $B=U\oplus\bar{U}$, a decomposition as N'-module into the ω, ω^2-eigenspaces of z, where $\omega=\frac{1}{2}(-1+\sqrt{-3})$. We denote by v_d, $\bar{v}_d$, the projective images of $v(d)$ in U, $\bar{U}$ respectively. Note that $v_{zd}=\omega v_d$, and $v_{z^2d}=\omega^2 v_d$. Now the following holds:

Theorem A. (Conway-Norton). *The group $N(n)$ preserves a symmetric bilinear map $\tau=\tau(n)$: $B\times B\to B$ such that*

$$\tau(U, \bar{U})=0,$$

$$\tau(v_d, v_d)=m\bar{v}_d, \qquad \tau(\bar{v}_d, \bar{v}_d)=mv_d,$$

$$\tau(v_d, v_e)=\bar{v}_d+\bar{v}_e, \qquad \tau(\bar{v}_d, \bar{v}_e)=v_d+v_e,$$

$$\tau(v_d, v_x)=\bar{v}_{d^x}, \qquad \tau(\bar{v}_d, \bar{v}_x)=v_{d^x},$$

for $d\in D$, $e\in D_d$, $x\in A_d$, and the integer $m=4$ if $n=6$, 7, $m=10$ if $n=24$. The dimension of U is 21 if $n=6$, 27 if $n=7$, and 783 if $n=24$.

The commutative (non-associative) algebra $(B(n), \tau(n))$ is called *the Conway-Norton algebra* for $N(n)$, or $N'/Z(N')$. The vectors $v_d, \bar{v}_d$ are called the *D-vectors.*

(1.2) Here we mention the existence and the uniqueness of the group N, for the existence of the Conway-Norton algebras follow them.

The existence and the uniqueness of $N(6)$ are proved in (3.6). As for $N(7)$ (resp. $N(24)$), the orders of the Schur multiplier and the outer automorphism group of $N(7)'/Z(N(7)')\simeq\Omega(7, 3)$ (resp. $N(24)'/Z(N(24)')\simeq F'_{24}$) are 6 and 2 (resp. 3 and 2) respectively. The uniqueness of $N(n)$, $n=7$ or 24, is an easy consequence of these facts. The existence of $N(7)$ is proved in (7.3).

The existence of $N(24)$ is not proved in this paper. The author refers the reader to Griess' paper [10: §13], in which we may find the group $N(24)$ as $N_{F_1}(\langle a_{3D}\rangle)$ where the symbol a_{3D} means some element of order 3 in F_1 as in [10].

2. Tits' theorem

(2.1) It seems that the following is an important key in Tits [17].

Theorem B. (Tits). *Let B be a finite dimensional vector space over $\mathbb{C}$, and $\tau: B \times B \to B$ be a bilinear symmetric map. Suppose that:*

(1) *for some involution s of $\mathrm{Aut}(B, \tau)$, the centralizer $C_{\mathrm{Aut}(B,\tau)}(s)$ is finite,*

(2) *any commutative normal subgroup of $\mathrm{Aut}(B, \tau)$ is finite.*

Then $\mathrm{Aut}(B, \tau)$ is finite.

In [17], the finiteness is proved under slightly different assumptions. But the theorem above is obtained by the same arguments as in [17: (iv) in p. 496].

(2.2) The condition (2) is easily verified in our cases.

Let $(B, \tau)=(B(n), \tau(n))$ be as in Sect. 1, and set $d=\dim(B)$. Let A be a commutative normal subgroup of $\mathrm{Aut}(B, \tau)$. Then the sum of all simple sub-A-modules of B is stable by $\mathrm{Aut}(B, \tau)$. Since N acts on B irreducibly, B is a semisimple A-module. Hence we may write $B=B_1 \oplus \ldots \oplus B_r$, a sum of isotypic components of A, and N permutes transitively the spaces B_i, $1 \leqq i \leqq r$.

Now one can see that N has no subgroups of index $\leqq d$ except N'. In fact, the tables of maximal subgroups of $\Omega^-(6, 3)$, $\Omega(7, 3)$ are found in [7]. The group $N(24)$ has a maximal subgroup $H \simeq 6.F_{22}.2$, since F_{24} has a maximal subgroup $\simeq 2.F_{22}.2$. The table of maximal subgroups of index $\leqq 12096$ of F_{22} is in [19], and by this table, F_{22} has no subgroups of index $\leqq 783 \times 2 = d$. Now the above assertion is an easy consequence of these facts.

Hence we have $r \leqq 2$, and $\{B_1, B_2\}=\{U, \bar{U}\}$ in the case $r=2$. Let $a \in A$. Then there exist some constants c_1, c_2 such that $a|_U = c_1 I_U$ and $a|_{\bar{U}} = c_2 I_{\bar{U}}$ (if $r=1$, then $c_1=c_2$). Let $b \in U$. Since $\tau(b, b) \in \bar{U}$, we have $c_2 \tau(b, b)=a(\tau(b, b))=\tau(a(b), a(b))=c_1^2 \tau(b, b)$, and thus $c_2=c_1^2$. Similarly we have $c_1=c_2^2$. Hence $c_1=c_1^4$, and so c_1 is some power of ω. This implies $A \leqq Z(N')$, and in particular, A is finite, as required.

Hence, by Theorem B, the finiteness of $\mathrm{Aut}(B, \tau)$ follows from the finiteness of the centralizer of some involution.

3. The algebra $(B(6), \tau(6))$ for $\Omega^-(6, 3)$

In this section, we construct the algebra $(B(6), \tau(6))$.

(3.1) First we introduce the Coxeter-Todd lattice in two ways.

Set $\theta=\sqrt{-3}$, and $\omega=\frac{1}{2}(-1+\theta)$. Let $\mathbb{C}^6$ be spanned by the fixed orthonormal basis $r_1={}^t(1, 0, 0, 0, 0, 0), \ldots$, and $r_6={}^t(0, 0, 0, 0, 0, 1)$. We define the Coxeter-Todd lattice $\Lambda=\Lambda^{(3)}$, or $\Lambda^{(4)}$ ([8]):

$$\Lambda^{(3)}=\left\{\sum_{i=1}^{6} a_i r_i \mid a_i \in \mathbb{Z}[\omega],\ a_1 \equiv \ldots \equiv a_6 \pmod{\theta},\ \text{and } \sum_{i=1}^{6} a_i \equiv 0 \pmod 3\right\},$$

$$\Lambda^{(4)}=\left\{\sum_{i=1}^{6} a_i r_i \mid a_i \in \mathbb{Z}[\omega],\ a_1 \equiv \ldots \equiv a_6 \equiv m(\text{say}) \pmod 2,\ \text{and } \sum_{i=1}^{6} a_i \equiv 2m\omega^2 \pmod 4\right\}.$$

Then Λ is a unimodular $\mathbb{Z}[\omega]$-lattice. Using these definitions, we will construct the algebras $(B^{(3)}, \tau^{(3)})$, $(B^{(4)}, \tau^{(4)})$ $(\simeq(B(6), \tau(6)))$. If we need not distinguish between $\Lambda^{(3)}$ and $\Lambda^{(4)}$, we write simply Λ, B, τ, and so on, dropping the superscript (3) and (4).

The automorphism group $\mathrm{Aut}(\Lambda)$, the group of the 6×6-unitary matrices preserving Λ, is generated by the set E of reflections with respect to the vectors of minimal length in Λ, which are the vectors:

$$\left.\begin{array}{ll} \theta\omega^{l_1}r_i+\theta\omega^{l_j}r_j & (1\leqq i\neq j\leqq 6) \\ \sum_{i=1}^{6}\omega^{l_i}r_i & \left(\sum_{i=1}^{6} l_i\equiv 0(\mathrm{mod}\ 3)\right) \end{array}\right\} \quad \text{in } \Lambda^{(3)},$$

$$\left.\begin{array}{ll} 2\omega^{l}(\varepsilon_i r_i+\varepsilon_j r_j) & (1\leqq i\neq j\leqq 6) \\ \omega^{l}((\prod_{j\neq i}\varepsilon_j)\theta r_i+\sum_{j\neq i}\varepsilon_j r_j) & (1\leqq i\leqq 6) \end{array}\right\} \quad \text{in } \Lambda^{(4)},$$

where l_i, $l\in\{0, 1, 2\}$, and $\varepsilon_i\in\{1, -1\}$. The set E is a conjugacy class of 3-transpositions. The group $\mathrm{Aut}(\Lambda)$ is a perfect extension of Z_6 by $\Omega^{-}(6, 3).2$.

(3.2) Now we present some elements of $\mathrm{Aut}(\Lambda)$ explicitly. They generate the group $\mathrm{Aut}(\Lambda)$.

$\mathrm{Aut}(\Lambda^{(3)})$;

T_{ij}=(the permutation matrix corresponding to the transposition (ij))
$(1\leqq i\neq j\leqq 6)$,

$D(\omega^{l_1}, \ldots, \omega^{l_6})$=(the diagonal matrix whose $i\times i$-entry equals ω^{l_i})
$\left(l_i\in\mathbb{Z}, \sum_{i=1}^{6} l_i\equiv 0(\mathrm{mod}\ 3)\right)$,

$\Pi=I-\frac{1}{3}J$, where I is the unit matrix and J is the matrix whose all entries equal 1,

$\mathrm{Aut}(\Lambda^{(4)})$;

T_{ij} $(1\leqq i\neq j\leqq 6)$,

$D(\varepsilon_1, \ldots, \varepsilon_6)$ $\left(\varepsilon_i=\pm 1, \sum_{i=1}^{6}\varepsilon_i\equiv 0(\mathrm{mod}\ 2)\right)$,

$$\Pi'=\frac{1}{2}\left(\begin{array}{cc|cccc} -1 & -\theta & & & & \\ \theta & -1 & & 0 & & \\ \hline & & 1 & -1 & -1 & -1 \\ & 0 & -1 & 1 & -1 & -1 \\ & & -1 & -1 & 1 & -1 \\ & & -1 & -1 & -1 & 1 \end{array}\right).$$

We set $S=-I$, $Z=\omega^2 I$. Then they generate the center of $\mathrm{Aut}(\Lambda)$. Moreover we set:

$$W^{(3)}=\langle T_{ij}, D(\omega^{l_1}, \ldots, \omega^{l_6}), S \mid 1\leqq i\neq j\leqq 6, l_i\in\mathbb{Z}\rangle,$$
$$W^{(4)}=\langle T_{ij}, D(\varepsilon_1, \ldots, \varepsilon_6), Z \mid 1\leqq i\neq j\leqq 6, \varepsilon_i=\pm 1\rangle.$$

Then $W^{(3)} \simeq Z_2 \times (3^5.\Sigma_6)$, $W^{(4)} \simeq Z_3 \times (2^5.\Sigma_6)$, which are the subgroups of the monomial matrices of $\mathrm{Aut}(\Lambda^{(3)})$, $\mathrm{Aut}(\Lambda^{(4)})$.

(3.3) We denote by ξ the complex conjugation on $\mathbb{C}^6$, that is, $\xi({}^t(a_1, \ldots, a_6)) = {}^t(\bar{a}_1, \ldots, \bar{a}_6)$ where $\bar{a}_i$ is the complex conjugate of a_i. Then ξ preserves the set $\Lambda^{(3)}$, preserving the length of vectors. Hence ξ acts on $\mathrm{Aut}(\Lambda^{(3)})$. We set $C_0^{(3)} = \mathrm{Aut}(\Lambda)\langle \xi^{(3)} \rangle$, and $R^{(3)} = (\mathbb{C} \otimes \Lambda^{(3)})^{C_0^{(3)}}$, the induced $C_0^{(3)}$-module. Moreover we set $\bar{r}_i = \xi(r_i)$ for $i = 1, \ldots, 6$ and set $R_1^{(3)} = \langle r_i \mid 1 \leqq i \leqq 6 \rangle$, $R_2^{(3)} = \langle \bar{r}_i \mid 1 \leqq i \leqq 6 \rangle$.

In the case of $\Lambda^{(4)}$, we define ξ' by $\xi'({}^t(a_1, \ldots, a_6)) = {}^t(-\bar{a}_1, \bar{a}_2, \ldots, \bar{a}_6)$. Then we can define $C_0^{(4)} = \mathrm{Aut}(\Lambda^{(4)})\langle \xi' \rangle$, and $R^{(4)} = (\mathbb{C} \otimes \Lambda^{(4)})^{C_0^{(4)}}$. This time we set $\bar{r}_1 = -\xi'(r_1)$, and $\bar{r}_i = \xi'(r_i)$ for $i \neq 1$, and set $R_1^{(4)} = \langle r_i \mid 1 \leqq i \leqq 6 \rangle$, $R_2^{(4)} = \langle \bar{r}_i \mid 1 \leqq i \leqq 6 \rangle$. Then for any element X of $\mathrm{Aut}(\Lambda^{(4)})$, the action of X on $R_2^{(4)}$ is given by $\bar{X}$, the complex conjugate of X, as well as the case of $\mathrm{Aut}(\Lambda^{(3)})$.

(3.4) We put $r_1 = {}^t(1, 0, \ldots, 0), \ldots, r_6 = {}^t(0, 0, 0, 0, 0, 1, 0, \ldots, 0)$, $\bar{r}_1 = {}^t(0, \ldots, 0, 1, 0, 0, 0, 0, 0), \ldots, \bar{r}_6 = {}^t(0, \ldots, 0, 1) \in \mathbb{C}^{12}$. Then C_0 can be embedded into $GL(12, \mathbb{C})$. We denote by A_0 the image of $\mathrm{Aut}(\Lambda)$ in $GL(12, \mathbb{C})$, in order to distinguish their dimensions as matrix groups, while we denote by C_0, the same letter, the image of C_0. For an element X of $\mathrm{Aut}(\Lambda)$, we denote by x, the corresponding small letter, the corresponding element $\begin{pmatrix} X & O \\ O & \bar{X} \end{pmatrix}$ of A_0. For example,

$$t_{ij} = \begin{pmatrix} T_{ij} & O \\ O & T_{ij} \end{pmatrix}, \quad z = \begin{pmatrix} \omega^2 I & O \\ O & \omega I \end{pmatrix}, \quad s = \begin{pmatrix} -I & O \\ O & -I \end{pmatrix},$$

and

$$d(\omega^{l_1}, \ldots, \omega^{l_6}) = \begin{pmatrix} D(\omega^{l_1}, \ldots, \omega^{l_6}) & O \\ O & D(\omega^{2l_1}, \ldots, \omega^{2l_6}) \end{pmatrix}.$$

Now we consider the symmetric tensors $U = S^2(R_1)$, $\bar{U} = S^2(R_2)$ which are spanned by the basis $u_{ij} = \frac{1}{2}(r_i \otimes r_j + r_j \otimes r_i)$, $\bar{u}_{ij} = \frac{1}{2}(\bar{r}_i \otimes \bar{r}_j + \bar{r}_j \otimes \bar{r}_i)$, $1 \leqq i \leqq j \leqq 6$, respectively. Clearly the spaces U, $\bar{U}$ are the ω, ω^2-eigenspaces of z, and their dimensions are 21, respectively. The group $C_0/\langle s \rangle$ acts on U (resp. $\bar{U}$) faithfully and irreducibly. We set $C_* = C_0/\langle s \rangle$, $A = A_0/\langle s \rangle$, and $B = B(6) = U \oplus \bar{U}$.

Notice that B means $B^{(3)}$ or $B^{(4)}$ according as $\Lambda = \Lambda^{(3)}$ or $\Lambda^{(4)}$.

(3.5) We present another description of $B (= B^{(3)}$ or $B^{(4)})$.

We consider the subspace of the linear space $M(12, \mathbb{C})$ consisting of the 12×12-matrices $\begin{pmatrix} O & X_2 \\ X_1 & O \end{pmatrix}$ such that X_1, X_2 are 6×6-matrices whose entries are in $\mathbb{C}$, and ${}^t X_i = X_i$ for $i = 1, 2$. We set $J_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $J_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and denote $\begin{pmatrix} O & O \\ X & O \end{pmatrix}$ by $J_1 \otimes X$, and $\begin{pmatrix} O & X \\ O & O \end{pmatrix}$ by $J_2 \otimes X$.

We can identify this space with B, under the identifications, $u_{ij} = J_1 \otimes (A_{ij} + A_{ji})$, $\bar{u}_{ij} = J_2 \otimes (A_{ij} + A_{ji})$, where A_{ij} is the 6×6-matrix with its $i \times j$-entry 1, and the other entries 0. The actions of elements of C_* are described as follows. Let X be an element of $\mathrm{Aut}(\Lambda)$, and $x = \begin{pmatrix} X & O \\ O & \bar{X} \end{pmatrix} \in C_0$, and x' be the projective

image of x in C_*. Then for $b=\begin{pmatrix} O & A' \\ A & O \end{pmatrix} \in B$, we have $x'(b)=x b x^{-1}$ $=\begin{pmatrix} O & {}^t\bar{X}A'\bar{X} \\ {}^tXAX & O \end{pmatrix}$. Moreover for $\xi=\begin{pmatrix} O & I \\ I & O \end{pmatrix}$, $\xi'=\begin{pmatrix} O & I-2A_{11} \\ I-2A_{11} & O \end{pmatrix}$, we have $\xi(b)=\xi b \xi^{-1}$, $\xi'(b)=\xi' b \xi'^{-1}$.

Note that each element of $B(=B^{(3)}$ or $B^{(4)})$ can be regarded as an endomorphism on $R(=R^{(3)}$ or $R^{(4)})$.

(3.6) *The group $N(6)$ is uniquely determined up to isomorphism by the conditions* (i), (ii) *in Sect.* 1, *and the property $N(6)'/Z(N(6)') \simeq \Omega^-(6, 3)$. Moreover C_* contains a subgroup N_0 isomorphic to $N(6)$.*

Proof. It is well known that $\mathrm{Out}(\Omega^-(6, 3)) \simeq D_8$, the dihedral group of order 8, and there exist (at least) two extensions of $\Omega^-(6, 3)$ by Z_2 which are generated by a class of 3-transpositions, as subgroups of $\mathrm{Aut}(\Omega^-(6, 3))$. In fact, such extensions are obtained as the groups generated by the reflections with respect to vectors of norm $\varepsilon(=\pm 1)$ on the 6-dimensional orthogonal space over $\mathbb{F}_3$.

Such extensions are isomorphic to each other, and so there exist only two extensions as above in $\mathrm{Aut}(\Omega^-(6, 3))$ by the structure of D_8. On the other hand, there exists a unique maximal subgroup of D_8 containing a fixed non-central involution, and it is generated by the fixed element and its conjugate element. Hence the structure of $C_*/\langle z\rangle$, an extension of $\Omega^-(6, 3)$ by 2^2, is uniquely determined and contains a subgroup $N_0/\langle z\rangle$ ($\neq A/\langle z\rangle$ defined in (3.4)) generated by a class of 3-transpositions as well as $A/\langle z\rangle$. Since $N_0 \neq A$ and ξ inverts z, the group N_0 satisfies the conditions (i), (ii) in (1.1). Denote by D_0 the conjugacy class which generates N_0 and whose image in $N_0/\langle z\rangle$ is a class of 3-transpositions.

Suppose $N=\langle D\rangle$ is a group with the conditions (i), (ii) in (1.1). By Theorem A, the group N' has a 21-dimensional faithful representation. There exist just two perfect extensions of Z_3 by $\Omega^-(6, 3)$ up to isomorphism. By the character tables of these groups in [3, 12], we have that the group N' is uniquely determined, and thus it is isomorphic to $N_0'=C_*'=A'$.

On the other hand, $N/Z(N')$ is isomorphic to $N_0/\langle z\rangle$. Hence for $d \in D$, there exists some $x \in D_0$ such that d and x induce the same automorphism on $N/Z(N') \simeq N_0/\langle z\rangle$. Moreover they invert $Z(N')$ and z, respectively. Now it is easily verified that N is isomorphic to N_0.

(3.7) Now we can determine the algebra $(B(6), \tau(6))$.

Theorem 1$^{(3)}$. *Let $\tau=\tau^{(3)}: B^{(3)} \times B^{(3)} \to B^{(3)}$ be a symmetric bilinear map defined by:*

$$\begin{aligned}
&\tau(u_{ii}, u_{ii})=2\bar{u}_{ii},\\
&\tau(u_{ii}, u_{jj})=-2\bar{u}_{ij},\\
&\tau(u_{ii}, u_{ij})=-\bar{u}_{jj},\\
&\tau(u_{ii}, u_{jk})=0,\\
&\tau(u_{ij}, u_{ij})=-2\bar{u}_{ij},\\
&\tau(u_{ij}, u_{jk})=0,\\
&\tau(u_{ij}, u_{kh})=-\bar{u}_{st},
\end{aligned}$$

where the letters i, j, k, h, s, t are mutually distinct, and moreover $\tau(U, \bar{U})=0$, *and the formulas about* $\tau(\bar{U}, \bar{U})$ *are given by changing the roles of* u_{ij} *and* $\bar{u}_{ij}$, *with the same structure constants.*

Then $(B^{(3)}, \tau^{(3)})$ *is a* $C_*^{(3)}$*-invariant commutative algebra, and is uniquely determined up to multiplication by a constant, as an* N_0*-invariant 42-dimensional non-trivial algebra. In particular,* $(B^{(3)}, \tau^{(3)})$ *is isomorphic to the algebra* $(B(6), \tau(6))$.

This algebra is also presented as follows, by using the notation with superscript (4).

Theorem 1$^{(4)}$. *The unique non-trivial* $C_*^{(4)}$*-invariant algebra* $(B^{(4)}, \tau^{(4)})$ *is given, up to constant, by:*

$$\begin{aligned}
&\tau(u_{ii}, u_{ii})=\tfrac{3}{2}(\bar{u}_{11}+\ldots+\bar{u}_{66}),\\
&\tau(u_{ii}, u_{jj})=\tau(u_{ij}, u_{ij})=\tfrac{1}{2}(3(\bar{u}_{ii}+\bar{u}_{jj})+\sum_{k\neq i,\, j}\bar{u}_{kk}),\\
&\tau(u_{ii}, u_{ij})=3\bar{u}_{ij},\\
&\tau(u_{ii}, u_{jk})=-\bar{u}_{jk},\\
&\tau(u_{ij}, u_{jk})=-\bar{u}_{ik},\\
&\tau(u_{ij}, u_{kh})=-\theta\bar{u}_{st},
\end{aligned}$$

where the letters i, j, k, h, s, t are mutually distinct, and moreover $\tau(U, \bar{U})=0$, *and the formulas about* $\tau(\bar{U}, \bar{U})$ *are given by changing the roles of* u_{ij} *and* $\bar{u}_{ij}$, *with the same structure constants, except the formula* $\tau(\bar{u}_{ij}, \bar{u}_{kh})=\theta u_{st}$ *(not* $-\theta u_{st}$*).*

The above theorems are verified by direct examinations using the explicit presentations of actions of elements of C_* on B mentioned above. So we omit the proofs, which are similar to the proof of Griess [10: Lemma 5.3].

Next we state one of our main results.

Theorem 2. $\mathrm{Aut}(B(6), \tau(6))=C_*$.

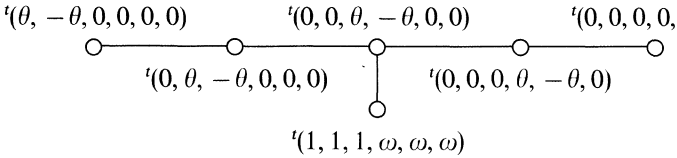
This theorem is used in the later sections. But our proof of Theorem 2 is quite similar to the proof of Theorem 5 in Sect. 6. So we do not prove Theorem 2 here, and present the outline of the proof in Sect. 7, comparing with the proof of Theorem 5.

(3.8) We determine the D-vectors v_d, $\bar{v}_d$ in $B^{(3)}$ and $B^{(4)}$.

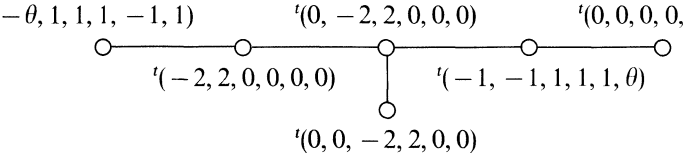
By the definition, the vectors v_d, $\bar{v}_d$ are in the 1-eigenspace of $C_{N'}(d)$. So we consider $C_A(d)$ (or $C_{A_0}(d)$), which contains $C_{N'}(d)$ with index 2. Moreover $C_A(d)$ is isomorphic to the Weyl group $W(E_6)$ of type E_6, since it is generated by a subset of E (a class of 3-transpositions) and $C_{N'}(d)/\langle d, s\rangle$ is isomorphic to $\Omega(5, 3)\simeq W(E_6)'$. This is uniquely determined up to conjugate in A.

We present some vectors of minimal length in Λ as follows. The reflections with respect to them generate $C_{A_0}(d)$, and satisfy the relations given by the Coxeter diagram of type E_6;

$B^{(3)}$:



$B^{(4)}$:



The 1-eigenspace of $C_{A'}(d)$ is as follows:

$$\left\langle \sum_{i=1}^{6} u_{ii} - \sum_{i<j} u_{ij}, \sum_{i=1}^{6} \bar{u}_{ii} - \sum_{i<j} \bar{u}_{ij} \right\rangle \quad \text{in } B^{(3)},$$

$$\langle u_{11} - \sum_{i \neq 1} u_{ii}, \bar{u}_{11} - \sum_{i \neq 1} \bar{u}_{ii} \rangle \quad \text{in } B^{(4)}.$$

Hence we may set v_d, $\bar{v}_d$ as follows, so that $\tau(v_d, v_d) = 4\bar{v}_d$.

$$v_d = \frac{1}{3}\left(\sum_{i=1}^{6} u_{ii} - \sum_{i<j}^{6} u_{ij}\right) \text{ in } B^{(3)}, \quad \frac{1}{4}\left(\sum_{i \neq 1} u_{ii} - u_{11}\right) \text{ in } B^{(4)},$$

$$\bar{v}_d = \frac{1}{3}\left(\sum_{i=1}^{6} \bar{u}_{ii} - \sum_{i<j}^{6} \bar{u}_{ij}\right) \text{ in } B^{(3)}, \quad \frac{1}{4}\left(\sum_{i \neq 1} \bar{u}_{ii} - \bar{u}_{11}\right) \text{ in } B^{(4)}.$$

By considering the action of C_*, we can list all the D-vectors as follows:

D-vectors in $U^{(3)}$ (Total 3×126 vectors):

$$\frac{1}{3}\left(\sum_{i=1}^{6} \omega^{2l_i} u_{ii} - \sum_{i \neq j} \omega^{(l_i + l_j)} u_{ij}\right)$$

$$\left(l_i \in \mathbb{Z}, \sum_{i=1}^{6} l_i \equiv 0 \pmod 3\right), \quad (3 \times 3^4 \text{ vectors})$$

$$-(\omega^{l(i,j)} u_{ij} + \omega^{l(k,h)} u_{kh} + \omega^{l(s,t)} u_{st})$$

$$(\{i, j, k, h, s, t\} = \{1, 2, 3, 4, 5, 6\},\ l(i, j) \in \mathbb{Z},$$

$$l(i, j) + l(k, h) + l(s, t) \equiv 0 \pmod 3), \quad (3 \times 45 \text{ vectors}).$$

D-vectors in $U^{(4)}$ (Total 3×126 vectors):

$$\frac{\omega^l}{4}\left(\sum_{j \neq i} u_{jj} + u_{ii}\right) \quad (1 \leqq i \leqq 6), \quad (3 \times 6 \text{ vectors})$$

$$\frac{\omega^l}{8}\left(\sum_{j \neq s,t} u_{jj} - u_{ss} - u_{tt} - \sum_{\substack{i,j \neq s,t \\ i<j}} 2\varepsilon_i \varepsilon_j u_{ij} + 2\theta\left(\prod_{i \neq s,t} \varepsilon_i\right) u_{st}\right)$$

$$(1 \leqq s \neq t \leqq 6,\ \varepsilon_i = \pm 1,\ l \in \mathbb{Z}), \quad (3 \times 120 \text{ vectors}).$$

Moreover if v_d is a D-vector in U, then the D-vector $\bar{v}_d$ is obtained by changing u_{ij} into $\bar{u}_{ij}$, and its coefficients into their complex conjugate. For example, if $v_d = -(u_{12}+\omega u_{34}+\omega^2 u_{56})$, then $\bar{v}_d = -(\bar{u}_{12}+\omega^2 \bar{u}_{34}+\omega \bar{u}_{56})$.

(3.9) By Theorem A, we can know the order of the product of any two elements a, b of D by calculating $\tau(v_a, v_b)$. We mention only the partial results for $D^{(4)}$, which are used later.

Let a, b, c be elements of $D^{(4)}$ with $v_a = \frac{1}{4}(\sum_{j\neq 1} u_{jj} - u_{11})$,

$$v_b = \tfrac{1}{4}(\sum_{j\neq k} u_{jj} - u_{kk}), \qquad k \neq 1,$$

and

$$v_c = \tfrac{1}{8}(\sum_{j\neq s,t} u_{jj} - u_{ss} - u_{tt} - \sum_{\substack{i,j\neq s,t \\ i<j}} 2\varepsilon_i\varepsilon_j u_{ij} + 2\theta(\prod_{i\neq s,t}\varepsilon_i)\, u_{st}).$$

Then by direct calculations, we have the following:

(3.9.1) $\tau(v_a, v_b) = \bar{v}_a + \bar{v}_b$, and so $|ab| = 2$.

(3.9.2) if $1 \in \{s,t\}$, then $\tau(v_a, v_c) = \bar{v}_a + \bar{v}_c$, and so $|ac| = 2$.

(3.9.2′) if $1 \notin \{s,t\}$, then

$$\tau(v_a, v_c) = \tfrac{1}{8}(\sum_{j\neq s,t} \bar{u}_{jj} - \bar{u}_{ss} - \bar{u}_{tt} - \sum_{\substack{i,j\neq s,t \\ i<j}} 2\varepsilon'_i\varepsilon'_j \bar{u}_{ij} + 2\theta(\prod_{i\neq s,t}\varepsilon'_i)\,\bar{u}_{st})$$

where $\varepsilon'_1 = -\varepsilon_1$, and $\varepsilon'_i = \varepsilon_i$ for $i \neq 1$, and in particular, $\tau(v_a, v_c) \neq \omega^l \bar{v}_a + \omega^{2l}\bar{v}_c$ for any l, and so $|ac| = 3$.

By using them, we can determine the action of elements of $D^{(4)}$ on $B^{(4)}$. We present their inverse images in $C_0^{(4)} \subset GL(R^{(4)})$.

Lemma 3.9.1. *Let $d \in D^{(4)}$, and d_0 be in the inverse image of d in $C_0^{(4)}$. Then $d_0 = \pm 4(v_d + \bar{v}_d)$. Here $B^{(4)}$ is regarded as a subspace of* $\mathrm{End}(R^{(4)})$.

Proof. Let a, b, c be as in (3.9). First we show that the action of a coincides with that of ξ'. By the definition of the algebra in Sect. 1, and the definition of ξ', we have $a(v_b) = \bar{v}_{b^a} = \bar{v}_b = \xi'(v_b)$, and $a(v_c) = \bar{v}_c = \xi'(v_c)$ if $i \in \{s,t\}$. Suppose $i \notin \{s,t\}$. Then $a(v_c) = \bar{v}_{c^a}$, and the right-hand side is as in (3.9.2′). Hence we have $a(v_c) = \xi'(v_c)$, and the assertion is proved, since B is spanned by the D-vectors.

Hence we have $a = \xi'$, and $\pm(J_1 \otimes (I - 2A_{11}) + J_2 \otimes (I - 2A_{11})) = \pm 4(v_a + \bar{v}_a)$ are the elements of the inverse image of a. Since D is a conjugacy class, and $g(v_a) = v_{a^g}$ for $g \in C_*$, we have the lemma.

4. The algebra $(B(7), \tau(7))$ for $\Omega(7,3)$

In this section, we determine the structure of the algebra $(B(7), \tau(7))$.

We use the notation with the superscript (3) in Sect. 3, especially $A_0^{(3)}$, $C_0^{(3)}$, $B^{(3)}$, $\tau^{(3)}$, $R^{(3)}$. In this section, (B, τ) always denotes $(B(7), \tau(7))$.

Set $K = N(7)$, $D = D(7)$, and $Z(K') = \langle z \rangle$. Then K' is generated by a con-

jugacy class E of 3-transpositions, since $K'/Z(K')$ is so. Set U, $\bar{U}$ be the ω, ω^2-eigenspaces of z in B respectively.

(4.1) First we consider $C_K(e)$ for $e \in E$.

Set $C = C_{K'}(e)$. Then by the structure of $\Omega(7, 3)$, the group $C/\langle z \rangle$ is a non-split extension of Z_2 by $\Omega^-(6, 3).2$. The group C acts on U, and acts faithfully on the (-1)-eigenspace of e in U.

We denote by $\chi(g)$ the trace of the action of $g \in K$ on U.

If C is split over $\langle z \rangle$, then the dimension of its faithful irreducible representation with its degree less than 27 is 20, and if C is non-split over $\langle z \rangle$, then that is 6 by [3, 12]. If the former case occurs, then $\chi(e) = 7 - 20 = -13$. But in this case, the 20-dimensional character has the value ± 4 on $C \cap E$ by [3, 12]. Hence for $e' \in C \cap E$, we have $|\chi(e')| \leqq 4 + 7 = 11$. This contradicts the fact that e is conjugate with e' in K'. Hence C is non-split, and thus C is isomorphic to $A_0^{(3)}$, and furthermore $C_K(e)$ is isomorphic to $C_0^{(3)}$ (see the proof of (3.6)). Hence we identify $C_K(e)$ with $C_0^{(3)}$.

Moreover, by the definition of the Conway-Norton algebras, the space U contains the 21-dimensional subspace $U_+ = \langle v_d \mid d \in C_D(e) \rangle$, on which $C/\langle e \rangle$ acts irreducibly. Hence the (-1)-eigenspace of e in U is 6-dimensional. We can identify the 1-eigenspace U_+ of e in B with $B^{(3)}$, and the (-1)-eigenspace U_- of e in B with $R^{(3)}$, with the action of $C_K(e)$ as described in Sect. 3.

(4.2) Let $f \in E - C_E(e)$. Then $K = \langle C_K(e), f \rangle$. Hence in order to determine the action of K on B, it is sufficient to determine that of f.

By the structure of $\Omega(7, 3)$, we may assume $C_K(e, f) = W^{(3)}$. Set $U_0 = \langle u_{ii} \mid 1 \leqq i \leqq 6 \rangle$, $U_1 = \langle u_{ij} \mid 1 \leqq i \neq j \leqq 6 \rangle$, the 6 and 15-dimensional subspaces of $B^{(3)} \cap U$. Then they are $K' \cap W^{(3)}$-irreducible, respectively. The group $K' \cap W^{(3)}$ acts on the (-1)-eigenspace V of $\langle e, f \rangle$. Since the 1-eigenspace of f is also 21-dimensional, we have $\dim(V) \geqq 21 + 21 - 27 = 15$. Hence we have $V = U_1$.

Since $K' \cap W^{(3)}$ acts irreducibly on U_0, U_- respectively, we can take the following basis x_i, $y_i (1 \leqq i \leqq 6)$ of $U_0 \oplus U_-$:

$$\begin{aligned} e(x_i) &= y_i, \quad & e(y_i) &= x_i, \\ f(x_i) &= \omega y_i, \quad & f(y_i) &= \omega^2 x_i. \end{aligned} \qquad (1 \leqq i \leqq 6) \tag{4.2.1}$$

The 1-eigenspace of e in $U_0 \oplus U_-$ is $\langle x_i + y_i \mid 1 \leqq i \leqq 6 \rangle$. Hence we may assume

$$u_{ii} = x_i + y_i \quad (1 \leqq i \leqq 6). \tag{4.2.2}$$

We show that

$$r_i = x_i - y_i \quad (1 \leqq i \leqq 6), \tag{4.2.3}$$

with suitable normalization.

Set $\delta = D(\omega^{l_1}, \ldots, \omega^{l_6})$ and $\alpha = T_{ij}$, which are defined in (3.2). Then $\delta(x_i + y_i) = \delta(u_{ii}) = \omega^{2l_i} u_{ii} = \omega^{2l_i}(x_i + y_i)$, and $\delta(\omega^2 x_i + \omega y_i) = \delta f(x_i + y_i) = f\delta(x_i + y_i) = \omega^{2l_i+2} x_i + \omega^{2l_i+1} y_i$. Hence it is easily seen that $\delta(x_i) = \omega^{2l_i} x_i$, $\delta(y_i) = \omega^{2l_i} y_i$ for each $i = 1, \ldots, 6$.

By similar calculations, we have $\alpha(x_i)=x_j$, $\alpha(y_i)=y_j$, and $\alpha(x_k)=x_k$, $\alpha(y_k)=y_k$ for each $k\neq i,j$.

Let W_i be the stabilizer of the vector u_{ii} in $W^{(3)}$ for $i=1,\ldots,6$. The above calculations imply that the 1-eigenspace of W_i in $U_0\oplus U_-$ is $\langle x_i, y_i\rangle$. Since $r_i\in\langle x_i, y_i\rangle$ and $e(r_i)=-r_i$, we have $r_i\in\langle x_i-y_i\rangle$. Since $T_{ij}(r_i)=r_j$, there exists some constant c with $r_i=c(x_i-y_i)$ for all i. Hence we may assume $r_i=x_i-y_i$ under the condition (4.2.1).

Similarly, there exist $\bar{x}_i$, $\bar{y}_i$ $(1\leqq i\leqq 6)$ such that:

$$\bar{u}_{ii}=\bar{x}_i+\bar{y}_i,\qquad \bar{r}_i=\bar{x}_i-\bar{y}_i\qquad (1\leqq i\leqq 6),$$
$$e(\bar{x}_i)=\bar{y}_i,\qquad e(\bar{y}_i)=\bar{x}_i,\qquad f(\bar{x}_i)=\omega^2\bar{y}_i,\qquad f(\bar{y}_i)=\omega\bar{x}_i.$$

Notice that we have taken the basis of $\bar{U}$ so that the representation of each element of K' on $\bar{U}$ is given as the complex conjugate of that on U.

(4.3) Now we determine $\tau=\tau(7)$. Clearly $\tau|_{U_+}$ is non-trivial. Hence we may assume $\tau|_{U_+}=\tau^{(3)}$ by Theorem $1^{(3)}$.

Set $u_{ii}^{(l)}=\omega^{2l}x_i+\omega^l y_i$, $\bar{u}_{ii}^{(l)}=\omega^l\bar{x}_i+\omega^{2l}\bar{y}_i$, $1\leqq i\leqq 6$, $l=0,1,2$. Then $u_{ii}^{(0)}=u_{ii}$, $u_{ii}^{(1)}=f(u_{ii})$, $u_{ii}^{(2)}=ef(u_{ii})$. By Theorem $1^{(3)}$ and the actions of f, ef, we have the following:

(4.3.1) $$\tau(u_{ii}^{(l)}, u_{ii}^{(l)})=2\bar{u}_{ii}^{(l)},$$

(4.3.2) $$\tau(u_{ii}^{(l)}, u_{jj}^{(l)})=2\bar{u}_{ij},$$

(4.3.3) $$\tau(u_{ii}^{(l)}, u_{ij})=-\bar{u}_{jj}^{(l)},$$

(4.3.4) $$\tau(u_{ii}^{(l)}, u_{jk})=0.$$

By using the above, we have the following:

(4.3.1′) $$\tau(x_i, y_i)=0,\qquad \tau(x_i, x_i)=2\bar{x}_i,\qquad \tau(y_i, y_i)=2\bar{y}_i,$$

(4.3.2′) $$\tau(x_i, x_j)=\tau(y_i, y_j)=0,$$

(4.3.2″) $$\tau(x_i, y_j)+\tau(x_j, y_i)=-2\bar{u}_{ij},$$

(4.3.3′) $$\tau(x_i, u_{ij})=-\bar{y}_j,\qquad \tau(y_i, u_{ij})=-\bar{x}_j,$$

(4.3.4′) $$\tau(x_i, u_{jk})=\tau(y_i, u_{jk})=0\qquad (i\neq j\neq k\neq i).$$

By using (4.3.2′), we have:

$$\tau(x_i+y_i, x_j-y_j)=\tau(x_j, y_i)-\tau(x_i, y_j)\in\langle\bar{x}_k-\bar{y}_k \mid 1\leqq k\leqq 6\rangle,$$

since the left-hand side is contained in the (-1)-eigenspace of e. Hence by (4.3.2″), we have that there exists some $\bar{v}\in\langle\bar{x}_k-\bar{y}_k\rangle$ such that $\tau(x_i, y_j)=-\bar{u}_{ij}+\bar{v}$, $\tau(x_j, y_i)=-\bar{u}_{ij}-\bar{v}$. Hence $\bar{v}=\bar{u}_{ij}+\tau(x_i, y_j)$, and the right-hand side is stable by $D(\omega^{l_1},\ldots,\omega^{l_6})$ with $l_i=1$, $l_j=2$. Since l_k $(k\neq i,j)$ are arbitary under $\sum_{h=1}^{6} l_h\equiv 0 \pmod 3$, we have $\bar{v}=0$, and thus:

(4.3.5) $$\tau(x_i, y_j)=-\bar{u}_{ij}\qquad (i\neq j).$$

Now we have determined τ completely. The formulas about $\tau(\bar{U}, \bar{U})$ are given by changing the roles of u_{ij}, x_i, y_i and $\bar{u}_{ij}$, $\bar{x}_i$, $\bar{y}_i$ with the same structure constants.

Theorem 3. *Let $B^{(3)}$, $R^{(3)}$, and $\tau^{(3)}$ be as in Sect. 3. Set $B=B^{(3)}\oplus R^{(3)}$ and $\tau: B\times B\to B$ be a symmetric bilinear map defined by the following:*

(i) $\tau|_{B^{(3)}}=\tau^{(3)}$,

(ii) $\tau(b, r)=b(r)$ *for any* $b\in B^{(3)}$, $r\in R^{(3)}$, *where $B^{(3)}$ is regarded as a subspace of* $\mathrm{End}(R^{(3)})$ *as in Sect. 3,*

(iii) $\tau(r_i, r_j)=2\bar{u}_{ij}$, $\tau(\bar{r}_i, \bar{r}_j)=2u_{ij}$, $\tau(r_i, \bar{r}_j)=0$ $(1\leqq i, j\leqq 6)$.

Then (B, τ) is a K-invariant algebra, and is isomorphic to the Conway-Norton algebra $(B(7), \tau(7))$.

This theorem is easily checked by the formulas (5.3.4′)–(5.3.8) together with (4.2.3). The following is one of our main theorems. We prove this in Sect. 7.

Theorem 4. $\mathrm{Aut}(B(7), \tau(7))=K=N(7)$.

(4.4) We give a simple description of $(B(7), \tau(7))$.

Set $u'_{ij}=-u_{ij}$, $\bar{u}'_{ij}=-\bar{u}_{ij}$. We consider the set $\mathscr{P}$ of the pairs of vectors:

$$\mathscr{P}=\{(u'_{ij}, \bar{u}'_{ij}), (x_i, \bar{x}_i), (y_i, \bar{y}_i) \mid 1\leqq i\neq j\leqq 6\}.$$

We define the set $\mathscr{L}$ of lines as a family of subsets of $\mathscr{P}$. Let P_1, P_2, $P_3\in\mathscr{P}$ with $P_i=(p^{(i)}, \bar{p}^{(i)})$. Then $\{P_1, P_2, P_3\}$ is a line if and only if $\tau(p^{(i)}, p^{(j)})=\bar{p}^{(k)}$ where $\{i, j, k\}=\{1, 2, 3\}$. Then the geometry $(\mathscr{P}, \mathscr{L})$ is well-known, that is, it is isomorphic to the geometry described in [5; Note H], which is isomorphic to the geometry of isotropic points and lines of $\Omega^-(6, 2)$-projective space, that is, the 5-dimensional projective orthogonal space of minus type over $\mathbb{F}_2$. Hence we have the following:

Proposition. *Let $\mathscr{P}$, $\mathscr{L}$ be the sets of the isotropic points and lines of the $\Omega^-(6, 2)$-projective space respectively. Let B' be a 54-dimensional vector space over $\mathbb{C}$ with basis $v_p^{(1)}$, $v_p^{(2)}$ $(p\in\mathscr{P})$, and $\tau': B'\times B'\to B'$ be a bilinear symmetric map defined by the following:*

$$\tau'(v_p^{(i)}, v_q^{(j)})=\begin{cases} 2v_p^{(k)} & (p=q, i=j\neq k),\\ v_r^{(k)} & (\{p, q, r\}\in\mathscr{L}, i=j\neq k),\\ 0 & (\textit{otherwise}),\end{cases}$$

where $i, j, k\in\{1, 2\}$.

Then (B', τ') is isomorphic to $(B(7), \tau(7))$.

(4.5) Finally we list all the $D(7)$-vectors, which are not used in the later sections.

$$\frac{1}{3}\left(\sum_{i=1}^{6}\omega^{2l_i}u_{ii}^{(m)}-\sum_{i<j}\omega^{l_i+l_j}u_{ij}\right)$$
$$\left(m, l_i\in\{0, 1, 2\}, \sum_{i=1}^{6}l_i\equiv 0\ (\mathrm{mod}\ 3)\right) \qquad (3\times 243 \text{ vectors})$$
$$-(\omega^{l(i,j)}u_{ij}+\omega^{l(k,h)}u_{kh}+\omega^{l(s,t)}u_{st})$$

$$(\{i,j,k,h,s,t\}=\{1,2,3,4,5,6\}, \quad l(i,j)\in\{0,1,2\},$$
$$l(i,j)+l(k,h)+l(s,t)\equiv 0 \pmod 3) \qquad (3\times 45 \text{ vectors})$$
$$\omega^{l_i}x_i+\omega^{l_j}y_j-\omega^{2(l_i+l_j)}u_{ij} \quad (i\neq j) \qquad (3\times 90 \text{ vectors})$$
$$(\text{total } 3\times 378 \text{ vectors})$$

5. The algebra $(B(24), \tau(24))$ for F_{24}'

In this section, we consider the algebra $(B(24), \tau(24))$. We set $G=N(24)$, $D=D(24)$, and $Z(G')=\langle z\rangle$.

(5.1) First we summarize some properties of G, which are deduced from the structure of F_{24} (e.g. see [9, 11]).

We fix a Sylow 2-subgroup of G. Let L be an intersection of D with the fixed Sylow 2-subgroup. Then L consists of 24 elements, which generate an elementary abelian group of order 2^{12}. The group $N_G(L)$ acts on L 5-transitively, and $C_G(L)=\langle L\rangle$. The quotient $N_G(L)/\langle L\rangle$ is isomorphic to the Mathieu group M_{24}. The group $N_{G'}(L)$ ($\simeq 2^{11}.M_{24}$) is perfect, and moreover for any $a\in L$, $N_{G'}(L)\cap C_G(a)$ ($\simeq 2^{11}.M_{23}$) is perfect. We set $M=N_G(L)\langle z\rangle$.

Let $\mathcal{O}$ be the family of 8-elements subsets X of L such that the product of the elements of X equals the unit. Then any 5-elements subset of L is contained in a unique member of $\mathcal{O}$, that is, the pair $(L, \mathcal{O})$ is the Steiner system $S(24, 8, 5)$ ([6, 17]). Members of $\mathcal{O}$ are called octads. We assume the readers are familiar with M_{24} and $S(24, 8, 5)$.

We fix a sextet $\{T_1, \ldots, T_6\}$ (i.e. $L=T_1\cup\ldots\cup T_6$, $|T_i|=4$ and $T_i\cup T_j\in\mathcal{O}$ $(i\neq j)$), and set $T_i=\{d_{4i-3}, d_{4i-2}, d_{4i-1}, d_{4i}\}$. Then we have $d_1d_2d_3d_4=\ldots=d_{21}d_{22}d_{23}d_{24}$. We set $s=d_1d_2d_3d_4$. Then $C=C_G(s)\simeq(3\times 2_+^{1+12}).3.\Omega^-(6,3).2^2$. We set $Q=O_2(C)\simeq 2_+^{1+12}$.

The group G has four classes of involutions with representatives d_1, d_1d_2, $d_1d_2d_3$, and s. We denote these classes by D_1, D_2, D_3, D_4 respectively. Note that D_2 and D_4 are contained in G', and D_1 and D_3 are not.

Concerning D_2, we have $xy=x'y'$ for $x, y, x', y'\in D$ with $x\neq y$, if and only if $\{x, y\}=\{x', y'\}^{z^i}$ for some i.

Concerning D_4, we have $s=abcd$ for $a, b, c, d\in D$, if and only if $a\in C_D(s)$ and $a^Q=\{a, b, c, d\}$.

We set $Q_2=\{xy \mid x\in C_D(s), y\in x^Q\}$. Then Q_2 is a conjugacy class of $C_G(s)$, and $Q=\langle Q_2\rangle$.

The group $C/Q\langle z\rangle$ is isomorphic to C_* defined in Sect. 3. Here the conjugacy class $C_D(s)Q\langle z\rangle/Q\langle z\rangle$ corresponds with E in Sect. 3, and thus $|C_D(s)|=126\times 3\times 4$, $|Q_2|=126\times\binom{4}{2}=756$. Let ζ be an element of the inverse image of the center of $C'/Q\langle z\rangle$. Here we can choose ζ as an element of order 3 in $N_G(L)$, and we may assume $d_{4i-3}^{\zeta}=d_{4i-3}$ and $d_{4i}=d_{4i-1}^{\zeta}=d_{4i-2}^{\zeta^2}$ for each $i=1, \ldots, 6$.

(5.2) Now we consider the Conway-Norton algebra for F_{24}'. A summary of our results is in (5.14), the end of this section.

Set $(B, \tau)=(B(24), \tau(24))$, and U, $\bar{U}$ be the ω, ω^2-eigenspaces of z, respectively. We write $v_i=v_{d_i}$, $\bar{v}_i=\bar{v}_{d_i}$, the D-vectors for $d_i \in L$ $(1 \leqq i \leqq 24)$. By Theorem A, we have $\tau(U, \bar{U})=0$, and:

$$\begin{aligned} &\tau(v_i, v_i)=10\bar{v}_i, \quad \tau(\bar{v}_i, \bar{v}_i)=10v_i, \\ &\tau(v_i, v_j)=\bar{v}_i+\bar{v}_j, \quad \tau(\bar{v}_i, \bar{v}_j)=v_i+v_j. \end{aligned} \qquad (1 \leqq i \neq j \leqq 24) \tag{5.2.1}$$

(5.3) M-invariant spaces

The group $G' \cap M$ acts on U and $\bar{U}$ respectively. Then we can write $U=\bigoplus_{\alpha} U_\alpha$, where U_α is an isotypic component of $G' \cap \langle L \rangle$ which affords a character α. Then $G' \cap M$ permutes the spaces U_α. By considering the subgroups of M_{24}, we have each orbit length of this permutation representation is 1, 24, or 759. The permutation representation of these degrees are uniquely determined, respectively: the trivial one, the one on L of degree 24, the one on $\mathcal{O}$ of degree 759. Since $C_{G' \cap M}(d_1)$ acts on $G' \cap \langle L \rangle$ irreducibly, the permutation representation of degree 24 does not appear here. On the other hand, the vectors $v_1, \ldots, v_{24}$ generate a 24-dimensional subspace. In fact, $v_1+\ldots+v_{24}$ $(=w$, say$) \neq 0$, since $\tau(v_1, w)=23\bar{v}_1+\bar{v}_2+\ldots+\bar{v}_{24}=22\bar{v}_1+d_1(w)$ by (5.2.1). Hence the 5-transitivity of $N_G(L)$ on L implies $\dim(\langle v_1, \ldots, v_{24} \rangle)=24$.

Since $783=24+759$, we have a decomposition of U:

$$U=\langle v_1, \ldots, v_{24} \rangle \oplus (\bigoplus_{x \in \mathcal{O}} U_X),$$

where the 1-dimensional space U_X affords a non-trivial character of $G' \cap \langle L \rangle$, and the stabilizer of U_X in $G' \cap M$ is $N_{G' \cap M}(X)$ for each $X \in \mathcal{O}$. Similarly we have

$$\bar{U}=\langle \bar{v}_1, \ldots, \bar{v}_{24} \rangle \oplus (\bigoplus_{X \in \mathcal{O}} \bar{U}_X).$$

We show that for each octad X, $a(u)=-b(u)$ for any a, b, u with $a \in X$, $b \notin X$, and $u \in U_X$. Clearly $a(u)=\pm b(u)$. Suppose $a(u)=b(u)$. Then the action of $N_{G' \cap M}(X)$ on L implies that $G' \cap \langle L \rangle$ preserves the vector u ([6]; §3, Theorem 10). Hence we have a contradiction and so the assertion.

Since $N_{G' \cap M}(X)/G' \cap \langle L \rangle$ $(\simeq 2^4.A_8$ ([6])) is perfect, we have $g(u)=\pm u$ for any $u \in U_X$ and $g \in N_{G' \cap M}(X)$. Hence by taking a suitable basis $v_X \in U_X$, $\bar{v}_X \in \bar{U}_X$, we have:

$$B=\langle v_i, \bar{v}_i \mid 1 \leqq i \leqq 24 \rangle \oplus \langle v_X, \bar{v}_X \mid X \in \mathcal{O} \rangle,$$

with the action for $m \in G' \cap M$, $a \in L$:

$$\begin{aligned} &m(v_i)=v_{(d_i)m}, \quad m(v_X)=\pm v_{Xm} \\ &a(v_i)=v_i, \quad a(v_X)=-\bar{v}_X (\text{if } a \in X), \quad \bar{v}_X (\text{if } a \notin X). \end{aligned} \tag{5.3.1}$$

(5.4) $C \cap M$-invariant spaces

Set $W=G' \cap C \cap M$. Then the following subspaces $U_1, \ldots, U_5$ of U are W-invariant respectively.

$$U_1=\langle v_1+v_2+v_3+v_4, \ldots, v_{21}+v_{22}+v_{23}+v_{24}\rangle,$$
$$U_2=\langle v_X \mid X=T_i\cup T_j \text{ for some } i\neq j\in\{1,\ldots,6\}\rangle,$$
$$U_3=\langle v_s-v_t \mid d_s, d_t\in T_i \text{ for some } i\in\{1,\ldots,6\}\rangle,$$
$$U_4=\langle v_X \mid X\in\mathcal{O}, |X\cap T_i|=0 \text{ or } 2 \text{ for all } i\in\{1,\ldots,6\}\rangle,$$
$$U_5=\langle v_X \mid X\in\mathcal{O}, |X\cap T_i|=1 \text{ or } 3 \text{ for all } i\in\{1,\ldots,6\}\rangle.$$

Here any octad satisfies one of the conditions in the definitions of the spaces U_2, U_4, and U_5, and thus we have $U=U_1\oplus\ldots\oplus U_5$. The dimensions of $U_1,\ldots,U_5$ are 6, 15, 18, 360, and 384 respectively. Similarly we define the spaces $\bar{U}_i$ by changing v_k, v_X into $\bar{v}_k$, $\bar{v}_X$. Then $a(U_i)=\bar{U}_i$ for any $a\in L$, and thus $U_i\oplus\bar{U}_i$ is $C\cap M$-invariant for each i.

(5.5) Character values of some involutions

We denote by $\chi_V(g)$ the trace of the action of $g\in G'$ on some subspace V of U which is stable by g. We write $\chi_U=\chi$.

Since U_5 is the (-1)-eigenspace of s in U, we have $\chi(s)=6+15+18+360-384=15$.

Since there exist 77 octads containing both d_1 and d_2, and 330 octads containing neither d_1 nor d_2, and 352 octads containing just one of d_1 and d_2, we have $\chi(d_1d_2)=24+77+330-352=79$.

The group $N_G(L)/\langle L\rangle$ has just two classes of involutions, and one consists of the elements which have eight fixed points (making an octad) on L, and the other consists of the elements whose actions on L are fixed point free. Let x be an involution in $N_{G'}(L)-\langle L\rangle$ such that x has no fixed points on L. Then one can check that the number of octads fixed by x is 15. Hence we have $|\chi(r)|\leqq 15$ and so x belongs to D_4. In particular $\chi(x)=15$, and thus for any octad X fixed by x, we have $x(v_X)=v_X$.

Hence any y of D_2 in $N_{G'}(L)-\langle L\rangle$ fixes eight elements of L. This time y fixes 71 octads, and so for any octad X fixed by y, we have $y(v_X)=v_X$ since $\chi(y)=79=8+71$.

Let $x\in G'\cap\langle L\rangle$. Then it is easily seen by using (5.3.1) that $\chi_{U_1\oplus U_2}(x)=5$ or 21, and $\chi_{U_1\oplus U_2}(x)=21$ if and only if $x\in Q_2\cap\langle L\rangle=\langle d_{4i-m}d_{4i-n} \mid 1\leqq i\leqq 6,\ 0\leqq m, n\leqq 3\rangle$.

By [17], the group $W/G'\cap\langle L\rangle$ has six classes of involutions. The numbers of the sets T_i and the octads $T_i\cup T_j$ $(i\neq j)$ fixed by these involutions are (6, 15), (4, 7), (2, 3), (6, 15), (2, 3), or (0, 3).

Hence for any involution x of W, we have $\chi_{U_1\oplus U_2}(x)=3$, 5, 11, or 21. Moreover it is easily seen that any $q\in Q$ satisfies $\chi_{U_1\oplus U_2}(q)=21=\dim(U_1\oplus U_2)$.

The similar facts hold for $\bar{U}$ and $\bar{U}_1\oplus\bar{U}_2$.

(5.6) The C-invariant space B_*

Set $B_*=U_1\oplus U_2\oplus\bar{U}_1\oplus\bar{U}_2$. Then Q acts trivially on B_* by (5.5). On the other hand, it is easily seen that Q acts trivially on no subspaces of $U_3\oplus U_4\oplus U_5$. Hence C acts on B_*.

The group $G'\cap C/Q\langle\zeta\rangle$ acts on the 1-eigenspace of ζ in $U_1\oplus U_2$, which

contains U_1. On the other hand, it is easily seen that U_1 affords two W-irreducible representations with their dimensions 1 and 5, and that U_2 is W-irreducible. Since any extension of Z_3 by $\Omega^-(6,3).2$ has no irreducible representations with their dimensions greater than 2 and less than 15, we have $U_1 \oplus U_2$ itself is the 1-eigenspace of ζ. Similarly ζ acts trivially on $\bar{U}_1 \oplus \bar{U}_2$.

Moreover by the character tables in [3, 12], and the values in (5.5), we have that $C/Q\langle\zeta\rangle$ acts on irreducibly on B_*, and is isomorphic to C_*, since the split extension of Z_3 by $\Omega^-(6,3).2$ does not have character values 3, 11 for any involutions.

Hence $(B_*, \tau|_{B_*})$ is isomorphic to $(B(6), \tau(6))$, since $\tau|_{B_*}$ is non-trivial. In fact, we have $\tau(v_1+v_2+v_3+v_4,\ v_1+v_2+v_3+v_4) = 16(\bar{v}_1+\bar{v}_2+\bar{v}_3+\bar{v}_4) \neq 0$ by (5.2.1). Hence we may set

$$B_* \cap U = \langle u_{ij} \mid 1 \leqq i \leqq j \leqq 6 \rangle, \quad B_* \cap \bar{U} = \langle \bar{u}_{ij} \mid 1 \leqq i \leqq j \leqq 6 \rangle$$

and we may assume $\tau|_{B_*}$ satisfies the formulas in Theorem $1^{(4)}$.

(5.7) The vectors correspondences

Under the isomorphism $C_* \simeq C/Q\langle\zeta\rangle$, the set $D(6)$ corresponds to $C_D(s)Q\langle\zeta\rangle/Q\langle\zeta\rangle$. We denote by $v^*_{\underline{x}}$ (resp. $\bar{v}^*_{\underline{x}}$) the $D(6)$-vector for $xQ\langle\zeta\rangle/Q\langle\zeta\rangle$ with $x \in C_D(s)$ in $B_* \cap U$ (resp. $B_* \cap \bar{U}$). Notice that $v^*_x = v^*_{x'}$ if $x' \in x^Q$.

Set $w^i_1 = \frac{1}{4}(v_{4i-3}+v_{4i-2}+v_{4i-1}+v_{4i})$ $(\in U)$, and $w^i_2 = \frac{1}{4}(\bar{v}_{4i-3}+\bar{v}_{4i-2}+\bar{v}_{4i-1}+\bar{v}_{4i})$ $(\in \bar{U})$. Then w^i_1 and w^i_2 are vectors invariant under the action of $C_{G' \cap C}(d_{4i})$ and satisfy that $\tau(w^i_s, w^i_s) = 4w^i_t$ $(\{s,t\} = \{1,2\})$ for each $i = 1, \ldots, 6$. Hence we have $v^*_{d_{4i}} = w^i_1$, and $\bar{v}^*_{d_{4i}} = w^i_2$ for each i.

By considering the action of C, we have in general,

$$\begin{aligned} v^*_x &= \tfrac{1}{4}(v_x + v_{x'} + v_{x''} + v_{x'''}), \\ \bar{v}^*_x &= \tfrac{1}{4}(\bar{v}_x + \bar{v}_{x'} + \bar{v}_{x''} + \bar{v}_{x'''}), \end{aligned} \tag{5.7.1}$$

for each $x \in C_D(s)$ with $x^Q = \{x, x', x'', x'''\}$.

Since $LQ\langle\zeta\rangle/Q\langle\zeta\rangle$ consists of mutually commuting elements, we may assume

$$\begin{aligned} v^*_{d_{4i}} &= \tfrac{1}{4}\Big(\sum_{j \neq i} u_{jj} - u_{ii}\Big), \\ \bar{v}^*_{d_{4i}} &= \tfrac{1}{4}\Big(\sum_{j \neq i} \bar{u}_{jj} - \bar{u}_{ii}\Big). \end{aligned} \qquad (1 \leqq i \leqq 6) \tag{5.7.2}$$

By using (5.7.1) and (5.7.2), we have

$$\begin{aligned} u_{ii} &= \tfrac{1}{8}\Big(\sum_{d_k \notin T_i} v_k - 3\sum_{d_h \in T_i} v_h\Big), \\ \bar{u}_{ii} &= \tfrac{1}{8}\Big(\sum_{d_k \notin T_i} \bar{v}_k - 3\sum_{d_h \in T_i} \bar{v}_h\Big). \end{aligned} \qquad (1 \leqq i \leqq 6) \tag{5.7.3}$$

Next we consider the vectors u_{ij}, $\bar{u}_{ij}$ and $v_{T_i \cup T_j}$, $\bar{v}_{T_i \cup T_j}$. The space $U_{T_i \cup T_j} = \langle v_{T_i \cup T_j} \rangle$ is invariant under the action of $N_C(\{d^Q_{4i}, d^Q_{4j}\})$, and $d_{4i}d_{4m}(v_{T_i \cup T_j}) = -v_{T_i \cup T_j}$ for $m \neq i, j$. On the other hand, we may identify $W/Q\langle\zeta\rangle$ with $W^{(4)}$

since they both are the normalizers of $\{d_{4i}Q\langle\zeta\rangle/Q\langle\zeta\rangle \mid 1\leqq i\leqq 6\}$ by (5.7.2). Now $N_{W^{(4)}}(\{d_{4i}Q\langle\zeta\rangle/Q\langle\zeta\rangle, d_{4j}Q\langle\zeta\rangle/Q\langle\zeta\rangle\})$-invariant 1-dimensional subspaces are $\langle u_{ii}-u_{jj}\rangle$ or $\langle u_{ij}\rangle$ or subspaces of $\langle u_{ii}+u_{jj}, \sum_{k\neq i,j} u_{kk}\rangle$. Since $d_{4i}d_{4m}$ acts trivially on $\langle u_{11}, \ldots, u_{66}\rangle$, we have $U_{T_i\cup T_j}=\langle u_{ij}\rangle$. Hence we can set

$$v_{T_i\cup T_j}=u_{ij}, \quad \bar{v}_{T_i\cup T_j}=\bar{u}_{ij} \quad (1\leqq i\neq j\leqq 6), \tag{5.7.4}$$

since $d_1(v_{T_1\cup T_i})=-\bar{v}_{T_1\cup T_i}$, and $d_1(u_{1i})=-\bar{u}_{1i}$.

(5.8) Product formulas for v_i, $\bar{v}_i$, v_X, and $\bar{v}_X$

By using (5.7) and Theorem $1^{(4)}$, we can get several formulas.

By $\tau(u_{ij}, u_{ij})$ in Theorem $1^{(4)}$ and (5.7.3), we have

$$\tau(v_{T_i\cup T_j}, v_{T_i\cup T_j})=\tfrac{1}{8}(-5\sum_{d_k\in T_i\cup T_j}\bar{v}_k+3\sum_{d_h\notin T_i\cup T_j}\bar{v}_h),$$

and thus by (5.3.1), we have

$$\tau(v_X, v_X)=\tfrac{1}{8}(-5\sum_{d_k\in X}\bar{v}_k+3\sum_{d_h\notin X}\bar{v}_h) \quad (X\in\mathcal{O}). \tag{5.8.1}$$

By $\tau(u_{ii}, u_{ij})=3\bar{u}_{ij}$, $\tau(u_{ii}, u_{jk})=-\bar{u}_{jk}$ (Theorem $1^{(4)}$), and (5.7.3), we have $\tau(\frac{1}{4}(v_{4i-3}+v_{4i-2}+v_{4i-1}+v_{4i}), u_{ij})=-\bar{u}_{ij}$, and $\tau(\frac{1}{4}(v_{4i-3}+v_{4i-2}+v_{4i-1}+v_{4i}), u_{jk})=\bar{u}_{jk}$. Hence we have

$$\tau(v_d, v_{T_i\cup T_j})=\tau(v_d, u_{ij})=\begin{cases}-\bar{u}_{ij} & (\text{if } d\in T_i\cup T_j),\\ \bar{u}_{ij} & (\text{if } d\notin T_i\cup T_j),\end{cases}$$

and thus in general:

$$\tau(v_i, v_X)=\begin{cases}-\bar{v}_X & (\text{if } d_i\in X),\\ \bar{v}_X & (\text{if } d_i\notin X).\end{cases} \tag{5.8.2}$$

By $\tau(u_{ij}, u_{jk})=-\bar{u}_{ik}$, $\tau(u_{ij}, u_{kh})=-\theta\bar{u}_{st}$ (Theorem $1^{(4)}$), we have

$$\tau(v_{T_i\cup T_j}, v_{T_j\cup T_k})=-\bar{v}_{T_i\cup T_k}, \quad \tau(v_{T_i\cup T_j}, v_{T_k\cup T_h})=-\theta\bar{v}_{T_s\cup T_t}.$$

Hence for any two octads X, Y with $|X\cap Y|=0$ or 4, we have

$$\tau(v_X, v_Y)=\begin{cases}\pm\bar{v}_{(X\cup Y)-(X\cap Y)} & (\text{if } |X\cap Y|=4)\\ \pm\theta\bar{v}_{L-(X\cup Y)} & (\text{if } |X\cap Y|=0).\end{cases} \tag{5.8.3}$$

For the ambiguous signs in the above, we shall give a little more comments in Lemma 5.8.1.

It remains to determine $\tau(v_X, v_Y)$ for $X, Y\in\mathcal{O}$ with $|X\cap Y|=2$. Let $\tau(v_X, v_Y)=\sum_{i=1}^{24}\alpha_i\bar{v}_i+\sum_{Z\in\mathcal{O}}\alpha_Z\bar{v}_Z$, and a, b, c be any elements of L such that $a, b\in X\not\ni c$, and $a, c\notin Y\ni b$. Then $ab(\tau(v_X, v_Y))=\tau(ab(v_X), ab(v_Y))=\tau(v_X, -v_Y)=-\tau(v_X, v_Y)$. Similarly $ac(\tau(v_X, v_Y))=-\tau(v_X, v_Y)$. Hence we have $\alpha_i=0$ for each i, and $a_Z=0$ unless Z contains one of a, b, and one of a, c. The arbitrariness of a, b, c shows

$\alpha_Z=0$ for each Z. Hence we have:

$$\tau(v_X, v_Y)=0 \quad (\text{if } |X\cap Y|=2). \tag{5.8.4}$$

Now the product formulas for v_i, v_X are obtained except the ambiguity of signs in (5.8.3). The formulas for $\bar{v}_i, \bar{v}_X$ are given by changing the roles of v_i, v_X and $\bar{v}_i, \bar{v}_X$. For signs in (5.8.3), we know the following:

Lemma 5.8.1. *Let X and Y be distinct octads.*

(1) *If $|X\cap Y|=4$ and $\tau(v_X, v_Y)=\varepsilon\bar{v}_Z$ $(Z=(X\cup Y)-(X\cap Y))$, then $\tau(\bar{v}_X, \bar{v}_Y)=\varepsilon v_Z$, and $\tau(v_X, v_Z)=\varepsilon\bar{v}_Y$.*

(2) *If $|X\cap Y|=0$ and $\tau(v_X, v_Y)=\varepsilon\theta\bar{v}_Z$ $(Z=L-(X\cup Y))$, then $\tau(\bar{v}_X, \bar{v}_Y)=-\varepsilon\theta v_Z$, and $\tau(v_X, v_Z)=\varepsilon\theta\bar{v}_Y$.*

Proof. If $|X\cap Y|=0$ or 4, then there exist some element g in M and some i, j, k, h of $\{1, \ldots, 6\}$ such that $X=(T_i\cup T_j)^g$ and $Y=(T_k\cup T_h)^g$. Let $g(u_{ij})=\varepsilon_1 v_X$, and $g(u_{kh})=\varepsilon_2 v_Y$. If $|X\cap Y|=4$, then we may assume $j=h$. Then $\tau(v_X, v_Y)=g(\tau(\varepsilon_1 u_{ij}, \varepsilon_2 u_{jh}))=g(-\varepsilon_1\varepsilon_2\bar{u}_{ik})=\varepsilon\bar{v}_Z$. Let a be an element contained in $X\cap Y$. Then a is not contained in Z, and we have $\tau(\bar{v}_X, \bar{v}_Y)=a(\tau(v_X, v_Y))=a(\varepsilon\bar{v}_Z)=\varepsilon v_Z$, and moreover $\tau(v_X, v_Z)=\tau(g(\varepsilon_1 u_{ij}),\ g(-\varepsilon_1\varepsilon_2\varepsilon u_{ik}))=g(\varepsilon_2\varepsilon\bar{u}_{jk})=\varepsilon\bar{v}_Z$. Hence we have (1). The assertion (2) is obtained similarly.

(5.9) The C-invariant space B_1

We set $B_1=U_5\oplus\bar{U}_5$, which is the (-1)-eigenspace of s. In order to formulate $\tau(B_*, B_1)$, we introduce new coordinates in B_1. We set $R=R^{(4)}$, and let V be a 2^6-dimensional space with the basis $v(\delta_1, \delta_2, \ldots, \delta_6)$ with $\delta_i\in\{0, 1\}$. Then as a linear space, the space B_1 can be identified with the space $R\otimes V$, with $B_1\cap U=R_1\otimes V$ and $B_1\cap\bar{U}=R_2\otimes V$. Here notice that the dimensions of these two spaces are equal to each other ($384\times 2=12\times 2^6$; see (5.4)).

We set $\Delta=\{(\delta_1, \delta_2, \ldots, \delta_6)\mid \delta_i\in\{0, 1\}\}$, the set of indices for basis of V. We will frequently denote by (δ_l) a member $(\delta_1, \delta_2, \ldots, \delta_6)$ of Δ. Set $U_i((\delta_l))=\langle r_i\otimes v((\delta_l))\rangle$ and $\bar{U}_i((\delta_l))=\langle\bar{r}_i\otimes v((\delta_l))\rangle$ for $(\delta_l)\in\Delta$, and $i=1, \ldots, 6$. Here the vectors r_i, $\bar{r}_i$ are the fixed basis of R as in Sect. 3. Moreover we relabel the elements of L as follows:

$$a^{(i)}_{00}=a_{4i-3},\quad a^{(i)}_{01}=a_{4i-2},\quad a^{(i)}_{10}=a_{4i-1},\quad a^{(i)}_{11}=a_{4i},$$

for $i=1, \ldots, 6$.

Let X be an octad with $v_X\in B_1$. We give a correspondence between $U_X=\langle v_X\rangle$ (resp. $\bar{U}_X=\langle\bar{v}_X\rangle$) and $U_i((\delta_l))$ (resp. $\bar{U}_i((\delta_l))$).

We set

$$X\# T_i=\begin{cases} X\cap T_i & (\text{if } |X\cap T_i|=1)\\ T_i-(X\cap T_i) & (\text{if } |X\cap T_i|=3).\end{cases} \tag{5.9.1}$$

Here by the definition of $U_1((5.4))$, there exists just one k with $|X\cap T_k|=3$, and then $|X\cap T_h|=1$ for each $h\neq k$. Since each T_i is a 4-elements subset of L, the set $X\# T_i$ always contains only one element. We define $\delta(X)=(\delta_1, \delta_2, \ldots, \delta_6)$, a member of Δ, by $X\# T_i=\{a^{(i)}_{\delta_{2i-1}\delta_{2i}}\}$ for $i=1, 2, 3$. Here notice that the octad X

is uniquely determined by $\delta(X)$ and k with $|X\cap T_k|=3$ (see the following Lemma 5.9.1). Then we identify B_1 with $R\otimes V$ by the following:

$$U_X=\langle v_X\rangle=U_k(\delta(X)),\qquad \bar{U}_X=\langle \bar{v}_X\rangle=\bar{U}_k(\delta(X)), \tag{5.9.2}$$

if $|X\cap T_k|=3$.

Lemma 5.9.1. *Let X, Y be octads with v_X, $v_Y\in B_1$, and $\delta(X)=\delta(Y)$. Then we have $X\#T_i=Y\#T_i$ for each $i=1,\ldots,6$. Moreover if $|X\cap T_k|=|Y\cap T_k|=3$ for some k, then we have $X=Y$.*

Proof. Let $U_X=U_k(\delta(X))$ and $U_Y=U_h(\delta(Y))$. If $k\neq 1$, we set $X'=X\cup(T_1\cup T_k)-X\cap(T_1\cup T_k)$, and if $k=1$, we set $X'=X$. Then X' is an octad with $|X'\cap T_1|=3$, and $X\#T_i=X'\#T_i$ for each i. In particular $\delta(X)=\delta(X')$. Similarly we define an octad Y' by using Y, h instead of X, k. Since $\delta(X')=\delta(X)=\delta(Y)=\delta(Y')$, we have $X'\#T_i=Y'\#T_i$ for $i=1,2,3$. Hence $X'\cap Y'$ contains five elements in $T_1\cup T_2\cup T_3$, and thus $X'=Y'$. Hence we have $X\#T_i=X'\#T_i=Y'\#T_i=Y\#T_i$ for all i, as required.

The rest of the statement is easily verified by the definition (5.9.1).

Before considering $\tau(B_*, B_1)$, we need an observation about the action of C on B_1.

Let X be an octad with $\delta(X)=(0,0,0,0,0,0)$. We write (0) instead of (0, 0, 0, 0, 0, 0). Since $T_i^\zeta=T_i$ for all i, we have $|X\cap T_i|=|X^\zeta\cap T_i|$ for all i. Since $C_L(\zeta)=\{a_1,a_5,\ldots,a_{21}\}=\{a_{00}^{(i)}\mid 1\leqq i\leqq 6\}$, we have $\delta(X)=\delta(X^\zeta)$. These together with Lemma 5.9.1 imply $X=X^\zeta$. In particular, we have $X\#T_i=\{a_{00}^{(i)}\}$ for each i. Since $C_C(\zeta)\cap N_G(L)$ preserves $C_L(\zeta)$, this group acts on the set $\{U_k(0)\mid 1\leqq k\leqq 6\}$.

Lemma 5.9.2. (1) *C contains a subgroup H such that:*

(i) *H acts transitively on the set*

$$\{(u_{ij}, U_i(0))\mid 1\leqq i\neq j\leqq 6\},$$

(ii) *for any octad X with $\delta(X)=(0)$, no elements of H exchange the sign of v_X.*

(2) *For each i, the group Q acts transitively on the set $\{U_i(\delta_l)\mid(\delta_l)\in\Delta\}$, and the stabilizer of $U_i(0)$ in Q is $Q\cap\langle L\rangle$.*

Proof. (1) Under the notation in Sect. 3, the group $\langle t_{ij}\mid 1\leqq i\neq j\leqq 6\rangle'$ $(=H_0$, say), a subgroup of C_* isomorphic to A_6, permutes transitively the vectors u_{ij}, $1\leqq i\neq j\leqq 6$. Moreover H_0 permutes 4-transitively the $D(6)$-vectors $\sum_{j\neq i}u_{ii}-u_{jj}$, $1\leqq j\leqq 6$.

Since $C_*\simeq C/Q\langle\zeta\rangle$, we may assume H_0 is a subgroup of $C/Q\langle\zeta\rangle$. Moreover since $C_C(\zeta)Q=C$ and $C_C(\zeta)\cap Q=\langle s\rangle$, we have $C_C(\zeta)/\langle\zeta,s\rangle\simeq C/Q\langle\zeta\rangle$. Hence we may also assume H_0 is a subgroup of $C_C(\zeta)/\langle\zeta,s\rangle$. Let H_1 be the inverse image of H_0 in $C_C(\zeta)$. Then H_1 permutes transitively the vectors $u_{ij}=v_{T_i\cup T_j}$, and 4-transitively the vectors $v_{d_1}^*$, $v_{d_5}^*$, $\ldots$, $v_{d_{21}}^*$, and so the sets $T_1,\ldots,T_6$. Now it is easily seen that H_1 satisfies the condition (i).

We show that H_1 has a complement H of $\langle s\rangle$. Suppose false. Then the order of some preimage t of $t_{12}t_{34}$ must be 4. Since $t^2=s$ and t preserves the

space $U_5(0)$, we have $t(r_5 \otimes v(0)) = \pm\sqrt{-1}(r_5 \otimes v(0))$. This contradicts the fact that t permutes the vectors $\bar{v}_i$, $1 \leqq i \leqq 24$, and the formula (5.8.1). Hence the assertion holds, and clearly H satisfies the condition (i).

The stabilizer of $U_1(0)$ in H is isomorphic to A_4 modulo the center $\langle\zeta\rangle$. Since it contains no subgroup of index 2, H satisfies the condition (ii).

(2) Suppose $q \in Q$ preserves the space $U_1(0)$. Let X be an octad with $U_X = U_1(0)$. Then $X^q = X$, and $X \# T_i = C_{T_i}(\zeta)$ for each i. Hence q commutes with the elements of $C_L(\zeta)$. Moreover since q is a 2-element and $|T_i| = 4$, q commutes with more than two elements of each T_i, that is, more than 12 elements of L. This implies $q \in \langle L \rangle \cap Q$. Since $\langle L \rangle \cap Q$ is an elementary abelian subgroup of Q, its order is less than 2^7. Clearly Q acts on the set $\{U_i(\delta_l) \mid (\delta_l) \in \Delta\}$, and its cardinality equals $2^6 = |Q|/2^7$. Hence the stabilizer of $U_1(0)$ is $\langle L \rangle \cap Q$ of order 2^7, and the statement (2) holds.

Let X be an octad with $U_X = U_1(0)$, and set $r_1 \otimes v(0) = v_X$, $\bar{r}_1 \otimes v(0) = \bar{v}_X$. Let $Y = (X \cup (T_1 \cup T_2)) - (X \cap (T_1 \cup T_2))$. Then Y is an octad and $U_Y = U_2(0)$. Hence $\tau(u_{12}, r_1 \otimes v(0)) \in U_2(0)$.

For each $i \neq 1$, we set $r_i \otimes v(0) = t(r_1 \otimes v(0))$ and $\bar{r}_i \otimes v(0) = t(\bar{r}_1 \otimes v(0))$, where t is an element of H satisfying $t(U_1(0)) = U_i(0)$. By Lemma 5.9.2(1), it is well-defined. Then we have $\tau(u_{12}, r_1 \otimes v(0)) = \varepsilon(\bar{r}_2 \otimes v(0))$ for some $\varepsilon = \pm 1$, and then $\tau(\bar{u}_{12}, \bar{r}_1 \otimes v(0)) = \varepsilon(\bar{r}_2 \otimes v(0))$ by Lemma 5.8.1. By Lemma 5.8.2, we have

$$\begin{aligned} &\tau(u_{ij}, r_i \otimes v(0)) = \varepsilon(\bar{r}_j \otimes v(0)), \\ &\tau(\bar{u}_{ij}, \bar{r}_i \otimes v(0)) = \varepsilon(r_j \otimes v(0)), \end{aligned} \tag{5.9.3}$$

for any $i \neq j$.

For each $(\delta_l) \in \Delta$ with $(\delta_l) \neq (0)$, we fix an element $q(\delta_l)$ of Q with $q(\delta_l)(U_1(0)) = U_1(\delta_l)$. Here $q(\delta_l)$ does exist by Lemma 5.9.2(2). Then it is easily seen that $q(\delta_l)(U_i(0)) = U_i(\delta_l)$ for each i. We set $r_i \otimes v(\delta_l) = q(\delta_l)(r_i \otimes v(0))$ and $\bar{r}_i \otimes v(\delta_l) = q(\delta_l)(\bar{r}_i \otimes v(0))$. Then we have

$$\begin{aligned} &\tau(u_{ij}, r_i \otimes v(\delta_l)) = \varepsilon(\bar{r}_j \otimes v(\delta_l)), \\ &\tau(\bar{u}_{ij}, \bar{r}_i \otimes v(\delta_l)) = \varepsilon(r_j \otimes v(\delta_l)), \end{aligned} \tag{5.9.4}$$

for any $i \neq j$.

We determine ε in the next section.

(5.10) $\tau(B_*, B_1)$

We determine the formulas about $\tau(B_*, B_1)$. We identify the space B_* with a subspace of $\mathrm{End}(R)$ as in Sect. 3.

For each b of B_*, we denote by τ_b the linear map $B_1 \to B_1$ defined by $\tau_b(x) = \tau(b, x)$ for $x \in B_1$. Let X be an octad with $|X \cap T_k| = 3$. Then $U_X = U_k(\delta(X))$.

First we consider $\tau_{u_{ii}}$. By using (5.7.3), (5.8.2), we have $\tau(u_{ii}, v_X) = 2\bar{v}_X$ if $i = k$, $-2\bar{v}_X$ if $i \neq k$. By easy examinations, we can write this as follows:

$$\tau(u_{ii}, r_j \otimes v(\delta_l)) = u_{ii}(r_j) \otimes v(\delta_l),$$

for any i, j and $(\delta_l) \in \Delta$. Here in the right-hand side, u_{ii} is regarded as an element of $\mathrm{End}(R)$. Hence $\tau_{u_{ii}} = u_{ii} \otimes I_V$, where we denote by I_V the identity on V. Similarly $\tau_{\bar{u}_{ii}} = \bar{u}_{ii} \otimes I_V$.

Next we consider $\tau_{u_{ij}}$, $i \neq j$. By using (5.9.3), (5.9.4), we have $\tau_{u_{ij}} = \varepsilon u_{ij} \otimes I_V$. We show $\varepsilon = 1$.

Let $b \in B_*$, $x \in B_1$, and $g \in C$. Then we have

$$\tau_{g(b)}(x) = \tau(g(b), x) = g(\tau(b, g^{-1}(x))) = (g|_{B_1})\, \tau_b (g|_{B_1})^{-1}(x). \tag{5.10.1}$$

Set $v_0^* = \frac{1}{8}\left(\sum_{j \neq 5,6} u_{jj} - (u_{55} + u_{66}) - \sum_{\substack{i,j \neq 5,6 \\ i<j}} 2u_{ij} + 2\theta u_{56} \right)$, and $v_1^* = v_{d_1}^* = \frac{1}{4}\left(\sum_{i \neq 1} u_{ii} - u_{11} \right)$. Since v_0^* and v_1^* are $D(6)$-vectors in B_*, there exists an element g of C such that $g(v_0^*) = v_1^*$. Hence the rank of $\tau_{v_0^*}$ coincides with that of $\tau_{v_1^*}$, and the latter equals 6×2^6. Suppose $\tau_{u_{ij}} = -u_{ij} \otimes I_V$. Then by direct calculations, we have that the rank of $\tau_{v_0^*}$ is 5×2^6, a contradiction. Hence we have $\tau_{u_{ij}} = u_{ij} \otimes I_V$. Similarly we have $\tau_{\bar{u}_{ij}} = \bar{u}_{ij} \otimes I_V$.

Now we have determined $\tau(B_*, B_1)$, that is:

$$\tau(b, r \otimes v) = b(r) \otimes v, \tag{5.10.2}$$

for any $b \in B_*$, $r \in R$, and $v \in V$.

(5.11) The action of C on B_1

Lemma 5.11.1. *Let g be an element of* $\mathrm{Aut}(B, \tau)$ *such that g preserves B_* and B_1, respectively. Then there exist $g_R \in GL(R)$ and $g_V \in GL(V)$ such that $g|_{B_1} = g_R \otimes g_V$.*

Proof. It is easily seen that the algebra generated by all τ_b for $b \in B_*$ is $\mathrm{End}(R) \otimes I_V$, which is stable by g. Let X be the representational matrix of g on B_1, and Σ be the set of all (12×12)-submatrices of X which give the restrictions of g to $R \otimes \langle v(\delta_l) \rangle$ followed by the projections $B_1 \to R \otimes \langle v(\delta'_l) \rangle$ for $(\delta_l), (\delta'_l) \in \Delta$.

For any $A \in \mathrm{End}(R)$, there exists some $A' \in \mathrm{End}(R)$ such that $X(A \otimes I_V) X^{-1} = A' \otimes I_V$. This means $YA = A'Y$ for each $Y \in \Sigma$. Now that A is arbitrary, it is easily seen that Y is singular if and only if $Y = O$. Moreover if Y_1, Y_2 of Σ are non-singular, then Y_2 is a scalar times Y_1. Hence each member of Σ is written as cY for a non-singular member Y of Σ and some constant c. Since X is non-singular, we have $X = Y \otimes X_V$ for some $X_V \in GL(V)$.

Since $Q\langle\zeta\rangle$ acts trivially on B_*, we may take $q_R = I_R$ for any $q \in Q\langle\zeta\rangle$. Then the map $q \to q_V$ on Q gives the unique non-linear irreducible representation of Q. Moreover the map $x \to x_R$ on $C/Q\langle\zeta\rangle$ gives the projective representation induced from the embedding $C_0 \subset GL(R)$, mentioned in Sect. 3. In fact, by using (5.10.1), (5.10.2), we have $\tau_{g(b)} = \tau_{g_R b g_R^{-1}}$. Hence g_R induces an automorphism on the algebra B_*, and thus Theorem 2 implies the assertion.

Now it is easily verified that C acts irreducibly on B_1.

(5.12) The C-irreducible space B_2, and $\tau(B_2, B_1)$

Set $w = \frac{1}{4}(v_1 + v_2 - v_3 - v_4)$. Then $w \in U_3$. By (5.8.2), we have

$$\tau(w, r_i \otimes v(\delta_l)) = \begin{cases} \frac{1}{2}\bar{r}_i \otimes v(\delta_l), \\ \quad \text{(if } i=1 \text{ and } (\delta_1, \delta_2)=(0,0) \text{ or } (0,1), \\ \quad \text{or if } i \neq 1 \text{ and } (\delta_1, \delta_2)=(1,0) \text{ or } (1,1)) \\ -\frac{1}{2}\bar{r}_i \otimes v(\delta_l), \\ \quad \text{(if } i=1 \text{ and } (\delta_1, \delta_2)=(1,0) \text{ or } (1,1), \\ \quad \text{or if } i \neq 1 \text{ and } (\delta_1, \delta_2)=(0,0) \text{ or } (0,1)) \end{cases}$$

for any $(\delta_l) \in \Delta$ and i. Here notice that the right-hand side equals $v^*_{d_1}(r_i) \otimes (d_1 d_2)_V(v(\delta_l))$. We set:

$$\begin{aligned} \varphi(v^*_x, xx') &= \tfrac{1}{4}(v_x + v_{x'} - v_{x''} - v_{x'''}), \\ \varphi(\bar{v}^*_x, xx') &= \tfrac{1}{4}(\bar{v}_x + \bar{v}_{x'} - \bar{v}_{x''} - \bar{v}_{x'''}), \end{aligned} \tag{5.12.1}$$

for each $x \in C_D(s)$ with $x^Q = \{x, x', x'', x'''\}$. Then we have

$$\begin{aligned} &\varphi(\omega^i v^*_x, xx') = \omega^i \varphi(v^*_x, xx'), \\ &\varphi(v^*_x, xx's) = -\varphi(v^*_x, xx') \quad \text{and} \\ &v_x = v^*_x + \varphi(v^*_x, xx') + \varphi(v^*_x, xx'') + \varphi(v^*_x, xx'''). \end{aligned} \tag{5.12.2}$$

Now by considering the action of C, and Lemma 5.11.1, we have

$$\begin{aligned} \tau(\varphi(v^*_x, xx'), r \otimes v) &= v^*_x(r) \otimes (xx')_V(v), \\ \tau(\varphi(\bar{v}^*_x, xx'), r \otimes v) &= \bar{v}^*_x(r) \otimes (xx')_V(v). \end{aligned} \tag{5.12.3}$$

We set $B_2 = \langle \varphi(v^*_x, xx'), \ \varphi(\bar{v}^*_x, xx') \mid x \in C_D(s), \ x' \in x^Q \rangle$. The spaces $\langle \varphi(v^*_x, xx') \rangle$ are stable by Q, respectively, and they afford mutually distinct characters of $Q/\langle s \rangle$. Since $G' \cap C$ permutes them transitively, B_2 is C-irreducible, and furthermore its dimension equals $\frac{1}{2}|Q_2| \times 2 = 756$. Hence $B_2 = U_3 \otimes \bar{U}_3 \oplus U_4 \oplus \bar{U}_4$.

(5.13) $\tau(B_*, B_2)$ and $\tau(B_2, B_2)$

Let $x \in C_D(s)$, and set $x^Q = \{x, x', x'', x'''\}$.

First we consider $\tau(B_*, B_2)$. By direct calculations using (5.7.2)–(5.7.4), (5.12.1), and Theorem $1^{(4)}$, we have

$$\tau(u_{ij}, \varphi(v^*_{d_1}, d_1 d_2)) = \begin{cases} -2\varphi(\bar{v}^*_{d_1}, d_1 d_2) & (\text{if } i=j=1) \\ 2\varphi(\bar{v}^*_{d_1}, d_1 d_2) & (\text{if } i=j \neq 1) \\ 0 & (\text{if } i \neq j). \end{cases}$$

We can summarize the above as follows:

$$\tau(b, \varphi(v^*_{d_1}, d_1 d_2)) = 2\,\mathrm{Tr}(b\, \bar{v}^*_{d_1})\, \varphi(\bar{v}^*_{d_1}, d_1 d_2),$$

for any $b \in B_*$. Hence in general,

$$\tau(b, \varphi(v^*_x, xx')) = 2\,\mathrm{Tr}(b \bar{v}^*_x)\, \varphi(\bar{v}^*_x, xx'), \tag{5.13.1}$$

for any $x \in C_D(s)$, $x' \in x^Q$, and $b \in B_*$.

Next we consider $\tau(B_2, B_2)$. By direct calculations using (5.7.1), (5.12.1) together with Theorem A, we have

$$\tau(\varphi(v_x^*, xx'), \varphi(v_x^*, xx')) = 2\bar{v}_x^*, \tag{5.13.2}$$

$$\tau(\varphi(v_x^*, xx'), \varphi(v_x^*, xx'')) = -2\varphi(\bar{v}_x^*, x'x'') = 2\varphi(\bar{v}_x^*, xx'''). \tag{5.13.3}$$

Let $y \in C_D(s) - x^Q$. Assume $|xy| = 2$ or 6. Then we have

$$\tau(\varphi(v_x^*, xx'), \varphi(v_y^*, yy')) = 0, \tag{5.13.4}$$

for any $y' \in y^Q$.

Next assume $|xy| = 3$, and set $y^Q = \{y, y', y'', y'''\}$. Then we have

$$\begin{aligned}
&\tau(\varphi(v_x^*, xx'), \varphi(v_y^*, yy')) \\
&\quad = \tfrac{1}{16}(\bar{v}_{x^y} + \bar{v}_{x^{y'}} - \bar{v}_{xy''} - \bar{v}_{x^{y'''}} + \bar{v}_{x'y} + \bar{v}_{x'y'} - \bar{v}_{x'y''} - \bar{v}_{x'y'''} \\
&\qquad + \bar{v}_{x''y} + \bar{v}_{x''y'} - \bar{v}_{x''y''} - \bar{v}_{x''y'''} + \bar{v}_{x'''y} + \bar{v}_{x'''y'} - \bar{v}_{x'''y''} - \bar{v}_{x'''y'''}).
\end{aligned}$$

Since $xx' \in Q$, we have $y^{xx'} \in y^Q$. If $y' \neq y^{xx'}$, then we may assume $y'' = y^{xx'}$. Then we have $x^y = x'^{y''}$, $x^{y'} = (x^y)^{y'} = (x'^{y''})^{y'} = x'^{y'''}$, and so on. After all we have

$$\tau(\varphi(v_x^*, xx'), \varphi(v_y^*, yy')) = 0 \quad (\text{if } y' \neq y^{xx'}). \tag{5.13.5}$$

If $y' = y^{xx'}$, then $x^y = x'^{y'}$, $x^{y''} = x'^{y'''}$. Moreover we may assume $y^{xx''} = y''$, and then we have

$$\tau(\varphi(v_x^*, xx'), \varphi(v_y^*, yy')) = \varphi(\bar{v}_{x^y}^*, x^y x^{y'}) \quad (\text{if } y' = y^{xx'}). \tag{5.13.6}$$

We do not describe $\tau(B_1, B_1)$ here. We consider it in Sect. 8.

(5.14) Summary

$$\begin{aligned}
B &= \langle v_i, \bar{v}_i, v_X, \bar{v}_X \mid 1 \leqq i \leqq 24,\ X \in \mathcal{O} \rangle && [(5.3)] \\
&= B_* \oplus B_2 \oplus B_1 && [(5.4)] \\
B_* &= B^{(4)} \subset \mathrm{End}(R) && [\text{Sect. } 3] \\
&= \langle u_{ij}, \bar{u}_{ij} \mid 1 \leqq i, j \leqq 6 \rangle && [(5.6), (5.7), (5.8)] \\
B_2 &= \langle \varphi(v_x^*, xx'), \varphi(\bar{v}_x^*, xx') \mid x \in C_D(s),\ x' \in x^Q \rangle && [(5.12)] \\
B_1 &= R \otimes V && [(5.9)] \\
&= \langle r_i, \bar{r}_i \mid 1 \leqq i \leqq 6 \rangle \otimes \langle v(\delta_1, \ldots, \delta_6) \mid \delta_l = 0, 1 \rangle
\end{aligned}$$

$$\tau(v_i, v_j) = \begin{cases} 10\bar{v}_i & (\text{if } i = j) \\ \bar{v}_i + \bar{v}_j & (\text{if } i \neq j) \end{cases}$$

$$\tau(v_i, v_X) = \begin{cases} -\bar{v}_X & (\text{if } d_i \in X) \\ \bar{v}_X & (\text{if } d_i \notin X) \end{cases}$$

$$\tau(v_X, v_Y) = \begin{cases} \frac{1}{8}(-5 \sum_{d_k \in X} \bar{v}_k + 3 \sum_{d_h \notin X} \bar{v}_h) & (\text{if } X = Y) \\ \pm \bar{v}_{(X \cup Y) - (X \cap Y)} & (\text{if } |X \cap Y| = 4) \\ \pm \theta \bar{v}_{L - (X \cup Y)} & (\text{if } |X \cap Y| = 0) \\ 0 & (\text{if } |X \cap Y| = 2) \end{cases}$$

$$\tau(B_*, B_*) = \tau^{(4)}(B^{(4)}, B^{(4)}) \qquad \text{[Sect. 3]}$$

$$\tau(B_*, B_2): \ \tau(b, \varphi(v_x^*, xx')) = 2\,\mathrm{Tr}(b\bar{v}_x^*)\,\varphi(\bar{v}_x^*, xx') \qquad \text{[(5.13)]}$$

$$\tau(B_*, B_1): \ \tau(b, r\otimes v) = b(r)\otimes v \qquad \text{[(5.10)]}$$

$$\tau(B_2, B_1): \ \tau(\varphi(v_x^*, xx'), r\otimes v) = v_x^*(r)\otimes(xx')_V(v) \qquad \text{[(5.12)]}$$

$$\tau(B_2, B_2): \ \tau(\varphi(v_x^*, xx'), \varphi(v_x^*, xx')) = 2\bar{v}_x^* \qquad \text{[(5.13)]}$$

$$\tau(\varphi(v_x^*, xx'), \varphi(v_x^*, xx'')) = -2\varphi(\bar{v}_x^*, x'x'')$$

$$\tau(\varphi(v_x^*, xx'), \varphi(v_x^*, yy'))$$

$$= \begin{cases} \varphi(\bar{v}_{x^y}^*, x^y x^{y'}) & (\text{if } |xy| = 3 \text{ and } y' = x^{yy'}) \\ 0 & (\text{if } |xy| = 2, 6, \text{ or } |xy| = 3 \text{ and } y' \neq x^{yy'}). \end{cases}$$

6. The automorphism group Aut$(B(24), \tau(24))$

In this section, we determine the full automorphism group of the algebra $(B, \tau) = (B(24), \tau(24))$. We use the same notation as in Sect. 5. We will prove the following:

Theorem 5. $\mathrm{Aut}(B, \tau) = G$.

We set $\Gamma = \mathrm{Aut}(B, \tau)$. Our first purpose is to show $C_\Gamma(s) = C$. If it is established, then we get the finiteness of Γ by Theorem B.

(6.1) *$C_\Gamma(s)$ preserves B_*, B_2, and B_1 respectively.*

Proof. Since B_1 is the (-1)-eigenspace of s, it is $C_\Gamma(s)$-invariant. Hence $B_* \oplus B_2$, is $C_\Gamma(s)$-invariant, since C preserves this subspace.

For $b \in B_* \oplus B_2$, we define the linear maps $\tau_b: B \to B$ and $\tau_b': B_1 \to B_1$ by $\tau_b(x) = \tau(b, x)$ and $\tau_b'(x) = \tau(b, y)$ for $x \in B$ and $y \in B_1$ respectively. Moreover we define the $C_\Gamma(s)$-invariant bilinear symmetric forms β, β' on $B_* \oplus B_2$ by $\beta(a, b) = \mathrm{Tr}(\tau_a \tau_b)$ and $\beta'(a, b) = \mathrm{Tr}(\tau_a' \tau_b')$ for $a, b \in B_* \oplus B_2$, respectively.

The group C acts irreducibly on B_*, B_2 (and B_1), respectively. Moreover these representations are self-dual, since the representation of G on B is self-dual ([15]), and since the dimensions of B_*, B_2, B_1 are mutually distinct. Hence the restrictions $\beta|_{B_*}$, $\beta|_{B_2}$ are some scalar multiple of $\beta'|_{B^*}$, $\beta'|_{B_2}$ respectively, if they are non-trivial.

First we consider β'. Let $a \in B_* \cap U$, $b \in B_* \cap \bar{U}$. Then under the identification $B_* = B(6) \subset \mathrm{End}(R)$ ((5.6), (3.5)), we can write $a = J_1 \otimes X$, $b = J_2 \otimes Y$ for some symmetric 6×6-matrices X, Y. Hence by (5.10.2), we have $\tau_a' = (J_1 \otimes X) \otimes I_V$, $\tau_b' = (J_2 \otimes Y) \otimes I_V$. Hence we have $\beta'(a, b) = 2^6\,\mathrm{Tr}((J_1 \otimes X)(J_2 \otimes Y)) = 2^6\,\mathrm{Tr}(XY)$. In particular, $\beta'(u_{11}, u_{11}) = 2^8$ by (3.5).

Let $\varphi = \varphi(v_x^*, q)$, $\varphi' = \varphi(\bar{v}_y^*, r) \in B_2$. Then $\tau_\varphi' = v_x^* \otimes q_V$, $\tau_{\varphi'}' = \bar{v}_y^* \otimes r_V$ by (5.12.3). Hence we have $\beta'(\varphi, \varphi') = \mathrm{Tr}(v_x^* \bar{v}_y^*)\,\mathrm{Tr}(q_V r_V)$. Here $\mathrm{Tr}(q_V r_V)$ is the value of the non-linear irreducible character of Q. Hence in particular,

$$\beta'(\varphi(v_{d_1}^*, d_1 d_2), \varphi(\bar{v}_{d_1}^*, d_1 d_2)) = \mathrm{Tr}(v_{d_1}^* \bar{v}_{d_1}^*)\,\mathrm{Tr}(I_V) = \tfrac{3}{2} \times 2^6 = 96,$$

by (5.7.2) and (3.5).

Concerning β, we have $\beta(u_{11}, \bar{u}_{11})=784$, and $\beta(\varphi(v^*_{d_1}, d_1 d_2), \varphi(\bar{v}^*_{d_1}, d_1 d_2))=196$, by straightforward calculations using several formulas in Sect. 5. We leave details to the reader.

By the above, we have $\beta|_{B_*}=\frac{784}{256}\beta'|_{B_*}$, and $\beta|_{B_2}=\frac{196}{96}\beta'|_{B_2}$. Since $\frac{784}{256}\neq\frac{196}{96}$, we have $B_*=(B_*\oplus B_2)\cap \mathrm{Rad}\left(\beta-\frac{784}{256}\beta'\right)$, $B_2=(B_*\oplus B_2)\cap \mathrm{Rad}\left(\beta-\frac{196}{96}\beta'\right)$. Hence B_*, B_2 are $C_\Gamma(s)$-invariant respectively.

(6.2) We let $\alpha\in C_\Gamma(s)$. We shall show $\alpha\in C$ in (6.7).

By Lemma 5.11.1, there exist some $\alpha_R\in GL(R)$ and $\alpha_V\in GL(V)$ such that $\alpha|_{B_1}=\alpha_R\otimes\alpha_V$. Here we may assume $\alpha_R=I_R$. In fact, $(B_*, \tau|_{B_1})\simeq(B(6), \tau(6))$ and $\mathrm{Aut}(B(6), \tau(6))=C_*\simeq C/Q\langle\zeta\rangle$. Hence by multiplying some element of C, we may assume α acts trivially on B_*. Then by (5.10.1), $b\otimes I_V=\tau'_b=\tau'_{\alpha(b)}=\alpha_R b(\alpha_R)^{-1}\otimes I_V$ for each $b\in B_*$. Since B_* as endomorphisms on R are irreducible, we have $\alpha_R=cI_R$ for some scalar c. Now by resetting α_V to $c\alpha_V$, we have $\alpha_R=I_R$, as required.

(6.3) *For each $x\in C_D(s)$, we set:*

$$B_2^+(x)=\langle\varphi(v^*_x, xx')\,|\,x'\in x^Q\rangle,$$
$$B_2^-(x)=\langle\varphi(\bar{v}^*_x, xx')\,|\,x'\in x^Q\rangle.$$

Then the space $B_2^+(x)$, $B_2^-(x)$ are stable by α.

Proof. First notice that (5.12.2) implies $B_2^+(x)=B_2^+(y)$, $B_2^-(x)=B_2^-(y)$ for each $y\in x^{\langle z\rangle Q}$.

For $b\in B_*$, we denote by τ''_b the linear map $B_2\to B_2$ defined by $\tau''_b(x)=\tau(b, x)$ for $x\in B_2$. By (6.2), α commutes with τ''_b for all $b\in B_*$.

Let $b, b'\in B_*$, and $u\in B_2^+(x)$, $w\in B_2^-(x)$ for $x\in C_D(s)$. By (5.13.1), we have

$$\tau''_b\,\tau''_{b'}(u)=\mathrm{Tr}(bv^*_x)\,\mathrm{Tr}(b'\bar{v}^*_x)\,u,$$
$$\tau''_b\,\tau''_{b'}(w)=\mathrm{Tr}(b\bar{v}^*_x)\,\mathrm{Tr}(b'v^*_x)\,w.$$

Hence u (resp. w) is contained in the $\mathrm{Tr}(bv^*_x)\,\mathrm{Tr}(b'\bar{v}^*_x)$-eigenspace (resp. $\mathrm{Tr}(b\bar{v}^*_x)\,\mathrm{Tr}(b'v^*_x)$-eigenspace) of $\tau''_b\tau''_{b'}$, which also contains $\alpha(u)$ (resp. $\alpha(w)$). If $y\notin x^{\langle z\rangle Q}$, then we can choose b, b' so that $\mathrm{Tr}(bv^*_x)\,\mathrm{Tr}(b'\bar{v}^*_x)\neq\mathrm{Tr}(bv^*_y)\,\mathrm{Tr}(b'\bar{v}^*_y)$. Moreover we can also choose b, b' so that $\mathrm{Tr}(bv^*_x)\,\mathrm{Tr}(b'\bar{v}^*_x)\neq\mathrm{Tr}(b\bar{v}^*_x)\,\mathrm{Tr}(b'v^*_x)$. Hence we have $\alpha(u)\in B_2^+(x)$, $\alpha(w)\in B_2^-(x)$, as required.

(6.4) *For each $x\in C_D(s)$, the set of vectors of B_2,*

$$\{\pm\varphi(v^*_x, xx')\,|\,x'\in x^Q-\{x\}\},$$

is stable by α.

Proof. It is sufficient to show this in the case of $x=d_1$.

For $m_1, m_2, m_3\in\mathbb{C}$, we set

$$\varphi(m_1, m_2, m_3)=\sum_{i=1}^{3} m_i\,\varphi(v^*_{d_1}, d_1 d_{i+1}).$$

Then $\tau_{\varphi(m_1, m_2, m_3)} = v_{d_1}^* \otimes (m_1(d_1 d_2)_V \oplus m_2(d_1 d_3)_V \oplus m_3(d_1 d_4)_V)$. We consider the rank of $\tau_{\varphi(m_1, m_2, m_3)}$. For $\varepsilon_1, \varepsilon_2 \in \{0, 1\}$, we set

$$V(\varepsilon_1, \varepsilon_2) = \langle v(\varepsilon_1, \varepsilon_2, \varepsilon_3, \ldots, \varepsilon_6) \mid \varepsilon_l \in \{0, 1\} (3 \leqq l \leqq 6) \rangle.$$

Then $V = V(0, 0) \oplus V(0, 1) \oplus V(1, 0) \oplus V(1, 1)$, and clearly

$$\begin{aligned} \tau_{\varphi(m_1, m_2, m_3)} = & -(m_1 + m_2 + m_3) I_{V(0, 0)} \oplus (-m_1 + m_2 + m_3) I_{V(0, 1)} \\ & \oplus (m_1 - m_2 + m_3) I_{V(1, 0)} \oplus (m_1 + m_2 - m_3) I_{V(1, 1)}. \end{aligned}$$

Hence $\operatorname{rank}(\tau_{\varphi(m_1, m_2, m_3)}) = \frac{1}{2} \dim(B_1) = 12 \times 5$ if and only if $m_i = 0$, $m_j = -m_k$, or $m_i = 0$, $m_j = m_k$ for some mutually distinct i, j, k of $\{1, 2, 3\}$. Hence α permutes the following six 1-dimensional subspaces of B_2:

$$\langle \varphi(v_{d_1}^*, d_1 d_i) \pm \varphi(v_{d_1}^*, d_1 d_j) \rangle \qquad (i \neq j \in \{2, 3, 4\}).$$

Since $\tau(\varphi(v_{d_1}^*, d_1 d_i) + \varphi(v_{d_1}^*, d_1 d_j), \varphi(v_{d_1}^*, d_1 d_i) + \varphi(v_{d_1}^*, d_1 d_j)) = 4\bar{v}_{d_1}^* + \varphi(\bar{v}_{d_1}^*, d_1 d_i)$, and $\tau(\bar{v}_{d_1}^*, \varphi(\bar{v}_{d_1}^*, d_1 d_i)) = 3\varphi(v_{d_1}^*, d_1 d_i)$, we have that α permutes the subspaces $\langle \varphi(v_{d_1}^*, d_1 d_i) \rangle$, $i = 2, 3, 4$. Finally $\tau(\varphi(v_{d_1}^*, d_1 d_i), \varphi(v_{d_1}^*, d_1 d_i)) = 2\bar{v}_{d_1}^*$ implies the assertion.

(6.5) *By multiplying some element of* $\langle \zeta \rangle$, *we may assume* $\alpha(\varphi(v_x^*, xx')) = \pm \varphi(v_x^*, xx')$ *for each* $x \in C_D(s)$, $x' \in x^Q - \{x\}$.

Proof. By multiplying some element of $\langle \zeta \rangle$, we may assume $\alpha(\varphi(v_{d_1}^*, d_1 d_2) = \pm \varphi(v_{d_1}^*, d_1 d_2)$. Let y_1 be an element of $C_D(s)$ with $|d_1 y_1| = 3$. By the structure of $\langle C_D(s) \rangle / Q \langle \zeta \rangle$, there exists $z_1 \in C_D(s)$ such that $\langle d_1, y_1, z_1 \rangle \cap D$ consists of nine elements, and the product of any two of them is of order 3 ([9]; (1.5)).

Set $y_i = y_1^{d_1 d_i}$ and $z_i = z_1^{d_1 d_i}$ $(2 \leqq i \leqq 4)$. Since $d_1 d_i \in Q$, we have $z_1 z_i \in Q$ and $y_1^{z_1 z_i} \in y_1^Q = \{y_1, y_2, y_3, y_4\}$. We show $y_2 \neq y_1^{z_1 z_2}$. In fact, if $y_2 = y_1^{z_1 z_2}$, then

$$y_1 = y_1^{d_1 d_2 z_1 z_2} = y_1^{(d_2 z_1 d_2)(d_1 z_1 d_1)},$$

since

$$\begin{aligned} d_1 d_2 z_1 z_2 &= d_1 d_2 z_1 z_1^{d_1 d_2} = d_1 (d_2 z_1 d_2)(d_1 z_1 d_1) d_2 \\ &= d_1 (z_1 d_2 z_1)(z_1 d_1 z_1) d_2 = d_1 z_1 d_1 d_2 z_1 d_2. \end{aligned}$$

This implies $z_1^{d_1} = z_1^{d_2}$, and so $d_1 = d_2$, a contradiction. Hence we may assume $y_3 = y_1^{z_1 z_2}$.

By (5.13), $\tau(\varphi(v_x^*, xx'), \varphi(v_y^*, yy')) \neq 0$ if and only if $y' = y^{xx'}$ (and so $x' = x^{yy'}$). Since α preserves $\langle \varphi(v_{d_1}^*, d_1 d_2) \rangle$, α also preserves $\langle \varphi(v_{y_1}^*, y_1 y_2) \rangle$ and $\langle \varphi(v_{z_1}^*, z_1 z_2) \rangle$. Hence α also preserves $\langle \varphi(v_{y_1}^*, y_1 y_3) \rangle$ and furthermore $\langle \varphi(v_{d_1}^*, d_1 d_3) \rangle$, since $y_3 = y_1^{z_1 z_2} = y_1^{d_1 d_3}$. Hence we have proved that $\alpha(\varphi(v_{d_1}^*, d_1 d_i)) = \pm \varphi(v_{d_1}^*, d_1 d_i)$ for each i, and $\alpha(\varphi(v_y^*, yy')) = \pm \varphi(v_y^*, yy')$ for any $y \in C_D(s)$ with $|d_1 y| = 3$ and any $y' \in y^Q - \{y\}$.

For $x \in C_D(s)$ with $|d_1 x| = 2$, there exists $y \in C_D(s)$ with $|d_1 y| = |xy| = 3$. By the same argument as above, we have $\alpha(\varphi(v_x^*, xx')) = \pm \varphi(v_x^*, xx')$ for any $x' \in x^Q - \{x\}$. This completes the proof of (6.5).

(6.6) Conclusion. $C_\Gamma(s) = C$, $|\Gamma| < \infty$

By (6.5), α_V centralizes Q_2 modulo $Z(Q)$ in $GL(V)$. Hence α_V centralizes Q modulo $Z(Q)$. In particular, α_V normalizes Q. The structure of $\mathrm{Aut}(Q)$ implies $\alpha_V \in \mathbb{C}^\times Q$, since $C_{GL(V)}(Q) = \mathbb{C}^\times Z(Q)$ by the irreducibility of Q on V. Hence by multiplying some element of Q, we may assume $\alpha_V = cI_V$ for some $c \in \mathbb{C}$.

Now α acts trivially on B_* and B_2, and $\alpha|_{B_1} = \alpha_R \otimes \alpha_V = cI_{B_1}$. By using (5.8.1) for $v_X \in B_1$ (or more precisely (8.4.2)), we have $c = \pm 1$, that is $\alpha \in \langle s \rangle$.

This completes the proof of $C_\Gamma(s) = C$, and $|\Gamma| < \infty$, as remarked in the beginning of this section.

(6.7) Next we prove $\Gamma = G$. This can be seen by applying Aschbacher's result [1: Sect. 8 (8.1)]. But then we will probably need (some part of) the classification theorem of finite simple groups. So we give an elementary proof. We need the following lemmas.

Lemma 6.7.1 (a weak form of Bender's theorem). *Let Y be a subgroup of even order of a finite group X, and S be a Sylow 2-subgroup of Y. Assume that $N_X(S) \subseteqq Y$ and $C_X(t) \subseteqq Y$ for any involution t of S. Then one of the following holds:*

(i) $X = Y$,

(ii) *all involutions of X consists of a single conjugacy class of X.*

Lemma 6.7.2 (Thompson's transfer lemma). *Let S be a Sylow 2-subgroup of a finite group X of even order, and M be a subgroup of S of index 2. Assume an involution of S is conjugate with no elements of M in X. Then X possesses a subgroup of index 2.*

Moreover we remark the following property of Γ.

Lemma 6.7.3. *The conjugacy classes D_1, D_2, D_4 of G are contained in mutually distinct conjugacy classes of Γ.*

Proof. See (5.5).

(6.8) Let S be a Sylow 2-subgroup of C. Since $Z(S) = \langle s \rangle$, S is a Sylow 2-subgroup of Γ. By Lemma 6.7.2, 6.7.3, Γ has a subgroup H_1 of index 2, and clearly $H_1 \cap G = G'$.

By [13], $N_{H_1}(\langle z \rangle)/\langle z \rangle$ is isomorphic to F'_{24}. In particular, $N_{H_1}(\langle z \rangle) = C_{H_1}(z) = G'$, and $S \cap G'$ is a Sylow 2-subgroup of H_1. By Lemma 6.7.1. and the fact $N_{H_1}(s) = C_{G'}(s)$, we have:

(#1) $H_1 = G'$ *(and thus $\Gamma = G$) or* $C_{H_1}(d_1 d_2) \not\subset G'$.

We consider $C_{H_1}(d_1 d_2)$. Set $H_2 = C_{H_1}(d_1 d_2)$ and $H_3 = C_{H_2}(z) = C_{G'}(d_1 d_2)$.

The group $H_3/\langle d_1 d_2, z \rangle$ is isomorphic to $\mathrm{Aut}(F_{22})$, and so H_3 has a unique subgroup H_4 of index 2 containing $d_1 d_2$. By [2], we have

(#2) $H_4/\langle d_1 d_2 \rangle$ *has just three classes of involutions with their representatives* $t_1 = d_1 d_3 \langle d_1 d_2 \rangle$, $t_2 = s\langle d_1 d_2 \rangle$, *and* $t_3 = d_1 d_3 d_4 d_5 \langle d_1 d_2 \rangle$.

Moreover let x be an element of D with $|d_1 x| = |d_2 x| = 3$, and $|d_3 x| = 2$. Then $t_4 = x x^{d_1 d_2} d_1 d_3 \langle d_1 d_2 \rangle$ is an involution of $H_3/\langle d_1 d_2 \rangle - H_4/\langle d_1 d_2 \rangle$. Let T_i be the inverse image of $\langle t_i \rangle$ in H_3 for each $i = 1, \ldots, 4$. Then T_i, $1 \leqq i \leqq 3$, are

elementary abelian respectively, and T_4 is cyclic. This together with the facts $|T_1 \cap D_4|=0$, $|T_2 \cap D_4|=1$, and $|T_3 \cap D_4|=2$, implies:

(#3) *any two of* t_1, t_2, t_3, t_4 *are not conjugate in* $H_2/\langle d_1 d_2\rangle$.

Let S_1 be a Sylow 2-subgroup of $N_{H_3}(T_2)$ $(=C_{H_3}(T_2))$. Then S_1 is a Sylow 2-subgroup of H_3, and $Z(S_1)=T_2$ by the structure of F_{22}. Hence $N_{H_2}(S_1) \subseteqq N_{H_2}(T_2) \subseteqq C_{H_2}(s) \subseteqq G \cap H_1 = G'$, and thus $N_{H_2}(S_1) \subseteqq H_3$. In particular, S_1 is a Sylow 2-subgroup of H_2.

By Lemma 6.7.2 and (#2), (#3), H_2 has a subgroup H_5 of index 2, and then $C_{H_5}(z)=H_4$. Set $S_2=H_4 \cap S_1$. Then S_2 is a Sylow 2-subgroup of H_5. By (#2), (#3), t_1, t_2, t_3 are representatives of conjugacy classes of involutions of $H_5/\langle d_1 d_2\rangle$, and we have

(#4) $N_{H_4}(T_i)$ *contains a Sylow 2-subgroup of* $N_{H_5}(T_i)$ *for each* $i=1, 2, 3$.

We consider $N_{H_5}(T_i)$, which is the preimage of $C_{H_5/\langle d_1 d_2\rangle}(t_i)$. Since $T_2 \cap D_4 = \{s\}$, we have $N_{H_5}(T_2) \subseteqq C \cap H_1 \subseteqq G'$, and thus $N_{H_5}(T_2) \subseteqq H_4$. By $|T_3 \cap D_4|=2$, we have $C_{H_5}(T_3) \subseteqq H_4$, and $|N_{H_5}(T_3): C_{H_5}(T_3)| \leqq 2$. Hence (#4) implies $N_{H_5}(T_3) \subseteqq H_4$, and thus by Lemma 6.7.1, we have

(#5) $H_5=H_4$ *or* $N_{H_5}(T_1) \not\subset H_4$.

Clearly:

(#6) $H_5=H_4$ *implies* $H_2=H_3 \subseteqq G'$.

We consider $N_{H_5}(T_1)$. Set $H_6=N_{H_5}(T_1)$ and $H_7=C_{H_6}(z)$. Then $H_7=N_{H_4}(T_1)$. By the structure of $H_4/\langle d_1 d_2, z\rangle$ $(\simeq F_{22})$, $H_7/T_1\langle z\rangle$ is isomorphic to $PSU(4, 2)$. Hence H_7/T_1 has jsut three classes of involutions with their representatives $t_5 = d_1 d_4 T_1$, $t_6=d_4 d_5 T_1$, $t_7=d_1 d_4 d_5 d_6 T_1$. Let T_i be the inverse image of $\langle t_i\rangle$ for each $i=5, 6, 7$, respectively. Then $|T_5 \cap D_4|=1$, $|T_6 \cap D_4|=3$, and $|T_7 \cap D_4| = 4$.

Let S_3 be a Sylow 2-subgroup of $N_{H_7}(T_5)$. Then S_3 is a Sylow 2-subgroup of H_7, and $Z(S_3)=T_5$ by the structure of $PSU(6, 2)$. Hence $N_{H_6}(S_3) \subseteqq C_{H_6}(s) \subseteqq G'$, and thus $N_{H_6}(S_3) \subseteqq H_7$. Hence S_3 is a Sylow 2-subgroup of H_6, and we have

(#7) $N_{H_7}(T_i)$ *contains a Sylow 2-subgroup of* $N_{H_6}(T_i)$ *for each* $i=5, 6, 7$.

Since $N_{H_6}(T_5) \subseteqq C$, and $N_{H_6}((T_6-T_1) \cap D_2) = C_{H_6}(d_4 d_5) \subset C_{H_6}(d_1 d_2 d_4 d_5) \subseteqq G'$, we have $N_{H_6}(T_5)$, $N_{H_6}(T_6) \subseteqq H_7$. Moreover for any element t of $T_7 \cap D_4$, we have $N_{H_6}(T_7) \cap C_\Gamma(t) \subseteqq G'$. Hence $|N_{H_6}(T_7): N_{H_6}(T_7) \cap C_\Gamma(T_7 \cap D_2)|=1, 2$, or 4, and thus (#7) implies $N_{H_6}(T_7) \subset H_7$. Hence by Lemma 6.7.1, we have $H_6=H_7$.

By (#5), we have $H_5=H_4$, and then by (#6), $H_2=H_3 \subset G'$. Hence (#1) implies $H_1=G'$, and thus $\Gamma=G$, and the proof of Theorem 5 is complete.

7. The automorphism group Aut$(B(6), \tau(6))$ and Aut$(B(7), \tau(7))$

(7.1) First we prove Theorem 4, by using Theorem 2. We use the notation in Sect. 4. We set $\tau=\tau(7)$, $B=B(7)=B^{(3)} \oplus R^{(3)}$, $K=N(7)=\langle D\rangle$, $K'=\langle E\rangle$, and $\Gamma = \mathrm{Aut}(B, \tau)$. Let $e \in E$, and set $C=C_K(e)$.

Let $\alpha \in C_K(e)$. Then α acts on $B^{(3)}$, $R^{(3)}$, the 1, (-1)-eigenspaces of e, respectively. By Theorem 2 and the facts $(B^{(3)}, \tau|_{B^{(3)}}) \simeq (B(6), \tau(6))$ and $C/\langle e\rangle \simeq C_*$, we may assume $\alpha|_{B^{(3)}} = I_{B^{(3)}}$ by multiplying some element of C. By similar arguments as in (6.2), we have $\alpha|_{R^{(3)}} = cI_{R^{(3)}}$ for some $c \in \mathbb{C}^\times$. Moreover Theorem 3(3) implies $c = \pm 1$, that is $\alpha \in \langle e\rangle$. Hence $C_\Gamma(e) = C$ and Γ is finite by Theorem B.

(7.2) Let S be a Sylow 2-subgroup of C. By character arguments as in (5.3), we have $e^K = e^\Gamma \cap K$, and $D^\Gamma \cap K' = \emptyset$. In particular, $|S \cap e^\Gamma| = 7$ and $Z(S) \cap e^\Gamma = \{e\}$ by the structure of $\Omega(7,3)$. By Lemma 6.2, Γ has a subgroup H of index 2, and $S \cap K'$ is a Sylow 2-subgroup of H. By [16], $H/O(H)$ is isomorphic to $\Omega(7,3)$.

Let P be an S-invariant Sylow p-subgroup of $O(H)$ for some prime $p \mid |O(H)|$, and set $S \cap e^\Gamma = \{e_0 = e, e_1, \ldots, e_6\}$. Then $e = e_1 e_2 \ldots e_6$ by the structure of $\Omega(7,3)$. Since $C_{O(H)}(e) = \langle z\rangle$, we have $C_P(e_i) \subseteqq \langle z\rangle$ for each i. Let $x \in Z(P)$. Then $x^{e_i} x \in \langle z\rangle$ for each i. Suppose $p \neq 3$. Then $x^{e_i} = x^{-1}$ for each i. On the other hand, $x^e = x^{e_1 e_2 \cdots e_6} = x$. Since p is odd, we have $x = 1$, and thus $P = \langle 1\rangle$. Suppose $p = 3$. By a similar argument as above, we may assume $Z(P) = \langle z\rangle$. Moreover concerning $C_P(\langle z\rangle)$, such argument imply $C_P(\langle z\rangle) = \langle z\rangle$. Hence we have $P = \langle z\rangle$, and thus $O(H) = \langle z\rangle$. Hence we have $H = K'$, and this completes the proof of Theorem 4.

(7.3) Here we mention the existence of the group $N(7)$. We can prove that $\mathrm{Aut}(B(7), \tau(7))$ satisfies the conditions (i), (ii) in (1.1) without the assumption that $N(7)$ acts on $(B(7), \tau(7))$.

First we define the algebra (B, τ) by the formulas in Theorem 3 (or equivalently Proposition in (4.4)). Then one can see that $\mathrm{Aut}(B, \tau)$ $(= \Gamma$, say) contains a subgroup isomorphic to $C_0^{(3)}$. In fact, the product formulas in Theorem 3 are determined so that $C_0^{(3)}$ does act on (B, τ). Moreover it is a proper subgroup of Γ. In fact, the automorphism denoted by f in (4.2) is outside $C_0^{(3)}$. This fact can be also proved by using Proposition in (4.4), that is, one can easily verify that Γ contains a subgroup of shape $3^6.\Omega^-(6,2)$ or $3^6.\Omega(5,3)$, which can not be contained in $C_0^{(3)}$. (This corresponds a maximal parabolic subgroup of $\Omega(7,3)$.) Then in view of (7.1), (7.2), we have $\Gamma \simeq 3.\Omega(7,3).2$. Moreover by the structure of $C_0^{(3)}$, the group Γ can not be written as a direct product, and so Γ satisfies the condition (i). The condition (ii) is easy consequence of the structure of $\mathrm{Aut}(\Omega(7,3))$.

(7.4) Next we prove Theorem 2. Hereafter we use the notation in Sect. 3. We set $B = B(6)^{(4)}$, $\tau = \tau(6)^{(4)}$, $G = C_*^{(4)}$, and $\Gamma = \mathrm{Aut}(B, \tau)$.

Set $s = d(-1, -1, 1, 1, 1, 1)$, and $C = C_G(s)$. Then C is generated by the following elements:

$$t_{ij}, \quad t'_{ij} = t_{ij} d(1, 1, \overset{i}{-1}, 1, \overset{j}{-1}, 1), \quad (3 \leqq i \neq j \leqq 6)$$

$$t_{12}, \quad s(\overset{i}{-1}, 1, \overset{j}{-1}, 1, 1, 1), \quad (1 \leqq i \leqq 2, 3 \leqq j \leqq 6)$$

$$z, \pi', \xi'.$$

The group $\langle t_{ij}, t'_{ij} \mid 3 \leqq i \neq j \leqq 6 \rangle$ is a subgroup of C and is isomorphic to the Weyl group $W(D_4)$. Set

$$Q_2 = \{t_{ij}t'_{ij}, t_{ij}t_{kh}, t_{ij}t'_{kh} \mid \{i, j, k, h\} = \{1, 2, 3, 4\}\},$$

and $Q = \langle Q_2 \rangle$. Then $Q = O_2(\langle t_{ij}, t'_{ij} \mid 3 \leqq i \neq j \leqq 6 \rangle) \simeq 2_+^{1+4}$.

Now the following are C-invariant (and C-irreducible) subspaces of B:

$$\begin{aligned}
B_\infty &= \langle u_{33}+u_{44}+u_{55}+u_{66}, \bar{u}_{33}+\bar{u}_{44}+\bar{u}_{55}+\bar{u}_{66} \rangle, \\
B_*^- &= \langle u_{11}-u_{22}, \bar{u}_{11}-\bar{u}_{22} \rangle, \\
B_*^+ &= \langle u_{11}+u_{22}, u_{12}, \bar{u}_{11}-\bar{u}_{22}, \bar{u}_{12} \rangle, \\
B_2 &= \langle u_{ii}-u_{jj}, u_{ij}, \bar{u}_{ii}-\bar{u}_{jj}, \bar{u}_{ij} \mid 3 \leqq i \neq j \leqq 6 \rangle, \\
B_1 &= \langle u_{ij}, \bar{u}_{ij} \mid 1 \leqq i \leqq 2, 3 \leqq j \leqq 6 \rangle.
\end{aligned}$$

(7.5) The space B_1 is the (-1)-eigenspace of s. We set $B_1 = R \otimes V$, where

$$\begin{aligned}
R = \langle & u_1^{(1)} = {}^t(1, 0, 0, 0), u_2^{(1)} = {}^t(0, 1, 0, 0), \\
& u_1^{(2)} = {}^t(0, 0, 1, 0), u_2^{(2)} = {}^t(0, 0, 0, 1) \rangle, \\
V = \langle & v_3, v_4, v_5, v_6 \rangle, \quad \text{and} \\
& u_{ij} = u_i^{(1)} \otimes v_j, \quad \bar{u}_{ij} = u_i^{(2)} \otimes v_j.
\end{aligned}$$

For any element $q \in Q$, there exists some $q_V \in GL(V)$ such that $q(r \otimes v) = r \otimes q_V(v)$ for any $r \in R$, $v \in V$. The map $q \to q_V$ gives the unique non-linear irreducible representation of Q.

Set $B_* = B_*^+ \oplus B_*^-$. Concerning B_* and B_2, we use the following notation in order to formulate the algebra products $\tau(B_*, B_1)$ and $\tau(B_2, B_1)$. First we regard B_* as a linear subspace of $\mathrm{End}(R)$ by the following identification:

$$\begin{aligned}
u_{11}-u_{22} &= J_1 \otimes \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix}, & \bar{u}_{11}-\bar{u}_{22} &= J_2 \otimes \begin{pmatrix} 4 & 0 \\ 0 & -4 \end{pmatrix}, \\
u_{11}+u_{22} &= J_1 \otimes \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, & \bar{u}_{11}+\bar{u}_{22} &= J_2 \otimes \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}, \\
u_{12} &= J_1 \otimes \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix}, & \bar{u}_{12} &= J_2 \otimes \begin{pmatrix} 0 & -1 \\ -1 & 0 \end{pmatrix},
\end{aligned}$$

where $J_1 = \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}$, $J_2 = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}$, and $J_1 \otimes X = \begin{pmatrix} O & O \\ X & O \end{pmatrix}$.

Next we define $v_j^{(i)}$ $(1 \leqq i \leqq 2,\ 1 \leqq j \leqq 3)$, elements of $\mathrm{End}(R)$, as follows:

$$\begin{aligned}
v_1^{(1)} &= \tfrac{1}{2}(u_{11}+u_{22}), & v_1^{(2)} &= \tfrac{1}{2}(\bar{u}_{11}+\bar{u}_{22}), \\
v_2^{(1)} &= \tfrac{1}{4}(-u_{11}-u_{22}+2\theta u_{12}), & v_2^{(2)} &= \tfrac{1}{4}(-\bar{u}_{11}-\bar{u}_{22}-2\theta \bar{u}_{12}), \\
v_3^{(1)} &= \tfrac{1}{4}(-u_{11}-u_{22}-2\theta u_{12}), & v_3^{(2)} &= \tfrac{1}{4}(-\bar{u}_{11}-\bar{u}_{22}+2\theta \bar{u}_{12}).
\end{aligned}$$

Next we set

$$\varphi(v_1^{(1)}, t_{ij}t'_{ij}) = -\tfrac{1}{4}(u_{ii}+u_{jj}-u_{kk}-u_{hh}),$$
$$\varphi(v_2^{(1)}, t_{ij}t_{kh}) = \tfrac{1}{2}(u_{ij}+u_{kh}),$$
$$\varphi(v_2^{(1)}, t'_{ij}t'_{kh}) = -\tfrac{1}{2}(u_{ij}+u_{kh}),$$
$$\varphi(v_3^{(1)}, t_{ij}t'_{kh}) = \tfrac{1}{2}(u_{ij}-u_{kh}),$$

where $\{i,j,k,h\}=\{3,4,5,6\}$. Here we set $Q_2(1)=\{t_{ij}t'_{ij}\}$, $Q_2(2)=\{t_{ij}t_{kh}, t'_{ij}t'_{kh}\}$, $Q_2(3)=\{t_{ij}t'_{kh}\}$. Then $Q_2=Q_2(1)\cup Q_2(2)\cup Q_2(3)$, and for $l\in\{1,2,3\}$ and $q\in Q_2$, the vector $\varphi(v_l^{(1)}, q)$ is defined if and only if $q\in Q_2(l)$. We define $\varphi(v_l^{(2)}, q)$ ($q\in Q_2(l)$) by changing the roles of u_{ij} and $\bar{u}_{ij}$. Notice that $\varphi(v_l^{(m)}, q) = -\varphi(v_l^{(m)}, sq)$ since $s=t_{ij}t'_{ij}t_{kh}t'_{kh}$.

Finally we set

$$a_1=\tfrac{1}{4}(u_{33}+u_{44}+u_{55}+u_{66}), \qquad a_2=\tfrac{1}{4}(\bar{u}_{33}+\bar{u}_{44}+\bar{u}_{55}+\bar{u}_{66}).$$

(7.6) Now we present the product formulas:

$$\tau(B_\infty, B_\infty):\ \tau(a_i, a_i)=3a_j,$$

$$\tau(B_\infty, B_*):\ \tau\left(a_i, J_i\otimes\begin{pmatrix} a+b & c \\ c & a-b\end{pmatrix}\right)=J_j\otimes\begin{pmatrix} a+2b & -c \\ -c & a-2b\end{pmatrix},$$

$$\tau(B_\infty, B_2):\ \tau(a_i, \varphi(v_l^{(i)}, q))=\varphi(v_l^{(j)}, q),$$

$$\tau(B_\infty, B_1):\ \tau(a_i, B_1)=0,$$

$$\tau(B_*, B_*):\ \tau\left(J_i\otimes\begin{pmatrix} a+b & c \\ c & a-b\end{pmatrix}, J_i\otimes\begin{pmatrix} a'+b' & c' \\ c' & a'-b'\end{pmatrix}\right)$$
$$=J_j\otimes\begin{pmatrix} a\,a'+b\,b' & a\,c'+c\,a' \\ a\,c'+c\,a' & a\,a'+c\,c'\end{pmatrix}+(2a\,a'-2c\,c'+b\,b')\,a_j,$$

$$\tau(B_*, B_2):\ \tau(X, \varphi(v_l^{(i)}, q))=\mathrm{Tr}(X v_l^{(j)})\,\varphi(v_l^{(j)}, q),$$

$$\tau(B_*, B_1):\ \tau(X, r\otimes v)=X(r)\otimes v,$$

$$\tau(B_2, B_2):\ \tau(\varphi(v_l^{(i)}, q), \varphi(v_m^{(i)}, q'))$$
$$=\begin{cases} v_l^{(j)}+a_j & (\text{if } l=m,\ q=q') \\ -\varphi(v_l^{(j)}, qq') & (\text{if } l=m,\ q\neq q') \\ 2\varphi(v_n^{(j)}, qq') & (\text{if } l\neq m,\ qq'\in Q_2(n)) \\ 0 & (\text{otherwise}),\end{cases}$$

$$\tau(B_2, B_1):\ \tau(\varphi(v_l^{(i)}, q), r\otimes v)=v_l^{(i)}(r)\otimes q_V(v),$$

where $\{i,j\}=\{1,2\}$, $q\in Q_2(l)$, $q'\in Q_2(m)$, $X\in B_*$, $r\in R$, and $v\in V$. In addition, $\tau(U,\bar{U})=0$ as mentioned in Sect. 3. We do not mention about $\tau(B_1, B_1)$. We need only:

$$\tau(u_{13}, u_{14}) = -\bar{u}_{34}\in B_2.$$

In general, $\tau(B_1, B_1)\subseteqq B_\infty\oplus B_*\oplus B_2$.

Now we shall show $C=C_\Gamma(s)$. If the arguments used here are similar to those in Sect. 6, we will only present the outline of the proof.

Let $\alpha \in C_\Gamma(s)$. We shall show $\alpha \in C$.

(7.7) *$C_\Gamma(s)$ preserves B_∞, B_*, B_2, and B_1 respectively.*

Proof. Since B_1 is the (-1)-eigenspace of s, B_1 and $B_\infty \oplus B_* \oplus B_2$ are stable by the action of $C_\Gamma(s)$, respectively.

For $b \in B_\infty \oplus B_* \oplus B_2$, we define the linear maps $\tau_b: B \to B$, and $\tau'_b: B_1 \to B_1$ by $\tau_b(x) = \tau(b, x)$ and $\tau'_b(y) = \tau(b, y)$ for $x \in B$ and $y \in B_1$. Moreover we define the $C_\Gamma(s)$-invariant bilinear symmetric forms β, β' by $\beta(a, b) = \mathrm{Tr}(\tau_a \tau_a)$ and $\beta'(a, b) = \mathrm{Tr}(\tau'_a \tau'_b)$.

Then by direct calculations (as in (6.1)), we have $B_\infty = \mathrm{Rad}(\beta')$, $B_*^+ = \mathrm{Rad}(\beta - \frac{3}{2}\beta')$, $B_*^- = \mathrm{Rad}(\beta - 6\beta')$, and $B_2 = \mathrm{Rad}(\beta - 3\beta')$. Hence we have the assertions.

(7.8) *There exist some $\alpha_R \in GL(R)$ and $\alpha_V \in GL(V)$ such that $a|_{B_1} = \alpha_R \otimes \alpha_V$.*

Proof. See Lemma 5.11.1.

(7.9) We consider the action of α on B_∞.

Let $\alpha(a_1) = c_1 a_1 + c_2 a_2$, and $\alpha(a_2) = c_3 a_1 + c_4 a_2$. Then $0 = \alpha(\tau(a_1, a_2)) = \tau(c_1 a_1 + c_2 a_2, c_3 a_1 + c_4 a_2) = 3(c_2 c_4 a_1 + c_1 c_3 a_2)$, and so $c_1 c_3 = c_2 c_4 = 0$. On the other hand, $3(c_3 a_1 + c_4 a_2) = \alpha(3a_2) = \alpha(\tau(a_1, a_1)) = \tau(c_1 a_1 + c_2 a_2, c_1 a_1 + c_2 a_2) = 3(c_2^2 a_1 + c_1^2 a_2)$, and so $c_3 = c_2^2$, $c_4 = c_1^2$. Similarly we have $c_1 = c_4^2$, $c_2 = c_3^2$. Hence we have $(c_1, c_2, c_3, c_4) = (\omega^i, 0, 0, \omega^{2i})$ or $(0, \omega^i, \omega^{2i}, 0)$ for some $i = 1, 2, 3$. Hence by multiplying some element of $\langle \xi', z \rangle$, we may assume $\alpha(a_1) = a_1$ and $\alpha(a_2) = a_2$.

(7.10) Let $b \in B_*^- \cap U$. Then $\alpha(b) \in B_*^-$ by (7.7), and $\tau(\alpha(b), a_2) = \alpha(\tau(b, a_2)) = 0$. Hence $\alpha(b) \in B_*^- \cap U$ by (7.6), that is, $B_*^- \cap U$ is stable by α. Similarly $B_*^- \cap \bar{U}$, $B_*^+ \cap U$, $B_*^+ \cap \bar{U}$, $B_2 \cap U$ and $B_2 \cap \bar{U}$ are stable by α, respectively.

(7.11) For each $i = 1, 2$, α permutes the three subspaces $\langle \varphi(v_l^{(i)}, q) \mid q \in Q_2^{(l)} \rangle$, $l = 1, 2, 3$ (see (6.3)). By multiplying some element of $\langle \pi', d(-1, 1, -1, 1, 1, 1) \rangle$, we may assume α preserves $\langle \varphi(v_l^{(i)}, q) \mid q \in Q_2^{(l)} \rangle$ for any $i = 1, 2$, and $l = 1, 2, 3$. Moreover for any i and l, α preserves the set of vectors $\{\varphi(v_l^{(i)}, q) \mid q \in Q_2^{(l)}\}$ (see (6.4)). Hence by multiplying some element of $\langle t_{34}, t_{56} \rangle$, we may assume $\alpha(\varphi(v_1^{(1)}, q)) = \pm \varphi(v_1^{(1)}, q)$ for each $q \in Q_2^{(1)}$. Then by (7.6), $\alpha(\varphi(v_l^{(i)}, q)) = \pm \varphi(v_l^{(i)}, q)$ for any i, l, and $q \in Q_2^{(l)}$ (see (6.5)).

(7.12) Now we consider the action of α on B_*.

Regarding B_* as a subspace of $\mathrm{End}(R)$, we have $\alpha(b) = \alpha_R b \alpha_R^{-1}$ for each $b \in B_*$.

Set $\alpha_R = \begin{pmatrix} X_{11} & X_{21} \\ X_{21} & X_{22} \end{pmatrix}$, where X_{ij} $(1 \leqq i, j \leqq 2)$ are (2×2)-matrices. By (7.10), α_R preserves $\mathbb{C} v_l^{(i)}$ for any i, l. By straight calculations, we have $X_{12} = X_{21} = O$, and $X_{11} = \begin{pmatrix} d & e \\ e & d \end{pmatrix} = c^{-1} X_{22}$ for some $c, d, e \in \mathbb{C}$. Now α preserves $B_*^- \cap U$, that is,

$$\alpha(u_{11} - u_{22}) = \alpha_R \left(J_1 \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right) \alpha_R^{-1}$$

$$= J_1 \otimes \left(c X_{11} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} X_{11}^{-1} \right) \in \mathbb{C} \left(J_1 \otimes \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right).$$

This implies $d=0$, or $e=0$. Hence by multiplying some element of $\langle u_{12}\rangle$, we may assume $e=0$. Then $\alpha(b)=cb$, and $\alpha(b')=c^{-1}b'$ for any $b\in B_*\cap U$, $b'\in B_*\cap \bar{U}$. By using formulas $\tau(B_*, B_*)$ in (7.6), we have $c=1$.

(7.13) Finally by similar arguments as in (6.6), we may assume $\alpha\in\langle s\rangle$ by multiplying some element of Q. Hence we have $C=C_\Gamma(s)$, and thus $|\Gamma|<\infty$ by Theorem B.

(7.14) Next we prove $\Gamma=G$. Here we need the following theorem.

Theorem C (Bender [4]). *Let X, Y, and S be groups as in the assumption of Lemma* 6.7.1. *Then one of the following holds:*

(i) $X=Y$,

(ii) *S is isomorphic to a cyclic group or a generalized quaternion group, and so S possesses a unique involution,*

(iii) *There exists a normal subgroup L of X containing $O(X)$ such that $|X/L|$ is odd, and $L/O(X)$ is isomorphic to one of $PSL(2, 2^m)$, $Sz(2^{2m-1})$, or $PSU(3, 2^m)$, $m\geqq 2$. Furthermore $Y=O(X)N_X(S)$ and in particular, $O(X)S$ is normal in Y.*

(7.15) Let S be a Sylow 2-subgroup of C. Then by the structure of $\Omega^-(6, 3)$, we have $Z(S)=\langle s\rangle$, and thus S is a Sylow 2-subgroup of Γ.

Set $S_1=A\cap S$, and $S_2=A'\cap S$, where A is the image of $\mathrm{Aut}(\Lambda)$ in $G=C_*$ (see Sect. 3). Then by a character argument as in (5.3), we have that any element of $S-S_1$ (resp. $S-S_2$) is not conjugate to elements of S_1 (resp. S_2) in Γ. Hence Lemma 6.7.2 implies that Γ has a normal subgroup G_0 of index 4. Then $C_{G_0}(s)=C\cap A'$. By [14], $N_{G_0}(\langle z\rangle)/\langle z\rangle$ is isomorphic to $\Omega^-(6, 3)$, and thus $N_{G_0}(\langle z\rangle)=A'=G'$. Since G' has just one class of involutions, the groups G_0, $N_{G_0}(\langle z\rangle)$ and S_2 satisfy the assumption of Theorem C. This implies $N_{G_0}(\langle z\rangle)=G_0$, and we have $\Gamma=G$.

8. Appendix: $\tau(24)(B_1, B_1)$ and the structure of $C_{N(24)}(s)$

In this section, we use the notation in Sects. 5, 6. In particular, the involution s is defined by the fixed sextet $\{T_1, \ldots, T_6\}$ as the product of the elements of any T_i.

We determine the algebra product $\tau(B_1, B_1)$, and construct the group $C_G(s)$. By this, the algebra $(B, \tau)=(B(24), \tau(24))$ can be constructed without using the original construction.

(I) $\tau(B_1, B_1)$

(8.1) First we define a bilinear symmetric form $f(\ ,\ )$ on V by

$$f(v(\delta_l), v(\delta_l'))=\begin{cases}1 & (\text{if } (\delta_l)=(\delta_l')),\\ 0 & (\text{otherwise}),\end{cases}$$

for $(\delta_l), (\delta_l')\in\Delta$. Then $f(\ ,\ )$ is C-invariant.

(8.2) Let X, Y be octads with v_X, $v_Y \in B_1$. Then there exist some i, j such that $v_X = r_i \otimes v(\delta(X))$, $v_Y = r_j \otimes v(\delta(Y))$. Then we define the vector $\Phi(v_X, v_Y)$ of B_2 as follows:

$$\Phi(v_X, v_Y) = \sum_{q \in Q_2} \mathrm{Tr}(\bar{u}_{ij} v_x^*) f(q_V(v(\delta(X))), v(\delta(Y))) \varphi(\bar{v}_x^*, q).$$

Here $q|_{B_1} = I_R \otimes q_V$, and $x \in C_D(s)$ is chosen so that $\varphi(\bar{v}_x^*, q)$ is defined, that is, $q = xx'$ for some $x' \in x^Q$. Then q is also written as $x^z x'^z$ or $x^{z^2} x'^{z^2}$. But then $v_{x^z}^* = \omega v_x^*$, $\bar{v}_{x^z}^* = \omega^2 \bar{v}_x^*$ imply that $\mathrm{Tr}(\bar{u}_{ij} v_{x^z}^*) = \omega \mathrm{Tr}(\bar{u}_{ij} v_x^*)$ and $\varphi(\bar{v}_{x^z}^*, q) = \omega^2 \varphi(v_x^*, q)$. Hence $\Phi(v_X, v_Y)$ is independent of the choice of x for each q.

We shall show the following:

(8.3) $\tau(v_X, v_Y) = f(v(\delta(X)), v(\delta(Y)))\, \bar{u}_{ij} + \frac{1}{6} \Phi(v_X, v_Y)$.

(8.4) *If* $\delta(X) = \delta(Y)$ *then* (8.3) *holds.*

Proof. Let $q \in Q_2$, and suppose $f(q_V(v(\delta(X))), v(\delta(Y))) \neq 0$. Then $q \in \langle L \rangle$ by Lemma 5.9.2, and so $q = ab$ for some $a \neq b$ of T_k for some $k \in \{1, \ldots, 6\}$. Then by using (5.7.2) and Lemma 3.9.1, we have

$$\mathrm{Tr}(\bar{u}_{ij} v_a^*) = \begin{cases} 0 & (\text{if } i \neq j) \\ 1 & (\text{if } i = j \neq k \text{ and so } X = Y) \\ -1 & (\text{if } i = j = k \text{ and so } X = Y). \end{cases}$$

Hence we have

(8.4.1)
$$\mathrm{Tr}(\bar{u}_{ij} v_a^*) f(q_V(v(\delta(X))), v(\delta(Y))) \varphi(\bar{v}_a^*, q)$$
$$= \begin{cases} -\varphi(v_a^*, q) & (\text{if } i = j = k \text{ and } a, b \in X, \\ & \quad \text{or if } i = j \neq k \text{ and } |\{a, b\} \cap X| = 1) \\ \varphi(v_a^*, q) & (\text{if } i = j = k \text{ and } |\{a, b\} \cap X| = 1, \\ & \quad \text{or if } i = j \neq k \text{ and } a, b \notin X) \\ 0 & (\text{otherwise}). \end{cases}$$

Suppose $i \neq j$. Then the right-hand side of (8.3) equals u_{ij}. Hence by Lemma 5.9.1, we have $|X \cap Y| = 4$, and $(X \cup Y) - (X \cap Y) = T_i \cup T_j$. Since $\tau(u_{ij}, v_X) = u_{ij}(r_i) \otimes v(\delta(X)) = \bar{r}_j \otimes v(\delta(X)) = \bar{v}_Y$, we have $\tau(v_X, v_Y) = \bar{u}_{ij}$ by (5.7.4) and Lemma 5.8.1. Hence (8.3) holds.

Next suppose $i = j$. Then $X = Y$. Set $X \# T_h = \{a_h\}$, and $T_h - \{a_h\} = \{a_h', a_h'', a_h'''\}$ for each h. Notice that $X \cap T_i = T_i - X \# T_i$ and $X \cap T_k = X \# T_k$ for each $k \neq i$. Then by (5.8.1) and (5.11), we have

(8.4.2)
$$\tau(v_X, v_X) = \tfrac{1}{8}(-\sum_{a_s \in X} \bar{v}_s + 3 \sum_{a_t \notin X} \bar{v}_t)$$
$$= \tfrac{1}{8}(3\bar{v}_{a_i} - 5(\bar{v}_{a_i'} + \bar{v}_{a_i''} + \bar{v}_{a_i'''}) + \sum_{k \neq i} (3(\bar{v}_{a_k'} + \bar{v}_{a_k''} + \bar{v}_{a_k'''}) - 5\bar{v}_{a_k}))$$
$$= \tfrac{1}{8}(-3(\bar{v}_{a_i} + \bar{v}_{a_i'} + \bar{v}_{a_i''} + \bar{v}_{a_i'''}) + 6\bar{v}_{a_i} - 2(\bar{v}_{a_i'} + \bar{v}_{a_i''} + \bar{v}_{a_i'''})$$
$$+ \sum_{k \neq i} ((\bar{v}_{a_k} + \bar{v}_{a_k'} + \bar{v}_{a_k''} + \bar{v}_{a_k'''}) - 6\bar{v}_{a_k} + 2(\bar{v}_{a_k'} + \bar{v}_{a_k''} + \bar{v}_{a_k'''})))$$
$$= \bar{u}_{11} + \varphi(\bar{v}_{a_i}^*, a_i a_i') + \varphi(\bar{v}_{a_i}^*, a_i a_i'') + \varphi(\bar{v}_{a_i}^*, a_i a_i''')$$
$$- \sum_{k \neq i} (\varphi(\bar{v}_{a_k}^*, a_k a_k') + \varphi(\bar{v}_{a_k}^*, a_k a_k'') + \varphi(\bar{v}_{a_k}^*, a_k a_k''')).$$

Hence by (8.4.1), we have (8.3).

(8.5) From now on, we suppose $\delta(X)\neq\delta(Y)$. Then by using the properties of $S(24,8,5)$ and the definitions of B_1 and B_2, it is easily seen that $\tau(v_X, v_Y)\in B_2$. On the other hand, by the definition, $f(v(\delta(X)), v(\delta(Y)))=0$. Hence in order to prove (8.3), it is sufficient to show $\tau(v_X, v_Y)=\frac{1}{6}\Phi(v_X, v_Y)$.

(8.6) *Let* $a, b\in L$, *and* $ab(v_X)=\varepsilon_X v_X$, $ab(v_Y)=\varepsilon_Y v_Y$. *Then* $ab(\Phi(v_X, v_Y))=\varepsilon_X\varepsilon_Y\Phi(v_X, v_Y)$.

Proof. Clearly $\varepsilon_X, \varepsilon_Y\in\{1, -1\}$, and it is easily seen that

$$(8.6.1)\quad \Phi(v_X, v_Y)=\sum_{q\in Q_2}\mathrm{Tr}(\bar{u}_{ij}v^*_{x^{ab}})f(q_V^{ab}(v(\delta(X))), v(\delta(Y)))\varphi(\bar{v}^*_{x^{ab}}, q^{ab}),$$

and

$$(8.6.2)\quad ab(\Phi(v_X, v_Y))=\sum_{q\in Q_2}\mathrm{Tr}(\bar{u}_{ij}v^*_x)f(q_V(v(\delta(X))), v(\delta(Y)))\varphi(\bar{v}^*_{x^{ab}}, q^{ab}).$$

Let $a\in T_k$ and $b\in T_h$. Then by Lemma 3.9.1, we may assume $a_R=\frac{1}{4}(\sum_{j\neq k}u_{jj}-u_{kk})$, $b_R=\frac{1}{4}(\sum_{j\neq h}u_{jj}-u_{hh})$. Then we have

$$\begin{aligned} ab(v_X)&=ab(r_i\otimes v(\delta(X))=a_R b_R(r_i)\otimes a_V b_V(v(\delta(X))\\ &=\varepsilon_{ki}\varepsilon_{hi}r_i\otimes a_V b_V(v(\delta(X)) \end{aligned}$$

where ε_{ij} is defined as follows:

$$\varepsilon_{ij}=\begin{cases} 1 & (\text{if } i=j)\\ -1 & (\text{if } i\neq j).\end{cases}$$

Hence we have $a_V b_V(v(\delta(X)))=\varepsilon_{ki}\varepsilon_{hi}\varepsilon_X v(\delta(X))$. Similarly $a_V b_V(v(\delta(X)))=\varepsilon_{kj}\varepsilon_{hj}\varepsilon_Y v(\delta(Y))$.

Let $q\in Q_2$, and $q_V(v(\delta(X)))\in\langle v(\delta(Y))\rangle$. Since q is an involution, $q_V(v(\delta(X)))=\varepsilon v(\delta(Y))$ for some $\varepsilon=\pm 1$. Then

$$\begin{aligned} (q^{ab})_V(v(\delta(X)))&=a_V b_V q_V a_V b_V(v(\delta(X)))\\ &=a_V b_V q_V(\varepsilon_{ki}\varepsilon_{hi}\varepsilon_X v(\delta(X)))\\ &=a_V b_V(\varepsilon_{ki}\varepsilon_{hi}\varepsilon_X\varepsilon v(\delta(Y)))\\ &=\varepsilon_{ki}\varepsilon_{hi}\varepsilon_X\varepsilon\varepsilon_{kj}\varepsilon_{hj}\varepsilon_Y v(\delta(Y)). \end{aligned}$$

Hence we have

$$\begin{aligned} (8.6.3)\qquad & f((q^{ab})_V(v(\delta(X))), v(\delta(Y)))\\ &=\varepsilon_{ki}\varepsilon_{hi}\varepsilon_{kj}\varepsilon_{hj}\varepsilon_X\varepsilon_Y f(q_V(v(\delta(X))), v(\delta(Y))). \end{aligned}$$

If $f(q_V(v(\delta(X))), v(\delta(Y)))=0$ for $q\in Q_2$, that is $q_V(v(\delta(X)))\notin\langle v(\delta(Y))\rangle$, then by similar arguments we have $f((q^{ab})_V(v(\delta(X))), v(\delta(Y)))=0$. Hence (8.6.3) holds for any $q\in Q_2$.

On the other hand, we have

$$\begin{aligned}(8.6.4)\qquad \mathrm{Tr}(\bar{u}_{ij} v^*_{x^{ab}}) &= \mathrm{Tr}(\bar{u}_{ij}(a_R b_R) v^*_x (a_R b_R)) \\ &= \mathrm{Tr}((a_R b_R)\bar{u}_{ij}(a_R b_R) v^*_x) \\ &= \varepsilon_{ki}\varepsilon_{hi}\varepsilon_{kj}\varepsilon_{hj}\mathrm{Tr}(\bar{u}_{ij} v^*_x).\end{aligned}$$

By using (8.6.1)-(8.6.4), we have the assertion.

(8.7) *If* $|X\cap Y|=0$ *or* 4, *then* $\Phi(v_X, v_Y)\in\langle\bar{v}_Z\rangle$, *where*

$$Z=\begin{cases}(X\cup Y)-(X\cap Y) & (\text{if } |X\cap Y|=4),\\ L-(X\cup Y) & (\text{if } |X\cap Y|=0).\end{cases}$$

Moreover if $|X\cap Y|=2$, *then* $\Phi(v_X, v_Y)=0$, *and hence, in particular,* (8.3) *holds. In addition, in each case,*

$$\tau(\Phi(v_X, v_Y), \bar{v}_Y)\in\langle v_Y\rangle.$$

Proof. By using (8.5) and (5.3.1), the assertions are easily verified. (For example, see the proof of (5.8.4).)

(8.8) *The cardinality of* $\{q\in Q_2 \mid f(q_V(v(\delta(X))), v(\delta(Y)))\neq 0\}$ *is* 16.

Proof. By Lemma 5.9.2, $f(q_V(v(\delta(X))), v(\delta(Y)))\neq 0$ implies $q\notin\langle L\rangle$. The number of the elements of $Q_2-\langle L\rangle$ is $756-6\times\binom{4}{2}=45\times 16$. On the other hand, if $q\notin Q_2-\langle L\rangle$, then q fixes each T_i and has eight fixed elements in L (see (5.5)). By [18], such elements are mutually conjugate in $N_G(L)\cap C_G(s)$, and fall into 45 elements in $(N_G(L)\cap C_G(s))/\langle L\rangle$. Hence we have the assertion.

(8.9) We consider $\tau(\Phi(v_X, v_Y), \bar{v}_X)$. By (8.7), it is sufficient to consider the coefficient of v_Y. By the definition and (5.11), we have

$$\begin{aligned}&\tau(\Phi(v_X, v_Y), \bar{v}_X)\\ &\quad=\sum_{q\in Q_2}\mathrm{Tr}(\bar{u}_{ij} v^*_x) f(q_V(v(\delta(X))), v(\delta(Y)))\,\tau(\varphi(\bar{v}^*_x, q), \bar{r}_i\otimes v(\delta(X)))\\ &\quad=\sum_{q\in Q_2}\mathrm{Tr}(\bar{u}_{ij} v^*_x) f(q_V(v(\delta(X))), v(\delta(Y)))(\bar{v}^*_x(\bar{r}_i)\otimes q_V(v(\delta(X))))\\ &\quad=\sum_{q\in Q_2}{}'\,\mathrm{Tr}(\bar{u}_{ij} v^*_x)(\bar{v}^*_x(\bar{r}_i)\otimes v(\delta(Y))).\end{aligned}$$

In summation $\sum'$, q ranges over the set of $q\in Q_2$ satisfying the condition $f(q_V(v(\delta(X))), v(\delta(Y)))\neq 0$. Since $u_{ij}=J_1\otimes(A_{ij}\oplus A_{ji})$, $v_Y=r_j\otimes v(\delta(Y))$, and r_i (resp. $\bar{r}_i$) has the unique non-zero component in i (resp. $6+i$)-component, the coefficient of v_Y in $\bar{v}^*_x(\bar{r}_i)\otimes v(\delta(Y))$ is the $(6+i, j)$-entry of $\bar{v}^*_x$. On the other hand, $\mathrm{Tr}(\bar{u}_{ij} v^*_x)$ is twice the $(i, 6+j)$-entry of v^*_x, which is the complex conjugate of the $(6+i, j)$-entry of $\bar{v}^*_x$. Here notice that all $v^*_x, \bar{v}^*_x$ are symmetric. Hence the coefficient of v_Y in $\tau(\Phi(v_X, v_Y), \bar{v}_X)$ is $\sum_q{}' \frac{1}{2}|\mathrm{Tr}(\bar{u}_{ij} v^*_x)|^2$.

(8.10) We fix $q\in Q_2$ such that $f(q(v(\delta(X))), v(\delta(Y)))\neq 0$. Moreover we fix $x\in C_D(s)$ such that $q=xx'$ for some $x'\in x^Q$. Then Lemma 5.9.2 and the assumption $\delta(X)\neq\delta(Y)$ imply $q\notin\langle L\rangle$, and thus $x\notin L$. By (3.9.2′), there exist some

$m, n \in \{1, \ldots, 6\}$, $l \in \mathbb{Z}$, $\varepsilon_k = \pm 1$ $(1 \leqq k \leqq 6)$ such that

$$\bar{v}_x^* = \frac{\omega^l}{8}\Big(\sum_{k \neq m,\, n} u_{kk} - u_{mm} - u_{nn} - \sum_{\substack{k,\, h \neq m,\, n \\ k < h}} 2\varepsilon_k \varepsilon_h u_{kh} + 2\theta\Big(\prod_{k \neq m,\, n} \varepsilon_k\Big) u_{mn}\Big).$$

Then the entries of $\bar{v}_x^*$ are as follows:

$$\text{the } (k, h+6)\text{-entry} = \begin{cases} \pm\frac{1}{4}\omega^l & (\text{if } k = h \text{ or } k, h \notin \{m, n\}), \\ \pm\frac{1}{4}\omega^l\theta & (\text{if } \{k, h\} = \{m, n\}), \\ 0 & (\text{otherwise}), \end{cases}$$

other entries $= 0$.

(8.11) We separate our proof into four subcases as follows:

Case 1: $i = j$,
Case 2: $i \neq j$, $X \# T_i = Y \# T_i$ and $X \# T_j = Y \# T_j$,
Case 3: $i \neq j$, $X \# T_i = Y \# T_i$ and $X \# T_j \neq Y \# T_j$,
Case 4: $i \neq j$, $X \# T_i \neq Y \# T_i$ and $X \# T_j \neq Y \# T_j$.

Case 1. The assumption implies $|X \cap Y| \neq 0$. By (8.9), $\mathrm{Tr}(u_{ii}\bar{v}_x^*) = \frac{1}{2} \times (\frac{1}{2})^2 = \frac{1}{8}$. Hence by (8.8), (8.9), we have

$$\tau(\Phi(v_X, v_Y), \bar{v}_X) = 16 \times \tfrac{1}{8} v_Y = 2v_Y. \tag{8.11.1}$$

In particular, $\Phi(v_X, v_Y) \neq 0$. By (8.7), we have $|X \cap Y| = 4$. Since $\Phi(v_X, v_Y) \in \langle \bar{v}_Z \rangle$ for $Z = (X \cup Y) - (X \cap Y)$, (5.8.3) implies that $\frac{1}{2}\Phi(v_X, v_Y)$ is $\pm\bar{v}_Z^*$. Hence by (8.11.1) and Lemma 5.8.1, we have $\tau(v_X, v_Y) = \frac{1}{2}\Phi(v_X, v_Y)$, as required.

Case 2. The assumption for this case and the assumption $q(v(\delta(X))) \in \langle v(\delta(Y)) \rangle$ imply that q commutes with the elements of $X \# T_i$ and $X \# T_j$. Hence x commutes with them, and thus by (3.9.2'), $\{m, n\} = \{i, j\}$ in (8.9). By (8.9), we have $\mathrm{Tr}(u_{ij}\bar{v}_x^*) = \frac{1}{2} \times \left|2 \times \frac{\theta}{4}\right|^2 = \frac{3}{8}$, and furthermore we have $\tau(\Phi(v_X, v_Y), \bar{v}_X) = 16 \times (\frac{3}{8}) v_Y = 6v_Y$. In particular, $\Phi(v_X, v_Y) \neq 0$, and so $|X \cap Y| \neq 2$.

Now we show $|X \cap Y| = 0$. By the assumption, we have $X \cap Y \cap T_i = X \cap Y \cap T_j = \emptyset$. Hence if $|X \cap Y| = 4$, then $X \cap T_k = Y \cap T_k$ for each $k \neq i, j$. This implies $\delta(X) = \delta(Y)$, a contradiction.

Hence by (5.8.3), $\tau(\bar{v}_Z, \bar{v}_X) = \pm\theta v_Y$, and $\Phi(v_X, v_Y) = \pm\frac{2}{\theta}\bar{v}_Z$ for $Z = L - (X \cup Y)$. By Lemma 5.1(2), $\tau(v_X, v_Y) = \pm\theta\bar{v}_Z = \frac{1}{2}\Phi(v_X, v_Y)$ as required.

Case 3. By the assumption, q commutes with the element of $X \cap T_i$ and does not commute with the one of $X \cap T_j$. By (3.9.2'), we have $i \in \{m, n\} \not\ni j$. Then by (8.10), we have $\tau(\Phi(v_X, v_Y), \bar{v}_X) = 0$. Hence $\Phi(v_X, v_Y) = 0$ by (8.7).

We show $|X \cap Y| = 2$. By the assumption, we have $|X \cap Y \cap T_i| = 0$, $|X \cap Y \cap T_j| = 1$. Suppose $i = m$. Then $X \# T_n = X \cap T_n = Y \cap T_n = Y \# T_n$, since $q(v(\delta(X))) \in \langle v(\delta(Y)) \rangle$, and q commutes with the elements of T_n. Let $k \neq i, j, m, n$, and set $X \# T_k = \{a\}$. Then $a^q \neq a$, and $Y \# T_k = \{a^q\}$. Hence $|X \cap Y \cap T_k| = 0$, and so we have $|X \cap Y| = 2$. By (5.8.4), $\tau(v_X, v_Y) = 0 = \Phi(v_X, v_Y)$, as required.

Case 4. By the assumption, we have $|X\cap Y\cap T_i|=|X\cap Y\cap T_j|=1$, and q does not commute with the elements of $T_i\cup T_j$. Hence $\{i,j\}\cap\{m,n\}=\emptyset$. The similar argument as in Case 1 implies that $\mathrm{Tr}(\bar{u}_{ij}v_x^*)=\frac{1}{8}$, $\tau(\Phi(v_X,v_Y),\bar{v}_X)=2v_Y$, $|X\cap Y|=4$, and $\tau(v_X,v_Y)=\frac{1}{2}\Phi(v_X,v_Y)$.

Now the proof of (8.3) is complete, and then the algebra product $\tau(24)$ is completely determined.

(II) Construction of the group C

We construct the group C, which is realized as a subgroup of $GL(R\otimes V)$, where $R=R^{(4)}$, and V is the representation space of the extraspecial 2-group 2_+^{1+12}, as in Sect. 5.

(8.12) We start from the group Q_∞, the extraspecial 2-group 2_+^{1+12}. We set $Z(Q_\infty)=\langle s_\infty\rangle$, and $\bar{Q}=Q_\infty/\langle s_\infty\rangle$. We regard $\bar{Q}$ as a subgroup of $\mathrm{Aut}(Q_\infty)$. Let ζ_∞ be an element of $\mathrm{Aut}(Q_\infty)$ of order 3 such that its action on $\bar{Q}$ is fixed point free. That is,

$$\bar{q}^{\zeta_\infty^{i+j}}=\bar{q}^{\zeta_\infty}\bar{q}^{\zeta_\infty^i}\quad\text{for any } i\neq j\in\{0,1,2\}.$$

It is easily proved that such ζ_∞ is uniquely determined up to conjugacy in $\mathrm{Aut}(Q_\infty)$.

(8.13) Set $\mathbb{F}_4=\{0,1,\lambda,\lambda^2\}$, $\lambda^3=1$, $1+\lambda+\lambda^2=0$. By using ζ_∞, we define the operation of $\mathbb{F}_4$ on $\bar{Q}$:

$$\lambda^i\bar{q}=\bar{q}^{\zeta_\infty^i}\quad\text{for } i=0,1,2,$$
$$0\cdot\bar{q}=1,\quad\text{the unit of }\bar{Q}.$$

Then it is easily verified that $\bar{Q}$ can be regarded as a vector space over $\mathbb{F}_4$ of dimension 6.

(8.14) We denote by κ the $\mathrm{Aut}(Q_\infty)$-invariant orthogonal form on $\bar{Q}$, that is, $\kappa(\bar{q})=q^2$ for $\bar{q}\in\bar{Q}$. Here q is an element of the inverse image of $\bar{q}$ in Q_∞, and $\langle s_\infty\rangle$ is regarded as the 2-elements field, and is also regarded as a subfield of $\mathbb{F}_4$. Let $(\cdot,\cdot)$ be the corresponding symplectic form, that is,

$$\kappa(\bar{q}\bar{q}')=\kappa(\bar{q})+\kappa(\bar{q}')+(\bar{q},\bar{q}')\quad\text{for } \bar{q},\bar{q}'\in\bar{Q}.$$

Then we define $\langle\cdot,\cdot\rangle$ as follows:

$$\langle\bar{q},\bar{q}'\rangle=(\bar{q},\bar{q}')+\lambda(\bar{q},\bar{q}'^{\zeta_\infty})+\lambda^2(\bar{q},\bar{q}'^{\zeta_\infty^2}),$$

for $\bar{q},\bar{q}'\in\bar{Q}$. Then it is easily verified that $\langle\cdot,\cdot\rangle$ is a non-degenerate hermitian form on $\bar{Q}$ over $\mathbb{F}_4$. In fact, by direct calculations, we have $\langle\bar{q},\bar{q}'\rangle+\langle\bar{q}',\bar{q}\rangle=(\bar{q},\bar{q}')$. Since $(\cdot,\cdot)$ is non-degenerate, $\langle\cdot,\cdot\rangle$ is so. Clearly $C_{\mathrm{Aut}(Q_\infty)}(\zeta_\infty)$ preserves $\langle\cdot,\cdot\rangle$ as a linear transformation group on $\bar{Q}$, and is isomorphic to $U(6,2)$. Moreover $N_{\mathrm{Aut}(Q_\infty)}(\zeta_\infty)\simeq\Gamma U(6,2)$.

(8.15) The group $C_{\mathrm{Aut}(Q_\infty)}(\zeta_\infty)$ has a subgroup H of index 2, which is isomorphic to $SU(6,2)$. The group $SU(6,2)$ is generated by a unique conjugacy class of 3-transpositions which is the class of unitary transvections. Let T be such a class of 3-transpositions of H, and let X be a subgroup of H, isomorphic to $\Omega^-(6,3).2$ and generated by a subset of T. Then by [9: (16.1.12), (16.1.20)], X is

uniquely determined up to conjugacy in $N_{\mathrm{Aut}(Q_\infty)}(\zeta_\infty)$. Let X_0 be the normalizer of X in $N_{\mathrm{Aut}(Q_\infty)}(\zeta_\infty)$. Then $|X_0 : X| = 2$.

(8.16) For $\bar{q} \in \bar{Q}$ with $\langle \bar{q}, \bar{q} \rangle = 0$, we denote by $t(\bar{q})$ the unitary transvection with respect to $\bar{q}$. Set $\bar{Q}_2$ be the set of $\bar{q} \in \bar{Q}$ with the property $t(\bar{q}) \in \bar{T}$, the image of T in $C_\infty / \langle s_\infty \rangle$. Let $Q_{\infty,2}$ be the inverse image of $\bar{Q}_2$ in C_∞. Then $Q_{\infty,2}$ is a conjugacy class of C_∞.

(8.17) The group Q_∞ has a unique non-linear irreducible representation over $\mathbb{Q}$ of dimension 2^6. Hence we may regard $Q_\infty \subset GL(2^6, \mathbb{Q})$. Let $x \in \mathrm{Aut}(Q_\infty)$. Then there exists some S_x of $GL(2^6, \mathbb{Q})$ such that $S_x q S_x^{-1} = q^x$ for all $q \in Q_\infty$. By irreduciblity of Q_∞, S_x is uniquely determined up to constant. Hence this correspondence gives an embedding $\mathrm{Aut}(Q_\infty) \subset PGL(2^6, \mathbb{Q})$. Then clearly $O_2(\mathrm{Aut}(Q_\infty))$ coincides with the image of Q_∞ in $PGL(2^6, \mathbb{Q})$. Let Y be a subgroup of $GL(2^6, \mathbb{Q})$ generated by the elements of determinant ± 1 of the inverse image of $\mathrm{Aut}(Q_\infty)$. Then $Y \cap Z(GL(2^6, \mathbb{Q})) = \langle \pm I \rangle$, and Y contains Q_∞. Clearly $Y / \langle s_\infty \rangle$ is isomorphic to $\mathrm{Aut}(Q_\infty)$. Hence we identify them, and then $Y / \langle s_\infty \rangle$ contains subgroups X, X_0 as in (8.15). Set X_∞, C_∞ be the inverse images of X, X_0 in Y. Then $X_\infty \subset GL(2^6, \mathbb{Q})$, and $C_\infty \simeq 2^{1+12}.3.\Omega^-(6,3).2^2$. We denote an element of the inverse image of ζ_∞ by the same symbol ζ_∞.

We may assume this representation space of Q_∞ is the space V. Then $C_\infty \subset GL(V)$.

(8.18) Next we consider C_0 defined in Sect. 3. The group C_0 is regarded as a subgroup of $GL(R)$, and we may assume $N(6) = \langle D(6) \rangle$ is a subgroup of C_0. We denote by s_0, z_0 the elements of $Z(C_0')$ instead of s, z. We denote by Ω the group $\Omega^-(6,3).2^2$, the unique extention of $\Omega^-(6,3)$ by the group 2^2 containing two 3-transpositions groups. Then there exist surjective homomorphisms:

$$\Psi_\infty : C_\infty \to \Omega,$$
$$\Psi_0 : C_0 \to \Omega.$$

Here we can choose them so that $\Psi_\infty(X_\infty) = \Psi_0(N(6))$.

(8.19) We define the group C_1 as follows:

$$C_1 = \{(g, h) \mid g \in C_0, h \in C_\infty, \Psi_0(g) = \Psi_\infty(h)\}.$$

Then C_1 acts the space $R \otimes V$ by $(g,h)(r \otimes v) = g(r) \otimes h(v)$ for $(g,h) \in C_1$, $r \in R$, and $v \in V$. The kernel of this action is the subgroup $\langle (s_0, s_\infty) \rangle$. Hence we may regard the group $C_1 / \langle (s_0, s_\infty) \rangle$ is a subgroup of $GL(R \otimes V)$.

This is a group isomorphic to C. But we omit the proof, since it seems that this fact is not essential in this paper.

(8.20) Finally we present some elements of C, and the definition of $(B(24), \tau(24))$, by using the notation for C_1. In the following, the left-hand sides indicate the notation used in the previous sections. In the right-hand sides, for elements of $C_1 / \langle (s_0, s_\infty) \rangle$, we present elements of their inverse images in C_1.

$$s = (s_0, 1) (\text{or } (1, s_\infty)), \quad z = (z_0, 1), \quad \zeta = (1, \zeta_\infty),$$
$$C_D(s) = \{(d_0, d_\infty) \mid d_0 \in D(6), d_\infty \in D_\infty, \Psi_0(d_0) = \Psi_\infty(d_\infty)\},$$
$$Q_2 = \{(1, q) \mid q \in Q_{\infty,2}\}.$$

Concerning $B(24)$, the spaces B_*, B_1 are constructed by using R, V, and C_0. The group C_1 acts on them by

$$(g, h)(b) = g(b), \qquad (g, h)(r \otimes v) = g(r) \otimes h(v),$$

where $g \in C_0$, $h \in C_\infty$, $b \in B_*$, $r \in R$, and $v \in V$.

The basis $\varphi(v^*_{d_0}, q)$ of B_2 is defined for $(d_0, d_\infty) \in C_D(s)$ and $q \in Q_2$ with $t(q) = d_\infty$. Here $\varphi(\omega v^*_d, q) = \omega\varphi(v^*_d, q)$ and $\varphi(v^*_d, qs_\infty) = -\varphi(v^*_d, q)$. The group C_1 acts on B_2 by

$$(g, h)(\varphi(v^*_d, q)) = \varphi(g(v^*_d), hqh^{-1}).$$

The structure of the algebra is defined by using the formulas indicated in (5.14) and (8.3).

