SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Inventiones Mathematicae
Verlag: Springer
Jahr: 1983
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN356556735_0073
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN356556735_0073

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 73 1983

Editores

J. F. Adams, Cambridge **M. Berger,** Paris **A. Borel,** Princeton
J. Coates, Orsay **A. Connes,** Bures-sur-Yvette **L. Hörmander,** Lund
A. A. Kirillov, Moskwa **B. Mazur,** Cambridge, Mass. **R. Remmert,** Münster
J-P. Serre, Paris **T. A. Springer,** Utrecht **S.-T. Yau,** Princeton

Editores quondam

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) N. H. Kuiper (1966–1974) B. Malgrange (1966–1976)
Y. Manin (1966–1978) J.K. Moser (1974–1982) J-P. Serre (1967–1979)
J. Tits (1966–1975) E. Vesentini (1966–1973)



Springer-Verlag
Berlin · Heidelberg New York · Tokyo

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 the authors and by the responsible authorities – tacitly or explicitly – in the institutes where the work was carried out and that, if accepted, it will not be published elsewhere in the same form, in either the same or another language, without the consent of the copyright holders. By submitting a manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints photographic reproductions, microform, electronic database, video-disks, or any other reproductions of similar nature, and translations.
Photographic reproduction, microform, electronic data-base, video-disks, or any other reproduction of text, figures or tables from this journal is prohibited without permission obtained from the publisher.
The use of general descriptive names, trade names, trade marks, etc., in this publication, even if the former are not specifically identified, is not to be interpreted that such names are exempt from the relevant protective laws and regulations.
Special Regulations for the USA. Photocopies may be made for personal or internal use beyond that permitted by Section 107 or 108 of the U.S. Copyright Law, provided a fee is paid. This fee is $ 0.20 per page or a minimum of $ 1.00 if an article consists of less than five pages. Please pay this fee to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISBN 0020-9910, volume, and first and last page numbers of each article copied.
The copyright owner's consent does not extend to copying for general distribution for promotion for creating new works, or for resale. Specific written permission must be obtained from the publisher for such copying.
Other Regulations. Authors of this journal can benefit from library and photocopy fees collected by VG WORT if certain conditions are met. If an author lives in the Federal Republic of Germany or in West Berlin it is recommended that the Federal Republic of Germany or in West Berlin it is recommended that he contact Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, for detailed information.

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 auf 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 deutsche oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebende Autor 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.

Printers: Universitätsdruckerei H. Stürtz AG, Würzburg

© Springer-Verlag Berlin Heidelberg 1983

Index

Allard, W.K.: An a Priori Estimate for the Oscillation of the Normal to a Hypersurface Whose First and Second Variation With Respect to an Elliptic Integrand is Controlled 287

Barth, W., Peters, C.: Automorphisms of Enriques Surfaces 383

Benson, M.: Growth Series of Finite Extensions of $\mathbb{Z}^n$ are Rational . . . 251

Benveniste, X.: Sur l'anneau canonique de certaines variétés de dimension 3 157

Bombieri, E., Vaaler, J.: On Siegel's Lemma 11

Carrell, J.B., Goresky, R.M.: A Decomposition Theorem for the Integral Homology of a Variety 367

Chmutov, S.V.: The Monodromy Groups of Critical Points of Functions II 491

Conley, C.C., Zehnder, E.: The Birkhoff-Lewis Point Theorem and a Conjecture of V.I. Arnold 33

Coquet, J.: A Summation Formula Related to the Binary Digits 107

Dummit, D., Labute, J.: On a New Characterization of Demuskin Groups 413

Evertse, J.-H.: On the Representation of Integers by Binary Cubic Forms of Positive Discriminant 117

Faltings, G.: Arakelov's Theorem for Abelian Varieties 337

Faltings, G.: Endlichkeitssätze für abelsche Varietäten über Zahlkörpern 349

Futaki, A.: An Obstruction to the Existence of Einstein Kähler Metrics 437

Gau, Y.-N., Lipman, J.: Differential Invariance of Multiplicity on Analytic Varieties 165

Goresky, R.M., see Carrell, J.B. 367

Grosshans, F.D.: The Invariants of Unipotent Radicals of Parabolic Subgroups 1

Han, C.K.: Analyticity of CR Equivalences Between Some Real Hypersurfaces in $\mathbb{C}^n$ with Degenerate Levi Forms 51