References

1. Aschbacher, M.: Overgroups of Sylow subgroups in sporadic groups. Mem. Am. Math. Soc. **343** (1986)
2. Aschbacher, M., Seitz, G.: On groups with a standard component of known type. Osaka J. Math. **13**, 439–482 (1976)
3. Benard, M.: Characters and Schur indices of the unitary reflection group $[3\,2\,1]^3$. Pacific J. Math. **58-2**, 309–321 (1975)
4. Bender, H.: Transitive Gruppen gerader Ordnung, in denen jede Involutionen genau einen Punkt festläßt. J. Algebra **17**, 527–554 (1971)
5. Burnside, W.: The theory of finite groups, 2nd edition. New York: Dover 1955
6. Conway, J.H.: Three lectures on exceptional groups. In: Finite simple groups, Powell, M.B., Higman, G. (eds.), pp. 215–246. New York: Academic Press 1971
7. Conway, J.H., Curtis, R.T., Norton, S.P., Parker, R.A., Wilson, R.A.: An ATLAS of finite groups. Oxford: Clarendon Press 1985
8. Conway, J.H., Sloane, N.J.: The Coxeter-Todd lattice, the Mitchell group, and related sphere packings. Math. Proc. Camb. Philos. Soc. **10**, 77–94 (1969)
9. Fischer, B.: Finite groups generated by 3-transpositions. Univ. of Warwick, Preprint, 1969
10. Griess, R.: The friendly giant. Invent. Math. **69**, 1–102 (1982)
11. Kitazume, M.: A 2-local geometry for the Fischer group F_{24}. J. Fac. Sci., Univ. Tokyo, Sect. I A **31-1**, 59–79 (1984)
12. Linsey, J.H., II: On a six dimensional projective representation of $PSU_4(3)$. Pacific J. Math. **36-2**, 407–425 (1971)
13. Parrot, D.: Characterizations of the Fischer groups I, II, III. Trans. Amer. Math. Soc. **265-2** (June) 303–347 (1981)
14. Phan, K.W.: A characterization of the finite simple group $U_4(3)$. Aust. J. Math. **10**, 77–94 (1969)
15. Smith, S.D.: Nonassociative commutative algebras for triple covers of 3-transposition groups. Michigan J. Math. **24**, 273–287 (1977)
16. Solomon, R.: Finite groups with Sylow 2-subgroups of type $\Omega(7, q)$, $q \equiv \pm 3 \pmod 8$. J. Algebra **28**, 174–181 (1974)
17. Tits, J.: On R. Griess' "Friendly giant". Invent. Math. **78**, 491–499 (1984)
18. Todd, J.A.: A representation of the Mathieu group M_{24} as a collineation group. Ann. Mat. Pura Appl., IV. Ser. **71**, 199–238 (1966)
19. Wilson, R.A.: On maximal subgroups of the Fischer group Fi_{22}. Math. Proc. Camb. Philos. Soc. **95**, 197–222 (1984)

Oblatum 10-V-1986

Invent. math. 88, 319-340 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

New volume ratio properties for convex symmetric bodies in $\mathbb{R}^n$

J. Bourgain and V.D. Milman

Institut des Hautes Etudes Scientifiques, 35, Route de Chartres, F-91440 Bures-sur-Yvette, France

1. Introduction and summary of results

For $x, y \in \mathbb{R}^n$, denote $|x| = \left(\sum_{i=1}^n x_i^2\right)^{1/2}$ the Euclidean norm and $\langle x, y\rangle = \sum x_i y_i$ the inner product. Let $B_n = \{x \in \mathbb{R}^n;\ |x| \leqq 1\}$ be the euclidean ball and $S_{n-1} = \{x \in \mathbb{R}^n;\ |x| = 1\}$ the euclidean sphere equipped with invariant normalized measure μ_{n-1}.

Let K stand for a convex centrally symmetric bounded and absorbing set in $\mathbb{R}^n$ and let $K^0 = \{x \in \mathbb{R}^n;\ \sup_{y \in K} |\langle x, y\rangle| \leqq 1\}$ be the polar body of K. Notice that if $\| \ \|$ is the norm on $\mathbb{R}^n$ induced by K, then K^0 coincides with the unit ball of the dual space E^*, where $E = \mathbb{R}^n$, $\| \ \|$.

Consider first the expression $\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0$, where $\mathrm{Vol} = \mathrm{Vol}_n$ is the standard n-dimensional volume of subsets of $\mathbb{R}^n$. This expression is clearly an affine invariant: for any invertible map $u: \mathbb{R}^n \to \mathbb{R}^n$, we have

$$\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 = \mathrm{Vol}\, u(K) \cdot \mathrm{Vol}\, [u(K)]^0.$$

K. Mahler ([Ma], 1939) proved that

$$4^n (n!)^{-2} \leqq \mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 \leqq 4^n \tag{1.1}$$

and he applied this result to geometric number theory (see also [L], p. 106). Several improvements of the inequalities (1.1) have been produced since then. The precise upper bound was obtained by L. Santalo [Sa, 1]. He proved the inequality

$$\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 \leqq (\mathrm{Vol}\, B_n)^2 \tag{1.2}$$

which obviously becomes equality for $K = B_n$. The reader may find an elegant proof of (1.2) in [S_1]. Mahler also conjectured that

$$\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 \geqq 4^n (n!)^{-1} \tag{1.3}$$

which means that the minimum is attained for $K = \{x \in \mathbb{R}^n;\ \sum |x_i| \leqq 1\}$ (then K is the unit ball of the normed space l_n^1 and $K^0 = \{x \in \mathbb{R}^n;\ \max |x_i| \leqq 1\}$ is the symmetric unit cube). He proved this fact for $n = 2$.

Several incomplete proofs of (1.3) appear in the literature. R. Bambah [Ba] used the maximal volume ellipsoid to get the minoration $n^{-n/2}$ $(\mathrm{Vol}\, B)^2$ in (1.3). More recently, using progress in the geometry of Banach spaces, this lower bound has been improved up to $(c\, n \log n)^{-n}$ (see [G-R]).

The first result of this paper is the following converse to Santalo's inequality (1.2).

Theorem 1. *There exists an absolute constant $c>0$ (thus independent of the dimension n) such that*

$$\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 \geqq c^n (\mathrm{Vol}\, B)^2. \tag{1.4}$$

In Sect. 6 of this paper, we discuss some applications of the above theorem to geometrical number theory as well as further developments related to the methods involved. Some natural problems find their solution through Theorem 1.

The approach to Theorem 1 makes essential use of methods and results from geometry of Banach spaces (the isoperimetric inequality on the sphere, random subspaces, the Banach-Mazur distance, the Rademacher projection, the l-ellipsoid etc. ...). Let us at this point recall an important notion in local Theory of Normed Spaces (the theory about asymptotic behavior of finite dimensional normed spaces for large dimension). The "Banach-Mazur distance" $d(E, l_n^2)$, $n=\dim E$, between a space E (with unit ball K) and the euclidean space l_n^2 is given by

$$d_E = \inf\{\|u\|\, \|u^{-1}\|;\ u: l_n^2 \to E \text{ is a linear isomorphism}\}$$

and where $\|u\|$, $\|u^{-1}\|$ are the corresponding operator norms. Thus d_E is by definition a linear invariant. Geometrically, it means thus that there exists an affine position $\tilde{K}=u(K)$ of K such that for any $x\in\mathbb{R}^n$

$$a^{-1}|x| \leqq \|x\|_{\tilde{K}} \leqq b|x| \tag{1.5}$$

holds, where $a, b>0$ satisfy $a\cdot b = d_E$. However, inequalities of type (1.5) will also appear when $a\cdot b$ does not equal d_E. Thus if $d_K \equiv a\cdot b$ for the smallest values a, b verifying (1.5), then

$$d_E = \inf\{d_{\tilde{K}} | \tilde{K} \text{ is an affine image of } K\}.$$

The proof of Theorem 1 permits to make the following remark on Mahler's conjecture about the optimal value of the constant c in (1.4). Denote

$$s(K) = \left\{\frac{\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0}{(\mathrm{Vol}\, B)^2}\right\}^{1/n} \quad \text{(the Santalo characteristic of } K)$$

and

$$s_n = \inf\{s(K);\ \dim K = n\}.$$

Then, given $\varepsilon>0$ there is a sequence K_n $(n=1, 2, \ldots)$ of bodies, $\dim K_n = n$, satisfying

$$d_{K_n} \leqq \alpha(\varepsilon) \quad \text{and} \quad s(K_n) < (1+\varepsilon)\, s_n$$

where α is some function on $]0, 1[$.

Another bi-product of the proof of Theorem 1 is a weak version of Santalo's sharp inequality (1.2): there exists an absolute constant C such that $S(K) \leqq C$ for any convex symmetric body K in $\mathbb{R}^n$ $(n=1, 2, \ldots)$.

Let $\mathscr{E}=\mathscr{E}_K$ be the ellipsoid of maximal volume contained in K. This ellipsoid, called the F. John ellipsoid (see [J]), has remarkable properties and plays an essential role in several problems of geometry and local Theory of Banach Spaces. The volume ratio v.r.E of a normed space E is defined as the volume ratio of its unit ball K,

$$\text{v.r.}\, E \equiv \text{v.r.}\, K = \left(\frac{\text{Vol}\, K}{\text{Vol}\, \mathscr{E}}\right)^{1/n} \qquad \text{where } n=\dim E.$$

Say that a family $\mathscr{F}$ of finite dimensional normed spaces has finite volume ratio if there exists a constant C satisfying

$$\text{v.r.}\, E \leqq C \qquad \text{for all } E \in \mathscr{F}.$$

An interesting example of such a family $\mathscr{F}$ is given by $\mathscr{F}=\{l_n^1 | n=1, 2, \ldots\}$. An easy computation shows indeed that

$$\text{v.r.}\, K \leqq \text{constant (independent of } n) \qquad \text{if } K=\left\{x \in \mathbb{R}^n \,\middle|\, \sum_{i=1}^{n} |x_i| \leqq 1\right\}.$$

Similar computations have been done for l_n^p-spaces and trace-classes C_n^p $(p \leqq 2)$, see [Ro] and [S2], in order to obtain an explicit expression for the volume ratio and in particular, to prove its boundedness. The notion of volume ratio plays an important role in the work of B.S. Kashin, S. Szarek, N. Tomczak. It turns out indeed that finite volume ratio of a family $\mathscr{F}$ has many interesting consequences for this family (see in particular [Sz], [Sz-T]).

Definition. A normed space E, $\| \ \|$ is said to have cotype 2 provided there is some constant C such that

$$C \operatorname*{Ave}_{\varepsilon_i = \pm 1} \left\| \sum \varepsilon_i x_i \right\| \geqq \left(\sum \|x_i\|^2\right)^{1/2}$$

whenever $\{x_i\}$ is a finite sequence of vectors in E. The smallest such constant $C_2(E)$ is called the cotype-2 constant of E.

The fact that $C_2(l_n^1)$ is uniformly bounded is just a consequence of Khintchine's inequality. Also the other families considered above fall in the category of "cotype-2 spaces".

Theorem 2.

$$\text{v.r.}\, E \leqq C\, C_2(E) \cdot [\log C_2(E)]^4. \tag{1.6}$$

The estimate (in particular 1.6) of the volume ratio in terms of the cotype-2 constant of the space gives an affirmative solution to a problem raised in [Pel] and [Sz-T] in the mid 70's and well studied by several people over the recent years. In [Pel], a rather extensive discussion of this problem appears.

This article is written in the following way. In Sect. 2, we develop the theory of euclidean sections of convex symmetric bodies which plays an important role in the proofs of both Theorems 1 and 2. These results are in some

sense relatives of Dvoretzky's theorem (see [FLM]). Section 3 contains some elementary volume estimates. The proof of Theorem 1 (resp. Theorem 2) is given in Sect. 4 (resp. Sect. 5). In Sect. 6, various applications are presented: consequences to geometric number theory (lattice point counting), mixed volumes, etc. In Sect. 7, comments are given and subsequent work mentioned. As was said before, concepts from local theory of Banach spaces play an essential role and also, of course, some results in this theory. In order to make the paper more accessible to non-experts, these prerequisites with quick proofs are summarized in a separate section (8) - Appendix on Local Theory. (See also [M-S] for more details.)

In fact, both Theorems 1 and 2 are obtained as a corollary to a more general new approach to inequalities on convex bodies, different from the classical variational and symmetrization arguments. This approach involves the *linear structure* of the body. One replaces with "small" loss of codimension the body by an euclidean *isomorph*, with essential volume preservation.

The main results of the paper were announced in [B-M]. The authors are grateful to G. Pisier for simplification of the original arguments. Throughout the paper, we use the letters $c>0$, $C<\infty$ for different numerical constants.

2. The M_*-result and its application to euclidean sections

For a normed space $E=\mathbb{R}^n$, $\|\ \|$, define the following quantities

$$M_E=\int_{S_{n-1}} \|x\|\, \mu_{n-1}(dx) \qquad M_{E^*}=\int_{S_{n-1}} \|x\|_*\, \mu_{n-1}(dx)$$

which are known to be of importance in finite dimensional Banach space theory (see [FLM] for instance).

We will make repeated use of the following fact.

Proposition 2.1. *For $0<\lambda<1$, there is a subspace F of $\mathbb{R}^n$, $\dim F\geqq \lambda n$, such that*

$$\|x\|\geqq c(1-\lambda)\, M_{E^*}^{-1}|x| \quad \text{for } x\in F. \tag{2.2}$$

To be useful, (2.2) will have to be combined with an information on M_{E^*}. It should also be mentioned that a more elaborated argument than given here permits to replace the factor $1-\lambda$ by $(1-\lambda)^{1/2}$ (see [P-T]). This is the "right" dependence.

We derive (2.2) from the isoperimetric inequality on the sphere. Let d be the metric on S_{n-1} given by the geodesic distance, i.e. $0\leqq d(x,y)<\pi$ is the angle of the lines $[0,x]$ and $[0,y]$. For A a subset of S_{n-1} and $\varepsilon>0$, denote $A_\varepsilon=\{x\in S_{n-1};\ d(x,A)\leqq\varepsilon\}$. As a consequence of the isoperimetric inequality on S_{n-1}, if follows that if $\mu_{n-1}(A)\geqq 1/2$, then

$$\mu(A_\varepsilon)\geqq \gamma_n \int_{-\pi/2}^{\varepsilon} \cos^{n-2} t\, dt \quad \text{where} \quad \gamma_n=\Gamma(n/2)/\Gamma(1/2)\,\Gamma\left(\frac{n-1}{2}\right)\sim\sqrt{n}. \tag{2.3}$$

As we will show, integration on the Grassmannian $G_{n,k}$ of k-dimensional subspaces enables us to derive from (2.3) the following

Lemma 2.4. *Let $A \subset S_{n-1}$, $\mu(A) \geqq 1/2$ and $\tau > 0$. Then there is a subspace F of $\mathbb{R}^n$, $\dim F \geqq (1 - C\tau)n$ satisfying*

$$(F \cap S_{n-1}) \subset A_{(\pi/2)-\tau}$$

where C is a constant.

Proof of Proposition 2.1. If $A = \{x \in S_{n-1} \mid \|x\|_* \leqq 2M_{E^*}\}$, then clearly $\mu(A) \geqq 1/2$. Let F be the subspace given by (2.4). Then for $y \in F \cap S_{n-1}$, there is $x \in A$ such that

$$\sin \tau \leqq |\langle x, y \rangle| \leqq \|x\|_* \, \|y\| \leqq 2M_{E^*} \cdot \|y\|.$$

Proof of Lemma 2.4. Fix $k = [\lambda n]$, $1/2 < \lambda < 1$. Recall that

$$\mu_{n-1}(A) = \int_{G_{n,k}} \mu_{k-1}(A \cap F)\, \sigma(dF)$$

where σ is the normalized measure on $G_{n,k}$ invariant under the orthogonal group. Since

$$\mu_{n-1}(A_{\pi/4}) \geqq \gamma_n \int_{-\pi/2}^{\pi/4} \cos^{n-2} t \, dt = 1 - \gamma_n \int_0^{\pi/4} \sin^{n-2} t \, dt$$

there is $F \in G_{n,k}$ satisfying

$$\mu_{k-1}(A_{\pi/4} \cap F) \geqq 1 - \gamma_n \int_0^{\pi/4} \sin^{n-2} t \, dt. \tag{2.5}$$

Assume $\tau > 0$ satisfies

$$\gamma_k \int_0^{(\pi/4)-\tau} \sin^{k-2} t \, dt > \gamma_n \int_0^{\pi/4} \sin^{n-2} t \, dt. \tag{2.6}$$

Since for $x \in F \cap S_{n-1}$ we have

$$\mu_{k-1}(B_{(\pi/4)-\tau}(x)) = \gamma_k \int_0^{(\pi/4)-\tau} \sin^{k-2} t \, dt$$

where

$$B_\varepsilon(x) = \{y \in F \cap S_{n-1} \mid d(x, y) < \varepsilon\} \tag{2.7}$$

it follows from (2.5)-(2.7) that

$$A_{\pi/4} \cap B_{(\pi/4)-\tau}(x) \neq \emptyset \Rightarrow x \in A_{(\pi/2)-\tau}.$$

It remains thus to analyse condition (2.6) on τ. The reader will easily check that (2.6) holds for $\tau \sim 1 - k/n$. This proves the lemma.

Next we consider bounds on M_*. The proofs of Theorem 2 and Theorem 1 are based on the following estimates. For the first, cf. [D-M-T].

Proposition 2.8. *If the euclidean ball B_n coincides with the maximal volume ellipsoid of K (= unit ball of E), then for every subspace F of E, $\dim F = k$ we have*

$$M_{F^*} \leqq C \cdot C_2(E)(1 + \log d_E)^{3/2} \sqrt{n/k}. \tag{2.8}$$

Proposition 2.9. *Every normed space E has an isometric realization X (which means that K is replaced by an affine image $\tilde{K}=u(K)$) such that*

$$M_X \cdot M_{X^*} \leqq C(1+\log d_E)^2. \tag{2.9}$$

The operator u appearing in Proposition 2.9 determines the so-called l-ellipsoid. This ellipsoid seems essential in the proof of Theorem 1. Notice that (2.8), (2.9) could be improved by writing $1+\log d_E$ instead of the corresponding power. Short proofs of (2.8), (2.9) will be given in the Appendix.

Proposition 2.1 and Lemma 2.8 give immediately

Lemma 2.10. *Under the hypothesis of Lemma 2.8, there exists for $1/2<\lambda<1$ a subspace F of E satisfying* $\dim F>\lambda \dim E$ *and*

$$c(1-\lambda)\, C_2(E)^{-1}(1+\log d_E)^{-3/2}|x| \leqq \|x\| \leqq |x| \quad \text{for } x\in F. \tag{2.10}$$

In other words, $d_F \leqq C(1-\lambda)^{-1}\ C_2(E)(1+\log d_E)^{3/2}$ and this estimate is obtained with respect to the maximal volume ellipsoid of K.

Lemma 2.11. *Let E be an n-dimensional space and $0<\lambda<1$. Then there exists a subspace F of E and a quotient space G of F with*

$$\dim G \geqq \lambda n \quad \text{and} \quad d_G \leqq C(1-\lambda)^{-2}(1+\log d_E)^2. \tag{2.11}$$

Proof. Apply twice Proposition 2.1. First F subspace of E is found, $\dim F \geqq \lambda_1 n$ and

$$\|x\| \geqq c(1-\lambda_1)\, M_{E^*}^{-1}|x| \quad \text{for } x\in F. \tag{2.12}$$

For $\lambda_1>1/2$, we may write $M_{(F^*)^*}=M_F \leqq \sqrt{2}M_E$ (the last inequality is explained in the Appendix).

Applying Proposition 2.1 to F^* now gives a subspace G of F^* such that

$$\|x\|_{F^*} \geqq c(1-\lambda_1)\, M_E^{-1}|x| \quad \text{for } x\in G\subset F^*. \tag{2.13}$$

But (2.12) implies

$$\|x\|_{F^*} \leqq C(1-\lambda_1)^{-1} M_{E^*}|x| \quad \text{if } x\in F^*. \tag{2.14}$$

It follows from (2.12), (2.13)

$$d_G \leqq C(1-\lambda_1)^{-2} M_E \cdot M_{E^*}.$$

If we replace E by its isometric representation X given by Lemma 2.9, then

$$d_G \leqq C(1-\lambda_1)^{-2}(1+\log d_E)^2$$

(the estimate is obtained with respect to the l-ellipsoid).

Let $\lambda=\lambda_1^2$. Then G^* is a quotient space of the subspace F of E with the required properties.

Remark. Using an iteration procedure, both Lemmas 2.10 and 2.11 lead to estimates on the euclidean distance which do not depend anymore on the

euclidean distance d_E of the initial space E. In particular, Lemma 2.11 yields the "subspace of quotient theorem" [Mi, 1], stating that G appearing in (2.11) has an euclidean distance only depending on the relative codimension. We do not need these improvements however and will derive directly Theorems 1 and 2 from (2.11), (2.10) respectively.

3. A geometric inequality

The following lemma (which simplifies the proofs of Theorem 1, Theorem 2 given in [B-M]) was kindly pointed out to the authors by G. Pisier.

Lemma 3.1. *Assume K a convex symmetric body in $\mathbb{R}^n$, F a k-dimensional subspace of $\mathbb{R}^n$ and $F^\perp$ the orthogonal complement of F. Then we have*

$$\binom{n}{k}^{-1} \mathrm{Vol}_k(K\cap F)\,\mathrm{Vol}_{n-k}[P_{F^\perp}(K)] \leqq \mathrm{Vol}_n K \leqq \mathrm{Vol}_k(K\cap F)\,\mathrm{Vol}_{n-k}[P_{F^\perp}(K)]. \quad (3.1)$$

This inequality appears in [R-S]. *The proof is straightforward and we include it here.*

Proof of Lemma 3.1. Obviously

$$\mathrm{Vol}_n K = \int_{P_{F^\perp}(K)} \mathrm{Vol}_k(K\cap(x+F))\,dx \quad (3.2)$$

and by the Brunn-Minkowski theorem

$$\mathrm{Vol}_k(K\cap(x+F)) \leqq \mathrm{Vol}_k(K\cap F)$$

from where the upper-estimate in (3.1). To get the lower estimate, denote $\|\ \|_1$ the norm on $F^\perp$ induced by $P_{F^\perp}(K)$ and notice that, by convexity, for $x\in P_{F^\perp}(K)$

$$\mathrm{Vol}(K\cap(x+F)) \geqq (1-\|x\|_1)^k\,\mathrm{Vol}(K\cap F).$$

Thus the right member of (3.2) dominates

$$\mathrm{Vol}(K\cap F)\,\mathrm{Vol}[P_{F^\perp}(K)]\,(n-k)\int_0^1 t^k(1-t)^{n-k-1}\,dt$$

from where the lemma.

Recall that

$$\mathrm{Vol}\,B_n = \frac{2\pi^{n/2}}{\Gamma\left(\frac{n}{2}+1\right)}; \quad (\mathrm{vol}\,B_n)^{1/n} = \frac{c_n}{\sqrt{n}} \quad \text{where } c_n \to c \text{ for } n\to\infty. \quad (3.3)$$

Assume

$$a^{-1}|x| \leqq \|x\|_K \leqq b|x|. \quad (3.4)$$

Then clearly, in the situation of Lemma 3.1, as a consequence of (3.4)

$$b^{-1}(B_n\cap F)\subset(K\cap F)\subset a(B_n\cap F). \quad (3.5)$$

Similarly,

$$b^{-1}(B_n \cap F^{\perp}) \subset P_{F^{\perp}}(K) \subset a(B_n \cap F^{\perp}). \tag{3.6}$$

Applying (3.1) and (3.5), (3.6), it follows

Lemma 3.7.

$$\binom{n}{k}^{-1} b^{-n+k} \left(\frac{c_{n-k}}{\sqrt{n-k}}\right)^{n-k} \operatorname{Vol}(K \cap F) \leqq \operatorname{Vol} K \leqq a^{n-k} \left(\frac{c_{n-k}}{\sqrt{n-k}}\right)^{n-k} \operatorname{Vol}(K \cap F) \tag{3.7'}$$

$$\binom{n}{k}^{-1} b^{-k} \left(\frac{c_k}{\sqrt{k}}\right)^{k} \operatorname{Vol}[P_{F^{\perp}}(K)] \leqq \operatorname{Vol} K \leqq a^{k} \left(\frac{c_k}{\sqrt{k}}\right)^{k} \operatorname{Vol}[P_{F^{\perp}}(K)]. \tag{3.7''}$$

Remark. In general, (3.1) is best possible, but can be improved in certain cases, for instance for uniformly convex bodies.

4. Proof of Theorem 1

By the affine invariance mentioned in the introduction and since trivially $s(K) \leqq d_K$, following estimate on the Santalo characteristic of K holds

$$s(K) \leqq d_E \quad \text{where } E = \mathbb{R}^n, \ \| \ \|_K. \tag{4.1}$$

Fix a positive integer N. Define for $n = 1, \ldots, N$ and $2 \leqq t \leqq N^{1/2}$ the class $\mathscr{C}_n(t)$ of convex symmetric bodies K in $\mathbb{R}^n$ such that $d_E \leqq t$ where $E = \mathbb{R}^n$, $\| \ \|_K$. Let further

$$\sigma(t) = \inf\{s(K) \mid K \in \mathscr{C}_n(t), \ n \leqq N\}.$$

Our purpose is thus to get a uniform lower estimate on $\sigma(t)$. Fix $K \in \mathscr{C}_n(t)$, $n \leqq N$. For $n/2 \leqq k \leqq n$, apply the subspace of quotient result (2.11) in order to obtain subspaces $G \subset F \subset \mathbb{R}^n$ satisfying

$$\dim F = k', \quad \dim G = k$$

and

$$K_1 \in \mathscr{C}_k\left(C\left(\frac{n}{n-k} \log t\right)^2\right) \quad \text{where } K_1 = P_G(K \cap F). \tag{4.2}$$

Apply twice (3.1) to get

$$\begin{aligned} \operatorname{Vol} K &\geqq \binom{n}{k'}^{-1} \operatorname{Vol}(K \cap F) \operatorname{Vol} P_{F^{\perp}}(K) \\ &\geqq \binom{n}{k'}^{-1} \binom{k'}{k}^{-1} \operatorname{Vol} K_1 \cdot \operatorname{Vol}(K \cap F \cap G^{\perp}) \cdot \operatorname{Vol} P_{F^{\perp}}(K). \end{aligned} \tag{4.3}$$

Similarly

$$\begin{aligned} \operatorname{Vol}(K^0) &\geqq \binom{n}{k'}^{-1} \operatorname{Vol}[P_F(K^0)] \operatorname{Vol}(K^0 \cap F^{\perp}) \\ &\geqq \binom{n}{k'}^{-1} \binom{k'}{k} \operatorname{Vol} K_1^0 \cdot \operatorname{Vol}[P_{F \cap G^{\perp}}(K^0)] \cdot \operatorname{Vol}(K^0 \cap F^{\perp}). \end{aligned} \tag{4.4}$$

Multiplying (4.3), (4.4), we find

$$s(K) \geqq \left(c\frac{n-k}{n}\right)^{2\frac{n-k}{n}} s(K_1)^{k/n} s(K \cap F \cap G^{\perp})^{\frac{k'-k}{n}} s(P_{F^{\perp}}(K))^{\frac{n-k'}{n}} \cdot \left\{\frac{\operatorname{Vol} B_k \cdot \operatorname{Vol} B_{k'-k} \cdot \operatorname{Vol} B_{n-k'}}{\operatorname{Vol} B_n}\right\}^{2/n}. \tag{4.5}$$

Using the euclidean distance estimate (4.1), which is inherited by subspaces and quotients, (4.5) implies

$$s(K) \geqq \left(c\frac{n-k}{n}\right)^{2\frac{n-k}{n}} \left(\frac{1}{t}\right)^{\frac{n-k}{n}} \sigma\left(C\left(\frac{n}{n-k}\log t\right)^2\right)^{k/n}. \tag{4.6}$$

Choose k such that $\dfrac{n}{n-k} \approx (\log t)^2$. Since $\sigma \leqq 1$, (4.6) becomes

$$s(K) \geqq c^{\frac{1}{\log t}} \sigma(C(\log t)^6) \tag{4.7}$$

and we therefore found that

$$\sigma(t) \geqq c^{\frac{1}{\log t}} \sigma(C(\log t)^6)$$

(where the constants c, C are independent of N).

It is an easy exercise now to deduce from (4.7) that σ is bounded by an absolute constant from below. This completes the proof of Theorem 1.

Remarks. 1. It is clear that reversing the previous procedure and using the (3.1) upper estimate, an absolute upperbound on $s(K)$ is obtained.

2. As a by-product of the reasoning made above, we get the following:

Given $\varepsilon > 0$, there is a sequence K_n of bodies $(n = 1, 2, \ldots)$, $\dim K_n = n$, satisfying

$$d_{K_n} \leqq \alpha(\varepsilon) \quad \text{and} \quad s(K_n) < (1+\varepsilon) s_n$$

where s_n stands for the worse n-dimensional characteristic.

5. Proof of Theorem 2

Fix an n-dimensional "cotype-2" space E, $\| \ \|$ (meaning that we assume a bound on $C_2(E)$). Let K be the unit ball of E and assume the standard euclidean ball B_n the maximal volume ellipsoid of K. Define now for $2 \leqq t \leqq \sqrt{n}$

$$\sigma(t) = \sup\left\{\left(\frac{\operatorname{Vol}(K \cap F)}{\operatorname{Vol} B_k}\right)^{1/k} \middle| k \leqq n,\ F \text{ } k\text{-dimensional subspace of } E \text{ and } d_{F \cap K} \leqq t\right\}.$$

We wish then to bound $\sigma(t)$ by an appropriate function of $C_2(E)$. Fix t and consider a k-dimensional subspace F of E. Notice that trivially

$$C_2(F) \leqq C_2(E).$$

By the subspace lemma (2.10), there is for given $k/2 \leqq k' \leqq k$ a subspace F' of F satisfying

$$c\frac{k-k'}{k} C_2(E)^{-1} (\log t)^{-3/2} |x| \leqq \|x\| \leqq |x| \quad \text{for } x \in F'. \tag{5.1}$$

Hence, by (3.7′)

$$\mathrm{Vol}(K \cap F) \leqq t^{k-k'} \left(\frac{C}{\sqrt{k-k'}}\right)^{k-k'} \mathrm{Vol}(K \cap F') \tag{5.2}$$

where by (5.1)

$$\left(\frac{\mathrm{Vol}(K \cap F')}{\mathrm{Vol}\, B_{k'}}\right)^{1/k'} \leqq \sigma\left(C\frac{k}{k-k'} C_2(E) (\log t)^{3/2}\right). \tag{5.3}$$

From (5.2) and (5.3)

$$\begin{aligned}\left\{\frac{\mathrm{Vol}(K \cap F)}{\mathrm{Vol}\, B_k}\right\}^{1/k} &\leqq t^{\frac{k-k'}{k}} \left(\frac{C}{\sqrt{k-k'}}\right)^{\frac{k-k'}{k}} \left(\frac{\mathrm{Vol}\, B_{k'}}{\mathrm{Vol}\, B_k}\right)^{1/k} \sigma\left(C\frac{k}{k-k'} C_2(E) (\log t)^{3/2}\right) \\ &\leqq t^{\frac{k-k'}{k}} \left(C\frac{k}{k-k'}\right)^{\frac{k-k'}{2k}} \sigma\left(C\frac{k}{k-k'} C_2(E) (\log t)^{3/2}\right).\end{aligned} \tag{5.4}$$

Choosing k' such that $\frac{k}{k-k'} \approx (\log t)^2$, (5.4) implies

$$\left(\frac{\mathrm{Vol}(K \cap F)}{\mathrm{Vol}\, B_k}\right)^{1/k} \leqq 2^{1/\log t}\, \sigma(C\, C_2(E) (\log t)^4).$$

In this inequality, F is any k-dimensional subspace of E satisfying

$$\frac{1}{t}|x| \leqq \|x\| \quad \text{for } x \in F.$$

Therefore we proved

$$\sigma(t) \leqq 2^{1/\log t}\, \sigma(C\, C_2(E) (\log t)^4). \tag{5.5}$$

Again, (5.5) implies that

$$\sigma(t) \leqq C\, C_2(E) [1 + \log C_2(E)]^4$$

and in particular

$$\mathrm{V.r.}\, E \leqq C\, C_2(E) [1 + \log C_2(E)]^4.$$

Remark. A more delicate argument permits to improve this estimate. In fact, it can be shown that

$$\mathrm{V.r.}\, E \leqq C(\varepsilon)\, C_2(E) [1 + \log C_2(E)]^{\frac{3}{2}+\varepsilon}, \quad \varepsilon > 0.$$

6. Applications

The purpose of this section is to derive some consequences of the reverse Santalo inequality

A. Non-symmetric case and mixed volumes

Corollary 6.1. *Let K be a bounded, convex and absorbing set in $\mathbb{R}^n$ containing the origin. Let $K^0=\{x\in\mathbb{R}^n \mid \sup_{y\in K}\langle x,y\rangle\leqq 1\}$. Then*

$$s(K)\equiv\left(\frac{\mathrm{Vol}\,K}{\mathrm{Vol}\,B}\right)^{1/n}\left(\frac{\mathrm{Vol}\,K^0}{\mathrm{Vol}\,B}\right)^{1/n}\geqq c_2 \tag{6.1}$$

where $c_2>0$ is a numerical constant.

Proof. The problem is to pass from the symmetric to the non-symmetric case. By Theorem 1, $s(K-K)\geqq C'$ while trivially

$$(K-K)^0\subset K^0$$

and by the Rogers-Shephard theorem [R-S]

$$\mathrm{Vol}(K-K)\leqq\binom{2n}{n}\mathrm{Vol}\,K\leqq 4^n\,\mathrm{Vol}\,K;$$

hence $s(K-K)\leqq 4s(K)$, thus (6.1).

(6.1) has the following consequence for mixed volumes (see [Sa, 2] or [B-Z] for definitions and main properties).

Corollary 6.2. *Let K_i, $i=1,\ldots,n$, be bounded convex and absorbing sets in $\mathbb{R}^n$ and K_i^0 the polar sets of K_i. Then*

$$V(K_1,\ldots,K_n)\cdot V(K_1^0,\ldots,K_n^0)\geqq c_2^n(\mathrm{Vol}\,B_n)^2 \tag{6.2}$$

where c_2 is the constant appearing in (6.1).

Proof. Use the following variant of the Alexandrov-Fenchel inequality [A], [Fe]

$$\mathrm{Vol}(K_1,\ldots,K_n)\geqq\Big(\prod_{1\leqq i\leqq n}\mathrm{Vol}\,K_i\Big)^{1/n}.$$

Remarks. 1. An important particular case of (6.2) appears when $K_{t+1}=\ldots=K_n=B$ $(1\leqq t\leqq n-1)$. Thus

$$V(K_1,\ldots,K_t,B,\ldots,B)\cdot V(K_1^0,\ldots,K_t^0,B,\ldots,B)\geqq c_2^t(\mathrm{Vol}\,B)^2. \tag{6.3}$$

Note that one special case of (6.3) was known [F]. This is the case when $K_i=K$ $(1\leqq 1\leqq t)$, $t\leqq\dfrac{n+1}{2}$. In this situation, c_2 may be replaced by 1.

2. The above formula (6.2) may be useful in estimating from below the typical number of solutions of a system of n algebraic equations with n variables. Such connection is obtained through results of [Be], [K] where it is shown that $n!\,V(K_1, \ldots, K_n)$ is the typical number of solutions of algebraic equations with Newton polyhedrons $\{K_i\}_{i=1}^n$ (for more details, see [Be], [K] or [B-Z]).

B. Minkowski's theorem and geometric number theory

Let Y be the lattice $\mathbb{Z}^n \subset \mathbb{R}^n$. Minkowski's fundamental theorem states that if K is a bounded, 0-symmetric convex body in $\mathbb{R}^n$, $\mathrm{Vol}\, K > 2^n$, then $K \cap (Y-0) \neq \emptyset$ (see [L]). The example for the cube $[|x_i| < 1;\ 1 \leqq i \leqq n]$ shows that this statement cannot be improved without additional hypothesis on K. A short proof of Minkowski's theorem is obtained as follows (see [L]). Denote $P = \{x \mid 0 \leqq x_i < 1\ (1 \leqq i \leqq n)\}$ the unit cell. Assume $K \cap Y = \{0\}$ and define

$$\varphi(x) = \sum_{u \in Y} \chi(x+u)$$

where χ stands for the indicator function of $1/2\,K$. Since the $\frac{1}{2}K + u$ $(u \in Y)$ are mutually disjoint, $\varphi^2 = \varphi$. By construction φ is 1-periodic in each variable separately. Applying Parseval's identity, it follows

$$\alpha(0) = \int_P \varphi(x)\,dx = \int_P \varphi(x)^2\,dx = \alpha(0)^2 + \sum_{\substack{u \in Y \\ u \neq 0}} |\alpha(u)|^2;$$

$$\alpha(u) = \int_{\mathbb{R}^n} \chi(x)\, e^{-2\pi i \langle u, x \rangle}\,dx.$$

Hence $\alpha(0) \leqq 1$ or $\mathrm{Vol}\, K \leqq 2^n$. The argument also shows that if $\mathrm{Vol}\, K = 2^n$ and $K \cap Y = \{0\}$, then K is an affine image of $[|x_i| \leqq 1]$. For ellipsoids, there are improvements of Minkowski's theorem, but the optimal result does not seem to be known.

As a consequence of Theorem 1,

$$\mathrm{Vol}\, K \cdot \mathrm{Vol}\, K^0 > \left(\frac{c}{n}\right)^n.$$

Therefore, by Minkowski's theorem

Corollary 6.4. *Either* $(C_1 \sqrt{n}\, K) \cap (Y-0) \neq \emptyset$ *or* $(C_1 \sqrt{n}\, K^0) \cap (Y-0) \neq \emptyset$, *where* C_1 *is an absolute constant.*

We will show that this is an optimal result (up to the value of C_1), already for ellipsoids. Latter fact will be deduced from

Proposition 6.5. *Let* χ_1 *and* χ_2 *be bounded Riemann integrable functions vanishing outside a bounded region and let* $\varepsilon > 0$ *be a positive number. Then there exist* $u \in GL(\mathbb{R}^n)$, $\det u = 1$, *such that*

$$\sum_{x \in Y-0} \chi_1(u(x)) + \sum_{x \in Y-0} \chi_2((u^{-1})^* x) \leqq \int_{\mathbb{R}^n} \chi_1(x)\,dx + \int_{\mathbb{R}^n} \chi_2(x)\,dx + \varepsilon. \tag{6.5}$$

Proposition 6.5 implies that (6.4) is optimal in the family of affine images of any K. Assume $\mathrm{Vol}\, K = \mathrm{Vol}\, B_n$. By Santalo's inequality,

$$\mathrm{Vol}(c\sqrt{n}\, K) < 1/2 \quad \text{and} \quad \mathrm{Vol}(c\sqrt{n}\, K^0) < 1/2$$

for some absolute constant c. Apply (6.5) with χ_1 (resp. χ_2) taken to be the characteristic function of $c\sqrt{n}\, K$ (resp. $c\sqrt{n}\, K^0$). Since the left hand side of (6.5) is 0 or at least 1, the right hand side less than 1 (for ε small enough), it follows that $Y - 0$ does not intersect $u^{-1}(K)$ nor $[u^{-1}(K)]^0 = u^*(K^0)$.

Proof of (6.5). This is an adaptation of the Davenport-Rogers Lemma (see [L], p. 146). It involves an averaging argument over $\bar{z} = (z_1, \ldots, z_{n-1}) \in [0,1]^{n-1}$ given by

$$u_{\bar{z}} = \begin{bmatrix} \alpha^{-1/n-1} & 0 & \ldots & 0 & \alpha^{-1/n-1} z_1 \\ 0 & \alpha^{-1/n-1} & \ldots & 0 & \alpha^{-1/n-1} z_2 \\ 0 & 0 & \ldots & 0 & \vdots \\ \vdots & \vdots & \ldots & \alpha^{-1/n-1} & \alpha^{-1/n-1} z_{n-1} \\ 0 & 0 & \ldots & 0 & \alpha \end{bmatrix},$$

$$(u_{\bar{z}}^{-1})^* = \begin{bmatrix} \alpha^{1/n-1} & 0 & \ldots & 0 & 0 \\ 0 & \alpha^{1/n-1} & \ldots & 0 & 0 \\ 0 & 0 & \ldots & \vdots & \vdots \\ \vdots & \vdots & \ldots & \alpha^{1/n-1} & 0 \\ -\alpha^{-1} z_1 & -\alpha^{-1} z_2 & \ldots & -\alpha^{-1} z_{n-1} & \alpha^{-1} \end{bmatrix}$$

and where we let the parameter $\alpha \to 0$.

In [L], the reader will find the proof of the inequality

$$\int_{[0,1]^{n-1}} \Big[\sum_{x \in Y - 0} \chi(u(x)) \Big] d\bar{z} < \int \chi(x)\, dx + \varepsilon \tag{6.6}$$

for α small enough. A similar argument, which follows next, also applies to the term involving $(u^{-1})^*$. Consider

$$\sum_{\substack{x = (\bar{v}, r) \neq 0 \\ \bar{v} \in \mathbb{Z}^{n-1},\, r \in \mathbb{Z}}} \chi((u^{-1})^* x)$$
$$= \sum \chi(\alpha^{1/n-1} v_1, \ldots, \alpha^{1/n-1} v_{n-1}, -\alpha^{-1} z_1 v_1 \ldots - \alpha^{-1} z_{n-1} v_{n-1} + \alpha^{-1} r). \tag{6.7}$$

The condition $(\bar{v}, r) \neq 0$ may be rewritten as $\bar{v} \neq 0$ for α small enough (because otherwise $\alpha \to 0$ gives a point out of the support of χ). If $\varphi(t)$ is a 1-periodic function on $\mathbb{R}$, $\bar{v} \in \mathbb{Z}^{n-1}$, $\bar{v} \neq 0$ then

$$\int_0^1 \varphi(t)\, dt = \int_0^1 \ldots \int_0^1 \varphi(v_1 z_1 + \ldots + v_{n-1} z_{n-1})\, dz_1 \ldots dz_{n-1}. \tag{6.8}$$

Integrate (6.7) in $z_1, \ldots, z_{n-1}$ from 0 to 1 and apply (6.8) to the function

$$\varphi(t) = \sum_{r \in \mathbb{Z}} \chi(\alpha^{1/n-1} v_1, \ldots, \alpha^{1/n-1} v_{n-1}, \alpha^{-1}(t + r)).$$

Thus we get

$$\sum_{\bar{v}\in\mathbb{Z}^{n-1}-0}\int_0^1\sum_{r\in\mathbb{Z}}\chi(\alpha^{1/n-1}v_1,\ldots,\alpha^{1/n-1}v_{n-1},\alpha^{-1}(t+r))\,dt$$

$$=\sum_{\bar{v}\in\mathbb{Z}^{n-1}-0}\alpha\int_{-\infty}^{\infty}\chi(\alpha^{1/n-1}v_1,\ldots,\alpha^{1/n-1}v_{n-1},t)\,dt\xrightarrow{\alpha\to 0}\int\chi.$$

Thus for α small enough

$$\int[\sum_{x\in Y-0}\chi((u_{\bar{z}}^{-1})^*(x))]\,d\bar{z}\leqq\int\chi+\varepsilon. \tag{6.9}$$

The existence of u satisfying (6.5) is now immediate from (6.6), (6.9).

Recall that the consecutive minima of K is the sequence of numbers

$$\lambda_1\leqq\lambda_2\leqq\ldots\leqq\lambda_n$$

defined by

$$\lambda_i=\inf\{\lambda>0\,|\,\dim(\lambda K\cap Y)\geqq i\}.$$

Minkowski's theorem has the following improvement (second theorem) [L], p. 52

$$\lambda_1\ldots\lambda_n\operatorname{Vol}K\leqq 2^n.$$

This fact together with Theorem 1 in this paper yields

Corollary 6.10. *Let $\lambda_1,\ldots,\lambda_n$ and $\lambda_1',\ldots,\lambda_n'$ be the consecutive minima of K, K^0 respectively. Then*

$$\lambda_1,\ldots,\lambda_n\,\lambda_1',\ldots,\lambda_n'\leqq(Cn)^n$$

$$\inf_{1\leqq i\leqq n}\lambda_i\lambda_{n+1-i}'\leqq Cn\quad\text{and}\quad\sup_{1\leqq i\leqq n}\lambda_i\lambda_{n+1-i}'\leqq(Cn)^n.$$

The last inequality in a consequence of $\lambda_i\lambda_{n+1-i}'\geqq 1$ for $i=1,\ldots,n$ (see [L], p. 107).

C. Harmonic analysis

The use of geometric methods in order to prove the existence of trigonometric polynomials satisfying certain prescribed conditions, has been exploited by several people (cf. in particular the work of B.S. Kashin related to the theory of Kolmogorov widths, uniformly convergent Fourier series etc.). Our purpose is to present some corollaries of the reverse Santalo-inequality in this context. Actually, the most interesting applications already appear in Kashin's work. Let G be a compact Abelian group and $\Gamma=G^*$ the dual group of G. For a subset Λ of Γ, denote P_Λ the set of polynomials f with Fourier transform $\hat{f}$ supported by Λ, i.e. the set of linear combinations of characters belonging to Λ and considered as functions on G. Consider the following pseudo-distances on G

$$d_\Lambda(x,y)=\sup_{\gamma\in\Lambda}|\gamma(x)-\gamma(y)|\quad\text{and}\quad\bar{d}_\Lambda(x,y)=\sup_{f\in P_\Lambda,\,|f|\leqq 1}|f(x)-f(y)|.$$

Obviously $d_\Lambda \leqq \bar{d}_\Lambda$. Less trivial is the fact that $\bar{d}_\Lambda \leqq C d_\Lambda$ for an absolute constant C. Latter inequality is deduced in [Bo, 2] from the spectral synthesis property of the one-point set. For $\varepsilon > 0$, let $N(\Lambda, \varepsilon)$ be the ε-entropy number with respect to the metric d_Λ, i.e. the minimal number of d_Λ-balls of radius ε needed to cover G. It can be shown that $\log N(\Lambda, \varepsilon) \leqq \log 1/\varepsilon \cdot \log N(\Lambda, 1/20)$ [Bo, 2]. It follows from the uniform equivalence of the pseudo-metrics d_Λ, $\bar{d}_\Lambda$ that if we let $N(\Lambda) \equiv N(\Lambda, c)$ for a sufficiently small (absolute) constant $c > 0$, then there are points $x_1, \ldots, x_{N(\Lambda)}$ in G satisfying

$$|f| \leqq 2 \sup_{1 \leqq j \leqq N(\Lambda)} |f(x_j)| \quad \text{if } f \in P_\Lambda. \tag{6.11}$$

Corollary 6.12. *Let Λ be a finite subset of Γ and $\{\Lambda_i\}$ subsets of Λ. Let $\{s_i\}$ be positive numbers. Then there exists $f \in P_\Lambda$ satisfying the following conditions*

$$\Big| \sum_{\gamma \in \Lambda_i} \hat{f}(\gamma)\, \gamma \Big| \leqq s_i |\Lambda_i|^{1/2} \quad \textit{for each } i \tag{6.12a}$$

$$\sum_{\gamma \in \Lambda} |\hat{f}(\gamma)| \geqq C |\Lambda|^2 \Big[\sum_i s_i^{-1} N(\Lambda_i)\Big]^{-1} \tag{6.12b}$$

where $c > 0$ is an absolute constant.

In particular, (6.12) easily implies

Corollary 6.13. *Let N be a positive integer and $\Lambda \subset \{0, 1, \ldots, N\}$. Then there exists a trigonometric polynomial $f \in P_\Lambda$ satisfying*

- $\|\hat{f}\|_\infty \leqq 1$,
- $\|D_k f\|_\infty \leqq k^{1/2}$; $\quad D_k f = \sum_{j=1}^{k} e^{ij\theta} \hat{f}(j)$,
- $\|f\|_\Lambda \equiv \|\hat{f}\|_1 \geqq c\, N^{-1} |\Lambda|^2$.

Proof of 6.12. For each i, let $\mathscr{E}_i$ be a subset of G such that

$$\# \mathscr{E}_i = N(\Lambda_i) \quad \text{and} \quad |f| \leqq 2 \sup |f(x)|, \quad x \in \mathscr{E}_i.$$

Thus $f \in P_\Lambda$ will satisfy (6.12a) provided $\{\hat{f}(\gamma) \mid \gamma \in \Lambda\}$ belongs to K^0, where K is the convex symmetric body in $\mathbb{C}^{|\Lambda|}$ generated by the set

$$\left\{ \frac{2}{s_i |\Lambda_i|^{1/2}} \Big(\sum_{\gamma \in \Lambda_i} \gamma(x)\, e_\gamma \Big) \mid i = 1, 2, \ldots \text{ and } x \in \mathscr{E}_i \right\}.$$

Since an easy computation shows that

$$(\operatorname{Vol} K)^{1/|\Lambda|} \leqq C |\Lambda|^{-2} \sum_i |\mathscr{E}_i|\, s_i^{-1}$$

if follows from the reverse Santalo inequality that

$$(\operatorname{Vol} K^0)^{1/|\Lambda|} \geqq c |\Lambda| \Big[\sum_i |\mathscr{E}_i|\, s_i^{-1}\Big]^{-1}$$

and (6.12b) now follows by comparing this volume with the volume of the $l^1_{|\Lambda|}$-unit ball.

7. Comments

(1) The simplifications of the proofs of Theorems 1 and 2 with respect to [B-M] are due to G, Pisier [Pi, 2]. We included the argument here with his permission.

(2) V. Milman and G. Pisier have characterized those Banach spaces X with the property that

$$\sup_{\substack{E \hookrightarrow X \\ \dim E < \infty}} (\text{V.r. } E) < \infty.$$

They are the so-called weak-cotype 2-spaces, defined as those spaces X such that any finite dimensional subspace E of X contains a subspace F,

$$\dim F > 1/2 \dim E$$

$$d_F \leqq C$$

where $C = C(X) < \infty$. Such Banach spaces are not necessarily of cotype 2.

(3) The (optimal) improvement of inequality (2.2) where $1-\lambda$ gets replaced by $(1-\lambda)^{1/2}$ was obtained by A. Pajor and N. Tomczak [P-T].

(4) Using the ideas and techniques of [B-M] and the present paper, V. Milman has proved more recently a "converse" to Brunn's theorem. A summary of his work appears in [Mi, 4].

(5) Recently, S. Reisner [R] proved (1.3) in the class of zonoids.

8. Appendix: Prerequisites from geometry of Banach spaces

Define for $k = 1, 2, \ldots$ and $0 \leqq t \leqq 1$

$$r_k(t) = \operatorname{sign} \sin 2\pi 2^k t.$$

This sequence of ± 1 valued functions are called Rademacher functions. It is often more convenient to see the sequence $\{r_i(t)\}_{i=1}^{\infty}$ as the coordinate projections on the probability space $\Omega = \{-1, 1\}^{\mathbb{N}}$ endowed with the natural product measure. Thus

$$r_i(t) = t_i \quad \text{for } t = (t_1, t_2, \ldots).$$

Notice that if X is a normed space and $\{x_i\} \subset X$, then

$$\operatorname*{Ave}_{\varepsilon_i = \pm 1} \| \sum \varepsilon_i x_i \| = \int \| \sum r_i(t) x_i \| \, dt.$$

Such averages (Rademacher averages) play an essential role in local Theory of Banach spaces (cf. [Ma-P]).

Given a measure space $(\Omega, \mathscr{F}, \mu)$ and a Banach space X, define the space $L^2(X) = L^2(X, \Omega)$ as the space of all measurable functions from $(\Omega, \mathscr{F})$ to X equipped with norm

$$\| f \|_{L^2(X)} = [\int_{\Omega} \| f(\omega) \|_X^2 \, d\omega]^{1/2} (< \infty).$$

For $f \in L^2(X)$, $g \in L^2(X^*)$, the duality relation is given by $\langle f, g \rangle = \int \langle f(\omega), g(\omega) \rangle \, \mu(d\omega)$. If X is finite dimensional, then $L^2(X^*)$ identifies with

the dual space of $L^2(X)$. The main measure spaces involved are the finite groups $\{1,-1\}^n$ with normalized counting measure. In this case, only n Rademacher functions $\{r_i(t)\}_{i=1}^n$ appear. Notice that the space $L^2(\{1,-1\}^n)$ is 2^n-dimensional. To extend $\{r_i(t)\}_{i=1}^n$ to an orthonormal basis of $L^2(\{1-1\}^n)$, consider for every subset A of $\{1,\ldots,n\}$ the function

$$w_A(t)=\prod_{i\in A} r_i(t)$$

(in the case $A=\emptyset$, let $w_\emptyset(t)=1$). The set of functions $\{w_A | A\subset\{1,\ldots,n\}\}$ is an orthonormal basis of $L^2(\{1,-1\}^n)$, called the Walsh system. Any X-valued function f on $\{1,-1\}^n$ has an expansion

$$f(t)=\sum_A w_A(t)\,x_A \quad \text{where} \quad x_A=\hat{f}(A)=\int f(t)\,w_A(t)\,dt.$$

The following special subspace of $L^2(X)$ is of importance

$$\operatorname{Rad} X=\{\textstyle\sum r_i(t)\,x_i \,|\, x_i\in X\}.$$

Recall the uniform equivalence

$$\|f\|_{L^p(X)}\sim\|f\|_{L^2(X)} \qquad (1\leqq p<\infty) \text{ for } f\in\operatorname{Rad} X,$$

by Kahane's inequalities.

There is a natural (orthogonal) projection on $\operatorname{Rad} X$. For

$$f=\sum_A w_A\,x_A, \qquad \text{let } R_1 f=\sum_i r_i(t)\,x_{\{i\}}.$$

Denote $\operatorname{Rad}_X=\|R_1\|$ as an operator on $L^2(X)$.

Lemma 8.1 (Pisier [Pi, 1]). $\operatorname{Rad}_X\leqq C\log(1+d_X)$, *for an absolute constant* C.

Proof. Define $R_j f=\sum_{|A|=j} w_A(t)\,x_A$ (projection on the order $-j$ Walsh functions). Fix $f=\sum w_A(t)\,\hat{f}(A)$, $\hat{f}(\emptyset)=\int f=0$. For $|s|\leqq 1$, convolute f with the measure $\mu_s=\prod(1+s r_i)$ (=biased-coin convolution), giving

$$\|f\|_{L^2(X)}\geqq\|f*\mu_s\|_{L^2(X)}=\|\sum_{j\geqq 1} s^j R_j f\|_{L^2(X)}. \tag{8.2}$$

Fix an integer J and estimate

$$\|f\|_{L^2(X)}\geqq\|\sum_{j\leqq J} s^j R_j f\|_{L^2(X)}-\sum_{j>J}|s|^j\,\|R_j f\|_{L^2(X)}. \tag{8.3}$$

Since $\|R_j f\|_{L^2(E)}\leqq\|f\|_{L^2(E)}$ if E is a Hilbert space, one gets

$$\|R_j f\|_{L^2(X)}\leqq d_X\|f\|_{L^2(X)} \quad \text{for each } j. \tag{8.4}$$

It follows from Bernstein's inequality (see [Zy])

$$|p'(0)|\leqq CJ\sup_{|t|\leqq 1/2}|p(t)|$$

if p is a polynomial of degree $\leqq J$, that there is a measure ν on $[-1/2, 1/2]$ satisfying the conditions

$$\|\nu\| \leqq CJ; \quad \int t\,\nu(dt)=1 \quad \text{and} \quad \int t^j \nu(dt)=0 \quad \text{for } 2\leqq j\leqq J. \tag{8.5}$$

It now follows from (8.3)–(8.5) that

$$\begin{aligned}\|f\|_{L^2(X)} &\geqq \frac{1}{CJ}\int \Big\| \sum_{j\leqq J} s^j R_j f\Big\|_{L^2(X)} \nu(ds) - 2^{-J} d_X \|f\|_{L^2(X)} \\ &\geqq \frac{1}{CJ}\|R_1 f\|_{L^2(X)} - 2^{-J} d_X \|f\|_{L^2(X)}.\end{aligned}$$

Choosing $J \sim \log d_X$, (8.1) clearly follows.

Remark. From (8.1), we deduce in particular that $\mathrm{Rad}_X \leqq C\log(\dim X+1)$, which is an optimal estimate in general (see [Bo, 1]).

Let $\{g_i(\omega)\}_{i>1}$ be a sequence of independent normalized Gaussian random variables on some probability space. If $u\colon l_n^2 \to X$, $\dim X = n$ and $\{e_i\}_{i=1}^n$ an orthonormal basis of l_n^2, then

$$M(u) \equiv \int_{S_{n-1}} \|u(x)\|\,\mu_{n-1}(dx) = \frac{c_n}{\sqrt{n}} \int \left\| \sum_{i=1}^n g_i(\omega)\, u(e_i)\right\| d\omega \qquad (c_n \sim 1)$$

and

$$\int_{S_{n-1}} \|u(x)\|^2\,\mu_{n-1}(dx) = \frac{1}{n}\int \left\| \sum_{i=1}^n g_i(\omega)\, u(e_i)\right\|^2 d\omega.$$

These facts follow from the invariance properties for the Gaussian system. Notice also that by Kahane's inequalities

$$\|\textstyle\sum g_i(\omega)\, u(e_i)\|_{L^1(X)} \sim \|\sum g_i(\omega)\, u(e_i)\|_{L^2(X)}.$$

The aim of the next lemma is to compare Rademacher and Gaussian averages.

Lemma 8.6.

$$\int \|\textstyle\sum g_i(\omega)\, x_i\|\, d\omega \leqq C\sqrt{\log(d_X+1)} \int \left\| \sum_{i=1}^n r_i(t)\, x_i\right\| dt \qquad \textit{for } \{x_i\} \subset X.$$

Proof. Decomposing each $g_i = h_i + h_i'$ where $h_i = g_i \chi_{[|g_i|<\lambda]}$, we have $\int |h_i'|^2 < e^{-c\lambda^2}$. Here λ is a constant to be specified later. Choose $u\colon X \to l_n^2$ satisfying

$$d_X = \|u\|\,\|u^{-1}\|.$$

Using the independence of the h_i

$$\begin{aligned}\|\textstyle\sum g_i(\omega)\, x_i\|_{L^2(X)} &\leqq \|\sum h_i(\omega)\, x_i\|_{L^2(X)} + \|\sum h_i'(\omega)\, x_i\|_{L^2(X)} \\ &\leqq \lambda \|\textstyle\sum r_i(t)\, x_i\|_{L^2(X,\,dt)} + \|u^{-1}\|\,\|\sum h_i'(\omega)\, u(x_i)\|_{L^2(l_n^2)}.\end{aligned} \tag{8.7}$$

The second term in (8.7) may further be estimated as (by the orthogonality of the h_i')

$$\begin{aligned}\|u^{-1}\|\,(\textstyle\sum \|h_i'\|_2^2\, |u(x_i)|^2)^{1/2} &\leqq \|u^{-1}\|\, e^{-c\lambda^2} \int |\sum r_i(t)\, u(x_i)|\, dt \\ &\leqq \|u\|\,\|u^{-1}\|\, e^{-c\lambda^2} \int \|r_i(t)\, x_i\|\, dt.\end{aligned}$$

Hence

$$\int \|\sum g_i(\omega)\, x_i\|\, d\omega \leqq (\lambda + e^{-c\lambda^2} d_X) \int \|\sum r_i(t)\, x_i\|\, dt$$

and the lemma follows taking $\lambda \sim (\log d_X)^{1/2}$.

Proof of Proposition 2.8. Let E be an n-dimensional normed space such that the maximum volume ellipsoid of its unit ball coincides with the euclidean ball B_n. Let F be a k-dimensional subspace of E. We have to prove that

$$M_{F^*} \leqq C\, C_2(E)\, (1 + \log d_E)^{3/2} \sqrt{n/k}. \tag{2.8}$$

If $e_1, \ldots, e_k, e_{k+1}, \ldots, e_n$ is an orthonormal basis with $F = [e_1, \ldots, e_k]$, it follows from the preceeding that

$$M_{F^*} \sim \frac{1}{\sqrt{k}} \int \left\| \sum_{j=1}^{k} g_i(\omega)\, e_j \right\|_{F^*} d\omega \leqq \frac{1}{\sqrt{k}} \int \left\| \sum_{j=1}^{n} g_j(\omega)\, e_j \right\|_{*} d\omega. \tag{8.8}$$

By the basic properties of the maximum volume ellipsoid (see [J]), there are positive numbers $\{\lambda_\alpha\}$ and points $\{x_\alpha\}$ satisfying

$$\sum \lambda_\alpha = n \tag{8.9}$$

$$\|x_\alpha\| = \|x_\alpha\|_* = |x_\alpha| = 1 \tag{8.10}$$

$$x = \sum \lambda_\alpha \langle x, x_\alpha \rangle\, x_\alpha \qquad \forall x \in \mathbb{R}^n. \tag{8.11}$$

Hence, by (8.11)

$$e_j = \sum_\alpha A_{j\alpha} (\lambda_\alpha^{1/2} x_\alpha); \qquad A_{j\alpha} = \lambda_\alpha^{1/2} \langle e_j, x_\alpha \rangle.$$

Considering the (non-square) matrix $(A_{j\alpha})_{1 \leqq j \leqq n, \alpha} \equiv A$, polar decomposition and invariance of the Gaussian system gives the estimate

$$\|\sum g_j(\omega)\, e_j\|_{L^2(E^*)} \leqq \|A\|_{l^2 \check{\otimes} l^2} \|\sum_\alpha g'_\alpha(\omega)\, \lambda_\alpha^{1/2} x_\alpha\|_{L^2(E^*)} \tag{8.12}$$

where $\{g'_\alpha\}$ is another Gaussian system. By (8.11)

$$\|A\|_{l^2 \check{\otimes} l^2} \equiv \sup_{|\xi|=1} (\sum_\alpha \lambda_\alpha |\langle \xi, x_\alpha \rangle|^2)^{1/2} = 1. \tag{8.13}$$

By (8.6)

$$\|\sum g'_\alpha(\omega)\, \lambda_\alpha^{1/2} x_\alpha\|_{L^2(E^*)} \leqq C(1 + \log d_E)^{1/2}\, \|\sum r_\alpha(t)\, \lambda_\alpha^{1/2} x_\alpha\|_{L^2(E^*)} \tag{8.14}$$

replacing the Gaussian average by a Rademacher average. The latter is evaluated by a duality reasoning. Take thus $f \in L^2(E)$, $\|f\|_{L^2(E)} = 1$ and write by (8.9), (8.10) and (8.1)

$$\begin{aligned} |\langle \sum_\alpha r_\alpha \lambda_\alpha^{1/2} x_\alpha, f \rangle| = |\sum_\alpha \lambda_\alpha^{1/2} \langle x_\alpha, \hat{f}(\alpha) \rangle| &\leqq \sum \lambda_\alpha^{1/2} \|x_\alpha\|_*\, \|\hat{f}(\alpha)\| \\ &\leqq n^{1/2} (\sum \|\hat{f}(\alpha)\|^2)^{1/2} \\ &\leqq n^{1/2}\, C_2(E)\, \|R_1 f\|_{L^2(E)} \\ &\leqq n^{1/2}\, C_2(E)\, (1 + \log d_E)\, \|f\|_{L^2(E)}. \end{aligned}$$

Hence

$$\|\sum r_\alpha \lambda_\alpha^{1/2} x_\alpha\|_{L^2(E^*)} \leqq n^{1/2} C_2(E)(1+\log d_E). \tag{8.15}$$

Collecting inequalities (8.8), (8.12)–(8.15), inequality (2.8) follows.

The next lemma (see [Lew]) is a generalization of the variational argument exploited by F. John in order to prove the basic properties (8.9)–(8.11) of the maximal volume ellipsoid, mentione above.

Lemma 8.16. *Let α be a norm on the space $\mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$ of linear operators on $\mathbb{R}^n$ and α^* the dual norm of α with respect to the usual trace duality*

$$\alpha^*(A) = \sup\{|\mathrm{Tr}\, AB|;\ \alpha(B) \leqq 1\}.$$

Then there exists $u \in GL(\mathbb{R}^n)$ satisfying

$$\alpha(u)\,\alpha^*(u^{-1}) = n. \tag{8.16}$$

Proof. Fix $u \in GL(\mathbb{R}^n)$ satisfying $\alpha(u)=1$ and $\det u$ maximum under the restriction $\alpha(u) \leqq 1$. Let $v \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$. By construction, it follows for all $t>0$

$$\det\left[\frac{u+tv}{\alpha(u+tv)}\right] \leqq \det u,$$

$$\det u \cdot \det(I + t u^{-1} v) \leqq \alpha(u+tv)^n \det u,$$

$$\det(I + t u^{-1} v)^{1/n} - 1 \leqq \alpha(u+tv) - \alpha(u) \leqq t\,\alpha(v).$$

Since

$$\frac{1}{n}\mathrm{Trace}\, u^{-1} v = \lim_{t \downarrow 0} \frac{1}{t}[\det(I + t u^{-1} v)^{1/n} - 1] \leqq \alpha(v)$$

we have

$$\alpha^*(u^{-1}) = \sup_{\alpha(v) \leqq 1} \mathrm{Tr}(u^{-1} v) \leqq n.$$

Notice that trivially always $\alpha(u)\,\alpha^*(u^{-1}) \geqq n$.

Proof of Proposition 2.9 cf. [F-T]. Our purpose is to find $u: l_n^2 \to E$ satisfying

$$M(u)\, M((u^{-1})^*) \leqq C(1+\log d_E)^2 \tag{8.17}$$

which is clearly an equivalent statement. By (8.6), this problem reduces to the existence of $u: l_n^2 \to E$ such that

$$\|\sum_{1 \leqq i \leqq n} r_i u(e_i)\|_{L^2(E)} \cdot \|\sum_{1 \leqq i \leqq n} r_i (u^{-1})^*(e_i)\|_{L^2(E^*)} \leqq C(1+\log d_E). \tag{8.18}$$

Apply Lemma 8.16 with respect to

$$\alpha(u) = \|\sum r_i u(e_i)\|_{L^2(E)}.$$

(Here $\{e_i\}_{i=1}^n$ is a fixed orthonormal basis.)

Consider u satisfying

$$\alpha(u)\,\alpha^*(u^{-1}) = n.$$

It remains to relate $\alpha^*(u^{-1})$ and the second factor is the left hand side of (8.18).

It follows from (8.1) that

$$\|\sum r_i(u^{-1})^*(e_i)\|_{L^2(E^*)} \leqq C(1+\log d_E)\cdot \sup|\sum \langle (u^{-1})^* e_i, x_i\rangle| \tag{8.19}$$

where the supremum is taken over all sequences $\{x_i\}_{i=1}^n$ in E satisfying

$$\|\sum r_i(t)\, x_i\|_{L^2(E)} \leqq 1.$$

Fix such a sequence $\{x_i\}_{i=1}^n$ and define $v: l_n^2 \to E$ by $v(e_i)=x_i$. Thus $\alpha(v)\leqq 1$ and

$$|\sum \langle (u^{-1})^* e_i, x_i\rangle| = |\sum \langle e_i, u^{-1} v(e_i)\rangle| = |\mathrm{Trace}\, u^{-1} v| \leqq \alpha^*(u^{-1}).$$

Hence u satisfies (8.18), which completes the proof.

The reader will find a proof of the isoperimetric inequality on the sphere in the appendix of [FLM].

References

[A] Alexandrov, A.: On the theory of mixed volumes of convex bodies. I. Math. Sbornik **2** (N5), 947-972 (1937); II. idem (N6) 1205-1238; III. idem **3** (NI) 27-46 (1938); IV. **3** (N2) 227-251 (1938)

[Ba] Bambah, R.P.: Polar reciprocal convex bodies. Proc. Camb. Philos. Soc. **51**, 377-378 (1954)

[Be] Bernstein, D.N.: The number of roots of a system of equations. Funkts. Anal. Prilozh. **9** (N3), 1-4 (1975)

[Bo, 1] Bourgain, J.: On Martingale transforms in finite dimensional lattices with an appendix on the K-convexity constant. Math. Nachr. **119**, 41-53 (1984)

[Bo, 2] Bourgain, J.: A remark on entropy on Abelian groups and the uniform approximation property. Studia Math. (in press) (1987)

[B-M] Bourgain, J., Milman, V.D.: Sections euclidiennes et volume des corps symétriques convexes dans $\mathbb{R}^n$. C.R. Acad. Sci., Paris, Sér. I, t 300, **13**, 435-438 (1985)

[B-Z] Burago, Y.D., Zalgaler, V.A.: Geometrical inequalities. Leningrad: Nauka 1980

[D-M-T] Davis, W.J., Milman, V.D., Tomczak-Jaegermann, N.: The distance between certain n-dimensional Banach spaces. Israel J. Math. **39**, 1-15 (1982)

[Fe] Fenchel, W.: Inégalités quadratiques entre les volumes mixtes des corps convexes. C.R. Acad. Sci., Paris, Ser. I., **203**, 647-650 (1936)

[F] Firey, W.M.J.: Support flags to convex bodies, Geom. Dedic **2**, 225-248 (1973)

[F-L-M] Figiel, T., Lindenstrauss, J., Milman, V.D.: The dimension of almost spherical sections of convex bodies. Acta Math. **139**, 53-94 (1977)

[F-T] Figiel, T., Tomczak-Jaegermann, N.: Projections onto Hilbertian subspaces of Banach spaces. Israel J. Math. **53**, 155-171 (1979)

[G-R] Gordon, Y., Reisner, S.: Volume estimates and applications to finite-dimensional Banach spaces. Preprint

[J] John, F.: Extremum problems with inequalities as subsidiary conditions. Courant anniversary volume, pp. 187-204. New York: Interscience 1948

[K] Kauchnirenko, A.G.: Polyèdres de Newton et nombres de Milnor. Invent. Math. **32**, 1-31 (1979)

[L] Lekkerkerker, C.G.: Geometry of numbers. Amsterdam: North-Holland 1969

[Lew] Lewis, D.R.: Ellipsoides defined by Banach ideal norms. Mathematica **26**, 18-29 (1979)

[Lu] Lutwak, E.: On cross-sectional measures of polar reciprocal convex bodies. Geom. Dedicata **1**, 79-80 (1976)

[Ma] Mahler, K.: Ein Übertragungsprincip für konvexe Körper. Casopis Pest. Mat. Fys. **68**, 93-102 (1909)

[Ma-P] Maurey, B., Pisier, G.: Séries de variables aléatoires vectorielles indépendantes et propriétés géométriques des espaces de Banach. Stud. Math. **58**, 45–90 (1976)

[Mi, 1] Milman, V.D.: Almost Euclidean quotient spaces of subspaces of finite dimensional normed spaces. Proc. Am. Math. Soc. **94**, 445–449 (1985)

[Mi, 2] Milman, V.D.: Random subspaces of proportional dimension of finite dimensional normed spaces; approach through isoperimetric inequality. Missouri Conf. 1984. Springer LNM 1166, 1985, pp. 106–116

[Mi, 3] Milman, V.D.: Volume approach and iteration procedures in local theory of normed spaces. Missouri Conf. 1984, Springer LNM 1166, 1985, pp. 99–105

[Mi, 4] Milman, V.D.: Inégalité de Brunn-Minkowski inverse et applications à la théorie locale des espaces normés. C.R. Acad. Sci., Paris, Sér. I., **7302**, 25–28 (1986)

[M-S] Milman, V.D., Schechtman, G.: Asymptotic theory of finite dimensional normed spaces. (Lect. Notes in Math., vol. 1200). Berlin-Heidelberg-New York: Springer 1986

[P-T] Pajor, A., Tomczak-Jaegerman, N.: Nombres de Gelfand et sections euclidiennes de grande dimension. Séminaire d'Analyse fonctionnelle, 1984/85, Univ. Paris VI et VII

[Pel] Pelczynski, A.: Geometry of infinite dimensional Banach spaces and operator ideals. Notes in Banach spaces, pp. 81–181. Austin and London: Univ. of Texas Press 1980

[Pi, 1] Pisier, G.: K-convexity, Proceeding of Research Workshop on Banach Spaces, University of Iowa, 1981, pp. 139–151
Algo: Remarque sur un résultat non-publié de B. Maurey, Sém. d'Analyse Fonctionnelle, 1980/81, Exp. v, Ecole Polytechnique

[Pi, 2] Pisier, G.: Private communications

[R] Reisner, S.: Random polytopes and volume product of symmetric bodies. Math. Scand. **57**, 386–392 (1985)

[Ro] Rogalski, M.: Sur le quotient volumique d'un espace de dimension fini. Publ. Math. Univ. Pierre et Marie Curie **46**, C3 (1980/81)

[R-S] Rogers, C.A., Shephard, G.C.: The difference body of a convex body. Arch. Math. **8**, 220–233 (1957)

[S, 1] Saint-Raymond, J.: Sur le volume des corps convexes symétriques. Séminaire d'initiation à l'analyse 1980/81, Paris VI University

[S, 2] Saint-Raymond, J.: Le volume des idéaux d'opérateurs classiques. Stud. Math. **80**, 63–75 (1984)

[Sa, 1] Santalo, L.A.: Un invariante afin pasa los cuerpos convexos del espacio de n-dimensiones. Port. Math. **8**, 155–161 (1949)

[Sa, 2] Santalo, L.A.: Integral geometry and geometric probability. Encyclopedia Math. and Appl., Reading, MA: Addison-Wesley 1976

[Sz] Szarek, S.: On Kashin's almost Euclidean orthogonal decomposition of l_n^1. Bull. Acad. Pol. Sci., Ser. Sci. Math. **26**, 691–694 (1978)

[Sz-T] Szarek, S., Tomczak-Jaegermann, N.: On nearly Euclidean decompositions for some classes of Banach spaces. Compos. Math. **40**, 367–385 (1980)

[Zy] Zygmund, A.: Trigonometric series, Cambridge UP, 1959

Oblatum 30-IV-1986 & 20-VIII-1986

Invent. math. 88, 341-374 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Rigid reparametrizations and cohomology for horocycle flows

Marina Ratner*

Department of Mathematics, University of California, Berkeley, CA 94720, USA

In this paper we establish further rigidity properties of the classical horocycle flows in addition to those found in [8-11]. Namely, we show that, for flows obtained from horocycle flows by C^1-time changes all their ergodic joinings as well as all their factors remain algebraic. Moreover, we show that every nontrivial joining of two smoothly time-changed horocycle flows is induced by certain cohomological relations between the time changes.

We begin with some basic concepts and definitions (see [10, 11]).

Let T and S be two measure preserving transformations (m.p.t.'s) on probability spaces $(X, \mathscr{B}_X, \mu)$ and $(Y, \mathscr{B}_Y, \nu)$ respectively. A $T \times S$-invariant measure m on $(X \times Y, \mathscr{B}_X \times \mathscr{B}_Y)$ is called a joining of T and S if m has marginals μ and ν, i.e., $m(A \times Y) = \mu(A)$ for all $A \in \mathscr{B}_X$ and $m(X \times B) = \nu(B)$ for all $B \in \mathscr{B}_Y$. It is clear that $\mu \times \nu = m$ is a joining of T and S called the trivial joining.

Let $J(T, S)$ denote the set of all ergodic joinings of T and S. We say that T and S are disjoint if $J(T, S)$ is either empty or consists of the trivial joining $\mu \times \nu$. (This notion of disjointness was introduced by Furstenberg in [2].)

Let T_t be an m.p. flow on a probability space $(X, \mathscr{B}_X, \mu)$ and let $T = T_p = S$, $p \neq 0$, be ergodic. For $s \in R$ let m_s be the probability measure on $(X \times X, \mathscr{B}_X \times \mathscr{B}_X)$ defined by $m\{(x, T_s x): x \in A\} = \mu(A)$ for all $A \in \mathscr{B}_X$. It is clear that $m_s \in J(T, T)$ for all $s \in R$. The measures m_s, $s \in R$ are called off-diagonal self-joinings of T. T is said to have trivial self-joinings if every $m \in J(T, T)$ is either off-diagonal or the product $\mu \times \mu$. (This terminology is due to Rudolph [13].)

Sometimes the word "joining" will be used for the pair $(T \times S, m)$, meaning $T \times S$ acting on $(X \times Y, m)$. We say that $(T \times S, m)$ is a finite extension of S if there is a $T \times S$-invariant subset $\Omega \subset X \times Y$ with $m(\Omega) = 1$ such that the intersection $\Omega(y) = \Omega \cap (X \times \{y\})$ is finite for ν-almost every (a.e.) $y \in Y$. If $(T \times S, m)$ is ergodic, the number of points in $\Omega(y)$ is the same for ν-a.e. $y \in Y$.

Similar definitions may be made for measure preserving flows and semi-flows.

* Partially supported by National Science Foundation grant DMS-84-20770 and the Miller Institute for Basic Research, Univ. of Calif., Berkeley

Let T be an m.p.t. on (X, μ) and S an m.p.t. on (Y, ν). We say that S is a factor of T if there is a measure-preserving $\psi: X \to Y$ such that $\psi T(x) = S\psi(x)$ for μ-a.e. $x \in X$. The map ψ is called a factor map or a conjugacy between T and S. T and S are called isomorphic ($T \sim S$) if there is an invertible conjugacy between T and S called an isomorphism. The conjugacy ψ induces a T-invariant measurable partition η of X into sets $\psi^{-1}\{y\}$, $y \in Y$, called ψ-fibers. The m.p.t. T induces an m.p.t. on the quotient space $(X/\eta, \mu_\eta)$ which is isomorphic to S via the map $\psi_\eta: X/\eta \to Y$, $\psi_\eta(\eta(x)) = \psi(x)$, where $\eta(x)$ denotes the ψ-fiber containing $x \in X$. A factor S is called trivial if $\nu\{y\} = 1$ for some $y \in Y$. Henceforth the word "factor" means non-trivial factor.

Let T_t be an m.p. flow on (X, μ) and let τ be a positive integrable function on X. Set

$$\int_X \tau \, d\mu = \bar{\tau}.$$

We say that a flow T_t^τ is obtained from T_t by the time change τ if

$$T_t^\tau(x) = T_{w(x,t)}(x)$$

for μ-a.e. $x \in X$ and all $t \in R$, where $w(x, t)$ is defined by

$$\int_0^{w(x,t)} \tau(T_u x) \, du = t.$$

The flow T_t^τ preserves the probability measure μ_τ on X defined by $d\mu_\tau(x) = (\tau/\bar{\tau}) \, d\mu(x)$, $x \in X$.

We say that two integrable functions $\tau_1, \tau_2: (X, \mu) \to R$ are cohomologous along T_t if there is a measurable $v: X \to R$ such that

$$\int_0^t (\tau_1 - \tau_2)(T_u x) \, du = v(T_t x) - v(x)$$

for μ-a.e. $x \in X$ and all $t \in R$. One can check that two time changes τ_1 and τ_2 are cohomologous via v if and only if the map $\chi_v: X \to X$ defined by

$$\chi_v(x) = T_{v(x)}^{\tau_2} x$$

is an isomorphism between $T_t^{\tau_1}$ and $T_t^{\tau_2}$, i.e.

$$\chi_v \, T_t^{\tau_1}(x) = T_t^{\tau_2} \chi_v(x)$$

for a.e. $x \in X$. We shall call χ_v the isomorphism induced by v. If T_t is ergodic and τ_1, τ_2 are cohomologous along T_t via different measurable functions v_1 and v_2 then $v_1 - v_2$ is equal to a constant a.e.

Let $T_t^{\tau_i}$ be obtained from an m.p. flow $T_t^{(i)}$ on $(X_i, \mathscr{B}_i, \mu_i)$ by a time change τ_i, $i = 1, 2$, and let $\phi: (X_1, \mu_1) \to (X_2, \mu_2)$ be an m.p. conjugacy between $T_t^{(1)}$ and $T_t^{(2)}$. Suppose that τ_1 and $\tau_2 \circ \phi$ are cohomologous along $T_t^{(1)}$ via v. It is clear that the map $\psi = \phi \circ \chi_v$ is an m.p. conjugacy between $T_t^{\tau_1}$ and $T_t^{\tau_2}$.

Now let I be an index set and let $\mathbb{F} = \{T_t^{(i)}: i \in I\}$ be a family of m.p. flows $T_t^{(i)}$ on probability spaces (X_i, μ_i). Let K_i, $i \in I$, be a class of positive integrable functions on (X_i, μ_i) and let $\mathbb{K} = \bigcup_{i \in I} K_i$.

Definition 1. The class $\mathbb{K}$ is said to be conjugacy (isomorphism)-rigid for the family $\mathbb{F}$ if given $\tau_i \in K_i$, $\tau_j \in K_j$, $\bar{\tau}_i = \bar{\tau}_j$, $i, j \in I$, and an m.p. conjugacy (isomorphism) ψ between $T_t^{\tau_i}$ and $T_t^{\tau_j}$ there are a measurable $v: X_i \to R$ and an m.p. conjugacy (isomorphism) ϕ between $T_t^{(i)}$ and $T_t^{(j)}$ such that

(1) τ_i and $\tau_j \circ \phi$ are cohomologous along $T_t^{(i)}$ via v;

(2) $\psi = \phi \circ \chi_v$ where χ_v is the isomorphism between $T_t^{\tau_i}$ and $T_t^{\tau_j \circ \phi}$ induced by v.

Next we consider a pair of time changes for two m.p. flows, and define what is meant by saying that these time changes are "jointly cohomologous".

Let T^{τ_i} and $T^{(i)}$ on (X_i, μ_i), $i = 1, 2$, be as above. Let $X = X_1 \times X_2$ and let $f_i: X \to R^+$ be defined by

$$f_i(x_1, x_2) = \tau_i(x_i), \qquad i = 1, 2.$$

Definition 2. Let m be a joining of $T_t^{(1)}$ and $T_t^{(2)}$. We say that τ_1 and τ_2 are jointly cohomologous via m if f_1 and f_2 are cohomologous along $S_t = T_t^{(1)} \times T_t^{(2)}$ on (X, m).

More specifically, if f_1 and f_2 are cohomologous along S_t via a measurable function $v: X \to R$ then we say that τ_1 and τ_2 are jointly cohomologous via (m, v).

We have

$$v(T_t^{(1)} x_1, T_t^{(2)} x_2) - v(x_1, x_2) = \int_0^t (f_1 - f_2)(T_u^{(1)} x_1, T_u^{(2)} x_2)\, du \tag{1}$$

for m-a.e. $(x_1, x_2) \in X$ and all $t \in R$.

There exists a unique family $\{m_{x_1} : x_1 \in X_1\}$ of probability measures on X_2 such that

$$\int_X f\, dm = \int_{X_1} f^{(m)}(x_1)\, d\mu_1(x_1)$$

for every $f \in L_1(X, m)$, where

$$f^{(m)}(x_1) = \int_{X_2} f(x_1, x_2)\, dm_{x_1}(x_2)$$

and

$$m_{T_t^{(1)} x_1}(A) = m_{x_1}(T_{-t}^{(2)} A) \tag{2}$$

for every $t \in R$, $A \in \mathcal{B}_2$ and μ_1-a.e. $x_1 \in X_1$. Set

$$\tau_2^{(m)}(x_1) = f_2^{(m)}(x_1), \qquad x_1 \in X_1.$$

Similarly we define $\tau_1^{(m)}(x_2)$, $x_2 \in X_2$.

Expressions (1) and (2) show that if $v(x_1, \cdot) \in L_1(X_2, m_{x_1})$ for μ-a.e. $x_1 \in X_1$ then τ_1 and $\tau_2^{(m)}$ are cohomologous along $T_t^{(1)}$ via $v^{(m)}: X_1 \to R$. We have just proved the following.

Proposition 1. *Let τ_i be a time change for $T_t^{(i)}$ on (X_i, μ_i), $i = 1, 2$. Suppose that τ_1 and τ_2 are jointly cohomologous via (m, v) with $v(x_1, \cdot) \in L_1(X_2, m_{x_1})$ for μ_1-a.e. $x_1 \in X_1$. Then τ_1 and $\tau_2^{(m)}$ are cohomologous along $T_t^{(1)}$.*

If τ_1 and τ_2 are jointly cohomologous via (m, v) then the flows $S_t^{f_1}$ on (X, m_{f_1}) and $S_t^{f_2}$ on (X, m_{f_2}) are isomorphic via the isomorphism χ_v induced by

v. Also the map $\psi_v: X \to X$ defined by

$$\psi_v(x) = (x_1, T^{\tau_2}_{v(x)} x_2), \qquad x = (x_1, x_2)$$

is a measurable conjugacy between $S_t^{f_1}$ and $T_t^{\tau_1} \times T_t^{\tau_2}$; that is,

$$\psi_v \circ S_t^{f_1} = (T_t^{\tau_1} \times T_t^{\tau_2}) \circ \psi_v. \tag{3}$$

Let $\nu = \nu_{(m, v)}$ be the measure on $\mathscr{B}_1 \times \mathscr{B}_2$ defined by

$$\nu(A) = m_{f_1}(\psi_v^{-1}(A)), \qquad A \in \mathscr{B}_1 \times \mathscr{B}_2.$$

It follows from (3) that ν is a joining of $T_t^{\tau_1}$ and $T_t^{\tau_2}$. The flow $T_t^{\tau_1} \times T_t^{\tau_2}$ on (X, ν) is a factor of $S_t^{f_1}$ on (X, m_{f_1}) via ψ_v and a factor of $S_t^{f_2}$ on (X, m_{f_2}) via $\psi_v \circ \chi_v^{-1}$. If m is ergodic, then so is ν. We shall call $\nu_{(m, v)}$ the joining induced by (m, v).

Definition 3. Let I, $\mathbb{F}$ and $\mathbb{K}$ be as in Definition 1. The class $\mathbb{F}$ is said to be joining-rigid for the family $\mathbb{F}$ is given $\tau_i \in K_i$, $\tau_j \in K_j$, $\bar{\tau}_i = \bar{\tau}_j$, $i, j \in I$ and a nontrivial ergodic joining ν of $T_t^{\tau_i}$ and $T_t^{\tau_j}$, there are a measurable $v: X_i \times X_j \to R$ and a nontrivial ergodic joining m of $T_t^{(i)}$ and $T_t^{(j)}$ such that

(1) τ_i and τ_j are jointly cohomologous via (m, v);
(2) $\nu = \nu_{(m, v)}$ is induced by (m, v).

Next we introduce an analog of Definition 2 for factors.

Let T_t^{τ} be obtained from an m.p. flow T_t on (X, μ) by a time change τ and let ζ and η be two measurable partitions of (X, μ) respectively induced by a factor U_t of T_t and by a factor V_t of T_t^{τ}. We shall identify U_t with the flow on the quotient space $(X/\zeta, \mu_\zeta)$ induced by T_t.

There exists a unique family $\{\mu_C, C \in \zeta\}$, of probability measures μ_C on atoms of ζ such that

$$\int_X f d\mu = \int_{X/\zeta} f_\zeta(C)\, d\mu_\zeta(C)$$

for every $f \in L_1(X, \mu)$ where

$$f_\zeta(C) = \int_C f(x)\, d\mu_C(x), \qquad C \in \zeta$$

and

$$\mu_{C(T_t x)}(T_t A) = \mu_{C(x)}(A)$$

for all measurable $A \subset C(x)$, all $t \in R$ and μ-a.e. $x \in X$. Set

$$\tau_\zeta(x) = \tau_\zeta(C(x)), \qquad x \in C(x) \in \zeta.$$

Definition 4. The factor-induced partitions ζ and η are called shift-related along T_t if there is a measurable function $v(x, y) = v_x(y)$, $x \in X$, $y \in \zeta(x)$, $v(x, x) = 0$ such that for μ-a.e. $x \in X$,

(1) v_x is integrable on $(\zeta(x), \mu_{\zeta(x)})$,
(2) $\eta(x) = \{T^{\tau}_{v(x, y)} y: y \in \zeta(x)\}$.

The following proposition can be deduced from Proposition 1. It is proved in Sect. 5.

Proposition 2. *Let ζ and η be the factor-induced partitions of (X, μ) described above. Suppose that ζ and η are shift-related along T_t via $v(x, y)$, $x \in X$, $y \in \zeta(x)$. Then*

1) *τ_ζ is cohomologous to τ via $\tilde{v}: X \to R$ defined by*

$$\tilde{v}(x) = -\int_{\zeta(x)} v(x, y)\, d\mu_{\zeta(x)}(y).$$

2) *$\eta(x) = \chi_{\tilde{v}}(\zeta(\chi_{\tilde{v}}^{-1}(x)))$ for μ-a.e. $x \in X$, where $\chi_{\tilde{v}}$ is the isomorphism between $T_t^{\tau_\zeta}$ and T_t^{τ} induced by $\tilde{v}$.*

Consequently, $U_t^{\tau_\zeta} \sim V_t$.

Definition 5. Let I, $\mathbb{F}$ and $\mathbb{K}$ be as in Definition 1. The class $\mathbb{K}$ is said to be factor-rigid for the family $\mathbb{F}$ if given $\tau_i \in K_i$, $i \in I$, and a $T_t^{\tau_i}$-invariant partition η of (X_i, μ_i) there are a measurable function $v: X_i \to R$ and a $T_t^{(i)}$-invariant partition ζ such that

(1) $(\tau_i)_\zeta$ is cohomologous to τ_i via v;

(2) $\eta(x) = \chi_v(\zeta(\chi_v^{-1}(x)))$ for μ_i-a.e. $x \in X_i$, where χ_v is the isomorphism between $T_t^{(\tau_i)_\zeta}$ and $T_t^{\tau_i}$ induced by v.

Now let G denote the group $SL(2, R)$ equipped with a left invariant Riemannian metric and let $\mathcal{T}$ be the set of all discrete subgroups Γ of G such that the quotient space $M = \Gamma \backslash G = \{\Gamma g : g \in G\}$ has finite volume. Recall that the horocycle flow h_t and the geodesic flow g_t on M are defined by

$$h_t(\Gamma g) = \Gamma g \begin{pmatrix} 1 & 0 \\ t & 1 \end{pmatrix}, \qquad g_t(\Gamma g) = \Gamma g \begin{pmatrix} e^t & 0 \\ 0 & e^{-t} \end{pmatrix},$$

$g \in G$, $t \in R$. The flows h_t and g_t preserve the normalized volume measure μ on M, are ergodic and mixing on (M, μ) and obey the commutation relation

$$g_t \circ h_s = h_{s \exp(2t)} \circ g_t \tag{4}$$

for all $s, t \in R$.

It was shown in [11] that there is a large class $K(M)$ of positive square-integrable functions on (M, μ) such that for $\tau \in K(M)$ the time-changed horocycle flow h_t^τ inherits the rigidity property [8] from h_t. The class $K(M)$ specifically contains all positive bounded functions τ such that τ^{-1} is also bounded and s.t. τ is Hölder continuous in the direction of the rotation subgroup of G (see [5, 12]).

In the present paper we consider the more restricted class $B^1(M)$ of all continuously differentiable positive functions on M, which are bounded with bounded reciprocals. We will show that for time changes $\tau \in B^1(M)$, the flows h_t^τ, in addition to the above mentioned rigidity property for conjugates, also inherit the factor and joining rigidity properties [9, 10].

The flows h_t^τ with $\tau \in B^1(M)$ are known [4] to be mixing in (M, μ_τ).

Restricting the time change τ to $B^1(M)$ enables us to show that h_t^τ inherits a fundamental property of h_t, the "H-property", which we introduced in [10].

An important geometrical fact about the horocycle flow h_t is that two initially close trajectories diverge polynomially. The H-property states that this

divergence develops much faster in the direction of the flow than in any other direction. (The exact description of the H-property is rather technical; it will be given in the proof of the first theorem.)

Theorem 1. *Let h_t be the horocycle flow on $(M=\Gamma\backslash G, \mu)$, $\Gamma\in\mathscr{T}$, and let h_t^τ be obtained from h_t by a time change $\tau\in B^1(M)$. Then h_t^τ possesses the H-property.*

The significance of the H-property is reflected in the following corollaries, proved in [10].

Corollary 1. *Let h_t^τ be as in Theorem 1 and let $h^\tau=h_1^\tau$, (the "time 1" map).*

(a) *If ν is a nontrivial ergodic joining of h^τ and some m.p.t U then $(h^\tau\times U, \nu)$ is a finite extension of U.*

(b) *If $U=U_1$ for some measure preserving flow U_t then $\nu\in J(h^\tau, U)$ if and only if $\nu\in J(h_t^\tau, U_t)$ for all $t\in R$.*

Corollary 2. *Let h_t^τ be as in Theorem 1 and let $h^\tau=h_1^\tau$.*

(a) *If an m.p.t V is a factor of h^τ via a factor map ψ then a.e. fiber of ψ is finite.*

(b) *If $V=V_1$ for some m.p. flow V_t and V is a factor of h^τ via a factor map ψ then the flow V_t is a factor of h_t^τ via ψ.*

Let $h_t^{(i)}$ be the horocycle flow on $(M=\Gamma_i\backslash G, \mu_i)$, $\Gamma_i\in\mathscr{T}$ and let $h_t^{\tau_i}$ be obtained from $h_t^{(i)}$ by a time change τ_i, $i=1, 2$.

Consider $\kappa(\Gamma_1, \Gamma_2)=\{\alpha\in G: \Gamma_1\cap\alpha\Gamma_2\alpha^{-1}\in\mathscr{T}\}$. For $\alpha\in\kappa(\Gamma_1, \Gamma_2)$, let $\Gamma_\alpha=\Gamma_1\cap\alpha\Gamma_2\alpha^{-1}$, $M_\alpha=\Gamma_\alpha\backslash G$ and let $h_t^{(\alpha)}$ be the horocycle flow on (M_α, μ_α). Let $\psi_i: M_\alpha\to M_i$, $i=1, 2$, be defined by

$$\psi_1(\Gamma_\alpha g)=\Gamma_1 g, \quad \psi_2(\Gamma_\alpha g)=\Gamma_2\alpha^{-1}g, \quad g\in G.$$

The flows $h_t^{(1)}$ and $h_t^{(2)}$ are factors of $h_t^{(\alpha)}$ via ψ_1 and ψ_2 respectively. Let $\tau_{\alpha,i}: M_\alpha\to R^+$ be defined by

$$\tau_{\alpha,i}(x)=\tau_i(\psi_i(x)),$$

$x\in M_\alpha$, $i=1, 2$. We call $\tau_{\alpha,i}$ the time changes associated to α.

Definition 6. τ_1 and τ_2 are said to be algebraically cohomologous via $\alpha\in\kappa(\Gamma_1, \Gamma_2)$ if $\tau_{\alpha,1}$ and $\tau_{\alpha,2}$ are cohomologous along $h_t^{(\alpha)}$.

The following theorem is a consequence of the joinings classification theorem for horocycle flows [10, Theorem 6].

Theorem 2. *Let $h_t^{\tau_i}$ be obtained from the horocycle flow $h_t^{(i)}$ on $(M_i=\Gamma_i\backslash G, \mu_i)$, $\Gamma_i\in\mathscr{T}$ by a time change τ_i, $i=1, 2$. The functions τ_1 and τ_2 are jointly cohomologous via a nontrivial $m\in J(h_t^{(1)}, h_t^{(2)})$ if and only if τ_1 and τ_2 are algebraically cohomologous via some $\alpha\in\kappa(\Gamma_1, \Gamma_2)$.*

Theorem 3 (Main Joinings Theorem). *Let $h_t^{\tau_i}$ be as in Theorem 2. Assume that $\tau_i\in B^1(M_i)$, $i=1, 2$ and $\bar{\tau}_1=\bar{\tau}_2$. Let ν be a nontrivial ergodic joining of $h_1^{\tau_1}$ and $h_1^{\tau_2}$. Then $\nu\in J(h_t^{\tau_1}, h_t^{\tau_2})$ for all $t\in R$. Moreover, there exist a measurable $v: M_1\times M_2\to R$ and a nontrivial $m\in J(h_t^{(1)}, h_t^{(2)})$ such that*

(1) τ_1 *and* τ_2 *are jointly cohomologous via* (m, v);
(2) $v = v_{(m,v)}$ *is induced by* (m, v);
(3) $(h_t^{\tau_1} \times h_t^{\tau_2}, v) \sim (S_t^{f_1}, m_{f_1}) \sim (S_t^{f_2}, m_{f_2})$, *where* $S_t = h_t^{(1)} \times h_t^{(2)}$ *and* $f_i: M_1 \times M_2 \to R$ *is defined by* $f_i(x_1, x_2) = \tau_i(x_i)$, $i = 1, 2$.

Corollary 3. *The class* $\mathbb{B}^1 = \bigcup_{\Gamma \in \mathscr{T}} B^1(\Gamma \backslash G)$ *is joining-rigid for the family* $\mathbb{H}$ *of all horocycle flows on* $\Gamma \backslash G$, *where* Γ *ranges over all lattices in* G.

We showed in [10] that if m is a nontrivial ergodic joining of $h_t^{(1)}$ and $h_t^{(2)}$ then there are $\sigma \in R$ and $\alpha \in \kappa(\Gamma_1, \Gamma_2)$ such that the set

$$\{(x_1, h_\sigma^{(2)} x_2): (x_1, x_2) \in \Omega_\alpha\}$$

has m-measure 1, where

$$\Omega_\alpha = \{(\Gamma_1 g, \Gamma_2 \alpha^{-1} \gamma_k g): g \in G, k = 1, \ldots, n\} \subset M_1 \times M_2,$$

$n = |\Gamma_\alpha \backslash \Gamma_1|$ is the cardinal number of $\Gamma_\alpha \backslash \Gamma_1 = \{\Gamma_\alpha \gamma_k: k = 1, \ldots, n\}$.

We obtain the following algebraic version of Theorem 3.

Theorem 4 (Joinings Classification Theorem). *Let* τ_i *and* $h_t^{\tau_i}$, $i = 1, 2$, *be as in Theorem* 3. *Let* ν *be a nontrivial ergodic joining of* $h_1^{\tau_1}$ *and* $h_1^{\tau_2}$. *Then there exist* $\alpha \in \kappa(\Gamma_1, \Gamma_2)$ *and a measurable* $v: M_1 \times M_2 \to R$ *such that*

(1) τ_1 *and* τ_2 *are algebraically cohomologous via* α;
(2) $(h_t^{\tau_1} \times h_t^{\tau_2}, \nu) \sim h_t^{\tau_{\alpha,1}} \sim h_t^{\tau_{\alpha,2}}$, *where* $h_t^{\tau_{\alpha,i}}$ *is obtained from the horocycle flow* $h_t^{(\alpha)}$ *on* (M_α, μ_α) *by the time change* $\tau_{\alpha,i}$ *associated to* α, $i = 1, 2$.
(3) $\nu(\psi_v(\Omega_\alpha)) = 1$, *where* $\psi_v: M_1 \times M_2 \to M_1 \times M_2$ *is defined by*

$$\psi_v(x) = (x_1, h_{v(x)}^{\tau_2} x_2), \qquad x = (x_1, x_2).$$

Theorems 3, 4 and Proposition 1 imply the following

Corollary 4. *Let* τ_i *and* $h_t^{\tau_i}$, $i = 1, 2$, *be as in Theorem* 3.

(1) *If* $h_1^{(1)}$ *and* $h_1^{(2)}$ *are disjoint, then so are* $h_1^{\tau_1}$ *and* $h_1^{\tau_2}$;
(2) *If* $h_t^{(1)}$ *has only trivial self-joinings, then so does* $h_t^{\tau_1}$;
(3) *If* τ_1 *and* τ_2 *are not algebraically cohomologous then* $h_1^{\tau_1}$ *and* $h_1^{\tau_2}$ *are disjoint. If, in addition,* M_1 *and* M_2 *are compact then* $h_1^{\tau_1} \times h_1^{\tau_2}$ *is uniquely ergodic;*
(4) *If* $\bar{\tau}_2 = 1$ *and* τ_2 *is not cohomologous to* 1 *along* $h_t^{(2)}$ *then* $h_1^{(1)}$ *and* $h_1^{\tau_2}$ *are disjoint. In this case* $h_1^{(1)} \times h_1^{\tau_2}$ *is uniquely ergodic, if* M_1 *and* M_2 *are compact.*

The following theorem is really a particular case of Theorem 3 (or Theorem 4) (see [10], where a similar situation is discussed in detail).

Theorem 5 (Rigidity of Time Changes Theorem). *Let* τ_i *and* $h_t^{\tau_i}$, $i = 1, 2$, *be as in Theorem* 3. *Suppose that there is an m.p.* $\psi: (M_1, \mu_{\tau_1}) \to (M_2, \mu_{\tau_2})$ *such that*

$$\psi h_t^{\tau_1}(x) = h_t^{\tau_2} \psi(x)$$

for μ_{τ_1}*-a.e.* $x \in M_1$ *and all* $t \in R$. *Then there are* $\alpha \in G$ *and a measurable* $v: M_2 \to R$ *such that*

(1) $\alpha^{-1}\Gamma_1\alpha\subset\Gamma_2$;

(2) $\psi(x)=h^{\tau_2}_{v(\psi_\alpha(x))}(\psi_\alpha(x))$ *a.e. where* $\psi_\alpha(\Gamma_1 g)=\Gamma_2\alpha^{-1}g$, $g\in G$;

(3) τ_1 *and* $\tau_2\circ\psi_\alpha$ *are cohomologous along* $h_t^{(1)}$.

In [11] we proved this theorem for τ_i belonging to the class $K(M_i)$ mentioned above. Since $K(M_i)$ is much larger than $B^1(M_i)$ the proof in [11] is necessarily rather technical.

In Sect. 3 we give a proof of Theorem 5 different from that in [11] and independent of Theorems 1–4. In fact, this proof combined with some arguments from [10] yields the proof of the Main Theorem 3.

To prove part (1) of Theorem 5 we use none of the techniques needed in [11], such as the combinatorial Lemma 2.2 or the estimate of the decay rate of the correlation function of τ_i for $h_t^{(i)}$. (It is only the cohomological parts (2) and (3) of the theorem that need these techniques.) It seems plausible that this method of proof can be extended to deal with various geometrical and algebraic generalizations.

Corollary 5. *The class* $\mathbb{K}=\bigcup_{\Gamma\in\mathscr{T}} K(\Gamma\backslash G)\supset\mathbb{B}^1$ *is conjugacy and isomorphism-rigid for* $\mathbb{H}=\{h_t \text{ on } \Gamma\backslash G\colon \Gamma\in\mathscr{T}\}$. *In particular if* $\tau\in K(M)\supset B^1(M)$ $M=\Gamma\backslash G$, *then* $h_t^\tau\sim h_t^{\bar\tau}$ *if and only if* τ *is cohomologous to* $\bar\tau$ *along* h_t.

Next we classify factors of h_t^τ, $\tau\in B^1(M)$, $M=\Gamma\backslash G$, $\Gamma\in\mathscr{T}$.

We showed in [9] that if U_t is a factor of h_t on (M,μ) then there exists $\Gamma'\supset\Gamma$, $\Gamma'\in\mathscr{T}$ such that the h_t-invariant partition ζ of M induced by U_t has the form

$$\zeta(\Gamma g)=\psi^{-1}_{\Gamma,\Gamma'}\{\Gamma' g\},$$

where

$$\psi_{\Gamma,\Gamma'}(\Gamma g)=\Gamma' g,\qquad g\in G.$$

We shall denote ζ by $\zeta(\Gamma,\Gamma')$.

Theorem 6 (Factors Theorem). *The class* $\mathbb{B}^1$ *is factor-rigid for the family* $\mathbb{H}=\{h_t \text{ on } \Gamma\backslash G\colon \Gamma\in\mathscr{T}\}$. *More precisely, let* h_t^τ *be obtained from the horocycle flow* h_t *on* $(M=\Gamma\backslash G,\mu)$, $\Gamma\in\mathscr{T}$ *by a time change* $\tau\in B^1(M)$. *Let* V *be a factor of* h_1^τ *and* η *be the* h_1^τ*-invariant partition of* M *induced by* V. *Then* η *is invariant under* h_t^τ *for all* $t\in R$. *Moreover, there exist* $\Gamma'\supset\Gamma$, $\Gamma'\in\mathscr{T}$ *and a measurable* $v\colon M\to R$ *such that*

(1) *the function* τ_ζ *defined by* $\tau_\zeta(x)=(\sum_{y\in\zeta(x)}\tau(y))/|\Gamma\backslash\Gamma'|$ *is cohomologous to* τ *via* v, *where* $\zeta=\zeta(\Gamma,\Gamma')$;

(2) $\eta(x)=\chi_v(\zeta(\chi_v^{-1}(x)))$ *for* μ*-a.e.* $x\in M$, *where* χ_v *is the isomorphism between* $h_t^{\tau_\zeta}$ *and* h_t^τ *induced by* v;

(3) $V_t\sim h_t^{\tau'}$, *where* $h_t^{\tau'}$ *is obtained from the horocycle flow* h'_t *on* $(M'=\Gamma'\backslash G,\mu')$ *by the time change* $\tau'(\Gamma' g)=\tau_\zeta(\psi^{-1}_{\Gamma,\Gamma'}\{\Gamma' g\})$, $g\in G$, *and* V_t *is the flow on the quotient space* M/η *induced by* h_t^τ, $V_1\sim V$.

Corollary 6. *Let* h_t^τ *be as in Theorem* 6.

(1) *The number of nonisomorphic factors of* h_1^τ *is finite.*

(2) *If* h_1 *has only trivial factors then so does* h_1^τ.

We assumed in Theorems 3–5 that $\bar{\tau}_1 = \bar{\tau}_2$. The case $\bar{\tau}_1 \neq \bar{\tau}_2$ can be derived from the case $\bar{\tau}_1 = \bar{\tau}_2$ using the commutation relation (4), just as in [11].

Indeed, let $a = \bar{\tau}_1 \neq \bar{\tau}_2 = b$ and let $s = \frac{(\log(b/a))}{2}$. Let σ_2 on M_2 be defined by $\sigma_2(x) = \frac{a\tau_2(g_{-s}^{(2)} x)}{b}$. We have $\bar{\sigma}_2 = a = \bar{\tau}_1$.

Relation (4) shows that $h_t^{\tau_2}$ and $h_t^{\sigma_2}$ are isomorphic via $g_s^{(2)}$, that is,

$$g_s^{(2)} \circ h_t^{\tau_2} = h_t^{\sigma_2} \circ g_s^{(2)}.$$

Let $\psi_s : M_1 \times M_2 \to M_1 \times M_2$ be defined by $\psi_s(x_1, x_2) = (x_1, g_s^{(2)} x_2)$. It is clear that a measure ν on $M_1 \times M_2$ is a nontrivial ergodic joining of $h_t^{\tau_1}$ and $h_t^{\tau_2}$ if and only if the measure $\nu_s = \psi_s^* \nu$ is a nontrivial ergodic joining of $h_t^{\tau_1}$ and $h_t^{\sigma_2}$.

Thus in order to state Theorem 3 for ν one should replace τ_2 by σ_2 in (1) and say "ν_s is induced by (m, ν)" in (2).

Similar changes must be made in Corollary 4 and Theorems 4 and 5 for the case $\bar{\tau}_1 \neq \bar{\tau}_2$.

Finally we pose the following problems.

Problem 1. Does there exist a function $\tau \in B^1(M)$ which is not cohomologous to $\bar{\tau}$ along h_t?

This problem is motivated by a theorem of Kolmogorov [3], stating that there is a subset $A \subset R$ of full Lebesgue measure such that if an irrational α belongs to A then every C^∞-function τ on the 2-torus is cohomologous to $\bar{\tau}$ along the irrational flow induced by α. (In fact, τ is cohomologous to $\bar{\tau}$ via a C^∞-function v.) This, however, is not true for many irrational $\alpha \notin A$.

Problem 2. Suppose that $\tau \in B^1(M)$ and τ is cohomologous to $\bar{\tau}$ along h_t via v. Is v integrable? continuous? smooth?

Problem 3. Are there other naturally arising families of ergodic m.p. flows (possibly consisting of just one m.p. flow) for which the class of smooth functions is conjugacy-rigid? joining-rigid? factor-rigid? Do unipotent flows on finite volume homogeneous spaces of semi-simple Lie groups form such a family? (see [14]). What about uniformly parameterized horocycle flows for surfaces of variable negative curvature? (see [1]). Is the class of smooth functions conjugacy-rigid for irrational flows on the 2-torus or more generally for nilflows on compact nilmanifolds? (see [6, 7] for rigidity properties of nilflows).

1. Basic lemma

Let h_t and g_t be the horocycle and the geodesic flows on $(M = \Gamma \backslash G, \mu)$, $\Gamma \in \mathcal{T}$ respectively. We shall also consider the flow k_t on (M, μ) defined by

$$k_t(\Gamma g) = \Gamma g \begin{pmatrix} 1 & t \\ 0 & 1 \end{pmatrix}$$

$g \in G$, $t \in R$.

One has

$$g_s \circ k_t = k_{t \exp(-2s)} \circ g_s$$

for all $s, t \in R$.

The flows k_t and h_t satisfy the following commutation relation

$$h_t \circ k_b = k_{c(b,t)} \circ g_{r(b,t)} \circ h_{\sigma(b,t)} \tag{1.1}$$

for all $b, t \in R$ with $bt > -1$, where

$$\begin{aligned} r(b,t) &= \log(1+bt), \\ \sigma(b,t) &= te^{-r(b,t)}, \\ c(b,t) &= be^{-r(b,t)}. \end{aligned} \tag{1.2}$$

Denote $\Delta(b,t) = \sigma(b,t) - t = \dfrac{-bt^2}{1+bt}$.

For $a, b \in R$, $t > 0$, $bt > -1$ we have

$$h_t \circ k_b \circ g_a = k_{c(b,t)} \circ g_{r(a,b,t)} \circ h_{\sigma(a,b,t)}$$

where

$$\begin{aligned} r(a,b,t) &= a + r(b,t), \\ \sigma(a,b,t) &= e^{-2a}\sigma(b,t). \end{aligned}$$

It follows from (1.2) that there exists $0 < r_0 \leqq \frac{1}{2}$ such that if $|a|, |bt| \leqq r_0$ then

$$\begin{gathered} |e^{-2a} - 1| \leqq 0.1 \\ |\Delta(b,t)|/2t \leqq |r(b,t)| \leqq 2|\Delta(b,t)|/t \\ |\Delta(b,t)|/2t^2 \leqq |b| \leqq 2|\Delta(b,t)|/t^2 \\ |\Delta(b,s) - s^2\Delta(b,t)/t^2| \leqq 0.01\,|\Delta(b,t)| \end{gathered} \tag{1.3}$$

for all $0 \leqq s \leqq t$.

Denote

$$\Delta(a,b,t) = \sigma(a,b,t) - t.$$

We have

$$\Delta(a,b,t) = e^{-2a}\Delta(b,t) - t(1 - e^{-2a}). \tag{1.4}$$

Lemma 1.1. *Let $a, b \in R$, $t > 0$ be such that $|a|, |b| \leqq r_0$. Then*

$$|\Delta(b,t)| \leqq 20 \max\{|\Delta(a,b,s)| : 0 \leqq s \leqq t\}.$$

Proof. Denote $\Delta_0 = \max\{|\Delta(a,b,s)| : 0 \leqq s \leqq t\}$. We have from (1.4) and (1.3),

$$\begin{aligned} \Delta(a,b,t/2) &= e^{-2a}\Delta(b,t/2) - t(1-e^{-2a})/2, \\ e^{-2a}0.1\,|\Delta(b,t)| &\leqq e^{-2a}(|\Delta(b,t)|/2 - |\Delta(b,t/2)|) \\ &\leqq |\Delta(a,b,t)/2 - \Delta(a,b,t/2)| \leqq 3\Delta_0/2 \end{aligned}$$

since $|\Delta(b,t/2)| \leqq 0.4\,|\Delta(b,t)|$ by (1.3). This implies that $|\Delta(b,t)| \leqq 20\Delta_0$ by (1.3). □

Corollary 1.1. *If* $|a|, |bt| \leqq r_0$ *and*

$$|\Delta(a, b, t)| = \max\{|\Delta(a, b, s)| : 0 \leqq s \leqq t\}$$

then

$$|\Delta(a, b, \beta t)| \geqq 0.4\,|\Delta(a, b, t)|$$

for all $\beta \in [0.99, 1]$.

Proof. We have

$$\begin{aligned}\Delta(a, b, \beta t) &= e^{-2a}\Delta(b, \beta t) + \beta t(e^{-2a} - 1)\\ &= e^{-2a}(\Delta(b, \beta t) - \beta\Delta(b, t)) + \beta\Delta(a, b, t)\end{aligned}$$

and

$$|\Delta(a, b, \beta t)| \geqq \beta|\Delta(a, b, t)| - e^{-2a}|\Delta(b, \beta t) - \beta\Delta(b, t)|. \tag{1.5}$$

Also,

$$\begin{aligned}|\Delta(b, \beta t) - \beta\Delta(b, t)| &\leqq |\Delta(b, \beta t) - \beta^2\Delta(b, t)| + |\Delta(b, t)|\,|\beta - \beta^2|\\ &\leqq 0.02\,|\Delta(b, t)| \leqq 0.4\,|\Delta(a, b, t)|\end{aligned}$$

by Lemma 1.1. Thus (1.3) and (1.5) imply

$$|\Delta(a, b, \beta t)| \geqq 0.4\,|\Delta(a, b, t)|$$

if $\beta \in [0.99, 1]$. □

Let τ be a positive bounded function on (M, μ) and let $x \in M$, $y = g_{-a}k_b x$ for some $a, b \in R$. For $t > -1/|b|$, denote

$$\xi(x; a, b, t) = \int_0^{\sigma(a, b, t)} \tau(h_u y)\,du,$$

$$\xi(x; 0, 0, t) = \xi(x, t) = \int_0^t \tau(h_u x)\,du,$$

$$\xi(x; 0, b, t) = \xi(x; b, t).$$

Also let

$$\Delta_\tau(x; a, b, t) = \xi(x; a, b, t) - \xi(x, t).$$

We have

$$\Delta_\tau(x; a, b, t) = \Delta_\tau(x; b, t) + \Delta_\tau(g_a y; a, 0, \sigma(b, t)). \tag{1.6}$$

Lemma 1.2 (Basic). *Assume* $\tau \in B^1(M)$. *Then given* $\varepsilon > 0$ *there are* $\gamma = \gamma(\varepsilon) > 0$, $l = l(\varepsilon) > 0$ *and a subset* $E = E(\varepsilon) \subset M$ *with* $\mu(E) > 1 - \varepsilon$ *such that if* $x, y \in E$, $x = k_b g_a y$ *for some* $|a| \leqq \gamma$, $|b| \leqq \gamma/l$ *then*

$$|\Delta_\tau(x; a, b, t) - \bar{\tau}\Delta(a, b, t)| \leqq \varepsilon \max\{|\Delta(a, b, s)| : s \in [0, t]\} \tag{1.7}$$

for all $l \leqq t \leqq \gamma/|b|$.

Proof. Denote

$$\tau_g(x) = \lim_{t \to 0} \frac{\tau(g_t x) - \tau(x)}{t},$$

$$\tau_k(x) = \lim_{t \to 0} \frac{\tau(k_t x) - \tau(x)}{t}.$$

The functions τ_g and τ_k are continuous on M and

$$|\tau(x)|, \quad |\tau_g(x)|, \quad |\tau_k(x)| \leqq L_\tau \tag{1.8}$$

for some $L_\tau \geqq 1$ and all $x \in M$. We have

$$\int_M \tau_g \, d\mu = \int_M \tau_k \, d\mu = 0.$$

Let $\varepsilon > 0$ be given and let $0 < \theta = \theta(\varepsilon) < 0.01\,\varepsilon$ be chosen later. Let $K \subset M$ be an open subset of M such that $\bar{K}$ is compact and

$$\mu(K) > 1 - \theta, \quad \mu(\partial K) = 0$$

where ∂K denotes the boundary of K.

Let $0 < \bar{\gamma} = \bar{\gamma}(\theta) < \min\{r_0, \theta\}$ be such that $\mu(O_{\bar{\gamma}}(\partial K)) \leqq \theta$ and if $u, v \in \bar{K}$, $d(u, v) \leqq \bar{\gamma}$ then

$$\begin{aligned} |\tau_g(u) - \tau_g(v)| &\leqq \theta, \\ |\tau_k(u) - \tau_k(v)| &\leqq \theta, \end{aligned} \tag{1.9}$$

where $O_{\bar{\gamma}}(\partial K)$ denotes the $\bar{\gamma}$-neighborhood of ∂K.

Denote

$$K_{\bar{\gamma}} = K - O_{\bar{\gamma}}(\partial K), \quad \mu(K_{\bar{\gamma}}) \geqq 1 - 2\theta.$$

It follows from (1.2) that there is $0 < \gamma \leqq 0.01\,\bar{\gamma}$ such that if $|bt| \leqq \gamma$ then

$$\begin{aligned} &|e^{-2\gamma}\sigma(b, t)| \geqq 0.99\,t \\ &\left|\frac{e^{-2r(b,s)} - 1}{\Delta(b, t)/t} - \frac{2s}{t}\right| \leqq \theta \end{aligned} \tag{1.10}$$

for all $0 \leqq s \leqq t$.

Since h_t is ergodic, there exist $t_1 = t_1(\varepsilon) > 0$ and a subset $E = E(\varepsilon) \subset M$ with $\mu(E) > 1 - \varepsilon$ such that if $u \in E$, $t \geqq t_1$ then the relative length measure of $K_{\bar{\gamma}}$ on the orbit interval $[u, h_t u]$ is at least $1 - 3\theta$ and

$$\begin{aligned} \left|\frac{1}{t}\int_0^t \tau(h_s u)\, ds - \bar{\tau}\right| &\leqq \theta, \\ \left|\frac{1}{t}\int_0^t \tau_g(h_s u)\, ds\right| &\leqq \theta. \end{aligned} \tag{1.11}$$

Choose $l = l(\varepsilon) > t_1$ such that

$$t_1/l \leqq \theta. \tag{1.12}$$

Let us show that if $x, y \in E$, $x = k_b g_a y$ for some $|a| \leqq \gamma$, $|b| \leqq \gamma/l$, and $l \leqq t \leqq \gamma|b|$ then (1.7) holds if θ is sufficiently small.

Indeed we have from (1.6),

$$\Delta_\tau(x; a, b, t) = \Delta_\tau(g_a y; a, 0, \sigma(b, t)) + \Delta_\tau(x; b, t) = \bar{\Delta}_\tau + \tilde{\Delta}_\tau.$$

Also,

$$\Delta(a, b, t) = \sigma(b, t)(e^{-2a} - 1) + (\sigma(b, t) - t) = \bar{\Delta} + \tilde{\Delta}.$$

We shall estimate $\bar{\Delta}_\tau/\bar{\Delta}$ and $\tilde{\Delta}_\tau/\tilde{\Delta}$.

1) First let us estimate $\bar{\Delta}_\tau/\bar{\Delta}$. We have

$$\bar{\Delta}_\tau = \int_0^{\sigma(a,b,t)} \tau(h_s y)\,ds - \int_0^{\sigma(b,t)} \tau(h_s g_a y)\,ds$$

$$\int_0^{\sigma(b,t)} \tau(h_s g_a y)\,ds = \int_0^{\sigma(b,t)} \tau(g_a h_{se^{-2a}} y)\,ds = e^{2a}\int_0^{\sigma(a,b,t)} \tau(g_a h_s y)\,ds$$

$$= e^{2a}\left[\int_0^{\sigma(a,b,t)} \tau(h_s y)\,ds + a\int_0^{\sigma(a,b,t)} \tau_g(y_s)\,ds\right]$$

where $y_s = g_{a(s)} h_s y$ for some $a(s)\in[0,a]$. This implies that

$$\bar{\Delta}_\tau = (1-e^{2a})\int_0^{\sigma(a,b,t)} \tau(h_s y)\,ds - ae^{2a}\int_0^{\sigma(a,b,t)} \tau_g(y_s)\,ds = I_1 + I_2.$$

We have

$$I_1 = \bar{\Delta}\left(\int_0^{\sigma(a,b,t)} \tau(h_s y)\,ds\right)\Big/\sigma(a,b,t) = \bar{\Delta}(\bar{\tau}+\theta_1) \tag{1.13}$$

where $|\theta_1|\leqq\theta$ by (1.11), since $\sigma(a,b,t)\geqq t_1$ by (1.10).

For I_2 we have

$$I_2 = ae^{2a}\int_0^{\sigma(a,b,t)} \tau_g(h_s y)\,ds + ae^{2a}\int_0^{\sigma(a,b,t)} (\tau_g(y_s)-\tau_g(h_s y))\,ds = I_2' + I_2''.$$

Also

$$I_2' = a\sigma(b,t)\,\theta_2 = \frac{1}{e^{-2a}-1}\,\bar{\Delta}\cdot\theta_2$$

where $|\theta_2|\leqq\theta$ by (1.11). Thus

$$|I'|\leqq|\bar{\Delta}|\,\theta/2. \tag{1.14}$$

We have

$$|I_2''|\leqq|a|\,e^{2a}\int_0^{\sigma(a,b,t)} |\tau_g(y_s)-\tau_g(h_s y)\,ds$$

$$\leqq|a|\,e^{2a}\int_{K_y} |\tau_g(y_s)-\tau_g(h_s y)|\,ds + 6|a|\,e^{2a}\,\sigma(a,b,t)\,L_\tau\theta$$

$$\leqq|\bar{\Delta}|\,\theta(1-3\theta+6L_\tau)/2 \tag{1.15}$$

by (1.8) and (1.9) since $y\in E$, $d(h_s y, y_s)\leqq|a|\leqq\gamma$ and $y_s\in\bar{K}$, if $h_s y\in K_{\bar{\gamma}}$, where $K_y = \{s\in[0,\sigma(a,b,t)]: h_s y\in K_{\bar{\gamma}}\}$.

It follows now from (1.13), (1.14) and (1.15) that we can choose $\theta=\theta(\varepsilon)>0$ such that

$$|\bar{\Delta}_\tau - \bar{\tau}\bar{\Delta}|\leqq\varepsilon\bar{\Delta}. \tag{1.16}$$

2) Now let us estimate $\tilde{\Delta}_\tau/\tilde{\Delta}$. For $s\in[0,t]$ denote

$$c(s)=c(b,s),\qquad r(s)=r(b,s),\qquad \sigma(s)=\sigma(b,s).$$

We have from (1.2),

$$c(s)=be^{-r(s)},\qquad \sigma'(s)=e^{-2r(s)},\qquad \tilde{\Delta}=t(e^{-r(t)}-1). \tag{1.17}$$

Thus

$$\tilde{\Delta}_\tau = \int_0^{\sigma(t)} \tau(h_s z)\, ds - \int_0^t \tau(h_s x)\, ds$$

where $z = k_{-b} x$. We have

$$\begin{aligned}\int_0^{\sigma(t)} \tau(h_s z)\, ds &= \int_0^t \tau(h_{\sigma(s)} z)\, \sigma'(s)\, ds \\ &= \int_0^t \tau(g_{-r(s)} k_{-c(s)} h_s x)\, e^{-2r(s)}\, ds \\ &= \int_0^t \tau(h_s x)\, e^{-2r(s)}\, ds - b \int_0^t e^{-3r(s)}\, \tau_k(x_s)\, ds \\ &\quad - \int_0^t r(s) \cdot \tau_g(z_s)\, e^{-2r(s)}\, ds\end{aligned}$$

where $x_s = k_{c_s} h_s x$ for some $c_s \in [-c(s), 0]$ and $z_s = g_{r_s} h_{\sigma(s)} z$ for some $r_s \in [0, r(s)]$. This implies

$$\begin{aligned}\tilde{\Delta}_\tau &= \int_0^t (e^{-2r(s)} - 1)\, \tau(h_s x)\, ds + b \int_0^t e^{-3r(s)}\, \tau_k(x_s)\, ds \\ &\quad + \int_0^t r(s)\, \tau_g(z_s)\, e^{-2r(s)}\, ds = J_1 + J_2 + J_3. \end{aligned} \tag{1.18}$$

Using (1.10) we have

$$J_1 = \frac{2\tilde{\Delta}(t)}{t^2} \int_0^t s\tau(h_s x)\, ds + \frac{\tilde{\Delta}(t)}{t} \int_0^t \theta_3(s)\, \tau(h_s x)\, ds = J_1' + J_1'' \tag{1.19}$$

where $|\theta_3(s)| \leqq \theta$ for all $0 \leqq s \leqq t$. Thus

$$|J_1''| \leqq |\tilde{\Delta}|\, \theta(\bar{\tau} + \theta) \tag{1.20}$$

by (1.11), since $x \in A_0$.

For J_1' we have integrating by parts

$$\begin{aligned}\frac{1}{t^2} \int_0^t s\tau(h_s x)\, ds &= \frac{1}{t} \int_0^t \tau(h_s x)\, ds - \frac{1}{t^2} \int_0^t \left(\int_0^s \tau(h_u x)\, du \right) ds \\ &= \bar{\tau} + \theta_1(t) - \frac{1}{t^2} \int_{t_1}^t \left(\int_0^s \tau(h_u x)\, du \right) ds + \theta_4(t) \\ &= \bar{\tau} + \theta_1(t) - \frac{1}{t^2} \int_{t_1}^t s(\bar{\tau} + \theta_1(s))\, ds + \theta_4(t)\end{aligned}$$

where $|\theta_1(s)| \leqq \theta$ for all $t_1 \leqq s \leqq t$ and $|\theta_4(t)| \leqq \theta \cdot L_\tau$ by (1.11), (1.8) and (1.12). This, (1.19) and (1.20) imply that

$$\left| \frac{J_1}{\tilde{\Delta}} - \bar{\tau} \right| \leqq \theta(L_\tau + \bar{\tau} + 10). \tag{1.21}$$

For J_2 we have, using (1.3),

$$|J_2| \leqq \frac{4|\tilde{\Delta}|}{t} L_\tau \leqq \theta L_\tau \tag{1.22}$$

by (1.17), since $|r(t)| \leqq \gamma \leqq 0.01\,\theta$.

For J_3 we have

$$\begin{aligned}\left| J_3 - \int_0^t r(s)\,\tau_g(h_s x)\,e^{-2r(s)}\,ds \right| &\leqq 2|r(t)|\Big[\int_{K_x} |\tau_g(z_s)-\tau_g(h_s x)|\,ds + 6L_\tau t\theta\Big] \\ &\leqq 2t|r(t)|(\theta(1-3\theta)+6L_\tau\theta) \\ &\leqq 2|\tilde{\Delta}|\,\theta(4+6L_\tau)\end{aligned} \tag{1.23}$$

by (1.8) and (1.9), since $d(h_s x, z_s) \leqq 2\gamma < \bar{\gamma}$, $z_s \in \bar{K}$ if $h_s x \in K_{\bar{\gamma}}$, where

$$K_x = \{s \in [0, t] : h_s x \in K_{\bar{\gamma}}\}.$$

We also have

$$\begin{aligned}\left| \int_0^t r(s)\,\tau_g(h_s x)\,e^{-2r(s)}\,ds - \int_0^t r(s)\,\tau_g(h_s x)\,ds \right| &\leqq |\tilde{\Delta}/t|(2+\theta)\cdot|r(t)|\cdot t\cdot L_\tau \\ &\leqq 3|\tilde{\Delta}|\,\theta L_\tau\end{aligned} \tag{1.24}$$

by (1.10).

Finally by using integration by parts we get

$$\int_0^t r(s)\,\tau_g(h_s x)\,ds = r(t)\int_0^t \tau_g(h_s x)\,ds - \int_0^t r'(s)\left(\int_0^s \tau_g(h_u x)\,du\right) ds = Q_1 + Q_2$$

where $r'(s) = be^{-r(s)}$ by (1.2). We have

$$|Q_1| \leqq 2\theta|\tilde{\Delta}| \tag{1.25}$$

by (1.3) and (1.11). Also,

$$|Q_2| \leqq |b|\left[\int_0^{t_1} e^{-r(s)}\left|\int_0^s \tau_g(h_u x)\,du\right| ds + 2\theta\int_{t_1}^t s\,ds\right] \leqq 2|\tilde{\Delta}|\,\theta(L_\tau+1)$$

by (1.3), (1.8) and (1.12). This and (1.18), (1.21)–(1.25), imply that

$$|\tilde{\Delta}_\tau - \bar{\tau}\tilde{\Delta}| \leqq 100\,\theta|\tilde{\Delta}|(L_\tau+\bar{\tau}+1).$$

This, (1.16) and Lemma 1.1 show that

$$|\Delta_\tau(x; a, b, t) - \bar{\tau}\Delta(a, b, t)| \leqq \varepsilon \max\{|\Delta(a, b, s)| : s \in [0, t]\}$$

if $\theta = \theta(\varepsilon)$ is sufficiently small. This completes the proof of the lemma. □

Remark 1.1. It follows from the proof of Lemma 1.2 that we can set $E = M$ in the lemma if M is compact, since in this case h_t is uniquely ergodic.

2. The H-property

In this section we shall prove Theorem 1. Let us state this theorem in full.

Theorem 2.1 (The H-property, see [10]). *Suppose* $\tau\in B^1(M)$, $M=\Gamma\backslash G$, $\Gamma\in\mathscr{T}$. *Then the flow* h_t^τ *possesses the following property: given* $p>0$, $\varepsilon>0$, $N>1$, *there are* $\alpha=\alpha(p,\varepsilon)>0$, $\delta=\delta(p,\varepsilon,N)>0$ *and a subset* $B=B(p,\varepsilon,N)\subset M$ *with* $\mu(B)>1-\varepsilon$ *such that if* $x, y\in B$, $d(x,y)\leqq\delta$ *and* y *is not on the* h_t^τ*-orbit of* x, *then there are* $L=L(x,y)>0$ *and* $Q=Q(x,y)\geqq N$ *with* $Q/L\geqq\alpha$ *such that either* $d(h_{np}^\tau, h_{(n+1)p}^\tau)\leqq\varepsilon$ *for all integers* $n\in[L, L+Q]$ *or* $d(h_{np}^\tau, h_{(n-1)p}^\tau)\leqq\varepsilon$ *for all integers* $n\in[L, L+Q]$, *where* d *denotes the distance in* M.

Proof. First let us show that Theorem 2.1 follows from the following property: given $p, \varepsilon, N>0$ there are $\tilde{\alpha}=\tilde{\alpha}(p,\varepsilon)>0$, $\tilde{\delta}=\tilde{\delta}(p,\varepsilon,N)>0$ and a subset $\tilde{B}=\tilde{B}(p,\varepsilon,N)\subset M$ with $\mu(\tilde{B})>1-\varepsilon$ such that if $x, y\in\tilde{B}$, $x=h_c k_b g_a y$ for some $|c|$, $|a|$, $|b|\leqq\delta$, $|a|+|b|\neq 0$ then there are $\tilde{L}=\tilde{L}(x,y)>0$ and $\tilde{Q}=\tilde{Q}(x,y)>N$ with $\tilde{Q}/\tilde{L}=\tilde{\alpha}$ such that

$$|b|(\tilde{L}+\tilde{Q})\leqq\varepsilon$$

and

$$\begin{aligned} &\text{either} \quad |\Delta_\tau(x;a,b,t)-p|\leqq\varepsilon \quad \text{for all} \quad t\in[\tilde{L}, \tilde{L}+\tilde{Q}] \\ &\text{or} \quad |\Delta_\tau(x;a,b,t)+p|\leqq\varepsilon \quad \text{for all} \quad t\in[\tilde{L}, \tilde{L}+\tilde{Q}], \end{aligned} \tag{2.1}$$

where $\Delta_\tau(x;a,b,t)$ is as in Sect. 1.

Indeed, let $p, \varepsilon>0$, $N>1$, be given and let $\varepsilon_1=\min\{\varepsilon/3, \tilde{\alpha}(p,\varepsilon/3)\}$.

Since h_t is ergodic, there are $t_1\geqq 2Np/\bar{\tau}$ and $B_1\subset M$, $\mu(B_1)>1-0.4\,\varepsilon_1/1+\varepsilon_1$ such that if $u\in B_1$, $t\geqq t_1$ then

$$\bar{\tau}t/2\leqq\int_0^t\tau(h_s u)\,ds\leqq 3\bar{\tau}t/2. \tag{2.2}$$

Also there are $t_2\geqq 2t_1$ and $B_2\subset M$, $\mu(B_2)>1-\varepsilon/3$ such that if $u\in B_2$, $t\geqq t_2$ then the relative length measure of B_1 on the orbit interval $[u, h_t u]$ is at least $1-0.5\,\varepsilon_1/1+\varepsilon_1$.

Now set $\alpha=\tilde{\alpha}(p,\varepsilon/3)/6=\tilde{\alpha}/6$, $\delta=\tilde{\delta}(p,\varepsilon/3,t_2)/3=\tilde{\delta}/3$, $B=B_1\cap B_2\cap\tilde{B}(p,\varepsilon/3,t_2)$, $\mu(B)>1-\varepsilon$, and show that α, δ and B satisfy Theorem 2.1.

Indeed, let $x, y\in B$, $d(x,y)\leqq\delta$ and y is not on the h_t^τ-orbit of x. Then $x=h_c k_b g_a y$ for some $|c|$, $|b|$, $|a|\leqq\tilde{\delta}$ and $|a|+|b|\neq 0$.

Since $x, y\in\tilde{B}(p,\varepsilon/3,t_2)$ there are $\tilde{L}>0$, $\tilde{Q}\geqq t_2$ with $\tilde{Q}/\tilde{L}=\tilde{\alpha}$ such that (2.1) holds with $\varepsilon/3$ instead of ε. We have

$$\alpha\tilde{L}=\tilde{Q}\geqq t_2\geqq 2t_1, \qquad \tilde{L}\geqq t_2. \tag{2.3}$$

Since $x\in B_2$ there is $s\in[\tilde{L}, \tilde{L}+\tilde{Q}]$ such that $h_s x\in B_1$. This and (2.3) imply that

$$\xi(h_{\tilde{L}}x, \tilde{Q})=\int_{\tilde{L}}^{\tilde{L}+\tilde{Q}}\tau(h_s x)\,ds\geqq\tilde{Q}\bar{\tau}/4\geqq t_1\bar{\tau}/2\geqq Np.$$

Also

$$\xi(x, \tilde{L})\leqq 3\bar{\tau}\tilde{L}/2 \tag{2.4}$$

by (2.2), since $x\in B_1$.

Set $L=\xi(x, \tilde{L})/p$, $Q=\xi(h_{\tilde{L}}x, \tilde{Q})/p$. We have from (2.3) and (2.4)

$$Q/L \geqq \tilde{Q}/6\tilde{L}=\tilde{\alpha}/6=\alpha, \qquad Q \geqq N.$$

It is clear that if $\xi(x, t)=np$ for some integer $n\in[L, L+Q]$ then $t\in[\tilde{L}, \tilde{L}+\tilde{Q}]$ and either $d(h^{\tau}_{np}, h^{\tau}_{(n+1)p}) \leqq \varepsilon$ or $d(h^{\tau}_{np}, h^{\tau}_{(n-1)p}) \leqq \varepsilon$ by (1.2) and (2.1) with $\varepsilon/3$ instead of ε.

Thus (2.1) implies Theorem 2.1. Now we shall prove (2.1).

It was shown in [9] that the horocycle flow h_t satisfies the following stronger form of (2.1): given $p, \varepsilon, N>0$ there are $0<\bar{\alpha}=\bar{\alpha}(p, \varepsilon)<1$ and $\bar{\delta}=\bar{\delta}(p, \varepsilon, N)>0$ such that if $|a|, |b| \leqq \delta$, $|a|+|b| \neq 0$ then there are $\bar{L}=\bar{L}(a, b)>0$, $\bar{Q}=\bar{Q}(a, b) \geqq N$ with $\bar{Q}/\bar{L}=\bar{\alpha}$ such that

$$|b|(\bar{L}+\bar{Q}) \leqq \varepsilon$$

$$\max\{|\Delta(a, b, s)|: s\in[0, \bar{L}]\} \leqq p$$

and

$$\begin{aligned} &\text{either} \quad |\Delta(a, b, t)-p| \leqq \varepsilon \quad \text{for all } t\in[\bar{L}, \bar{L}+\bar{Q}] \\ &\text{or} \quad |\Delta(a, b, t)+p| \leqq \varepsilon \quad \text{for all } t\in[\bar{L}, \bar{L}+\bar{Q}]. \end{aligned}$$

We use this and Lemma 1.2 to prove (2.1). Indeed, let $p, \varepsilon, N>0$ be given. Let $q=p/\bar{\tau}$ and let

$$\begin{aligned} &\gamma=\gamma(\varepsilon/4q)<\min\{\varepsilon/4q, \varepsilon/4\bar{\tau}, \varepsilon/4\}, \\ &l=l(\varepsilon/4q), \\ &E=E(\varepsilon/4q)\subset M, \qquad \mu(E)>1-\varepsilon \end{aligned} \tag{2.5}$$

be as in Lemma 1.2.

Set $\tilde{\alpha}=\bar{\alpha}(q, \gamma)$, $\tilde{\delta}=\min\{\gamma/l, \bar{\delta}(q, \gamma, l/\tilde{\alpha})\}$, $\tilde{B}=E(\varepsilon/4q)$, $\mu(\tilde{B})>1-\varepsilon$. We claim that $\tilde{\alpha}, \tilde{\delta}, \tilde{B}$ satisfy (2.1).

Indeed, let $x, y\in\tilde{B}$, $x=h_c k_b g_a y$ for some $|c|, |a|, |b| \leqq \tilde{\delta}$, $|a|+|b| \neq 0$. Set $\tilde{L}=\tilde{L}(x, y)=\bar{L}(a, b)>0$, $\tilde{Q}=\tilde{Q}(x, y)=\bar{Q}(a, b)>0$.

We have, by our choice of $\tilde{\delta}$,

$$\tilde{Q} \geqq t_0/\tilde{\alpha} \geqq N, \qquad \tilde{\alpha}\tilde{Q}=\tilde{L} \geqq t_0. \tag{2.6}$$

Also,

$$|b|(\tilde{L}+\tilde{Q}) \leqq \gamma \leqq \varepsilon/4$$

$$\max\{|\Delta(a, b, s)|: s\in[0, \tilde{L}]\} \leqq q$$

and

$$\begin{aligned} &\text{either} \quad |\Delta(a, b, t)-q| \leqq \gamma \quad \text{for all } t\in[\tilde{L}, \tilde{L}+\tilde{Q}] \\ &\text{or} \quad |\Delta(a, b, t)+q| \leqq \gamma \quad \text{for all } t\in[\tilde{L}, \tilde{L}+\tilde{Q}]. \end{aligned} \tag{2.7}$$

It follows from (2.6), (2.7) and Lemma 1.2 that

$$|\Delta_{\tau}(x; a, b, t)-\bar{\tau}\Delta(a, b, t)| \leqq \varepsilon/4$$

for all $t\in[\tilde{L}, \tilde{L}+\tilde{Q}]$. This, (2.5) and (2.7) imply (2.1). This completes the proof of Theorem 2.1. □

3. The rigidity of time changes theorem

In this section we give a proof of Theorem 5 independent of Theorems 1–4.

So let $\tau_i \in B^1(M)$, $i=1, 2$, $\bar{\tau}_1 = \bar{\tau}_2$. We can assume without loss of generality that $\bar{\tau}_1 = \bar{\tau}_2 = 1$. Also,

$$Q^{-1} \leqq |\tau_i(x)| \leqq Q \tag{3.1}$$

for some $Q \geqq 1$, $i=1, 2$.

Let $\psi : (M_1, \mu_{\tau_1}) \to (M_2, \mu_{\tau_2})$ be m.p. and let

$$\psi h_t^{\tau_1}(x) = h_t^{\tau_2} \psi(x)$$

for all $t \in R$ and μ_{τ_1}-a.e. $x \in M_1$.

Let $p : G \to M_2$ be the covering projection $p(g) = \Gamma_2 g$, $g \in G$. Since Γ_2 is discrete, there are a compact $K \subset M_2$ with $\mu_1(\psi^{-1}(K)) > 0.999$ and $\rho = \rho(K) > 0$ such that if $g \in p^{-1}(K)$ then p is an isometry on the ball of radius ρ centered at g.

Let $\varepsilon_n = 10^{-n}$, $n = 4, 5, \ldots$ and let $0 < \bar{\varepsilon}_n \leqq 0.1\, \varepsilon_n$ be such that

$$\mu_1(\psi^{-1}(B)) \leqq 0.1\, \varepsilon_n$$

whenever $\mu_2(B) \leqq \bar{\varepsilon}_n$, $B \subset M_2$.

Let $\gamma_n^{(i)} = \gamma^{(i)}(\bar{\varepsilon}_n)$, $l_n^{(i)} = l^{(i)}(\bar{\varepsilon}_n)$ and $E_n^{(i)} = E^{(i)}(\bar{\varepsilon}_n)$ be as in Lemma 1.2 for τ_i, $i = 1, 2$. Denote

$$\gamma_n = \min \{\gamma_n^{(1)}, \gamma_n^{(2)}, 0.1\, \varepsilon_n\},$$
$$l_n = \max \{l_n^{(1)}, l_n^{(2)}\},$$
$$E_n = E_n^{(1)} \cap \psi^{-1}(E_n^{(2)}).$$

We have

$$\mu_1(E_n) > 1 - 0.2\, \varepsilon_n.$$

Since ψ is measurable, there is $\Lambda_n \subset M_1$, $\mu_1(\Lambda_n) > 1 - 0.1\, \varepsilon_n$ such that ψ is uniformly continuous on Λ_n. Given $\varepsilon > 0$ let $\delta(\varepsilon) > 0$ be such that $d(\psi(u), \psi(v)) \leqq \varepsilon/4$ whenever $u, v \in \Lambda_n$, $d(u, v) \leqq \delta(\varepsilon)$.

Denote

$$D = \rho/2Q.$$

Since $h_t^{(i)}$ is ergodic, $i=1, 2$, there are $t_n \geqq \max \{l_n, 20 D \gamma_n^{-1}\}$ and a subset $A_n \subset M_1$ with $\mu_1(A_n) > 1 - 0.1\, \varepsilon_n$ such that if $u \in A_n$, $t \geqq t_n$ then the relative length measure of $\Lambda_n \cap \psi^{-1}(K)$ on the orbit interval $[u, h_t^{(1)} u]$ is at least 0.998 and

$$\begin{aligned} \left| \int_0^t \tau_1(h_s^{(1)} u)\, ds - t \right| &\leqq 0.1\, \varepsilon_n t, \\ \left| \int_0^t \tau_2(h_s^{(2)} \psi(u))\, ds - t \right| &\leqq 0.1\, \varepsilon_n t. \end{aligned} \tag{3.2}$$

Denote

$$V_n = \Lambda_n \cap A_n \cap E_n, \qquad \mu_1(V_n) \geqq 1 - \varepsilon_n = 1 - 10^{-n}. \tag{3.3}$$

For $a, b \in R$, $|a| + |b| \neq 0$, let $L(a, b)$ denote the first $t > 0$ with $\Delta(a, b, t) = D/4$, where $\Delta(a, b, t)$ is defined in (1.4).

Let $0<\omega_n<\gamma_n$ be so small that if $|a|, |b|\leqq\omega_n$, $|a|+|b|\neq 0$, then

$$L(a, b)\geqq\max\{5t_n, 2D(\delta(\varepsilon_n))^{-1}\}. \tag{3.4}$$

Denote

$$\begin{aligned}&\delta_n=\min\{\omega_n, \delta(\omega_n)\}/4,\\ &r_n=\max\{-(\log\delta_n), \log t_n\},\\ &V=\bigcup_{m=1}^{\infty}\bigcap_{n=m}^{\infty} g^{(1)}_{-r_n}(V_n).\end{aligned} \tag{3.5}$$

It follows from (3.3) that

$$\mu_1(V)=1.$$

Lemma 3.1. *Let $x, y\in V$ and $x=k_b^{(1)}y$ for some $b\in R$. Then $d(\bar{x}_n, k_b^{(2)}\bar{y}_n)\to 0$ when $n\to\infty$, where $\bar{x}_n=g^{(2)}_{-r_n}\psi g^{(1)}_{r_n}x$, $\bar{y}_n=g^{(2)}_{-r_n}\psi g^{(1)}_{r_n}y$, $n=1, 2, \ldots$.*

Proof. Denote

$$x_n^{(1)}=g^{(1)}_{r_n}x_n, \qquad y_n^{(1)}=g^{(1)}_{r_n}y.$$

Since $x, y\in V$, there is $m\geqq 4$ such that $x_n^{(1)}, y_n^{(1)}\in V_n$ for all $n\geqq m$.

Henceforth we assume that $n\geqq m$. Denote

$$b_n=e^{-2r_n}b, \qquad x_n^{(2)}=\psi(x_n^{(1)}), \qquad y_n^{(2)}=\psi(y_n^{(1)}).$$

We have

$$\begin{aligned}&x_n^{(1)}=k^{(1)}_{b_n}y_n^{(1)}\\ &|b_n|\leqq|b|\,\delta_n^2\leqq\delta_n\leqq\omega_n,\end{aligned}$$

if n is sufficiently large. Also,

$$x_n^{(2)}=k^{(2)}_{\beta_n}g^{(2)}_{\alpha_n}h^{(2)}_{c_n}y_n^{(2)}$$

for some $|\alpha_n|, |\beta_n|, |c_n|\leqq\eta_n\leqq\gamma_n$. We can assume without loss of generality that $c_n=0$. Denote

$$L_n=\min\{L(0, b_n), L(\alpha_n, \beta_n)\}\geqq 5t_n.$$

Since $x_n^{(1)}, y_n^{(1)}\in V_n$ and $L_n\geqq 5t_n$ there is

$$2t_n\leqq 0.99\,L_n\leqq\lambda_n\leqq 0.995\,L_n \tag{3.6}$$

such that

$$\begin{aligned}&h^{(1)}_{\lambda_n}x_n^{(1)}\in\Lambda_n\cap\psi^{-1}(K),\\ &h^{(1)}_{\sigma(b_n,\lambda_n)}y_n^{(1)}\in\Lambda_n\cap\psi^{-1}(K)\end{aligned} \tag{3.7}$$

if δ_n is sufficiently small, where $\sigma(b_n, \lambda_n)$ is defined in (1.2). Also

$$\begin{aligned}&d(h^{(1)}_{\lambda_n}x_n^{(1)}, h^{(1)}_{\sigma(b_n,\lambda_n)}y_n^{(1)})\leqq\frac{2|\Delta(b_n, L_n)|}{L_n}+2|b_n|\leqq\frac{D}{L_n}+2\delta_n\leqq\delta(\varepsilon_n),\\ &|b_n\lambda_n|\leqq 2|\Delta(b_n, \lambda_n)|/\lambda_n\leqq D/\lambda_n\leqq\gamma_n\end{aligned}$$

by (1.3), (3.4) and (3.5), since $t_n\leqq\lambda_n\leqq L_n$.

This implies via Lemma 1.2 that

$$|\Delta_{\tau_1}(x_n^{(1)}; b_n, \lambda_n) - \Delta(b_n, \lambda_n)| \leqq 0.1\, D\varepsilon_n \tag{3.8}$$

since $x_n^{(1)} \in E_n$ and $\lambda_n \geqq t_n \geqq l_n$.

Denote

$$z_n = h_{s_n}^{(2)} x_n^{(2)}, \qquad w_n = h_{u_n}^{(2)} y_n^{(2)},$$

where s_n and u_n are defined by

$$\xi^{(1)}(x_n^{(1)}, \lambda_n) = \xi^{(2)}(x_n^{(2)}, s_n)$$
$$\xi^{(1)}(y_n^{(1)}, \sigma(b_n, \lambda_n)) = \xi^{(2)}(y_n^{(2)}, u_n)$$

and

$$\xi^{(i)}(v^{(i)}, t) = \int_0^t \tau_i(h_s^{(i)} v^{(i)})\, ds, \qquad v^{(i)} \in M_i,\ i = 1, 2.$$

We have using (3.2) and (3.6),

$$\begin{gathered} 2t_n \leqq 0.98\, L_n \leqq s_n \leqq 0.996\, L_n \leqq L(\alpha_n, \beta_n), \\ |s_n - \lambda_n| \leqq 0.8\, \varepsilon_n s_n. \end{gathered} \tag{3.9}$$

Also,

$$|\beta_n s_n| \leqq \frac{20D}{s_n} \leqq \frac{20D}{t_n} \leqq \gamma_n$$

by Lemma 1.1, (1.3) and the definition of t_n. This implies via Lemma 1.2 that

$$|\Delta_{\tau_2}(x_n^{(2)}; \alpha_n, \beta_n, s_n) - \Delta(\alpha_n, \beta_n, s_n)| \leqq 0.1\, D\varepsilon_n$$

since $|\alpha_n| \leqq \omega_n \leqq \gamma_n$, $l_n \leqq t_n \leqq s_n$. This and (3.8) imply that

$$\begin{aligned} &|(\xi^{(2)}(y_n^{(2)}; \sigma(\alpha_n, \beta_n, s_n)) - \xi^{(2)}(y_n^{(2)}; u_n)) \\ &\quad - (\Delta(\alpha_n, \beta_n, s_n) - (\Delta(b_n, \lambda_n))| \leqq 0.1\, D\varepsilon_n \end{aligned} \tag{3.10}$$

and therefore

$$\int_{u_n}^{\sigma(\alpha_n, \beta_n, s_n)} \tau_2(h_s^{(2)} y_n^{(2)})\, ds = |\xi^{(2)}(y_n^{(2)}; \sigma(\alpha_n, \beta_n, s_n)) - \xi^{(2)}(y_n^{(2)}; u_n)| \leqq 2D.$$

This implies that

$$|u_n - \sigma(\alpha_n, \beta_n, s_n)| \leqq 2DQ = \rho. \tag{3.11}$$

We have

$$d(h_{\sigma(\alpha_n, \beta_n, s_n)}^{(2)} y_n^{(2)}, z_n) \leqq 3\gamma_n \leqq \varepsilon_n/3. \tag{3.12}$$

Also,

$$\begin{gathered} z_n = \psi(h_{\lambda_n}^{(1)} x_n^{(1)}) \in K, \\ w_n = \psi(h_{\sigma(b_n, \lambda_n)}^{(1)} y_n^{(1)}) \in K, \\ d(z_n, w_n) \leqq \varepsilon_n/4. \end{gathered} \tag{3.13}$$

This and (3.12) show that

$$d(h_{\sigma(\alpha_n, \beta_n, s_n)}^{(2)} y_n^{(2)}, w_n) \leqq \varepsilon_n$$

and therefore

$$|u_n - \sigma(\alpha_n, \beta_n, s_n)| \leqq \varepsilon_n \tag{3.14}$$

by (3.11) and the definition of ρ. This implies that

$$|\xi^{(2)}(y_n^{(2)}; \sigma(\alpha_n, \beta_n, s_n)) - \xi^{(2)}(y_n^{(2)}; u_n)| \leqq Q\varepsilon_n$$

and therefore

$$|\Delta(\alpha_n, \beta_n, s_n) - \Delta(b_n, \lambda_n)| \leqq \varepsilon_n(D+Q) \tag{3.15}$$

by (3.10). Also

$$|\Delta(b_n, s_n) - \Delta(b_n, \lambda_n)| \leqq 10\varepsilon_n \Delta(b_n, \lambda_n) \leqq 5\varepsilon_n D$$

by (1.2) and (3.9). This and (3.15) give

$$|\Delta(\alpha_n, \beta_n, s_n) - \Delta(b_n, s_n)| \leqq \varepsilon_n(6D+Q). \tag{3.16}$$

We have

$$x_n^{(2)} = k_{b_n}^{(2)}(k_{p_n}^{(2)} g_{\alpha_n}^{(2)} y_n^{(2)})$$

where $p_n = \beta_n - b_n$. Expression (3.16) says that

$$|\Delta(\alpha_n, p_n, \sigma(b_n, s_n)| \leqq \varepsilon_n(6D+Q)$$

and therefore

$$\begin{aligned} |p_n| &\leqq \frac{2\varepsilon_n(6D+Q)}{[\sigma(b_n, s_n)]^2} \leqq c_1 \varepsilon_n / L_n^2, \\ |\alpha_n| &\leqq \frac{2\varepsilon_n(6D+Q)}{|\sigma(b_n, s_n)|} \leqq c_1 \varepsilon_n / L_n \end{aligned} \tag{3.17}$$

by (1.2), (1.3), (3.6) and (3.9), where $c_1 > 0$ is a constant.

It follows now from Corollary 1.1, expressions (3.6), (3.9) and the definition of L_n that

$$\max\{|\Delta(b_n, \lambda_n)|, |\Delta(\alpha_n, \beta_n, s_n)|\} \geqq 0.1\, D.$$

This implies via (3.15) that

$$\min\{|\Delta(b_n, \lambda_n)|, |\Delta(\alpha_n, \beta_n, s_n)|\} \geqq 0.05\, D$$

if n is sufficiently large. This gives

$$\frac{0.05\, D}{2L_n^2} \leqq \frac{|\Delta(b_n, \lambda_n)|}{2\lambda_n^2} \leqq |b_n|$$

by (1.3) and therefore

$$\begin{aligned} |p_n| &\leqq c_2 \varepsilon_n |b_n| \\ |\alpha_n| &\leqq c_2 \varepsilon_n |b_n|^{\frac{1}{2}} \end{aligned} \tag{3.18}$$

for some $c_2 > 0$ by (3.17). We have

$$\bar{x}_n = g_{-r_n}^{(2)} x_n^{(2)} = k_b^{(2)}(k_{p_n e^{2r_n}}^{(2)} g_{\alpha_n}^{(2)} \bar{y}_n)$$

where

$$|p_n| e^{2r_n} \leqq c_1 \varepsilon_n |b_n| e^{2r_n} = c_1 \varepsilon_n |b|.$$

This and (3.18) show that

$$d(\bar{x}_n, k_b^{(2)} \bar{y}_n) \to 0 \quad \text{when} \quad n \to \infty.$$

This completes the proof of the lemma. □

Lemma 3.2. *If* $x \in V$ *and* $y = h_p^{(1)} x$ *for some* $p \in R$ *then* $d(\bar{y}_n, h_p^{(2)} \bar{x}_n) \to 0$ *when* $n \to \infty$, *where* $\bar{x}_n$, $\bar{y}_n$ *are as in Lemma* 2.1.

Proof. Let $x_n^{(i)}$, $y_n^{(i)}$, $i = 1, 2$, be as in the proof of Lemma 2.1. We have

$$y_n^{(1)} = h^{(1)}_{pe^{2r_n}}(x_n^{(1)}), \qquad y_n^{(2)} = h^{(2)}_{q_n}(x_n^{(2)})$$

where q_n is given by

$$\xi^{(1)}(x_n^{(1)}, pe^{2r_n}) = \xi^{(2)}(x_n^{(2)}, q_n)$$

and

$$pe^{2r_n} \geqq pt_n^2 \geqq t_n$$

if n is sufficiently large. It follows from (3.2) that

$$|pe^{2r_n} - q_n| \leqq \varepsilon_n pe^{2r_n} \tag{3.19}$$

if n is sufficiently large. We have

$$\bar{y}_n = g^{(2)}_{-r_n} y_n^{(2)} = h^{(2)}_{q_n e^{-2r_n}}(\bar{x}_n)$$

where

$$|p - q_n e^{-2r_n}| \leqq \varepsilon_n p$$

by (3.19). This completes the proof. □

Lemma 3.3. *There is an* $h_t^{(1)}$*-invariant subset* $U \subset M_1$ *with* $\mu_1(U) = 1$ *and a subsequence* $\{n_k : k = 1, 2, \ldots\} \subset \{n : n = 1, 2, \ldots\}$ *such that if* $u \in U$ *then* $\lim_{k \to \infty} \bar{u}_{n_k} = \zeta(u) \in M_2$ *exists and* $\zeta(h_p^{(1)} u) = h_p^{(2)} \zeta(u)$ *for all* $p \in R$, $u \in U$.

Proof. Let $M_2 = \bigcup_{n=1}^{\infty} K_n$, where K_n are compact and $\mu_2(M_2 - K_n) < 2^{-n}$, $n = 1, 2, \ldots$. Denote

$$\tilde{K}_n = M_2 - K_n$$

$$F_n = g^{(1)}_{-r_n} \psi^{-1} g^{(2)}_{r_n} \tilde{K}_n, \qquad n = 1, 2, \ldots.$$

We have

$$\sum_{n=1}^{\infty} \mu_1(F_n) < \infty.$$

Let

$$F = \{u \in M_1 : u \text{ belongs to finitely many } F_n\}.$$

By the Borel-Cantelli lemma,

$$\mu_1(F) = 1.$$

If $u \in F$ then $\bar{u}_n \in \tilde{K}_n$ for finitely many n. This implies that there is a subsequence $n_k(u)$, $k = 1, 2, \ldots$, such that $\bar{u}_{n_k(u)}$ converges in M_2.

For $x\in M_1$ let $W(x)=\{g_t^{(1)}k_s^{(1)}x\colon s,t\in R\}$ be the stable leaf of x for $g_t^{(1)}$. Let $I(x)$ denote the orbit interval $[x,k_{-1}^{(1)}x]$ and let $\pi_x\colon I(x)\to W(h_1^{(1)}x)$ be defined by

$$\pi_x(y)=h^{(1)}_{\sigma(b,\,1)}y$$

if $y=k_{-b}^{(1)}x$, $0\leqq b\leqq 1$, where $\sigma(b,1)$ is defined in (1.1). The map π_x is a diffeomorphism from $I(x)$ onto $\pi_x(I(x))$.

Denote $\tilde{x}=h_1^{(1)}x$. We have from (1.1),

$$\tilde{x}=k^{(1)}_{c(b,\,1)}g^{(1)}_{r(b,\,1)}\pi_x(y)$$

if $y=k_{-b}^{(1)}x\in I(x)$. Define $q_x\colon I(x)\to[\tilde{x},g^{(1)}_{-r(1,\,1)}\tilde{x}]=J(\tilde{x})$ by

$$q_x(y)=g^{(1)}_{-r(b,\,1)}\tilde{x},\qquad y=k_{-b}^{(1)}x.$$

The map q_x is a diffeomorphism from $I(x)$ onto $J(\tilde{x})$. Let

$$T(x)=\{z\in J(x)\colon \text{the length measure of } M_1-V \text{ on the } k_t^{(1)}\text{-orbit of } z \text{ is } 0\},$$

$$S'=\{x\in M_1\colon \text{the relative length measure of } T(x) \text{ on } J(x) \text{ is } 1\},$$

$$S=\{x\in V\colon \tilde{x}\in S' \text{ and the relative length measure of the set } \tilde{J}(\tilde{x})=T(\tilde{x})\cap q_x(V\cap\pi_x^{-1}(V)\cap I(x)) \text{ on } J(\tilde{x}) \text{ is } 1\}.$$

Standard measure theoretic arguments show that $\mu_1(S')=\mu_1(S)=1$.

Now pick $x\in S\cap F$ and let $n_k=n_k(x)$, $k=1,2,\ldots$ be such that $\bar{x}_{n_k}$ converges in M_2 when $k\to\infty$.

Let $\tilde{J}=\tilde{J}(\tilde{x})$ and let $u\in\tilde{J}$. Then $q_x^{-1}(u)\in V\cap I(x)$ and therefore $\overline{(q_x^{-1}(u))}_{n_k}$ converges in M_2 by Lemma 2.1 and $\overline{(\pi_x(q_x^{-1}(u)))}_{n_k}$ converges in M_2 by Lemma 2.2. This implies via Lemma 2.1 that $\bar{z}_{n_k}$ converges in M_2 for every $z\in V$ on the $k_t^{(1)}$-orbit of u, since $\pi_x(q_x^{-1}(u))$ lies on this orbit and belongs to V.

Thus $\bar{z}_{n_k}$ converges for every z in the set

$$\tilde{V}=V\cap\Big(\bigcup_{u\in\tilde{J}}I(u)\Big).$$

It follows from the definition of $\tilde{J}$ that the Riemannian volume of $\tilde{V}$ on the leaf $W(\tilde{x})$ is positive. Let

$$U=\{h_t^{(1)}v\colon v\in\tilde{V}\}.$$

The set U is $h_t^{(1)}$-invariant, $\mu_1(U)=1$ and $\lim_{k\to\infty}\bar{z}_{n_k}=\zeta(z)\in M_2$ exists for every $z\in U$ by Lemma 2.2. It also follows from that lemma that $\zeta(h_p^{(1)}z)=h_p^{(2)}\zeta(z)$ for every $z\in U$. This completes the proof. □

Proof of Theorem 5. Let $U\subset M_1$, $\mu_1(U)=1$ and a subsequence $\{n_k\}\subset\{n\}$ be as in Lemma 3.3. We can assume without loss of generality that $U=M_1$ and $\{n_k\}=\{n\}$. Thus

$$\lim_{n\to\infty}\bar{u}_n=\zeta(u)\in M_2$$

exists for all $u\in M_1$ and

$$\zeta(h_p^{(1)}u)=h_p^{(2)}\zeta(u)$$

for all $p \in R$, $u \in M_1$. This says that the map

$$\zeta : (M_1, \mu_1) \to (M_2, \mu_2)$$

is a measurable conjugacy between $h_t^{(1)}$ and $h_t^{(2)}$. It follows then from the rigidity theorem [8] that there are $C \in G$, $a \in R$ such that

$$\begin{aligned} C\Gamma_1 C^{-1} &\subset \Gamma_2 \quad \text{and} \\ \zeta(u) &= h_a^{(2)} \psi_C(u) \end{aligned} \tag{3.20}$$

for μ_1-a.e. $u \in M_1$, where $\psi_C(\Gamma_1 g) = \Gamma_2 Cg$, $g \in G$. It follows from Lemma 3.1 that if $u, v \in V$, $v = k_b^{(1)} u$ then

$$\zeta(v) = k_b^{(2)} \zeta(u).$$

This implies that $a = 0$ in (3.20) and therefore

$$\zeta(u) = \psi_C(u)$$

for μ_1-a.e. $u \in M_1$. This completes the proof of part (1) of the theorem.

Part (3) follows from part (2). In order to prove part (2) one has to repeat the proof of the corresponding part of the main theorem in [11].

Let us outline the main point of this proof. The proof makes essential use of the decay rate of the correlation function of $\tau_i - 1$ for $h_t^{(i)}$, $i = 1, 2$, found in [5] and [12].

Indeed, using this decay rate we proved in [11] (Lemma 3.1) that given $\varepsilon > 0$ there are $P = P(\varepsilon) \subset M_1$, $\mu_1(P) > 1 - \varepsilon$ and $m = m(P) > 0$ such that if $u \in P$, $t \geqq m$, then

$$\begin{aligned} \left| \int_0^t \tau_1(h_s^{(1)} u)\, ds - t \right| &\leqq t^\alpha \\ \left| \int_0^t \tau_2(h_s^{(2)} \psi(u))\, ds - t \right| &\leqq t^\alpha \end{aligned} \tag{3.21}$$

for some $0 < \alpha < 1$ and all $t \geqq m$.

Now let $\bar{X} \subset M_1$, $\mu(\bar{X}) > 1 - \varepsilon$ and $n_0 > 0$ be such that if $u \in \bar{X}$ and $n \geqq n_0$ then

$$d(\bar{u}_n, \zeta(u)) \leqq \varepsilon.$$

Fix $n \geqq n_0$ with $e^{2r_n} \geqq 2Q^2 m$, where $Q \geqq 1$ is as in (3.1), and denote

$$v(u) = \bar{u}_n.$$

Let $\tilde{X}$ be the generic set of $\bar{X} \cap g_{-r_n}^{(1)}(P)$ for $h_t^{(1)}$, $\mu_1(\tilde{X}) = 1$. This means that if $u \in \tilde{X}$ then the relative length measure of $\bar{X} \cap g_{-r_n}^{(1)}(P)$ on $[u, h_t^{(1)} u]$ tends to $\mu_1(\bar{X} \cap g_{-r_n}^{(1)}(P)) \geqq 1 - 2\varepsilon$ when $t \to \infty$.

Let

$$X = \tilde{X} \cap \bar{X} \cap g_{-r_n}^{(1)}(P), \qquad \mu_1(X) > 0$$

and for $u \in X$ let

$$A(u) = \{s \in R^+ : h_s^{(1)} u \in \bar{X} \cap g_{-r_n}^{(1)}(P)\}.$$

For $s\in R^+$ define $t(s)>0$ by

$$h^{(2)}_{t(s)}\, v(u)=v(h^{(2)}_s u).$$

We have for $s\in A(u)$,

$$g^{(1)}_{r_n} h^{(1)}_s u\in P$$
$$\psi(g^{(1)}_{r_n} h^{(1)}_s u)=g^{(2)}_{r_n} v(h^{(1)}_s u)=g^{(2)}_{r_n} h^{(2)}_{t(s)}\, v(u).$$

It follows from (3.21) that if $s, s'\in A(u)$ and $|s-s'|\geqq 1$ then

$$e^{2r_n}|(t(s)-t(s'))-(s-s')|\leqq |s-s'|^{\alpha} e^{2\alpha r_n}$$

and therefore

$$|(t(s)-t(s'))-(s-s')|\leqq |s-s'|^{\alpha}. \tag{3.22}$$

Also,

$$d(h^{(2)}_{t(s)}\, v(u), h^{(2)}_s \zeta(u))\leqq \varepsilon \tag{3.23}$$

for all $s\in A(u)$ and

$$l(A(u)\cap[0, t])/t\geqq 1-2\varepsilon$$

for all sufficiently large t, where $l(A)$ denotes the length measure of A. This, (3.22) and (3.23) are the main conditions of the basic Lemma 2.1 in [11]. It follows then from that lemma that $v(u)=\bar{u}_n$ lies on the $h^{(2)}_t$-orbit of $\zeta(u)$ for every $u\in X$, if $\varepsilon>0$ is chosen sufficiently small.

We have

$$\zeta(g^{(1)}_{r_n} u)=g^{(2)}_{r_n}\zeta(u)$$

for μ_1-a.e. $u\in M_1$. This implies that if we denote

$$X_n=g^{(1)}_{r_n} X, \qquad \mu(X_n)>0,$$

then

$$\psi(u)=h^{(2)}_{\sigma(u)}\zeta(u)$$

for some $\sigma(u)\in R$ and all $u\in X_n$. The set

$$\hat{X}=\{u\in M_1\colon \psi(u)=h^{(2)}_{\sigma(u)}\zeta(u) \text{ for some } \sigma(u)\in R\}$$

is $h^{(1)}_t$-invariant and contains X_n. Therefore $\mu_1(\hat{X})=1$, since $h^{(1)}_t$ is ergodic and $\mu_1(X_n)>0$. This completes the proof of the theorem. □

4. The Joinings theorem

In this section we prove Theorem 3. The proof follows that of Theorem 5 in the previous section using the techniques from [10].

So let $\tau_i\in B^1(M)$, $i=1, 2$, $\bar{\tau}_1=\bar{\tau}_2=1$ be as above and let ν be a nontrivial ergodic joining of $h^{\tau_1}_1$ and $h^{\tau_2}_1$. By Corollary 1(b), ν is an ergodic joining of $h^{\tau_1}_t$ and $h^{\tau_2}_t$ for all $t\in R$.

By Corollary 1(a), $(h^{\tau_1}_t\times h^{\tau_2}_t, \nu)$ is a finite extension of $h^{\tau_1}_t$. This means that there are an integer $N\geqq 1$ and an $(h^{\tau_1}_t\times h^{\tau_2}_t)$-invariant subset $\bar{\Omega}\subset M_1\times M_2$, $\nu(\bar{\Omega})$

$=1$ such that for μ_{τ_1}-a.e. $x\in M_1$ the set $\bar\eta(x)=\bar\Omega\cap(\{x\}\times M_2)$ consists of exactly N points $\{(x,z_1),\dots,(x,z_N)\}$ with $\nu_x\{(x,z_j)\}=1/N$, $j=1,\dots,N$, where ν_x is the probability measure on $\{x\}\times M_2$ obtained from the Fubini theorem applied to ν. We can assume without loss of generality that this holds for all $x\in M_1$.

Let $\bar\eta$ denote the partition of $\bar\Omega$ into sets $\bar\eta(x)$, $x\in M_i$. For $A\subset\bar\Omega$ we write $A<\bar\eta$ if A consists of atoms of $\bar\eta$.

Let $\pi_i\colon M_1\times M_2\to M_i$, $i=1,2$, be the projection $\pi_1(x,z)=x$, $\pi_2(x,z)=z$.

It follows from the Fubini theorem that if $A\subset\Omega$, $\nu(A)>1-\alpha/N^2$ for some $0<\alpha\leqq 1$ then there is $A'\subset\Omega$ such that $\nu(A')>1-\alpha/N$ and $A'<\bar\eta$.

This implies that if $B\subset M_2$, $\mu_2(B)>1-\alpha/N^2$, then there is $\tilde B\subset M_1$, $\mu_1(\tilde B)>1-Q^2\alpha$ such that $\pi_2(\bar\eta(x))\subset B$ for all $x\in\tilde B$, where Q is as in (3.1). We say that $\tilde B$ is induced by B.

Let a compact $K\subset M_2$, $\mu_2(K)>1-10^{-4}/Q^2N^2$ and $0<\rho=\rho(K)<1$ be as in Sect. 3 and let $\tilde K\subset M_1$, $\mu_1(\tilde K)>1-10^{-4}$ be induced by K.

Since atoms of $\bar\eta$ are finite there are $\rho_1>0$ and $K_1<\bar\eta$ with $\nu(K_1)>1-10^{-4}$ such that if $w_1\neq w_2$ and $w_1,w_2\in K_1\cap\bar\eta(x)$ for some $x\in M_1$ then $d(w_1,w_2)\geqq 4\rho_1$.

Let

$$\hat K=\pi_1(K_1)\cap\tilde K,\qquad \mu_1(\hat K)>0.999,\qquad \hat\rho=\min\{\rho,\rho_1\}.$$

Let $\varepsilon_n=10^{-n}$, $n=5,6,\dots$ and let $\bar\varepsilon_n=\min\{0.1\,\varepsilon_n,\varepsilon_n/Q^2N^2\}$.

Let $\gamma_n^{(i)}=\gamma_n^{(i)}(\bar\varepsilon_n)$, $l_n^{(i)}=l_n^{(i)}(\bar\varepsilon_n)$ and $E_n^{(i)}=E^{(i)}(\bar\varepsilon_n)$ be as in Lemma 1.2 for τ_i, $i=1,2$. Denote

$$\gamma_n=\min\{\gamma_n^{(1)},\gamma_n^{(2)},0.1\,\varepsilon_n\},$$
$$l_n=\max\{l_n^{(1)},l_n^{(2)}\},$$
$$E_n=E_n^{(1)}\cap\tilde E_n^{(2)},\qquad \mu_1(E_n)\geqq 1-0.2\,\varepsilon_n$$

where $\tilde E_n^{(2)}$ is induced by $E_n^{(2)}$.

As in [10] there are $\Omega'<\bar\eta$, $\nu(\Omega')=1$ and pairwise disjoint measurable subsets $\bar\Omega_j\subset\Omega'$, $\nu(\bar\Omega_j)=1/N$, $j=1,\dots,N$ such that for every $x\in\pi_1(\Omega')$, the intersection $\bar\eta(x)\cap\bar\Omega_j$ consists of exactly one point and the map $\psi_j\colon\pi_1(\Omega')\to\bar\Omega_j$ defined by $\psi_j(x)=\bar\eta(x)\cap\bar\Omega_j$ is measurable, $j=1,\dots,N$. We can assume without loss of generality that $\Omega'=\bar\Omega$.

Let $\Lambda_n\subset M_1$, $\mu_1(\Lambda_n)>1-0.1\,\varepsilon_n$ be such that ψ_j are uniformly continuous on Λ_n for $j=1,\dots,N$. Given $\varepsilon>0$ let $\delta(\varepsilon)>0$ be such that $d(\psi_j(u),\psi_j(v))\leqq\varepsilon/4$ for all $j=1,\dots,N$, whenever $u,v\in\Lambda_n$, $d(u,v)\leqq\delta(\varepsilon)$.

Denote

$$D=\hat\rho/2Q.$$

Since the flows $h_t^{(i)}$, $i=1,2$, are ergodic, there are $t_n\geqq\max\{l_n,20D\gamma_n^{-1}\}$ and a subset $\Lambda_n\subset M_1$ with $\mu_1(\Lambda_n)>1-0.1\,\varepsilon_n$ such that if $u\in\Lambda_n$, $t\geqq t_n$ then the relative length measure of $\Lambda_n\cap\hat K$ on the orbit interval $[u,h_t^{(1)}u]$ is at least 0.998 and

$$\left|\int_0^t\tau_1(h_s^{(1)}u)\,ds-t\right|\leqq 0.1\,\varepsilon_n t$$

$$\left|\int_0^t\tau_2(h_s^{(2)}\pi_2\psi_j(u))\,ds-t\right|\leqq 0.1\,\varepsilon_n t$$

for all $j=1, \ldots, N$. Denote

$$V_n = \Lambda_n \cap A_n \cap E_n, \qquad \mu_1(V_n) \geqq 1-\varepsilon_n = 1-10^{-n}.$$

Now let $L(a, b)$, ω_n, δ_n, r_n and V, $\mu_1(V)=1$, be as in the proof of Theorem 5 in Sect. 3.

For $x \in M_1$ denote

$$\bar{x}_{n,j} = g^{(2)}_{-r_n} \pi_2 \psi_j g^{(1)}_{r_n} x, \qquad j=1, \ldots, N.$$

Lemma 4.1. *Let $x, y \in V$ and $x = k^{(1)}_b y$ for some $b \in R$. Then $d(\bar{x}_{n,j}, k^{(2)} \bar{y}_{n,j}) \to 0$ when $n \to \infty$ for all $j=1, \ldots, N$.*

Proof. We shall prove the lemma for $j=1$. For $j=2, \ldots, N$, the proof is the same. Denote

$$\pi_2 \psi_1 = \psi, \qquad \bar{x}_{n,1} = \bar{x}_n$$

and repeat the proof of Lemma 3.1 up to expression (3.13), substituting $\psi^{-1}(K)$ in (3.7) by $\hat{K}$ and replacing ρ in (3.11) by $\hat{\rho}$. Instead of expression (3.13) we now have

$$z_n = \pi_2 \psi_k (h^{(1)}_{\lambda_n} x^{(1)}_n) \in K$$
$$w_n = \pi_2 \psi_j (h^{(1)}_{\sigma(b_n, \lambda_n)} y^{(1)}_n) \in K$$

for some $k, j \in \{1, \ldots, N\}$. Also

$$d(z_n, \pi_2 \psi_k (h^{(1)}_{\sigma(b_n, \lambda_n)} y^{(1)}_n)) \leqq \varepsilon_n/4 \tag{4.1}$$

since $h^{(1)}_{\lambda_n} x^{(1)}_n$, $h^{(1)}_{\sigma(b_n, \lambda_n)} y^{(1)}_n \in \Lambda_n$. This, (3.7) and (3.12) imply

$$d(w_n, \pi_2 \psi_k (h^{(1)}_{\sigma(b_n, \lambda_n)} y^{(1)}_n)) \leqq \hat{\rho} + \varepsilon_n \leqq 2\rho_1$$

if n is sufficiently large. This implies by the definition of ρ_1 that

$$w_n = \pi_2 \psi_k (h^{(1)}_{\sigma(b_n, \lambda_n)} y^{(1)}_n)$$

and

$$d(w_n, h^{(2)}_{\sigma(\alpha_n, \beta_n, s_n)} y^{(2)}_n) \leqq \varepsilon_n$$

by (3.12) and (4.1). This implies

$$|u_n - \sigma(\alpha_n, \beta_n, s_n)| \leqq \varepsilon_n$$

by (3.11) and the definition of ρ. Thus we obtained expression (3.14).

Now we repeat the proof of Lemma 3.1 from (3.14) to the end. This completes the proof of the lemma. □

Denote

$$\eta(x) = \pi_2 \bar{\eta}(x), \qquad \eta_n(x) = g^{(2)}_{-r_n} \eta(g^{(1)}_{r_n} x).$$

For $A, B \subset M_2$ define

$$d(A, B) = \max \{d_{A,B}, d_{B,A}\}$$

where

$$d_{A,B} = \sup_{x \in A} \inf_{y \in B} d(x, y).$$

Corollary 4.1. *If $x, y \in V$ and $x = k_b^{(1)} y$ for some $b \in R$ then $d(\eta_n(x), k_b^{(2)} \eta_n(y)) \to 0$ when $n \to \infty$.*

Lemma 4.2. *If $x \in V$ and $y = h_p^{(1)} x$ for some $p \in R$, then $d(\eta_n(y), h_p^{(2)} \eta_n(x)) \to 0$ when $n \to \infty$.*

Proof. The proof is completely analogous to that of Lemma 3.2.