Heitsch, J.L.: Flat Bundles and Residues for Foliations 271

Hsia, J.S., Hung, D.C.: Theta Series of Ternary and Quaternary Quadratic Forms 151

Hung, D.C., see Hsia, J.S. 151

Karoubi, M.: Homology of the Infinite Orthogonal and Symplectic Groups Over Algebraically Closed Fields. An Appendix to the Paper of A. Suslin 247

Koiso, N.: Einstein Metrics and Complex Structures 71
Labute, J., see Dummit, D. 413
Lee, R., Schwermer, J.: The Lefschetz Number of an Involution on the Space of Harmonic Cusp Forms of SL_3 189
Lipman, J., see Gau, Y.-N. 165
Meuser, D.: On the Poles of a Local Zeta Function for Curves 445
Moussu, R.: Classification C^∞ des èquations de pfaff complétement intégrables a singularité isolée 419
Müller, D.: Calderón-Zygmund Kernels Carried by Linear Subspaces of Homogeneous Nilpotent Lie Algebras 467
Peters, C., see Barth, W. 383
Ran, Z.: On Projective Varieties of Codimension 2 333
Schwermer, J., see Lee, R. 189
Siu, Y.-T.: Every K3 Surface is Kähler 139
Suslin, A.: On the *K*-Theory of Algebraically Closed Fields 241
Vaaler, J., see Bombieri, E. 11
Zehnder, E., see Conley, C.C. 33

Covered by Zentralblatt für Mathematik and Current Mathematical Publications

A20900

Inventiones mathematicae

1-3 TJ. 2A 25965

Vol. 73 Fasc. 1 1983

F.D. Grosshans	The Invariants of Unipotent Radicals of Parabolic Subgroups	1
E. Bombieri, J. Vaaler	On Siegel's Lemma	11
C.C. Conley, E. Zehnder	The Birkhoff-Lewis Fixed Point Theorem and a Conjecture of V.I. Arnold	33
C.K. Han	Analyticity of CR Equivalences Between Some Real Hypersurfaces in $\mathbb{C}^n$ with Degenerate Levi Forms	51
N. Koiso	Einstein Metrics and Complex Structures	71
J. Coquet	A Summation Formula Related to the Binary Digits	107
J.-H. Evertse	On the Representation of Integers by Binary Cubic Forms of Positive Discriminant	117
Y.-T. Siu	Every K3 Surface is Kähler	139
J.S. Hsia, D.C. Hung	Theta Series of Ternary and Quaternary Quadratic Forms	151
X. Benveniste	Sur l'anneau canonique de certaines variétés de dimension 3	157

Covered by Zentralblatt für Mathematik and Current Mathematical Publications



Springer-Verlag
Berlin Heidelberg New York Tokyo

222 Inventiones math. ISSN 0020-9910 INVMBH 73(1) 1–164 (1983)
July 1983

Niedersächsische Staats- u. Universitätsbibliothek Göttingen
28 JUL 1983

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

It is a fundamental condition that submitted manuscripts have not been published and will not be simultaneously submitted or published elsewhere. By submitting a manuscript, the authors agree that the copyright for their article is transferred to the publisher if and when the article is accepted for publication. The copyright covers the exclusive rights to reproduce and distribute the article, including reprints, photographic reproductions, microform or any other reproductions of similar nature, and translations.

Photographic reproduction, microform, or any other reproduction of text, figures, or tables from this journal is prohibited without permission obtained from the publisher.

The use of general descriptive names, trade names, trade marks, etc., in this publication, even if the former are not specifically identified, is not to be interpreted that such names are exempt from the relevant protective laws and regulations and may accordingly be used freely by anyone.

Special Regulations for the USA

Photocopies may be made for personal or internal use beyond that permitted by Section 107 or 108 of the U.S. Copyright Law, provided a fee is paid. This fee is $0.20 per page or a minimum of $1.00 if an article consists of less than five pages. Please pay this fee to the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA, stating the ISSN 0020-9910, volume, and first and last page numbers of each article copied.

The copyright owner's consent does not extend to copying for general distribution, for promotion, for creating new works, or for resale. Specific written permission must be obtained from the publisher for such copying.

Other Regulations

Authors of this journal can benefit from library and photocopy fees collected by VG WORT if certain conditions are met. If an author lives in the Federal Republic of Germany or in West Berlin it is recommended that he contact 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 Bänden vereinigt werden. Es ist das Ziel der Zeitschrift, Arbeiten innerhalb von 4 Monaten nach Annahme zu publizieren.

Urheberrecht

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 deutsche oder ständig in der Bundesrepublik Deutschland oder Berlin (West) lebende Autor 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

Reinhold Remmert
Berliner Straße 7
D-4540 Lengerich

Les manuscrits en deux exemplaires doivent être envoyés à

or to any of the editors
ou à un éditeur

Manuskripte in doppelter Anfertigung erbeten an

oder an einen der Herausgeber

Das Heft enthält eine Beilage des Springer-Verlages Berlin Heidelberg New York Tokyo

Invent. math. 73, 1-9 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

The Invariants of Unipotent Radicals of Parabolic Subgroups

F.D. Grosshans

Department of Mathematical Sciences, West Chester University of Pennsylvania, West Chester, PA 19380, USA

In this paper, we shall be concerned with obtaining positive results relating to Hilbert's fourteenth problem. In particular, our purpose, here, is to prove the following

Theorem. *Let k be an algebraically closed field. Let G be a reductive algebraic group over k and let U_1 be the unipotent radical of a parabolic subgroup of G. If X is any affine variety on which G acts (rationally), then*

$$k[X]^{U_1}=\{f\in k[X]:\ u\cdot f=f \text{ for all } u\in U_1\}$$

is a finitely generated k-algebra.

This theorem was proved for maximal unipotent subgroups in characteristic 0 in [2] and arbitrary characteristic in [1]. Some of the details of the latter proof will be used here so it is outlined in (2.1). The theorem, itself, was proved in characteristic 0 in [3] and in arbitrary characteristic for GL_n in [8].

A proof of the theorem in characteristic 0 is given in (2.2) since it points to a construction used later. This proof depends on the complete reducibility of reductive algebraic groups. For arbitrary characteristic, the argument requires the geometric properties of reductive groups found in [7; Theorem 3.5, p. 61], for example.

1. The Algebra H'

(1.1) **Definitions.** *Let k be an algebraically closed field and let G be an affine algebraic group over k. Let $k[G]$ denote the algebra of regular functions on G. The group G operates on itself via $g\cdot g_1=g_1g^{-1}$ and this gives rise to an action on $k[G]$ by $(g\cdot f)(g_1)=f(g_1g)$.*

For H a subgroup of G, we put

$$H'=k[G]^H=\{f\in k[G]:\ h\cdot f=f \text{ for all } h\in H\}$$

and for R a subset of $k[G]$, we put

$$R' = \{g \in G\colon g \cdot r = r \text{ for all } r \in R\}.$$

Then, H' is a k-subalgebra of $k[G]$ and R' is an algebraic subgroup of G.

Note 1. If $g \in G$ is in the normalizer of H, then $g \cdot H' = H'$.

The subgroups H of G which are of interest in invariant theory satisfy the condition $H = H''$ [1; Lemma 1, p. 242]. Such subgroups are said to be *observable* in G. The next theorem gives an important condition on H which is equivalent to $H = H''$.

Theorem [1; Theorem 2, p. 235 or 8; (§1), p. 359]. *Let G be an affine algebraic group over k and let H be a closed subgroup of G. The following conditions on H are equivalent:*

(i) $H = H''$;

(ii) *there is a rational representation of G on a finite-dimensional vector space V and a $v \in V$ so that*

$$H = G_v = \{g \in G\colon g \cdot v = v\}.$$

Note 2. If U is a unipotent subgroup of G, then (i) holds and, so, U is observable in G [1; Corollary, p. 234].

(1.2) **Theorem.** *Let G be a reductive algebraic group and let H be a closed subgroup of G. The following conditions on H are equivalent:*

(i) *H' is a finitely generated k-algebra;*

(ii) *if X is any affine variety on which G acts, then $k[X]^H$ is a finitely generated k-algebra.*

Proof. We shall denote the algebra of regular functions on a variety V by $k[V]$. Now, let us assume that condition (i) holds and prove that condition (ii) must hold. (The other implication is immediate if we take $X = G$.) The idea here is to consider the morphism $\phi\colon X \to G/H \times X$ given by $x \to (H, x)$ and show that its comorphism induces an isomorphism from $k[G/H \times X]^G$ to $k[X]^H$. This will solve the problem. Indeed, $k[G/H \times X] = k[G/H] \otimes k[X]$ (this result is usually stated only for affine varieties, but is true in general [10; Theorem 1, p. 400]). But $k[G/H]$ is finitely generated by assumption and $k[X]$ also, as X is affine. Hence, $k[G/H] \otimes [k[X]$ is finitely generated and so is $k[G/H \times X]^G$ because of the reductivity of G.