Lemma 4.3. *There are an $h_t^{(1)}$-invariant subset $U \subset M_1$ with $\mu_1(U) = 1$ and a subsequence $\{n_k : k = 1, 2, \ldots\} \subset \{n : n = 1, 2, \ldots\}$ such that if $u \in U$ then $\lim_{k \to \infty} \eta_{n_k}(u) = \zeta(u) \subset M_2$ exists and $\zeta(h_p^{(1)} u) = h_p^{(2)} \zeta(u)$ for all $p \in R$, $u \in U$.*

Proof. Let $\tilde{K}_n \subset M_2$, $\mu_2(\tilde{K}_n) \leqq 2^{-n}$, $n = 1, 2, \ldots$ be as in the proof of Lemma 3.3 and let

$$F_n = \{u \in M_1 : \eta_n \cap g_{r_n}^{(2)}(\tilde{K}_n) \neq \emptyset\},$$

$$F = \{u \in M_1 : g_{r_n}^{(1)} u \in F_n \text{ for finitely many } n\}.$$

We have, using the Fubini theorem,

$$\mu_1(F_n) \leqq 2^{-n+1}$$

if n is sufficiently large. This implies by the Borel-Cantelli lemma that $\mu_1(F) = 1$. Thus if $u \in F$ then there is a subsequence n_k such that $\eta_{n_k}(u)$ converges to a finite subset of M_2 when $k \to \infty$.

From now on we proceed exactly as in the proof of Lemma 3.3 to construct an $h_t^{(1)}$-invariant subset $U \subset M_1$ as required in the lemma. □

Let $U \subset M_1$, $\mu_1(U) = 1$, and a subsequence $\{n_k\} \subset \{n\}$ be as in Lemma 4.3. We can assume without loss of generality that $U = M_1$ and $\{n_k\} = \{n\}$. Thus

$$\lim_{n \to \infty} \eta_n(u) = \zeta(u) \subset M_2$$

exists for all $u \in M_1$ and

$$\zeta(h_p^{(1)} u) = h_p^{(2)} \zeta(u) \tag{4.2}$$

for all $p \in R$, $u \in M_1$.

The set $\zeta(u)$ is a finite subset of M_2 and the cardinal number $|\zeta(u)|$ of $\zeta(u)$ is the same for μ_1-a.e. $u \in M_1$, since $h_t^{(1)}$ is ergodic. It follows from Corollary 4.1 that

$$\zeta(k_b^{(1)} u) = k_b^{(2)} \zeta(u) \tag{4.3}$$

for all $b \in R$, μ_1-a.e. $u \in M_1$. This implies via (1.1) that

$$\zeta(g_s^{(1)} u) = g_s^{(2)} \zeta(u) \tag{4.4}$$

for μ_1-a.e. $u \in M_1$ and all $s \in R$. Let

$$\Omega = \{(u, z) : u \in M_1, z \in \zeta(u)\}$$

and let m be the probability measure on Ω defined by

$$m(A \times B) = \int_A |B \cap \zeta(u)| / \ell \, d\mu_1(u)$$

for any measurable subsets $A \subset M_1$, $B \subset M_2$, where $\ell = |\zeta(u)|$.

Expressions (4.2), (4.3) and (4.4) show that m is invariant under $h_t^{(1)} \times h_t^{(2)}$, $g_t^{(1)} \times g_t^{(2)}$, $k_t^{(1)} \times k_t^{(2)}$. It is clear that the M_1-marginal of m is μ_1. Let m_2 be the M_2-marginal of m. Then m_2 is preserved by $h_t^{(2)}$, $g_t^{(2)}$ and $k_t^{(2)}$ and therefore $m_2 = \mu_2$. Thus m is a nontrivial G-invariant joining of $h_t^{(1)}$ and $h_t^{(2)}$.

We shall show that τ_1 and τ_2 are jointly cohomologous via m and that m is an ergodic joining of $h_t^{(1)}$ and $h_t^{(2)}$. To do so we follow the proof of part (1) of Theorem 5, which makes essential use of combinatorial Lemma 2.2 in [11] and the decay rate of the correlation function of τ_i for $h_t^{(i)}$, $i=1, 2$, found in [5] and [12].

Indeed, using these decay rates we were able to prove Lemma 3.1 in [11] which implies for the case in question that there is $0<\alpha<1$ such that given $\omega>0$ there are $P \subset M_1$ with $\mu_1(P)>1-\omega$ and $\bar{t}=\bar{t}(P)$ such that if $u \in P$, $t \geqq \bar{t}$. Then

$$\begin{aligned} \left| \int_0^t \tau_1(h_s^{(1)} u)\, ds - t \right| &\leqq t^\alpha \\ \left| \int_0^t \tau_2(h_s^{(2)} z)\, ds - t \right| &\leqq t^\alpha \end{aligned} \tag{4.5}$$

for all $z \in \eta(u)$.

The following lemma is analogous to the Basic Lemma 2.1 in [11].

Lemma 4.4. *Given $0<\alpha$, $\omega<1$ there are $Y \subset M_1$ with $\mu_1(Y)>1-\omega$, $\theta=\theta(\alpha)>0$ and $\varepsilon=\varepsilon(Y)>0$ such that if $u \in Y$, $z \in \zeta(u)$, $w \in M_2$ and $d(z, w) \leqq \varepsilon$ then w lies on the $h_t^{(2)}$-orbit of z whenever there is a subset $A \subset R^+$ with the following properties:*

(1) $0 \in A$;

(2) *If $s \in A$ then $h_s^{(1)} u \in Y$ and there is $t(s)>0$ increasing in s such that $d(h_{t(s)}^{(2)} w, z_s) \leqq \varepsilon$ for some $z_s \in \zeta(h_s^{(1)} u)$;*

(3) *$|(t(s')-t(s))-(s'-s)| \leqq |s-s'|^\alpha$ for all $s', s \in A$ with $|s'-s| \geqq 1$;*

(4) *$l(A \cap [0, \lambda])/\lambda \geqq 1-\theta$ for all $\lambda \in A$ with $\lambda \geqq \lambda_0$, where $l(A)$ denotes the length measure of A.*

Proof. We begin as in the proof of the Basic Lemma 2.1 in [11]. Given $0<\alpha<1$ let $0<\gamma=\gamma(\alpha)<\alpha/2$ and $\theta=\theta(\gamma)>0$ be as in (2.1) of [11], where the combinatorial Lemma 2.2 has been used.

Denote $\bar{\gamma}=\min\{\gamma, \omega\}$. Let a compact $K \subset M_2$ with $\mu_2(K)>1-0.1\,\bar{\gamma}/\mathscr{k}^2$ and $0<\rho(K)<1$ be as in Sect. 3. Let $\bar{K} \subset M_1$, $\mu_1(\bar{K})>1-0.1\,\bar{\gamma}$ be such that $\zeta(u) \subset K$ for all $u \in \bar{K}$.

Let $Y_0 \subset M_1$, $\mu_1(Y_0)>1-0.1\,\bar{\gamma}$ and $\rho(Y_0)>0$ be such that if $u \in Y_0$, $z_1, z_2 \in \zeta(u)$ and $z_1 \neq z_2$, then $d(z_1, z_2) \geqq 4\rho(Y_0)$.

Since $g_t^{(1)}$ is ergodic, there are $\bar{Y} \subset M_1$ with $\mu_1(\bar{Y})>1-0.1\,\bar{\gamma}$ and $t_0>0$ such that if $u \in \bar{Y}$, $t \geqq t_0$, then the relative length measure of $\bar{K} \cap Y_0$ on $[u, g_t^{(1)} u]$ is at least $1-0.2\,\bar{\gamma}$. Set

$$Y=\bar{Y} \cap \bar{K} \cap Y_0, \qquad \mu_1(Y)>1-\omega.$$

Now in order to show that $\varepsilon=\varepsilon(\rho(K), \rho(Y_0), t_0)>0$ can be chosen so small as to satisfy the requirements of the lemma, we just have to repeat the proof of Lemma 2.1 in [11]. □

Proof of Theorem 3. (1) Let us show that there exists an $h_t^{(1)}$-invariant subset $\hat{X} \subset M_1$ with $\mu_1(\hat{X})=1$ such that if $u \in \hat{X}$, $w \in \eta(u)$ then there is $z(w) \in \zeta(u)$ such that w lies on the $h_t^{(2)}$-orbit of $z(w)$.

Indeed, let $0<\alpha<1$ be as in (4.5) and $\theta=\theta(\alpha)>0$ be as in Lemma 4.4. For $0<\omega<\theta$ let $P=P(\omega)\subset M_1$, $\mu_1(P)>1-0.1\,\omega$ and $\bar{t}=\bar{t}(P)>0$ be as in (4.5). Also let $Y\subset M_1$, $\mu_1(Y)>1-0.1\,\omega$ and $\varepsilon=\varepsilon(Y)>0$ be as in Lemma 4.4.

Let $\bar{X}\subset M_1$, $\mu_1(\bar{X})>1-0.1\,\omega$ and $n_0>0$ be such that if $u\in\bar{X}$, $n\geqq n_0$, then

$$d(\eta_n(u), \zeta(u))\leqq\varepsilon.$$

Fix $n\geqq n_0$ with $e^{2r_n}\geqq 2Q^2\bar{t}(P)$, where $Q\geqq 1$ is as in (3.1). Let $\tilde{X}\subset M_1$ be the generic set of $\bar{X}\cap g^{(1)}_{-r_n}P\cap Y$ for $h^{(1)}_t$, $\mu_1(\tilde{X})=1$ and let

$$X=\tilde{X}\cap\bar{X}\cap g^{(1)}_{-r_n}(P)\cap Y,\qquad \mu_1(X)>0.$$

Let $u\in X$ and $w\in\eta_n(u)$. Then there exists $z\in\zeta(u)$ such that $d(w, z)\leqq\varepsilon$. Let

$$A=\{s\in R^+: h^{(1)}_s u\in Y\cap\bar{X}\cap g^{(1)}_{-r_n}(P)\}.$$

We have $0\in A$ and

$$l(A\cap[0, \lambda])/\lambda\geqq 1-\omega\geqq 1-\theta \tag{4.6}$$

for all $\lambda\geqq\lambda_0$, since $u\in\tilde{X}$.

For $s\in R^+$ let $t(s)>0$ be defined by

$$\xi^{(1)}(g^{(1)}_{r_n}u, e^{2r_n}s)=\xi^{(2)}(g^{(2)}_{r_n}w, e^{2r_n}t(s))$$

where $\xi^{(i)}$, $i=1, 2$, are as in Sect. 3. We have

$$h^{(2)}_{t(s)}v\in\eta_n(h^{(1)}_s u)$$

since $\bar{\eta}$ is $h^{\tau_1}_t\times h^{\tau_2}_t$-invariant. This implies that if $s\in A$ then

$$d(h^{(2)}_{t(s)}v, z_s)\leqq\varepsilon \tag{4.7}$$

for some $z_s\in\zeta(h^{(1)}_s u)$. We have

$$g^{(2)}_{r_n}h^{(2)}_{t(s)}w\in\eta(g^{(1)}_{r_n}h^{(1)}_s u).$$

Also

$$g^{(1)}_{r_n}h^{(1)}_s u\in P$$

if $s\in A$. This implies via (4.5) and our choice of n that if $s, s'\in A$ and $|s-s'|\geqq 1$ then

$$e^{2r_n}|(t(s)-t(s'))-(s-s')|\leqq|s-s'|^\alpha e^{2\alpha r_n}$$

and therefore

$$|(t(s)-t(s'))-(s-s')|\leqq|s-s'|^\alpha.$$

This, (4.6) and (4.7) imply via Lemma 4.4 that w lies on the $h^{(2)}_t$-orbit of z or that $g^{(2)}_{r_n}w\in\eta(g^{(1)}_{r_n}u)$ lies on the $h^{(2)}_t$-orbit of $g^{(2)}_{r_n}z\in\zeta(g^{(1)}_{r_n}u)$.

Now let

$$X_n=g^{(1)}_{r_n}(X),\qquad \mu_1(X_n)>0$$

and let

$$\hat{X}=\{u\in M_1:\text{ for every } w\in\eta(u)\text{ there is } z(w)\in\zeta(u)$$
$$\text{such that } w \text{ lies on the } h^{(2)}_t\text{-orbit of } z(w)\}.$$

The set $\hat{X}$ is $h_t^{(1)}$-invariant and $X_n \subset \hat{X}$. This implies that $\mu_1(\hat{X})=1$, since $h_t^{(1)}$ is ergodic. This completes the proof of claim (1).

We can assume without loss of generality that $\hat{X}=M_1$.

(2) Now let us show that τ_1 and τ_2 are jointly cohomologous via m.

Indeed, it follows from (4.3) that for μ_1-a.e. $u \in M_1$ no two distinct points of $\zeta(u)$ can lie on the same orbit of $h_t^{(2)}$. This implies that the map $w \to z(w)$ from $\eta(u)$ to $\zeta(u)$ is well defined for μ_1-a.e. $u \in M_1$.

Also for μ_1-a.e. $u \in M_1$ no two distinct points of $\eta(u)$ can lie on the same orbit of $h_t^{(2)}$, since $\bar{\eta}$ is invariant under $h_t^{\tau_1} \times h_t^{\tau_2}$ and ν is an ergodic joining of $h_t^{\tau_1}$ and $h_t^{\tau_2}$. This implies that the map $w \to z(w)$ is bijective and the map $\phi: (\bar{\Omega}, \nu) \to (\Omega, m_{f_1})$ defined by

$$\phi(u, w)=(u, z(w))$$

is an isomorphism between $h_t^{\tau_1} \times h_t^{\tau_2}$ and $S_t^{f_1}$, where $S_t = h_t^{(1)} \times h_t^{(2)}$ and $f_1(u, z) = \tau_1(u)$, $(u, z) \in \Omega$. This implies that m is an ergodic joining of $h_t^{(1)}$ and $h_t^{(2)}$.

Let $f_2(u, z)=\tau_2(z)$, $(u, z) \in \Omega$ and let $v(u, z)$ be such that $w = h_{v(u, z)}^{\tau_2} z$, where $z = z(w)$. It is now clear that f_1 and f_2 are cohomologous via v along S_t acting on (Ω, m) and that $v = v_{(m, v)}$ is induced by (m, v). This completes the proof of Theorem 3. □

5. The factors theorem

In this section we prove Proposition 2 and Theorem 6.

Proof of Proposition 2. Let ζ and η be shift-related partitions of (X, μ) invariant under T_t and T_t^{τ} respectively. We have

$$\eta(x)=\{T_{v(x, y)}^{\tau} y: y \in \zeta(x)\}$$

for μ-a.e. $x \in X$, where $v(x, \cdot)=v_x \in L_1(\zeta(x), \mu_{\zeta(x)})$. We shall assume without loss of generality that this holds for all $x \in X$. We have

$$\begin{aligned} v(x, y) &= -v(y, x) \\ v(x, z) &= v(x, y)+v(y, z) \end{aligned} \tag{5.1}$$

and

$$v(T_t x, T_t y)-v(x, y)=\int_0^t (\tau(T_s x)-\tau(T_s y))\, ds \tag{5.2}$$

for all $y, z \in \zeta(x)$, $x \in X$ and all $t \in R$, since η is T_t^{τ}-invariant.

We have

$$\tau_{\zeta}(x)=\int_{\zeta(x)} \tau(y)\, d\mu_{\zeta(x)}(y),$$

$$\tilde{v}(x)=-\int_{\zeta(x)} v(x, y)\, d\mu_{\zeta(x)}(y).$$

Expression (5.2) gives

$$\tilde{v}(T_t x)-\tilde{v}(x)=\int_0^t (\tau_{\zeta}-\tau)(T_s x)\, ds \tag{5.3}$$

for all $x\in X$, $t\in R$. This shows that τ_ζ and τ are cohomologous along T_t via $\tilde{v}$. Also

$$\tilde{v}(x)=\tilde{v}(y)-v(x, y) \tag{5.4}$$

for all $y\in\zeta(x)$, $x\in X$ by (5.1).

Let $\chi_{\tilde{v}}$ be the isomorphism between $T_t^{\tau_\zeta}$ and T_t^{τ} induced by $\tilde{v}$, that is

$$\chi_{\tilde{v}}(x)=T^{\tau}_{\tilde{v}(x)}x, \qquad x\in X.$$

We have

$$\begin{aligned}\chi_{\tilde{v}}(\zeta(x))&=\{T^{\tau}_{\tilde{v}(y)}y\colon y\in\zeta(x)\}\\&=\{T^{\tau}_{\tilde{v}(x)+v(x, y)}y\colon y\in\zeta(x)\}\\&=T^{\tau}_{\tilde{v}(x)}\eta(x)=\eta(T^{\tau}_{\tilde{v}(x)}x)=\eta(\chi_{\tilde{v}}(x))\end{aligned}$$

for all $x\in X$ by (5.4), since η is T_t^{τ}-invariant. This completes the proof of the proposition. □

Proof of Theorem 6. Let $\tau\in B^1(M)$, V be a factor of h_1^{τ} and η be the h_1^{τ}-invariant partition of M induced by V. It follows from Corollary 2 that a.e. atom of η is finite and η is invariant under h_t^{τ} for all $t\in R$.

Let ν be the probability measure on $M\times M$ defined by

$$\nu(A\times B)=\int_{M/\eta}\mu_C(A)\,\mu_C(B)\,d\mu_{\tau,\eta}(C)$$

for every measurable $A, B\subset M$, where $(M/\eta, \mu_{\tau,\eta})$ is the quotient space induced by η, $\mu_C(A)=|A\cap C|/|C|$ for every $C\in\eta$ and $|C|$ denote the cardinal number of C. The set

$$\bar{\Omega}=\{(u, w)\colon w\in\eta(u), u\in M\}$$

is $h_t^{\tau}\times h_t^{\tau}$-invariant and $\nu(\bar{\Omega})=1$.

The measure ν is a nontrivial joining of h_t^{τ} with itself. It might not be ergodic. Since atoms of η are finite, ν is a finite convex sum of nontrivial ergodic joinings of h_t^{τ} with itself (see [10] for this).

It follows from Theorem 3 that there exists $\zeta(u)\subset M$, $u\in M$ and a measurable v on $M\times M$ such that

$$\begin{aligned}|\zeta(u)|&=|\eta(u)|\\\eta(u)&=\{h^{\tau}_{v(u, w)}w\colon w\in\zeta(u)\}\end{aligned} \tag{5.5}$$

for μ-a.e. $u\in M$ and ζ is invariant under the action of G on M.

Henceforth we use without loss of generality the word "everywhere" instead of "almost everywhere".

We claim that

$$\zeta(w)=\zeta(u) \quad \text{if } w\in\zeta(u), \qquad u\in M. \tag{5.6}$$

Inded, it follows from Lemma 4.3 that

$$\zeta(u)=\lim_{n\to\infty}g_{-r_n}(\eta(g_{r_n}u)) \tag{5.7}$$

for some sequence $r_n\to\infty$ when $n\to\infty$. This implies that $u\in\zeta(u)$ for all $u\in M$, since $u\in\eta(u)$, $u\in M$.

Denote $u_n = g_{r_n}(u)$. We have

$$u_n, w_n \in \zeta(u_n), \qquad \eta(u_n) = \{h^{\tau}_{v(u_n, z)} z : z \in \zeta(u_n)\}.$$

This implies that

$$\eta(w_n) = h^{\tau}_{-v(u_n, w_n)} \eta(u_n)$$

since η is h_t^{τ}-invariant. We have

$$g_{-r_n} \eta(u_n) = \{h_{e^{-2r_n} s(z_n)} z_n : z \in \zeta(u)\}$$

where $s(z_n)$ is defined by

$$\int_0^{s(z_n)} \tau(h_t z_n)\, dt = v(u_n, z_n)$$

and

$$e^{-2r_n} s(z_n) \to 0, \qquad \text{when } n \to \infty \tag{5.8}$$

for all $z \in \zeta(u)$, $u \in M$ by (5.7). Also

$$g_{-r_n} \eta(w_n) = \{h_{e^{-2r_n} \bar{s}(z_n)} z : z \in \zeta(u)\} \tag{5.9}$$

where

$$\bar{s}(z_n) = s(z_n) - \int_0^{v(u_n, w_n)} \tau^{-1}(h_t(h_{s(z_n)} z_n))\, dt = s(z_n) - \tilde{s}(z_n)$$

and

$$|\tilde{s}(z)| \leqq Q^2 s(w_n)$$

for all $z \in \zeta(u)$, $u \in M$, where $Q \geqq 1$ is as in (3.1) for $\tau_i = \tau$, $i = 1, 2$. This, (5.8) and (5.9) imply that

$$\zeta(w) = \lim_{n \to \infty} g_{-r_n} \eta(w_n) = \zeta(u).$$

This proves (5.7).

Denote

$$\Gamma' = \{g \in G : \Gamma g \in \zeta(\Gamma e)\}$$

where e denotes the unity in G.

Let $g_1, g_2 \in \Gamma'$ and $z = g_1^{-1} g_2$. We have

$$\Gamma g_2 \in \zeta(\Gamma z)$$

since ζ is G-invariant. Also,

$$\Gamma z \in \zeta(\Gamma g_2) = \zeta(\Gamma e)$$

by (5.6). Therefore $z \in \Gamma'$.

This proves that Γ' is a discrete subgroup of G. Also $\Gamma \subset \Gamma'$ and hence $\Gamma' \in \mathscr{T}$. We have

$$\zeta(\Gamma g) = \psi^{-1}_{\Gamma, \Gamma'} \{\Gamma' g\}$$

where $\psi_{\Gamma, \Gamma'}(\Gamma g) = \Gamma' g$, $g \in G$.

Expression (5.5) shows that η and ζ are shift related along h_t. The theorem now follows from Proposition 2. □

References

1. Feldman, J., Ornstein, D.: Semirigidity of horocycle flows over compact surfaces of variable negative curvature. Ergodic Theory Dyn. Syst. 1987 (to appear)
2. Furstenberg, H.: Disjointness in ergodic theory, minimal sets and a problem in diophantine approximation. Math. Syst. Theory **1**, 1–49 (1967)
3. Kolmogorov, A.N.: On dynamical systems with integral invariant on the torus. Doklady Akad. Nauk. SSSR **93**, 763–766 (1953) [Russian]
4. Marcus, B.: Ergodic properties of horocycle flows for surfaces of negative curvature. Ann. Math. **105**, 81–105 (1977)
5. Moore, C.: Exponential decay of correlation coefficients for geodesic flow. In: Moore, C.C. (ed.) Group Representations, Ergodic Theory, Operator Algebras, and Mathematical Physics. Berlin, Heidelberg, New York: Springer 1987
6. Parry, W.: Metric classification of ergodic nilflows and unipotent affines. Am. J. Math. **93**, 819–828 (1971)
7. Parry, W.: Dynamical representations in nilmanifolds. Compos. Math. **26**, 159–174 (1973)
8. Ratner, M.: Rigidity of horocycle flows. Ann. Math. **115**, 587–614 (1982)
9. Ratner, M.: Factors of horocycle flows. Ergodic Theory Dyn. Syst. **2**, 465–489 (1982)
10. Ratner, M.: Horocycle flows, joinings and rigidity of products. Ann. Math. **118**, 277–313 (1983)
11. Ratner, M.: Rigidity of time changes for horocycle flows. Acta Math. **156**, 1–32 (1986)
12. Ratner, M.: The rate of mixing for geodesic and horocycle flows. Ergodic Theory Dyn. Syst. 1987 (to appear)
13. Rudolph, D.: An example of a measure preserving map with minimal self-joinings, and applications. J. Anal. Math. **35**, (1979)
14. Witte, D.: Rigidity of some translations on homogeneous spaces. Invent. Math. **81**, 1–27 (1985)

Oblatum 6-VIII-1986

Invent. math. 88, 375-394 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Jets de fonctions différentiables sur le dual d'un groupe de Lie nilpotent

Fokko Du Cloux

Centre de Mathématiques de l'Ecole Polytechnique, Plateau de Palaiseau, F-91128 Palaiseau Cedex, France

Introduction

Soit G un groupe de Lie nilpotent, connexe et simplement connexe, $G\hat{}$ son dual unitaire, $C^*(G)$ la C^*-algèbre de G. Depuis le célèbre théorème de Gelfand, il est classique de considérer la théorie des C^*-algèbres comme une généralisation non-commutative de la théorie des espaces localement compacts. Si maintenant on veut passer du niveau «continu» au niveau «différentiable», il est naturel de dire (cf. [2], introduction), qu'une structure différentiable sera définie par la donnée d'une sous-algèbre involutive dense de la C^*-algèbre considérée, pouvant jouer le rôle de l'algèbre des «fonctions différentiables» de la théorie. Bien entendu, un certain arbitraire existe au niveau du choix d'une telle sous-algèbre; toutefois, dans le cas qui nous intéresse il ne fait guère de doute que l'algèbre $\mathscr{S}(G)$ des fonctions de Schwartz sur G (cf. n° 1.1 ci-après) fournit une théorie satisfaisante.

Soit donc $\mathscr{S}=\mathscr{S}(G)$, et soit (E, π) un point de $G\hat{}$; soit M le noyau de l'action de $\mathscr{S}$ dans E. Une fois définie la notion de fonction différentiable, il est naturel de se demander ce que pourront être les développements limités en un point de $G\hat{}$. Prenant exemple sur le cas commutatif, et sur la géométrie algébrique, nous appellerons algèbre des jets d'ordre n (resp. d'ordre infini) de fonctions différentiables au point (E, π) l'algèbre $\mathscr{S}_n=\mathscr{S}/M^{n+1}$ (resp. l'algèbre $\hat{\mathscr{S}}=\varprojlim \mathscr{S}_n$). De façon imagée, on pourra dire que l'application canonique $\mathscr{S}\to\hat{\mathscr{S}}$ associe à $\varphi\in\mathscr{S}$ sa «série de Taylor» au point (E, π). Naturellement, les coefficients de cette série de Taylor ne sont en général pas scalaires; ils se trouvent dans l'algèbre $\mathscr{S}/M$, que nous notons $\mathscr{A}$.

Notre objectif dans cet article est l'étude de ces algèbres $\mathscr{S}_n$ et $\hat{\mathscr{S}}$. Nous montrerons que l'algèbre $\mathscr{A}$ se relève toujours dans $\hat{\mathscr{S}}$, et que l'on obtient ainsi une équivalence de Morita (cf. n° 1.3) entre $\hat{\mathscr{S}}$ et une algèbre C que nous avions introduite précédemment dans [1] pour décrire une certaine catégorie

«Unité Associée au C.N.R.S. n° 169»

de G-modules de longueur finie (cf. n° 1.2). Ayant ainsi éliminé la partie inessentielle de $\hat{\mathcal{S}}$, on obtient bien cette fois une algèbre de «séries formelles non-commutatives», sur laquelle on trouvera de plus amples renseignements dans [1]; l'algèbre C_n associée de manière analogue à $\mathcal{S}_n$ est une $\mathbb{C}$-algèbre de dimension finie.

La théorie des G-modules de longueur finie joue un rôle central dans tout l'article: le point crucial de notre argument consiste en effet à montrer que sous l'action de $G \times G$, $\mathcal{S}_n$ est un module de longueur finie au sens ci-dessus, ce qui permet de lui appliquer les résultats de [1]. Mais ce travail est aussi une «application de la théorie des catégories abéliennes à l'analyse»: en effet nous n'aurions pas pu conclure sans un théorème de Gabriel [5] affirmant l'unicité de l'algèbre «sobre» associée à une catégorie abélienne finie (cf. n° 1.2).

Au cours de la démonstration nous obtenons une description de l'idéal M (th. 3.5), et des résultats sur les G-modules de longueur finie (n° 4), qui ont peut-être un intérêt indépendant.

1. Enonce du resultat

Gardons les notations de l'introduction.

1.1. Soit $\mathfrak{g} = \mathrm{Lie}(G)$ l'algèbre de Lie de G, et identifions G à $\mathfrak{g}$ à l'aide de l'application exponentielle. Cela permet de définir l'espace $\mathcal{S}$ des fonctions de Schwartz sur G (rappelons que $\mathcal{S}(\mathfrak{g})$ est l'espace des fonctions C^∞ sur $\mathfrak{g}$, nulles à l'infini ainsi que leur image par n'importe quel opérateur différentiel à coefficients polynomiaux). Faisons choix d'une mesure de Haar dg sur G; alors $\mathcal{S}$ est une sous-algèbre involutive dense de $L^1(G, dg)$, donc de $C^*(G)$. La restriction à $\mathcal{S}$ de l'action de $C^*(G)$ dans E définit un homomorphisme que nous notons encore π, de $\mathcal{S}$ vers $\mathcal{L}(E)$, dont M est le noyau et $\mathcal{A}$ l'image.

1.2. Rappelons maintenant brièvement ce que nous entendons par G-module de longueur finie. Soit $\mathbf{mod}_G$ la catégorie des G-modules continus de Fréchet, avec pour morphismes les entrelacements linéaires continus. Soit E_∞ l'espace des vecteurs C^∞ de E. On note $\mathbf{Ext}_G(E_\infty)$ la sous-catégorie pleine de $\mathbf{mod}_G$ formée des V possédant une suite de composition forte:

$$V = V_m \supset V_{m-1} \supset \ldots \supset V_1 \supset V_0 = 0$$

telle que $V_j/V_{j-1} \simeq E_\infty$ pour $1 \leqq j \leqq m$. (ici «forte» veut dire que V_{j-1} est un sous-G-module fermé de V_j, et que l'espace de Fréchet sous-jacent à V_j est somme directe de V_{j-1} et de V_j/V_{j-1} (nous dirons que V_{j-1} est topologiquement supplémentable dans V_j).) Les objets de $\mathbf{Ext}_G(E_\infty)$ seront appelés G-modules de longueur finie à sous-quotients simples E_∞.

Il est prouvé dans [1] partie III th. 1.2.1 que la catégorie $\mathbf{Ext}_G(E_\infty)$ est abélienne et ne possède, à isomorphisme près, qu'un seul objet simple, à savoir E_∞.

Rappelons, puisque nous en aurons besoin par la suite, comment on construit une équivalence de catégories entre $\mathbf{Ext}_G(E_\infty)$ et une catégorie de

modules. Soit $U=U(\mathfrak{g})$ l'algèbre enveloppante complexifiée de $\mathfrak{g}$, I le noyau de l'action de U dans E_∞, $A=U/I$. On sait d'après [1] partie II cor. 4.6.2 que A se relève en une sous-algèbre R de $\hat{U}=\varprojlim U/I^{n+1}$; soit C le commutant de R dans $\hat{U}$, $\mathfrak{m}$ l'unique idéal maximal de C, de sorte que $C/\mathfrak{m}\simeq\mathbb{C}$, et notons $\mathbf{Ext}_C(\mathbb{C})$ la catégorie des C-modules de longueur finie (nécessairement à sous-quotients simples $C/\mathfrak{m}$). Alors on obtient une équivalence de catégories $\mathbf{Ext}_G(E_\infty)\xrightarrow{\sim}\mathbf{Ext}_C(\mathbb{C})$ par le foncteur $V\to\mathrm{Hom}_R(E_\infty, V)$, d'inverse $W\to E_\infty\otimes W$ (cf. [1] partie III, th. 2.1.1).

Il est important de remarquer que cette propriété caractérise C à isomorphisme près parmi les $\mathbb{C}$-algèbres qui sont des anneaux locaux de radical $\mathrm{rad}(C)=\mathfrak{m}$, de corps résiduel $C/\mathfrak{m}=\mathbb{C}$, séparés et complets pour la topologie $\mathfrak{m}$-adique (ceci est un cas très particulier des résultats de Gabriel, cf. [5] ch. IV, §4, th. 4 et cor. 4). En particulier, le commutant C est indépendant du choix de R à isomorphisme près, ce que l'on peut d'ailleurs démontrer directement ([1] partie II, prop. 4.3.2).

1.3. *Equivalence de Morita*

Soit $\mathscr{W}$ l'algèbre des matrices infinies $(a_{mn})_{m,n\in\mathbb{N}}$ $a_{mn}\in\mathbb{C}$, à décroissance rapide (i.e. $\sup_{m,n}(|a_{mn}|(m+n)^k)<+\infty$ pour tout $k\in\mathbb{N}$). On munit $\mathscr{W}$ de sa topologie évidente d'espace de Fréchet, qui en fait un espace nucléaire. On note $\hat{\otimes}$ le produit tensoriel projectif (complété) d'espaces de Fréchet.

Nous dirons avec Connes que deux algèbres de Fréchet $\mathscr{B}_1$ et $\mathscr{B}_2$ sont Morita-équivalentes, si $\mathscr{W}\hat{\otimes}\mathscr{B}_1\simeq\mathscr{W}\hat{\otimes}\mathscr{B}_2$ comme algèbres de Fréchet (l'algèbre $\mathscr{W}$ joue donc pour les algèbres de Fréchet le rôle de l'algèbre des opérateurs compacts pour les C^*-algèbres).

Nous verrons plus loin (cf. lemme 2.3.2) d'autres réalisations intéressantes de l'algèbre $\mathscr{W}$. En particulier nous verrons que $\mathscr{W}\hat{\otimes}\mathscr{W}\simeq\mathscr{W}$, donc $\mathscr{B}$ et $\mathscr{W}\hat{\otimes}\mathscr{B}$ sont Morita-équivalentes pour toute algèbre de Fréchet $\mathscr{B}$.

1.4. Dans les notations du n° 1.2, soit $U_n=U/I^{n+1}$, $C_n=H^0(R, U_n)$ le commutant de R dans U_n. D'après [1] partie II, lemme 4.4.1. On a $\hat{I}^{n+1}\cap C=\mathfrak{m}^{n+1}$, où $\hat{I}=\hat{U}I$ est le noyau de l'application canonique $\hat{U}\to A$, donc $C_n=C/\mathfrak{m}^{n+1}$, et $C=\varprojlim C_n$. De plus, chaque C_n est une $\mathbb{C}$-algèbre de dimension finie.

Voici maintenant le résultat que nous avons en vue:

1.5. **Théorème.** (i) *Pour tout* $n\in\mathbb{N}$, M^{n+1} *est un idéal fermé de* $\mathscr{S}$.

(ii) *Les systèmes projectifs d'algèbres de Fréchet* $(\mathscr{S}_n)$ *et* $(\mathscr{A}\otimes C_n)$ *sont isomorphes, et on a* $\mathscr{A}\simeq\mathscr{W}$ (*cf.* n° 1.3) *si* $\dim E=+\infty$, $\mathscr{A}=\mathbb{C}$ *si* $\dim E=1$; *on a donc* $\hat{\mathscr{S}}\simeq\mathscr{W}\hat{\otimes}C$ *si* $\dim E=+\infty$, $\hat{\mathscr{S}}\simeq C$ *si* $\dim E=1$.

1.6. *Remarques.* a) Si $\dim E=1$, C est en fait le complété I-adique de U, isomorphe au complété de U en l'idéal d'augmentation (cf. n° 7 ci-dessous).

b) Si l'orbite $\mathcal{O}$ associée à (E,π) par la bijection de Kirillov $G^\wedge\to\mathfrak{g}^*/G$ est de dimension maximale $2r$, C est isomorphe à un anneau de séries formelles commutatives $\mathbb{C}[\![x_1,\ldots,x_d]\!]$ en $d=\dim\mathfrak{g}-2r$ indéterminées ([1] partie II, démonstration du cor. 6.1.2). Donc $\hat{\mathscr{S}}\simeq\mathscr{W}[\![x_1,\ldots,x_d]\!]$ si G est non-commutatif, $\hat{\mathscr{S}}\simeq\mathbb{C}[\![x_1,\ldots,x_d]\!]$ si G est commutatif.

1.7. *Plan de la démonstration*

Pour démontrer le th. 1.5, nous procéderons en deux temps. Nous montrerons d'abord que l'on a une décomposition du système projectif $(\mathscr{S}_n)$ de la forme $(\mathscr{S}_n) \simeq (\mathscr{A} \otimes C'_n)$, où (C'_n) est un certain système projectif de $\mathbb{C}$-algèbres de dimension finie avec unité; puis nous montrerons qu'en fait $(C_n) \simeq (C'_n)$ à l'aide du résultat d'unicité signalé au n° 1.2.

L'assertion (i) du théorème est une partie de la proposition 5.6; la décomposition $\mathscr{S}_n \simeq \mathscr{A} \otimes C'_n$ est établie à la proposition 5.12, et l'isomorphisme $C_n \simeq C'_n$ au n° 6.

2. Preliminaires

Dans ce n°, on étudie le comportement de $\mathscr{S}$ au cours de la récurrence, dans le schéma de démonstration classique pour les groupes de Lie nilpotents. Si $\mathfrak{h}$ est une sous-algèbre de Lie de $\mathfrak{g}$, on note $\exp(\mathfrak{h})$ le sous-groupe analytique de $\mathfrak{g}$ d'algèbre de Lie $\mathfrak{h}$.

2.1. *Transformation de Fourier partielle le long d'un idéal central*

Soient $\mathfrak{c}$ un idéal central de $\mathfrak{g}$, $\varphi \in \mathscr{S}$. Pour tous $g \in G$, $\lambda \in \mathfrak{c}^*$, on pose:

$$\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(g, \lambda) = \int_{\mathfrak{c}} \varphi(g \exp(\xi))\, e^{2i\pi\langle \lambda, \xi \rangle}\, d\xi$$

où $d\xi$ est une mesure de Lebesgue sur $\mathfrak{c}$, fixée pour la suite. Si z est un élément du centre $\mathfrak{z}$ de $\mathfrak{g}$, on note aussi $\bar{\mathscr{F}}_z$ pour $\bar{\mathscr{F}}_{\mathbb{R}z}$. On obtient ainsi une fonction C^∞ sur $G \times \mathfrak{c}^*$, à valeurs complexes, vérifiant $\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(g \exp(\xi), \lambda) = e^{-2i\pi\langle \lambda, \xi \rangle} \bar{\mathscr{F}}_{\mathfrak{c}} \varphi(g, \lambda)$ pour tout $\xi \in \mathfrak{c}$; en d'autres termes, pour tout $\lambda \in \mathfrak{c}^*$ fixé, la fonction $\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(\cdot, \lambda)$ appartient à $C^\infty \operatorname{Ind}_{\exp(\mathfrak{c})} e^{2i\pi\lambda} = \{\varphi \in C^\infty(G)$ t.q. $\varphi(g \exp(\xi)) = e^{-2i\pi\langle \lambda, \xi \rangle} \varphi(g)$ pour tous $g \in G$, $\xi \in \mathfrak{c}\}$. Plus précisément, soit Y un relèvement de la variété algébrique réelle $G/\exp(\mathfrak{c})$ dans G, défini par exemple à l'aide d'une base coexponentielle à $\mathfrak{c}$ dans $\mathfrak{g}$ (cf. [6], ch. XII, th. 2.2). Alors $\bar{\mathscr{F}}_{\mathfrak{c}} \varphi$ est entièrement déterminée par sa restriction à $Y \times \mathfrak{c}^*$, et on obtient ainsi un isomorphisme d'espaces de Fréchet entre $\mathscr{S}$ et $\mathscr{S}(Y \times \mathfrak{c}^*)$.

Posons pour tout $\lambda \in \mathfrak{c}^*$:

$$\mathscr{S}_\lambda = \mathscr{S} \operatorname{Ind}_{\exp(\mathfrak{c})}^G e^{2i\pi\lambda} = \{\varphi \in C^\infty \operatorname{Ind}_{\exp(\mathfrak{c})}^G e^{2i\pi\lambda} \text{ t.q. } \varphi|_Y \in \mathscr{S}(Y)\}$$

(cela ne dépend pas du choix de Y). On munit $\mathscr{S}_\lambda$ d'une structure d'algèbre de Fréchet involutive par la formule:

$$\varphi_1 * \varphi_2(g) = \int_{G/\exp(\mathfrak{c})} \varphi_1(g g_1^{-1})\, \varphi_2(g_1)\, d\dot{g}_1$$

où $d\dot{g}$ est le quotient de dg par la mesure de Lebesgue sur $\mathfrak{c} \simeq \exp(\mathfrak{c})$ qui a été choisie plus haut. On vérifie aussitôt que l'expression sous le signe $\int$ passe au quotient par $\exp(\mathfrak{c})$, que l'intégrale converge et définit un élément de $\mathscr{S}_\lambda$.

Pour tout $\lambda \in \mathfrak{c}^*$, notons $\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(\lambda)$ l'élément $\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(\cdot, \lambda)$ de $\mathscr{S}_\lambda$. Alors $\bar{\mathscr{F}}_{\mathfrak{c}}$ établit un isomorphisme de l'algèbre $\mathscr{S}$ sur l'algèbre des sections $\mathscr{S}$ de la famille d'algèbres $(\mathscr{S}_\lambda)_{\lambda \in \mathfrak{c}^*}$ (les sections $\mathscr{S}$ étant définies après trivialisation par le choix de Y comme ci-dessus, disons). Dans cette écriture, la multiplication se fait point par point par rapport à λ, i.e.

$$\bar{\mathscr{F}}(\varphi_1 * \varphi_2)(\lambda) = \bar{\mathscr{F}} \varphi_1(\lambda) * \bar{\mathscr{F}} \varphi_2(\lambda) \qquad \forall \varphi_1, \varphi_2 \in \mathscr{S}.$$

2.2. Les considérations qui précèdent admettent aussi une version relative, qui nous sera utile, au sens suivant:

Soit $\mathfrak{c}_1$ un idéal central de $\mathfrak{g}$ contenu dans $\mathfrak{c}$. Fixons $\lambda_1 \in \mathfrak{c}_1^*$. Alors on peut définir $\bar{\mathscr{F}}_{\mathfrak{c}}$ pour les éléments de $\mathscr{S}_{\lambda_1}$ par la formule suivante:

$$\bar{\mathscr{F}}_{\mathfrak{c}} \varphi(g, \lambda) = \int_{\mathfrak{c}/\mathfrak{c}_1} \varphi(g \exp(\xi))\, e^{2i\pi\langle \lambda, \xi \rangle}\, d\dot{\xi}$$

pour tous $g \in G$, $\lambda \in \lambda_1 + \mathfrak{c}_1^\perp$ (dans cette formule, on suppose que l'on a fait choix d'une mesure de Lebesgue $d\xi_1$ sur $\mathfrak{c}_1$, et on note $d\dot{\xi} = d\xi / d\xi_1$). Nous laissons au lecteur le soin d'étendre au cas relatif les remarques et définitions du n° 2.1.

2.3. Les actions de G à gauche et à droite sur $\mathscr{S}(G)$ donnent naissance à des actions de $\mathfrak{g}$ que nous notons par des convolutions:

$$\xi * \varphi(g) = \frac{d}{dt} \varphi(\exp(-t\xi)\, g)|_{t=0},$$

$$\varphi * \xi(g) = \frac{d}{dt} \varphi(g \exp(-t\xi))|_{t=0}.$$

Ces actions se prolongent en des actions de U notées $\varphi \to u * \varphi$, $\varphi \to \varphi * u$ pour tout $u \in U$.

On a alors de manière évidente, à droite par exemple:

$$\bar{\mathscr{F}}_{\mathfrak{c}}(\varphi * \xi)(\lambda) = 2i\pi\lambda(\xi)\, \bar{\mathscr{F}}_{\mathfrak{c}} \varphi(\lambda) \qquad \text{pour tous } \lambda \in \mathfrak{c}^*,\ \xi \in \mathfrak{c}.$$

2.4. **Lemme.** *Pour tout $\lambda \in \mathfrak{c}^*$, l'application canonique $\varphi \to \bar{\mathscr{F}}(\varphi)(\lambda)$ de $\mathscr{S}$ vers $\mathscr{S}_\lambda$ est une surjection scindée, de noyau $N_\lambda = \mathscr{S} * P_\lambda$, où P_λ est l'idéal de U engendré par les $(\xi - 2i\pi\lambda(\xi))$, $\xi \in \mathfrak{c}$. En particulier, pour $\lambda = 0$, on obtient une surjection canonique de $\mathscr{S}$ vers $\mathscr{S}(G/\exp(\mathfrak{c}))$, de noyau $\mathscr{S} * \mathfrak{c}$.*

Résultat analogue dans le cas relatif.

Démonstration. Limitons-nous au cas absolu, le cas relatif étant analogue. Comme ici λ est fixé, notons λ' le point courant de $\mathfrak{c}^*$. Soit $e_1, \ldots, e_m$ une base de $\mathfrak{c}$, $\lambda_j = \lambda(e_j)$ pour $1 \leqq j \leqq m$; alors les $(e_j - 2i\pi\lambda_j)$ engendrent P_λ comme idéal à gauche ou à droite, de sorte que $\mathscr{S} * P_\lambda = \sum_{j=1}^{m} \mathscr{S} * (e_j - 2i\pi\lambda_j)$ (somme non directe, en général).

Choisissons comme au n° 2.1 un relèvement Y de $G/\exp(\mathfrak{c})$ dans G, et identifions $\mathscr{S}$ à $\mathscr{S}(Y \times \mathfrak{c}^*)$ par $\bar{\mathscr{F}}_{\mathfrak{c}}$. Dans cette écriture, comme on l'a vu au n° 2.3, e_j agit par multiplication par $2i\pi\lambda'_j$. Donc pour voir que $N_\lambda = \mathscr{S} * P_\lambda$, il

suffit de montrer que toute $\varphi \in \mathscr{S}(Y \times \mathfrak{c}^*)$ telle que $\varphi|_{Y \times \{\lambda\}} = 0$ a une expression de la forme $\varphi = \sum_{j=1}^{m} 2i\pi(\lambda'_j - \lambda_j)\,\varphi_j$, $\varphi_j \in \mathscr{S}(Y \times \mathfrak{c}^*)$, ce qui est évident.

Pour scinder l'application $\mathscr{S} \to \mathscr{S}_\lambda$, il suffit de choisir $\alpha \in C_c^\infty(\mathbb{R}, [0, 1])$, $\alpha \equiv 1$ dans un voisinage de 0, et de poser pour toute $\varphi \in \mathscr{S}_\lambda \simeq \mathscr{S}(Y)$:

$$a(\varphi)(y, \lambda') = \varphi(y) \prod_{j=1}^{m} \alpha(\lambda'_j - \lambda_j), \quad \text{cqfd.}$$

2.5. Nous allons maintenant préparer le cas de récurrence où π ne passe au quotient par aucun idéal central non nul de $\mathfrak{g}$, et où $\dim \mathfrak{g} > 1$.

Cela suppose que le centre $\mathfrak{z}$ de $\mathfrak{g}$ est de dimension 1, et que $f|_{\mathfrak{z}} \neq 0$. Soit $e_3 \in \mathfrak{z}$ l'unique élément tel que $f(e_3) = 1$. Soit $\mathfrak{z}_2$ le deuxième terme de la suite centrale ascendante de $\mathfrak{g}$; soit $e_2 \in \mathfrak{z}_2 \setminus \mathfrak{z}$ tel que $f(e_2) = 0$. Soit $\mathfrak{g}_1$ le centralisateur de e_2 dans $\mathfrak{g}$, $G_1 = \exp(\mathfrak{g}_1)$, et $e_1 \in \mathfrak{g}$ tel que $[e_1, e_2] = e_3$. Alors $\mathfrak{g} = \mathbb{R}e_1 \ltimes \mathfrak{g}_1$, $G = \exp(\mathbb{R}e_1) \ltimes G_1$. Soit $\tilde{\mathfrak{g}} = \mathfrak{g}_1 / \mathbb{R}e_2$, $\tilde{G} = G_1 / \exp(\mathbb{R}e_2)$, $\tilde{e}_3$ l'image de e_3 dans $\tilde{\mathfrak{g}}$. Soient $f_1 = f|_{\mathfrak{g}_1}$, $\tilde{f}$ la forme linéaire sur $\tilde{\mathfrak{g}}$ définie par f_1 par passage au quotient. Soit $\lambda = f|_{\mathfrak{z}} = e_3^*$, $\tilde{\lambda} = \tilde{e}_3^* \in \tilde{\mathfrak{c}}^*$, où $\tilde{\mathfrak{c}} = \mathbb{R}\tilde{e}_3$.

Soit (E_1, π_1) la représentation unitaire irréductible de G_1 correspondant à f_1. Alors $E = \mathrm{Ind}_{G_1}^{G} E_1$, de sorte que E se réalise dans $L^2(\mathbb{R}, E_1)$, et E_∞ dans $\mathscr{S}(\mathbb{R}, E_{1\infty})$ par les formules (pour l'action de $\mathfrak{g}$):

$$e_1 \to -\frac{d}{dt}; \qquad \xi \to \pi_1(\exp(-te_1)\,\xi) \text{ si } \xi \in \mathfrak{g}_1$$

(ce qui veut dire que ξ agit par l'opérateur décomposé défini par la fonction $t \to \pi_1(\exp(-te_1)\,\xi) \in \mathrm{End}(E_{1\infty})$, polynomiale en t puisque $\mathrm{ad}\, e_1$ est un endomorphisme nilpotent de $\mathfrak{g}_1$). On a en particulier: $e_2 \to -2i\pi t$, $e_3 \to 2i\pi$.

2.6. Ceci posé, nous voulons établir un parallèle entre là décomposition de U et celle de $\mathscr{S}$. Posons $U_1 = U(\mathfrak{g}_1)$, $\tilde{U} = U(\tilde{\mathfrak{g}})$, $U_\lambda = U/(e_3 - 2i\pi)\,U$, et définissons de même $U_{1,\lambda}$, $\tilde{U}_{\tilde{\lambda}}$. Rappelons d'abord comment on établit un isomorphisme: (cf. [3], ch. 4, lemme 4.7.8)

$$U_\lambda \simeq W_1 \otimes \tilde{U}_{\tilde{\lambda}}$$

où W_1 est l'algèbre de Weyl d'indice 1 sur $\mathbb{C}$. On peut l'exprimer en disant que si dans la formule $\xi \to \pi_1(\exp(-te_1)\,\xi)$) étendue à U_1, on remplace π_1 par la projection canonique $p: U_{1,\lambda} \to \tilde{U}_{\tilde{\lambda}}$ on obtient un isomorphisme de $U_{1,\lambda}$ avec $\tilde{U}_{\tilde{\lambda}}[t]$; dans cette écriture l'action de e_1 devient $-\dfrac{d}{dt}$, d'où l'isomorphisme cherché en prenant $-\dfrac{d}{dt}$ et $-t = p\left(\dfrac{1}{2i\pi}\, e_2\right)$ comme générateurs canoniques de W_1.

Passons maintenant à l'algèbre $\mathscr{S}$. L'analogue de W_1 est ici l'algèbre $\mathscr{S}(\mathbb{R}^2)$, munie du «produit des noyaux» $\varphi_1 * \varphi_2(t_1, t_2) = \int_{\mathbb{R}} \varphi_1(t_1, s)\,\varphi_2(s, t_2)\,ds$. Posons $\mathscr{S}_1 = \mathscr{S}(G_1)$, $\tilde{\mathscr{S}} = \mathscr{S}(\tilde{G})$.

2.7. **Lemme.** *Avec les notations ci-dessus on a un isomorphisme, canonique après les choix de e_1 et e_2:*

$$\mathscr{S}_\lambda \simeq \mathscr{S}(\mathbb{R}^2, \tilde{\mathscr{S}}_{\tilde{\lambda}})$$

où le membre de droite est muni du «produit des noyaux»:

$$\varphi_1 * \varphi_2(t_1, t_2) = \int_{\mathbb{R}} \varphi_1(t_1, s) * \varphi_2(s, t_2)\, ds.$$

Démonstration. On fait la même construction que pour l'algèbre enveloppante.

a) L'analogue de la surjection canonique $p: U_{1,\lambda} \to \tilde{U}_{\tilde{\lambda}}$ est l'homomorphisme (encore noté p) de $\mathscr{S}_{1,\lambda}$ vers $\tilde{\mathscr{S}}_{\tilde{\lambda}}$ défini par:

$$p(\varphi) = \bar{\mathscr{F}}_{e_2} \varphi(0)$$

dans les notations du n° 2.1 (noter que $\mathfrak{c} = \mathbb{R} e_2$ et $\mathfrak{c}^* = \mathbb{R} e_2^*$ sont identifiés à $\mathbb{R}$, d'où une mesure de Lebesgue sur $\mathfrak{c}$). Donc on a un isomorphisme d'algèbres $\rho: \mathscr{S}_{1,\lambda} \xrightarrow{\sim} \mathscr{S}(\mathbb{R}, \tilde{\mathscr{S}}_{\tilde{\lambda}})$ en posant:

$$\rho(\varphi)(t) = p(\exp(-t e_1)\, \varphi).$$

Soit, en identifiant $\tilde{\mathscr{S}}_{\tilde{\lambda}}$ à $\mathscr{S}\, \mathrm{Ind}_{\exp(\mathfrak{a})}^{G_1} e^{2i\pi e_3^*}$, $\mathfrak{a} = \mathbb{R} e_2 \oplus \mathbb{R} e_3$:

$$\begin{aligned}
\rho(\varphi)(t, g_1) &= \int_{\mathbb{R}} \varphi(\exp(t e_1)\, g_1 \exp(\tau e_2) \exp(-t e_1))\, d\tau \\
&= \int_{\mathbb{R}} \varphi(\exp(t e_1)\, g_1 \exp(-t e_1) \exp(\tau e_2))\, e^{-2i\pi\tau} d\tau \\
&= \bar{\mathscr{F}}_{e_2} \varphi(\exp(t e_1)\, g_1 \exp(-t e_1), -t)
\end{aligned}$$

(au besoin, cette écriture permet de s'assurer de la bijectivité de ρ).

b) Si dans cette formule on remplace φ par $\exp(t_1 e_1)\, \varphi$:

$$g_1 \to \varphi(\exp(-t_1 e_1)\, g_1 \exp(t_1 e_1))$$

on trouve:

$$\begin{aligned}
&\rho(\exp(t_1 e_1)\, \varphi)(g_1, t) \\
&\quad = \int_{\mathbb{R}} \varphi(\exp(t - t_1) e_1 \cdot g_1 \cdot \exp -(t - t_1) e_1 \\
&\qquad \cdot \exp(\tau e_2) \cdot \exp(-t_1 \tau e_3))\, e^{-2i\pi t\tau}\, d\tau \\
&\quad = \bar{\mathscr{F}}_{e_2} \varphi(\exp(t - t_1) e_1 \cdot g_1 \cdot \exp -(t - t_1) e_1, -(t - t_1)) \\
&\quad = \rho(\varphi)(g_1, t - t_1).
\end{aligned}$$

Donc ρ transforme l'action de $\mathbb{R} e_1$ sur $\mathscr{S}_{1,\lambda}$ en l'action par translation à gauche. Or $\mathscr{S}_\lambda$ n'est autre que le «produit semi direct» en un sens évident de $\mathscr{S}(\mathbb{R})$ avec $\mathscr{S}_{1,\lambda}$ pour l'action de $\mathbb{R} e_1$ sur $\mathscr{S}_{1,\lambda}$ (qui est tempérée au sens de la déf. 4.1 ci-dessous, ce qui donne un sens au produit semi direct); la formule est:

$$\varphi_1 * \varphi_2(t_1, g_1) = \int_{\mathbb{R}} \varphi_1(t_1 - s, \exp(s e_1)\, g_1 \exp(-s e_1)) * \varphi_2(s, g_1)\, ds.$$

Par l'application ρ, l'action se transforme en translation à gauche sur $\mathscr{S}(\mathbb{R}, \tilde{\mathscr{S}}_{\tilde{\lambda}})$, ce qui donne le produit:

$$\varphi_1 * \varphi_2(t_1, t_2) = \int_{\mathbb{R}} \varphi_1(t_1 - s, t_2 + s) * \varphi_2(s, t_2)\, ds$$

et en posant $\psi(\tau_1, \tau_2) = \varphi(\tau_1 - \tau_2, \tau_2)$ on trouve bien:

$$\begin{aligned}\psi_1 * \psi_2(\tau_1, \tau_2) &= \int_{\mathbb{R}} \psi_1(\tau_1, \tau_2 + s) * \psi_2(\tau_2 + s, \tau_2)\, ds \\ &= \int_{\mathbb{R}} \psi_1(\tau_1, s) * \psi_2(s, \tau_2)\, ds.\end{aligned}$$

c) En conclusion, on obtient l'isomorphisme cherché en associant à $\varphi \in \mathscr{S}_\lambda$ le noyau $k \in \mathscr{S}(\mathbb{R}^2, \tilde{\mathscr{S}}_{\tilde{\lambda}})$ défini par:

$$k(t_1, t_2, g_1) = \int_{\mathbb{R}} \varphi(\exp(t_1 e_1) \cdot g_1 \cdot \exp(\tau e_2) \cdot \exp(-t_2 e_1))\, d\tau.$$

2.8. *Remarques.* a) D'après le lemme 3.2 ci-dessous, l'algèbre $\mathscr{S}(\mathbb{R}^2)$ avec le produit des noyaux est en fait isomorphe à l'algèbre $\mathscr{W}$ du n° 1.3. Donc le lemme dit que les algèbres $\mathscr{S}_\lambda$ et $\tilde{\mathscr{S}}_{\tilde{\lambda}}$ sont Morita-équivalentes.

b) Soit $(\tilde{E}, \tilde{\pi})$ la représentation de $\tilde{G}$ définie par $\tilde{f}$; c'est la représentation déduite de (E_1, π_1) par passage au quotient. Comme on le vérifie immédiatement, l'action de $\varphi \in \mathscr{S}_\lambda$ dans E se fait par le noyau $\tilde{\pi} \circ k \in \mathscr{S}(\mathbb{R}^2, \mathscr{L}(\tilde{E}))$, où k est associé à φ par la formule donnée en c) ci-dessus; c'est d'ailleurs tout l'intérêt de la construction.

De même, si $\tilde{u}$ est un élément de $\tilde{U}_{\tilde{\lambda}}$, $u = 1 \otimes \tilde{u}$ l'élément de U_λ correspondant dans la décomposition $U_\lambda = W_\lambda \otimes \tilde{U}_{\tilde{\lambda}}$ du n° 2.6, on a $\varphi * u(t_1, t_2) = \varphi(t_1, t_2) * \tilde{u}$ pour tout $\varphi \in \mathscr{S}_\lambda = \mathscr{S}(\mathbb{R}^2, \tilde{\mathscr{S}}_{\tilde{\lambda}})$.

3. Description du noyau et de l'image de π

3.1. Nous pouvons maintenant revenir à l'étude de l'application $\pi: \mathscr{S} \to \mathscr{L}(E)$. Remarquons que π commute à l'action naturelle du groupe $G \times G$ sur les deux membres. De plus, il est facile de voir (par exemple par récurrence sur $\dim G$ en utilisant l'analyse faite au n° précédent) qu'en fait π prend ses valeurs dans l'idéal $\mathscr{L}^2(E)$ des opérateurs de Hilbert-Schmidt. Comme l'action de $G \times G$ sur $\mathscr{S}$ est différentiable, on en déduit que $\pi(\mathscr{S}) = \mathscr{A}$ est contenue dans $\mathscr{L}^2(E)_\infty$, espace des vecteurs C^∞ de $\mathscr{L}^2(E)$ pour l'action de $G \times G$.

3.2. **Lemme.** *L'algèbre $\mathscr{S}(\mathbb{R}^2)$ avec le produit des noyaux* (cf. n° 2.6) *est isomorphe à l'algèbre $\mathscr{W}$ du* n° 1.3. *De même, $\mathscr{W} \hat{\otimes} \mathscr{W} \simeq \mathscr{W}$.*

Démonstration. Il suffit de remarquer que ces algèbres sont obtenues de la façon suivante: On a un espace de Fréchet F muni d'un produit scalaire continu $(\,|\,)$; on munit alors l'espace de Fréchet $\mathscr{B} = F \hat{\otimes} F^\vee$ (où $F^\vee$ est le $\mathbb{C}$-espace vectoriel conjugué de F) d'une structure d'algèbre de Fréchet involutive en posant:

$$a_1 \otimes b_1 \cdot a_2 \otimes b_2 = (a_2 \,|\, b_1)\, a_1 \otimes b_2$$

pour les tenseurs décomposés, ce qui s'étend en une application bilinéaire (continue parce que séparément continue) de $\mathscr{B} \times \mathscr{B} \to \mathscr{B}$. L'involution est donnée par $(a \otimes b)^* = b \otimes a$.

Dans le cas de $\mathscr{S}(\mathbb{R}^2)$, F est l'espace $\mathscr{S}(\mathbb{R})$, le produit scalaire provient de son inclusion dans $L^2(\mathbb{R})$; dans le cas de $\mathscr{W}$, F est l'espace $\mathscr{S}(\mathbb{N})$ des suites à décroissance rapide, et le produit scalaire provient encore de son inclusion dans $L^2(\mathbb{N})$. Pour démontrer que $\mathscr{S}(\mathbb{R}^2) \simeq \mathscr{W}$, il suffit donc d'établir un isomorphisme entre $\mathscr{S}(\mathbb{R})$ et $\mathscr{S}(\mathbb{N})$ respectant les produits scalaires, ce qui peut se faire en considérant par exemple la base de $L^2(\mathbb{R})$ formée des fonctions d'Hermite. De même, pour voir que $\mathscr{W} \otimes \mathscr{W} \simeq \mathscr{W}$ il suffit de montrer que $\mathscr{S}(\mathbb{N}^2) \simeq \mathscr{S}(\mathbb{N})$ avec conservation des produits scalaires, ce qui est facile.

3.3. Rappelons le résultat suivant de *R.* Howe, que nous redémontrons pour la commodité du lecteur, vu que nous disposons de tout ce qu'il faut pour cela.

Théorème (cf. [7], th. 3.4). *L'application* $\pi: \mathscr{S} \to \mathscr{L}^2(E)_\infty$ *est une surjection scindée. On a donc* $\mathscr{A} = \mathscr{L}^2(E)_\infty$, *et l'algèbre* $\mathscr{A}$ *est isomorphe à l'algèbre* $\mathscr{W}$ *du* n° 1.2 *si* $\dim E = +\infty$, *à* $\mathbb{C}$ *si* $\dim E = 1$.

Démonstration. Par récurrence sur $\dim(G)$. Comme $\mathscr{L}^2(E)_\infty$ est l'algèbre obtenue avec $F = E_\infty$ dans la démonstration du lemme 3.2, et que $E_\infty \simeq \mathscr{S}(\mathbb{R}^r)$ avec le produit scalaire induit par $L^2(\mathbb{R}^r)$, où $r = \frac{1}{2}\dim \mathcal{O}$, l'assertion sur la forme de $\mathscr{A}$ découle aussitôt de la surjectivité et du lemme 3.2 (en faisant une récurrence sur r).

Reste à montrer que $\pi: \mathscr{S} \to \mathscr{L}^2(E)_\infty$ est une surjection scindée. On peut supposer $\dim(G) > 1$.

Si π passe au quotient par un idéal central $\mathfrak{c}$ de dimension >0 en une représentation de G': $G/\exp(\mathfrak{c})$, l'application $\pi: \mathscr{S} \to \mathscr{L}^2(E)_\infty$ se factorise par la surjection canonique $\mathscr{S} \to \mathscr{S}(G')$ du lemme 2.4, et on conclut en appliquant l'hypothèse de récurrence.

Sinon, on est dans la situation du n° 2.5, dont on adopte les notations. Le même argument que ci-dessus prouve qu'il suffit de voir que $\pi: \mathscr{S}_\lambda \to \mathscr{L}^2(E)_\infty$ est une surjection scindée. Or, d'après la remarque 2.8. b), dans les décompositions $\mathscr{S}_\lambda = \mathscr{S}(\mathbb{R}^2, \tilde{\mathscr{S}}_\lambda)$ (lemme 2.7) et $\mathscr{L}^2(E)_\infty \simeq \mathscr{S}(\mathbb{R}^2, \mathscr{L}^2(\tilde{E})_\infty)$, π s'écrit $\varphi \to \tilde{\pi} \circ \varphi$. Donc π est une surjection scindée dès lors que $\tilde{\pi}$ l'est, cqfd.

3.4. **Corollaire** ([7], cor. 3.4.1). *Le* $\mathscr{S}$*-module* E_∞ *est (algébriquement) simple et à commutant scalaire.*

Démonstration. Il est déjà simple et à commutant scalaire sous l'action de $E_\infty \otimes E_\infty \subset \mathscr{L}^2(E)_\infty$ (produit tensoriel non complété).

3.5. **Théorème** *(description du noyau de* π*). Dans les notations du* n° 2.3, *on a* $M = \mathscr{S} * I = I * \mathscr{S}$. *Plus précisément, soit* $\mathfrak{c}$ *un idéal central de* $\mathfrak{g}$, $\lambda = f|_{\mathfrak{c}}$. *Introduisons les notations du lemme* 2.4, *et soient* $M_\lambda = M/N_\lambda$, $U_\lambda = U/P_\lambda$, $I_\lambda = I/P_\lambda$. *Alors si* $m = \dim(\mathfrak{c})$, $d = \dim \mathfrak{g}(f)$, *on peut construire une suite centralisante régulière* $(x_{m+1}, \ldots, x_d)$ *d'éléments de* U_λ *engendrant* I_λ, *et des applications linéaires continues* $u_j: M_\lambda \to \mathscr{S}_\lambda$, $m+1 \leqq j \leqq d$, *telles que:*

$$\varphi = \sum_{j=m+1}^{d} u_j(\varphi) * x_j$$

pour toute $\varphi \in M_\lambda$, *de sorte que* $M_\lambda = \mathscr{S}_\lambda * I_\lambda = I_\lambda * \mathscr{S}_\lambda$.

Démonstration. Par récurrence sur $\dim(G)$. Comme on a de façon évidente $\pi(\varphi * u) = \pi(\varphi)\pi(u)$ pour tous $\varphi \in \mathscr{S}$, $u \in U$, l'inclusion $\mathscr{S} * I \subset M$ est immédiate; de même $I * \mathscr{S} \subset M$. On peut supposer $\dim(G) > 1$.

a) Montrons d'abord que si l'on sait démontrer l'assertion du théorème pour un certain idéal $\mathfrak{c}$, on sait aussi la démontrer pour tout idéal $\mathfrak{c}_1$ de codimension 1 dans $\mathfrak{c}$. Pour cela, on choisit $e_m \in \mathfrak{c} \setminus \mathfrak{c}_1$, et on pose $x_m = e_m - 2i\pi\lambda_m$, $\lambda_m = \lambda(e_m)$; on identifie x_m à son image dans U_{λ_1}, où $\lambda_1 = \lambda|_{\mathfrak{c}_1}$. Alors x_m est central dans U_{λ_1} et engendre $P_\lambda / P_{\lambda_1}$; d'après le lemme 2.4 on a donc $N_\lambda / N_{\lambda_1} = \mathscr{S}_{\lambda_1} * x_m = x_m * \mathscr{S}_{\lambda_2} = P_\lambda * \mathscr{S}_{\lambda_1}$, ce qui prouve déjà en appliquant l'hypothèse de récurrence que $\mathscr{S}_{\lambda_1} * I_{\lambda_1} = I_{\lambda_1} * \mathscr{S}_{\lambda_1}$.

Soient $x'_{m+1}, \ldots, x'_d$; $u'_{m+1}, \ldots, u'_d$ une suite centralisante régulière de U_λ engendrant I_λ, et des applications linéaires continues $M_\lambda \to \mathscr{S}_\lambda$ vérifiant

$$\varphi = \sum_{j=m+1}^{d} u'_j(\varphi) * x'_j \quad \text{pour toute } \varphi \in M_\lambda.$$

Pour $m+1 \leqq j \leqq d$, soit x_j un relèvement quelconque de x'_j dans U_{λ_1}. Alors les x_j pour $m \leqq j \leqq d$ forment une suite centralisante régulière de U_{λ_1} engendrant I_{λ_1}. Soit a une scission de la surjection canonique $\mathscr{S}_{\lambda_1} \to \mathscr{S}_\lambda$ (lemme 2.4), et soit, pour $m+1 \leqq j \leqq d$, u_j l'application $M_{\lambda_1} \to \mathscr{S}_{\lambda_1}$ obtenue en composant u'_j avec le relèvement a et la projection canonique $M_{\lambda_1} \to M_\lambda$. Alors pour toute $\varphi \in M_{\lambda_1}$ on a:

$$\varphi - \sum_{j=m+1}^{d} u_j(\varphi) * x_j = 0 \bmod N_\lambda / N_{\lambda_1}.$$

Or comme on l'a vu on a $N_\lambda / N_{\lambda_1} = \mathscr{S}_{\lambda_1} * x_m$. Soit $v: N_\lambda / N_{\lambda_1} \to \mathscr{S}_{\lambda_1}$ l'inverse de l'application $\varphi \to \varphi * x_m$, et posons:

$$u_m(\varphi) = v\left(\varphi - \sum_{j=m+1}^{d} u_j(\varphi) * x_j\right).$$

Alors il est clair que $x_m, \ldots, x_d, u_m, \ldots, u_d$ répondent à la question.

b) D'après a), on peut supposer que $\mathfrak{c} = \mathfrak{z}$. Si $\operatorname{Ker}\lambda \neq 0$, on est aussitôt ramené à l'hypothèse de récurrence. Supposons que $\operatorname{Ker}\lambda = 0$; alors nous sommes dans la situation du n° 2.5 dont on adopte les notations.

Par hypothèse de récurrence, on a une suite centralisante régulière $\tilde{x}_2, \ldots, \tilde{x}_d$ de $\tilde{U}_{\tilde{\lambda}}$ engendrant $\tilde{I}_{\tilde{\lambda}}$, et des applications $\tilde{u}_2, \ldots, \tilde{u}_d$: $\tilde{M}_{\tilde{\lambda}} \to \tilde{\mathscr{S}}_{\tilde{\lambda}}$ vérifiant l'assertion du théorème. Dans la décomposition $U_\lambda = W_1 \otimes \tilde{U}_{\tilde{\lambda}}$ du n° 2.6, posons $x_j = 1 \otimes \tilde{x}_j$ pour $2 \leqq j \leqq d$; et soit $u_j: M_\lambda \to \mathscr{S}_\lambda$ l'application définie par $\varphi \to \tilde{u}_j \circ \varphi$ dans la décomposition $\mathscr{S}_\lambda = \mathscr{S}(\mathbb{R}^2, \tilde{\mathscr{S}}_{\tilde{\lambda}})$ (rappelons que dans cette décomposition $M_\lambda = \mathscr{S}(\mathbb{R}^2, \tilde{M}_{\tilde{\lambda}})$). D'après la remarque 2.8b), les $x_2, \ldots, x_d$, $u_2, \ldots, u_d$ répondent à la question.

3.6. **Corollaire.** *On a* $M^{n+1} = \mathscr{S} * I^{n+1} = I^{n+1} * \mathscr{S}$ *pour tout* $n \in \mathbb{N}$.

3.7. **Corollaire.** *On a* $\mathscr{S} * M^{n+1} = M^{n+1}$ $\forall\, n \in \mathbb{N}$.

Démonstration. (dans les deux cas): D'après [4], th. 7.2 on a $\mathscr{S} * \mathscr{S} = \mathscr{S}$.

4. Action de $\mathscr{S}$ dans les modules de longueur finie

4.1. **Définition.** Soit W un G-module de Fréchet. Nous dirons que W est tempéré, si pour toute seminorme continue q sur W il existe une seminorme continue q', une constante $a>0$ et un entier $m\in\mathbb{N}$ tels que:

$$q(\exp(\xi)\,w)\leqq a(1+|\xi|^m)\,q'(w)$$

pour tous $\xi\in\mathfrak{g}$, $w\in W$, où $|\ |$ est une norme quelconque sur $\mathfrak{g}$ (par exemple $|\xi|=|\xi_1|+\ldots+|\xi_m|$ si ξ s'écrit $\xi_1 e_1+\ldots+\xi_m e_m$ dans une base $e_1,\ldots,e_m$ de $\mathfrak{g}$).

(On peut bien sûr restreindre les seminormes à appartenir à un ensemble fondamental de semi normes définissant la topologie de W).

4.2. **Lemme.** *Soit (W,ρ) un G-module tempéré. Alors pour tout $\varphi\in\mathscr{S}$ et tout $w\in W$ l'intégrale $\int_G \varphi(g)\,\rho(g)\,w\,dg$ converge, et définit un endormorphisme $\rho(\varphi)$ de W. On obtient ainsi un homomorphisme d'algèbres $\rho:\mathscr{S}\to\mathrm{End}(W)$ faisant de W un $\mathscr{S}$-module topologique.*

Démonstration. Vérification immédiate.

4.3. **Lemme.** *Soit $(V,\rho)\in\mathbf{Ext}_G(E_\infty)$* (cf. n° 1.2). *Alors V est un G-module tempéré.*

Démonstration. Par récurrence sur la longueur $l(V)$ de V. Si $l(V)=1$, on a $V\simeq E_\infty$, dont on vérifie aussitôt qu'il est tempéré (de façon générale, un G-module unitaire est évidemment tempéré, et l'espace des vecteurs C^∞ d'un G-module tempéré est un G-module tempéré).

Supposons donc $l(V)>1$, et considérons une suite exacte forte

$$0\to V'\to V\to E_\infty\to 0$$

avec $l(V')=l(V)-1$. Soit $\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty,V')$ l'espace des vecteurs algébriques du $\mathfrak{g}$-module $\mathrm{Hom}(E_\infty,V')$ (i.e. ceux qui engendrent sous l'action de $\mathfrak{g}$ un sous-module de dimension finie à sous-quotients simples triviaux). Alors l'inclusion canonique $\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty,V')\subset\mathrm{Hom}(E_\infty,V')$ induit un isomorphisme en cohomologie. (Démonstration: si $V'\simeq E_\infty$, c'est le th. 4.1 de [1], partie I, en remarquant que l'image de U/I dans $\mathrm{End}(E_\infty)$ n'est autre que $\mathrm{End}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty)$. Dans le cas général on considère une suite exacte

$$0\to V''\to V'\to E_\infty\to 0$$

avec $l(V'')=l(V')-1$; en réalisant comme ci-après V' à l'aide d'un cocycle polynomial à valeurs dans $\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty,V'')$ on en déduit une suite exacte:

$$0\to\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty,V'')\to\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty,V')\to\mathrm{End}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty)\to 0$$

et on conclut par récurrence sur $l(V')$.) On peut donc choisir un relèvement de l'espace de Fréchet E_∞ dans V tel que ρ se réalise dans l'espace $V'\oplus E$ par la formule:

$$g\to\begin{pmatrix}\rho'(g) & \zeta(g)\,\pi(g)\\ 0 & \pi(g)\end{pmatrix}$$

avec $\zeta \in Z^1(G, \mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty, V'))$; de plus il résulte de [1] partie II n°s 2.2 et 2.3 que l'on peut choisir le relèvement pour que ζ soit un polynôme.

En d'autres termes, si $e_1, \ldots, e_n$ est une base de $\mathfrak{g}$, $\xi = \xi_1 e_1 + \ldots + \xi_n e_n$, on peut écrire:

$$\zeta(\exp(\xi)) = \sum_{\alpha q w} \xi^\alpha \zeta_\alpha$$

où les ζ_α sont des éléments fixés de $\mathrm{Hom}_{(\mathfrak{g};\,\mathrm{alg})}(E_\infty, V')$, presque tous nuls. Il est alors clair que si q est une seminorme continue sur V', on a une majoration:

$$q(\zeta(\exp(\xi))\, \pi(\exp(\xi))\, x) \leqq a(1 + |\xi|^m)\, q'(x)$$

pour tous $\xi \in \mathfrak{g}$, $x \in E_\infty$, où q' est une seminorme continue sur E_∞; le résultat s'en déduit aussitôt.

4.4. Le lemme précédent nous permet de définir une action continue de $\mathscr{S}$ sur tout $V \in \mathbf{Ext}_G(E_\infty)$; remarquons que d'après le cor. 3.4 on obtient ainsi un $\mathscr{S}$-module de longueur finie au sens de l'algèbre, à sous-quotients simples E_∞.

4.5. **Proposition.** *Soit $V \in \mathbf{Ext}_G(E_\infty)$. Alors tout sous-$\mathscr{S}$-module de V est fermé, et se trouve dans $\mathbf{Ext}_G(E_\infty)$.*

Démonstration. Par récurrence sur la longueur $l(V)$ de V (au sens de $\mathbf{Ext}_G(E_\infty)$). Si $l(V) = 1$, $V \simeq E_\infty$ est un $\mathscr{S}$-module simple, et il n'y a rien à démontrer.

Supposons $l(V) > 1$, et soit W un sous-$\mathscr{S}$-module de V. Considérons une suite exacte forte $0 \to V' \to V \to E_\infty \to 0$, avec $l(V') = l(V) - 1$. Si $W' = W \cap V' \neq 0$, par hypothèse de récurrence W' est dans $\mathbf{Ext}_G(E_\infty)$, et est fermé dans V', donc dans V. Alors W est dans $\mathbf{Ext}_G(E_\infty)$ et fermé dans V si et seulement si W/W' l'est dans V/W', ce qui est vrai par hypothèse de récurrence.

On peut donc supposer que $W \cap V' = 0$, et $W \neq 0$. Alors la projection canonique $p: V \to E_\infty$ se restreint en un isomorphisme de $\mathscr{S}$-modules de W sur E_∞. Soit w un élément non nul de W. Comme w s'identifie à un élément de E_∞ et que l'action de $\mathscr{S}$ dans E_∞ est continue, l'annulateur L de w dans $\mathscr{S}$ est un idéal à gauche maximal et fermé de $\mathscr{S}$. Alors l'application $(v', \varphi) \to (v', \varphi w)$ de $V' \oplus \mathscr{S}$ sur V passe au quotient en une bijection continue de l'espace de Fréchet $V' \oplus \mathscr{S}/L$ sur V, dans laquelle W s'identifie à $\mathscr{S}/L$. D'après le théorème du graphe fermé, cette bijection est bicontinue, ce qui prouve bien que W est un sous-espace fermé de V.

Or l'utilisation d'une unité approchée dans $\mathscr{S}$ (formée par exemple d'éléments de $C_c^\infty(G)$) montre aussitôt que tout sous-$\mathscr{S}$-module fermé d'un G-module tempéré est en fait un sous-G-module. Donc W est un sous-G-module fermé de V, isomorphe à E_∞ (par p), cqfd.

4.6. **Corollaire.** *Tout sous-G-module fermé de $V \in \mathbf{Ext}_G(E_\infty)$ est dans $\mathbf{Ext}_G(E_\infty)$. En particulier V est un G-module topologiquement de longueur finie au sens «naïf» du terme* (cf. [8], §8, n° 8.1 *par exemple*), *ce qui veut dire que la longueur des suites croissantes de sous-espaces fermés G-stables est bornée (par $l(V)$, ici). En outre, puisque ces suites se trouvent en fait dans $\mathbf{Ext}_G(E_\infty)$; elles vérifient le théorème de Jordan-Hölder.*

Démonstration. Tout sous-G-module fermé de V est manifestement un sous-$\mathscr{S}$-module.

5. Decomposition des algebres $\mathscr{S}/M^{n+1}$

5.1. Les résultats des n°s précédents vont nous permettre de démontrer que l'algèbre $\mathscr{S}_n = \mathscr{S}/M^{n+1}$ est Morita-équivalente à une certaine $\mathbb{C}$-algèbre de dimension finie (avec unité). Pour cela, nous allons voir que, sous l'action de $G \times G$, $\mathscr{S}_n$ est un $G \times G$-module de longueur finie au sens du n° 1.2; puis nous utiliserons l'action de $G \times G$ pour obtenir la décomposition voulue.

Notons $\mathscr{T}$ l'algèbre $\mathscr{S}(G \times G) = \mathscr{S} \hat{\otimes} \mathscr{S}$, et $\mathscr{B}$ l'image de $\mathscr{T}$ dans $\mathscr{L}(\mathscr{L}^2(E))$; on a donc $\mathscr{B} = \mathscr{L}^2(\mathscr{L}^2(E))_\infty$ d'après le th. 3.3, les vecteurs C^∞ étant pris pour l'action de $G^4 = G \times G \times G \times G$.

5.2. **Lemme.** *Soit H un sous-groupe fermé connexe de G; alors l'action de G sur $\mathscr{S}(G/H)$ par translations à gauche fait de $\mathscr{S}(G/H)$ un G-module tempéré* (déf. 4.1).

(Démonstration laissée au lecteur).

5.3. En particulier, $\mathscr{S}$ est un $G \times G$-module tempéré, et donc un $\mathscr{T}$-module via la formule du lemme 4.2; et l'application $\pi: \mathscr{S} \to \mathscr{A}$ commute aux actions de $\mathscr{T}$. Par conséquent M est un sous-$\mathscr{T}$-module fermé de $\mathscr{S}$.

5.4. **Définition.** *Soit (V, ρ) un $G \times G$-module tempéré, X un G-module de dimension finie à sous-quotients simples triviaux. On munit $V \otimes X$ d'une structure de $G \times G$-module tempéré en faisant agir G à gauche par $g \cdot (v \otimes x) = gv \otimes x$ et à droite par $(v \otimes x)\, g = vg \otimes g^{-1} x$.*

(Ici et dans la suite nous ferons passer librement l'action du deuxième facteur G dans un $G \times G$-module en une action à droite, i.e. vg signifie $(1, g^{-1})\, v$.)

5.5. **Lemme.** *Pour tout $n \in \mathbb{N}$, M^{n+1} est un sous-$\mathscr{T}$-module de $\mathscr{S}$.*

Démonstration. D'après le cor. 3.6 on a $M^{n+1} = \mathscr{S} * I^{n+1}$. Or il est clair que pour tout sous-G-module de dimension finie X de U (pour l'action adjointe), l'application $\mathscr{S} \otimes X \to \mathscr{S}$ définie par $\varphi \otimes x \to \varphi * x$ commute aux actions de $G \times G$ si l'on fait agir $G \times G$ sur $\mathscr{S} \otimes X$ comme à la déf. 5.4. Comme il s'agit de $G \times G$-modules tempérés, cette application commute donc aux actions de $\mathscr{T}$. Si maintenant on prend pour X un sous-G-module de dimension finie de U engendrant I^{n+1} comme idéal à gauche, on a clairement $\mathscr{S} * X = \mathscr{S} * I^{n+1} = M^{n+1}$, donc M^{n+1} est un sous-$\mathscr{T}$-module de $\mathscr{S}$ comme image d'un $\mathscr{T}$-morphisme.

5.6. **Proposition.** *Pour tout $n \in \mathbb{N}$, M^{n+1} est un idéal fermé de $\mathscr{S}$, et $\mathscr{S}_n \in \mathbf{Ext}_{G \times G}(\mathscr{A})$.*

Démonstration. On raisonne par récurrence sur n. Pour $n = 0$ le résultat est trivial; supposons donc $n > 0$. Démontrons d'abord que M^{n+1} est fermé dans $\mathscr{S}$, et pour cela qu'il est fermé dans M.

Choisissons comme ci-dessus un sous-G-module de dimension finie X de U engendrant I comme idéal à gauche; on a alors une surjection de $\mathscr{T}$-modules $a: \mathscr{S} \otimes X \to M$. Comme il s'agit d'espaces de Fréchet, la bijection $\mathscr{S} \otimes X/\mathrm{Ker}\, a \to M$ correspondante est un isomorphisme. Donc pour montrer que

M^{n+1} est fermé dans M, il suffit de montrer que $a^{-1}(M^{n+1})$ est fermé dans $\mathscr{S}\otimes X$. Or $a^{-1}(M^{n+1})$ contient certainement $M^n\otimes X$, qui est fermé dans $\mathscr{S}\otimes X$ par hypothèse de récurrence; donc il suffit de prouver que $a^{-1}(M^{n+1})/M^n\otimes X$ est fermé dans $\mathscr{S}_{n-1}\otimes X$. Comme $\mathscr{S}_{n-1}\otimes X$ (avec l'action de $G\times G$ donnée par la déf. 5.4) est un $G\times G$-module de longueur finie, et que $a^{-1}(M^{n-1})/M^n\otimes X$ en est un sous-$\mathscr{T}$-module d'après le lemme 5.5, la conclusion résulte de la prop. 4.5.

Le même argument montre de plus que $M/M^{n+1}\in\mathbf{Ext}_{G\times G}(\mathscr{A})$. Vu la suite exacte forte $0\to M/M^{n+1}\to\mathscr{S}_n\to\mathscr{A}\to 0$, on en déduit bien que $\mathscr{S}_n\in\mathbf{Ext}_{G\times G}(\mathscr{A})$, cqfd.

5.7. Nous avons rappelé au n° 1.2 comment le choix d'un relèvement R de A dans $\hat{U}$ (notations du n° 1.2) permet d'établir une équivalence de catégories entre $\mathbf{Ext}_G(E_\infty)$ et $\mathbf{Ext}_C(\mathbb{C})$.

Pour toute $\mathbb{C}$-algèbre B, notons B^0 l'algèbre opposée; identifions U à U^0 par l'antiautomorphisme principal ([3], ch. 2, n° 2.2.18), noté $u\to u^\tau$ (c'est l'unique autiautomorphisme de U tel que $\xi^\tau=-\xi$ si $\xi\in\mathfrak{g}$). Si l'on note $u\to\bar{u}$ la conjugaison de U, on remarque que $I^0=\bar{I}=I^\vee$, noyau de l'action de U dans $E_\infty^\vee$.

Par composition avec $u\to u^\tau$, on obtient donc un relèvement $R\otimes R^0$ de $A\otimes A^0$ dans $U\otimes U^0\simeq U\otimes U$. Faisant alors la construction du n° 1.2 pour la catégorie $\mathbf{Ext}_{G\times G}(\mathscr{A})$, on obtient une équivalence:

$$\mathbf{Ext}_{G\times G}(\mathscr{A})\xrightarrow{\sim}\mathbf{Ext}_{C\otimes C^0}(\mathbb{C})$$

par le foncteur $\mathrm{Hom}_{R\otimes R^0}(\mathscr{A},\cdot)$, et pour tout $V\in\mathbf{Ext}_{G\times G}(\mathscr{A})$ une décomposition:

$$V\simeq\mathscr{A}\otimes\mathrm{Hom}_{R\otimes R^0}(\mathscr{A},V).$$

Appliquant cela à $V=\mathscr{S}_n$, ce qui est loisible d'après la prop. 5.6, on obtient une décomposition $\mathscr{S}_n=\mathscr{A}\otimes C'_n$, et vu la fonctorialité de la construction, les C'_n forment un système projectif à flèches surjectives de $\mathbb{C}$-espaces vectoriels de dimension finie.

Nous allons maintenant munir les C'_n d'une structure d'algèbre, de telle sorte que la décomposition ci-dessus soit un isomorphisme d'algèbres.

5.8. **Lemme.** *Identifions E_∞ à $\mathscr{S}(\mathbb{R}^r)$ de telle sorte que A s'identifie à*

$$\mathbb{C}\left[\frac{\partial}{\partial t_1},\ldots,\frac{\partial}{\partial t_r},t_1,\ldots,t_r\right];$$

alors $\mathscr{A}$ s'identifie à $\mathscr{S}(\mathbb{R}^{2r})$ avec le produit des noyaux. Soit $\mathfrak{a}$ l'algèbre de Lie commutative $\mathbb{R}^{2r}$, que l'on fait agir sur $\mathscr{A}$ par $e_j\to\mathrm{ad}\left(\frac{\partial}{\partial t_j}\right)$ pour $1\leqq j\leqq r$, $e_j\to\mathrm{ad}(t_{j-r})$ pour $r+1\leqq j\leqq 2r$; et soit $\mathrm{Tr}:\mathscr{A}\to\mathbb{C}$ la trace. Alors le complexe d'homologie augmenté:

$$0\leftarrow\mathbb{C}\leftarrow\mathscr{A}\leftarrow\mathfrak{a}\otimes\mathscr{A}\leftarrow\ldots\leftarrow\Lambda^{2r}\mathfrak{a}\otimes\mathscr{A}\leftarrow 0$$

admet une homotopie contractante continue.

Démonstration. Si on écrit $a=a(t_1, \ldots, t_r, s_1, \ldots, s_r)$, on voit que $\mathrm{ad}\left(\frac{\partial}{\partial t_j}\right)$ agit par $\frac{\partial}{\partial t_j}+\frac{\partial}{\partial s_j}$, et $\mathrm{ad}(t_j)$ par multiplication par t_j-s_j. Ecrivant a comme fonction de $\tau_1, \ldots, \tau_r, \sigma_1, \ldots, \sigma_r$, avec $\tau_j=t_j+s_j$, $\sigma_j=t_j-s_j$, on voit que e_j agit par $\frac{\partial}{\partial \tau_j}$ pour $1 \leqq j \leqq r$, par σ_{j-r} pour $r+1 \leqq j \leqq 2r$. Faisant une transformation de Fourier par rapport aux variables σ_j, on remplace σ_j par $\frac{-1}{2i\pi}\frac{\partial}{\partial \sigma_j}$, puis remplaçant e_j par $-2i\pi e_j$ pour $r+1 \leqq j \leqq 2r$, on voit finalement que le complexe de $\mathfrak{a}$-homologie de $\mathscr{A}$ est isomorphe au complexe de de Rham usuel de $\mathscr{S}(\mathbb{R}^{2r})$, la trace étant devenue l'intégrale. D'où le lemme.

Corollaire. *On a* $H_j(A, \mathscr{A})=\mathbb{C}$ *si* $j=0$, 0 *si* $j>0$ *(où* H_j *est l'homologie de Hochschild).*

Démonstration. Comme le montre la démonstration de [1] partie II, lemme 1.4.1, on a $H_j(A, \mathscr{A})=H_j(\mathfrak{a}, \mathscr{A})\ \forall_j \geqq 0$.

5.9. Nous allons maintenant définir un bifoncteur exact

$$\mathbf{Ext}_{G\times G}(\mathscr{A}) \times \mathbf{Ext}_{G\times G}(\mathscr{A}) \to \mathbf{Ext}_{G\times G}(\mathscr{A}),$$

noté $(V_1, V_2) \to V_1 \hat{\otimes}_R V_2$, de la façon suivante.

Tout d'abord, la décomposition $V=\mathscr{A}\otimes \mathrm{Hom}_{R\otimes R^0}(\mathscr{A}, V)$, que nous noterons $V=\mathscr{A}\otimes X(V)$ pour abréger, permet de définir sur $V \in \mathrm{Ext}_{G\times G}(\mathscr{A})$ une nouvelle action de $G\times G$, dite non naturelle, en faisant agir $G\times G$ seulement sur le facteur $\mathscr{A}$. L'action de $\mathscr{T}$ correspondante passe au quotient en une action de $\mathscr{B}$ et fait de V un $\mathscr{B}$-module semi-simple de longueur finie; nous dirons pour simplifier que V est un $\mathscr{B}$-module.

Posons $G^4=G\times G\times G\times G$, et soient $V_1, V_2 \in \mathbf{Ext}_{G\times G}(\mathscr{A})$. On a sur $V_1 \hat{\otimes} V_2$ une action de G^4, qui fait de $V_1 \hat{\otimes} V_2$ un objet de $\mathbf{Ext}_{G^4}(\mathscr{B})$, de longueur $l(V_1)\cdot l(V_2)$; on obtient ainsi un foncteur exact $\mathbf{Ext}_{G\times G}(\mathscr{A}) \times \mathbf{Ext}_{G\times G}(\mathscr{A}) \to \mathbf{Ext}_{G^4}(\mathscr{B})$. Faisons agir les deux facteurs G centraux par leur action non-naturelle; on obtient alors une action de $U\otimes U$ passant au quotient en une action de $A^0\otimes A$: c'est l'action obtenue en identifiant $A^0\otimes A$ à la sous-algèbre $R^0\otimes R$ de $U\otimes U$. Pour cette action, $V_1 \hat{\otimes} V_2$ est isomorphe à une somme directe de $l(V_1)\, l(V_2)$ copies de $\mathscr{B}$.

Or l'application canonique $\mathscr{A}\otimes\mathscr{A} \to \mathscr{A}$ définie par $\varphi_1\otimes\varphi_2 \to \varphi_1\varphi_2$ se prolonge en une «contraction» notée τ, de $\mathscr{B}=\mathscr{A}\hat{\otimes}\mathscr{A}$ vers $\mathscr{A}$.

Si on écrit $\mathscr{B}=E_\infty \hat{\otimes} E_\infty^\vee \hat{\otimes} E_\infty \hat{\otimes} E_\infty^\vee = E_\infty \hat{\otimes} \mathscr{A}^\vee \hat{\otimes} E_\infty^\vee$, on voit que τ s'obtient en prenant la trace des éléments de $\mathscr{A}^\vee$. Dans les notations du lemme 5.8, on en déduit que le complexe:

$$0 \leftarrow \mathscr{A} \leftarrow \Lambda\mathfrak{a}\otimes\mathscr{B}$$

admet une homotopie contractante continue, et en particulier, par le même argument que pour le corollaire dudit lemme, que $H_0(A, \mathscr{B})$ est séparé et isomorphe à $\mathscr{A}$, comme $G\times G$-module si on fait agir $G\times G$ sur $H_0(R, \mathscr{B})$ via l'action des deux facteurs extrêmes de G^4.

Nous pouvons maintenant définir $V_1 \hat{\otimes}_R V_2$; ce sera $H_0(R, V_1 \hat{\otimes} V_2)$, muni comme ci-dessus d'une structure canonique de $G \times G$-module. Comme le $R^0 \otimes R$-module $V_1 \hat{\otimes} V_2$ est isomorphe à la somme directe de $l(V_1)\, l(V_2)$ copies de $\mathscr{B}$, on voit que $V_1 \hat{\otimes}_R V_2 \in \mathbf{Ext}_{G\times G}(\mathscr{A})$ et est de longueur $l(V_1)\, l(V_2)$.

5.10. Gardons les notations ci-dessus, et soit $V \in \mathbf{Ext}_{G\times G}(\mathscr{A})$. Il est clair que toute application bilinéaire continue $\beta: V_1 \times V_2 \to V$ vérifiant $\beta(x_1 \xi, x_2) = \beta(x_1, \xi x_2)$ pour tout $\xi \in \mathfrak{g}$ (ou même seulement pour tout $\xi \in R$) passe en une application linéaire continue (encore notée β):

$$V_1 \hat{\otimes}_R V_2 \to V.$$

Si β vérifie $\beta(\xi x_1, x_2) = \xi \beta(x_1, x_2)$, $\beta(x_1, x_2 \xi) = \beta(x_1, x_2)\xi$ pour tout $\xi \in \mathfrak{g}$, alors l'application $V_1 \hat{\otimes}_R V_2 \to V$ correspondante est un $G \times G$-morphisme, et induit donc une application:

$$X(V_1 \hat{\otimes}_R V_2) \to X(V).$$

Mais $X(V_1 \hat{\otimes}_R V_2) \simeq X(V_1) \otimes X(V_2)$ de façon évidente, donc on voit que β comme ci-dessus induit un couplage: $X(V_1) \times X(V_2) \to X(V)$, noté $\bar{\beta}$. Par définition, si on écrit $V_1 = \mathscr{A} \otimes X(V_1)$, $V_2 = \mathscr{A} \otimes X(V_2)$, on a alors

$$\beta(a_1 \otimes x_1, a_2 \otimes x_2) = a_1 a_2 \otimes \bar{\beta}(x_1, x_2) \in \mathscr{A} \otimes X(V).$$

5.11. **Lemme.** *Le produit $\mathscr{S}_n \times \mathscr{S}_n \to \mathscr{S}_n$ induit sur l'espace C'_n du* n° 5.7 *une structure de $\mathbb{C}$-algèbre associative avec unité.*

Démonstration. L'application $\beta: \mathscr{S}_n \hat{\otimes} \mathscr{S}_n \to \mathscr{S}_n$ prolongeant la multiplication vérifie les hypothèses du n° 5.10, et définit donc une application $C'_n \otimes C'_n \to C'_n$, i.e. une loi d'algèbre sur C'_n. Par naturalité de la construction, l'associativité de $\mathscr{S}_n$ entraîne celle de C'_n.

Reste à prouver que C'_n possède une unité. Faisons-le d'abord pour $n=0$. Comme C'_0 est de dimension 1, tout revient à montrer que l'application $C'_0 \otimes C'_0 \to C'_0$ est bijective; cela résulte de ce que l'application $\mathscr{A} \hat{\otimes}_R \mathscr{A} \to \mathscr{A}$ qui l'induit est bijective (parce que non nulle). On peut donc identifier C'_0 à $\mathbb{C}$.

Dans le cas général, soit $\mathfrak{m}'_n$ le noyau de l'application canonique $C'_n \to \mathbb{C}$ induite par $\mathscr{S}_n \to \mathscr{A}$. Si l'on pose $M_n = M/M^{n+1} \subset \mathscr{S}_n$, il est clair que $\mathfrak{m}'_n = X(M_n)$. Comme l'algèbre M_n est nilpotente, il en va de même pour l'algèbre $\mathfrak{m}'_n$. Donc C'_n possède un idéal nilpotent de codimension 1.

Soit $x \in C'_n$ se projetant sur 1 dans $\mathbb{C}$. Si l'on filtre C'_n par les puissances de $\mathfrak{m}'_n$, il est clair que la multiplication à gauche ou à droite par x induit l'identité sur le gradué; c'est donc un endomorphisme inversible de C'_n. Il existe donc un unique $y \in C'_n$ tel que $xy = x$; alors pour tout $z \in C'_n$ $xyz = xz$ et comme x est simplifiable à gauche on voit que y est élément neutre à gauche; De même, on démontre l'existence d'un élément neutre à droite y_1; alors $yy_1 = y = y_1$, cqfd.

La proposition suivante résume les résultats de ce n°:

5.12. **Proposition.** *Après choix d'un relèvement R de A dans $\hat{U}$ on a une décomposition d'algèbres $\mathscr{S}_n = \mathscr{A} \otimes C'_n$ où C'_n est une certaine $\mathbb{C}$-algèbre de dimension finie avec unité. En particulier $\mathscr{A}$ s'identifie à une sous-algèbre de $\mathscr{S}_n$ par $a \to a \otimes 1$.*

En passant à la limite projective, on a une décomposition $\hat{\mathscr{S}} = \mathscr{A} \hat{\otimes} C'$, *où* $C' = \varprojlim C'_n$.

Démonstration. Le seul point qui reste à vérifier, compte tenu du n° 5.7 et du lemme 5.11, est le fait que la décomposition $\mathscr{S}_n = \mathscr{A} \otimes C'_n$ respecte les structures d'algèbre; mais cela est évident, puisque la loi d'algèbre sur C'_n est précisément définie par le fait que

$$(a_1 \otimes x_1) * (a_2 \otimes x_2) = a_1 a_2 \otimes x_1 x_2 \qquad \text{pour tous } a_1, a_2 \in \mathscr{A},\ x_1, x_2 \in C'_n.$$

6. Isomorphisme des algebres C et C'

6.1. Notons $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$ la catégorie des $\mathscr{S}$-modules fréchétiques possédant une suite de composition forte $V = V_m \supset V_{m-1} \supset \ldots \supset V_1 \supset V_0 = 0$ à sous-quotients simples E_∞; en particulier, comme E_∞ est un $\mathscr{S}$-module simple (cf. cor. 3.4), V est de longueur finie au sens de l'algèbre. Les morphismes de $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$ sont les entrelacements linéaires continus. Comme pour les catégories $\mathbf{Ext}_G$ du n° 1.2, on vérifie sans mal que la catégorie $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$ est abélienne. Définissons de manière analogue $\mathbf{Ext}_{\mathscr{A}}(E_\infty)$.

Comme on l'a vu au n° 4.4, tout $V \in \mathbf{Ext}_G(E_\infty)$ est canoniquement muni d'une action de $\mathscr{S}$ qui en fait un objet de $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$, et le foncteur $\mathbf{Ext}_G(E_\infty) \to \mathbf{Ext}_{\mathscr{S}}(E_\infty)$ ainsi obtenu est trivialement exact et pleinement fidèle. Comme il est clair qu'inversement tout $V \in \mathbf{Ext}_{\mathscr{S}}(E_\infty)$ peut être muni d'une action de G définissant celle de $\mathscr{S}$ (par exemple en écrivant V comme quotient de $\mathscr{S} \otimes \mathbb{C}^m$ par un sous-$\mathscr{S}$-module fermé), c'est en fait une équivalence de catégories.

Notons $\mathbf{Ext}_C(\mathbb{C})$ la catégorie des C-modules de longueur finie, et de même $\mathbf{Ext}_{C'}(\mathbb{C})$. Compte tenu de ce qui précède et du résultat d'unicité mentionné au n° 1.2, pour prouver que les systèmes projectifs (C_n) et (C'_n) sont isomorphes, il suffit de montrer que les catégories $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$ et $\mathbf{Ext}_{C'}(\mathbb{C})$ sont équivalentes.

6.2. Nous allons pour cela imiter la construction du n° 5.9. Pour tout $V \in \mathbf{Ext}_{\mathscr{S}}(E_\infty)$, posons $Y(V) = \mathrm{Hom}_R(E_\infty, V)$. Comme on l'a rappelé au n° 1.2 l'application canonique $E_\infty \otimes Y(V) \to V$ est un isomorphisme d'espaces de Fréchet.

Avec des notations évidentes, posons pour $W \in \mathbf{Ext}_{G \times G}(\mathscr{A})$, $V \in \mathbf{Ext}_G(E_\infty)$:

$$W \hat{\otimes}_R V = H_0(R, W \hat{\otimes} V).$$

On obtient ainsi un bifoncteur exact de $\mathbf{Ext}_{G \times G}(\mathscr{A}) \times \mathbf{Ext}_G(E_\infty)$ vers $\mathbf{Ext}_G(E_\infty)$, et il est clair que $Y(W \hat{\otimes}_R V) \simeq X(W) \otimes Y(V)$, de sorte que $l(W \hat{\otimes}_R V) = l(W) \cdot l(V)$.

Raisonnant comme au n° 2.5.10, on voit que la structure de $\mathscr{S}_n$-module (n assez grand) sur V définit une structure de C'_n-module sur $Y(V)$, et que dans les décompositions $\mathscr{S}_n = \mathscr{A} \otimes C'_n$, $V = E_\infty \otimes Y(V)$ on a $(a \otimes x) \cdot (v \otimes y) = av \otimes xy$. (En fait, il n'est pas tout à fait évident que $1 \in C'_n$ agit par 1 dans $Y(V)$. Il suffit pour cela de voir que $C'_n Y(V) = Y(V)$, car $Y(V)$ est alors un quotient de $C'_n \otimes \mathbb{C}^m$; et

il suffit de vérifier cela pour $V=E_\infty$, auquel cas on peut prendre $n=0$. Or si $C'_0\, Y(E_\infty)=0$, on aurait $\mathscr{A}E_\infty=0$ d'après la décomposition de l'action de $\mathscr{A}$, ce qui est absurde).

Ainsi, Y est un foncteur exact de $\mathbf{Ext}_{\mathscr{S}}(E_\infty)$ vers $\mathbf{Ext}_{C'}(\mathbb{C})$. Il est clair que ce foncteur est pleinement fidèle et essentiellement surjectif, d'inverse $W\to E_\infty\otimes W$, où l'on fait agir $\hat{\mathscr{S}}$, donc $\mathscr{S}$, sur $E_\infty\otimes W$ via la décomposition $\hat{\mathscr{S}}=\mathscr{A}\hat{\otimes}C'$.

Ceci achève la démonstration du th. 1.5.

6.3. *Remarque.* A posteriori, on obtient un isomorphisme explicite de C_n sur C'_n de la manière suivante: si on pose $U_n=U/I^{n+1}$, la restriction à $\mathscr{A}\otimes C_n$ de l'application $\mathscr{S}_n\otimes U_n\to\mathscr{S}_n$ définissant l'action de U est un homomorphisme d'algèbres $\mathscr{A}\otimes C_n\to\mathscr{S}_n$ (noter que dans la décomposition $\mathscr{S}_n=\mathscr{A}\otimes C'_n$ l'action de C_n se fait sur le facteur C'_n). Comme ce morphisme commute aux actions de $R\otimes R^0$, il est en fait induit par un morphisme d'algèbres de C_n vers C'_n; cela montre déjà que son image est fermée dans $\mathscr{S}_n$. Mais il est clair qu'en fait c'est même un morphisme de U-bimodules, donc de $G\times G$-modules (rappelons que d'après [1] partie III, n° 1.2.1 la structure naturelle de $\mathfrak{g}\times\mathfrak{g}$-module sur $\mathscr{A}\otimes C_n$ provient d'une action de $G\times G$). Donc l'image est un sous-$G\times G$-module fermé de $\mathscr{S}_n$, et appartient à $\mathbf{Ext}_{G\times G}(\mathscr{A})$. Comme le cosocle de $\mathscr{S}_n$ est $\mathscr{A}$, et que l'image de $\mathscr{A}\otimes C_n$ est transverse au cosocle, l'application $\mathscr{A}\otimes C_n\to\mathscr{S}_n$ ci-dessus est surjective. Alors il en va de même pour l'application $C_n\to C'_n$, et comme ces deux algèbres ont même dimension, c'est un isomorphisme.

7. Exemple: Serie de Taylor d'un element de $\mathscr{S}$ en la representation triviale

7.1. Dans ce n°, on suppose que $E=\mathbb{C}$ est la représentation triviale de G. L'action de $\mathscr{S}$ s'écrit donc $\varphi\to\pi(\varphi)=\int_G\varphi(g)\,dg$; et M est l'ensemble des $\varphi\in\mathscr{S}$ d'intégrale nulle. Soit P l'algèbre des fonctions polynômes sur G (identifié à $\mathfrak{g}$ par l'application exponentielle). Il est facile de voir que P est exactement l'espace des $a\in C^\infty(G)$ qui sont algébriques, i.e. qui engendrent sous l'action régulière gauche de $\mathfrak{g}$ un sous-module de dimension finie à sous-quotients simples triviaux. Soit $\mathfrak{m}$ l'idéal d'augmentation de U (i.e. le noyau de la représentation triviale), de sorte que $M=\mathfrak{m}*\mathscr{S}=\mathscr{S}*\mathfrak{m}$. Par définition, tout élément de P est annulé par une puissance de $\mathfrak{m}$; notons $P(n)=\{a\in P \text{ t.q. } \mathfrak{m}^{n+1}a=0\}$, $n\in\mathbb{N}$.

7.2. **Proposition.** *Soit $n\in\mathbb{N}$. On a $\varphi\in M^{n+1}$ si et seulement si $\int_G a(g)\,\varphi(g)\,dg=0$ pour tout $a\in P(n)$.*

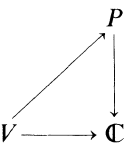
Démonstration. Il est clair que le couplage $(a,\varphi)\to\int_G a(g)\,\varphi(g)\,dg$ de $P(n)\times\mathscr{S}$ vers $\mathbb{C}$ est $\mathfrak{g}$-équivariant. Montrons qu'il passe au quotient en un couplage $P(n)\times\mathscr{S}_n\to\mathbb{C}$, i.e. que $\int_G a(g)\,\varphi(g)\,dg=0$ si $a\in P(n)$, $\varphi\in M^{n+1}$. Cela est évident, vu que φ est somme finie de fonctions au type $\xi_1*\ldots*\xi_{n+1}*\psi$ dans les notations du

n° 2.3, avec $\xi_1, \ldots, \xi_{n+1} \in \mathfrak{g}$, et

$$\int_G a(g)(\xi_1 * \ldots * \xi_{n+1} * \varphi)(g)\, dg = (-1)^{n+1} \int_G (\xi_{n+1} * \ldots * \xi_1 * a)(g)\, \varphi(g)\, dg = 0$$

puisque $\xi_{n+1} * \ldots * \xi_1 * a = 0$ par définition de $P(n)$.

En d'autres termes, on a un $\mathfrak{g}$-morphisme $u: P(n) \to \mathscr{S}_n^*$. Or les invariants de $\mathfrak{g}$ dans $P(n)$ sont réduits aux scalaires, et $u(1) = dg \neq 0$, donc u est injectif. Or il est bien connu que P s'identifie canoniquement en tant que $\mathfrak{g}$-module aux éléments algébriques du dual U^* de U (en remarquant par exemple que, munis respectivement des applications $a \to a(1)$, $f \to f(1)$, P et U^*_{alg} vérifient tous deux la propriété universelle suivante, que nous énonçons pour P: pour tout $\mathfrak{g}$-module algébrique (i.e. dont tous les vecteurs sont algébriques) V, et toute application linéaire $V \to \mathbb{C}$, il existe un unique $\mathfrak{g}$-morphisme $V \to P$ rendant commutatif le diagramme:



où la flèche verticale est $a \to a(1)$). D'après la remarque 1.6a) et le th. 1.5, $P(n)$ et $\mathscr{S}_n^*$ ont donc même dimension; alors u est un isomorphisme, donc le couplage $P(n) \times \mathscr{S}_n \to \mathbb{C}$ est non-dégénéré, et la proposition est démontrée.

7.3. **Corollaire.** *La fonction $\varphi \in \mathscr{S}$ est plate en la représentation triviale (i.e. $\varphi \in M^\infty = \bigcap_{n \geq 1} M^n$) si et seulement si $\int_G a(g)\, \varphi(g)\, dg = 0$ pour tout $a \in P$.*

7.4. *Remarques.* a) On peut s'affranchir de la dépendance en la mesure de Haar dg en considérant l'algèbre des «densités tempérées», i.e. l'algèbre des expressions de la forme $\varphi(g)\, dg$, $\varphi \in \mathscr{S}$. On a alors une construction complètement intrinsèque, et vu la définition du produit de convolution des densités, il est clair que l'application $\hat{\mathscr{S}} \xrightarrow{\sim} P^*$ qui se déduit de la proposition est un isomorphisme d'algèbres, si l'on munit P^* de la structure d'algèbre duale de la structure de cogèbre de P. Mais on a vu dans la démonstration que $P^* = \varprojlim P(n)^* \simeq \hat{U}$, et là aussi c'est un isomorphisme d'algèbres. D'où un isomorphisme $\hat{\mathscr{S}} \xrightarrow{\sim} \hat{U}$, qui est le même que celui de la remarque 6.3.

b) La condition $\int_G a(g)\, \varphi(g)\, dg = 0 \ \forall a \in P$ ne dépend que de la structure de variété algébrique de G. En particulier, si on munit G de la structure de groupe commutatif provenant de son identification avec $\mathfrak{g}$ par l'application exponentielle, on obtient le critère suivant: $\varphi \in \mathscr{S}$ est plate en la représentation triviale, si et seulement si la transformée de Fourier ordinaire de l'application $\xi \to \varphi(\exp(\xi))$ est plate au point 0 de $\mathfrak{g}^*$.

Références

1. Du Cloux, F.: Extensions entre représentations unitaires irréductibles des groupes de Lie nilpotents, in Homologie, groupes Ext^n, représentations de longueur finie des groupes de Lie. Astérisque **124-125**, 129-212 (1985)

2. Connes, A.: Non-commutative differential geometry. Publ. Math., Inst. Hautes Etud. Sci. **62**, 257–360 (1986)
3. Dixmier, J.: Algèbres enveloppantes. Paris: Gauthier-Villars 1974
4. Dixmier, J., Malliavin, P.: Factorisations de fonctions et de vecteurs indéfiniment différentiables. Bull. Sci. Math. **102** (2), 305–330 (1978)
5. Gabriel, P.: Des catégories abéliennes. Bull. Soc. Math. France **90**, 323–448 (1962)
6. Hochschild, G.: La structure des groupes de Lie. Paris: Dunod 1968
7. Howe, R.: On a connection between nilpotent groups and oscillatory integrals associated to singularities. Pac. J. Math. **73**, (n° 2) 329–363 (1977)
8. Kirillov, A.A.: Eléments de la théorie des représentations. Ed. Mir, Moscou, 1974

Oblatum 20-V-1986

Invent. math. 88, 395–403 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

L-functions of *p*-adic characters and geometric Iwasawa theory

Richard Crew

Department of Mathematics, Boston University, Boston, MA 02215, USA

§ 1. Introduction

Let $X/\mathbf{F}_q$ be a smooth affine geometrically connected curve defined over a finite field of characteristic p. In this paper we will study the analytic and arithmetic properties of the L-functions attached to p-adic characters of $\pi_1(X)$. By a p-adic character, we mean a continuous homomorphism

$$\rho\colon \pi_1(X) \to \Lambda^\times$$

where Λ is a complete local noetherian ring of mixed characteristic p with maximal ideal m and *finite* residue field $k=\Lambda/m$. Corresponding to ρ is an étale sheaf of Λ-modules $\mathscr{G}$ on X, and the cohomology group $H^1_c(X\otimes\bar{\mathbf{F}}_q, \mathscr{G})$ is known to be a *free* Λ-module (Lemma (2.1) below). It therefore makes sense to define

$$H(\rho, T)=\det_\Lambda(1-TF\,|\,H^1_c(X\otimes\bar{\mathbf{F}}_q, \mathscr{G}))\in\Lambda[T]$$

and we want to compare this function with the L-function

$$L(\rho, T)=\prod_x (1-\rho(\mathrm{Frob}_x)\,T^{\deg x})^{-1}\in\Lambda[\![T]\!].$$

The result is the following, in which $\Lambda[T]_m$ denotes the m-adic completion of $\Lambda[T]$:

(1.1) **Theorem.** $L(\rho, T)/H(\rho, T)\in\Lambda[T]^\times_m$.

In other words the formal power series $L(\rho, T)$ converges for $|T|\leqq 1$ and has the same zeroes on the unit circle as $H(\rho, T)$. Theorem (1.1) is what a conjecture of Katz ([9] 6.1) would assert in this situation, except that the original conjecture would predict that $L(\rho, T)$ is actually p-adically meromorphic in T. Of course one cannot expect in general that $L(\rho, T)$ is rational, as is the case for l-adic representations. The (possibly nonrational) cases in which $L(\rho, T)$ is known to be meromorphic are relatively few, although the ones we discuss in

§4 are of this type (cf. [1, 4]). As to the statement about the zeroes, this seems to be the first time that the conjecture has been checked in an interesting case (e.g. characters of infinite order). When ρ is of finite order, Theorem (1.1) is a consequence of a theorem of Bloch and Illusie [8] identifying the p-adic étale cohomology of a proper smooth variety with the unit-root subspace of the crystalline cohomology (cf. also [5]). We will reduce the general case to this one.

One important application arises when Λ is the completed group ring of a torsion-free finitely generated abelian pro-p-group Γ, and ρ is a product of a character $\rho_0: \pi_1(X) \twoheadrightarrow \Gamma \hookrightarrow \Lambda^\times$ and a character $\chi: \pi_1(X) \to \Lambda^\times$ of finite prime-to-p order. Then the special value $L(\rho, 1)$ is readily seen to be a "Stickelberger element" (cf. [14] IV 1.5, 1.6) corresponding to the "Γ-tower" of coverings of X, while $H(\rho, 1)$ is, as we shall see in § 2, a characteristic element, in the sense of Iwasawa theory, of a Λ-module $M(\rho)$ defined in a manner analogous to the inverse limits of p-class groups studied in the classical cyclotomic theory. Setting $T=1$ in Theorem (1.1) yields $L(\rho, 1)/H(\rho, 1) \in \Lambda^\times$, i.e. the Stickelberger element is the same, up to a unit, as the characteristic element of the Λ-module $M(\rho)$, an assertion analogous to the "Main Conjecture" of Iwasawa theory proven by Mazur-Wiles.

The last section is devoted to the special case when $\Gamma = \mathbf{Z}_p$ and the Γ-tower in question is the tower of *Igusa curves* studied by Mazur and Wiles [12]. The interest of this case is that the characteristic element of $M(\rho)$ is divisible by a twist of the Kubota-Leopoldt L-function. On the other hand it is not difficult to give congruences ((4.0.4) below) for $L(\rho, 1)$ in terms of special values of Hecke polynomials, and this in turn gives arithmetic information about the Igusa tower.

Similar considerations ought to apply to the towers one obtains from division points of Drinfeld modules. The interest of *this* case is that David Hayes has obtained very beautiful explicit formulas for $L(\rho, 1)$ analogous to those of Kronecker-Siegel-Ramachandra [6, 7]. In fact, the general motivation for Theorem (1.1) is that $L(\rho, 1)$ is a more accessible and computable quantity than the arithmetically more transparent $H(\rho, 1)$.

It is a pleasure to thank Barry Matur for directing my interest to these topics, and I would also like to thank Nicholas Katz and Ofer Gabber for some helpful observations.

§ 2. Proof of the main theorem

We retain the notation of the introduction: $X/\mathbf{F}_q$ is a smooth and affine, geometrically connected curve, Λ is complete, noetherian, and local with maximal ideal m, Λ/m is finite, and $\mathscr{G}$ is the lisse sheaf of Λ-modules on X corresponding to $\rho: \pi_1(X) \to \Lambda^\times$.

(2.1) **Lemma.** *$H^1_c(X \otimes \bar{\mathbf{F}}_q, \mathscr{G})$ is a free Λ-module of finite rank, and $H^i_c(X \otimes \bar{\mathbf{F}}_q, \mathscr{G}) = 0$ if $i \neq 1$.*

Proof. Since Λ is local and noetherian, it is enough to show that

(2.1.1) $\mathrm{Tor}_1^{\Lambda}(H_c^1(X\otimes\bar{\mathbf{F}}_q, \mathscr{G}), \Lambda/m)=0.$

(2.1.2) $H_c^1(X\otimes\bar{\mathbf{F}}_q, \mathscr{G})$ is finitely generated over Λ.

Since X is an affine curve, it follows from Artin-Schreier theory that $R\Gamma_c(\mathscr{M})=H_c^1(\mathscr{M})[-1]$ for any p-adic étale sheaf $\mathscr{M}$ on $X\otimes\bar{\mathbf{F}}_q$ (cf. SGA 4 1/2 pp. 120-1 for the situation mod p). Since $\mathscr{G}$ is Λ-flat, we have

$$H_c^1(\mathscr{G})\overset{L}{\otimes}\Lambda/m^n[-1]\simeq R\Gamma_c(\mathscr{G})\overset{L}{\otimes}\Lambda/m^n \tag{2.1.3}$$
$$\simeq R\Gamma_c(\mathscr{G}\overset{L}{\otimes}\Lambda/m^n)$$
$$\simeq H_c^1(\mathscr{G}\otimes\Lambda/m)[-1].$$

For $n=1$, this proves (2.1.1). As for (2.1.2), this follows SGA 4 1/2 Rapport 4.9 when Λ is finite. In general, we see that $H_c^1(\mathscr{G}\otimes\Lambda/m^n)$ is free over Λ/m^n and $H_c^1(\mathscr{G}\otimes\Lambda/m^n)\simeq H_c^1(\mathscr{G})\otimes\Lambda/m^n$ by (2.1.3). Then (2.1.2) follows by taking inverse limits.

(2.2) We will prove (1.1) by reducing to the case when Λ is the ring of integers in a local field. Observe first that by (2.1), the formation of $H_c^1(X\otimes\bar{\mathbf{F}}_q, \mathscr{G})$ commutes with any change of complete local noetherian rings $\Lambda\to\Lambda'$ with finite residue field; therefore if the theorem holds for the character $\pi_1(X)\to\Lambda^\times$, it holds for the composite $\pi_1(X)\to\Lambda^\times\to(\Lambda')^\times$. Furthermore, it is enough to prove (1.1) with Λ replaced by Λ/m^n for each n, and then take inverse limits. Now $\pi_1(X)\to\Lambda\to\Lambda/m^n$ factors through a finite quotient Γ_n of $\pi_1(X)$, yielding a factorization

$$\pi_1(X)\to\mathbf{Z}_p[\Gamma_n]\to\Lambda/m^n.$$

The semilocal ring $\mathbf{Z}_p[\Gamma]$ can be replaced by its localization at a maximal ideal, so it is enough to prove (1.1) in the following case: $\Lambda=R[\Gamma]$, where R is the ring of integers in a local field, Γ is a finite p-group, and $\rho\colon\pi_1(X)\to R[\Gamma]$ has finite order. One sees immediately, however, that the theorem will be true for $\pi_1(X)\to R[\Gamma]^\times$ if it is true for every composite character $\pi_1(X)\to R[\Gamma]^\times\to S^\times$ where $R[\Gamma]\to S$ is induced by a character $\Gamma\to S^\times$; in fact on $R[\Gamma]$ the m-adic and p-adic topologies coincide, and to recover an element of $R[\Gamma]$ from its images in S involves dividing by at most a factor of $|\Gamma|$. One can take S to be the ring of integers in a local field; this completes the reduction.

Assume therefore that Λ is an integer ring of a local field and that ρ has finite order. Let $G=\ker\rho$ and let $Y\to X$ be the principal G-cover corresponding to ρ. Choose nonsingular compactifications $X\subset\bar{X}$, $Y\subset\bar{Y}$ of X, Y; then the action of G on Y/X extends to $\bar{Y}/\bar{X}$. Writing j for the inclusion $j\colon X\hookrightarrow\bar{X}$, we see that the series $L(\rho, T)$ is the L-function $L(j_!\mathscr{G})$ of the sheaf $j_!\mathscr{G}$ on $\bar{X}$, while for the Artin L-function $L(j_*\mathscr{G}, T)$ we have the expression

$$L(j_*\mathscr{G}, T)=\prod_{i=0}^{2}\det(1-TF|H^i_{\mathrm{cris}}(\bar{Y}\otimes\bar{\mathbf{F}}_q)^\rho)^{(-1)^{i+1}}.$$

$(H^i(\ )^\rho$ denoting the ρ-isotypical part of $H^i_{\text{cris}}(\ )\otimes\mathbf{Q})$ which follows from the usual l-adic formula and the comparison theorem of Katz-Messing [11]. The excision exact sequence in p-adic étale cohomology then yields

$$\frac{L(\rho, T)}{H(\rho, T)} = \frac{\det(1-TF|H^1_{\text{cris}}(\bar{Y}\otimes\bar{\mathbf{F}}_q)^\rho)}{\det(1-TF|H^1_{\text{et}}(\bar{Y}, \mathbf{Q}_p)^\rho)\det(1-TF|H^2_{\text{cris}}(\bar{Y}\otimes\bar{\mathbf{F}}_q)^\rho)}.$$

Now the action of F on $H^2_{\text{cris}}(\bar{Y}\otimes\bar{\mathbf{F}}_q)^\rho$ is divisible by q, while the eigenvalues F on $H^1_{\text{et}}(\bar{Y}\otimes\bar{\mathbf{F}}_q, \mathbf{Q}_p)^\rho$ are, by a well-known theorem of Bloch and Illusie, equal to the p-adic unit eigenvalues of F on $H^1_{\text{cris}}(\bar{Y}\otimes\bar{\mathbf{F}}_q)^\rho$ ([8] II 5.2.2; in fact when G is trivial, the assertion goes back to Serre [13]). Therefore $L(\rho, T)/H(\rho, T)$ is a quotient of products of terms like $(1-\alpha T)$ with α a non-unit, and so $L(\rho, T)/H(\rho, T)\in\Lambda[T]^\times_m$.

§ 3. $H(\rho, 1)$ as a characteristic function

We now study the case when Λ is a completed group ring: $\Lambda = R[\![\Gamma]\!]$ with $\Gamma \simeq \mathbf{Z}_p\times\ldots\times\mathbf{Z}_p$ and R is the integer ring of a local field, and $\rho\colon \pi_1(X)\to\Lambda^\times$ is the product of a character

$$\rho_0\colon \pi_1(X)\twoheadrightarrow\Gamma\hookrightarrow\Lambda^\times$$

by a character $\chi\colon \pi_1(X)\to R^\times$ of finite, prime-to-p order. The constructions in this section are a slight generalization of Mazur-Wiles [12] § 2.

Set $\Delta = \mathrm{Im}\,\chi$. For any finite quotient H of Γ, let $\pi_H\colon X_H\to X$ be the principal $\Delta\times H$-cover of X obtained from ρ_0 and χ, and set

$$\Lambda_H = R[H] \qquad S = \mathbf{Z}_p[\Delta]$$
$$\mathscr{G}_H = \pi_{H*}\mathbf{Z}_p\otimes_S R_\chi$$

where R_χ is R viewed as an S-module via χ. Then

$$\mathscr{G} = \varprojlim_H \mathscr{G}_H \quad\text{and}\quad \mathscr{G}_H = \mathscr{G}\otimes_\Lambda\Lambda_H$$

where the projection maps in the inverse system $\{\mathscr{G}_H\}$ are obtained from the trace maps.

Choose now nonsingular compactification $X\subset\bar{X}$, $X_H\subset\bar{X}_H$, and set

$$M_H(\rho) = \mathrm{Hom}(\mathrm{Pic}^0\,\bar{X}_H(\mathbf{F}_q), \mathbf{Q}_p/\mathbf{Z}_p)\otimes_S R_\chi \tag{3.0.1}$$
$$M_H(\rho/_{\mathbf{F}_q}) = \mathrm{Hom}(\mathrm{Pic}^0\,\bar{X}_H(\bar{\mathbf{F}}_q), \mathbf{Q}_p/\mathbf{Z}_p)\otimes_S R_\chi.$$

The natural functoriality of Pic^0 induces maps

$$M_{H'}(\rho)\to M_H(\rho) \qquad M_{H'}(\rho_{/\mathbf{F}_q})\to M_H(\rho_{/\mathbf{F}_q})$$

for every surjective map $H'\to H$ of quotients of Γ, and we set

$$M(\rho) = \varprojlim_H M_H(\rho) \qquad M(\rho_{/\mathbf{F}_q}) = \varprojlim_H M_H(\rho_{/\mathbf{F}_q}). \tag{3.0.2}$$

Then $M(\rho)$, $M(\rho_{/\mathbf{F}_q})$ (resp. $M_H(\rho)$, $M_H(\rho_{/\mathbf{F}_q})$) are naturally Λ-modules (resp. Λ_H-modules), although it is *not* in general true that $M_H(\rho) \simeq M(\rho) \otimes_\Lambda \Lambda_H$ (cf. Remark 3.2).

Since $\mathrm{Pic}^0 \bar{X}_H(\mathbf{F}_q) = \mathrm{Pic}^0 \bar{X}_H(\bar{\mathbf{F}}_q)^{\mathbf{F}}$ and R_χ is a flat S-module (by the assumption on χ) we have exact sequences

$$0 \to M_H(\rho_{/\mathbf{F}_q}) \xrightarrow{1-F} M_H(\rho_{/\mathbf{F}_q}) \to M_H(\rho) \to 0 \tag{3.0.3}$$
$$0 \to M(\rho_{/\mathbf{F}_q}) \xrightarrow{1-F} M(\rho_{/\mathbf{F}_q}) \to M(\rho) \to 0.$$

The structure of $M(\rho_{/\mathbf{F}_q})$ is determined by the next result:

(3.1) **Proposition.** *If ρ is ramified at every point of $\bar{X} - X$ then there is a natural isomorphism*

$$M(\rho_{/\mathbf{F}_q}) \simeq H^1_c(X \otimes \bar{\mathbf{F}}_q, \mathscr{G})$$

of Λ-modules with Frobenius action.

(3.1.1) *Remarks.* In particular $M(\rho_{/\mathbf{F}_q})$ is a free Λ-module of finite type, and by (3.0.1) $M(\rho)$ is a torsion Λ-module of finite type whose characteristic function is

$$H(\rho, 1) = \det_\Lambda(1 - F \mid H^1_c(X \otimes \bar{\mathbf{F}}_q, \mathscr{G})).$$

On the other hand $M_H(\rho_{/\mathbf{F}_q})$ is *not* in general a free Λ_H-module (cf. Remark (3.2)).

Proof of (3.1). Since Λ is a domain, $\mathscr{G}$ has by hypothesis no inertia invariants of the points of $\bar{X} - X$. Thus $j_! \mathscr{G} \xrightarrow{\sim} j_* \mathscr{G}$ and

$$H^1_c(X \otimes \bar{\mathbf{F}}_q, \mathscr{G}) \xrightarrow{\sim} H^1(\bar{X} \otimes \bar{\mathbf{F}}_q, j_* \mathscr{G}) = \varprojlim_H H^1(\bar{X} \otimes \bar{\mathbf{F}}_q, j_* \mathscr{G}_H),$$

so it is enough to produce isomorphisms

$$H^1(\bar{X} \otimes \bar{\mathbf{F}}_q, j_* \mathscr{G}_H) \xrightarrow{\sim} M_H(\rho_{/\mathbf{F}_q})$$

compatible with the projection maps for $H' \twoheadrightarrow H$. Since R_χ is flat over S we have

$$j_* \mathscr{G}_H \simeq (\bar{\pi}_{H*} \mathbf{Z}_p) \otimes_S R_\chi$$

where $\bar{\pi}_H \colon \bar{X}_H \to \bar{X}$ is the compactification of π_H; thus

$$\begin{aligned} H^1(\bar{X} \otimes \bar{\mathbf{F}}_q, j_* \mathscr{G}_H) &\simeq H^1(\bar{X}_H \otimes \bar{\mathbf{F}}_q, \mathbf{Z}_p) \otimes_S R_\chi \\ &\simeq \mathrm{Hom}_{\mathbf{F}_q}(\mu_{p^\infty}, \mathrm{Pic}^0 \bar{X}_H) \otimes_S R_\chi \\ &\simeq \mathrm{Hom}(\mathrm{Pic}^0 \bar{X}_H(\bar{\mathbf{F}}_q), \mathbf{Q}_p/\mathbf{Z}_p) \otimes_S R_\chi \simeq M_H(\rho_{/\mathbf{F}_q}) \end{aligned}$$

is the desired isomorphism. Compatibility with the projection maps amounts to observing that in the standard isomorphism

$$H^1(X, \mathbf{Z}/p^r\mathbf{Z}) \simeq \mathrm{Hom}_{\mathbf{F}_q}(\mu_{p^r}, \mathrm{Pic}^0 X) \tag{3.1.2}$$

(X a smooth proper curve over $\bar{\mathbf{F}}_q$) the trace map for a finite map $\pi: X \to Y$ corresponds to the dual of the "Pic functoriality" $\mathrm{Pic}^0 Y \to \mathrm{Pic}^0 X$, which is the same as the map $\mathrm{Pic}^0 X \to \mathrm{Pic}^0 Y$ induced by the norm map $N: \pi_* \mathbf{G}_{mX} \to \mathbf{G}_{mY}$. But (3.1.2) is obtained by applying $H^1(X,)$ to the isomorphism $\mathbf{Z}/p^r\mathbf{Z} \xrightarrow{\sim} \mathrm{Hom}(\mu_{p^r}, \mathbf{G}_m)$, so the desired compatibility follows by applying H^1 to the commutative diagram

$$\begin{array}{ccccc} \pi_* \mathbf{Z}/p^r\mathbf{Z} & \xrightarrow{\sim} & \pi_* \mathrm{Hom}(\mu_{p^r}, \mathbf{G}_m) & \xrightarrow{\sim} & \mathrm{Hom}(\mu_{p^r}, \pi_* \mathbf{G}_{mX}) \\ \downarrow \text{Trace} & & & \swarrow \text{Norm} & \\ \mathbf{Z}/p^r\mathbf{Z} & \xrightarrow{\sim} & \mathrm{Hom}(\mu_{p^r}, \mathbf{G}_{mY}). & & \end{array}$$

(3.2) *Remarks.* When Λ_H is not a domain the map $j_! \mathscr{G}_H \to j_* \mathscr{G}_H$ is not in general an isomorphism. Defining K_H by the exact sequence

$$0 \to j_! \mathscr{G}_H \to j_* \mathscr{G}_H \to K_H \to 0$$

and taking cohomology, we get

$$0 \to H^0(K_H)/H^0(j_* \mathscr{G}_H) \to M(\rho/_{\mathbf{F}_q}) \otimes \Lambda_H \to M_H(\rho/_{\mathbf{F}_q}) \to 0$$

by (3.1), (2.1), and the definitions.

Finally, from (3.0.3), (3.1), and the Hochschild-Serre spectral sequence, we get the curious isomorphism

$$M(\rho) \simeq H^2_c(X, \mathscr{G}).$$

§ 4. The Igusa Tower

In this section we study the representations arising on the Tate module of an elliptic curve with invariant j over the mod p j-line for $p > 2$, and their associated L-series and Iwasawa modules. In this case we will have $\Gamma \simeq \mathbf{Z}_p$, and the corresponding $\mathbf{Z}_p$-extension $\mathbf{F}_p(j)$ is essentially the Igusa Tower ([12] § 3). Most of what we do goes through with little or no change for the more general case of a "modular" family of elliptic curves, but we will confine our treatment to this case since it is typical, and it is the one discussed by Mazur-Wiles [12].

Let $\bar{X}/\mathbf{F}_p$ be the projective j-line and let $f: E \to \bar{X}$ be the Neron model of an elliptic curve over $\mathbf{F}_q(j)$ with invariant j, good reduction outside 0, 1728, ∞, and multiplicative at ∞. If we let $U = \bar{X} - \{j = 0, \infty, 1728,$ and the supersingular j-values$\}$ then $E|U$ is a family of ordinary elliptic curves, and the p-adic Tate module $(T_p E)^{\text{et}}$ is a twisted form of $\mathbf{Z}_p$ on U. It gives rise to a representation $\psi: \pi_1(U) \to \mathbf{Z}_p^\times$, which by a theorem of Igusa is surjective ([10] Theorem 4.3). It is simple to produce a $L(\rho, T)$ of the sort studied in § 3 which interpolates the Artin L-functions $L(\psi^k, T)$ for k congruent mod $p-1$ to a fixed $i \in \mathbf{Z}/(p-1)\mathbf{Z}$. In fact for any $k \equiv i$, the ψ^k are ramified outside of the same open subset $X \subset \bar{X}$

and have the same image in $\mathbf{Z}_p^\times$, so that there are characters

$$\rho_0^{(i)}\colon \pi_1(X) \twoheadrightarrow \Gamma \simeq \mathbf{Z}_p$$
$$\chi^{(i)}\colon \pi_1(X) \to \mathbf{Z}_p^\times \quad \text{of finite order}$$

and for each $k>0$ a character $\psi_k\colon \Gamma \to \mathbf{Z}_p^\times$ such that

$$(\psi_k \circ \rho_0^{(i)})\chi^{(i)} | \pi_1(U) = \psi^k$$

and

$$\psi_k(L(\rho^{(i)}, T)) = L(\psi^k, T)$$

where $\rho^{(i)} = \rho_0^{(i)}\chi^{(i)}$.

We should point out there are *two* elliptic curves over $\mathbf{F}_p(j)$ with good reduction outside $j=0, \infty, 1728$ and multiplicative at ∞ ([3] IV 1.6). Either one is a quadratic twist of the other, so if ψ_1, ψ_2 are the corresponding characters of $\pi_1(U)$, then $\psi_1^2 = \psi_2^2$. Thus when k or i is even, $L(\psi^k, T)$, and $\rho^{(i)}$ are independent of which curve has been chosen.

Denote by $M^{(i)}$ the module $M(\rho^{(i)})$ of § 3. When i is even, $M^{(i)}$ is exactly the Iwasawa module that Mazur-Wiles call $M(\mathbf{F}_p, \omega^{-i})$, where ω is the Teichmuller character. In fact when $\bar{X}_n \to \bar{X}$ is the cover corresponding to $\pi_1(U) \xrightarrow{\psi} \mathbf{Z}_p^\times \to (\mathbf{Z}/p^n\mathbf{Z})^\times$, Mazur-Wiles define $M(\mathbf{F}_p, \theta^i)$ with the same formulas (3.0.1), (3.0.2), but with $\bar{X}_n$ replaced by the Igusa curve $\mathrm{Ig}(p^n)$ of level p^n (they also take the χ-part of Pic^0 rather than of its Pontryagin dual). We must check that $\bar{X}_n$ is a double cover of $\mathrm{Ig}(p^n)$ and that the map $\bar{X}_n \to \mathrm{Ig}(p^n)$ is the quotient map for the subgroup $(\pm 1) \subset (\mathbf{Z}/p^n\mathbf{Z})^\times$ acting on $\bar{X}_n$. By definition $\mathrm{Ig}(p^n)$ is the object representing the moduli problem of elliptic curves provided with a generator of the nth power of the Verschiebung $V^n\colon E^{(p^n)} \to E$. The isomorphism $V^n E^{(p^n)} \xrightarrow{\sim} (p^n E)^{\mathrm{et}}$ shows that E has such a generator P, and since $(E, P) \simeq (E, -P)$, the classifying map $\bar{X}_n \to \mathrm{Ig}(p^n)$ factors through $\bar{X}_n/(\pm 1) \to \mathrm{Ig}(p^n)$. The latter map is necessarily an isomorphism, since both curves have degree $\phi(p^n)/2$ over $\bar{X}$.

Denote by $j\colon U \to \bar{X}$ the inclusion. If for $k>0$ we set

$$\mathscr{F}_k = j_* \operatorname{Symm}^k R^1 f_* \mathbf{Q}_l$$

then it is not difficult to check that for every $x \in \bar{X}$ (not just $x \in U$) we have

$$\psi^k(\mathrm{Frob}_x) = \text{the } p\text{-adic unit eigenvalue of Frobenius on } \mathscr{F}_x$$

where either side is understood to be 0 if it is undefined. Then by comparing Euler factors, we see that for all $k>2$

$$L(\psi^k, T) \equiv \frac{L(\mathscr{F}_k, T)}{L(\mathscr{F}_{k-2}, pT)}(1-pT)^{-1} \bmod p^k. \tag{4.0.3}$$

If we denote by $H_k(T)$ the p-Hecke polynomial of weight k and level one, i.e.

$$H_k(T) = \det(1 - T \cdot T_p + p^{k-1}T^2 | S_k)$$

where S_k is the space of cusp forms of weight k and level one on $SL_2(\mathbf{Z})$, and T_p is the pth Hecke operator, then it is shown in [2] 5.7 that for k even and positive,

$$L(\mathscr{F}_k, T) = H_k(T).$$

Combining this with (4.0.2) and (4.0.3) yields

$$\psi_k(L(\rho^{(i)}, T)) \equiv \frac{H_{k+2}(T)}{H_k(pT)}(1-pT)^{-1} \bmod p^k \tag{4.0.4}$$

for all positive even $k \equiv i(p-1)$.

This gives an analogue for $M^{(i)}$ of the Herbrand-Ribet criterion, as was found (by different methods) by Mazur-Wiles. Since Λ is in our present case a power series ring in one variable, we can define in the usual way the μ and λ invariants of the Λ-module $M^{(i)}$.

(4.1) **Theorem** ([12] Proposition 6). *For any nonzero even $i \in \mathbf{Z}/(p-1)\mathbf{Z}$, the following statements are equivalent:*

(4.1.1) $\mu(M^{(i)}) = \lambda(M^{(i)}) = 0$ (i.e. $M^{(i)} = 0$).

(4.1.2) $H_{k+2}(1) \equiv 0$ (p) for any $k \equiv i(p-1)$.

(4.1.3) *The Atkin U-operator acting on the space of* mod p *cusp forms of weight $k+2$ and level one has a fixed vector.*

Proof. The equivalence of (4.1.1) and (4.1.2) follows from Theorem 1 of § 1 and (4.0.4). The equivalence of (4.1.2) and (4.1.3) is immediate.

From (4.0.3) and a simple induction we get the Kummer-type congruences

$$H_{k+2}(T) \equiv H_{k'+2}(T) \bmod p^n \quad \text{when } k, k' \geqq n \text{ and } k \equiv k' \bmod (p-1)p^{n-1}$$

which were found (for level two) by Adolphson [1] and Koike. On the other hand, there is the following *noncongruence* result:

(4.2) **Theorem.** *For any nonzero even $i \in \mathbf{Z}/(p-1)\mathbf{Z}$, the following are equivalent:*

(4.2.1) $\mu(M^{(i)}) = 0$, $\lambda(M^{(i)}) = 1$,

(4.2.2) $H_{k+1}(1) \equiv 0 \bmod p$ *and*

$$\frac{H_{k+p+1}(1)}{H_{k+p-1}(p)} \not\equiv \frac{H_{k+2}(1)}{H_k(p)} \bmod p^2$$

for any $k \equiv i(p-1)$.

Proof. Again, this follows from Theorem 1 of § 1, (4.0.3), and standard arguments (e.g. [15] Corollary 10.17).

Since $k > 0$ one can, in (4.1.2) and (4.2.2), replace $H_k(T)$ by $\det(1 - T_p | S_k)$.

(4.3) *Remarks.* When $p \equiv 3(4)$, $i = \dfrac{p-3}{2}$, and the class number of $\mathbf{Q}(\sqrt{-p})$ is large, one does not expect (4.2.1), (4.2.2) to hold; in fact there are at least as

many cusp forms mod p of weight $\frac{p+1}{2}$ on $SL_2(\mathbf{Z})$ fixed by U as there are independent θ-series for $\mathbf{Q}(\sqrt{-p})$ mod p (cf. [12] Proposition 10). For other values of (p, i) nothing is known.

References

1. Adolphson, A.: A p-adic theory of Hecke polynomials. Duke Math. J. **43** (No. 1), 115–145 (1976)
2. Deligne, P.: Formes modulaires et representations l-adiques. Sém. Bourb. 355. Lect. Notes Math., vol. 179, Berlin-Heidelberg-New York: Springer 1971
3. Deligne, P., Rapoport, M.: Les schemas de nodules de courbes elliptiques. In: Modular forms of one variable II. Lect. Notes Math., vol. 349. Berlin-Heidelberg-New York: Springer 1973
4. Dwork, B.: On Hecke polynomials. Invent. Math. **12**, 249–256 (1971)
5. Goss, D., Sinnot, W.: Class-groups of function fields. Duke Math. J. **52** (No. 2), 507–516 (1985)
6. Hayes, D.: Analytic class number formulas in global function fields. Invent. Math. **65**, 49–69 (1981)
7. Hayes, D.: Stickelberger elements in function fields. Compos. Math. **55**, 209–239 (1985)
8. Illusie, L.: Complexe de De Rham-Witt et cohomologie cristalline. Ann. Sci. Ec. Norm. Super., IV. Ser. **12**, 501–661 (1979)
9. Katz, N.: Travaux de Dwork. Sem. Bourb. 409. Lect. Notes Math., vol. 317. Berlin-Heidelberg-New York: Springer 1973
10. Katz, N.: P-adic properties of modular schemes and modular forms. In: Modular forms of one variable III. Lect. Notes Math., vol. 350. Berlin-Heidelberg-New York: Springer 1973
11. Katz, N., Messing, W.: Some consequences of the Riemann hypothesis for varieties over finite fields. Invent. Math. **23**, 73–77 (1974)
12. Mazur, B., Wiles, A.: Analogies between function fields and number fields. Am. J. Math. **105** (No. 2), 507–521 (1983)
13. Serre, J-P.: Quelque propriétés des varietés abéliennes en caracteristique p. Am. J. Math. **80** (No. 3), 714–739 (1958)
14. Tate, J.: Les Conjectures de Stark sur les Fonctions L d'Artin en $s=0$. Boston: Birkhäuser 1984
15. Washington, L.: Introduction to Cyclotomic Fields. Graduate texts in mathematics. Berlin-Heidelberg-New York: Springer 1982

Oblatum 26-VII-1986

Invent. math. 88, 405–422 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Local units, elliptic units, Heegner points and elliptic curves

Karl Rubin*

Department of Mathematics, The Ohio State University, Columbus, OH 43210, USA

Introduction

Suppose E is an elliptic curve defined over $\mathbf{Q}$ with complex multiplication by the ring of integers $\mathcal{O}$ of an imaginary quadratic field K. Let $\mathfrak{p}$ be a prime of K lying over the rational prime p, and form the extension $K(E_{\mathfrak{p}^\infty})/K(E_{\mathfrak{p}})$ with Galois group $\mathbf{Z}_p^2$. In this tower one can study, following Coates and Wiles [5], $U_\infty/\mathscr{C}_\infty$ – the inverse limit of the local units at $\mathfrak{p}$ modulo the inverse limit of the elliptic units attached to E, restricted to a certain eigenspace under the action of $\mathrm{Gal}(K(E_{\mathfrak{p}})/K)$. When p splits in K, Yager [23] has shown that as a module over the Iwasawa algebra $\Lambda_2 = \mathbf{Z}_p[\![\mathrm{Gal}(K(E_{\mathfrak{p}^\infty})/K(E_{\mathfrak{p}}))]\!] \cong \mathbf{Z}_p[\![S, T]\!]$,

$$U_\infty/\mathscr{C}_\infty \cong \Lambda_2/F\Lambda_2 \tag{1}$$

where F is a 2-variable $\mathfrak{p}$-adic L-function attached to E. This isomorphism also gives evidence for the "Main Conjecture" [3] which identifies F as the characteristic power series of a Selmer group.

When p does not split in K the picture is not nearly so satisfactorily developed. One difficulty is that U_∞ has Λ_2-rank 2 while $\mathscr{C}_\infty$ has Λ_2-rank 1, so $U_\infty/\mathscr{C}_\infty$ is not a torsion Λ_2-module. In this paper we begin with the observation (see §2) that the "restriction" V_∞ of U_∞ to the anticyclotomic $\mathbf{Z}_p$-extension of $K_{\mathfrak{p}}$ has 2 natural Λ-submodules V^+ and V^- of rank 1 over Λ, the Iwasawa algebra attached to the anticyclotomic $\mathbf{Z}_p$-extension of $K_{\mathfrak{p}}$. These modules depend only on p, *not* on E or on K (see §3). A further remarkable fact is that the "restriction" of $\mathscr{C}_\infty$ always lies in V^+ or V^-, depending on the sign in the functional equation of $L(E, s)$. Conjecture 2.2 (§2) states that $V_\infty \cong V^+ \oplus V^-$.

Most of this paper is devoted to proving a criterion (Theorem 8.4) under which Conjecture 2.2 is true. This criterion (which seems likely to be true for all primes) can be checked for many primes p; see §8. It is interesting that although Conjecture 2.2 is purely a statement about local fields, the methods used in this paper to attack it are very global: we make crucial use of elliptic

* Partially supported by NSF grant DMS-8501937

units and Heegner points on elliptic curves with complex multiplication. It would be interesting to have a completely local proof.

The last two sections give applications to elliptic curves. In §9 we prove the counterpart for supersingular primes of a theorem of Greenberg [7]. Our result states that if E is an elliptic curve defined over $\mathbf{Q}$ with complex multiplication, and the sign in the functional equation of $L(E, s)$ is -1 then either $E(\mathbf{Q})$ is infinite or the p-part of the Tate-Safarevic group of E is infinite for every prime $p \geqq 5$ where E has good, supersingular reduction. (With the word "supersingular" replaced by "ordinary" this is Greenberg's theorem.) In §10 we discuss the analogue of (1), showing that if $\varepsilon = \pm 1$ is the sign in the functional equation of $L(E, s)$ then $V^{\varepsilon}/\mathscr{C}_{\infty} \cong \Lambda/F\Lambda$ where F has some properties of a p-adic L-function.

Acknowledgement. I would like to thank David Rohrlich for several helpful conversations.

§ 1. Preliminaries

Fix a prime $p \geqq 5$. Denote by Φ the unique unramified quadratic extension of $\mathbf{Q}_p$, O its ring of integers, and let $\mathbf{E}$ denote a fixed Lubin-Tate formal group over O for the uniformizing parameter $-p$. Write $\Phi_n = \Phi(\mathbf{E}_{p^{n+1}})$, the extension of Φ generated by the p^{n+1}-torsion of $\mathbf{E}$, and $\Phi_\infty = \cup \Phi_n$. Then we have a natural isomorphism

$$\mathrm{Gal}(\Phi_\infty/\Phi) \to \mathrm{Aut}(\mathbf{E}_{p^\infty}) \cong O^* \cong \Delta \times O \tag{2}$$

where $\Delta = \mathrm{Gal}(\Phi_0/\Phi) \cong (O/pO)^*$ is a cyclic group of order $p^2 - 1$. Write κ for the isomorphism of $\mathrm{Gal}(\Phi_\infty/\Phi)$ into O^* given by (2), and let ω denote the restriction of κ to Δ.

From (2) we see that $\mathrm{Gal}(\Phi_\infty/\Phi_0) \cong \mathbf{Z}_p^2$. The natural action of $\mathrm{Gal}(\Phi/\mathbf{Q}_p)$ on this group gives a decomposition $\mathrm{Gal}(\Phi_\infty/\Phi_0) \cong G^+ \oplus G^-$, $G^+ \cong G^- \cong \mathbf{Z}_p$, where G^+ (resp. G^-) is the maximal subgroup on which $\mathrm{Gal}(\Phi/\mathbf{Q}_p)$ acts via the trivial (resp. nontrivial) character. Fix topological generators σ of G^+ and τ of G^-.

Define the Iwasawa algebras

$$\Lambda_2 = O[\![\mathrm{Gal}(\Phi_\infty/\Phi_0)]\!] = \varprojlim O[\mathrm{Gal}(\Phi_n/\Phi_0)] \quad \text{and} \quad \Lambda = O[\![G^-]\!].$$

As usual we will identify Λ_2 with $O[\![S, T]\!]$ and Λ with $O[\![T]\!]$ by sending σ to $1+S$ and τ to $1+T$.

For every n let U_n denote the units of Φ_n which are congruent to 1 modulo the maximal ideal and define

$$U_\infty = (\varprojlim U_n \otimes_{\mathbf{Z}_p} O)^\omega,$$

where the inverse limit is taken with respect to the norm maps, and for any $O[\Delta]$-module A, A^ω denotes the submodule of A on which Δ acts via ω. Also define

$$V_\infty = U_\infty/(\sigma - \kappa(\sigma))\, U_\infty.$$

This quotient does not depend on the choice of σ.

Proposition 1.1. *$U_\infty \cong \Lambda_2^2$, and $V_\infty \cong \Lambda^2$.*

Proof. The first assertion can be found in [22] and the second follows immediately.

§ 2. Logarithmic derivatives

Fix a generator $\{v_n\}$ for the Tate module $T_p(\mathbf{E})$. For any $u=\{u_n\}\in\varprojlim U_n$ Coleman [6] has shown the existence of a unique power series $f_u\in O[\![X]\!]^*$ with the property that $f_u(v_n)=u_n$ for every n. For every n we define the Coates-Wiles logarithmic derivatives $\delta\colon U_\infty\to O$, $\delta_n\colon U_\infty\to\Phi_n$ by

$$\delta(u)=\frac{f_u'(0)}{f_u(0)} \qquad \delta_n(u)=\frac{1}{\lambda'(v_n)}\,\frac{f_u'(v_n)}{f_u(v_n)},$$

where λ denotes the logarithm map of $\mathbf{E}$. (Note that we have extended these maps O-linearly to $\varprojlim U_n\otimes_{\mathbf{Z}_p} O$). For any character $\chi\colon \mathrm{Gal}(\Phi_\infty/\Phi)\to\mathbf{Q}_p^*$ of finite order, choose n large enough so that χ factors through $\mathrm{Gal}(\Phi_n/\Phi)$ (i.e. the conductor of χ divides p^{n+1}) and define an O-linear homomorphism $\delta_\chi\colon U_\infty\to\Phi_n$ by

$$\delta_\chi(u)=p^{-n-1}\sum_{\gamma\in\mathrm{Gal}(\Phi_n/\Phi)}\chi(\gamma)\,\delta_n(u)^\gamma.$$

The properties of f_u [6] show that the right hand side of this definition does not depend on the choice of n.

Lemma 2.1. (i) *If $u\in U_\infty$ and $\gamma\in\mathrm{Gal}(\Phi_\infty/\Phi)$ then $\delta_n(u^\gamma)=\kappa(\gamma)\,\delta_n(u)^\gamma$.*

(ii) *If χ is a character of G^- then $\delta_\chi((\sigma-\kappa(\sigma))\,U_\infty)=0$.*

(iii) *If χ is a character of $\mathrm{Gal}(\Phi_\infty/\Phi_0)$, $u\in U_\infty$ and $F(S,T)\in\Lambda_2$ then*

$$\delta_\chi(F\cdot u)=F(\kappa\chi^{-1}(\sigma)-1,\,\kappa\chi^{-1}(\tau)-1)\,\delta_\chi(u).$$

(iv) *$\delta=-p/(p+1)\,\delta_{\chi_0}$, where χ_0 denotes the trivial character.*

(v) *If χ is a character of G^-, $u\in U_\infty$ and $\delta_\chi(u)=0$ then $\delta_\rho(u)=0$ for all characters ρ of G^- with the same conductor as χ.*

Proof. Statements (i) and (iv) follow from the basic properties of f_u; see for example [6]. Then for any $\beta\in\mathrm{Gal}(\Phi_\infty/\Phi)$,

$$\delta_\chi(u^\beta)=p^{-n-1}\sum\chi(\gamma)\,\delta_n(u^\beta)^\gamma=p^{-n-1}\sum\kappa(\beta)\,\chi(\gamma)\,\delta_n(u)^{\gamma\beta}=\kappa\chi^{-1}(\beta)\,\delta_\chi(u).$$

If χ is a character of G^- then $\chi(\sigma)=1$, so applying this with $\beta=\sigma$ proves (ii). Taking $\beta=\sigma^m\tau^n$ proves (iii) for $F(S,T)=(1+S)^m(1+T)^n$, and the general case follows since these power series generate Λ_2 topologically. To prove (v), observe that since G^- is cyclic, any such ρ can be written $\rho=\chi^\beta$ with some $\beta\in G^+$ and then

$$\delta_\rho(u)=p^{-n-1}\sum\chi^\beta(\gamma\beta)\,\delta_n(u)^{\gamma\beta}=\chi^\beta(\beta)[\delta_\chi(u)]^\beta=0.$$

We call the characters of G^- anticyclotomic characters. Note that (ii) of the lemma shows that if χ is an anticyclotomic character then δ_χ induces a well-defined homomorphism of V_∞. Define

$$\Xi^+ = \{\text{anticyclotomic characters } \chi: \text{ the conductor of } \chi \text{ is an even power of } p\}$$

$$\Xi^- = \{\text{anticyclotomic characters } \chi: \text{ the conductor of } \chi \text{ is an odd power of } p\}$$

and

$$V^+ = \{v \in V_\infty : \delta_\chi(v) = 0 \text{ for every } \chi \in \Xi^-\}$$

$$V^- = \{v \in V_\infty : \delta_\chi(v) = 0 \text{ for every } \chi \in \Xi^+\}.$$

Conjecture 2.2. *$V^+ \cong V^- \cong \Lambda$ and $V_\infty = V^+ \oplus V^-$.*

§ 3. Elliptic curves and elliptic units

For the basic facts about elliptic curves with complex multiplication, formal groups and elliptic units stated in this section see [9], [4], [15].

Let $\mathscr{E}$ denote the collection of elliptic curves E with the following properties:

(i) E has complex multiplication by the ring of integers of some imaginary quadratic field K in which p does *not* split.

(ii) E is a $\mathbf{Q}$-curve in the sense of Gross [9] (i.e. E is defined over the Hilbert class field H of K and E is isogenous over H to all of its $\mathrm{Gal}(H/\mathbf{Q})$-conjugates) and further there is a grossencharacter $\varphi = \varphi_E$ of K such that the grossencharacter $\psi = \psi_E$ of H attached to E is given by $\psi = \varphi \circ \mathbf{N}_{H/K}$.

(iii) E has good reduction at all primes of H above p.

If E is any element of $\mathscr{E}$ condition (ii) above implies that for all ideals $\mathfrak{a}$ of K,

$$\varphi(\bar{\mathfrak{a}}) = \overline{\varphi(\mathfrak{a})}. \tag{3}$$

We have the L-function identities

$$L(E_{/H}, s) = L(\psi, s)^2 = \prod L(\varphi\chi, s)^2 \tag{4}$$

where the product is taken over characters χ of $\mathrm{Gal}(H/K)$, and by (3),

$$L(\bar{\varphi}, s) = L(\varphi, s),$$

so the constant $w = w_E$ in the functional equation of $L(\varphi, s)$ is ± 1. Thus we can write $\mathscr{E} = \mathscr{E}^+ \cup \mathscr{E}^-$ where $\mathscr{E}^\pm = \{E \in \mathscr{E} : w_E = \pm 1\}$.

Note that condition (ii) above is automatically satisfied if E is defined over $\mathbf{Q}$. Also, we will make use of the $\mathbf{Q}$-curve $A(q)$ defined by Gross in [9] for primes $q \equiv 3 \pmod 4$: $A(q)$ has complex multiplication by the ring of integers of $\mathbf{Q}(\sqrt{-q})$, good reduction outside q, and the sign in the functional equation of $L(\varphi_{A(q)}, s)$ is $(-1)^{(q+1)/4}$. If $-q$ is not a square mod p then $A(q) \in \mathscr{E}$.

Lemma 3.1. *If $E\in\mathscr{E}$ then $\varphi_E((p))=-p$.*

Proof. Form the Dirichlet character $\varepsilon(n)=\varphi_E((n))/n$ and let f be its conductor. If q is a rational prime not dividing f and $q=\mathfrak{q}\bar{\mathfrak{q}}$ in K, then $\varepsilon(q)=|\varphi_E(\mathfrak{q})|^2/q=1$ by (3). Since $\varepsilon(-1)=-1$, ε is not the trivial character. Thus on integers prime to f, ε must be the quadratic character attached to $K/\mathbf{Q}$, so $\varepsilon(p)=-1$.

Now suppose $E\in\mathscr{E}$ and let H, K be as above. For any n set $H_n=H(E_{p^{n+1}})$, and $H_\infty=\cup H_n$. The principal prime ideal (p) of K splits completely in H/K, so H has $[H:K]$ primes above p, all with absolute norm p^2. Fix one of these primes and denote it by $\mathfrak{p}$; then $H_\mathfrak{p}=K_p=\Phi$. The theory of complex multiplication shows that the formal group $\hat{E}$ over O giving the kernel of reduction modulo $\mathfrak{p}$ on E is a Lubin-Tate group for the uniformizing parameter $\psi(\mathfrak{p})=\varphi((p))=-p$, so $\hat{E}$ is isomorphic over O to the formal group $\mathbf{E}$ of § 1. It follows that for every n, $\mathfrak{p}$ is totally ramified in H_n/H, and the completion of H_n at the unique prime above $\mathfrak{p}$ is

$$\Phi(E_{p^{n+1}})=\Phi(\hat{E}_{p^{n+1}})=\Phi(\mathbf{E}_{p^{n+1}})=\Phi_n.$$

This gives a canonical embedding (once $\mathfrak{p}$ is chosen) of H_n into Φ_n for each n, and we can identify $\mathrm{Gal}(\Phi_\infty/\Phi)$ with $\mathrm{Gal}(H_\infty/H)$ by restriction.

Let $\Omega=\Omega_E$ be a period of E, i.e. $\Omega\in\mathbf{C}^*$ satisfies

$$\Omega^{-1}L\subset\mathcal{O} \tag{5}$$

where L is the period lattice of a Weierstrass model of E with discriminant prime to p and $\mathcal{O}$ is the ring of integers of K.

As in [4] we can construct from E elliptic units in each of the fields H_n. Viewing these as units of Φ_n we obtain elements of U_∞. Fix once and for all an embedding of $\bar{\mathbf{Q}}$ into $\bar{\mathbf{Q}}_p$, extending the map above from H into Φ, so that we can identify (among other things) complex and p-adic valued characters. With this identification, we have the following result due to Coates and Wiles.

Theorem 3.2. *Suppose $E\in\mathscr{E}$. There is an elliptic unit $\xi=\xi(E,\Omega_E)\in U_\infty$ such that*

$$\delta(\xi)=\Omega_E^{-1}L(\varphi_E,1)$$

and for all characters χ of $\mathrm{Gal}(\Phi_\infty/\Phi)$

$$\delta_\chi(\xi)=-\Omega_E^{-1}(1-\bar{\varphi}_E\chi(p)\,p^{-2})\,L(\bar{\varphi}_E\chi,1).$$

Proof. [4] when $H=K$, or [15] in general.

Corollary 3.3. *Suppose $E\in\mathscr{E}^\pm$ and let ξ denote the projection of $\xi(E,\Omega_E)$ to V_∞. Then $\xi\in V^\pm$.*

Proof. Let w denote the sign in the functional equation of $L(\varphi,s)$. Greenberg [7] showed that if χ is a character of G^- of conductor p^n, then Lemma 3.1 implies that the sign in the functional equation of $L(\bar{\varphi}\chi,s)$ is $(-1)^n w$. Thus $L(\bar{\varphi}\chi,1)=0$ if $(-1)^n\neq w$, and the theorem follows from Theorem 3.2.

Corollary 3.4. $\mathrm{rank}_\Lambda V^+\geqq 1$.

Proof. Choose a rational prime $q \equiv 7 \pmod 8$ such that $-q$ is not a square mod p. Rohrlich and Montgomery [13] have shown that the L-function of the $\mathbf{Q}$-curve $A(q)$ does not vanish at 1. Therefore if ξ is the unit of Theorem 3.2 for $A(q)$, $\delta(\xi) \neq 0$ so ξ is a nonzero element of V^+. The corollary follows since V_∞ is torsion-free (Proposition 1.1).

Remark. Using theorems of Rohrlich [14] or Waldspurger [20] one can find elliptic units $\xi \in V^-$ with $\delta_\chi(\xi) \neq 0$ for some χ, so $\operatorname{rank}_\Lambda V^- \geqq 1$ as well. We will obtain an alternate proof of this fact in § 7 (Corollary 7.5).

§ 4. Background: Λ-modules, formal groups

In this section we gather the results we will need concerning formal groups and Iwasawa theory. We let $\Lambda = O[\![\Gamma]\!]$, where Γ is a subgroup of $\operatorname{Gal}(\Phi_\infty/\Phi)$ isomorphic to $\mathbf{Z}_p$, and let γ be a topological generator of Γ.

Lemma 4.1. *Suppose $X \cong \Lambda^2$, Y is a rank-1 submodule of X and X/Y is torsion-free. Then Y is free.*

Proof. Fix a Λ-basis α, β of X, and let $Z = \Lambda(F\alpha + G\beta)$ be a maximal free submodule of Y. Then $F, G \in \Lambda$ have no common factor. Since Λ is a UFD, X/Z is easily seen to be torsion-free, so Y/Z, which is a torsion submodule of X/Z, must be 0.

Lemma 4.2. *Suppose X is a free, rank-1 Λ-module, and ϕ is a continuous homomorphism from X to O. If $x \in X$ is such that $\phi(x) \in O^*$ then x generates X over Λ.*

Proof. Writing $\mathbf{m}$ for the maximal ideal of Λ, the continuity of ϕ implies that ϕ maps $\mathbf{m}X$ into pO. Thus $x \notin \mathbf{m}X$, so by Nakayama's Lemma x generates X.

Lemma 4.3. *If X is a rank-1 Λ-module and $\operatorname{Hom}_O(X, \mathbf{E}_{p^\infty})^\Gamma \cong \Phi/O$, then X is free.*

Proof. $\operatorname{Hom}_O(X, \mathbf{E}_{p^\infty})^\Gamma \cong \operatorname{Hom}_O(X/(\gamma - \kappa(\gamma))X, \mathbf{E}_{p^\infty})$, so we conclude

$$X/(\gamma - \kappa(\gamma))X \cong O.$$

By Nakayama's Lemma it follows that X is generated by a single element, hence free.

Lemma 4.4. *Let A be an $O[\Gamma]$-module with no p-torsion and write $D_p = \mathbf{Q}_p/\mathbf{Z}_p$. Then*

$$(A \otimes D_p)^\Gamma / (A^\Gamma \otimes D_p) \cong H^1(\Gamma, A).$$

Proof. From the exact sequence

$$0 \to A \to A \otimes \mathbf{Q}_p \to A \otimes D_p \to 0$$

we get a Γ-cohomology exact sequence

$$A^\Gamma \to A^\Gamma \otimes \mathbf{Q}_p \to (A \otimes D_p)^\Gamma \to H^1(\Gamma, A) \to H^1(\Gamma, A \otimes \mathbf{Q}_p) = 0$$

which proves the Lemma.

Proposition 4.5. *Let* $\mathbf{G}$ *be a Lubin-Tate formal group of height* 2 *defined over* O, *and let* F_∞/Φ *be a ramified extension with Galois group* $\mathbf{Z}_p^d$, $d \geqq 1$. *Write* F_n *for the subfield of* F_∞ *with* $\mathrm{Gal}(F_n/F) \cong (\mathbf{Z}/p^n\mathbf{Z})^d$ *and write* $\mathbf{N}_n$ *for the norm map from* $\mathbf{G}(F_n)$ *to* $\mathbf{G}(\Phi)$. *Then*

(i) $\bigcap_n \mathbf{N}_n \mathbf{G}(F_n) = 0$.

(ii) $\mathbf{G}(F_\infty)$ *is torsion-free.*

Proof. The first statement is due to Hazewinkel [12]. The second follows from the fact that since $\mathbf{G}$ is a Lubin-Tate group, $\mathrm{Gal}(\Phi(\mathbf{G}_p)/\Phi) \cong (O/pO)^*$, which has order prime to p.

§ 5. Kummer theory and the dual situation

For all $n \leqq \infty$ define $\Theta_n = \Phi_n^{\Delta}$, the subfield of Φ_n fixed by Δ, and $\Psi_n = \Theta_n^{G^+}$. Then we have restriction isomorphisms

$$\mathrm{Gal}(\Theta_n/\Phi) \cong \mathrm{Gal}(\Phi_n/\Phi_0) \cong O/p^n O \qquad \mathrm{Gal}(\Theta_\infty/\Phi) \cong G^+ \times G^-$$
$$\mathrm{Gal}(\Psi_n/\Phi) \cong G^-/p^n G^- \cong \mathbf{Z}/p^n\mathbf{Z} \qquad \mathrm{Gal}(\Psi_\infty/\Phi) \cong G^-.$$

Then Ψ_∞ is the anticyclotomic $\mathbf{Z}_p$-extension of Φ in Θ_∞, the maximal subfield on which $\mathrm{Gal}(\Phi/\mathbf{Q}_p)$ acts via the nontrivial character.

For any algebraic extension $\mathscr{F}$ of Φ write $\mathbf{E}(\mathscr{F})$ for the maximal ideal of $\mathscr{F}$ endowed with the group law of $\mathbf{E}$. For each n the exact sequence

$$0 \to \mathbf{E}_{p^{n+1}} \to \mathbf{E}(\bar{\Phi}) \xrightarrow{p^{n+1}} \mathbf{E}(\bar{\Phi}) \to 0$$

gives rise to a $\mathrm{Gal}(\bar{\Phi}/\Theta_n)$ cohomology exact sequence

$$0 \to \mathbf{E}(\Theta_n)/p^{n+1}\mathbf{E}(\Theta_n) \to H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_n), \mathbf{E}_{p^{n+1}})$$
$$\to H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_n), \mathbf{E}(\bar{\Phi}))_{p^{n+1}} \to 0. \tag{6}$$

Lemma 5.1. (i) $\varinjlim_n H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_n), \mathbf{E}(\bar{\Phi}))_{p^{n+1}} = 0$.

(ii) $\varinjlim_n H^1(\mathrm{Gal}(\bar{\Phi}/\Psi_n), \mathbf{E}(\bar{\Phi}))_{p^{n+1}} = 0$.

Proof. If E is any elliptic curve over Φ whose formal group $\hat{E}$ is isomorphic to $\mathbf{E}$ ($E \in \mathscr{E}$, for example) then

$$\varinjlim_n H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_n), \mathbf{E}(\bar{\Phi}))_{p^{n+1}} \cong \varinjlim_n H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_n), E(\bar{\Phi}))_{p^{n+1}}.$$

By Tate duality [19] applied to E, this direct limit is dual to $\varprojlim \mathbf{E}(\Theta_n)$, the inverse limit with respect to the norm maps, and since $\mathbf{E}$ is a height 2 formal group Proposition 4.5(i) shows that this inverse limit is 0. The proof of (ii) is identical.

Recall $D_p = \mathbf{Q}_p/\mathbf{Z}_p$.

Proposition 5.2. *The exact sequences* (6) *induce an isomorphism*

$$\mathbf{E}(\Theta_\infty)\otimes D_p \cong \mathrm{Hom}_O(U_\infty, \mathbf{E}_{p^\infty}).$$

Equivalently, we have a nondegenerate, O-linear, Galois equivariant Kummer pairing

$$\langle\ ,\ \rangle : \mathbf{E}(\Theta_\infty)\otimes D_p \times U_\infty \to \mathbf{E}_{p^\infty}.$$

Proof. Taking the direct limit over n of the exact sequences (6) and using Lemma 5.1(i) shows

$$\begin{aligned}\mathbf{E}(\Theta_\infty)\otimes D_p &\cong H^1(\mathrm{Gal}(\bar{\Phi}/\Theta_\infty), \mathbf{E}_{p^\infty})\\ &\cong \mathrm{Hom}(\mathrm{Gal}(\bar{\Phi}/\Phi_\infty), \mathbf{E}_{p^\infty})^{\Delta}\\ &\cong \mathrm{Hom}_O((\mathrm{Gal}(\bar{\Phi}/\Phi_\infty)\otimes O)^\omega, \mathbf{E}_{p^\infty})\\ &\cong \mathrm{Hom}_O(U_\infty, \mathbf{E}_{p^\infty}),\end{aligned}$$

where the last isomorphism is given by local class field theory.

We will need the following formula for the pairing $\langle\ ,\ \rangle$ due to Wiles. Recall that λ is the logarithm map of $\mathbf{E}$ and $\{v_n\}$ is the generator of the Tate module $T_p(\mathbf{E})$ used to define the logarithmic derivative maps. Write Tr_n for the trace map from Φ_n to Φ.

Theorem 5.3 (Explicit Reciprocity Law). *If* $y\in\mathbf{E}(\Theta_n)$ *and* $u\in U_\infty$, *then*

$$\langle y\otimes p^{-k}, u\rangle = [p^{-n-1}\,\mathrm{Tr}_n(\delta_n(u)\,\lambda(y))]\,v_k.$$

Proof. Wiles [21].

Proposition 5.4. *The Kummer pairing of Proposition* 5.2 *induces a nondegenerate pairing*

$$\langle\ ,\ \rangle : \mathbf{E}(\Psi_\infty)\otimes D_p \times V_\infty \to \mathbf{E}_{p^\infty}.$$

Proof. From Proposition 5.2 we get

$$\begin{aligned}(\mathbf{E}(\Theta_\infty)\otimes D_p)^{G^+} &\cong (\mathrm{Hom}_O(U_\infty, \mathbf{E}_{p^\infty}))^{G^+}\\ &= \mathrm{Hom}_O(U_\infty/(\sigma-\kappa(\sigma))\,U_\infty, \mathbf{E}_{p^\infty})\\ &= \mathrm{Hom}_O(V_\infty, \mathbf{E}_{p^\infty}).\end{aligned}$$

By Lemmas 4.4 and 4.5(ii), $(\mathbf{E}(\Theta_\infty)\otimes D_p)^{G^+}/(\mathbf{E}(\Psi_\infty)\otimes D_p) \cong H^1(G^+, \mathbf{E}(\Theta_\infty))$, so to prove the Proposition we need only show that this cohomology group vanishes.

The inflation-restriction sequence of Galois cohomology identifies $H^1(G^+, \mathbf{E}(\Theta_\infty))$ with a subgroup of $H^1(\mathrm{Gal}(\bar{\Phi}/\Psi_\infty), \mathbf{E}(\bar{\Phi}))$. But

$$H^1(\mathrm{Gal}(\bar{\Phi}/\Psi_\infty), \mathbf{E}(\bar{\Phi})) \cong \varinjlim_n H^1(\mathrm{Gal}(\bar{\Phi}/\Psi_n), \mathbf{E}(\bar{\Phi}))$$

which is 0 by Lemma 5.1(ii).

For any character χ of $\mathrm{Gal}(\Psi_\infty/\Psi) = G^-$ define a map λ_χ from $\mathbf{E}(\Psi_\infty)$ to Φ_∞ by

$$\lambda_\chi(y) = p^{-n} \sum_{\gamma\in\mathrm{Gal}(\Psi_n/\Phi)} \chi^{-1}(\gamma)\,\lambda(y)^\gamma$$

where n is any integer large enough so that $y \in \mathbf{E}(\Psi_n)$ and the conductor of χ divides p^{n+1}. Now define

$$A^+ = \{y \in \mathbf{E}(\Psi_\infty) : \lambda_\chi(y) = 0 \text{ for all } \chi \in \Xi^+\}$$

and

$$A^- = \{y \in \mathbf{E}(\Psi_\infty) : \lambda_\chi(y) = 0 \text{ for all } \chi \in \Xi^-\}.$$

The next Lemma follows immediately from elementary properties of characters, and we omit the proof. In particular (iii) shows that G^- preserves A^+ and A^-.

Lemma 5.5. (i) *If $y \in \mathbf{E}(\Psi_n)$ and the conductor of χ is greater than p^{n+1} then $\lambda_\chi(y) = 0$.*

(ii) *If $y \in \mathbf{E}(\Psi_\infty)$ then $\lambda(y) = \sum \lambda_\chi(y)$, summing over all characters χ of G^-.*

(iii) *If $y \in \mathbf{E}(\Psi_\infty)$, χ is a character of G^- and $\gamma \in G^-$ then $\lambda_\chi(y^\gamma) = \chi(\gamma)\, \lambda_\chi(y)$.*

(iv) *If $m \geqq n$, $y \in \mathbf{E}(\Psi_m)$ and χ is a character of $\mathrm{Gal}(\Psi_n/\Phi)$ then $\lambda_\chi(\mathbf{N}_{m/n} y) = p^{m-n}\, \lambda_\chi(y)$.*

(v) $A^+ \cap A^- = 0$.

(vi) *For every n, $A^+ \cap \mathbf{E}(\Psi_n) + A^- \cap \mathbf{E}(\Psi_n) \supseteq p^n\, \mathbf{E}(\Psi_n)$.*

There are natural injections $A^\pm \otimes D_p \to \mathbf{E}(\Psi_\infty) \otimes D_p$, and by (vi) of the Lemma,

$$\mathbf{E}(\Psi_\infty) \otimes D_p = A^+ \otimes D_p + A^- \otimes D_p. \tag{7}$$

Proposition 5.6. *The annihilator of $A^\pm \otimes D_p$ in V_∞ is $V^\pm$.*

Proof. If $y \in \mathbf{E}(\Psi_\infty)$ and $v \in V_\infty$ then using Theorem 5.3 and Lemma 5.5 yields

$$\begin{aligned} \langle y \otimes p^{-k}, v \rangle &= [p^{-n-1}\, \mathrm{Tr}_n\, \delta_n(v)\, \lambda(y)]\, v_k \\ &= [p^{-n-1} \sum_\gamma \delta_n(v)^\gamma\, \lambda(y^\gamma)]\, v_k \\ &= [\sum_\chi p^{-n-1} \sum_\gamma \delta_n(v)^\gamma\, \chi(\gamma)\, \lambda_\chi(y)]\, v_k \\ &= [\sum_\chi \delta_\chi(v)\, \lambda_\chi(y)]\, v_k \end{aligned} \tag{8}$$

where the sums are taken over $\gamma \in \mathrm{Gal}(\Phi_n/\Phi)$ and all characters χ of G^-. Now it follows from the definitions that $V^\pm$ annihilates $A^\pm \otimes D_p$.

Now suppose that $v \in V_\infty$ and v annihilates $A^\pm \otimes D_p$, and $\chi \in \Xi^\mp$. Choose any $y \in A^\pm$ such that $\lambda_\chi(y) \neq 0$ (such a y must exist because $\mathbf{E}(\Psi_n)$ has a submodule of finite index isomorphic to $O[\mathrm{Gal}(\Psi_n/\Phi)]$). Then using (8) for every k shows that

$$0 = \sum_\rho \delta_\rho(v)\, \lambda_\rho(y^\gamma)$$

for every $\gamma \in \mathrm{Gal}(\Psi_n/\Phi)$. Thus

$$0 = \sum_\rho \sum_\gamma \chi^{-1}(\gamma)\, \delta_\rho(v)\, \lambda_\rho(y^\gamma) = p^n\, \delta_\chi(v)\, \lambda_\chi(y).$$

Since $\lambda_\chi(y) \neq 0$ this shows $\delta_\chi(v) = 0$. Therefore $v \in V^\pm$ as claimed.

Corollary 5.7. *The Kummer pairing of Proposition 5.2 induces a nondegenerate pairing*

$$\langle\ ,\ \rangle\colon A^{\pm}\otimes D_p\times V_\infty/V^{\pm}\to \mathbf{E}_{p^\infty}.$$

Corollary 5.8. $V^+\cap V^-=0$ *and* $\operatorname{rank}_\Lambda V^{\pm}\leqq 1$.

Proof. The first assertion follows from Propositions 5.4 and 5.6 and (7). For the second, since $\operatorname{rank}_{\mathcal{O}} A^{\pm}$ is infinite, Corollary 5.7 implies that $\operatorname{rank}_\Lambda V^{\pm}<\operatorname{rank}_\Lambda V_\infty=2$.

§ 6. Heegner points

For the facts we will use concerning Heegner points see [10]. For this section we fix an elliptic curve $E\in\mathscr{E}$, and let K be its field of complex multiplication, D the discriminant of K, H the Hilbert class field of K, and $\mathfrak{f}$ the conductor of the grossencharacter φ_E defined in § 3. Define $N=D\mathbf{N}_{K/\mathbf{Q}}\mathfrak{f}$. It is shown in [17] that there is a weight 2 newform f on $X_0(N)$ whose associated L-function $L(f,s)$ is equal to $L(\varphi_E,s)$. It follows that A_f, the factor of the Jacobian $J_0(N)$ associated to f by Shimura [17], is isogenous (over $\mathbf{Q}$) to the restriction of scalars of E from $F=\mathbf{Q}(j(E))$ to $\mathbf{Q}$ (see [9], § 18), so there are nontrivial maps defined over H from A_f to E, and therefore from $J_0(N)$ to E as well. Using the fixed embedding of H into Φ (see § 3) we can view any such map as being defined over Φ.

We will use Heegner points obtained in this way to obtain information about $A^{\pm}\otimes D_p$. Choose an auxiliary quadratic field $\mathscr{K}$ such that

(i) all primes dividing N split into two distinct primes in $\mathscr{K}$.
(ii) p does not split in $\mathscr{K}$. (9)

There will be infinitely many such $\mathscr{K}$ because N is prime to p. Let $\mathcal{O}$ denote the ring of integers of $\mathscr{K}$, and for every n let $\mathcal{O}_n$ denote the order of conductor p^{n+1} in $\mathcal{O}$. Also, fix an ideal $\mathfrak{n}$ of $\mathcal{O}$ such that $\mathcal{O}/\mathfrak{n}\cong\mathbf{Z}/N\mathbf{Z}$; the existence of such an ideal is assured by (9)(i).

Following the construction in [10], we define the Heegner point $x_n\in X_0(N)(\mathbf{C})$ to be the point corresponding to the moduli data $(\mathbf{C}/\mathfrak{n}\cap\mathcal{O}_n,\ \mathcal{O}_n/\mathfrak{n}\cap\mathcal{O}_n)$, and x to be the point corresponding to $(\mathbf{C}/\mathfrak{n},\ \mathcal{O}/\mathfrak{n})$. Then in fact $x\in X_0(N)(\mathscr{H})$ and $x_n\in X_0(N)(\mathscr{H}_n)$ where $\mathscr{H}$ is the Hilbert class field of $\mathscr{K}$ and $\mathscr{H}_n=\mathscr{K}(j(\mathcal{O}_n))$ is the ring class field of $\mathscr{K}$ of conductor p^{n+1}. Class field theory shows that

$$\operatorname{Gal}(\mathscr{H}_n/\mathscr{K})\cong\operatorname{Pic}(\mathcal{O}_n)\cong\operatorname{Pic}(\mathcal{O})\times\mathbf{Z}/(p+1)\,\mathbf{Z}\times\mathbf{Z}/p^n\,\mathbf{Z}$$

and $\operatorname{Gal}(\mathscr{K}/\mathbf{Q})$ acts on $\operatorname{Gal}(\mathscr{H}_n/\mathscr{K})$ via the nontrivial character. It follows that $\mathscr{H}_n=\mathscr{H}_0\,\mathscr{K}_n$, where $\mathscr{K}_n$ is the extension of $\mathscr{K}$ of degree p^n inside $\mathscr{K}_\infty$, the anticyclotomic $\mathbf{Z}_p$-extension of $\mathscr{K}$. Thus (now using the fact that p does not split in $\mathscr{K}$, and therefore splits completely in $\mathscr{H}/\mathscr{K}$) any embedding of $\mathscr{H}$ into Φ extends uniquely to an embedding of $\mathscr{H}\mathscr{K}_n$ into Ψ_n, and we can identify $\operatorname{Gal}(\Psi_n/\Phi)$ with $\operatorname{Gal}(\mathscr{H}\mathscr{K}_n/\mathscr{H})$. Let $\mathbf{N}'$ denote the norm map from $J_0(N)(\mathscr{H}_\infty)$ to $J_0(N)(\mathscr{H}\mathscr{K}_\infty)$, which we can identify with the norm map from $J_0(N)(\mathscr{H}_n)$ to $J_0(N)(\mathscr{H}\mathscr{K}_n)$ for every n.

Fixing a map θ from $J_0(N)$ to E defined over Φ as above, we define y_n to be the image of the divisor $[x_n]-[\infty]$ under the composition

$$J_0(N)(\mathscr{H}_n) \xrightarrow{\mathbf{N}'} J_0(N)(\mathscr{H}\mathscr{K}_n) \to J_0(N)(\Psi_n) \xrightarrow{\theta} E(\Psi_n)$$
$$\xrightarrow{(p+1)^2} E_1(\Psi_n) \to \hat{E}(\Psi_n) \to \mathbf{E}(\Psi_n). \tag{10}$$

Here E_1 denotes the kernel of the reduction map: $(p+1)^2 E(\Psi_n) \subset E_1(\Psi_n)$ because (4), Lemma 3.1 and the fact that Ψ_n/Φ is totally ramified shows that $E(\Psi_n)/E_1(\Psi_n)$ has order $(p+1)^2$. The fifth map is the canonical map from the kernel of reduction to the formal group, and the final map is any isomorphism from $\hat{E}$ to $\mathbf{E}$ (see § 3). Similarly we define y to be the image of $[x]-[\infty]$. The points y, y_n depend on the choices of $\mathscr{K}$, θ, and the embedding of $\mathscr{H}$ into Φ but this will not concern us. Fix some such choice for the remainder of this section.

The following proposition is crucial in using these Heegner points to study $A^{\pm} \otimes D_p$. Recall that $\mathbf{N}_{m/n}$ denotes the norm map from $\mathbf{E}(\Psi_m)$ to $\mathbf{E}(\Psi_n)$.

Proposition 6.1. (i) $\mathbf{N}_{n+1/n} y_{n+1} = -y_{n-1}$ *if* $n \geqq 1$.

(ii) $\mathbf{N}_{1/0} y_1 = -y$.

(iii) $y_0 = 0$.

(iv) $\mathbf{N}_{n+2/n} y_{n+2} = -p y_n$ *if* $n \geqq 0$.

Proof. Let T_p denote the pth Hecke operator acting on $X_0(N)$ or $J_0(N)$. By Lemma 3.1, the Euler factor at p in $L(\varphi, s) = L(f, s)$ is $(1+p^{1-2s})^{-1}$. Therefore a_p, the eigenvalue of T_p acting on f, is 0. It follows that $\theta T_p = 0$ on $J_0(N)$.

In terms of the moduli data,

$$T_p(x_n) = \sum (\mathbf{C}/\mathfrak{n} \cap L, L/\mathfrak{n} \cap L)$$

where the sum ranges over the $p+1$ lattices L of index p in $\mathcal{O}_n$. If $n \geqq 1$ one of these lattices is $\mathcal{O}_{n+1}$, one is $p\mathcal{O}_{n-1}$, and the other $p-1$ represent the nontrivial classes in $\mathrm{Pic}(\mathcal{O}_{n+1})$ which, when multiplied by $\mathcal{O}_n$, become trivial in $\mathrm{Pic}(\mathcal{O}_n)$. Thus using (4.2) of [10],

$$T_p(x_n) = x_{n-1} + \sum_{\gamma \in \mathrm{Gal}(\mathscr{H}_{n+1}/\mathscr{H}_n)} (x_{n+1})^{\gamma}. \tag{11}$$

Now applying the maps (10) to (11) we conclude

$$0 = y_{n-1} + \mathbf{N}_{n+1/n}(y_{n+1})$$

since $\theta T_p = 0$. Applying T_p to x_0 yields in exactly the same way

$$0 = y + \mathbf{N}_{1/0}(y_1).$$

The $p+1$ sublattices of $\mathcal{O}$ of index p represent the $p+1$ elements of $\mathrm{Pic}(\mathcal{O}_0)$ which become trivial in $\mathrm{Pic}(\mathcal{O})$ when multiplied by $\mathcal{O}$. Thus, just as in (11) we get

$$T_p(x) = \sum_{\gamma \in \mathrm{Gal}(\mathscr{H}_0/\mathscr{H})} (x_0)^{\gamma}.$$

Now since $y_0 \in \mathbf{E}(\Phi)$ applying the maps (10) gives

$$0 = \sum (y_0)^\gamma = (p+1)\, y_0.$$

This completes the proof of (i), (ii) and (iii), and (iv) is immediate from (i).

Corollary 6.2. *$y_n \in A^+$ if n is even; $y_n \in A^-$ if n is odd.*

Proof. Suppose $k \geqq 1$ and χ is a character of $\mathrm{Gal}(\Psi_k/\Phi)$ which is not trivial on $\mathrm{Gal}(\Psi_k/\Psi_{k-1})$, i.e. χ is a character of conductor p^{k+1}. If $n < k$ then $\lambda_\chi(y_n) = 0$. If $n \geqq k$ then $\lambda_\chi(\mathbf{N}_{n/k}\, y_n) = p^{n-k}\, \lambda_\chi(y_n)$. If $n-k$ is odd then applying (i) and (iv) of Proposition 6.1,

$$N_{n/k}\, y_n = -(-p)^{(n-k-1)/2}\, y_{k-1}$$

so $\lambda_\chi(\mathbf{N}_{n/k}\, y_n) = 0$. Similarly if χ is the trivial character and n is even,

$$p^n\, \lambda_\chi(y_n) = \lambda_\chi(\mathbf{N}_{n/0}\, y_n) = \lambda_\chi((-p)^{n/2}\, y_0) = 0.$$

This proves the Corollary.

Proposition 6.3. *For every $E \in \mathscr{E}$ and map θ as above there is an auxiliary quadratic field $\mathscr{K}$ satisfying (9) such that the corresponding Heegner point $y \in \mathbf{E}(\Phi)$ is nonzero.*

Proof. Write θ' for the map $\theta'(z) = \theta([z] - [\infty])$. Then θ' is a nonconstant map (defined over H) from $X_0(N)$ to E, and therefore finite-to-one. For any field $\mathscr{K}$ satisfying (9), $x = x(\mathscr{K}) \in X_0(N)(\mathscr{H})$ so the map $\mathbf{N}'$ of (10) on $[x] - [\infty]$ is multiplication by $p+1$. Thus the only way $y = y(\mathscr{K})$ can be zero is if $\theta'(x)$ is killed by $(p+1)^2$. But $\theta'^{-1}(E_{(p+1)^2})$ is finite, so there are at most finitely many $\mathscr{K}$ for which $y(\mathscr{K})$ can be zero. Since there are infinitely many $\mathscr{K}$ satisfying (9), this proves the Proposition.

Remark. One could also prove this proposition by appealing to deep results of Gross-Zagier [11], Rohrlich [unpublished], and Waldspurger [20]. We did not need to do this because our Heegner point $y \in E(\Phi)$ comes directly from a global point in $A_f(\mathscr{H})$ *without* taking the norm to $A_f(\mathscr{K})$.

§ 7. Cohomology computations

A reference for the Galois cohomology used below is [1].

Write $G_n = \mathrm{Gal}(\Psi_n/\Phi)$. For any $\mathcal{O}[G_n]$-module Z write $h_n(Z)$ for the Herbrand quotient

$$h_n(Z) = |\hat{H}^0(G_n, Z)|/|H^1(G_n, Z)|.$$

Standard results show that $h_n(Z_1/Z_2) = h_n(Z_1)/h_n(Z_2)$ and that $h_n(Z) = 1$ if Z is finite.

Fix E, $\mathscr{K}$, θ, and a collection $y, y_1, y_2, \ldots$ of Heegner points as in § 6. Define $A_n^- = A^- \cap \mathbf{E}(\Psi_n)$, and write $\mathbf{N}_n$ for the norm map $\mathbf{N}_{n/0}$. In this section we will use these Heegner points to study $H^1(G^-, A^-) = \varinjlim H^1(G_n, A_n^-)$.

Lemma 7.1. *Suppose* $y \neq 0$. *Then for all odd* n, $h_n(A_n^-) = p^{n-1}$.

Proof. Since $\mathbf{E}(\Psi_n)$ contains a submodule of finite index isomorphic to $O[G_n]$, A_n^- contains a submodule C of finite index,

$$C \cong O[G_1] \bigoplus_{\substack{k \text{ odd} \\ 3 \leq k \leq n}} O[G_k]/O[G_{k-1}].$$

An easy computation shows that $\hat{H}^0(G_n, O[G_k]) = O/p^{n-k}O$ and $H^1(G_n, O[G_k]) = 0$. Therefore

$$h_n(A_n^-) = p^{2(n-1)} \prod_{\substack{k \text{ odd} \\ 3 \leq k \leq n}} p^{2(n-k)}/p^{2(n-k+1)} = p^{n-1}.$$

Lemma 7.2. *For all odd* n, $|\hat{H}^0(G_n, A_n^-)| \leq p^{n-1}[\mathbf{E}(\Phi): Oy]$.

Proof. $\hat{H}^0(G_n, A_n^-) = (A_n^-)^{G_n}/\mathbf{N}_n(A_n^-)$; $(A_n^-)^{G_n} = \mathbf{E}(\Phi)$ and by Proposition 6.1 and Corollary 6.2, $\mathbf{N}_n(A_n^-) \supseteq O\,\mathbf{N}_n(y_n) = p^{(n-1)/2}Oy$.

Remark. This Lemma is actually the only use we will make of Heegner points. The only fact we need is the existence of *one* sequence of points $z_n \in A_n^-$ such that $[p^{(n-1)/2}\mathbf{E}(\Phi): O\,\mathbf{N}_n(z_n)]$ is 1 (to prove Conjecture 2.2) or is bounded independent of n (to prove Theorem 9.1).

Proposition 7.3. $|H^1(G_n, A_n^-)| \leq [\mathbf{E}(\Phi): Oy]$.

Proof. We have $|H^1(G_n, A_n^-)| = |\hat{H}^0(G_n, A_n^-)|/h_n(A_n^-)$, so the statement follows from Lemmas 7.1 and 7.2.

Proposition 7.4. $|(A^- \otimes D_p)^{G^-}/\mathbf{E}(\Phi) \otimes D_p|$ *is finite. If* $Oy = \mathbf{E}(\Phi)$ *then*

$$(A^- \otimes D_p)^{G^-} = \mathbf{E}(\Phi) \otimes D_p.$$

Proof. Since Proposition 7.3 holds for *all* choices of Heegner point y, this is immediate from Propositions 7.3 and 6.3 and Lemma 4.4.

Corollary 7.5. $\operatorname{rank}_\Lambda V^- \geq 1$.

Proof. If not, since V_∞ is free of rank 2 we would have $A^- \otimes D_p \cong \operatorname{Hom}_O(\Lambda^2, \mathbf{E}_{p^\infty})$ by Corollary 5.7. But then $(A^- \otimes D_p)^{G^-} \cong (\Phi/O)^2$, which would contradict Proposition 7.4.

§ 8. Conclusions

Proposition 8.1. $V^+ \cong V^- \cong \Lambda$ *and* $V^+ \cap V^- = 0$.

Proof. Define $W^+ = \{v \in V_\infty$: for some nonzero $F \in \Lambda$, $F \cdot v \in V^+\}$, and similarly for W^-. By Corollaries 3.4, 5.8 and 7.5, $V^+ \cap V^- = 0$ and $\operatorname{rank}_\Lambda V^+ = \operatorname{rank}_\Lambda V^- = 1$, so by Lemma 4.1 W^+ and W^- are both free of rank 1 over Λ. Now if $w_\pm$ is a generator of $W^\pm$, by Lemma 2.1(iii) we conclude $V^\pm = F_\pm W^\pm$, where $F_\pm$ is a power series with simple zeroes at $\kappa\chi^{-1}(\tau) - 1$ for all $\chi \in \Xi^\mp$ such that $\delta_\chi(w_\pm) \neq 0$ and no other zeroes. Thus $V^+ \cong V^- \cong \Lambda$ as well.

Define two sets of primes

$\mathscr{P}_+ = \{p$: for some $E \in \mathscr{E}$ and period Ω_E, $\Omega_E^{-1} L(\varphi_E, 1) \not\equiv 0 \pmod{p}\}$

$\mathscr{P}_- = \{p$: for *some* choice of $E \in \mathscr{E}$, auxiliary field $\mathscr{K}$ satisfying (9) and map $\theta: J_0(N) \to E$, the corresponding Heegner point $y \notin p\mathbf{E}(\Phi)\}$.

Proposition 8.2. *$\mathscr{P}_+$ has density 1.*

Proof. For any prime $q \equiv 7 \pmod 8$ fix a period Ω_q of the $\mathbf{Q}$-curve $A(q)$ which satisfies (5) for all p. Rohrlich and Montgomery [13] have shown that for all of these q, $L(\varphi_{A(q)}, 1) \neq 0$. If $p \geqq 5$ does not split in $\mathbf{Q}(\sqrt{-q})$ then $A(q) \in \mathscr{E}$.

Let S denote the complement of $\mathscr{P}_\infty$, and for each $q \equiv 7 \pmod 8$ write

$$S_q = \{p: p \text{ splits in } \mathbf{Q}(\sqrt{-q}) \text{ or } p \text{ divides } \Omega_q^{-1} L(\varphi_{A(q)}, 1)\}.$$

Then each S_q has density 1/2, and furthermore if $q_1, q_2, \ldots, q_n$ are distinct primes then $\bigcap_{i=1}^{n} S_{q_i}$ has density 2^{-n}. Therefore S, which is contained in $\bigcap_q S_q$, has density 0.

Remark. Unfortunately the analogue for $\mathscr{P}_-$ of Proposition 8.2 is not obviously true: the question of for how many primes p a given point $y \in E(H\mathscr{K})$ lies in $pE(\Phi)$ seems very difficult. However, given y it is very easy (with a computer) to check this condition. Using computations of Stephens (unpublished, but for similar calculations see [2]) involving curves with CM by $\mathbf{Q}(\sqrt{-1})$ and $\mathbf{Q}(\sqrt{-3})$ *I* have verified the following:

Computation. *If $5 \leqq p \leqq 1000$ and $p \not\equiv 1 \pmod{12}$ then $p \in \mathscr{P}_+ \cap \mathscr{P}_-$.*

Proposition 8.3. *If $p \in \mathscr{P}_-$ then $V_\infty / V^- \cong \Lambda$. If $p \in \mathscr{P}_+$ then V_∞ / V^+ is a torsion-free, rank-1 Λ-module.*

Proof. If $p \in \mathscr{P}_-$ then by Proposition 7.4 and Corollary 5.7 $\mathrm{Hom}_O(V_\infty / V^-, \mathbf{E}_{p^\infty})^{G^-} \cong \Phi / O$. Therefore $V_\infty / V^- \cong \Lambda$ by Proposition 8.1 and Lemma 4.3.

Now suppose $p \in \mathscr{P}_+$. Recall from the proof of Proposition 8.1 that

$$W^+ = \{v \in V_\infty: \text{for some nonzero } F \in \Lambda,\ F \cdot v \in V^+\}$$

is a free, rank 1 Λ-module.

Choose an $E \in \mathscr{E}$ such that $\Omega^{-1} L(\varphi, 1) \not\equiv 0 \pmod p$ and let $\xi = \xi(E, \Omega)$ be the elliptic unit of Theorem 3.2 corresponding to E. Since $L(\varphi, 1) \neq 0$, $w_E = 1$ so $\xi \in V^+ \subset W^+$ by Corollary 3.3. Since $\delta(\xi) = \Omega^{-1} L(\varphi, 1) \in O^*$, Lemma 4.2 shows that ξ generates W^+, so $V^+ = W^+$, and the second assertion follows.

Theorem 8.4. *Suppose $p \in \mathscr{P}_+ \cap \mathscr{P}_-$. Then Conjecture 2.2 is true:*

$$V^+ \cong V^- \cong \Lambda \quad \text{and} \quad V_\infty = V^+ \oplus V^-.$$

Proof. By Propositions 8.1 and 8.3, V_∞ / V^-, V^+ and V^- are free of rank 1. Since $p \in \mathscr{P}_+$, Theorem 3.2 and Corollary 3.3 show that there is a unit $\xi \in V^+$ such

that $\delta(\xi)\in O^*$. Since δ factors through V_∞/V^-, Lemma 4.2 shows that ξ generates V_∞/V^-, so $V_\infty = V^+ \oplus V^-$.

The next theorem is needed for the strengthening of Greenberg's theorem in §9. Recall $W^- = \{v\in V_\infty$: for some nonzero $F\in\Lambda$, $F\cdot v\in V^-\}$.

Theorem 8.5. *For all* p, $\delta|_{W^-}=0$.

Proof. This can be deduced from Proposition 7.4 and Corollary 5.7 using standard methods of Iwasawa theory, but we can also argue directly as follows. By Proposition 7.4 and Corollary 5.7, $\mathrm{Hom}_O(V_\infty/V^-, \mathbf{E}_{p^\infty})^{G^-}$ has a unique divisible subgroup $D\cong\Phi/O$. By Lemma 2.1(i) we can identify this subgroup as

$$D=\{v\to\delta(v)\,z\colon z\in\mathbf{E}_{p^\infty}\}.$$

But since V_∞/W^- has rank 1, $\mathrm{Hom}_O(V_\infty/W^-, \mathbf{E}_{p^\infty})^{G^-}$ has a nontrivial divisible subgroup. Thus D must lie in the subgroup $\mathrm{Hom}_O(V_\infty/W^-, \mathbf{E}_{p^\infty})^{G^-}$ of $\mathrm{Hom}_O(V_\infty, \mathbf{E}_{p^\infty})^{G^-}$, i.e. δ is 0 on W^-.

§9. Application: Greenberg's theorem

With the word "supersingular" replaced by "ordinary", the following theorem was proved by Greenberg in [7]. In this section we use Theorem 8.5 to prove the supersingular version. It is worth noting that the methods of this paper do *not* yield a proof for ordinary primes.

Theorem 9.1. *Let E be an elliptic curve defined over $\mathbf{Q}$ with complex multiplication, and suppose that the sign in the functional equation of $L(E,s)$ is -1. Then either $E(\mathbf{Q})$ is infinite or the p-part of the Tate-Safarevic group of E over $\mathbf{Q}$ is infinite for every prime $p\geqq 5$ where E has good, supersingular reduction.*

Proof. Fix a prime p where E has supersingular reduction (i.e. p does not split in the field K of complex multiplication). If F is any number field, Galois cohomology gives an exact sequence

$$\begin{aligned} 0\to E(F)\otimes D_p &\to H^1(G(\bar{F}/F), E_{p^\infty}) \\ &\to H^1(G(\bar{F}/F), E(\bar{F}))_{p^\infty}\to 0 \end{aligned} \tag{12}$$

which is a global version of (6). The Tate-Safarevic group $Ш(F)$ of E over F is the kernel of the localization maps

$$H^1(G(\bar{F}/F), E(\bar{F}))\to\bigoplus_{\mathfrak{q}} H^1(G(\bar{F}_{\mathfrak{q}}/F_{\mathfrak{q}}), E(\bar{F}_{\mathfrak{q}}))$$

and the Selmer group $S_{p^\infty}(F)$ is the inverse image of $Ш(F)_{p^\infty}$ in $H^1(G(\bar{F}/F), E_{p^\infty})$ under (12). Thus

$$0\to E(F)\otimes D_p\to S_{p^\infty}(F)\to Ш(F)_{p^\infty}\to 0 \tag{13}$$

and Galois cohomology shows that $S_{p^\infty}(\mathbf{Q})\cong S_{p^\infty}(K)^{\mathrm{Gal}(K/\mathbf{Q})}$.

We first show that $S_{p^\infty}(K)$ is infinite. Set $K_\infty = K(E_{p^\infty})$. It is shown in [16], §1 that

$$S_{p^\infty}(K)=\ker(\mathrm{Hom}_O(X_\infty, E_{p^\infty})^{\mathrm{Gal}(K_\infty/K)}\to H^1(G(\bar{\Phi}/\Phi), E(\bar{\Phi}))) \tag{14}$$

where X_∞ is the Galois group of the maximal abelian p-extension of K_∞ which is unramified outside the prime above p. Consider the diagram

$$\begin{array}{ccccc} & & \mathrm{Hom}_O(X_\infty, E_{p^\infty})^{\mathrm{Gal}(K_\infty/K)} & & \\ & & \downarrow f & \searrow h & \\ 0 \to \mathbf{E}(\Phi)\otimes D_p & \xrightarrow{g} & \mathrm{Hom}_O(U_\infty, \mathbf{E}_{p^\infty})^{\mathrm{Gal}(\Phi_\infty/\Phi)} & \to & H^1(G(\bar{\Phi}/\Phi), E(\bar{\Phi})). \end{array}$$

The bottom row is obtained from (6) using the isomorphism

$$H^1(\mathrm{Gal}(\bar{\Phi}/\Phi), \mathbf{E}_{p^\infty}) \cong \mathrm{Hom}_O(U_\infty, \mathbf{E}_{p^\infty})^{\mathrm{Gal}(\Phi_\infty/\Phi)}$$

which is deduced immediately from (6) and (7) of [16] by class field theory. Thus the bottom row is exact. By (14), $S_{p^\infty}(K) = \ker(h)$, so $f(S_{p^\infty}(K)) = \mathrm{image}(f) \cap \mathrm{image}(g)$. We will show that $\mathrm{image}(g)$ is contained in $\mathrm{image}(f)$, so that $S_{p^\infty}(K)$ must be infinite.

For each n let $\mathscr{U}_n$ denote the global units of the field $K(E_{p^{n+1}})$ which are congruent to 1 modulo the unique prime above p, and define

$$\mathscr{U}_\infty = (\varprojlim \overline{\mathscr{U}_n} \otimes O)^\omega \subset U_\infty$$

where $\overline{\mathscr{U}_n}$ denotes the closure of $\mathscr{U}_n$ in Ψ_n^*. Let $\xi \in \mathscr{U}_\infty$ be an elliptic unit attached to E as in Theorem 3.2. As a Λ_2-module, $\mathscr{U}_\infty$ has rank 1 (see [8]), so $\mathscr{U}_\infty/\Lambda_2 \xi$ is Λ_2-torsion. Projecting into V_∞ this implies (since ξ maps into V^-) $\mathscr{U}_\infty/\mathscr{U}_\infty \cap (\sigma - \kappa(\sigma))\, U_\infty \subset W^-$, so by Theorem 8.5 $\delta|_{\mathscr{U}_\infty} = 0$. It then follows from the Explicit Reciprocity Law (Theorem 5.3) and Lemma 2.1(iv) that the image of the map g above is contained in

$$\mathrm{Hom}_O(U_\infty/\mathscr{U}_\infty, \mathbf{E}_{p^\infty})^{\mathrm{Gal}(\Phi_\infty/\Phi)},$$

which by class field theory is precisely the image of f.

Since $S_{p^\infty}(K)$ is infinite it contains a divisible submodule $D \cong \Phi/O$. Therefore $S_{p^\infty}(K)^{\mathrm{Gal}(K/\mathbf{Q})}$ contains a copy of $\mathbf{Q}_p/\mathbf{Z}_p$, so $S_{p^\infty}(\mathbf{Q})$ is infinite as well. Now the theorem follows from (13).

§ 10. Application: p-adic L-functions

For this section fix $\varepsilon = \pm 1$, and fix a generator v of the free, rank-1 Λ-module V^ε. Suppose $E \in \mathscr{E}^\varepsilon$ and let $\xi \in V^\varepsilon$ be an elliptic unit as in Theorem 3.2. Write $\mathscr{C}_\infty = \Lambda\xi$ for the Λ-module of elliptic units attached to E. There is a unique power series $F_E(T) \in \Lambda$ such that $F_E \cdot v = \xi$, and then $V^\varepsilon/\mathscr{C}_\infty \cong \Lambda/F_E\Lambda$. From Theorem 3.2 and Lemma 2.1(iii) we see that for any anticyclotomic character $\chi \in \Xi^\varepsilon$

$$\delta_\chi(v)\, F_E(\kappa\chi^{-1}(\tau) - 1) = \Omega^{-1}\, L(\bar{\varphi}\chi, 1).$$

In other words, the power series F_E interpolates p-adically the values $\Omega^{-1}\, L(\bar{\varphi}\chi, 1)$, with the normalization factor $\delta_\chi(v)$. Unfortunately we do not

know what the values $\delta_\chi(v)$ are, or even what their p-adic valuations are. Ideally one would like to show that $\delta_\chi(v)$ is some sort of p-adic period. However at least these normalization constants are *independent of* $E \in \mathscr{E}^\varepsilon$, and one can view F_E as a sort of p-adic L-function attached to E in the anti-cyclotomic $\mathbf{Z}_p$-extension. By a theorem of Rohrlich [14] we can conclude that if E is defined over $\mathbf{Q}$, then $F_E \neq 0$. We also have the following:

Lemma 10.1. *If Conjecture* 2.2 *is true for p and v is a generator of* V^ε *then for all* $\chi \in \Xi^\varepsilon$, $\delta_\chi(v) \neq 0$.

Proof. Fix such a χ. If $\delta_\chi(v)=0$ then δ_χ must be 0 on all of $V^+ \oplus V^- = V_\infty$, and by Lemma 2.1(v) the same must be true for all characters χ' with the same conductor as χ. Choose a nonzero $y \in \mathbf{E}(\Psi_\infty)$ such that $\lambda_\rho(y)=0$ if the conductor of ρ is different from the conductor of χ. (If the conductor of χ is p^{n+1} this just amounts to choosing $y \in \mathbf{E}(\Psi_n)$ such that $\mathbf{N}_{n/n-1}(y)=0$.) Then by (8) $\langle y \otimes p^{-k}, w \rangle = 0$ for every k and every $w \in V_\infty$, which contradicts Proposition 5.4.

Thus if Conjecture 2.2 is true for p (for example if $p \in \mathscr{P}_+ \cap \mathscr{P}_-$) then we can write

$$F_E(\kappa\chi^{-1}(\tau)-1) = \delta_\chi(v)^{-1}(\Omega^{-1} L(\bar{\varphi}\chi, 1)) \quad \text{for all } \chi \in \Xi^\varepsilon.$$

It would be interesting to understand the significance of the special values $F_E(\kappa\chi^{-1}(\tau)-1)$ for $\chi \in \Xi^{-\varepsilon}$. One guess would be that they contain information about the Λ-module of Heegner points lying inside the local points. More specifically (recall $A_n^\pm = A^\pm \cap \mathbf{E}(\Psi_n)$), we can define the Λ-modules

$$A_\infty^\varepsilon = \varprojlim A_n^\varepsilon,$$

where the inverse limit is taken over n with $(-1)^n = \varepsilon$, with respect to the maps $x \to -\mathbf{N}_{n/n-1}(x)$ from A_n^ε to A_{n-2}^ε. Using the techniques of this paper one can show that A_∞^ε is a free rank-1 Λ-module. Further, by Proposition 6.1 and Corollary 6.2,

$$\mathbf{y} = \{y_n\}_{(-1)^n=\varepsilon} \in A_\infty^\varepsilon$$

for any choice of Heegner points $y_1, y_2, \ldots$ as in §6. Then the question raised above is related to the following:

Question. What is the relation between $V^\varepsilon/\mathscr{C}_\infty$ and $A_\infty^\varepsilon/\Lambda\mathbf{y}$?

This is the analogue of a question asked by Mazur when the curve E has ordinary reduction. For some evidence that there is a relation, see [10].

References

1. Atiyah, M., Wall, C.: Cohomology of groups. In: Algebraic number theory. Cassels, J.W.S., Fröhlich, A. (eds.), pp. 94–115. London: Academic Press 1967
2. Birch, B., Stephens, N.: Computation of Heegner points. In: Modular forms. Rankin, R.A. (ed.), pp. 13–41. Chichester: Ellis Horwood Ltd (1984)
3. Coates, J.: Elliptic curves and Iwasawa theory. In: Modular forms. Rankin, R.A. (ed.), pp. 51–73. Chichester: Ellis Horwood Ltd (1984)
4. Coates, J., Wiles, A.: On the conjecture of Birch and Swinnerton-Dyer. Invent. Math. **39**, 223–251 (1977)

5. Coates, J., Wiles, A.: On p-adic L-functions and elliptic units. J. Austr. Math. Soc. **26**, 1–25 (1978)
6. Coleman, R.: Division values in local fields. Invent. Math. **53**, 91–116 (1979)
7. Greenberg, R.: On the Birch and Swinnerton-Dyer conjecture. Invent. Math. **72**, 241–265 (1983)
8. Greenberg, R.: On the structure of certain Galois groups. Invent. Math. **47**, 85–99 (1978)
9. Gross, B.: Arithmetic on elliptic curves with complex multiplication. Lect. Notes Math. vol. **776**. Berlin-Heidelberg-New York: Springer 1980
10. Gross, B.: Heegner points on $X_0(N)$. In: Modular forms. Rankin, R.A. (ed.), pp. 87–105. Chichester: Ellis Horwood Ltd (1984)
11. Gross, B., Zagier, D.: Points de Heegner et dérivées de fonctions L. C.R. Acad. Sci. Paris **297**, 85–87 (1983)
12. Hazewinkel, M.: On norm maps for one-dimensional formal groups III. Duke Math. J. **44**, 305–314 (1977)
13. Montgomery, H., Rohrlich, D.: On the L-functions of canonical Hecke characters of imaginary quadratic fields II. Duke Math. J. **49**, 937–942 (1982)
14. Rohrlich, D.: On L-functions of elliptic curves and anticyclotomic towers. Invent. Math. **75**, 383–408 (1984)
15. Rubin, K.: Elliptic curves with complex multiplication and the conjecture of Birch and Swinnerton-Dyer. Invent. Math. **64**, 455–470 (1981)
16. Rubin, K.: Elliptic curves and $\mathbf{Z}_p$-extensions. Compos. Math. **56**, 237–250 (1985)
17. Shimura, G.: On elliptic curves with complex multiplication as factors of the jacobians of modular function fields. Nagoya Math. J. **43**, 199–208 (1971)
18. Shimura, G.: Introduction to the arithmetic theory of automorphic forms. Princeton: Princeton Univ. Press 1971
19. Tate, J.: WC-groups over p-adic fields. Séminaire Bourbaki (1957/1958) exposé 156
20. Waldspurger, J-L.: Sur les valeurs de certaines fonctions L automorphes en leur centre de symétrie. Compos. Math. **54**, 173–242 (1985)
21. Wiles, A.: Higher explicit reciprocity laws. Ann. Math. **107**, 235–254 (1978)
22. Wintenberger, J-P.: Structure galoisienne de limites projectives d'unités locales. Compos. Math. **42**, 89–103 (1981)
23. Yager, R.: On two variable p-adic L-functions. Ann. Math. **115**, 411–449 (1982)

Oblatum 18-IV-1986 & 6-X-1986

Invent. math. 88, 423-425 (1987)

Inventiones mathematicae

© Springer-Verlag 1987

A correction to a paper of H. King

Les Reid

Mathematics Department, University of Utah, Salt Lake City, UT 84112, USA

In his paper [3], H. King shows that a conjecture of Thom's is false, if one can find a smooth compact connected n-dimensional submanifold N of S^n such that $NK_0(\mathbb{Z}[\pi_1 N]) \neq 0$ and $\tilde{K}_0(\mathbb{Z}[\pi_1(\partial N \times S^1)]) = 0$. His candidate had $\pi_1 N = \mathbb{Z}_2 \oplus \mathbb{Z}_2$, $\partial N = \partial_1 N \cup \partial_2 N$, $\partial_i N$ connected and $\pi_1(\partial_i N) = \mathbb{Z}_2$, $i = 1, 2$. Such a manifold doesn't exist, but one having the desired properties can be constructed.

We first show that a manifold such as King's cannot exist. Assume that it does. Note that we must have $n \geqq 4$ since $\pi_1(\partial_i N) = \mathbb{Z}_2$ and $\mathbb{R}P^3$ does not embed in S^3. Let i denote the inclusion of ∂N into N. An easy calculation with the Mayer-Vietoris sequence for cohomology with $\mathbb{Z}_2$ coefficients for the pair $N, S^n \setminus N$ shows that i^*: $H^1(N, \mathbb{Z}_2) \to H^1(\partial N, \mathbb{Z}_2)$ is an isomorphism and i^*: $H^2(N, \mathbb{Z}_2) \to H^2(\partial N, \mathbb{Z}_2)$ is an injection. Let $a, b \in H^1(N, \mathbb{Z}_2)$ be the elements mapped to the generators of $H^1(\partial_1 N, \mathbb{Z}_2)$ and $H^1(\partial_2 N, \mathbb{Z}_2)$ respectively by i^*. We then have $i^*(a \cup b) = i^*(a) \cup i^*(b) = 0$, which implies $a \cup b = 0$ by the injectivity of i^*. On the other hand since $\pi_1(N)$ is abelian, there is a map f: $T^2 \to N$ such that f^*: $H^1(N, \mathbb{Z}_2) \to H^1(T^2, \mathbb{Z}_2)$ sends a and b to α and β, the standard generators for $H^1(T^2, \mathbb{Z}_2)$. Thus $f^*(a \cup b) = \alpha \cup \beta \neq 0$, so $a \cup b \neq 0$. This is a contradiction. The argument above is essentially due to John Morgan.

We now construct a manifold having the required properties. According to Fox [2, p. 319] there is a smooth embedding f: $S^2 \to S^4$ such that $\pi_1(S^4 \setminus f(S^2)) = G$, where G is the metacyclic group $\langle a, b: a^{2n+1} = 1, aba^{-1} = b^{-1} \rangle$. Let $2n+1 = p^2$, p an odd prime. Let N be the complement of an open tubular neighborhood of $f(S^2)$. We have $\tilde{K}_0(\mathbb{Z}[\pi_1(\partial N \times S^1)]) = \tilde{K}_0(\mathbb{Z}[\mathbb{Z} \oplus \mathbb{Z}]) = 0$ by regularity. It remains to show that $NK_0(\mathbb{Z}[\pi_1 N]) \neq 0$.

First, we note that $\mathbb{Z}[\pi_1 N] = \mathbb{Z}[\mathbb{Z}_{p^2}]_\alpha[T]$ (the twisted group ring [1, p. 193]) where $\alpha(g) = g^{-1}$, g being a generator of $\mathbb{Z}_{p^2}$. We have the following cartesian squares of rings.

$$\begin{array}{ccc} \mathbb{Z}[\mathbb{Z}_p]_\alpha[T] & \xrightarrow{\bmod \phi_p(g)} & \mathbb{Z}[\zeta_p]_\alpha[T] \\ \Big\downarrow{\scriptstyle \bmod g-1} & & \Big\downarrow{\scriptstyle \bmod \zeta_p - 1} \\ \mathbb{Z}_\alpha[T] & \xrightarrow[\bmod p]{} & (\mathbb{F}_p)_\alpha[T] \end{array}$$

$$\begin{array}{ccc} \mathbb{Z}[\pi_1 N] = \mathbb{Z}[\mathbb{Z}_{p^2}]_\alpha[T] & \xrightarrow{\bmod \phi_{p^2}(g)} & \mathbb{Z}[\zeta_{p^2}]_\alpha[T] \\ \Big\downarrow \bmod g^p - 1 & & \Big\downarrow \bmod \zeta_p^p - 1 \\ \mathbb{Z}[\mathbb{Z}_p]_\alpha[T] & \xrightarrow[\bmod p]{} & \mathbb{F}_p[\mathbb{Z}_p]_\alpha[T] \end{array}$$

Now $\mathbb{Z}$, $\mathbb{F}_p$, $\mathbb{Z}[\zeta_p]$ and $\mathbb{Z}[\zeta_{p^2}]$ are regular so $\mathbb{Z}_\alpha[T]$, $(\mathbb{F}_p)_\alpha[T]$, $\mathbb{Z}[\zeta_p]_\alpha[T]$ and $\mathbb{Z}[\zeta_{p^2}]_\alpha[T]$ are also regular [1, p. 213, Thm. 25]. Therefore

$$NK_i(\mathbb{Z}_\alpha[T]) = NK_i((\mathbb{F}_p)_\alpha[T]) = NK_i(\mathbb{Z}[\zeta_p]_\alpha[T]) = NK_i(\mathbb{Z}[\zeta_{p^2}]_\alpha[T]) = 0.$$

Applying Milnor's Mayer-Vietoris sequence to the first square, we conclude that $NK_i(\mathbb{Z}[\mathbb{Z}_p]_\alpha[T]) = 0$, $i = 0, 1$. Applying it to the second square, we conclude that $NK_0(\mathbb{Z}[\pi_1 N]) \cong NK_1(\mathbb{F}_p[\mathbb{Z}_p]_\alpha[T])$.

We must show that $NK_1(\mathbb{F}_p[\mathbb{Z}_p]_\alpha[T]) \neq 0$. To do this we will use Farrell and Hsiang's exact sequence for K_1 of twisted Laurent extensions [1, p. 204, Thm. 14]. The functoriality of that sequence allows us to replace the K_1 terms in it by NK_1 terms, namely

$$0 \to I(NK_1(\mathbb{F}_p[\mathbb{Z}_p]), \alpha_*) \to NK_1(\mathbb{F}_p[\mathbb{Z}_p]) \to NK_1(\mathbb{F}_p[\mathbb{Z}_p]_\alpha[T])$$

is exact, where $I(H, \beta) = \mathrm{im}(d)$, $d\colon H \to H$, $d(h) = h\beta(h)^{-1}$. Now

$$NK_1(\mathbb{F}_p[\mathbb{Z}_p]) = NK_1(\mathbb{F}_p[g]/(g^p - 1)) \cong NK_1(\mathbb{F}_p[x]/x^p) = \{1 + xtf \mid f \in \mathbb{F}_p[\mathbb{Z}_p][t]\},$$

where $x = g - 1$. We claim that $1 + xt \in NK_1(\mathbb{F}_p[\mathbb{Z}_p]) \setminus I(NK_1(\mathbb{F}_p[\mathbb{Z}_p]), \alpha_*)$, i.e. $1 + xt$ represents a non-zero element of $NK_1(\mathbb{F}_p[\mathbb{Z}_p]_\alpha[T])$. Assume not; then there are $y, z \in NK_1(\mathbb{F}_p[\mathbb{Z}_p])$ such that

$$yz = 1 \qquad y\bar{z} = 1 + xt$$

where a bar denotes the action of α_*. Writing

$$y = 1 + a_1 t + \ldots \qquad z = 1 + b_1 t + \ldots \qquad a_1, b_i \in x\mathbb{F}_p[\mathbb{Z}_p]$$

and comparing linear coefficients we obtain

$$a_1 + b_1 = 0 \qquad a_1 + \bar{b}_1 = x;$$

but this implies $0 = x + \bar{x} = g - 1 + g^{-1} - 1$. Since g, g^{-1} and 1 are linearly independent over $\mathbb{F}_p$, this is impossible.

The above gives a counterexample to Thom's conjecture in dimension 7. For higher dimensions take $N' = N \times T^k$. We have $\tilde{K}_0(\mathbb{Z}[\pi_1(\partial N' \times S^1)]) = \tilde{K}_0(\mathbb{Z}[\mathbb{Z}^{k+2}]) = 0$ by regularity and $NK_0(\mathbb{Z}[\pi_1 N'])$ contains $NK_0(\mathbb{Z}[\pi_1 N])$ as a direct summand, so it is also non-zero.

Bibliography

1. Farrell, F.T., Hsiang, W.C.: A formula for $K_1 R_\alpha[T]$. Applications of categorical algebra. Proceedings of Symposia in Pure Mathematics, Vol. XVII, AMS, 1970
2. Fox, R.H.: A quick trip through Knot theory. Topology of 3-Manifolds and related topics, Prentice-Hall, Inc., 1962
3. King, H.: Real analytic germs and their varieties at isolated singular points. Invent. Math. **37**, 193–199 (1976)

Oblatum 2-VI-1986

Invent. math. 88, 427–438 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

On complex vector bundles on projective threefolds

Constantin Bănică and Mihai Putinar

Department of Mathematics-INCREST, B-dul Păcii 220, Bucharest, Romania

Introduction

One of the first questions which arises in the study of (complex) vector bundles on complex projective manifolds is to determine those topological bundles which admit algebraic, or equivalently, holomorphic structures.

In the case of curves, every topological vector bundle has an algebraic structure. Schwarzenberger [15] has shown for surfaces that a topological vector bundle admits an algebraic structure if and only if its first Chern class is algebraic.

From the papers of Atiyah-Rees [3], Horrocks [10] and Vogelaar [18] it follows that every topological vector bundle on the projective space $\mathbb{P}^3$ has an algebraic structure. The same conclusion holds on a homogeneous rational threefold, see [5] and [6]. What about the general case of (complex, smooth) projective threefolds?

One can expect that in this case a topological vector bundle admits an algebraic structure if and only if its Chern classes are algebraic. The "only if" part is clear, and for the proof of the "if" part only rank 2 and 3 vector bundles should taken into account.

First, the topological classification is needed. This is discussed in the second section of the present paper for any compact connected three-dimensional complex manifold. The rank 3 bundles are parametrized, via Chern classes, by the cohomology classes (c_1, c_2, c_3) subject to an integrality condition; the rank 2 vector bundles are determined, roughly speaking, by the Chern classes (c_1, c_2) and by two other "finite" invariants $\beta(c_1, c_2)$ and $\alpha(c_1, c_2, \beta(c_{1,c2}))$, see Theorem 1 for the precise statement.

In the third section we answer into affirmative the above question in stable rank, i.e. $\geqq 3$, by proving the following:

Theorem 2. *A topological vector bundle of rank* 3 *on a projective* 3*-fold admits an algebraic structure if and only if its Chern classes are algebraic.*

The remaining case of rank 2 vector bundles is treated in the fourth section, where we prove a weaker result:

Theorem 3. *For every rank 2 topological vector bundle on a projective 3-fold X with algebraic Chern classes (c_1, c_2) there exists an algebraic vector bundle with the same Chern classes.*

Moreover, if $c_1 = c_1(K_X)$, then this algebraic bundle can be chosen to be K_X-symplectic, with a prescribed arbitrary value in $\mathbb{Z}_2$ for the Atiyah-Rees invariant.

Recall that if E is a holomorphic K_X-symplectic bundle, then the (analytical) Atiyah-Rees invariant [3] is

$$\alpha(E) = h^0(E) + h^2(E) \pmod 2.$$

In Theorems 2 and 3 above one constructs algebraic bundles by the method of extensions [16], in the case of rank 3 bundles by a variant given in [5]. The fact that 1-dimensional cycles are smoothable and that they are very movable, due to the residual intersection device, plays an important role in these constructions. Non-reduced curves (of smooth support) are also intensively used. Accordingly, a key point in the proof of Theorem 3 is played by a method of "tripling" curves, due to Forster and the first author [4]: it was very interesting to us to see how the triple structures are connected with the integrality condition $(c_1 + c_1(X))c_2 \equiv 0 \pmod 2$ and how they relate the Atiyah-Rees invariant with θ-characteristics.

In a final section we combine the previous results in order to prove the next:

Theorem 4. *Every topological vector bundle an a projective rational threefold admits algebraic structures.*

In this case an essential step in the proof is the fact that the analytical Atiyah-Rees invariant for K_X-symplectic bundles corresponds to the topological ones [3]. We were not able to prove the result for rank 2 vector bundles in its full generality, mainly due to the fact that we failed to understand the intricate question whether the second topological invariant β has an analytic counterpart.

The first section contains preliminaries concerning extensions, double and triple structures, cycles of dimension 1 and a statement derived from the Atiyah-Hirzebruch spectral sequence in K-theory.

We adopt the following conventions. A 3-fold means in the sequel a complex, connected, smooth, projective manifold of dimension 3. We abbreviate l.c.i. for locally complete intersection. We simply say bundle for a topological complex vector bundle, and if the context is clear, for an algebraic vector bundle. Sometimes we denote by the same letter an integer cohomology class as well as its image by the reduction $\mathbb{Z} \to \mathbb{Z}_2$.

1. Preliminaries

1.1. Let us first recall the construction of rank 2 bundles by extensions ([16], [14] Ch. I. § 5). Denote by X a 3-fold, Y a l.c.i. curve in X and by L_1, L_2 line bundles on X. Assume that $\det(N_Y) \cong L_1^{-1} \otimes L_2$ and that $H^2(X, L_1 \otimes L_2^{-1}) = 0$.

Then there exists a bundle E of rank 2, given by an extension

$$0 \to L_1 \to E \to \mathcal{I}_Y L_2 \to 0.$$

One has $\det(E) \cong L_1 \otimes L_2$ and the next formulas for the Chern classes are true:

$$c_1(E) = c_1(L_1) + c_1(L_2),$$
$$c_2(E) = c_1(L_1)c_1(L_2) + [Y].$$

We shall consider for our purposes L_1 to be sufficiently positive and L_2 to be sufficiently negative. Then the condition $H^2(X, L_1 \otimes L_2^{-1}) = 0$ is satisfied, and, we notice that in this case the equality

$$h^0(E) + h^2(E) = h^0(L_1) + h^1(L_2 | Y)$$

holds, a relation which will be needed in the computation of the Atiyah-Rees invariant. Due to the condition on det, the difficulty in performing this construction lies in finding the right curves: we shall consider convenient disjoint unions of double and triple structures.

In order to obtain bundles of rank 3 we need the following variant of the above construction [5]. Let $Y \subset X$ be as above, let F be a rank 2 bundle on X and L be a line bundle on X. Assume that $H^2(X, F) = 0$ and that $\det N_Y \otimes F$ (regarded as a 2 bundle on Y) has a nowhere vanishing global section. Then there exists a bundle E of rank 3 on X, given by an extension

$$0 \to F \to E \to \mathcal{I}_Y \to 0.$$

The Chern classes of E are given by the formulas:

$$c_1(E) = c_1(F),$$
$$c_2(E) = c_2(F) + [Y],$$
$$c_3(E) = (c_1(F) + c_1(X))[Y] - 2\chi(\mathcal{O}_Y).$$

1.2. Ferrand's method of "doubling" a curve is as follows, [7]. Again X is a 3-fold and $Y \subset X$ is a l.c.i. curve. Assume that one has an epimorphism $N_Y^* = \mathcal{I}_Y / \mathcal{I}_Y^2 \xrightarrow{\varepsilon} L \to 0$, where L is a line bundle on Y. Consider the closed subscheme Z of X given by the ideal

$$\mathcal{I}_Z = \ker(\mathcal{I}_Y \to \mathcal{I}_Y / \mathcal{I}_Y^2 \xrightarrow{\varepsilon} L).$$

Then Z is a l.c.i. Z has the same support as Y, and

$$(\det N_Z) | Y = (\det N_Y) \otimes L^{-1}.$$

At the level of cohomology classes one has $[Z] = 2[Y]$. The exact sequence

$$0 \to L \to \mathcal{O}_Z \to \mathcal{O}_Y \to 0$$

gives rise to the exact sequence

$$H^1(Y, L) \to \operatorname{Pic} Z \to \operatorname{Pic} Y \to 1.$$

Notice also that, if L is sufficiently positive, then there exist such epimorphisms ε. In that case, $\operatorname{Pic} Z \cong \operatorname{Pic} Y$.

Next we briefly recall the "tripling" method [4]. One starts with a doubled curve as above. The inclusions of ideals yield the exact sequence

$$0 \to \mathscr{I}_Y^2/\mathscr{I}_Y\mathscr{I}_Z \to \mathscr{I}_Z/\mathscr{I}_Y\mathscr{I}_Z \to \mathscr{I}_Y/\mathscr{I}_Y^2 \to \mathscr{I}_Y/\mathscr{I}_Z \to 0,$$

and the isomorphisms

$$\mathscr{I}_Y/\mathscr{I}_Z \cong L, \quad \mathscr{I}_Y^2/\mathscr{I}_Y\mathscr{I}_Z \cong L^2, \quad \mathscr{I}_Z/\mathscr{I}_Y\mathscr{I}_Z \cong N_Z^* \mid Y.$$

Using these, one finds an exact sequence of bundles on Y:

$$0 \to L^2 \xrightarrow{i} N_Z^* \mid Y \to (\det N_Y^*) \otimes L^{-1} \to 0.$$

Suppose that this sequence splits and let τ be a retract of i. Take the subscheme T of X associated to the ideal

$$\mathscr{I}_T = \ker(\mathscr{I}_Z \to \mathscr{I}_Z/\mathscr{I}_Y\mathscr{I}_Z \xrightarrow{\tau} L^2).$$

Then T is a l.c.i. curve with the same support as Y, verifying $[T]=3[Y]$ and $\det N_T \mid Y = (\det N_Y) \otimes L^{-2}$. The exact sequences

$$0 \to L \to \mathscr{O}_Z \to \mathscr{O}_Y \to 0, \quad 0 \to L^2 \to \mathscr{O}_T \to \mathscr{O}_Z \to 0$$

allow to connect $\operatorname{Pic} T$ to $\operatorname{Pic} Y$. They also are needed in the computation of the Atiyah-Rees invariant.

The obstruction of the above splitting lies in $H^1(Y, \det N_Y \otimes L^3)$; hence, if L is sufficiently positive, the above construction may be performed.

As we have seen, by doubling and tripling procedures one can vary the cohomology classes of curves. More important is to remark that, due to the degree of freedom allowed by the choice of the line bundle L, one can also vary the determinant of the normal bundle.

1.3. We shall need the residual intersection device and the smoothness of 1-dimensional cycles, see ([9], § 7) for an account of these topics. We explain here for later use the meaning of the statement that, on 3-folds, the (smooth) 1-cycles are very movable.

We rely on the following form of Bertini's theorem: if L is a line bundle on a quasi-projective manifold X and V is a finite dimensional subspace of $H^0(X, L)$ which generates L, then for s general in V the scheme of zeroes of s is non-singular, and connected if $\dim X \geqq 2$.

Let Y be a non-singular curve in Y, not necessarily connected. Take $k \gg 0$, so that $\mathscr{I}_Y \otimes L^k$ is globally generated. Then the general section s in $H^0(X, \mathscr{I}_Y \otimes L^k)$ defines a non-singular, connected surface H_s passing through Y (one uses Bertini's theorem on $X \setminus Y$ and the fact that the image in $H^0(Y, N_Y^* \otimes L^k \mid Y)$ of the general section has no zeroes). Now, if one considers another general section t in $H^0(X, \mathscr{I}_Y \otimes L^k)$, then H_t is still non-singular and connected, and the ideal-theoretically equality $H_s \cap H_t = Y \cup Z$ holds, where Z is a second non-singular curve (we choose t with the additional condition that

it smoothly vanishes in the quotient bundle of $N_Y^* \otimes L^k | Y$ by the sub-bundle generated by s; the points of $Y \cap Z$ are exactly these zeroes). Moreover, we can choose s and t such that H_s and H_t avoid a given finite subset of $X \setminus Y$ and Z avoid a given finite subset of Y. In particular Z may avoid a given curve which has no common irreducible component with Y.

Let denote $h = c_1(L)$. One has $[H_s] = [H_t] = k.\ h$ and consequently $k^2 h^2 = [Y] + [Z]$.

Notice also that if H' and H'' are general in $|H^0(X, L)|$, then $H' \cap H''$ is a smooth connected curve and $h^2 = [H' \cap H'']$. Here one can again choose a representative curve for h^2, which is disjoint of a given curve.

Let assume now that ξ is a 1-dimensional cycle on X. By a result of Hironaka and Kleiman, ξ is smoothable in its rational equivalence class. In particular, passing to the associated cohomology classes in $H^4(X, \mathbb{Z})$ one gets

$$[\xi] = \sum_{i=1}^{k} n_i [Y_i], \qquad n_i \in \mathbb{Z},$$

where Y_i are smooth and connected curves. As we have already seen one can write $[Y_1] = k_1^2 h^2 - [Z_1]$ with $k_1 \gg 0$ and Z_1 disjoint of $Y_2, Y_3, \ldots, Y_k$. By repeating this procedure we can finally assume that ξ may be expressed as above, with new curves and coefficients, such that the curves Y_i are pairwise disjoint. This fact, and some information on the parity of n_i, is what we shall use in the sequel.

1.4. Let Y be a finite polyhedron and let d_{2k+1} denote the (possible) non-trivial coboundaries in the Atiyah-Hirzebruch spectral sequence for $K(Y)$, associated to the simplicial decomposition of Y. The following assertion is valid ([2] or [8], Theorem I in Chap. VIII):

Fix an integer $q \geqq 1$. If $\alpha \in H^{2q}(Y, \mathbb{Z})$ satisfies $d_{2k+1}\alpha = 0$ for all $k \geqq 1$, then there exists a class ξ in $K(Y)$, such that ξ is trivial on the skeleton Y^{2q-1}, and $c_q(\xi) = (q-1)!\,\alpha$.

2. Toward the topological classification

2.1. Proposition. *Let Y be a finite polyhedron of dimension less or equal to 7 and assume $H^7(Y, \mathbb{Z})$ has no 2-torsion. The cohomology classes (c_1, c_2, c_3) are the Chern classes of a rank 3 vector bundle on Y iff*

$$c_3 \equiv (c_1 + Sq^2) c_2 \quad \text{in } H^6(Y, \mathbb{Z}_2).$$

Proof. The necessity of the condition follows by pulling-back a formula of Wu, see [13], which asserts that the Stiefel-Whitney classes of the tautological bundle on $BU(3)$ satisfy

$$w_6 = (w_2 + Sq^2) w_4.$$

Conversely, assume that the congruence condition holds for a triple (c_1, c_2, c_3). Every class $\xi \in K(Y)$ decomposes into $\xi = [E] + k[\mathbb{1}]$, where E is a rank 3 bundle on Y, $k \in \mathbb{Z}$ and $\mathbb{1}$ the trivial line bundle, see [11]. Consequently, it suffices to find a $\xi \in K(Y)$, with $c_i(\xi) = c_i$, $i = 1, 2, 3$.

First, we choose a line bundle L with $c_1(L)=c_1$, and we put $\xi_1=[L]$, so that $c_1(\xi_1)=c_1$, $c_2(\xi_1)=0$ and $c_3(\xi_1)=0$. The only possible non-trivial coboundary in the Atiyah-Hirzebruch spectral sequence, starting at $E_r^{4,q}$, is d_3: $H^4(Y,\mathbb{Z})\to H^7(Y,\mathbb{Z})$. As the values of d_3 lie in the 2-torsion part of $H^7(Y,\mathbb{Z})$ one gets $d_3=0$. Therefore there exists by 1.4 a class $\xi_2\in K(Y)$, such that $c_1(\xi_2)=0$, $c_2(\xi_2)=c_2$; let us denote $c_3(\xi_2)=\tau$. Similarly one finds $\xi_3\in K(Y)$ with $c_1(\xi_3)=0$, $c_2(\xi_3)=0$ and $c_3(\xi_3)=2\sigma$, where σ is a given, arbitrary, class in $H^6(Y,\mathbb{Z})$.

Take $\xi=\xi_1+\xi_2+\xi_3$. Then $c_1(\xi)=c_1$, $c_2(\xi)=c_2$, $c_3(\xi)=c_1c_2+\tau+2\sigma$. Since $\tau\equiv Sq^2c_2$ and $c_3\equiv c_1c_2+Sq^2c_2$ in $H^6(X,\mathbb{Z}_2)$, one can find a cohomology class σ such that $c_3(\xi)=c_3$.

2.2. Theorem 1. *Let X be a compact connected 3-dimensional complex manifold.*

(1) *The cohomology classes (c_1,c_2,c_3) are the Chern classes of a rank 3 topological vector bundle on X iff $(c_1+c_1(X))c_2\equiv c_3 \pmod 2$. Moreover, the triple (c_1,c_2,c_3) determines the bundle up to an isomorphism.*

(2) *The pair (c_1,c_2) corresponds to the Chern classes of a rank 2 topological vector bundle on X iff $(c_1+c_1(X))c_2\equiv 0 \pmod 2$. Moreover,*

$$\mathrm{Vect}^2_{\mathrm{top}}(X)\cong\{(c_1,c_2,\beta(c_1,c_2),\alpha(c_1,c_2,\beta(c_1,c_2)))\},$$

where (c_1,c_2) are as above, β is an arbitrary element in the quotient

$$H^5(X,\mathbb{Z}_2)/(c_1+c_1(X))H^3(X,\mathbb{Z})+c_2H^1(X,\mathbb{Z}),$$

and α is an arbitrary element in some quotient of

$$H^6(X,\mathbb{Z}_2)/(c_1+c_1(X))H^4(X,\mathbb{Z}).$$

(Notice the abuse of notation in writing the denominators.)

For (1) see [6], Theorem 1, or the above proposition. The statement (2) follows by analysing the first two steps of the Postnikov tower of the map $BU(2)\to BU$, as described in [17]. For the first step $E_7\to BU$ one uses [12], namely Formula 1.5 written for Chern classes, and further one applies the above proposition to the space $Y=SX$, the suspension of X. By remarking that the cup product $H^2\times H^4\to H^6$ is trivial on SX, one finds that the set of homotopy classes of liftings $X\to E_7$ of a given class $\xi\in[X,BU]$, with $c_3(\xi)=0$, is an affine space over $H^5(X,\mathbb{Z}_2)/\Delta(\xi)$, where $\Delta(\xi)$ is the range of the map

$$H^1(X,\mathbb{Z})\times H^3(X,\mathbb{Z})\to H^5(X,\mathbb{Z}_2),$$

defined by $(\tau_1,\tau_3)\mapsto c_2(\xi)\tau_1+(c_1(\xi)+Sq^2)\tau_3$. Furthermore $[X,BU(2)]\cong[X,E_8]$, and the enumeration of the liftings relative to $E_8\to E_7$ is done in [17], p. 101. All these facts imply (2).

It turns out that the right interpretation of the parametrization in (2) is given by actions of the groups $H^6(X,\mathbb{Z}_2)$ and $H^5(X,\mathbb{Z}_2)$ on the fibres in the decomposition

$$[X,BU(2)]\cong[X,E_8]\to[X,E_7]\to[X,BU].$$

The actions are transitive and those corresponding to the second arrow have precise isotropy groups. We do not know the isotropy groups for the first arrow, or equivalently, whether the parameter space for α is exactly

$$H^6(X, \mathbb{Z}_2)/(c_1 + c_1(X))H^4(X, \mathbb{Z}).$$

The question is related, via K_X-symplectic K-theory [3], to the problem of lifting (up to homotopy) the determinants of automorphisms of the bundles, from the skeleton X^5 to X (a problem which "lies" in $H^1(X, \mathbb{Z})$, in fact in $H^1(X, \mathbb{Z}_2)$).

Let us assume in addition that $H_1(X, \mathbb{Z}) = 0$. Then one has no β in the parametrization of rank 2 bundles, and, if $c_1 \not\equiv c_1(X) \pmod 2$, then α disappears too. In the K_X-symplectic case, i.e. when $c_1 \equiv c_1(X) \pmod 2$, the space of values for α is precisely $H^6(X, \mathbb{Z}_2) = \mathbb{Z}_2$, see [6] or the above remark. As we have already mentioned in the introduction, in this case the analytical invariant α distinguishes the topological type; more precisely, if E_1, E_2 are holomorphic K_X-symplectic bundles with $\alpha(E_1) \neq \alpha(E_2)$, then E_1 is not topologically isomorphic with E_2, see [3].

Finally, let us remark that the statements (1) and (2) in the theorem are valid (replacing $c_1(X)$ by Sq^2) for an arbitrary finite polyhedron X of dimension less or equal to 6 and without 2-torsion in $H^6(X, \mathbb{Z})$.

3. The proof of Theorem 2

3.1. We follow the general framework of [5]. We have to prove that if c_1, c_2, c_3 are algebraic cohomology classes, satisfying $(c_1 + c_1(X))c_2 \equiv c_3 \pmod 2$, then there exists a rank 3 algebraic bundle with these Chern classes.

It is useful to write down the transformation formulas for the Chern classes of a rank 3 bundle G, tensorized with a line bundle P:

$$\begin{aligned}
c_1(G \otimes P) &= c_1(G) + 3c_1(P), \\
c_2(G \otimes P) &= c_2(G) + 2c_1(G)c_1(P) + 3c_1(P)^2, \\
c_3(G \otimes P) &= c_3(G) + c_2(G)c_1(P) + c_1(G)c_1(P)^2 + c_1(P)^3.
\end{aligned}$$

We can tensorize the topological bundle having Chern classes (c_1, c_2, c_3) with a sufficiently positive algebraic line bundle, such that c_1 is the Chern class of an ample line bundle. By tensorizing again with a sufficiently high power of this line bundle, and by using 1.3, we can assume in addition that $c_2 = \sum k_i [Y_i]$, where $k_i > 0$ and Y_i are pairwise disjoint smooth curves. The ample line bundle L appearing in the sequel satisfies $c_1(L) = c_1$. Put $h = c_1$. Let us consider general elements H_1 and H_2 in $|H^0(X, L)|$ and denote $C = H_1 \cap H_2$, so that $[C] = h^2$. By the exact sequence

$$0 \to \mathcal{O}(-H_1 - H_2) \to \mathcal{O}(-H_1) \oplus \mathcal{O}(-H_2) \to \mathcal{I}_C \to 0, \tag{1}$$

we get $N_C \cong \mathcal{O}(H_1) \oplus \mathcal{O}(H_2) | Y = 2L | Y$. Accordingly, for every $k \geqq 0$, one finds epimorphisms

$$\mathcal{I}_C / \mathcal{I}_C^2 \to \det N_C \otimes L^{2k-1} | C \to 0,$$

and consequently one can use Ferrand's method. Since $H^1(C, \det N_C \otimes L^{2k-1} \mid C) = 0$ for large values of k (but independent of the realization C of h^2), one gets l.c.i. curves C' supported by C, so that $[C'] = 2[C]$ and $\det N_{C'} \cong L^{1-2k} \mid C'$. Using this fact and the method of Serre, one can find integers - and we fix such a value, say k_0 - so that the next statement holds: for an arbitrary disjoint union Y of such double structures (with $k = k_0$) there exists an extension

$$0 \to L^{k_0} \to F \to \mathscr{I}_Y L^{1-k_0} \to 0, \tag{2}$$

with F locally free. We shall consider only curves Y which are disjoint of the curves Y_i which appear in the expression of c_2. One has

$$c_1(F) = c_1(L) = h = c_1,$$
$$c_2(F) = (k_0(1-k_0) + 2d)h^2,$$

where d is the number of double curves considered in Y. The integer $d \geqq 1$ is arbitrary and it will be conveniently choosen at the end of the proof.

3.2. Start with an F as above. We seek for integers $t \gg 0$ and curves Z such that there exist extensions of the form

$$0 \to F \otimes L^{3t} \to E \otimes L^{2t} \to \mathscr{I}_Z \to 0, \tag{3}$$

with E locally free of rank 3, and such that the Chern classes of E are the given c_1, c_2, c_3. According to the conditions in Serre's construction one has to verify firstly that $H^2(X, F \otimes L^{3t}) = 0$ and that $\det N_Z \otimes F \otimes L^{3t}$ has a section without zeroes on Z. There is a t_0, such that for every $t \geqq t_0$ one has $H^2(X, F \otimes L^{3t}) = 0$ for every F as above (use (1), the exact sequence $0 \to L^{k_0 - 1} \otimes \det N_C \to \mathscr{O}_{C'} \to \mathscr{O}_C \to 0$, and (2)). If we can find an extension (3) with E locally free, then $c_1(E) = c_1(F) = c_1$. For the second Chern class one finds

$$c_2(E \otimes L^{2t}) = c_2(F \otimes L^{3t}) + [Z].$$

As we claim that $c_2(E) = c_2 = \sum k_i [Y_i]$, one derives that Z has the associated cohomology class

$$[Z] = \sum k_i [Y_i] + (t(3t+1) - k_0(k_0 - 1) - 2d)h^2.$$

Take $t \geqq t_0$ such that $t(3t+1) - k_0(k_0-1) - 2d > 1$. Let us explain how such curves Z may be chosen so that they satisfy for large values of t the condition on $\det N_Z \otimes F \otimes L^{3t}$, keeping in mind that we have to take care also of c_3.

Before doing this, we do the following observation. Let C be a non-singular curve, $C = H_1 \cap H_2$, as above. If P' is a line bundle and if the degree of P' exceeds a suitable integer, which depends only on $c_1(X)$ and h, then there exist epimorphisms $\mathscr{I}_C / \mathscr{I}_C^2 \to P' \to 0$. Let us fix such a degree. An epimorphism as above determines a double structure C', and let us define $\varepsilon = \chi(\mathscr{O}_{C'})$. For the sake of simplicity we call such a curve an ε-curve. If epimorphisms $\mathscr{I}_C / \mathscr{I}_C^2 \to P'' \to 0$ are taken into account, with $\deg P'' = \deg P' + 1$, then we obtain curves C' with $\chi(\mathscr{O}_{C'}) = \varepsilon + 1$. They are simply called $(\varepsilon+1)$-curves.

Let us come back to the construction of Z. We can assume that $\sum k_i[Y_i] = [T]$, where T is a l.c.i. curve supported by the union of Y_i (choose smooth surfaces S_i which contain correspondingly Y_i, then replace Y_i by its (k_i-1)-infinitesimal neighbourhood in S_i, and take the union). Fix such a T.

Take the arbitrary decomposition

$$t(3t+1)-k_0(k_0-1)-2d=2\alpha+2\beta, \qquad \alpha\geqq 0,\ \beta\geqq 0. \tag{4}$$

The desired curve Z will be a disjoint union of T with α copies of curves of type ε and β copies of type $\varepsilon+1$. Moreover, we may assume that Z is disjoint of the curve Y which appears in (2). There is an integer t_1 such that for every $t\geqq t_1$ the bundle $\det N_Z\otimes F\otimes L^{3t}$ possesses a nonvanishing section: for this purpose restrict (2) to any component of Z and so on....

All in all, we can find an integer t_2, so that for $t\geqq t_2$ there exist extensions (3) with E locally free and $c_1(E)=c_1$, $c_2(E)=c_2$, for every curve Z as above. The integer t_2 depends only on $c_1(X)$, $h=c_1$, k_0, the decomposition $c_2=\sum k_i[Y_i]=[T]$ and $\deg P'$.

3.3. Let us take an E as above. By the variation of α and β in (4) we would like to obtain $c_3(E)=c_3$. One has:

$$c_3(E\otimes L^{2t})=(c_1(F\otimes L^{3t})+c_1(X))[Z]-2\chi(\mathcal{O}_Z).$$

hence

$$\begin{aligned}8h^3t^3+4h^3t^2+2hc_2t+c_3(E)\\ =(6ht+h+c_1(X))(c_2+2(\alpha+\beta)h^2)-2(\chi(\mathcal{O}_T)+\alpha\varepsilon+\beta(\varepsilon+1)).\end{aligned}$$

By requiring $c_3(E)=c_3$, one finds an equation in α and β, which must be solved. Due to the integrality condition $(h+c_1(X))c_2\equiv c_3 \pmod 2$, we can divide by 2 and write this equation as

$$P(t)=\alpha A(t)+\beta(A(t)-1),$$

where $A(t)=6h^3t+h^3+h^2c_1(X)-\varepsilon$ and $P(t)=4h^3t^3+$ lower terms (which depend only on $c_1(X)$, c_1, c_2, c_3 and the fixed curve T).

From now on, we reverse the way of the argument. Thus, $A(t)$ and $A(t)-1$ are relatively prime numbers, hence for $t\gg 0$ it is possible to find a decomposition

$$P(t)=\alpha_t A(t)+\beta_t(A(t)-1),$$

where α_t, β_t are positive integers. Since

$$P(t)=(\alpha_t+\beta_t)A(t)-B(t),$$

we asymptotically find $\alpha_t+\beta_t\sim\frac{2}{3}t^2$. Thus $t(3t+1)-k_0(k_0-1)-2(\alpha+\beta)\sim\frac{5}{3}t^2$ for $t\gg 0$. Consequently, there exist integers $t\geqq t_2$, sufficiently large such that $P(t)$ can be expressed as above and the number $t(3t+1)-k_0(k_0-1)-2(\alpha_t+\beta_t)$ is positive. Take d to be half of this number. Choose Y, construct F, then construct Z associated to α_t and β_t and finally one obtains the desired bundle E.

4. The proof of Theorem 3

4.1. Let us consider a topological bundle F of rank 2, with algebraic Chern classes c_1 and c_2. We want to construct an algebraic bundle E of rank 2 and with the same Chern classes. We begin as in 3.1. By successive convenient twistings of E with algebraic line bundles, we can assume (and we do) that $c_1 = c_1(L)$, where L is an ample line bundle, and $c_2 = \sum n_i[Y_i]$, with $n_i \geqq 3$ and Y_i disjoint connected smooth curves. Let us denote $h = c_1(L)$. Let $Y_{i_1}, \ldots, Y_{i_r}$ be those components whose coefficient n_i is odd. We choose an even integer $m \gg 0$, so that $m^2h^2 = [Y_{i_1}] + \ldots + [Y_{i_r}] + [U]$, where U is smooth and disjoint of the other components, see 1.3. We also choose an even integer $n \gg 0$, such that $n^2h^2 = [U] + [Z]$, with Z smooth and disjoint of all Y_i. Accordingly, $[Y_{i_1}] + \ldots + [Y_{i_r}] = (m^2 - n^2)h^2 + [Z]$. Further we tensorize once more F with L^q. One gets $c_2(F \otimes L^q) = c_2(F) + q(q+1)h^2$. If we take $q \gg 0$, then the new bundle $F \otimes L^q$ has $c_2 = \sum 2n_i[Y_i] + 3[Z]$, with new coefficients and new components Y_i, where $n_i > 0$, Y_i and Z are smooth, connected and pairwisely disjoint. The new c_1 is still the first Chern class of an ample line bundle; we keep the same notation L for it, and h for $c_1(L)$. After these normalizations of the initial topological data, we can look for extensions. Let us choose, as in § 3, an l.c.i. curve Y supported by the union of the curves Y_i and satisfying $[Y] = \sum n_i[Y_i]$; it follows that $c_2 = 2[Y] + 3[Z]$.

We seek for integers k and for l.c.i. curves T, which give rise to extensions

$$0 \to L^k \to E \to \mathscr{I}_T L^{1-k} \to 0,$$

with E locally free. One has $c_1(E) = c_1(L) = h = c_1$ and $c_2(E) = k(1-k)h^2 + [T]$. If we show that for $k \gg 0$, one can choose T so that there exist extensions with $k(1-k)h^2 + [T] = c_2$, then the proof will be finished. We take T to be a disjoint union of a convenient double structure Y' on Y with a triple structure Z'' on Z and with $\frac{1}{2}k(k-1)$ double structure C' supported on pairwise disjoint curves of the form $C = H' \cap H''$, where H' and H'' are general elements of $|H^0(X, L)|$.

First of all, notice that $H^2(X, L^{2k-1}) = 0$ for $k \gg 0$. We have to accomplish Serre's condition: $\det N_T \cong L^{1-2k}|T$. For an arbitrary l.c.i. curve S(e.g., for Y or for the previously described curves C) one finds, for large values of k, epimorphisms

$$\mathscr{I}_S/\mathscr{I}_S^2 \to \det N_S \otimes L^{2k-1} \to 0,$$

hence double structures S' with $\det N_{S'}|S \cong L^{1-2k}|S$. Since for $k \gg 0$ the restriction map $\operatorname{Pic} S' \to \operatorname{Pic} S$ is bijective, one obtains exactly the desired relation $\det N_{S'} \cong L^{1-2k}|S'$. It remains to prove that (again for $k \gg 0$) there are triple structures Z'' on Z with $\det N_{Z''} \cong L^{1-2k}|Z''$. Because $\det N_Z \cong \omega_Z \otimes (\omega_X|Z)^*$, $[Z] \equiv c_2 \pmod 2$ and $\deg \omega_Z \equiv 0 \pmod 2$, we get that the degree of $\det N_Z \otimes L^{2k-1}|Z$ has the same parity as $(c_1 + c_1(X))c_2$, therefore it is even by the integrality condition. Consequently, for every k, one finds a line bundle P_k on Z, so that $P_k^2 \cong \det N_Z \otimes L^{2k-1}|Z$. There exists an integer k_0 with the following property: if P is a line bundle on Z and $\deg P \geqq k_0$, then there exists an epimorphism $\mathscr{I}_Z/\mathscr{I}_Z^2 \to P \to 0$ and a retract $\mathscr{I}_{Z'}/\mathscr{I}_Z\mathscr{I}_{Z'} \xrightarrow{\tau} P^2$ for the inclusion $0 \to P^2 \xrightarrow{i} \mathscr{I}_{Z'}/\mathscr{I}_Z\mathscr{I}_{Z'}$, where Z' is the associated double structure. This retract defines a triple structure Z'' on Z. By enlarging k_0, if necessary, we have isomorphisms

$\operatorname{Pic} Z'' \xrightarrow{\sim} \operatorname{Pic} Z' \xrightarrow{\sim} \operatorname{Pic} Z$ and consequently $\det N_{Z''} \cong L^{1-2k}|Z''$. We take $P = P_k$ for $k \gg 0$ and thus we get the needed curve Z''.

4.2. Let F be a rank 2 topological bundle with Chern classes c_1, c_2, and in addition, $c_1 \equiv c_1(X) \pmod 2$. As described above, we may assume after a normalization that $c_1 = c_1(L)$, with L ample on X and $c_2 = \sum 2n_i[Y_i] + 3[Z]$, where $n_i > 0$, and Y_i and Z are connected, pairwise disjoint smooth curves. By the exponential exact sequence, the condition $c_1 + c_1(X) \equiv 0 \pmod 2$ implies the existence of an algebraic line bundle A on X, with the property $L \otimes K_X^{-1} \cong A^2$.

We have already seen that extensions like

$$0 \to L^k \to E \to \mathscr{I}_T L^{1-k} \to 0$$

give rise to algebraic bundles E of rank 2 with prescribed Chern classes c_1 and c_2. Moreover, $\det E \cong L$, hence for $E_{\text{norm}} := E \otimes A^{-1}$ one gets $\det(E_{\text{norm}}) \cong K_X$. We shall show how to vary the extension parameters in order to obtain for the invariant

$$\alpha(E_{\text{norm}}) = h^0(E \otimes A^{-1}) + h^2(E \otimes A^{-1}) \pmod 2$$

both values 0 and 1. This will conclude the proof of the K_X-symplectic part of Theorem 3. Let us come back to the construction presented in 4.1. The triple structure Z'' on Z involves an integer $k \gg 0$ and an epimorphism $\mathscr{I}_Z/\mathscr{I}_Z^2 \to P \to 0$, where P is a line bundle on Z with $P^2 \cong \det N_Z \otimes L^{2k-1}|Z$. Using the extension, the exact sequences

$$0 \to P \to \mathscr{O}_{Z'} \to \mathscr{O}_Z \to 0, \qquad 0 \to P^2 \to \mathscr{O}_{Z''} \to \mathscr{O}_{Z'} \to 0$$

and choosing a sufficiently high value of k (which depends only on the given data Y, Z and A) in order to have vanishing terms in computations, one gets

$$\begin{aligned} h^0(E \otimes A^{-1}) + h^2(E \otimes A^{-1}) &= h^0(L^k \otimes A^{-1}) + h^1(L^{1-k} \otimes A^{-1} \mid T) \\ &= h^0(L^k \otimes A^{-1}) + h^1(L^{1-k} \otimes A^{-1} \mid Y') + \tfrac{1}{2}k(k-1)h^1(L^{1-k} \otimes A^{-1} \mid C') \\ &\quad + h^1(L^{1-k} \otimes A^{-1} \mid Z) + h^1(P \otimes (L^{1-k} \otimes A^{-1}) \mid Z). \end{aligned}$$

Let us denote $Q = P \otimes (L^{1-k} \otimes A^{-1}) | Z$. We have

$$Q^2 \cong \det N_Z \otimes (L \otimes A^{-2}) | Z \cong \det N_Z \otimes K_X | Z \cong K_Z,$$

hence Q is a θ-characteristic on Z. Now, we modify the curve T by changing the triple structure on Z. more precisely we start with an epimorphism $\mathscr{I}_Z/\mathscr{I}_Z^2 \to P \otimes R \to 0$, where $R^2 \cong \mathscr{O}_Z$. If $\tilde{T}$ is the new curve and $\tilde{E}$ is the new bundle, then $\alpha(E_{\text{norm}}) - \alpha(\tilde{E}_{\text{norm}}) = h^1(Z, Q) - h^1(Z, Q \otimes R)$. Because $h^1 \pmod 2$ takes both values 0 and 1 when the θ-characteristic changes (see [1] for the precise result), we can modify the value $\alpha(E_{\text{norm}})$, and the proof is complete.

5. The case of rational manifolds

We prove Theorem 4. Let X be a complex projective, smooth rational threefold. Since X is simply connected, $H^1(X, \mathbb{Z}) = H^5(X, \mathbb{Z}) = 0$ and $H^2(X, \mathbb{Z})$ has no torsion. By ([6], th. 1, or § 2 above) the topological classification of rank 2 bundles is given by the Chern classes and, when $c_1 + c_1(X) \equiv 0 \pmod 2$ (in

which case we can normalize so that $c_1 = c_1(K_X)$), by the Atiyah-Rees invariant. Moreover, the analytical invariant α distinguishes between the topological types.

By using these facts and Theorems 2 and 3, it remains to prove that every cohomology class in $H^2(X, \mathbb{Z})$ or $H^4(X, \mathbb{Z})$ is algebraic. For $H^2(X, \mathbb{Z})$ this is clear by the exponential exact sequence. Because we have not found a reference for the analogous statement for $H^4(X, \mathbb{Z})$, we sketch below a proof of it. As it is well known, there exists a rational threefold together with two morphisms $Z \to X$, $Z \to \mathbb{P}^3$, which are both compositions of monoidal transformations with smooth center. Thus we are reduced to prove the following fact. If $X \xrightarrow{f} Y$ is a monoidal transformation of 3-folds with smooth center, then every cohomology class in $H^4(X, \mathbb{Z})$ is algebraic if and only if the same property is true on Y.

For the proof, if X has the property, then Y has it too, as $f_* f^* = \mathrm{id}$, in cohomology as well as in the Chow ring. Assume that Y has the property and let be an element of $H^4(X, \mathbb{Z})$. Then $\xi - f^* f_* \xi$ is of the form $i_*(\eta)$, where $\eta \in H^2(E, \mathbb{Z})$, E denoting the exceptional divisor and $i: E \to X$ the inclusion. But η is algebraic in virtue of $H^2(E, \mathcal{O}) = 0$, which concludes the proof.

References

1. Atiyah, M.F.: Riemann surfaces and Spin structures. Ann. Ecole Norm. Sup. **4**, 47–62 (1971)
2. Atiyah, M.F., Hirzebruch, F.: Vector bundles and homogeneous spaces. Proceedings of Symposia of A.M.S., vol. III, 20–104 (1960)
3. Atiyah, M.F., Rees, E.: Vector bundles on projective 3-space. Invent. Math. **35**, 131–153 (1976)
4. Bănică, C., Forster, O.: Multiplicity structures on space curves. (To appear)
5. Bănică, C., Coandă, J.: Existence of rank 3 vector bundles with given Chern classes on homogeneous rational 3-folds. Manuscripta Math. **51**, 121–143 (1985)
6. Bănică, C., Putinar, M.: On complex vector bundles on rational threefolds. Math. Proc. Camb. Phil. Soc. **97**, 279–288 (1985)
7. Ferrand, D.: Courbes gauches et fibrés de rang 2. C.R. Acad. Sci. Paris **281**, A, 345–347 (1975)
8. Griffiths, P.A., Adams, J.: Topics in algebraic and analytic geometry. Princeton University Press 1974
9. Hartshorne, R.: Equivalence relations on algebraic cycles and subvarieties of small codimension. In: Algebraic Geometry, Arcata 1974, Proceedings of Symposia of A.M.S. **29**, 129–164 (1975)
10. Horrocks, G.: A construction for locally free sheaves. Topology **7**, 117–120 (1968)
11. Husemoller, D.: Fibre bundles. McGraw-Hill (1966)
12. James, I-M., Thomas, E.: An approach to the enumeration problem for non-stable vector bundles. J. Math. and Mech. **14**, 485–506 (1965)
13. Milnor, J., Stasheff, J.: Characteristic classes. Ann. Math. Studies **76**, Princeton University Press 1974
14. Okonek, C., Schneider, M., Spindler, H.: Vector bundles on complex projective spaces. Basel, Boston: Birkhäuser 1980
15. Schwarzenberger, R.L.E.: Vector bundles on algebraic surfaces. Proc. London. Math. Soc. **11**, 601–622 (1961)
16. Serre, J.P.: Sur les modules projectifs. Sém. Dubreil-Pisot 1960/1961 exposé 2
17. Switzer, R.: Postnikov towers associated with complex 2-plane and symplectic line bundles. Math. Z. **168**, 87–103 (1979)
18. Vogelaar, J.A.: Constructing vector bundles from codimension two subvarieties. Thesis, Leiden (1978)

Oblatum 19-VI-1986

Invent. math. 88, 439-446 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Variational Torelli implies generic Torelli

David Cox[1,*], Ron Donagi[2,*], and Loring Tu[3]

[1] Department of Mathematics, Amherst College, Amherst, MA 01002, USA
[2] Department of Mathematics, Northeastern University, Boston, MA 02115, USA
[3] Department of Mathematics, Tufts University, Medford, Ma 02155, USA

Introduction

The methods of Infinitesimal Variation of Hodge Structure (cf. [3]) provide a powerful tool for proving generic Torelli theorems (see [6] for the case of projective hypersurfaces). However, the success of this method depends on the implication

$$\text{Variational Torelli} \Rightarrow \text{Generic Torelli.}$$

In the literature, various forms of this implication have appeared, sometimes with proof (cf. [2, 3, 5, 7-9, 14 and 17]). All previous precise formulations have required several auxiliary hypotheses, including:

- The existence of a coarse moduli space for the varieties in question.
- The existence of a regular value for the period map, i.e., a *smooth* point which is the image of a regular point of the map.
- The generic injectivity of the differential of the period map, i.e., Infinitesimal Torelli.

Checking that these conditions are satisfied in any given case can require considerable effort. For instance, the existence of a regular value requires that the isotropy group of a generic member of the family be trivial (cf. [14], §9).

This paper will prove that whenever an analytic period map exists, variational Torelli always implies generic Torelli. There are no extra conditions to be verified. In addition, variational Torelli implies a stronger result: the generic injectivity of the period map to D/Γ, where Γ is *any* discrete subgroup of the automorphism group $G_{\mathbb{R}}$ of the period domain D, for which the period map is well defined. Thus Γ must contain the monodromy group, but it need not preserve the integral, or even rational, structure of the Hodge structures in question.

The converse, that generic Torelli implies variational Torelli, is false. This was demonstrated for elliptic surfaces in [4], and can also be verified for $K3$

* Research partially supported by the National Science Foundation

surfaces. It seems plausible, though, that some version of variational Torelli might be equivalent to the above strong generic Torelli, i.e., to generic Torelli for all possible Γ.

In § 1 we state and prove a "principle of prolongation" for analytic maps. In § 2 we review the existence of generic quotients for algebraic equivalence relations. Finally, in § 3 we give precise statements of variational and generic Torelli, and prove our main result, that variational Torelli implies generic Torelli.

§ 1. The principle of prolongation

Let $f: X \to Y$ be an analytic map of complex manifolds of constant rank r. The prolongation $f_{(*)}$ of f is the map

$$f_{(*)}: X \to \tilde{Y} = \mathrm{Gr}(r, T(Y))$$

defined by sending $x \in X$ to $f_*(T_x(X)) \subset T_{f(x)}(Y)$. Here, $T(Y)$ is the tangent bundle of Y, and $\mathrm{Gr}(r, T(Y))$ is its bundle of Grassmannians of r-dimensional subspaces.

Recall that a subset T of an analytic space X is thin if T is contained in a countable union of proper analytic subsets of open subsets of X. In case f is not proper, so that $f(X)$ need not be an analytic subspace, we say that a subset T of $f(X) \subset Y$ is thin if $f^{-1}(T)$ is thin in X.

Here is a precise form of the "principle of prolongation" (cf. [5] and [6]).

Theorem. *Let X be a complex manifold, not necessarily connected, and let $f: X \to Y$ be an analytic map to another complex manifold, and assume that f has constant rank. If $f_{(*)}: X \to \tilde{Y}$ is the prolongation of f, then there is a thin subset $T \subset f(X) \subset Y$ such that*

$$x \in X - f^{-1}(T), \qquad y \in X, f(x) = f(y) \Rightarrow f_{(*)}(x) = f_{(*)}(y).$$

Remark. We summarize this situation by saying that outside of a thin subset of $f(X)$, f and $f_{(*)}$ have the same fibers.

Proof. Let Δ_Y denote the diagonal of $Y \times Y$, and consider the closed analytic subsets

$$\Delta_f = (f \times f)^{-1}(\Delta_Y), \qquad \Delta_{f_{(*)}} = (f_{(*)} \times f_{(*)})^{-1}(\Delta_{\tilde{Y}})$$

in $X \times X$, and the open subset

$$E = \Delta_f - \Delta_{f_{(*)}} \subset \Delta_f.$$

Let r be the rank of f, which is assumed to be constant. It follows that for $y \in Y$, the fibers $f^{-1}(y)$ are smooth of pure codimension r when nonempty. Let

$$\pi = f \circ p_1 = f \circ p_2: E \to Y,$$

where $p_1, p_2: E \to X$ are the projections. Since $\dim(p_1^{-1}(x)) = \dim(X) - r$, $i = 1, 2$, it follows that π has codimension $2r$ fibers (when nonempty) in $X \times X$.

We claim that E has codimension $>r$ in $X \times X$. To prove this, consider a point $(x, x') \in E$ with $f(x)=f(x')=y$. We can choose small complex neighborhoods U of x and U' of x' such that

$$f: U \to V \subset Y, \quad f: U' \to V' \subset Y$$

are submersions onto r dimensional submanifolds V, V' of a neighborhood of y in Y. Then

$$\begin{aligned}(x, x') \in E &\Rightarrow f_{(*)}(x) \neq f_{(*)}(x') \\ &\Rightarrow T_y(V) \neq T_y(V') \\ &\Rightarrow \dim_y(V \cap V') < r.\end{aligned}$$

This proves our claim, since π maps $E \cap (U \times U')$ into $V \cap V'$ with fibers that have codimension $2r$.

The proof of the theorem is now completed by taking T to be the subset $\pi(E) \subset f(X)$. T is a countable union of locally analytic sets of dimension $<r$, and hence is thin in $f(X)$. QED

Remark. Being a *countable union* of proper analytic subsets, a thin set could very well be dense. Thus, one may ask if the thinness of the exceptional set can be improved to analyticity. The answer is negative without a properness assumption on f.

For example, take an elliptic curve $E=\mathbb{C}/\Lambda$, where $\Lambda \subset \mathbb{C}$ is a lattice, and choose a complex number α such that $\alpha\Lambda \cap \Lambda = \{0\}$. Then define $f: \mathbb{C} \to \mathrm{Sym}^2(E)$ to be the composition

$$f: \mathbb{C} \xrightarrow{(1, \alpha)} \mathbb{C} \times \mathbb{C} \to E \times E \to \mathrm{Sym}^2(E),$$

where the map $E \times E \to \mathrm{Sym}^2(E)$ is the natural map sending the ordered pair $(x, y) \in E \times E$ to the unordered pair $\{x, y\}$. Since $\alpha\Lambda \cap \Lambda = \{0\}$, one sees that for $z \neq w$ in $\mathbb{C}$, we have

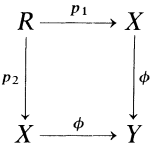
$$f(z)=f(w) \Leftrightarrow z=(1-\alpha^2)^{-1}(\alpha\lambda+\omega), \quad w=(1-\alpha^2)^{-1}(\lambda+\alpha\omega), \quad \lambda \neq \omega \in \Lambda.$$

If follows that f fails to be injective on a dense subset of $\mathbb{C}$. However, f has constant rank 1, so that $f_{(*)}$ is defined on all of $\mathbb{C}$, and in fact $f_{(*)}$ is injective. To see this, assume that $z \neq w$ and $f(z)=f(w)$, so that z and w are as described above. One can then find local coordinates around $f(z)=f(w)$ in $\mathrm{Sym}^2(E)$ such that $f_*(T_z(\mathbb{C}))$ is the subspace spanned by $(1, \alpha)$, while $f_*(T_w(\mathbb{C}))$ is spanned by $(\alpha, 1)$. Thus $f_{(*)}(z) \neq f_{(*)}(w)$, and injectivity is proved.

§ 2. The existence of generic quotients

Let X be a variety over an algebraically closed field of characteristic zero, and let $R \subset X \times X$ be a constructible equivalence relation. The notions of categorical quotient and geometric quotient of X by R are straightforward extensions of the corresponding notions for the group case (cf. [13], pp. 3–4). Thus a

morphism $\phi: X \to Y$ is a categorical quotient if the diagram



commutes, and any map $\psi: X \to Y$ with a similar commutative diagram factors uniquely through ϕ. Also, $\phi: X \to Y$ is a geometric quotient if:

(1) ϕ is surjective

(2) $R = X \times_Y X$

(3) ϕ is submersive

(4) $\mathcal{O}_Y = (\phi_* \mathcal{O}_X)^R$, the R-invariant functions on X. (A function $f \in \Gamma(U, \phi_* \mathcal{O}_X) = \Gamma(\phi^{-1}(U), \mathcal{O}_X)$ is R-invariant if it is constant on equivalence classes.)

As in the group case, a geometric quotient is always a categorical quotient (cf. [13], Prop. (0.1)).

The following seems to be a folk theorem:

Theorem. *Let $R \subset X \times X$ be a constructible equivalence relation on an irreducible variety X over an algebraically closed field of characteristic zero. Then there is a nonempty open subset $X_0 \subset X$ and a morphism $\phi: X_0 \to Z$ which is a categorical and geometric quotient of X_0 by $R|_{X_0}$. Furthermore, X_0 may be chosen so that Z and $\phi: X_0 \to Z$ are smooth.*

Proof. The case where R is given by the action of an algebraic group is due to Rosenlicht (cf. [16]). In general we may assume, by shrinking X, that the projections $p_1, p_2: R \to X$ are flat. A well-known result of Artin (cf. [1], Cor. (6.3), or [15], Cor. (8.10)) then asserts that the categorical quotient exists as an algebraic space. Since every algebraic space contains a dense open subscheme (cf. [12], Prop. (6.7)), we can then find $X_0 \subset X$ so that the categorical quotient of X_0 exists as a scheme.

A more elementary argument, which constructs a geometric quotient, can be given as follows.

Given $x \in X$, let Rx be its equivalence class. If y is a smooth point of $Rx = Ry$, then $p_1: R \to X$ is smooth at (y, y) since $p_1^{-1}(y) = \{y\} \times Ry$. It follows that by shrinking that we may assume that p_1 is smooth at all points of $\Delta_X \subset R$. Since R is an equivalence relation, this implies that all equivalence classes are smooth of pure dimension, say r.

Choose $x \in X$ and a codimension r subvariety $Y \subset X$ which is nonsingular at x and transversal to Rx there (e.g., embed a neighborhood of x in affine space and intersect it with a codimension r linear subspace transverse to $T_x(Rx)$). Let $Y_0 \subset Y$ be the open subset

$$Y_0 = \{y \in Y: Y \text{ is nonsingular and transversal to } Ry \text{ at } y\}.$$

Let $R_0 = R \cap (Y_0 \times Y_0)$ be the restriction of R to Y_0. The projections $p_1, p_2: R_0 \to Y_0$ are étale by construction, and by shrinking Y_0 we may assume

that they are finite étale, say of degree k. Then $p_1: R_0 \to Y_0$ induces a map $\psi: Y_0 \to \mathrm{Sym}^k(Y_0)$. Let $Z = \overline{\psi(Y_0)}$, so that we get a dominating map $\psi: Y_0 \to Z$.

By shrinking X to X_0 we may assume that $X_0 = RY_0$, i.e., that every equivalence class in X_0 meets Y_0, necessarily in an R_0 orbit. This gives us a morphism

$$\phi: X_0 \to Z$$

defined by sending $x \in X_0$ to $(Rx) \cap Y_0$. One easily checks that the fibers of ϕ are exactly the equivalence classes of $R|_{X_0}$. By shrinking further, we may assume that Z is smooth and ϕ is surjective and smooth (remember that the equivalence classes are smooth). Then the argument of [13], Prop. (0.2), shows that $\phi: X_0 \to Z$ is the desired geometric quotient. QED

Remark. When working over the complex numbers, the map $\phi: X_0 \to Z$ is also a geometric quotient in the category of analytic spaces. The proof is identical.

§ 3. The main result

Let $f: \mathscr{X} \to S$ be a flat family of polarized varieties over a smooth base S. We will assume that the cohomology of the fibers has a pure Hodge structure which varies analytically with $s \in S$ and depends only on the isomorphism class of the polarized variety X_s. For example, this holds if the X_s are smooth (or, by [18], quasi-smooth) projective varieties.

This data gives us a holomorphic period map

$$\tilde{p}: \tilde{S} \to D$$

where D is the appropriate period domain and $\tilde{S}$ is the universal cover of S. If a typical polarized Hodge structure is $(H_{\mathbb{Z}}, F^{\cdot}, Q)$, then we also get the monodromy representation

$$\pi_1(S) \to G_{\mathbb{Z}} = \mathrm{Aut}(H_{\mathbb{Z}}, Q)$$

and the image Γ_0 of this map is the monodromy group. Let Γ be any discrete subgroup of $G_{\mathbb{R}} = \mathrm{Aut}(H_{\mathbb{R}}, Q)$ (the group of automorphisms of D) which contains Γ_0. Then the period map $\tilde{p}$ induces a holomorphic map

$$p: S \to D/\Gamma.$$

While it is customary to let $\Gamma = \Gamma_0$ or $\Gamma = G_{\mathbb{Z}}$, we will make no such restrictions. In particular, Γ need not be contained in $G_{\mathbb{Z}}$.

Let $\approx$ denote the equivalence relation given by isomorphism of polarized varieties. Then generic Torelli for the family $f: \mathscr{X} \to S$ and the group Γ asserts that on the complement $S_0 \subset S$ of a proper analytic subvariety, the period map p is injective modulo $\approx$, i.e.,

$$s, t \in S_0, \qquad p(s) = p(t) \Rightarrow X_s \approx X_t.$$

Using standard results about $\mathrm{Isom}_S^p(X, Y)$ and $\mathrm{Pic}^{\tau}_{X/S}$ (cf. Technique de descente et théorèmes d'existence en geometrie algébrique, IV–VI, in [11]), it follows that $\approx$ induces a constructible equivalence relation on S.

Thus we may let $\phi: S_0 \to \mathscr{M}$ be a smooth generic quotient of S by the equivalence relation $\approx$ (this is where we use §2). Since $\mathscr{M}$ is a categorical quotient, we obtain a holomorphic map

$$q: \mathscr{M} \to D/\Gamma.$$

Generic Torelli is equivalent to the generic injectivity of this map, i.e., to injectivity away from a proper analytic subset of $\mathscr{M}$.

If $s \in S$, recall that the Infinitesimal Variation of Hodge Structure (IVHS) of X_s is $(H_{\mathbb{Z}}, F^\bullet, Q, T, \delta)$, where $(H_{\mathbb{Z}}, F^\bullet, Q)$ is the polarized Hodge structure of X_s, T is the image of $T_s(S)$ in the tangent space of D under of local lifting of the period map, and δ is the inclusion of the image into the tangent space of D. The basic idea of variational Torelli is to ask whether X_s is determined modulo $\approx$ by the algebraic part of its IVHS, i.e., by $(F^\bullet, Q, T, \delta)$. This is the notion used in [4], [5], [8] and elsewhere.

For the purposes of this paper, we need only a weaker notion of variational Torelli, namely the assertion that X_s is determined modulo $\approx$ by $\tilde{p}_{(*)}$ modulo the action of the group $G_{\mathbb{R}}$ of automorphisms of D. Here

$$\tilde{p}_{(*)}: \tilde{S} \to \mathrm{Gr}(r, T(D))$$

is the generically defined prolongation of the period map $\tilde{p}$. This is equivalent to saying that X_s is determined modulo $\approx$ by $(H_{\mathbb{R}}, F^\bullet, Q, T, \delta)$, i.e., by the algebraic and real structures of the IVHS of X_s.

We can now prove our main result.

Theorem. *If variational Torelli holds for a family $f: \mathscr{X} \to S$, then generic Torelli holds for f and for any discrete subgroup $\Gamma \subset G_{\mathbb{R}}$ containing the monodromy group Γ_0.*

Proof. First, we can shrink S so that the period map $\tilde{p}$ has constant rank r on $\tilde{S}$ and so that a geometric quotient $\phi: S \to \mathscr{M}$ exists, with ϕ a smooth map.

We saw above that we have a map

$$q: \mathscr{M} \to D/\Gamma.$$

Since $\phi: S \to \mathscr{M}$ is smooth, it has local sections, and thus q has local liftings to D since $p: S \to D/\Gamma$ does. This gives us a map

$$\tilde{q}: \tilde{\mathscr{M}} \to D$$

where $\tilde{\mathscr{M}}$ is the universal cover of $\mathscr{M}$.

Furthermore, if $s \in S$ and $m = \phi(s)$, then working in $\tilde{S}$ and $\tilde{\mathscr{M}}$ above these points, we have

$$\tilde{p}_*(T_s(S)) = \tilde{q}_*(T_m(\mathscr{M}))$$

since the fibers of ϕ consist of isomorphic polarized varieties, which forces $\tilde{p}$ to be constant above them. It follows that $\tilde{q}$ has constant rank and that the IVHS of X_s is determined by $m = \phi(s)$.

Now consider the map

$$\tilde{Q}: \tilde{\mathscr{M}} \times \Gamma \to D$$

defined by $\tilde{Q}(\tilde{m}, \gamma) = \gamma(\tilde{q}(\tilde{m}))$, which has

$$\tilde{Q}_{(*)}: \tilde{\mathscr{M}} \times \Gamma \to \mathrm{Gr}(r, T(D))$$

as its prolongation. By our assumption of variational Torelli and the analysis of the previous paragraph, the fibers of $\tilde{Q}_{(*)}$ are contained in the fibers of the map $\tilde{\mathscr{M}} \times \Gamma \to \mathscr{M}$. By the principle of prolongation (from § 1), it follows that the fibers of $\tilde{Q}$ are contained in the fibers of $\tilde{\mathscr{M}} \times \Gamma \to \mathscr{M}$, at least outside of a thin set. This means exactly that $q: \mathscr{M} \to D/\Gamma$ is injective outside of a thin set. By a basic result of Griffiths [10] (cf. also [3], Sect. (1.a)), the map q can be extended to a proper map

$$\bar{q}: \bar{\mathscr{M}} \to D/\Gamma$$

where $\bar{\mathscr{M}}$ is an appropriate partial compactification of $\mathscr{M}$. Since $\bar{q}$ is proper, its image Y is an analytic variety, and then $\bar{q}: \bar{\mathscr{M}} \to Y$ is proper, surjective, and is injective outside of a thin set. We claim that $\bar{q}$ (and hence q) is injective outside of an *analytic* set. To prove this, first note that $\dim(\bar{\mathscr{M}}) = \dim(Y)$ (this follows by analyzing the differential $\bar{q}_*$ at a point of maximal rank). Standard arguments then show that a proper surjective map between varieties of the same dimension has a well-defined degree (and the map is a finite covering space over the complement of a subvariety of the range). Since $\bar{q}$ is injective outside of a thin set, the degree must be one, and $\bar{q}$ is thus generically injective. This proves the theorem. QED

Remarks. 1) Note that the theorem remains true if we only assume that variational Torelli holds generically on S.

2) In our definition of generic Torelli, we talked about injectivity modulo $\approx$ outside a proper *analytic* subset. It would be nice to know if this exceptional set were always an *algebraic* subvariety.

References

1. Artin, M.: Versal deformations of algebraic stacks. Inv. math. **27**, 165–189 (1974)
2. Carlson, J., Griffiths, P.: Infinitesimal variations of Hodge structure and the global Torelli problem. Journées de géométrie algébrique d'Angers, pp. 51–76. Groningen: Sijthoff and Noordhoff 1980
3. Carslon, J., Green, M., Griffiths, P., Harris, J.: Infinitesimal variations of Hodge structures I. Comp. Math. **50**, 109–205 (1983)
4. Cox, D., Donagi, R.: On the failure of variational Torelli for regular elliptic surfaces with a section. Math. Ann. **273**, 673–683 (1986)
5. Donagi, R.: Generic Torelli and variational Schottky. Topics in Transcendental Algebraic Geometry, pp. 239–258. Princeton 1984
6. Donagi, R.: Generic Torelli for projective hypersurfaces. Comp. Math. **50**, 325–353 (1983)
7. Donagi, R., Green, M.: A new proof of the symmetrizer lemma and a stronger weak Torelli theorem for projective hypersurfaces. J. Diff. Geom. **20**, 459–461 (1984)
8. Donagi, R., Tu, L.: Generic Torelli for weighted hypersurfaces. Math. Ann., to appear
9. Green, M.: The period map for hypersurface sections of high degree of an arbitrary variety. Comp. Math. **55**, 135–156 (1985)
10. Griffiths, P.: Periods of integrals on algebraic manifolds III. Publ. Math. IHES **38**, 125–180 (1970)
11. Grothendieck, A.: Fondements de la géométrie algébrique (FGA). Extraits du Séminaire BOURBAKI 1957–1962, Secrétariat mathématique, 11 rue Pierre Curie, Paris (1962)

12. Knutson, D.: Algebraic spaces. Lect. Notes Math. **203**. Berlin Heidelberg New York: Springer 1971
13. Mumford, D., Fogarty, J.: Geometric invariant theory, 2nd enlarged edn. Ergeb. Math. **34**. Berlin Heidelberg New York: Springer 1982
14. Peters, C., Steenbrink, J.: Infinitesimal variation of Hodge structure and the generic Torelli problem for projective hypersurfaces, Classification of algebraic and analytic manifolds, pp. 399–464. Basel Boston: Birkhäuser 1983
15. Popp, H.: Moduli theory and classification theory of algebraic varieties. Lect. Notes Math. **620**. Berlin Heidelberg New York: Springer 1977
16. Rosenlicht, M.: A remark on quotient spaces. An. Acad. Bras. Cienc. **35**, 487–489 (1963)
17. Saito, M.-H.: Weak global Torelli theorem for certain weighted projective hypersurfaces. Duke Math. J. **53**, 67–111 (1986)
18. Steenbrink, J.: Intersection forms for quasihomogeneous singularities. Comp. Math. **34**, 211–223 (1977)

Oblatum 18-II-1986 & 1-IX-1986

Invent. math. 88, 447-493 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Functional determinants and geometry

Robin Forman

Department of Mathematics, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

Summary. In this paper we study the zeta-function determinant in the context of elliptic boundary value problems. Our main technique is to relate the determinant of an operator, or a ratio of determinants, to the boundary values of the solutions of the operator. This has the advantage of restricting attention to the solutions of the operator, which do not depend on the boundary conditions and can often be written down explicitly, rather than the eigenvalues, which are usually difficult to work with. In addition, the problem is reduced to a calculation over the boundary of the manifold which is a closed manifold of dimension one less than the original manifold. This has special significance in the case that the manifold is a finite interval. In this case the boundary is a pair of points and the determinant of an ordinary differential operator is expressed in terms of the determinant of a finite matrix.

The results are then applied to some geometric operators. In Sect. 4 we study the Jacobi operator acting along a geodesic segment and the covariant derivative operator acting along a loop. In Sect. 2 we calculate the determinant of the Laplacian acting on sections of a flat bundle over a flat torus. This can be related to an Eisenstein series and thus we have presented a new geometric method of summing such series. This sum is known as Kroneker's Second Limit formula. We then consider operators on a product manifold $M \times S^1$.

Introduction

In recent years, in particular with the advent of the path integral formalism of quantum physics, physicists have been faced with the task of extending to infinite dimensions some purely finite dimensional concepts. One of the most problematic is the notion of the determinant of a linear map. The case in which we are usually interested is when the linear map is a self-adjoint differential operator acting on sections of a vector bundle. In this case we cannot simply take the product of the eigenvalues, as they are unbounded, so

we need a new approach. One candidate is provided by the zeta-function determinant of Ray and Singer (see [R-S1]). This is defined by forming the zeta-function of the operator, $\zeta(s)=\sum_i \lambda_i^{-s}$, where the sum is over all eigenvalues, analytically continuing this to a meromorphic function of s, and letting the determinant equal $\exp(-\zeta'(0))$. This yields the normal answer for finite dimensional operators, and theorems of Seeley imply that this process is always possible. Furthermore, in many situations (e.g. Ray and Singer's analytic torsion) the zeta-function determinant has proven itself to be the correct generalization of the finite dimensional determinant. By "correct" I mean that it provides the answer that we would like and expect in situations in which there is one natural answer.

This paper presents some work in a program of studying the properties of the zeta-function determinant. Much analysis has been done on the analytic torsion T which, for an n-dimensional Riemannian manifold, is defined by

$$\log T=\sum_q (-1)^q q \log \det(\Delta_q)$$

where Δ_q is the Laplacian acting on q-forms with values in a fixed flat vector bundle. The alternating sum causes some simplification so that, for example, T does not depend on the metric. A broader question is to investigate the determinant of a *general* elliptic operator acting on sections of a vector bundle. By considering operators other than the Laplacian we gain insight into the properties of the determinant. Similarly, we consider general elliptic boundary problems and investigate the effect of varying the boundary conditions.

The original motivation for this work came from an attempt to gain a deeper understanding of the main result of [L-S], which states that the hessian of the energy function along a geodesic is essentially equal to the determinant of the differential of the exponential map along the geodesic (see Sect. 4.1 for the precise statement). This result is intriguing for two reasons. First, it indicates that the determinant of a "geometric" operator may provide geometric information, or more importantly, we may able to calculate the determinant directly from the geometric data. Second, the result expresses the determinant of a differential operator over a manifold with boundary (the geodesic with its endpoints) in terms of a very natural map acting on sections of a vector bundle over the boundary. This is the guiding principle of this paper.

Although it was not apparent at the time, with hindsight it has become clear that this approach should not be surprising. In 1963 Calderon introduced a new method of studying elliptic boundary problems (see [Ca], [Se3] [Hö]). The basic idea is best seen by example. If one is given a manifold M with boundary, one can solve the Neumann problem

$$Lf=0 \qquad \frac{\partial}{\partial v} f\Big|_{\partial M}=g$$

where v is an inward normal vector field along ∂M and L is a second order elliptic operator as follows: Suppose, for example that one understands well

the boundary value problem

$$Lf=0 \qquad \left(\frac{\partial}{\partial v} f+\mu f\right)\Big|_{\partial M}=h.$$

Then given a function h on ∂M, we can find the unique function f_h such that $Lf=0$ and $\left(\frac{\partial}{\partial v} f_h+\mu f_h\right)\Big|_{\partial M}=h$. Solving the Neumann problem is clearly equivalent to finding a function h such that $\frac{\partial}{\partial v} f_h\Big|_{\partial M}=g$. Hörmander showed in [Hö] that the map $h\to\frac{\partial}{\partial v} f_h\Big|_{\partial M}$, and all similar maps, is a pseudo-differential operator over the boundary. Thus, solving the Neumann boundary value problem has been reduced to a question about a pseudo-differential operator which exists over the boundary, a closed manifold.

In this paper we prove that, in a well-defined sense, $\frac{\operatorname{Det} L_N}{\operatorname{Det} L_\mu}$ is equal to the determinant of this pseudo-differential operator, where $L_N(L_\mu)$ is the operator L restricted to functions f which satisfy $\frac{\partial}{\partial v} f\Big|_{\partial M}=0$ $\left(\left(\frac{\partial}{\partial v} f+\mu f\right)\Big|_{\partial M}=0\right)$. More precisely, it is not clear how to define the determinant of the pseudodifferential operator, so we work instead with the determinant of a ratio of such operators (see Theorem 1). What may be surprising is that we can use this idea to explicitly calculate many determinants. In fact, almost all explicitly known determinants fit nicely into this framework. For example, in Sect. 2.1.1 we calculate the determinant of the Laplacian acting on sections of a flat unitary bundle over a flat torus. As shown in [R-S2] this determinant is equal to the sum of a certain Eisenstein series. Thus we have presented a new geometric method of summing such series, the sum of which is known as Kroneker's Second Limit formula. Using the same methods, we can make precise statements about the determinant of the operator $L_M+\frac{d^n}{dx^n}$ on $M\times S^1$, where L_M is any elliptic operator of order n on M, acting on sections of a twisted bundle.

The advantage of this approach to determinants is immediate. The pseudodifferential operator can be expressed entirely in terms of the boundary values of the solution space of our original elliptic differential operator. If one has a good understanding of the solution space (which is independent of the boundary conditions) one can determine how the determinant varies as a function of the operator and the boundary conditions. Furthermore, considerations can be restricted completely to the boundary, which is a closed manifold of dimension one less than the original manifold. This should be contrasted with terms such as $\operatorname{Det} L_N$ and $\operatorname{Det} L_\mu$, which are defined in terms of eigenvalues which are manifestly global quantities and can almost never be known precisely.

These simplifications have special significance in the case of an ordinary differential operator on a finite interval. Here the operator acts on an infinite dimensional space of functions, but the space of possible boundary values is

finite dimensional. Determinants of ordinary differential operators were studied previously in [D-S]. Our results for partial differential operators, restricted to the case of ordinary differential operators, are more general than those of [D-S] as they apply to a wider class of perturbations of the operator as well as to perturbations of the boundary conditions.

This paper is organized as follows: We begin Sect. 1 by presenting notation, definitions and the basic properties of the determinant function. Some of the preliminary results are well-known and are presented without proof. We conclude Sect. 1 with a proof of our main results, Theorems 1 and 2. In Sect. 2 we apply these theorems to calculate the determinants of some operators on closed manifolds as well as manifolds with boundary. In Sect. 3 we refine the results in the case of ordinary differential operators, and then we calculate the determinant of some ordinary differential operators. With our choice of examples we have tried to emphasize the possible applications of these ideas to geometry.

Much of this paper appeared in the author's doctoral thesis [Fo]. I wish to thank my advisor Raoul Bott. The influence of his never-ending enthusiasm, patience and humor, as well as his mathematical insight, extends far beyond the results of this paper.

1. Elliptic operators and determinants

In this section we introduce the subject matter at hand. We begin with a brief description of the requirements which we place on the operators we will consider. We then proceed to define the determinant of such operators and to demonstrate the basic properties of the determinant function. We conclude this section with the statement and proof of our main theorems.

1.1. Elliptic operators

We should mention that we make no attempt to find the most general class of operators such that our later analysis will go through. Rather, we will restrict ourselves to the class of linear partial differential operators, as there are some steps which do not appear to generalize, at least in a straightforward way, to pseudo-differential operators.

Let M be a smooth k-dimensional manifold, possibly with smooth $(k-1)$-dimensional boundary, and E a complex vector bundle of rank r over M. A linear map L from $C^\infty(M,E)$, C^∞ sections of E over M, to itself is a partial differential operator of order n if in any local coordinate patch U over which E is trivial and any section f with compact support in U, Lf has the form

$$Lf=\sum_{|\alpha|\leqq n} a_\alpha(x)\,D_x^\alpha$$

where $\alpha=(\alpha_1,\alpha_2,\ldots,\alpha_k)\in(\mathbf{Z}^+)$ $|\alpha|=\sum_{i=1}^{k}\alpha_i$, $D_x^\alpha=(-i)^{|\alpha|}\dfrac{\partial}{\partial x_1^{\alpha_1}}\cdots\dfrac{\partial}{\partial x_k^{\alpha_k}}$ and $a_\alpha(x)$ is an $r\times r$ complex matrix valued function.

The symbol of L, denoted by $\sigma(L)$, is defined by

$$\sigma(L)=\sum_{j\leqq n}\sigma_j(x,\xi)=\sum_{j\leqq n}\Big(\sum_{|\alpha|=j}a_\alpha(x)\,\xi^\alpha\Big)$$

where $\xi^\alpha=\xi^{\alpha_1}\dots\xi^{\alpha_k}$. The symbol of L depends on the choice of coordinates, but σ_n has an invariant interpretation as a map from the cosphere bundle of M (the sphere bundle in the cotangent bundle of M) into the space of linear transformations on the fibers of E. That is, as a bundle map

$$\sigma_n\colon\ \pi^*E\to\pi^*E$$

where $\pi\colon S^*M\to M$ is the natural projection. This corresponds to thinking of x as a coordinate on M and ξ as a coordinate in the fiber of the cotangent bundle.

Suppose M is closed. Then L is elliptic if σ_n is invertible. We will also require that there be a ray $\{re^{2\pi i\theta}\colon\ r\in R^+\}\subset\mathbf{C}$ containing no eigenvalues of $\sigma_n(x,\xi)$. This will be true whenever the spectrum of L is not the whole complex plane. Such a ray is called a ray of minimal growth. The need for such a ray of minimal growth will be made clear when we define the determinant.

If M is not closed then we must include the notion of a boundary condition in our definition of ellipticity. Let v be any vector field along the boundary of M which is transverse to the boundary at each point. We have a map T from $C^\infty(M,E)$ to $C^\infty(\partial M,nE_{\partial M})$, where nE is the direct sum of n copies of E, defined by

$$[T(f)](x)=\left(f(x),\frac{\partial}{\partial v}f(x),\dots,\frac{\partial^{n-1}}{\partial v^{n-1}}f(x)\right)\quad\text{for } x\in\partial M.$$

A boundary condition is a linear subspace $B\subset C^\infty(\partial M,nE)$. We say f satisfies B if $T(f)\in B$.

We will not be precise in our definition of elliptic boundary conditions for L. Our condition for B is only that L_B (L restricted to functions which satisfy the boundary condition B) have a well behaved resolvent $(L_B-\lambda)^{-1}$. To be accurate, by L_B we mean the closure in L^2 of L acting on C^∞ sections of E which satisfy B.

We present three classes of boundary conditions (no others will appear in this paper) which satisfy this criterion.

1. Normal boundary conditions (in the sense of Agmon [Ag]): These boundary conditions are local and include the standard elliptic boundary conditions, i.e. where L is the Laplacian and B is the Dirichlet (f is zero along ∂M) or Neumann ($\partial_v f$ is zero along ∂M) boundary conditions. The precise definition is as follows. Suppose the boundary condition B is given by

$$B_j(f)=0\qquad j=1,\dots,p,$$

where

$$B_j=\sum_{k=0}^{n_j}b_j^k\left(x,\frac{\partial}{\partial x}\right)\left(\frac{\partial}{\partial v}\right)^{n_j-k}$$

and b_j^k is an $r_j \times m$ system of differential operators on ∂M of order $\leqq k$ and $n_1 < n_2 < \ldots < n_p < n$. If we write

$$\sigma(B_j)\left(x', \xi', \frac{\partial}{\partial v}\right) = \sum_{k=0}^{n_j} \sigma_k(b_j^k)(x', \xi')\left(\frac{\partial}{\partial v}\right)^{n_j - k}.$$

Then the condition is that there exist a ray $\rho = \{re^{2\pi i\theta}: r \in \mathbf{R}^+\}$ such that

(a) $\sigma_n(x, \xi)$ has no eigenvalues along ρ.

(b) For each x, each $(\xi, \lambda) \neq 0$ with $\lambda \in \rho$, and each $g = (g_1, \ldots, g_p)$ there is a unique solution to

- $\sigma_n(L)\left(x, \xi, \frac{\partial}{\partial t}\right) u = \lambda$
- $\lim_{t \to \infty} u(t) = 0$
- $\sigma(B_j)\left(x, \xi, \frac{\partial}{\partial t}\right) u = g_j \quad$ for $j = 1, \ldots, p$.

Seeley has shown in [Se3] that with these restrictions the spectrum is discrete and all of the necessary analysis goes through.

2. Periodic boundary conditions: The simplest example is $\frac{d}{dx}$ acting on functions on the unit interval which satisfy the boundary condition $f(0) = f(1)$. This is equivalent, in a sense to be made precise later, to $\frac{d}{dx}$ acting on functions on the circle of unit length so that we know the resolvent is well behaved (see [Se2]). This type of boundary condition is discussed in greater detail in Sect. 2.1.

3. Atiyah-Patodi-Singer boundary conditions: Atiyah, Patodi and Singer have shown in [A-P-S] that we get a resolvent satisfying all necessary estimates when considering the Dirac operator on a spin manifold M with boundary if we take the boundary condition that $f|_{\partial M}$ lie in the span of the eigenfunctions of the Dirac operator on ∂M with positive eigenvalue.

Given an elliptic operator and one such boundary condition, we can now define its determinant.

1.2. The determinant of an elliptic operator

The motivation for our definition is very simple. If L is a finite dimensional square matrix with eigenvalues λ_i, then

$$-\log \operatorname{Det} L = \frac{d}{ds}\left(\sum_i \lambda_i^{-s}\right)\Big|_{s=0} = \frac{d}{ds} \operatorname{Tr}(L^{-s})\Big|_{s=0}.$$

We take this to be our definition of the determinant of a differential operator. Of course, if we let $\{\lambda_i\}$ be the set of eigenvalues of a differential operator, then the sum will not converge near $s = 0$. However, $\operatorname{Tr}(L^{-s})$ does converge for $\operatorname{Re}(s)$

(the real part of s) large enough. Theorems of Seeley insure that this function can be analytically continued to a meromorphic function of the entire complex plane which is holomorphic at $s=0$. We can therefore define $\operatorname{Det} L$ by the same formula as in the finite dimensional case.

Now let us proceed more precisely. Suppose L is an elliptic operator and B an elliptic boundary condition such that $\operatorname{sp}(L_B)$ (the spectrum of L restricted to B) misses some cone $(\theta-\varepsilon<\arg\lambda<\theta+\varepsilon)$ in $\mathbf{C}$. Let $\{re^{i\theta}: r\in R^+\}$ be a ray in this cone. Then, in analogy with Cauchy's Integral formula

$$f(z)=-\frac{1}{2\pi i}\int_\Gamma \frac{f(\lambda)}{\lambda-z}\,d\lambda$$

where f is analytic in the interior of the closed curve Γ and z is inside Γ, we define

$$f(L)=-\frac{1}{2\pi i}\int_\Gamma f(\lambda)(L-\lambda)^{-1}\,d\lambda.$$

In this formula we have written L for L_B and Γ for $\Gamma_1\cup\Gamma_2\cup\Gamma_3$ where

- $\Gamma_1=\{re^{i\theta}: r \text{ goes from } +\infty \text{ to } \varepsilon\}$
- $\Gamma_2=\{\varepsilon e^{i\phi}: \phi \text{ goes from } \theta \text{ to } \theta+2\pi\}$
- $\Gamma_3=\{re^{i(\theta-2\pi)}: r \text{ goes from } \varepsilon \text{ to } +\infty\}$.

Then Γ is a curve which "contains" the spectrum of L and $(L-\lambda)^{-1}$ is analytic and bounded in a neighborhood of Γ. We are mainly interested in

$$L^{-s}=-\frac{1}{2\pi i}\int_\Gamma \lambda^{-s}(L-\lambda)^{-1}\,d\lambda.$$

For $\operatorname{Re}(s)>0$ this gives a well-defined operator. This analytic family of operators has an extension to the entire s-plane satisfying

1. $L^{z_1}L^{z_2}=L^{z_1+z_2}$.
2. $L^k=$ the usual definition of L^k for $k=1,2,3,\ldots$.
3. If L has no zero modes, then $L^k=$ the usual definition of L^k for all $k\in Z$. Otherwise $L^0=I-P_0$ where I is the identity map and P_0 is the projection onto the zero modes of L.

For $\operatorname{Re}(s)$ large, the operator L^{-s} has a continuous kernel $L^{-s}(x,y)$, where $L^{-s}(x,y)$ is an element of $\operatorname{End}(x,y)$, and therefore is of trace class. $\operatorname{Tr}(L^{-s})$ has an analytic extension to the complex plane as a meromorphic function. Furthermore, given our assumptions on L and B, $\operatorname{Tr}(L^{-s})$ is holomorphic at $s=0$ [Se2, 3]. We can therefore define $-\log\operatorname{Det} L$ to be $\left.\frac{d}{ds}\right|_{s=0}\operatorname{Tr}(L^{-s})$. In analogy with the zeta functions of number theory, $\operatorname{Tr}(L^{-s})=\sum_i \lambda_i^{-s}$ is denoted by $\zeta_L(s)$.

We note here that for any operator L we will use the notation $\operatorname{Det} L$ whenever we are defining the determinant of L by a normalization process, and $\det L$ if the determinant of L is a well-defined finite number without any normalization.

Another common approach is to work with the heat kernel. If the eigenvalues of L, $\{\lambda_i\}$, all lie in the cone $(\frac{\pi}{2}-\varepsilon<\arg\lambda<-\frac{\pi}{2}+\varepsilon)$ then one can consider the fundamental solution e^{-tL} of the operator $\frac{\partial}{\partial t}+L$. The operator e^{-tL} has a smooth kernel $k(t,x,y)$ for $t>0$, with trace $\sum_i e^{-t\lambda_i}$. Then

$$L^{-s}=\frac{1}{\Gamma(s)}\int_0^\infty t^{s-1}e^{-tL}\,dt$$

and therefore

$$\zeta_L(s)=\mathrm{Tr}(l^{-s})=\frac{1}{\Gamma(s)}\int_0^\infty t^{s-1}\,\mathrm{Tr}(e^{-tL})\,dt.$$

For any $\varepsilon>0$, $\frac{1}{\Gamma(s)}\int_\varepsilon^\infty t^{s-1}\,\mathrm{Tr}(e^{-tL})\,dt$ is entire holomorphic in s, so to study the poles of $\mathrm{Tr}(L^{-s})$ one need only study the asymptotics of e^{-tL} as t goes to 0. The popularity of this approach can be explained by the numerous geometric (and not so geometric) techniques available to aid in the analysis of the asymptotics.

1.3. Properties of the determinant

In this section we present the basic properties of the determinant function. These results will form the foundation of our later analysis.

We begin with theorem which shows that our definition of determinant is on the right track. For a one-parameter family of finite dimensional matrices A_t, we have

$$\frac{d}{dt}\log\det A=\frac{d}{dt}\,\mathrm{tr}\log A=\mathrm{tr}\left(\frac{d}{dt}A\right)A^{-1}.$$

The same is true for differential operators. Let L_t be a one parameter family of operators and B a boundary condition which is elliptic for each member of the family. If the L_t's have the same top order symbol, this is equivalent to saying that B is elliptic for L_0.

Proposition 1.1. *Suppose* $\left(\frac{d}{dt}L_0\right)L_0^{-1}$ *is trace class. Then, in fact,*

$$\frac{d}{dt}\log\mathrm{Det}\,L_0=\mathrm{tr}\left(\frac{d}{dt}L_0\right)L_0^{-1}$$

where we have written L_0 *for* $L_{0,B}$.

Proof. The hypothesis implies that $\left(\frac{d}{dt}L_0\right)(L_0-\lambda)^{-1}$ is trace for every λ in the resolvent set of L_0.

$$\begin{aligned}\frac{d}{dt}\log\mathrm{Det}\,L&=\frac{d}{dt}\frac{d}{ds}\bigg|_{s=0}\mathrm{tr}-\frac{1}{2\pi i}\int_\Gamma\lambda^{-s}(L-\lambda)^{-1}\,d\lambda\\&=\frac{d}{ds}\bigg|_{s=0}\mathrm{tr}\,\frac{1}{2\pi i}\int_\Gamma\lambda^{-s}(L-\lambda)^{-1}\left(\frac{d}{dt}L\right)(L-\lambda)^{-1}\,d\lambda.\end{aligned}$$

$(L-\lambda)^{-1}\left(\frac{d}{dt}L\right)(L-\lambda)^{-1}$ is trace class since $(L-\lambda)^{-1}$ is bounded along Γ. This also implies that $\mathrm{tr}(L-\lambda)^{-1}\left(\frac{d}{dt}L\right)(L-\lambda)^{-1}=\mathrm{tr}\left(\frac{d}{dT}L\right)(L-\lambda)^{-2}$. Therefore the last integral is equal to

$$\begin{aligned}\frac{d}{ds}\Big|_{s=0}&\frac{1}{2\pi i}\int_\Gamma \lambda^{-s}\,\mathrm{tr}(L-\lambda)^{-1}\left(\frac{d}{dt}L\right)(L-\lambda)^{-1}\,d\lambda\\ &=\frac{d}{ds}\Big|_{s=0}\frac{1}{2\pi i}\int_\Gamma \lambda^{-s}\,\mathrm{tr}\left(\frac{d}{dt}L\right)(L-\lambda)^{-2}\,d\lambda\\ &=\frac{d}{ds}\Big|_{s=0}\mathrm{tr}\left(\frac{d}{dt}L\right)\frac{1}{2\pi i}\int_\Gamma \lambda^{-s}(L-\lambda)^{-2}\,d\lambda.\end{aligned}$$

We can now integrate by parts, since for $s>-1$ the norm of $\lambda^{-s}(L-\lambda)^{-1}$ goes to 0 at the ends of Γ, to get

$$\begin{aligned}\frac{d}{ds}\Big|_{s=0}&\mathrm{tr}\left(\frac{d}{dt}L\right)\frac{1}{2\pi i}\int_\Gamma(-s)\,\lambda^{-s-1}(L-\lambda)^{-1}\,d\lambda\\ &=\frac{d}{ds}\Big|_{s=0}\mathrm{tr}\left[s\left(\frac{d}{dt}L\right)L^{-s-1}\right]\end{aligned}$$

$\mathrm{tr}\left(\frac{d}{dt}L\right)L^{-s-1}$ is finite for $s\leqq 0$ so this last term is equal to

$$\mathrm{tr}\left(\frac{d}{dt}L\right)L^{-1}. \qquad \square$$

If $\left(\frac{d}{dt}L_t\right)L_t^{-1}$ is trace class for each t then $(L_tL_0^{-1})-1$ is trace class for each t. This implies that $L_tL_0^{-1}$ is of "determinant class" (see [G-K]) and one can define the determinant of $L_tL_0^{-1}$ in a natural way, called the Fredholm determinant or the perturbation determinant. It is known that

$$\frac{d}{dt}\log\det(L_tL_0^{-1})=\mathrm{tr}\left(\frac{d}{dt}L_t\right)L_t^{-1}.$$

Therefore, since they are equal at $t=0$ and have the same derivative we have proven

Corollary 1.2. *If* $\left(\frac{d}{dt}L_t\right)L_t^{-1}$ *is trace class for every* t *then* $L_tL_0^{-1}$ *is of determinant class and*

$$\frac{\mathrm{Det}\,L_t}{\mathrm{Det}\,L_0}=\det(L_tL_0^{-1})$$

where the left side is defined by a functional integral, and the right side is a perturbation determinant.

Unfortunately, it is rare in important problems for $\left(\frac{d}{dt}L\right)L^{-1}$ to be trace class, but we can find a similar result for more general problems. Suppose A

and B are boundary conditions such that $L'(L_A^{-1}-L_B^{-1})$ is "trace class" (where we have written L' for $\frac{d}{dt}L$). Here we need to do some explaining since we are comparing operators which act on different spaces. However, we mean for this to be interpreted in the most obvious fashion. L_A^{-1} and L_B^{-1} have kernels which are smooth off the diagonal and have the same singularity along the diagonal. Therefore $L_A^{-1}-L_B^{-1}$ has a kernel which extends smoothly across the diagonal in the interior of M. The condition that $L'(L_A^{-1}-L_B^{-1})$ be trace class is intended to mean that

$$\int_M |\mathrm{tr}(L'(L_A^{-1}-L_B^{-1})(x,y)|_{y=x})|\,dx<\infty$$

where L' acts on the kernel as a function of x. We emphasize that this is a restriction on the behavior of the data at the boundary. If this condition, and two others which are implied by the first under normal circumstances, is satisfied, we can state a result analogous to Proposition 1.1.

Proposition 1.3. *If L_t is a one-parameter family of operators, and A and B are two boundary conditions elliptic for the L_t such that for each t*

1. $\int_M |\mathrm{tr}\, L'(L_A^{-1}-L_B^{-1})|<\infty$
2. $\int_{\mathrm{supp}\,L'} |\mathrm{tr}(L_A^{-1}-L_B^{-1})|<\infty$
3. *$L'L_B^{-1}$ is bounded (or, equivalently, $L'L_A^{-1}$ is bounded)*

then for each value of t

$$\frac{d}{dt}\log\frac{\mathrm{Det}\,L_A}{\mathrm{Det}\,L_B}=\mathrm{tr}\,L'(L_a^{-1}-L_B^{-1})$$

where we must use a single ray of minimal growth to define both $\mathrm{Det}\,L_A$ and $\mathrm{Det}\,L_B$. (Remark: Unless L' or L blow up at the boundary, $0\leqq(\mathrm{order}\,L')\leqq(\mathrm{order}\,L)$ is sufficient to imply (3) and for (1) to imply (2).)

Proof.

$$\begin{aligned}\frac{d}{dt}\log\frac{\mathrm{Det}\,L_A}{\mathrm{Det}\,L_B}&=-\frac{d}{dt}\frac{d}{ds}\bigg|_{s=0}\mathrm{tr}\,\frac{1}{2\pi i}\int_\Gamma\lambda^{-s}[(L_a-\lambda)^{-1}-(L_B-\lambda)^{-1}]\,d\lambda\\&=\frac{d}{ds}\,\mathrm{tr}\,\frac{1}{2\pi i}\int_\Gamma\lambda^{-s}((L_A-\lambda)^{-1}L'(L_A-\lambda)^{-1}-(L_B-\lambda)^{-1})\\&\quad+((L_A-\lambda)^{-1}-(L_B-\lambda)^{-1}L'(L_B-\lambda)^{-1})\,d\lambda.\end{aligned}$$

Now $L'((L_A-\lambda)^{-1}-(L_B-\lambda)^{-1})$ is trace class and $(L_A-\lambda)^{-1}$ is bounded so we can reverse the order of these factors. Similarly, we can reverse the order of the factors in the second term since $L'(L_B-\lambda)^{-1}$ is bounded and $((L_A-\lambda)^{-1}-(L_B-\lambda)^{-1})$ is trace class on the support of L'. Thus we get

$$\frac{d}{ds}\bigg|_{s=0}\mathrm{tr}\,L'\frac{1}{2\pi i}\int_\Gamma\lambda^{-s}((L_A-\lambda)^{-2}-(L_B-\lambda)^{-2})\,d\lambda.$$

We integrate by parts and follow the same argument that we used to finish the proof of Theorem 1 to find that this integral is equal to

$$\operatorname{tr} L'(L_A^{-1} - L_B^{-1}). \qquad \square$$

1.4. Boundary value problems

In this section we exploit the duality between two types of boundary value problems to lay the groundwork for our main theorems. With all notation as in Sect. 1.1, let L be an elliptic operator with boundary condition B. Recall that B is a linear subspace of $C^\infty(\partial M, nE)$. Let β be any linear complement to B. Then for any section f of E we have projections of $T(f)$, the boundary values of f, onto B and β, which we denote by $T_B(f)$ and $T_\beta(f)$. So that, in particular, $T(f) = T_B(f) + T_\beta(f)$ and $T(f) \in B \leftrightarrow T_\beta(f) = 0$.

The two types of boundary value problems referred to above are

1. Given $g \in C^\infty(M, E)$, find $f \in C^\infty(M, E)$ such that

$$Lf = g$$
$$T(f) \in B.$$

2. Given $h \in C^\infty(\partial M, nE)$, find $f \in C^\infty(M, E)$ such that

$$Lf = 0$$
$$T_\beta(f) = h.$$

In either case, such an f will be unique as long as there is no non-trivial solution to the homogeneous problem: $Lf = 0$, $T_\beta(f) = 0$. We will now assume this.

The standard tool for problems of type (1) is the Green's function for L and B, which is defined as the unique function

$$L^{-1}\colon\ M \times M \to \operatorname{End}(E)$$

(where $L^{-1}(x, y)$ is a linear map from the fiber of E over y to the fiber of E over x), satisfying

1. For all y, $L_x L^{-1}(x, y) = \delta(x, y)$, where δ is the Dirac delta function and the symbol L_x means that we fix y and let L act on L^{-1} as a function of x.

2. For all y, $T_x L^{-1}(x, y) \in B$, where T_x means that we fix y and consider the boundary values of $L^{-1}(x, y)$ as a function of x.

To be more precise, we can write $L^{-1}(x, y) = \sum_i v_i(x)\, u_i^*(y)$ where $v_i(x) \in E_x$ and $u_i^*(y) \in E_y^*$ (the space of linear maps from E_y to $\mathbf{C}$). If H is any operator acting on sections of E then $H_x L^{-1}(x, y) = \sum_i H(v_i(x))\, u_i^*(y)$.

The standard tool for the problems of type (2) is what we will call the Poisson map of L and B. The Poisson map is a linear map

$$P\colon\ \beta \to C(M, E)$$

such that for all $h \in \beta$, $P(h)$ is a solution of L and $T_\beta[P(h)] = h$. The map P can be represented by a kernel $P(x, y)$ for $x \in M$ and $y \in \partial M$, so that

$$[P(h)](x) = \int_{\partial M} P(x, y)\, h(y)\, dy.$$

In our subsequent discussion we will express $P(x, y)$ in terms of the kernel $L^{-1}(x, y)$, thereby arriving at the above representation.

The proofs in the next section will rest upon the idea that there is a very close relationship between the Green's function of L and B and the Poisson map of L and B. The fact which relates them is Lagrange's Identity (or at least it is called Lagrange's Identity in the case of ordinary differential operators). To state Lagrange's Identity we need to choose a Hermitian metric on the bundle E, that is, a Hermitian inner product $\langle\ ,\ \rangle$ on each fiber of E. This induces an L^2 inner product $(\ ,\ )$ on sections of E defined by

$$(f, g) = \int_M \langle f(x), g(x) \rangle\, dx.$$

Let L^* be the adjoint of L. L^* is defined as the unique differential operator such that for any two sections f, g with support in the interior of M

$$(Lf, g) - (f, L^* g) = 0.$$

There is an Hermitian inner product on the space of boundary values $G(\ ,\ )$ (called the Green's form in [Pa]), such that for any two C^∞ sections f, g of E

$$(Lf, g) - (f, L^* g) = G(Tf, Tg).$$

Note that G is linear in Tf and conjugate linear in Tg. We will sometimes write $G(f, g)$ for $G(Tf, Tg)$ as there is no ambiguity.

At this point we pause to make an important comment. It will be necessary for later purposes to extend the definition of the inner products $\langle\ ,\ \rangle$ and $G(\ ,\ )$ to apply to linear maps as well as to sections. For any linear maps C from vector space U to vector space V and D from W to V, where U, V and W are all equipped with inner products, define $\langle C, D \rangle$ to be $D^* C$, where D^* is the adjoint map to D. Thus $\langle C, D \rangle$ is a map from U to W. Note that this agrees with the usual definition if we think of a vector in U as a linear map from C to U. Similarly, we extend the domain of $G(\ ,\ )$. In particular, if we again write $L^{-1}(x, y) = \sum v_i(x)\, u_i^*(y)$ then

$$\begin{aligned} G(f(x), L^{-1}(x, y)) &= G(f(x), \textstyle\sum v_i(x)\, u_i^*(y)) \\ &= \textstyle\sum u_i(y)\, G(f(x), v_i(x)) \end{aligned}$$

where we use the Hermitian inner product to define the section $u_i(y)$ which is dual to $u_i^*(y)$. That is, $u_i(y)$ is the unique element in E_y such that for all $\gamma \in E_y$, $\langle \gamma, u_i(y) \rangle = u_i^*(y)\, \gamma$.

For any boundary condition B of L, define the adjoint boundary condition B^* by

$$\text{for } k \in C(\partial M, nE), \qquad k \in B^* \leftrightarrow G(h, k) = 0 \quad \text{for all } h \in B.$$

The purpose of introducing the adjoint boundary condition is made clear by the following observation. We note that Lagrange's identity implies that for all f and g such that $T(f)\in B$ and $T(g)\in B^*$,

$$\int_M \langle Lf, g\rangle = \int_M \langle f, L^* g\rangle.$$

In particular, let $f=L^{-1}(x, y)$, the Green's function for L and B, and $g=(L^*)^{-1}(x, y)$ the Green's function for L^* and B^*. Then, for all y and z, the left-hand side of the above equality is equal to

$$\begin{aligned}\int_M \langle L_x L^{-1}(x, y), (L^*)^{-1}(x, z)\rangle &= \int_M \langle \delta(x, y), (L^*)^{-1}(x, z)\rangle \\ &= \int_M ((L^*)^{-1}(x, z))^* \delta(x, y) = ((L^*)^{-1}(y, z))^*.\end{aligned}$$

The right-hand side is equal to

$$\begin{aligned}\int_M \langle L^{-1}(x, y), L^*(L^*)^{-1}(x, z)\rangle &= \int_M \langle L^{-1}(x, y), \delta(x, z)\rangle \\ &= \int_M (\delta(x, z))^* L^{-1}(x, y) = L^{-1}(z, y).\end{aligned}$$

Therefore, $L^{-1}(z, y)=((L^*)^{-1}(y, z))^*$ for all y and z. That is $(L^{-1}(z, y))^* = (L^*)^{-1}(y, z)$.

Before arriving at our final result of this section, we will make one more remark. Let $h\in C(\partial M, nE)$ and $k\in B^*$. Then

$$G(h, k)=G(T_B(h)+T_\beta(h), k)=G(T_B(h), k)+G(T_\beta(h), k)=G(T_\beta(h), k).$$

Now use Lagrange's identity with f a solution to L and $g(x)=(L^*)^{-1}(x, y)$, the Green's function of L^* and B^*. Using the previous remark, we have for all $y\in M$

$$\int_M \langle Lf, (L^*)^{-1}(x, y)\rangle - \langle f, L^*(L^*)^{-1}(x, y)\rangle = G(T_\beta f, T_y(L^*)^{-1}(x, y)).$$

The left-hand side is equal to $-f(y)$. Therefore, switching x and y,

$$f(x)=-G(T_{\text{beta}} f, T_y(L^*)^{-1}(y, x))=-G(T_\beta f, T_y L^{-1}(x, y)).$$

This provides us with a solution to boundary problem (2) in terms of the Green's function of L and B: To find f such that $Lf=0$ and $T_\beta(f)=h$, let $f(x) = -G(h, T_y L^{-1}(x, y))$. However, we already know that the solution is given by $f(x)=[P(h)](x)$, where P is the Poisson map of L and B, and this yields the promised relationship between the Poisson map and the Green's function:

$$[P(h)](x)=-G(h, T_y L^{-1}(x, y)).$$

With these techniques in hand, we can now tackle the problem of determinants.

1.5. The main theorems

In this section we see how the variation of the determinant can be expressed in terms of an operator acting on sections over the boundary. Let L be an elliptic operator, and A and B two boundary conditions with linear complements α and β. Suppose further that L_B, L restricted to functions which satisfy condition B, has no null space. Then there is a natural map Φ_L: $\beta \to \alpha$ given by $\Phi_L = T_\alpha P_\beta$ where P_β is the Poisson map for L and B. As for the nature of this map, Hörmander has shown in [Hö] that the map $TP_\beta T_\beta$ acting on the entire space of boundary values is pseudo-differential. We are just restricting this map to β and projecting to α. If A is a local boundary condition, then we can choose α so that this projection is a differential operator.

Note that Φ_L is constructed out of the solution space of L and the boundary conditions, but does not at all involve the space of sections satisfying a particular boundary condition, unlike L_A or L_B. Yet we have the following theorem.

Theorem 1. *Let L_t be a one-parameter family of operators such that all hypotheses of Proposition 1.3 are satisfied for all t. Then*

$$\frac{d}{dt} \log \frac{\operatorname{Det} L_{t,A}}{\operatorname{Det} L_{t,B}} = \frac{d}{dt} \log \det \Phi_{L_t}.$$

In particular, the right hand side is finite.

We begin with a lemma that will allow the proof to go smoother.

Lemma 1.4. *1.* $\dfrac{d}{dt} P_\beta = -L_B^{-1} L' P_\beta$.

2. $P_\alpha T_\alpha (-L_B^{-1}) = L_A^{-1} - L_B^{-1}$.

Proof. 1. Suppose, for some $h \in \beta$, that $P_{t,\beta} h = f_t$. That is, $L_t f_t = 0$ and $P_{t,\beta} f_t = h$. Then, differentiating these equalities yields $Lf' = -L'f$ and $P_\beta f' = 0$. Thus $f' = -L_B^{-1} L' f$, or equivalently $P_\beta' h = -L_B^{-1} L' P_\beta h$. This is true for every $h \in \beta$ so $P_\beta' = -L_B^{-1} L' P_\beta$.

2. $P_\alpha Y_\alpha(-L_B^{-1})$ is the unique smooth solution of L with the same α-boundary values as L_B^{-1}. But $T_\alpha L_A^{-1} = 0$ implies that $L_A^{-1} - L_B^{-1}$ satisfies the criteria.

Proof of the theorem. The right-hand side is equal to

$$\begin{aligned}
\frac{d}{dt} \log \operatorname{Det}(T_\alpha P_\beta) &= \operatorname{tr}(T_\alpha P_\beta)^{-1} \frac{d}{dt}(T_\alpha P_\beta)\Big|_\beta \\
&= \operatorname{tr} T_\beta P_\alpha T_\alpha \left(\frac{d}{dt} P_\beta\right)\Big|_\beta \\
&= \operatorname{tr} T_\beta P_\alpha T_\alpha (-L_B^{-1}) L' P_\beta|_\beta \\
&= \operatorname{tr} T_\beta (L_A^{-1} - L_B^{-1}) L' P_\beta|_\beta \\
&= \operatorname{tr} P_\beta T_\beta (L_A^{-1} - L_B^{-1}) L'|_{\operatorname{Ker} L} \\
&= \operatorname{tr} (L_A^{-1} - L_B^{-1}) L'|_{\operatorname{Ker} L}.
\end{aligned}$$

Now let $\{f_i\}$ be an L^2 orthonormal basis for the solution space of L. $(L_A^{-1} - L_B^{-1})(x, y)$ is a solution of L as a function of x so it can be expressed as $\sum f_i(x) g_i^*$ where g_i^* is an element of the dual space of $C^\infty(M, E)$. Then

$$\operatorname{tr}(L_A^{-1} - L_B^{-1}) L'|_{\operatorname{Ker} L} = \sum_i g^*(L' f_i)$$

$$= \operatorname{tr} L'(L_A^{-1} - L_B^{-1}) = \frac{d}{dt} \log \frac{\operatorname{Det} L_A}{\operatorname{Det} L_B}$$

where the last equality follows from Proposition 1.3. □

Corollary 1.5. *With the hypotheses of the preceeding theorem*

$$\frac{\operatorname{Det} L_{0,A}}{\operatorname{Det} L_{0,B}} \frac{\operatorname{Det} L_{t,B}}{\operatorname{Det} L_{t,A}} = \operatorname{Det} \Phi_{L_t}^{-1} \Phi_{L_0} \colon \beta \to \beta.$$

Proof. This is just the integrated version of the statement in the previous theorem.

Theorem 1 concerns variations of the elliptic operator L. We can make a similar statement concerning variations of the boundary conditions. Let U be any automorphism of E (which need not preserve M). That is, U can be thought of as a diffeomorphism θ of M which maps the boundary to itself and a linear map F_x from E_x to $E_{\theta(x)}$ for each $x \in M$. Then U maps an elliptic boundary condition A to another elliptic boundary condition UA. That is, $f(x)$ satisfies UA if and only if $U^{-1} f$ satisfies A (where $(U^{-1} f)(x) = F_x^{-1}[f(\theta(x))]$). There is an automorphism of nE over ∂M, which we also call U (as our meaning will be clear from the context) so that $T(Uf) = U(Tf)$. Thus, if we think of A as a subspace of $C^\infty(\partial M, nE)$, UA is the space of sections $UA = \{Ua \colon a \in A\}$. Let U_t be a one-parameter family of automorphisms of E with the corresponding automorphisms of $nE|\partial M$. We wish to study

$$\frac{\operatorname{Det} L_{U_t A}}{\operatorname{Det} L_{U_t B}}.$$

The main tool is again the map

$$\Phi_{t,L} \colon U_t \beta \to \operatorname{Ker} L \to U_t \alpha$$

where $\Phi_{t,L} = T_{U_t \alpha} P_{U_t \beta}$. Unfortunately, as t varies, the Φ's map between varying spaces. To remedy this situation we consider instead the family of maps

$$U^{-1} \Phi_{t,L} U \colon \beta \to \alpha$$

where U and U^{-1} act on α and β as defined earlier. Now, in analogy with Theorem 1

Theorem 2. *Let $L_t = U_t^{-1} L U_t$. If L_t and the boundary conditions A and B satisfy the hypotheses of Proposition 1.3 then*

$$\frac{d}{dt} \log \frac{\operatorname{Det} L_{U_t A}}{\operatorname{Det} L_{U_t B}} = \frac{d}{dt} \log \det U^{-1} \Phi_{t,L} U.$$

Proof. First we simplify the left-hand side. We begin by noticing that

$$\operatorname{Det} L_{UB} = \operatorname{Det}(U^{-1} L U)_B$$

and similarly for the boundary condition A. (In what follows, we will write U for U_t and a prime will always denote $\frac{d}{dt}$.) Therefore

$$\begin{aligned}
\frac{d}{dt} \log \frac{\operatorname{Det} L_{UA}}{\operatorname{Det} L_{UB}} &= \frac{d}{dt} \log \frac{\operatorname{Det}(U^{-1} L U)_A}{\operatorname{Det}(U^{-1} L U)_B} \\
&= \operatorname{tr}\left[\frac{d}{dt}(U^{-1} L U)\right] [(U^{-1} L U)_A^{-1} - (U^{-1} L U)_B^{-1}] \\
&= \operatorname{tr}[(-U^{-1} U' U^{-1} L U) + (U^{-1} L U')] [(U^{-1} L_{UA}^{-1} U) - (U^{-1} L_{UB}^{-1} U)] \\
&= \operatorname{tr}[(-U^{-1} U' U^{-1} L (L_{UA}^{-1} - L_{UB}^{-1} U) + (U^{-1} L U' U^{-1} (L_{UA}^{-1} - L_{UB}^{-1}) U)]
\end{aligned}$$

$(L_{UA}^{-1} - L_{UB}^{-1})$ is a solution of L so the first term is 0. The second term is equal to

$$\operatorname{tr} L U' U^{-1} (L_{UA}^{-1} - L_{UB}^{-1}).$$

Now we use some facts about Green's functions. Let $\{f_i\}$ be an L^2 basis of the solution space of L. We know that $(L_{UA}^{-1} - L_{UB}^{-1})(x, y)$ is in this space as a function of x so it can be written as $\sum_i f_i(x) g_i^*(y)$, where g_i^* is a map from sections of E to $\mathbf{C}$ given by

$$h \to \int_M \langle h(y), g_i(y) \rangle \, dy$$

for some section $g(y)$, but $L^*(L_{UA} - L_{UB})^* = 0$ so $L^* g_i = 0$. Therefore

$$\begin{aligned}
&\operatorname{tr} L U' U^{-1} (L_{UA}^{-1} - L_{UB}^{-1}) \\
&\quad = \sum_i \int_M \langle L U' U^{-1} f_i(x), g_i(x) \rangle \, dx \\
&\quad = \sum_i \int_M \langle U' U^{-1} f_i(x), L^* g_i(x) \rangle + \sum_i G(U' U^{-1} f_i(x), g_i(x)) \\
&\qquad \text{(Using Lagrange's identity)} \\
&\quad = \sum_i G(U' U^{-1} f_i(x), g_i(x)).
\end{aligned}$$

Now we will simplify the right-hand side to show that it is equal to the same expression. We begin with a Lemma.

Lemma 1.6. *1.* $(U^{-1} T_{U\alpha})' = U^{-1} T_{U\alpha} U' U^{-1}$.

2. $(P_{U\beta} U)' = P_{U\beta} T_{U\beta} U' U^{-1} P_{U\beta} U$.

Proof. 1. $U^{-1} T_{U\alpha} = T_{U\alpha} U^{-1}$. Therefore

$$(U^{-1} T_{U\alpha})' = -T_\alpha U^{-1} U' U^{-1} = -U^{-1} T_{U\alpha} U' U^{-1}.$$

2. Suppose $f=[P_{U\beta}U]b$ for some $b\in\beta$, so that $Lf=0$ and $U^{-1}T_{U\beta}f=T_\beta U^{-1}f$. Differentiating with respect to t gives $Lf'=0$ and $T_\beta U^{-1}f'=T_\beta U^{-1}U'U^{-1}f$. Therefore

$$f'=[U^{-1}P_\beta T_\beta U^{-1}U'U^{-1}P_{U\beta}U]b=[P_{U\beta}T_{U\beta}U'U^{-1}P_{U\beta}U]b$$

and thus

$$(P_{U\beta}U)'=[P_{U\beta}T_{U\beta}U'U^{-1}P_{U\beta}U].$$

Continuation of the proof of the theorem:

$$\begin{aligned}\frac{d}{dt}\log\operatorname{Det}U^{-1}\Phi_{t,L}U&=\frac{d}{dt}\operatorname{tr}\log(U^{-1}T_{U\alpha}P_{U\beta}U)\\&=\operatorname{tr}(U^{-1}T_{U\alpha}P_{U\beta}U)'(U^{-1}T_{U\alpha}P_{U\beta}U)^{-1}\\&=\operatorname{tr}((U^{-1}T_{U\alpha})'P_{U\beta}U+U^{-1}T_{U\alpha}(P_{U\beta}U)')U^{-1}T_{U\beta}P_{U\alpha}U|_\alpha\\&=\operatorname{tr}(U^{-1}T_{U\alpha}U'U^{-1}P_{U\alpha}U+U^{-1}T_{U\alpha}P_{U\beta}T_{U\beta}U'U^{-1}P_{U\alpha}U)|_\alpha\\&=\operatorname{tr}(-T_{U\alpha}U'U^{-1}P_{U\alpha}+T_{U\alpha}P_{U\beta}T_{U\beta}U'U^{-1}P_{U\alpha})|_{U\alpha}\\&=\operatorname{tr}(-P_{U\alpha}T_{U\alpha}U^{-1}+P_{U\beta}T_{U\beta}U'U^{-1})|_{\operatorname{Ker}L}.\end{aligned}$$

Now recall that for any boundary condition B and any b in the chosen complement β of B, $[P_\beta b](x)=-G(b,T_y(L_B^{-1}(x,y))^*)$. Therefore

$$\begin{aligned}(P_\beta T_\beta U'U^{-1})f&=-G(T_\beta U'U^{-1}f,T_y(L_B^{-1}(x,y))^*)\\&=-G(TU'U^{-1}f,T_y(L_B^{-1}(x,y))^*).\end{aligned}$$

Thus, if $\{f_i\}$ is the same basis of $\operatorname{Ker}L$ as before

$$\begin{aligned}\operatorname{tr}(-P_{U\alpha}T_{U\alpha}U'U^{-1}+P_{U\beta}T_{U\beta}U'U^{-1})&=\sum_i\langle R(TU'U^{-1}f_i,T_y(L_{UA}^{-1}-L_{UB}^{-1})^*),f_i\rangle\\&=\sum_i\langle G(U'U^{-1}f_i,T_y\sum_k(f_k(x)g_k^*(y))^*),f_i\rangle\\&=\sum_{i,k}\langle f_k(x)G(TU'U^{-1}f_i,Tg_k),f_i(x)\rangle\\&=\sum_i G(TU'U^{-1}f_i,Tg_i).\end{aligned}$$

□

Integrating with respect to t and the exponentiating the identity stated in the theorem gives the following formula.

Corollary 1.7.

$$\frac{\operatorname{Det}L_{U_tA}}{\operatorname{Det}L_{U_0A}}\frac{\operatorname{Det}L_{U_0B}}{\operatorname{Det}L_{U_tB}}=\det U_0^{-1}(\Phi_{0,L})^{-1}U_0U_t^{-1}\Phi_{t,L}U_t.$$

We will now use these theorems to calculate some determinants.

2. Applications

2.1. Closed manifolds

Here we describe a technique for studying the determinant of an elliptic operator on a closed manifold. This technique also serves to provide many interesting examples of operators and pairs of boundary conditions satisfying the hypotheses of Proposition 1.4.

Suppose L_t is a one-parameter family of elliptic operators of order n acting on sections of a vector bundle E over a closed orientable manifold M. Let N be a smooth codimension-one submanifold. We get a manifold $\tilde{M}$ with boundary by "cutting" M along N ($\tilde{M}$ might not be connected). We then have a vector bundle $\tilde{E}$ over $\tilde{M}$, which has boundary $N_1 \cup N_2$, where N_1 and N_2 are copies of N. There is an obvious map π: $\tilde{E} \to E$, mapping both $\tilde{E}_{N_1}$ and $\tilde{E}_{N_2}$ to E_N, which projects to a map, which we shall also call π, from $\tilde{M}$ to M. Given a section f of E over M we get a section $\tilde{f} = \pi^* f$ of $\tilde{E}$ over $\tilde{M}$. We can define a boundary condition A by declaring that $\tilde{f} \in \Gamma(\tilde{M}, \tilde{E})$ satisfies A if $\tilde{f} = \pi^* f$ for some C^{n-1} section f of E. In other words, if v is the unit inward normal to $N_1 \cup N_2$ and ϕ is the "identity" map from N_1 to N_2, $\tilde{f}$ satisfies A if and only if

$$\frac{\partial^k}{\partial v^k} \tilde{f}(x) = (-1)^k \frac{\partial^k}{\partial v^k} \tilde{f}(\phi x) \quad \text{for all } x \in N_1.$$

More generally, we can let Ψ be any vector bundle map from $\tilde{E}_{N_1}$ to $\tilde{E}_{N_2}$. Then by "glueing" according to Ψ, we get a vector bundle E_Ψ over a closed manifold M_Ψ. Suppose each L_t, which is defined over $M_\Psi - N$ extends smoothly to all of M_Ψ. We can define a boundary condition B on sections of $\tilde{E}$ by requiring that $\tilde{f} \in \Gamma(\tilde{M}, \tilde{E})$ be equal to $\pi^* f$ for some C^{n-1} section of E_Ψ over M_Ψ.

If G_t and $G_{\Psi,t}$ are the Green's functions of L_t over M and M_Ψ, then $G_{t,A} = \pi^* G_t$ and $G_{t,B} = \pi^* G_{\Psi,t}$. The same relationship holds among the heat kernels, if they exist. This implies that A and B are, in fact, elliptic boundary conditions, and that $G_{t,A}$ and $G_{t,B}$ have the same singularity at the boundary of $\tilde{M}$ as they do in the interior (since they are pulled back from a manifold with no boundary). Therefore $G_{t,A} - G_{t,B}$ is smooth on the closure of $\tilde{M}$ and we can apply Theorems 1 and 2. Furthermore, for each t, $\operatorname{Det} L_{t,A} = \operatorname{Det} L_t|_M$ and $\operatorname{Det} L_{t,B} = \operatorname{Det} L_t|_{M_\Psi}$.

We now present some applications of these ideas.

2.1.1. Complex analytic torsion of the torus and Kroneker's Second Limit formula. In [R-S2] Ray and Singer define an analytic torsion for complex manifolds. It is defined relative to a flat unitary bundle (which is given by a unitary representation of the fundamental group). The torsion for a Riemann surface is just the square root of the determinant of the Laplacian acting on sections of the bundle. We will calculate the torsion of a complex torus with the techniques of the preceding section. We will then relate this determinant to an Eisenstein series, thereby providing a new method of summing such series. The formula for this sum is known as Kroneker's Second Limit formula.

Let M be the flat torus given by $\mathbf{C}/\Gamma$, where $\Gamma=\pi_1(M)$ is generated by 1 and τ (with $\mathrm{Im}(\tau)>0$, where $\mathrm{Im}(\tau)$ is the imaginary part of τ). Since $\pi_1(M)$ is abelian it is sufficient to consider 1-dimension representations. Let E be the flat unitary line bundle given by the representation

$$\chi(m\tau+n)=e^{2\pi i(m\mu+n\nu)} \qquad m,n\in\mathbf{Z}.$$

Let Δ be the Laplacian acting on sections of E. We will prove

Proposition 2.1. *With all notation as above*

$$\mathrm{Det}\,\Delta=e^{2\pi\,\mathrm{Im}(\tau)(-\nu^2+\nu-\frac{1}{6})}\left|\prod_{k=-\infty}^{\infty}(1-e^{2\pi i(|k|\tau-\varepsilon_k(\mu-\tau\nu))})\right|^2$$

where $\varepsilon_k=\mathrm{sign}(k+\frac{1}{2})$.

Proof. Consider first the case where $\tau=si$ is purely imaginary and $\chi_0(m\tau+n)=e^{2\pi inv}$ for $m,n\in\mathbf{Z}$. Now cut M along the curve c represented by the line segment $[0,1]$ (which is identified with $[si,si+1]$) and glue M back together by mapping the fiber at $x+si$ ($x\in\mathbf{R}$) to the fiber at $x-t$ by multiplication by $e^{2\pi i\mu}$. This results in the torus with π_1 generated by $\{1,t+si\}$ and the flat unitary bundle induced by the representation $\chi_1(m(t+si)+n)=e^{2\pi i(m\mu+n\nu)}$. We can also think not of glueing the curves back together, but rather of the cylinder $\tilde{M}$ which is obtained by cutting along c, and imposing the boundary condition

$$A\colon\ f(x+si)=f(x), \qquad \frac{\partial}{\partial v}f(x+si)=-\frac{\partial}{\partial v}f(x) \qquad \text{for } x\in[0,1]$$

to recapture the original M and the bundle induced by χ_0, and

$$B\colon\ f(x+t+si)=e^{-2\pi i\mu}f(x), \qquad \frac{\partial}{\partial v}f(x+t+si)=-e^{-2\pi i\mu}\frac{\partial}{\partial v}f(x) \qquad \text{for } x\in[0,1]$$

to recapture M with the twisting and the bundle induced by χ_1, where v is the unit inward normal along the curve c. As discussed in the introduction to this section, because these boundary conditions correspond to glueing along a submanifold, Theorem 2 can be applied.

Let $U_{r,\theta}$, with $r,\theta\in\mathbf{R}$, be a two parameter family of automorphisms of E over $\tilde{M}$, where each $U_{r,\theta}$ is the identity near $[0,1]$ and maps the fiber at $x+yi$ to the fiber at $x+r+yi$ by multiplication by $e^{2\pi i\theta}$ for y near 1. For simplicity of notation, in the future we will leave the subscripts off of the U. UA and UB are the boundary conditions

$$UA\colon\ f(x+r+si)=e^{-2\pi i\theta}f(x), \qquad \frac{\partial}{\partial v}f(x+r+si)=-e^{2\pi i\theta}\frac{\partial}{\partial v}f(x)$$

$$UB\colon\ f(x+t+r+si)=e^{-2\pi i(\theta+\mu)}f(x), \qquad \frac{\partial}{\partial v}f(x+t+r+si)=-e^{-2\pi i(\theta+\mu)}\frac{\partial}{\partial v}f(x).$$

Thus, $\mathrm{Det}\,L_{UA}$ is equal to the determinant of the Laplacian on the torus given by the lattice $\{1,\tau=r+si\}$ with the flat bundle given by the representation $(m(r+si)+n)\to\exp(2\pi i(m\theta+n\nu))$. The analogous statement is true for UB with r

replaced by $r+t$ and θ replaced by $\theta+\mu$. Theorem 2 says that for any complementary subspaces α and β

$$\frac{d}{d\theta}\log\frac{\operatorname{Det} L_{UA}}{\operatorname{Det} L_{UB}}=\frac{d}{d\theta}\log\operatorname{Det} U^{-1}\Phi_{r,\theta}U$$

and the same statement is true with $\frac{d}{d\theta}$ replaced by $\frac{d}{dr}$.

We can choose coordinates for α and β so that

$$T_\alpha f=\begin{bmatrix} f(x)-f(x+si)\\ f_v(x)+f_v(x+si)\end{bmatrix}\qquad x\in[0,1]$$

$$T_\beta f=\begin{bmatrix} f(x)-e^{2\pi i\mu}f(x+t+si)\\ f_v(x)+e^{2\pi i\mu}f_v(x+t+si)\end{bmatrix}\qquad x\in[0,1]$$

where f_v denotes $\frac{d}{dv}f$.

A basis for the solution space of L acting on the cylinder $\tilde{M}$ is given by

$$\{e^{-2\pi i(k+v)(x+yi)},\ e^{-2\pi i(k+v)(x-yi)}\}\qquad k\in\mathbf{Z}.$$

For each r and θ, $U^{-1}\Phi_{r,\theta}U$ is diagonalized with respect to this basis. That is, given a basis function, say

$$f=e^{-2\pi i((k+v)(x+yi))}$$

$U^{-1}\Phi_{r,\theta}U$ maps $U^{-1}T_{U\beta}f$ to $U^{-1}T_{U\alpha}f$.

$$U^{-1}T_{U\beta}f=T_\beta U^{-1}f=\begin{bmatrix} f(x)-e^{2\pi i(\theta+\mu)}f(x+t+r+si)\\ f_v(v)+e^{2\pi i(\theta+\mu)}f_v(x+t+r+si)\end{bmatrix}$$

$$=(1-e^{2\pi i(\theta+\mu)-2\pi i(k+v)(t+r+si)})\begin{bmatrix} f(x)\\ f_v(x)\end{bmatrix}.$$

Similarly

$$U^{-1}T_{U\alpha}f=(1-e^{2\pi i\theta-2\pi i(k+v)(r+si)})\begin{bmatrix} f(x)\\ f_v(x)\end{bmatrix}.$$

Therefore, $U^{-1}\Phi_{r,\theta}U$ taking $U^{-1}T_{U\beta}f$ to $U^{-1}T_{U\alpha}f$ is multiplication by

$$\frac{1-e^{-2\pi i(-\theta+(k+v)(r+si))}}{1-e^{-2\pi i(-(\mu+\theta)+(k+v)(t+r+si))}}.$$

Thus

$$\operatorname{Det} U^{-1}\Phi_{r,\theta}U$$

$$=\prod_k\frac{(1-e^{-2\pi i(-\theta+(k+v)(r+si))})(1-e^{-2\pi i(-\theta+(k+v)(r-si))})}{(1-e^{-2\pi i(-(\theta+\mu)+(k+v)(t+r+si))})(1-e^{-2\pi i(-(\theta+\mu)+(k+v)(t+r-si))})}$$

$$=\prod_k e^{-2\pi i\mu}\cdot\frac{(1-e^{2\pi i(|k|(r+si)-\varepsilon_k(\theta+v(r+si)))})(1-e^{-2\pi i(|k|(r-si)-\varepsilon_k(\theta-v(r-si)))})}{(1-e^{2\pi i(|k|(t+r+si)-\varepsilon_k(\theta+\mu-v(r+si)))})(1-e^{-2\pi i(|k|(r-si)-\varepsilon_k(\theta+\mu-v(t+r-si)))})}$$

where ε_k is equal to the sign of $(k+\frac{1}{2})$. This product does not converge with $\mu \neq 0$ because of the presence of the terms $e^{-2\pi i\mu}$. However, by Theorem 2, $\frac{d}{d\theta}$ of the log of this product is finite and rigorous sense can be made of this statement whether or not we have chosen coordinates which make $\det U^{-1}\Phi U$ itself finite. To be pedantic, we could alleviate this problem by multiplying our coordinates for α by $e^{-2\pi i\mu}$, thereby always working with a finite product, but the conclusion, statement (1) below, would remain exactly the same.

By Theorem 2, the derivative with respect to θ of the log of this product is finite and is equal to $\frac{d}{d\theta}\log\frac{\operatorname{Det} L_{UA}}{\operatorname{Det} L_{UB}}$ and the same is true with θ replaced by r. Note that when we take the derivative with respect to θ or r the $e^{-2\pi i\mu}$ terms disappear. Now integrating back shows that there is a non-zero constant c (independent of θ and r) such that

$$\frac{\operatorname{Det} L_{UA}}{\operatorname{Det} L_{UB}} = c\prod_k \frac{(1-e^{2\pi i(|k|(r+si)-\varepsilon_k(\theta-\nu(r+si)))})(1-e^{-2\pi i(|k|(r-si)-\varepsilon_k(\theta-\nu(r-si)))})}{(1-e^{2\pi i(|k|(r+si)-\varepsilon_k(\theta+\mu-\nu(r+si)))})(1-e^{-2\pi i(|k|(r-si)-\varepsilon_k(\theta+\mu-\nu(r-si)))})}. \tag{1}$$

From this we can deduce the following key observation.

Lemma 2.2. *There is a function h of* $\operatorname{Im}(\tau)$ *and ν such that*

$$\begin{aligned}\det \Delta &= h(\operatorname{Im}(\tau),\nu)\prod_k (1-e^{2\pi i(|k|\tau-\varepsilon_k(\mu-\nu\tau))})(1-e^{-2\pi i(|k|\bar{\tau}-\varepsilon_k(\mu-\nu\bar{\tau}))})\\ &= h(\operatorname{Im}(\tau),\nu)\Big|\prod_k (1-e^{2\pi i(|k|\tau-\varepsilon_k(\mu-\nu\tau))})\Big|^2.\end{aligned} \tag{2}$$

Proof of the lemma. Let $f(r,s,\mu,\nu)$ denote the determinant of Δ acting on the torus given by the lattice $\{1,\tau=r+si\}$ and the bundle induced by the representation $(m\tau+n)\to e^{2\pi i(m\mu+n\nu)}$, and $g(r,s,\mu,\nu)$ the infinite product on the right hand side of (2) (again with $\tau=r+si$). Then (1) states that there is a function $c(t,s,\mu,\nu)$ such that for every k, θ

$$\frac{f(r,s,\theta,\nu)}{f(t+r,s,\mu+\theta,\nu)} = c(r,s,\mu,\nu)\frac{g(r,s,\theta,\nu)}{g(t+r,s,\mu+\theta,\nu)}$$

or

$$\frac{g(t+r,s,\mu+\theta,\nu)}{f(t+r,s,\mu+\theta,\nu)} = c(t,s,\mu,\nu)\frac{g(r,s,\theta,\nu)}{f(r,s,\theta,\nu)}.$$

We will see that this function $c(t,s,\mu,\nu)$ is the constant 1. It is clear that c is a continuous, real-valued function (as everything involved is continuous and real-valued). Suppose μ, ν and t are rational numbers and choose $d\in\mathbf{Z}$ such that $d\mu$, dt, $dt\,\nu\in\mathbf{Z}$. Then applying the formula d times

$$\begin{aligned}\frac{g(r,s,\theta,\nu)}{f(r,s,\theta,\nu)} &= [c(t,s,\theta,\nu)]^d\frac{g(r-dt,s,\theta-d\mu,\nu)}{f(r-dt,s,\theta-d\mu,\nu)}\\ &= [c(t,s,\theta,\nu)]^d\frac{g(r,s,\theta-d\mu+dt\,\nu,\nu)}{f(r,s,\theta-d\mu+dt\,\nu,\nu)}.\end{aligned} \tag{3}$$

This last equality follows from the observation that for any $k \in \mathbf{Z}$ the torus generated by $\{1, \tau\}$ with the bundle induced by $(m\tau+n) \to e^{2\pi i(m\mu+n\nu)}$ is exactly the same as the torus generated by $\{1, \tau+k\}$ and the bundle induced by $(m(\tau+k)+n) \to e^{2\pi i(\mu+k\nu)+n\nu)}$. The expression in statement (3) is equal to

$$[c(t,s,\theta,\nu)]^d \frac{g(r,s,\theta,\nu)}{f(r,s,\theta,\nu)}$$

since everything depends only on $\theta \pmod 1$. Thus $[c(t,s,\theta,\nu)]^d = 1$, which implies that for t, θ, ν rational $c(t,s,\theta,\nu) = \pm 1$. Continuity of c combined with the fact that it must equal 1 when $t=\theta=0$ implies that $c \equiv 1$. That is

$$\frac{g(t+r,s,\mu+\theta,\nu)}{f(t+r,s,\mu+\theta,\nu)} = \frac{g(r,s,\theta,\nu)}{f(r,s,\theta,\nu)}.$$

The right hand side, and hence the left hand side, is independent or μ and t. This implies that the left hand side, and hence the right hand side, is independent of r and θ. Thus there is some function h of s and ν such that

$$f(r,s,\theta,\nu) = h(s,\nu)\, g(r,s,\theta,\nu)$$

as claimed. □

Continuation of the proof of the proposition: Thus, without much work, we have been able to discover the basic form of the determinant function.

To determine h we can proceed in a couple different fashions, but we follow the most thematic approach. That is, to get more information, we cut the torus M along a different curve and repeat the above process. In particular, switching the roles of x and y (i.e. cutting the torus M along the straight line $[0, \tau]$ rather than $[0,1]$) allows us to express $\operatorname{Det} \Delta$ as some function of τ and μ multiplied by an infinite product. This will completely determine the dependence of h on ν.

More precisely, a calculation following exactly along the lines of the previous calculation, except reversing the roles of x and y, results in the conclusion that there is some non-zero function g such that

$$\operatorname{Det} \Delta = g\left(\frac{\operatorname{Im}(\tau)}{|\tau|^2}, \mu\right) \left|\prod_{k \in \mathbf{Z}} \left(1 - e^{-2\pi i\left(\frac{|k|}{\tau} + \frac{\varepsilon_k}{\tau}(\mu - \nu\tau)\right)}\right)\right|^2.$$

Therefore, letting $z = \mu - \nu\tau$

$$\begin{aligned} \operatorname{Det} \Delta &= h(\operatorname{Im}(\tau), \nu) \left|\prod_k (1 - e^{2\pi i(|k|\tau - \varepsilon_k z)})\right|^2 \\ &= g\left(\frac{\operatorname{Im}(\tau)}{|\tau|^2}, \mu\right) \left|\prod_k \left(1 - e^{-2\pi i\left(\frac{|k|}{\tau} + \frac{\varepsilon_k}{\tau} z\right)}\right)\right|^2. \end{aligned}$$

Let $\Theta(\tau, z) = \prod_k (1 - e^{2\pi i(|k|\tau - \varepsilon_k z)})$ and $\hat{\Theta}(\tau, z) = \prod_k \left(1 - e^{2\pi i\left(\frac{|k|}{\tau} + \frac{\varepsilon_k}{\tau} z\right)}\right)$. Fixing τ, these two theta functions are entire functions with no zeroes. Furthermore, it is easy to check that the ratio has the following quasi-periodic behavior

$$\frac{\Theta}{\hat{\Theta}}(\tau, z+1) = -e^{-2\pi i \frac{z}{\tau}} \frac{\Theta}{\hat{\Theta}}(\tau, z)$$

$$\frac{\Theta}{\hat{\Theta}}(\tau, z+\tau) = -e^{-2\pi i(z+\tau)} \frac{\Theta}{\hat{\Theta}}(\tau, z).$$

This behavior is also exhibited by $e^{-\pi i\left(\frac{1}{\tau}(z^2-z)+z\right)}$ which implies that for some non-zero constant $c \in \mathbf{C}$

$$\frac{\Theta}{\hat{\Theta}} = c\, e^{-\pi i\left(\frac{1}{\tau}(z^2-z)+z\right)}.$$

Thus

$$\frac{g\left(\frac{\mathrm{Im}(\tau)}{|\tau|^2}, \mu\right)}{h(\mathrm{Im}(\tau), v)} = |c\, e^{-\pi i\left(\frac{1}{\tau}(z^2-z)+z\right)}|^2 = |c|^2\, e^{-2\pi \mathrm{Im}(\tau)(-v^2-v)}\, e^{-2\pi \frac{\mathrm{Im}(\tau)}{|\tau|^2}(\mu^2-\mu)}.$$

Therefore

$$h(\mathrm{Im}(\tau), v) = \hat{h}(\mathrm{Im}(\tau))\, e^{-2\pi \mathrm{Im}(\tau)(v^2-v)}$$

and

$$g\left(\frac{\mathrm{Im}(\tau)}{|\tau|^2}, \mu\right) = \hat{g}\left(\frac{\mathrm{Im}(\tau)}{|\tau|^2}\right) e^{-2\pi \frac{\mathrm{Im}(\tau)}{|\tau|^2}(\mu^2-\mu)}$$

for some functions $\hat{h}$, and $\hat{g}$. Combining this with our earlier information gives

$$\begin{aligned}\mathrm{Det}\, \Delta &= \hat{h}(\mathrm{Im}(\tau))\, e^{-2\pi \mathrm{Im}(\tau)(v^2-v)} \left|\prod_k (1 - e^{2\pi i(|k|\tau - \varepsilon_k(\mu - v\tau))})\right|^2 \\ &= \hat{g}\left(\frac{\mathrm{Im}(\tau)}{|\tau|^2}\right) e^{-2\pi \frac{\mathrm{Im}(\tau)}{|\tau|^2}(\mu^2-\mu)} \left|\prod_k (1 - e^{-2\pi i\left(\frac{|k|}{\tau} + \frac{\varepsilon_k}{\tau}(\mu - v\tau)\right)})\right|^2.\end{aligned}$$

So far we have used very general ideas which can work for many manifolds, but to calculate $\hat{h}$ and $\hat{g}$ we will have to use techniques specific to the torus. Namely, we will use the fact proven in [R-S2] that the log of the determinant is related to the sum of an Eisenstein series. Namely

$$\zeta'_{\Delta}(0) = \frac{\mathrm{Im}(\tau)}{\pi} \lim_{s \to 0} {\sum}' \frac{e^{2\pi i(m\mu + nv)}}{|m\tau + n|^{2(1-s)}}$$

where the notation $\sum'$ means that we sum over all pairs $(m, n) \in \mathbf{Z}^2$ with $(m, n) \neq (0, 0)$. Recall that $\zeta_{\Delta}(s) = \sum \lambda_k^{-s}$, where the λ_k's are the eigenvalues of Δ, and that $\log \mathrm{Det}\, \Delta = -\zeta'_{\Delta}(0)$. To completely determine $\log \mathrm{Det}\, \Delta$ we need to sum the right hand side of the above equality for arbitrary values of $\mathrm{Im}(\tau)$. We will do this for the values $\mu = 0$, $v = \frac{1}{2}$ and $\tau = ci$ with c an arbitrary real number. In this case, the sum we must consider is

$$\lim_{s \to 0} -\frac{c}{\pi} {\sum}' \frac{(-1)^n}{(c^2 m^2 + n^2)^{1-s}}.$$

For $s < 0$ this sum converges absolutely and we may rearrange the order of the terms.

$$-\frac{c}{\pi}\sum{}'\frac{(-1)^n}{(c^2m^2+n^2)^{1-s}}=-\frac{c}{\pi}\sum_{n\neq 0}\frac{(-1)^n}{n^{2-2s}}-\frac{c}{\pi}\sum_{m\neq 0}\sum_{n=-\infty}^{\infty}\frac{(-1)^n}{(c^2m^2+n^2)^{1-s}}$$

$$=\sum_{n\neq 0}\frac{(-1)^n}{n^{2s-2}}+2\sum_{m>0}\sum_{n=-\infty}^{\infty}\frac{(-1)^n}{(c^2m^2+n^2)^{1-s}}.$$

The only fact we will use is that $\sum_{n=-\infty}^{\infty}\frac{(-1)^n}{r^2+n^2}=\frac{\pi\operatorname{cosech}\pi r}{r}$. Thus, letting s go to 0 in the first sum yields

$$-\frac{c}{\pi}\lim_{r\to\infty}\left(\frac{\pi\operatorname{cosech}\pi r}{r}-\frac{1}{r^2}\right)=-\frac{c}{\pi}\left(-\frac{\pi^2}{6}\right)=\frac{c\pi}{6}.$$

Letting s go to 0 in the second sum yields

$$-\frac{2c}{\pi}\sum_{m>0}\frac{\pi\operatorname{cosech}\pi cm}{cm}=-4\sum_{m>0}\frac{e^{-\pi cm}}{m}\frac{1}{1-e^{-2\pi cm}}$$

$$=-4\sum_{m>0}\frac{1}{m}e^{-\pi cm}\sum_{k>0}(e^{-2\pi ck})^m=4\sum_{m>0}-\frac{1}{m}\sum_{k>0}(e^{-2\pi ck-\pi c})^m$$

$$=4\sum_{k>0}\log(1-e^{-2\pi ck-\pi c})=\sum_{k=-\infty}^{\infty}2\log(1-e^{-2\pi\left(|k|c+\frac{\varepsilon_k}{2}\right)}).$$

This sum is just the log of our infinite product. Therefore, comparing terms, we find that

$$e^{\frac{c\pi}{6}}=\hat{h}(c)\,e^{\frac{c\pi}{2}}$$

or

$$\hat{h}(c)=e^{-\frac{c\pi}{3}}.$$

Similarly, $\hat{g}(c)=e^{-\frac{c\pi}{3}}$. Therefore

$$\operatorname{Det}\Delta=e^{2\pi\operatorname{Im}(\tau)(-\nu^2+\nu-\frac{1}{6})}\left|\prod_{k=-\infty}^{\infty}(1-e^{2\pi i(|k|\tau-\varepsilon_k(\mu-\nu\tau))})\right|^2$$

$$=e^{2\pi\frac{\operatorname{Im}(\tau)}{|\tau|^2}(-\mu^2+\mu-\frac{1}{6})}\left|\prod_{k=-\infty}^{\infty}(1-e^{-2\pi i\left(\frac{|k|}{\tau}+\frac{\varepsilon_k}{\tau}(\mu-\nu\tau)\right)})\right|^2.$$

We can also restate this result in terms of Eisenstein series.

$$\lim_{s\to 0}\sum{}'\frac{e^{2\pi i(m\mu+n\nu)}}{|m\tau+n|^{2(1-s)}}=-\frac{\pi}{\operatorname{Im}(\tau)}\log\operatorname{Det}\Delta$$

$$=-2\pi^2(-\nu^2+\nu-\tfrac{1}{6})-\frac{\pi}{\operatorname{Im}(\tau)}\sum_{k=-\infty}^{\infty}\log|(1-e^{2\pi i(|k|\tau-\varepsilon_k(\mu-\nu\tau))})|^2$$

$$=-\frac{2\pi^2}{|\tau|^2}(-\mu^2+\mu-\tfrac{1}{6})-\frac{\pi}{\operatorname{Im}(\tau)}\sum_{k=-\infty}^{\infty}\log|(1-e^{2\pi i\left(\frac{|k|}{\tau}+\frac{\varepsilon_k}{\tau}(\mu-\nu\tau)\right)})|^2.$$

These equalities are known as Kroneker's Second Limit formula. Thus, along the way we have provided a new proof (although by no means simpler

than the original) of this formula. Our proof does have the advantage of showing explicitly where the terms in the series come from, namely they are log Det(1−(the holonomy of the space of harmonic functions around the torus)).

2.1.2. The Laplacian on $M \times S^1$. Here we present a very simple example to demonstrate that the method used to analyse the torus provides information in more general situations. Let M be a compact manifold without boundary with a complex line bundle E and a self-adjoint elliptic operator L_M of order n acting on sections of E. We can consider the bundle $E \times I$ over the manifold $M \times I$, where I is the unit interval, and the operator $L = -\frac{d^n}{dx^n} + L_M$ (x is the coordinate on I), acting on sections of $E \times I$. We let A denote the boundary condition

$$A: \left.\frac{d^r}{dx^r}\right|_{x=1} f(y,x) = e^{2\pi i\theta} \left.\frac{d^r}{dx^r}\right|_{x=0} f(y,x) \quad \text{for } 0 \leqq r \leqq n-1,\ y \in M$$

(where the first coordinate is in M and the second in I). Let $\{f_k\}_{k \geqq 0}$ be a complete set of eigenfunctions of L_M with eigenvalues $\{\lambda_k\}$, where $|\lambda_0| \leqq |\lambda_1| \leqq |\lambda_2| \leqq \ldots$. Let η be a primitive n^{th} root of 1. Then there is a basis for the solution space of L acting on $E \times I$ of the form

$$\{e^{\eta^j \sqrt[n]{\lambda_k}\, x} f_k(y)\} \qquad k \in \mathbf{Z},\ j = 1, 2, \ldots, n$$

where, for each value of k we have chosen one particular value of $\sqrt[n]{\lambda_k}$. Following the same logic as for the torus we find

Proposition 2.3. *With all notation as above*

$$\operatorname{Det} L_A = c \prod_k \left(\prod_{j=1}^{n} (1 - e^{\varepsilon_{j,k}(2\pi i\theta - \eta^j \sqrt[n]{\lambda_k})}) \right)$$

where $\varepsilon_{j,k} = \operatorname{sign}(\operatorname{Re}(\eta^j \sqrt[n]{\lambda_k}))$, *and* c *is independent of* θ.

For example, if $n=2$ and L_M is non-negative (each λ_k is non-negative) then

$$\operatorname{Det} L_A = c \prod_k (1 - e^{2\pi i\theta - \sqrt{\lambda_k}})(1 - e^{-2\pi i\theta - \sqrt{\lambda_k}}).$$

2.2. The unit disc

Now we present our last example of this section, that of $L_\lambda = \Delta - \lambda^2$ acting on functions on the unit disc D with boundary S^1. This example is perhaps somewhat less geometric than the others, but it serves to demonstrate the results of this paper. It also shows that we were lucky in Sect. 2.1.1 to be able to say as much as we did (in particular, Lemma 2.2). In more general situations there will be no such lemma, and the results will be decidedly less precise.

Recall that in polar coordinates

$$\Delta-\lambda^2=\frac{\partial^2}{\partial r^2}+\frac{1}{r}\frac{\partial}{\partial r}+\frac{1}{r^2}\frac{\partial^2}{\partial\theta^2}-\lambda^2.$$

Using separation of variable to find a basis for the solution space of $\Delta-\lambda^2$, we arrive at a basis $\{f_{\lambda,k}\}_{k\in\mathbf{Z}}$ where $f_{\lambda,k}=J_k(\lambda r)e^{ik\theta}$. In this expression, J_k is the kth Bessel function. For $\lambda=0$ we can take our basis to be $f_{0,k}=r^{|k|}e^{ik\theta}$. Let a and b be any constants and A and B the boundary conditions

$$A\colon\ af(x)+f_v(x)=0 \quad \text{for } x\in S^1$$
$$B\colon\ bf(x)+f_v(x)=0 \quad \text{for } x\in S^1$$

where v is the inward unit normal along S^1 and $f_v=\frac{\partial}{\partial v}f$. Now let c be another constant and U any function on the disc with $U(x)=1$ and $U_v(x)=-c$ for all $x\in S^1$. Multiplication by U maps any function satisfying A, B to a function satisfying UA, UB where

$$UA\colon\ (a+c)f(x)+f_v(x)=0 \quad \text{for } x\in S^1$$
$$UB\colon\ (b+c)f(x)+f_v(x)=0 \quad \text{for } x\in S^1.$$

Simple estimates show that the hypotheses of Proposition 1.3 are satisfied and thus the results of Sect. 1 can be applied. Theorem 1 and Theorem 2 combine to imply (after integrating the conclusion of the theorems) that for any complements α and β to A and B

$$\frac{\operatorname{Det}L_{\lambda,UA}}{\operatorname{Det}L_{\lambda,UB}}\frac{\operatorname{Det}L_{0,B}}{\operatorname{Det}L_{0,A}}=\det U^{-1}\Phi_{\lambda,c}U\Phi_{0,0}^{-1}=\det U^{-1}T_{U\alpha}P_{\lambda,U\beta}UT_\beta P_{0,\alpha}$$

where the right hand side is a well-defined, finite determinant. To calculate the right hand side, note that we can choose coordinates on α and β so that

$$T_\alpha f=[af(x)+f_v(x)] \quad x\in S^1$$
$$T_\beta f=[bf(x)+f_v(x)] \quad x\in S^1.$$

Then $T_\beta P_{0,\alpha}$ takes $T_\alpha f_{0,k}$ to $T_\beta f_{0,k}$.

$$T_\alpha f_{0,k}=(a+|k|)e^{ik\theta}$$
$$T_\beta f_{0,k}=(b+|k|)e^{ik\theta}.$$

$U^{-1}T_{U\alpha}P_{\lambda,U\beta}U$ takes $U^{-1}T_{U\alpha}f_{\lambda,k}$ to $U^{-1}T_{U\beta}f_{\lambda,k}$.

$$U^{-1}T_{U\alpha}f_{\lambda,k}=T_\alpha U^{-1}f_{\lambda,k}=\left(b+c+\lambda\frac{J_k'(\lambda)}{J_k(\lambda)}\right)e^{ik\theta}$$
$$U^{-1}T_{U\beta}f_{\lambda,k}=T_\beta U^{-1}f_{\lambda,k}=((b+c)J_k(\lambda)+\lambda J_k'(\lambda))e^{ik\theta}$$
$$U^{-1}T_{U\alpha}f_{\lambda,k}=T_\alpha U^{-1}f_{\lambda,k}=((a+c)J_k(\lambda)+\lambda J_k'(\lambda))e^{ik\theta}.$$

Therefore $U^{-1}\Phi_{\lambda,c}U\Phi_{0,0}^{-1}$ acting on $e^{ik\theta}\in\alpha$ is multiplication by

$$\frac{\left(a+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}{\left(b+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}\frac{(b+|k|)}{(a+|k|)}$$

and

$$\det U^{-1}\Phi_{\lambda,c}U\Phi_{0,0}^{-1}=\prod_{k\in\mathbf{Z}}\frac{\left(a+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}{\left(b+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}\frac{(b+|k|)}{(a+|k|)}.$$

Let $L_{\lambda,a}$ denote the operator $\Delta-\lambda^2$ acting on functions on the unit disc which satisfy $af(x)+f_v(x)=0$. Then we have proven

Proposition 1.4.

$$\frac{\operatorname{Det}L_{\lambda,a+c}}{\operatorname{Det}L_{\lambda,b+c}}\frac{\operatorname{Det}L_{0,b}}{\operatorname{Det}L_{0,a}}=\prod_{k\in\mathbf{Z}}\frac{\left(a+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}{\left(b+c+\lambda\dfrac{J_k'(\lambda)}{J_k(\lambda)}\right)}\frac{(b+|k|)}{(a+|k|)}.$$

Unfortunately, there is no periodicity in the variables to enable us to zero in on the dependence on λ and c, as we did in the case of the torus. It appears that this is the best one can do with the techniques of Sect. 1. On the other hand, this is still great progress. The left hand side is equal to $\prod_k\left(\prod_i\frac{x_{i,k}}{z_{i,k}}\frac{y_{i,k}}{w_{i,k}}\right)$ where the $x_{i,k}$'s are the eigenvalues of $L_{\lambda,a+c}$ whose eigenfunctions restrict to $e^{ik\theta}$ along S^1, and the y's, z's and w's are defined similarly. These x's, y's, z's and w's can be expressed as zeros of Bessel functions, using separation of variables, but for generic λ, a, b, c not a single one can be written down explicitly.

Our reduction of this doubly infinite product to a singly infinite product is a direct result of working over the boundary circle which is one dimensional, as opposed to the two dimensional disc. Furthermore, our product is an explicit one, as we work with the solution space rather than eigenvalues.

As a last remark, we note that the equality between the two infinite products actually holds independently for each k. This is due to the fact that separation of variables allows one to prove result one value of k at a time. No such simplification would occur if there was no circular symmetry.

Suppose for example, that the c in the boundary conditions was not constant along S^1. That is, the boundary conditions UA and UB are

$$UA\colon\ (a+c(x))f(x)+f_v(x)=0 \quad \text{for } x\in S^1$$
$$UB\colon\ (b+c(x))f(x)+f_v(x)=0 \quad \text{for } x\in S^1.$$

In this case, the problem has lost its symmetry so separation of variables cannot be used to solve for eigenvalues. Thus we stand little chance of saying

anything explicit about the eigenvalues, but the solution space of L_λ has not changed and can still be written down. Thus, the right hand side will still be the determinant of an explicitly given map.

Very little changes because c is now a function. $T_\beta P_{0,\alpha}$ is still the map which sends $e^{ik\theta}$ to $\dfrac{(b+|k|)}{(a+|k|)} e^{ik\theta}$ and $(U^{-1} T_{U\alpha} P_{\lambda, U\beta} U)$ still sends

$$\left(b+c(x)+\lambda \frac{J_k'(\lambda)}{J_k(\lambda)}\right) e^{ik\theta}$$

to

$$\left(a+c(x)+\lambda \frac{J_k'(\lambda)}{J_k(\lambda)}\right) e^{ik\theta}.$$

Expand $c(x)$ in Fourier series $c(x)=\sum_n c_n e^{in\theta}$. Then $(U^{-1} T_{U\alpha} P_{\lambda, U\beta} U)$ sends

$$\sum_n c_n e^{i(k+n)\theta} + \left(b+\lambda \frac{J_k'(\lambda)}{J_k(\lambda)}\right) e^{ik\theta}$$

to

$$\sum_n c_n e^{i(k+n)\theta} + \left(a+\lambda \frac{J_k'(\lambda)}{J_k(\lambda)}\right) e^{ik\theta}.$$

Define for any constant λ and any function $c(x)$ the infinite square matrix $M_{\lambda, c(x)}(p, q)$, with $p, q \in \mathbf{Z}$, and $-\infty < p, q < \infty$, by

$$M_{\lambda, c(x)}(k, k+n) = \begin{cases} c_0 + \lambda \dfrac{J_k'(\lambda)}{J_k(\lambda)} & \text{if } n=0 \\ c_n & \text{if } n \neq 0. \end{cases}$$

Then we have

Proposition 2.5.

$$\frac{\operatorname{Det} L_{\lambda, b+c(x)}}{\operatorname{Det} L_{\lambda, a+c(x)}} \frac{\operatorname{Det} L_{0,a}}{\operatorname{Det} L_{0,b}} = \det M_{\lambda, a+c(x)} M^{-1}_{\lambda, b+c(x)} M_{0,b} M^{-1}_{0,a}$$

where the right hand side is an infinite square matrix with a well-defined determinant.

Note that $M_{0,a} M^{-1}_{0,b}$ is the diagonal matrix whose k^{th} entry is $\dfrac{(b+|k|)}{(a+|k|)}$.

3. Ordinary differential operators

3.1. *A closer look*

In the case of ordinary differential operators, we can make much more precise statements than those in the previous sections. This is mainly due to the fact that we can work more concretely. For example, any vector bundle over an interval is trivial so we can take E to be $[a, b] \times \mathbf{C}^r$ over the interval $[a, b]$.

Sections of E are therefore r-tuples of complex functions. A differential operator L of order n is of the form

$$L = P_0(x)\frac{d^n}{dx^n} + P_1(x)\frac{d^{n-1}}{dx^{n-1}} + \ldots + P_n(x)$$

where the $P_i(x)$ are $r \times r$ complex matrices. L is elliptic if $\det P_0(x) \neq 0$ for each $x \in [a, b]$. The solution space of L has dimension nr. The space of boundary values has dimension $2nr$, where for any section f of E

$$Tf = \begin{bmatrix} f(a) \\ f'(a) \\ \ldots\ldots \\ f^{(n-1)}(a) \\ f(b) \\ \ldots\ldots \\ f^{(n-1)}(b) \end{bmatrix}.$$

Any elliptic boundary condition is a linear subspace of dimension nr in the space of boundary values. For L as above, and any elliptic boundary condition B, $L^{-1}(x, y)$ is the unique $r \times r$ matrix-valued function satisfying

1. $L_x(L^{-1}(x, y)) = 0$ for $x \neq y$.
2. $T_x(L^{-1}(x, y)) \in B$.
3. $\dfrac{d^k}{dx^k} L^{-1}(x, y)$ is continuous for $k \leqq n-2$ and

$$\lim_{y \to x+} \frac{d^{n-1}}{dx^{n-1}} L^{-1}(x, y) + P_0(x) = \lim_{y \to x-} \frac{d^{n-1}}{dx^{n-1}} L^{-1}(x, y).$$

L_x and T_x can be thought of as acting independently on each column of $L^{-1}(x, y)$.

We have seen that if $L_t' L_t^{-1}$ is trace class, then

$$\frac{d}{dt} \log \operatorname{Det} L_t = \operatorname{tr} L' L^{-1} = \int_M \operatorname{tr}[L_x' L^{-1}(x, y)]|_{y=x} dx.$$

If L is an ordinary differential operator then we can go a step further. $L' L^{-1}$ is trace class, i.e. has a kernel which is continuous along the diagonal, if the order of L' is at least 2 less than the order of L (this is apparent from property 3 of L^{-1}). If L' is of order 1 less than L then $L' L^{-1}$ is not trace class, as the kernel $L_x' L^{-1}(x, y)$ jumps across the diagonal. We will see, however, that $\dfrac{d}{dt} \log \operatorname{Det} L$ can easily be written in terms of the limits of $L' L^{-1}$ as (x, y) approaches the diagonal from the two sides.

To do this explicitly, we need the following lemma. We know that

$$\lim_{s \to 0} L^{-s} f = f$$

for any differentiable function f. We need to know what happens at a jump discontinuity. This lemma may have some interest in its own right, if only because of the dependence of the answer on the order of L. To set the scene, let L be an elliptic ordinary differential operator of order n with an elliptic boundary condition acting on sections of E. Suppose further that L restricted to the boundary condition has no zero modes. We require, as we have all along, that the spectrum of the top order symbol of L, $P_0(x)(-i)^n\xi^n$ misses some cone C in $\mathbf{C}$ (which implies that it misses $-C$ if n is odd). Choose a ray of minimal growth in C to define L^{-s}.

Lemma 3.1. *With all notation as above, let f be a piecewise differentiable function with only finitely many singularities, all of them jump discontinuities in the interior of the interval. Then*

1. *If n is even, then*

$$\lim_{s\to 0}(L^{-s}f)(x)=\tfrac{1}{2}[f(x+)+f(x-)].$$

2. *If n is odd then for each x there is some projection π_x such that*

$$\lim_{s\to 0}(L^{-s}f)(x)=\pi_x f(x+)+(1-\pi_x)f(x-).$$

Furthermore, the rank of π_x is constant along the interval. The generalized eigenspaces of $(-i)^nP_0(x)$ can be divided into two sets depending on the sector of $\mathbf{C}-(C\cup -C)$ in which their eigenvalue falls. π_x and $(1-\pi_x)$ are projections onto these spaces. So that, for example, if $P_0(x)=I$ then $\pi_x=0$ or I.

Proof. Fix a point $x_0\in(a,b)$. The proof is divided into two steps.

(i) $\lim_{s\to 0}(L^{-s}f)(x_0)$ depends only on the value of the highest order term of L at x_0 and the ray of minimal growth, and in particular, it is independent of the boundary condition: Let L_0 be any other elliptic operator of the same order as L such that the highest order terms agree at x_0 and the spectrum of the top order symbol of L_0 also misses the cone C. Choose any elliptic boundary condition for L_0 such that L_0 restricted to this boundary condition has no zero modes. L and L_0 have only finitely many eigenvalues in C. Moving the ray of minimal growth past finitely many eigenvalues does not affect the limit of L^{-s} as s goes to 0. Therefore we may choose a ray in C common to both L and L_0. Defining L^{-s} and L_0^{-s} with respect to this ray we will see that

$$(\lim_{s\to 0}L_0^{-s}f)(x_0)-(\lim_{s\to 0}L^{-s}f)(x_0)=0.$$

The left hand side is equal to 0 at all points at which f is differentiable, so it is enough to see that it is continuous at x_0.

$$\begin{aligned}(\lim_{s\to 0}L_0^{-s}f)-(\lim_{s\to 0}L^{-s}f)&=\lim_{s\to 0}(L_0L_0^{-1-s}-LL^{-1-s})f\\&=\lim_{s\to 0}(L_0(L_0^{-1-s}-L^{-1-s}))f+\lim_{s\to 0}(L_0-L)L^{-1-s}f\\&=L_0(L_0^{-1}-L^{-1})f+(L_0-L)L^{-1}f.\end{aligned}$$

It is enough to see that these two functions are continuous at x_0. It can be seen from property 3 of L^{-1} that both $L_0(L_0^{-1}-L^{-1})$ and $(L_0-L)L^{-1}$ are of the form $P(x)\delta(x-y)+K(x,y)$ where P is continuous, $P(x_0)=0$ and K has a jump discontinuity along the diagonal. It follows then that both $L_0(L_0^{-1}-L^{-1})f$ and $(L_0-L)L^{-1}f$ are continuous at x_0.

(ii) From (i) it is enough to consider, for example, $L=P\dfrac{d^n}{dx^n}+P\varepsilon$ (the point of the ε is to make L invertible) where P is a constant $r\times r$ matrix, acting on functions satisfying the boundary condition

$$f(a)=f(b),\ f'(a)=f'(b),\ \ldots,\ f^{(n-1)}(a)=f^{(n-1)}(b).$$

Suppose that P is diagonalizable. Treating one dimension at a time, we may assume that P is a scalar. For the operator $L=P\dfrac{d^n}{dx^n}+P\varepsilon$ and the given boundary condition the eigenfunctions are e^{ikx}, $n\in\mathbf{Z}$, with eigenvalues $P[(ik)^n+\varepsilon]$. We will now assume, for simplicity that $a=0$, $b=2\pi$ and $P=1$. We can make this assumption about P because $(PL)^s=P^sL^s$ as long as we take the ray of minimal growth of PL to be $P\times$(the ray of minimal growth for L).

Consider first the function

$$h(x)=\begin{cases} x & x\leqq\pi \\ x-2\pi & x\geqq\pi. \end{cases}$$

The Fourier series for $h(x)$ is $\sum_{0\neq k\in\mathbf{Z}}\dfrac{(-1)^k e^{ikx}}{ik}$ so

$$L^{-s}h=\sum_{0\neq k\in\mathbf{Z}}\frac{(-1)^k e^{ikx}}{ik[(ik)^n+\varepsilon]^{-s}}.$$

We will now investigate $\lim_{s\to 0}(L^{-s}h)(\pi)$.

$$\begin{aligned}\lim_{s\to 0}(L^{-s}h)(\pi)&=\lim_{s\to 0}\sum_{0\neq k\in\mathbf{Z}}\frac{1}{ik[(ik)^n+\varepsilon]^s}\\&=\lim_{s\to 0}\sum_{k=1}^{\infty}\frac{1}{ik}\left(\frac{1}{[(ik)^n+\varepsilon]^s}-\frac{1}{[(-ik)^n+\varepsilon]^s}\right).\end{aligned}$$

We can rearrange terms since the sum converges absolutely for $s>0$. If n, the order of L, is even then each term in the series vanishes and we have

$$\lim_{s\to 0}(L^{-s}h)(\pi)=0=\tfrac{1}{2}[\lim_{x\to\pi+}h(x)+\lim_{x\to\pi-}h(x)].$$

If n is odd we must work a little harder. The same proof as in part (i) shows that we can take $\varepsilon=0$ in this sum with out affecting the limit (as h is orthogonal to the kernel of L). If we choose a ray of minimal growth ρ with $\frac{\pi}{2}<\arg\rho<\dfrac{3\pi}{2}$ and choose the branch so that λ^s is real if λ is real then the sum equals

$$\lim_{s\to 0}\sum_{k=1}^{\infty}\frac{1}{i(k)^{1+sn}}(e^{-i\frac{\pi}{2}sn}-e^{i\frac{\pi}{2}sn})=\lim_{s\to 0}\left(-2\sin\left(\frac{\pi}{2}sn\right)\sum_{k=1}^{\infty}\frac{1}{k^{1+sn}}\right).$$

The simple estimate

$$\frac{1}{sn}=\int_1^{\infty}x^{1+sn}dx<\sum_{k=1}^{\infty}\frac{1}{k^{1+sn}}<1+\int_1^{\infty}x^{1+sn}dx=1+\frac{1}{sn}$$

shows that

$$\lim_{s\to 0}(L^{-s}h)(\pi)=\lim_{s\to 0}\left(-2\sin\left(\frac{\pi}{2}sn\right)\sum_{k=1}^{\infty}\frac{1}{k^{1+sn}}\right)=-\pi=\lim_{x\to\pi+}h(x).$$

If we had chosen our ray in the other half plane then we would have found

$$\lim_{s\to 0}(L^{-s}h)(\pi)=\pi=\lim_{x\to\pi-}h(x).$$

Now let f be any function which is continuously differentiable except at a set of points $x_0, x_1, \ldots, x_q$. Then

$$f(x)=g(x)+\sum_{i=0}^{q}\frac{f(x_i-)-f(x_i+)}{2\pi}h(x-x_i(\mathrm{mod}\,2\pi))$$

for some continuously differentiable function g. Applying $\lim_{s\to 0}L^{-s}$ to each term in the right hand side proves the lemma for general f.

Now suppose that P is not diagonalizable. It is enough to restrict attention to a minimal invariant subspace of P, where we can choose a basis so that P has the form

$$P=\begin{pmatrix}\lambda & 1 & \ldots & 0\\ 0 & \lambda & \ldots & 0\\ 0 & 0 & \ldots & 1\\ 0 & 0 & \ldots & \lambda\end{pmatrix}.$$

Perturbing P slightly to a diagonalizable matrix with eigenvalues all near λ, and applying the above proof shows that $\lim_{s\to 0}(L^{-s}f)(x_0)$ is either $f(x_0+)$ or $f(x_0-)$ depending only on which sector of $\mathbf{C}-(C\cup -C)$ contains λ. This proves the lemma in general □

Now we make a definition for notational convenience. Given an operator L of order n and a ray of minimal growth, define a map $R\colon [a,b]\to \mathrm{End}(\mathbf{C}^r)$ by

$$R(x)=\begin{cases}\frac{1}{2} & \text{if } n \text{ is even}\\ \pi_x & \text{if } n \text{ is odd}\end{cases} \tag{4}$$

where π_x is defined as in Lemma 3.1. Lemma 3.1 can now be restated as

$$\lim_{s\to 0}(L^{-s}f)(x_0)=R(x_0)f(x_0+)+(1-R(x_0))f(x_0-).$$

We can now apply this result to study the determinant of an ordinary differential operator.

Proposition 3.2. *Let L_t be a one-parameter family of ordinary differential operators on $[a, b]$ with the same highest order term. Let B be a boundary condition which is elliptic for the L_t. Then (writing L for $L_{t,B}$)*

$$\frac{d}{dt}\log \operatorname{Det} L = \int_a^b \operatorname{tr}[R(x) \lim_{y\to x-} (L' L^{-1}(x, y)) + (1-R(x)) \lim_{y\to x+} (L' L^{-1}(x, y))]\, dx.$$

Proof.

$$\frac{d}{dt}\log \operatorname{Det} L = \frac{d}{dt}\frac{d}{ds}\Big|_{s=0} \operatorname{tr} \frac{1}{2\pi i}\int_\Gamma \lambda^{-s}(L-\lambda)^{-1}\, d\lambda$$

$$= \frac{d}{ds}\Big|_{s=0} \frac{1}{2\pi i}\int_\Gamma \lambda^{-s} \operatorname{tr}(L-\lambda)^{-1} L'(L-\lambda)^{-1}\, d\lambda.$$

To simplify this expression we note that

$$\operatorname{tr}(L-\lambda)^{-1} L'(L-\lambda)^{-1} = \int_a^b \int_a^b \operatorname{tr}[(L-\lambda)^{-1}(x, y)]\, [L'_y (L-\lambda)^{-1}(y, x)]\, dy\, dx$$

$$= \int_a^b \int_a^b \operatorname{tr}[L'_y (L-\lambda)^{-1}(x, y)]\, [(L-\lambda)^{-1}(x, y)]\, dy\, dx.$$

We can switch the order of integration since $L'_y(L-\lambda)^{-1}$ and $(L-\lambda)^{-1}$ have, at worst, jump discontinuities.

$$= \operatorname{tr} L'_y (L-\lambda)^{-2}.$$

Thus

$$\frac{d}{dt}\log \operatorname{Det} L = \frac{d}{ds}\Big|_{s=0} \operatorname{tr} \frac{1}{2\pi i}\int_\Gamma \lambda^{-s} L'(L-\lambda)^{-2}\, d\lambda.$$

Now integrating by parts we get

$$\lim_{s\to 0} \operatorname{tr} L' L^{-1-s} = \lim_{s\to 0} \operatorname{tr} L^{-s} L' L^{-1}$$

(since for s large, L^{-s} is trace class and $L' L^{-1}$ is bounded)

$$= \lim_{s\to 0} \operatorname{tr} \int_a^b [L^{-s}(L'_x L^{-1}(x, y))]|_{y=x}\, dx$$

$$= \operatorname{tr} \int_a^b [\lim_{z\to x+} R(x) L'_z L^{-1}(z, y) + \lim_{z\to x-} (1-R(x)) L'_z L^{-1}(z, y)]|_{y=x}\, dx$$

$$= \int_a^b \operatorname{tr}[\lim_{z\to x+} R(x) L'_z L^{-1}(z, x) + \lim_{z\to x-} (1-R(x)) L'_z L^{-1}(z, x)]\, dx$$

$$= \int_a^b \operatorname{tr}[R(x) \lim_{y\to x-} L'_x L^{-1}(x, y) + (1-R(x)) \lim_{y\to x+} L'_x L^{-1}(x, y)]\, dx. \qquad \square$$

We should point out that the expression

$$\int_a^b \operatorname{tr}[R(x) \lim_{y\to x+} L' L^{-1}(x, y) + (1-R(x)) \lim_{y\to x-} L' L^{-1}(x, y)]\, dx$$

makes sense for operators for which the determinant is not defined. We will now *define* for one-parameter family L_t of ordinary differential operators with the same highest order term, and *any* boundary condition B (any linear subspace in the space of boundary values of half the dimension)

$$\frac{d}{dt}\log \operatorname{Det} L_{t,B} = \int_a^b \operatorname{tr}[R(x) \lim_{y\to x+} L' L^{-1}(x,y) + (1-R(x)) \lim_{y\to x-} L' L^{-1}(x,y)]\,dx \quad (5)$$

where, in the right hand side, L stands for $L_{t,B}$. This allows us to give meaning to $\dfrac{\operatorname{Det} L_{0,B}}{\operatorname{Det} L_{1,B}}$ but not necessarily to $\operatorname{Det} L_{0,B}$ alone. From our previous discussion we know that this agrees with the notion of Fredholm determinant whenever it exists (i.e. when $[\text{order } L'] \leqq [(\text{order } L)-2]$), and with the notion of the zeta function determinant whenever we have elliptic operators. Furthermore, using definition (5) we have

$$\frac{d}{dt}(\log \operatorname{Det} L_{t,A} - \log \operatorname{Det} L_{t,B}) = \int_a^b \operatorname{tr}[L'_x(L_A^{-1}(x,y) - L_B^{-1}(x,y))]|_{y=x}\,dx$$

which is all that is required for the proof of Proposition 1.3 to go through. This proposition will enable us to make conclusions about the determinant of elliptic operators by examining the corresponding ratios of determinants (using the above definition) for some non-elliptic boundary conditions. Thus, summarizing, we have

Lemma 3.3. *If for any family of elliptic differential operators L_t with the same highest order term and any (not necessarily elliptic) boundary condition we define $\dfrac{d}{dt}\log \operatorname{Det} L_t$ by* (5) *then, for any two boundary conditions A and B with linear complements α, β*

$$\frac{\operatorname{Det} L_{0,A}}{\operatorname{Det} L_{1,A}} \frac{\operatorname{Det} L_{1,B}}{\operatorname{Det} L_{0,B}} = \det \Phi_{L_1}^{-1} \Phi_{L_0} \colon \beta \to \beta$$

where Φ_L is defined as in Sect. 1.5.

Note that in this case Φ_L is a map between *finite dimensional* vector spaces.

Now we will choose a boundary condition B such that $\dfrac{\operatorname{Det} L_{1,B}}{\operatorname{Det} L_{0,B}}$ is easily computed from the coefficients of the operators. This will leave us with a formula for $\dfrac{\operatorname{Det} L_{0,A}}{\operatorname{Det} L_{1,A}}$. For any "reasonable" boundary condition, we will be able to choose some operator L_1 whose highest order term agrees with that of L_0, such that we can calculate $\operatorname{Det} L_{1,A}$ and $\operatorname{Det} \Phi_{L_1}$ explicitly, and this will leave us with a formula for $\operatorname{Det} L_{0,A}$ in terms of the boundary values of L_0. This will be demonstrated in the next section.

The boundary condition B that we choose is the Cauchy boundary condition. Considering B as a subspace of $\mathbf{C}^{2nr}$, let $B = \{b\colon b = [0, v] \text{ with } v \in \mathbf{C}^{nr}\}$.

That is

$$f \text{ satisfies } B \leftrightarrow f(a)=f'(a)=\ldots=f^{(n-1)}(a)=0.$$

The purpose of choosing this boundary condition is made clear by the following proposition.

Proposition 3.4. *Let* $L=P_0(x)\frac{d^n}{dx^n}+P_1(x)\frac{d^{n-1}}{dx^{n-1}}+$(*lower order terms*), $\hat{L}=P_0(x)\frac{d^n}{dx^n}+P_1(x)\frac{d^{n-1}}{dx^{n-1}}+$(*lower order terms*) *and B the Cauchy boundary condition, then*

$$\frac{\operatorname{Det} L_B}{\operatorname{Det} \hat{L}_B}=\frac{\exp\left(\int_a^b \operatorname{tr} R(x)P_1(x)P_0^{-1}(x)\,dx\right)}{\exp\left(\int_a^b \operatorname{tr} R(x)\hat{P}_1(x)P_0^{-1}(x)\,dx\right)}.$$

Proof. Let

$$L_t=(1-t)L+t\hat{L}=P_0(x)\frac{d^n}{dx^n}+((1-t)P_1(x)+t\hat{P}_1(x))\frac{d^{n-1}}{dx^{n-1}}+(\text{lower order terms}).$$

Then the following properties are true (see the discussion on Green's functions in the beginning of this section)

(i) $\lim_{y\to x+}\frac{d^k}{dx^k}L_{t,B}^{-1}(x,y)=\lim_{y\to x-}\frac{d^k}{dx^k}L_{t,B}^{-1}(x,y)=0 \quad \text{for } k\leqq n-2.$

(ii) $\lim_{y\to x+}\frac{d^{n-1}}{dx^{n-1}}L_{t,B}^{-1}(x,y)=0.$

(iii) $\lim_{y\to x-}\frac{d^{n-1}}{dx^{n-1}}L_{t,B}^{-1}(x,y)=P_0^{-1}(x).$

Now, applying the definition (5), we find

$$\frac{d}{dt}\log \operatorname{Det} L_{t,B}=\int_a^b \operatorname{tr} R(x)\,[\hat{P}_1(x)-P_1(x)]\,P_0^{-1}(x)\,dx.$$

Integrating this expression yields the proposition. □

We can use this result to state the following corollary to Theorem 1.

Corollary 3.5. *Let A be any boundary condition, B the Cauchy boundary condition and* Φ_L: $\beta\to\alpha$ *defined as in Sect. 1. If L and* $\hat{L}$ *are any operators of the form*

$$L=P_0(x)\frac{d^n}{dx^n}+P_1(x)\frac{d^{n-1}}{dx^{n-1}}+(\text{lower order terms})$$

and

$$\hat{L}=P_0(x)\frac{d^n}{dx^n}+\hat{P}_1(x)\frac{d^{n-1}}{dx^{n-1}}+(\text{lower order terms})$$

then

$$\frac{\operatorname{Det} L_A}{\operatorname{Det} \hat{L}_A} = \frac{\exp\left(\int_a^b \operatorname{tr} R(x) P_1(x) P_0^{-1}(x)\, dx\right) \det \Phi_L}{\exp\left(\int_a^b \operatorname{tr} R(x) \hat{P}_1(x) P_0^{-1}(x)\, dx\right) \det \Phi_{\hat{L}}}.$$

Similarly, we can state a corollary to Theorem 2 concerning the variation of the determinant of ordinary differential operators as the boundary conditions are varied. As before, the variations we will consider are those that are induced by automorphisms of the bundle. Each automorphism induces a map from the boundary of the interval $[a, b]$ to itself, so it either fixes a and b, or inverts them. We will restrict ourselves in this section to automorphisms which leave a and b fixed. As before, let B be the Cauchy boundary condition. If U is any automorphism of the bundle E over $[a, b]$ which fixes a, then $UB = B$. We can thus simplify the conclusion to Theorem 2.

Corollary 3.6. *With all notation as above, B the Cauchy boundary condition, and U_t a one-parameter family of automorphisms of E which leave a fixed, then*

$$\frac{\operatorname{Det} L_{U_t A}}{\operatorname{Det} L_{U_0 A}} = \frac{\det U_t^{-1} \Phi_L U_t}{\det \Phi_L}$$

where $U_t^{-1} \Phi_L U_t$, Φ_L: $\beta \to \alpha$ are defined as in Theorem 2.

We can now restate these results in a very useful fashion. First we slightly reorganize our objects of study. Instead of using operator

$$L = P_0(x) \frac{d^n}{dx^n} + \ldots + P_n(x)$$

we will work with the equivalent first order operator on nr-tuples os functions

$$\Gamma = \frac{d}{dx} - \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ 0 & 0 & 0 & \ldots & 0 \\ \ldots\ldots & \ldots\ldots & \ldots\ldots & \ldots & \ldots\ldots \\ -P_0^{-1} P_n & -P_0 P^{-1} P_{n-1} & -P_0^{-1} P_{n-2} & \ldots & -P_0^{-1} P_1 \end{bmatrix}.$$

Let $Y_L(x)$ be the fundamental solution of Γ, i.e. the $nr \times nr$ matrix solution such that $Y_L(a) = I$, the $nr \times nr$ identity matrix. Then, for any f such that $Lf = 0$

$$\begin{bmatrix} f(x) \\ f'(x) \\ \ldots \\ f^{(n-1)} \end{bmatrix} = Y_L(x) \begin{bmatrix} f(a) \\ f'(a) \\ \ldots \\ f^{(n-1)}(a) \end{bmatrix}.$$

The advantage of this approach is that the columns of Y_L provide us with a convenient basis of the solution space of L. Let B be the Cauchy boundary

condition and A be any other boundary condition. $A \subset \mathbf{C}^{2nr}$ is a linear subspace of dimension nr so there exist $nr \times nr$ matrices M and N such that the rank of $[M, N]$ is nr and

$$A = \left\{ \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} : \; M v_1 + N v_2 = 0 \right\}$$

i.e.

$$f \text{ satisfies } A \leftrightarrow M \begin{pmatrix} f(a) \\ \cdots\cdots \\ f^{(n-1)}(a) \end{pmatrix} + N \begin{pmatrix} f(b) \\ \cdots\cdots \\ f^{(n-1)}(b) \end{pmatrix} = 0.$$

We then say that A is represented by (M, N), so that, for example, the Cauchy boundary condition is represented by $(I, 0)$. Note that the pair (M, N) is not uniquely determined.

If β is any complement to B then we can choose coordinates for β so that

$$T_\beta(f) = \begin{pmatrix} f(a) \\ \cdots\cdots \\ f^{(n-1)}(a) \end{pmatrix}.$$

Let α be any complementary subspace to A, then

$$\Phi_L = T_\alpha P_\beta : \; \beta \to \{\text{solution space of } L\} \to \alpha$$

is given by

$$\Phi_L(v) = T_\alpha([Y_L(x)]\, v) = \pi_\alpha \begin{bmatrix} v \\ [Y_L(b)]\, v \end{bmatrix} = \left[\pi_\alpha \begin{pmatrix} I \\ Y_L(b) \end{pmatrix} \right] v$$

where π_α is the projection from $\mathbf{C}^{2nr}$ to α. Now choose a basis for α. It is easy to see that there is some non-zero constant $c \in \mathbf{C}$ (depending on our choice of basis as well as our choice of M and N) such that

$$\det \Phi_L = \det \left[\pi_\alpha \begin{pmatrix} I \\ Y_L(b) \end{pmatrix} \right] = c \det(M + N Y_L(b)).$$

We can now restate Corollary 3.5 as follows

Corollary 3.7. *Let L and $\hat{L}$ be two operators of the form*

$$L = P_0(x) \frac{d^n}{dx^n} + P_1(x) \frac{d^{n-1}}{dx^{n-1}} + (\textit{lower order terms})$$

$$\hat{L} = P_0(x) \frac{d^n}{dx^n} + \hat{P}_1(x) \frac{d^{n-1}}{dx^{n-1}} + (\textit{lower order terms})$$

and suppose A is a boundary condition represented by (M, N). Then

$$\frac{\operatorname{Det} L_A}{\operatorname{Det} \hat{L}_A} = \frac{\exp\left(\int_a^b \operatorname{tr} R(x) \hat{P}_1(x) P_0^{-1}(x)\, dx \right) \det(M + N Y_{\hat{L}}(b))}{\exp\left(\int_a^b \operatorname{tr} R(x) P_1(x) P_0^{-1}(x)\, dx \right) \det(M + N Y_L(b))}.$$

We will now incorporate the variation of the boundary conditions into this framework. As before we restrict attention to those variations in the boundary conditions induced by automorphisms of the bundle. In addition we require, for simplicity only, that the automorphisms we consider leave fixed the underlying interval. Any such automorphism is specified by a smooth invertible $r \times r$-matrix valued function $U(x)$ defined on $[a, b]$. Given such a $U(x)$ there are invertible $nr \times nr$ matrices $u(a)$ and $u(b)$ so that for any section f,

$$T(U(x) f(x)) = \begin{pmatrix} u(a) & 0 \\ 0 & u(b) \end{pmatrix} T(f).$$

There is more that can be said about $u(a)$ and $u(b)$ but we will not need to know very much. We will use only that they are lower triangular matrices with 1's along the diagonal, so in particular

$$\det u(a) = \det u(b) = 1.$$

For any boundary condition A represented by (M, N), UA is represented by $(M u^{-1}(a), N u^{-1}(b))$, so that if f satisfies A then Uf satisfies UA. If L is any operator and A and B are any boundary conditions, then according to Theorem 2 we need to consider

$$U^{-1} \Phi_L U = (U^{-1} T_{U\alpha} P_{U\beta} U) \colon \beta \to \alpha.$$

If B is the Cauchy boundary condition then

$$U^{-1} \Phi_L U = U^{-1} T_{U\alpha} \begin{bmatrix} u(a) \\ Y_L(b)\, u(a) \end{bmatrix}.$$

In analogy with our discussion of the previous corollary, after choosing bases there is a non-zero constant $c \in \mathbf{C}$, independent of U, such that

$$\begin{aligned} \det U^{-1} \Phi_L U &= \det U^{-1} T_{U\alpha} \begin{bmatrix} u(a) \\ Y_L(b)\, u(a) \end{bmatrix} = \det T_\alpha U^{-1} \begin{bmatrix} u(a) \\ Y_L(b)\, u(a) \end{bmatrix} \\ &= \det T_\alpha \begin{bmatrix} I \\ u^{-1}(b)\,(Y_L(b))\, u(a) \end{bmatrix} = c \det(M + N u^{-1}(b)\, Y_L(b)\, u(a)) \\ &= c \det(M u^{-1}(a) + N u^{-1}(b)\, Y_L(b)). \end{aligned}$$

In this language, Corollary 3.6 becomes

Corollary 3.8. *Let L, $U(x)$, $u(a)$ and $u(b)$ be as above and A any boundary condition represented by (M, N). Then*

$$\frac{\operatorname{Det} L_{UA}}{\operatorname{Det} L_A} = \frac{\det(M u^{-1}(a) + N u^{-1}(b)\, Y_L(b))}{\det(M + N Y_L(b))}.$$

Proof. For any such function $U(x)$, we can find a one-parameter family $U_t(x)$ with $U_1(x) = U(x)$ and $U_0(x) =$ identity. Apply Corollary 3.6 to this family.

We can combine the two previous corollaries to arrive at our most general statement.

Proposition 3.9. *Let L and $\hat{L}$ be two operators of the form*

$$L = P_0(x)\frac{d^n}{dx^n} + P_1(x)\frac{d^{n-1}}{dx^{n-1}} + (\textit{lower order terms})$$

$$\hat{L} = P_0(x)\frac{d^n}{dx^n} + \hat{P}_1(x)\frac{d^{n-1}}{dx^{n-1}} + (\textit{lower order terms}).$$

Let A be any boundary condition represented by (M, N) and $U(x)$ any smooth invertible $r \times r$-matrix valued function with corresponding matrices $u(a)$ and $u(b)$. Then

$$\frac{\operatorname{Det} \hat{L}_{UA}}{\operatorname{Det} L_A} = \frac{\exp\left(\int_a^b \operatorname{tr} R(x)\hat{P}_1(x)P_0^{-1}(x)\,dx\right)\det(Mu^{-1}(a) + Nu^{-1}(b)\,Y_{\hat{L}}(b))}{\exp\left(\int_a^b \operatorname{tr} R(x)P_1(x)P_0^{-1}(x)\,dx\right)\det(M + NY_L(b))}.$$

We will now apply these results to some examples.

3.2. *Applications to ordinary differential operators*

3.2.1. The hessian of the energy function at a geodesic. Our first application will be to the determinant of the hessian of the energy function at a geodesic. The motivation for looking at such a thing comes from the problem of understanding the asymptotics of the heat kernel of a Riemannian manifold M^n. The heat kernel is the unique function $k(t, x, y)$ on $R^+ \times M^2$ such that $\left(\frac{\partial}{\partial t} + \frac{1}{2}\Delta\right)k(t, x, y) = 0$ and $\lim_{t \to 0} k(t, x, y) = \delta(x - y)$, where Δ is the (positive) Laplacian acting on functions on M. The heat kernel has become, of late, one of the basic tools for relating geometric, analytic and topological information. The main technique is to consider the asymptotic behavior of $k(t, x, y)$ as t approaches 0. The variation of this quantity from the heat kernel for $\mathbf{R}^n$, $\frac{1}{(2\pi t)^{n/2}}\exp\left(-\frac{|x-y|^2}{2t}\right)$, can be determined in a precise manner from the local geometric data.

One approach is to reason that for very small t and $x \neq y$, the heat that travels from x to y in time t must do so very efficiently, i.e. it must stay near a geodesic which travels from x to y. Thus, as t approaches 0, $k(t, x, y)$ should be expressible as a sum of contributions, one from each geodesic from x to y. This is made precise by the following statement [E1]

$$k(t, x, y) = \left(\sum_\gamma \frac{1}{(2\pi t)^{n/2}} e^{-\frac{|\gamma|^2}{2t}} D^{-\frac{1}{2}}(\gamma)\right)(1 + O(t))$$

where the sum is over all geodesics from x to y, $|\gamma|$ is the length of γ and $D(\gamma)$ is the determinant of the differential of the exponential map along γ.

One can arrive at this result almost effortlessly if one is willing to accept some purely formal manipulations. If we let $P_t(x, y)$ be the space of maps γ: $[0, t] \to M$ such that $\gamma(0) = x$ and $\gamma(t) = y$, then the amount of heat which travels from x to y in time t is equal to the integral over $P_t(x, y)$ of the amount of heat which travels along each path. From statistical mechanics we know that the amount of heat which travels along any given path γ is proportional to $\exp(-S(\gamma))$, where $S(\gamma) = \frac{1}{2}\int_0^t \left|\frac{d\gamma}{dt}\right|^2 dt$ is the energy it takes to travel along γ. Thus

$$k(t, x, y) = \frac{1}{N} \int_{P_t(x,y)} e^{-S(\gamma)} d\mu(\gamma)$$

where N is a normalizing constant. One should be warned that the measure $d\mu(\gamma)$ does not really exist, although $\exp(-S(\gamma))d\mu(\gamma)$ does. Reparameterizing γ proportionately so that it travels from x to y in time 1, we find

$$S(\gamma) = \frac{1}{2t}\int_0^t \left|\frac{d\gamma}{dt}\right|^2 dt.$$

Thus

$$k(t, x, y) = \frac{1}{N} \int_{P_1(x,y)} e^{-\frac{S(\gamma)}{2t}} dt.$$

Here we apply stationary phase methods. The critical points of S are the geodesics from x to y. Thus, applying a standard result of asymptotic analysis and using the fact that for a geodesic γ, $S(\gamma) = |\gamma|^2$, we should have

$$k(t, x, y) = \left(\frac{1}{N}\sum_{\gamma} e^{-\frac{|\gamma|^2}{2t}} \operatorname{Det}^{-\frac{1}{2}}(-2\pi t S'')\right)(1 + O(t))$$

where the sum is taken over all geodesics and S'' is the second variation of S, the energy functional. Of course, we need to make sense out of $\operatorname{Det}(-2\pi t S'')$. The second variation of the energy functional is a quadratic form on the tangent space to $P_1(x, y)$ at γ, i.e. on the space of all vector fields along γ which vanish at both endpoints.

Here is where we end our discussion of the heat kernel. We now "localize" to the problem of understanding the term $\operatorname{Det}(-2\pi t S'')$. If we are ever to consider ourselves successful, our answer should be closely related to $D(\gamma)$. This idea is carried out in [L-S] where they prove using Fredholm determinants that the corresponding ratios of determinants are equal. We verify their formula now and calculate the appropriate constants of proportionality.

Suppose γ is a geodesic which travels from a point x to a point y in time t. Given two vector fields V and W along γ we have the hessian applied to V and W, $H(V, W)$ (also called the index), which is the second derivative of the energy function in the direction of V and W. To be more precise, let $\gamma_{\mu,\nu}(t)$ be a two parameter family of paths such that $\gamma_{0,0} = \gamma$, $\left.\frac{\partial}{\partial\mu}\right|_{0,0} \gamma = V$ and $\left.\frac{\partial}{\partial\nu}\right|_{0,0} \gamma = W$. Then

$$H(V, W) = \left.\frac{\partial}{\partial\mu}\frac{\partial}{\partial\nu}\right|_{0,0} (\text{Energy of } \gamma_{\mu,\nu}) = \left.\frac{\partial}{\partial\mu}\frac{\partial}{\partial\nu}\right|_{0,0} \left(\frac{1}{2}\int_0^t \left|\frac{\partial}{\partial s}\gamma_{\mu,\nu}(s)\right|^2 ds\right).$$

This is independent of our two-parameter family.

Now we restrict our attention to those vector fields along γ which are elements of the tangent space of paths from x to y. These are exactly the vector fields which vanish at x and y. For such fields

$$H(V, W) = \int_\gamma \langle J(V), W \rangle$$

where J, the Jacobi operator, is defined by

$$J(V) = -\nabla_T \nabla_T V + R(T, V)\, T$$

where ∇ denotes covariant differentiation, T is the tangent vector along γ and R is the curvature operator (see [C-E]). Therefore, the determinant of the hessian of the energy function at γ is equal to the determinant of the Jacobi operator acting on vector fields along γ which vanish at x and y. Choose a frame for the tangent space at x and parallel translate to get a frame along γ. This allows us to write the Jacobi operator as $-\dfrac{d^2}{dx^2} + R(x)$, where $R(x)$ is an $n \times n$-matrix valued function and x is the parameter along γ and runs from 0 to t.

Note that in the notation of Corollary 3.7 the boundary conditions, which we will denote by A, can be represented by $\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix}\right)$ where I is the $n \times n$ identity matrix. If

$$\hat{L} = c\left(-\frac{d^2}{dx^2} + R(x)\right) \text{ and } L = c\left(-\frac{d^2}{dx^2}\right)$$

and $Y_{\hat{L}}$ and Y_L are the fundamental solutions of the corresponding first order equations, then Corollary 3.7 implies

$$\operatorname{Det} \hat{L}_A = \operatorname{Det} L_A \frac{\det\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix} Y_{\hat{L}}(t)\right)}{\det\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix} Y_L(t)\right)}.$$

We will now calculate each of the terms on the right hand side.

(i) $\operatorname{Det} L_A$: $L_A = -c\dfrac{d^2}{dx^2}$ acting on n-tuples of functions on the interval $[0, t]$ which vanish at the endpoints. The eigenvalues are $\left[\sin\left(\dfrac{\pi k}{t}\right)x\right]_i$, $k = 1, 2, \ldots$ and $i = 1, 2, \ldots, n$ where $[f(x)]_i$ is the n-tuple of functions which has $f(x)$ in the ith coordinate and 0 elsewhere. Therefore, the eigenvalues of L_A are $c\left(\dfrac{\pi k}{t}\right)^2$, each with multiplicity n. Thus

$$\zeta_L(s) = n \sum_k \left[c\left(\frac{\pi k}{t}\right)^2\right]^{-s} = n\left[c\left(\frac{\pi}{t}\right)^2\right] \sum_k k^{-2s} = n\left[\frac{t^2}{\pi^2 c}\right]^s \zeta_R(2s)$$

where $\zeta_R(s)$ is the Riemann zeta function $\sum_k k^{-s}$.

$$\zeta_L'(s)=n\log\left(\frac{t^2}{\pi^2 c}\right)\left(\frac{t^2}{\pi^2 c}\right)^s\zeta_R(s)+2n\left(\frac{t^2}{\pi^2 c}\right)^s\zeta_R'(2s)$$

$$\zeta_L'(0)=n\log\left(\frac{t^2}{\pi^2 c}\right)\zeta_R(s)+2n\zeta_R'(0)$$

$$=-\frac{n}{2}\log\left(\frac{t^2}{\pi^2 c}\right)-n\log(2\pi)=-\frac{n}{2}\log\left(\frac{(2t)^2}{c}\right).$$

Therefore

$$\operatorname{Det} L_A=e^{-\zeta'(0)}=\left[\frac{(2t)^2}{c}\right]^{n/2}.$$

(ii) $\det\left(\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}+\begin{bmatrix}0 & 0\\ I & 0\end{bmatrix}Y_L(t)\right)$: It is easily seen that $Y_L(x)=\begin{bmatrix}I & xI\\ 0 & I\end{bmatrix}$. Therefore

$$\det\left(\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}+\begin{bmatrix}0 & 0\\ I & 0\end{bmatrix}Y_L(t)\right)=\det\begin{bmatrix}I & 0\\ I & tI\end{bmatrix}=t^n.$$

(iii) $\det\left(\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}+\begin{bmatrix}0 & 0\\ I & 0\end{bmatrix}Y_{\tilde{L}}(t)\right)$: Write $Y_{\tilde{L}}$ as $\begin{bmatrix}Y_1(x) & Y_2(x)\\ Y_1'(x) & Y_2'(x)\end{bmatrix}$. Then

$$\det\left(\begin{bmatrix}I & 0\\ 0 & 0\end{bmatrix}+\begin{bmatrix}0 & 0\\ I & 0\end{bmatrix}Y_{\tilde{L}}\right)=Y_2(t).$$

Note that $Y_2(x)$ are the solutions to the Jacobi operator (i.e. the Jacobi fields) which vanish at x.

Let $v\in T_xM$, the tangent space to M at x, be the initial vector of γ, that is $\gamma=\exp_x(sv)$, $s\in[0,t]$. Then we can define the determinant of $d\exp_x(tv)$ by thinking of $d\exp_x(tv)$ as a map from T_xM to itself. This is done by using the canonical identification of $T_{tv}(T_xM)$ with T_xM and identifying $T_{\exp_x(tv)}M$ with T_xM by using the parallel framing along γ that we chose in order to view the Jacobi operator as an ordinary differential operator.

It is not hard to see that in these bases, $Y_2(t)=t(d\exp_x(tv))$. Note that this checks with the fact that in the flat case exp is an isometry so $d\exp=I$, where I is the identity map, while, as we saw in (ii), in this case $T_2(t)=tI$. To verify this in general, note that for $w\in T_{tv}(T_xM))$

$$[d\exp_x(tv)](w)=\frac{d}{d\alpha}\exp_x(tv+\alpha w)\Big|_{\alpha=0}$$

$$=\left[\frac{d}{d\alpha}\exp_x s\left(v+\alpha\left(\frac{w}{t}\right)\right)\Big|_{\alpha=0}\right]\Big|_{s=t}.$$

The last term is nothing more than the value at t of the Jacobi field which has derivative w/t at x, i.e.

$$\frac{1}{t}[Y_2(t)](w)=[Y_2(t)]\left(\frac{w}{t}\right)=[d\exp_x(tv)](w).$$

Therefore

$$\det\left(\begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ I & 0 \end{bmatrix} Y_{\hat{L}}(t)\right) = \det Y_2(t) = t^m \det(d\exp_x(t\,v)).$$

Combining (i), (ii) and (iii) yields

Proposition 3.10. *Let J = the Jacobi operator acting on vector fields along a geodesic $\gamma = \exp_x(s\,v)$ $s \in [0, t]$, which vanish at the endpoints. Then for any $c \in \mathbf{C}$*

$$\mathrm{Det}(c\,J) = 2^n c^{-\frac{n}{2}} \det(d\exp_x(t\,v)).$$

3.2.2. The covariant derivative operator around a loop. Let M be an n-dimensional Riemannian manifold and γ: $S^1 \to M$ a smooth map. Then using the Levi-Civita connection ∇ on M we get a covariant derivative operator ∇_T, where T is the tangent vector field of γ, which acts on vector fields along γ. In [At] Atiyah presents an argument of Witten in which the determinant of ∇_T is used to "prove" (although not rigorously) the index theorem for the Dirac operator. In [At] a formula for this determinant is given, and it is proved by explicitly calculating each side of the formula. We will verify this formula by using Corollary 3.7.

Choosing a frame along γ, ∇_T becomes $\frac{d}{dt}$ + (a skew-hermitian matrix-valued function). Using this frame we can think of vector fields V along γ as sections of the trivial vector bundle over $[0, 2\pi]$ which satisfy $V(0) = V(2\pi)$. This boundary condition is represented by the pair $(I, -I)$ where I is the $n \times n$ identity matrix. We apply Corollary 3.7 and note that $\mathrm{tr}\, R P_1 P_0^{-1} = 0$ since $P_0 = I$, and hence $R = 0$ or 1 by definition (4) and Lemma 3.1, and P_1 is skew-hermitian. Thus we learn

Proposition 3.11. *There is some constant c (depending only on n) such that*

$$\mathrm{Det}\,\nabla_T = c \det(I - Y(2\pi))$$

where $Y(2\pi)$ is the fundamental solution of ∇_T (since ∇_T is already a first order operator).

Note that $Y(2\pi)$ is, by definition, the parallel transport matrix.

To find the constant c in the proposition we calculate one example. Since $Y' Y^{-1}$ is skew-hermitian and $Y(0) = I$, $Y(x)$ is an orthogonal $n \times n$ matrix, and therefore $\det(I - Y(2\pi)) = 0$ if n is odd. Suppose $n = 2$ and let the skew-hermitian matrix-valued function be the constant $\begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix}$. We will calculate the constant c in this case. The constant when $n = 2m$ will then be c^m.

(i) The eigenvalues of $\nabla_T = \frac{d}{dt} + \begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix}$ are $\{i(k+\frac{1}{2}), i(k-\frac{1}{2})\}$, $k \in \mathbf{Z}$. Therefore

$$\mathrm{Det}\,\nabla_T = \prod_{k \in \mathbf{Z}} \left(ik + \frac{i}{2}\right)\left(ik - \frac{i}{2}\right) = \prod_{k \in \mathbf{Z}} (\tfrac{1}{4} - k^2)$$

$$= \tfrac{1}{4} \prod_{k>0} (k^2 - \tfrac{1}{4})^2 = \frac{1}{4}\left(\prod_{k>0} \frac{k^2 - \frac{1}{4}}{k^2}\right)^2 \left(\prod_{k>0} k^2\right)^2.$$

The first product converges absolutely.

$$\left(\prod_{k>0}\frac{k^2-\frac{1}{4}}{k^2}\right)^2=\left(\prod_{k>0}\left(1-\frac{1}{4k^2}\right)\right)^2=\left[\frac{\sin\frac{\pi}{2}}{\frac{\pi}{2}}\right]^2.$$

$(\prod_{k>0}k^2)^2$ can be calculated using the zeta function. That is, $(\prod_{k>0}k^2)^2=e^{-2\zeta'(0)}$ where $\zeta(s)=\sum(k^2)^{-s}=\zeta_R(2s)$ and ζ_R is the Riemann zeta function $\sum k^{-s}$. Therefore

$$-\zeta'(0)=-2\zeta_R'(0)=\log(2\pi)$$

so $e^{-2\zeta'(0)}=(2\pi)^2$. Thus

$$\operatorname{Det}\nabla_T=\frac{1}{4}\left[\frac{\sin^2(\frac{\pi}{2})}{\frac{\pi}{2}}\right]^2=4\sin^2(\tfrac{\pi}{2})=4.$$

(ii) It is easy to check that

$$Y(t)=\begin{bmatrix}\cos\left(\frac{t}{2}\right) & -\sin\left(\frac{t}{2}\right)\\ \sin\left(\frac{t}{2}\right) & \cos\left(\frac{t}{2}\right)\end{bmatrix}.$$

Substituting this into the formula yields

$$\det(I-Y(2\pi))=\det\begin{bmatrix}1-\cos(\pi) & \sin(\pi)\\ -\sin(\pi) & 1-\cos(\pi)\end{bmatrix}=4.$$

Thus the two determinants are equal and $c=1$. This agrees with Lemma 2 of [At].

Proposition 3.12. *With all notation as above*

$$\operatorname{Det}\nabla_T=\det(I-Y)$$

where Y is the parallel transport matrix around the loop.

3.2.3. Self-adjoint second order operators. Now we will investigate the problem of varying the boundary conditions while we vary the operator in order to always have a self-adjoint operator. A good reference for such operators is [Bo]. Let L be a positive, symmetric, second order operator on the interval $[a,b]$. That is, L is of the form

$$Ly=-(py'+qy)'-q^*y'+ry$$

where a prime denotes $\frac{d}{dx}$, * denotes the conjugate transpose, and p, q and r are smooth $m\times m$ matrix-valued functions satisfying

1. $p^*=p$, $r^*=r$.
2. p is positive definite.

There is a simple characterization of those boundary conditions A such that L_A is self-adjoint. If A is represented by (M, N) then let S and T be $2m \times 2m$ matrices such that $MS+NT=0$ and $\begin{bmatrix} S \\ T \end{bmatrix}$ has rank $2m$. Define S_L, T_L and J by

$$S_L=\begin{pmatrix} 1 & 0 \\ q(a) & p(a) \end{pmatrix}, \quad T_L=\begin{pmatrix} 1 & 0 \\ q(b) & p(b) \end{pmatrix}, \quad J=\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

then L_A is self-adjoint if and only if

$$S_L^* J S_L = T_L^* J T_L.$$

This can be seen as follows: From the choice of S and T it follows that for every function F such that

$$M\begin{pmatrix} f(a) \\ f'(a) \end{pmatrix}+N\begin{pmatrix} f(b) \\ f'(b) \end{pmatrix}=0$$

there exists a vector $v_f \in \mathbf{C}^{2m}$ such that

$$\begin{pmatrix} f(a) \\ f'(a) \end{pmatrix}=S v_f \begin{pmatrix} f(b) \\ f'(b) \end{pmatrix}=T v_f.$$

$$\begin{aligned} \int_a^b \langle Lf, g\rangle - \langle f, Lg\rangle \, dx &= \left\langle J\begin{pmatrix} 1 & 0 \\ q(a) & p(a) \end{pmatrix}\begin{pmatrix} f(a) \\ f'(a) \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ q(a) & p(a) \end{pmatrix}\begin{pmatrix} g(a) \\ g'(a) \end{pmatrix}\right. \\ &= \langle J S_L v_f, S_L V_g\rangle - \langle J T_L v_f, T_L v_g\rangle \\ &= \langle S_L^* J S_L v_f, v_g\rangle - \langle T_L^* J T_L v_f, v_g\rangle. \end{aligned}$$

This is 0 for every f, g satisfying A if and only if

$$S_L^* J S_L = T_L^* J T_L.$$

If this is true, $\operatorname{rank}[M, N]=2m$ implies that L_A is actually self-adjoint.

Now we work backwards. Let S, T be any pair of $2m \times 2m$ matrices such that $\begin{bmatrix} S \\ T \end{bmatrix}$ has rank $2m$ and $S^*JS=T^*JT$. Given an operator L there is a natural way to associate a boundary condition A_L to S and T such that L restricted to A_L is self-adjoint. Namely, we choose M and N such that $[M, N]$ has rank $2m$ and $MS+NT=0$, let A_L be the boundary condition represented by

$$(M_L, N_L)=\left(M\begin{pmatrix} 1 & 0 \\ q(a) & p(a) \end{pmatrix}, N\begin{pmatrix} 1 & 0 \\ q(b) & p(b) \end{pmatrix}\right).$$

Note that A_L depends only on S and T, not on our choice of M and N. If we have another operator $\hat{L}$ with

$$\hat{L}=(p y'+\hat{q} y)'-\hat{q}^* y'+\hat{r} y$$

then

$$M_{\hat{L}} = M_L \begin{pmatrix} 1 & 0 \\ \hat{q}(a) - q(a) & 1 \end{pmatrix}, \quad N_{\hat{L}} = N_L \begin{pmatrix} 1 & 0 \\ \hat{q}(b) - q(b) & 1 \end{pmatrix}.$$

It is easy to see that there is some function $U(x)$ on $[a, b]$ such that

$$u^{-1}(a) = \begin{pmatrix} 1 & 0 \\ \hat{q}(a) - q(a) & 1 \end{pmatrix}, \quad u^{-1}(b) = \begin{pmatrix} 1 & 0 \\ \hat{q}(b) - q(b) & 1 \end{pmatrix}.$$

(See the discussion following Corollary 3.7 for notation.) Therefore we can apply Proposition 3.9 to find

Proposition 3.13. *With all notation as above*

$$\operatorname{Det} \hat{L}_{A_{\hat{L}}} = \operatorname{Det} L_{A_L} \frac{\exp\left(-\frac{1}{2}\int_a^b \operatorname{tr}((-p' - \hat{q} - \hat{q}^*) p^{-1})\right) \det(M_{\hat{L}} - N_{\hat{L}} Y_{\hat{L}}(b))}{\exp\left(-\frac{1}{2}\int_a^b \operatorname{tr}((-p' - q - q^*) p^{-1})\right) \det(M_L + N_L Y_L(b))}.$$

References

[Ag] Agmon, S.: On the eigenfunctions and eigenvalues of general elliptic boundary value problems. Commun. Pure Appl. Math. **15**, 119–147 (1962)

[At] Atiyah, M.F.: Circular symmetry and stationary phase approximation. Proceedings of the Conference in Honor of L. Schwartz. Astérisque **131**, 43–60 (1985)

[A-P-S] Atiyah, M.F., Patodi, V., Singer, I.M.: Spectral asymmetry and Riemannian geometry I. Math. Proc. Camb. Philos. Soc. **77**, 43–69 (1975)

[Bo] Bott, R.: On the iteration of closed geodesics and the Sturm intersection theory. Commun. Pure Appl. Math. **9**, 171–206 (1956)

[Ca] Calderon, A.P.: Boundary value problems for elliptic operators. Outline of the Joint Soviet-American Symposia on Partial Differential Equations, Novosibrush, August 1963

[C-E] Cheeger, J., Ebin, D.G.: Comparison theorems in Riemannian geometry. New York: North Holland 1975

[D-D] Dreyfuss, T., Dym, H.: Product formulas for the eigenvalues of a class of boundary value problems. Duke Math. J. **45**, 15–37 (1978)

[El] Elie, L.: Équivalent de la densité d'une diffusion en temps petit. Cas des points proches. Géodésiques et Diffusion en Temps Petit. Astérique **84–85**, 55–71 (1981)

[Fo] Forman, R.: Functional determinants and applications to geometry. Doctoral Thesis, Harvard University 1985

[G-K] Gohberg, I.C., Krein, M.G.: Introduction to the theory of linear nonselfadjoint operators, Vol. 18 (Translation of Mathematical Monographs). Providence, Rhode Island, American Mathematical Society 1969

[Hö] Hörmander, L.: Pseudo-differential operators and non-elliptic boundary problems. Ann. Math. **83**, 129–209 (1966)

[L-S] Levit, S., Smilansky, U.: A theorem on infinite products of eigenvalues of Sturm-Liouville type operators. Proc. Am. Math. Soc. **65**, 299–302 (1977)

[Pa] Palais, R.: Seminar on the Atiyah-Singer index theorem. Princeton, Princeton University Press 1965

[R-S1] Ray, D.B., Singer, I.M.: *R*-Torsion and the Laplacian on Riemannian manifolds. Adv. Math. **7**, 145–210 (1971)

[R-S2] Ray, D.B., Singer, I.M.: Analytic torsion for complex manifolds. Ann. Math. **98**, 154–177 (1973)

[Se1] Seeley, R.: Singular integrals and boundary value problems. Am. J. Math. **88**, 154–177 (1966)

[Se2] Seeley, R.: Complex powers of an elliptic operator. Singular Integrals (Proceedings of the Symposia in Pure Mathematics, Chicago, Illinois 1966), Providence, Am. Math. Soc. pp. 288–307

[Se3] Seeley, R.: Analytic extension of the trace associated with an elliptic boundary problems. Am. J. Math. **91**, 963–983 (1969)

Oblatum 4-VI-1986

Invent. math. 88, 495–520 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Local rigidity of discrete groups acting on complex hyperbolic space

W.M. Goldman and J.J. Millson* **

Department of Mathematics, University of California, Los Angeles, CA 90024, USA

Introduction

The superrigidity theorem of Margulis, see Zimmer [17], classifies finite dimensional representations of lattices in semi-simple Lie groups of real rank strictly larger than 1. It is a fundamental problem to obtain the classification of finite dimensional representations of lattices in rank 1 semi-simple Lie groups. That this problem will be considerably harder than the previous one is suggested by the existence of continuous families of inequivalent representations or, in other words, the existence of non-trivial deformations. In Johnson-Millson [6] and Kourouniotis [8] such deformations were constructed for certain representations of lattices in $SO(n, 1)$ based on a construction of Thurston called bending. The deformation space of the representation of a lattice $\Gamma \subset SO(n, 1)$ obtained by restricting an inclusion $SO(n, 1) \to SO(n+1, 1)$ to Γ is of particular interest. If $n > 2$ the space of infinitesimal deformations is $H^1(\Gamma, \mathbf{R}^{n+1})$ where Γ acts on $\mathbf{R}^{n+1}$ by the restriction of the standard action of $SO(n, 1)$. The space of infinitesimal deformations is non-zero for the standard arithmetic examples and the main point of the papers cited above was to establish that some of these infinitesimal deformations are integrable (in [6] it is also shown that some are not).

In this paper, we study the complex analogue of the above example. We let Γ be a cocompact torsion free lattice in $SU(n, 1)$ and consider the deformation space of the representation of Γ obtained by restricting an inclusion $SU(n, 1) \to SU(n+1, 1)$. If $n > 1$ the space of infinitesimal deformations is $H^1(\Gamma, \mathbf{R}) \oplus H^1(\Gamma, \mathbf{C}^{n+1})$. In the first summand Γ acts trivially and the infinitesimal deformations are tangent to the obvious deformations obtained by deforming Γ in $U(n, 1)$ by a curve of homomorphisms into the center of $U(n, 1)$ (observe that the above inclusion factors as $SU(n, 1) \to U(n, 1) \to SU(n+1, 1)$). In the second summand Γ acts by the restriction of the standard action of $SU(n, 1)$. This summand is non-zero for the standard arithmetic examples,

* *Present address:* J.J. Millson, Department of Mathematics, University of Maryland, College Park, MD 20742, USA

** Both authors were partially supported by NSF grants

Borel-Wallach [1], 5.9 and is the complex analogue of the space of infinitesimal deformations in the real case considered above. Thus, it is rather surprising that *none* of these latter infinitesimal deformations are integrable (Theorem 4.2 of this paper). This fact is the main technical result behind our main theorem. We now give two statements of our main theorem; the first one geometric, the second one group-theoretic.

Let D^{n+1} be the unit ball in $\mathbf{C}^{n+1}$ equipped with the metric of constant holomorphic sectional curvature-1 (complex hyperbolic space).

Theorem 1. *Let Γ be a torsion free group acting isometrically and properly discontinuously on D^{n+1} in such a way as to stabilize a totally geodesic n-ball D^n. Assume the quotient $\Gamma\backslash D^n$ is compact. Then all nearby isometric actions of Γ on D^{n+1} also stabilize a totally geodesic n-ball.*

We now give the group-theoretic version of our theorem. We define a representation ρ from Γ into $SU(n+1,1)$, see Sect. 1 for notation, to be *Fuchsian* if it is discrete and faithful and leaves invariant a line in $\mathbf{C}^{n+2}$ containing a vector of positive length for the hermitian form of signature $(n+1,1)$. Equivalently, ρ is Fuchsian if it leaves invariant a totally geodesic n-ball contained in the $n+1$-ball. We can now state our main theorem.

Theorem 2. *The set of Fuchsian representations is a connected component in the strong (or classical) topology on* $\mathrm{Hom}(\Gamma, SU(n+1,1))$.

In fact we give a complete description of the component. It is a locally algebraic subset diffeomorphic to the smooth manifold $(SU(n+1,1)/S^1)\times T$ where T is the torus $\mathrm{Hom}(\Gamma, S^1)$ and S^1 is embedded as the center Z of the subgroup (isomorphic to $U(n,1)$) of $SU(n+1,1)$ stabilizing the first standard basis vector. In particular, we observe that if ρ_t is any curve in $\mathrm{Hom}(\Gamma, SU(n+1,1))$ containing a Fuchsian representation ρ then there exist g_t in $SU(n+1,1)$ and χ_t in $\mathrm{Hom}(\Gamma, Z)$ such that $\rho_t = \mathrm{Ad}\, g_t \circ (\rho\chi_t)$.

We now sketch the proof of our theorem. We let $\mathscr{R}_0$ denote the set of Fuchsian representations. In the first section we assume $\mathscr{R}_0$ is not open and use the Curve Selection Lemma of F. Bruhat and H. Cartan, see Milnor [11], to find a curve of representations ρ_t intersecting $\mathscr{R}_0$ at a single point ρ. We also prove that $\mathscr{R}_0$ is closed in $\mathrm{Hom}(\Gamma, SU(n+1,1))$.

In Sect. 2 we construct a quadratic form Q on $H^1(\Gamma, \mathbf{C}^{n+1})$. Here $\mathbf{C}^{n+1}$ is identified with the orthogonal complement in $\mathbf{C}^{n+2}$ of the invariant line L of ρ and Γ operates by the action induced by ρ twisted by a unitary character. The form Q is defined as follows. The given Hermitian form on $\mathbf{C}^{n+2}$ induces a form of signature $(n,1)$ on the orthogonal complement of L. Using the imaginary part of this form we construct a symmetric bilinear form $A(\cdot,\cdot)$ on $H^1(\Gamma, \mathbf{C}^{n+1})$ with values in $H^2(\Gamma, \mathbf{R})$. But $H_2(\Gamma, \mathbf{R})$ contains a canonical class Z, the $(n-1)$-fold self-intersection of the hyperplane section class. We define $Q(c)$, for $c\in H^1(\Gamma, \mathbf{C}^{n+1})$, to be the Kronecker index of $A(c,c)$ and Z. In Sect. 2 we prove by examining the Taylor expansion of ρ_t around ρ that the assumption that $\mathscr{R}_0$ is not open implies that Q is isotropic.

In the last two sections we prove that Q is in fact positive definite contradicting the assumption that $\mathscr{R}_0$ was not open. The fact that Q is positive

definite turns out to be equivalent to a special case of a general vanishing theorem of Matsushima-Murakami [9]. In the case $n=1$, it was proved in Goldman [5] by a different method that Q was positive definite. Also Theorem 2 was proved in [5] again by a different method for the case $n=1$. For this reason we may make the assumption in this paper that $n \geqq 2$.

As remarked at the end of this paper, our results generalize to the case of a torsion-free group acting isometrically and properly discontinuously on D^{n+k} stabilizing a totally geodesic D^n such that $\Gamma \backslash D^n$ is compact. We again call such an action of Γ *Fuchsian*. It seems reasonable to look for a "strong rigidity" theorem corresponding to our "local rigidity" results. However, it is important to observe that there is an invariant (under conjugation by $SU(n+k, 1)$) of isometric actions of Γ on D^{n+k} obtained as follows. The Kahler form ω on D^{n+k} (properly normalized) corresponds, by the van Est Theorem, to a continuous Eilenberg-MacLane cohomology class $\omega \in H^2(SU(n+k, 1), \mathbf{R})$ which is the image of an integral class. Taking exterior powers we get a class $\omega^n \in H^{2n}(SU(n+k, 1), \mathbf{R})$. If ρ is a representation of Γ in $SU(n+k, 1)$ we can pull back ω^n by ρ to obtain a class $\rho^* \omega^n$ in $H^{2n}(\Gamma, \mathbf{R})$. Evaluating $\rho^* \omega^n$ on the fundamental class in $H_{2n}(\Gamma, \mathbf{Z})$ we obtain an integer which we denote $v(\rho)$ - in the case ρ is Fuchsian $v(\rho)$ is the volume of $\Gamma \backslash D^n$. We observe that v is constant on each component of $\mathrm{Hom}(\Gamma, SU(n+k, 1))$.

We now conjecture that *if* $\rho \in \mathrm{Hom}(\Gamma, SU(n+k, 1))$ *satisfies* $v(\rho) = \mathrm{vol}(\Gamma \backslash D^n)$ *then* ρ *is Fuchsian*. There is considerable evidence in addition to the main result of this paper that this strong rigidity result should hold. The conjecture has been proved for the case $n=1$ and all k in unpublished work of D. Toledo. Also if ρ is a representation of Γ into $SU(n+k, 1)$ there is a Γ-equivariant smooth map $f: D^n \to D^{n+k}$. Results of Faran [3] and Webster [15] show that it suffices to prove that f is a holomorphic embedding.

We thank E. Bierstone and Y.T. Siu for helpful conversations.

1. Fuchsian representations and their deformations

In this section we prove that the set of Fuchsian representations is a locally algebraic subset of $\mathrm{Hom}(\Gamma, G)$. That is, we construct a real algebraic subset X of $\mathrm{Hom}(\Gamma, G)$ and a strongly open subset U of $\mathrm{Hom}(\Gamma, G)$ such that $X \cap U$ is the set of Fuchsian deformations. X will consist of the representations which have an invariant line and U will consist of the representations which do not leave invariant a null-line or a negative line (in fact U is slightly smaller than this). We then assume that the set of Fuchsian representations is not open in $\mathrm{Hom}(\Gamma, G)$. We apply the Curve Selection Lemma to deduce the existence of a real analytic curve ρ_t in U meeting the space of Fuchsian representations at a single point.

We now establish some notation. In what follows G will denote the Lie group $SU(n+1, 1)$ the subgroup of the unimodular matrices consisting of the (unimodular) isometries of the hermitian form $F: \mathbf{C}^{n+2} \to \mathbf{R}$ given by:

$$F(z_1, \ldots, z_{n+2}) = |z_1|^2 + \ldots + |z_{n+1}|^2 - |z_{n+2}|^2.$$

We let $\langle , \rangle$ denote the sesquilinear form associated to F. We will assume throughout that $n \geqq 2$. We will sometimes let W denote $\mathbf{C}^{n+2}$.

We now give a structure of a real algebraic group to the complexification $\underline{G}$ of G so that the associated real points are G. We note $\underline{G} = SL_{n+2}(\mathbf{C})$. It is sufficient to define an antilinear action of $\mathbf{Z}/2$ (the Galois group of $\mathbf{C}$ over $\mathbf{R}$) on $\underline{G}$. We let r be the $n+2$ by $n+2$ diagonal matrix with diagonal entries $(1, \ldots, 1, -1)$. We make the non-trivial element of $\mathbf{Z}/2$ act by the anti-holomorphic involution σ defined by:

$$\sigma(g) = r(g^*)^{-1} r.$$

Here g^* is the usual adjoint of g; that is, $g^* = {}^t\bar{g}$ where the superscript t denotes transpose. We have now given a real structure to $SL_{n+2}(\mathbf{C})$ - a polynomial function f on $SL_{n+2}(\mathbf{C})$ is real if:

$$\overline{f(\sigma(g))} = f(g).$$

Here $\overline{f(\sigma(g))}$ denotes the complex conjugate of the complex number $f(\sigma(g))$. Clearly the set of real points of $SL_{n+2}(\mathbf{C})$ for the above real structure is $SU(n+1, 1)$. We often abuse notation by letting the same symbol denote both the set of complex points and the algebraic group. We may use the involution σ to give a real structure to $\underline{G}^N$ for any integer $N > 1$. Then if Γ is a group generated by N elements we obtain an induced real structure on $\mathrm{Hom}(\Gamma, \underline{G})$.

In what follows $\underline{G}_0$ will denote the real algebraic subgroup of $\underline{G}$ consisting of those elements which leave invariant the line through e_1 where $\{e_1, e_2, \ldots, e_{n+2}\}$ is the standard basis for $\mathbf{C}^{n+2}$. We will denote by V the subvector space of $\mathbf{C}^{n+2}$ consisting of those vectors with first component zero. We let G_0 denote the real points of $\underline{G}_0$. Then G_0 is isomorphic to $U(n, 1)$ and the induced action of G_0 on V is equivalent to the standard representation of $U(n, 1)$ on $\mathbf{C}^{n+1}$. We will often identify G_0 with $U(n, 1)$ and V with $\mathbf{C}^{n+1}$. We will also use $\langle , \rangle$ to denote the induced sesquilinear form on V.

Let Γ be a torsion-free group and $\rho_0 : \Gamma \to U(n, 1)$ be a faithful representation of Γ as a cocompact discrete subgroup of $U(n, 1)$. Then Γ operates via ρ_0 on the unit ball D^n in $\mathbf{C}^n$ as isometries of the Bergmann metric, see Kobayashi-Nomizu [7], p. 282. In [7], D^n, together with the Bergmann metric is called *complex hyperbolic space*. We have adopted this terminology in our title. The quotient space $M = \Gamma \backslash D^n$ is a compact Kahler manifold. A suitable multiple of ω, the imaginary part of the Bergmann metric, is the Chern form of a line bundle over M which is a fortiori positive. Hence, by the Kodaira Embedding Theorem, there exists a complex analytic embedding from M into a complex projective space $\mathbf{P}^N(\mathbf{C})$. A suitable multiple of ω is the Poincare dual of the hyperplane section class in M, the homology class determined by the intersection of the image of M with a generic hyperplane in $\mathbf{P}^N(\mathbf{C})$. We now investigate the connected component of $\mathrm{Hom}(\Gamma, U(n, 1))$ containing ρ_0.

Since $\rho_0(\Gamma)$ is torsion free we see that the intersection of $\rho_0(\Gamma)$ and the center of $U(n, 1)$ is $\{e\}$. Consequently $\rho_0(\Gamma)$ projects isomorphically into the simple Lie group $PU(n, 1)$. We now prove a general theorem which will imply that for any Lie group G the set of discrete faithful representations of Γ into G

form a closed subset of $\mathrm{Hom}(\Gamma, G)$. We observe that $\rho_0(\Gamma)$ is Zariski dense in $PU(n, 1)$ by the Borel Density Theorem [12], Chap. V.

Theorem 1.1. *Let Γ be a torsion free subgroup of a connected semi-simple linear Lie group H. Assume Γ is Zariski dense in H. Let G be any linear Lie group. Then the discrete faithful representations of Γ into G form a closed subset of* $\mathrm{Hom}(\Gamma, G)$.

The theorem is a consequence of the next two lemmas.

Lemma 1.1. *Suppose Γ is a torsion free group containing no non-trivial nilpotent normal subgroups. Let G be a linear Lie group. Then the set of discrete faithful representations is closed in* $\mathrm{Hom}(\Gamma, G)$.

Proof. Choose a neighbourhood U of the identity in G such that if Γ is any discrete subgroup of G then $\Gamma \cap U$ is contained in a connected nilpotent Lie subgroup of G, see Raghunathan [12], Theorem 8.16.

Now let $\{\rho_n\}$ be a sequence of discrete faithful representations converging to a representation ρ. We first show that ρ is faithful. Let $K = \mathrm{Ker}\,\rho$. Then K is normal. We claim K is nilpotent. Since ρ_n embeds Γ into a linear Lie group and Γ is torsion-free, there is an upper bound to the length of the descending central series and it is sufficient to show every finitely generated subgroup of K is nilpotent. Let $\{\gamma_1, \gamma_2, \ldots, \gamma_r\}$ be a finite subset of K. Then there exists n so that $\{\rho_n(\gamma_1), \ldots, \rho_n(\gamma_r)\} \subset U$. Consequently the image of the subgroup K' generated by $\{\gamma_1, \gamma_2, \ldots, \gamma_r\}$ is nilpotent. Hence K' is nilpotent and consequently K is also. Hence $K = \{e\}$ and ρ is faithful.

We now prove ρ is discrete. Suppose $\rho(\Gamma)$ is not discrete. Let N be the closure of $\rho(\Gamma)$ in G so N is a Lie group. Let N^0 be the connected component of the identity in N. Then $N^0 \neq \{e\}$ for otherwise N would be discrete and consequently $\rho(\Gamma)$ would also be discrete. Now N^0 is generated by any neighborhood of the identity hence by $N^0 \cap U$. Consequently $U \cap \Gamma$ generates a dense subgroup of N^0 and N^0 is nilpotent. Since N^0 is normal in N we see $\rho^{-1}(N^0)$ is normal in Γ. But since ρ is faithful $\rho^{-1}(N^0)$ is nilpotent and consequently $N^0 = \{e\}$, a contradiction. With this the lemma is proved.

Lemma 1.2. *Let H be a connected semi-simple linear Lie group with center Z_H. Let Γ be a closed subgroup of H which is Zariski dense in H and such that $\Gamma \cap Z_H = \{e\}$. Then Γ has no non-trivial nilpotent normal subgroups.*

Proof. Suppose $A \subset \Gamma$ is a nilpotent normal subgroup of H. Let M be the Zariski closure of A. Then M is normalized by Γ. Indeed consider the set $N = \{x \in H : \gamma x \gamma^{-1} \in M, \forall\, \gamma \in \Gamma\}$. Then N is an algebraic set containing A so $M \subset N$. Since Γ is Zariski dense in M we find that G normalizes M. But M is nilpotent so M is discrete hence central. Hence $A \subset Z_H$ so $A = \{e\}$.

We now examine the connected component of ρ_0 in $\mathrm{Hom}(\Gamma, U(n, 1))$ using the rigidity theorems of Weil and Mostow. We let $\{\mathscr{R}_i : i = 0, 1, \ldots, l\}$ be the set of components of $\mathrm{Hom}(\Gamma, U(n, 1))$. We number the components so that $\mathscr{R}_0$ is the connected component of ρ_0.

Theorem 1.2. *The connected component of ρ_0 coincides with the set of discrete faithful representations. The group $PU(n,1)\times \mathrm{Hom}(\Gamma, S^1)$ acts simply transitively on it.*

The theorem is a consequence of the next two lemmas.

Lemma 1.3. *The set of discrete faithful representations of Γ into $U(n,1)$ is a union of components.*

Proof. We show that the set of discrete faithful representations of Γ into $U(n,1)$ is open and closed in $\mathrm{Hom}(\Gamma, U(n,1))$. We observe that since $\rho_0(\Gamma)$ is cocompact we have $H_{2n}(\Gamma, \mathbf{R}) \neq \{0\}$. Consequently any discrete faithful representation of Γ is cocompact. By Weil [16] the set of discrete faithful representations of Γ is open in the strong topology. But by Theorem 1.1 the set of discrete faithful representations of Γ into $U(n,1)$ is closed and the lemma is proved.

We observe that we have an action of $PU(n,1)\times \mathrm{Hom}(\Gamma, S^1)$ on $\mathrm{Hom}(\Gamma, U(n,1))$ by:

$$(g, \chi)\cdot \rho = \mathrm{Ad}\, g \circ \chi\rho.$$

Here $\chi\rho$ is the representation given by:

$$\chi\rho(\gamma) = \chi(\gamma)\,\rho(\gamma).$$

Lemma 1.4. *$PU(n,1)\times \mathrm{Hom}(\Gamma, S^1)$ acts simply transitively on the set of discrete faithful representations of Γ in $U(n,1)$.*

Proof. We have a central extension $S^1 \to U(n,1) \xrightarrow{\pi} PU(n,1)$. We observe that if $\rho: \Gamma \to U(n,1)$ is discrete and faithful then $\rho(\Gamma)\cap S^1 = \{e\}$. Hence $\pi\circ\rho$ is also discrete and faithful. But by the Mostow Rigidity Theorem the group $PU(n,1)$ acts transitively on the discrete faithful representations of Γ into $PU(n,1)$. Hence any two elements ρ_1, ρ_2 in $\mathrm{Hom}(\Gamma, U(n,1))$ satisfy $\pi(\rho_1) = \mathrm{Ad}\, g \circ \pi(\rho_2)$ for some g in $U(n,1)$. Hence $\rho_1 = \chi g \circ \rho_2$ for some character χ. This proves that $PU(n,1)\times \mathrm{Hom}(\Gamma, S^1)$ acts transitively. To prove that the action is free it is sufficient to compute the isotropy at ρ_0. Suppose there exist $h \in PU(n,1)$ and $\chi \in \mathrm{Hom}(\Gamma, S^1)$ such that:

$$\mathrm{Ad}\, h \circ \chi\rho_0 = \rho_0.$$

Taking determinants we see that χ has order $n+1$. Letting Γ' be the kernel of χ we find that Γ' is Zariski dense and hence $h = 1$. Hence $\chi = 1$ and the lemma is proved.

Corollary 1.4.1. *The set of discrete faithful representations of Γ into $U(n,1)$ consists of a single component.*

We now define an embedding j from $U(n,1)$ into $SU(n+1,1)$ as follows. Suppose $h \in U(n,1)$ has matrix (h_{ik}) for $1 \leqq i, k \leqq n+1$. Then we define $j(h) \in SU(n+1,1)$ by the formula:

$$j(h)\, e_1 = (\det h)^{-1}\, e_1,$$
$$j(h)\, e_k = h_{i-1,\,k-1}\, e_i \quad \text{for } 2 \leqq i,\ k \leqq n+2.$$

That is:

$$j(h)=\begin{pmatrix}\det h^{-1} & 0\\ 0 & h\end{pmatrix}.$$

Lemma 1.5. *j is a proper map.*

Proof. j is clearly a homeomorphism onto its image and the image of j is described by the linear equations $\{z_{1k}=0\colon 2\leqq k\leqq n+2\}$. Hence the image of j is closed and j is proper.

We let j_* denote the induced map

$$j_*\colon \mathrm{Hom}(\Gamma, U(n,1))\to \mathrm{Hom}(\Gamma, SU(n+1,1)$$

defined by $j_*(\rho)=j\circ\rho$. The image of j_* is again cut out by linear equations and we obtain a lemma.

Lemma 1.6. *j_* is a proper map.*

Corollary 1.6.1. *The decomposition* $j_*\,\mathrm{Hom}(\Gamma, U(n,1))=\bigcup_{i=0}^{l} j_*(\mathscr{R}_i)$ *is the decomposition into components.*

We let ρ_0 denote also the composition $j\circ\rho_0$. This composition gives rise to an action of Γ on the unit ball D^{n+1} in $\mathbf{C}^{n+1}$ leaving invariant the linearly embedded n-ball defined by the equation $z_1=0$.

We will be concerned in this paper with the deformations of ρ_0 in $\mathrm{Hom}(\Gamma, G)$. Such deformations would be the analogues for complex hyperbolic space of classical quasi-Fuchsian groups in hyperbolic 3-space.

There are certain obvious deformations of ρ_0. First, there are the trivial deformations. Let g_t be a curve in $SU(n+1,1)$. Then we obtain a deformation ρ_t of ρ_0 by defining $\rho_t=\mathrm{Ad}\,g_t\circ\rho_0$. Such a deformation is said to be a trivial deformation. We may also deform ρ_0 by characters of Γ. Suppose that the dimension of $H^1(\Gamma, \mathbf{R})$ is greater than zero. Then we may find a curve χ_t in $\mathrm{Hom}(\Gamma, S^1)$ and we obtain a deformation ρ_t of ρ_0 to be denoted $\rho_0 * \chi_t$ by the formula:

$$\rho_t(\gamma)\,e_1=\chi_t(\gamma)^{-(n+1)}\rho_0(\gamma)\,e_1,$$
$$\rho_t(\gamma)\,e_j=\chi_t(\gamma)\,\rho_0(\gamma)\,e_j \quad \text{for } j\geqq 2.$$

We have $\rho_0 * \chi_t = j(\chi_t\,\rho_0)$.

Definition. We say a deformation ρ_t of ρ_0 is Fuchsian if there exists a curve g in G and a curve χ_t in $\mathrm{Hom}(\Gamma, S^1)$ such that:

$$\rho_t=\mathrm{Ad}\,g_t\circ(\rho_0 * \chi_t).$$

Thus the space of Fuchsian deformations of ρ_0 is just the saturation by G of the space $j_*\,\mathrm{Hom}(\Gamma, U(n,1))$. We wish to analyse this space which will be denoted $R(\rho_0)$. The following general lemma will be very useful to us. Let $\mathbf{P}(W)$ denote the projective space of complex lines in W.

Lemma 1.7. *Let X be compact subset of $\mathbf{P}(W)$. The subset of representations which have a fixed point in X is closed in* $\mathrm{Hom}(\Gamma, G)$.

Proof. We consider the map $F: G^N \times X \to X^{N+1}$ given by:

$$F(g_1, \ldots, g_N, x) = (x, g_1 x, \ldots, g_N x).$$

If Δ denotes the diagonal in X^{N+1}, the condition that ρ has a fixed point in X is that there exists $x \in X$ (the fixed point) such that $(g_1, \ldots, g_N, x) \in F^{-1}(\Delta)$. Let $p_1: G^N \times X \to G^N$ denote the projection. Since the fiber of p_1 is compact, p_1 is a proper and hence a closed map. Hence $p_1(F^{-1}(\Delta))$ is closed and the lemma is proved.

We apply the lemma in two cases. In the first case we let C denote the light cone in W; that is, the set described by:

$$C = \{x \in W: \langle x, x \rangle = 0\}.$$

We let $X = \mathbf{P}C$, the image of C in $\mathbf{P}(W)$.

Corollary 1.7.1. *The set of $\rho \in \mathrm{Hom}(\Gamma, G)$ which do not leave a point in $\mathbf{P}C$ fixed is open.*

For our second corollary, we observe that $\mathbf{P}(W)$ is separated by $\mathbf{P}C$ into the disjoint union of $\mathbf{P}(W)_+$, the set of positive lines in W and $\mathbf{P}(W)_-$, the set of negative lines in W. Here a line L in W is said to be positive (resp. negative) if $\langle , \rangle | L$ is positive definite (resp. negative definite). By taking $X = \mathbf{P}(W)_- \cup \mathbf{P}C$ we obtain the following corollary.

Corollary 1.7.2. *The set of $\rho \in \mathrm{Hom}(\Gamma, G)$ which do not leave a point in $\mathbf{P}(W)_- \cup \mathbf{P}C$ fixed is open in $\mathrm{Hom}(\Gamma, G)$.*

We let U' denote this open set. We need a variation on Lemma 1.7 using the Zariski topology.

Lemma 1.8. *Let X be the subset of $\mathrm{Hom}(\Gamma, \underline{G})$ consisting of all representations that fix a line. Then X is Zariski closed in $\mathrm{Hom}(\Gamma, \underline{G})$. Moreover X is defined over $\mathbf{R}$.*

Proof. We consider the map $F: \underline{G}^N \times \mathbf{P}(W) \to \mathbf{P}(W)^{N+1}$ given by the formula:

$$F(g_1, \ldots, g_N, x) = (x, g_1 x, \ldots, g_N x).$$

Then $X = p_1 F^{-1}(\Delta)$. Since p_1 is proper in the Zariski topology X is Zariski closed.

To see that X is defined over $\mathbf{R}$ we have only to check that if $\rho \in \mathrm{Hom}(\Gamma, G)$ fixes a line then $\sigma(\rho)$ fixes a line. Suppose ρ fixes L. Choose g so that $gL = \mathbf{C}e_1$. Then $\mathrm{Ad}\, g \circ \rho$ fixes $\mathbf{C}e_1$. Hence $\mathrm{Ad}\, g \circ \rho$ has the block diagonal form an upper 1 by 1 block with an $n+1$ by $n+1$ block below. But then $(\mathrm{Ad}\, g \circ \rho)^{*-1}$ has the same block diagonal form and hence $\mathrm{Ad}\, g^{*-1} \circ (\rho^*)^{-1}$ has the same block diagonal form and fixes $\mathbf{C}e_1$. Consequently $(\rho^*)^{-1}$ fixes $\mathbf{C}g^* e_1$. But then $\sigma(\rho) = \mathrm{Ad}\, r \circ (\rho^*)^{-1}$ fixes $\mathbf{C}rg^* e_1$. With this the lemma is proved.

Notation. If Y is a subset of $\mathrm{Hom}(\Gamma, G)$ we let $S(Y)$ denote its saturation by G:

$$S(Y) = \{\mathrm{Ad}\, g \circ y: y \in Y, g \in G\}.$$

The following lemma is obvious since G acts transitively on the positive lines.

Lemma 1.9.

$$X \cap U' = S(\mathrm{Hom}(\Gamma, U(n, 1))).$$

We now prove the final lemma of this section.

Lemma 1.10. *The closure of* $S\left(j_*\left(\bigcup_{i=1}^{l} \mathscr{R}_i\right)\right)$ *in* $\mathrm{Hom}(\Gamma, G)$ *does not intersect* $R(\rho_0)$.

Proof. Suppose for the purpose of contradiction that the closure of $S\left(j_*\left(\bigcup_{i=1}^{l} \mathscr{R}_i\right)\right)$ in $\mathrm{Hom}(\Gamma, G)$ intersects $R(\rho_0)$. Then there exist sequences $\{g_n\}$ in G, $\{\rho_n\}$ in $\bigcup_{i=1}^{l} \mathscr{R}_i$ and $\rho \in S(\mathscr{R}_0)$ such that $\lim_{n\to\infty} \mathrm{Ad}\, g_n \circ \rho_n = \rho$. By conjugating by a fixed element of G we may assume ρ has $\mathbf{C}e_1$ as its fixed line. The representation $\mathrm{Ad}\, g_n \circ \rho$ fixes the line $L_n = \mathbf{C}g_n e_1$. By compactness the sequence of fixed lines $\{L_n\}$ in W has a line L of accumulation. By passing to a subsequence we may assume $\lim_{n\to\infty} L_n = L$. But we have $\mathrm{Ad}\, g_n \circ \rho_n L_n = L_n$. Passing to the limit we have $\rho L = L$. But ρ has a unique fixed line. Hence $L = \mathbf{C}e_1$. Now choose a unit vector $v_n \in L_n$ for each n. The sequence $\{L_n\}$ is contained in a closed ball B inside $\mathbf{P}(W)_+$. Hence the sequence $\{v_n\}$ is contained in a compact set - the tautological circle bundle over B. By passing to a subsequence we may assume $\lim_{n\to\infty} v_n = ce_1$ with $c \in S^1$. We redefine v_n to be $c^{-1} v_n$ and we find a sequence of unit vectors $\{v_n\}$ such that v_n is fixed by $\mathrm{Ad}\, g_n \circ \rho_n$ and $\lim_{n\to\infty} v_n = e_1$. We now construct a sequence of elements $\{h_n\}$ in G such that $h_n e_1 = v_n$ and $\lim_{n\to\infty} h_n = \mathrm{id}$. If infinitely many of the v_n's are multiples of e_1 then we are done because $\bigcup_{i=1}^{l} j_*(\mathscr{R}_i)$ is closed by Lemma 1.6. Hence we may assume that none of the v_n's are multiples of e_1. Let

$$a_n = (v_n, e_1), \qquad b_n = \| v_n - \langle v_n, e_1 \rangle e_1 \|$$

and

$$f_n = \frac{1}{b_n} (v_n - \langle v_n, e_1 \rangle e_1).$$

Then $v_n = a_n e_1 + b_n f_n$. Let h_n be the element of G defined by:

$$\begin{aligned} h_n(e_1) &= a_n e_1 + b_n f_n, \\ h_n(f_n) &= \bar{b}_n e_1 + \bar{a}_n f_n, \\ h_n(u) &= u \quad \text{if } \langle u, e_1 \rangle = 0 \quad \text{and} \quad \langle u, f_n \rangle = 0. \end{aligned}$$

Then clearly $\{h_n\}$ satisfies the required properties. We now consider the sequence $\{h_n^{-1} g_n \rho_n g_n^{-1} h_n\}$. The elements of this sequence all fix the line $\mathbf{C}e_1$, hence are contained in $j_*\left(\bigcup_{i=1}^{l} \mathscr{R}_i\right)$. But this sequence converges to $\rho \in \mathscr{R}_0$. This is a contradiction since $j_*\left(\bigcup_{i=1}^{l} \mathscr{R}_i\right)$ is closed by Lemma 1.6.

We now define U'' to be the complement of the closure of $S\left(j_*\left(\bigcup_{i=1}^{l} \mathscr{R}_i\right)\right)$ in $\mathrm{Hom}(\Gamma, G)$. We define $U = U' \cap U''$. Then $X \cap U = R(\rho_0)$ and we have proved the following theorem.

Theorem 1.3. *There exists a saturated strongly open subset U of $\mathrm{Hom}(\Gamma, G)$ and a saturated real algebraic subset X of $\mathrm{Hom}(\Gamma, G)$ such that $X \cap U = R(\rho_0)$.*

The idea of studying the limiting line of the fixed lines of a sequence of Fuchsian representations will allow us to prove the following theorem:

Theorem 1.4. *The set of Fuchsian representations $R(\rho_0)$ is closed in $\mathrm{Hom}(\Gamma, G)$.*

Proof. Suppose $\{\rho_n\}$ is a sequence of representations in $R(\rho_0)$ converging to a representation ρ in $\mathrm{Hom}(\Gamma, G)$. Then ρ is discrete and faithful by Theorem 1.1. Let $\{L_n\}$ to be a sequence of fixed lines as in the previous theorem and L a limit line. If L is a positive line then we are done since ρ fixes L and ρ is discrete and faithful. We may then assume that L is a null-line, for the set of representations fixing a non-negative line is closed by Corollary 1.7.2. Hence ρ is a discrete faithful embedding of Γ into a parabolic subgroup P of G. But P is amenable so Γ is amenable. But this implies Γ is virtually solvable by Tits [14]. With this the lemma is proved.

Thus to prove our main theorem, Theorem 1.2 (Theorem 2 of the introduction), it remains to prove that $R(\rho_0)$ is open in U. We assume for the purpose of contradiction that this is not the case. Hence there exists a sequence of points $\{y_n\}$ in U converging to some $\rho \in R(\rho_0)$. By conjugating by some $g \in G$ we may assume $\rho \in j_* \mathrm{Hom}(\Gamma, G_0)$. We may now apply the Curve Selection Lemma to find a curve ρ_t such that:

(i) $\rho_t \in Y - X$ for $t \neq 0$.
(ii) $\rho_t = \rho$ for $t = 0$.

In the next section we study the Taylor series of ρ_t.

We conclude this section with a theorem concerning $R(\rho_0)$.

Theorem 1.5. *$R(\rho_0)$ is diffeomorphic to $(SU(n+1, 1)/S^1) \times \mathrm{Hom}(\Gamma, S^1)$ where S^1 is embedded in $SU(n+1, 1)$ as the center of $j(U(n, 1))$.*

Proof. If $\rho \in R(\rho_0)$ we let $Z(\rho)$ denote the isotropy subgroup of ρ in G. If $g \in Z(\rho)$ then g stabilizes the invariant line of ρ so $Z(\rho)$ is contained in a conjugate of $j(U(n, 1))$. Applying Lemma 1.4 we obtain the theorem.

2. Construction of a normal cocycle with bounding square

In this section we use the curve ρ_t constructed in the previous section to construct a cocycle $c \in Z^1(\Gamma, \mathfrak{g})$ which is not tangent to the space of Fuchsian deformations, that is $c \notin B^1(\Gamma, \mathfrak{g}) + Z^1(\Gamma, \mathfrak{g}_0)$, and with cup square $[c, c] \in B^2(\Gamma, \mathfrak{g})$. We recall $[c, c]$ is the 2-cocycle defined by:

$$[c, c](\gamma, \eta) = [c(\gamma), \mathrm{Ad}\, \rho(\gamma) \circ c(\eta)].$$

We then use c to construct a zero for a certain real quadratic form on $H^1(\Gamma, V)$. Here $\mathfrak{g}$ denotes the Lie algebra of G and $\mathfrak{g}_1$, is the orthogonal complement relative to the Killing form of $\mathfrak{g}_0$, the Lie algebra of G_0, in $\mathfrak{g}$. If $c \in H^1(\Gamma, \mathfrak{g})$, the class of $[c, c]$ is the first obstruction to finding a curve ρ_t tangent to c.

We will need to generalize the material in Johnson-Millson [6], Sect. 2 to the case in which the tangent vector of a deformation vanishes at $t=0$. We will also need to break up the deformation into a block diagonal and off diagonal part and generalize the two main lemmas, Lemma 2.1 and Lemma 2.3 of [6] to the leading coefficient of the off diagonal part.

We begin with an easy lemma as a bridge to the more detailed generalizations we will require.

Definition. Let X be an affine variety in $\mathbf{R}^m$ and $\alpha: (-\varepsilon, \varepsilon) \to X$ be a real analytic curve such that $\alpha(0) = x$. Let $\alpha(t) = \sum_{k=0}^{\infty} \alpha_k(0)\, t^k$ be the Taylor series for α about $t=0$. We define the leading coefficient of α to be α_n if $n>0$, $\alpha_n \neq 0$ and $\alpha_m = 0$ for $0 < m < n$.

In other words α_n is the first non-zero term after the constant term. We observe that $\alpha_n \in \mathbf{R}^m$.

Lemma 2.1. *The leading coefficient α_n of α is tangent to X at x.*

Proof. Let f be a polynomial function vanishing on X. Then $f(\alpha(t)) = 0$ for all t. We may assume $x=0$. We expand $f(\alpha(t))$ around $t=0$ and find that $f(\alpha(t)) \equiv \mathrm{d}f_0(\alpha_n)\, t^n \bmod t^{n+1}$. Hence $\mathrm{d}f_0$ annihilates α_n and the lemma is proved.

Corollary. *Let $G \subset GL(k, \mathbf{C})$ be a closed subgroup. Suppose ρ_t is a curve in $\mathrm{Hom}(\Gamma, G)$ and ρ_n is the leading term of ρ_t. Then the function $c: \Gamma \to \mathrm{gl}(k, \mathbf{C})$ defined by:*

$$c(\gamma) = \rho_n(\gamma)\, \rho_0(\gamma)^{-1}$$

is a cocycle on Γ with values in $\mathfrak{g}$. Furthermore if ρ_t is a trivial deformation then c is a coboundary.

Let U be a neighborhood of O in $\mathfrak{g}$ such that the restriction of $\exp: \mathfrak{g} \to G$ to U is injective. Let $F \subset \Gamma$ be an arbitrary finite set. Then for sufficiently small $\varepsilon > 0$ we have $\rho_t(\gamma)\, \rho_0(\gamma)^{-1} \in \exp(U)$ for all $\gamma \in F$ and $|t| < \varepsilon$. Thus we may write for $\gamma \in F$:

$$\rho_t(\gamma)\, \rho_0(\gamma)^{-1} = \exp u_t(\gamma)$$

where $u_t(\gamma)$ is curve in $\mathfrak{g}$ with $u_0(\gamma) = 0$.

We write out the Taylor series for $u_t(\gamma)$ with $\gamma \in \Gamma$ as:

$$u_t(\gamma) = \sum_{k=0}^{\infty} u_k(\gamma)\, t^k.$$

Substituting into the power series for exp we find that $u_k(\gamma) = 0$ for $k < n$ and $u_n(\gamma) = c(\gamma)$ (where $\rho_n(\gamma)$ is the leading coefficient of $\rho_t(\gamma)$ and $c(\gamma) = \rho_n(\gamma)\, \rho_0(\gamma)^{-1}$) as above).

We next observe that if F generates Γ then there exists k so that $u_k(\gamma) \notin \mathfrak{g}_0$. Otherwise $u_t(\gamma) \in \mathfrak{g}_0$ for all t and $\gamma \in F$ and hence $\rho_t(\gamma) = \exp u_t(\gamma)\, \rho_0(\gamma)$ is an element of G_0 for all t and $\gamma \in F$. Since F generates Γ this is a contradiction.

We observe that as a consequence of the identity $\rho_t(\gamma\eta) = \rho_t(\gamma)\, \rho_t(\eta)$ we obtain:

$$\exp u_t(\gamma\eta) = \exp u_t(\gamma) \exp \operatorname{Ad} \rho_0(\gamma)\, u_t(\eta).$$

We now derive some consequences of the Campbell-Baker-Hausdorff formula. We have

$$\exp u_t(\gamma) \exp \operatorname{Ad} \rho_0(\gamma)\, u_t(\eta) = \exp \left(\sum_{n=0}^{\infty} C_n(u_t(\gamma), \operatorname{Ad} \rho_0(\gamma)\, u_t(\eta)) \right)$$

where C_n is the nth Campbell-Baker-Hausdorff polynomial, Bourbaki [2], p. 55. We obtain an equality of power series:

$$\begin{aligned} \sum_{n=1}^{\infty} u_n(\gamma\eta)\, t^n = & \sum_{n=1}^{\infty} u_n(\gamma)\, t^n + \sum_{n=1}^{\infty} \operatorname{Ad} \rho_0(\gamma)\, u_n(\eta)\, t^n \\ & + \frac{1}{2} \sum_{n=1}^{\infty} \sum_{i+j=n} [u_i(\gamma), \operatorname{Ad} \rho(\gamma)\, u_j(\eta)]\, t^n \\ & + \sum_{n=3}^{\infty} C_n \left(\sum_{i=1}^{\infty} u_i(\gamma)\, t^i, \sum_{i=1}^{\infty} \operatorname{Ad} \rho_0(\gamma)\, u_i(\eta)\, t^i \right). \end{aligned}$$

Bringing the first two terms to the left-hand side and equating coefficients of t^n we obtain a multilinear function N_n on the Lie algebra such that:

$$\delta u_n(\gamma, \eta) = N_n(u_1(\gamma), \ldots, u_{n-1}(\gamma); \operatorname{Ad} \rho_0(\gamma)\, u_1(\eta), \ldots, \operatorname{Ad} \rho_0(\gamma)\, u_{n-1}(\eta)).$$

We observe that N_n is a sum of monomials which are brackets of the $u_i(\gamma)$'s and the $\operatorname{Ad} \rho_0(\gamma)\, u_j(\eta)$'s. In particular if u_i takes values in an abelian Lie algebra for $i < n$ then u_n is a cocycle.

We now observe that $\mathfrak{g}$ is $\mathbf{Z}/2$ graded by the decomposition described earlier. That is, we have a decomposition $\mathfrak{g} = \mathfrak{g}_0 \oplus \mathfrak{g}_1$ with the bracket relations:

(i) $[\mathfrak{g}_0, \mathfrak{g}_0] \subset \mathfrak{g}_0$.
(ii) $[\mathfrak{g}_0, \mathfrak{g}_1] \subset \mathfrak{g}_1$.
(iii) $[\mathfrak{g}_1, \mathfrak{g}_1] \subset \mathfrak{g}_0$.

In particular any word $[[[x_1, x_2], x_3], \ldots, x_n]$ in $\mathfrak{g}$ with the x_j's in either $\mathfrak{g}_0$ or $\mathfrak{g}_1$ will be in $\mathfrak{g}_0$ if an even number of the x_j's are in $\mathfrak{g}_1$ and will be in $\mathfrak{g}_1$ if an odd number of the x_j's are in $\mathfrak{g}_1$. Corresponding to the decomposition of $\mathfrak{g}$ we have a decomposition of cochains $u: \Gamma \to \mathfrak{g}$ as $u = w + v$ with $w: \Gamma \to \mathfrak{g}_0$ and $v: \Gamma \to \mathfrak{g}_1$.

We also assume we have a decomposition $\mathfrak{g}_0 = \mathfrak{g}_0' \oplus \mathfrak{z}_0$ of $\mathfrak{g}_0$ as the direct sum of two ideals such that the map $H^1(\Gamma, \mathfrak{z}_0) \to H^1(\Gamma, \mathfrak{g}_0)$ is an isomorphism. If $x \in \mathfrak{g}_0$ we let x' denote the projection of x on $\mathfrak{g}_0'$. Clearly this projection is a Lie algebra homomorphism and commutes with the group coboundary δ.

We observe that this last assumption is satisfied when $\mathfrak{g}$ is the Lie algebra of $SU(n+1, 1)$, $\mathfrak{g}_0$ is the Lie algebra of $U(n, 1)$ and $\mathfrak{z}_0$ is the center of $\mathfrak{g}_0$. The preceding statement is a consequence of Weil's vanishing theorem, see [12], VII-5, recall that we are assuming $n \geqq 2$. However for later applications we continue in the more general situation of arbitrary Γ and arbitrary $\mathfrak{g}$ subject to the hypotheses of the previous two paragraphs.

We define the *leading coefficient* mod $\mathfrak{g}_0$ of a deformation ρ_t to be v_m where u_m is the first term such that $u_m(\gamma) \notin \mathfrak{g}_0$ for some $\gamma \in \Gamma$ and $u_m = w_m + v_m$ is the above decomposition. Here u_m is the m-th Taylor coefficient of the curve u_t associated to ρ_t as above.

Lemma 2.2. *The leading coefficient* mod $\mathfrak{g}_0$ *is a cocycle.*

Proof. To see that v_m is a cocycle compare $\mathfrak{g}_1$ parts of degree t^m on the two sides of the equation:

$$\exp u_t(\gamma\eta) = \exp u_t(\gamma) \exp(\rho_0(\gamma)\, u_t(\eta)\, \rho_0(\gamma)^{-1}).$$

We have seen that δu_m is a sum of iterated brackets of lower order terms. But these lower order terms are all in $\mathfrak{g}_0$. Hence $\delta(w_m + v_m)$ is in $\mathfrak{g}_0$ and consequently $\delta v_m = 0$. With this the lemma is proved.

We say a deformation ρ_t is *normalized* if all the coefficients u_k of $\log(\rho_t \rho_0^{-1})$ take values in $\mathfrak{z}_0$ for $k \leqq m$. Here m is such that v_m is the leading coefficient mod $\mathfrak{g}_0$.

Lemma 2.3. *If ρ_t is any deformation of ρ_0 then there exists an analytic curve h_t in G_0 with $h_0 = e$ such that* $\mathrm{Ad}\, h_t \circ \rho_t$ *is normalized.*

Proof. We observe that by assumption $H^1(\Gamma, \mathfrak{z}_0)$ maps onto $H^1(\Gamma, \mathfrak{g}_0)$. Hence if $u \in Z^1(\Gamma, \mathfrak{g}_0)$ then there exists $\eta \in Z^1(\Gamma, \mathfrak{z}_0)$ and $b \in B^1(\Gamma, \mathfrak{g}_0)$ such that $u = \eta + b$.

We will prove the lemma by induction on the degree. We first normalize the leading coefficient u_n. We choose η_n and b_n as above such that $u_n = \eta_n + b_n$. We choose a curve h_t in G_0 such that if $x = \dot{h}$ then $b_n(\gamma) = \mathrm{Ad}\, \rho_0(\gamma)\, x - x$. Then the leading coefficient of $\mathrm{Ad}\, h_{t^n} \circ \rho_t$ is the coefficient of t^n in the polynomial $(1 + xt^n)(\rho_0 + u_n t^n \rho_0)(1 - xt^n)$, that is $x\rho_0 - \rho_0 x + u_n \rho_0$. The corresponding cocycle is η_n and we have normalized the leading term.

Let us suppose that the first k terms are normalized. We wish to normalize u_{k+1}. By the remarks preceding this lemma we have $\delta u_{k+1} = N_{k+1}(u_1, \ldots, u_k)$. But $u_1, \ldots, u_k$ take values in the Lie algebra $\mathfrak{z}_0$ so $N_{k+1}(u_1, \ldots, u_k)' = 0$ and u'_{k+1} is a cocycle. Hence there exist b_{k+1} and η_{k+1} with $b_{k+1} \in B^1(\Gamma, \mathfrak{g}_0)$ and $\eta_{k+1} \in C^1(\Gamma, \mathfrak{z}_0)$ such that $u_{k+1} = \eta_{k+1} + b_{k+1}$. We now construct a curve h_t as before so that $h_t \equiv 1 + xt^{k+1} \bmod t^{k+2}$ and $b_{k+1}(\gamma) = \mathrm{Ad}\, \rho(\gamma)\, x - x$. We claim that the curve $\rho'_t = \mathrm{Ad}\, h_t \circ \rho_t$ has its first $k+1$ terms normalized. Indeed we have:

$$\begin{aligned} \rho'_t &\equiv (1 + xt^{k+1})(\exp u_t)\, \rho_0 (1 - xt^{k+1}) \bmod t^{k+2} \\ &\equiv (\exp u_t)\, \rho_0 + (x\rho_0 - \rho_0 x)\, t^{k+1} \bmod t^{k+2}. \end{aligned}$$

Hence:

$$\rho'_t \rho_0^{-1} \equiv \exp u_t - b_{k+1}\, t^{k+1} \bmod t^{k+2}.$$

We see that $\log(\rho_t' \rho_0^{-1})$ has the same coefficients as u_t up to and including the terms of order k. The coefficient of t^{k+1} is $u_{k+1} - b_{k+1} = \eta_{k+1}$ and the lemma is proved.

We are now ready to prove our main lemma. We now assume $\varkappa_0$ is abelian, hence central in $\mathfrak{g}_0$.

Lemma 2.4. *If ρ_t is a normalized deformation then we have:*

$$\delta w_{2m} = [v_m, v_m].$$

Proof. We examine the Campbell-Baker-Hausdorff formula more carefully to obtain:

$$\delta u_{2m}(\gamma, \eta) = \frac{1}{2} \sum_{k=1}^{2m-1} [u_k(\gamma), \operatorname{Ad} \rho_0(\gamma)\, u_{2m-k}(\eta)] + M_{2m}(u_1(\gamma), \ldots, u_{2m-1}(\gamma); \operatorname{Ad} \rho_0(\gamma)\, u_1(\eta), \ldots, \operatorname{Ad} \rho_0(\gamma)\, u_{2m-1}(\eta)).$$

Here M_{2m} is the sum of all monomials in N_{2m} involving at least three brackets.

We now decompose each u_k according to $u_k = w_k + v_k$. Hence

$$\operatorname{Ad} \rho_0(\gamma)\, u_k = \operatorname{Ad} \rho_0(\gamma)\, u_k + \operatorname{Ad} \rho_0(\gamma)\, v_k$$

for all $\gamma \in \Gamma$. We now expand the first term on the right-hand side of the above equation.

$$\begin{aligned} \frac{1}{2} \sum_{k=1}^{2m-1} [u_k(\gamma), \operatorname{Ad} \rho_0(\gamma)\, u_{2m-k}(\eta)] = {} & \frac{1}{2} \sum_{k=1}^{2m-1} [w_k(\gamma), \operatorname{Ad} \rho_0(\gamma)\, w_{2m-k}(\eta)] \\ & + \frac{1}{2} \sum_{k=1}^{m} [w_k(\eta), \operatorname{Ad} \rho_0(\gamma)\, v_{2m-k}(\eta)] \\ & + \frac{1}{2} \sum_{k=m}^{2m} [v_k(\gamma), \operatorname{Ad} \rho_0(\gamma)\, w_{2m-k}(\eta)] \\ & + \frac{1}{2} [v_m(\gamma), \operatorname{Ad} \rho_0(\gamma)\, v_m(\eta)]. \end{aligned}$$

We see that the two middle terms on the right are annihilated by the projection on $\mathfrak{g}_0$. Also the first term is zero for in each of the brackets one of the terms is in $\varkappa_0$ and the other in $\mathfrak{g}_0$. We now analyze M_{2m}.

We expand M_{2m} by the total number of brackets to obtain:

$$\begin{aligned} & M_{2m}(w_i(\gamma), v_i(\gamma); \operatorname{Ad} \rho_0(\gamma)\, w_i(\eta), \operatorname{Ad} \rho_0(\gamma)\, v_i(\eta)) \\ & \quad = \sum_{j=3}^{2m} A_j(w_i(\gamma), v_i(\gamma); \operatorname{Ad} \rho_0(\gamma)\, w_i(\eta), \operatorname{Ad} \rho_0(\gamma)\, v_i(\eta)). \end{aligned}$$

Here A_j is a sum of monomials each of which is a j-fold bracket. Since we have projected onto $\mathfrak{g}_0$ each monomial is an element of $\mathfrak{g}_0$. The total t-degree is $2m$. We claim that the A_j's involve only the w's. Indeed any monomial occurring in the A_j's is in $\mathfrak{g}_0$ and hence must involve an even number of v's by the discussion preceding Lemma 2.1. But if any v's occur then the monomial

will have t-degree larger than or equal to $n+2m$. Thus, the A_j's do not depend on the v's.

Since $j \geqq 3$ it is clear that at least one w_k or $\mathrm{Ad}\,\rho_0(\gamma)\,w_k$ occurring in a monomial must satisfy $k<m$. Say w_{k_0} is such a term. Then w_{k_0} takes values in $\mathfrak{z}_0$. But w_{k_0} occurs bracketed with a monomial in the w's which take values in $\mathfrak{g}_0$. The resulting bracket with w_{k_0} must be zero since w_{k_0} takes values in $\mathfrak{z}_0$. With this the lemma is proved.

We now modify the curve ρ_t so that the leading coefficient mod $\mathfrak{g}_0$ is not a coboundary. We have the following lemma.

Lemma 2.5. *If ρ_t is any deformation of ρ_0 then there exists an analytic curve g_t in G with $g_0=e$ such that either $\mathrm{Ad}\,g_t\circ\rho_t$ is normalized with leading coefficient* mod $\mathfrak{g}_0$ *not a coboundary or $\mathrm{Ad}\,g_t\circ\rho_t$ is contained in* $\mathrm{Hom}(\Gamma, G_0)$.

Proof. By Lemma 2.3 we may assume that ρ_t is normalized. Suppose v_{m_1} is the leading coefficient mod $\mathfrak{g}_0$ of ρ_t. Suppose v_{m_1} is a coboundary. Then there exists $x\in\mathfrak{g}$ such that $v_{m_1}(\gamma)=\mathrm{Ad}\,\rho_0(\gamma)\,x-x$. Let $g_{m_1,t}$ be the one parameter group in G with $\dot{g}_{m_1}=x$. We consider the curve ρ_t' given by $\rho_t'=\mathrm{Ad}\,g_{m_1,t^{m_1}}\circ\rho_t$. We define a curve u_t' with values in $\mathfrak{g}$ as before with $u_t'=\log(\rho_t'\circ\rho_0^{-1})$ and write $u_t'(\gamma)=\sum_{k=1}^{\infty}u_k'(\gamma)\,t^k$. Then a calculation similar to that of Lemma 2.3 shows that $u_k'=u_k$ for $k<m_1$ and $u_{m_1}'=u_{m_1}-v_{m_1}$. We see then that the curve ρ_t' has leading coefficient mod $\mathfrak{g}_0$ equal to v_{m_2} with $m_2>m_1$. We normalize ρ_t' and repeat the above process. Either we arrive at a deformation $\mathrm{Ad}\,g_t\circ\rho_t$ with leading coefficient mod $\mathfrak{g}_0$ which is not a coboundary or we obtain a *formal* curve $g_t'=\prod_{k=1}^{\infty}g_{m_k,t^{m_k}}$ which satisfies the analytic equations given by $\mathrm{Ad}\,g_t'\circ\rho_t\in\mathrm{Hom}(\Gamma,Z_0)$ where Z_0 is the analytic subgroup of G_0 corresponding to $\mathfrak{z}_0$. But then by Artin [18] there exists an *analytic* curve g_t such that $\mathrm{Ad}\,g_t\circ\rho_t\in\mathrm{Hom}(\Gamma,Z_0)$. With this the lemma is proved.

We now examine more closely the cup-product of the beginning of this section. The cup-product of cochains gives rise to a map:

$$H^1(\Gamma,\mathfrak{g})\times H^1(\Gamma,\mathfrak{g})\xrightarrow{\otimes}H^2(\Gamma,\mathfrak{g}\otimes\mathfrak{g})$$

given by

$$c\otimes c(\gamma,\eta)=c(\gamma)\otimes\mathrm{Ad}\,\rho(\gamma)\cdot c(\eta).$$

But the Lie bracket gives a Γ-module map $\mathfrak{g}\otimes\mathfrak{g}$ into $\mathfrak{g}$ inducing a map from $H^2(\Gamma,\mathfrak{g}\otimes\mathfrak{g})$ to $H^2(\Gamma,\mathfrak{g})$. The composition of this map with the cup-product gives rise to the map

$$H^1(\Gamma,\mathfrak{g})\times H^1(\Gamma,\mathfrak{g})\xrightarrow{[,]}H^2(\Gamma,\mathfrak{g})$$

described in the beginning of this section given by:

$$[\alpha,\alpha](\gamma,\eta)=[\alpha(\gamma)\,\mathrm{Ad}\,\rho(\gamma)\cdot\alpha(\eta)].$$

We observe that $[,]$ induces a map

$$H^1(\Gamma,\mathfrak{g}_1)\otimes H^1(\Gamma,\mathfrak{g}_1)\to H^2(\Gamma,\mathfrak{g}_0).$$

We may apply the considerations of this section to the curve ρ_t constructed in Sect. 1. We let $c = v_m$ where v_m is the leading coefficient of $\rho_t \bmod \mathfrak{g}_0$. We have observed that since ρ_t is not contained in $\mathrm{Hom}(\Gamma, G_0)$ there exists m so that $v_m \neq 0$. Moreover, since ρ_t is not contained in $S(\mathrm{Hom}(\Gamma, G_0))$ for $t > 0$, there is no curve g_t such that $\mathrm{Ad}\, g_t \circ \rho_t \in \mathrm{Hom}(\Gamma, G_0)$. By Lemma 2.5, we may assume that v_m is not a coboundary. But $[v_m, v_m]$ is a coboundary by Lemma 2.4. We let Γ act on $\mathfrak{g}$ by ρ_0. We observe that $\rho_0(\Gamma) \subset j(U(n, 1))$ and consequently we may identify Γ with $j^{-1} \circ \rho_0(\Gamma)$ a cocompact discrete subgroup of $U(n, 1)$. We obtain the following theorem.

Theorem 2.1. *Suppose $R(\rho_0)$ is not open in $\mathrm{Hom}(\Gamma, G)$. Then there exists a realization of Γ as a cocompact torsion-free discrete subgroup of $U(n, 1)$ and a non-zero class $c \in H^1(\Gamma, \mathfrak{g}_1)$ such that $[c, c]$ is zero in $H^2(\Gamma, \mathfrak{g}_0)$.*

We now observe that the mapping trace: $\mathfrak{g}_0 \to \mathbf{C}$ induces a map $H^2(\Gamma, \mathfrak{g}_0) \to H^2(\Gamma, \mathbf{C})$. Moreover, a calculation shows that trace takes pure imaginary values on $\mathfrak{g}_0$. Multiplying by $i/2$ we obtain a map from $H^2(\Gamma, \mathfrak{g}_0)$ to $H^2(\Gamma, \mathbf{R})$. By composition we obtain a map:

$$\Psi: H^1(\Gamma, \mathfrak{g}_1) \times H^1(\Gamma, \mathfrak{g}_1) \to H^2(\Gamma, \mathbf{R}).$$

We now use some considerations from differential geometry to get a real quadratic form on $H^1(\Gamma, \mathfrak{g}_1)$. We may identify the homology and cohomology groups of Γ with real coefficients with the corresponding homology and cohomology groups of the compact manifold $M = \Gamma \backslash D$. But M is a projective variety and consequently has a distinguished class Z in $H_2(M, \mathbf{Z})$. This class is obtained as follows. Let $H \subset M$ be the hyperplane section class of the projective embedding given by applying the Kodaira Embedding Theorem to the Kahler form ω normalized to have integral periods. ω is (a multiple of) the imaginary part of the Bergmann metric. We observe that $H \in H_{2n-2}(M, \mathbf{Z})$ and ω is the Poincare dual of H. Then Z, the $(n-1)$-fold self-intersection class of H, is a non-zero element of $H_2(M, \mathbf{Z})$ which is Poincare dual to ω^{n-1}.

We now define a quadratic form Q on $H^1(\Gamma, \mathfrak{g}_1)$ by:

$$Q(c) = \Psi(c, c) \cdot Z.$$

Here we use $\cdot$ to denote the Kronecker index.

We now observe that there is a vector space isomorphism from V to $\mathfrak{g}_1$ given by sending $(z_1, z_2, \ldots, z_{n+1})$ to the $n+2$ by $n+2$ matrix with first column $(0, z_1, z_2, \ldots, z_{n+1})$ and first row $(0, -\bar{z}_1, \ldots, -\bar{z}_n, \bar{z}_{n+1})$ and all other entries zero. The action of $\rho_0(\Gamma)$ on $\mathfrak{g}_1$ corresponds to the standard action of $\Gamma \subset U(n, 1)$ on $\mathbf{C}^{n+1}$ twisted by the restriction of the determinant. We then have the following consequence of Theorem 2.1.

Theorem 2.1 (bis). *Suppose $R(\rho_0)$ is not open in $\mathrm{Hom}(\Gamma, G)$. Then there exists a non-zero class c in $H^1(\Gamma, V)$ which is a zero of the real quadratic form Q defined by the composition:*

$$H^1(\Gamma, V) \times H^1(\Gamma, V) \to H^2(\Gamma, \mathbf{R}) \to \mathbf{R}.$$

Here the first arrow is the cup-product induced by the real bilinear map given by the imaginary part of $\langle , \rangle$. The second arrow is evaluation on the class Z.

3. Some hermitian linear algebra

The purpose of the last two sections of our paper is to prove that the quadratic form Q on $H^1(\Gamma, V)$ defined at the end of the last section is in fact positive definite. Thus the assumption that $R(\rho_0)$ is not open in $\mathrm{Hom}(\Gamma, G)$ leads to a contradiction and our main theorem is proved.

In this section we prove a definiteness theorem for a certain quadratic form. This theorem will imply that certain integrands considered in Sect. 4 are pointwise either positive or negative. This will in turn imply that Q will have a fixed sign on each of the four Hodge groups associated to $H^1(\Gamma, V)$. We will then invoke a theorem of Matsushima-Murakami which implies that the two Hodge groups on which Q is negative definite are in fact zero.

Let U be a real vector space with a complex structure J and an inner product $(\,,\,)$ such that for all u_1, u_2 in U we have:

$$(Ju_1, Ju_2) = (u_1, u_2).$$

The J acts on U^*, the dual space of U, by

$$J \cdot \alpha(u) = -\alpha(Ju).$$

We define $\omega \in \Lambda^2 U^*$, the Kahler form associated to $(\,,\,)$ and J by:

$$\omega(u_1, u_2) = (Ju_1, u_2).$$

We can choose an orthonormal basis $\{q_j, p_j : j = 1, 2, \ldots, n\}$ for U^* such that for all j:

$$Jq_j = p_j$$
$$Jp_j = -q_j.$$

We find that the Kahler form ω is given by:

$$\omega = \sum_{j=1}^{n} q_j \wedge p_j.$$

We now consider elements of $U^* \otimes_{\mathbf{R}} \mathbf{C}$. We define an element $\alpha \in U^* \otimes_{\mathbf{R}} \mathbf{C}$ to be of type $(1, 0)$ if α satisfies for all $u \in U$:

$$\alpha(Ju) = i\alpha(u) \quad \text{or equivalently} \quad J\alpha = -i\alpha.$$

We denote the subspace of $U^* \otimes \mathbf{C}$ (we drop the subscript $\mathbf{R}$ on the tensor products henceforth) of type $(1, 0)$ linear functionals by $(U^* \otimes \mathbf{C})^{(1,0)}$.

We define an element β of $U^* \otimes \mathbf{C}$ to be of type $(0, 1)$ if β satisfies for all $u \in U$:

$$\beta(Ju) = -i\beta(u) \quad \text{or equivalently} \quad J\beta = i\beta.$$

Now if $\alpha \in U^* \otimes \mathbf{C}$ then we define its complex conjugate $\bar{\alpha} \in U^* \otimes \mathbf{C}$ by:

$$\bar{\alpha}(v) = \overline{\alpha(v)}.$$

Clearly the operation of complex conjugation interchanges $(U^* \otimes \mathbf{C})^{(1,0)}$ and $(U^* \otimes \mathbf{C})^{(0,1)}$. Also we note that $\{p_j + iq_j : j = 1, 2, \ldots, n\}$ is a basis for $(U^* \otimes \mathbf{C})^{(1,0)}$ considered as a complex vector space and $\{p_j - iq_j : j = 1, 2, \ldots, n\}$ is a basis for $(U^* \otimes \mathbf{C})^{(0,1)}$ considered as a complex vector space.

We define $\tau \in \Lambda^2 U^*$ to be of type (1, 1) if $\tau(Ju_1, Ju_2) = \tau(u_1, u_2)$. We observe that ω is of type (1, 1).

We will also need to consider $U^* \otimes_{\mathbf{R}} V$ with V a complex vector space equipped with a non-singular sesquilinear form $\langle , \rangle$. We let τ denote the imaginary part of $\langle , \rangle$. Since V is a complex vector space we may define type (1, 0) forms, type (0, 1) forms and complex conjugation using the same formulas as above, noting the difference that $\alpha(u)$ is now an element of V. Observe that if $\alpha \in U^*$ and $v \in V$ we have the dyad $\alpha \otimes v \in U^* \otimes V$ defined by:

$$\alpha \otimes v(u) = \alpha(u)\, v.$$

We now define a real bilinear map B on $U^* \otimes V$ with values in the type (1, 1) elements of $\Lambda^2 U^*$ by the formula:

$$B(\alpha, \beta)(u_1, u_2) = \tau(\alpha(u_1), \beta(u_2)) - \tau(\alpha(u_2), \beta(u_1)).$$

We note that this formula is the same as:

$$B(\alpha, \beta)(u_1, u_2) = \operatorname{Im}(\langle \alpha(u_1), \beta(u_2) \rangle - \langle \alpha(u_2), \beta(u_1) \rangle).$$

It is important to observe that B *is symmetric*, that is, $B(\alpha, \beta) = B(\beta, \alpha)$.

Lemma 3.1. *Suppose α, β are in $U^* \otimes \mathbf{C}$ and v, v' are in V. Then:*

(i) *If v and v' are orthogonal we have:*

$$B(\alpha \otimes v, \beta \otimes v') = 0.$$

(ii) $B(\alpha \otimes v, \alpha \otimes v) = \langle v, v \rangle \operatorname{Im} \alpha \wedge \bar{\alpha}$.

Here if α, β are in $U^* \otimes \mathbf{C}$ then $\operatorname{Im}(\alpha \wedge \beta)$ is the element of $\Lambda^2 U^*$ whose value at (u_1, u_2) is the imaginary part of the complex number $\alpha(u_1)\beta(u_2) - \alpha(u_2)\beta(u_1)$.

Proof. From the defining formula for B we have for $u_1, u_2 \in U$:

$$\begin{aligned} B(\alpha \otimes v, \beta \otimes v')(u_1, u_2) &= \operatorname{Im}(\langle \alpha(u_1)\, v, \beta(u_2)\, v' \rangle - \langle \alpha(u_2)\, v, \beta(u_1)\, v' \rangle) \\ &= \operatorname{Im}(\alpha(u_1)\, \bar{\beta}(u_2) \langle v, v' \rangle - \alpha(u_2)\, \bar{\beta}(u_1) \langle v, v' \rangle). \end{aligned}$$

The lemma is now obvious.

We now define a real-valued symmetric form $((,))$ on $U^* \otimes V$ by:

$$((\alpha, \beta)) = (B(\alpha, \beta), \omega).$$

Here the inner product on the right is the inner product on $\Lambda^2 U^*$ induced by $(,)$.

Lemma 3.2. (i) *Suppose* $\alpha \in (U^* \otimes \mathbf{C})^{(1,0)}$, *then:*

$$(\mathrm{Im}\,\alpha \wedge \bar{\alpha}, \omega) = -\|\alpha\|^2.$$

(ii) *Suppose* $\alpha \in (U^* \otimes \mathbf{C})^{(0,1)}$, *then:*

$$(\mathrm{Im}\,\alpha \wedge \bar{\alpha}, \omega) = \|\alpha\|^2.$$

Proof. The second statement follows from the first since $\alpha \to \bar{\alpha}$ interchanges $(U^* \otimes \mathbf{C})^{(1,0)}$ and $(U^* \otimes \mathbf{C})^{(0,1)}$. To prove the first statement note that both sides are bilinear functions on the *real* vector space $(U^* \otimes \mathbf{C})^{(1,0)}$. Hence it is sufficient to check the identity on the real basis

$$\{q_j + ip_j, i(q_j + ip_j) : 1 \leqq j \leqq n\}.$$

But observing that both sides are unchanged upon replacing α by $i\alpha$ we see that is sufficient to prove for $1 \leqq j \leqq n$:

$$\mathrm{Im}\,((q_j + ip_j) \wedge (q_j - ip_j), \omega) = -2.$$

But this is clear for

$$(q_j + ip_j) \wedge (q_j - ip_j) = -2iq_j \wedge p_j.$$

With this the lemma is proved.

Let $V = V^+ \otimes V^-$ be a splitting of V into the orthogonal sum of a positive definite subspace and a negative definite subspace. We obtain a corresponding decomposition:

$$U^* \otimes V = (U^* \otimes V^+)^{(1,0)} \oplus (U^* \otimes V^-)^{(1,0)} \oplus (U^* \otimes V^+)^{(0,1)} \oplus (U^* \otimes V^-)^{(0,1)}.$$

Theorem 3.1. (i) *The summands in the above decomposition are orthogonal for* $((\,,\,))$.

(ii) $((\,,\,))$ *is negative definite on the first and last summands and positive definite on the other two summands.*

Proof. The first statement follows from Lemma 3.1. It is sufficient to prove all dyads from one space are orthogonal to all dyads from another. Let us consider $U^* \otimes V^+)^{(1,0)}$. Clearly it is orthogonal to the second and fourth summands by Lemma 3.1. But it is orthogonal to the third summand since if α has type $(1,0)$ and β has type $(0,1)$ then $\mathrm{Im}\,\alpha \wedge \bar{\beta}$ will have no type $(1,1)$ component and will consequently be orthogonal to ω. The orthogonality of the other summands may be verified in a similar fashion. We now prove (ii).

Let $\{v_j' : 1 \leqq j \leqq r\}$ be an orthonormal basis for V^+ and $\{v_j'' : 1 \leqq j \leqq s\}$ be an orthogonal basis for V^- satisfying $\langle v_j'', v_j'' \rangle = -1$ for $j = 1, 2, \ldots, s$. Let $\alpha \in (U^* \otimes V^-)^{(1,0)}$. Then α may be written as $\alpha = \sum_{j=1}^{s} \alpha_j \otimes v_j''$ where $\alpha_j \in (U^* \otimes \mathbf{C})^{(1,0)}$. Then we have (by Lemma 3.1):

$$B(\alpha \wedge \alpha) = \sum_j \langle v_j'', v_j'' \rangle\, \mathrm{Im}\,(\alpha_j \wedge \bar{\alpha}_j).$$

Hence we obtain:

$$((\alpha, \alpha)) = (B(\alpha \wedge \alpha), \omega) = \sum_j \langle v_j'', v_j'' \rangle (\mathrm{Im}(\alpha_j \wedge \bar{\alpha}_j), \omega) = -\sum_j \langle v_j'', v_j'' \rangle \|\alpha_j\|^2.$$

The other statements can be proved in a similar fashion, all dyads from one space are orthogonal to all dyads from another. Let us consider $(U^* \otimes V^+)^{(1,0)}$. Clearly it is orthogonal to the second and fourth summands by Lemma 3.1. But it is orthogonal to the third summand since if α has type $(1,0)$ and β has type $(0,1)$ then $\mathrm{Im}\, \alpha \wedge \bar{\beta}$ will have no type $(1,1)$ component and will consequently be orthogonal to ω. The orthogonality of the other summands may be verified in a similar fashion. We now prove (ii).

Let $\{v_j' : 1 \leqq j \leqq r\}$ be an orthonormal basis for V^+ and $\{v_j'' : 1 \leqq j \leqq s\}$ be an orthogonal basis for V^- satisfying $\langle v_j'', v_j'' \rangle = -1$ for $j = 1, 2, \ldots, s$. Let $\alpha \in (U^* \otimes V^-)^{(1,0)}$. Then α may be written as $\alpha = \sum_{j=1}^{s} \alpha_j \otimes v_j''$ where $\alpha_j \in (U^* \otimes \mathbf{C})^{(1,0)}$. Then we have (by Lemma 3.1):

$$B(\alpha \wedge \alpha) = \sum_j \langle v_j'', v_j'' \rangle \, \mathrm{Im}(\alpha_j \wedge \bar{\alpha}_j).$$

Hence we obtain:

$$((\alpha, \alpha)) = (B(\alpha \wedge \alpha), \omega) = \sum_j \langle v_j'', v_j'' \rangle (\mathrm{Im}(\alpha_j \wedge \bar{\alpha}_j), \omega) = -\sum_j \langle v_j'', v_j'' \rangle \|\alpha_j\|^2.$$

The other statements can be proved in a similar fashion.

4. Group cohomology, harmonic forms and the Matsushima-Murakami theorem

In the section we show that the results of the last section imply that the restrictions of Q to the non-zero Hodge pieces of $H^1(\Gamma, V)$ are given by the integral of an everywhere positive function.

We first describe the decomposition of $H^1(\Gamma, V)$ using Hodge theory. If E is a vector bundle over M we let $\mathscr{A}^r(M, E)$ denote the E-valued r-forms on M. An element of $\mathscr{A}^r(M, E)$ assigns an element of the fiber of E over x to a p-tuple of tangent vectors at $x \in M$. If E has a connection we may construct an exterior differential d on $\mathscr{A}^*(M, E)$. However $d^2 = 0$ if and only if the connection is flat.

Now let $\tilde{V}$ be the flat bundle over M associated to V. Since $\tilde{V}$ is flat we obtain a complex $\{\mathscr{A}^*(M, \tilde{V}), d\}$ with cohomology groups canonically isomorphic to the cohomology groups $H^*(M, \tilde{V})$ of M with local coefficients in V - see Raghunathan [12], Chap. VIII. We observe that the preceding remarks apply when V is replaced by any Γ-module W and $\tilde{V}$ is replaced by the associated flat bundle $\tilde{W}$.

We will be especially concerned with the case in which $W = \mathfrak{g}$. In this case we have the cup-product $H^1(\Gamma, \mathfrak{g}) \times H^1(\Gamma, \mathfrak{g}) \to H^2(\Gamma, \mathfrak{g})$ studied in Sect. 2. But we also have a multiplicative structure on $\mathscr{A}^r(M, \tilde{\mathfrak{g}})$. In the case $r = 1$ we define a bilinear function $[\,,\,]$ on $\mathscr{A}^1(M, \tilde{\mathfrak{g}})$ with values in $\mathscr{A}^2(\mathrm{M}, \tilde{\mathfrak{g}})$ by the formula

for X, Y tangent vectors to M:

$$[\omega, \tau](X, Y) = [\omega(X), \tau(Y)] - [\omega(Y), \tau(X)].$$

The reader will verify that the above bilinear function is symmetric. In the next lemma π will denote the universal covering $\pi: \tilde{M} \to M$.

Lemma 4.1. *The above isomorphism carries the bilinear form* $[\ ,\ ]$ *on* $H^1(M, \tilde{g})$ *to the cup-product on* $H^1(\Gamma, g)$.

Proof. This is standard material but since it is essential in our proof we give the details. We define $\mathscr{A}^r(\tilde{M}, g)$ to be the set of smooth g-valued r-forms ω on $\tilde{M}$ with the action:

$$\gamma\omega = \mathrm{Ad}\,\rho(\gamma)(\gamma^{-1})^*\,\omega.$$

Clearly one obtains an isomorphism from $\mathscr{A}^r(M, \tilde{g})$ to $\mathscr{A}^r(\tilde{M}, g)^\Gamma$ by pulling back $\tilde{g}$ to $\tilde{M}$ and composing with parallel translation to the standard fiber.

We consider the double complex:

$$\{C^p(\Gamma, \mathscr{A}^q(\tilde{M}, g)): p, q \text{ non-negative integers}\}$$

of Eilenberg-MacLane p-cochains with values in $\mathscr{A}^q(\tilde{M}, g)$ equipped with the obvious differentials. Then the above isomorphism may be realized be studying the two spectral sequences attached to the double complex.

The map from $H^1(M, \tilde{g})$ to $H^1(\Gamma, g)$ is constructed as follows. Let ω be a closed 1-form on M with values in $\tilde{g}$. Choose f a smooth section of $\pi^*\tilde{g}$ such that $\mathrm{d}f = \pi^*\omega$. Define c: $\Gamma \to \mathscr{A}^0(\tilde{M}, g)$ by $c(\gamma) = \mathrm{Ad}\,\rho(\gamma)^{-1} f \circ \gamma - f$. Then $c(\gamma)$ is constant, $\delta c = 0$ and so c gives rise to an element of $H^1(\Gamma, g)$.

We now compute the image of $[\omega, \omega]$ in $H^2(\Gamma, g)$. We note that for f as above $[f, \omega]$ is an element of $C^0(\Gamma, \mathscr{A}^1(\tilde{M}, g))$ satisfying $d([f, \omega]) = [\omega, \omega]$. But then $a = \delta[f, \omega]$ is an element of $C^1(\Gamma, \mathscr{A}^1(\tilde{M}, g))$ cohomologous to $[\omega, \omega]$ in the double complex. We find:

$$a(\gamma) = [c(\gamma), \omega].$$

But then $-[c, f]$ satisfies $-d(-[c, f]) = a$ and hence $-\delta[c, f]$ is an element of $C^2(\Gamma, g)$ cohomologous to $[\omega, \omega]$ in the double complex. But a direct computation shows $\delta[c, f] = -[c, c]$. With this the lemma is proved.

We use the bilinear form B of the previous section to define a *symmetric* bilinear mapping from $\mathscr{A}^1(M, \tilde{V}) \times \mathscr{A}^1(M, \tilde{V})$ to $\mathscr{A}^2(M)$ also denoted B by:

$$B(\alpha, \beta)|_x = B(\alpha|_x, \beta|_x).$$

Here $x \in M$ and we take $U = T_x(M)$ in the discussion of the previous section.

We use the real valued bilinear form $((\ ,\ ))$ to define a symmetric bilinear form with values in $C^\infty(M)$ also denoted $((\ ,\ ))$ by:

$$((\alpha, \beta)) = (B(\alpha, \beta), \omega) \qquad \text{here } (\ ,\ ) \text{ is the Riemannian metric.}$$

Both B and $((\ ,\ ))$ factor through cohomology.

We compose $B: H^1(M, \tilde{V}) \times H^1(M, \tilde{V}) \to H^2(M, \mathbf{R})$ with integration over the cycle Z to get a quadratic form Q' on $H^1(M, \tilde{V})$.

Lemma 4.2. *Under the isomorphism between $H^1(\Gamma, V)$ and $H^1(M, \tilde{V})$ the quadratic forms Q and Q' coincide.*

Proof. The lemma follows immediately from Lemma 4.1.

We now redefine Q to be the quadratic form on $H^1(M, \tilde{V})$ constructed above. We see that our theorem will follow if we can prove that Q is positive definite.

We can give another formula for Q. Since ω^{n-1} is the Poincare dual of Z we have:

$$Q(\alpha) = \int_Z B(\alpha, \alpha) = \int_M B(\alpha, \alpha) \wedge \omega^{n-1}.$$

But $\omega^{n-1} = {}_*\omega$ hence we obtain

$$Q(\alpha) = \int_M B(\alpha, \alpha) \wedge {}_*\omega = \int_M (B(\alpha, \alpha), \omega) \text{ vol} = \int_M ((\alpha, \alpha)) \text{ vol}.$$

Here vol denotes the Riemannian volume element. Our main theorem is reduced to proving that we may choose a representative for a class α such that $((\alpha, \alpha))$ is a positive function. We will now see that the harmonic representative will work.

We now recall the Hodge theory of $H^1(M, \tilde{V})$. A good reference for the real Hodge theory is [12], Chap. VIII. For the Hodge decomposition see Murakami [10]. We have a Riemannian metric on M, we need a metric along the fibers for $\tilde{V}$. The standard positive definite Hermitian form on V is admissible, Borel-Wallach [1] [1], p. 47, and consequently we get an induced metric along the fibers and we may form a Laplacian Δ. We apply [12], Chap. VIII to conclude that each class in $H^1(M, \tilde{V})$ has a unique harmonic representative. We observe that the Hodge Theorem in [12] is stated for G semi-simple but the same proof works for G reductive with compact center.

Let us now consider the projective model for D, that is, D is the set of negative lines in $\mathbf{P}(V)$. Then we have the tautological line bundle E_- over D and the perpendicular bundle E_+ which assigns to a negative line L the vectors in $L^\perp$, the orthogonal complement of L in V for $\langle\ ,\ \rangle$. The bundles E_+ and E_- descend to M. We denote the corresponding bundles again by E_+ and E_-. We have then a decomposition as bundles over M:

$$\tilde{V} = E_+ \oplus E_-.$$

We let p_+ and p_- denote the corresponding bundle projections. We have an associated decomposition:

$$\mathscr{A}^*(M, \tilde{V}) = \mathscr{A}^*(M, E_+) \oplus \mathscr{A}^*(M, E_-).$$

Since E_+ and E_- are not flat bundles it would seem impossible to define cohomology groups with values in E_+ and E_-. At this point we review the general theory of Matsushima-Murakami [9] which will in fact allow us to

make such a definition and get a decomposition of cohomology groups. The point will be that although the above decomposition is not compatible with d, it is compatible with Δ. To see this we will need more notation and some lemmas. Henceforth we let G denote $U(n, 1)$. The symbol K will denote $U(n) \times U(1)$, a maximal compact subgroup of G.

Let $\mathscr{A}^r(\Gamma\backslash G, K, \rho)$ denote the subspace of all V-valued smooth r-forms η on $\Gamma\backslash G$ satisfying:

(i) $\eta \circ R_k = \rho(k)^{-1}\eta$ for all $k \in K$.
(ii) $\iota(X)\eta = 0$ for all $X \in \mathfrak{k}$, the Lie algebra of K.

Here R_k denote the operation of right translation by $k \in K$ and ι denotes interior multiplication. The following fact is quite standard, Raghunathan [12], Matsushima-Murakami [9], but we include the proof for the sake of clarity.

Lemma 4.3. $\mathscr{A}^r(M, \tilde{V})$ is canonically isomorphic to $\mathscr{A}^r(\Gamma\backslash G, K, \rho)$.

Proof. We construct the map from $\mathscr{A}^r(M, \tilde{V})$ to $\mathscr{A}^r(\Gamma\backslash G, K, \rho)$.

Let $\omega \in \mathscr{A}^r(M, \tilde{V})$ and $\pi: \tilde{M} \to M$ be the universal cover. Then since $\tilde{V}$ pulled back to $\tilde{M}$ has a global parallelization we obtain a V-valued r-form ω' on $\tilde{M}$ satisfying:

$$\gamma^* \omega' = \rho(\gamma)\omega'.$$

We pull ω' back to G via the fibering $K \to G \to \tilde{M}$ and we obtain a V-valued form ω'' on G satisfying:

(i) $L_\gamma^* \omega'' = \rho(\gamma)\omega''$.
(ii) $\iota(X)\omega'' = 0$ for $X \in \mathfrak{k}$.

We now define ω''' a V-valued r-form on G by the formula:

$$\omega'''|_g = \rho(g)^{-1}\omega''|_g.$$

We leave the reader to verify that the map sending ω to ω'' is the required isomorphism.

Remark. In the last formula above it is essential that the representation ρ of Γ extend to a representation of G.

We may now write out ω''' in terms of the invariant parallelization of G to obtain an element η of $\Lambda^r \mathfrak{p}^* \otimes C^\infty(G) \otimes V$ satisfying:

(i) $[1 \otimes L_\gamma^* \otimes \rho(\gamma)] \cdot \eta = \eta$.
(ii) $[(\mathrm{Ad}\, k)^* \otimes R_k^* \otimes \rho(k)] \cdot \eta = \eta$.

Here $\mathfrak{p}$ is the orthogonal complement of $\mathfrak{k}$ in $\mathfrak{g}$ relative to the Killing form.

Let us denote the isomorphism $\mathscr{A}^r(M, \tilde{V}) \to \Lambda^r \mathfrak{p}^* \otimes C^\infty(G) \otimes V$ sending ω to η by Φ. We let q_+ and q_- denote the projections of V onto V_+ and V_-. It is apparent that:

(i) $\Phi \circ p_+ \circ \Phi^{-1} = 1 \otimes 1 \otimes q_+$.
(ii) $\Phi \circ p_- \circ \Phi^{-1} = 1 \otimes 1 \otimes q_-$.

The main ingredient in the work of Matsushima-Murakami is *Kuga's Formula* which expresses $\Phi \circ \Delta \circ \Phi^{-1}$ in terms of an operator coming from the

center of the universal enveloping algebra $U(\mathfrak{g})$ of $\mathfrak{g}$. We now describe this formula.

Let $\{X_i : i=1, 2, \ldots, s\}$ be an orthonormal basis for the restriction of the Killing form $A(\cdot, \cdot)$ to $\mathfrak{p}$ and $\{Y_j : j=1, 2, \ldots, r\}$ be a basis for $\mathfrak{k}$ such that $A(Y_i, Y_j) = -\delta_{ij}$. Consider the element C of $U(\mathfrak{g})$ given by:

$$C = \sum_{i=1}^{s} X_i^2 - \sum_{j=1}^{r} Y_j^2.$$

Then C is in the center of $U(\mathfrak{g})$; that is, it gives rise to a bi-invariant differential operator in $C^\infty(G)$. We may also consider $\rho(C) \in \operatorname{End} V$ defined by

$$\rho(C) = \sum_{i=1}^{s} \rho(X_i)^2 - \sum_{j=1}^{r} \rho(Y_j)^2.$$

Then $\rho(C)$ commutes with $\rho(X)$ for any $X \in \mathfrak{g}$. Here and in the line above we are using the symbol ρ to denote the induced representation of $\mathfrak{g}$.

Kuga's Formula then states:

$$\Phi \circ \Delta \circ \Phi^{-1} = -1 \otimes C \otimes 1 + 1 \otimes 1 \otimes \rho(C).$$

We are now in a position to decompose $H^*(M, \tilde{V})$. We first interpret the almost complex structure on D in terms of the Lie algebra $\mathfrak{k}$. We observe that we may identify $\mathbf{C}^n$ with $\mathfrak{p}$ by sending $(z_1, z_2, \ldots, z_n)$ to the matrix with the last column $(z_1, z_2, \ldots, z_n, 0)$ and last row $(\bar{z}_1, \bar{z}_2, \ldots, \bar{z}_n, 0)$ and all other entries equal to zero. We let J be the diagonal matrix in $\mathfrak{k}$ with diagonal entries

$$\left(\frac{i}{n+1}, \ldots, \frac{i}{n+1}, \frac{-ni}{n+1}\right).$$

Then the action of $\operatorname{ad} J$ on $\mathfrak{p}$ coincides with multiplication by i on $\mathbf{C}^n$. We observe that $\mathfrak{p}$ is to be considered as a $2n$ dimensional real vector space with J operating by the usual $2n$ by $2n$ matrix.

Since V is a complex vector space we may decompose V into the eigenspaces of $\rho(J)$. Clearly this is just the decomposition $V = V^+ \oplus V^-$. The projections q_+ and q_- are then polynomials in $\rho(J)$. Since $J \in \mathfrak{g}$ we know $\rho(J)$ commutes with $\rho(C)$ and we obtain the following critical lemma.

Lemma 4.4. *Δ commutes with p_+ and p_-.*

Proof. The lemma follows since $1 \otimes C \otimes 1$ and $1 \otimes 1 \otimes \rho(C)$ commute with $1 \otimes 1 \otimes q_+$ and $1 \otimes 1 \otimes q_-$.

Corollary. *$\omega \in \mathscr{A}^r(M, \tilde{V})$ is harmonic if and only if $p_+ \omega$ and $p_- \omega$ are harmonic.*

In this way we obtain a decomposition:

$$H^r(M, \tilde{V}) = H^r(M, E_+) \oplus H^r(M, E_-).$$

We may decompose the harmonic r-forms still further by decomposing $\Lambda^r \mathfrak{p}^* \otimes \mathbf{C}$ into Hodge types.

Lemma 4.5. *Δ commutes with the projection of $\mathscr{A}^r(M, \tilde{V})$ on any Hodge type.*

Proof. After conjugation by Φ such a projection operator will be of the form $p \otimes 1 \otimes 1$ so the lemma follows from Kuga's Formula.

Returning to the case in hand we obtain a decomposition:

$$H^1(M, \tilde{V}) = H^{1,0}(M, E_+) \oplus H^{1,0}(M, E_-) \oplus H^{0,1}(M, E_+) \oplus H^{0,1}(M, E_-).$$

We now state one of our main results.

Theorem 4.1. (i) *The above decomposition is an orthogonal splitting for Q.*

(ii) *Q is negative definite on the first and last summands and positive definite on the other two.*

Proof. By Lemma 4.2 the form Q may be computed by integrating a function over M. By Theorem 3.1 this function is identically zero, positive or negative as required in the theorem. The result follows by integrating the pointwise statements of Theorem 3.1.

The positive definiteness of Q is then equivalent to the following theorem which follows from a general result of Matsushima-Murakami [9]. We leave to the reader the task of verifying that the results of [9] generalize to reductive groups with compact center.

Theorem 4.2. *$H^{1,0}(M, E_+) = H^{0,1}(M, E_-) = 0$ and consequently Q is positive definite on $H^1(M, \tilde{V})$.*

Proof. The eigenvalues of $\rho(J)$ are $i/n+1$ on V_+ and $-in/n+1$ on V_-. By Murakami [10], Theorem 6.1, or [9], Theorem 6.1 (see the discussion preceding the theorem), we find that the $(0, q)$ cohomology of M with values in $\tilde{V}$ occurs only in the highest weight space of $\rho(J)$ and the $(p, 0)$ cohomology occurs only in the lowest weight space of $\rho(J)$. With this the theorem is proved.

Corollary. *Let c be a non-zero element of $H^1(\Gamma, \mathbf{C}^{n+1})$. Then $[c, c] \neq 0$.*

Remark. The results of this paper go over to the case

$$G_0 = S(U(n, 1) \times U(k)) \quad \text{and} \quad G = SU(n+k, 1).$$

In this case $\mathfrak{g}_1$ is isomorphic to V^k twisted by $U(k)$ acting from the right. We take for a Hermitian form on $\mathfrak{g}_1$ the direct sum of $\langle\ ,\ \rangle$ with itself k times. The results in Sects. 1 and 4 generalize in a straightforward manner to this case. However a modification of Lemma 2.4 is required since the centralizer of $SU(n, 1)$ in G_0 is no longer abelian. For x in $\mathfrak{g}_0$ we let x'' denote the projection of x on the subalgebra $u(n, 1)$ of $\mathfrak{g}_0$ (the projection is relative the obvious product decomposition). But (in the notation of Lemma 2.4) since ρ_t is normalized w_k'' is in the center of $u(n, 1)$ for $k \leqq m$. The proof of Lemma 2.4 then shows that $\delta w_{2m}'' = [v_m, v_m]''$. Since $Q(v_m)$ factors through $[v_m, v_m]''$ the above formula implies Q is isotropic. We leave the details to the reader.

Corresponding results seem to hold for embeddings of $U(n, 1)$, in the automorphism groups of other bounded dymmetric domains but we have not yet studied the question in detail.

References

1. Borel, A., Wallach, N.: Continuous cohomology, discrete subgroups, and representations of reductive groups. Ann. Math. Stud. **94**. Princeton University Press 1980
2. Bourbaki, N.: Element de mathematiques. Groupes et algebres de Lie. Chapitres 2 et 3. Paris: Hermann 1972
3. Faran, J.: Maps from the two-ball to the three-ball. Invent. Math. **80**, 441–475 (1982)
4. Goldman, W.M.: Characteristic classes and representations of discrete subgroups of Lie groups. Bull. Am. Math. Soc. **6**, 91–94 (1982)
5. Goldman, W.M.: Representations of fundamental groups of surfaces. In: Geometry and Topology, Proceedings, University of Maryland. Lect. Notes Math., vol. 1167, pp. 95–117. Berlin-Heidelberg-New York: Springer 1987
6. Johnson, D., Millson, J.J.: Deformation spaces associated to compact hyperbolic manifolds. In: Discrete groups in geometry and analysis. Proceedings of a conference held at Yale University in honor of G.D. Mostow, Progress in Math. Boston: Birkhäuser 1987
7. Kobayashi, S., Nomizu, K.: Foundations of differential geometry. Interscience tracts in pure and applied mathematics, Number 15, Vol. II
8. Kourouniotis, C.: Deformations of hyperbolic structures on manifolds of several dimensions, Math. Proc. Camb. Phil. Soc. **98**, 247–261 (1985)
9. Matsushima, Y., Murakami, S.: On vector bundle valued harmonic forms and automorphic forms on symmetric spaces. Ann. Math. **78** (2), 365–416 (1963)
10. Murakami, S.: Cohomology groups of vector-valued forms on symmetric spaces. Lecture notes, University of Chicago, 1966
11. Milnor, J.: Singular points of complex hypersurfaces. Ann. Math. Stud., vol. 61. Princeton University Press 1968
12. Raghunathan, M.S.: Discrete subgroups of Lie groups. Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 68. Berlin-Heidelberg-New York: Springer 1972
13. Thurston, W.P.: The geometry and topology of three-manifolds. Princeton University Lecture Notes, 1978
14. Tits, J.: Free subgroups in linear groups. J. Algebra **20**, 250–270 (1972)
15. Webster, S.M.: On mapping an n-ball into an $(n+1)$-ball in complex space. Pacific J. Math. **81**, 267–272 (1979)
16. Weil, A.: On discrete subgroups of Lie groups. Ann. Math. **72** (2), 369–384 (1960)
17. Zimmer, R.: Ergodic theory and semisimple groups. Monographs in Mathematics. Boston: Birkhäuser, 1985
18. Artin, M.: On the solutions of analytic equations. Invent. Math. **5**, 277–291 (1968)

Oblatum 9-XI-1985 & 5-V-1986

Note added in proof

The special case of our conjecture in which ρ is discrete and faithful has been proved by K. Corlette in his PhD thesis at Harvard.

Invent. math. 88, 521-554 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Determinacy and unipotency

J.W. Bruce[1], A.A. du Plessis[2], and C.T.C. Wall[3]

[1] Department of Pure Mathematics, Merz Caut, The University, Newcastle-on-Tyne NE1 7RU, UK
[2] Matematisk Institut, Universitet Aarhus, ny Munkegåde, Bygning 530, 8000 Aarhus C, Denmark
[3] Department of Pure Mathematics, The University, Liverpool L 69 3BX, UK

Introduction

One of the key ideas of singularity theory has been that smooth map-germs are frequently determined (up to some convenient equivalence relation E) by the terms of degree $\leqq k$ in their Taylor expansion; or, as we shall say, $k-E$-determined. A crucial contribution to the theory was made by Mather in [20] where he defined the equivalence relations $\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ and $\mathscr{A}$ which have become standard, and obtained necessary and sufficient conditions for finite determinacy with respect to them. Since then, similar conditions have been obtained for many other equivalence relations: a very general statement, which seems to encompass all known examples, appears in [10].

Mather's results in the $\mathscr{L}$ and $\mathscr{A}$ cases give extremely poor estimates for the order of determinacy; improvements leading to effective estimates have since been given in [13-15] and [22] (see [26] for a survey of known results on determinacy up to 1980).

The notion of determinacy was generalised in [22, §4] and some effective results on determinacy in this more general setting are obtained in [22, §4], [17, 16] (though at the cost of much hard labour).

In this paper we develop a new technique for these questions. The general idea remains that we seek information about orbits of group actions by studying their tangent spaces. Moreover, since we can regard the question of finite determinacy as already solved, it is enough to work at the level of formal power series, so it is natural to proceed inductively on degree. Thus for example if f is known to be finitely E-determined, then to prove that f is $k-E$-determined it is sufficient to be able to construct, for any $l \geqq k$ and any g with the same l-jet as f, a germ E-equivalent to g with the same $(l+1)$-jet as f.

Our investigations began with the observation by the first author that for the equivalence relations defined by the groups $\mathscr{G}_s$ ($\mathscr{G}=\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ or $\mathscr{A}$; $s \geqq 1$), if the natural condition for f to be $k-\mathscr{G}_s$-determined (shown by Mather [20] to be necessary) holds, then a surprisingly short and simple argument shows that the construction alluded to above can be made, so the condition is also sufficient. This is already a new result.

It was clear that the method applied to several other situations under conditions which could be formalised. The second author then observed that the essential condition on the group $\mathscr{G}$ needed for the argument to work was that certain natural representations of the Lie algebra $L(J^1\mathscr{G})$ were nilpotent. The third author then pointed out that this condition held if $J^1\mathscr{G}$ was a unipotent affine algebraic group. Since, for well-behaved $\mathscr{G}$, this is equivalent to having $J^r\mathscr{G}$ a unipotent group, it then seemed worth re-casting the method in the context of general results on orbits of unipotent groups. This, then, is the main thrust of this paper.

Our results yield, in particular, a precise estimate of the order of determinacy of a map-germ with respect to any appropriately well-behaved subgroup $\mathscr{H}$ of $\mathscr{K}$ (not only those with $J^1\mathscr{H}$ nilpotent). However, a more important aspect of the results is generalised determinacy: we obtain astonishingly large ranges of determinacy for map-germs with respect to all (well-behaved) subgroups $\mathscr{H}$ of $\mathscr{K}$.

In some cases this even gives a precise determination of the equivalence class, and in many others reduces further calculation to a handful of low order terms. Results of this power can be used to obtain classification and recognition theorems, and indeed this is the main reason for their interest. In this paper we do not give extensive results of this type, but content ourselves by illustrating the strength of our results with numerous examples drawn from a wide range of situations.

The plan of this paper is as follows. In §1 we provide a technical introduction describing the nature of our results more fully and exhibiting the rôle played by results on algebraic actions of unipotent groups. By contrast, in §2 we give an 'elementary' proof of the characterisation of $k-\mathscr{G}$-determinacy for appropriate $\mathscr{G}$: the arguments here are closer in style to earlier papers in this area. In §3 we collect results on unipotent algebraic groups and their actions. Our general references here are [3] and [19]: we refer to them for background and basic results, but hope to include enough connecting material for the reader to follow the main proofs. The next §4 contains the precise formulation and proof of our main theorems. These are illustrated by §5, showing how they can be applied to a wide variety of situations, and §6, which contains a number of calculations for particular $\mathscr{A}$-determinacy questions.

Thanks are due to H.H. Andersen, for helpful discussions; to R.W. Richardson, for correspondence about §3; to the referee, for simplifying the proof of (3.6); to SERC for supporting the third author, and paying for him to visit Aarhus, and to the university of Aarhus, for hospitality during that visit.

Notations and terminology

We shall work throughout over a field $\mathbb{F}$, which is either $\mathbb{R}$ or $\mathbb{C}$; and in one of the categories C^τ where either $\tau=\infty$ ($\mathbb{F}=\mathbb{R}$, C^∞-maps), $\tau=\mathbb{R}-\omega$ ($\mathbb{F}=\mathbb{R}$, real analytic maps) or $\tau=\mathbb{C}-\omega$ ($\mathbb{F}=\mathbb{C}$, complex analytic map). Write C_n for the ring of germs at the origin of C^τ-functions $\mathbb{F}^n\to\mathbb{F}$, and $C_n^{\times p}$ for the free C_n-module of germs at of C^τ-maps $\mathbb{F}^n\to\mathbb{F}^p$.

The ideal in C_n of functions vanishing at 0 is denoted by $\mathfrak{m}_n$. Thus $\mathfrak{m}_n . C_n^{\times p}$ denotes the space of germs with target $0 \in \mathbb{F}^p$; and for any $r \in \mathbb{N}$, $\mathfrak{m}_n^r \cdot C_n^{\times p}$ the subspace of those whose Taylor expansions at 0 contain no terms of degree $< r$. The quotient $C_n^{\times p}/(\mathfrak{m}_n^{r+1} . C_n^{\times p})$ is thus the space $J^r(n, p)$ of r-jets of map-germs $(\mathbb{F}^n, 0) \to \mathbb{F}^p$. For any subspace M of $C_n^{\times p}$, we write $J^r M$ for its image in $J^r(n, p)$; the r-jet of a map f is denoted by $j^r f$.

If E is an equivalence relation on C^τ map-germs $(\mathbb{F}^n, 0) \to \mathbb{F}^p$ (perhaps only defined on some subset), and M is a vector subspace of $\mathfrak{m}_n . C_n^{\times p}$, we say that a C^τ map-germ $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ is *$M-E$-determined* if $f + M$ is contained in the E-equivalence class of f; we also say that M is a *range of E-determinacy* of f. In the case when $M = \mathfrak{m}_n^{r+1} . C_n^{\times p}$ we speak of *$r-E$-determinacy*.

In discussions of (finite dimensional) Lie groups and algebraic groups and their actions, we shall denote the group by a letter such as G and its Lie algebra by LG. Given a smooth or algebraic action of G on a manifold M, we write the group on the left. For $x \in M$, $T_x M$ denotes the tangent space to M at x; $LG . x \subset T_x M$ is the image of LG by the differential of the action. This coincides with the tangent space $T_x(G . x)$ of the orbit of x under G (recall in the algebraic case that we assume characteristic 0).

We use similar notations in the infinite dimensional case. $\mathscr{K}$ will denote Mather's [20] group of contact equivalences, viz. the group of diffeomorphisms of $\mathbb{F}^n \times \mathbb{F}^p$ which respect the subspace $\mathbb{F}^n \times 0$ and the projection on $\mathbb{F}^n$. It acts on the space $\mathfrak{m}_n . C_n^{\times p}$ of map-germs $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ via its action on the graph of f. We have the subgroups $\mathscr{C}$ (projecting to the identity on $\mathbb{F}^n$) and the groups $\mathscr{R}$, $\mathscr{L}$ of diffeomorphism-germs of $\mathbb{F}^n$ and $\mathbb{F}^p$, extended to $\mathbb{F}^n \times \mathbb{F}^p$ as products with the identity on the other factor; we also have $\mathscr{A} = \mathscr{R} \times \mathscr{L}$. In the cases of $\mathscr{R}$, $\mathscr{L}$, the action on map-germs reduces to composition of f on the right (resp. left) with a diffeomorphism-germ (in the latter case, the inverse of the diffeomorphism above). The action of $\mathscr{K}$ on $\mathfrak{m}_n \cdot C_n^{\times p}$ induces an action on the space $(\mathfrak{m}_n . C_n^{\times p}/\mathfrak{m}_n^{r+1} . C_n^{\times p})$ of r-jets: we write $\mathscr{K}_r$ for the kernel of this action, and $J^r \mathscr{K}$ for its image: similarly for any subgroup $\mathscr{G}$ of $\mathscr{K}$, set $\mathscr{G}_r = \mathscr{K}_r \cap \mathscr{G}$, $J^r \mathscr{G}$ for the image of $\mathscr{G}$ in $J^r \mathscr{K}$. Lie algebras of these groups will be considered more fully in §4 below: we identify (following Mather, but not notationally) $L\mathscr{R}$ with the Lie algebra of germs at 0 of C^τ-vector fields on $\mathbb{F}^n$; similarly, $L\mathscr{K}$ is a Lie algebra of germs of vector fields on $\mathbb{F}^{n+p}$. For a subgroup $\mathscr{G}$ of $\mathscr{K}$, $L\mathscr{G}$ denotes the set of derivatives at the identity of smooth 1-parameter families in $\mathscr{G} \subset \mathscr{K}$. Thus, for example, we have an exact sequence

$$0 \to L\mathscr{K}_s \to L\mathscr{K} \to L(J^s \mathscr{K}) \to 0.$$

An affine space A is a principal homogeneous space of a vector space V_A: its group of translations, which we identify with any tangent space to A. In general, if V is an algebraic variety, $V(\mathbb{F})$ denotes the set of points of V rational over $\mathbb{F}$: however, we write $\mathbb{F}^\times$ for $GL_1(\mathbb{F})$.

§ 1. The main ideas, and first version of results

We first observe that – for the standard equivalence relations and obvious modifications of them – the equivalence relation is generated by the action of a

group of diffeomorphism-germs which induces an action on the space of k-jets of map-germs. Moreover, this action can be regarded as an algebraic group action of an affine algebraic group. (This observation is by no means new - it goes back at least as far as [25]).

Now in [20], Mather showed how to calculate the tangent spaces to the orbits of these group actions. Thus the following became the basis for work in this area.

(1.1) **Lemma.** *Let G be a Lie group acting smoothly on a finite-dimensional manifold V. Let W be a connected submanifold of V. Then W is contained in a single orbit of G if and only if*

(i) *for each $x \in W$, $T_x W \subset T_x(G \cdot x)$, and*
(ii) *$\dim T_x(G \cdot x)$ is constant for $x \in W$.*

The proof [21, 3.1] is a straightforward exercise in Lie group theory. We refer to this hereinafter as "Mather's Lemma".

To obtain determinacy theorems, we are of course concerned with the case when V is an affine space and W an affine subspace - so that $T_x W$ is 'constant' along W. This explains why in practice (1.1) has always been applied via the following special case.

(1.2) **Corollary.** *Let G be a Lie group acting smoothly on an affine space A; let B be a vector subspace of V_A. If $x \in A$, then $x+B$ is contained in a single orbit of G if*

(i) *$B \subset LG \cdot x$,*
(ii) *for all $y \in (x+B)$, $LG \cdot y = LG \cdot x$.*

We have observed above that in the cases of interest, G is an algebraic group and the action on A is also algebraic. Our main idea is to concentrate further on the case when G is unipotent. The background is recalled in § 3: here we remind the reader that equivalent conditions are that for one faithful (or that for every) linear representation ρ, $\rho(G)$ lies up to conjugacy in the group of upper unitriangular matrices. Now we have a useful result (for proof and discussion see (3.3)):

(1.3) **Proposition.** *Let U be a unipotent affine algebraic group over $\mathbb{F}$, acting algebraically on an affine variety V. Then the orbits of $U(\mathbb{F})$ are Zariski-closed in $V(\mathbb{F})$.*

We obtain immediately an avatar of Mather's lemma.

(1.4) **Lemma.** *Let U be a unipotent affine algebraic group over $\mathbb{F}$, acting algebraically on an affine variety V. Let W be an irreducible subvariety of V. Then $W(\mathbb{F})$ is contained in a single orbit of $U(\mathbb{F})$ if and only if there is a non-empty subset N of $W(\mathbb{F})$, open in the usual (not the Zariski) topology, such that*

(i) *for each $x \in N$, $T_x W \subset LU \cdot x$ and*
(ii) *$\dim LU \cdot x$ is constant for $x \in N$.*

For by Mather's lemma, N is contained in a single orbit of $U(\mathbb{F})$. Since this orbit is Zariski-closed, it contains the Zariski-closure of N. But this is the whole of $W(\mathbb{F})$, since W is irreducible and N does not lie in a lower dimensional subset.

As with Mather's lemma, the applications of (1.4) to determinacy come from the following special case.

(1.5) **Corollary.** *Let U be a unipotent affine algebraic group over $\mathbb{F}$, acting algebraically on an affine space A. Let B be a vector subspace of V_A. If $x \in A$, then $(x+B)$ is contained in a single orbit of U if*

(i) $B \subset LU \,.\, x$ *and*

(ii) *for all* $y \in (x+B)$, $LU \,.\, y \subset LU \,.\, x$.

To see this, note that $\dim LU \,.\, y$ is upper semi-continuous. It thus follows from (ii) that for all y in some open neighbourhood N of x in $(x+B)$, $LU \,.\, y = LU \,.\, x$. The result thus follows from (1.4).

The statement of (1.5) differs only slightly from that of (1.2): however, we will see that the difference is crucial.

Our converse results stem from the following, which is again proved in §3. It is rather deeper than the preceding, requiring the full force of the structure theory of affine algebraic groups in characteristic zero.

(1.6) **Proposition.** *Let G be an affine algebraic group over $\mathbb{F}$ acting algebraically on an affine space A, with $G(\mathbb{F})$ transitive on $A(\mathbb{F})$. Then there is a unipotent affine algebraic subgroup U of G such that $U(\mathbb{F})$ is transitive on $A(\mathbb{F})$.*

(1.7) **Corollary.** *Let G be an affine algebraic group over $\mathbb{F}$ acting algebraically on an affine space A. Let B be a vector subspace of V_A such that there is an induced action of G on A/B. If, for some $x \in A$, $(x+B) \subset G \,.\, x$, then there is a unipotent subgroup U of G such that $(x+B) = U \,.\, x$.*

Proof. Let S be the stabiliser of $x+B$ for the action of G on A/B: then S is an affine algebraic subgroup of G. Since $(x+B) \subset G \,.\, x$, it follows at once that $(x+B) = S \,.\, x$. Applying (1.6) to the action of S on the affine space $(x+B)$, we see that S has a unipotent subgroup U such that $x+B = U \,.\, x$.

Up to this point, the reader may feel that unipotent algebraic groups are rather special and not of particular interest. That they are indeed very common in our context is shown by the next result.

(1.8) **Proposition.** *Let H be a closed connected subgroup of the group $J^s\mathscr{K}$ for some $s \geqq 1$. Then H is a unipotent affine algebraic group if and only if J^1H is unipotent.*

Here J^1H denotes the projection of H on $J^1\mathscr{K}$, which in turn is naturally identified with $GL(n) \times GL(p)$, and acts linearly on $\mathbb{F}^{n+p}$.

Proof. Necessity is clear, since any quotient of a unipotent group is unipotent. For sufficiency we regard $\mathscr{K}$ as a group of diffeomorphism-germs of $\mathbb{F}^{n+p}$: then $J^s\mathscr{K}$ (and hence H) acts faithfully on the space V of s-jets at 0 of smooth functions on $\mathbb{F}^{n+p}$, which we identify with the space of polynomials of degree $\leqq s$. By (3.0) it will suffice to show that each element of H acts unipotently.

Choose a basis $x_1, \ldots, x_{n+p}$ of $(\mathbb{F}^{n+p})^*$ with respect to which the representation of J^1H is lower unitriangular. Then arrange all monomials $\mathbf{x}^{\mathbf{a}} = x_1^{a_1} \ldots x_{n+p}^{a_{n+p}}$ in order of increasing degree, and lexicographical order for those of equal degree. Then the matrices representing elements of H with respect to

this basis are lower unitriangular, for the image of a monomial of degree r is given, modulo monomials of higher degree, by its image under the rth symmetric power of the representation on $\mathbb{F}^{n+p}$.

It follows from (1.8) that if J^1H is trivial, H is already unipotent; and any H has unipotent subgroups of rather low codimension.

We have now summarised our techniques: to conclude this introduction we give as an example a rather definitive theorem on "standard" determinacy.

Definition. Let $\mathscr{G}$ be one of $\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ or $\mathscr{A}$. A subgroup $\mathscr{H}$ of $\mathscr{G}$ is *strongly closed* in $\mathscr{G}$ if $\mathscr{G}_s \subset \mathscr{H}$ for some s, and $J^s\mathscr{H}$ is a closed subgroup of $J^s\mathscr{G}$. If moreover $J^s\mathscr{H}$ is an algebraic subgroup we will call $\mathscr{H}$ *strongly Z-closed.*

It is not hard to show that if $\mathscr{H}$ is strongly closed in $\mathscr{G}$ as above, we have $J^r(L\mathscr{H}) = L(J^r\mathscr{H})$. We defer presenting the argument till § 4, where it is given in a more general context.

(1.9) **Theorem.** *Let $\mathscr{G}$ be one of $\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ or $\mathscr{A}$. Let $\mathscr{H}$ be a strongly Z-closed subgroup of $\mathscr{G}$. Let $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ be a C^τ map-germ. Then for any $r < \infty$, f is $r-\mathscr{H}$-determined if and only if there is a strongly closed subgroup $\mathscr{U} \subset \mathscr{H}$ of $\mathscr{G}$, with $J^1\mathscr{U}$ unipotent, for which*

$$\mathfrak{m}_n^{r+1} \cdot C_n^{\times p} \subset L\mathscr{U} \cdot f.$$

Proof. "*If*". Since $\mathfrak{m}_n^{r+1} \cdot C_n^{\times p} \subset L\mathscr{U} \cdot f \subset L\mathscr{G} \cdot f$, f is finitely $\mathscr{G}$-determined and even (see [20] or [26, 4.4]) finitely $\mathscr{G}_s$-determined for any s. Choose s with $\mathscr{G}_s \subset \mathscr{U}$ and k such that f is $k-\mathscr{G}_s$-determined. Then it is enough to show that the affine space of k-jets with the same r-jet as f is contained in the $J^k\mathscr{U}$-orbit of $j^k f$.

Now by the standard approximation lemmas (see [20] or [26, 1.3]), if g has the same r-jet as f,

$$\begin{aligned} L\mathscr{U} \cdot g &\subset L\mathscr{U} \cdot f + \mathfrak{m}_n^{r+1} \cdot C_n^{\times p} \\ &= L\mathscr{U} \cdot f \quad \text{by hypothesis.} \end{aligned}$$

Passing to k-jets, we obtain

$$L(J^k\mathscr{U}) \cdot j^k g \subset L(J^k\mathscr{U}) \cdot j^k f,$$

for all g having the same r-jet as f. Now $J^k\mathscr{U}$ is unipotent by (1.8) so (1.5) applies, completing the argument.

"*Only if.*" Since f is $r-\mathscr{H}$-determined, it is $r-\mathscr{G}$-determined, hence (choosing as above s with $\mathscr{G}_s \subset \mathscr{H}$) $k-\mathscr{G}_s$-determined for some k. Clearly $k \geqq s$ (the $k-\mathscr{G}_s$-determinacy implies $\mathfrak{m}_n^{k+1} \cdot C_n^{\times p} \subset L\mathscr{G}_s \cdot f$ while obviously $L\mathscr{G}_s \cdot f \subset \mathfrak{m}_n^{s+1} \cdot C_n^{\times p}$). Now $J^k\mathscr{H} \subset J^k\mathscr{G}$ acts on $J^k(n,p)$, with an induced action on $J^r(n,p) = J^k(n,p)/J^k(\mathfrak{m}_n^{r+1} \cdot C_n^{\times p})$. We can thus apply (1.7) with $A = J^k(n,p)$, $B = J^k(\mathfrak{m}_n^{r+1} \cdot C_n^{\times p})$, $G = J^k\mathscr{H}$ to see that there exists a unipotent subgroup U of $J^k\mathscr{H}$ such that $U \cdot j^k f = j^k f + J^k(\mathfrak{m}_n^{r+1} \cdot C_n^{\times p})$. Now define $\mathscr{U}$ to be the subgroup of $\mathscr{H}$ containing $\mathscr{G}_k$ and with $J^k\mathscr{U} = U$. Since J^1U is unipotent, so is $J^1\mathscr{U}$. Since f is $k-\mathscr{G}_s$-determined,

$$L\mathscr{U} \cdot f \supset L\mathscr{G}_s \cdot f \supset \mathfrak{m}_n^{k+1} \cdot C_n^{\times p}$$

and since also $LU \,.\, j^k f = J^k(\mathfrak{m}_n^{r+1} \,.\, C_n^{\times p})$, it follows that

$$L\mathscr{U} \,.\, f \supset \mathfrak{m}_n^{r+1} \,.\, C_n^{\times p}$$

as required.

§ 2. Alternative (traditional) proof

Although we feel unipotent actions form the natural setting for our results, it is possible to prove most of them without recourse to algebraic geometry. In this paragraph we illustrate this by proving directly a slightly weaker version of (1.9), viz.

(2.1) **Theorem.** *Let $\mathscr{G}$ be one of $\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ and $\mathscr{A}$, and $\mathscr{H}$ a strongly closed subgroup of $\mathscr{G}$ such that $L(J^1\mathscr{H})$ acts nilpotently on $\mathbb{F}^{n+p}$. Then a C^r map-germ $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ is $k-\mathscr{H}$-determined if and only if $\mathfrak{m}_n^{k+1} \,.\, C_n^{\times p} \subset L\mathscr{H} \,.\, f$.*

We begin with an elementary lemma on Lie group actions.

(2.2) **Lemma.** *Let G be a connected Lie group acting smoothly on an affine space A; let B be a vector subspace of V_A. Then the action of G induces an action on A/B if and only if, for all $f \in A$, $b \in B$, $l \in LG$,*

$$\{l(f+b) - l(f)\} \in B.$$

Proof. First suppose there is an induced action on the quotient. Now l is tangent to a smooth curve $\phi: (\mathbb{R}, 0) \to (G, 1)$ and by hypothesis

$$\{\phi(t)(f+b) - \phi(t)(f)\} \in B$$

for all $t \in \mathbb{R}$. Differentiating and setting $t = 0$ yields the desired condition.

For the converse, let g be any element of G: choose a smooth path $\psi: [0, 1] \to G$ with $\psi(0) = 1_G$, $\psi(1) = g$. Differentiating this, and (left) translating these tangent vectors to the identity, yields a smooth path $\phi: [0, 1] \to LG$ where

$$\phi(t) = (L\psi(t))^{-1} \psi'(t).$$

Now $g(f+b)$ and $g(f)$ are the end points of the smooth paths $\psi(t)(f+b)$, $\psi(t)(f)$ obtained by integrating the ordinary differential equation

$$\frac{d}{dt} a(t) = \phi(t) \,.\, a(t)$$

with initial points $f+b$, f respectively. These paths project to the paths in A/B obtained by integrating the corresponding equation there, which is well-defined by hypothesis. Thus the projections in A/B coincide, so $\{g(f+b) - g(f)\} \in B$, which proves the result.

We return to the theorem. It will be convenient to write, for any $r, s \in \mathbb{N}$,

$$M_{r,s}(\mathscr{H}) = \sum_{i \geqq s} (L\mathscr{H})^i \,.\, (\mathfrak{m}_n^r \, C_n^{\times p}) + \mathfrak{m}_n^{r+1} \,.\, C_n^{\times p}.$$

It is important to observe that (see proof of (1.8)) $L\mathscr{H}$ acts nilpotently on $\mathfrak{m}_n^r C_n^{\times p}/\mathfrak{m}_n^{r+1} C_n^{\times p}$, so for each r there exists $t_r < \infty$ with

$$M_{r,s}(\mathscr{H}) = \mathfrak{m}_n^{r+1} C_n^{\times p} = M_{r+1,0}(\mathscr{H})$$

for all $s \geqq t_r$. In particular, the above sum is finite.

(2.3) **Lemma.** *Let* $f \in \mathfrak{m}_n . C_n^{\times p}$, $b \in M_{r,s}(\mathscr{H})$, $l \in L\mathscr{H}$. *Then*

$$\{l(f+b) - lf\} \in M_{r,s+1}(\mathscr{H}).$$

Proof. Write $l = l_0 + l_1$ with

$$l_0 \in \mathrm{End}(\mathbb{F}^n) \oplus \mathrm{End}(\mathbb{F}^p)$$
$$l_1 \in L\mathscr{K}_1.$$

Since l_0 acts linearly, we have

$$\begin{aligned} l(f+b) - l(f) &= l_0(f+b) - l_0(f) + l_1(f+b) - l_1(f) \\ &= l_0(b) + l_1(f+b) - l_1(f) \\ &= l(b) + l_1(f+b) - l_1(f) - l_1(b). \end{aligned}$$

Now as $b \in \mathfrak{m}_n^r . C_n^{\times p}$, it follows from standard approximation lemmas (see e.g. [26, 1.3]) that $\{l_1(f+b) - l_1(f)\} \in \mathfrak{m}_n^{r+1} . C_n^{\times p}$, hence also (taking $f=0$) that $l_1(b) \in \mathfrak{m}_n^{r+1} . C_n^{\times p}$. Finally

$$l(b) \in L\mathscr{H} . M_{r,s}(\mathscr{H}) \subset M_{r,s+1}(\mathscr{H})$$

by definition.

Proof of (2.1). Choose s so that $\mathscr{H} \supset \mathscr{G}_s$. Now if f is $k-\mathscr{H}$-determined, it is à fortiori $k-\mathscr{G}$-determined so (by [20] or see [26]) also finitely $\mathscr{G}_s$-determined: say $l-\mathscr{G}_s$-determined. Hence (same references)

$$\mathfrak{m}_n^{l+1} . C_n^{\times p} \subset L\mathscr{G}_s . f \subset L\mathscr{H} . f.$$

But also the set of l-jets with the same k-jet as f is contained in the $J^l\mathscr{H}$-orbit of $j^l f$. Taking tangent spaces at $j^l f$ yields

$$J^l(\mathfrak{m}_n^{k+1} . C_n^{\times p}) \subset L(J^l\mathscr{H}) . j^l f = J^l(L\mathscr{H}) . j^l f$$

i.e.

$$\begin{aligned} \mathfrak{m}_n^{k+1} . C_n^{\times p} &\subset L\mathscr{H} . f + \mathfrak{m}_n^{l+1} C_n^{\times p} \\ &\subset L\mathscr{H} . f \end{aligned}$$

by the above.

Conversely, suppose this condition satisfied. Then (usual references) f is finitely $\mathscr{G}$-determined and indeed finitely $\mathscr{G}_s$-determined, so à fortiori finitely $\mathscr{H}$-determined: say $l-\mathscr{H}$-determined.

Now for any r, $\mathscr{H}$ acts (via $J^r\mathscr{H}$) on $\mathfrak{m}_n . C_n^{\times p}/\mathfrak{m}_n^{r+1} . C_n^{\times p}$. It follows from (2.2) and (2.1) that we have an induced action on $\mathfrak{m}_n . C_n^{\times p}/M_{r,s}(\mathscr{H})$ for any r, s. We apply (1.2) to show that, provided $r \geqq k+1$, $f + M_{r,s}(\mathscr{H})/M_{r,s+1}(\mathscr{H})$ lies in a single orbit of the corresponding action. Indeed, hypothesis (i) of (1.2) follows from the hypothesis of (2.1) and (ii) follows at once from (2.3). Thus any g_s with

$g_s - f \in M_{r,s}(\mathscr{H})$ lies in the same $\mathscr{H}$-orbit as a g_{s+1} with $g_{s+1} - f \in M_{r,s+1}(\mathscr{H})$. It follows (by induction on s) that any g with $g - f \in M_{r,0}(\mathscr{H}) = \mathfrak{m}_n^r \cdot C_n^{\times p}$ lies in the same $\mathscr{H}$-orbit as some h with $h - f \in M_{r,t_r}(\mathscr{H}) = \mathfrak{m}_n^{r+1} \cdot C_n^{\times p}$, provided still that $r \geqq k+1$. Now an induction on r shows that any g with the same k-jet as f is $\mathscr{H}$-equivalent to some h with the same l-jet as f and hence (by the first part of the proof) is $\mathscr{H}$-equivalent to f.

(2.4) *Examples.* (i) The function-germ

$$f_1(x, y) = x^3 + xy^3$$

satisfies (as is easily checked)

$$L\mathscr{R}_1 \cdot f_1 + \mathbb{F}\left(x \cdot \frac{\partial f_1}{\partial y}\right) \supset \mathfrak{m}_2^5 \cdot C_2,$$

so is $4-\mathscr{R}$-determined.

(ii) The map-germs

$$f_2(x, y) = (x, y^3 + xy), \quad f_3(x, y) = (x, y^4 + xy)$$

satisfy respectively

$$L\mathscr{A}_1 \cdot f_2 + \mathbb{F}(y^3 + xy, 0) \supset \mathfrak{m}_2^4 \cdot C_2^{\times 2}$$

$$L\mathscr{A}_1 \cdot f_3 + \mathbb{F}(y^4 + xy, 0) + \mathbb{F}\left(x \cdot \frac{\partial f_2}{dy}\right) \supset \mathfrak{m}_2^5 \cdot C_2^{\times 2}.$$

Thus f_2 is $3-\mathscr{A}$-determined and f_3 is $4-\mathscr{A}$-determined.

A special case of (2.1) whose proof is much simpler ((2.3) can be omitted) is

(2.5) **Theorem.** *A C^r map-germ $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ is $r-\mathscr{G}_s$-determined ($s \geqq 1$) if and only if*

$$\mathfrak{m}_n^{r+1} \cdot C_n^{\times p} \subset L\mathscr{G}_s \cdot f.$$

Necessity here follows from [20]. The cases $\mathscr{G} = \mathscr{R}$, $\mathscr{C}$ and $\mathscr{K}$ of (2.6) are due to Kucharz for $p = 1$ and to Gervais in general (see [26, p. 506]). The case $\mathscr{G} = \mathscr{L}$ is fully discussed (with a simpler proof) in [9]. The most interesting case $\mathscr{G} = \mathscr{A}$ is new.

(2.5.1) **Corollary.** *A C^r map-germ f is $r-\mathscr{G}_s$-determined ($\mathscr{G} = \mathscr{L}$ or $\mathscr{A}$, $s \geqq 1$) if and only if*

$$\mathfrak{m}_n^{r+1} \cdot C_n^{\times p} \subset L\mathscr{G}_s \cdot f + \mathfrak{m}_n^{r+1} \cdot (f^* \mathfrak{m}_p \cdot C_n + \mathfrak{m}_n^{r+1})\, C_n^{\times p}.$$

This follows from (2.5) and the following variant of [26, 1.6]:

(2.6) **Lemma.** *Let C be a finitely generated C_n-module, $B \subset C$ a finitely generated C_n-submodule, $A \subset f^* \mathfrak{m}_p \cdot C$ a finitely generated C_p-submodule (via f), and M a proper, finitely generated ideal in C_n. If*

$$M \cdot C \subset A + B + M \cdot (f^* \mathfrak{m}_p + M) \cdot C,$$

then

$$M \cdot C \subset A + B.$$

Proof. Multiplying by M, and applying Nakayama's lemma, we obtain

$$M^2 . C \subset M . (A+B) + M^2 . (f^* \mathfrak{m}_p + M) . C,$$

$$\begin{aligned} M^2 . C &\subset M . (A+B) \\ &\subset M . f^* \mathfrak{m}_p . C + M . B \end{aligned}$$

whence by substitution,

$$M . C \subset A + B + M . f^* \mathfrak{m}_p . C.$$

The conclusion now follows from [26, 1.6(i)], applied to

$$M . C / ((M . C) \cap B).$$

(2.5.2.) **Corollary.** *If f satisfies*

$$\mathfrak{m}_n^l . C_n^{\times p} \subset L\mathscr{K} . f$$
$$\mathfrak{m}_n^{r+1} . C_n^{\times p} \subset L\mathscr{A}_1 . f + \mathfrak{m}_n^{l+r+1} . C_n^{\times p}$$

then f is $r-\mathscr{A}_1$-determined.

For the given inclusions imply

$$\mathfrak{m}_n^{r+1} . C_n^{\times p} \subset L\mathscr{A}_1 . f + \mathfrak{m}_n^{r+1} L\mathscr{K} . f$$

and hence the hypothesis of (2.5.1) for this case.

We observe that for $\mathscr{G} = \mathscr{C}$ or $\mathscr{L}$, the order of $\mathscr{G}_1$-determinacy equals the order of $\mathscr{G}$-determinacy [9], and for $\mathscr{G} = \mathscr{R}$ or $\mathscr{K}$, the two orders differ by at most 1, by an old result of Mather. For $\mathscr{G} = \mathscr{A}$, the difference can be arbitrarily large (see Ex (2.8) below). It can be bounded using the argument of [26, 4.4.4]. Since there is an error in the formulation of that result, we recapitulate here.

Suppose in fact that

$$\mathfrak{m}_n^l . C_n^{\times p} \subset tf(C_n^{\times n}) + (f^* \mathfrak{m}_p . C_n)^{\times p} \tag{2.7.1}$$

and that, for some $r \geqq s+1$,

$$\mathfrak{m}_n^r . C_n^{\times p} \subset tf((\mathfrak{m}_n^s . C_n)^{\times n}) + \omega f((\mathfrak{m}_p^s . C_p)^{\times p}). \tag{2.7.2}$$

Multiply (2.7.1) by $\mathfrak{m}_n^r$, and substitute for $(\mathfrak{m}_n^r)^{\times p}$ in the final term, using (2.7.2). We find

$$\begin{aligned} \mathfrak{m}_n^{l+r} . C_n^{\times p} &\subset tf((\mathfrak{m}_n^r . C_n)^{\times n}) + f^* \mathfrak{m}_p . (tf((\mathfrak{m}_n^s . C_n)^{\times n}) + \omega f((\mathfrak{m}_p^s . C_p)^{\times p}) \\ &\subset tf((\mathfrak{m}_n^{s+1} . C_n)^{\times n}) + \omega f((\mathfrak{m}_p^{s+1} . C_p)^{\times p}) \end{aligned}$$

since $f^* \mathfrak{m}_p \subset \mathfrak{m}_n$ and $r \geqq s+1$.

Now suppose f is $k-\mathscr{A}$-determined. Then [26, 1.2] $L\mathscr{A} . f \supset (\mathfrak{m}_n^{k+1} . C_n)^{\times p}$, so (2.4) holds with $s=1$, $r=k+1$. By the above, $(\mathfrak{m}_n^{k+l+1} . C_n)^{\times p} \subset L\mathscr{A}_1 . f$, so by (2.5) f is $(k+l)-\mathscr{A}_1$-determined. Thus the difference in orders of determinacy mentioned above is bounded by the number l in (2.7.1).

In general write r_s for the minimum value of $r \geqq s+1$ for which (2.7.2) holds, then a simple induction yields

(2.7.3) $$r_{s+k} \leqq r_s + k \,.\, l \quad \text{for all } s, k \geqq 0.$$

In particular,

$$r_{s+1} \leqq r_0 + (s+1) \,.\, l$$

so that, by (2.5), if f satisfies (2.7.1) and (2.7.2) with $s=0$, $r=r_0$, then f is $(r_0 + l(s+1) - 1) - \mathscr{A}_s$-determined for all $s \geqq 1$ (correcting the slip in [26, p. 507, $l. -8$).

By [15], such an f is also $(r_0 + l) - \mathscr{A}$-determined.

All these results are in fact best possible with the given hypotheses.

(2.8) *Example.* Consider the germ

$$f\colon (u_1, \ldots, u_{l-1}, y) \mapsto \left(u_1, \ldots, u_{l-1}, y^{l+1} + \sum_{i-1}^{l-1} u_i y^i\right).$$

This is infinitesimally stable, so $r_0 = 1$, whilst it is easy to see that (2.7.1) holds (with the given l).

By the above,

$$r_s \leqq 1 + s \,.\, l \quad \text{for all } s \geqq 0$$

and f is

$$\begin{cases} l+1 - \mathscr{A}\text{-determined} \\ (s+1)\, l - \mathscr{A}_s\text{-determined } (s \geqq 1). \end{cases}$$

The $\mathscr{A}$-determinacy estimate is obviously exact, and the $\mathscr{A}_s$-determinacy results are also exact by (2.5), since we have

$$r_s = 1 + s \,.\, l \quad \text{for all } s \geqq 0.$$

To see this, note that, for all $s \geqq 0$,

$$L\mathscr{A}_s f \cap (0, \mathfrak{m}_n) = (0, \mathfrak{m}_n^{s+1}) \left((l+1)\, y^l + \sum_{i=1}^{l-1} i u_i\, y^{i-1}\right) + f^* \mathfrak{m}_p^{s+1} \{y, \ldots, y^{l-1}\}$$

(where $n = p = l$).

Allot weight 1 to y and weight $l+1-i$ to u_i $(i = 1, \ldots, l-1)$. Writing elements of $L\mathscr{A}_s \,.\, f \cap (0, \mathfrak{m}_n)$ uniquely modulo

$$(0, \{\text{terms of wt.} \geqq (s+1)\, l + 1)$$

as

$$(0, \text{sum of monomials of wt.} < (s+1)\, l + 1),$$

we see that the only terms in $L\mathscr{A}_s \,.\, f \cap (0, \mathfrak{m}_n)$ in which

$$\{(0, u_1^r\, y^{(s+1-r)l}) \colon r = 0, \ldots, s\}$$

occur as summands are

$$tf\left(u_1^r\, y^{(s-r)l} \,.\, \frac{\partial}{\partial y}\right) \equiv (0, k \,.\, u_1^r\, y^{(s-r+1)l} + u_1^{r+1} \,.\, y^{(s-r)l})$$

(modulo $(0$, terms of wt. $\geqq (s+1)\, l + 1)$) for $r = 0, \ldots, s-1$ (*not* $r = s$); so that

$$(0, u_1^r\, y^{(s+1-r)l}) \notin L\mathscr{A}_s . f \quad \text{for } r=0, \ldots, s$$

and in particular,

$$(0, y^{(s+1)l}) \notin L\mathscr{A}_s . f$$

so that

$$\mathfrak{m}_n^{(s+1)l} . \theta_f \not\subset L\mathscr{A}_s . f$$

i.e.

$$r_{s+1} > (s+1)\, l.$$

Note in particular that, as previously mentioned, the difference between the orders of $\mathscr{A}$- and $\mathscr{A}_1$-determinacy of a map-germ can be arbitrarily large (for this f, we obtain the difference $l-1$).

We conclude this section by remarking that the method of proof of (2.1) can be adapted to prove our first main theorem. However, this introduces complications which tend to obscure the argument which is better expressed in general terms as explained below.

§ 3. Actions of unipotent groups

We first recall the basic definitions: for these we refer to Borel [3] and Hochschild [19]. An endomorphism α of an $\mathbb{F}$-vector space V ($\mathbb{F}$ a field) is called *nilpotent*, and $I+\alpha$ is called *unipotent*, if $\alpha^r=0$ for some r. If L is a Lie algebra (or indeed a vector space) of nilpotent endomorphisms of V, then [19, VII, 1.5] L is nilpotent in the sense that $L^r V=0$ for some r. Also [19, V, 2.1] or [3, I, 4.8] if G is a group of unipotent automorphisms of V, the Lie algebra generated by the endomorphisms $g-I$ ($g \in G$) is nilpotent.

The following, proved in [19, XVI, 4.2], was already used in § 1.

(3.0) **Lemma.** *If $\mathbb{F}=\mathbb{R}$ or $\mathbb{C}$, and G is a connected $\mathbb{F}$-Lie group of unipotent automorphisms of the finite-dimensional $\mathbb{F}$-vector space V, then there is an algebraic subgroup H of $GL(V)$ with $G=H(\mathbb{F})$.*

In general, an algebraic group of unipotent automorphisms of a $\mathbb{F}$-vector space is called a *unipotent algebraic group*. Such a group is always irreducible (as algebraic variety). If $\mathbb{F}$ has characteristic zero, an algebraic group G is unipotent if and only if its Lie algebra LG is nilpotent.

Any morphism of algebraic groups maps unipotent elements to unipotent elements [3, I, 4.4.4] [19, V, 2.3]; thus any (algebraic) representation of a unipotent group is unipotent, and hence [3, I, 4.8] can be put in upper unitriangular form by an appropriate choice of basis. Similarly any morphism of Lie algebras preserves nilpotent elements, so any (algebraic) representation of a nilpotent Lie algebra is nilpotent.

Although the next result is not essential for us, it helps to clarify what is going on.

(3.1) **Lemma.** *If U is a unipotent algebraic group over a field $\mathbb{F}$ of characteristic 0, then the exponential map*

$$\exp: LU \to U$$

is an isomorphism of affine varieties.

Proof. (See [19, VIII, 1.1]): Since LU is nilpotent, the exponential power series reduces to a polynomial map; similarly, so does the power series for $\log(I+g)$. These are mutually inverse by the usual identities.

The same idea is the basis for the following, which is due to [7]: a proof is also given in [4, (11.13)].

(3.2) **Proposition.** *Suppose U a unipotent algebraic group over a field $\mathbb{F}$ of characteristic 0, V an algebraic subgroup. Then U/V is isomorphic as variety to the affine space LU/LV.*

Proof. The lower central series of LU is defined by setting $L_1 = LU$ and, by induction on r, $L_r = [L_1, L_{r-1}]$. Since LU is nilpotent we have $L_{r+1} = \{0\}$ for some integer l: the least such is the *length* of LU. As is well known (and easily verified by induction), for all $r, s \geqq 1$

$$[L_r, L_s] \subset L_{r+s}.$$

Write $M_r = (LV) \cap L_r$, and choose (by downward induction on r) $C_r \supset C_{r+1}$ to be an additive complement to M_r in L_r. Set $C = C_1$, and define ϕ to be the composite

$$\phi\colon C \subset LU \xrightarrow{\exp} U \to U/V.$$

We shall show that ϕ is an isomorphism.

The exponential isomorphism (3.1) can be used to transport multiplication in U to a composition law in LU, say

$$x * y = \log(\exp x \exp y).$$

This law is given explicitly by the Campbell-Hausdorff formula [9, p. 228]: again, since LU is nilpotent, the series terminate and define polynomial maps

$$x * y = x + y + \tfrac{1}{2}[x, y] + \text{higher commutators.}$$

It will suffice for our purposes to know that if $x \in L_r$, $y \in L_s$ then $((x * y) - x - y) \in L_{r+s}$, and more generally, for any x, z, $(x * (y+z) - (x * z) - y) \in L_{s+1}$.

We now show that ϕ is injective. Suppose in fact that $\phi(x) = \phi(y)$. Then $(\exp x) \cdot V = (\exp y) \cdot V$, so for some $v \in LV$, $\exp x = \exp y \exp v$: thus $x = y * v$. Now suppose inductively $v \in M_r$. As $y \in L_1$, we have $((y * v) - y - v) \in L_{r+1}$. Since $x = y * v$ and y belong to C, we have $v \in C + L_{r+1} = C \oplus M_{r+1}$. As $v \in M$, it follows that $v \in M_{r+1}$. Thus, by induction, $v = 0$.

A similar idea allows us to construct a right inverse to ϕ, and thus conclude the proof. Write π_C, π_M for the projections of $LU = C \oplus LV$ on C and on LV respectively. Given $x \in LU$, write

$$y_1 = \pi_C(x) \qquad v_1 = \pi_M(x)$$

and, by induction on s,

$$y_{s+1} = y_s + \pi_C(x - (y_s * v_s))$$
$$v_{s+1} = v_s + \pi_M(x - (y_s * v_s)).$$

Each of these is a polynomial function of x: we set $\psi(x)=y_l$. Clearly $y_s \in C$, $v_s \in LU$ and, by induction on s,

$$x-(y_s * v_s) \in L_{s+1}.$$

Thus $x=\psi(x) * V_l$, and $\exp x \,.\, V = \exp \psi(x) \,.\, V$.

This gives the required right inverse to ϕ.

As already noted in § 1, the following is crucial to this paper.

(3.3) **Proposition.** *Let U be unipotent affine algebraic group over $\mathbb{F}$ ($=\mathbb{R}$ or $\mathbb{C}$), acting algebraically on an affine variety V. Then the orbits of $U(\mathbb{F})$ are Zariski-closed in $V(\mathbb{F})$.*

Proof. First suppose $\mathbb{F}$ algebraically closed. In this case we reproduce, for the reader's convenience, the proof given in [19, XII, 4.1]. I am indebted to the referee for pointing out that the result is due to Rosenlicht [28]. For the algebraically closed case, we need make no distinction between a variety and the set of its rational points.

First observe that U is irreducible. Indeed, for any $g \in U$, the 1-parameter group $\langle g \rangle$ consisting of the elements $\exp(t \log g)$ ($t \in \mathbb{F}$) is irreducible (it is isomorphic to the additive group of $\mathbb{F}$), 1-dimensional, Zariski-closed (we can see this by putting g in Jordan normal form), and contains all powers of g. As $\mathrm{char}(\mathbb{F})=0$, these are all distinct. Thus the powers g^n are Zariski-dense in $\langle g \rangle$, so $\langle g \rangle \subset U$. Thus any $g \in U$ lies in the identity component.

Let T be a U-orbit in V. Then, by Chevalley's theorem [19, X, 4.3], T contains a subset T_0 which is (Zariski-)open in the Zariski-closure $\bar{T}$ of T (in fact, T is constructible). We may thus suppose $V=\bar{T}$ for the rest of the proof. If T is not closed, $T_0 \neq V$ so there exists a regular rational function f on V which vanishes on $V-T_0$.

Since U acts algebraically on V, the left translates f^u ($u \in U$) of f by elements of U

$$f^u(x)=f(u \,.\, x)$$

generate a finite dimensional subspace M of the ring $\mathcal{O}(V)$. Now M is a U-module; as U is unipotent, there exists a nonzero function $\phi \in M$ which is fixed under the action of U.

As ϕ is fixed under U, it is constant on the U-orbit T; and since T is dense in V, ϕ is constant on V. But as f vanishes on $V-T_0$, hence on $V-T$ (which is U-invariant), so does every translate f^u, and hence every element of M, in particular ϕ. But ϕ is constant and nonzero, hence $V-T$ is empty. Thus $T=V$ is Zariski-closed.

We now consider the case $\mathbb{F}=\mathbb{R}$. First complexify the action and apply the result proved above: we see that the orbits of $U(\mathbb{C})$ are Zariski-closed in $V(\mathbb{C})$: we may thus replace V by such an orbit. By (3.2), $V(\mathbb{C})$ is isomorphic as variety to affine space over $\mathbb{C}$, and $V(\mathbb{R})$ is the fixed point set of complex conjugation which is an involution acting on $V(\mathbb{C})$.

Now by Smith theory (see e.g. [12, III, 4.6]) if a group of order 2 acts on a finite dimensional $\mathbb{Z}_2$-aspherical space (e.g. $V(\mathbb{C})$), the fixed point set (here $V(\mathbb{R})$) is again $\mathbb{Z}_2$-aspherical; in particular, connected in the usual topology.

On the other hand, for each $z \in V(\mathbb{R})$ we have $T_z(U(\mathbb{R}).z) = LU(\mathbb{R}).z$, which is the real subspace of $LU(\mathbb{C}).z = T_z(U(\mathbb{C}).z) = T_z(V(\mathbb{C}))$, and thus equals $T_z(V(\mathbb{R}))$. It follows that each orbit of $U(\mathbb{R})$ is open (in the usual topology) in $V(\mathbb{R})$. As $V(\mathbb{R})$ is connected, it consists of a single orbit.

An alternative proof of the last step, pointed out by the referee, goes as follows. We have to show that $V(\mathbb{F}) = U/U_x(\mathbb{F})$ is a single orbit of $U(\mathbb{F})$, for $\mathbb{F} = \mathbb{R}$. But as U_x is unipotent, for any $\mathbb{F}$ of characteristic zero we have $H^1(\mathrm{Gal}(\bar{\mathbb{F}}/\mathbb{F}), U_x(\bar{\mathbb{F}})) = 0$, and then the assertion follows from [6, 1.12].

We have already noted that (1.4) follows from this result. Further consequences of the same kind are

(3.4) **Corollary.** *Let U be a unipotent algebraic group over $\mathbb{F}(=\mathbb{R}$ or $\mathbb{C})$, acting algebraically on the irreducible affine variety V. If, for some $x \in V$, $LU.x = T_x V$, then $U(\mathbb{F})$ is transitive on $V(\mathbb{F})$.*

For the orbit $U(\mathbb{F}).x$ is Zariski-closed in $V(\mathbb{F})$ by (3.3), but the hypothesis implies that it also contains a subset open in the usual topology, so has the same dimension as $V(\mathbb{F})$. The result follows since V is irreducible.

(3.5) **Corollary.** *Let U be a unipotent algebraic group over $\mathbb{F}$ acting algebraically on the affine space V. Let W be a vector subspace of the group of translations of V, such that there is an induced action of U on the quotient V/W. Then, for any $x \in V$, $x + W$ is contained in an orbit of U if and only if $LU.x$ contains W.*

Proof. Let H be the stabiliser in U of the affine subspace $x + W$: then H acts on this affine space. Clearly H is algebraic and unipotent, and

$$LH = \{l \in LU : l.x \in W\},$$

so $LH.x = W$. The result now follows from (3.4).

The above results are all characteristic of unipotent groups, and we now make this assertion precise by proving explicit converse results. Recall that any affine algebraic group G has a unique maximal normal unipotent subgroup R, called its unipotent radical [3, IV, 11.21]. Over a field of characteristic 0, G is a semidirect product $R \rtimes P$, with P a reductive algebraic group [19, VIII, 4.3]. The following, which includes (1.6), is our deepest result, and the next few pages are devoted to its proof.

(3.6) **Theorem.** *Let G be an affine algebraic group over $\mathbb{F}(=\mathbb{R}$ or $\mathbb{C})$ acting algebraically on an affine space A, with $G(\mathbb{F})$ transitive on $A(\mathbb{F})$. Then if R is the unipotent radical of G, $R(\mathbb{F})$ is already transitive on $A(\mathbb{F})$.*

Proof of (3.6). *Case* $\mathbb{F} = \mathbb{C}$. Here we need not distinguish between a variety and its set of rational points. Choose $x \in A$: write G_x for the isotropy group of $x \in G$, and let K_x be a maximal compact subgroup of G_x (this is a Lie subgroup, not of course algebraic). Then G/G_x (with the usual topology) is homeomorphic to the euclidean space A, and G_x/K_x is homeomorphic to euclidean space [18, XV, 3.1]; hence so is G/K_x, so K_x is a maximal compact subgroup of G.

Since [18, XVII, 5.3] the complex reductive groups are the universal complexifications of their maximal compact subgroups, the complexification $K_x^{\mathbb{C}}$ is

a maximal reductive subgroup of G, hence [19, VIII, 4.3] conjugate to P and complementary to R. Since $K_x^{\mathbb{C}}$, as the Zariski-closure of K_x in G fixes x, it follows that $G_x \cdot R = G$. So the action of R is indeed transitive.

After the first draft of this paper was produced, an alternative proof of this result was given by Borel [5].

The result in the real case lies deeper.

Proof of (3.6). *The case* $\mathbb{F} = \mathbb{R}$. $G(\mathbb{R})$ has an "algebraic" Iwasawa decomposition K.A.N. (i.e. there is an analytic diffeomorphism $K \times A \times N \to G(\mathbb{R})$ given by multiplication) where K is a maximal compact subgroup, A the connected component of 1 (in the standard topology) of $S(\mathbb{R})$, where S is a maximal $\mathbb{R}$-split torus, and $N = U(\mathbb{R})$, where U is a unipotent subgroup of G normalised by S. That this decomposition exists follows from [4, (11.18), (11.19)] (which treats the case G reductive), together with the structure theorem [19, VIII, (4.3)].

Since K is compact, it follows by the argument already given for the case $\mathbb{F} = \mathbb{C}$ that K is contained in the isotropy subgroup of a point; so that $S(\mathbb{R}) \cdot U(\mathbb{R}) = (S \cdot U)(\mathbb{R})$ acts transitively on $V(\mathbb{R})$.

By choosing appropriate $x \in V(\mathbb{R})$, we can suppose that $(S \cdot U)_x = S_x \cdot U_x$; to see this, write, by [19, VIII, (4.3)], $(S \cdot U)_x = P \cdot N$, where P is a maximal reductive subgroup, defined over $\mathbb{R}$, and N is the unipotent radical, of $(S \cdot U)_x$. Since the unipotent radical of a soluble group (which $S \cdot U$, $(S \cdot U)_x$ are) consists of all its unipotent elements ([19, VI, (1.2)]), it follows that $U_x = N$. Now, by [19, VIII, (4.2)], P is conjugate to a subgroup of S by an element of $U(\mathbb{R})$; so, transforming x, we can suppose that $P = S_x$.

Write M for $S \cdot U$. By [19, XII, (2.2)] there is a natural isomorphism of varieties $M/M_x \to M \cdot x = V$; also, since $M_x = S_x \cdot U_x$ is a semi-direct product, we have a further isomorphism of varieties

$$M/M_x \to S/S_x \times U/U_x;$$

both these isomorphisms, hence their composite, are defined over $\mathbb{R}$. Thus

$$S/S_x(\mathbb{R}) \times U/U_x(\mathbb{R}) \cong V(\mathbb{R})$$

is connected (in the usual topology), hence so is $S/S_x(\mathbb{R})$. But S/S_x is a quotient of S, hence is an $\mathbb{R}$-split torus; as $S/S_x(\mathbb{R})$ is connected, it is trivial. Hence $U(\mathbb{R})$ acts transitively on $V(\mathbb{R})$.

For any $x \in V(\mathbb{R})$, $LU(\mathbb{R}) \cdot x = T(V(\mathbb{R}))_x$, so, complexifying, $LU(\mathbb{C}) \cdot x = T(V(\mathbb{C}))_x$. Hence, by (3.4), $U(\mathbb{C})$ acts transitively on $V(\mathbb{C})$, so, a fortiori, $G(\mathbb{C})$ acts transitively on $V(\mathbb{C})$. Hence, by the result in the complex case, $R(\mathbb{C})$ acts transitively on $V(\mathbb{C})$. Thus, taking $x \in V(\mathbb{R})$, $LR(\mathbb{C}) \cdot x = T(V(\mathbb{C}))_x$. Since $R(\mathbb{C})$ is the complexification of $R(\mathbb{R})$, $LR(\mathbb{R}) \cdot x = T(V(\mathbb{R}))_x$. Now, by Mather's lemma, or by (3.4), it follows that $R(\mathbb{R})$ acts transitively on $V(\mathbb{R})$.

Remark. The proof remains valid if the hypothesis that V is affine is weakened to having $V(\mathbb{C})$ and $V(\mathbb{R})$ contractible; the conclusion then implies that V is affine.

Remark. The following stronger result has been conjectured but the above methods will not prove it.

Any algebraic action of a reductive algebraic group on an affine space has a fixed point.

(3.8) **Corollary.** (Lazard; see [11] and [2].) *An affine algebraic group over* $\mathbb{R}$ *or* $\mathbb{C}$ *is unipotent if and only if it is isomorphic (as variety) to affine space.*

Necessity follows from (3.1) and the sufficiency by applying (3.6) to the group acting on itself by translation.

We can now show the 'necessity' of the unipotency hypothesis in (3.5).

(3.9) **Theorem.** *Let G be an algebraic group acting algebraically on an affine space A; let W be a vector subspace of V_A such that there is an induced action of G on the quotient A/W. If $\mathbf{x}+W$ is contained in an orbit of G, then there is a unipotent subgroup U of G such that $LU \,.\, \mathbf{x} = W$.*

Proof. As before, write H for the stabiliser in G of the coset $\mathbf{x}+W$: then H acts transitively on $\mathbf{x}+W$. By (3.7) there is a unipotent subgroup U of H which also acts transitively. Clearly $LU \,.\, \mathbf{x} = W$.

Another useful application of (3.6) is the following.

(3.10) **Theorem.** *Let G be an affine algebraic group defined over* $\mathbb{R}$, *which acts algebraically on an affine space A. Then $G(\mathbb{R})$ acts transitively on $A(\mathbb{R})$ if and only if $G(\mathbb{C})$ acts transitively on $A(\mathbb{C})$.*

Proof. If $G(\mathbb{R})$ acts transitively on $A(\mathbb{R})$, there is a unipotent subgroup $U \subset G$ (also defined over $\mathbb{R}$) such that $U(\mathbb{R})$ also acts transitively. Thus for any $x \in A$, $LU \,.\, x = V_A$. These vector spaces agree over $\mathbb{R}$, hence over $\mathbb{C}$; thus the orbit of $U(\mathbb{C})$ through x is open in $A(\mathbb{C})$. But this orbit is closed (by (3.3)), and hence is the whole of $A(\mathbb{C})$.

Conversely if $G(\mathbb{C})$ acts transitively on $A(\mathbb{C})$ then, for any $x \in A(\mathbb{R})$, we have $LG(\mathbb{C}) \,.\, x = V_{A(\mathbb{C})}$ and hence $LG(\mathbb{R}) \,.\, x = V_{A(\mathbb{R})}$. Thus by Mather's lemma, $G(\mathbb{R})$ acts transitively on $A(\mathbb{R})$.

§ 4. Groups of equivalences and the main theorems

We recall the basic philosophy of this paper. $\mathscr{G}$ is to be a group of equivalences of map-germs; f a finitely $\mathscr{G}$-determined germ: then we can obtain precise results on the range of $\mathscr{G}$-determinacy of f by studying the induced action of $\mathscr{G}$ on r-jets (where f is $r-\mathscr{G}$-determined). This is a finite dimensional problem, where we can apply the results of § 3 to reduce to calculations of tangent spaces.

In order to state precise results, we must now impose conditions on $\mathscr{G}$. In all cases, $\mathscr{G}$ will be a subgroup of Mather's [20] group $\mathscr{K}$ of equivalences acting on map-germs belonging to some affine subspace N of $\mathfrak{m}_n \,.\, C_n^{\times p}$. Denote by $J^r\mathscr{G}$ the group of r-jets of elements of G. We call $\mathscr{G}$ *jet-closed* if for each $r \geqq 1$, $J^r\mathscr{G}$ is a closed subgroup of $J^r\mathscr{K}$ and *jet Z-closed* if each $J^r\mathscr{G}$ is closed in the Zariski topology. Observe that this holds for all the examples commonly

considered: for example Mather's groups $\mathscr{R}$, $\mathscr{A}$, etc. To obtain further examples, we define (cf. § 1) a subgroup $\mathscr{H}$ of $\mathscr{G}$ to be *strongly closed* in $\mathscr{G}$ if $\mathscr{H}_s = \mathscr{G}_s$ for some s, and $J^s\mathscr{H}$ is closed in $J^s\mathscr{G}$. If follows at once that a strongly closed subgroup $\mathscr{H}$ of a jet-closed group $\mathscr{G}$ is itself jet-closed. Correspondingly we have *strongly Z-closed* subgroups and an analogous assertion. By (1.8), if $J^s\mathscr{H}$ is closed and connected and $J^1\mathscr{H}$ is a unipotent algebraic group, so is $J^s\mathscr{H}$; thus $J^s\mathscr{H}$ is also Z-closed. If this holds for each s, we call $\mathscr{H}$ jet-unipotent.

Our basic determinacy result can now be stated as follows.

(4.1) **Proposition.** *Let $\mathscr{G}$ be a jet-unipotent subgroup of $\mathscr{K}$. Let $A \subset \mathfrak{m}_n \cdot C_n^{\times p}$ be a $\mathscr{G}$-invariant affine subspace, M a vector subspace of V_A and $f \in A$ be $s-\mathscr{G}$-determined. Suppose $L = L(J^s\mathscr{G})$ satisfies*

(i) $J^s M \subset L \cdot j^s f$,

(ii) $L \cdot j^s(f+m) \subset L \cdot j^s f$ *for all* $m \in M$;

then f is $M-\mathscr{G}$-determined.

Proof. We apply (1.5) to deduce from (i) and (ii) that $j^s f + J^s M$ is contained in a single orbit of $J^s\mathscr{G}$. Thus if $g \in f+M$, g is $\mathscr{G}$-equivalent to some h with $j^s h = j^s f$ and hence (since f is $s-\mathscr{G}$-determined) to f.

The next formulation is rather more convenient for applications.

(4.2) **Theorem.** *Let $\mathscr{G}$ be a jet-closed subgroup of $\mathscr{K}$. Let $A \subset \mathfrak{m}_n \cdot C_n^{\times p}$ be a $\mathscr{G}$-invariant affine subspace, M a vector subspace of V_A, and $f \in A$ be $s-\mathscr{G}$-determined. Suppose there exists a subspace L of $L(J^s\mathscr{G})$ such that*

(0) $J^1 L$ *acts nilpotently on* $\mathbb{F}^n \oplus \mathbb{F}^p$,

(i) $J^s M \subset L \cdot j^s f$,

and either

(iia) L *is a Lie algebra and* $L \cdot j^s(f+m) \subset L \cdot j^s f$ *for all* $m \in M$ *or*

(iib) $\{l \cdot j^s(f+m) - l \cdot j^s f\} \in J^s M$ *for all* $l \in L$, $m \in M$.

Then f is $M-\mathscr{G}$-determined.

Proof. As in the proof of (4.1) the result will follow from (1.5) if we can construct a closed connected subgroup H of $J^s\mathscr{G}$ such that $J^1 H$ is unipotent and LH satisfies (i) and (ii) of (4.1). In case (a), this is easy: arguing as in (1.8) we see that the action of L is represented by lower unitriangular matrices, so that (by (3.0)) L is already the Lie algebra of a connected closed (and unipotent) subgroup H of $J^s\mathscr{K}$. The result in case (b) is a consequence of the following lemma.

(4.3) **Lemma.** *Let the Lie group G act smoothly on the affine space A. Let B be a vector subspace of V_A. Suppose given a point $x \in A$ and a vector subspace L of LG such that*

(i) $B \subset L \cdot x$, *and*

(ii) *for any* $y \in (x+B)$ *and* $l \in L$, $(l \cdot y - l \cdot x) \in B$.

Then there exists a closed connected Lie subgroup H of G such that

(iii) $B = LH \cdot x$, *and*

(iv) $(x+B)$ *is H-invariant.*

Moreover, given a linear representation ρ of LG on a finite dimensional vector space W such that $\rho(L)$ acts nilpotently, we may suppose that $\rho(LH)$ is nilpotent.

Proof. Let L' be a subspace of L such that

(v) $B = L' . x$.

It follows from (ii) and (v) that

(vi) for any $y \in (x+B)$, $L' . y \subset B$.

Let L'' be the Lie subalgebra of LG generated by L', H'' the corresponding connected Lie subgroup of G, and H the closure of H'' in G. We claim that H has the desired properties.

Indeed, let ϕ be any function vanishing on $x+B$. Then if $X, Y \in L'$, $X\phi$ and $Y\phi$ vanish on $x+B$ by (vi). Hence also $[X, Y]\phi = X(Y\phi) - Y(X\phi)$ vanishes on $x+B$. Thus for any $y \in (x+B)$, $[X, Y] . y \subset B$. It follows that for $y \in (x+B)$, $(L'' . y) \subset B$. Hence H'' leaves $(x+B)$ invariant, and hence so does H. Thus $LH . x \subset B$, and the converse inclusion follows already from (v).

For the final clause in the theorem, it suffices to note that under the hypothesis, there exists a flag in W, stabilised by $\rho(L)$, such that $\rho(L)$ is trivial on the successive quotients: e.g. set $F^i = (\rho(L))^i W$. Then it follows in turn that each of $\rho(L')$, $\rho(L'')$, $\rho(H'')$ and $\rho(H)$ stabilise the flag and act trivially on its quotients: so then does $\rho(LH)$.

Remark. The result in case (b) can also be proved by the direct method of § 2.

By imposing more stringent hypotheses we can also obtain a condition which is necessary as well as sufficient for determinacy.

(4.4) **Theorem.** *Let $\mathscr{G}$ be a jet-Z-closed subgroup of $\mathscr{K}$. Let $A \subset \mathfrak{m}_n . C_n^{\times p}$ be a $\mathscr{G}$-invariant affine subspace, M a vector subspace of V_A such that there is an induced action of $\mathscr{G}$ on A/M.*

Let $f \in A$ be $s-\mathscr{G}$-determined in A $(s < \infty)$. Then f is $M-\mathscr{G}$-determined if and only if there exists a subspace L of $L(J^s\mathscr{G})$ such that $J^1 L$ acts nilpotently on $\mathbb{F}^n \oplus \mathbb{F}^p$ and $J^s M \subset L . j^s f$.

Proof. Sufficiency follows from (4.2): observe that condition (iib) follows from the fact that there is an induced action of $\mathscr{G}$ on A/M, and hence one of $J^s\mathscr{G}$ on $J^s A/J^s M$.

Conversely if f is $M-\mathscr{G}$-determined, $f + J^s M$ is contained in an orbit of $J^s\mathscr{G}$. By (3.9), there is a unipotent subgroup U of $J^s\mathscr{G}$ with $LU . j^s f = J^s M$ (this is the step where the jet-Z-closed hypothesis is needed). It thus suffices to take $L = LU$.

The above results are presented in terms of conditions on the jet groups $J^s\mathscr{G}$: it is clearly desirable to have a formulation directly in terms of a tangent space to $\mathscr{G}$. This approach requires more structure, and we present here only a preliminary version of the axioms that are needed.

A map-germ $\phi: (M^m, x) \to (\mathscr{K}, 1)$ is called C^τ if the induced map

$$\Phi: (M^m, x) \times (\mathbb{F}^{n+p}, 0) \to (\mathbb{F}^{n+p}, 0)$$

is of class C^τ ($\tau = \mathbb{R} - \infty$, $\mathbb{R} - \omega$ or $\mathbb{C} - \omega$). Taking in particular $M = \mathbb{F}$, with coordinate t,

$$z \mapsto \left.\frac{\partial \Phi(t, z)}{\partial t}\right|_{t=0}$$

defines a C^r vector field germ on $\mathbb{F}^{n+p}$. The set of such vector fields is denoted by $L\mathscr{K}$. There are standard identifications of $L\mathscr{R}$, $L\mathscr{L}$, $L\mathscr{C}$ (and hence $L\mathscr{A}$, $L\mathscr{K}$) in terms of coordinates: see [20]. Now if $\mathscr{H}$ is a subgroup of $\mathscr{K}$, we define $L\mathscr{H}$ to be the set of vector fields arising from C^r curves ϕ with image in $\mathscr{H}$.

In the corresponding finite dimensional situation, $L\mathscr{H}$ is the Lie algebra of $\mathscr{H}$. Here it is not clear that $L\mathscr{H}$ is a Lie algebra in general, though this is true for the examples below. However we shall not have occasion to use a Lie bracket in $L\mathscr{H}$ in this paper.

If $\mathscr{H}$ is jet-closed, it follows that

$$J^s(L\mathscr{H}) \subset L(J^s\mathscr{H}).$$

Equality need not hold. It does so if any C^r curve-germ in $J^s\mathscr{H}$ (for any s) can be lifted to a C^r curve-germ in $\mathscr{H}$. When this condition holds, we say that $\mathscr{H}$ is *fibrant*.

It is clear that $\mathscr{R}$, $\mathscr{L}$, $\mathscr{A}$, $\mathscr{C}$ and $\mathscr{K}$ itself are all fibrant in this sense. Moreover if $\mathscr{G}$ is fibrant and $\mathscr{H}$ is strongly closed in $\mathscr{G}$ then $\mathscr{H}$ is fibrant. For suppose $\mathscr{H} \supset \mathscr{G}_s$, and we have a C^r curve in $J^r\mathscr{H}$. If $r<s$, we lift first to a C^r curve in $J^s\mathscr{H}$, using the fact that the epimorphism $J^s\mathscr{H} \to J^r\mathscr{H}$ is a C^r fibration. Thus we may suppose $r \geqq s$. By hypothesis, the given curve in $J^r\mathscr{H} \subset J^r\mathscr{G}$ lifts to a C^r curve in $\mathscr{G}$. But the image of this lift lies in $\mathscr{H}$, since membership of $\mathscr{H}$ is detected by s-jets.

We observe that the notation $L\mathscr{H}.f$ means the same as the $T(\mathscr{H}.f)$ used, for example, in [26]. The old notation was defined ad hoc in each case: the new one thus provides natural generalisations.

Now let $\mathscr{G}$ be a fibrant subgroup of $\mathscr{K}$; let A be an affine subspace of $C_n^{\times p}$ which is G-invariant. We say that $(\mathscr{G}, A)$ satisfies the *Mather condition* if the following conditions on $f \in A$ are equivalent:

(i) f is finitely $\mathscr{G}$-determined in A – i.e. for some $r<\infty$; $g \in A$, $f - g \in \mathfrak{m}_n^r . C_n^{\times p}$ imply f, g $\mathscr{G}$-equivalent.

(ii) $\dim_{\mathbb{F}}(A/L\mathscr{G}.f) < \infty$.

(iii) For some $r<\infty$, $L\mathscr{G}.f \supset V_A \cap \mathfrak{m}_n^r . C_n^{\times p}$.

If $\mathscr{G}$ is any fibrant subgroup of $\mathscr{K}$, (i)$\Rightarrow$(iii)$\Rightarrow$(ii). If $\mathscr{H} \subset \mathscr{G}$ has $\dim(L\mathscr{G}/L\mathscr{H}) < \infty$ (e.g. $\mathscr{H}$ strongly closed in $\mathscr{G}$) and $\mathscr{G}$ is fibrant, then if $(\mathscr{H}, A)$ satisfies the Mather condition, so does $(\mathscr{G}, A)$. The converse is not clear: we say that $(\mathscr{G}, A)$ satisfies the *strong Mather condition* if, for each s, $(\mathscr{G}_s, A)$ satisfies the Mather condition. Thus if $(\mathscr{G}, A)$ satisfies the strong Mather condition and $\mathscr{H}$ is strongly closed in $\mathscr{G}$, $(\mathscr{H}, A)$ also satisfies the (strong) Mather condition. As Mather showed in [20], for $\mathscr{G} = \mathscr{R}$, $\mathscr{L}$, $\mathscr{A}$, $\mathscr{C}$ or $\mathscr{K}$, $(\mathscr{G}, \mathfrak{m}_n . C_n^{\times p})$ satisfies the strong Mather condition.

We can now give the alternative versions of our main theorems.

(4.5) **Theorem.** *Let $\mathscr{G}$ be a fibrant subgroup of $\mathscr{K}$. Let $A \subset \mathfrak{m}_n . C_n^{\times p}$ be a $\mathscr{G}$-invariant affine subspace such that $(\mathscr{G}, A)$ satisfies the Mather condition, M a*

vector subspace of V_A *containing* $\mathfrak{m}_n^{r+1}.C_n^{\times p}\cap V_A$ *for some* $r<\infty$. *Let* $f\in A$. *Suppose there exists a subspace* L *of* $L\mathscr{G}$ *such that*

(0) J^1L *acts nilpotently on* $\mathbb{F}^n\oplus\mathbb{F}^p$,
(i) $M\subset L.f$, *and either*
(iia) L *is a Lie algebra and* $L.(f+m)\subset L.f$ *for all* $m\in M$, *or*
(iib) $\{l.(f+m)-l.f\}\in M$ *for all* $l\in L$, $m\in M$.

Then f *is* $M-\mathscr{G}$*-determined*

Proof. Since $M\subset L.f$, we have à fortiori

$$\mathfrak{m}_n^{r+1}.C_n^{\times p}\cap V_A\subset L\mathscr{G}.f$$

so by the Mather condition f is finitely $\mathscr{G}$-determined, say s-determined. Now since f is fibrant, $L(J^s\mathscr{G})=J^s(L\mathscr{G})$, so J^sL is a subspace of this which satisfies the hypothesis of (4.2). By that result, f is $M-\mathscr{G}$-determined.

(4.6) **Theorem.** *Let* $\mathscr{G}$ *be a jet* Z*-closed fibrant subgroup of* $\mathscr{K}$, $A\subset\mathfrak{m}_n.C_n^{\times p}$ *a* $\mathscr{G}$*-invariant affine subspace such that* $(\mathscr{G}, A)$ *satisfies the strong Mather condition. Let* M *be a vector subspace of* V_A, *containing* $\mathfrak{m}_n^{r+1}.C_n^{\times p}\cap V_A$ *for some* $r<\infty$, *and such that there is an induced action of* $\mathscr{G}$ *on* A/M. *Then* $f\in A$ *is* $M-\mathscr{G}$*-determined if and only if there exists a subspace* $\tilde{L}$ *of* $L\mathscr{G}$ *such that* $J^1\tilde{L}$ *acts nilpotently on* $\mathbb{F}^n\oplus\mathbb{F}^p$, *and* $\tilde{L}.f\supset M$.

Proof. The sufficiency of the condition follows from (4.5) (making the same observation as in the proof of (4.4)).

Conversely if f is $M-\mathscr{G}$-determined, it is $r-\mathscr{G}$-determined in A, so (by the Mather condition) $\dim_{\mathbb{F}}(A/L\mathscr{G}.f)<\infty$. But now $\dim_{\mathbb{F}}(A/L\mathscr{G}_1.f)<\infty$: since $(\mathscr{G}, A)$ satisfies the strong Mather condition,

$$L\mathscr{G}_1.f\supset V_A\cap\mathfrak{m}_n^{s+1}.C_n^{\times p}$$

for some $s<\infty$. We may suppose that $s\geqq r$, and that f is $s-\mathscr{G}$-determined in A. By (4.4) there is a subspace L of $L(J^s\mathscr{G})$ such that J^1L acts nilpotently on $\mathbb{F}^n\oplus\mathbb{F}^p$ and $J^sM\subset L.j^sf$.

Let $\tilde{L}$ be the preimage in $L\mathscr{G}$ of $J^1L\subset J^1L\mathscr{G}$. The nilpotency condition is then satisfied. Since $L\mathscr{G}_1\subset\tilde{L}$,

$$\tilde{L}.f\supset V_A\cap\mathfrak{m}_n^{s+1}.C_n^{\times p}.$$

Now as A is $\mathscr{G}$-invariant, $LG.f\subset V_A$. Since

$$J^s(\tilde{L}.f)\supset L.j^sf\supset J^sM$$

it follows that $\tilde{L}.f\supset M$, thus completing the proof.

Remark. If $\mathscr{H}$ is a subgroup of $\mathscr{K}$, and M_1, M_2 are ranges of $\mathscr{H}$-determinacy for some $f\in m_n.C_n^{\times p}$, it is *not* necessarily the case that M_1+M_2 is a range of $\mathscr{H}$-determinacy for f (in particular, it follows that there will not usually exist any maximum range of $\mathscr{H}$-determinacy, for any given $\mathscr{H}, f$).

However, it is *sometimes* possible to add ranges of determinacy; and since this can be very convenient, we record the possibility here.

(4.6) **Lemma.** *Let $\mathscr{H}_1 \subset \mathscr{H}$ be subgroups of $\mathscr{K}$. Let $A \subset m_n C_n^{\times p}$ be an $\mathscr{H}$-invariant affine subspace let M_1, M be subspaces of V_A, and suppose that there is an induced $\mathscr{H}_1$-action on A/M.*

Suppose that $f \in A$ is $M - \mathscr{H}$-determined and $M_1 - \mathscr{H}_1$-determined.

Then f is $(M_1 + M) - \mathscr{H}$-determined.

Proof. Let $g \in (f + M_1 + M)$; and write $g = f + m_1 + m$, with $m_1 \in M_1$, $m \in M$. Since f is $M_1 - \mathscr{H}_1$-determined, there exists $h_1 \in \mathscr{H}$, with $h_1 . f = f + m_1$. Since f is $M - \mathscr{H}$-determined we have $f + M \subset \mathscr{H} . f$. Applying h_1 yields $h_1(f + M) \subset h_1 . (\mathscr{H} . f) = \mathscr{H} . (h_1 . f)$. Since there is an induced $\mathscr{H}_1$-action on A/M, $h_1 . (f + M) = h_1 . f + M$; so $h_1 . f$ is $M - \mathscr{H}$-determined, and so in particular is $\mathscr{H}$-equivalent to $h_1 . f + m$. So g is $\mathscr{H}$-equivalent to f_1 as required. □

§ 5. Applications

In this chapter we describe some applications of our main results to give new general results. The examples chosen reflect the predilections of the authors but, we feel, illustrate well the scope of our results. For clarity, we do not seek maximum generality here. Calculations for particular examples are reserved for § 6.

Weighted determinacy

For any sequence $\alpha = (\alpha_1, \ldots, \alpha_n)$ of positive integers, we define the monomial $\underline{x}^k = \prod x_i^{k_i}$ to have *weight* $\sum_1^n k_i \alpha_i$ (with respect to α). We filter C_n by writing $F_\alpha^r C_n$ (or $F^r C_n$ for short) for the ideal generated by monomials of weight $\geqq r$ with respect to α. One sees easily that $F_\alpha^r C_n$ has finite codimension: indeed, it contains $\mathfrak{m}_n^k$ for all large enough k.

It is natural, after the work of Arnold [1] to seek criteria for determinacy of functions with respect to the $F^r C_n$: a result of this nature is given in [8]. For map-germs we need further structure.

Given a further sequence $\beta = (\beta_1, \ldots, \beta_p)$ of positive integers, we define a filtration of $C_n^{\times p}$ by letting $\phi = (\phi_1, \ldots, \phi_p)$ belong to $F_{\alpha,\beta}^r M$ (say) if, for each j, $\phi_j \in F^{r+\beta_j} C_n$. Indeed, if each ϕ_j contains only monomials of weight $r + \beta_j$ we say that ϕ is homogeneous of weight r. The following can now be proved by elementary inductive arguments.

(5.1) **Lemma.** (a) *$\phi \in F_{\alpha,\beta}^r M$ if and only if, for all k and for all $g \in F_\beta^k C_p$, $g \circ \phi \in F_\alpha^{r+k} C_n$.*

(b) *If $\phi, \psi \in F_{\alpha,\beta}^0 M$ with $(\phi - \psi) \in F_{\alpha,\beta}^r M$, and $\chi \in F_{\beta,\gamma}^s M$, then $(\chi \circ \phi - \chi \circ \psi) \in F_{\alpha,\gamma}^{r+s} M$.*

(c) *If $\xi \in F_{\alpha,\alpha}^r M$ and $\phi \in F_{\alpha,\beta}^s M$, then the Lie derivative $\xi . \phi \in F_{\alpha,\beta}^{r+s} M$.*

We shall now define, for each of $\mathscr{B} = \mathscr{R}$, $\mathscr{L}$, $\mathscr{A}$, $\mathscr{C}$ or $\mathscr{K}$ (it suffices to consider $\mathscr{R}$, $\mathscr{L}$ and $\mathscr{C}$) a filtration of $\mathscr{B}$ by strongly closed subgroups $F^r \mathscr{B}$ –

which thus satisfy the Mather condition - each acting on $F^0_{\alpha,\beta}M$ respecting the filtration, and inducing actions on the quotients $F^s_{\alpha,\beta}M/F^{r+s}_{\alpha,\beta}M$. It will be trivial in each case that $F^1\mathscr{B}$ is unipotent. Moreover, they satisfy an approximation lemma (we omit the proof - easily deduced from (5.1c))

(*) *If* $\xi\in L(F^s\mathscr{B})$ *and* $\phi,\psi\in F^0_{\alpha,\beta}M$ *with* $\phi-\psi\in F^l_{\alpha,\beta}M$ *then* $\xi.\phi-\xi.\psi\in F^{l+s}_{\alpha,\beta}M$.

Applying (4.3), we deduce in each case

(5.2) **Theorem.** *A map-germ* $\phi\in F^0_{\alpha,\beta}M$ *is* $F^s_{\alpha,\beta}M-F^r\mathscr{B}$*-determined* ($s\geqq 0$, $r\geqq 1$) *if and only if*

$$F^s_{\alpha,\beta}M\subset L(F^r\mathscr{B}).\phi.$$

For $\rho\in\mathscr{R}$, we set $\rho\in F^r\mathscr{R}$ if, for all $f\in C_n$ (it is enough in fact to consider the x_i), $f\in F^s C_n$ implies $(f\circ\rho-f)\in F^{r+s}C_n$. We can encapsulate this definition as

$$F^r\mathscr{R}=(1_n+F^r_{\alpha,\alpha}M)\cap\mathscr{R}.$$

Since $\mathscr{R}$ acts C_n-linearly on $C_n^{\times p}$, it is immediate that $F^0\mathscr{R}$ respects the filtration, and induces actions on the quotients as claimed.

Now $\mathscr{C}$ is the group of diffeomorphisms of $(\mathbb{F}^{n+p},0)$ leaving $\mathbb{F}^n$ fixed and projecting to the identity on $\mathbb{F}^n$: symbolically,

$$\mathscr{C}=1_{n+p}+\mathfrak{m}_p.C^{\times p}_{n+p}.$$

We set

$$F^r\mathscr{C}=(1_{n+p}+F^r_{\alpha\cup\beta,\beta}M)\cap\mathscr{C}:$$

here $\alpha\cup\beta$ denotes the sequence $(\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_p)$. It is clear that if $F^r\mathscr{L}$ is defined analogously to $F^r\mathscr{R}$, then $F^r\mathscr{L}\subset F^r\mathscr{C}$ so it remains only to establish the properties for $\mathscr{C}$. Now if $\phi,\psi\in F^0_{\alpha,\beta}M$ with $(\phi-\psi)\in F^t_{\alpha,\beta}M$, and $\gamma\in F^r\mathscr{C}$ then as the action of γ is defined by $\gamma\circ(1,\phi)=(1,\phi^\gamma)$ we have $(\phi^\gamma-\phi)-(\psi^\gamma-\psi)$ equal to the second component (the first is 0) of

$$(\gamma-1_{n+p})\circ(1,\phi)-(\gamma-1_{n+p})\circ(1,\psi)$$

which belongs to $F^{r+t}_{\alpha,\beta}M$ by (5.1(b)).

A more careful argument gives a much sharper result.

(5.3) **Proposition.** *Let* $f\in\mathfrak{m}_n.C_n^{\times p}$ *be homogeneous of weight zero. Let* $\mathscr{B}$ *be one of* $\mathscr{R}$, $\mathscr{L}$, $\mathscr{A}$, $\mathscr{C}$ *or* $\mathscr{K}$. *Let* N *be a vector space such that for some* $r<\infty$,

$$F^{2r}_{\alpha,\beta}M\subset N\subset F^r_{\alpha,\beta}M.$$

Then f *is* $N-\mathscr{B}$*-determined if and only if*

$$r>0\quad\text{and}\quad N\subset L\mathscr{B}.f.$$

Proof. The condition $r>0$ is necessary since if $r=0$, we have $f\in N$: the other is the standard necessary condition.

For sufficiency we apply (4.4) with $\mathscr{U}=F^r\mathscr{B}$ and with N replacing M. Since $r>0$, $\mathscr{U}$ is unipotent; it is strongly closed in $\mathscr{B}$, hence satisfies the Mather condition. By (5.1*), for any $m\in F^rM$ and $\xi\in L\mathscr{U}$,

$$\xi . (f+m) - \xi . f \in F^{2r} M \subset N.$$

It thus remains only to verify that $N \subset L\mathscr{U} . f$.

Here we use homogeneity of f. If $\xi \in L\mathscr{B}$ is homogeneous of weight s, so is $\xi . f$. Now we can write

$$L\mathscr{B} = X \oplus L\mathscr{U},$$

where X is the sum of homogeneous components of weight $< r$; and hence $X . f$ is also contained in such a sum, whereas $L\mathscr{U} . f \subset F^r M$ (as above). Hence

$$\begin{aligned} L\mathscr{U} . f &= (L\mathscr{B} . f) \cap F^r M \\ &\supset N \cap F^r M = N. \end{aligned}$$

Corollary 1. *The hypothesis $N \subset L\mathscr{B} . f$ above can be replaced by $N \subset T_e \mathscr{B} . f$ (in the notation of* [26]).

For $T_e \mathscr{B} . f$ is generated by $L\mathscr{B} . f$ and the image of f under certain maps of weight $\leqq 0$, so the above argument remains valid.

Corollary 2. *If f is homogeneous of weight zero, and $T_e \mathscr{B} . f \supset F^{2r} M$, then f is $F^r M \cap T_e \mathscr{B} . f - \mathscr{B}$-determined.*

Take $N = F^r M \cap T_e \mathscr{B} . f$ in the proposition, as modified by Corollary 1.

We can also use the idea in the proof of (5.3) to give an easy estimate for standard determinacy of a weighted-homogeneous germ.

(5.3.1) **Proposition.** *Let $f: (\mathbb{F}^n, 0) \to (\mathbb{F}^p, 0)$ be weighted-homogeneous. If $\mathfrak{m}_n^{r+1} . C_n^{\times p} \subset T_e \mathscr{B} . f \cap F^1 M + L\mathscr{B}_1 . f$, then f is $r - \mathscr{B}$-determined.*

Proof. Arguing as in (5.3), we have

$$T_e \mathscr{B} . f \cap F^1 M = L(F^1 \mathscr{B}) . f,$$

so that

$$\mathfrak{m}_n^{r+1} . C_n^{\times p} \subset L(F^1 \mathscr{B} . \mathscr{B}_1) . f.$$

Since $J^1(F^1 \mathscr{B} . \mathscr{B}_1) = J^1(F^1 \mathscr{B})$, which is nilpotent, it follows from (1.9) that f is $r - \mathscr{B}$-determined. □

(5.3.2) *Remark.* Dropping the $L\mathscr{B}_1 . f$-term in (5.3.1) gives an inclusion which is easier to check; but which gives weaker results, for there are (many) weighted-homogeneous map-germs which are $r - \mathscr{B}$-determined whilst $\mathfrak{m}_n^{r+1} . C_n^{\times p} \not\subset F^1 M$ - see (6.1)-(6.3) for examples. In particular, it follows that the range of $\mathscr{B}$-determinacy provided by (5.3.1) is not necessarily contained in that implied by (5.3); note, however, that the sum of these ranges is a range of $\mathscr{B}$-determinacy, by (4.6).

The equivariant case

Suppose given a compact Lie group G acting (we may suppose) linearly on $\mathbb{R}^n$ and $\mathbb{R}^p$. Then there is an induced action of G on $\mathscr{K}$ such that the invariant subgroup $\mathscr{K}^G$ acts on the equivariant map-germs $(\mathfrak{m}_n . C_n^{\times p})^G$.

If $\mathscr{B}$ is a subgroup of $\mathscr{K}$, A a $\mathscr{B}$-invariant affine subspace of $\mathfrak{m}_n . C_n^{\times p}$, then the G-fixed part $\mathscr{B}^G = \mathscr{B} \cap \mathscr{K}^G$ acts on A^G, so we can try applying our main theorems. Unfortunately none of our main conditions on $\mathscr{B}$ (jet-closed, fibrant, Mather condition) appears to be automatically inherited by $\mathscr{B}^G$: indeed, even the equalities

$$(J^k \mathscr{B})^G = J^k(\mathscr{B}^G), \qquad (L\mathscr{B})^G = L(\mathscr{B}^G)$$

are not guaranteed in general.

However, these properties are verified for the examples that tend to arise in practice. In particular,

(5.4) **Lemma.** *If $\mathscr{B} = \mathscr{R}$, $\mathscr{L}$, $\mathscr{A}$, $\mathscr{C}$ or $\mathscr{K}$, then $\mathscr{B}^G$ is jet-closed and fibrant, and its action on $(\mathfrak{m}_n . C_n^{\times p})^G$ satisfies the strong Mather condition.*

For since G acts linearly, a k-jet belongs to $J^k \mathscr{B}^G$ if and only if its preferred polynomial representative is equivariant. This implies the jet-closed and fibrant properties: the Mather condition holds by appropriate modifications of Mather's arguments: see [24] or [10].

Our results thus apply to these cases; and to some further ones in view of the remark that if $\mathscr{H}$ is strongly closed in $\mathscr{B}$, then $\mathscr{H}^G$ is strongly closed in $\mathscr{B}^G$ for $\mathscr{B}$ as above. In particular, we obtain natural analogues of the results of § 2.

These results however are not the most useful for applications. Instead of $(\mathfrak{m}_n^k . C_n^{\times p})^G$-determinacy, it is more convenient to work with $(\mathfrak{m}_n^G)^k . (C_n^{\times p})^G$-determinacy. Clearly

$$(\mathfrak{m}_n^G)^k . (C_n^{\times p})^G \subset (\mathfrak{m}_n^k . C_n^{\times p})^G,$$

and it is shown in [23] for each k how to find an $l (= Ak + B$ for suitable $A, B)$ such that

$$(\mathfrak{m}_n^l . C_n^{\times p})^G \subset (\mathfrak{m}_n^G)^k . (C_n^{\times p})^G.$$

We are also interested in modifications of the above groups of equivalences: these were introduced partly in [24] and partly in [27, Appendix]. Set

$$\mathscr{R}_s^{*G} = \mathscr{R} \cap (1_n + (\mathfrak{m}_n^G)^{s+1} (C_n^{\times n})^G)$$

and similarly for $\mathscr{L}$, and

$$\mathscr{C}_s^{*G} = 1_{n+p} + (\mathfrak{m}_n^G)^s . (C_{n+p}^{\times p})^G$$

with

$$\mathscr{A}_s^{*G} = \mathscr{R}_s^{*G} \times \mathscr{L}_s^{*G}, \qquad \mathscr{K}_s^{*G} = \mathscr{R}_s^{*G} . \mathscr{C}_s^{*G}.$$

For $\mathscr{B} = \mathscr{R}$, $\mathscr{L}$ or $\mathscr{A}$, $\mathscr{B}_s^{*G}$ is strongly closed in $\mathscr{B}^G$, so is fibrant and satisfies the Mather condition. For $\mathscr{B} = \mathscr{C}$ or $\mathscr{K}$, $\mathscr{B}_s^{*G}$ is not strongly closed in $\mathscr{B}^G$. However the fibrant property is easy to verify, and the Mather condition follows by standard arguments.

We deduce in particular

(5.5) **Proposition.** *A G-equivariant map-germ is $(\mathfrak{m}_n^G)^{r+1} (C_n^{\times p})^G - \mathscr{B}_s^{*G}$-determined $(r, s \geqq 0$; $\mathscr{B} = \mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ or $\mathscr{A})$ if and only if*

$$L\mathscr{B}_s^{*G}.f+(\mathfrak{m}_n^G)^{\varepsilon r+2}.(C_n^{\times p})^G=(\mathfrak{m}_n^G)^{r+1}.(C_n^{\times p})^G$$

(*where* $\varepsilon=1$ for $\mathscr{B}=\mathscr{R}$, $\mathscr{C}$ or $\mathscr{K}$; $\varepsilon=2$ for $\mathscr{B}=\mathscr{L}$ or $\mathscr{A}$).

For we can delete the second term on the left hand side by Nakayama's lemma (if $\varepsilon=1$) or by [26, Lemma 1.6] (if $\varepsilon=2$), and the conclusion then follows from Theorem (4.3).

A less routine application is the following.

(5.6) **Proposition.** *Any* $\mathscr{R}^G$-stable *equivariant map-germ is* $(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G).(C_n^{\times p})^G$ $-\mathscr{R}^G$-*determined.*

Proof. Let $\mathscr{U}$ be the unipotent group with Lie algebra $(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G).(C_n^{\times n})^G$: this is somewhat smaller than $\mathscr{R}^G\cap\mathscr{R}_1$, and can be defined as

$$\mathscr{U}=1_n+(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G).(C_n^{\times n})^G.$$

Then $\mathscr{U}$ is strongly closed in $\mathscr{R}^G$. We can thus apply Theorem (4.4): it remains to verify (i) and (ii). Now as f is $\mathscr{R}^G$-stable,

$$(C_n^{\times n})^G.f=T_e\mathscr{R}^G.f=(C_n^{\times p})^G.$$

Multiplying by $\mathfrak{m}_n^2\cap\mathfrak{m}_n^G$, we deduce

$$L\mathscr{U}.f=(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G).(C_n^{\times p})^G.$$

Since $L\mathscr{U}$ is generated by products $\phi\xi$ with $\phi\in(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G)$, and for any g

$$\phi\xi.(f+g)-\phi\xi.f=\phi\xi.g=\phi(\xi.g)$$

belongs to $(\mathfrak{m}_n^2\cap\mathfrak{m}_n^G)$, (ii) holds also.

Example. G of order 2, generated by T

$\mathbb{R}^n$ with coordinates x_i (fixed by T), y_j (changed in sign by T),

$\mathbb{R}^p$ with coordinates X_i (fixed by T), Y_j (changed in sign by T).

We can write f as

$$X_i\circ f=\sum_j a_{ij}x_j+\text{higher terms}$$

$$Y_j\circ f=\sum_k b_{jk}y_k+\sum_{k,l}c_{jkl}x_l y_k+\text{higher terms.}$$

Then if f is $\mathscr{R}^G$-stable, it is $\mathscr{R}^G$-determined by the terms exhibited: the "higher terms" may be omitted.

A general range for $\mathscr{A}$-determinacy

(5.7) **Proposition.** *Let*

$$I.\theta_f\subset M\subset L\mathscr{A}.f$$

where $I \subset C_n$ is a (proper) ideal of finite codimension, and M is a linear subspace of θ_f such that $(I+f^*\mathfrak{m}_p).M$ is a C_n-module.

Then f is $I.M-\mathscr{A}_1$-determined.

Proof. Multiplying the inclusion $M \subset L\mathscr{A}.f$ by $I+f^*\mathfrak{m}_p$ gives

$$\begin{aligned}(I+f^*\mathfrak{m}_p).M &\subset tf((I+f^*\mathfrak{m}_p).\mathfrak{m}_n C_n^{\times n})+f^*\mathfrak{m}_p.I.C_n^{\times p}+\omega f(\mathfrak{m}_p^2 C_p^{\times p})\\ &= tf((I+f^*\mathfrak{m}_p).\mathfrak{m}_n C_n^{\times n})+\omega f(\mathfrak{m}_p^2.C_p^{\times p}) \qquad (*)\end{aligned}$$

(since $I.\theta_f \subset L\mathscr{A}.f$).

Now define

$$L=(I+f^*\mathfrak{m}_p).\mathfrak{m}_n.C_n^{\times n}\oplus\mathfrak{m}_p^2\theta_p;$$

this is a Lie subalgebra of $L\mathscr{A}_1$.

If g is an $I.M$-approximation to f, then, by [17, (1.14)],

$$I+f^*\mathfrak{m}_p.C_n=I+g^*\mathfrak{m}_p.C_n,$$

and it is easy to see that, for any $l\in L$,

$$\begin{aligned}(l.f-l.g)&\in\mathfrak{m}_n(I+f^*\mathfrak{m}_p).M;\\ &\subset L.f, \quad \text{by (*)}.\end{aligned}$$

So $L.g \subset L.f$. It now follows at once from (4.5) ((i) and (ii)(a)) that f is indeed $I.M-\mathscr{A}_1$-determined.

(5.7.1) **Corollary.** *If $I.\theta_f \subset L\mathscr{A}.f$, where $I\subset C_n$ is an ideal of finite codimension, then f is $I^2.\theta_f-\mathscr{A}_1$-detrmined.*

Proof. Take $M=I.\theta_f$.

Remark. This result can be extended to cases where I is not of finite codimension.

Real and complex determinacy

The following result appears to be new, even for $\mathscr{H}=\mathscr{G}$.

(5.8) **Theorem.** *Let $\mathscr{H}$ be a strongly Z-closed subgroup of $\mathscr{G}$ ($\mathscr{G}=\mathscr{R}$, $\mathscr{C}$, $\mathscr{K}$, $\mathscr{L}$ or $\mathscr{A}$); let $f:(\mathbb{R}^n,0)\to(\mathbb{R}^p,0)$ be an analytic map-germ. Then f is $r-\mathscr{H}$-determined if and only if $f_{\mathbb{C}}$ is $r-\mathscr{H}_{\mathbb{C}}$-determined.*

Here we let $\mathscr{G}_s\subset\mathscr{H}$ and define $\mathscr{H}_{\mathbb{C}}$ to be the unique subgroup of $\mathscr{G}_{\mathbb{C}}$ containing $(\mathscr{G}_{\mathbb{C}})_s$ and with s-jet group $(J^s\mathscr{H})_{\mathbb{C}}$.

Proof. f is finitely $\mathscr{H}$-determined

$\Leftrightarrow$ f is finitely $\mathscr{G}$-determined

$\Leftrightarrow$ $\mathfrak{m}_n^r.C_n^{\times p}\subset L\mathscr{G}.f+\mathfrak{m}_n^{2r}.C_n^{\times p}$ for some $r<\infty$

$\Leftrightarrow$ $\mathfrak{m}_n^r.C_n(\mathbb{C})^{\times p}\subset L\mathscr{G}_{\mathbb{C}}.f+\mathfrak{m}_n^{2r}.C_n(\mathbb{C})^{\times p}$ for some $r<\infty$

$\Leftrightarrow$ $f_{\mathbb{C}}$ is finitely $\mathscr{G}_{\mathbb{C}}$-determined

$\Leftrightarrow$ $f_{\mathbb{C}}$ is finitely $\mathscr{H}_{\mathbb{C}}$-determined.

Let N be an integer such that f is $N-\mathscr{H}$-determined and $f_{\mathbb{C}}$ is $N-\mathscr{H}_{\mathbb{C}}$-determined. Write H for the subgroup of $J^N\mathscr{H}$ fixing $j^r f$: this is an algebraic subgroup, defined over $\mathbb{R}$. Now f is $r-\mathscr{H}$-determined

$\Leftrightarrow$ the action of $H(\mathbb{R})$ on $J^N(f+\mathfrak{m}_n^{r+1}.C_n^{\times p})$ is transitive

$\Leftrightarrow$ the action of $H(\mathbb{C})$ on $J^N(f_{\mathbb{C}}+\mathfrak{m}_n^{r+1}.C_n(\mathbb{C})^{\times p}$ is transitive (by 3.10)

$\Leftrightarrow f_{\mathbb{C}}$ is $r-H_{\mathbb{C}}$-determined.

§ 6. Examples

In this final section we return to determinacy with respect to $\mathscr{A}$-equivalence and give examples to illustrate how the use of our theorems vastly simplifies the calculations needed to obtain these results.

To describe vector subspaces of $C_n^{\times p}$, some notation is needed. If $\{u_1, \ldots, u_r\}$ is a finite list of monomials in C_n, of degrees $\leqq k$, we write (as in [22, p. 131]) $C_n \div \{u_1, \ldots, u_r\}$ to denote the vector subspace of C_n spanned by $\mathfrak{m}_n^{k+1}$ together with all monomials of degrees $\leqq k$ *other* than $u_1, \ldots, u_r$.

(6.1) *Example.* Let $n=2$, $p=3$ and consider the map-germs $f_{i,r}: (\mathbb{F}^2, 0) \to (\mathbb{F}^3, 0)$ given (for $i \geqq 1$, $r \geqq 2$) by

$$f_{i,r}(x, y) = (x, y^2, y^{2i+1} + x^r y).$$

The map $f_{i,r}$ is homogeneous of weight zero with respect to the weights $\alpha = (2i, r)$, $\beta = (2i, 2r, (2i+1)r)$. We find that

$$\begin{aligned} T_e\mathscr{A}.f_{i,r} &= (C_n, C_n, C_n \div \{x^k y^{2s+1} : 0 \leqq k \leqq r-2,\ 0 \leqq s \leqq i-1\}) \\ &\supset F^{2(i-1)(r-2)-3} M \end{aligned}$$

so that by (5.3) (with its corollaries), $f_{i,r}$ is $N-\mathscr{A}$-determined, where

$$\begin{aligned} N &= F^1 M \quad \text{if } (i-1)(r-2) \leqq 2, \\ N &= F^{(i-1)(r-2)-1} M \cap T_e\mathscr{A}.f_{i,r} \quad \text{otherwise.} \end{aligned}$$

Also we see that

$$T_e\mathscr{A}.f_{i,r} \supset \mathfrak{m}_n^{2i+r-2}.C_n^{\times p}.$$

Setting $t = \max\{r+1, 2i+1, 2i+r-3\}$, we see easily that

$$y.\mathfrak{m}_n^t.C_n^{\times p} \subset F^1 M$$

and so

$$\begin{aligned} \mathfrak{m}_n^{t+1}.C_n^{\times p} &\subset T_e\mathscr{A}.f_{i,r} \cap F^1 M + (f^* \mathfrak{m}_p^{t+1})^{\times p} \\ &\subset T_e\mathscr{A}.f_{i,r} \cap F^1 M + L\mathscr{A}_1.f_{i,r} \end{aligned}$$

whence, by (5.3.1), $f_{i,r}$ is $t-\mathscr{A}$-determined. This recaptures the results of [22, 3.12, 3.16, 3.20, 3.29].

Note, in connection with remark (5.3.2), that $(0, 0, x^{2i+r-2})$ has weight $4i^2 - 4i - r$ and so does not belong to $F^1 M$ if $r \geqq 4i^2 - 4i$. We deduce that if $i \geqq 2$, $r \geqq 4i^2 - 4i$, then $\mathfrak{m}^{t+1}.C_n^{\times p} \not\subset F^1 M$.

(6.2) *Example.* Let $n=3$, $p=3$ and consider

$$f_{r,s}(x, y, z)=(x, y, z^3 \pm z(x^r \pm y^s)) \qquad (r, s \geqq 2)).$$

This is homogeneous of weight zero with respect to the weights $\alpha=(2s, 2r, rs)$, $\beta=(2s, 2r, 3rs)$. We have

$$\begin{aligned} T_e\mathscr{A} \cdot f_{r,s} &= (C_n, C_n, C_n \div \{x^i y^j z : 0 \leqq i \leqq r-2, 0 \leqq j \leqq s-2\}) \\ &\supset F^{2(r-2)(s-2)-7} M. \end{aligned}$$

Again by (5.3) and corollaries, $f_{r,s}$ is

$$F^1 M - \mathscr{A}\text{-determined if } (r-2)(s-2) \leqq 4,$$
$$F^{(r-2)(s-2)-3} M \cap T_e\mathscr{A} \cdot f_{r,s} - \mathscr{A}\text{-determined otherwise.}$$

Also, we see that

$$T_e\mathscr{A} \cdot f_{r,s} \supset \mathfrak{m}_n^{r+s-3} \cdot C_n^{\times p}.$$

Setting $t=\max\{r+1, s+1, r+s-3\}$, we see easily that

$$z \cdot \mathfrak{m}_n^t \cdot C_n^{\times p} \subset F^1 M$$

and so

$$\mathfrak{m}_n^{t+1} \cdot C_n^{\times p} \subset T_e\mathscr{A} \cdot f_{r,s} \cap F^1 M + L\mathscr{A}_1 \cdot f_{r,s}$$

whence, by (5.3.1), $f_{r,s}$ is $t-\mathscr{A}$-determined. This recovers the results of [22, 3.13, 3.17, 3.21].

In this case $(0, 0, y^{r+s-2})$ has weight $r(2r-4-s)$ so that it does not belong to $F^1 M$ if $s \geqq 2r-4$. Thus if $r \geqq 4$, $s \geqq 2r-4$, $\mathfrak{m}_n^{t+1} \cdot C_n^{\times p} \not\subset F^1 M$ (cf. (5.3.2)).

(6.3) *Example.* Let $n=p=2$, and let $f=f_{k,l}: (\mathbb{F}^2, 0) \to (\mathbb{F}^2, 0)$ be given (for $k \geqq 3$, $l \geqq 1$) by

$$f(x, y)=(x, y^k \pm x^l y).$$

This is weighted homogeneous of weight zero with respect to

$$\alpha=(k-1, l), \qquad \beta=(k-1, kl).$$

We claim that

$$\begin{aligned} T_e\mathscr{A} \cdot f = (&C_2 \div \{x^{al+i} y^{b(k-1)+j} : 0 \leqq a, b; 0 \leqq i < l, 0 \leqq j < k-2, a+b<j\}, \\ &C_2 \div \{x^{al+i} y^{b(k-1)+j} : 0 \leqq a, b; 0 \leqq i < l, 0 \leqq j < k-1, a+b<j-1 \\ &\qquad \text{or } a+b=j-1, i<l-1\}) + tf(C_2^{\times 2}), \end{aligned}$$

and that $T_e\mathscr{A} \cdot f$ does not contain any of the monomials in braces. To see this, we first note that we can reduce to considering the second coordinate; then arrange the monomials in series

$$S_{i,j,c} = \{(0, x^{al+i} y^{b(k-1)+j}) : 0 \leqq a, b; a+b=c\}$$

for $0 \leqq i < l$, $0 \leqq j < k-1$, $0 \leqq c$. Each element of $S_{i,j,c}$ has weight $c(k-1)l+(k-1)i+lj-kl$. It is then easy to see that $S_{i,j,c} \subset T_e\mathscr{A} \cdot f$ if and only if either $c \geqq j$ or $c=j-1$ and $i=l-1$.

Thus the greatest weight amongst the monomials *not* in $T_e\mathscr{A}.f$ accurs for $S_{l-2,k-2,k-3}$, and

$$T_e\mathscr{A}.f \supset F^{(k^2-3k)l-2k+3}M.$$

Now (5.3) shows that $f_{k,l}$ is

$$F^1M - \mathscr{A}\text{-determined if } k=3 \text{ or } k=4,\ l=1$$
$$F^{\frac{1}{2}(k^2-3k)l-k+2}M \cap (T_e\mathscr{A}.f) - \mathscr{A}\text{-determined otherwise.}$$

We can also compute the order of determinacy of f. If $k=3$, we note that

$$y.\mathfrak{m}_n^{l+1}.C_n^{\times p} \subset F^1M$$

so that

$$\begin{aligned}\mathfrak{m}_n^{l+2}.C_n^{\times p} &\subset F^1M + (f^*\mathfrak{m}_p^{l+2})^{\times p}\\ &\subset T_e\mathscr{A}.f \cap F^1M + L\mathscr{A}_1.f\end{aligned}$$

and so, by (5.3.1), $f=f_{3,l}$ is $(l+1)-\mathscr{A}$-determined. This is the exact order of $\mathscr{A}$-determinacy, for $(x,y)\mapsto(x,y^3)$ is not finitely $\mathscr{A}$-determined. (In connection with Remark (5.3.2), note that $\mathfrak{m}_n^{l+2}.C_n^{\times p} \not\subset F^1M$ if $l\geqq 4$, for $(0,x^{l+2})$ has weight $4-l$).

If $k=4$ and $l=1$, we have

$$\begin{aligned}\mathfrak{m}_n^5.C_n^{\times p} &\subset F^1M;\\ &\subset T_e\mathscr{A}.f_{k,1} \cap F^1M\end{aligned}$$

and so, by (5.3.1), f if $4-\mathscr{A}$-determined.

This is the exact order of $\mathscr{A}$-determinacy, for $(x,y)\mapsto(x,y^4)$ is not finitely $\mathscr{A}$-determined.

In the remaining cases it follows from our description of $T_e\mathscr{A}.f_{k,l}$ that the minimum value t_0 of s such that

$$\mathfrak{m}_n^{s+1}.C_n^{\times p} \subset T_e\mathscr{A}.f_{k,l}$$

is given by

$$t_0 = \max\left\{ al+b(k-1)+i+j \;\middle|\; \begin{array}{l} either\ 0\leqq j<k-1,\ 0\leqq i<l,\ a+b<j-1\\ or\quad 0\leqq j<k-1,\ 0\leqq i<l-1,\ a+b=j-1\end{array}\right\}$$

$$= \begin{cases}(k-2)l+k-4 & \text{if } l\geqq k-1\\ l+k^2-3k-1 & \text{if } 2\leqq l\leqq k-1\\ k^2-4k+2 & \text{if } l=1\end{cases}$$

It is easy to check that (for $k\geqq 5$, or $k=4$ and $l\geqq 2$)

$$\mathfrak{m}_n^{t_0+1}.C_n^{\times p} \subset F^1M,$$

so we deduce from (5.3.1) that $f_{k,l}$ is $t_0-\mathscr{A}$-determined, and also see that this is the exact order of $\mathscr{A}$-determinacy. (This recaptures the results of [22, 3.11 ($k=3$), 3.30, 3.32 ($k=4$) and 3.33 ($k=5$)].)

Remark that for $k\geqq 5$ or $k=4$, $l\geqq 2$,

$$\mathfrak{m}_n^{t_0+1}.C_n^{\times p} \subset F^{\frac{1}{2}(k^2-3k)l-k+2}M \cap T_e\mathscr{A}.f;$$

and that the range of $\mathscr{A}$-determinacy on the RHS is very much larger than that on the LHS. For example, $f_{5,1}$ is $(\langle x^2, xy^3, y^7\rangle . C_n, \langle x^2, xy^4, y^8\rangle . C_n)$ $-\mathscr{A}$-determined, not just $7-\mathscr{A}$-determined.

For further examples, it is convenient to have

(6.4) **Lemma.** *Let* $\eta\in\mathfrak{m}_p^{r+1} . C_{n+p}^{\times p}, f\in\mathfrak{m}_n . C_n^{\times p}$, I *be an ideal in* C_n *and* $m\in I . C_n^{\times p}$. *Then*

$$\eta\circ(1, f+m)-\eta\circ(1, f)\subset I . (f^*\mathfrak{m}_p+I)^r . C_n^{\times p}.$$

Proof. We argue by induction. The case $r=0$ follows from [17, 1.14]. Assume inductively that the result holds for $(r-1)$, and write

$$\eta=\sum_1^p y_i\eta_i$$

where the y_i are coordinates in $\mathbb{F}^p$ and $\eta_i\in\mathfrak{m}_p^r . C_{n+p}^{\times p}$. Then

$$\begin{aligned}\eta\circ(1, f+m)-\eta\circ(1, f)&=\sum_1^p (f_i+m_i)\,\eta_i\circ(1, f+m)-\sum_1^p f_i\eta_i\circ(1, f)\\&=\sum_1^p (f_i+m_i)\{\eta_i\circ(1, f+m)-\eta_i\circ(1, f)\}+\sum_1^p m_i\eta_i\circ(1, f).\end{aligned}$$

Here the term in braces belongs to $I . (f^*\mathfrak{m}_p+I)^{r-1} . C_n^{\times p}$ by the inductive hypothesis, and this is multiplied by (f_i+m_i), which belongs to $f^*\mathfrak{m}_p+I$. As to the remaining terms, we observe that $m_i\in I$ and $\eta_i\in\mathfrak{m}_p^r . C_{n+p}^{\times p}$ so that $(\eta_i\circ(1, f))\in(f^*\mathfrak{m}_p)^r . C_n^{\times p}$. The result follows.

(6.5) *Example.* Take $n=p=2$ and define f by

$$f(x, y)=(x, y^4+xy^2+x^2y)$$

where we use (x, y) for coordinates in the source and (X, Y) for those in the target. Define $L\subset L\mathscr{A}_1$ by

$$L=\left((\mathbb{R} . x^2+\mathfrak{m}_n^3)\frac{\partial}{\partial x}+\mathfrak{m}_n^2\frac{\partial}{\partial y}, \mathfrak{m}_p^2\frac{\partial}{\partial X}+\mathfrak{m}_p^2\frac{\partial}{\partial Y}\right).$$

Then $L . f=M$, where

$$M=(\{x^2, xy^2, y^4\} . C_n, \{x^3, x^2y^2, xy^3+2y^5, xy^4\} . C_n),$$

and an easy calculation using (6.4) shows that

$$(l . (f+m)-l . f)\in M \quad \text{for all } l\in L, m\in M.$$

Thus by (4.5), f is $M-\mathscr{A}$-determined. Indeed, L is the Lie algebra of a subgroup $\mathscr{A}_L$ (with $\mathscr{A}_2\subset\mathscr{A}_L\subset\mathscr{A}_1$) and it follows that f is $M-\mathscr{A}_L$-determined. This result is obtained in [17, 3.5], but only after a great deal more calculation.

Remark. It is claimed in [22, p. 153, Remark 4] that f is $L\mathscr{A}_1 . f-\mathscr{A}$-determined. However, this is not true. For we have

$$tf\left(y^2\frac{\partial}{\partial x}\right)=(y^2, y^4+2xy^3),$$

so that an $L\mathscr{A}_1 . f$-approximation to f is given by

$$\begin{aligned} g_a(x, y)&=f(x, y)+a(y^2, y^4+2xy^3)\\ &=(x+ay^2, (1+a)y^4+xy^2+x^2y+2axy^3). \end{aligned}$$

Substituting x for $x+ay^2$ shows this $\mathscr{R}_1$-equivalent to

$$g'_a(x, y)=(x, x^2y+xy^2+y^4-a^2y^5).$$

By [22, p. 153, Remark 2],

$$\mathfrak{m}_n^4 . C_n^{\times p}\not\subset L\mathscr{A} . g'_a+\mathfrak{m}_n^6 . C_n^{\times p}\Leftrightarrow a=\pm 2.$$

(The proof is a direct calculation.)

Since this inclusion is an $\mathscr{A}$-invariant property of g'_a, $f=g_0$ cannot be $\mathscr{A}$-equivalent to g_2.

(6.6) *Example.* Let $n=2$, $p=3$ and $f: (\mathbb{F}^2, 0)\to(\mathbb{F}^3, 0)$ be given by

$$f(x, y)=(x, y^3+x^2y, y^4+xy^2).$$

Here we define $M\subset\mathfrak{m}_n . C_n^{\times p}$, $L\subset L\mathscr{A}$ by

$$M=(\mathfrak{m}_n^3, \{x^3, xy^2, y^4\} . C_n, \quad \{x^3, x^2y^2, xy^3, y^5\} . C_n).$$

$$L=\left(\mathfrak{m}_n^2 . C_n^{\times n}, (\mathbb{R}Y+\mathbb{R}Z)\frac{\partial}{\partial X}+\mathbb{R}Z\frac{\partial}{\partial Y}+\mathfrak{m}_p^2 . C_p^{\times p}\right).$$

Now J^1L is upper triangular, hence nilpotent, and we can verify easily that $M\subset L . f$. Moreover, using (6.4), we have

$$l . (f+m)-l . f\in M \quad \text{for all } l\in L, m\in M.$$

Thus by (4.5), f is $M-\mathscr{A}$-determined. Indeed, L is the Lie algebra of a jet-unipotent subgroup $\mathscr{A}_L$ (with $\mathscr{A}_1\subset\mathscr{A}_L\subset\mathscr{A}$), and it follows that f is $M-\mathscr{A}_L$-determined. This result is obtained in [17, 3.6], but with much more effort.

(6.7) *Example.* Consider the map-germ $f: (\mathbb{F}^2, 0)\to(\mathbb{F}^2, 0)$ given by

$$f(x, y)=(x^2+y^{2r+1}, y^2+x^{2s+1}).$$

Define

$$M=(\{x^3, x^2y, xy^{r+1}, y^{2r+2}\} . C_n, \{y^3, y^2x, yx^{s+1}, x^{2s+2}\} . C_n\subset m_n . C_n^{\times p});$$

$$L=\left(C_n\{x^2, xy, y^{r+1}\}\cdot\frac{\partial}{\partial x}+C_n\cdot\{y^2, xy, x^{s+1}\}\cdot\frac{\partial}{\partial y},\right.$$

$$\left.\mathbb{R}\cdot\{X^k, Y^l: k\geqq 2, l\geqq r+1\}\cdot\frac{\partial}{\partial X}+\mathbb{R}\cdot\{X^m, Y^n; m\geqq s+1, n\geqq 2\}\cdot\frac{\partial}{\partial Y}\right)\subset L\mathscr{A}_1.$$

It is not hard to see that

$$M = L.f$$

and, moreover,

$$l.(f+m)-l.f \in M \quad \text{for all } l \in L,\ m \in M.$$

It follows from (4.5) that f is $M-\mathscr{A}_1$-determined.

It is perhaps interesting (in connection with the discussion on p. 14) to notice that the above implies at once that the order of $\mathscr{A}$-determinacy $-2\max\{r,s\}+1-$calculated for f in [22], (3.18) is in fact the order of $\mathscr{A}_1$-determinacy also.

References

1. Arnold, V.I.: Normal forms of functions in neighbourhoods of degenerate critical points. Russ. Math. Surv. **29**, 11-49 (1974)
2. Bialynicki-Birula, A.: On fixed point schemes of actions of multiplicative and additive groups. Topology **12**, 99-103 (1973)
3. Borel, A.: Linear algebraic groups. New York: Benjamin 1969
4. Borel, A.: Introduction aux groupes arithmétiques. Paris: Hermann 1969
5. Borel, A.: On affine algebraic homogeneous spaces. Arch. Math. **45**, 74-78 (1985)
6. Borel, A., Serre, J-P.: Théorèmes de finitude en cohomologie galoisienne. Comment. Math. Helv. **39**, 111-164 (1964)
7. Borel, A., Springer, T.A.: Rationality properties of linear algebraic groups. In: Algebraic groups and discontinuous subgroups. Proc. Symp. Pure Math. Amer. Math. Soc. **9**, 26-32 (1966)
8. Bruce, J.W.: On the degree of determinacy of smooth functions. Bull. Lond. Math. Soc. **13**, 55-58 (1981)
9. Bruce, J.W., Gaffney, T., du Plessis, A.A.: On left equivalence of map-germs. Bull. Lond. Math. Soc. **16**, 303-306 (1984)
10. Damon, J.N.: The unfolding and determinacy theorems for subgroups of $\mathscr{A}$ and $\mathscr{K}$. Mem. Am. Math. Soc. no 306 (1984): see also Singularities. Proc. Symp. Pure Math. Amer. Math. Soc.: Providence, R.I. **40**, (part 1) (1983)
11. Demazure, M., Gabriel, P.: Groupes algébriques I. North-Holland: Amsterdam, 1970
12. Floyd, E.E.: Periodic maps via Smith theory. In: Borel, A. (ed.), Seminar on transformation groups, Chapter III. Ann. Math. Stud. no 46, Princeton University Press: Princeton, N.J. 1960
13. Gaffney, T.: Properties of finitely determined germs. Thesis, Brandeis University, 1975
14. Gaffney, T.: On the order of determination of a finitely determined germ. Invent. Math. **37**, 83-92 (1976)
15. Gaffney, T.: A note on the order of determination of a finitely determined germ. Invent. Math. **52**, 127-130 (1979)
16. Gaffney, T.: The structure of $T\mathscr{A}(f)$, classification, and an application to differential geometry. In: Orlik, P. (ed.), Singularities. Proc. Symp. Pure Math. **40**, (part 1) 409-428 (1983)
17. Gaffney, T., du Plessis, A.A.: More on the determinacy of smooth map-germs. Invent. Math. **66**, 137-163 (1982)
18. Hochschild, G.P.: The structure of Lie groups. San Francisco-London-Amsterdam: Holden-Day 1965
19. Hochschild, G.P.: Basic theory of algebraic groups and Lie algebras. Berlin-Heidelberg-New York: Springer 1981
20. Mather, J.N.: Stability of C^∞ mappings III: finitely determined map-germs. Publ. Math., Inst. Hautes Etud. Sci. **35**, 127-156 (1969)
21. Mather, J.N.: Stability of C^∞ mappings IV: classification of stable germs by $\mathbb{R}$-algebras. Publ. Math., Inst. Hautes Etud. Sci. **37**, 223-248 (1970)
22. du Plessis, A.A.: On the determinacy of smooth map-germs. Invent. Math. **58**, 107-160 (1980)

23. Poenaru, V.: Singularités C^∞ en présence de symétrie. Lect. Notes in Math., vol. 510. Berlin-Heidelberg-New York: Springer 1976
24. Roberts, R.M.: Characterisations of finitely determined map-germs. Math. Ann. **275**, 573–582 (1986)
25. Thom, R., Levine, H.: Singularities of differentiable mappings. Lect. Notes in Math., vol. 192, pp. 1–89. Berlin-Heidelberg-New York: Springer 1971
26. Wall, C.T.C.: Finite determinacy of smooth map-germs. Bull. Lond. Math. Soc. **13**, 481–539 (1981)
27. Wall, C.T.C.: Equivariant jets. Math. Ann. **272**, 41–65 (1985)
28. Rosenlicht, M.: On quotient varieties and the affine embedding of certain homogeneous spaces. Trans. Am. Math. Soc. **101**, 211–223 (1961)

Oblatum 14-XI-1985 & 14-VIII-1986

Invent. math. 88, 555–569 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Newton polyhedra and the degree of the *L*-function associated to an exponential sum

Alan Adolphson [1, a] and Steven Sperber [2, b]

[1] Department of Mathematics, Oklahoma State University, Stillwater, OK 74078, USA
[2] School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA

§ 1. Introduction

The main results of this paper sharpen and extend some results of Bombieri ([2, 3]) on exponential sums. Perhaps more important, we shall connect, in this paper and in several subsequent papers, the study of exponential sums with singularity theory.

Bombieri's work gives bounds on the degree and total degree of the *L*-function associated to an exponential sum (where the degree of a rational function is the number of zeros minus the number of poles and the total degree is the number of zeros plus the number of poles). These bounds are expressed in terms of the degree and number of variables of the polynomial f appearing in the exponential sum. Our improvement gives sharper bounds by taking into account which monomials actually occur in f. More precisely, the bounds will be expressed in terms of the Newton polyhedron of f. The connection with singularity theory is apparent from the work of Kouchnirenko [12], who showed that our upper bound for the degree is, under certain restrictions on f, related to the Milnor numbers of the isolated singular points of the hypersurface $f=0$. We intend to pursue this in future work.

Let p be a prime, $q=p^a$, k the field of q elements, and k_m its extension of degree m. Fix a nontrivial additive character Ψ of k. For any Laurent polynomial $f\in k[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]$ we may form the exponential sums

$$S_m^*(f)=\sum \Psi\circ Tr_m(f(x_1,\ldots,x_n)), \tag{1.1}$$

where Tr_m: $k_m\to k$ is the trace map and the sum is over all $(x_1,\ldots,x_n)\in(k_m^\times)^n$. The associated L-function is

$$L^*(f,t)=\exp\left(\sum_{m=1}^{\infty} S_m^*(f)\frac{t^m}{m}\right). \tag{1.2}$$

[a] Partially supported by NFS Grant N. DMS-8401723.
[b] Partially supported by NFS Grant N. DMS-8301453.

When f is actually a polynomial, we can also define sums $S_m(f)$ by taking the sum on the right-hand side of (1.1) over all $(x_1,\dots,x_n)\in(k_m)^n$. The associated L-function will be denoted $L(f,t)$. By general theorems of Dwork and Grothendieck, both $L^*(f,t)$ and $L(f,t)$ are rational functions.

The sums $S_m^*(f)$ and $S_m(f)$ are related in the following way. For any subset $A\subseteq\{1,2,\dots,n\}$, let f_A be the polynomial obtained from f by setting $x_i=0$ for $i\in A$. Let $|A|$ denote the cardinality of A. Then

$$S_m(f)=\sum_A S_m^*(f_A) \tag{1.3}$$

$$S_m^*(f)=\sum_A(-1)^{|A|}S_m(f_A). \tag{1.4}$$

In terms of the corresponding L-functions, these relations are expressed as

$$L(f,t)=\prod_A L^*(f_A,t) \tag{1.5}$$

$$L^*(f,t)=\prod_A L(f_A,t)^{(-1)^{|A|}}. \tag{1.6}$$

Thus information about $L^*(f,t)$ will yield information about $L(f,t)$, and vice versa.

Write $j=(j_1,\dots,j_n)\in\mathbf{Z}^n$, $x=(x_1,\dots,x_n)$, $x^j=x_1^{j_1}\dots x_n^{j_n}$. The Newton polyhedron $\Delta(f)$ of a Laurent polynomial $f\in k[x_1,x_1^{-1},\dots,x_n,x_n^{-1}]$ is defined as follows: Write

$$f=\sum_{j\in J}a_j x^j, \tag{1.7}$$

where the sum is over a finite subset J of $\mathbf{Z}^n$. Then $\Delta(f)$ is the convex hull in $\mathbf{R}^n$ of the set $J\cup\{(0,\dots,0)\}$. Let $V(f)$ be the volume of $\Delta(f)$ with respect to Lebesgue measure on $\mathbf{R}^n$. We define the dimension $\tilde{n}$ of $\Delta(f)$ to be the dimension of the smallest subspace $E(f)$ of $\mathbf{R}^n$ containing $\Delta(f)$. Let $\tilde{V}(f)$ be the $\tilde{n}$-dimensional volume of $\Delta(f)$ relative to the induced lattice, i.e., $\tilde{V}(f)$ is the volume of $\Delta(f)$ with respect to Haar measure on $E(f)$ normalized so that a fundamental domain for the lattice $E(f)\cap\mathbf{Z}^n$ has volume 1.

We shall assume throughout that $\tilde{n}\geqq 1$, since $\tilde{n}=0$ implies f is constant and the corresponding exponential sum is trivial. Our main result is:

Theorem 1.8. *The following inequalities hold:*

$$0\leqq \text{degree } L^*(f,t)^{(-1)^{n-1}}\leqq n!\,V(f) \tag{1.9}$$

$$\text{total degree } L^*(f,t)\leqq 2^{\tilde{n}}(1+2^{(\tilde{n}+1)/\tilde{n}})^n\,\tilde{n}!\,\tilde{V}(f). \tag{1.10}$$

Observe that if $\tilde{n}<n$ we may conclude that degree $L^*(f,t)=0$, whereas if $\tilde{n}=n$, then $\tilde{V}(f)=V(f)$.

Note that for any polynomial $f\in k[x_1,\dots,x_n]$ of degree d, $\Delta(f)$ is contained in the simplex with vertices $(0,\dots,0)$, $(d,0,\dots,0)$, $(0,d,0,\dots,0),\dots,(0,\dots,0,d)$, hence $n!\,V(f)\leqq d^n$. Thus (1.9) implies the earlier estimate of Bombieri [2, Theorem 1(iv)]. If $\Delta(f)$ equals this simplex, then (1.10) is weaker than [3, Theorem 1A] by a factor that lies between $(3/2)^n$ and $(5/2)^n$, depending on n.

Thus (1.10) is not an improvement over [3] in all cases. However, it is an improvement in most cases and it does make clear certain properties that are not apparent from [3]. For example, (1.10) implies that the total degree of $L^*(f,t)$ grows at most linearly as a function of the degree of f in each variable.

Corresponding estimates for $L(f,t)$ are straightforward to derive. It follows from (1.5) that when f is a polynomial,

$$L(f,t)^{(-1)^{n-1}} = \prod_A (L^*(f_A,t)^{(-1)^{n-|A|-1}})^{(-1)^{|A|}}. \tag{1.11}$$

Let

$$\mu_1 = \sum_{|A| \text{ even}} (n-|A|)!\, V(f_A)$$

$$\mu_2 = \sum_{|A| \text{ odd}} (n-|A|)!\, V(f_A).$$

Then (1.9) and (1.11) imply

$$-\mu_2 \leqq \text{degree } L(f,t)^{(-1)^{n-1}} \leqq \mu_1. \tag{1.12}$$

From (1.5) and the estimate, total degree $L^*(f,t) \leqq 2^{\tilde{n}} 5^n \tilde{n}!\, \tilde{V}(f)$ (which is weaker than (1.10)), we have immediately

$$\text{total degree } L(f,t) \leqq 2^{\tilde{n}} 6^n \tilde{n}!\, \tilde{V}(f). \tag{1.13}$$

There is a standard technique [3, Sect. IV] for extending these estimates to exponential sums on a variety. If V is a variety and f is a regular function on V, one can define exponential sums

$$S_m(V,f) = \sum \Psi \circ Tr_m(f(x)),$$

where the sum is over the k_m-rational points $x \in V$. The associated L-function, defined as in (1.2), will be denoted $L(V,f;t)$. If V is defined by the vanishing of polynomials $g_1, \ldots, g_r \in k[x_1, \ldots, x_n]$ and if $f \in k[x_1, \ldots, x_n]$, then it is clear that

$$q^{rm} S_m(V,f) = S_m\left(f + \sum_{i=1}^{r} y_i g_i\right),$$

where $y_1, \ldots, y_r$ are new indeterminates and we regard $f + \sum_{i=1}^{r} y_i g_i$ as a polynomial in $x_1, \ldots, x_n, y_1, \ldots, y_r$. Thus

$$L(V,f;q^r t) = L\left(f + \sum_{i=1}^{r} y_i g_i, t\right),$$

and we can apply (1.12) and (1.13) to get corresponding estimates for $L(V,f;t)$. Note that adding the extra variables $y_1, \ldots, y_r$ does not cause the estimate to blow up as was the case in passing from Theorem 1 to Theorem 2 in [3], where the upper bound grows exponentially with the total number of variables.

For example, if $\deg f = d$ and $\deg g_i = d_i$ for $i = 1, 2, \ldots, r$, then $\Delta(f + \sum y_i g_i) \subset \mathbf{R}^{n+r}$ is contained in the polyhedron with vertices $(x_1, \ldots, x_n; y_1, \ldots, y_r)$ of the following three types: $x_i = 0$ for all i and $y_j = 0$ for all but at most one index j_0

where $y_{j_0}=1$; $y_j=0$ for all j and $x_i=0$ for all but one index i_0 where $x_{i_0}=d$; $x_i=0$ for all but one index i_0 where $x_{i_0}=d_k$ and $y_j=0$ for all but one index j_0 where $y_{j_0}=1$, $k=1,2,\ldots,r$. It is not difficult to check that the volume of this polyhedron is

$$\frac{1}{(n+r)!}\sum_{i_0+i_1+\ldots+i_r=n} d^{i_0}d_1^{i_1}\ldots d_r^{i_r}.$$

Thus (1.9) implies

$$\text{degree } L^*(f+\textstyle\sum y_i g_i,t)^{(-1)^{n-1}} \leqq \displaystyle\sum_{i_0+i_1+\ldots+i_r=n} d^{i_0}d_1^{i_1}\ldots d_r^{i_r},$$

which reproves one of the main results of [1].

The estimate (1.9) is sharp. In a future article we shall show that if f is nondegenerate with respect to $\Delta(f)$ (in the sense of [12]), then the degree of $L^*(f,t)^{(-1)^{n-1}}$ is exactly $n!\,V(f)$ (in particular, if $\tilde{n}<n$, the degree is zero). If in addition $\tilde{n}=n$, then $L^*(f,t)^{(-1)^{n-1}}$ is actually a polynomial. For example, if $f=x_1+\ldots+x_n+(x_1\ldots x_n)^{-1}$, then $\Delta(f)$ is the simplex with vertices $(1,0,\ldots,0)$, $(0,1,0,\ldots,0),\ldots,(0,\ldots,0,1)$, $(-1,-1,\ldots,-1)$, and one checks easily that $n!\,V(f)=n+1$. This gives the well-known result ([7, 14]) that the L-function associated to the hyperkloosterman sum is a polynomial of degree $n+1$. If f is a polynomial that is nondegenerate with respect to $\Delta(f)$, then (1.11) implies that the degree of $L(f,t)^{(-1)^{n-1}}$ equals $\mu_1-\mu_2$. If in addition f is commode in the sense of [12] then $L(f,t)^{(-1)^{n-1}}$ is a polynomial. For example, if f is a polynomial of degree d prime to p whose homogeneous part of highest degree is nonsingular, then after an invertible change of coordinates f is nondegenerate and commode, and one calculates $\mu_1-\mu_2=(d-1)^n$, which is a result first proved by Deligne [6, Lemma 8.5(i)].

As a by-product of the method, we obtain p-adic information about the roots of $L^*(f,t)^{(-1)^{n-1}}$ when this L-function is a polynomial, generalizing [2, Theorem 2]. Suppose that $\Delta(f)$ is n-dimensional and let $S(f)$ denote the sum of the $(n-1)$-dimensional volumes (relative to the induced lattices) of those $(n-1)$-dimensional faces of $\Delta(f)$ that contain the origin. Let ord_q be an additive valuation on the algebraic closure of $\mathbf{Q}$, lying over p and normalized by $\text{ord}_q q=1$.

Theorem 1.14. *Suppose that* $\dim \Delta(f)=n$ *and that* $L^*(f,t)^{(-1)^{n-1}}$ *is a polynomial of degree* $n!\,V(f)$. *Let* $\{\rho_i\}_{i=1}^{n!V(f)}$ *be its reciprocal roots. Then*

$$\sum_{i=1}^{n!V(f)} \text{ord}_q \rho_i \geqq \frac{n}{2} n!\,V(f) - \tfrac{1}{2}(n-1)!\,S(f).$$

In particular, if the origin is an interior point of $\Delta(f)$, *then*

$$\sum_{i=1}^{n!V(f)} \text{ord}_q \rho_i \geqq \frac{n}{2} n!\,V(f).$$

This theorem will be applied in a future article to obtain archimedian estimates for exponential sums satisfying the conditions of the preceding paragraph.

Our method of proving Theorems 1.8 and 1.14 parallels very closely that of Bombieri in [2] and [3]. The main new idea, described in Sect. 2, is to study

more closely the growth conditions satisfied by the p-adic analytic function $F(x_1,\dots,x_n)$ that Dwork's theory associates to the exponential sum. This leads to the definition of a new weight function $w(r_1,\dots,r_n)$, which we believe will play a fundamental role in future studies of p-adic properties of exponential sums. We also note that our procedure (Sect. 5) for estimating the number of monomials of a given weight is different from Bombieri's, since explicit formulas are not possible in our case.

§ 2. p-Adic theory

The purpose of this section is to present our version of Dwork's theory describing $L^*(f,t)$ in terms of a completely continuous operator on a p-adic Banach space. This Banach space is a set of p-adic Laurent series satisfying certain growth conditions. Our main contribution is a sharper formulation of these growth conditions in terms of a new weight function w. This weight function has a geometric description in terms of the Newton polyhedron of f, which will eventually lead to our main results.

Let $\mathbf{Q}_p$ denote the p-adic numbers and let Ω be the completion of an algebraic closure of $\mathbf{Q}_p$. Let K denote the unramified extension of $\mathbf{Q}_p$ in Ω of degree a, where $q=p^a$. Let $\pi\in\Omega$ satisfy $\pi^{p-1}=-p$. Then $\Omega_1=\mathbf{Q}_p(\pi)$ is a totally ramified extension of $\mathbf{Q}_p$ of degree $p-1$, in fact, $\Omega_1=\mathbf{Q}_p(\zeta_p)$, where ζ_p is a primitive p-th root of unity. Let Ω_0 be the compositum of Ω_1 and K. Then Ω_0 is an unramified extension of Ω_1 of degree a. The residue class fields of Ω_0 and K are both k, and the residue fields of Ω_1 and $\mathbf{Q}_p$ are both $\mathbf{F}_p$. The Frobenius automorphism $x\mapsto x^p$ of $\mathrm{Gal}(k/\mathbf{F}_p)$ lifts to a generator τ of $\mathrm{Gal}(K/\mathbf{Q}_p)$ which is extended to Ω_0 by requiring $\tau(\pi)=\pi$. If ζ is a $(q-1)$-st root of unity in Ω_0, then $\tau(\zeta)=\zeta^p$. Denote by "ord" the additive valuation on Ω normalized by $\mathrm{ord}\,p=1$, and denote by "ord_q" the additive valuation normalized by $\mathrm{ord}_q\,q=1$.

Let $E(t)$ be the Artin-Hasse exponential series:

$$E(t)=\exp\left(\sum_{m=0}^{\infty}\frac{t^{p^m}}{p^m}\right)\in(\mathbf{Z}_p\cap\mathbf{Q})[[t]]. \tag{2.1}$$

Let $\gamma\in\Omega_1$ be a root of $\sum_{m=0}^{\infty} t^{p^m}/p^m=0$ satisfying $\mathrm{ord}\,\gamma=1/(p-1)$. The series

$$\theta(t)=E(\gamma t)=\sum_{m=0}^{\infty}\lambda_m t^m \tag{2.2}$$

is a splitting function in Dwork's terminology [8, Sect. 4] and its coefficients satisfy

$$\mathrm{ord}\,\lambda_m\geqq\frac{m}{p-1},\qquad \lambda_m\in\Omega_1. \tag{2.3}$$

For f as in (1.7), let

$$\hat{f}=\sum_{j\in J}\hat{a}_j x^j\in K[x_1,x_1^{-1},\dots,x_n,x_n^{-1}] \tag{2.4}$$

be its Teichmüller lifting, i.e., $(\hat{a}_j)^q = \hat{a}_j$. Set

$$F(x) = \prod_{j \in J} \theta(\hat{a}_j x^j) \tag{2.5}$$

$$F_0(x) = \prod_{i=0}^{a-1} \tau^i F(x^{p^i}). \tag{2.6}$$

Note that (2.3) implies $F(x)$ and $F_0(x)$ are well-defined as formal Laurent series in $x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}$ with coefficients in Ω_0.

Our goal is a description of the growth conditions satisfied by F. Write

$$F(x) = \sum_{r \in \mathbf{Z}^n} F_r x^r. \tag{2.7}$$

The coefficients F_r can be described as follows: Let A be the $(n \times |J|)$-matrix whose columns are the $j = (j_1, \ldots, j_n) \in J$. Then from (2.2) and (2.5),

$$F_r = \sum_u \Big(\sum_{j \in J} \lambda_{u_j} (\hat{a}_j)^{u_j} \Big), \tag{2.8}$$

where the outer sum is over all $|J|$-tuples $u = (u_j)_{j \in J}$ of nonnegative integers such that (thinking of u and r as column vectors)

$$Au = r. \tag{2.9}$$

Thus $F_r = 0$ if (2.9) has no such solutions. Otherwise, (2.3) implies

$$\operatorname{ord} F_r \geqq \inf_u \Big\{ \sum_{j \in J} u_j/(p-1) \Big\}, \tag{2.10}$$

where the inf is taken over all such solutions.

A weaker but more convenient estimate is obtained by considering rational solutions to (2.9). Set

$$w(r) = \inf_u \Big\{ \sum_{j \in J} u_j \Big\} \tag{2.11}$$

where now the inf is taken over all $|J|$-tuples $u = (u_j)_{j \in J}$ of nonnegative rational numbers satisfying (2.9). Set $w(r) = +\infty$ if (2.9) has no such solutions. Then (2.10) implies

$$\operatorname{ord} F_r \geqq w(r)/(p-1), \tag{2.12}$$

with the obvious convention that $F_r = 0$ if $w(r) = +\infty$.

Our main point is that the weight function w can be described geometrically. Let $\mathbf{R}_+\langle J \rangle$ denote the subset of $\mathbf{R}^n$ consisting of all linear combinations with nonnegative real coefficients of elements of J. Then $\mathbf{R}_+\langle J \rangle$ is the union of all rays emanating from the origin and passing through $\Delta(f)$. Equation (2.9) has a solution u whose components are nonnegative rational numbers if and only if $r \in \mathbf{R}_+\langle J \rangle$. Thus $w(r) = +\infty$ if and only if $r \notin \mathbf{R}_+\langle J \rangle$. If $r \in \mathbf{R}_+\langle J \rangle$, the ray emanating from the origin and passing through r intersects $\Delta(f)$ in a face that does not contain the origin. Let $\sum_{i=1}^{n} \alpha_i X_i = 1$ be the equation of a hyperplane passing through this face (this hyperplane is not uniquely determined

unless the face has dimension $n-1$). Then by standard arguments in linear programming,

$$w(r)=\sum_{i=1}^{n}\alpha_i r_i. \tag{2.13}$$

From this description of w, the following result is immediate:

Lemma 2.14. (a) *If k is a nonnegative integer, then $w(kr)=kw(r)$.*

(b) *If $w(r), w(r')<+\infty$, then $w(r+r')<+\infty$. Furthermore, $w(r+r')\leqq w(r)+w(r')$.*

(c) *There exists a positive integer D such that $w(\mathbf{Z}^n)\subseteq\frac{1}{D}\mathbf{Z}_+\cup\{+\infty\}$, where $\mathbf{Z}_+$ denotes the set of nonnegative integers.*

We now apply Dwork's theory to the space of all formal Laurent series satisfying growth conditions of the form (2.12). For a positive real number b, let $L(b)$ be the space of all Laurent series $\sum_{r\in\mathbf{Z}^n} A_r x^r\in\Omega[[x_1, x_1^{-1}, \ldots, x_n, x_n^{-1}]]$ satisfying

$$\operatorname{ord} A_r\geqq b\, w(r)+O(1). \tag{2.15}$$

In particular, this means $A_r=0$ if $w(r)=+\infty$. Thus

$$F\in L\left(\frac{1}{p-1}\right),\qquad F_0\in L\left(\frac{p}{q(p-1)}\right). \tag{2.16}$$

Lemma 2.14(b) and an easy calculation imply that $L(b)$ is a ring. Define an operator ψ on Laurent series by

$$\psi\left(\sum A_r x^r\right)=\sum A_{pr} x^r.$$

From Lemma 2.14(a), one immediately deduces that $\psi(L(b))\subseteq L(pb)$.

Let ι: $L(p/(p-1))\hookrightarrow L(p/q(p-1))$ be the canonical injection and denote by α the composition

$$L\left(\frac{p}{p-1}\right)\xrightarrow{\iota} L\left(\frac{p}{q(p-1)}\right)\xrightarrow{F_0} L\left(\frac{p}{q(p-1)}\right)\xrightarrow{\psi^a} L\left(\frac{p}{p-1}\right),$$

where the middle arrow means "multiplication by F_0". It follows from Serre [13] that the trace $Tr(\alpha^m)$ and characteristic series $\det(I-t\alpha)$ are well-defined and that $\det(I-t\alpha)$ is a p-adic entire function, i.e., it converges for all $t\in\Omega$. The Dwork trace formula asserts that

$$(q^m-1)^n\, Tr(\alpha^m)=S_m^*(f), \tag{2.17}$$

where the nontrivial additive character implicit on the right-hand side is derived from the splitting function $\theta(t)$ [8, Sect. 4]. Equivalently, one can define an operator δ on formal power series with constant term 1 by

$$g(t)^\delta=g(t)/g(qt),$$

and then (2.17) and the relation $\det(I-t\alpha)=\exp\left(-\sum_{m=1}^{\infty} Tr(\alpha^m)\frac{t^m}{m}\right)$ imply

$$L^*(f,t)^{(-1)^{n-1}}=\det(I-t\alpha)^{\delta^n}. \tag{2.18}$$

§ 3. The Newton polygon of $\det(I-t\alpha)$

The first step in the proof of Theorem 1.8 is to find a good lower bound for the Newton polygon of $\det(I-t\alpha)$. Let α_0 be the Ω_1-linear endomorphism of $L(p/(p-1))$ defined by $\alpha_0=\psi\circ\tau^{-1}\circ F$. Arguing exactly as in [9, Lemma 7.1] one shows that a point $(x,y)\in\mathbf{R}^2$ is a vertex of the Newton polygon of $\det(I-t\alpha)$ computed with respect to the valuation ord_q if and only if (ax,ay) is a vertex of the Newton polygon of $\det_{\Omega_1}(I-t\alpha_0)$ computed with respect to the valuation ord. Hence it suffices to estimate the Newton polygon of $\det_{\Omega_1}(I-t\alpha_0)$.

To do this we consider the matrix representation of α_0 as Ω_1-linear endomorphism of $L(b)$. Let $\xi_1,\dots,\xi_a$ be an integral basis for Ω_0 over Ω_1 that has the property of p-adic directness [8, Sect. 3c], i.e., for any $\beta_1,\dots,\beta_a\in\Omega_1$,

$$\mathrm{ord}\left(\sum_{i=1}^{a}\beta_i\xi_i\right)=\min_i\{\mathrm{ord}\,\beta_i\}. \tag{3.1}$$

We represent α_0 by an infinite matrix $C(r,i;r',i')$ (where $1\leqq i,i'\leqq a, r,r'\in\mathbf{Z}^n$, $w(r),w(r')<+\infty$) defined in the usual way:

$$\alpha_0(\xi_i x^r)=\sum_{r',i'}C(r,i;r',i')\,\xi_{i'}x^{r'}. \tag{3.2}$$

For F_r as in (2.7), write

$$F_r=\sum_{i=1}^{a}F_{r,i}\,\xi_i,\qquad F_{r,i}\in\Omega_1 \tag{3.3}$$

$$\tau^{-1}(\xi_i\xi_j)=\sum_{k=1}^{a}s_{ijk}\,\xi_k,\qquad s_{ijk}\in\Omega_1. \tag{3.4}$$

It follows easily from (3.1) that

$$\mathrm{ord}\,s_{ijk}\geqq 0 \tag{3.5}$$

and from (3.1) and (2.12) that

$$\mathrm{ord}\,F_{r,i}\geqq w(r)/(p-1). \tag{3.6}$$

One calculates directly from (3.2), (3.3), (3.4), and the definition of α_0 that

$$C(r,i;r',i')=\sum_{k=1}^{a}s_{iki'}F_{pr'-r,k}, \tag{3.7}$$

hence by (3.5) and (3.6)

$$\mathrm{ord}\,C(r,i;r',i')\geqq w(pr'-r)/(p-1). \tag{3.8}$$

There is a well-known relation [13, Sect. 5] between this infinite matrix and $\det(I-t\alpha_0)$. Write

$$\det(I-t\alpha_0)=\sum_{m=0}^{\infty}c_m t^m. \tag{3.9}$$

Then

$$c_m=(-1)^m\sum_{\sigma}\varepsilon(\sigma)\prod_{k=1}^{m}C(r(k),i(k);r(\sigma(k)),i(\sigma(k))), \tag{3.10}$$

where the outer sum is over all sets $\{(r(k), i(k))\}_{k=1}^{m}$ of m distinct elements satisfying $1 \leqq i(k) \leqq a$, $r(k) \in \mathbf{Z}^n$, $w(r(k)) < +\infty$, and where the inner sum is over all permutations σ on m letters, $\varepsilon(\sigma)$ being the sign of the permutation. It follows from (3.8) and (3.10) that

$$\operatorname{ord} c_m \geqq (p-1)^{-1} \inf\left\{\sum_{k=1}^{m} w(pr(\sigma(k)) - r(k))\right\},$$

where the inf is over all sets $\{(r(k), i(k))\}_{k=1}^{m}$ as above and over all permutations σ of m letters. By Lemma 2.1 (a) and (b),

$$\operatorname{ord} c_m \geqq \inf\left\{\sum_{k=1}^{m} w(r(k))\right\}, \tag{3.11}$$

where the inf is over all sets $\{(r(k), i(k))\}_{k=1}^{m}$ as above.

Thus the problem is reduced to counting the number of $r \in \mathbf{Z}^n$ of a given weight. Using Lemma 2.14(c), we may define integers $W(k)$ for $k = 0, 1, 2, \ldots$ by

$$W(k) = \operatorname{card}\{r \in \mathbf{Z}^n \mid w(r) = k/D\}. \tag{3.12}$$

Estimate (3.11) and the argument of [9, Sect. 7] then show that the Newton polygon of $\det_{\Omega_1}(I - t\alpha_0)$ lies above the polygon with vertices $(0,0)$ and

$$\left(a \sum_{k=0}^{M} W(k), \frac{a}{D} \sum_{k=0}^{M} k W(k)\right), \qquad M = 0, 1, 2, \ldots.$$

The remarks at the beginning of this section then imply:

Proposition 3.13. *The Newton polygon of* $\det(I - t\alpha)$ *computed with respect to* ord_q *lies above the polygon with vertices* $(0,0)$ *and*

$$\left(\sum_{k=0}^{M} W(k), \frac{1}{D} \sum_{k=0}^{M} k W(k)\right), \qquad M = 0, 1, 2, \ldots.$$

§ 4. Degree of $L^*(f,t)^{(-1)^{n-1}}$

We now apply Bombieri's method [2, Sect. IV] to deduce the bounds for the degree of $L^*(f,t)^{(-1)^{n-1}}$. Since the lower bound of (1.9) involves nothing new, we give details for the upper bound only.

Since $L^*(f,t)^{(-1)^{n-1}}$ is a rational function, we may write

$$L^*(f,t)^{(-1)^{n-1}} = \prod_{i=1}^{R} (1 - \rho_i t) / \prod_{j=1}^{S} (1 - \eta_j t). \tag{4.1}$$

Inverting (2.18) yields

$$\det(I - t\alpha) = D_1(t)/D_2(t),$$

where

$$D_1(t)=\prod_{i=1}^{R}\prod_{m=0}^{\infty}(1-q^m\rho_i t)^{c(m)}$$

$$D_2(t)=\prod_{i=1}^{S}\prod_{m=0}^{\infty}(1-q^m\eta_j t)^{c(m)}$$

with $c(m)=\binom{m+n-1}{n-1}$.

Lemma 4.2. [2, Corollary to Lemma 3]. *For any real number* $x\geqq 0$,

$$\sum_{i=1}^{R}\sum{}'(x-\operatorname{ord}_q(q^m\rho_i))\,c(m)-\sum_{j=1}^{S}\sum{}'(x-\operatorname{ord}_q(q^m\eta_j))\,c(m)$$
$$\leqq x\sum_{k\leqq Dx}W(k)-D^{-1}\sum_{k\leqq Dx}kW(k), \tag{4.3}$$

where the sums $\sum'$ *are over all* m *such that the summands are positive.*

Since $\operatorname{ord}_q(q^m\rho_i)=m+\operatorname{ord}_q\rho_i$ and

$$\sum_{m\leqq x}(x-m)\,c(m)=\frac{x^{n+1}}{(n+1)!}+\frac{1}{2}\frac{x^n}{(n-1)!}+O(x^{n-1})$$

as $x\to+\infty$, the left-hand side of (4.3) equals

$$\frac{(R-S)\,x^{n+1}}{(n+1)!}+\left(\frac{n(R-S)}{2}-\sum_{i=1}^{R}\operatorname{ord}_q\rho_i+\sum_{j=1}^{S}\operatorname{ord}_q\eta_j\right)\frac{x^n}{n!}+O(x^{n-1}). \tag{4.4}$$

We now determine the asymptotic growth of the right-hand side of (4.3), which, when combined with (4.4), will lead to (1.9) and Theorem 1.14.

In what follows, a "face" of $\Delta(f)$ always means a closed face of arbitrary dimension. For each face Γ of $\Delta(f)$ not containing the origin, let $C(\Gamma)$ be the union of all half-lines emanating from the origin and passing through Γ. Let $\hat{\Gamma}$ be the convex hull of $\Gamma\cup\{(0,\ldots,0)\}$. Let $\tilde{V}(\Gamma)$ (resp. $\tilde{V}(\hat{\Gamma})$) be the volume of Γ (resp. $\hat{\Gamma}$) with respect to Haar measure on the smallest affine space (resp. linear space) containing Γ (resp. $\hat{\Gamma}$), normalized so that a fundamental domain for the induced lattice has volume 1. Let $E(f)$ be the smallest subspace of $\mathbf{R}^n$ containing $\Delta(f)$ and set $\tilde{n}=\dim E(f)$.

Define $W_\Gamma(k)$, $k=0,1,2,\ldots$, by

$$W_\Gamma(k)=\operatorname{card}\{r\in\mathbf{Z}^n\cap C(\Gamma)\mid w(r)=k/D\}.$$

Let $D(\Gamma)$ be the smallest positive integer such that

$$w(\mathbf{Z}^n\cap C(\Gamma))\subseteq\frac{1}{D(\Gamma)}\mathbf{Z}_+\cup\{+\infty\}.$$

Clearly, if $D\not\mid D(\Gamma)k$, then $W_\Gamma(k)=0$. If $D\mid D(\Gamma)k$, then there is a positive integer k' such that $k'/D(\Gamma)=k/D$. For each integer r, $0\leqq r\leqq D(\Gamma)-1$, if we

consider only those $k \geqq 1$ for which $k' \equiv r \pmod{D(\Gamma)}$, then $W_\Gamma(k)$ is a polynomial in k' of degree $\tilde{n}-1$ with leading coefficient $\tilde{V}(\Gamma)/D(\Gamma)^{\tilde{n}-1}$ [10, Corollaire 4.2.3]:

$$W_\Gamma(k) = \frac{\tilde{V}(\Gamma)}{D(\Gamma)^{\tilde{n}-1}} k'^{\tilde{n}-1} + \frac{a_{\Gamma,r}}{D(\Gamma)^{\tilde{n}-2}} k'^{\tilde{n}-2} + O(k'^{\tilde{n}-3}). \tag{4.5}$$

This implies that for x integral

$$\sum_{k \leqq Dx} W_\Gamma(k) = \sum_{k' \leqq D(\Gamma)x} \frac{\tilde{V}(\Gamma)}{D(\Gamma)^{\tilde{n}-1}} k'^{\tilde{n}-1} + \sum_{r=0}^{D(\Gamma)-1} \sum_{l=0}^{x-1} a_{\Gamma,r} \frac{(lD(\Gamma)+r+1)^{\tilde{n}-2}}{D(\Gamma)^{\tilde{n}-2}} + O(x^{\tilde{n}-2}). \tag{4.6}$$

A straightforward calculation then shows that

$$\sum_{k \leqq Dx} W_\Gamma(k) = \frac{\tilde{V}(\Gamma) D(\Gamma)}{\tilde{n}} x^{\tilde{n}} + \left(\frac{\tilde{V}(\Gamma)}{2} + \frac{1}{\tilde{n}-1} \sum_{r=0}^{D(\Gamma)-1} a_{\Gamma,r} \right) x^{\tilde{n}-1} + O(x^{\tilde{n}-2}). \tag{4.7}$$

It is not difficult to check that $\tilde{V}(\Gamma) D(\Gamma)/\tilde{n} = \tilde{V}(\hat{\Gamma})$. In fact, since $\sum_{k \leqq Dx} W_\Gamma(k)$ is the number of lattice points in $x\hat{\Gamma}$, the dilation of $\hat{\Gamma}$ by the factor x, it follows by [10, Corollaire B.4] and [11] that for x integral,

$$\sum_{k \leqq Dx} W_\Gamma(k) = \tilde{V}(\hat{\Gamma}) x^{\tilde{n}} + \frac{S'(\hat{\Gamma})}{2} x^{\tilde{n}-1} + O(x^{\tilde{n}-2}), \tag{4.8}$$

where $S'(\hat{\Gamma})$ denotes the sum of the volumes of the $(\tilde{n}-1)$-dimensional faces of $\hat{\Gamma}$, each volume being computed relative to the induced lattice. Comparing (4.7) and (4.8) we see that

$$\frac{1}{\tilde{n}-1} \sum_{r=0}^{D(\Gamma)-1} a_{\Gamma,r} = \tfrac{1}{2} S(\hat{\Gamma}), \tag{4.9}$$

where $S(\hat{\Gamma})$ denotes the sum of the volumes (relative to the induced lattices) of those $(\tilde{n}-1)$-dimensional faces of $\hat{\Gamma}$ that contain the origin.

It is clear that

$$W(k) = \sum_{\Gamma}{}' W_\Gamma(k) - \sum_{\sigma}{}'' W_\sigma(k) + O(k^{\tilde{n}-3}), \tag{4.10}$$

where $\sum_{\Gamma}{}'$ denotes a sum over all $(\tilde{n}-1)$-dimensional faces Γ of $\Delta(f)$ not containing the origin and $\sum_{\sigma}{}''$ denotes a sum over all $(\tilde{n}-2)$-dimensional faces σ of $\Delta(f)$ that are the intersection of two $(\tilde{n}-1)$-dimensional faces occuring in $\sum_{\Gamma}{}'$. We can now use (4.7), (4.8), and (4.9) to evaluate the first term on the right hand side of (4.3):

$$x \sum_{k \leqq Dx} W(k) = \sum_{\Gamma}{}' \tilde{V}(\tilde{\Gamma}) x^{\tilde{n}+1} + \tfrac{1}{2} \sum_{\Gamma}{}' (\tilde{V}(\Gamma) + S(\tilde{\Gamma})) x^{\tilde{n}} - \sum_{\sigma}{}'' \tilde{V}(\hat{\sigma}) x^{\tilde{n}} + O(x^{\tilde{n}-1}). \tag{4.11}$$

But $\sum_{\Gamma}{}' \tilde{V}(\hat{\Gamma}) = \tilde{V}(f)$. Furthermore, it is clear that

$$\tfrac{1}{2} \sum_{\Gamma}{}' S(\hat{\Gamma}) - \sum_{\sigma}{}'' \tilde{V}(\hat{\sigma}) = \tfrac{1}{2} S(f),$$

where $S(f)$ denotes the sum of the volumes (relative to the induced lattices) of those $(\tilde{n}-1)$-dimensional faces of $\Delta(f)$ that contain the origin. We have shown:

$$x \sum_{k \leqq Dx} W(k) = \tilde{V}(f)\, x^{\tilde{n}+1} + \tfrac{1}{2}(S(f) + \sum_{\Gamma}{}' \tilde{V}(\Gamma))\, x^{\tilde{n}} + O(x^{\tilde{n}-1}). \tag{4.12}$$

A similar argument based on (4.5) and (4.9) evaluates the second summand on the right hand side of (4.3):

$$D^{-1} \sum_{k \leqq Dx} kW(k) = \frac{\tilde{n}\,\tilde{V}(f)}{\tilde{n}+1} x^{\tilde{n}+1} + \frac{1}{2}\left(\frac{\tilde{n}-1}{\tilde{n}} S(f) + \sum_{\Gamma}{}' \tilde{V}(\Gamma)\right) x^{\tilde{n}} + O(x^{\tilde{n}-1}). \tag{4.13}$$

Combining (4.3), (4.4), (4.12), and (4.13) gives

$$\begin{aligned} &\frac{(R-S)\,x^{n+1}}{(n+1)!} + \left(\frac{n(R-S)}{2} - \sum_{i=1}^{R} \operatorname{ord}_q \rho_i + \sum_{j=1}^{S} \operatorname{ord}_q \eta_j\right) \frac{x^n}{n!} + O(x^{n-1}) \\ &\leqq \frac{\tilde{n}!\,\tilde{V}(f)}{(\tilde{n}+1)!} x^{\tilde{n}+1} + \tfrac{1}{2}(\tilde{n}-1)!\, S(f) \frac{x^{\tilde{n}}}{\tilde{n}!} + O(x^{\tilde{n}-1}). \end{aligned} \tag{4.14}$$

Let $V(f)$ be the volume of $\Delta(f)$ with respect to Lebesgue measure on $\mathbf{R}^n$. If $\tilde{n} < n$, then (4.14) implies $R - S = 0 (= V(f))$. If $\tilde{n} = n$, then $\tilde{V}(f) = V(f)$ and (4.14) implies $R - S \leqq n!\, V(f)$. This completes the proof of (1.9). If $\tilde{n} = n$ and $L^*(f,t)^{(-1)^{n-1}}$ is a polynomial of degree $n!\, V(f)$ (i.e., $R = n!\, V(f)$, $S = 0$ in (4.1)), then (4.14) implies

$$\left(\frac{n}{2} n!\, V(f) - \sum_{i=1}^{n!\,V(f)} \operatorname{ord}_q \rho_i\right) \frac{x^n}{n!} + O(x^{n-1}) \leqq \tfrac{1}{2}(n-1)!\, S(f) \frac{x^n}{n!} + O(x^{n-1}),$$

from which Theorem 1.14 follows immediately.

§ 5. Total degree of $L^*(f,t)$

Unlike the estimate for $\deg L^*(f,t)^{(-1)^{n-1}}$, the estimate for the total degree of $L^*(f,t)$ is not asymptotic in nature and consequently requires a different description of the function $W(k)$.

Lemma 5.1. *With the notation of the previous section,*

$$\sum_{k=0}^{\infty} W(k)\, t^k = \frac{P(t)}{(1-t^D)^{\tilde{n}}}, \tag{5.2}$$

where

$$P(t) = a_0 + a_1 t + \ldots + a_{\tilde{n}D}\, t^{\tilde{n}D}$$

is a polynomial with nonnegative integer coefficients satisfying $P(1) = \tilde{n}!\, \tilde{V}(f)$.

Proof. Equation (5.2) for some polynomial $P(t)$ satisfying $P(1) = \tilde{n}!\, \tilde{V}(f)$ is the content of [12, Lemme 2.9]. The other properties of $P(t)$ follow by the method of [4, Lemma 1.5]. □

It follows immediately from (5.2) that

$$W(k)=\sum_{\substack{i=0\\ i\equiv k\,(\mathrm{mod}\,D)}}^{\min\{k,\tilde{n}D\}} a_i\binom{\frac{k-i}{D}+\tilde{n}-1}{\tilde{n}-1}. \tag{5.3}$$

Given a nonnegative integer M, let $k(M,i)$ be the largest integer $\leqq M$ that is congruent to i modulo D and put $M(i)=(k(M,i)-i)/D$. Interchanging the order of summation and using the identity $\sum_{r=0}^{s}\binom{r+\tilde{n}-1}{\tilde{n}-1}=\binom{s+\tilde{n}}{\tilde{n}}$ gives

$$\sum_{k=0}^{M} W(k)=\sum_{i=0}^{\min\{M,\tilde{n}D\}} a_i\binom{M(i)+\tilde{n}}{\tilde{n}}. \tag{5.4}$$

Using (5.3), writing $k/D=(k-i)/D+i/D$, and using the identity $\sum_{r=0}^{s} r\binom{r+\tilde{n}-1}{\tilde{n}-1}=\frac{\tilde{n}s}{\tilde{n}+1}\binom{s+\tilde{n}}{\tilde{n}}$ we have

$$D^{-1}\sum_{k=0}^{M} kW(k)=\sum_{i=0}^{\min\{M,\tilde{n}D\}} a_i\left(\frac{\tilde{n}M(i)}{\tilde{n}+1}+\frac{i}{D}\right)\binom{M(i)+\tilde{n}}{\tilde{n}}. \tag{5.5}$$

Since $k(M,i)\geqq M-D+1$, it follows that

$$\frac{\tilde{n}M(i)}{\tilde{n}+1}+\frac{i}{D}\geqq\max\left\{0,\frac{\tilde{n}(M-D+1)}{(\tilde{n}+1)D}\right\}.$$

We have proved:

Lemma 5.6. *The slope of the line connecting the origin to the point* $\left(\sum_{k=0}^{M} W(k),\ D^{-1}\sum_{k=0}^{M} kW(k)\right)$ *is* $\geqq\max\left\{0,\frac{\tilde{n}(M-D+1)}{(\tilde{n}+1)D}\right\}$.

We can now apply the method of Bombieri [3] to estimate the total degree of $L^*(f,t)$. By [5, Corollary 5.5.3(iii)] we have $0\leqq\mathrm{ord}_q\rho_i,\ \mathrm{ord}_q\eta_j\leqq n$, where ρ_i, η_j are as in (4.1). On the other hand, writing out the right-hand side of (2.18) gives

$$\frac{\prod(1-\rho_i t)}{\prod(1-\eta_j t)}=\prod_{m=0}^{n}\det(I-q^m t\alpha)^{(-1)^m\binom{n}{m}}.$$

Hence the zeros and poles of $L^*(f,t)$ all occur among the zeros of $\prod_{m=0}^{n}\det(I-q^m t\alpha)^{\binom{n}{m}}$ of $\mathrm{ord}_q\leqq n$. Let N_m be the number of zeros of $\det(I-q^m t\alpha)$ of $\mathrm{ord}_q\leqq n$. Then

$$\text{total degree } L^*(f,t)\leqq\sum_{m=0}^{n}\binom{n}{m}N_m. \tag{5.7}$$

Now N_m is the number of zeros of $\det(I-t\alpha)$ of $\operatorname{ord}_q \leqq n-m$, hence equals the total length of the projections on the x-axis of the sides of slope $\leqq n-m$ of the Newton polygon of $\det(I-t\alpha)$. Let M_{n-m} be the least integer $\geqq (n-m)(\tilde{n}+1)D\tilde{n}^{-1}+D-1$. Then it is easily checked that

$$\frac{\tilde{n}(M_{n-m}-D+1)}{(\tilde{n}+1)D} \geqq n-m,$$

so by Lemma 5.6 and Proposition 3.13,

$$N_m \leqq \sum_{k=0}^{M_{n-m}} W(k). \tag{5.8}$$

Note that from the definition, $M(i) \leqq [M/D]$, so from (5.4) and Lemma 5.1 it follows that

$$\sum_{k=0}^{M} W(k) \leqq \tilde{n}!\, \tilde{V}(f) \binom{[M/D]+\tilde{n}}{\tilde{n}}. \tag{5.9}$$

Combining (5.7), (5.8), and (5.9) gives

$$\text{total degree } L^*(f,t) \leqq \tilde{n}!\, \tilde{V}(f) \sum_{m=0}^{n} \binom{n}{m} \binom{[M_{n-m}/D]+\tilde{n}}{\tilde{n}}. \tag{5.10}$$

A straightforward calculation shows that

$$[M_{n-m}/D] \leqq n-m+\frac{n-m}{\tilde{n}}+\frac{\tilde{n}D-1}{\tilde{n}D}. \tag{5.11}$$

Using the elementary estimate $\binom{[M_{n-m}/D]+\tilde{n}}{\tilde{n}} \leqq 2^{[M_{n-m}/D]+\tilde{n}-1}$ it follows from (5.10) and (5.11) that

$$\text{total degree } L^*(f,t) \leqq 2^{\tilde{n}}(1+2^{(\tilde{n}+1)/\tilde{n}})^n \tilde{n}!\, \tilde{V}(f), \tag{5.12}$$

which establishes (1.10).

References

1. Adolphson, A., Sperber, S.: On the degree of the L-function associated with an exponential sum. (to appear)
2. Bombieri, E.: On exponential sums in finite fields. Am. J. Math. **88**, 71–105 (1966)
3. Bombieri, E.: On exponential sums in finite fields, II. Invent. Math. **47**, 29–39 (1978)
4. Danilov, V.I.: The geometry of toric varieties. Russ. Math. Surv. **33**(2), 97–154 (1979)
5. Deligne, P.: Théorèmes d'intégralité. In: Deligne, P., Katz, N. (eds.), Groupes de monodromie en géométrie algébrique (SGA 7 II). Lect. Notes Math. vol. 340, pp. 384–400. Berlin-Heidelberg-New York: Springer 1973
6. Deligne, P.: La conjecture de Weil I. Publ. Math. Inst. Hautes Etud. Sci. **43**, 273–307 (1974)
7. Deligne, P.: Applications de la formule des traces aux sommes trigonométriques. In: Deligne, P. (ed.) Cohomologie Etale (SGA $4\frac{1}{2}$). Lect. Notes Math., vol. 569, pp. 168–232. Berlin-Heidelberg-New York: Springer 1977

8. Dwork, B.: On the zeta function of a hypersurface. Publ. Math. Inst. Hautes Etud. Sci. **12**, 5–68 (1962)
9. Dwork, B.: On the zeta function of a hypersurface, II. Ann. Math. **80**, 227–299 (1964)
10. Ehrhart, E.: Sur un problème de géométrie diophantienne linéare, I: Polyèdres et réseaux. J. Reine Angew. Math. **226**, 1–29 (1967)
11. Ehrhart, E.: Démonstration de la loi de réciprocité du polyèdre rationnel. C.R. Acad. Sci. Paris Sér. A-B **265**, A91–A94 (1967)
12. Kouchnirenko, A.G.: Polyèdres de Newton et nombres de Milnor. Invent. Math. **32**, 1–31 (1976)
13. Serre, J-P.: Endomorphismes complètement continus des espaces de Banach p-adiques. Publ. Math. Inst. Hautes Etud. Sci. **12**, 69–85 (1962)
14. Sperber, S.: p-Adic hypergeometric functions and their cohomology, I. Duke Math. J. **44**, 535–589 (1977)

Oblatum 24-VI-1986

Invent. math. 88, 571-602 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

L-groups of crystallographic groups

Masayuki Yamasaki

Department of Mathematics, Josai University, Sakado, Saitama, Japan

Introduction

In their paper [6], Farrell and Hsiang proved the following:

Theorem. *Let N^n be a closed connected flat Riemannian manifold where $n \neq 3, 4$ and let M^n be an aspherical manifold such that $\pi_1(M^n)$ is isomorphic to $\pi_1(N^n)$, then M^n and N^n are homeomorphic.*

This is equivalent to the statement that the structure set $\mathcal{S}_{\mathrm{TOP}}(N)$ consists of a single element. The structure set appears in the surgery exact sequence [16]: $\ldots \to [\Sigma N, G/\mathrm{TOP}] \to L_{n+1}(\pi_1 N) \to \mathcal{S}_{\mathrm{TOP}}(N) \to [N, G/\mathrm{TOP}] \to L_n(\pi_1 N)$, and the result is proved by showing that the homomorphisms $[\Sigma^i N, G/\mathrm{TOP}] \to L_{n+i}(\pi_1 N)$ are bijections.

We would like to prove a similar result for certain stratified spaces. Let Γ be a crystallographic group acting on $\mathbb{R}^n$ and consider the orbit space $\mathbb{R}^n/\Gamma$. If Γ is torsion-free, $\mathbb{R}^n/\Gamma$ is an n-dimensional closed flat Riemannian manifold, and Farrell and Hsiang's result will apply to this. If Γ has torsion, $\mathbb{R}^n/\Gamma$ is a stratified space. The following is our conjecture.

Conjecture. *If a stratified space is homotopy equivalent to $\mathbb{R}^n/\Gamma$ in some nice way, then it is homeomorphic to $\mathbb{R}^n/\Gamma$.*

This paper is the first step toward this conjecture. As with the Farrell-Hsiang theorem this conjecture is approached by showing the functions in appropriate "stratified" exact sequences are bijections. Our main result is a partial computation of one of the terms in these exact sequences, specifically the L-groups of Γ. Actually we need to do the computation for a slightly larger class of groups. Let Λ be a finitely generated group which maps onto a crystallographic group Γ of rank n with a finite kernel; in other words, Λ acts by isometries on $\mathbb{R}^n$ discretely, virtually faithfully, with a compact orbit space. Let W_Λ be a free contractible Λ-space. Then the map p: $(\mathbb{R}^n \times W_\Lambda)/\Lambda \to \mathbb{R}^n/\Lambda$ has point inverses $p^{-1}(x) = W_\Lambda/\Lambda_x$, which are classifying spaces for "isotropy" subgroups Λ_x. Quinn has defined Ω-spectra $\mathbb{L}(X)$ whose homotopy groups are the surgery obstruction groups $L_i(\pi_1 X)$ [7]. This functor can be applied fibrewise to obtain a "sheaf" of spectra $\mathbb{L}(p) \to \mathbb{R}^n/\Lambda$, with fibre over x, $\mathbb{L}(P^{-1}(x))$. Next Quinn has defined [9] homology groups with spectral sheaf

coefficients $H_*(\mathbb{R}^n/\Lambda; \mathbb{L}(p))$. For technical reasons we use a definition of $\mathbb{L}$ using the Poincaré chain complex of Ranicki [12, 13]. The homotopy groups are the limits $L_i^{-\infty}$ of Ranicki's lower L-theory $L_i^{(-m)}$ [11] which may differ from L_i possibly by 2-torsion. The following is our main theorem.

Theorem (4.11). *If Λ is a finitely generated group which acts by isometries on $\mathbb{R}^n$ discretely, virtually faithfully, with compact quotient, then there is a natural isomorphism modulo 2-torsion:*

$$a\colon\ H_*(\mathbb{R}^n/\Lambda; \mathbb{L}(p)) \to L_*^{-\infty}(\Lambda).$$

The map a is essentially Quinn's assembly map. To explain the calculational significance of the theorem, let us recall that there is an Atiyah-Hirzebruch type spectral sequence [9, Theorem 8.7] $H_i(\mathbb{R}^n/\Lambda; L_j^{-\infty}(p)) \Rightarrow H_{i+j}(\mathbb{R}^n/\Lambda; \mathbb{L}(p))$. Here $L_j^{-\infty}(p)$ is obtained by applying π_j fibrewise to $\mathbb{L}(p)$, or equivalently applying $L_j^{-\infty}$ to the sheaf of isotropy subgroups Λ_x. Since the isotropy subgroups are finite, the theorem provides a "calculation" of $L^{-\infty}(\Lambda)$ in terms of $L^{-\infty}$ of finite subgroups.

This is the first systematic calculation of surgery obstruction groups for a class of infinite groups with torsion. Previously purely algebraic techniques have been mostly limited to finite groups, and geometric techniques limited to torsion free groups. Here we refine the geometric techniques to reduce calculations for these infinite groups to those for finite groups, which are accessible to the algebra.

The organization of this paper is as follows. In § 1, we prove the splitting lemma for quadratic Poincaré complexes and pairs. In § 2, we introduce quadratic complexes with geometric control in the sense of [8–10]. Using the results in § 1, we prove the stable splitting lemma for geometric quadratic Poincaré complexes and pairs (2.5) and the stable splitting lemma over a manifold (2.11). In § 3, we construct the spectra $\mathbb{L}(X; p)$ and $\mathbb{H}(X; \mathbb{L}(p))$ for stratified systems of fibrations $p\colon E \to X$ [9], define the assembly maps, and prove the characterization theorem (3.9), which characterizes elements of homology groups as certain objects with geometric control. At the same time the shrinking lemma (3.10) can be proved. The key ingredient of the proof of these is an application of the stable splitting lemma of the previous section. We also show in this section that the homotopy groups of $\mathbb{L}(*; B_\Lambda \to *)$ are the limits of Ranicki's lower L-groups of Λ, where B_Λ is the classifying space of Λ. In § 4, the main theorem (4.11) is proved. The proof is like that in [6]. We use induction on the "size" of the action of Λ (= the size of Γ), Farrell and Hsiang's structure theorem of crystallographic groups (4.2), and Dress's induction theory [2]. When the induction on the size of Γ does not work, we need to use the characterization theorem and the shrinking lemma.

This is a work done basically in the author's Ph.D thesis at Virginia Polytechnic Institute and State University. I would like to thank Professor F.S. Quinn for suggesting this problem to me and for providing many invaluable comments.

1. Splitting lemma for quadratic Poincaré complexes and pairs

In this section, we give a sufficient condition to splitting quadratic Poincaré pairs defined by Ranicki. We use notations and sign conventions used in [13]. R denotes a ring with 1 and an involution. Let $c=(f\colon C\to D, (\delta\psi,\psi))$ be an $(n+1)$-dimensional quadratic Poincaré pair over R, and C' (resp. D') be an n-dimensional subcomplex of C (resp. an $(n+1)$-dimensional subcomplex of D). We assume that

(1.1) C/C' is n-dimensional, D/D' is $(n+1)$-dimensional, and

(1.2) $f(C')\subset D'$.

Let i_C denote the inclusion map of C' into C, and p_C denote the projection map of C onto C/C'. We fix splittings $j_C\colon C/C'\to C$ and $q_C\colon C\to C'$ of p_C and i_C. These give an identification of C with $C'\oplus(C/C')$, and if we define a chain map $\rho_C\colon (C/C')_r\to(SC')_r=C'_{r-1}$ by $(-)^{r-1}q_C dj_C$, then the boundary map of C is given by a matrix

$$\begin{pmatrix} d & (-)^{r-1}\rho_C \\ 0 & d \end{pmatrix}\colon C'_r\oplus(C/C')_r\to C'_{r-1}\oplus(C/C')_{r-1}$$

under this identification. Here S denotes the suspension of a chain complex. Maps i_D, p_D, j_D, q_D, ρ_D are defined in the same way for D.

Let C'' denote the chain complex $(C/C')^{n-*}$. By the assumption 1.2, f induces a chain map $f'=q_D f i_C\colon C'\to D'$ such that $i_D f'=f i_C$. The algebraic mapping cone $C(f')$ of f' is a subcomplex of the algebraic mapping cone $C(f)$ of f. Define D'' by $(C(f)/C(f'))^{n+1-*}$, or equivalently $C(p_D f j_C)^{n+1-*}$. There is a chain map (inclusion map)

$$f''=\begin{pmatrix}0\\1\end{pmatrix}\colon C''_r=(C/C')^{n-r}\to D''_r=(D/D')^{n+1-r}\oplus(C/C')^{n-r}$$

from C'' to D''. Recall that $(1+T)\psi_0\colon C^{n-*}\to C$, and $((1+T)\delta\psi_0, (-)^r(1+T)\psi_0 f^*)\colon D^{n+1-r}\to C(f)_r=D_r\oplus C_{r-1}$ are chain equivalences. Here $*$ denotes the dual.

(1.3) **Lemma** *(Splitting lemma for quadratic Poincaré pairs). Let c, C', D', C'', D'' be as above. We further assume that*

$$H_i(p_C(1+T)\psi_0 p_C^*)=0 \quad \text{for } i<0, \tag{1.4}$$

and

$$H_i(p_D((1+T)\delta\psi_0 p_D^*,\ f(1+T)\psi_0 p_C^*))=0 \quad \text{for } i<0. \tag{1.5}$$

Then there are two adjoining $(n+1)$-dimensional quadratic Poincaré triads over R:

$$\left(\mathscr{T}_1\colon \begin{array}{ccc} B & \longrightarrow & C^! \\ \downarrow & \searrow & \downarrow \\ C' & \longrightarrow & D' \end{array}, \Psi_1\right) \quad \text{and} \quad \left(\mathscr{T}_2\colon \begin{array}{ccc} B & \longrightarrow & C'' \\ \downarrow & \searrow & \downarrow \\ C^! & \longrightarrow & D'' \end{array}, \Psi_2\right)$$

whose union is homotopy equivalent to c.

Proof. We give explicit matrices describing the quadratic structures and the equivalence. This will be necessary later when we will have to check that these satisfy certain size estimate. B and $C^!$ in the lemma are defined to be $S^{-1}C((i_C, (1+T)\psi_0 p_C^*)\colon C' \oplus C'' \to C)$ and $S^{-1}C((i_D, (1+T)\delta\psi_0 p_D^*, f(1+T)\psi_0 p_C^*)\colon D' \oplus D'' \to D)$ respectively. Here S^{-1} denotes the desuspension. We have a chain map $g^! = f' \oplus f \oplus f''\colon B_r = C'_r \oplus C_{r+1} \oplus C''_r \to C^!_r = D'_r \oplus D_{r+1} \oplus D''_r$. Notice that, since B and $C^!$ are chain equivalent to $S^{-1}C(p_C(1+T)\psi_0 p_C^*)$ and $S^{-1}C(p_D((1+T)\delta\psi_0 p_D^*, f(1+T)\psi_0 p_C^*))$, the assumptions 1.4 and 1.5 imply that B and $C^!$ are chain equivalent to non-negative chain complexes. We define an n-dimensional quadratic structure $(\delta\bar{\psi}, -\bar{\psi})$ on the pair $g^!\colon B \to C^!$ as follows:

$$\delta\bar{\psi}_s = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (-)^{n-r-s} T\delta\psi_{s-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (s>1)$$

$$: C'^{n-r-s} = D'^{n-r-s} \oplus D^{n+1-r-s} \oplus C''^{n-r-s} \to C^!_r = D'_r \oplus D_{r+1} \oplus D''_r$$

$$\delta\bar{\psi}_0 = \begin{pmatrix} 0 & q_D(1+T)\delta\psi_0 & (-)^{nr+r}\rho' \\ 0 & 0 & (-)^{nr+r+1} j' \\ 0 & 0 & 0 \end{pmatrix}$$

$$: C'^{n-r} = D'^{n-r} \oplus D^{n+1-r} \oplus D''^{n-r} \to C^!_r = D'_r \oplus D_{r+1} \oplus D''_r$$

$$\bar{\psi}_s = \begin{pmatrix} 0 & 0 & 0 \\ 0 & (-)^{n-r-s-1} T\psi_{s-1} & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (s>1)$$

$$: B^{n-1-r-s} = C'^{n-1-r-s} \oplus C^{n-r-s} \oplus C''^{n-1-r-s} \to B_r = C'_r \oplus C_{r+1} \oplus C''_r$$

$$\bar{\psi}_0 = \begin{pmatrix} 0 & q_C(1+T)\psi_0 & (-)^{nr}\rho_C \\ 0 & 0 & (-)^{nr+1} j_C \\ 0 & 0 & 0 \end{pmatrix}$$

$$: B^{n-1-r} = C'^{n-1-r} \oplus C^{n-r} \oplus C''^{n-1-r} \to B_r = C'_r \oplus C_{r+1} \oplus C''_r,$$

where $\rho' = (\rho_D, q_D f j_C)\colon D''^{n-r} = (D/D')_{r+1} \oplus (C/C')_r \to D'_r$ and $j' = (j_D, 0)\colon D''^{n-r} \to D_{r+1}$. A direct calculation shows that this is Poincaré. And the duality implies that B is $(n-1)$-dimensional and $C^!$ is n-dimensional. Thus $(g^!\colon B \to C^!)$ is an n-dimensional quadratic Poincaré pair. Now the triads in the lemma are given by

$$\left(\mathscr{T}_1\colon\ \begin{array}{ccc} B & \xrightarrow{g'} & C' \\ {\scriptstyle g'_C}\downarrow & \circlearrowright 0 & \downarrow{\scriptstyle g'_D} \\ C' & \xrightarrow[f']{} & D' \end{array},\ \Psi_1=(\bar{\psi},0,-\delta\bar{\psi},0)\right)$$

$$\left(\mathscr{T}_2\colon\ \begin{array}{ccc} B & \xrightarrow{g''_C} & C'' \\ {\scriptstyle g^!}\downarrow & \circlearrowright 0 & \downarrow{\scriptstyle f''} \\ C^! & \xrightarrow[g''_D]{} & D'' \end{array},\ \Psi_2=(\bar{\psi},\delta\bar{\psi},0,0)\right)$$

where $g'_C=(-1,0,0)$, $g'_D=(-1,0,0)$, $g''_C=(0,0,1)$, $g''_D=(0,0,1)$. These are $(n+1)$-dimensional quadratic Poincaré triads. There union is

$$\begin{aligned}(F=f'\oplus g^!\oplus(-f'')\colon (C'\cup_B C'')_r &= C'_r\oplus B_{r-1}\oplus C''_r\to(D'\cup_{C^!}D'')_r\\ &= D'_r\oplus C^!_{r-1}\oplus D''_r,\qquad (0\cup_{\delta\bar{\psi}}0,0\cup_{\bar{\psi}}0))\end{aligned}$$

where $C'\cup_B C''$ and $D'\cup_{C^!}D''$ are push-outs of

$$C' \xleftarrow{g'_C} B \xrightarrow{g''_C} C''$$

$$D' \xleftarrow{g'_D} C^! \xrightarrow{g''_D} D'',$$

i.e., $C({}^t(g'_C,g''_C))$ and $C({}^t(g'_D,g''_D))$ respectively, and

$$\begin{aligned}(0\cup_{\delta\bar{\psi}}0)_s=0\oplus 0\oplus\delta\psi_s\oplus 0\oplus 0\colon\ & D'^{n+1-r}\oplus D'^{n-r}\oplus D^{n+1-r}\oplus D''^{n-r}\oplus D''^{n+1-r}\\ &\to D'_r\oplus D'_{r-1}\oplus D_r\oplus D''_{r-1}\oplus D''_r\\ (0\cup_{\bar{\psi}}0)_s=0\oplus 0\oplus\psi_s\oplus 0\oplus 0\colon\ & C'^{n-r}\oplus C'^{n-1-r}\oplus C^{n-r}\oplus C''^{n-1-r}\oplus C''^{n-r}\\ &\to C'_r\oplus C'_{r-1}\oplus C_r\oplus C''_{r-1}\oplus C''_r.\end{aligned}$$

Chain equivalences

$${}^t(0,0,1,0,0)\colon\ C\to C'\cup_B C'',\ \text{and}$$

$${}^t(0,0,1,0,0)\colon\ D\to D'\cup_{C^!}D''$$

make the following diagram commute:

$$\begin{array}{ccc} C & \xrightarrow{f} & D \\ \downarrow & & \downarrow \\ C'\cup_B C'' & \xrightarrow[F]{} & D'\cup_{C^!}D'', \end{array}$$

and

$$^t(0,0,1,0,0)\,\psi_s(0,0,1,0,0)=(0\cup_{\bar{\psi}}0)_s$$
$$^t(0,0,1,0,0)\,\delta\psi_s(0,0,1,0,0)=(0\cup_{\delta\bar{\psi}}0)_s,$$

for each $s\geqq 0$. Therefore the union is homotopy equivalent to the original quadratic Poincaré pair, and the lemma is proved. □

(1.6) *Remarks.* (1) If c is strictly n-dimensional, then the conditions 1.4 and 1.5 are satisfied when the images of C'^n and D'^{n+1} by $(1+T)\psi_0 q_C^*$ and $(1+T)\delta\psi_0 q_D^*$ lie in C'_0 and D'_0, respectively.

(2) If C, C', D', C/C', D/D' are free, then the above argument can be carried through in the category of free chain complexes.

(3) If the splitting of the boundary is already given, we can construct a splitting with the given splitting of the boundary.

2. Geometric chain complexes and the stable splitting lemma

Geometric modules, geometric morphisms, and geometric chain complexes are defined in [10, § 1]. Since these are essential to this paper, we review them at first. Let $p\colon E\to X$ be a map, where X is a metric space. This is called the control map.

(2.1) **Definitions** (Quinn). A *geometric* $\mathbf{Z}$*-module on* E *(generated by a set* S*)* is a free module $\mathbf{Z}[S]$ together with a map of the basis $f\colon S\to E$. We usually require that geometric $\mathbf{Z}$-modules be *locally finite* in the sense that every point in E has a neighborhood whose preimage by f in S is finite.

A *geometric morphism* $h\colon \mathbf{Z}[S]\to\mathbf{Z}[T]$ of geometric $\mathbf{Z}$-modules with $f\colon S\to E$ and $g\colon T\to E$ is a sum $\Sigma m_j\rho_j^{(x,y)}$, where $m_j\in\mathbf{Z}$, $x\in S$, $y\in T$, and $\rho_j^{(x,y)}$ is a path: $[0,t_j^{(x,y)}]\to E$ which starts at $f(x)$ and ends at $g(y)$. Here $t_j^{(x,y)}$ is a non-negative real number. We require that the sum is locally finite in the sense that for each $x\in S$ there are only finitely many paths $\rho_j^{(x,y)}$ with non-zero coefficient, and for each $y\in T$ there are only finitely many paths $\rho_j^{(x,y)}$ with non-zero coefficient. In a morphism we allow a deletion or an insertion of a path with coefficient 0. Morphisms are composed by composing the component paths and multiplying coefficients: let $h=\Sigma m_j\rho_j^{(x,y)}\colon \mathbf{Z}[S]\to\mathbf{Z}[T]$, $k=\Sigma n_i\mu_i^{(y,8)}\colon \mathbf{Z}[T]\to\mathbf{Z}[U]$, then $kh=\Sigma n_i m_j\mu_i^{(y,8)}\rho_j^{(x,y)}$. Here the Moore composition of paths is used, and we write compositions of paths from right to left.

A *homotopy* ($\sim$) of a morphism is obtained by changing all the paths in the morphism by homotopy keeping the endpoints fixed.

A *morphism has radius* ε *(in* X*)* if each path with non-zero coefficient in the morphism has image in X (via p) inside the closed ball of radius ε about its starting point. Similarly, a *homotopy of morphism has radius* ε if the homotopy of each path has image in X inside the closed ball of radius ε about the starting point of the path. The symbol $\underset{\varepsilon}{\sim}$ denotes a homotopy of radius ε.

Suppose W is a subset of X. The *restriction* $\mathbf{Z}[S]|W$ *of a geometric module* $\mathbf{Z}[S]$, $f\colon S\to E$, *to* W is $\mathbf{Z}[(pf)^{-1}(W)]$. The *restriction* $h|W$ *of a morphism* h is obtained by throwing away all the paths with initial point outside of $p^{-1}(W)$.

A geometric morphism $h\colon \mathbf{Z}[S]\to\mathbf{Z}[T]$ of radius ε is an *ε-isomorphism (with support W)* if there is a geometric morphism $k\colon \mathbf{Z}[T]\to\mathbf{Z}[S]$ of radius ε such that there are ε homotopies (=homotopies of radius ε) $kh\underset{\varepsilon}{\sim}1$, $hk\underset{\varepsilon}{\sim}1 (kh|W\underset{\varepsilon}{\sim}1,\ hk|W\underset{\varepsilon}{\sim}1)$. Here 1's denote the "identity" morphisms made up of appropriate constant paths with coefficient 1. According to this definition, an ε-isomorphism with support W may *not* be an ε-isomorphism. It is an ε-isomorphism if $W=X$.

(2.2) *Remarks.* (1) We sometimes do not mention the map of the basis $f\colon S\to E$ of a geometric $\mathbf{Z}$-module $\mathbf{Z}[S]$ and pretend as if the basis points were in E.

(2) Suppose E has a universal cover $\tilde{E}$. Then by taking the pull-back $\tilde{S}\to E$ of geometric $\mathbf{Z}$-module $\mathbf{Z}[S]$ on E, we get a geometric $\mathbf{Z}$-module on $\tilde{E}$, which is naturally a free $\mathbf{Z}[\pi_1 E]$-module generated by S. A geometric morphism between geometric $\mathbf{Z}$-modules on E induces a $\mathbf{Z}[\pi_1 E]$-module homomorphism.

(2.3) **More Definitions.** A *geometric $\mathbf{Z}$-module chain complex C on E (with support $K\subset X$)* is a sequence of morphisms of geometric $\mathbf{Z}$-modules on E:

$$C\colon \ldots\to C_i\xrightarrow{d_i} C_{i-1}\xrightarrow{d_{i-1}}\ldots$$

such that $d_{i-1}d_i\sim 0 (d_{i-1}d_i|K\sim 0)$. Here 0 denotes the empty morphism. Such C *has radius* ε if all d_i's have radius ε and there are homotopies $d^2\sim 0 (d^2|K\sim 0)$ of radius ε. The word "support" may not give a correct impression. It does *not* mean that C lies over K, but it means that, roughly speaking, C is a chain complex when restricted to K and may *not* be so outside of K. Thus C is a geometric chain complex if it has support X. Similar remarks will apply to the followings.

A *chain map* f (with support K) between geometric $\mathbf{Z}$-module chain complexes C, D (with support K) consists of geometric morphisms $f_i\colon C_i\to D_i$ such that $df\sim fd (df|K\sim fd|K)$. It has *radius* ε if all f_i's have radius ε and there are homotopies $df\sim fd (df|K\sim fd|K)$ of radius ε.

A *chain homotopy (with support $K\subset X$)* between two chain maps $f,g\colon C\to D$ (with support K) is a collection $\{H_i\}$ of geometric morphisms $H_i\colon C_i\to D_{i+1}$ such that $dH_i+H_{i-1}d\sim f_i-g_i ((dH_i+H_{i-1}d)|K\sim (f_i-g_i)|K)$ for all i. It has radius ε if all the H_i's have radius ε and there are ε homotopies $dH_i+H_{i-1}d\underset{\varepsilon}{\sim}f_i-g_i ((dH_i+H_{i-1}d)|K\underset{\varepsilon}{\sim}(f_i-g_i)|K)$.

A chain map $f\colon C\to D$ (with support K) is a *chain equivalence (with support K)* if there are a chain map $g\colon D\to C$ (with support K) and chain homotopies $gf\sim 1$ and $fg\sim 1$ (with support K). It is an *ε-chain equivalence (with support K)* if f, g and the two chain homotopies have radius ε.

A *chain contraction of C with support $K\subset X$ (of radius ε)* is a chain homotopy $\{s_i\colon C_i\to C_{i+1}\}$ with support K between 1 and 0: $C\to C$ (of radius ε).

The *dual* $\mathbf{Z}[S]^*$ of $\mathbf{Z}[S]$ is $\mathbf{Z}[S]$ itself. When we have a geometric morphism $h=\Sigma m_j\rho_j^{(x,y)}\colon \mathbf{Z}[S]\to\mathbf{Z}[T]$, its *dual* $h^*\colon \mathbf{Z}[T]^*\to\mathbf{Z}[S]^*$ is defined to be $h^*=\Sigma m_j\mu_j^{(y,x)}$, where $\mu_j^{(y,x)}(t)=\rho_j^{(x,y)}(t_j^{(x,y)}-t)$.

The *tensor product over* $\mathbb{Z}$ of a geometric $\mathbb{Z}$-module $\mathbb{Z}[S]$, $f\colon S\to E$, on E and a geometric $\mathbb{Z}$-module $\mathbb{Z}[T]$, $g\colon T\to F$, on F, denoted $\mathbb{Z}[S]\otimes\mathbb{Z}[T]$, is a geometric $\mathbb{Z}$-module $\mathbb{Z}[S\times T]$, $f\times g\colon S\times T\to E\times F$, on $E\times F$. The *tensor product* $h\otimes k$ of geometric morphisms $h=\Sigma m_j\rho_j^{(x,x')}\colon \mathbb{Z}[S]\to\mathbb{Z}[S']$ (on E) and $k=\Sigma n_i\mu_i^{(y,y')}\colon \mathbb{Z}[T]\to\mathbb{Z}[T']$ (on F) is defined to be $\Sigma m_j n_i\rho_j^{(x,x')}\otimes\mu_i^{(y,y')}$, where $\rho_j^{(x,x')}\otimes\mu_i^{(y,y')}$ is a path from (x,y) to (x',y') in $E\times F\colon[0,t_j+u_i]\to E\times F$ sending t to $(\rho_j^{(x,x')}(t),y)$ if $0\leqq t\leqq t_j$ and to $(x',\mu_i^{(y,y')}(t-t_j))$ if $t_j\leqq t\leqq t_j+u_i$ ($\rho_j^{(x,x')}\colon[0,t_j]\to E$ and $\mu_i^{(y,y')}\colon[0,u_i]\to F$). (See Remark 2.2(1).) The morphism $(hh')\otimes(kk')$ is homotopic to $(h\otimes k)(h'\otimes k')$.

In § 1, we studied quadratic Poincaré complexes. Now we define geometric quadratic Poincaré complexes. Again $p\colon E\to X$ is the control map.

(2.4) **Definitions.** An *n-dimensional geometric* $\mathbb{Z}$*-module quadratic complex* $c=(C,\psi)$ *on* E with support K consists of the underlying strictly n-dimensional geometric $\mathbb{Z}$-module chain complex C on E with support K and a quadratic structure $\psi=\{\psi_s|s\geqq 0\}$. Here ψ_s is a set of geometric morphisms: $C^{n-r-s}=(C_{n-r-s})^*\to C_r$ ($r\in\mathbb{Z}$) such that

$$(*)\qquad [d\psi_s+(-)^r\psi_s d^*+(-)^{n-s-1}(\psi_{s+1}+(-)^{s+1}T\psi_{s+1})]|K\sim 0 : C^{n-r-s-1}|K\to C_r \quad (s\geqq 0).$$

T above sends $f\colon C^p\to C_q$ to $(-)^{pq}f^*\colon C^q\to C_p$. Such a complex c is *Poincaré* if

$$(1+T)\psi_0\colon C^{n-*}\to C$$

is a chain equivalence with support K. When the support is X, we do not mention the support. It *has radius* ε if C, all the morphisms ψ_s, and all the homotopies (*) above have radius ε (and $(1+T)\psi_0$ is an ε-chain equivalence with support K when it is Poincaré).

A *map* (resp. *homotopy equivalence*) of n-dimensional geometric $\mathbb{Z}$-module quadratic complexes on E (with support K)

$$f\colon (C,\psi)\to(C',\psi')$$

is a chain map (resp. chain equivalence) $f\colon C\to C'$ (with support K) such that

$$(**)\qquad f\psi_s f^*\sim\psi'_s(f\psi_s f^*|K\sim\psi'_s|K).$$

A homotopy equivalence f is an *ε-homotopy equivalence (with support K)* if $f\colon C\to C'$ is an ε-chain equivalence (with support K) and the homotopies (**) have radius ε.

Similarly, *geometric symmetric Poincaré complexes (of radius ε)* and *geometric quadratic* or *symmetric Poincaré pairs* and *triads (of radius ε)* can be defined by replacing everything in the standard definition by geometric objects (of radius ε) and replacing the necessary identities by homotopies (of radius ε). If M is a PL manifold of dimension m, then a PL triangulation of M produces an m-dimensional geometric $\mathbb{Z}$-module symmetric Poincaré complex on M. See Ranicki [13] and Quinn [10]. This will be denoted $\sigma^*(M)$. *Unions* and *tensor products* of geometric complexes and pairs are defined by the same formulae as

the standard cases. Then the following is the main technical result of this section.

(2.5) **Lemma.** (Stable splitting lemma for geometric quadratic Poincaré pairs). *Fix n and p: $E \to X$. Given any $\varepsilon > 0$, there exists $\delta > 0$ such that for any $(n+1)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré pair c on E of radius δ in X and a subset Y of X, the $(n+2)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré pair $\sigma^*(S^1) \otimes c$ on $S^1 \times E$ is ε-homotopy equivalent to a union of two $(n+2)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré triads of radius ε lying over Y^ε and $X - Y^{-\varepsilon}$ respectively. Here we use the control map $p' = (\text{constant map}) \times p$: $S^1 \times E \to \{pt.\} \times X = X$.*

To prove this, we need the following local "folding" argument (cf. [1] § 14).

(2.6) **Lemma.** *If a geometric $\mathbb{Z}$-module chain complex (C, d) on E of radius δ in X has a chain contraction with support $W (\subset X)$ of radius δ and satisfies the following:*

$$C_l | W = 0 \quad \text{for } l < k,$$

then C is 2δ-chain equivalent to a geometric $\mathbb{Z}$-module chain complex (C', d') on E of radius δ in X such that

(1) $C'_l | W^{-\delta} = 0$ *for* $l \leqq k$,
(2) $C' | X - W = C | X - W$, *and*
(3) *there exists a chain contraction of C' with support $W^{-2\delta}$ of radius 3δ.*

Proof. Let $\{s_i\}$ denote the chain contraction of C with support W, and let i, j, r, q denote the following canonical inclusion morphisms and projection morphisms:

$$C_k | X - W^{-\delta} \underset{r}{\overset{i}{\leftrightarrows}} C_k \underset{j}{\overset{q}{\leftrightarrows}} C_k | W^{-\delta}.$$

We have homotopies $d_{k+1} s_k j \sim j$ and $d_k j \sim 0$. Define (C', d') as follows: $C'_l = C_l$ $(l \neq k,\ k+2)$, $C'_k = C_k | X - W^{-\delta}$, $C'_{k+2} = C_{k+2} \oplus C_k | W^{-\delta}$; $d'_k = d_k i$, $d'_{k+1} = p\, d_{k+1}$, $d'_{k+2} = (d_{k+2}, s_k j)$, $d'_{k+3} = {}^t(d_{k+3}, 0)$, $d'_l = d_l$ otherwise. Let T be a chain complex with $T_{k+2} = T_{k+1} = C_k | W^{-\delta}$, $T_l = 0$ otherwise, and $d_T = 1$: $T_{k+2} \to T_{k+1}$; and let T' be the desuspension of T. There are obvious chain equivalences f: $C \to C \oplus T$ and h: $C' \oplus T' \to C'$. Define a chain map g: $C \oplus T \to C' \oplus T'$ by

$$g_l = 1 \quad \text{if } l \neq k+1$$

$$g_{k+1} = \begin{pmatrix} 1 & s_k j \\ q\, d_{k+1} & 0 \end{pmatrix} : C_{k+1} \oplus C_k | W^{-\delta} \to C_{k+1} \oplus C_k | W^{-\delta}.$$

Since g_{k+1} can be decomposed as

$$g_{k+1} \sim \begin{pmatrix} 1 & 0 \\ q\, d_{k+1} & -1 \end{pmatrix} \begin{pmatrix} 1 & s_k j \\ 0 & 1 \end{pmatrix}$$

g is a δ-isomorphism. Composing f, g, and h, we obtain a 2δ-chain equivalence between C and C'. A desired chain contraction $\{s'_l\}$ of C' is defined by $s'_l = 0$ for $l \leqq k$, $s'_{k+1} = {}^t(s_{k+1},\ q\, d_{k+1})$, $s'_{k+2} = (s_{k+2},\ -s_{k+2} s_{k+1} s_k j)$, and $s'_l = s_l$ for $l \geqq k+3$. □

Using this lemma repeatedly, we get the following:

(2.7) **Corollary.** *Fix n and p. Given any $\varepsilon>0$, there exists $\delta>0$ such that for any n-dimensional geometric $\mathbb{Z}$-module chain complex C on E of radius δ, C is ε chain equivalent to an n-dimensional geometric $\mathbb{Z}$-module chain complex C' on E of radius ε satisfying:*

(1) $C'_0|W^{-\varepsilon}=\ldots=C'_{n-2}|W^{-\varepsilon}=0$

(2) $C|X-W=C'|X-W$

(3) *there exists a geometric morphism $s\colon C'_{n-1}\to C'_n$ of radius ε such that*

$$d_n s|W^{-\varepsilon}\underset{\varepsilon}{\sim}1 \quad \text{and} \quad s d_n|W^{-\varepsilon}\underset{\varepsilon}{\sim}1.$$

It is in general impossible to finish this elimination in the last two layers, but if we "stabilize" everything, we can avoid the difficulty as follows. Let (A,d_A) be the underlying geometric $\mathbb{Z}$-module chain complex on S^1 of $\sigma^*(S^1)$: $0\to A_1\to A_0\to 0$. Then $A\otimes C'$ is a strictly $(n+1)$-dimensional $\mathbb{Z}$-module chain complex on $S^1\times E$ of radius ε in X:

$$0\to A_1\otimes C'_n\to(A_0\otimes C'_n)\oplus(A_1\otimes C'_{n-1})\to(A_0\otimes C'_{n-1})\oplus(A_1\otimes C'_{n-2})\to$$
$$\begin{pmatrix}(-)^n d_A^n\otimes 1\\ 1\otimes d_n\end{pmatrix} \qquad \begin{pmatrix}1\otimes d_n & (-)^n d_A\otimes 1\\ 0 & 1\otimes d_{n-1}\end{pmatrix}$$

and $(A\otimes C')_i|W^{-\varepsilon}=0$ if $i\leqq n-2$. The morphisms

$$(1\otimes s'_i)\oplus(1\otimes s'_{i-1})\colon (A_0\otimes C'_i)\oplus(A_1\otimes C'_{i-1})\to(A_0\otimes C'_{i+1})\oplus(A_1\otimes C'_i),$$

where $s'_{n-1}=s$, and $s'_i=0$ if $i\neq n-1$, define a chain contraction of $A\otimes C'$ with support $W^{-\varepsilon}$ and radius ε. Use 2.6 again to get a strictly $(n+1)$-dimensional geometric $\mathbb{Z}$-module chain complex $\bar{C}$ on $S^1\times E$ of radius ε in X such that $\bar{C}_i|W^{-2\varepsilon}=0$ if $i\leqq n-1$; the boundary map $d\colon \bar{C}_{n+1}\to\bar{C}_n$ is given by the following matrix:

$$d=\begin{pmatrix}(-)^n d_A\otimes 1 & 1\otimes sj\\ 1\otimes d_n & 0\end{pmatrix}$$
$$:(A_1\otimes C'_n)\oplus(A_0\otimes C'_{n-1}|W^{-2\varepsilon})\to(A_0\otimes C'_n)\oplus(A_1\otimes C'_{n-1}),$$

where j is the inclusion morphism of $C'_{n-1}|W^{-2\varepsilon}$ into C'_{n-1}. Now we consider the following ε-isomorphism of $\bar{C}_{n+1}$ to itself:

$$h=\begin{pmatrix}1 & (-)^n d_A\otimes sj\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & -1\otimes sj\\ 0 & 1\end{pmatrix}\begin{pmatrix}1 & 0\\ 1\otimes q\,d_n & 1\end{pmatrix}\begin{pmatrix}1 & 1\otimes sj\\ 0 & 1\end{pmatrix},$$

where q is the projection morphism of C'_{n-1} onto $C'_{n-1}|W^{-2\varepsilon}$. Then $h|W^{-3\varepsilon}\sim d|W^{-3\varepsilon}$. If we replace the boundary map $d\colon \bar{C}_{n+1}\to\bar{C}_n$ by dh^{-1}, we get a new geometric $\mathbb{Z}$-module chain complex $\bar{C}'$, which is ε-isomorphic to $\bar{C}$ and the boundary map dh^{-1} is homotopic to the identity when restricted to $W^{-3\varepsilon}$. Now we can delete $\bar{C}'_{n+1}|W^{-4\varepsilon}$ and $\bar{C}'_n|W^{-4\varepsilon}$ from $\bar{C}'$. Thus we obtain the following:

(2.8) **Lemma.** *Fix n and p. Given any $\varepsilon>0$, there exists $\delta>0$ such that for any strictly n-dimensional geometric $\mathbb{Z}$-module chain complex C on E of radius δ in X which has a chain contraction with support W and radius δ, there exists a strictly $(n+1)$-dimensional geometric $\mathbb{Z}$-module chain complex $\tilde{C}$ on $S^1\times E$ of radius ε in X satisfying*

(1) *$\tilde{C}$ lies over $X-W^{-\varepsilon}$,*
(2) *$\tilde{C}$ is ε-chain equivalent to $A\otimes C$, and*
(3) *$\tilde{C}|X-W=(A\otimes C)|X-W$.*

Proof of 2.5. Suppose an $(n+1)$-dimensional pair $(f\colon C\to D, (\delta\psi,\psi))$ of radius δ is given. Define subcomplexes C', D' of C, D as follows:

$$C'_i=C_i|Y^{-(i+1)\delta}$$
$$D'_i=D_i|Y^{-i\delta}.$$

Then 1.1 and 1.2 are automatic, and 1.4 and 1.5 are also satisfied. (See Remark 1.6(1).) So formulae in the previous section will give a splitting into two $(n+1)$-dimensional triads. The only defect is that the common boundary pieces B and C' of Lemma 1.3 may lie all over X. It is easy to cut off the portion of B lying over $Y^{-n\delta}$; B is chain equivalent to $S^{-1}C(p_C(1+T)\psi_0 p_C^*)$. Similarly for C'. Next notice that $p_C(1+T)\psi_0 p_C^*$ is a chain equivalence with support $X-Y^{\delta}$. The formula on p. 167 of [15] produces a chain contraction of $S^{-1}C(p_C(1+T)\psi_0 p_C^*)$ with support $X-Y^{3\delta}$ and radius 3δ. Now, using Lemma 2.8, we can eliminate the portion of $S^{-1}C(p_C(1+T(\psi_0 p_C^*)$ lying over $X-Y^{\delta'}$ for some $\delta'>0$ which depends only on δ and n. Apply the same argument to C'. This ends the proof of Lemma 2.5. □

The notion of pairs(=2-ads) and triads(=3-ads) naturally extends to "$(k+2)$-ads" (Ranicki [13]) or "higher algebraic bordisms" (Weiss [17]). A $(k+2)$-ad x of chain complexes is a collection of

(1) the *underlying chain complex* $|x|$,
(2) $k+1$ $(k+1)$-ads $\partial_0 x,\dots,\partial_k x$ satisfying

$$\partial_j\partial_i x=\partial_i\partial_{j+1}x \quad \text{if } 0\leqq i\leqq j<k,$$

(3) a pair $\bigcup\partial_i x\to|x|$.

Let x be a $(k+2)$-ad. When $i_1<\dots<i_l$, we define $\partial_{\{i_1,\dots,i_l\}}x$ by $\partial_{i_1}\partial_{i_2}\dots\partial_{i_l}x$.

An *n-dimensional quadratic Poincaré $(k+2)$-ad* x is a $(k+2)$-ad x of chain complexes together with structure maps

$$(\psi_\alpha)_s\colon |\partial_\alpha x|^{n-|\alpha|-r-s}\to|\partial_\alpha x|, \qquad \alpha\subset\{0,1,\dots,k\}$$

such that

(1) $\partial_i x$ is an $(n-1)$-dimensional quadratic Poincaré $(k+1)$-ad for each i, and
(2) $(\bigcup\partial_i x\to|x|, (\psi_\phi, \bigcup\psi_i))$ is an n-dimensional quadratic Poincaré pair.

Here $|\alpha|$ denotes the size of the subset α.

Let X be a metric space and $p\colon E\to X$ be a map. If we use geometric $\mathbb{Z}$-module chain complexes on E, we can define *geometric $\mathbb{Z}$-module quadratic Poincaré n-ads on E*. Such a thing has *radius* ε (resp. *support* $K\subset X$) if all the

involved chain complexes, chain maps, chain homotopies, and structure maps are replaced by those of radius ε in X (resp. support K). An n-ad x is called *special* if $\partial_{\{0,1,\ldots,n-2\}}x=0$. Recall that an "$n$-ad" in the usual sense is a topological space together with $n-1$ subsets. We call this a *topological n-ad* to distinguish it from n-ads of (quadratic) chain complexes. Let $(E,\partial_* E)$ be a topological n-ad. A *geometric $\mathbb{Z}$-module quadratic Poincaré n-ad x on* $(E,\partial_* E)$ is a geometric $\mathbb{Z}$-module quadratic Poincaré n-ad on E such that $\partial_i x$ lies over $\partial_i E$ for each i.

Let us consider the following problem. Let M be an n-dimensional compact manifold with a PL-triangulation L, and fix a metric on M. Also let $p\colon E\to M$ be a map. Suppose each n-simplex Δ of L is given a geometric $\mathbb{Z}$-module quadratic Poincaré $(n+2)$-ad x_Δ of dimension m on a topological $(n+2)$-ad $(p^{-1}(\Delta), p^{-1}(\partial_*\Delta))$ such that

(2.9) (compatibility) if two distinct n-simplices Δ and Δ' have a common $(n-1)$-face, $\partial_i\Delta=\partial_j\Delta'$, then $\partial_i x_\Delta=-\partial_j x_{\Delta'}$.

We would like to glue all these $(n+2)$-ads to get an m-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré pair on E. We can also consider a problem of the inverse direction. Notice that there is a small difficulty. Ranicki's formula allows us to glue things only along a codimension 1 boundary piece, so we have to be careful about the order of glueing. There are several ways to avoid this difficulty; we do it in the following way. First we glue locally so that the local blocks behave nicely, and then we glue the blocks (which can be glued in any order). When we split something, we first split it into several blocks so that each block is over a union of simplices in a controlled way, and we split each block into the desired pieces. More precisely, assume that the triangulation L of M is the first barycentric subdivision of another triangulation K. For each vertex v of K, consider its star $S(v)$ in L, or the dual cone. Two such dual cones are either disjoint or meet along codimension 1 cell(s). The glueing and splitting problem over $S(v)$ can be solved by looking at the link $L(v)$ of v in L. Note that $L(v)$ is an $(n-1)$-dimensional sphere and the triangulation is the first barycentric subdivision of another. Thus we can keep on reducing the dimension until the link becomes a circle, and in this case there is an obvious order of 2-simplices and glueing and splitting can be done. Thus we have:

(2.10) **Theorem.** (Glueing over a manifold). *Let L be the barycentric subdivision of a PL-triangulation of a compact n-dimensional manifold M and $p\colon E\to M$ be a map. And suppose each n-simplex Δ is given an m-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special $(n+2)$-ad on $(p^{-1}(\Delta), p^{-1}(\partial_*\Delta))$ which are compatible on common faces (in the sense of* 2.9*). Then one can glue them together to get an m-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré pair on $(E,p^{-1}(\partial M))$.*

(2.11) **Theorem.** (Stable splitting lemma over a manifold). *Let us fix m and $p\colon E\to M$ and let L be the first barycentric subdivision of a PL-triangulation of the compact manifold M. Fix a metric of M. Let ε be any positive number. Then there exists $\delta>0$ such that any m-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré pair on $(E,p^{-1}(\partial M))$ of radius δ can be split into pieces each of which*

has radius ε and lies over an ε-neighborhood of the corresponding simplex of L, after tensored with a sufficiently large number of $\sigma^*(S^1)$'s.

(2.12) *Remark.* If a splitting of the boundary is already given, then the result can be arranged to have the given splitting of the boundary.

3. Surgery spaces and assembly maps

Let $p\colon E \to X$ be a map, where X is a metric space. In this section we first construct an Ω-spectrum $\mathbb{L}(X;p)$ which might be called a "controlled $L^{-\infty}$-theory spectrum". If X is a single point $*$, then $\mathbb{L}(*;\ E \to *)$ is homotopy equivalent to the $L^{-\infty}$-theory spectrum $\mathbb{L}(E)$, whose homotopy groups are the limits $L^{-\infty}_*(\pi_1 E)$ of Ranicki's lower L-groups $L^{(-j)}_*(\pi_1 E)$ [11]. Next we describe homology $H_*(X;\mathbb{L}(p))$. It is a sort of generalized homology with local coefficients, defined by Quinn [9, §8]. Given a space X and a spectrum $\mathscr{S}$, the usual (constant coefficient) generalized homology groups $H_*(X;\mathscr{S})$ is the homotopy groups of an Ω-spectrum $\lim_{n\to\infty} \Omega^n(\mathscr{S}_n \times X/\{\text{base point}\} \times X)$ $= \lim_{n\to\infty} \Omega^n((\bigcup_{x\in X} \mathscr{S}_n \times \{x\})/(\bigcup_{x\in X} \{\text{base point}\} \times \{x\}))$. In our case, we apply the $\mathbb{L}^{-\infty}$-theory spectrum functor $\mathbb{L}(-)$ to each fiber of $p\colon E \to X$, and define $H_*(X;\mathbb{L}(p))$ to be the homotopy groups of an Ω-spectrum $\mathbb{H}(X;\mathbb{L}(p))$ $= \lim_{n\to\infty} \Omega^n((\bigcup_{x\in X} \mathbb{L}_{-n}(p^{-1}(x)) \times \{x\})/(\bigcup_{x\in X} \{\text{base point}\} \times \{x\}))$. Actually, we assume that X is a polyhedron and apply $\mathbb{L}(-)$ blockwise to p and use $\mathbb{L}(p^{-1}(\Delta)) \times \Delta$ as a building block of $\mathbb{H}(X;\mathbb{L}(p))$, where Δ is a simplex of X. When p is sufficiently close to being a fibration, these two approaches are equivalent [ibid.]. Lastly we show that the homology spectrum $\mathbb{H}(X;\mathbb{L}(p))$ is homotopy equivalent to the controlled $L^{-\infty}$-theory spectrum $\mathbb{L}(X;p)$, for certain maps p (Theorem 3.9). This will be used to prove the main theorem (4.11) in section 4.

Now let us begin defining $\mathbb{L}(X;p)$. A point in $\mathbb{L}_n(X;p)$ is, roughly speaking, an $(n+l)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré complexes on $\mathbb{R}^l \times E$ whose sufficiently high suspension (defined below) is cobordant to one with arbitrarily small radius measured in X. For example, an n-dimensional surgery problem between PL-manifolds $M \to N$ together with a map $N \to E$ inducing an isomorphism on π_1 produces such a quadratic Poincaré complex on E; one can make the radius as small as one likes by choosing sufficiently fine PL triangulations. Actually, by successively taking barycentric subdivisions of M and N, one gets a sequence of quadratic Poincaré complexes with radius converging to 0. This defines a "path" in the "space of quadratic Poincaré complexes" $\mathbb{P}_n(X;p)$.

There are cases in which we need a restriction map $\mathbb{L}(X,p) \to \mathbb{L}(W;p|W)$ for an open subset W of X. Note that, when we restrict a geometric quadratic Poincaré complex with support the whole space X to W, the result is generally *not* a quadratic Poincaré complex on $p^{-1}(W)$ with support W because it is damaged near the frontier of W in X. So, when we define $\mathbb{L}(X;p)$ and $\mathbb{P}(X;p)$, we need to consider complexes with support not necessarily the whole space.

This surely complicates the construction. Fortunately, we can simplify the construction, as long as we deal with compact X's. See Remark (3.3) below. One place where we might use restriction maps is the proof of theorem 3.9. There are at least two ways of establishing a homotopy equivalence between $\mathbb{H}(X;\mathbb{L}(p))$ and $\mathbb{L}(X;p)$, one employed in this article and one employed by Quinn in [9]. In Quinn's way, one needs to check that $\mathbb{L}$ satisfy a certain restriction axiom. In this article, theorem 3.9 will be proved without using restriction maps, but the full definition will be given for completeness.

Fix $p\colon E\to X$ and an integer n. A *primitive k-simplex x of degree n with radius ε and support $K\subset X$* is an $(n+k+l)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special $(k+2)$-ad on $\mathbb{R}^l\times E$ with radius ε in X and support $K\subset X$, for some non-negative integer l. The composition $\mathbb{R}^l\times E\to E\to X$ is the control map. We require that x be locally finite and have bounded radius in $\mathbb{R}^l\times X$ via $1\times p$. If x has radius ε and support K, then its faces $\partial_0 x,\dots,\partial_k x$ are primitive $(k-1)$-simplices of degree n with radius ε and support K.

For a primitive k-simplex of degree n with radius ε and support K, we have the following operations.

(1) *Reduction.* Suppose $\varepsilon'\geqq\varepsilon$ and $K'\subset K\subset X$. Then x can be regarded as a primitive k-simplex of degree n with radius ε' and support K'. This is called a reduction of x.

(2) *External suspension.* $\sigma^*(S^1)\otimes x$ gives an $(n+k+l+1)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special $(k+2)$-ad on $S^1\times\mathbb{R}^l\times E$. Lifting everything into the infinite cyclic cover $\mathbb{R}\times\mathbb{R}^l\times E=\mathbb{R}^{l+1}\times E$, we obtain a new primitive k-simplex of degree n. The result is denoted by Σx, and called the external suspension. $\Sigma^m x=\Sigma\Sigma\dots\Sigma x$ has the same radius and support as x.

Now we define the space of quadratic Poincaré ads.

(3.1) **Definition.** $\mathbb{P}_n(X;p)$ is the Δ-set with simplices (which will be called *elaborate simplicies*) defined inductively: an elaborate 0-simplex is a primitive 0-simplex of degree n, i.e., a strictly $(n+l)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special 2-ad (=complex) on $\mathbb{R}^l\times E$ for some l (with unrestricted compact support and radius). An elaborate k-simplex σ consists of an underlying primitive k-simplex $|\sigma|$ of degree n, together with $k+1$ elaborate $(k-1)$-simplices $\partial_0\sigma,\dots,\partial_k\sigma$. We require these to satisfy the usual $\partial_i\partial_j$ identities, and in addition require that the external suspension of a reduction of the underlying primitive $(k-1)$-simplex $|\partial_i\sigma|$ of $\partial_i\sigma$ be equal to the i-th face $\partial_i|\sigma|$ of the underlying primitive k-simplex $|\sigma|$. The *support* and *radius* of an elaborate simplex are those of its underlying primitive simplex.

We are not interested in $\mathbb{P}_n(X;p)$ itself; it is contractible, because there is no restriction on radius and support in the definition. If we required that the support be equal to X and the radius be arbitrary, then we would have the usual $L^{-\infty}$-group spectrum. In the following definition, we are going to put a restriction on both the support and the radius. Suppose K is a compact subset of X, ε is a positive number, and N is a positive integer, then $\mathbb{P}_n(X,K,p,\varepsilon)$ ($\mathbb{P}_n(X,K,p,\varepsilon)^{(N)}$) denotes the subset of $\mathbb{P}_n(X;p)$ made up of all the simplices with support containing K, radius not exceeding ε (and the dimension of the underlying primitive simplex less than or equal to N).

(3.2) **Definition.** We define a Δ-set $\mathbb{L}_n(X;p)$ as follows. Let Δ^k be a k-simplex with the obvious Δ-set structure and let $[0,\infty)$ have the triangulation with vertices at integer points. The obvious ordering of vertices makes $[0,\infty)$ into a Δ-set. A k-simplex of $\mathbb{L}_n(X;p)$ is defined to be a Δ-map $\Delta^k \otimes [0,\infty) \to \mathbb{P}_n(X;p)$ which satisfies the following condition: there are an increasing sequence of compact sets $K_i \subset X$ with $\bigcup$ (interior of K_i) $=X$, a sequence ε_i of numbers monotone decreasing to 0, and a positive integer N, such that the image of $\Delta^k \otimes [i,\infty)$ lies in $\mathbb{P}_n(X, K_i, p, \varepsilon_i)^{(N)}$. Here $\otimes$ denotes the geometric product of Δ-sets [14].

(3.3) *Remarks.* (1) When X is compact, we may replace $\mathbb{P}_n(X, K_i, p, \varepsilon)^{(N)}$ by $\mathbb{P}_n(X, X, p, \varepsilon)^{(N)}$ in the above. In this case, we write $\mathbb{P}'_n(X, p, \varepsilon)$ (resp. $\mathbb{P}'_n(X, p, \varepsilon)^{(N)}$) instead of $\mathbb{P}_n(X, X, p, \varepsilon)$ (resp. $\mathbb{P}_n(X, X, p, \varepsilon)^{(N)}$). The union $\bigcup\{\mathbb{P}'_n(X, p, \varepsilon) | \varepsilon > 0\}$ (resp. $\bigcup\{\mathbb{P}'_n(X, p, \varepsilon)^{(N)}\}$) is denoted $\mathbb{P}'_n(X;p)$ (resp. $\mathbb{P}'_n(X;p)^{(N)}$).

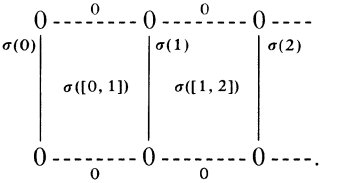
(2) $\mathbb{P}_n(X;p)$ satisfies the Kan condition. As was mentioned above , this fact itslef has no importance at all. What is important is the construction used to prove it. Note that the same proof works for $\mathbb{P}'_n(X;p)$. For example, suppose we have a "horn" of k-simplices of $\mathbb{P}_n(X;p)$, $y_0, \ldots, y_k$, satisfying $\partial_j y_i = -\partial_i y_{i+1}$ $(0 \leqq i \leqq j < k)$. After some necessary suspensions, we can fit $y_0, \ldots, y_k$ together to produce a new k-simplex y_{k+1}. There is an obvious cobordism (product cobordism) between y_{k+1} and itself, and this gives a $(k+1)$-simplex x such that $\partial_i x = y_i$ $(i = 0, 1, \ldots, k+1)$. The radius and support of x depend on those of y_i's $(i = 0, 1, \ldots, k)$. Therefore, $\mathbb{L}_n(X;p)$ also satisfies the Kan condition.

The next result describes the spectrum structure. Unfortunately, our subscripts for $\mathbb{L}$ do not coincide with the indexing for spectra.

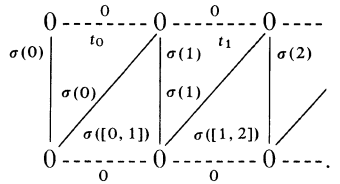
(3.4) **Theorem.** *There is a natural homotopy equivalence* T: $\Omega\mathbb{L}_n(X;p) \to \mathbb{L}_{n+1}(X;p)$.

Proof. A k-simplex of $\Omega\mathbb{L}_n(X;p)$ is a Δ-map σ: $\Delta^k \otimes [0,\infty) \otimes I \to \mathbb{P}_n(X;p)$. We define $T\sigma$: $\Delta^k \otimes [0,\infty) \to \mathbb{P}_{n+1}(X;p)$ as follows. Let τ be an m-simplex of $\Delta^k \otimes [0,\infty)$. $\sigma(\tau \otimes I)$ consists of several simplices of $\mathbb{P}_n(X;p)$. We can fit these together after some necessary suspensions and the result is an m-simplex of $\mathbb{P}_{n+1}(X;p)$, since $\sigma(\tau \otimes 0) = \sigma(\tau \otimes 1) = 0$. This defines a k-simplex $T\sigma$ of $\mathbb{L}_{n+1}(X;p)$.

We will show that T is a homotopy equivalence. First of all, note that each 0-simplex of $\mathbb{P}_{n+1}(X;p)$ can be naturally regarded as a 1-simplex of $\mathbb{P}_n(X;p)$ with two 0 faces, and a 0-simplex σ: $[0,\infty) \to \mathbb{P}_{n+1}(X;p)$ of $\mathbb{L}_{n+1}(X;p)$ can be expressed as in the following picture.



This itself is not a 0-simplex of $\Omega\mathbb{L}_n(X;p)$, since this picture is not triangulated. By inserting trivial cobordisms as in 3.3(2), we can triangulate this picture:



Here t_i denotes the trivial (=product) cobordism between $\sigma(i)$ and itself. Or one can supply a more formal construction: suppose $\sigma([i,i+1]) \subset \mathbb{P}_{n+1}(X, K_i, p, \varepsilon_i)$, then one can easily construct a "simplicial" map $\Delta^1 \otimes [i,i+1] \to \mathbb{P}_n(X, K_i, p, \varepsilon_i)$ by sending one of the two 2-simplices of $\Delta^1 \otimes [i,i+1]$ to the 2-simplex $\sigma([i,i+1])$ whose edges are $\sigma(i)$, $\sigma(i+1)$ and 0, and the other to the 1-simplex $\sigma(i)$. Then use the relative Δ-map approximation theorem of Rourke and Sanderson [14, theorem 5.3] to get a Δ-map. Here we use the fact that $\mathbb{P}_n(X, K_i, p, \varepsilon_i)$ is Kan (3.3). Observe that this gives the same map as above (up to homotopy). Similar argument will be used often later, and will be called the "triangulation argument". Now this defines a 0-simplex $[0,\infty) \otimes I \to \mathbb{P}_n(X; p)$ of $\Omega \mathbb{L}_n(X, p)$. If we apply T to this, then the result is different from the original only by trivial cobordisms; therefore these two can be connected by a 1-simplex of $\mathbb{L}_{n+1}(X; p)$; i.e., T maps into every component. Next consider an element of the relative homotopy group $\pi_j(T)$. By the Kan condition it is represented by a map $\rho\colon \Delta^j \otimes [0,\infty) \to \mathbb{P}_{n+1}(X; p)$ such that $\rho | \partial_i \Delta^j = 0$ for $i<j$ and $\rho | \partial_j \Delta^j = T\sigma$ for some $\sigma\colon \Delta^{j-1} \otimes [0,\infty) \otimes I \to \mathbb{P}_n(X; p)$. We need a deformation of ρ rel $\partial_j \rho$ to a map in the image of T. An extension $\sigma'\colon \Delta^j \otimes [0,\infty) \otimes I \to \mathbb{P}_n(X; p)$ of σ can be constructed by first letting $\sigma'(\tau) = 0$ for τ in $(\bigcup_{i<j} \partial_i \Delta^j) \otimes [0,\infty) \otimes I \cup \Delta^j \otimes [0,\infty) \otimes \{0,1\}$ and then using the triangulation argument. There is an obvious cobordism which gives a simplex connecting ρ and $T\sigma'$. □

Now let us consider a special case when X is a single point $*$. In this case, $\mathbb{L}_j(*; E \to *)$ is homotopy equivalent to $\mathbb{P}'_j(*; E \to *)$. We denote $\mathbb{P}'_j(*; E \to *)$ by $\mathbb{L}_j(E)$.

(3.5) **Proposition.** *There is a natural isomorphism*

$$\theta\colon L_n^{-\infty}(G) \to \pi_n \mathbb{L}(E)$$

where G is the fundamental group of E.

Here $L_n^{-\infty}$ is the direct limit $\varinjlim_{j \to \infty} L_n^{(-j)}$ of Ranicki's lower L-groups [11]. $L_n^{(-j)}(G)$ is defined to be the kernel of the product of projection maps

$$L^{(1)}_{n+j+1}(\mathbb{Z}^{j+1} \times G) \to \prod^{j+1} L^{(1)}_{n+j+1}(\mathbb{Z}^j \times G).$$

The map $L_n^{(-j)} \to L_n^{(-j-1)}$ is induced by the map

$$\sigma^*(S^1) \otimes :\ L^{(1)}_{n+j+1}(\mathbb{Z}^{j+1} \times G) \to L^{(1)}_{n+j+2}(\mathbb{Z}^{j+2} \times G).$$

Proof of 3.5. First let us define θ. An element of $L_n^{(-j)}(G)$ can be represented by a free $(n+j+1)$-dimensional quadratic Poincaré complex over $\mathbb{Z}[\mathbb{Z}^{j+1} \times G]$.

Represent this by a geometric $\mathbb{Z}$-module quadratic Poincaré complex on $T^{j+1} \times E$, where T^{j+1} is a $(j+1)$-torus. The pull-back of this on $\mathbb{R}^{j+1} \times E$ has a bounded radius in the $\mathbb{R}^{j+1}$ coordinates. So we can regard this as an n-simplex of $\mathbb{L}_0(E)$ with zero boundary, and hence as an element of $\pi_n \mathbb{L}_0(E)$. $(\sigma^*(S^1) \otimes)$ on the left side corresponds to Σ on the right, so this defines a homomorphism θ of the direct limit.

Next we show that θ is onto. Take an element of $\pi_n \mathbb{L}(E)$ and represent it by a locally finite geometric $\mathbb{Z}$-module quadratic Poincaré complex c on $\mathbb{R}^j \times E$ with bounded radius on $\mathbb{R}^j$ for some j. Split c along $\{0\} \times \mathbb{R}^{j-1} \times E$ without taking the external suspension; we obtain a splitting $c \sim c_+ \cup_b c_-$, where b may lie all over $\mathbb{R}^j \times E$. We proved that the geometric $\mathbb{Z}$-module quadratic Poincaré complex $\sigma^*(S^1) \otimes b$ on $S^1 \times \mathbb{R}^j \times E$ is homotopy equivalent to another complex d which lies on $S^1 \times \{0\} \times \mathbb{R}^{j-1} \times E$. So Σc on $\mathbb{R} \times \mathbb{R}^j \times E$ has a splitting $c'_+ \cup_{d'} c'_-$, where d' is the pull-back of d in $\mathbb{R} \times \{0\} \times \mathbb{R}^{j-1} \times E$. Note that $\Sigma d'$, the pull-back of $\sigma^*(S^1) \otimes d'$ (on $\mathbb{R} \times S^1 \times \mathbb{R}^{j-1} \times E$) on $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^{j-1} \times E$ also has a splitting $d'_+ \cup_{d'} d'_-$. Here $+$ and $-$ are with respect to the second $\mathbb{R}$. We claim that Σc and $\Sigma d'$ are cobordant. $\Sigma c \oplus (-\Sigma d')$ is cobordant to $(c'_+ \cup_{d'} (-d'_+)) \oplus (c'_- \cup_{-d'} (-d'_-))$. Since $c'_+ \cup_{d'} (-d'_+)$ (resp. $c'_- \cup_{-d'} (-d'_-)$) lies over $\mathbb{R} \times [0, \infty) \times \mathbb{R}^{j-1} \times E$ (resp. $\mathbb{R} \times (-\infty, 0] \times \mathbb{R}^{j-1} \times E$), the next lemma implies that these are cobordant to 0, and hence Σc is cobordant to $\Sigma d'$. Repeat this process until one gets $\Sigma^j e'$, where e is a geometric $\mathbb{Z}$-module quadratic Poincaré complex on $T^j \times E$, and e' is the pull-back of e on $\mathbb{R}^j \times E$. This e represents an element of $L_n^{-\infty}(G)$. Obviously θ sends this element to $[c] \in \pi_n \mathbb{L}(E)$.

(3.6) **Lemma.** *Any locally finite $\mathbb{Z}$-module quadratic Poincaré complex on $\mathbb{R}^j \times E$ which lies over $[0, \infty) \times \mathbb{R}^{j-1} \times E$ is locally finitely cobordant to zero.*

Proof. Let c be such a complex, and t denote the parrallel translation of $\mathbb{R} \times \mathbb{R}^{j-1} \times E$ defined by $t(x, y, z) = (x+1, y, z)$. Then c is cobordant to a locally finite complex

$$c \oplus [(-t\,c) \oplus (t^2\,c)] \oplus [(-t^3\,c) \oplus (t^4\,c)] \oplus \ldots$$
$$= [c \oplus (-t\,c)] \oplus [(t^2\,c) \oplus (-t^3\,c)] \oplus \ldots$$

which is cobordant to zero. □

Let us go back to the proof of 3.5. We will prove the injectivity of θ. Pick an element x in the kernel of θ. Its image is cobordant to 0. Apply the same argument to this null cobordism as in the onto part. This will show that $x = 0$. □

This justifies the following notation:

3.7 *Notation.* $$L_n^{-\infty}(E) := \pi_n \mathbb{L}(E).$$

Next we describe homology $H_*(X; \mathbb{L}(p))$ defined by Quinn [9, § 8], in terms of Δ-sets. We assume X is a finite polyhedron and p is a simplicial stratified system of fibrations. We fix a Δ-set K with realization $|K|$ equal to X; for example, the first derived of a triangulation of X gives such a Δ-set. $H_*(X; \mathbb{L}(p))$ does not depend on the choice of K [ibid.]. Recall that $\mathbb{L}(-)$ is a

covariant functor from spaces to spectra of Δ-sets (3.4). We apply $\mathbb{L}(-)$ blockwise to p. Define a Δ-set $\mathbb{L}_n(p)$ by:

$$\{\bigcup_{\Delta\in K} \mathbb{L}_n(p^{-1}(\Delta))\otimes\Delta\}/\sim,$$

where a simplex $\Delta\in K$ is given the obvious Δ-set structure, and the equivalence relation $\sim$ is generated by : a simplex in $\mathbb{L}_n(p^{-1}(\partial_j\Delta))\otimes\partial_j\Delta$ is identified with its image in $\mathbb{L}_n(p^{-1}(\Delta))\otimes\Delta$. The realization $|\mathbb{L}_n(p)|$ is homeomorphic to:

$$\{\bigcup_{\Delta\in K} |\mathbb{L}_n(p^{-1}(\Delta))|\times|\Delta|\}/\sim,$$

since $\mathbb{L}_n(p^{-1}(\Delta))$ is Kan. Here $\sim$ is induced by the identification of simplices given above. Denote by p_* the natural projection of $|\mathbb{L}_n(p)|$ to X. Also the section i: $X\to|\mathbb{L}_n(p)|$ is defined by fitting together the base points of the pieces. Further the structure maps $|\mathbb{L}_{-n}(p^{-1}(\Delta))|\to|\Omega\mathbb{L}_{-n-1}(p^{-1}(\Delta))|\simeq\Omega|\mathbb{L}_{-n-1}(p^{-1}(\Delta))|$ fit together to make $|\mathbb{L}_{-n}(p)|$ into an ex-spectrum ([9, p. 423]):

$$|\mathbb{L}_{-n}(p)|\to\Omega_X(|\mathbb{L}_{-n-1}(p)|).$$

For a subpolyhedron Y of X, the composition

$$|\mathbb{L}_{-n}(p)|\to\Omega_X(|\mathbb{L}_{-n-1}(p)|)\xrightarrow{\text{inclusion}}\Omega(|\mathbb{L}_{-n-1}(p)|/i(X))$$

induces a map:

$$|\mathbb{L}_{-n}(p)|/i(X)\cup p_*^{-1}(Y)\to\Omega(|\mathbb{L}_{-n-1}(p)|/i(X)\cup p_*^{-1}(Y)).$$

Taking Ω^{n-j} of this we have a map:

$$\Omega^{n-j}(|\mathbb{L}_{-n}(p)|/i(X)\cup p_*^{-1}(Y))\to\Omega^{n-j+1}(|\mathbb{L}_{-n-1}(p)|/i(X)\cup p_*^{-1}(Y)).$$

We replace this by a Δ-map between Δ-sets, applying singular complex functor S and then applying forgetful functor F [14].

(3.8) **Definition.** The homology spectrum $\mathrm{H}(X, Y;\mathbb{L}(p))$ is an Ω-spectrum of Δ-sets defined by

$$\mathbb{H}_j(X, Y;\mathbb{L}(p))=\lim_{n\to\infty} FS\Omega^{n-j}(|\mathbb{L}_{-n}(p)|/i(X)\cup p_*^{-1}(Y)).$$

The homology groups are the homotopy groups of this spectrum.

The functor $\mathbb{L}(-)$ which was used to construct $\mathbb{H}(X, Y;\mathbb{L}(p))$ is homotopy invariant; in fact, the homotopy type of $\mathbb{L}(E)$ depends only on the fundamental group $\pi_1 E$ (3.5). Therefore, according to Quinn [9, p. 421], $\mathbb{H}(-;\mathbb{L}(-))$ is a homology theory on the category of polyhedra with stratified systems of fibrations. From now on we always assume p as such. The following is the main theorem of this section.

(3.9) **Theorem.** (Characterization theorem). *Let p: $E\to X$ be a polyhedral stratified system of fibrations on a finite polyhedron X. Then there is a homotopy equivalence A_j: $\mathbb{H}_j(X;\mathbb{L}(p))\to\mathbb{L}_{-j}(X;p)$.*

Proof. One way to prove this is to use Characterization theorem (8.5) of Quinn [9]. But, since we will also need something more explicit (3.10 below), we give a definition of the map A, called the assembly map, and prove that it is a homotopy equivalence.

A k-simplex of $FS\Omega^{n-j}(|\mathbb{L}_{-n}(p)|/i(X))$ is a map $\rho\colon S^{n-j}\times\Delta^k\to|\mathbb{L}_{-n}(p)|/i(X)$. By modifying ρ a little, if necessary, we may assume that there exist a compact codimension 0 submanifold V of $S^{n-j}\times\Delta^k$ and a cellular map $\rho'\colon V\to|\mathbb{L}_{-n}(p)|$ with respect to a triangulation of V such that ρ sends the complement of $\mathrm{int}(V)$ to the base point $[i(X)]$ and $\rho|V$ factors through ρ'. We may assume the triangulation of V is a derived of another, and regard V as a Δ-set. By a simplex-wise inductive application of the relative Δ-map approximation theorem of Rourke and Sanderson [14, theorem 5.3], we may assume ρ' is a (realization of) a Δ-map, because $\mathbb{L}_{-n}(p^{-1}(\Delta))\otimes\Delta$ is Kan. A further application of the approximation theorem produces a homotopic Δ-map, also denoted by ρ', into a subcomplex $\bigcup\mathbb{L}_{-n}(p^{-1}(\Delta))\times FG(\Delta)$ of $\mathbb{L}_{-n}(p)$, because $\mathbb{L}_{-n}(p^{-1}(\Delta))$ is Kan. For each $(n-j+k)$-dimensional simplex Δ of V, $\rho'(\Delta)$ is given a structure of $(-j+k+l)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special $(n-j+k+2)$-ad on $\mathbb{R}^l\times E$ with radius measured in $X=$the diameter of $p_*\rho'(\Delta)$. Glueing all these, after taking external suspensions if necessary, we obtain a $(-j+k+l')$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré special $(k+2)$-ad on $\mathbb{R}^{l'}\times E$ with radius max {diameter of $p_*\rho'(\Delta)|\Delta\in V\}$, which is a simplex of $\mathbb{P}_{-j}(X;p)$. Note that, since p is a stratified system of fibrations, each piece can be shrunk to produce a copy of a smaller radius lying over a simplex Δ' of a barycentric subdivision of $p_*\rho'(\Delta)$ (use a lift to E of a shrinking map $p_*\rho'(\Delta)\to\Delta'$), and we fill in the resultant gaps with fine product cobordisms and glue these together to obtain a simplex with a smaller radius. More precisely, let V' denote the barycentric subdivision of V. There is an obvious simplicial map $f\colon V'\to V$ homotopic to the identity which maps, for each $(n-j+k)$-simplex Δ of V, one of the $(n-j+k)$-simplex Δ' of V' contained in Δ to Δ and all the others to faces of Δ. Apply the argument above to approximate (relatively) the composition $\rho' f\colon V'\to|\mathbb{L}_{-n}(p)|$ by a Δ-map. Then use a lift of the homotopy $1_V\simeq f$ to shrink things. This is called the "barycentric subdivision shrinking argument". There is a cobordism between the original and the new simplex, and repeated application of this barycentric subdivision shrinking argument, together with the triangulation argument, generates a map $\Delta^k\otimes[0,\infty)\to\mathbb{P}_{-j}(X;p)$. Since the radius goes to 0 as $t\to\infty$, this defines a simplex of $\mathbb{L}_{-j}(X;p)$. This defines the desired map A_j.

In the following proof, we will not mention taking barycentric subdivisions or external suspensions when we glue or split things for simplicity. First we will show that A_j maps into every component. Let σ be a 0-simplex of $\mathbb{L}_{-j}(X;p)$; since X is compact, we may assume that σ is a map from $[0,\infty)$ to $\mathbb{P}'_{-j}(X;p)^{(N)}$ with a sequence ε_i monotone decreasing to 0 such that $\sigma([i,\infty))\subset\mathbb{P}_{-j}(X,p,\varepsilon_i)^{(N)}$ for each i, where N is some positive integer. Embed X in S^{n-j} for a sufficiently large n, and let W and r denote a regular neighborhood of X in S^{n-j} and the retraction: $W\to X$. Let $p'\colon E'\to W$ be the stratified system of fibrations obtained as the pull-back of p by r; E' retracts to E, and p' is an extension of p and has the advantage that the base space is a

manifold. $\sigma(0)$ can now be regarded as an $(l-j)$-dimensional geometric $\mathbb{Z}$-module quadratic Poincaré complex on $\mathbb{R}^l \times E'$ with radius ε_0 (with respect to p'). Let us split (stably) $\sigma(0)$ into pieces lying over the simplices of W. Let ε be any positive number; the stable splitting lemma gives $\delta > 0$ such that if $\sigma(0)$ has radius δ, then each split piece lies over an ε-neighborhood of the corresponding simplex of W. We may assume that this is the case. If $\varepsilon_0 > \delta$, then we can replace σ by another simplex $\sigma'(t) = \sigma(t+i)$ in the same component as σ, where i is chosen to be sufficiently large so that $\varepsilon_i < \delta$. If we had chosen ε sufficiently small at the beginning, we can construct a map of p' to itself of radius ε which sends $(p')^{-1}$ (ε-neighborhood of Δ) to $(p')^{-1}(\Delta)$ for each simplex $\Delta \in W$. This map induces ε-isomorphisms which make each split piece lie exactly over the corresponding simplex of W. The retraction $E' \to E$, then, induces isomorphisms that make each piece lying on $\mathbb{R}^l \times E$ for some l. Thus we have a triangulation of W such that each simplex of W is given a geometric $\mathbb{Z}$-module quadratic Poincaré ad on $\mathbb{R}^l \times E$. Since there are only trivial ads over the simplices in ∂W, we can associate trivial ads to the simplices in $S^{n-j} - \mathrm{int}(W)$ and define a 0-simplex of $FS\Omega^{n-j}(|\mathbb{L}_{-n}(p)|/i(X))$. We will show that A_j sends ρ into the same component as σ. First of all, by construction, $(A_j\rho)(0)$ and $\sigma(0)$ can be joined by a 1-simplex in $\mathbb{P}'_{-j}(X;p)$. This is the first step of the inductive construction of a 1-simplex τ connecting σ and $A_j\rho$. Assume we have constructed a map from $\partial_0 \Delta^1 \otimes [0,\infty) \cup \Delta^1 \otimes [0,m] \cup \partial_1 \Delta^1 \otimes [0,\infty)$ for some integer m, giving σ and $A_j\rho$ on each end. We can extend this over $\Delta^1 \otimes [m,m+1]$ by applying the "barycentric subdivision" shrinking argument to the union $\sigma([m,m+1]) \cup \tau(\Delta^1 \otimes \{m\}) \cup A_j\rho([m,m+1])$ to get $\tau(\Delta^1 \otimes \{m+1\})$ and then applying the triangulation argument to the resulted cobordism to fill in the square $\Delta^1 \otimes [m,m+1]$. This inductively constructs the desired homotopy. Thus A_j maps into every component.

Next we will show that the relative homotopy groups $\pi_k(A_j)$ vanish for all k. Its element is represented by a map $\sigma\colon \Delta^k \otimes [0,\infty) \to \mathbb{P}'_{-j}(X;p)$ such that $\sigma|\partial_i \Delta^k = 0$ for $i<k$ and $\sigma|\partial_k \Delta^k = A_j\rho$ for some $\rho\colon S^{n-j} \times \partial_k \Delta^k \to |\mathbb{L}_{-n}(p)|/i(X)$ such that $\rho(S^{n-j} \times \partial\partial_k \Delta^k) = [i(X)]$. We may assume that ρ maps the complement of the interior of some codimension 0 submanifold V of $S^{n-j} \times \partial_k \Delta^k$ to $i(X)$, and hence restricts to a map $\rho'\colon V \to |\mathbb{L}_{-n}(p)|$. Roughly speaking, $\sigma(0)$ is a Poincaré pair whose boundary is split into pieces lying over simplices of V. To find a k-simplex ρ_σ of $FS\Omega^{n-j}(|\mathbb{L}_{-n}(p)|i(X))$ whose image $A_j\rho_\sigma$ by A_j is homotopic to σ fixing the boundary, we want to apply the stable splitting lemma to $\sigma(0)$. As before we assume that X is a subcomplex of S^{n-j}. This time let W denote a regular neighborhood of X in $S^{n-j} \times \Delta^k$, where S^{n-j} is identified with $S^{n-j} \times \{\text{the center of the } (k-1)\text{-simplex } \partial_k \Delta^k\}$. We extend p to the pull-back p': $E' \to W$ using the retraction $W \to X$ and split things here as before. Unfortunately, the splitting of the boundary of $\sigma(0)$ is not over $W_0 = W \cap (S^{n-j} \times \partial_k \Delta^k)$ but over V. We remedy this as follows, changing ρ by homotopy. First apply the "barycentric subdivision" shrinking argument to each split piece to get a homotopic map, denoted by ρ again, for which the radius of each split piece is very small. Consider the composition $p_*\rho'\colon V \to X$. Choose sufficiently large m, so that the mapping cylinder M of $p_*\rho'$ can be embedded in $\Sigma^m S^{n-j} \times \partial_k \Delta^k \times [0,1]$ so that $V \subset \Sigma^m S^{n-j} \times \partial_k \Delta^k \times \{0\}$ and $X \subset \Sigma^m S^{n-j} \times \partial_k \Delta^k \times \{1\}$.

Take a regular neighborhood N of M in $\Sigma^m S^{n-j} \times \partial_k \Delta^k \times [0,1]$, then N retracts to M and then to X. Use this to get a pull-back $E^* \to N$ of $E \to X$. At 0, $\Sigma^m \rho'$ gives a geometric $\mathbb{Z}$-module quadratic Poincaré complex on E^* split over simplices of $\Sigma^m S^{n-j} \times \partial_k \Delta^k \times \{0\}$, and at 1, consider Σ^m of the boundary of $\sigma(0)$ regarded to be on E^*. Since these are essentially the same thing, these can be connected by a cobordism. Actually, if one places sufficiently many layers of product cobordisms along the mapping cylinder, we may assume that the radius of the cobordism is very small. If everything is sufficiently small, then we can apply the stable splitting lemma over N, to get a homotopy of ρ to a new map for which the splitting is over a regular neighborhood of X in $\Sigma^m S^{n-j} \times \partial_k \Delta^k$. Thus, from the beginning, we may assume that the splitting of the boundary of $\sigma(0)$ is over W_0. Now we can apply the stable splitting lemma to obtain a map $\rho_\sigma : S^{n-j} \times \Delta^k \to |\mathbb{L}_{-n}(p)|/i(X)$. ρ_σ is a k-simplex of $FS\Omega^{n-j}(|\mathbb{L}_{-n}(p)|/i(X))$ with all faces at the base point except ρ. The same argument as in the first part gives a homotopy between σ and $A_j \rho_\sigma$ fixing the boundary. This completes the proof. □

The proof actually gives the following:

(3.10) **Corollary.** (Shrinking lemma). *Let p be as in* 3.9. *Then, given any positive integer n, there exists a positive number ε such that for any $0 < \delta < \varepsilon$, there is a function ("shrinking function") S: $\mathbb{P}'_j(X,p,\delta)^{(n)} \to \mathbb{L}_j(X;p)$ such that the map $\mathbb{P}'_j(X,p,\delta)^{(n)} \to \mathbb{P}'_j(X;p)$ induced by the following composition is homotopic to the inclusion map:*

$$\mathbb{P}'_j(X,p,\delta)^{(n)} \xrightarrow{S} \mathbb{L}_j(X;p) \xrightarrow{\text{restriction to } 0} \mathbb{P}_j(X;p).$$

4. Crystallographic groups

We begin this section by reviewing some work on crystallographic groups by Farrell and Hsiang in [6]. Γ is *crystallographic* if it is a discrete co-compact subgroup of $E(n)$, the group of rigid motions of Euclidean n-space. Identify $\mathbb{R}^n$ with the group of translations of $\mathbb{R}^n$, then $E(n) = \mathbb{R}^n \rtimes O(n)$. The intersection of Γ and $\mathbb{R}^n$ is the maximal abelian subgroup of Γ with finite index, which is denoted by A and is called the *translation subgroup* of Γ. The finite factor group Γ/A is called the *holonomy group* of Γ. The *rank* of Γ is the rank of A. For any positive integer s, $\Gamma_s = \Gamma/sA$ and $A_s = A/sA$. T and T_n denote the infinite cyclic group and the finite cyclic group of order n respectively.

4.1 Examples. (1) D_∞ will denote the infinite dihedral group; i.e., $D_\infty \subset E(1)$ is the subgroup generated by $x \mapsto x+1$ and $x \mapsto -x$ (where $x \in \mathbb{R}$). It is a semi-direct product $T \rtimes T_2$, where T_2 acts on T via multiplication by -1.

(2) See [6, p. 658] for the definition of 2-dimensional crystallographic groups of type 1, 2 and 3. The holonomy group of a crystallographic group of type 2 or 3 is $T_2 \oplus T_2$.

(3) A crystallographic group Γ of rank $n \geqq 2$ with holonomy group G is called *special* if there exist a crystallographic group $\tilde{\Gamma}$ of rank $m \geqq 1$ and an infinite sequence of positive integers s such that, if H is a maximal hyper-elementary subgroup of Γ_s, either

(4.1.1) H projects to a proper subgroup of G via the canonical map $\Gamma_s \to G$; or

(4.1.2) H is conjugate to a hyperelementary subgroup H' and there is a group surjection η: $q^{-1}(H') \to \tilde{\Gamma}$ together with a η-equivariant affine surjection h: $\mathbb{R}^n \to \mathbb{R}^m$ such that

$$|dh(X)| \leqq (2/\sqrt{s})\,|X|$$

for each vector X tangent to $\mathbb{R}^n$, where q: $\Gamma \to \Gamma_s$ is the natural projection and $|\ |$'s are the Euclidean metrics on $\mathbb{R}^m$ and $\mathbb{R}^n$. Note that $\tilde{\Gamma}$ is required to be independent of the choice of s and H.

Farrell and Hsiang showed that 2-dimensional crystallographic groups of type 1, 2 or 3 are special [6, theorem 4.2].

On the other hand, T and D_∞ are not special, because they have rank 1. But a similar statement as above holds true for D_∞ when we replace "hyperelementary" by "elementary". In fact, let $\tilde{\Gamma} = D_\infty$ and $\{s\}$ be the set of all odd primes; then the only maximal elementary subgroups of $(D_\infty)_s = T_s \rtimes T_2$ (dihedral group of order $2s$) are $H_1 = T_s$ and $H_2 = T_2$ up to conjugacy. The image of H_1 in T_2 is $\{1\}$, so H_1 satisfies (4.1.1). For H_2, $q^{-1}(H_2) = (sT) \rtimes T_2$, so we can let η be the obvious isomorphism: $(st) \rtimes T_2 \to T \rtimes T_2 = \tilde{\Gamma}$, and h be the linear map $h(x) = (1/s)\,x$ $(x \in \mathbb{R})$. Thus H_2 satisfies (4.1.2).

The following results of Farrell and Hsiang will play a key role in the proof of our main theorem.

(4.2) **Theorem.** ([6, theorem 1.1], [5, theorem 3.1]). *Let Γ be a crystallographic group of rank n and holonomy group G, then either*

(i) *$\Gamma = \Gamma' \rtimes T$ where Γ' is a crystallographic subgroup of rank $n-1$; or*

(ii) *there is an infinite sequence of positive integers $s \equiv 1 \bmod |G|$ such that any hyperelementary subgroup of Γ_s which projects onto G (via the canonical map) projects isomorphically onto G; or*

(iii) *G is an elementary abelian 2-group and*

(a) *if $|G| = 2$, then $\Gamma = A \rtimes T_2$ and $T_2 = G$ acts on A via multiplication by -1, and*

(b) *if $|G| > 2$, then Γ maps epimorphically onto a crystallographic group of type 2 or 3.*

(4.3) **Lemma.** ([6, lemma 1.2]). *Let ϕ: $\Gamma \to \Gamma'$ be an epimorphism between crystallographic groups Γ of rank n and Γ' of rank m. Then there exists a ϕ-equivariant affine surjection F: $\mathbb{R}^n \to \mathbb{R}^m$.*

Theorem 4.2 above is stated in a slightly different form from [6, theorem 1.1]; this stronger version is implicitly used in the proof of [6, theorem 5.1]. On p. 665 of [6], a crystallographic group $\hat{\Gamma}$ is replaced by $\bar{\Gamma}$, via an epimorphism ψ: $\hat{\Gamma} \to \bar{\Gamma}$. This epimorphism should not increase the order of the holonomy group. In the case of possibility (ii) of 4.2 (with $\hat{\Gamma}$ used as Γ), ψ is the identity map. In the case of possibility (iii)-(a), one uses the obvious epimorphism from $\hat{\Gamma}$ to $(T \oplus T) \rtimes T_2$; so both have the same holonomy group. In the case of possibility (iii)-(b), $\hat{\Gamma}$ has holonomy group of order $\geqq 4$; so the epimorphism does not increase the order of holonomy group.

Suppose Γ satisfies 4.2(ii). Theorem 4.1 in [6] and an argument in the proof of theorem 4.4 [ibid.] shows that Γ is special; Γ itself is used as $\tilde{\Gamma}$. If Γ satisfies the possibility (iii) of 4.2, then Γ maps epimorphically onto a crystallographic group of type 1, 2 or 3 with the order of the holonomy group $\leqq |G|$, or Γ is $T \rtimes T_2$. Therefore we can rephrase 4.2 as follows:

(4.4) **Corollary.** *Let Γ be a crystallographic group of rank $n \geqq 1$, then either*

(1) *$\Gamma = \Gamma' \rtimes T$ where Γ' is a crystallographic subgroup of rank $n-1$; or*

(2) *there is an epimorphism from Γ onto a special crystallographic group which does not increase the order of the holonomy group; or*

(3) *Γ is isomorphic to D_∞.*

Next let us consider a slightly larger class of groups. Suppose Λ is a finitely generated group which maps onto a crystallographic group Γ of rank n with a finite kernel. For example, if Λ is a virtually abelian group of rank n, then, by lemma 1.2 in [5], there is a surjection $\Lambda \to \Gamma$ onto a crystallographic group of rank n with a finite kernel. Via this surjection, Λ acts by isometries on $\mathbb{R}^n$ discretely, virtually faithfully, with compact orbit space. The orbit space $\mathbb{R}^n/\Lambda$ is equal to $\mathbb{R}^n/\Gamma$. The action of Γ may not be free, since Γ may have torsion, but its translation subgroup A acts on $\mathbb{R}^n$ freely and the orbit space is a flat torus T^n. The holonomy group G of Γ acts on T^n as a group of isometries so that $\mathbb{R}^n/\Gamma = T^n/G$. Therefore the orbit space can be viewed as the orbit space of a finite group action on a compact smooth manifold. If (H) is the conjugacy class of a subgroup H of G, then $Y_{(H)}$ denotes the subset of T^n/G consisting of the points x such that the isotropy subgroup at a point of T^n lying in the orbit x is in (H). It is observed in [6] that $\{Y_{(H)}\}$ gives a stratification of T^n/G.

Let W_Λ denote a free contractible Λ-complex. Λ acts freely on $\mathbb{R}^n \times W_\Lambda$ diagonally.

(4.5) **Proposition.** *The projection p: $(\mathbb{R}^n \times W_\Lambda)/\Lambda \to \mathbb{R}^n/\Lambda$ is a stratified system of fibrations.*

Proof. Let β: $\Lambda \to \Gamma$ denote the surjection and r: $\mathbb{R}^n \to \mathbb{R}^n/\Lambda$ the quotient map. For each orbit $x \in \mathbb{R}^n/\Lambda = \mathbb{R}^n/\Gamma$, define Λ_x and Γ_x to be the isotropy subgroups of Λ and Γ at a point $\bar{x} \in r^{-1}(x)$. Λ_x and Γ_x are well-defined up to conjugacy. The map p has point inverses $p^{-1}(x) = W_\Lambda/\Lambda_x$, which are classifying spaces of Λ_x. Since $\Lambda_x = \beta^{-1}(\Gamma_x)$, this proposition will be proved if one can show that (Γ_x) is constant on each component of each stratum $Y_{(H)}$. Suppose x is in $Y_{(H)}$, and let $[\bar{x}] \in T^n$ denote the orbit of $\bar{x}$ by A. The isotropy subgroup of G at $[\bar{x}]$ is H. Let π denote the quotient map: $\Gamma \to G$. Then $\pi|\Gamma_x$: $\Gamma_x \to H$ is an isomorphism. This implies that Γ_x is locally constant on $Y_{(H)}$, and it in turn implies the proposition. □

Again let β be the surjection $\Lambda \to \Gamma$. Let us study the geometric implication of corollary 4.4(1), (2), and (3). First suppose (1) $\Gamma = \Gamma' \rtimes T$. Then by 4.3, the epimorphism ϕ: $\Gamma \to T$ induces a ϕ-equivariant affine surjection F: $\mathbb{R}^n \to \mathbb{R}$. F is also $(\phi\beta)$-equivariant and defines a map $\bar{F}$: $\mathbb{R}^n/\Lambda \to \mathbb{R}/T = S^1$; $\bar{F}$ is a fibre bundle with fibre $\mathbb{R}^{n-1}/\Lambda'$, where $\mathbb{R}^{n-1}$ is a fibre of F and $\Lambda' = \ker(\phi\beta) = \beta^{-1}(\Gamma')$. $(\mathbb{R}^n \times W_\Lambda)/\Lambda$ also fibres over S^1 with fibre $(\mathbb{R}^{n-1} \times W_\Lambda)/\Lambda'$. Then p

restricts to a stratified system of fibrations between the fibres p': $(\mathbb{R}^{n-1} \times W_\Lambda)/\Lambda' \to \mathbb{R}^{n-1}/\Lambda'$.

Next suppose Γ satisfies 4.4(2) (resp. (3)), and let ϕ be the surjection $\Gamma \to \Gamma'$ (resp. the identity map: $D_\infty \to D_\infty$). Let $\tilde{\Gamma}$, s, η, h be as in 4.1(3) with Γ replaced by Γ'. Suppose H is a maximal hyperelementary (resp. elementary) subgroup of Γ_s'. Define a subgroup Λ_H of Λ (Γ_H of Γ) by $\Lambda_H = \beta^{-1}\phi^{-1}q^{-1}(H)$ ($\Gamma_H = \phi^{-1}q^{-1}(H)$). Γ_H is a subgroup of Γ of finite index, and hence a crystallographic group. Thus Λ_H maps onto a crystallographic group, and we can consider p_H: $(\mathbb{R}^n \times W_\Lambda)/\Lambda_H \to \mathbb{R}^n/\Lambda_H$. Here the complex W_Λ is used as W_{Λ_H}.

If H satisfies (4.1.1), then the holonomy group of $q^{-1}(H)$ has a strictly smaller order than Γ'.

Next suppose that H satisfies (4.1.2). Replace H by H'. The surjection ϕ induces a ϕ-equivariant affine surjection f: $\mathbb{R}^n \to \mathbb{R}^l$. The maps f and h induce a surjection

$$\alpha\colon\ \mathbb{R}^n/\Lambda_H = \mathbb{R}^n/\Gamma_H \to \mathbb{R}^l/q^{-1}(H) \to \mathbb{R}^m/\tilde{\Gamma}.$$

The composition hf: $\mathbb{R}^n \to \mathbb{R}^m$ will be denoted by $\tilde{\alpha}$. We consider the composition αp_H; the map p_H can be understood by studying αp_H and the restriction of p_H to the preimages of points of $\mathbb{R}^m/\tilde{\Gamma}$. For an orbit $x \in \mathbb{R}^m/\tilde{\Gamma}$, $\tilde{\Gamma}_x$ denotes the isotropy subgroup of $\tilde{\Gamma}$ at a point $\bar{x} \in \mathbb{R}^m$ in the orbit x, as usual. Define a subgroup $\Lambda_H(x)$ of $\Lambda_H(\Gamma_H(x)$ of Γ_H) by $\Lambda_H(x) = (\eta\phi\beta|\Lambda_H)^{-1}(\tilde{\Gamma}_x)$ ($\Gamma_H(x) = (\eta\phi|\Gamma_H)^{-1}(\tilde{\Gamma}_x)$); i.e., $\Lambda_H(x)(\Gamma_H(x))$ is the set of elements of $\Lambda_H(\Gamma_H)$ which leaves the affine subspace $\mathbb{R}^{n-m} = \tilde{\alpha}^{-1}(\bar{x})$ of $\mathbb{R}^n$ invariant. $\Lambda_H(x)$ and $\Gamma_H(x)$ are well-defined up to conjugacy. The point inverses $(\alpha p_H)^{-1}(x)$ of αp_H are $(\mathbb{R}^{n-m} \times W_\Lambda)/\Lambda_H(x)$. Note that αp_H is a stratified system of fibrations, and that we can use the same filtration of $\mathbb{R}^m/\tilde{\Gamma}$ and the same neighborhoods of strata as those for the projection $\tilde{p}$: $(\mathbb{R}^m \times W_{\tilde{\Gamma}})/\tilde{\Gamma} \to \mathbb{R}^m/\tilde{\Gamma}$, since Λ_H is determined by $\tilde{\Gamma}_x$. The actions of $\Lambda_H(x)$ and $\Gamma_H(x)$ on $\mathbb{R}^{n-m}$ give a homomorphism

$$\Lambda_H(x) \xrightarrow{\beta'} \Gamma_H(x) \xrightarrow{\gamma} E(n-m),$$

where β' is the restriction of β. Let $\Gamma_H^*(x)$ denote the image of this in $E(n-m)$.

(4.6) **Proposition.** *$\Gamma_H^*(x)$ is a crystallographic subgroup of $E(n-m)$ and the kernel of $\Lambda_H(x) \to \Gamma_H^*(x)$ is finite.*

Proof. Let K be the kernel of $\eta\phi|\Gamma_H$: $\Gamma_H \to \tilde{\Gamma}$. Since K is a normal subgroup of a crystallographic group Γ_H, K is also a crystallographic group. In fact there is a K-invariant $(n-m)$-dimensional affine subspace V of $\mathbb{R}^n$ on which K acts discretely and faithfully with compact quotient. (See Farkas [3], theorem 17.) Since all the parallels of $\mathbb{R}^{n-m}$ in $\mathbb{R}^n$ are of the form $\tilde{\alpha}^{-1}(p\,t)$, they are all invariant under the action of K, and hence V is contained in one of these, say $\tilde{\alpha}^{-1}(x_0)$ for some $x_0 \in \mathbb{R}^m$. Both V and $\tilde{\alpha}^{-1}(x_0)$ have the same dimension, so they are actually equal. Therefore $\tilde{\alpha}^{-1}(x_0)/K$ is compact. On the other hand, the actions of K on $\tilde{\alpha}^{-1}(\bar{x})$ and on $\tilde{\alpha}^{-1}(x_0)$ are affinely equivalent. So $\mathbb{R}^{n-m}/\gamma(K)$ is also compact. Since $\Gamma_H^*(x)$ contains $\gamma(K)$, $\mathbb{R}^{n-m}/\Gamma_H^*(x)$ is compact. $\Gamma_H^*(x)$ is obviously discrete in $E(n-m)$; therefore, $\Gamma_H^*(x)$ is a crystallographic subgroup of $E(n-m)$. Now $|\mathrm{Ker}(\gamma)| = [K \cdot \mathrm{Ker}(\gamma)\colon K] \leqq [\Gamma_H(x)\colon K] = |\tilde{\Gamma}_x| < \infty$, so

$\mathrm{Ker}(\gamma)$ is finite. Since the kernel of β is finite by assumption, so is the kernel of $\gamma\beta'$. □

Thus p_H restricts to p_x: $(\mathbb{R}^{n-m} \times W_\Lambda)/\Lambda_H(x) \to \mathbb{R}^{n-m}/\Lambda_H(x)$, where Λ_H maps onto a crystallographic group $\Gamma_H^*(x)$ with finite kernel and acts on $\mathbb{R}^{n-m}$ via the action of $\Gamma_H^*(x)$. Note that $n-m$ is strictly smaller than n.

Next we observe that $H_*(\mathbb{R}^n/\Lambda; \mathbb{L}(p))$ and $L_*^{-\infty}((\mathbb{R}^n \times W_\Lambda)/\Lambda)$ satisfy elementary and hyperelementary induction. Let β: $\Lambda \to \Gamma$ be as before, and suppose we have an epimorphism ψ: $\Gamma \to G$ onto a finite group G (not necessarily the holonomy group of Γ). Of course, what we have in mind is the composition $q\phi$: $\Gamma \to \Gamma' \to \Gamma'_s$ above, where Γ' is a special crystallographic group. For a subgroup H of G, let $\Lambda_H = \beta^{-1}\psi^{-1}(H)$ and $\Gamma_H = \psi^{-1}(H)$. Let $\mathcal{S}$ denote the category of the subgroups of G and conjugations and Ab denote the category of abelian groups.

Let us define a Mackey functor M: $\mathcal{S} \to Ab$. For a subgroup H of G, $M(H)$ is defined to be $H_j(\mathbb{R}^n/\Lambda_H; \mathbb{L}(p_H))$, where p_H is the projection $(\mathbb{R}^n \times W_\Lambda)/\Lambda_H \to \mathbb{R}^n/\Lambda_H$. Suppose $f = (H, g, K)$ is a morphism from H to K, i.e., g is an element of G such that $g^{-1}Hg \subset K$. Pick an element $\lambda \in \Lambda$ such that $\psi\beta(\lambda) = g$. Then the actions of λ on $\mathbb{R}^n \times W_\Lambda$ and $\mathbb{R}^n$ induce a map $f_\#$: $p_H \to p_K$ between stratified systems of fibrations; i.e., we have a commutative diagram:

$$\begin{array}{ccc} (\mathbb{R}^n \times W_\Lambda)/\Lambda_H & \xrightarrow{f_\#} & (\mathbb{R}^n \times W_\Lambda)/\Lambda_K \\ \downarrow{\scriptstyle p_H} & & \downarrow{\scriptstyle p_K} \\ \mathbb{R}^n/\Lambda_H & \xrightarrow[\bar{f}_\#]{} & \mathbb{R}^n/\Lambda_K. \end{array}$$

We have the following two operations corresponding to $f_\#$:

(1) (functorial image f_*: $\mathbb{P}(\mathbb{R}^n/\Lambda_H; p_H) \to \mathbb{P}(\mathbb{R}^n/\Lambda_K; p_K)$). If $\mathbb{Z}[S]$, h: $S \to (\mathbb{R}^n \times W_\Lambda)/\Lambda_H$, is a geometric $\mathbb{Z}$-module on $(\mathbb{R}^n \times W_\Lambda)/\Lambda_H$, then a geometric $\mathbb{Z}$-module $f_*(\mathbb{Z}[S], h)$ on $(\mathbb{R}^n \times W_\Lambda)/\Lambda_K$ is defined to be $\mathbb{Z}[S]$ with $f_\# h$: $S \to (\mathbb{R}^n \times W_\Lambda)/\Lambda_K$. If $k = \Sigma m_\rho \rho$ is a morphism between geometric $\mathbb{Z}$-modules on $(\mathbb{R}^n \times W_\Lambda)/\Lambda_H$, then $f_* k = \Sigma m_\rho f_\# \rho$ is the functorial image of k. These induce a map f_*: $\mathbb{P}(\mathbb{R}^n/\Lambda_H; p_H) \to \mathbb{P}(\mathbb{R}^n/\Lambda_K; p_K)$.

(2) (pullback f^*: $\mathbb{P}(\mathbb{R}^n/\Lambda_K; p_K) \to \mathbb{P}(\mathbb{R}^n/\Lambda_H; p_H)$). If $\mathbb{Z}[S]$, h: $S \to (\mathbb{R}^n \times W_\Lambda)/\Lambda_K$, is a geometric $\mathbb{Z}$-module on $(\mathbb{R}^n \times W_\Lambda)/\Lambda_K$, then $f^*(\mathbb{Z}[S], h)$ is $(\mathbb{Z}[S^*], h^*)$, where h^*: $S^* \to (\mathbb{R}^n \times W_\Lambda)/\Lambda_H$ is the pullback of h. Pullbacks of morphisms are also defined by pullbacks of paths in the obvious way. These define a map f^*: $\mathbb{P}(\mathbb{R}^n/\Lambda_K; p_K) \to \mathbb{P}(\mathbb{R}^n/\Lambda_H; p_H)$.

Obviously, f_* does not increase the radius. On the other hand, f^* may increase the radius in general; but f^* does not increase the radius measured in $\mathbb{R}^n$, so it does not increase the radii of things which have sufficiently small radius. Therefore f_* and f^* induce maps $\mathbb{L}(\mathbb{R}^n/\Lambda_H; p_H) \to \mathbb{L}(\mathbb{R}^n/\Lambda_K; p_K)$ and $\mathbb{L}(\mathbb{R}^n/\Lambda_K; p_K) \to \mathbb{L}(\mathbb{R}^n/\Lambda_H; p_H)$. By the characterization theorem, these induce the desired maps f_*: $M(H) \to M(K)$ and f^*: $M(K) \to M(H)$. M is a bifunctor.

(4.7) **Proposition.** *M satisfies the double coset formula, and hence is a Mackey functor.*

Proof. Let L and L' be subgroups of a subgroup H of G, and suppose H has a double coset decomposition $H=\bigcup_{i=1}^{k} Lg_iL'$, $g_i\in H$. Let $\Lambda_H=\bigcup_{i=1}^{k}\Lambda_L\bar{g}_i\Lambda_L$, be a corresponding double coset decomposition of Λ_H, where $\bar{g}_i$ is an element of Λ_H such that $\psi\beta(\bar{g}_i)=g_i$. Let P be the pullback:

$$\begin{array}{ccc} P & \longrightarrow & (\mathbb{R}^n\times W_\Lambda)/\Lambda_{L'} \\ \downarrow & & \downarrow{\scriptstyle (L',e,H)_*} \\ (\mathbb{R}^n\times W_\Lambda)/\Lambda_L & \longrightarrow & (\mathbb{R}^n\times W_\Lambda)/\Lambda_H\,, \end{array}$$

i.e.,

$$P=\{([x],[y])\in(\mathbb{R}^n\times W_\Lambda)/\Lambda_L\times(\mathbb{R}^n\times W_\Lambda)/\Lambda_{L'}$$
$$[x]_{\Lambda_H}=[y]_{\Lambda_H}\in(\mathbb{R}^n\times W_\Lambda)/\Lambda_H\},$$

where $x\in\mathbb{R}^n\times W_\Lambda$ and [ ] denotes the corresponding orbit. Then it is easily verified that the map

$$\bigcup_{i=1}^{k}(\mathbb{R}^n\times W_\Lambda)/\Lambda_{L\cap g_iL'g_i^{-1}}\to P;$$
$$[x]_{\Lambda_{L\cap g_iL'g_i^{-1}}}\mapsto([x],[\bar{g}_i^{-1}x])$$

is a Λ_H-isomorphism. Therefore we obtain a pullback diagram:

$$\begin{array}{ccc} \bigcup_{i=1}^{k}(\mathbb{R}^n\times W_\Lambda)/\Lambda_{L\cap g_iL'g_i^{-1}} & \xrightarrow{(L\cap g_iL'g_i^{-1},\,g_i^{-1},\,L')_*} & (\mathbb{R}^n\times W_\Lambda)/\Lambda_{L'} \\ \downarrow{\scriptstyle (L\cap g_iL'g_i^{-1},e,L)_*} & & \downarrow{\scriptstyle (L',e,H)_*} \\ (\mathbb{R}^n\times W_\Lambda)/\Lambda_L & \xrightarrow[(L,e,H)_*]{} & (\mathbb{R}^n\times W_\Lambda)/\Lambda_H \end{array}$$

and the double coset formula is easily derived from this. □

Dress's equivariant Witt ring $GW(H,\mathbb{Z})$ [2] acts on $M(H)$ by tensor product. Recall that $GW(H,\mathbb{Z})$ is constructed using H-spaces. An H-space is a $\mathbb{Z}$-free (left) $\mathbb{Z}H$-module N together with a symmetric H-invariant non-singular form f: $N\times N\to\mathbb{Z}$. Let $N^*=\mathrm{Hom}_{\mathbb{Z}}(N,\mathbb{Z})$, then N^* is also a (left) $\mathbb{Z}H$-module. An element $h\in H$ acts on N^* by $h\cdot\alpha(y)=\alpha(h^{-1}\cdot y)$ for $\alpha\in N^*$, $y\in N$. By letting $(C_N)_0=N^*$ and $(C_N)_i=0$ for $i\neq 0$, we have a $\mathbb{Z}$-module chain complex C_N. We define a 0-dimensional symmetric Poincaré structure ϕ_f: $N\to N^*$ by $\phi_f(x)=f(x,-)$. By assumption ϕ_f is an isomorphism. If $\{e_1,\ldots,e_m\}$ is a (free $\mathbb{Z}$-module) basis of N^* and $\mathbb{Z}[S]$, h: $S\to(\mathbb{R}^n\times W_\Lambda)/\Lambda_H$, is a geometric $\mathbb{Z}$-module, then $N^*\otimes_{\mathbb{Z}}\mathbb{Z}[S]$ is a geometric $\mathbb{Z}$-module $\mathbb{Z}[\{e_1,\ldots,e_m\}\times S]$ on $(\mathbb{R}^n\times W_\Lambda)/\Lambda_H$, where a basis element $e_i\otimes x$ is sent to $h(x)$ for $x\in S$. Since Λ_H maps onto H, Λ_H acts on N^*. Λ_H is supposed to act diagonally on the free $\mathbb{Z}\Lambda_H$-module corresponding to this geometric $\mathbb{Z}$-module on $(\mathbb{R}^n\times W_\Lambda)/\Lambda_H$. This can be done through the following definition of morphisms. Suppose a: $M\to N$ is a $\mathbb{Z}H$-homomorphism of $\mathbb{Z}$-free $\mathbb{Z}H$-modules with $\mathbb{Z}$-bases $\{d_i\}$, $\{e_k\}$ and b:

$\mathbb{Z}[T] \to \mathbb{Z}[S]$ is a morphism of geometric $\mathbb{Z}$-modules on $(\mathbb{R}^n \times W_\Lambda)/\Lambda_H$. Let $\rho_j^{(x,y)}$ be a path in the morphism b. Pick a point $\tilde{x}$ (resp. $\tilde{y}) \in \mathbb{R}^n \times W_\Lambda$ in the orbit x (resp. $y) \in (\mathbb{R}^n \times W_\Lambda)/\Lambda_H$. Lift this path to a path in $\mathbb{R}^n \times W_\Lambda$ from $\tilde{x}$ to $\tilde{y}'$. Let $g \in \Lambda_H$ be the unique covering transformation which sends $\tilde{y}$ to $\tilde{y}'$. Write $g^{-1} \cdot a(d_i) = \Sigma\, a_{ijk} e_k$. Then consider a sum of paths $\Sigma\, a_{ijk} \rho_{ijk}^{(d_i \otimes x, e_k \otimes y)}$, where $\rho_{ijk}^{(d_i \otimes x, e_k \otimes y)} = \rho_j^{(x,y)}$. The tensor product $a \otimes b$ is a morphism from $M \otimes_{\mathbb{Z}} \mathbb{Z}[T]$ to $N \otimes_{\mathbb{Z}} \mathbb{Z}[S]$ defined as the sum of these sums for all $\rho_j^{(x,y)}$'s in b. The tensor product formula in [13, § 1.9], viewed as above, allows us to take tensor products with (C_N, ϕ_f), and this induces the desired action of $GW(H, \mathbb{Z})$ on $M(H)$. The following theorem can be proved in the same way as the proof of [4, theorem 2.3].

(4.8) **Theorem.** *M is a $GW(-, \mathbb{Z})$-module.*

Another Mackey functor M': $\mathscr{S} \to Ab$ can be defined by setting $M'(H) = L_j^{-\infty}((\mathbb{R}^n \times W_\Lambda)/\Lambda_H)$. f_* and f^* are defined in the same way as for M. We can prove:

(4.9) **Theorem.** *M' is a $GW(-, \mathbb{Z})$-module.*

As an immediate consequence of 4.8 and 4.9, we have the following theorem. See [2].

(4.10) **Theorem.** *If F is the family of the conjugacy classes of maximal hyperelementary subgroups of G. Then the following sequences are exact.*

$$0 \to H_j(\mathbb{R}^n/\Lambda; \mathbb{L}(p)) \xrightarrow{(\mathrm{res}_H)_H} \bigoplus_{H \in F} H_j(\mathbb{R}^n/\Lambda_H; \mathbb{L}(p_H)) \longrightarrow \bigoplus_{H,K,g} H_j(\mathbb{R}^n/\Lambda_{H \cap gKg^{-1}}; \mathbb{L}(p_{H \cap gKg^{-1}}))$$

$$0 \to L_j^{-\infty}((\mathbb{R}^n \times W_\Lambda)/\Lambda) \xrightarrow{(\mathrm{res}_H)_H} \bigoplus_{H \in F} L_j^{-\infty}((\mathbb{R}^n \times W_\Lambda)/\Lambda_H) \longrightarrow \bigoplus_{H,K,g} L_j^{-\infty}((\mathbb{R}^n \times W_\Lambda)/\Lambda_{H \cap gKg^{-1}}).$$

If F is the family of the conjugacy classes of maximal elementary subgroups of G, then these are exact when each term is tensored with $\mathbb{Z}[1/2]$.

The following is the main result of this paper.

(4.11) **Theorem.** *Suppose Λ is a finitely generated group which acts by isometries on $\mathbb{R}^n$ discretely, virtually faithfully, with compact quotient. Let p be the projection $(\mathbb{R}^n \times W_\Lambda)/\Lambda \to \mathbb{R}^n/\Lambda$, where W_Λ is a contractible free Λ-complex. Then there is a natural isomorphism*

$$1 \otimes a\colon\ \mathbb{Z}[1/2] \otimes H_j(\mathbb{R}^n/\Lambda; \mathbb{L}(p)) \to \mathbb{Z}[1/2] \otimes L_j^{-\infty}(\Lambda).$$

Proof. The composition

$$\mathbb{H}_{-j}(\mathbb{R}^n/\Lambda; \mathbb{L}(p)) \xrightarrow{A_{-j}} \mathbb{L}_j(\mathbb{R}^n/\Lambda; p) \xrightarrow{F} \mathbb{P}_j(\mathbb{R}^n/\Lambda; p)$$

has image in $\mathbb{P}'_j(\mathbb{R}^n/\Lambda; p) = \mathbb{P}_j(*; (\mathbb{R}^n \times W_\Lambda)/\Lambda \to *) = \mathbb{L}_j((\mathbb{R}^n \times W_\Lambda)/\Lambda)$, where A_{-j} is the assembly map (3.9) and F is the restriction to 0; so it defines a map

$\mathbb{H}_{-j}(\mathbb{R}^n/\Lambda;\mathbb{L}(p))\to\mathbb{L}_j((\mathbb{R}^n\times W_\Lambda)/\Lambda)$. This induces the desired map a: $H_j(\mathbb{R}^n/\Lambda;\mathbb{L}(p))\to L_j^{-\infty}(\Lambda)$.

Note that the theorem is a statement on the action of Λ on $\mathbb{R}^n$ and not on the group itself. We will prove the theorem inductively on the "size" $(n,h(\Gamma))$ of the action of Λ, where Γ is the image of Λ in $E(n)$ and $h(\Gamma)$ is the holonomy number of Γ; i.e., $h(\Gamma)$ is the order of the holonomy group of Γ if $\operatorname{rank}(\Gamma)=1$, and $h(\Gamma)$ is the minimum order of the holonomy group of a crystallographic group of rank $\geqq 2$ onto which Γ can map epimorphically if $\operatorname{rank}(\Gamma)\geqq 2$. We use the lexicographic order for the pairs $(n,h(\Gamma))$. If $n=0$, then $\mathbb{R}^n/\Lambda$ is a single point. Since $\mathbb{H}_{-j}(*;\mathbb{L}(p))\simeq\mathbb{L}_j(*;p)$ $(=\mathbb{L}_j(W_\Lambda/\Lambda))$ by 3.9, the theorem is obvious in this case. So assume that $n\geqq 1$.

First suppose that Γ satisfies (1) of 4.4. We use the notation in the remarks after 4.4. Recall that $\mathbb{R}^n/\Lambda$ fibres over S^1 with fibre diffeomorphic to $\mathbb{R}^{n-1}/\Lambda'$. The homology group $H_j(\mathbb{R}^n/\Lambda,\ \mathbb{R}^{n-1}/\Lambda';\mathbb{L}(p))$ is isomorphic to $H_{j-1}(\mathbb{R}^{n-1}/\Lambda';\mathbb{L}(p'))$. Thus we have a commutative ladder:

$$\begin{array}{ccccc}
\ldots\to & H_j(\mathbb{R}^{n-1}/\Lambda';\mathbb{L}(p')) & \longrightarrow & H_j(\mathbb{R}^n/\Lambda;\mathbb{L}(p)) & \\
 & \downarrow & & \downarrow & \\
\ldots\to & L_j^{-\infty}(\Lambda') & \longrightarrow & L_j^{-\infty}(\Lambda) & \\
\end{array}$$

$$\begin{array}{ccccc}
\to H_{j-1}(\mathbb{R}^{n-1}/\Lambda';\mathbb{L}(p')) & \longrightarrow & H_{j-1}(\mathbb{R}^{n-1}/\Lambda';\mathbb{L}(p'))\to\ldots & & \\
\downarrow & & \downarrow & & \\
\to L_{j-1}^{-\infty}(\Lambda') & \longrightarrow & L_{j-1}^{-\infty}(\Lambda') & \to\ldots &
\end{array}$$

The first row is the exact sequence for the pair $(\mathbb{R}^n/\Lambda,\ \mathbb{R}^{n-1}/\Lambda')$; and the second row is induced from the well-known exact sequence due to Wall, Shaneson, Farrell and Hsiang. By induction hypothesis and 5-lemma, $1\otimes a$ is proved to be an isomorphism.

Next suppose Γ satisfies 4.4(2). Choose an epimorphism $\phi\colon\Gamma\to\Gamma'$ onto a crystallographic group Γ' of rank $l\geqq 2$ whose holonomy group G' has order equal to $h(\Gamma)$. Apply corollary 4.4 to Γ'. As $\operatorname{rank}(\Gamma')\geqq 2$, the possibility (3) does not occur. In the case of possibility (1), Γ' maps epimorphically onto T, and hence so does Γ. We have already observed that the theorem holds true in this case. In the case of possibility (2), Γ' maps epimorphically onto a special crystallographic group Γ'' with holonomy group G'', and $|G''|\leqq|G|$. Γ'' is special, so it has rank $\geqq 2$. As $|G'|$ is the minimum, $|G''|=|G'|=h(\Gamma)$. So we may assume from the beginning that Γ' is special. So there exist a crystallographic group $\tilde{\Gamma}$ of rank $m\geqq 1$ and an infinite sequence of positive integers s such that any maximal hyperelementary subgroup of Γ_s' satisfies (4.1.1) or (4.1.2) with Γ and G replaced by Γ' and G'.

We first show that the map being considered is injective. Suppose y is an element of the kernel. We will show that $y=0$. We can regard $2^r y$ to be an element of $H_j(\mathbb{R}^n/\Lambda;\mathbb{L}(p))$, for some r. Represent $2^r y$ by a 0-cell ρ of $\mathbb{H}_{-j}(\mathbb{R}^n/\Lambda;\mathbb{L}(p))$. $A_{-j}\rho(0)$ represents the image $a(2^r y)$ by a, which represents 0

in $\mathbb{Z}[1/2]\otimes L_j^{-\infty}((\mathbb{R}^n\times W_\Lambda)/\Lambda)$. By choosing a sufficiently large r, we may assume that $A_{-j}\rho(0)$ represents 0 in $L_j^{-\infty}((\mathbb{R}^n\times W_\Lambda)/\Lambda)$. The Kan condition implies that there exists a 1-simplex σ of $\mathbb{L}_j((\mathbb{R}^n\times W_\Lambda)/\Lambda)$ which connects $A_{-j}\rho(0)$ and 0. Pull σ back to $\mathbb{R}^n\times W_\Lambda$ and δ be the radius measured in $\mathbb{R}^n$. It is a finite number. Choose a positive number s in (2) sufficiently large so that

$$2\delta K/\sqrt{s}<\varepsilon$$

where K is the Lipschitz constant of the affine surjection $f\colon \mathbb{R}^n\to\mathbb{R}^l$ induced by the epimorphism $\phi\colon \Gamma\to\Gamma'$ and ε is the positive number posited in 3.10, where we consider $\mathbb{P}_j'(\mathbb{R}^m/\tilde{\Gamma},\tilde{p},-)^{(\dim\sigma)}$. Note that ε depends not only on $\tilde{\Gamma}$ but also on the dimension of σ, so we cannot choose s which works for all y's. Now we have a commutative diagram:

$$\begin{array}{ccccc}
0\to & H_j(\mathbb{R}^n/\Lambda;\mathbb{L}(p)) & \xrightarrow{(\mathrm{res}_H)_H} & \bigoplus_H H_j(\mathbb{R}^n/\Lambda_H;\mathbb{L}(p_H)) & \to\dots \\
& \downarrow{\scriptstyle a} & & \downarrow{\scriptstyle \oplus a_H} & \\
0\to & L_j^{-\infty}((\mathbb{R}^n\times W_\Lambda)/\Lambda) & \xrightarrow[(\mathrm{res}_H)_H]{} & \bigoplus_H L_j^{-\infty}((\mathbb{R}^n\times W_\Lambda)/\Lambda_H) & \to\dots
\end{array},$$

where each row comes from the restriction maps corresponding to the conjugacy classes of the maximal hyperelementary subgroups H of Γ_s' (4.10). For H which projects to a proper subgroup of G', the size of the action of Λ_H is strictly smaller than the size of the action of Λ, because Λ_H maps onto $q^{-1}(H)$ and the order of the holonomy group of $q^{-1}(H)$ is strictly smaller than $|G'|=h(\Gamma)$. So, by induction hypothesis, $1\otimes a_H$ is an isomorphism, and hence $\mathrm{res}_H(2^r y)=0$. For H which projects onto G', we have a shrinking map $\alpha\colon \mathbb{R}^n/\Lambda_H\to\mathbb{R}^m/\tilde{\Gamma}$, after replacing H by H'. Notice that the map a_H factors through $H_j(\mathbb{R}^m/\tilde{\Gamma};\mathbb{L}(\alpha p_H))$, by regarding things controlled in $\mathbb{R}^n/\Lambda_H$ to be controlled in $\mathbb{R}^m/\tilde{\Gamma}$ via α, and that the image of $\mathrm{res}_H(2^r y)$ in $H_j(\mathbb{R}^m/\tilde{\Gamma};\mathbb{L}(\alpha p_H))$ which is represented by $\mathrm{res}_H(A_{-j}\rho(0))$ is 0, due to smallness of σ in $\mathbb{R}^m/\tilde{\Gamma}$ and the shrinking lemma. Now we claim that the map

$$\mathbb{Z}[1/2]\otimes H_j(\mathbb{R}^n/\Lambda_H;\mathbb{L}(p_H))\to\mathbb{Z}[1/2]\otimes H_j(\mathbb{R}^m/\tilde{\Gamma};\mathbb{L}(\alpha p_H))$$

is an isomorphism; this will imply that $\mathrm{res}_H(2^r y)=0$.

Recall that

$$H_*(X;\mathbb{L}(p))=[S^*,\lim FS\Omega^j(|\mathbb{L}_{-j}(p)|/i(X))].$$

To prove the claim, we show that the map

$$\lim FS\Omega^j(|\mathbb{L}_{-j}(p_H)|/(\mathbb{R}^n/\Lambda_H))\to\lim FS\Omega^j(|\mathbb{L}_{-i}(\alpha p_H)|/(\mathbb{R}^m/\tilde{\Gamma}))$$

is a homotopy equivalence up to 2-torsion. For the convenience of the proof, we replace $\mathbb{L}(p)$ by

$$\mathbb{L}'(p)=\{\bigcup_\Delta \mathbb{L}(p^{-1}(v_\Delta))\otimes\Delta\}/\sim,$$

where v_Δ is a carefully chosen vertex of a simplex Δ of the control space. See [9, proof of 8.6]. Recall that $(\alpha p_H)^{-1}(v_\Delta)=(\mathbb{R}^{n-m}\times W_\Delta)/\Lambda_H(v_\Delta)$, and that the size of the action of $\Lambda_H(v_\Delta)$ is strictly smaller than that of Λ. Therefore, by induction, $\mathbb{L}((\alpha p_H)^{-1}(v_\Delta))\simeq\mathbb{H}(\mathbb{R}^{n-m}/\Lambda_H(v_\Delta);\mathbb{L}(p_\Delta))$ modulo 2-torsion. Here p_Δ is the stratified system of fibrations $(\mathbb{R}^{n-m}\times W_\Delta)/\Lambda_H(v_\Delta)\to\mathbb{R}^{n-m}/\Lambda_H(v_\Delta)$. So, modulo 2-torsion,

$$
\begin{aligned}
&\lim_j FS\Omega^j(|\mathbb{L}'_{-j}(p_H)|/(\mathbb{R}^n/\Lambda_H))\\
&\quad=\lim_{i+k} FS\Omega^{i+k}\{(\bigcup_{\Delta'\subset\mathbb{R}^n/\Lambda_H}|\mathbb{L}_{-i-k}((p_H)^{-1}(v_{\Delta'}))|\times\Delta')/\sim\}/(\mathbb{R}^n/\Lambda_H)\\
&\quad\simeq\lim_i FS\Omega^i\{\bigcup_{\Delta\subset\mathbb{R}^m/\tilde\Gamma}\lim_k\Omega^k(|\mathbb{L}'_{-k-i}(p_\Delta)|/(\mathbb{R}^{n-m}/\Lambda_H(v_\Delta)))\times\Delta/\sim\}/(\mathbb{R}^m/\tilde\Gamma)\\
&\quad\simeq\lim_i FS\Omega^i((\bigcup_\Delta|\mathbb{H}_i(\mathbb{R}^{n-m}/\Lambda_H(v_\Delta);\mathbb{L}(p_\Delta))|\times\Delta)/\sim)/(\mathbb{R}^m/\tilde\Gamma)\\
&\quad\simeq\lim_i FS\Omega^i((\bigcup|\mathbb{L}_{-i}((\alpha p_H)^{-1}(v_\Delta))|\times\Delta)/\sim)/(\mathbb{R}^m/\tilde\Gamma)\\
&\quad=\lim_i FS\Omega^i(|\mathbb{L}'_{-i}(\alpha p_H)|/(\mathbb{R}^m/\tilde\Gamma)).
\end{aligned}
$$

Thus the claim is proved. Now we have proved that $\mathrm{res}_H(2^r y)=0$ for all H. Since the product of restriction maps $(\mathrm{res}_H)_H$ is injective, this implies that $2^r y$, and hence y, is 0; i.e., $1\otimes a$ is injective.

The onto part is similar. Pick an element y of $\mathbb{Z}[1/2]\otimes L_j^{-\infty}((\mathbb{R}^n\times W_\Delta)/\Lambda)$. Without loss of generality we may assume that y belongs to $L_j^{-\infty}((\mathbb{R}^n\times W_\Delta)/\Lambda)$. Represent y by a geometric $\mathbb{Z}$-module quadratic Poincaré complex y^* on $(\mathbb{R}^n\times W_\Delta)/\Lambda$. We want to show that the restriction image $\mathrm{res}_H(y)$ of y in each $\mathbb{Z}[1/2]\otimes L_j^{-\infty}((\mathbb{R}^n\times W_\Delta)/\Lambda_H)$ lies in the image of $1\otimes a_H$. If the size of the action of Λ_H is strictly smaller than that of Λ, then by induction hypothesis this is the case. If not, then there is a shrinking map α: $\mathbb{R}^n/\Lambda_H\to\mathbb{R}^m/\tilde\Gamma$ for some crystallographic group $\tilde\Gamma$ of rank $m\geqq 1$, and we can make the radius of the restriction image $\mathrm{res}_H(y^*)$ of y^* arbitrarily small on $\mathbb{R}^m/\tilde\Gamma$, by choosing a very large integer s. (Thus s is chosen after y^* is picked up and $\tilde\Gamma$ is independent of s and H, as in the injectivity part.) Now the characterization theorem implies that $\mathrm{res}_H(y)$ comes from an element of $\mathbb{Z}[1/2]\otimes H_j(\mathbb{R}^m/\tilde\Gamma;\mathbb{L}(\alpha p_H))$, and hence from an element y_H of $\mathbb{Z}[1/2]\otimes H_j(\mathbb{R}^n/\Lambda_H;\mathbb{L}(p_H))$. Finally $(y_H)_H$ is the image of an element x of $\mathbb{Z}[1/2]\otimes H_j(\mathbb{R}^n/\Lambda;\mathbb{L}(p))$ by the restriction map. Here we need to use the next column of the diagram, which is already known to be isomorphic or at least injective. (This is why we proved injectivity part first.) Since restriction maps are injective, $1\otimes a$ sends x to y, and this completes the onto part in case 4.4(2).

Lastly, if Γ satisfies (3), we use elementary induction instead of hyperelementary induction. This is the only place where we need to use $\mathbb{Z}[1/2]\otimes$. Now the induction step is completed, and the theorem is proved. □

Note that, if Γ contains no element of order 2, then D_∞ never shows up in the induction steps. Therefore, for Λ acting on $\mathbb{R}^n$ with no elements of period 2, the theorem holds true without taking tensor products $\mathbb{Z}[1/2]\otimes$.

As a final remark, let us relate this result to the usual assembly map. Consider the following diagram:

$$\begin{array}{ccccc} H_*(B_\Lambda;\mathbb{L}(1)) & \xrightarrow{f} & H_*(\mathbb{R}^n/\Lambda;\mathbb{L}(p)) & \xrightarrow{a} & L_*^{-\infty}(\Lambda) \\ & \searrow^{p_*} & \downarrow g & & \\ & & H_*(\mathbb{R}^n/\Lambda;\mathbb{L}(1)) & & \end{array}$$

where $B_\Lambda = (\mathbb{R}^n \times W_\Lambda)/\Lambda$ is a classifying space of Λ. The composition af is the usual assembly map. This diagram is induced by a diagram of morphisms of stratified system of fibrations:

$$\begin{array}{ccc} (1: B_\Lambda \to B_\Lambda) \to & (p: B_\Lambda \to \mathbb{R}^n/\Lambda) & \to (B_\Lambda \to *) \\ & \searrow \quad \downarrow & \\ & (1: \mathbb{R}^n/\Lambda \to \mathbb{R}^n/\Lambda). & \end{array}$$

Note that the point inverses of p are the classifying spaces of the isotropy subgroups of Λ, which are finite, and hence are $\mathbb{Z}[1/k]$-acyclic, where k is any integer such that k and the order of these finite subgroups are coprime. This implies that p is an ordinary-$\mathbb{Z}[1/k]$-homology isomorphism, and hence p_* is a $(\mathbb{Z}[1/k] \otimes)$-isomorphism. And this, in return, implies that f is a split $(\mathbb{Z}[1/k] \otimes)$-injection. The main theorem states that a is a $(\mathbb{Z}[1/2] \otimes)$-isomorphism. Therefore we get:

(4.12) **Corollary.** *Suppose Λ is a group which acts by isometries on $\mathbb{R}^n$ discretely, virtually faithfully, with compact quotient, and k is an integer which is coprime with the order of each isotropy subgroup of Λ. Then the assembly map*

$$\mathbb{Z}[1/2, 1/k] \otimes H_*(B_\Lambda;\mathbb{L}(1)) \to \mathbb{Z}[1/2, 1/k] \otimes L_*^{-\infty}(\Lambda)$$

is a split injection.

This is a stronger form of the Novikov conjecture, which states that the assembly map is injective when tensored with $\mathbb{Q}$.

References

1. Cohen, M.: A course in simple-homotopy theory. Graduate texts in mathematics, vol. 10. Berlin-Heidelberg-New York: Springer 1973
2. Dress, A.W.M.: Induction and structure theorems for orthogonal representations of finite groups. Ann. Math. **102**, 291–325 (1975)
3. Farkas, D.: Crystallographic groups and their mathematics. Rockey Mountain J. Math. **11**, 511–551 (1981)
4. Farrell, F.T., Hsiang, W.C.: Rational L-groups of Bieberbach groups. Comment. Math. Helv. **52**, 89–109 (1977)
5. Farrell, F.T., Hsiang, W.C.: The Whitehead group of poly-(finite or cyclic) groups. J. Lond. Math. Soc. **24**, (2) 308–324 (1981)
6. Farrell, F.T., Hsiang, W.C.: Topological characterization of flat and almost flat Riemannian manifolds M^n $(n \neq 3, 4)$. Am. J. Math. **105**, 641–672 (1983)
7. Quinn, F.S.: A geometric formulation of surgery. Thesis, Princeton Univ., 1969

8. Quinn, F.S.: Ends of maps I. Ann. Math. **110**, 275–331 (1979)
9. Quinn, F.S.: Ends of maps II. Invent. Math. **68**, 353–424 (1982)
10. Quinn, F.S.: Geometric algebra. Lect. Notes Math., vol. 1126, pp. 182–198. Berlin-Heidelberg-New York-Tokyo: Springer 1985
11. Ranicki, A.A.: Algebraic L-theory, II: Laurent extensions. Proc. Lond. Math. Soc. **27**, (3) 126–158 (1973)
12. Ranicki, A.A.: The algebraic theory of surgery I, Foundations. Proc. Lond. Math. Soc. **40**, (3) 87–192 (1980)
13. Ranicki, A.A.: Exact sequences in the algebraic theory of surgery. Math. Notes, vol. 26. Princeton: Princeton Univ. Press 1981
14. Rourke, C.P., Sanderson, B.J.: Δ-sets I: Homotopy theory. Quart. J. Math. **22**, 321–338 (1971)
15. Spanier, E.H.: Algebraic topology. New York: McGraw Hill 1966
16. Wall, C.T.C.: Surgery on compact manifolds. New York-London: Academic Press 1970
17. Weiss, M.: Surgery and the generalized Kervaire invariant, I. Proc. Lond. Math. Soc. **51**, (3) 146–192 (1985)

Oblatum 17-X-1984 & 14-X-1985 & 21-X-1986

Invent. math. 88, 603-618 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Topology of the complement of real hyperplanes in $\mathbb{C}^N$

M. Salvetti

Dipartimento di Matematica, Via F. Buonarrotti 2, I-56100 Pisa, Italy

Introduction

Consider a union of real[1] affine hyperplanes in $\mathbb{C}^N$ with complement Y. In [1] it was conjectured that if W is a finite Coxeter group generated by reflections in $\mathbb{R}^N$ and the hyperplanes are the fixed sets of the reflections in W, then Y is a $K(\pi, 1)$. A more general result was proved in [2], that is Y is a $K(\pi, 1)$ when the chambers of the subdivision of $\mathbb{R}^N$ induced by the hyperplanes are simplicial cones. Later works have considered subdivisions of $\mathbb{R}^N$, from many points of view (see [3] for references) and recently, another result ([4]) along the lines of [2] appeared.

In this paper we also consider the problem of finding the homotopy groups of Y. In the first part we construct explicitly a CW-complex $X \subset Y$ of dimension N, and prove (Theorem 1) that X and Y are homotopically equivalent. The real part of X is the dual cellular complex of the subdivision of $\mathbb{R}^N$ induced by the hyperplanes. Moreover, there are as many 0-cells in X as the chambers, and in general for each facet F^k of codimension k of the subdivision there are as many k-cells in X as the chambers which contain F^k in their closure.

When the hyperplanes are those associated to some Coxeter group W, then the 1-skeleton of X turns out to be the Cayley graph of W.

In the second part we give some general results on $\pi_1(Y)$, and describe some procedures for finding a presentation of it, complementary to the Van Kampen method and [7, 8]. In practice, all relations can be deduced from the fact that two minimal and "positive" paths with the same ends are homotopic.

In the case of a Coxeter group W, CW-complex (or semicomplex) decompositions were used in [10] (for type A_n) and [11] (for affine Weyl groups) to find a presentation of the orbit space Y/W. Our construction could be used also to that extent. In fact, X can be chosen invariantly for the action of W, and the homotopy equivalence between Y and X induces a homotopy equivalence between Y/W and X/W.

[1] i.e., with real equation

As an application to higher homotopy groups we consider the general position case, already considered in [9]. In this case the complex X is a union of cellularly decomposed N-tori ($(S^1)^N$) attached by $(N-1)$-cells, and we show that it can be obtained from its real part by suitable identifications (Lemma 14). This yields a more compact proof of the fact that, if the number of hyperplanes is greater than the dimension N, then Y is not a $K(\pi, 1)$. Precisely, the universal cover of Y is homotopy equivalent to a bouquet of infinitely many N-spheres, so $\pi_i(Y)=0$ for $1 \leqq i < N$ and $\pi_N(Y)$ is infinitely generated free abelian (Theorem 2).

Because of the very explicit description of the regular complex X, we believe that it should be possible to compute the homotopy of Y in other not yet solved cases.

Deligne's work can very likely be interpreted in a natural way as a study of the complex X (in the simplicial case) and we hope we can return to this in the future. The point is that the equivalence relation between "galeries" used in [2] is related to homotopy equivalence between paths in $\pi_1(X)$.

The construction of X was also inspired by [6], in which counting-facets formulas are given. Since these formulas depend only on the geometrical lattice associated to the arrangement of hyperplanes, it is a natural conjecture that the homotopy type of Y depends only on the associated lattice (this is true in the general position case).

The notation and nomenclature are in part derived from [2] and [6].

Part One (construction of the complex $X \subset Y$ and of the homotopy $Y \to X$)

Let $\{M_j\}_{j \in J}$ be a locally-finite family of affine hyperplanes in $\mathbb{R}^N$. For brevity, the hyperplanes of the family will be called cuts ([6]). There is a subdivision of $\mathbb{R}^N$ into facets: chambers are codimension 0 facets, faces are codimension 1 facets. Let $\mathscr{M}$ denote the set of all cuts and facets, and $\mathscr{M}'$ the set of all facets. If $\bigcap_{j \in J} M_j \neq \emptyset$ $\mathscr{M}$ is said to be central. The support $|F|$ of a facet F is the affine subspace generated by F; each facet is open in its support. Given F, let $pr_F: \mathbb{R}^N \to \mathbb{R}^N/|F|$ be the affine projection, which sends cuts containing F into homogeneous hyperplanes of $\mathbb{R}^N/|F|$. These give a central decomposition $\mathscr{M}_F$ of $\mathbb{R}^N/|F|$. Let $\pi'_F: \mathscr{M}' \to \mathscr{M}'_F$ be the map sending the facet E into the smallest facet containing $pr_F(E)$. If ${}^F\mathscr{M} \subset \mathscr{M}$ is the set formed by all the cuts containing F and by all facets E having F as a facet (i.e. $E^- \supset F$) then π'_F induces a bijection (called again π'_F) $\pi'_F: {}^F\mathscr{M} \to \mathscr{M}_F$. If $E \in {}^F\mathscr{M}'$, then $\operatorname{codim}_{\mathbb{R}^N}(E) = \operatorname{codim}_{\mathbb{R}^N/|F|}(\pi'_F(E))$, and in general $\operatorname{codim}_{\mathbb{R}^N}(E) \geqq \operatorname{codim}_{\mathbb{R}^N/|F|}(\pi'_F(E))$; hence π'_F takes chambers into chambers.

For each facet F, choose one point ${}^Fw \in F$ and let $\mathscr{V}^i$ denote the set of such points belonging to codimension i facets. Clearly, there is a bijection between the sets $\mathscr{M}'$ and $\mathscr{V} = \bigcup_i \mathscr{V}^i$. Let G be the (undirected) graph which has one vertex for each chamber, and one edge joining two vertices when they belong to adjacent chambers. We shall consider G as embedded in $\mathbb{R}^N$ in the following way: the vertex-set VG is $\mathscr{V}^0$; next, if C and C' are adjacent chambers having the common face F, take as edge joining ${}^Cw$ to ${}^{C'}w$ the join in $\mathbb{R}^N$:

$${}^F w \vee \{{}^C w, {}^{C'} w\} = \{x = (1-\lambda)\,{}^F w + \lambda\,{}^C w,\ \lambda \in [0,1]\} \cup \{x = (1-\lambda)\,{}^F w + \lambda\,{}^{C'} w,\ \lambda \in [0,1]\}.$$

Denote by EG the edge-set of G, and by CG the set of edge-paths of G. A length $l(c)$ is associated with a path $c \in CG$ (namely, the number of edges in c); call c minimal if its length is a minimum among those of the paths joining the ends of c.

Next three lemmas relate $\mathscr{M}$ to G. Proofs of the first two can be found in [2; Lemma 1.2, Lemma 1.3].

Lemma 1. *Let C_1, C_2, C_3 be chambers and $\mathscr{M}(C_i, C_j)$ be the set of cuts separating C_i from C_j. Then*

$$\mathscr{M}(C_1, C_2) = [\mathscr{M}(C_1, C_3) \setminus \mathscr{M}(C_2, C_3)] \cup [\mathscr{M}(C_2, C_3) \setminus \mathscr{M}(C_1, C_3)].$$

Lemma 2. *A path c joining ${}^C w \in C$ to ${}^{C'} w \in C'$ is minimal if and only if it crosses once and only once the cuts separating C from C', and none of the other cuts, so that $l(c) = \#\mathscr{M}(C, C')$.*

For each facet F, let ${}^F VG$ be the set of the vertices belonging to a chamber of ${}^F\mathscr{M}$, and ${}^F EG$ be the set of edges crossing faces of ${}^F\mathscr{M}$.

Lemma 3. *Given $\bar{w} \in VG$, for each facet F there is a unique vertex $w_{F,\bar{w}} \in {}^F VG$ having the shortest distance from $\bar{w}$. If $w \in {}^F VG$, one minimal path in CG between $\bar{w}$ and w is the composition of a minimal path joining $\bar{w}$ to $w_{F,\bar{w}}$ (which does not cross cuts containing F) with a minimal path joining $w_{F,\bar{w}}$ to w (which is composed with edges of ${}^F EG$).*

Proof. Let $\bar{C}$ be the chamber containing $\bar{w}$. There is a unique chamber $C_{F,\bar{w}} \in {}^F\mathscr{M}$ such that $\pi'_F(C_{F,\bar{w}}) = \pi'_F(\bar{C})$. Choose $w_{F,\bar{w}}$ as the vertex of G belonging to $C_{F,\bar{w}}$; then using Lemma 1 and 2 the proof is easily achieved. Q.E.D.

By Lemma 3 one can construct a map $\mathscr{M}' \to VG: F \to w_{F,\bar{w}}$, depending on the vertex $\bar{w} \in VG$.

Now order $\mathscr{M}'$ by the relation: $E \geqq F$ iff $E^- \supset F$ (this is the opposite of [6]). The bijection between $\mathscr{M}'$ and $\mathscr{V}$ induces an ordering on V (namely ${}^E w \geqq {}^F w$ iff $E \geqq F$).

Lemma 4. *Suppose $\mathscr{M}$ central and $\{0\} \in \mathscr{M}$ (i.e. $\bigcap_{j \in J} M_j$ is not empty and it is given by the only point 0). Then if $F \neq \{0\}$ is a facet and $x \in F$, for any point $y \in F$ there exists a number $t_0 > 0$ such that the half-line starting from x and containing ty, $t \geqq 0$, meets ∂F if $t < t_0$, and does not meet any cut of $\mathscr{M} \setminus {}^F\mathscr{M}$ if $t \geqq t_0$.*

Proof. Since $\{0\} \in \mathscr{M}$ and $F \neq \{0\}$ there are cuts in $\mathscr{M} \setminus {}^F\mathscr{M}$. If $M \in \mathscr{M} \setminus {}^F\mathscr{M}$, there exists $0 < t(M) < +\infty$ such that the half-line from x to ty meets M if $t < t(M)$, is parallel to M if $t = t(M)$, and does not meet M if $t > t(M)$. (If y is in the half line $\{tx : t > 0\}$ then the half-line from x to ty will degenerate to a point for $t = t(M)$). Let $t_0 = \max\{t(M)\}$, $t_0 = t(M_0)$. Then the segment $[0, t_0 y)$ projects from x onto a half-line contained in $M_0 \cap \partial F$. Q.E.D.

In particular, if $\{0\} \in \mathscr{M}$, the closure of each facet is a polyedrical convex cone with vertex at 0. More precisely, one has the following fact.

Lemma 5. *Suppose again $\mathscr{M}$ central and $\{0\}\in\mathscr{M}$. To each $x\in\mathbb{R}^N\setminus\{0\}$ is associated a unique chain ${}^{F^{j_0}}w>\ldots>{}^{F^{j_k}}w$, where $\operatorname{codim}(F^{j_i})=j_i$ and $j_k<N$, such that the half-line $s(x)=\{tx: t>0\}$ meets ${}^{F^{j_0}}w\vee\ldots\vee{}^{F^{j_k}}w$ in a unique point x', with $x'=\sum\lambda_i\,{}^{F^{j_i}}w$, $\sum\lambda_i=1$, $\lambda_i\in(0,1]$ for $i=1,\ldots,k$.*

Proof. Let $x\in\mathbb{R}^N\setminus\{0\}$ and let F^{j_0} be the facet containing x. If ${}^{F^{j_0}}w\in s(x)$, then $x'={}^{F^{j_0}}w$ and the chain will be ${}^{F^{j_0}}w$. If ${}^{F^{j_0}}w\notin s(x)$, project $s(x)$ from ${}^{F^{j_0}}w$ onto ∂F^{j_0}, as in Lemma 4, obtaining a half-line $s_1(x)\subset F^{j_1}$ ($F^{j_0}>F^{j_1}$). If ${}^{F^{j_1}}w\in s_1(x)$ then x' is a convex combination of ${}^{F^{j_0}}w$, ${}^{F^{j_1}}w$, and the chain will be ${}^{F^{j_0}}w>{}^{F^{j_1}}w$. Otherwise, project $s_1(x)$ from ${}^{F^{j_1}}w$ onto ∂F^{j_1} obtaining a half-line $s_2(x)\subset F^{j_2}$ ($F^{j_1}>F^{j_2}$), and consider ${}^{F^{j_2}}w$. So, the chain is found after at most N steps. Q.E.D.

By completeness sake, to the point $x=0={}^{F^N}w$ will be associated the chain ${}^{F^N}w$.

Following lemma applies with no restrictions on $\mathscr{M}$.

Lemma 6. *Let $F_0^k\in\mathscr{M}$; then $D^k_{F_0^k}=\bigcup[{}^{F^0}w\vee\ldots\vee{}^{F^{k-1}}w\vee{}^{F_0^k}w]$, where the union is taken over all chains ${}^{F^0}w>\ldots>{}^{F^{k-1}}w>{}^{F_0^k}w$ ($\operatorname{codim}(F^i)=i$) is a triangulated k-cell in $\mathbb{R}^N$, with boundary $S^{k-1}_{F_0^k}=\bigcup[{}^{F^0}w\vee\ldots\vee{}^{F^{k-1}}w]$ the union being taken over the same chains.*

Proof. Consider the bijection $\pi'_{F_0^k}\colon {}^{F_0^k}\mathscr{M}\to\mathscr{M}_{F_0^k}$. Since $\mathscr{M}_{F_0^k}$ is central and clearly $\mathscr{M}_{F_0^k}\ni\{0\}=\pi'_{F_0^k}(F_0^k)$, we can apply Lemma 5 to $\mathscr{M}_{F_0^k}$. For every facet $\bar F^j=\pi'_{F_0^k}(F^j)$ of $\mathscr{M}_{F_0^k}$ select the point ${}^{\bar F^j}w=pr_{F_0^k}({}^{F^j}w)$. By Lemma 5, $\bar D^k_{F_0^k}=\bigcup[{}^{\bar F^0}w\vee\ldots\vee{}^{\bar F^{k-1}}w\vee{}^{\bar F_0^k}w]$, where ${}^{\bar F_0^k}w=0$, is a k-cell with boundary $\bar S^{k-1}_{F_0^k}=\bigcup[{}^{\bar F^0}w\vee\ldots\vee{}^{\bar F^{k-1}}w]$, the union being taken over all chains ${}^{\bar F^0}w>\ldots>{}^{\bar F_0^k}w$. But clearly $pr_{F_0^k}(D^k_{F_0^k})=\bar D^k_{F_0^k}$ and $pr_{F_0^k}|D^k_{F_0^k}$ is a homeomorphism; so the proof is complete. Q.E.D.

We can look at $D^k_{F_0^k}$ as at the "dual" cell of F_0^k, in the sense explained after Lemma 8.

Lemma 7. *Let $\bar k$ be the maximum of the codimensions of the facets in $\mathscr{M}'$, and $F_0^{\bar k}$ a facet of codimension $\bar k$. Then all cuts are parallel to $F_0^{\bar k}$, or contain it (so $F_0^{\bar k}=|F_0^{\bar k}|$). Conversely, if all cuts are parallel to or contain a facet F, then $\operatorname{codim}(F)=\bar k$ and F is parallel or equal to $F_0^{\bar k}$.*

Proof. If there exists a cut M not parallel to $F_0^{\bar k}$, then

$$\dim(M\cap|F_0^{\bar k}|)=\dim|F_0^{\bar k}|-1,$$

so there are facets of codim $>\bar k$. Conversely, if all cuts are parallel to or contain F, a translation takes $|F|$ into $|F_0^{\bar k}|$, and by the maximality of $\bar k$ one has the thesis. Q.E.D.

By Lemma 7, the minimal elements of $\mathscr{M}'$ are parallel facets (affine subspaces) of the same codimension. Moreover, following lemma holds.

Lemma 8. *Let $\bar{k}$ be as in Lemma 7. Given ${}^{F^j}w \in \mathscr{V}$, there exist (in general not unique) two chains ${}^{F^0}w, \ldots, {}^{F^{j-1}}w$ and ${}^{F^{j+1}}w, \ldots, {}^{F^{\bar{k}}}w$ such that ${}^{F^0}w > \ldots > {}^{F^{j-1}}w > {}^{F^j}w > {}^{F^{j+1}}w > \ldots > {}^{F^{\bar{k}}}w$.*

Proof. First we prove the existence of the chain ${}^{F^j}w > \ldots > {}^{F^{\bar{k}}}w$. If all cuts are parallel to or contain F^j then $j = \bar{k}$ by Lemma 7; so the chain is ${}^{F^j}w$ only. Otherwise, the cuts not parallel to F^j determine on $|F^j|$ a subdivision such that F^j is a chamber. Let F^{j+1} be a face of F^j. Now we repeat the same argument for F^{j+1}, and so on, until we find a facet of codimension $\bar{k}$, $F^{\bar{k}}$.

The existence of the chain ${}^{F^0}w > \ldots > {}^{F^{j-1}}w$ is reduced to the precedent case. Indeed, consider $\mathscr{M}_{F^j}$, whose minimum element is $\{0\}$, and choose any chain ${}^{F^0}w > \ldots > {}^{F^{j-1}}w > 0$, with $\pi'_{F^j}(F^i) = \bar{F}^i$ and ${}^{\bar{F}^i}w = pr_{F^j}({}^{F^i}w)$, $i < j$. Q.E.D.

Let $Q = Q^{\bar{k}} = \bigcup_{F \in \mathscr{M}'} D^j_{F^j}$, where $D^j_{F^j}$ is constructed similar to $D^k_{F^k_0}$ in Lemma 6 ($\bar{k}$ as above). Then Q is a cellular complex: it is a realization in $\mathbb{R}^N$ of the dual complex of (the complex associated to) $\mathscr{M}'$, endowed with a barycentrical triangulation. One has $\partial D^j_{F^j} = \bigcup_{F^{j-1} > F^j} D^{j-1}_{F^{j-1}}$, and by Lemma 8 each $D^{j-1}_{F^{j-1}}$ is in the boundary of some $D^j_{F^j} (j < \bar{k})$. So $Q^{\bar{k}}$ is a regular $\bar{k}$-cell complex.

Now let F^k be a facet of maximal codimension. If S is any subspace complementary to $|F^{\bar{k}}|$ in $\mathbb{R}^N$, the cuts determine over S a subdivision: by Lemma 7 each facet intersects S along a facet of the same codimension (in S). One says that $F \in \mathscr{M}'$ is relatively bounded (see [6; p. 25]) iff $S \cap F$ is bounded. This definition does not depend on the complementary S. Set B as the union of the bounded facets of $\mathscr{M}$, $B^{(r)}$ as the union of the relatively bounded facets, and note that $B \neq \emptyset$ iff $\bar{k} = N$ (in which case $B = B^{(r)}$). When $\bar{k} = N$, obviously Q is a neighbourhood of B; furthermore, Q contracts over B. In fact, all $\mathbb{R}^N$ contracts onto B: because $F^- \cap B \neq \emptyset$ for each F, by Lemma 8, and F^- is the join between $F^- \cap B$ and $F^- \cap (\infty\text{-hyperplane})$. So each F^- contracts onto $F^- \cap B$, giving a contraction of all $\mathbb{R}^N$. If $\bar{k} < N$, choose a complementary space $S^{\bar{k}}$ as above. The cuts will determine on S a subdivision of maximal codimension $\bar{k}$ with bounded space B_S, and S contracts onto B_S by the previous case. But it is easy to see that there is a contraction of the pair $(\mathbb{R}^N, B^{(r)})$ onto (S, B_S). So we have the following result (see [6, p. 33–34], in which only partial results are proved).

Lemma 9. *Let $B^{(r)}$ be as above; then $B^{(r)}$ is contractible.*

It follows that Q is contractible.

Now let $\mathbb{C}^N$ be the space obtained by complexification of $\mathbb{R}^N$ (with points $x + iy$, $x, y \in \mathbb{R}^N$) and $\mathscr{R}, \mathscr{I} : \mathbb{C}^N \to \mathbb{R}^N$ the real and imaginary parts. Let $Y = \mathbb{C}^N \setminus \bigcup_j M_{j,\mathbb{C}}$, where $M_{j,\mathbb{C}}$, $j \in J$, are the hyperplanes of $\mathbb{C}^N$ complexification of the M_j. The following obvious lemma is useful.

Lemma 10. *A point $x + iy \in \mathbb{C}^N$ is in Y iff, for each cut M, $x \in M \Rightarrow y \notin M_0$, where M_0 is the direction of M (i.e. the hyperplane parallel to M which passes through the origin).*

We now construct a CW-complex $X \subset Y$. Let $\bar{w} \in VG$, $F^k \in \mathscr{M}'$. Define an embedding $\phi_{F^k, \bar{w}}: D^k \to Y$ as follows. Identify, by Lemma 6, the k-cell D^k with $D^k_{F^k}$. Each $x \in D^k_{F^k}$ is written as $x = \sum_{i=0}^{k} \lambda_i {}^{F^i}w$, $\sum \lambda_i = 1$, $\lambda_i \in [0, 1]$, where ${}^{F^0}w > \ldots > {}^{F^k}w$ and $\operatorname{codim}(F^i) = i$. We set:

$$\mathscr{R}\phi_{F^k, \bar{w}}(x) = x$$

$$\mathscr{I}\phi_{F^k, \bar{w}}(x) = \sum_{i=0}^{k} \lambda_i (w_{F^i, \bar{w}} - {}^{F^i}w), \quad \text{if } x = \sum_{i=0}^{k} \lambda_i {}^{F^i}w.$$

If x belongs to the cut M of the chamber F^0, so that $\lambda_0 = 0$ but some $\lambda_i > 0$, then $y = \mathscr{I}\phi_{F^k, \bar{w}}(x) \notin M_0$. Indeed, y is a convex combination with coefficients $\geqq 0$, not all 0, of the points $w_{F^i, \bar{w}} - {}^{F^i}w$, $i \geqq 1$. By Lemma 3 the points $w_{F^i, \bar{w}}$ stand all on the same side as $\bar{w}$ with respect to M. So $w_{F^i, \bar{w}} - {}^{F^i}w$ stand all on the same side with respect to M_0. Hence $\phi_{F^k, \bar{w}}(D^k_{F^k}) \subset Y$, by Lemma 10. Moreover, $\phi_{F^k, \bar{w}}$ is obviously an embedding. Denote its image $\phi_{F^k, \bar{w}}(D^k_{F^k})$ (a k-cell in Y) by $D^k_{F^k, \bar{w}}$. Note that if $\bar{w}$, $\bar{\bar{w}}$ are such that $w_{F^k, \bar{w}} = w_{F^k, \bar{\bar{w}}}$ then $D^k_{F^k, \bar{w}} = D^k_{F^k, \bar{\bar{w}}}$. Therefore we shall obtain, relatively to F^k, as many distinct $D^k_{F^k, \bar{w}}$ as the $\bar{w} \in {}^{F^k}VG$.

One has $D^0_{F^0, \bar{w}} = {}^{F^0}w + i\,(w_{F^0, \bar{w}} - {}^{F^0}w) = {}^{F^0}w$, for all $\bar{w}$. Diagram 1 shows the two 1-cells $D^1_{F^1, \bar{w}_1}$, $D^1_{F^1, \bar{w}_2}$ in the case that $N = 1$, the only cut is $\{0\} \subset \mathbb{R}^1$, $\bar{w}_1 = 1 \in F_1$ and $\bar{w}_2 = -1 \in F_2$.

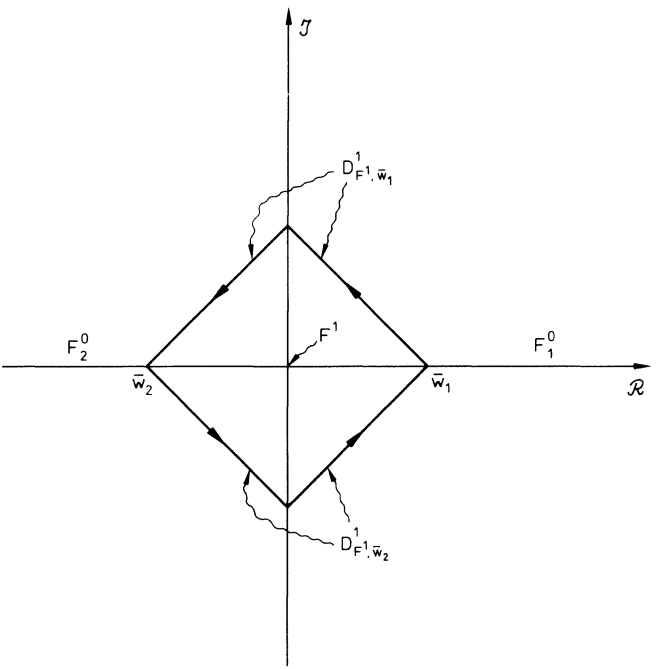


Fig. 1

We shall give a geometrical description of the higher dimensional cells in part two.

Let X be the complex obtained by using $\phi_{F^k,\bar{w}}$ to attach the j-cells $D^j_{F^j}$, $j=0, 1, \ldots, \bar{w}\in VG$, so $X=\bigcup_{F^j,\bar{w}} D^j_{F^j,\bar{w}}$. By construction, the real part of X is the complex Q. Analogous to the formula for Q, $\partial D^j_{F^j,\bar{w}}=\bigcup_{F^{j-1}>F^j} D^{j-1}_{F^{j-1},\bar{w}}$ (the same $\bar{w}$) and each $D^{j-1}_{F^{j-1},\bar{w}}$ is in the boundary of some $D^j_{F^j,w}$ if $j\leqq\bar{k}$/($\bar{k}$ as in Lemma 7). So X is a regular $\bar{k}$-cell complex, with i-skeleton $X_i=\bigcup_{F^i,\bar{w}} D^i_{F^i,\bar{w}}$. There is a bijection between X_0 and VG, and the number v_i of i-cells of X is (if finite) $v_i=\sum_{F^i}\#({}^{F^i}VG)$. In particular $v_0=\#\{\text{chambers}\}$, $v_1=2\#\{\text{faces}\}$, and $\chi(X)=\sum_{F\in\mathscr{M}'}(-1)^k\#({}^{F^k}VG)$.

Now we turn to the construction of a homotopy equivalence Φ_t, $t\in[0, 1]$, between Y and X. Let $\bar{k}$ be as above, $F^{\bar{k}}_0\in\mathscr{M}$. It is easy to see (by Lemma 7) that the affine projection $\mathbb{C}^N\to\mathbb{C}^N/|F^{\bar{k}}_0|_{\mathbb{C}}$ where $|F^{\bar{k}}_0|_{\mathbb{C}}=$complexification of $|F^{\bar{k}}_0|$, induces a contraction of Y onto $Y\cap|F^{\bar{k}}_0|^{\perp}_{\mathbb{C}}$, where $|F^{\bar{k}}_0|^{\perp}_{\mathbb{C}}\cong\mathbb{C}^{\bar{k}}$ is the orthogonal complement (in $\mathbb{C}^N$) of $|F^{\bar{k}}_0|_{\mathbb{C}}$. One can choose all the points of $\mathscr{V}$ in $|F^{\bar{k}}_0|^{\perp}\subset\mathbb{R}^N$, so that $X\subset|F^{\bar{k}}|^{\perp}_{\mathbb{C}}$. Therefore we are reduced to find Φ_t when $\bar{k}=N$, i.e. the facets of maximal codimension are points. We shall assume this for the remaining of the paper.

First, it is easy to construct a contraction $\Phi^{(1)}_t\colon Y\to Y$, $t\in[0, 1]$, of Y onto $\tilde{Q}=(\mathscr{R}\mid Y)^{-1}(Q^N)$ which works only on the real part, contracting the facet F onto $F\cap Q$. It will be similar to that which contracts $\mathbb{R}^N$ onto B, used before Lemma 9. Thus $\Phi^{(1)}_0=\mathrm{id}$, $\Phi^{(1)}_1(Y)=\tilde{Q}$, $\Phi^{(1)}_t\mid\tilde{Q}=\mathrm{id}$.

Now, for each facet $F\in\mathscr{M}$ and $\bar{w}\in VG$, consider the set

$$C(F,\bar{w})=\{[pr_F^{-1}(\pi'_F(C_{\bar{w}}))]^- - {}^Fw\}\subset\mathbb{R}^N$$

where $C_{\bar{w}}$ is the chamber containing $\bar{w}$ (minus sign means difference between sets; so, a translation of the first set by $-{}^Fw$). Observe that, if $\bar{w},\bar{w}'\in{}^FVG$ and $\bar{w}\neq\bar{w}'$ then the interior parts int $C(F,\bar{w})$, int $C(F,\bar{w}')$ have empty intersections. Moreover $\bigcup_{\bar{w}\in{}^FVG} C(F,\bar{w})=\mathbb{R}^N$, and $F>F'$ implies $C(F,\bar{w})\supset C(F',\bar{w})$.

It is possible to parametrize $C(F,\bar{w})$ as follows: if $y\in C(F,\bar{w})$, then $y=\mu(w_{F,\bar{w}}-{}^Fw)+\alpha$, $\mu\geqq0$, $\alpha\in\partial C(F,\bar{w})$. The pair (μ,α) is uniquely determined since $C(F,\bar{w})$ is the disjoint union of the translates of $\partial C(F,\bar{w})$ into the direction of $w_{F,\bar{w}}-{}^Fw$. If $y\in C(F,\bar{w})\cap C(F,\bar{w}')$, $\bar{w},\bar{w}'\in{}^FVG$, $\bar{w}\neq\bar{w}'$, then necessarily $y\in\partial C(F,\bar{w})\cap\partial C(F,\bar{w}')$ and in the expressions

$$y=\mu(w_{F,\bar{w}}-{}^Fw)+\alpha=\mu'(w_{F,\bar{w}'}-{}^Fw)+\alpha'$$

one has $\mu=\mu'=0$, $\alpha=\alpha'=y$.

For each full chain ${}^{F^0}w>\ldots>{}^{F^N}w$ and $\bar{w}\in VG$, introduce the subset of Y given by:

$$Y({}^{F^0}w>\ldots>{}^{F^N}w,\bar{w})=\{x+iy\in Y\colon x\in{}^{F^0}w\vee\ldots\vee{}^{F^N}w,\ y\in C(F^N,\bar{w})\}.$$

Clearly $\bigcup Y({}^{F^0}w>\ldots>{}^{F^N}w,\bar{w})=\tilde{Q}$, where the union is taken over all full chains and all $\bar{w}$. Using this, we shall construct $\Phi^{(2)}_t\colon\tilde{Q}\to\tilde{Q}$, $t\in[0, 1]$, such that $\Phi^{(2)}_0$

$=\mathrm{id}$, $\Phi_1^{(2)}(\tilde{Q})=X$. Let $x+iy\in Y(^{F^0}w>\ldots>{}^{F^N}w, \bar{w})$, where $x=\sum_{j=0}^{N}\lambda_j\,{}^{F^j}w$. Since $y\in C(F^N, \bar{w})\subset C(F^j, \bar{w})$, one can write

$$y=\mu_j(w_{F^j,\bar{w}}-{}^{F^j}w)+\alpha_j, \qquad \mu_j\geqq 0, \qquad \alpha_j\in\partial C(F^j, \bar{w}), \qquad j=0, \ldots, N.$$

For $j=0$ $w_{F^0,\bar{w}}={}^{F^0}w$, so $\alpha_j=y$ and μ_j would be undefined. In the following we shall always set $\mu_0=\lambda_0$. The numbers $\lambda_0, \ldots, \lambda_N$, $\mu_0, \ldots, \mu_N$ are uniquely determined by $x+iy$. In addition, $x+iy\in Y$ implies that not all the μ_j vanish. Precisely: $\lambda_0=\ldots=\lambda_{j-1}=0$, $\lambda_j\neq 0\Rightarrow\mu_j\neq 0$. In fact, if $\lambda_0\neq 0$ then $\mu_0=\lambda_0\neq 0$. If $\lambda_0=\ldots=\lambda_{j-1}=0$, $\lambda_j\neq 0$, $j>0$, then $x\in F^j$ and so $y\notin\partial C(F^j, \bar{w})$ (by Lemma 10). This means that $\mu_j\neq 0$. In particular

$$S(\lambda, \mu)=\sum_j \min\{\lambda_j, \mu_j\}\neq 0.$$

Define the vector

$$V(\lambda, \mu)=(1/S(\lambda, \mu))[\sum_j \min\{\lambda_j, \mu_j\}\,{}^{F^j}w+i\sum_j \min\{\lambda_j, \mu_j\}(w_{F^j,\bar{w}}-{}^{F^j}w)]\in X.$$

We will show that the set of maps

$$Y(^{F^0}w>\ldots>{}^{F^N}w, \bar{w})\to X: x+iy\to V(\lambda, \mu)$$

defines a map $\bigcup Y(^{F^0}w>\ldots>{}^{F^N}w, \bar{w})=\tilde{Q}\to X$. Take

$$x+iy\in Y(^{F^0}w>\ldots>{}^{F^N}w, \bar{w})\cap Y(^{F'^0}w>\ldots>{}^{F'^N}w, \bar{w}')$$

and let λ, μ and λ', μ' be the parameters defined by $x+iy$ in the two sets; we have to show that $V(\lambda, \mu)=V(\lambda', \mu')$. Since Q is a simplicial complex in $\mathbb{R}^N$, if $F^j\neq F'^j$ then $\lambda_j=\lambda'_j=0$, while if $F^j=F'^j$ then $\lambda_j=\lambda'_j$; so the sets λ, λ' coincide. Let then $y\in C(F^N, \bar{w})\cap C(F'^N, \bar{w}')$. If $\bar{w}=\bar{w}'$ then either $F^j\neq F'^j$, so that $\min\{\lambda_j, \mu_j\}=\min\{\lambda'_j, \mu'_j\}=0$, or $F^j=F'^j$, so that $C(F^j, \bar{w})=C(F'^j, \bar{w}')$ and $\mu_j=\mu'_j$, $\min\{\lambda_j, \mu_j\}=\min\{\lambda'_j, \mu'_j\}$. Now let $\bar{w}\neq\bar{w}'$; as above, if $F^j\neq F'^j$ then $\min\{\lambda_j, \mu_j\}=\min\{\lambda'_j, \mu'_j\}=0$. Let $F^j=F'^j$; if $w_{F^j,\bar{w}}\neq w_{F'^j,\bar{w}'}$, then $y\in\partial C(F^j, \bar{w})\cap\partial C(F'^j, \bar{w}')$, so $\mu_j=\mu'_j=0=\min\{\lambda_j, \mu_j\}=\min\{\lambda'_j, \mu'_j\}$. If $w_{F^j,\bar{w}}=w_{F'^j,\bar{w}'}$ then $C(F^j, \bar{w})=C(F'^j, \bar{w}')$, so $\mu_j=\mu'_j$ and the map is well defined.

Define: $\Phi_t^{(2)}(x+iy)=(1-t)(x+iy)+tV(\lambda, \mu)$, $t\in[0, 1]$ where $x=\sum\lambda_j^{F^j}w$, $y=\mu_j(w_{F^j,\bar{w}}-{}^{F^j}w)+\alpha_j$, $j=0, \ldots, N$. We must prove that $\Phi_t^{(2)}(x+iy):=x_t+iy_t\in\tilde{Q}$, $0\leqq t\leqq 1$. Since x_t belongs clearly to Q, this is equivalent to say $x_t+iy_t\in Y$.

If $\lambda_0\neq 0$, also $\mu_0=\lambda_0\neq 0$ and the thesis is obvious (by Lemma 10) since x_t belongs to a chamber. Suppose now $\lambda_0=\ldots=\lambda_{j-1}=0$, $\lambda_j\neq 0$, so $\mu_j\neq 0$ (as said before) and $x\in F^j$. Then $x_t\in F^j$, for all t, and it will be sufficient to prove that y_t does not belong to the direction of any cut containing F^j, $0\leqq t\leqq 1$. Since $\mu_j\neq 0$, $y\in\mathrm{int}\,C(F^j, \bar{w})$. By substituting the expression of V into that of $\Phi^{(2)}$ it follows that

$$y_t=(1-t)\,y+t\sum_{i=j}^{N}(1/S(\lambda, \mu))\min\{\lambda_i, \mu_i\}(w_{F^i,\bar{w}}-{}^{F^i}w).$$

Since $C(F^i, \bar{w})\subset C(F^j, \bar{w})$ for $i\geqq j$, the points $w_{F^i,\bar{w}}-{}^{F^i}w\in\mathrm{int}\,C(F^j, w)$, $i=j, \ldots, N$. But $C(F^j, \bar{w})$ is convex, and it follows that $y_t\in\mathrm{int}\,C(F^j, \bar{w})$, $0\leqq t\leqq 1$. So $x_t+iy_t\in\tilde{Q}$, as we had to prove.

If $x=\sum \lambda_j {}^{F^j}w$ and $y\in \operatorname{int} C(F^N, \bar{w})$ is far enough from F^N, then $\min\{\lambda_j, \mu_j\} = \lambda_j$, $j=0, \ldots, N$, so that $\Phi_1^{(2)}$ is surjective. Next, we show that $\Phi_1^{(2)} | X: X \to X$ is homotopy equivalent to the identity map. In fact, by definition of $V(\lambda, \mu)$, $\Phi_1^{(2)}$ sends the j-skeleton X_j into itself, $j=0, \ldots, N$. But it is easy to verify that the centers of the cells of X remain fixed, i.e.

$$\Phi_1^{(2)}({}^{F^j}w + i(w_{F^j, \bar{w}} - {}^{F^j}w)) = {}^{F^j}w + i(w_{F^j, \bar{w}} - {}^{F^j}w);$$

hence $\Phi_1^{(2)}$ is a cellular map, cellular homotopic to the identity. Thus we have proved the following theorem.

Theorem 1. *The map $\Phi_1^{(2)} \circ \Phi_1^{(1)}: Y \to X$ is a homotopy equivalence.* Q.E.D.

Part Two (on $\pi_1(Y)$ and the general position case)

It is now standard to find a presentation of $\pi_1(Y)=\pi_1(X)=\pi_1(X_2)$. First, one has to look at the complexes X_0 and X_1.

As we have seen above, X_0 is given by all vertices ${}^0w$, so it corresponds to VG. Next, let $F^1 \in \mathscr{M}'$ and F_1^0, F_2^0 be the chambers separated by F^1. Diagram 1 shows a projection (homeomorphism) of the two 1-cells $D^1_{F^1, \bar{w}_i}$, $i=1, 2$, in $\mathbb{C}^N/|F^1|_{\mathbb{C}}$, where $\bar{w}_i = {}^{F_i^0}w$. Note that when $x \in F^1$, y is $w_{F^1, \bar{w}_i}$ $(i=1, 2)$. Let

$$\tilde{G} = \bigcup_{j\in\{0,1\};\, w\in VG} D^j_{F^j, \bar{w}} (\cong X_1),$$

which is a graph in Y. $\tilde{G}$ can be (abstractly) obtained from G by putting $V\tilde{G} = VG$, while $E\tilde{G}$ is obtained from EG by substituting to each edge of G a pair of edges l', l'' joining again the ends of l. A natural orientation can be given to $\tilde{G}$ (so it is a directed graph): orient l' and l'' in an opposite way, so that while x runs from ${}^{F_1^0}w[{}^{F_2^0}w]$ to ${}^{F_2^0}w[{}^{F_1^0}w]$ the positive edge is that for which $x \in F^1$ implies $y = {}^{F_1^0}w[{}^{F_2^0}w]$ (see Fig. 1). We shall always denote a vertex of $\tilde{G}$ by the same symbol as the corresponding vertex of G.

Let $C\tilde{G}$ be the set of edge-paths of $\tilde{G}$, $C^+\tilde{G}[C^-\tilde{G}]$ the subset of the paths such that each edge is traversed along the positive [negative] direction. There is an obvious map $|\ |: C\tilde{G} \to CG$, which takes c into the path $|c|$ given by the real part of c. If $|c| = (l_1, \ldots, l_k)$ then we will write $c = (\varepsilon_1 l_1, \ldots, \varepsilon_k l_k)$, where $\varepsilon_i = 1$ if the corresponding edge of G is traversed along the positive sense, otherwise $\varepsilon_i = -1$. (If $l_1 = \ldots = l_k$ it is nevertheless necessary to specify the first end of the path, while this is determined if there is at least one $l_j \neq l_1$). Clearly the restriction of $|\ |$ to $C^+\tilde{G}$ (or $C^-\tilde{G}$) is an isomorphism, as well as the map $\Gamma = (+l_1, \ldots, +l_k) \to -\Gamma = (-l_1, \ldots, -l_k)$ between $C^+\tilde{G}$ and $C^-\tilde{G}$. Any $c \in C\tilde{G}$ induces a path in the topological space X_1, in the obvious way: we shall confuse c with this path (as well as X_1 with $\tilde{G}$).

Now we can give a geometrical picture of the 2-skeleton X_2, namely, of the attaching map of each 2-cell $D^2_{F^2, \bar{w}}$. Let $\bar{w}' \in {}^{F^2}VG$ be the vertex opposite to $\bar{w}$ with respect to F^2 (Fig. 2), and $\Gamma_1, \Gamma_2 \in C^+\tilde{G}$ the two minimal paths going from $\bar{w}$ to $\bar{w}'$ (their real parts are composed by edges of ${}^{F^2}EG$). Then, by construction, $\partial D^2_{F^2, \bar{w}} = \Gamma_1 \cup \Gamma_2$. (In particular, Γ_1 is homotopic to Γ_2 rel $\dot{I}$).

Similarly, let $D_{F^3, \bar{w}}$ be a 3-cell in X, and $\bar{w}'$ the vertex opposite to $\bar{w}$ with respect to F^3. Then $X_1 \cap D^3_{F^3, \bar{w}}$ is given by all minimal paths in $C^+\tilde{G}$ between

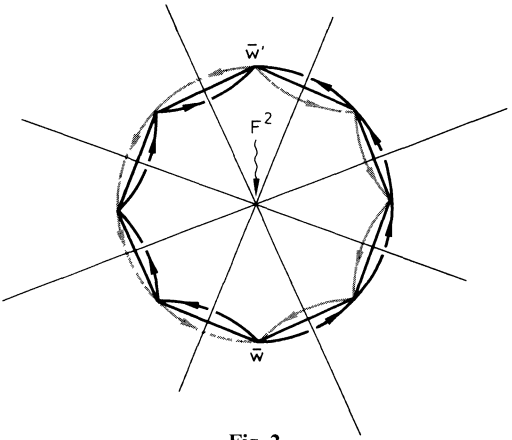


Fig. 2

$\bar{w}$ and $\bar{w}'$ (their real parts have edges in ${}^{F^3}EG$), and $X_2 \cap D^3_{F^3, \bar{w}}$ by all the 2-cells having their boundary in $X_1 \cap D^3_{F^3, \bar{w}}$. An analogous picture can be given for the higher dimensional cells of X.

Consider now the fundamental groupoid of X_2 ([5; p. 116]), which we call $\mathscr{C}$. It is the category whose objects are the vertices of G, and whose morphisms $\mathscr{C}_{w, w'}$ are equivalence classes of paths in G between w and w'. Such equivalence (which we denote by $\approx$) is generated by: $c_1 \sim c_2$ iff, in the reduced form, $c_1 c_2^{-1}$ is the boundary of one 2-cell of X_2, or is the empty path. Obviously $c \approx c'$ iff the corresponding paths in X_1 are homotopic in X_2 (rel $\dot{I}$), so there is an isomorphism between $\mathscr{C}_{w, w}$ and $\pi_1(X_2, w)$. Note that we can again associate a "length" to each class $[c] \in \mathscr{C}_{w, w'}$: if $c = (\varepsilon_1 l_1, \ldots, \varepsilon_k l_k)$ represents $[c]$, then put $l([c]) = \sum \varepsilon_j$. This does not depend on the representative of $[c]$ by the definition of equivalence and by the fact that the boundary of each 2-cell of X_2 has length zero. In particular,

$$l\colon \pi_1(X_2, w) \to \mathbb{Z} \quad \text{is a homomorphism.}$$

Next lemma is fundamental for the rest of this paper.

Lemma 11. *Two paths $u_1, u_2 \in C^+ \tilde{G}$ with the same ends w, w', which have minimal length (among all paths with these ends), are homotopic* (rel $\dot{I}$) *in X, and consequently in X_2.*

Proof. If $\mathscr{M}$ is central, the proof is easy. In fact, both u and u' stand in the boundary $\partial D^N_{F^N, w}$, where F^N is the only codimensional N facet of $\mathscr{M}$ (see the preceding description of the cells of X and note that by Lemma 2 a minimal path between w and w' can be extended to a minimal path between w and its opposite with respect to F^N).

In the general case, let $M_1, \ldots, M_p$ be the cuts separating w from w'. They give a subdivision $\mathscr{M}(w, w')$ of $\mathbb{R}^N$, where the maximum of the codimensions is (say) k. Let $F^k_1, \ldots, F^k_q \in \mathscr{M}$ be all the codimension k facets such that $|F^k|$ is an intersection of cuts of $\mathscr{M}(w, w')$ and F^k is in the same side as w and w' with respect to all cuts which are not in $\mathscr{M}(w, w')$. Then all minimal paths in $C^+ \tilde{G}$

joining w and w' are contained into $\bigcup_{j=1}^{q} D^k_{F^k_j, w}$, which is homeomorphic (through the real part) to $Q^k(w, w') := \bigcup_{j=1}^{q} D^k_{F^k_j}$. But $Q^k(w, w')$ is contractible, by a reasoning similar to that of Lemma 9, so we have the proof. Q.E.D.

Thus the minimal paths in $C^+\tilde{G}$ between w and w' determine a unique class which we call $u(w, w')$.

Now fix a vertex w_0. We want to investigate what happens around a codimension 2 facet of $\mathscr{M}$. For each path $\Gamma = (\varepsilon_1 l_1, \ldots, \varepsilon_k l_k)$, starting from a vertex w, put $\Gamma_{(t)} = (\varepsilon_1 l_1, \ldots, \varepsilon_t l_t)$, and

$$^{\Gamma}\beta_t = \Gamma_{(t-1)}(+l_t)^2(\Gamma_{(t-1)})^{-1} \in \mathscr{C}_{w,w} \qquad t \leqq k.$$

Here we have $(+l_t)^2 = (+l_t)\cdot(+l_t) = (+l_t, +l_t)$, which geometrically represents a loop around the complexification of the cut crossed by l_t (see also Fig. 1). Let F^2 be a facet which is contained in s cuts of $\mathscr{M}$, and choose an orientation for $\partial D^2_{F^2} \subset CG$. Let $l_1, \ldots, l_{2s}$ be the edges of $\partial D^2_{F^2}$, following the orientation and starting from $w = w_{F^2, w_0}$, and let

$$\Gamma_0 = (+l_1, \ldots, +l_{2s}); \qquad \Gamma_1 = (+l_1, \ldots, +l_s); \qquad \Gamma_2 = (+l_{2s}, \ldots, +l_{s+1})$$

(so Γ_1, Γ_2 are the two minimal paths in $C^+\tilde{G}$ joining w to the vertex opposite to it with respect to F^2). Then take $^{F^2}u \in u(w_0, w)$. Using these paths, construct the following circuits:

$$\alpha = {}^{F^2}u\,\Gamma_0\,{}^{F^2}u^{-1}; \qquad \beta_t = {}^{F^2}u\,{}^{\Gamma_1}\beta_t\,{}^{F^2}u^{-1}, \qquad t = 1, \ldots, s;$$

$$\beta_t = {}^{F^2}u\,{}^{\Gamma_2}\beta_{2s-t+1}\,{}^{F^2}u^{-1}, \qquad t = s+1, \ldots, 2s;$$

$$\beta_t^- = {}^{F^2}u\,{}^{-\Gamma_1}\beta_t\,{}^{F^2}u^{-1}, \qquad t = 1, \ldots, s;$$

$$\beta_t^- = {}^{F^2}u\,{}^{-\Gamma_2}\beta_{2s-t+1}\,{}^{F^2}u^{-1}, \qquad t = s+1, \ldots, 2s.$$

By Lemma 11, the classes of these paths do not depend on $^{F^2}u \in u(w_0, w)$, and $^{F^2}u \cdot (\Gamma_j)_{(t-1)}$ $(j = 1, 2)$ can be replaced by another $u' \in u(w_0, w_{F^1_t, w_0})$. Here $F^1_t > F^2$ is the facet crossed by l_t if $j = 1$, by l_{2s-t+1} if $j = 2$. Using the boundaries of the 2-cells $D^2_{F^2, \tilde{w}}$ of X one easily finds:

$$\beta_t^- \approx (\beta_{t-1} \cdots \beta_1)^{-1}\,\beta_t(\beta_{t-1} \cdots \beta_1);$$

$$\beta_{t+s}^- \approx (\beta_{t+s} \cdots \beta_{2s})^{-1}\,\beta_{t+s}(\beta_{t+s} \cdots \beta_{2s}) \qquad (1 \leqq t \leqq s).$$

In practice the use of the $\partial D_{F^2, \tilde{w}}$ consists in adding (or removing) segments of the kind ee^{-1} or substituting segments as follows: one of the two minimal paths in $C^+\tilde{G}$ joining two opposite vertices of $^{F^2}VG$ is changed into the other one. So we can easily verify also the following relations, where we define σ as the cyclic permutation $(1 \ldots s)$ (and σ^t is its t-th power):

$$\alpha \approx \beta_{\sigma^t(s)} \cdots \beta_{\sigma^t(1)}[\approx \beta_{\sigma^t(1)+s} \cdots \beta_{\sigma^t(s)+s}], \qquad t = 0, \ldots, s-1, \qquad [(1)'_t],$$

$$\beta_{t+s} \approx \beta_t^-\,[\beta_t \approx \beta_{t+s}^-], \qquad t = 1, \ldots, s \qquad [(2)'_t],$$

$$\beta_t\,\alpha\,\beta_t^{-1}\,\alpha^{-1} \approx 1, \qquad t = 1, \ldots, 2s \qquad [(3)'_t].$$

These relations are not independent: e.g., those in brackets can be deduced from the others, and (3)′ depend on (1)′. Let J_2 be an index-set for the codimension 2 facets, $I_2 \subset J_2$ be such that $|\alpha_i|_{i\in I}$ is a (free) basis of $\pi_1(G, w_0)$, where α_i, $i\in I_2$, is the analogous of α for $F_i(\#I_2 = \#\{F^1\} - \#\{F^0\} + 1$, if finite). Let J_1 be an index set for the codimension 1 facets, and β_j, $j\in J_1$, be constructed similar to the β_t. Then α_i, β_j, $i\in I_2$, $j\in J_1$, is a basis for $\pi_1(X_1, w_0)$. But it is easy to verify that $R'(F^2)$ are exactly the relations obtained by setting $\partial D_{F^2,\bar{w}} = 1$, $\bar{w}\in {}^{F^2}VG$. So one presentation of $\pi_1(X_2, w_0)$ is:

$$(\alpha_i, \beta_j, i\in I_2, j\in J_1 \mid R'(F^2), i\in I_2).$$

Relations $R'(F_{i'}^2)$, $i'\notin I_2$, depend on $R'(F_i^2)$, $i\in I_2$, because they are obtained expressing $\alpha_{i'}$ as a product in the α_i's.

Relations (1)′ containing α say simply that α is in the group generated by the β's, so the latter is all $\pi_1(X_2, w_0)$. Thus we have a presentation with generators β_j, $j\in J_1$, and relations (deleting those which depend on the others):

$$\beta_s \dots \beta_1 \approx \beta_{\sigma^t(s)} \dots \beta_{\sigma^t(1)}, \quad t = 1, \dots, s-1 \qquad (1)_t,$$

$$\beta_{t+s} \approx \beta_t^-, \quad t = 1, \dots, s \qquad (2)_t.$$

This presentation can be greatly simplified: we will do this in the following.

Let $\mathscr{A}$ be the abelian free group generated by the cuts M_i, $i\in J$, and n: $C\tilde{G} \to \mathscr{A}$ be the (groupoid) homomorphism

$$n((\varepsilon_1 l_1, \dots, \varepsilon_k l_k)) = \sum \varepsilon_j M(l_j)$$

where $M(l_j)$ is the cut crossed by l_j. Call $\rho'_{M_j}: \mathscr{A} \to \langle M_j \rangle$ the projection onto the factor generated by M_j, $\rho''_{M_j}: \langle M_j \rangle \to \mathbb{Z}$ the isomorphism sending M_j into 1, $\rho_{M_j} = \rho'_{M_j} \circ \rho''_{M_j}$. A path $c\in C^+\tilde{G}$ is minimal iff $(\rho_{M_j} \circ n)(c)\in\{0, 1\}$, $j\in J$ (Lemma 2). Then define a map Δ: $C^+\tilde{G} \to F(EG)$, where $F(EG)$ is the free group with basis EG. If $c = (+l_1, \dots, +l_k)$, the word $\Delta(c)$ is obtained starting from the word $l_1 \dots l_k$, and deleting those l_i such that

$$(\rho_{M(l_i)} \circ n)((+l_i, \dots, l_k)) \equiv 1 \pmod 2.$$

Then $\Delta(c) = l_{i_1} \dots l_{i_t}$, where $t < k$, and $\Delta(c) = 1$ iff c is minimal. In fact, clearly c minimal implies $\Delta(c) = 1$; conversely, if $\Delta(c) = 1$ then c crosses a cut either zero times or one time.

Lemma 12. *Let* $\Gamma = (+l_1, \dots, +l_k)$, *where* $+l_i$ *has first end* w_i, *second end* w_{i+1}, $i = 1, \dots, k$. *If* $\Delta(\Gamma) = l_{i_1} \dots l_{i_t}$, *then*

$$\Gamma \approx {}^{\Gamma}\beta_{i_t} \dots {}^{\Gamma}\beta_{i_1} u, \quad \text{where } u\in u(w_1, w_{k+1}).$$

Proof. By induction on t. If $t = 0$ then Γ is minimal and $\Gamma \approx u$ by Lemma 11. Let then $t > 0$. By definition of Γ the path $(+l_{i_t+1}, \dots, +l_k)$ is minimal, while there is a cut M crossed by both l_{i_t} and l_{i_t+s}, for some $0 < s \leqq k - i_t$. Let $u'\in u(w_{i_t}, w_{i_t+s+1})$. Since $(+l_{i_t+1}, \dots, +l_{i_t+s})$ and $(+l_{i_t})u'$ are minimal, one has

$$\begin{aligned}\Gamma &\approx (+l_1, \dots, +l_{i_t})(+l_{i_t})u'(+l_{i_t+s+1}, \dots, +l_k)\\ &\approx {}^{\Gamma}\beta_{i_t}(+l_1, \dots, +l_{i_t-1})u'(+l_{i_t+s+1}, \dots, +l_k).\end{aligned}$$

The path $\Gamma'=(+l_1, \ldots, +l_{i_t-1})\,u'(+l_{i_t+s+1}, \ldots, +l_k)$ is such that $u'(+l_{i+s+1}, \ldots, +l_k)$ is minimal; Γ' crosses all cuts which $(+l_{i_t}, \ldots, +l_{i_t+s})$ crosses except M, which is crossed by Γ' two times less than Γ. So $\Delta(\Gamma')=l_{i_1} \ldots l_{i_t-1}$, and by induction the proof is complete. Q.E.D.

Corollary 12. *Let $\Gamma=(+l_1, \ldots, +l_k)$ be minimal, w_i, w_{i+1} the first and the second end of l_i, M_i the cut crossed by l_i, $i=1, \ldots, k$. Let $l \in EG$ cross M_k, let w be the end of l which is on the same side as w_1 with respect to M_k, and $\Gamma_i \in u(w_i, w)$, $i=1, \ldots, k$. Let $\beta=\Gamma_1(+l)^2\Gamma_1^{-1}$ and define $\sigma_i=0$ or 1 according to whether M_i separates or not w_i from w, for $i=1, \ldots, k-1$. Then:*

$$\beta \approx \left[\prod_{i=k-1}^{1} ({}^{\Gamma}\beta_i)^{\sigma_i}\right]^{-1} \quad {}^{\Gamma}\beta_k \quad \left[\prod_{i=k-1}^{1} ({}^{\Gamma}\beta_i)^{\sigma_i}\right].$$

Proof. Let $\Gamma' \in u(w_{k+1}, w')$, where w' is the other end of l. It is easy to see that $\Delta(\Gamma\Gamma')=l_1^{\sigma_1} \ldots l_{k-1}^{\sigma_{k-1}}$, and $\Delta(\Gamma\Gamma'(+l))=l_1^{\sigma_1} \ldots l_{k-1}^{\sigma_{k-1}} l_k$, so by the preceding lemma there are equivalences:

$$\Gamma\Gamma' \approx \left[\prod_{i=k-1}^{1} ({}^{\Gamma}\beta_i)^{\sigma_i}\right] \Gamma_1(+l),$$

$$\Gamma\Gamma'(+l) \approx {}^{\Gamma}\beta_k \left[\prod_{i=k-1}^{1} ({}^{\Gamma}\beta_i)^{\sigma_i}\right] \Gamma_1.$$

By multiplying the first of these by $(+l)$, and considering that $\Gamma_1(+l)^2 \approx \beta\Gamma_1$, it follows that:

$$\Gamma\Gamma'(+l) \approx \left[\prod_{i=k-1}^{1} ({}^{\Gamma}\beta_i)^{\sigma_i}\right] \beta\Gamma_1.$$

This expression and the preceding one give the corollary. Q.E.D.

In particular, we have that two generators corresponding to codimension 1 facets with the same support are conjugate.

The corollary implies relations (2); conversely (2) imply the corollary, as one can see by dividing the circuit $\Gamma\Gamma'(-l)\Gamma_1^{-1}$ into elementary circuits (in the notations of the corollary).

Consider the following set $\mathcal{T}$ of trees in G: $T \in \mathcal{T}$ iff: i) $w_0 \in T$; ii) for each vertex w of T the path in T joining w_0 to w is minimal; iii) T crosses a given cut at most once.

Lemma 13. *There is a maximal tree T_0 in $\mathcal{T}$ which crosses all cuts.*

Proof. If $\mathcal{M}$ is central, one can take T_0 as a minimal path joining w_0 to the vertex which is in the chamber opposite to that (C_{w_0}) containing w_0.

In general, choose a line which passes from C_{w_0} and is generic with respect to the subdivision. Then the (at most two) positive paths starting from w_0 and "following" this line give a tree with the required properties. Q.E.D.

Let T_0 be a tree as in Lemma 13. Construct all the possible ${}^{\Gamma}\beta_t$, where the paths $\Gamma \in C^+\tilde{G}$ start from w_0 and $|\Gamma| \in T_0$. By Corollary 12 (or by relations (2)), the ${}^{\Gamma}\beta_t$ generate $\pi_1(X_2, w_0)$. So there are as many generators as there are cuts. Indicate by ${}^{0}\beta_j$, $j \in J$, the one relative to (the tree T_0) and the cut M_j.

Let us now simplify the relations.

Let F^2, F'^2 be facets with $|F^2|=|F'^2|=M_1\cap\ldots\cap M_s$. Let $F_i^1>F^2$, $F_i'^1>F'^2$, $\pi'_{F^2}(F_i^1)=\pi'_{F^2}(F_i'^1)$ (so $|F_i^1|=|F_i'^1|$), and $w=w_{F^2,w_0}$, $w'=w_{F'^2,w_0}$, $w_i=w_{F_i^1,w_0}$, $w_i'=w_{F_i'^1,w_0}$, $i=1,\ldots,2s$. If $l_i[l_i']$ is the edge crossing $F_i^1[F_i'^1]$, then let $l_1,\ldots,l_{2s}$ $[l_1',\ldots,l_{2s}']$ follow each other starting from $w[w']$, as after Lemma 11. Now let $\beta_i[\beta_i']$ be the circuits related to $F_i^1[F_i'^1]$ (generators in the last given presentation) and ${}^0\beta_i$ the generators relative to the cut $|F_i^1|=|F_i'^1|$, $i=1,\ldots,s$. Relations $R(F^2)$, (1), $R(F'^2)$, (1) for β_i, β_i' translate into relations among the ${}^0\beta_i$ by substituting (through Corollary 12) to β_i, β_i' the right conjugate of ${}^0\beta_i$. Call these relations ${}^0R(F^2)$, ${}^0R(F'^2)$. We shall prove that there is one circuit δ, starting from w_0, such that $\beta_i'=\delta\beta_i\delta^{-1}$, $i=1,\ldots,s$. One has:

$$\beta_i=\Gamma(+l_1,\ldots,+l_{i-1})(+l_i)^2(+l_1,\ldots,+l_{i-1})^{-1}\Gamma^{-1};$$
$$\beta_i'=\Gamma'(+l_1',\ldots,+l_{i-1}')(+l_i')^2(+l_1',\ldots,+l_{i-1}')^{-1}\Gamma'^{-1}$$

where $\Gamma\in u(w_0,w)$, $\Gamma'\in u(w_0,w')$. Take $\gamma_i\in u(w_i',w_i)$, $i=1,\ldots,s-1$. By Corollary 12 $(+l_i')^2\approx\gamma_i(+l_i)^2\gamma_i^{-1}$ and by Lemma 11

$$(+l_1',\ldots,+l_{i-1}')\gamma_i\approx\gamma_1(+l_1,\ldots,+l_{i-1}).$$

By substituting into the expression of β_i' one obtains:

$$\beta_i'=\Gamma'\gamma_1(+l_1,\ldots,+l_{i-1})(+l_i)^2(+l_1,\ldots,+l_{i-1})^{-1}\gamma_1^{-1}\Gamma'^{-1}.$$

By Lemma 12 there is a circuit δ such that $\Gamma'\gamma_1\approx\delta\Gamma$; thus $\beta_i'\approx\delta\beta_i\delta^{-1}$, as it was to be proved.

We have so obtained the following presentation of $\pi_1(X_2,w_0)$: *the generators are the* ${}^0\beta_j$, $j\in J$, *which are as many as the cuts of* $\mathscr{M}$; *the relations are the* ${}^0R(F^2)$, *which are as many as the codimension* 2 *intersections of cuts, having to choose one* F^2 *for each such intersection.*

As an immediate application of this presentation one obtains that $H_1(Y,\mathbb{Z})$ is an abelian free group with $\#(J)$ generators, given by the classes of ${}^0\beta_j$, $j\in J$. If $\mathscr{A}$ is the precedently defined abelian free group generated by the cuts, one can identify $H_1(Y,\mathbb{Z})$ with the subgroup $2\mathscr{A}<\mathscr{A}$ generated by $2M_j$, $j\in J$. In fact, the homomorphism $n\colon C\tilde{G}\to\mathscr{A}$ given above induces a morphism $\mathscr{C}=C\tilde{G}/\approx\to\mathscr{A}$. In particular, there is a homomorphism $\mathscr{C}_{w,w}=\pi_1(X,w_0)\to\mathscr{A}$ which has image $2\mathscr{A}$, and which is just the abelianization homomorphism.

As another immediate application suppose that every codimension 2 facet is contained in exactly two cuts, and no two cuts are parallel. Then $\pi_1(Y)$ is free abelian, with $\#(J)$ generators. In fact, relations $R(F^2)$, (1), become $\beta_1\beta_2\approx\beta_2\beta_1$, while $R(F^2)$, (2), become $\beta_{t+2}\approx\beta_t$, $t=1,2$. Let F^2 be a facet contained in the cuts M_i and M_j, and let β_1,β_3 and β_2,β_4 be the circuits relative to F^2 and respectively to M_i and M_j. By a repeated application of relations (2) (or by Corollary 12) we find ${}^0\beta_i\approx\beta_1\approx\beta_3$, ${}^0\beta_j\approx\beta_2\approx\beta_4$, so ${}^0\beta_i\,{}^0\beta_j\approx{}^0\beta_j\,{}^0\beta_i$, which proves the assertion.

We now consider the following situation: *each facet* F^j *is contained in exactly* j *cuts* ($j=0,\ldots,N$). This is equivalent to say (by Lemma 8) that each maximal codimension facet F^N, $i=1,\ldots,r$, is contained in N cuts. Let $X_i=\bigcup_{\bar{w}\in F_i^N VG} D^N_{F_i^N,\bar{w}}$; then $X=\bigcup_i X_i$, and each X_i is homotopy equivalent to a N-

torus $T^N=(S^1)^N$. This can be seen by taking the decomposition of $\mathbb{R}^N$ given by the coordinate hyperplanes, which yields $Y=(\mathbb{C}^*)^N$, and using Theorem 1. Actually, X_i is homeomorphic to T^N. Indeed, take the N-cube $C^N=[-1,1]^N \subset \mathbb{R}^N$, so T^N is obtained by identification of the opposite faces of C^N. The coordinate hyperplanes give a cellular decomposition on C^N, and consequently on T^N, the latter having $\binom{N}{k} 2^N$ k-cells. Then one can easily see by induction on N that X_i is cellularly isomorphic to T^N with this decomposition.

Let T_0^N be the N-torus obtained by identifying the opposite faces of C^N, with the cellular subdivision induced by the standard subdivision of C^N (in which the only N-cell is C^N). There is a homotopy equivalence $\bar{h}: T^N \to T_0^N$, which sends cells into cells (not necessarily of the same dimension). For instance, that obtained from a homotopy $h_t: C^N \to C^N$, defined by:

$$(\ldots, x_i, \ldots) \to (\ldots, (1+\operatorname{sgn}(x_i)\, t)\, x_i - t, \ldots), \qquad i=1, \ldots, N.$$

This sends the N-cell $\{0 \leqq x_i \leqq 1,\ i=1, \ldots, N\}$ onto C^N, while the other N-cells are crushed onto $\partial C^N \cap [\bigcup_i \{x_i=-1\}]$. Then one sets $\bar{h}$ as the composition of h_1 with the projection over the quotient.

Lemma 14. *Suppose that each facet F^j is contained in exactly j cuts, $j=0, \ldots, N$. Then X is homotopy equivalent to the cellular complex K obtained from the complex Q^N by identifying two j-cells $D^j_{F^j}$, $D^j_{F'^j}$, iff $|F^j|=|F'^j|$.*

Proof. Let $F_1^N, \ldots, F_r^N$ be the codimension N facets of $\mathcal{M}$. If $r=1$ (so $\mathcal{M}$ is central) we have already seen that both X and K are N-tori, so lemma is true. In general, using the notations above, $X=\bigcup_i X_i$, where each X_i is glued to the remaining tori by some $(N-1)$-tori of its $(N-1)$-skeleton. But the map $\bar{h}$ constructed before identifies two such $(N-1)$-tori if they correspond to $(N-1)$-cells $D^{N-1}_{F^{N-1}}$, $D^{N-1}_{F'^{N-1}}$ such that $|F^{N-1}|=|F'^{N-1}|$. So the thesis follows by induction on r and N. Q.E.D.

Now let $m=\#(J)$, and let $I^m \subset \mathbb{R}^m$ be the standard m-cube, with its standard cellular decomposition. By identifying its opposite faces we obtain a m-torus, which we call again T_0^m, with the induced cellular decomposition. It is easy to see (e.g. by induction on N) that K is cellularly isomorphic to a subcomplex of the complex $(T_0^m)_N=T(N,m)$, the N-skeleton of T_0^m. Moreover, suppose that the cuts are in general position (that is, each F^j is contained in exactly j cuts, and there are no parallel facets). Then K is cellularly isomorphic to $T(N,m)$. The universal cover of $T(N,m)$ is given ($N \geqq 2$) by the N-skeleton $\mathcal{U}_N$ of the universal cover $\mathcal{U}$ of T^m. But $\mathcal{U}$ is $\mathbb{R}^m$ equipped with its standard cellular decomposition: namely, the standard one of I^m translated along all vectors with integers coordinates. Thus, if $N<m$, $\mathcal{U}_N$ is homotopy equivalent to a bouquet of infinitely many N-spheres, and we have the following theorem.

Theorem 2. *Suppose that the hyperplanes $M_j \subset \mathbb{R}^N$, $j \in J$, $\#J=m$, are in general position. Then Y is a $K(\pi, 1)$ iff $N \geqq m$ or $N=1$. If $1<N<m$, then $\pi_i(Y)=0$ for $1<i<N$, while $\pi_N(Y)$ is an infinitely generated free abelian group.* Q.E.D.

References

1. Brieskorn, E.: Sur les groupes de tresses. Sém. Bourbaki (1971/72); Lect. Notes Math., **317**, 21–44 (1973)
2. Deligne, P.: Les immeubles des groupes de tresses généraligés. Invent. Math. **17**, 273–302 (1972)
3. Cartier, P.: Arrangements d'hyperplans: un chapitre de géométrie combinatoire. Sém. Bourbaki (1980/81). Lect. Notes Math., vol. 901. Berlin-Heidelberg-New York: Springer, 1977, pp. 1–22
4. Hendriks, H.: Hyperplane complements of large type. Invent. Math. **79**, 375–381 (1985)
5. Lyndon, R.C., Schupps, P.E.: Combinatorial group theory. Ergebnisse der Mathematik. Berlin-Heidelberg-New York: Springer 1977
6. Zavlasky, T.: Facing up to arrangements: face-count formulas for partitiones of space by hyperplanes. Mem. Am. Math. Soc. **154** (1975)
7. Randell, R.: The fundamental group of the complement of a union of complex hyperplanes. Invent. Math. **69**, 103–108 (1982)
8. Randell, R.: The fundamental group of the complement of a union of complex hyperplanes: correction. Invent. Math. **80**, 467–468 (1985)
9. Hattori, A.: Topology of $\mathbb{C}^N$ minus a finite number of affine hyperplanes in general position. J. Fac. Sci., Univ. Tokio, Sect. IA **22**, 205–219 (1975)
10. Fox, R.H., Neuwirth, L.: The braid groups. Scand. Math. **10**, 119–126 (1962)
11. Viet Dung, N.: The fundamental groups of the space of regular orbits of the affine Weyl groups. Topology **22**, 425–435 (1983)

Oblatum 13-III-1986

Invent. math. 88, 619-634 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

Equisingular deformations of isolated 2-dimensional hypersurface singularities

Klaus Altmann

Sektion Mathematik der Humboldt-Universität zu Berlin, PSF 1297, DDR-1086 Berlin, German Democratic Republic

In [Wa] Wahl has introduced the functor ES of infinitesimal equisingular deformations. Starting with a good resolution π: $Y \to \operatorname{Spec} R$ of an isolated normal surface singularity, he defines ES-deformations as those deformations of Y and the components of $\pi^{-1}(0)$ which blow down to $\operatorname{Spec} R$. Taking hypersurface singularities for $\operatorname{Spec} R$, we can use the embedded resolution given by the Newton polyhedron of the defining equation f - some properties of the functor ES can then be expressed in terms of $\Gamma(f)$.

After introducing some notations, in §2 and §4 we describe embedded equisingular deformations (ESE). So we give a cohomological characterization of when the deformations induced by such embedded equisingular ones are exactly the over-$\Gamma(f)$-deformations of $\operatorname{Spec} R$.

In §5 we begin with a comparison of both equisingularity functors; Proposition 5.4 gives an easy geometric condition which is sufficient for $ESE \to ES$ to be formally smooth. This allows us to lift the results of §2 and §4 to the functor ES. In particular, we get the formal smoothness of $\overline{ES}$ (defined by using the minimal good resolution) if each of the coordinate axes of $\mathbb{R}^3$ is intersected by exactly one 2-dimensional face of the Newton boundary $\Gamma(f)$. In §3, an example illustrates the results.

Now we want to fix some notations:

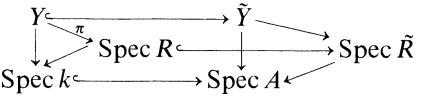
$\mathscr{C} :=$ category of local Artinian algebras of finite type over k
($k =$ fixed algebraic closed field with $\operatorname{char} k = 0$),
Def_R: $\mathscr{C} \to \mathrm{Ens}$ denotes the deformation functor of the singularity $\operatorname{Spec} R$,
$k[\varepsilon] \in \mathscr{C}$ denotes the ring $k[\varepsilon] := k[\varepsilon]/\varepsilon^2$;

all schemes considered here are assumed to be essentially of finite type over k.

§ 1. Preliminaries

Let R be a normal k-algebra with the only singular point $0 \in \operatorname{Spec} R$, let π: $Y \to \operatorname{Spec} R$ be a good resolution with the exceptional divisor $\pi^{-1}(0)_{\mathrm{red}} = E_1 \cup \ldots \cup E_N$ (that means, the E_ν are smooth and intersect each other transversally).

(1.1) **Definition.** Suppose $\tilde{Y}|A \in \mathscr{C}$ is a deformation of Y. We say that $\tilde{Y}$ blows down to a deformation $\tilde{R}$ of R if there exists a commutative diagram of A-deformations



(1.2) **Proposition.** *$\tilde{Y}$ blows down to R iff $\Gamma(\tilde{Y}, \mathcal{O}_{\tilde{Y}})$ is flat over A iff $\Gamma(\tilde{Y}, \mathcal{O}_{\tilde{Y}}) \to \Gamma(Y, \mathcal{O}_Y)$ is surjective. In this case we have $\tilde{R} = \Gamma(\tilde{Y}, \mathcal{O}_{\tilde{Y}})$* (cf. [Wa]).

(1.3) For the fixed good resolution π: $Y \to \operatorname{Spec} R$ we can define ES: $\mathscr{C} \to$ Ens by

$$ES(A) := \{\text{isomorphism classes of } A\text{-deformations } \tilde{E}_1, \ldots, \tilde{E}_N \hookrightarrow \tilde{Y} \text{ of } E_1, \ldots, E_N \hookrightarrow Y \text{ such that } \tilde{Y} \text{ blows down to } R\}.$$

Suppose $\dim R = 2$. Then the image of ES in Def_R does not depend on the special choice of the resolution π and coincides with the equisingularity functor $\overline{ES}$ defined by using the minimal good resolution (cf. [Wa]).

(1.4) The aim of this paper is to investigate the functor ES for hypersurface singularities under use of the resolution induced by the Newton boundary of the defining equation f.

Let $f \in (\underline{x})^2 k[\underline{x}]$ $(\underline{x} := (x_1, \ldots, x_n)$, $n \geqq 3)$ be an irreducible, complete polynomial (i.e., $f(0, \ldots, x_i, \ldots, 0) \not\equiv 0$ for all i); suppose also that f is not degenerated on its Newton boundary $\Gamma(f)$. These are the usual conditions when dealing with Newton diagrams (cf. [Va]).

Let us shortly fix the notations:

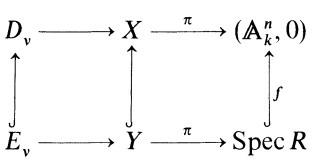
$$M := \mathrm{Hom}_{\text{alg. gr.}}(\mathbb{G}_m^n, \mathbb{G}_m) \cong \mathbb{Z}^n \quad \text{(hence } \Gamma_+(f) \subseteq M_{\mathbb{R}})$$
$$N := \mathrm{Hom}_{\text{alg. gr.}}(\mathbb{G}_m, \mathbb{G}_m^n) \cong \mathbb{Z}^n.$$

The elements of $M_{\mathbb{R}}$ are denoted by $r = (r_1, \ldots, r_n)$, the elements of $N_{\mathbb{R}}$ by $a = (a_1, \ldots, a_n)$.

$$(\mathbb{A}_k^n, 0) := \operatorname{Spec} k[\underline{x}]_{(\underline{x})}; \qquad R := k[\underline{x}]_{(\underline{x})}/(f).$$

By subdividing (without destroying the coordinate planes) the finite rational partial polyhedral (f.r.p.p.) decomposition $\Sigma_0 < \mathbb{R}^n_{\geqq 0} \subseteq N_{\mathbb{R}}$ corresponding to $\Gamma(f)$ we obtain a smooth f.r.p.p. decomposition $\Sigma < \mathbb{R}^n_{\geqq 0} \subseteq N_{\mathbb{R}}$ together with a proper map π: $X = X_\Sigma \to (\mathbb{A}_k^n, 0)$.

With $Y := \overline{\pi^{-1}(\operatorname{Spec} R \setminus 0)}$, $D_v :=$ components of $\pi^{-1}(0)_{\mathrm{red}}$ and $E_v := Y \cap D_v$ there is an embedded good resolution



of the singularity $0 \in \operatorname{Spec} R$ (cf. [Va]).

(1.5) It is much easier to describe embedded equisingular deformations in Def_R - so we define the functor

ESE: $\mathscr{C} \to \mathrm{Ens}$ (**E**qui**S**ingular **E**mbedded)

$ESE(A) := \{$isomorphism classes of A-deformations

$$\tilde{Y}, \tilde{D}_v \hookrightarrow \tilde{X} \text{ of } Y, D_v \hookrightarrow X\}.$$

Let $\tilde{Y} \hookrightarrow \tilde{X}$ be a deformation of the embedding $Y \hookrightarrow X$. We look at the commutative diagram

$$\begin{array}{ccc} \Gamma(\tilde{X}, \mathcal{O}_{\tilde{X}}) & \longrightarrow & \Gamma(X, \mathcal{O}_X) \\ \downarrow & & \downarrow \\ \Gamma(\tilde{Y}, \mathcal{O}_{\tilde{Y}}) & \longrightarrow & \Gamma(Y, \mathcal{O}_Y), \end{array}$$

and since $H^1(X, \mathcal{O}_X) = 0$, the deformation $\tilde{Y} \hookrightarrow \tilde{X}$ blows down to $\operatorname{Spec} \tilde{R} \hookrightarrow \widetilde{(\mathbb{A}_k^n, 0)}$. In particular, we get a natural transformation $ESE \to ES$.

§ 2. $\Gamma(f)$ – a frontier for embedded equisingular deformations?

(2.1) Given a resolution π: $X \to (\mathbb{A}_k^n, 0)$ described above, it is obvious that over-$\Gamma(f)$-deformations

$$\tilde{f} = \sum_{r \in \Gamma_+(f)} \tilde{\lambda}_r x^r \qquad (\tilde{\lambda}_r \in A \in \mathscr{C})$$

of f are induced by elements of $ESE(A)$: Indeed, take the trivial deformation $D_v \times_k A \hookrightarrow X \times_k A$ together with $\tilde{Y} :=$ strict transform of $\tilde{f}$ in $X \times_k A$.

For quasihomogeneous singularities it follows from Varchenko's results that all $\overline{ES}$-deformations are of this type. We give now a cohomological description of those singularities for which ESE induces exactly the over-$\Gamma(f)$-deformations.

(2.2) Let $V := X \setminus \pi^{-1}(0) = \mathbb{A}_k^n \setminus 0$. We have a commutative diagram

$$\begin{array}{ccccccc} 0 \to H^0(X, \mathcal{O}_X) & \longrightarrow & H^0(X, \mathcal{O}_X(Y)) & \longrightarrow & H^0(X, \mathcal{N}_{Y|X}) \to 0 \\ \downarrow \wr & & \downarrow \text{(ii)} & & \downarrow \text{(iii)} \\ 0 \to H^0(V, \mathcal{O}_X) & \xrightarrow{\text{(i)}} & H^0(V, \mathcal{O}_X(Y)) & \longrightarrow & H^0(V, \mathcal{N}_{Y|X}) \to 0. \end{array}$$

Let us comment the maps (i)–(iii):

(i) With $m(a) := \min_{r \in \Gamma_+(f)} \langle a, r \rangle$ we have

$\mathcal{O}_X(Y) = \mathcal{O}_X((f) - \sum_{a \in \Sigma^{(1)}} m(a) D_a)$ (*cf. the beginning of the proof of* (5.4)), hence

$$\mathcal{O}_V(Y) = \mathcal{O}_V((f)) \cong \mathcal{O}_V.$$

Therefore, the map $H^0(V, \mathcal{O}_X) \hookrightarrow H^0(V, \mathcal{O}_X(Y))$ corresponds to the multiplication by f f: $k[\underline{x}]_{(\underline{x})} \hookrightarrow k[\underline{x}]_{(\underline{x})}$.

(ii) By using the general formula for the computation of cohomology groups of T-invariant sheaves on toroidal embeddings (cf. [Ke], I/§ 3) we get

$$[H^0(X,\mathcal{O}_X(Y))\hookrightarrow H^0(V,\mathcal{O}_X(Y))]$$
$$\hat{=}\bigoplus_{r\in M}[H^0(\mathbb{R}^n_{\geqq 0},A_r,k)\to H^0(\partial\mathbb{R}^n_{\geqq 0},A_r\cap\partial\mathbb{R}^n_{\geqq 0},k)]$$

with $A_r:=\{a\in\mathbb{R}^n_{\geqq 0}/\langle a,r\rangle<m(a)\}$.

Now
$$A_r\cap\partial\mathbb{R}^n_{\geqq 0}=\emptyset \quad \text{iff } \{a\in\partial\mathbb{R}^n_{\geqq 0}/\langle a,r\rangle<0\}=\emptyset$$
$$\text{iff } r\geqq 0,$$

and we obtain again that $H^0(V,\mathcal{O}_X(Y))\cong k[\underline{x}]_{(\underline{x})}$.

For the other cohomology group we have

$$A_r=\emptyset \quad \text{iff } \forall a\in\mathbb{R}^n_{\geqq 0}:\ \langle a,r\rangle\geqq m(a)$$
$$\text{iff } r \text{ lies over } \Gamma(f).$$

So we see that $H^0(X,\mathcal{O}_X(Y))\subseteq H^0(V,\mathcal{O}_X(Y))$ is exactly the ideal in $k[\underline{x}]_{(\underline{x})}$ generated by all monomials over the Newton boundary $\Gamma(f)$.

(iii) Finally, it is clear that $H^0(V,\mathcal{N}_{Y|X})\cong R$, and $H^0(X,\mathcal{N}_{Y|X})\subseteq H^0(V,\mathcal{N}_{Y|X})$ corresponds to the ideal generated by all monomials over $\Gamma(f)$.

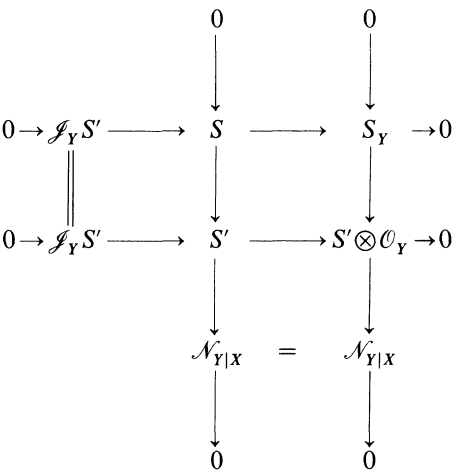
(2.3) Deformations from *ESE* are locally trivial – information about this functor can be obtained by considering the automorphism sheaf

$$S:=\mathrm{Ker}[\theta_X\to\mathcal{N}_{Y|X}\oplus(\bigoplus_\nu\mathcal{N}_{D_\nu|X})]=\Omega_X(\log(Y+\sum_\nu D_\nu))^{\mathrm{dual}}$$

of objects of $ESE(k[\varepsilon])$. For instance, the tangent space of *ESE* is $ESE(k[\varepsilon])=H^1(X,S)$.

Let
$$S':=\mathrm{Ker}[\theta_X\to\bigoplus_\nu\mathcal{N}_{D_\nu|X}]=\Omega_X(\log(\sum_\nu D_\nu))^{\mathrm{dual}}$$
$$S_Y:=\mathrm{Ker}[\theta_Y\to\bigoplus_\nu\mathcal{N}_{E_\nu|Y}]=\Omega_Y(\log(\sum_\nu E_\nu))^{\mathrm{dual}}.$$

Then we have a commutative diagram on X:



($\mathcal{J}_Y:=$ ideal sheaf of Y in X).

(2.4) The injections $S_Y \hookrightarrow \theta_Y$ and $S' \hookrightarrow \theta_X$ are isomorphisms on $V = X \setminus \pi^{-1}(0)$. Hence, the standard sequence

$$0 \to \mathrm{Def}_R(k[\varepsilon]) \to H^1(V, \theta_Y) \to H^1(V, \theta_X \otimes \mathcal{O}_Y)$$

yields the diagram

$$\begin{array}{ccccccc}
 & & & & 0 & & \\
 & & & & \downarrow & & \\
 & & & & \mathrm{Def}_R(k[\varepsilon]) & & \\
 & & & & \downarrow & & \\
H^1(V, \mathcal{J}_Y S') & \longrightarrow & H^1(V, S) & \longrightarrow & H^1(V, S_Y) & \longrightarrow & H^2(V, \mathcal{J}_Y S') \\
 & & \downarrow & & \downarrow & & \| \\
 & & H^1(V, S') & \longrightarrow & H^1(V, S' \otimes \mathcal{O}_Y) & \longrightarrow & H^2(V, \mathcal{J}_Y S').
\end{array}$$

Since the sheaves $\mathcal{J}_Y S'$ and S' are trivial on V and

$$H^1(\mathbb{A}_k^n \setminus 0, \mathcal{O}_{\mathbb{A}^n}) = \bigoplus_{i \in \mathbb{Z}} H^1(\mathbb{P}_k^{n-1}, \mathcal{O}_{\mathbb{P}^{n-1}}(i)) = 0,$$

we get $H^1(V, \mathcal{J}_Y S') = H^1(V, S') = 0$.

Therefore, we obtain an isomorphism $\mathrm{Def}_R(k[\varepsilon]) \xrightarrow{\sim} H^1(V, S)$.

(2.5) Now we have the following commutative diagram:

$$\begin{array}{ccccc}
 & & H^1_D(X, S) & \longrightarrow & H^1_D(X, S') \\
 & & \downarrow & \searrow & \downarrow \\
H^0(X, \mathcal{N}_{Y|X}) & \xrightarrow{\alpha} & H^1(X, S) & \twoheadrightarrow & H^1(X, S') \\
\downarrow & & \downarrow \gamma & & \\
H^0(V, \mathcal{N}_{Y|X}) & \xrightarrow{\beta} & H^1(V, S) & & (D := \pi^{-1}(0) \subseteq X).
\end{array}$$

The map α shows again how over-$\Gamma(f)$-deformations are induced by elements of ESE.

The map β corresponds to $R \twoheadrightarrow \mathrm{Def}_R(k[\varepsilon]) \cong R \Big/ \left(\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n}\right)$, and γ is the blow down map $ESE(k[\varepsilon]) \to \mathrm{Def}_R(k[\varepsilon])$.

(2.6) **Proposition.** 1) *Elements $\xi \in ESE(k[\varepsilon])$ with trivial deformation of X blow down to over-$\Gamma(f)$-deformations.*

2) *All $ESE(k[\varepsilon])$-deformations blow down to deformations over $\Gamma(f)$ iff $H^1_D(X, S) \to H^1(X, S')$ is surjective.*

Proof. 1) One derives the proof from the diagram of (2.5) and the injectivity of $H^1(X, S') \hookrightarrow H^1(X, \theta_X)$.

2) By the discussion above we see:

$ESE(k[\varepsilon])$-deformations blow down to deformations over $\Gamma(f)$ iff $\mathrm{Im}(\gamma\circ\alpha)=\mathrm{Im}(\gamma)$ in the diagram (2.5). □

(2.7) If $\xi=[\tilde{D}_v, \tilde{Y}\hookrightarrow\tilde{X}]\in ESE(k[\varepsilon])$ is given explicitly by a 1-cocycle $\{\xi_{\alpha\beta}\}\in H^1(X,S)$, the induced deformation $\tilde{f}(x,\varepsilon)=f(x)-\varepsilon d(x)$ can be studied by something like a formula for d. We must introduce some notations: By construction, X is covered by open, affine, $\mathbb{G}_m^n$-invariant subsets $\mathrm{Spec}\,B_\alpha$ ($\alpha\in\Sigma$, $\dim\alpha=n$). There we have coordinates $y_1^\alpha,\ldots,y_n^\alpha\in B_\alpha$ corresponding to the 1-dimensional generators $a^1,\ldots,a^n$ of the cone α.

All rings B_α are contained in $k[|\underline{x}|][\underline{x}^{-1}]$. For elements $a\in N_{\mathbb{R}}$ and $h\in k[|\underline{x}|][\underline{x}^{-1}]$ we define

$$\mathrm{ord}^a h := \min_{r\in\mathrm{supp}\,h}\langle a,r\rangle.$$

In particular, we have

1) $\mathrm{ord}^a f=m(a)$ and
2) $\mathrm{ord}^{a^i} y_j^\alpha=\delta_{ij}$ with $\alpha=\langle a^1,\ldots,a^n\rangle$.

Remark. Because of $H^1(X,\mathcal{O}_X)=0$, there are elements $b_i^\alpha\in B_\alpha$ with $b_i^\beta-b_i^\alpha=\xi_{\alpha\beta}(x_i)$ for $i=1,\ldots,n$.

The choice of such elements corresponds to taking coordinates in the induced deformation $(\widetilde{\mathbb{A}_k^n,0})$, that means, fixing an isomorphism

$$\Gamma(X,\mathcal{O}_X)\otimes_k k[\varepsilon]\xrightarrow{\sim}\Gamma(\tilde{X},\mathcal{O}_{\tilde{X}}).$$

The function $d(x)$ is (modulo f) characterized by the vanishing of $f(x)-\varepsilon d(x)$ in $\Gamma(\tilde{Y},\mathcal{O}_{\tilde{Y}})$. Therefore, we get

$$d(x)=\sum_{i=1}^n\frac{\partial f}{\partial x_i}b_i^\alpha \qquad \text{in } B_\alpha/(f_\alpha)\ (f_\alpha=\text{equation of } Y),\ \text{hence}$$

$$d(x)=\sum_{i=1}^n\frac{\partial f}{\partial x_i}b_i^\alpha-g_\alpha\cdot f \qquad \text{in } B_\alpha \text{ for some}$$

$$g_\alpha\in\frac{f_\alpha}{f}\cdot B_\alpha=y_1^{-m(a^1)}\cdot\ldots\cdot y_n^{-m(a^n)}B_\alpha.$$

(2.8) **Lemma.** *$g_\alpha\in B_\alpha$ for n-dimensional cones $\alpha\in\Sigma$. In particular, monomials x^r with $\langle a,r\rangle<m(a)$ for some $a\in\alpha$ appear in $d(x)$ and $\sum_{i=1}^n\frac{\partial f}{\partial x_i}b_i^\alpha$ with the same coefficient.*

Proof. Step 1. Let $c\in\Sigma^{(1)}$ be a 1-dimensional cone with $m(c)>0$. Then for $\alpha,\beta\in\Sigma$ with $c\in\alpha\cap\beta$ we have $\mathrm{ord}^c\,\xi_{\alpha\beta}\geqq 0$, hence

$$\begin{aligned}\mathrm{ord}^c(g_\beta-g_\alpha)+m(c)&=\mathrm{ord}^c((g_\beta-g_\alpha)f)\\&=\mathrm{ord}^c\left(\sum_{i=1}^n\frac{\partial f}{\partial x_i}(b_i^\beta-b_i^\alpha)\right)\\&=\mathrm{ord}^c\left(\sum_{i=1}^n\frac{\partial f}{\partial x_i}\xi_{\alpha\beta}(x_i)\right)\geqq m(c).\end{aligned}$$

The inequality $\mathrm{ord}^c(g_\beta - g_\alpha) \geqq 0$ means that monomials x^s with $\langle c, s\rangle < 0$ appear in g_α and g_β with the same coefficient.

Step 2. Let x^s be a monomial in g_α with $\langle a, s\rangle < 0$ for an element $a \in \alpha \cap \Sigma^{(1)}$. Then we can find a basis vector $e^i \in N$ (corresponding to a coordinate axis of $\mathbb{R}^n$) and a chain $a = c^0, \ldots, c^M = e^i$ of adjacent elements of $\Sigma^{(1)}$ with $\langle c^\nu, s\rangle < 0$ ($\nu = 0, \ldots, M$). In particular, there is a cone $\beta \in \Sigma$ containing e^i such that x^s appears in g_β (cf. 1-st step), and by $\langle e^i, s\rangle < 0 = -m(e^i)$ we get a contradiction. □

(2.9) **Proposition.** *For $i = 1, \ldots, n$ assume that the convex hull $\mathrm{conv}(a, b)$ of arbitrary elements $a, b \in \Sigma_0^{(1)}$ $(a, b \neq e^i)$ is contained in the union of the closed cones $\alpha \in \Sigma$ with $e^i \notin \alpha$. Then every element of $ESE(k[\varepsilon])$ blows down to an over-$\Gamma(f)$-deformation of* $\mathrm{Spec}\, R$.

Proof. As in the proof of (2.8) we start with the following remark: Let $i \in \{1, \ldots, n\}$ and let $c \in \Sigma^{(1)}$, $\alpha, \beta \in \Sigma$ be cones with $c \neq e^i$, $c \in \alpha \cap \beta$. Then we have $\mathrm{ord}^c(b_i^\beta - b_i^\alpha) \geqq c_i$.

Now we fix an element $i \in \{1, \ldots, n\}$ and a monomial x^r which occurs in some b_i^α. Assume that there is a 1-dimensional face $a \in \Sigma_0^{(1)} \cap \alpha$ with $m(a) > 0$ and $\langle a, r\rangle < a_i$. Then for each $b \in \Sigma_0^{(1)}$ with $\langle b, r\rangle < b_i$ there exists a chain $a = c^0, \ldots, c^M = b$ of adjacent elements $c^\nu \in \Sigma^{(1)}$ with

$$\langle c^\nu, r\rangle < c_i^\nu \ (\nu = 0, \ldots, M) \quad \text{and} \quad c^\nu \neq e^i \quad (\nu = 0, \ldots, M-1).$$

Hence, we can find a cone β containing b such that the monomial x^r occurs in b_i^α and b_i^β with the same coefficient. In particular, we get $\langle b, r\rangle \geqq 0$.

Now we have to distinguish between two cases:

Case 1. $r \ngeqq 0$

There exists some $j \in \{1, \ldots, n\}$ with $r_j = \langle e^j, r\rangle < 0 \leqq e_i^j$. The discussion above implies $\langle e^j, r\rangle \geqq 0$, and we get a contradiction.

Case 2. $r \geqq 0$

Then x^r is a global section of $\mathcal{O}_X$, and if $b \in \Sigma_0^{(1)}$, $\beta \in \Sigma$ are cones with $m(b) > 0$, $b \in \beta$ and $\langle b, r\rangle < b_i$, this monomial will occur in b_i^α and b_i^β with the same coefficient λ. Hence, we are able to remove λx^r from all functions b_i^γ $(\gamma \in \Sigma)$ without losing the properties of (2.7) – we change only the trivialisation of $\widetilde{(\mathbb{A}_k^n, 0)}$.

Finally, let $a \in \Sigma_0^{(1)}$ be an arbitrary 1-dimensional cone (corresponding to some top dimensional face of $\Gamma_+(f)$).

For cones $\alpha \in \Sigma$ containing a it holds:

$$\mathrm{ord}^a d(x) \geqq \min\left\{m(a), \mathrm{ord}^a \sum_{i=1}^n \frac{\partial f}{\partial x_i} b_i^\alpha\right\} \quad \text{and}$$

$$\mathrm{ord}^a \frac{\partial f}{\partial x_i} b_i^\alpha = \mathrm{ord}^a \frac{\partial f}{\partial x_i} + \mathrm{ord}^a b_i^\alpha$$

$$\geqq \begin{cases} (m(a) - a_i) + a_i & \text{for } m(a) > 0 \\ 0 \quad\quad +0 & \text{for } m(a) = 0 \end{cases} = m(a). \quad \square$$

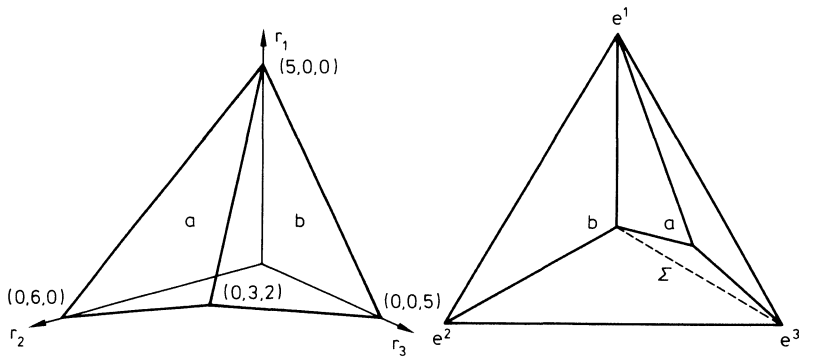
(2.10) **Corollary.** 1) *By subdividing Σ_0 sufficiently fine, we obtain an embedded resolution $\pi\colon X_\Sigma \to (\mathbb{A}_k^n, 0)$ such that all elements of $ESE(k[\varepsilon])$ blow down to over-$\Gamma(f)$-deformations.*

2) *We denote by K_i the union of all closed Σ_0-cones not containing e^i. If all polyhedra K_i $(i=1, \ldots, n)$ are convex, 1) holds for each $\Sigma < \Sigma_0$.*

§ 3. An example

(3.1) Let $\Delta := \{a \in \mathbb{R}^n_{\geq 0} \subseteq N_{\mathbb{R}} / a_1 + \ldots + a_n = 1\}$. We will look at the cones of Σ_0 and Σ only by regarding the intersections with Δ; as coordinates for points $a \in \Delta \cap \mathbb{Q}^n$ we take integers with greatest common divisor 1.

(3.2) The polynomial $f(x, y, z) := x^5 + y^6 + z^5 + y^3 z^2$ defines the f.r.p.p. decomposition Σ_0 shown in Fig. 1; let $a := (12, 10, 15)$, $b := (1, 1, 1) \in N$.



Newton boundary $\Gamma(f)$ f.r p.p. decomposition $\Sigma_0 \cap \Delta$

Now we can construct a subdivision $\Sigma < \Sigma_0$ by

(i) drawing the line $\overline{be^3}$ and

(ii) arbitrary subdividing without destroying the cones $\overline{be^1}$ and $\mathrm{conv}(b, e^2, e^3)$.

(3.3) For $\alpha \subseteq \mathrm{conv}(b, e^1, e^2)$, $\beta \subseteq \mathrm{conv}(b, e^2, e^3)$ and $\gamma \subseteq \mathrm{conv}(b, e^3, e^1)$ (with $\alpha, \beta, \gamma \in \Sigma$) let

$$\xi_{\alpha\beta} := (30x^{-2}y^2z^2 - 6x^{-2}y^2z)\frac{\partial}{\partial x} + 10x^2z^{-1}\frac{\partial}{\partial y} - 30x^2y^2z^{-2}\frac{\partial}{\partial z} \in \Gamma(\mathrm{Spec}\, B_{\alpha\beta}, S'),$$

$$\xi_{\beta\gamma} := (6x^{-2}y^2z - 30x^{-2}y^2z^2)\frac{\partial}{\partial x} + 25x^2y^{-3}z^2\frac{\partial}{\partial y} - 15x^2y^{-1}\frac{\partial}{\partial z} \in \Gamma(\mathrm{Spec}\, B_{\beta\gamma}, S')$$

and

$$\xi_{\gamma\alpha} := -(25x^2y^{-3}z^2+10x^2z^{-1})\frac{\partial}{\partial y}$$
$$+(30x^2y^2z^{-2}+15x^2y^{-1})\frac{\partial}{\partial z}\in\Gamma(\operatorname{Spec}B_{\gamma\alpha},S').$$

Because of

$$\xi_{\alpha\beta}+\xi_{\beta\gamma}+\xi_{\gamma\alpha}=0 \quad\text{and}\quad \xi_{\alpha\beta}(f)=\xi_{\beta\gamma}(f)=\xi_{\gamma\alpha}(f)=0,$$

this defines a 1-cocycle ξ of the sheaf $S\subseteq\theta_X$ (if $\sigma,\tau\in\Sigma$ are contained in a common cone $\operatorname{conv}(b,e^i,e^{i+1})$, let $\xi_{\sigma\tau}:=0$). We obtain

$$b_1^\alpha=0,\quad b_2^\alpha=-10x^2z^{-1},\quad b_3^\alpha=30x^2y^2z^{-2}$$
$$b_1^\beta=30x^{-2}y^2z^2-6x^{-2}y^2z,\quad b_2^\beta=b_3^\beta=0$$
$$b_1^\gamma=0,\quad b_2^\gamma=25x^2y^{-3}z^2,\quad b_3^\gamma=-15x^2y^{-1},\quad\text{hence}$$
$$d(x,y,z)=\frac{\partial f}{\partial x}b_1^\bullet+\frac{\partial f}{\partial y}b_2^\bullet+\frac{\partial f}{\partial z}b_3^\bullet=150x^2y^2z^2-30x^2y^2z.$$

The monomial x^2y^2z lies under the Newton boundary:

$$\langle a,(2,2,1)\rangle=59<60=m(a)$$
$$\langle b,(2,2,1)\rangle=5=m(b).$$

It is induced by the element $b_3^\gamma=-15x^2y^{-1}$. Here we see

$$\langle a,(2,-1,0)\rangle=14<15=a_3$$
$$\langle e^2,(2,-1,0)\rangle=-1<0=e_3^2,$$

but there is no chain $a=c^0,\dots,c^M=e^2$ of adjacent elements $c^\nu\in\Sigma^{(1)}$ with

$$\langle c^\nu,(2,-1,0)\rangle<c_i^\nu \quad\text{and}\quad c^\nu\neq e^3\ (\nu=0,\dots,M)$$

(cf. proof of (2.9)).

(3.4) *Remark.* The example together with the first part of (2.10) implies that even for $\dim R=2$ the image of the natural transformation $ESE\to\operatorname{Def}_R$ depends on the resolution π.

§ 4. *ESE* is formally smooth ($n=3$)

The obstructions for the formal smoothness of *ESE* are contained in $H^2(X,S)$. We can describe the cohomology groups of the sheaves S,S' (cf. (2.3)) by the methods of [Ke].

(4.1) **Lemma.** 1) *S' splits into a direct sum of invertible sheaves $S'=\bigoplus_{i=1}^n S_i'$; if we denote the embedding of the n-dimensional torus by j: $T^n\hookrightarrow X$, S_i' is given by*

$$S_i'=S'\cap j_*\mathcal{O}_T\frac{\partial}{\partial x_i}\qquad\left(S'\subseteq\theta_X\subseteq\bigoplus_{i=1}^n j_*\mathcal{O}_T\frac{\partial}{\partial x_i}\right).$$

2) $\frac{\partial}{\partial x_i} \mapsto 1$ *gives an embedding* $S_i' \hookrightarrow j_* \mathcal{O}_T$, *and* S_i' *becomes an invertible, T-invariant sheaf on X. The corresponding order function is linear on the cones of* Σ *and can be given by*

$$\begin{aligned} \operatorname{ord} S_i' \colon |\Sigma| = \mathbb{R}^n_{\geqq 0} &\to \mathbb{R} \\ a \in \Sigma^{(1)},\ a \neq e^i &\longmapsto a_i \\ e^i &\longmapsto 0. \end{aligned}$$

Proof. Each $\alpha \in \Sigma$ corresponds to local coordinates $y_1, \ldots, y_n$; in particular, let $\alpha = \langle e^1, \ldots, e^m, a^{m+1}, \ldots, a^n \rangle$ with $a^{m+1}, \ldots, a^n > 0$, then we obtain the following local description of $\pi \colon X \to (\mathbb{A}^n_k, 0)$:

$$\begin{aligned} x_i &= y_i \cdot y_{m+1}^{a_i^{m+1}} \cdot \ldots \cdot y_n^{a_i^n} \qquad (i \leqq m) \\ x_i &= y_{m+1}^{a_i^{m+1}} \cdot \ldots \cdot y_n^{a_i^n} \qquad (i \geqq m+1). \end{aligned}$$

It follows that

(i) $x_i \frac{\partial}{\partial x_i} = y_i \frac{\partial}{\partial y_i}$ for $i \leqq m$

(ii) $\sum_{i=m+1}^{n} k \cdot x_i \frac{\partial}{\partial x_i} = \sum_{i=m+1}^{n} k \cdot y_i \frac{\partial}{\partial y_i}$.

From this we get

$$\begin{aligned} S'|_{X_\alpha} &= \left\langle \frac{\partial}{\partial y_1}, \ldots, \frac{\partial}{\partial y_m}, y_{m+1} \frac{\partial}{\partial y_{m+1}}, \ldots, y_n \frac{\partial}{\partial y_n} \right\rangle \\ &= \left\langle \frac{x_1}{y_1} \frac{\partial}{\partial x_1}, \ldots, \frac{x_m}{y_m} \frac{\partial}{\partial x_m}, x_{m+1} \frac{\partial}{\partial x_{m+1}}, \ldots, x_n \frac{\partial}{\partial x_n} \right\rangle. \end{aligned}$$

In general, for n-dimensional cones α we see that

$$S_i'|_{X_\alpha} = \begin{cases} \left[x_i \frac{\partial}{\partial x_i} \right] & \text{for } e^i \notin \alpha \\ \left[\frac{x_i}{y_i} \frac{\partial}{\partial x_i} \right] & \text{for } e^i \in \alpha \ (y_i \hat{=} e^i). \end{cases} \quad \square$$

(4.2) **Proposition.**

$$\begin{aligned} H^{n-1}(X, S') &= H^n(X, S') = 0 \\ H^{n-1}(X, S) &= H^n(X, S) = 0. \end{aligned}$$

Proof. The sheaves S and S' differ only in $\mathcal{N}_{Y|X}$; since $H^\nu(X, \mathcal{N}_{Y|X}) = 0$ for $\nu \geqq 1$ (this can easily be shown by the same methods used below), it will be enough to show the first statement.

From [Ke] we know

$$H^\nu(X, S_i') = \bigoplus_{r \in M} \tilde{H}^{\nu-1}(B_r, k) \quad \text{with}$$

$$B_r := \{a \in \Delta / \langle a, r \rangle < (\operatorname{ord} S_i')(a)\}$$

(cf. (3.1) for the definition of Δ).

Let $\Delta' \subseteq \Delta$ be the subset corresponding to the union of closed cones $\alpha \in \Sigma$ not containing e^i (cf. Fig. 2).

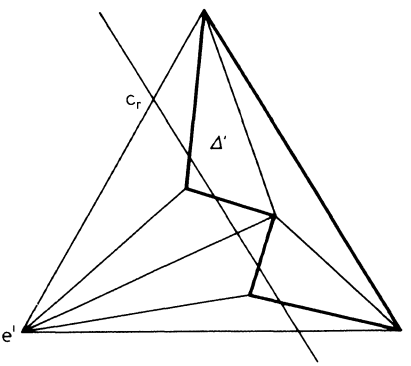
We denote $C_r := \{a \in \Delta / \langle a, r \rangle < a_i\}$, then $B_r \subseteq C_r$ and $B_r \cap \Delta' = C_r \cap \Delta'$.

Case 1: $e^i \in B_r$. Here B_r can be contracted to the point e^i, and we get $\tilde{H}^*(B_r, k) = 0$.

Case 2: $e^i \notin C_r$. Then, B_r and C_r can be contracted to the common subspace $B_r \cap \Delta' = C_r \cap \Delta'$, and we get

$$\tilde{H}^*(B_r, k) = \tilde{H}^*(C_r \cap \Delta', k) = \tilde{H}^*(C_r, k) = 0.$$

Case 3: $e^i \notin B_r$, $e^i \in C_r$. Again we have to look at the cohomology groups $\tilde{H}^*(C_r \cap \Delta', k)$:



By the projection from e^i onto the face $\sigma = \operatorname{conv}\{e^j / j \neq i\}$ we see that $C_r \cap \Delta'$ is homotopy equivalent to an open subset of $\mathbb{R}^{n-2}$; in particular, we get

$$\tilde{H}^{\nu-1}(B_r, k) = \tilde{H}^{\nu-1}(C_r \cap \Delta', k) = 0 \quad \text{for } \nu = n-1, n. \quad \square$$

(4.3) As a corollary we get

Proposition. *For* $\dim R = 2$ ($\hat{=} n = 3$) *the functor ESE* (*defined by a resolution* π: $X \to (\mathbb{A}_k^3, 0)$ *described in* (1.4)) *is formally smooth.*

§ 5. The natural transformation $ESE \to ES$

In § 2–§ 4 we only considered the functor of embedded equisingular deformations, but our main interest lies in the description of the subfunctor $\overline{ES} \subseteq \operatorname{Def}_R$ – the image of the natural transformation $ES \to \operatorname{Def}_R$. So it remains

to compare both functors, ESE and ES: Under which conditions is the natural transformation $ESE \to ES$ smooth?

(5.1) From (2.3) we get the commutative diagram of sheaves on X

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \mathscr{J}_Y S' & \longrightarrow & S & \longrightarrow & S_Y & \longrightarrow & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \longrightarrow & \mathscr{J}_Y S' & \longrightarrow & S' & \longrightarrow & S' \otimes \mathscr{O}_Y & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \oplus x_i} & & \downarrow{\scriptstyle \oplus x_i} & & \downarrow{\scriptstyle \oplus x_i} & & \\
0 & \longrightarrow & \mathscr{J}_Y^n & \longrightarrow & \mathscr{O}_X^n & \longrightarrow & \mathscr{O}_Y^n & \longrightarrow & 0
\end{array}$$

(the second vertical maps are defined by applying the derivations to the global sections $x_1, \ldots, x_n \in \Gamma(X, \mathscr{O}_X)$). On the cohomological level we obtain

$$\begin{array}{ccccc}
H^1(X, S) & \longrightarrow & H^1(X, S_Y) & \longrightarrow & H^2(X, \mathscr{J}_Y S') \\
\downarrow & & \downarrow{\scriptstyle \oplus x_i} & & \downarrow{\scriptstyle \oplus x_i} \\
0 & \longrightarrow & H^1(X, \mathscr{O}_Y^n) & \longrightarrow & H^2(X, \mathscr{J}_Y^n).
\end{array}$$

Given a $k[\varepsilon]$-deformation $[\tilde{E}_v \hookrightarrow \tilde{Y}] \in H^1(X, S_Y)$, the condition for lifting the section $x_i \in \Gamma(Y, \mathscr{O}_Y)$ to $\Gamma(\tilde{Y}, \mathscr{O}_{\tilde{Y}})$ is exactly the vanishing of the map x_i: $H^1(X, S_Y) \to H^1(X, \mathscr{O}_Y)$. In particular, we obtain

$$ES(k[\varepsilon]) = \operatorname{Ker}[H^1(X, S_Y) \to H^1(X, \mathscr{O}_Y^n)],$$

hence the obstructions for the smoothness of $ESE \to ES$ are contained in

$$\operatorname{Ker}[H^2(X, \mathscr{J}_Y S') \xrightarrow{\oplus x_i} H^2(X, \mathscr{J}_Y^n)].$$

(5.2) **Definition.** Let $a \in \Delta \cap (N \otimes_{\mathbb{Z}} \mathbb{Q})$. For $i = 1, \ldots, n$ we denote by $P_i(a)$ the point of $\Delta \cap (N \otimes_{\mathbb{Z}} \mathbb{Q})$ characterized by the following conditions:

1) $P_i(a) \in \overline{a e^i}$.

2) If $P_i(a)$ is given by integer coordinates with greatest common divisor 1, then $\{P_i(a), e^i\}$ will be a part of a $\mathbb{Z}$-basis of N.

3) The distance between a and $P_i(a)$ is minimal.

Remark. Let $a = (a_1, \ldots, a_n)$ with $a_i \in \mathbb{Z}$ and $\gcd(a_1, \ldots, a_n) = 1$; we denote $d := \gcd(a_1, \ldots, \hat{a}_i, \ldots, a_n)$. Then $P_i(a)$ is given by

$$P_i(a) = \left(\frac{a_1}{d}, \ldots, \frac{a_i + k}{d}, \ldots, \frac{a_n}{d} \right)$$

with $k \in \mathbb{Z}$, $0 \leqq k < d$, $d \mid a_i + k$.

(5.3) **Definition.** For the polynomial f with the corresponding f.r.p.p. decomposition Σ_0, a smooth subdivision $\Sigma < \Sigma_0$ is called "good" if the following condition is satisfied for all $i = 1, \ldots, n$:

Whenever an element $a\in\Sigma_0^{(1)}$ is contained, with e^i, in a common cone $\alpha\in\Sigma_0$, then there exists a cone $\tilde{\alpha}\in\Sigma$ containing $P_i(a)$ and e^i.

(5.4) **Proposition.** *For good subdivisions $\Sigma<\Sigma_0$ the map*

$$H^{n-1}(X, \mathscr{J}_Y S') \xrightarrow{\oplus x_i} H^{n-1}(X, \mathscr{J}_Y^n)$$

is injective.

Hence, for $\dim R=2$ the natural transformation $ESE\to ES$ is formally smooth.

Proof. The exceptional divisors $D_1, \ldots, D_{N'}\hookrightarrow X$ correspond to the elements $a\in\Sigma^{(1)}$ with $m(a)>0$, and from the construction of π: $X\to(\mathbb{A}_k^n, 0)$ it is clear that

$$Y+\sum_{a\in\Sigma^{(1)}} m(a)\, D_a=(f) \quad \text{in } \operatorname{Div} X.$$

So we get an isomorphism

$$\mathscr{O}_X(-Y)\cong\mathscr{O}_X(\sum_{a\in\Sigma^{(1)}} m(a)\, D_a),$$

and $\mathscr{J}_Y$ becomes an invertible, T-invariant subsheaf of $j_*\mathscr{O}_T$ with order function $-m(a)$. As we know the order function of S_i' too (cf. (4.1)), by [Ke] it remains to prove the following fact ($i=1, \ldots, n$):

$$\varphi(a):=\begin{cases} m(a)-a_i & \text{for } a\in\Sigma^{(1)},\ a\neq e^i \\ 0 & \text{for } a=e^i, \end{cases}$$

$$B_r:=\{a\in\Delta/\langle a, r\rangle>\varphi(a)\} \quad \text{and}$$

$$C_r:=\{a\in\Delta/\langle a, r\rangle>m(a)\}\subseteq B_r,$$

then the maps $H^{n-2}(B_r, k)\to H^{n-2}(C_r, k)$ are injective for all $r\in M$.

Step 1: Under the assumption of the proposition the sets $\Delta\setminus B_r$ are connected.

We can assume $i=1$, let us fix some notations:

As in (4.2) $\Delta'\subseteq\Delta$ is the subset corresponding to the union of closed Σ-cones not containing e^1;

$$\partial\Delta':=(\text{boundary of } \Delta')\setminus\operatorname{int}(\operatorname{conv}(e^2, \ldots, e^n)),$$

$$\Delta_r:=\{a\in\Delta/\langle a, r\rangle\leqq m(a)-a_1\}\subseteq\Delta\setminus B_r.$$

Then Δ_r is convex (hence connected) and

$$\Delta_r\cap\Delta'=(\Delta\setminus B_r)\cap\Delta'.$$

We have to distinguish two cases:

1) $e^1\notin\Delta\setminus B_r$. φ is linear on the cones of Σ, therefore each $a\in\Delta\setminus B_r$ is connected, in $\Delta\setminus B_r$, with some element of $\Sigma^{(1)}\cap(\Delta\setminus B_r)$.

From $e^1\notin\Delta\setminus B_r$ we get $\Sigma^{(1)}\cap(\Delta\setminus B_r)\subseteq\Delta'$, hence

$$\Sigma^{(1)}\cap(\Delta\setminus B_r)=\Sigma^{(1)}\cap\Delta_r.$$

So $a \in \Delta \setminus B_r$ is connected, in $\Delta \setminus B_r$, with some element of Δ_r.

2) $e^1 \in \Delta \setminus B_r$. It remains to show that e^1 and Δ_r belong to the same connected component of $\Delta \setminus B_r$. If this were not the case, we would get $\emptyset \neq \Delta_r \subseteq \Delta' \setminus \partial \Delta'$, and therefore $r = (0, r_2, \ldots, r_n)$ with $r_2, \ldots, r_n > 0$.

(i) For $c, d \in \Delta \cap (N \otimes_{\mathbb{Z}} \mathbb{Q})$ the existence of an element $s \in \Gamma_+(f)$ with $\langle c, s \rangle = m(c)$ and $\langle d, s \rangle = m(d)$ is equivalent to the existence of a Σ_0-cone containing both c and d.

(ii) There exists an element $a \in \Sigma_0^{(1)} \cap \Delta_r$ lying with e^1 in a common cone of Σ_0:

Let $b \in \Delta_r \subseteq \Delta'$ be an arbitrary element. Then there exists an n-dimensional cone $\alpha \in \Sigma_0$ (with $e^1 \in \alpha$) such that $\overline{be^1}$ has a non-zero intersection with a closed face $\beta < \alpha$ not containing e^1; we denote the intersection point by a'. a' belongs to a closed Σ_0-cone not containing e^1, so (i) implies the existence of $s \in \Gamma_+(f)$ with $s_1 > 0$ and $\langle a', s \rangle = m(a')$. Now there are positive reals λ_1, λ_2 with $a' = \lambda_1 e^1 + \lambda_2 b$, and from $\langle b, r \rangle \leqq m(b) - b_1$ ($b \in \Delta_r$) we obtain

$$\begin{aligned} \langle a', r \rangle &= \lambda_2 \langle b, r \rangle \leqq \lambda_2 m(b) - \lambda_2 b_1 \leqq \lambda_2 \langle b, s \rangle - \lambda_2 b_1 \\ &\leqq \langle a', s \rangle - \lambda_1 s_1 - a_1' + \lambda_1 \\ &\leqq m(a') - a_1' - \lambda_1 (s_1 - 1) \leqq m(a') - a_1', \end{aligned}$$

hence $a' \in \Delta_r$. Finally, m is linear on all Σ_0-cones, so one of the vertices a of β must be contained in Δ_r.

(iii) $a \in \Sigma_0^{(1)} \cap \Delta_r$ and e^1 belong to a common Σ_0-cone, so (i) implies that there is an element $t \in \Gamma_+(f)$ with $t_1 = 0$ and $\langle a, t \rangle = m(a)$. Hence, $m(a) = a_2 t_2 + \ldots + a_n t_n$, and with $d := \gcd(a_2, \ldots, a_n)$ we obtain $d \mid m(a)$.

Now $a \in \Delta_r$ implies

$$a_2 r_2 + \ldots + a_n r_n = \langle a, r \rangle \leqq m(a) - a_1,$$

and $d \mid m(a)$ gives the stronger inequality

$$\begin{gathered} a_2 r_2 + \ldots + a_n r_n \leqq m(a) - (a_1 + k) \\ (\text{with } k \in \mathbb{Z},\ 0 \leqq k < d \text{ and } d \mid a_1 + k). \end{gathered}$$

But this means

$$d \cdot \langle P_1(a), r \rangle \leqq m(a) - (a_1 + k) = d \cdot m(P_1(a)) - (a_1 + k),$$

hence $P_1(a) \in \Delta_r$. By the assumption of the proposition the point $P_1(a)$ is not contained in $\Delta' \setminus \partial \Delta'$, and we get a contradiction.

Step 2: By $\Delta \hookrightarrow \mathbb{R}^{n-1} \hookrightarrow S^{n-1}$ we get an embedding of Δ into an oriented closed manifold.

$$B_r^* := S^{n-1} \setminus (B_r \cap \operatorname{int} \Delta) = (S^{n-1} \setminus \operatorname{int} \Delta) \cup (\Delta \setminus B_r) \quad \text{and}$$

$$C_r^* := S^{n-1} \setminus (C_r \cap \operatorname{int} \Delta) = (S^{n-1} \setminus \operatorname{int} \Delta) \cup (\Delta \setminus C_r)$$

are compact subsets of S^{n-1}, both are unions of two connected sets. Now the following is true:

$$h^0(B_r^*)=1 \quad \text{iff } [B_r\cap\partial\Delta\neq\partial\Delta \quad \text{or } B_r=\Delta]$$
$$h^0(B_r^*)=2 \quad \text{iff } [B_r\cap\partial\Delta=\partial\Delta \text{ and } B_r\neq\Delta],$$

and the same formula holds for C_r instead of B_r.

Since $B_r\cap\partial\Delta=C_r\cap\partial\Delta$, the induced map $H^0(C_r^*,k)\to H^0(B_r^*,k)$ must be surjective; the Alexander-duality completes the proof of the proposition. □

For the last part of this paper we assume $\dim R=2$ $(n=3)$

(5.5) **Theorem.** 1) *If the 3-dimensional "boundary-cones" of* Σ_0 *are simplicial (that means, each of the coordinate axes of* $\mathbb{R}^3$ *is intersected by exactly one top dimensional face of the Newton boundary* $\Gamma(f)$*), it will always be possible to find a good subdivision* $\Sigma<\Sigma_0$.

2) *The existence of a good subdivision* $\Sigma<\Sigma_0$ *implies the formal smoothness of* $\overline{ES}$.

Proof. The first part is quite clear, and the rest is a direct consequence of (4.3) and (5.4). □

Remark. Laufer shows in [La] that analytic families (over a reduced parameter space) of homeomorphic isolated 2-dimensional hypersurface singularities admit weak simultaneous resolutions. By his formula for the Milnor number it is clear that such deformations are μ-constant. Together with the announced result of Perron (μ-constant-deformations do not change the topological type), this would imply that, over reduced parameter spaces, $\overline{ES}$ describes exactly the μ-constant-deformations.

Recently, Karras has shown the existence of an analytic semiuniversal deformation for $\overline{ES}$. Then the formal smoothness of $\overline{ES}$ implies the smoothness of the μ-constant-stratum of f.

(5.6) *Example.* In §3, the cone $\operatorname{conv}(e^2,e^3,a,b)$ is not simplicial. But by avoiding to destroy the cones $\overline{P_1(a)\,e^1}$, $\overline{P_3(a)\,e^3}$, $\overline{b\,e^1}$ and $\operatorname{conv}(b,e^2,e^3)$, a good subdivision $\Sigma<\Sigma_0$ will be obtained. Hence, $\overline{ES}$ is formally smooth; indeed, the element $f(x,y,z)+2\varepsilon x^2y^2z\in\overline{ES}(k[\varepsilon])$ can be lifted to the μ-constant $\mathbb{C}\{t\}$-deformation $f+2tx^2y^2z+t^2x^4y$.

(5.7) As a corollary of (2.6) and (5.4) we obtain the

Theorem. *Let* $\pi\colon X_\Sigma\to(\mathbb{A}_k^3,0)$ *be the resolution corresponding to a good subdivision* $\Sigma<\Sigma_0$. *Then,* $\overline{ES}$ *is exactly the functor of over-*$\Gamma(f)$*-deformations iff*

$$\operatorname{Coker}(H_D^1(X,S)\to H^1(X,S'))=0.$$

(5.8) If the unions K_i of all Σ_0-cones not containing e^i are convex for $i=1,2,3$, the "boundary-cones" of Σ_0 are simplicial automatically. Therefore, (2.10) implies the

Theorem. *If the sets* K_i *are convex* $(i=1,2,3)$, $\overline{ES}$ *will be exactly the functor of over-*$\Gamma(f)$*-deformations.*

References

[Ha] Hartshorne, R.: Algebraic geometry. Graduate Texts in Mathematics, vol. 52. Berlin-Heidelberg-New York: Springer 1977

[Ke] Kempf, G., Knudsen, F., Mumford, D., Saint-Donat, B.: Toroidal embeddings I. Lect. Notes Math., vol. 339. Berlin-Heidelberg-New York: Springer 1973

[Ku] Kurke, H.: Vorlesungen über algebraische Flächen. Teubner-Texte zur Mathematik, vol. 43. Leipzig: Teubner 1982

[KPR] Kurke, H., Pfister, G., Roczen, M.: Henselsche Ringe und algebraische Geometrie. Berlin: Deutscher Verlag der Wissenschaften 1975

[La] Laufer, H.B.: Weak simultaneous resolution for deformations of Gorenstein surface singularities. Proc. Symp. Pure Math. **40**/II, 1–30 (1983)

[Me] Merle, M.: Polyèdre de Newton, éventail et désingularisation, d'après A.N. Varchenko. In: Lect. Notes Math., vol. 777. Berlin-Heidelberg-New York: Springer 1980

[Sch] Schlessinger, M.: Functors of Artin rings. Trans. Am. Math. Soc. **130**, 208–222 (1968)

[Te] Teissier, B.: Résolution simultanée (I, II). In: Lect. Notes Math., vol. 777. Berlin-Heidelberg-New York: Springer 1980

[Va] Varchenko, A.N.: Zeta-function of monodromy and Newton's diagram. Invent. Math. **37**, 253–262 (1976)

[Wa] Wahl, J.: Equisingular deformations of normal surface singularities, I. Ann. Math. **104**, 325–356 (1976)

Oblatum 4-II-1986 & 3-XI-1986

Invent. math. 88, 635-667 (1987)

Inventiones mathematicae
© Springer-Verlag 1987

1-formes fermées singulières et groupe fondamental

Gilbert Levitt

Département de Mathématiques, Université Paris Sud, F-91405 Orsay Cedex, France

Summary. We study the influence of the π_1 of a closed manifold M^n ($n \geqq 3$) on the foliations of M defined by closed differential 1-forms with Morse singularities (of index $\neq 0, n$). Every nonexact form is cohomologous to a weakly complete one, that is one whose leaf space is of the same type as that of a nonsingular form. Generically, a form has compact leaves or is weakly complete. If $\pi_1 M$ has no quotient isomorphic to $\mathbb{Z} * \mathbb{Z}$, then every nonexact form on M is weakly complete. We also say a form ω is complete if every path in M is homotopic to either a path transverse to ω or a path contained in a leaf of ω. Completeness of ω depends only on its de Rham cohomology class. The set of complete cohomology classes depends only on $\pi_1 M$ and is related to finitely generated normal subgroups of $\pi_1 M$ with quotient $\simeq \mathbb{Z}$. If $\pi_1 M$ is nilpotent (or even polycyclic), every nonexact form on M is complete. On irreducible 3-manifolds, a form is complete iff it is cohomologous to a nonsingular one.

Introduction

Cet article est consacré aux 1-formes différentielles fermées singulières sur les variétés fermées de dimension $n \geqq 3$ (on notera que seules les variétés fibrées sur le cercle possèdent des 1-formes fermées non singulières [Ti]). Nous étudions principalement les trois questions suivantes:

1. Certains résultats sur les feuilletages définis par les 1-formes fermées non singulières s'étendent-ils aux formes singulières (en général, ou sur certaines variétés)?

2. Pour les formes non singulières, la classe de cohomologie de De Rham constitue un invariant très fort. En est-il de même pour les formes singulières? Des formes cohomologues ont-elles des propriétés topologiques en commun?

3. Si M^3 est une variété fermée irréductible de dimension 3, la norme sur l'homologie définie par Thurston [Th1] et le théorème de fibration de Stallings [St] décrivent le cône ouvert $\mathscr{C}(M) \subset H^1(M, \mathbb{R}) - \{0\}$ des classes de cohomologie représentables par une 1-forme fermée non singulière, et le relient à $\pi_1 M$. Cette théorie a-t-elle un équivalent en dimension plus grande?

Nous avons remarqué dans [AL] que, contrairement à ce qui se passe pour les formes non singulières, deux formes singulières cohomologues peuvent être très dissemblables; ainsi l'une peut être à feuilles régulières compactes alors que l'autre est à feuilles denses.

Toutefois nous allons voir que, sur certaines variétés et dans certaines classes de cohomologie, les formes se ressemblent beaucoup. Pour cela nous introduirons deux propriétés, la complétude et la faible complétude, que possèdent toutes les formes non singulières et certaines formes singulières (que, le plus souvent, nous supposerons à singularités de Morse, d'indice différent de 0 ou n).

Les *formes faiblement complètes*, sur une variété fermée M, seront essentiellement celles qui ont le même espace des feuilles qu'une forme non singulière.

Nous montrerons d'abord que *toute classe de cohomologie de De Rham* $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ *contient une forme faiblement complète;* en d'autres termes, *toute forme non exacte est cohomologue à une forme faiblement complète.*

En étudiant les perturbations d'une forme ω de Morse et minimale (c'est à dire à feuilles denses, ou de façon équivalente sans feuille L avec $L \cup \mathrm{Sing}\,\omega$ compact), nous verrons que, *si* ω *est minimale et si l'homomorphisme d'intégration* $[\omega]: \pi_1 M \to \mathbb{R}$ *ne peut pas se factoriser par un groupe libre, alors* ω *est faiblement complète.*

Nous montrerons aussi que, *génériquement, une forme est faiblement complète ou possède des feuilles compactes.*

Nous déterminerons également les *classes de cohomologie* $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ *dans lesquelles toutes les formes sont faiblement complètes:* ce sont exactement celles pour lesquelles l'homomorphisme de $\pi_1 M$ dans $\mathbb{R}$, obtenu en intégrant les formes de Ω, ne peut pas se *factoriser de façon non triviale par un produit libre* $\mathbb{Z}^a * \mathbb{Z}^b$ $(a, b \geqq 1)$.

En particulier, *s'il n'existe pas d'homomorphisme surjectif de* $\pi_1 M$ *sur* $\mathbb{Z} * \mathbb{Z}$, *alors toute forme sur* M *est faiblement complète.*

Sous une hypothèse plus forte sur $\pi_1 M$ (par exemple $\pi_1 M$ est nilpotent), nous montrerons que les formes fermées sur M sont en fait complètes.

Les *formes complètes* peuvent être définies de plusieurs façons équivalentes. En voici une: ω est complète si et seulement si tout chemin est homotope (à extrémités fixes) à un chemin contenu dans une feuille ou à un chemin transverse à ω.

Nous verrons que le fait pour une forme d'être ou non complète *ne dépend que de sa classe de cohomologie.* Ceci nous permettra de définir le cône $U(M) \subset H^1(M, \mathbb{R}) - \{0\}$ des *classes de cohomologie complètes.*

Nous montrerons que $U(M)$ est *ouvert, ne dépend que de* $\pi_1 M$, et que ses points à coordonnées entières correspondent aux homomorphismes non triviaux de $\pi_1 M$ dans $\mathbb{Z}$ dont le noyau est *de type fini.* Nous verrons également que $U(M)$, qui contient toujours $\mathscr{C}(M)$, lui est égal si M est une 3-variété irréductible.

Ces résultats nous font penser que $U(M)$, plus maniable et moins souvent vide que $\mathscr{C}(M)$, est un objet intéressant à étudier sur les variétés de dimension >3 (et sur certaines 3-variétés si la conjecture de Poincaré est fausse ...). On peut en fait en donner des définitions algébriques en fonction de $\pi_1 M$ (cf. partie V).

Énoncé des résultats

Commençons par rappeler quelques propriétés bien connues des 1-formes fermées non singulières sur les variétés fermées. Le lecteur peu familier avec les 1-formes fermées trouvera quelques généralités dans les préliminaires (sections I.1 à I.3).

Si le groupe des périodes $P(\omega)\subset\mathbb{R}$ de la forme non singulière ω est cyclique, toutes les feuilles du feuilletage défini par ω sont compactes (et sont les fibres d'une fibration sur S^1). Si au contraire ω possède deux périodes rationnellement indépendantes, toutes les feuilles sont denses; nous dirons que ω est *minimale*.

Deux formes cohomologues et proches sont isotopes [Mo], et en dimension $\leqq 3$ la classe de cohomologie détermine en fait la forme à isotopie près ([BL], [QR]). Notons enfin une propriété importante des formes non singulières: tout chemin sur lequel l'intégrale de ω est nulle est homotope à un chemin contenu dans une feuille.

Considérons maintenant des formes singulières. Nous travaillerons toujours sur une *variété fermée orientable* M^n ($n\geqq 3$), et nous étudierons des formes ω à *singularités de Morse, sans centre* (i.e. tout point de M a un voisinage sur lequel ω est la différentielle d'une fonction de Morse sans point critique d'indice 0 ou n; noter qu'une telle forme ne peut pas être exacte).

Les résultats se généralisent en fait aux formes de Morse avec centres, et même aux formes dont le lieu singulier est quelconque mais de codimension $\geqq 3$ (voir partie VI).

Une forme de Morse sans centre ω est *faiblement complète* si, dès que deux points de $M-\mathrm{Sing}\,\omega$ peuvent être joints par un chemin sur lequel l'intégrale de ω est nulle, alors ils appartiennent à la même feuille; cela revient à dire que l'espace des feuilles du feuilletage est du même type que si ω était non singulière: les feuilles sont les images réciproques des points par l'application $\psi\colon M-\mathrm{Sing}\,\omega\to\mathbb{R}/P(\omega)$ définie par

$$\psi(x)=\int_{x_0}^{x}\omega \qquad (x_0 \text{ étant un point base}).$$

Les formes faiblement complètes sont également celles pour lesquelles, pour toute singularité s d'indice 1 ou $n-1$, les deux bouts singuliers issus de s sont contenus dans la même feuille (voir figure 1).

Soit ω une forme faiblement complète. Si $P(\omega)$ est cyclique, toutes les feuilles régulières de ω sont compactes (et homologues entre elles); si par contre $P(\omega)$ est de rang $\geqq 2$, alors le feuilletage est minimal (et uniquement ergodique).

Nous commencerons par montrer (grâce à une construction de Arnoux-Levitt [AL]):

Théorème 1. *Toute classe de cohomologie* $\Omega\in H^1(M,\mathbb{R})-\{0\}$ *est représentable par une forme faiblement complète.*

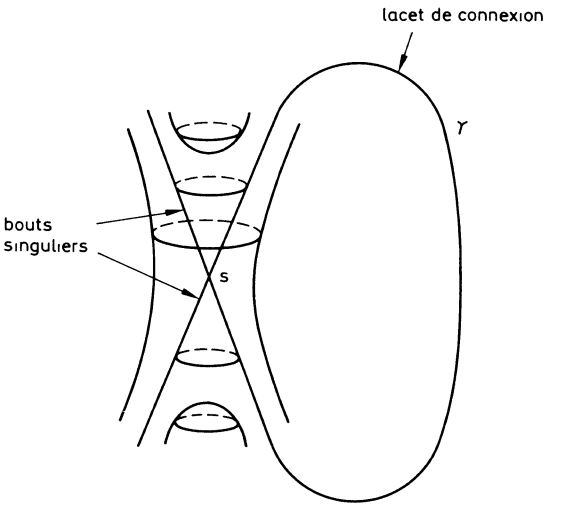


Fig. 1

Corollaire. *Tout homomorphisme surjectif de $\pi_1 M$ dans $\mathbb{Z}$ peut se représenter par une application p de M dans un cercle, à singularités de Morse, et à images réciproques $p^{-1}(c)$ connexes.*

Nous caractériserons les classes de cohomologie dans lesquelles toutes les formes de Morse sans centre sont faiblement complètes. Soit ω une forme de Morse sans centre, et $[\omega]$: $\pi_1 M \to \mathbb{R}$ l'homomorphisme obtenu par intégration de ω sur les lacets. Nous avons la dichotomie suivante:

Théorème 2. *Si $[\omega]$ ne peut pas se factoriser par un épimorphisme α: $\pi_1 M \twoheadrightarrow \mathbb{Z}^a * \mathbb{Z}^b$ $(a, b \geqq 1)$, alors ω (et toute forme cohomologue à ω) est faiblement complète.*

Si au contraire il y a factorisation par un $\mathbb{Z}^a * \mathbb{Z}^b$, alors ω est cohomologue à une forme (de Morse sans centre) *non faiblement complète* (possédant même une feuille compacte qui sépare M).

En particulier:

Corollaire. *S'il n'existe pas d'épimorphisme de $\pi_1 M$ sur $\mathbb{Z} * \mathbb{Z}$, toute forme sur M est faiblement complète, et réciproquement.*

Dans la partie VI nous déduirons du théorème 2 un résultat sur les formes dont les singularités ne sont pas de Morse:

Corollaire. *S'il n'existe pas d'épimorphisme de $\pi_1 M$ sur $\mathbb{Z} * \mathbb{Z}$, une forme sur M dont le lieu singulier est de codimension $\geqq 3$ ne possède pas de feuille exceptionnelle.*

Revenons aux formes de Morse sans centre. Avec une hypothèse sur $[\omega]$ plus faible que dans le théorème 2, nous obtenons la faible complétude pour les formes minimales (améliorant le théorème 5 de [AL]):

Théorème 3. *Si ω est minimale et si $[\omega]: \pi_1 M \to \mathbb{R}$ ne peut pas se factoriser par un groupe libre, alors ω est faiblement complète.*

Rappelons qu'une forme de Morse est minimale ssi il n'existe pas de feuille L avec $L \cup \mathrm{Sing}\,\omega$ compact (cf. I.3).

Les théorèmes 2 et 3 permettent, sous certaines hypothèses de non-factorisation de l'homomorphisme naturel $p_M: \pi_1 M \to H_1(M, \mathbb{Z})/\text{torsion}$, de prouver la faible complétude des formes (resp. des formes minimales) dans presque toute classe de cohomologie (voir partie III).

Les théorèmes 2 et 3 se déduisent de l'étude des formes ω' cohomologues à ω et proches de ω (dans la topologie C^∞). En fait:

Théorème 4. 1) *Si ω n'est pas faiblement complète, elle est approchable par des formes cohomologues possédant des feuilles compactes.*

2) *Si ω est minimale mais pas faiblement complète, elle est approchable par des formes cohomologues dont toutes les feuilles régulières sont compactes, et aussi par des formes cohomologues faiblement complètes.*

3) *Si ω est faiblement complète, il en est de même des formes cohomologues à ω et proches de ω.*

Soit $\Omega \in H^1(M, \mathbb{R}) - \{0\}$, et soit $F(\Omega)$ l'espace de toutes les 1-formes fermées représentant Ω, muni de la topologie C^∞. On déduit du théorème 4 (assertion 1):

Corollaire. *Les formes de Morse qui possèdent des feuilles compactes, ou qui n'ont pas de centre et sont faiblement complètes, forment un ouvert dense dans $F(\Omega)$.*

Passons maintenant aux *formes complètes*. On peut les définir de plusieurs façons équivalentes (voir propositions II.2 et IV.6). Par exemple: tout chemin d'intégrale nulle est homotope à un chemin contenu dans une feuille; pour toute feuille L, la suite $\pi_1 L \to \pi_1 M \xrightarrow{[\omega]} \mathbb{R}$ est exacte; pour toute singularité s d'indice 1 ou $n-1$, on peut joindre les deux bouts singuliers issus de s par un lacet de connexion γ (voir figure 1) homotope à 0 dans M.

Le fait qu'une forme soit complète ou non ne dépend que de sa classe de cohomologie:

Théorème 5. *Soient ω et ω' deux formes cohomologues. Alors ω' est complète si et seulement si ω l'est.*

Si $\pi_1 M$ est abélien, ou nilpotent, ou plus généralement si son sous-groupe des commutateurs est de type fini, *toute forme sur M est complète.* Cela se déduit du résultat suivant:

Théorème 6. *Si le noyau de $[\omega]: \pi_1 M \to \mathbb{R}$ est de type fini, alors ω est complète. La réciproque est vraie si $P(\omega)$ est cyclique.*

En général, soit $U(M) \subset H^1(M, \mathbb{R}) - \{0\}$ l'ensemble des *classes de cohomologie complètes* (c'est-à-dire représentées par des formes complètes).

Théorème 7. a) *Le cône $U(M) \subset H^1(M, \mathbb{R}) - \{0\}$ est ouvert.*

b) *$U(M)$ ne dépend que de $\pi_1 M$: si $\rho: \pi_1 M \to \pi_1 M'$ est un isomorphisme, alors $\rho^*: H^1(M', \mathbb{R}) \to H^1(M, \mathbb{R})$ envoie $U(M')$ sur $U(M)$.*

Si ω est à périodes rationnelles, alors d'après le théorème 6 elle est complète si et seulement si $\mathrm{Ker}[\omega]$ est de type fini. L'assertion a du théorème 7 a donc une conséquence algébrique (déjà connue):

Corollaire [Ne]. *Soit G un groupe de présentation finie. Dans* $\mathrm{Hom}(G, \mathbb{Q}) - \{0\}$, *l'ensemble des homomorphismes dont le noyau est de type fini forme un ouvert.*

Nous verrons que, *si M est une 3-variété irréductible, une forme est complète si et seulement si elle est cohomologue à une forme non singulière;* en d'autres termes, $U(M) = \mathscr{C}(M)$. Dans ce cas $U(M)$ est donc de type polyédral [Th1].

Ceci contraste fortement avec des exemples [BNS] où $U(M) \cup \{0\}$ contient $H^1(M, \mathbb{Q})$ sans être égal à $H^1(M, \mathbb{R})$ (nous donnerons à la fin de la partie V la relation entre $U(M)$ et l'invariant Σ défini algébriquement dans [BNS] à partir du groupe $G = \pi_1 M$). En particulier, il existe des variétés sur lesquelles toutes les classes de cohomologie de rang 1 sont complètes sans que toutes les classes de rang $\geqq 2$ le soient.

Remerciements. Le point de départ de cette recherche a été un article [Hen] dans lequel Damir Henč montre qu'une forme de Morse ω_0 sans feuille compacte, cohomologue à une forme ω_1 minimale non singulière, est minimale et uniquement ergodique. Après avoir étudié cet article avec Pierre Arnoux, je me suis aperçu que la démonstration de Henč donnait plus que l'unique ergodicité de ω_0; elle donne en fait la complétude, et marche également si ω_1 est une forme de Morse sans singularité d'indice 1 ou $n-1$. Ceci, ainsi qu'un autre texte de Henč et des conversations avec lui, m'a conduit à étudier les formes complètes et à montrer que la complétude d'une forme ne dépend que de sa classe de cohomologie. Henč a ainsi été une grande source d'inspiration, et je l'en remercie.

Le présent article s'est développé en même temps que celui avec Arnoux sur l'unique ergodicité [AL]. D'ailleurs la démonstration du théorème III.1 est essentiellement une construction de [AL], et ma première démonstration du théorème III.4 reprenait dans un contexte différent l'idée d'induire la forme sur une sous-variété. Pierre Arnoux a été pour moi un interlocuteur patient et attentif, qui m'a beaucoup aidé à développer mes idées. Je lui suis très reconnaissant pour toutes les conversations que nous avons eues.

Plan de l'article

I. Préliminaires 640
II. Définition des formes faiblement complètes et complètes 644
III. Formes et classes de cohomologie faiblement complètes 647
IV. Structure transverse et π_1 des feuilles 655
V. Classes de cohomologie complètes 658
VI. Généralisation 663

Certains résultats de cet article ont été annoncés dans [Le]. Les parties IV et V sont pratiquement indépendantes de la partie III. Les théorèmes 1 à 4 sont montrés dans la partie III, les théorèmes 5 à 7 dans la partie V.

I. Préliminaires

Soit ω une 1-forme fermée sur une variété compacte sans bord M^n de dimension $n \geqq 3$.

I.1. L'*homomorphisme* de $\pi_1 M$ dans $\mathbb{R}$ obtenu en intégrant ω le long des lacets sera noté $[\omega]$. Deux formes définissent le même homomorphisme si et seulement si leur différence est exacte, et nous identifierons l'homomorphisme $[\omega]$ à la *classe de cohomologie de De Rham* $[\omega] \in H^1(M, \mathbb{R})$. D'autre part nous confondrons souvent un lacet et l'élément qu'il détermine dans $\pi_1 M$.

L'image de $[\omega]$ est le *groupe des périodes* $P(\omega) \subset \mathbb{R}$. Ce groupe est de type fini, donc isomorphe à un $\mathbb{Z}^{r(\omega)}$; $r(\omega)$ s'appelle le *rang de* ω (ou de sa classe $[\omega]$). Le rang est nul pour les formes exactes, égal à 1 si $P(\omega)$ est infini cyclique, et $\geqq 2$ dès que ω possède deux périodes rationnellement indépendantes ($P(\omega)$ est alors dense dans $\mathbb{R}$).

Une forme ω (ou sa classe de cohomologie) est *totalement irrationnelle* si son rang est maximal, c'est à dire égal au premier nombre de Betti $b_1 M$. Une autre caractérisation est que le noyau de $[\omega]$: $\pi_1 M \to \mathbb{R}$ est égal au noyau de l'homomorphisme naturel p_M: $\pi_1 M \to H_1(M, \mathbb{Z})$/torsion (bien sûr on a toujours $\operatorname{Ker}[\omega] \supset \operatorname{Ker} p_M$). Dans $H^1(M, \mathbb{R})$ *l'ensemble des classes totalement irrationnelles est un* G_δ *dense de complémentaire négligeable* (pour la mesure de Lebesgue).

I.2. Une *primitive locale* de ω est une fonction f à valeurs réelles, définie sur un ouvert de M et vérifiant $df = \omega$. Sur tout ouvert simplement connexe ω admet une primitive locale, unique à une constante additive près.

Les *singularités* de ω sont les points x où $\omega_x = 0$; ils correspondent aux points critiques des primitives locales. L'ensemble des singularités de ω est noté $\operatorname{Sing}\omega$, et son complémentaire M^*.

Une forme ω est *de Morse* si tout point singulier x de ω est un point critique non dégénéré pour les primitives locales de ω. Les singularités d'une forme de Morse sont en nombre fini et ont un *indice* compris entre 0 et n. Une singularité d'indice 0 ou n est un *centre*; une singularité d'indice 1 ou $n-1$ sera dite *conique* (voir I.4). On notera qu'une forme de Morse sans centre ne peut pas être exacte.

I.3. Sur $M^* = M - \operatorname{Sing}\omega$ la forme ω est non singulière et définit un *feuilletage de codimension* 1, donné par les niveaux des primitives locales de ω.

La forme ω détermine une *mesure transverse invariante* du feuilletage, obtenue en associant à un intervalle (ou à une courbe fermée) transverse I sa *longueur* $|I| = |\int_I \omega|$. Le nombre $|I|$ sera parfois appelé *distance* entre les extrémités de I.

Le feuilletage possède également une *orientation transverse*, donnée par le sens de croissance des primitives locales. En particulier toute feuille compacte est transversalement orientée.

Une *feuille de* ω sera par définition une feuille du feuilletage ci-dessus. L'*espace des feuilles* M^*/ω de ω est le quotient de M^* par la relation d'équivalence «être sur la même feuille».

Si $r(\omega) \leqq 1$, on montre que toute feuille L de ω est *fermée* dans M^*, ou de façon équivalente que $L \cup \operatorname{Sing}\omega$ est compact (si ω est de Morse, toutes les feuilles sauf un nombre fini sont compactes).

En général, nous notons $N(\omega)$ l'union des feuilles L telles que $L \cup \operatorname{Sing}\omega$ ne soit pas compact. Si ω est de Morse, c'est un ouvert possédant un nombre fini

de composantes, dites *composantes minimales* car toute feuille de $N(\omega)$ est dense dans la composante qui la contient ([Im1]; voir appendice de [AL]).

Soit ω une forme de Morse sans centre, et x_0 un point base. On peut associer à tout $x \in M^*$ un élément bien défini $\psi(x) = \int_{x_0}^{x} \omega$ dans $\mathbb{R}/P(\omega)$. On obtient ainsi une surjection $\psi: M^* \to \mathbb{R}/P(\omega)$, qui se factorise par l'espace des feuilles M^*/ω. Les formes faiblement complètes seront celles pour lesquelles ψ induit une bijection entre M^*/ω et $\mathbb{R}/P(\omega)$.

I.4. Soit ω une forme de Morse. Si s est une singularité d'indice compris entre 2 et $n-2$, il existe des voisinages arbitrairement petits U de s tels que tous les niveaux $f^{-1}(c)$ des primitives locales $f: U \to \mathbb{R}$ soient *connexes*. La singularité s n'affecte donc pas la structure transverse du feuilletage défini par ω, et nous verrons que les formes n'ayant que des singularités d'indice compris entre 2 et $n-2$ se comportent dans une large mesure comme des formes non singulières.

Au contraire, comme l'avait remarqué Henč, les singularités d'indice 1 ou $n-1$ (que nous appelons *coniques*) jouent un rôle particulier car il en part *deux bouts singuliers* (voir Fig. 1 plus haut).

Il se peut que ces deux bouts singuliers soient contenus dans la même feuille de ω, c'est à dire qu'il existe un *lacet de connexion* γ avec $\gamma - \{s\}$ contenu dans une feuille de ω (voir Fig. 1). Ce sera une caractérisation des formes faiblement complètes; pour une forme complète, on peut trouver des lacets de connexion homotopes à 0 dans M (voir partie II).

I.5. Étant donné une forme ω, nous considérerons le *revêtement d'intégration* $\hat{p}: \hat{M} \to M$ associé à l'homomorphisme $[\omega]: \pi_1 M \to \mathbb{R}$. Le π_1 de $\hat{M}$ est ainsi isomorphe à $\mathrm{Ker}[\omega]$, et le groupe de transformations du revêtement est isomorphe à $P(\omega)$. Nous noterons

$$\hat{M}^* = \hat{p}^{-1}(M^*) = \hat{M} - \mathrm{Sing}(\hat{p}^* \omega).$$

La forme $\hat{p}^* \omega$ est exacte, et nous choisirons une *primitive globale* $\hat{f}: \hat{M} \to \mathbb{R}$ (unique à une constante additive près). Si η est une transformation du revêtement, la fonction $\hat{f} \circ \eta - \hat{f}$ est une constante $h(\eta) \in P(\omega)$; l'application $\eta \mapsto h(\eta)$ est un isomorphisme entre le groupe du revêtement et $P(\omega)$.

Quand aucune confusion ne sera possible, nous désignerons par $\hat{M}_{a,b}$ l'ensemble des points de $\hat{M}$ où la valeur de $\hat{f}$ appartient à $]a, b[$ (pour $-\infty \leqq a < b \leqq +\infty$).

Soit ω' cohomologue à ω, et $\hat{p}^* \omega' = d\hat{f}'$. Alors la différence $\hat{f}' - \hat{f}$ est laissée invariante par toute transformation du revêtement, et est donc *bornée* sur $\hat{M}$.

En quelques occasions nous considérerons également le *revêtement universel* $\tilde{p}: \tilde{M} \to M$ et une primitive $\tilde{f}: \tilde{M} \to \mathbb{R}$ de $\tilde{p}^* \omega$.

Pour une forme ω non singulière, le revêtement d'intégration $\hat{M}$ s'appelle revêtement des feuilles, car $\hat{M}$ est difféomorphe à $L \times \mathbb{R}$ feuilleté comme un produit (L est une feuille quelconque de ω).

I.6. Nous utiliserons souvent des *champs de vecteurs* X sur M^* tels que $\omega(X) \equiv 1$ (on construit de tels champs à l'aide d'une partition de l'unité). Le flot ϕ_t de X est partiellement défini sur M^*, et il préserve ω. En particulier, soit

$\tau>0$ tel que $\phi_\tau(x)$ soit défini pour tous les points d'un chemin σ contenu dans une feuille de ω; pour chaque $t\in[0,\tau]$, le chemin $\phi_t(\sigma)$ est contenu dans une feuille de ω. On a ainsi *relevé σ sur les feuilles voisines.*

Soit ω de Morse sans centre, et X comme ci-dessus. Supposons que près des singularités X est proportionnel à un champ à singularités hyperboliques. Alors pour tout $A\in\mathbb{R}$ l'ensemble des points où ϕ_A n'est pas défini est de codimension au moins 1. On en déduit que tout point de M^* appartient à l'image d'une immersion i: $\mathbb{R}\to M^*$ telle que $i^*\omega=dt$, et donc que tout point de $\hat{M}^*$ appartient à l'image d'un plongement $\hat{i}$: $\mathbb{R}\to\hat{M}^*$ tel que $\hat{f}\circ\hat{i}=\mathrm{id}_{\mathbb{R}}$.

I.7. L'espace des 1-formes fermées sur M sera muni de la topologie C^∞ (on pourrait également utiliser la topologie C^r, $r\geqq 1$); dans cet espace, les formes de Morse constituent un ouvert dense. Nous considérons ici certaines approximations d'une forme de Morse ω.

I.7.1 Fixons un voisinage U de $\mathrm{Sing}\,\omega$, et un champ de vecteurs Y sur M tel que $\omega(Y)>0$ sur M^*. Une fonction g: $M\to\mathbb{R}$, de classe C^∞, sera dite admissible si:

- $g\equiv 0$ sur $M-U$.
- $dg\equiv 0$ sur un voisinage de $\mathrm{Sing}\,\omega$ (dépendant de g).
- $dg(Y)>-\omega(Y)$ sur M^*.

Si g est admissible, alors $\omega_g=\omega+dg$ est une forme de Morse cohomologue à ω, qui a exactement les mêmes singularités que ω.

On peut en fait montrer que toute forme cohomologue à ω et assez proche de ω est isotope à une forme ω_g, et que deux formes ω_g et ω_h sont isotopes rel. $\mathrm{Sing}\,\omega$ ssi $g-h$ est constante sur $\mathrm{Sing}\,\omega$.

Nous n'aurons pas besoin de ces résultats, mais nous utiliserons la remarque suivante: soit V un voisinage de la fonction nulle dans la topologie C^∞; *l'ensemble des valeurs sur* $\mathrm{Sing}\,\omega$ *des fonctions admissibles contenues dans V est un voisinage de* 0 *dans* $\mathbb{R}^{\mathrm{Sing}\,\omega}$.

I.7.2. Géométriquement, le passage de ω à ω_g revient, pour chaque singularité s telle que $g(s)\neq 0$, à changer la feuille qui porte s. Si l'indice de s est compris entre 2 et $n-2$, tout se passe comme si s avait été déplacée de $g(s)$ transversalement au feuilletage (voir une représentation très schématique sur la fig. 2). Si s est une singularité conique, l'opération a eu pour effet de remplacer des feuilles à une nappe par des feuilles à deux nappes, ou vice-versa (voir fig. 3).

Grâce à cette interprétation on voit facilement que, *si L est une feuille* (non compacte) *de ω telle que $L\cup\mathrm{Sing}\,\omega$ est compact, alors on peut approcher ω par des formes cohomologues ayant des feuilles compactes proches de $\bar{L}$.*

I.7.3. Soit s une singularité conique telle qu'il existe un lacet de connexion γ (cf. I.4). Si ω' est cohomologue à ω et proche de ω, elle aura une singularité conique s' proche de s, et nous affirmons qu'*il existe pour s' un lacet de connexion γ' proche de γ.*

En effet on peut, en partant d'un bout singulier de s', relever le chemin γ sur une feuille de ω'. Lorsqu'on termine de parcourir γ, l'égalité $\int_\gamma\omega'=0$ entraîne

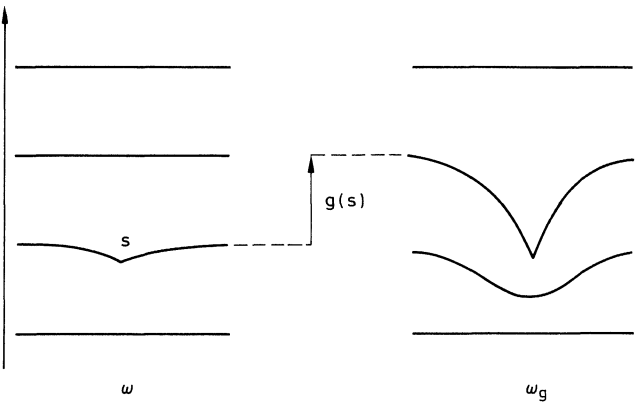


Fig. 2. Indice de s entre 2 et $n-2$, $g(s)>0$

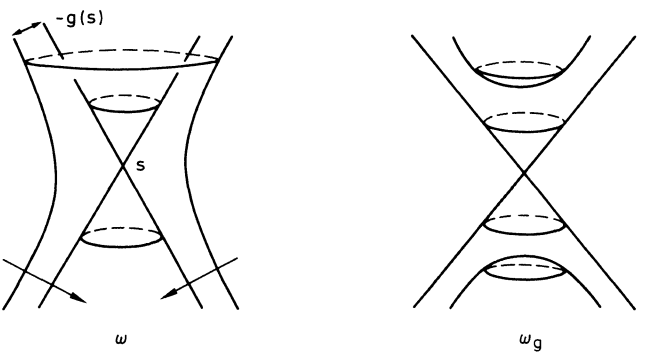


Fig. 3. s d'indice $n-1$, $g(s)<0$ (coupure de feuilles)

que ce chemin relevé revient près de s' *sur l'autre bout singulier.* Ceci montre l'existence de γ'.

Notons que ce raisonnement s'applique même si ω' n'est pas cohomologue à ω, à condition que $\int_\gamma \omega' = 0$; ce sera en particulier le cas si γ est homotope à 0.

I.8. Sauf mention explicite du contraire, nous désignerons par M une *variété fermée orientable de dimension* $n \geqq 3$, et par ω une *forme de Morse sans centre* sur M. Dans la partie VI nous supposerons simplement que Sing ω est de codimension $\geqq 3$.

II. Définition des formes faiblement complètes et complètes

Soient M et ω comme dans I.8.

Proposition II.1. *Les propriétés suivantes sont équivalentes:*

FC1. *L'espace des feuilles M^*/ω s'identifie à $\mathbb{R}/P(\omega)$*: deux points $x, y \in M^*$ sont sur la même feuille de ω si et seulement si $\psi(x)=\psi(y)$, où $\psi: M^* \to \mathbb{R}/P(\omega)$ est définie par $\psi(x)=\int_{x_0}^{x} \omega$ (x_0 étant un point base).

FC2. *Si $\theta \subset M^*$ est un chemin d'intégrale nulle ($\int_\theta \omega = 0$), ses extrémités appartiennent à la même feuille.*

FC3. *Pour toute singularité s conique* (i.e. d'indice 1 ou $n-1$), *les deux bouts singuliers issus de s sont situés sur la même feuille* (cf. I.4).

FC4. *Pour une (donc toute) primitive $\hat{f}$ de $\hat{p}^*\omega$ sur le revêtement d'intégration $\hat{M}$, tous les niveaux $\hat{f}^{-1}(c) \cap \hat{M}^*$ sont connexes.*

FC5. *Il existe une courbe fermée transverse C coupant toute feuille, telle que deux points de C soient sur la même feuille si et seulement si ils bordent un arc $J \subset C$ avec $|J| \in P(\omega)$.*

Proposition II.2. *Les propriétés suivantes sont équivalentes:*

C1. *Tout chemin $\theta \subset M^*$ d'intégrale nulle est homotope* (à extrémités fixes) *à un chemin contenu dans une feuille.*

C2. *Pour toute singularité conique s, les deux bouts singuliers issus de s sont situés sur la même feuille, et il existe un lacet de connexion γ homotope à 0 dans M* (cf. I.4).

C3. *Pour une (donc toute) primitive $\tilde{f}$ de $\tilde{p}^*\omega$ sur le revêtement universel $\tilde{M}$, tous les niveaux $\tilde{f}^{-1}(c) \cap \tilde{M}^*$ sont connexes.*

Définition. Une forme ω vérifiant les propriétés FC1 à FC5 (resp. C1 à C3) sera dite *faiblement complète* (resp. *complète*). Une *classe de cohomologie* $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ est (faiblement) complète si toute forme de Morse sans centre représentant Ω est (faiblement) complète.

Nous donnerons dans la partie IV d'autres caractérisations des formes faiblement complètes (resp. complètes).

Faisons quelques commentaires avant de montrer ces propositions. Tout d'abord une forme non singulière, ou plus généralement sans singularité conique, est complète.

D'autre part on voit facilement que, pour une forme ω faiblement complète de rang 1, toutes les feuilles régulières (qui sont compactes) sont homologues entre elles, et donc Poincaré-duales d'un multiple de $[\omega]$.

Si ω est faiblement complète et $r(\omega) \geqq 2$, alors le feuilletage défini par ω est *minimal* et *uniquement ergodique:* cela résulte de FC5 et des propriétés correspondantes des rotations irrationnelles du cercle.

De I.7.3 on déduit que toute forme assez proche d'une forme complète est complète; toute forme cohomologue à une forme faiblement complète, et assez proche de celle-ci, est faiblement complète.

Notre notion de «faiblement complète» coïncide avec celle de «quasi-complète» introduite par Imanishi dans [Im3]. Par contre notre notion de «complète» ne coïncide pas avec celle de [Im3]; elle se rapproche plutôt de celle considérée par Thurston pour les $\mathscr{H}$-feuilletages [Th2, §4.8]: pour une

forme complète, l'application développante $\tilde{f}$: $\tilde{M} \to \mathbb{R}$ est aussi proche que possible d'une fibration.

Démonstration des propositions II.1 et II.2

On vérifie facilement les équivalences FC1$\Leftrightarrow$FC2$\Leftrightarrow$FC4$\Leftrightarrow$FC5 et C1$\Leftrightarrow$C3, ainsi que les implications FC2$\Rightarrow$FC3 et C1$\Rightarrow$C2. Les réciproques FC3$\Rightarrow$FC2 et C2$\Rightarrow$C1 vont être montrées simultanément, dans un cadre un peu plus général qui nous sera utile dans la partie IV.

Supposons que ω vérifie FC3, et choisissons pour chaque singularité conique s_i un lacet de connexion γ_i. Soit $\mathcal{N}$ le plus petit sous-groupe distingué de $\pi_1 M$ contenant tous les γ_i.

Lemme II.3. *Soit θ un chemin d'intégrale nulle joignant deux points x, y de M^*. Alors θ est homotope à un chemin formé d'un lacet en x contenu dans $\mathcal{N}$, suivi d'un chemin dans une feuille (joignant x à y).*

Ce lemme achèvera de démontrer les propositions II.1 et II.2: si ω vérifie C2, nous pouvons en effet choisir les γ_i de façon que $\mathcal{N}=\{1\}$.

Démonstration du lemme II.3

Soit X un champ de vecteurs sur M^* tel que $\omega(X)\equiv 1$ et que près des singularités X soit proportionnel à un champ à singularité hyperbolique. On note ϕ_t le flot (partiellement défini) de X, cf. I.6. Nous commençons par le cas simple où ω n'a pas de singularité conique.

Soit λ: $[0,1]\to M^*$ une paramétrisation lisse de θ, et j: $[0,1]\to\mathbb{R}$ la fonction telle que $dj=\lambda^*\omega$ et $j(0)=j(1)=0$. Soit $A>\|j\|_\infty$. Puisqu'il n'y a pas de singularité conique (ni de centre), l'ensemble des points de M^* où ϕ_A (resp. ϕ_{-A}) n'est pas défini est de codimension $\geqq 2$.

En déplaçant légèrement θ (sans que ses extrémités quittent leurs feuilles respectives), nous pouvons donc supposer que $\phi_\tau(x)$ est défini pour tout $x\in\theta$ et tout $\tau\in[-A,A]$. Posons alors $\lambda_s(t)=\phi_{-sj(t)}(\lambda(t))$, pour $0\leqq s,t\leqq 1$. Cela définit une homotopie entre le chemin $\lambda_0=\lambda$ (qui représente θ) et un chemin λ_1 contenu dans une feuille, CQFD.

La démonstration dans le cas général utilise une technique classique (cf. [Ro] théorème 5, ou [FLP] proposition p. 81). On déforme θ en un chemin anguleux composé d'arcs I_k contenus dans des feuilles et d'arcs J_k contenus dans des orbites de X (voir Fig. 4), et on essaie de simplifier θ en relevant sur les feuilles voisines (cf. I.6) les I_k où θ fait demi-tour; si θ ne fait jamais demi-tour, il est contenu dans une feuille puisque $\int_\theta \omega = 0$.

Cette opération de relèvement se bloque lorsqu'on arrive à une singularité conique s_i. On remplace alors θ par un chemin formé d'un lacet θ_1 en x librement homotope à γ_i ou γ_i^{-1} (donc dans $\mathcal{N}$) suivi d'un nouveau chemin θ_2 (voir Fig. 5), et on continue en essayant de simplifier θ_2.

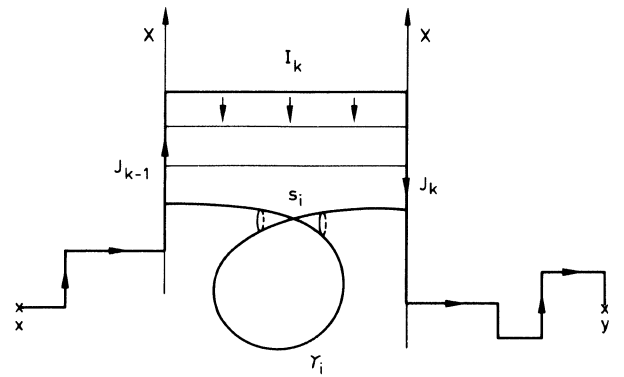


Fig. 4

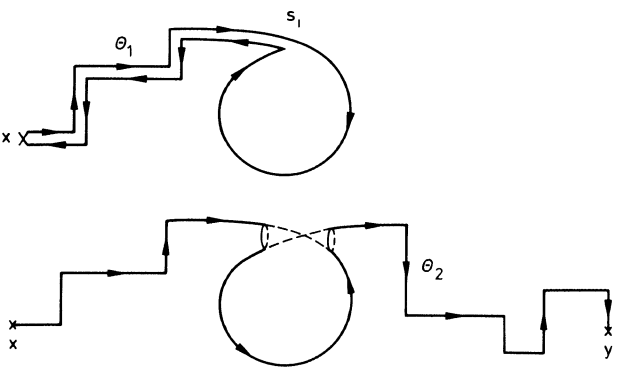


Fig. 5

Le point important est que le processus se termine après un nombre fini de ces remplacements. En effet il existe un $\varepsilon>0$ (indépendant de θ) tel que $\phi_t(x)$ soit défini pour tout $t\in[-\varepsilon, +\varepsilon]$ et tout $x\in\bigcup_i(\gamma_i-\{s_i\})$.

Si q est le nombre de points de I_k pour lesquels ϕ_ε ou $\phi_{-\varepsilon}$ n'est pas défini, on aura réussi après au plus q remplacements soit à supprimer un coin de I_k soit à diminuer la somme des $|J_k|$ d'au moins 2ε. Nous laissons les détails au lecteur. □

Remarque. Si ω est complète, cette technique permet de montrer que tout chemin $\theta\subset M^*$ est homotope soit à un chemin contenu dans une feuille soit à un chemin transverse (selon que $\int_\theta\omega$ est nul ou non).

III. Formes et classes de cohomologie faiblement complètes

Grâce à une construction de [AL], nous avons:

Théorème III.1. *Toute classe de cohomologie non nulle $\Omega\in H^1(M,\mathbb{R})-\{0\}$ est représentable par une forme (de Morse sans centre) faiblement complète.*

Démonstration

Nous ne donnons qu'une brève esquisse, car c'est essentiellement la section 1.2 de [AL]. Soit $B \subset M$ un bouquet de cercles tel que l'inclusion induise un isomorphisme $H^1(M, \mathbb{R}) \to H^1(B, \mathbb{R})$, et soit Y un voisinage régulier de B. On construit sur Y, dans la classe de cohomologie correspondant à Ω, une forme de Morse vérifiant FC5 par rapport à une courbe transverse $C \subset Y$ (voir [AL]).

On étend ensuite cette forme en une forme de Morse ω_1 sur M (pouvant posséder des centres), et on considère l'union U des feuilles de ω_1 qui rencontrent C. Si $U = M - \mathrm{Sing}\,\omega_1$, on a fini. Sinon, soit L une feuille de la frontière de U. Alors $L \cup \mathrm{Sing}\,\omega_1$ est compact et sépare M en deux composantes; de plus ω_1 est exacte dans la composante $\mathscr{A}$ qui ne contient pas C. On modifie ω_1 dans un voisinage de $\mathscr{A}$ de façon à faire disparaître tous les centres (il y en a au moins un), et après un nombre fini d'opérations on arrive à la forme sans centre cherchée. □

Remarque. En fait toute forme de Morse sans centre ω_0 peut être jointe à une forme faiblement complète ω_1 par un chemin de formes cohomologues, de Morse, ayant toutes les mêmes singularités. Ce chemin est du type $\omega_t = \omega_0 + t \cdot dg$, où $g: M \to \mathbb{R}$ est une fonction convenablement choisie, nulle aux singularités de ω_0 d'indice différent de $n-1$, et prenant des valeurs très grandes aux singularités d'indice $n-1$.

Corollaire III.2. *Soit M^n une variété fermée orientable, avec $n \geqq 3$. Tout homomorphisme surjectif $\pi: \pi_1 M \twoheadrightarrow \mathbb{Z}$ se représente par une application p de M dans un cercle, à singularités de Morse, telle que toutes les images réciproques $p^{-1}(c)$ soient connexes.*

Remarque. L'hypothèse $n \geqq 3$ est en fait inutile.

Démonstration

Soit ω une forme faiblement complète telle que $[\omega] = \pi$. L'application $p: M \to \mathbb{R}/\mathbb{Z}$ définie par $p(x) = \int_{x_0}^{x} \omega$ induit π; d'après FC1 elle est à fibres connexes. □

Nous allons maintenant étudier les formes qui ne sont pas faiblement complètes, et caractériser les classes de cohomologie faiblement complètes (c'est à dire dans lesquelles toute forme de Morse sans centre est faiblement complète). Nous énonçons d'abord les résultats sans les démonstrations, en commençant par les formes minimales.

Théorème III.3. *Toute forme ω minimale non faiblement complète peut être approchée dans sa classe de cohomologie par des formes dont toutes les feuilles régulières sont compactes, et aussi par des formes faiblement complètes* (donc minimales).

Remarques. 1) Rappelons que ω est de Morse sans centre, et que l'espace des formes fermées sur M est muni de la topologie C^∞ (ou C^r, $r \geqq 1$, cf. I.7).

2) Si au contraire ω est minimale et faiblement complète, il en est de même pour les formes cohomologues à ω et proches de ω (cf. partie II).

3) Ce théorème montre que, dans le théorème 2 de [AL], on peut imposer à ω_1 et ω_3 d'être des perturbations de ω_2.

Corollaire 1. *Soit ω une forme minimale telle que l'homomorphisme $[\omega]$: $\pi_1 M \to \mathbb{R}$ ne puisse pas se factoriser par un groupe libre. Alors ω est faiblement complète.*

Corollaire 2. *Si p_M: $\pi_1 M \to H_1(M, \mathbb{Z})/\text{torsion}$ ne peut pas se factoriser par un groupe libre, alors dans presque toute classe de cohomologie les formes minimales sont faiblement complètes* (en fait, *toute forme minimale totalement irrationnelle est faiblement complète*).

Remarque. Les corollaires 1 et 2 ci-dessus sont plus forts que les résultats de [AL], qui montraient seulement l'unique ergodicité.

Corollaire 3. *Soit ω une forme de rang $\geqq 2$. Les conditions suivantes sont équivalentes:*

a) *ω n'est pas faiblement complète.*

b) *ω est approchable dans sa classe de cohomologie par des formes possédant des feuilles compactes.*

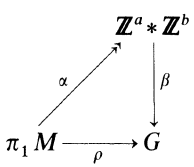
Soit $\Omega \in H^1(M, \mathbb{R})$ de rang $\geqq 2$, et $F_m(\Omega)$ l'espace des formes de Morse sans centre représentant Ω. Dans $F_m(\Omega)$, considérons les deux ouverts disjoints $FC(\Omega)$, constitué des formes faiblement complètes, et $C(\Omega)$, constitué des formes possédant des feuilles compactes.

D'après le corollaire 3, l'union de ces ouverts est dense (et même $\overline{C(\Omega)} = F_m(\Omega) - FC(\Omega)$):

Corollaire 4. *Soit $\Omega \in H^1(M, \mathbb{R})$ de rang $\geqq 2$. Génériquement, une forme de Morse sans centre représentant Ω ou bien est faiblement complète ou bien possède des feuilles compactes.* □

D'après le théorème III.1, on a toujours $FC(\Omega) \neq \emptyset$. Nous allons maintenant déterminer pour quelles classes Ω on a $C(\Omega) = \emptyset$, c'est à dire $FC(\Omega) = F_m(\Omega)$.

Nous dirons qu'un homomorphisme de groupes ρ: $\pi_1 M \to G$ *se factorise de façon non triviale à travers un* $\mathbb{Z}^a * \mathbb{Z}^b$ s'il existe un diagramme commutatif



avec α surjective et $a, b \geqq 1$.

Théorème III.4. *Soit $\Omega \in H^1(M, \mathbb{R}) - \{0\}$. Les conditions suivantes sont équivalentes:*

1) Ω *n'est pas faiblement complète.*

2) Ω: $\pi_1 M \to \mathbb{R}$ *se factorise de façon non triviale à travers un* $\mathbb{Z}^a * \mathbb{Z}^b$.

3) Ω *est représentable par une forme de Morse sans centre possédant une feuille compacte qui sépare* M.

Corollaire 1. *S'il n'existe pas d'homomorphisme surjectif de* $\pi_1 M$ *sur* $\mathbb{Z} * \mathbb{Z}$ *(par exemple si* $\pi_1 M$ *est résoluble), toute forme de Morse sans centre sur* M *est faiblement complète. La réciproque est également vraie.*

Corollaire 2. *Si* p_M: $\pi_1 M \to H_1(M, \mathbb{Z})/\text{torsion}$ *ne se factorise pas de façon non triviale par un* $\mathbb{Z}^a * \mathbb{Z}^b$, *presque toute classe* $\Omega \in H^1(M, \mathbb{R})$ *est faiblement complète.*

Si au contraire il y a factorisation de p_M, *aucune classe n'est faiblement complète.*

Corollaire 3. *Si* $\operatorname{Ker} \Omega_1 \subset \operatorname{Ker} \Omega_2$ *et si* Ω_2 *est faiblement complète, alors* Ω_1 *l'est aussi.*

Corollaire 4. *L'ensemble* $V(M) \subset H^1(M, \mathbb{R}) - \{0\}$ *des classes faiblement complètes ne dépend que de* $\pi_1 M$: *si* ρ: $\pi_1 M_1 \to \pi_1 M_2$ *est un isomorphisme, alors l'isomorphisme* ρ^*: $H^1(M_2, \mathbb{R}) \to H^1(M_1, \mathbb{R})$ *envoie* $V(M_2)$ *sur* $V(M_1)$.

Remarques. 1. Si ω possède une feuille compacte séparant M, il en est de même pour toute forme cohomologue à ω et proche de ω; on ne peut donc pas approcher ω par des formes cohomologues faiblement complètes. Il existe d'ailleurs des formes ω telles qu'aucune forme proche de ω, cohomologue ou non à ω, ne soit faiblement complète.

2. À une forme de Morse sans centre ω on peut associer un entier $d(\omega) \geqq 0$, qui est nul si et seulement si ω est faiblement complète; pour une forme «générique», $d(\omega)$ est simplement le nombre de singularités d'indice 1 (resp. $n-1$) pour lesquelles il n'existe pas de lacet de connexion. On peut montrer qu'il existe toujours un épimorphisme de $\pi_1 M$ sur le groupe libre à $d(\omega)+1$ générateurs (ce qui généralise le corollaire 1 ci-dessus), et que pour une forme minimale on a soit $d(\omega)=0$ soit $d(\omega) \geqq r(\omega)-1$.

Nous allons maintenant montrer successivement les corollaires, puis les théorèmes III.3 et III.4.

Démonstration des corollaires

Pour montrer le corollaire 1 du théorème III.3, soit ω minimale non faiblement complète. D'après le théorème il existe une forme cohomologue à ω dont toutes les feuilles régulières sont compactes. On en déduit facilement ([AL], lemme 2.2) que $[\omega]$ se factorise par un groupe libre.

Pour le corollaire 2, soit ω minimale totalement irrationnelle (cf. I.1). Si p_M ne se factorise pas par un groupe libre, alors $[\omega]$ non plus, et ω est faiblement complète par le corollaire 1. On en déduit le corollaire 2 car presque toute classe de cohomologie est totalement irrationnelle.

Pour montrer $a \Rightarrow b$ dans le corollaire 3, considérons une forme ω qui n'est pas faiblement complète. Si ω est minimale, on applique le théorème; si ω

possède des feuilles compactes, il n'y a rien à démontrer; sinon, il existe une feuille L avec $L \cup \mathrm{Sing}\,\omega$ compact (cf. I.3), et on applique I.7.2.

Si au contraire ω est faiblement complète et de rang $\geqq 2$ (donc minimale), nous avons déjà remarqué que les formes cohomologues à ω et proches de ω sont minimales.

Enfin les corollaires du théorème III.4 se déduisent immédiatement de l'équivalence $1 \Leftrightarrow 2$ du théorème. □

Démonstration du théorème III.3

Soit ω une forme minimale. Pour chaque singularité s_i d'indice $n-1$ de ω, fixons une petite boule B_i autour de s_i. Nous pouvons approcher ω par des formes cohomologues ω_g, comme dans I.7.1, où $g: M \to \mathbb{R}$ est nulle en dehors des B_i et prend en chaque s_i la même valeur *négative* $-\tau_g < 0$. Nous allons prouver la première partie du théorème en montrant que, si ω n'est pas faiblement complète, ces formes ω_g sont à feuilles régulières compactes.

Soit C une courbe disjointe des B_i, transverse à ω. Nous l'identifions à un cercle $\mathbb{R}/a\mathbb{Z}$ grâce à la mesure «de Lebesgue» induite par ω. Cette courbe est également transverse à ω_g, et deux points de C qui sont sur la même feuille de ω_g sont aussi sur la même feuille de ω: puisque les $g(s_i)$ sont négatifs, on passe de ω à ω_g en «coupant» des feuilles dans les B_i (cf. I.7.2).

On a en fait plus: *si deux points x et y de C sont sur la même feuille de ω_g, alors $x+t$ et $y+t$ sont sur la même feuille de ω pour tout $t \in [0, \tau_g]$*. En effet les singularités coniques constituent le seul obstacle au relèvement sur les feuilles voisines d'un chemin contenu dans une feuille (cf. démonstration du lemme II.3).

Remarquons enfin que nous pouvons supposer que C rencontre toute feuille de ω_g: il suffit pour cela d'imposer à C de passer au moins deux fois près de chaque s_i, de façon qu'elle rencontre près de s_i chacun des bouts singuliers, et de choisir les B_i assez petits.

Considérons maintenant les pseudogroupes d'holonomie $\Gamma \supset \Gamma'$ induits sur C par ω et ω_g respectivement. Ce sont des *pseudogroupes de rotations*, en ce sens que tout élément de Γ (resp. Γ') est, sur chaque composante de son domaine de définition, la restriction d'une rotation de C.

L'orbite d'un point $x \in C$ par Γ ou Γ', notée $\Gamma(x)$ ou $\Gamma'(x)$, est égale à l'intersection avec C de la feuille de ω ou ω_g passant par x. Comme ω est minimale, Γ *agit minimalement*, c'est à dire que toute orbite $\Gamma(x)$ est dense.

Grâce aux caractérisations FC3 et FC5 des formes faiblement complètes (voir prop. II.1), on voit facilement que ω *est faiblement complète si et seulement si Γ est un groupe de rotations*, c'est à dire si $y \in \Gamma(x) \Rightarrow R\,y \in \Gamma(R\,x)$ pour toute rotation R de C; une caractérisation équivalente est qu'il existe un sous-groupe G du groupe des rotations de C, tel qu'un homéomorphisme $\gamma: U \to V$ entre ouverts de C soit dans Γ ssi, pour toute composante U' de U, $\gamma|_{U'}$ est la restriction d'un élément de G.

Nous voulons montrer que, si ω n'est pas faiblement complète, les feuilles régulières de ω_g sont compactes. Supposons le contraire. Comme C rencontre

toute feuille de ω_g, elle rencontre une composante minimale W de ω_g, et toute orbite de Γ' dans $W \cap C$ est localement dense. Il ne nous reste donc qu'à montrer le lemme suivant:

Lemme III.5. *Soit Γ un pseudogroupe de rotations d'un cercle C, et $\Gamma' \subset \Gamma$ un sous-pseudogroupe. On suppose qu'il existe $\tau > 0$ tel que, si x et y sont dans la même orbite de Γ', alors $x+t$ et $y+t$ sont dans la même orbite de Γ pour tout $t \in [0, \tau]$. Si Γ agit minimalement et si Γ' possède une orbite localement dense* (en particulier si Γ' agit minimalement), *alors Γ est un groupe de rotations.*

Remarque. En particulier, un pseudogroupe qui contient une rotation irrationnelle définie sur tout le cercle est un groupe de rotations.

Démonstration

Supposons d'abord que Γ' possède une orbite dense et que, si x et y sont dans la même orbite de Γ', alors $x+t$ et $y+t$ sont dans la même orbite de Γ pour *tout* $t \in \mathbb{R}$. Il existe alors dans le groupe des rotations de C un sous-groupe dense G_0 tel que $x \in C$ et $R_0 \in G_0 \Rightarrow R_0 x \in \Gamma(x)$.

Soit γ un élément de Γ, défini sur un intervalle ouvert $I \subset C$, et appelons R la rotation dont γ est la restriction. Étant donné $y \in C$, soit $R_0 \in G_0$ avec $R_0 y \in I$. On a alors $R_0 y \in \Gamma(y)$, d'où $RR_0 y \in \Gamma(y)$ et $Ry = R_0^{-1} RR_0 y \in \Gamma(y)$, ce qui montre que Γ est un groupe de rotations.

Passons maintenant au cas général. Soit $J \subset C$ un intervalle dans lequel une orbite $\Gamma'(p)$ est dense, et soient a, b deux points de $J \cap \Gamma'(p)$ dont la distance est $< \tau$. On vérifie que Γ et Γ' induisent sur le cercle $\bar{C} = [a, b]/a \sim b$ des pseudogroupes $\bar{\Gamma}$ et $\bar{\Gamma}'$.

Le choix de a et b entraîne de plus que $\bar{\Gamma}$ et $\bar{\Gamma}'$ vérifient sur $\bar{C}$ les hypothèses du cas particulier ci-dessus, donc que $\bar{\Gamma}$ est un groupe de rotations. De la minimalité de Γ on déduit que Γ est aussi un groupe de rotations. □

Pour la deuxième partie du théorème III.3, approchons ω par des formes ω_g, où maintenant g prend en chaque s_i la même valeur *positive* τ_g. On voit que ω_g est faiblement complète en appliquant de nouveau le lemme III.5, mais cette fois-ci Γ' est le pseudogroupe d'holonomie de ω et Γ celui de ω_g. □

Démonstration du théorème III.4

Soit $\Omega \in H^1(M, \mathbb{R}) - \{0\}$, et $L^{n-1} \subset M^n$ une sous-variété compacte, connexe, orientable, sans bord, de codimension 1. Nous dirons que L est *Ω-exacte* si Ω induit sur L la classe nulle. Par exemple, une feuille compacte d'une forme représentant Ω est une sous-variété Ω-exacte.

Une sous-variété Ω-exacte L est *triviale* si ou bien L sépare M et Ω induit la classe nulle dans une des deux composantes de $M - L$ (par exemple si L est une sphère bordant une boule), ou bien $r(\Omega) = 1$ et L est Poincaré-duale d'un multiple non nul de Ω (i.e. $\Omega(u)$ est proportionnel au nombre d'intersection algébrique de $u \in H_1(M, \mathbb{Z})$ avec L).

Notons que, s'il existe une sous-variété Ω-exacte non triviale L, il en existe aussi une qui sépare M: si L ne sépare pas, on choisit un chemin θ d'intégrale non nulle coupant L en exactement un point (transversalement), et on remplace L par le bord d'un voisinage régulier de $L \cup \theta$.

Nous pouvons maintenant énoncer une quatrième condition équivalente à celles du théorème III.4:

4) *Il existe une sous-variété Ω-exacte non triviale.*

La démonstration du théorème va consister à prouver successivement $1 \Rightarrow 4 \Rightarrow 2 \Rightarrow 4 \Rightarrow 3$ (bien sûr $3 \Rightarrow 1$ est trivial).

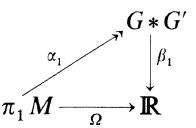
Démonstration de $1 \Rightarrow 4$

Supposons $r(\Omega) \geqq 2$ (nous laissons au lecteur le cas facile où $r(\Omega) = 1$). Si Ω n'est pas faiblement complète, alors d'après le corollaire 1 du théorème III.3 on peut la représenter par une forme de Morse sans centre ω possédant une feuille compacte L.

Cette feuille définit une sous-variété Ω-exacte, qui ne peut pas être triviale: si L séparait M et si ω était exacte dans une composante de $M - L$, alors ω devrait posséder un centre dans cette composante. □

Démonstration de $4 \Rightarrow 2$

Soit L une sous-variété Ω-exacte non triviale. D'après ce qui précède nous pouvons supposer que L sépare M. Le groupe fondamental de l'espace M/L obtenu en écrasant L en un point est de façon naturelle un produit libre $G * G'$, et on a une factorisation

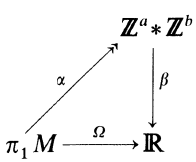


Puisque L n'est pas triviale, $\beta_1(G)$ et $\beta_1(G')$ ne sont ni l'un ni l'autre réduits à $\{0\}$, et on factorise β_1 par

$$(G/G \cap \operatorname{Ker} \beta_1) * (G'/G' \cap \operatorname{Ker} \beta_1) \simeq \mathbb{Z}^a * \mathbb{Z}^b. \quad \square$$

Démonstration de $2 \Rightarrow 4$

Lemme III.6. *S'il existe une factorisation*



(α surjective, $a, b \geqq 1$), *alors il en existe aussi une pour laquelle β est injective sur chacun des facteurs $\mathbb{Z}^a$ et $\mathbb{Z}^b$.*

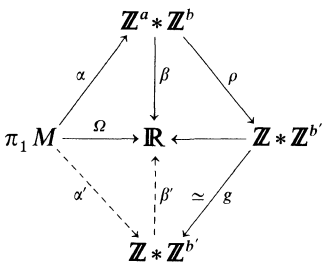
Démonstration

Si ni $\beta(\mathbb{Z}^a)$ ni $\beta(\mathbb{Z}^b)$ n'est réduit à $\{0\}$, on factorise simplement par

$$(\mathbb{Z}^a/\mathbb{Z}^a \cap \operatorname{Ker}\beta) * (\mathbb{Z}^b/\mathbb{Z}^b \cap \operatorname{Ker}\beta).$$

Sinon, nous pouvons supposer $\beta(\mathbb{Z}^a)=\{0\}$ et $\beta(\mathbb{Z}^b)$ isomorphe à $\mathbb{Z}^{b'}$, $b' \geqq 1$. Soit $\rho_1 : \mathbb{Z}^a \to \mathbb{Z}$ un épimorphisme. On commence par factoriser β à travers l'homomorphisme surjectif $\rho : \mathbb{Z}^a * \mathbb{Z}^b \to \mathbb{Z} * (\mathbb{Z}^b/\mathbb{Z}^b \cap \operatorname{Ker}\beta) \simeq \mathbb{Z} * \mathbb{Z}^{b'}$ obtenu à partir de ρ_1 et de la projection canonique.

On remarque ensuite qu'il existe un automorphisme g de $\mathbb{Z} * \mathbb{Z}^{b'}$ qui induit l'identité sur $\mathbb{Z}^{b'}$ et envoie un générateur z du facteur $\mathbb{Z}$ sur zu, où $u \in \mathbb{Z}^{b'}$ est non nul. On en déduit la factorisation cherchée, avec $\alpha' = g \circ \rho \circ \alpha$ (voir diagramme).



□

Pour montrer $2 \Rightarrow 4$, nous partons donc d'une factorisation de Ω par $\mathbb{Z}^a * \mathbb{Z}^b$, avec β injective sur chacun des facteurs. Nous commençons par le cas où $r(\Omega) \geqq 2$.

Soit H l'espace obtenu à partir d'un segment $J=[0, 1]$ et de tores T^a et T^b en identifiant une extrémité de J avec un point de T^a et l'autre avec un point de T^b. C'est un $K(\mathbb{Z}^a * \mathbb{Z}^b, 1)$, et nous pouvons représenter α par une application $p : M \to H$.

Soit $c \in \mathring{J}$ une valeur régulière. Alors $p^{-1}(c)$ est une collection finie non vide de sous-variétés Ω-exactes, et nous choisissons p et c de manière à en minimiser le nombre. Nous allons conclure en montrant que les composantes de $p^{-1}(c)$ ne sont pas toutes triviales.

Supposons au contraire qu'elles le soient. Il existe alors une composante $\mathscr{A}$ de $M - p^{-1}(c)$, sur laquelle Ω est exacte, et dont l'adhérence ne contient qu'une seule composante T de $p^{-1}(c)$. Pour tout lacet v contenu dans $\mathscr{A}$, l'image $p(v)$ représente un élément de $\operatorname{Ker}\beta \subset \pi_1 H$; de plus $p(v)$ est librement homotope à un lacet contenu dans T^a ou T^b.

Puisque β est injective sur chacun des facteurs $\mathbb{Z}^a$ et $\mathbb{Z}^b$, on en déduit que $p(v)$ est homotope à 0 dans H. Ceci permet de modifier p au voisinage de $\bar{\mathscr{A}}$, sans changer l'application induite sur le π_1, de façon à retirer de $p^{-1}(c)$ la composante T. On contredit ainsi le choix de p et c.

Considérons maintenant le cas où $r(\Omega)=1$. Nous avons alors $a=b=1$, et nous représentons α par une application p de M dans un bouquet de deux cercles orientés $C_1 \vee C_2$. Soit $c \in C_1$ une valeur régulière; $p^{-1}(c)$ est une collection non vide de sous-variétés Ω-exactes, et nous affirmons qu'elles ne sont pas toutes triviales.

Supposons le contraire. Il existe alors $t \in \mathbb{R}$ tel que, pour tout lacet v dans M, le nombre de passages de $p(v)$ en c soit égal à $t\Omega(v)$; le nombre de passages est bien sûr compté algébriquement, en comparant le sens de passage de $p(v)$ en c et l'orientation de C_1. On obtient la contradiction cherchée en considérant un lacet v_2 tel que $p(v_2)$ soit librement homotope à C_2 (ce qui donne $t=0$), et un lacet v_1 tel que $p(v_1)$ soit homotope à C_1 (ce qui donne $t \neq 0$). □

Démonstration de 4 ⇒ 3

Si Ω vérifie 4, il existe dans M une sous-variété Ω-exacte non triviale L qui sépare M. Nous cherchons à représenter Ω par une forme de Morse sans centre dont L soit une feuille. En considérant séparément l'adhérence de chaque composante de $M-L$, nous sommes ramenés au lemme suivant.

Lemme III.7. *Soit ω' une 1-forme fermée sur une variété compacte orientable P^n dont le bord est non vide et connexe ($n \geqq 3$). On suppose que ω' n'est pas exacte, mais que la forme induite sur δP l'est. Alors ω' est cohomologue à une forme de Morse sans centre dont δP est une feuille sortante (resp. rentrante).*

Par feuille sortante (resp. rentrante), nous voulons dire que δP est une feuille le long de laquelle l'orientation transverse du feuilletage pointe vers l'extérieur de P (resp. vers l'intérieur).

Démonstration

On raisonne comme dans la démonstration du théorème III.1, et nous nous bornons à signaler les différences.

Le bouquet de cercles B est choisi de façon qu'il y ait un isomorphisme $H^1(P, \delta P; \mathbb{R}) \to H^1(B, \mathbb{R})$. Après avoir construit la forme au voisinage de B, on choisit un chemin orienté θ allant d'un point de C (courbe transverse coupant toute feuille de la forme définie sur Y) à un point de δP; dans Y ce chemin doit être transverse au feuilletage, dans le sens positif (resp. négatif) si l'on veut δP sortant (resp. rentrant).

L'extension ω_1 est construite de façon que δP soit une feuille et que θ soit transverse, et on élimine les centres en considérant l'union U des feuilles rencontrant $C \cup \theta$. □

IV. Structure transverse et π_1 des feuilles

Nous établissons ici des critères pour qu'une forme soit faiblement complète ou complète (propositions IV.1 et IV.6). Étant donné une forme ω, soit $\mathscr{L}(\omega)$ le

plus petit sous-groupe distingué de $\pi_1 M$ contenant, pour toute feuille L, l'image de l'homomorphisme $\pi_1 L \to \pi_1 M$ induit par l'inclusion. On a bien sûr $\mathscr{L}(\omega) \subset \mathrm{Ker}[\omega]$.

Proposition IV.1. *Une forme de Morse sans centre ω est faiblement complète si et seulement si $\mathscr{L}(\omega) = \mathrm{Ker}[\omega]$.*

Remarque. Si ω est faiblement complète, le quotient $\pi_1 M/\mathscr{L}(\omega)$ est donc abélien libre. Si au contraire ω n'est pas faiblement complète, on peut montrer que ce quotient est un produit libre non trivial de groupes abéliens libres.

Corollaire IV.2. *Soit ω une forme de Morse sans centre, de rang $r(\omega) \geqq 2$. Les conditions suivantes sont équivalentes:*

1) *ω est faiblement complète.*

2) *Il existe une feuille L telle que $\mathrm{Ker}[\omega]$ soit le plus petit sous-groupe distingué de $\pi_1 M$ contenant l'image de $\pi_1 L$.*

3) *Pour toute feuille L, $\mathrm{Ker}[\omega]$ est le plus petit sous-groupe distingué de $\pi_1 M$ contenant l'image de $\pi_1 L$.*

Corollaire IV.3. *Soit ω comme dans le corollaire IV.2. Si ω possède une feuille simplement connexe, alors ou bien $\pi_1 M$ est abélien libre de rang $r(\omega)$ ou bien il existe un épimorphisme de $\pi_1 M$ sur $\mathbb{Z} * \mathbb{Z}$.*

Nous démontrons d'abord les corollaires.

Démonstration du corollaire IV.2

Les implications $3 \Rightarrow 2 \Rightarrow 1$ sont vraies même si $r(\omega) = 1$: $3 \Rightarrow 2$ est trivial, et $2 \Rightarrow 1$ résulte immédiatement de la proposition. Pour $1 \Rightarrow 3$, on remarque qu'une forme faiblement complète de rang $\geqq 2$ est minimale, et donc que pour toute feuille L le sous-groupe distingué de $\pi_1 M$ engendré par l'image de $\pi_1 L$ est égal à $\mathscr{L}(\omega)$. □

Remarque. Il existe des formes faiblement complètes de rang 1 pour lesquelles 3 est faux.

Démonstration du corollaire IV.3

Il résulte immédiatement du corollaire IV.2 ($1 \Rightarrow 3$) et du corollaire 1 du théorème III.3. □

Démonstration de la proposition IV.1

La partie «seulement si» résulte immédiatement du lemme II.3, compte tenu de l'inclusion évidente $\mathscr{N} \subset \mathscr{L}(\omega)$.

Pour la réciproque, considérons une forme ω, et une primitive $\hat{f}: \hat{M} \to \mathbb{R}$ de $\hat{p}^*\omega$ (cf. I.5). Rappelons que nous notons $\hat{M}_{-\infty,c}$ l'ouvert $\hat{f}^{-1}(]-\infty, c[)$; de même $\hat{M}_{c,+\infty} = \hat{f}^{-1}(]c, +\infty[)$. Dans le lemme suivant nous ne supposons rien sur les singularités de ω.

Lemme IV.4. *Soit ω une forme non exacte. Pour tout $c \in \mathbb{R}$, chacun des ouverts $\hat{M}_{-\infty,c}$ et $\hat{M}_{c,+\infty}$ possède exactement une composante sur laquelle $\hat{f}$ n'est pas bornée.*

Démonstration

Soit ω' une forme de Morse sans centre, faiblement complète, cohomologue à ω (cf. théorème III.1). Le lemme est vrai pour ω', puisque ω' vérifie FC4.

Soit $\hat{f}': \hat{M} \to \mathbb{R}$ une primitive de $\hat{p}^*\omega'$. Nous savons que $\hat{f}' - \hat{f}$ est bornée (cf. I.5), et on en déduit facilement que le lemme est également vrai pour ω. □

Remarque. On peut montrer ce lemme sans faire appel au théorème III.1, par exemple en utilisant une forme harmonique cohomologue à ω.

Du lemme IV.4 on déduit:

Corollaire IV.5. *Soit ω une forme de Morse sans centre. Pour tout $c \in \mathbb{R}$, les ouverts $\hat{M}_{-\infty,c}$ et $\hat{M}_{c,+\infty}$ sont connexes.*

Démonstration

Cela résulte du lemme IV.4 et du fait que, pour une forme de Morse sans centre, tout point de $\hat{M}^*$ appartient à l'image d'un plongement $\hat{i}: \mathbb{R} \to \hat{M}^*$ tel que $\hat{f} \circ \hat{i} = \mathrm{id}_{\mathbb{R}}$ (voir I.6). □

Terminons maintenant de montrer la proposition IV.1. Soit ω une forme de Morse sans centre, non faiblement complète. Il existe un niveau $\hat{f}^{-1}(c) \cap \hat{M}^*$ possédant au moins deux composantes T_1 et T_2. Le corollaire IV.5 nous permet de construire dans $\hat{M}^*$ un lacet rencontrant chaque T_i en exatement un point (et transversalement). La projection de ce lacet dans M représente un élément de $\mathrm{Ker}[\omega]$ qui ne peut pas être dans $\mathscr{L}(\omega)$. □

Proposition IV.6. *Soit ω une forme de Morse sans centre. Les conditions suivantes sont équivalentes:*

1) *ω est complète.*
2) *Il existe une feuille L telle que la suite $\pi_1 L \to \pi_1 M \xrightarrow{[\omega]} \mathbb{R}$ soit exacte.*
3) *Pour toute feuille L, la suite $\pi_1 L \to \pi_1 M \xrightarrow{[\omega]} \mathbb{R}$ est exacte.*
4) *Pour un (donc tout) $c \in \mathbb{R}$, les ouverts connexes $\hat{M}_{-\infty,c}$ et $\hat{M}_{c,+\infty}$ engendrent $\pi_1 \hat{M}$.*

Comme d'habitude, $\hat{f}: \hat{M} \to \mathbb{R}$ est une primitive de $\hat{p}^*\omega$. Nous disons qu'une sous-variété $T \subset \hat{M}$ *engendre* $\pi_1 \hat{M}$ si l'inclusion induit une surjection

$\pi_1 T \to \pi_1 \hat{M}$. Si $\hat{M}_{-\infty,c_0}$ engendre $\pi_1 \hat{M}$, il en est de même pour tout $\hat{M}_{-\infty,c}$ puisque $\hat{M}_{-\infty,c}$ contient l'image de $\hat{M}_{-\infty,c_0}$ par une transformation convenable du revêtement.

Démonstration de la proposition IV.6

Nous montrons successivement $1 \Rightarrow 2 \Rightarrow 4 \Rightarrow 3 \Rightarrow 1$. L'implication $1 \Rightarrow 2$ est claire puisqu'une forme complète vérifie C1. Si d'autre part L est une feuille telle que l'image de $\pi_1 L$ dans $\pi_1 M$ soit égale à $\mathrm{Ker}[\omega]$, n'importe quelle composante de $\hat{p}^{-1}(L)$ engendre $\pi_1 \hat{M}$ et est contenue dans un $\hat{f}^{-1}(c)$; ceci montre $2 \Rightarrow 4$.

Pour $4 \Rightarrow 3$, nous considérons une composante T d'un $\hat{f}^{-1}(c) \cap \hat{M}^*$, et nous devons montrer qu'elle engendre $\pi_1 \hat{M}$. L'argument est classique (cf. [Hem] pp. 105–106), et nous ne faisons que l'esquisser.

Prenons un point base $x \in T$, et un élément $e \in \pi_1(\hat{M}, x)$. L'hypothèse 4 nous permet de représenter e par deux lacets $e^+ \subset \hat{f}^{-1}([c, +\infty[)$ et $e^- \subset \hat{f}^{-1}(]-\infty, c])$. On considère une homotopie entre e^+ et e^- dans $\hat{M}^*$, et on conclut en étudiant la trace sur T de cette homotopie.

Supposons enfin que ω vérifie 3. Si un chemin $\theta \subset M^*$ est d'intégrale nulle, ses extrémités peuvent être jointes par un chemin θ_0 contenu dans une feuille L (car d'après la proposition IV.1 la forme ω est faiblement complète). D'après 3, le lacet formé par θ et θ_0 est homotope à un lacet de L, ce qui signifie que θ est homotope (à extrémités fixes) à un chemin contenu dans L. □

Remarque. On peut prouver $3 \Rightarrow 1$ directement, sans passer par la proposition IV.1: on montre successivement que toute feuille est coupée par une transversale fermée, puis que ω vérifie C2. Cette démonstration marche encore si on suppose seulement que pour toute feuille L l'image de $\pi_1 L$ contient le sous-groupe des commutateurs de $\pi_1 M$ (en fait ω est complète dès qu'il existe une feuille L telle que l'image de $\pi_1 L$ contienne le sous-groupe des commutateurs).

V. Classes de cohomologie complètes

L'équivalence $1 \Leftrightarrow 4$ de la proposition IV.6 va nous permettre de montrer que la complétude ou la non-complétude d'une forme ne dépend que de sa classe de cohomologie.

Théorème V.1. *Soient ω et ω' deux formes de Morse sans centre, cohomologues. Alors ω' est complète si et seulement si ω est complète.*

Démonstration

Supposons ω complète, et montrons que ω' l'est aussi. Soient $\hat{f}$ et $\hat{f}'$ des primitives de $\hat{p}^* \omega$ et $\hat{p}^* \omega'$ respectivement, et $A > \|\hat{f} - \hat{f}'\|_\infty$. Pour $c \in \mathbb{R}$, on a $\hat{f}^{-1}(]-\infty, c-A[) \subset \hat{f}'^{-1}(]-\infty, c[)$ et $\hat{f}^{-1}(]c+A, +\infty[) \subset \hat{f}'^{-1}(]c, +\infty[)$, d'où le résultat. □

Remarques. D. Henč a montré dans [Hen] qu'une forme ω de Morse sans feuille compacte, cohomologue à une forme minimale non singulière, est minimale et uniquement ergodique. Sa démonstration donne en fait la complétude de ω, et peut se généraliser en une démonstration du théorème V.1.

On peut imaginer d'autres propriétés des formes fermées ne dépendant que de leur classe de cohomologie. Par exemple on peut demander que les $\hat{M}_{-\infty,c}$ et les $\hat{M}_{c,+\infty}$ engendrent $H_i\hat{M}$ ($i \geqq 2$), ou représentent toute classe d'homotopie libre de lacets. On peut également considérer séparément les $\hat{M}_{-\infty,c}$ et les $\hat{M}_{c,+\infty}$, ou bien remplacer $\hat{M}$ par d'autres revêtements (par exemple celui correspondant au sous-groupe des commutateurs de $\pi_1 M$, ou au noyau de p_M).

Théorème V.2. *Soit ω une forme de Morse sans centre. Si* $\mathrm{Ker}[\omega]$ *est de type fini, alors ω est complète. La réciproque est vraie si ω est de rang* 1.

Démonstration

Elle utilise un petit argument bien connu (cf. [Hem] p. 105). Si $\mathrm{Ker}[\omega]$ est de type fini, il existe un compact $K \subset \hat{M}$ qui engendre $\pi_1\hat{M}$. Soit $\hat{f}$: $\hat{M} \to \mathbb{R}$ une primitive de $\hat{p}^*\omega$, et $c \in \mathbb{R}$. En appliquant à K des transformations convenables du revêtement, on peut l'envoyer dans $\hat{M}_{-\infty,c}$ ou dans $\hat{M}_{c,+\infty}$. D'après la proposition IV.6, la forme ω est complète.

Réciproquement, soit ω complète de rang 1, et L une feuille régulière de ω. D'après la proposition IV.6 ($1 \Rightarrow 3$), $\mathrm{Ker}[\omega]$ est l'image de $\pi_1 L$ dans $\pi_1 M$; il est de type fini car L est compacte. □

Remarque. La réciproque peut être fausse si $r(\omega) \geqq 2$. Il existe sur certaines variétés de dimension 3 des formes ω et ω' totalement irrationnelles (donc telles que $\mathrm{Ker}[\omega] = \mathrm{Ker}[\omega'] = \mathrm{Ker}\, p_M$), avec ω complète mais pas ω' (cf. [Th 1], et la discussion ci-dessous sur la dimension trois).

Du théorème V.2 on déduit immédiatement:

Corollaire V.3. *Si le noyau de* p_M: $\pi_1 M \to H_1(M, \mathbb{Z})/\text{torsion}$ *est de type fini, toute forme sur M est complète.* □

Corollaire V.4. *Si $\pi_1 M$ est abélien, ou nilpotent, ou contient un sous-groupe polycyclique d'indice fini, toute forme sur M est complète.* □

Remarques. Comme nous l'a fait remarquer E. Ghys, le corollaire V.4 ne s'étend pas au cas résoluble. Par contre nous pouvons montrer que toute forme sur M est complète si $\pi_1 M$ ne contient pas de sous-monoïde libre non abélien (en particulier si $\pi_1 M$ est à croissance sous-exponentielle). En fait, d'après le théorème B1 de [BNS], les variétés sur lesquelles toute forme est complète sont exactement celles avec $\mathrm{Ker}\, p_M$ de type fini (ou de façon équivalente $[\pi_1 M, \pi_1 M]$ de type fini).

Le cas où $\pi_1 M$ est abélien a été traité par Imanishi [Im2].

Considérons maintenant dans $H^1(M, \mathbb{R}) - \{0\}$ *l'ensemble $U(M)$ formé des classes de cohomologie complètes.*

Théorème V.5. *$U(M)$ ne dépend que de $\pi_1 M$: si ρ est un isomorphisme de $\pi_1 M_1$ sur $\pi_1 M_2$, alors ρ^*: $H^1(M_2, \mathbb{R}) \to H^1(M_1, \mathbb{R})$ envoie $U(M_2)$ sur $U(M_1)$. En particulier, $U(M)$ est laissé invariant par l'action sur $H^1(M, \mathbb{R})$ du groupe des automorphismes de $\pi_1 M$.*

Démonstration

Considérons deux classes $\Omega_2 \in U(M_2)$ et $\Omega_1 = \rho^* \Omega_2 \in H^1(M_1, \mathbb{R})$. Nous devons montrer $\Omega_1 \in U(M_1)$. Pour $i=1$ ou 2, représentons Ω_i par une forme de Morse sans centre ω_i sur M_i. Considérons comme d'habitude le revêtement $\hat{p}_i$: $\hat{M}_i \to M_i$ et une primitive $\hat{f}_i$ de $\hat{p}_i^* \omega_i$.

Soit $K_i \subset M_i$ le 2-squelette d'une triangulation de M_i, et $\hat{K}_i = \hat{p}_i^{-1}(K_i)$ celui de la triangulation induite sur $\hat{M}_i$. Puisque les inclusions $K_i \subset M_i$ induisent des isomorphismes sur le π_1, l'isomorphisme ρ^{-1}: $\pi_1 M_2 \to \pi_1 M_1$ induit un isomorphisme de $\pi_1 K_2$ sur $\pi_1 K_1$, que nous pouvons représenter par une application g: $K_2 \to K_1$. Choisissons un relevé $\hat{g}$: $\hat{K}_2 \to \hat{K}_1$, et considérons $\hat{f}_1 \circ \hat{g}$: $\hat{K}_2 \to \mathbb{R}$.

La différence $\hat{f}_1 \circ \hat{g} - \hat{f}_2$: $\hat{K}_2 \to \mathbb{R}$ est laissée invariante par les transformations du revêtement $\hat{p}_2$. Elle est donc bornée, et nous choisissons $A > \|\hat{f}_1 \circ \hat{g} - \hat{f}_2\|_\infty$. D'autre part, puisqu'il n'y a qu'un nombre fini de simplexes dans M_2, il existe un nombre B tel que $|\hat{f}_2(x) - \hat{f}_2(y)| < B$ dès que x et y sont des points de $\hat{M}_2$ appartenant au même simplexe.

Soit maintenant $c \in \mathbb{R}$, et fixons un point base $x_2 \in \hat{K}_2$ tel que $\hat{f}_2(x_2) < c - A - B$. Soit $e \in \pi_1(\hat{M}_2, x_2)$. Puisque ω_2 est complète, nous pouvons représenter e par un lacet v_e (basé en x_2) contenu dans $\hat{f}_2^{-1}]-\infty, c-A-B[$. Soit $\hat{M}_2(e)$ l'union des simplexes qui rencontrent v_e. Alors v_e est homotope dans $\hat{M}_2(e)$ à un lacet $v'_e \subset \hat{K}_2$. D'après le choix de A et B, on a $v'_e \subset \hat{f}_2^{-1}]-\infty, c-A[\subset (\hat{f}_1 \circ \hat{g})^{-1}]-\infty, c[$.

Le lacet $\hat{g}(v'_e)$, basé en $\hat{g}(x_2)$, représente $\hat{g}_*(e)$ et est contenu dans $\hat{f}_1^{-1}]-\infty, c[$. Puisque e était quelconque, nous voyons que $\hat{f}_1^{-1}]-\infty, c[$ engendre $\pi_1 \hat{M}_1$. On montre de même que $\hat{f}_1^{-1}]c, +\infty[$ engendre $\pi_1 \hat{M}_1$, et donc que ω_1 est complète. □

Remarque. Plus généralement, si ρ: $\pi_1 M_1 \to \pi_1 M_2$ est un épimorphisme et si $\Omega_1 = \rho^* \Omega_2$ est complète, alors Ω_2 est complète.

Théorème V.6. *L'ensemble $U(M) \subset H^1(M, \mathbb{R}) - \{0\}$ des classes complètes est un cône ouvert. Une classe rationnelle $\Omega \in H^1(M, \mathbb{Q}) - \{0\}$ est dans $U(M)$ si et seulement si le noyau de Ω: $\pi_1 M \to \mathbb{Q}$ est de type fini.*

Démonstration

Il est clair que $U(M)$ est un cône (i.e. $\Omega \in U(M)$ et $t \in \mathbb{R}^* \Rightarrow t\Omega \in U(M)$). D'autre part la deuxième assertion est simplement une reformulation du théorème V.2.

Il reste à voir que $U(M)$ est ouvert. La raison essentielle en est que la propriété C2 (voir proposition II.2) est préservée par perturbation de ω (cf.

I.7.3). Pour le lecteur que cet argument ne satisferait pas, nous procédons comme suit.

Soit $\Omega_1, \ldots, \Omega_q$ une base de $H^1(M, \mathbb{Z})$, et $H_i = p_i^{-1}(c_i)$, où p_i: $M \to S^1$ représente Ω_i et $c_i \in S^1$ est une valeur régulière. On peut représenter chaque Ω_i par une forme ω_i qui est nulle en dehors d'un voisinage arbitrairement petit U_i de l'hypersurface H_i.

Soit ω une forme complète de Morse sans centre. Supposons d'abord $n \geqq 5$. Dans ce cas les lacets de connexion γ_j (homotopes à 0) associés aux singularités coniques s_j de ω bordent des disques plongés disjoints D_j, et on peut par isotopie supposer que les $\overline{U_i}$ sont disjoints des D_j et de $\operatorname{Sing} \omega$.

Si les réels a_i sont assez proches de 0, la forme $\omega + \sum_{i=1}^{q} a_i \omega_i$ a les mêmes singularités que ω, et est donc complète (elle vérifie C2).

Pour le cas où $n = 3$ ou 4, nous utilisons le théorème V.5, qui identifie $U(M)$ à $U(M \times S^2)$. □

Du théorème V.6 on déduit immédiatement un résultat de W. Neumann [Ne]:

Corollaire V.7. *Soit G un groupe de présentation finie. Dans* $\operatorname{Hom}(G, \mathbb{Q}) - \{0\}$, *les homomorphismes dont le noyau est de type fini forment un ouvert.* □

Remarques. La topologie sur $\operatorname{Hom}(G, \mathbb{Q})$ est bien sûr celle de la convergence uniforme sur un système fini (quelconque) de générateurs.

En utilisant des formes non singulières, D. Fried et R. Lee ont montré qu'avoir un noyau de présentation finie est également une propriété ouverte [FrL].

Corollaire V.8. *Il existe sur M une forme complète si et seulement si il existe un épimorphisme de $\pi_1 M$ dans $\mathbb{Z}$ dont le noyau est de type fini.* □

Remarque. L'existence sur M d'une forme complète équivaut aussi à l'existence dans $\pi_1 M$ d'un sous-groupe distingué de type fini, d'indice infini, et contenant $\operatorname{Ker} p_M$ (ou contenant simplement le sous-groupe des commutateurs). Elle implique que $\operatorname{Ker} p_M$ est l'intersection d'une famille finie de sous-groupes distingués de type fini, et que presque toute classe de cohomologie est faiblement complète.

Appelons $U'(M)$ le cône ouvert des $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ tels que toute classe rationnelle assez proche de Ω soit complète. D'après le théorème V.6 on a $U(M) \subset U'(M)$, et ces deux cônes ouverts contiennent exactement les mêmes points rationnels.

Malheureusement l'inclusion $U(M) \subset U'(M)$ peut être stricte: on déduit de [BNS] (section 8) des exemples où $U'(M) = H^1(M, \mathbb{R}) - \{0\}$, alors que $H^1(M, \mathbb{R}) - U(M)$ se compose de deux droites ne contenant aucun point rationnel (à part 0). Il existe ainsi des variétés sur lesquelles certaines formes ne sont pas complètes bien que toute forme de rang 1 le soit.

Nous allons voir cependant qu'on a toujours $U(M) = U'(M)$ si $\pi_1 M$ est le *groupe fondamental d'une variété de dimension* 3. Cela résultera du théorème de

fibration de Stallings [St], et de la description due à Thurston ([Th1], [Fr]) du *cône ouvert* $\mathscr{C}(N) \subset H^1(N^3, \mathbb{R}) - \{0\}$ *formé des classes représentables par une forme non singulière* sur une variété N^3 irréductible (en fait $\mathscr{C}(N) = U(N)$, voir ci-dessous). Bien sûr toutes les variétés considérées ici sont fermées et orientables.

D'une façon générale, le théorème V.2 nous permet de représenter tout épimorphisme π: $\pi_1 M^n \to \mathbb{Z}$ avec $\operatorname{Ker}\pi$ de type fini, par une application p: $M \to S^1$ telle que pour tout $c \in S^1$ l'image réciproque $p^{-1}(c)$ soit connexe et engendre $\operatorname{Ker}\pi$ (i.e. la suite $\pi_1(p^{-1}(c)) \to \pi_1 M \xrightarrow{\pi} \mathbb{Z}$ est exacte).

Malheureusement nous ne pouvons pas en général espérer que l'application $\pi_1(p^{-1}(c)) \to \pi_1 M$ soit injective, car par exemple $\operatorname{Ker}\pi$ n'a aucune raison d'être de présentation finie.

La situation est bien meilleure si M est une variété irréductible de dimension 3; dans ce cas le théorème de Stallings dit que π est représentable par une fibration, ou en d'autres termes que toute forme complète de rang 1 est cohomologue à une forme non singulière.

Nous allons en fait déduire de la théorie de Thurston que le même résultat est vrai si le rang est >1: *sur toute variété* N^3 *irréductible. on a* $\mathscr{C}(N) = U(N)$: *une forme est complète si et seulement si elle est cohomologue à une forme non singulière.*

Thurston définit pour toute variété N^3 irréductible une semi-norme sur $H^1(N, \mathbb{R})$ (pour simplifier la discussion, nous supposerons que c'est en fait une norme). La sphère unité de cette norme est un polyèdre à sommets rationnels, et Thurston montre que $\mathscr{C}(N)$ est une union de cônes sur certaines faces ouvertes X_i (de dimension maximale) de ce polyèdre.

On en déduit que $\mathscr{C}(N) = U(N)$. Bien sûr $\mathscr{C}(N) \subset U(N)$. Si d'autre part $\Omega \notin \mathscr{C}(N)$, la structure de $\mathscr{C}(N)$ impose l'existence arbitrairement près de Ω de classes rationnelles qui ne sont pas dans $\mathscr{C}(N)$, donc pas dans $U(N)$; ceci empêche Ω d'être dans $U(N)$.

On a également $U'(N) = U(N)$: ceci résulte de la structure polyédrale de $U(N) = \mathscr{C}(N)$.

Si d'autre part M^3 est une 3-variété quelconque, et si $\pi_1 M$ n'est pas isomorphe à un $\pi_1 N^3$, avec N irréductible, alors ou bien $U(M) = \emptyset$, ou bien $\pi_1 M \simeq \mathbb{Z}$ ([Hem], lemmes 10.1 et 11.2). On a donc toujours $U(M) = U'(M)$ si M est de dimension 3.

Exemple: formes complètes sur $M_1 \times M_2$

Considérons une variété produit $M^n = M_1 \times M_2$. Si ω_1 (resp. ω_2) est une forme de Morse sans centre sur M_1 (resp. M_2), alors $\delta_1^* \omega_1 + \delta_2^* \omega_2$ est une forme de Morse sur M dont toutes les singularités ont un indice compris entre 2 et $n-2$ (δ_i est bien sûr la projection de M sur M_i); c'est donc une forme complète.

Cette remarque montre le fait suivant, qui est d'ailleurs vrai si G_1 et G_2 sont seulement de type fini: *soient* G_1 *et* G_2 *deux groupes de présentation finie, et* π: $G_1 \times G_2 \to \mathbb{Z}$ *un homomorphisme; si ni* $\pi(G_1)$ *ni* $\pi(G_2)$ *n'est réduit à* $\{0\}$, *alors* $\operatorname{Ker}\pi$ *est de type fini.*

On peut en fait déterminer complètement $U(M_1 \times M_2)$ en fonction de $U(M_1)$ et $U(M_2)$: les formes non complètes sur $M_1 \times M_2$ sont exactement celles pour lesquelles $\omega | M_1$ est exacte et $\omega | M_2$ non complète, ou vice-versa. D'après le théorème V.5, cette description de $U(M)$ se généralise à toute variété dont le π_1 est un produit direct.

Caractérisations algébriques de $U(M)$

Nous avons vu que $U(M)$ ne dépend que de $G = \pi_1 M$. On peut en donner plusieurs définitions algébriques. Nous en esquissons trois ci-dessous.

a) Soit $G' \subset G$ le sous-groupe des commutateurs. Étant donné un homomorphisme Ω: $G \to \mathbb{R}$, soit $G_\Omega \subset G$ le sous-monoïde formé des g tels que $\Omega(g) \geqq 0$. On appelle Σ l'ensemble des homomorphismes non nuls Ω: $G \to \mathbb{R}$ possédant la propriété suivante: il existe un sous-groupe de type fini $H \subset G'$, et un sous-monoïde de type fini $J \subset G_\Omega$, tels que tout $g \in G'$ soit un produit d'éléments de la forme $j^{-1} h j$, avec $h \in H$ et $j \in J$.

On déduit alors de [BNS] (cf. [Si]) que $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ appartient à $U(M)$ si et seulement si Ω et $-\Omega$ appartiennent à Σ.

b) À une présentation de G avec p générateurs et q relations est associée une suite exacte de Λ-modules $\Lambda^q \to \Lambda^p \to \Lambda$, où $\Lambda = \mathbb{Z}[G]$. Étant donné Ω: $G \to \mathbb{R}$, on appelle Λ_Ω l'anneau des séries formelles $\sum_g n_g \cdot g$ ($g \in G$, $n_g \in \mathbb{Z}$) telles que, pour tout $c \in \mathbb{R}$, l'ensemble des g tels que $n_g \neq 0$ et $\Omega(g) > c$ soit fini.

On montre [Si] que $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ est dans $U(M)$ si et seulement si les deux suites $\Lambda_\Omega^q \to \Lambda_\Omega^p \to \Lambda_\Omega$ et $\Lambda_{-\Omega}^q \to \Lambda_{-\Omega}^p \to \Lambda_{-\Omega}$ obtenues par extension des scalaires sont exactes.

c) D'après [Me], $\Omega \in H^1(M, \mathbb{R}) - \{0\}$ n'est pas dans $U(M) \Leftrightarrow G$ a une infinité de bouts dans la direction Ω ou dans la direction $-\Omega$. De façon explicite, Ω n'est pas dans $U(M)$ ssi il existe une partie $Q \subset G$ telle que ni $\Omega(Q)$ ni $\Omega(G - Q)$ ne soit majoré (resp. minoré) dans $\mathbb{R}$, mais que $\Omega(Q \triangle Q g)$ soit majoré (resp. minoré) pour tout $g \in G$ ($\triangle$ est la différence symétrique).

Mentionnons pour finir deux résultats de [BNS]. Si $\pi_1 M$ admet une présentation par p générateurs et q relations, avec $q \leqq p - 2$, alors $U(M) = \emptyset$. Si par contre $\pi_1 M$ ne contient pas de sous-groupe isomorphe à $\mathbb{Z} * \mathbb{Z}$, et si $b_1 M \geqq 2$, alors $U(M) \neq \emptyset$.

VI. Généralisation

Le but de cette partie est d'étendre nos résultats aux formes de Morse pouvant posséder des centres, et plus généralement aux formes ω dont le lieu singulier est de codimension $\geqq 3$; de façon précise, nous supposons que $\operatorname{Sing} \omega$ *est contenu dans une réunion finie de sous-variétés compactes de codimension au moins 3.*

Notons qu'alors $\pi_1 M^*$ est isomorphe à $\pi_1 M$, et donc que $N(\omega)$ (union des feuilles L telles que $L \cup \operatorname{Sing} \omega$ ne soit pas compact) est *ouvert* ([Ha], théorème p. 386). Nous notons $N(\omega)/\omega$ l'espace des feuilles de $N(\omega)$.

Remarque. Certains de nos résultats sont probablement des cas particuliers de résultats valables pour les formes non singulières sur les variétés ouvertes (cf. [Im2], [Me]).

VIa. Formes faiblement complètes

Dans la partie III nous avons caractérisé algébriquement les classes de cohomologie dans lesquelles toute forme de Morse sans centre est faiblement complète. Nous allons maintenant étudier les formes plus générales contenues dans ces classes.

Théorème VI.1. *Soit ω une forme dont le lieu singulier est de codimension au moins 3. On suppose que ω est de rang au moins 2, et que $[\omega]: \pi_1 M \to \mathbb{R}$ ne se factorise pas de façon non triviale par un $\mathbb{Z}^a * \mathbb{Z}^b$. Alors l'ouvert $N(\omega)$ est non vide, connexe, et l'espace des feuilles $N(\omega)/\omega$ s'identifie à $\mathbb{R}/P(\omega)$* (au sens de la proposition II.1).

Remarque. Ce théorème peut être faux si $\operatorname{Sing}\omega$ est de codimension 2; il ne semble pas admettre de bonne généralisation au cas où $r(\omega) \leqq 1$.

Si ω est comme dans le théorème VI.1, le feuilletage induit sur $N(\omega)$ est minimal et uniquement ergodique. En particulier toute feuille de ω est fermée (dans M^*) ou localement dense, c'est à dire que ω n'a *pas de feuille exceptionnelle.* Nous obtenons donc:

Corollaire VI.2. *S'il n'y a pas d'homomorphisme surjectif de $\pi_1 M$ dans $\mathbb{Z} * \mathbb{Z}$, une forme sur M dont le lieu singulier est de codimension $\geqq 3$ n'a pas de feuille exceptionnelle.*

On ignore si une forme fermée sur une variété compacte peut avoir des feuilles exceptionnelles (cf. [Im1] et [Me]).

Faisons encore quelques remarques sur le théorème VI.1 avant de le démontrer.

Comme pour les formes de Morse sans centre, on peut exprimer la conclusion $N(\omega)/\omega \sim \mathbb{R}/P(\omega)$ du théorème sous d'autres formes (cf. proposition II.1): tout chemin d'intégrale nulle à extrémités dans $N(\omega)$ a ses extrémités sur la même feuille; dans $\hat{M}$, tous les niveaux $\hat{f}^{-1}(c) \cap \hat{p}^{-1}(N(\omega))$ sont connexes; il existe une courbe fermée transverse $C \subset N(\omega)$ coupant toute feuille de $N(\omega)$, telle que deux points quelconques de C dont la distance est une période soient sur la même feuille.

Les formes ω du théorème VI.1 méritent d'être appelées faiblement complètes; en effet elles possèdent les deux propriétés suivantes, que nous prenons comme *définition générale des formes faiblement complètes* (de rang $\geqq 2$):

- si L est une feuille avec $L \cup \operatorname{Sing}\omega$ compact, alors $\bar{L}$ sépare M en deux composantes, et ω est exacte dans l'une d'elles;
- $N(\omega)/\omega \sim \mathbb{R}/P(\omega)$.

Bien sûr cette définition coïncide avec celle de la partie II pour les formes de Morse sans centre. Intuitivement, une forme faiblement complète se compo-

se d'un *noyau* $N(\omega)$ dans lequel tout se passe comme pour une forme de Morse sans centre faiblement complète, et d'*appendices de feuilles fermées* greffés sur ce noyau.

Démonstration du théorème VI.1

Quitte à considérer $M \times S^2$ au lieu de M, nous pouvons supposer $n \geqq 5$. L'hypothèse sur $\mathrm{Sing}\,\omega$ nous permet de trianguler M de façon que le 2-squelette soit disjoint de $\mathrm{Sing}\,\omega$. Le bord d'un voisinage régulier de ce 2-squelette fournit alors une sous-variété $V^{n-1} \subset M^n$, disjointe de $\mathrm{Sing}\,\omega$, telle que l'inclusion induise un isomorphisme $\pi_1 V \simeq \pi_1 M$ (c'est ici que nous utilisons $n \geqq 5$).

En bougeant un peu V nous pouvons supposer que ω induit sur V une forme de Morse ω' (pouvant avoir des centres) dont les singularités sont sur des feuilles distinctes de ω.

L'hypothèse de non-factorisation faite sur $[\omega]$ entraîne qu'il n'existe pas dans V de sous-variété $[\omega']$-exacte non triviale (au sens de la partie III); on en déduit que $N(\omega')$ est non vide, connexe, et que toute période est l'intégrale de ω' sur un lacet contenu dans $N(\omega')$.

Soit $C \subset N(\omega')$ une courbe transverse, et X un champ de vecteurs sur $M - \mathrm{Sing}\,\omega$ tel que $\omega(X) \equiv 1$ et que C soit une orbite. Puisque V est compact, il existe $\tau > 0$ tel que le flot $\phi_t(x)$ de X soit défini pour tout $x \in V$ et $t \in [0, \tau]$.

Ceci nous permet d'appliquer le lemme III.5, en prenant pour Γ et Γ' les pseudogroupes d'holonomie induits sur C par ω et ω' respectivement; comme ω' est de Morse, nous savons que $N(\omega')$ est une composante minimale (cf. I.3) et donc que Γ' agit minimalement.

Comme toute période est représentée dans $N(\omega')$, on en déduit $W/\omega \sim \mathbb{R}/P(\omega)$, où W est le saturé de C pour ω. Il ne nous reste donc plus qu'à montrer $W = N(\omega)$, c'est à dire que *toute feuille de* $N(\omega)$ *coupe* C.

C'est certainement vrai si ω est de Morse: en effet dans ce cas nous savons que $N(\omega)$ est connexe et que toute feuille contenue dans $N(\omega)$ y est dense. Nous avons donc montré le théorème pour les formes de Morse.

Dans le cas général, supposons qu'il existe dans $N(\omega)$ une feuille L disjointe de C, et choisissons un relevé $\hat{L}$ de L dans $\hat{M}$. Dans le niveau $\hat{f}^{-1}(c) \cap M^*$ qui contient $\hat{L}$, il existe une composante $\hat{L}'$ (en fait unique) dont la projection L' rencontre C.

Puisque L et L' sont coupées par des transversales fermées, les feuilles $\hat{L}$ et $\hat{L}'$ adhèrent à des composantes de $\hat{M}_{-\infty, c}$ (resp. $\hat{M}_{c, +\infty}$) sur lesquelles $\hat{f}$ n'est pas bornée.

Compte tenu du lemme IV.4, on en déduit comme dans la démonstration de la proposition IV.1 que $\mathscr{L}(\omega)$ est strictement contenu dans $\mathrm{Ker}[\omega]$; puisque $\mathrm{codim}\,\mathrm{Sing}\,\omega \geqq 3$, deux lacets de $\hat{M}^*$ homotopes dans $\hat{M}$ ont le même nombre d'intersection algébrique avec $\hat{L}$ ou $\hat{L}'$.

Nous sommes ainsi ramenés à montrer le résultat suivant:

Lemme VI.3. *Si* ω *est comme dans le théorème VI.1, on a* $\mathscr{L}(\omega) = \mathrm{Ker}[\omega]$.

Démonstration

Reprenons V et ω' comme ci-dessus. Puisque $\mathscr{L}(\omega') \subset \mathscr{L}(\omega)$, il nous suffit de montrer le résultat pour ω'.

Les feuilles non compactes L de ω' telles que $L \cup \mathrm{Sing}\,\omega'$ soit compact sont en nombre fini. Leur union avec $\mathrm{Sing}\,\omega'$ est un compact K, et les composantes de $V-K$ autres que $N(\omega')$ sont feuilletées comme un produit (feuille compacte $\times$ intervalle ouvert).

Si L est une feuille contenue dans K, $\bar{L}$ sépare V en deux composantes et ω' est exacte dans l'une d'elles (car il n'y a pas de sous-variété $[\omega']$-exacte non triviale). On peut donc représenter tout élément de $\mathrm{Ker}[\omega']$ par un produit fini de lacets d'intégrale nulle v_i librement homotopes à des lacets $v_i' \subset V-K$.

Les v_i' qui ne sont pas dans $N(\omega')$ représentent évidemment des éléments de $\mathscr{L}(\omega')$. Il en est de même pour les v_i' contenus dans $N(\omega')$: en effet nous savons que $N(\omega')/\omega'$ s'identifie à $\mathbb{R}/P(\omega')$, car le théorème VI.1 est montré pour les formes de Morse; il nous suffit alors d'appliquer à v_i' les opérations décrites dans la démonstration du lemme II.3, en remarquant qu'elles ne font pas sortir de $N(\omega')$. Finalement on a bien $\mathrm{Ker}[\omega'] \subset \mathscr{L}(\omega')$. □

VIb. Formes complètes

Nous avons étudié dans la section V le cône ouvert $U(M) \subset H^1(M, \mathbb{R}) - \{0\}$ des classes de cohomologie complètes, et nous avons caractérisé géométriquement les formes de Morse sans centre dont la classe est complète (propositions II.2 et IV.6). Nous étendons maintenant ces caractérisations aux formes à singularités quelconques.

Théorème VI.4. *Soit ω une forme de rang $\geqq 2$ dont le lieu singulier est de codimension $\geqq 3$. Les conditions suivantes sont équivalentes:*

1) *La classe de cohomologie $[\omega] \in H^1(M, \mathbb{R})$ est complète.*

2) *Il existe $c \in \mathbb{R}$ tel que, pour chacun des ouverts $\hat{M}_{-\infty,c}$ et $\hat{M}_{c,+\infty}$, la composante* (unique) *sur laquelle $\hat{f}$ n'est pas bornée engendre $\pi_1 \hat{M}$* (c'est alors vrai $\forall c$).

3) *$N(\omega)$ n'est pas vide, et tout chemin d'intégrale nulle à extrémités dans $N(\omega)$ est homotope à un chemin contenu dans une feuille.*

4) *$N(\omega)$ n'est pas vide, et sur le revêtement universel $\tilde{M}$ on a $\tilde{f}^{-1}(c) \cap \tilde{p}^{-1}(N(\omega))$ connexe pour tout $c \in \mathbb{R}$.*

5) *Il existe une feuille $L \subset N(\omega)$ telle que la suite $\pi_1 L \to \pi_1 M \xrightarrow{[\omega]} \mathbb{R}$ soit exacte.*

6) *$N(\omega)$ n'est pas vide, et pour toute feuille $L \subset N(\omega)$ la suite $\pi_1 L \to \pi_1 M \xrightarrow{[\omega]} \mathbb{R}$ est exacte.*

Démonstration

On a clairement $3 \Leftrightarrow 4$, $3 \Rightarrow 6 \Rightarrow 5 \Rightarrow 2$, et $1 \Leftrightarrow 2$ (d'après la proposition IV.6). Pour conclure, supposons que ω vérifie 2. Nous savons, par le théorème VI.1,

que $N(\omega)$ est non vide, connexe, et que $N(\omega)/\omega \sim \mathbb{R}/P(\omega)$. Ceci permet de montrer successivement que ω vérifie 6 et 3, comme dans la démonstration de $4 \Rightarrow 3 \Rightarrow 1$ dans la proposition IV.6. □

Remarque. On déduit de $1 \Rightarrow 5$ que, si $[\omega]$ est complète, alors $N(\omega)$ engendre $\pi_1 M$.

Bibliographie

[AL] Arnoux, P., Levitt, G.: Sur l'unique ergodicité des 1-formes fermées singulières. Invent. Math. **84**, 141–156 (1986)

[BL] Blank, S., Laudenbach, F.: Isotopie de formes fermées en dimension trois. Invent. Math. **54**, 103–177 (1979)

[BNS] Bieri, R., Neumann, W., Strebel, R.: A geometric invariant for discrete groups. Preprint

[FLP] Fathi, A., Laudenbach, F., Poenaru, V.: Travaux de Thurston sur les surfaces. Astérisque **66–67** (1979), SMF Paris

[Fr] Fried, D.: Fibrations over S^1 with pseudo-Anosov monodromy. Exposé 14 de [FLP]

[FrL] Fried, D., Lee, R.: Realizing group automorphisms, in Group actions on manifolds. Contemp. Math. **36**, 427–432 (1985)

[Ha] Haefliger, A.: Variétés feuilletées. Ann. Sci. Norm. Super. Pisa, Cl. Sci., IV. Ser. **16**, 367–397 (1962)

[Hem] Hempel, J.: 3-manifolds. Ann. Math. Stud. **86** (1976) Princeton Univ. press

[Hen] Henč, D.: Ergodicity of foliations with singularities. Preprint IHES 1982

[Im1] Imanishi, H.: On codimension one foliations defined by closed one forms with singularities. J. Math. Kyoto Univ. **19**, 285–291 (1979)

[Im2] Imanishi, H.: Structure of codimension 1 foliations without holonomy on manifolds with abelian fundamental group. J. Math. Kyoto Univ. **19**, 481–495 (1979)

[Im3] Imanishi, H.: Denjoy-Siegel theory of codimension one foliations. Sûgaku **32**, 119–132 (1980) (en japonais). MR 82k:57017

[Le] Levitt, G.: Geometry and ergodicity of singular closed 1-forms. Proc. V Escola Geom. Dif., São Paulo 1984, 109–118

[Me] Meigniez, G.: Bouts d'un groupe dans une direction et feuilletages par 1-formes fermées (preprint)

[Mo] Moser, J.: On the volume elements on a manifold. Trans. Am. Math. Soc. **120**, 286–294 (1965)

[Ne] Neumann, W.: Normal subgroups with infinite cyclic quotient. Math. Sci. **4**, 143–148 (1979)

[QR] Que, N., Roussarie, R.: Sur l'isotopie des formes fermées en dimension 3. Invent. Math. **64**, 69–87 (1981)

[Ro] Rosenberg, H.: Foliations by planes. Topology **7**, 131–138 (1968)

[Si] Sikorav, J.C.: Thèse d'État, Orsay 1987

[St] Stallings, J.: On fibering certain 3-manifolds, Topology of 3-manifolds and related topics. Prentice Hall 1961, 95–100

[Th1] Thurston, W.: A norm for the homology of 3-manifolds. Mem. Am. Math. Soc. **339**, 99–130 (1986)

[Th2] Thurston, W.: The geometry and topology of 3-manifolds. Princeton University notes

[Ti] Tischler, D.: On fibering certain foliated manifolds over S^1. Topology **9**, 153–154 (1970)

Oblatum 3-II-1986 & 9-X-1986