We introduce an action of $G \times H$ on $G \times X$ by defining $(g, h)(g_1, x) = (gg_1h^{-1}, gx)$. Next, we calculate the $G \times H$ invariant functions in two different ways. First, let us take the invariants of H and then of G. The invariants of H (or $\{e\} \times H$) are $k[G]^H \otimes k[X]$ [6; Lemma 1, p. 7]. Therefore

$$k[G \times X]^{G \times H} \simeq k[G/H \times X]^G.$$

Second, let us take the invariants of G and then of H. The restriction map from $k[G \times X]^G$ to $k[X]$ is injective and, furthermore, each $f \in k[X]$ extends to

an invariant function F by defining $F(g,x)$ to be $f(g^{-1}x)$. Therefore

$$k[G\times X]^{G\times H}\simeq k[X]^H.$$

Note 1. This theorem, in a somewhat weaker form, was first proved in [1; p. 245]. The proof above, whose generality we shall require later, was communicated to me by K. Pommerening and is due to an (unknown) referee.

Note 2. In the course of the proof, we showed that the restriction map from $G/H\times X$ to $H\times X$ gives an isomorphism from $k[G/H\times X]^G$ to $k[X]^H$. We shall use this later.

Criterion [1; Theorem 2, p. 239 or 8; §1, p. 359]. Let G be an affine algebraic group over k and let H be an observable subgroup of G. The following conditions on H are equivalent:

(i) H' is a finitely generated k-algebra;

(ii) there is a rational representation of G on a finite-dimensional vector space V and a $v\in V$ so that $H=G_v$ and, if Z is the closure of the orbit Gv, then $\dim(Z-Gv)\leqq\dim Z-2$.

For a moment, let G be a reductive algebraic group and let U_1 be the unipotent radical of a parabolic subgroup of G. We shall show that U_1 satisfies condition (ii) in the criterion above; the main result in this paper (stated in the introduction) will then follow from Theorem (1.2).

V.L. Popov and K. Pommerening have conjectured that if H is any closed subgroup (of the reductive group G) whose normalizer contains a maximal torus of G, then H satisfies condition (ii) in the criterion [9]. The result in this paper, of course, is a special case of this conjecture.

(1.3) *Extending invariant functions.* Let us start by stating an important lemma of Nagata.

Lemma [7; Lemma 3.4.1, p. 54]. *Let G be a reductive algebraic group acting rationally on a finitely generated k-algebra R. Let J be an ideal in R which is invariant under G and let $f+J\in(R/J)^G$. There is a positive integer t and an $F\in R^G$ so that $f^t-F\in J$.*

Theorem. *Let G be a reductive algebraic group acting rationally on a finitely generated k-algebra R having no nilpotent elements. Let J be a radical ideal in R which is invariant under G. Let H be a closed subgroup of G which satisfies the criterion in* (1.2). *If $f+J\in(R/J)^H$, then there is a positive integer t and an $F\in R^H$ so that $f^t-F\in J$.*

Proof. According to condition (ii) of the criterion of (1.2) and routine normalization arguments, we may assume that there is a rational representation of G on a finite-dimensional vector space V and a $v\in V$ so that

(i) $H=G_v$;

(ii) if Z is the closure of the orbit Gv, then $\dim(Z-Gv)\leqq\dim Z-2$;

(iii) $k[Z]=H'=k[G]^H$.

Let X be the affine variety corresponding to R and let Y be the subvariety of X defined by J. We consider the restriction maps in the following diagram:

$$\begin{array}{ccc} X\times Z & \longrightarrow & X\times\{v\} \\ \downarrow & & \downarrow \\ Y\times Z & \longrightarrow & Y\times\{v\} \end{array}$$

The H-invariant function f on $Y\times\{v\}$ corresponds to a G-invariant function F_1 on $Y\times Z$ (Note 2 in (1.2)). By Nagata's lemma, there is a positive integer t and a G-invariant function F_2 on $X\times Z$ so that the restriction of F_2 to $Y\times Z$ is F_1^t. To prove the theorem, we may take F to be the restriction of F_2 to $X\times\{v\}$.

2. Examples

(2.1) **Theorem.** *Let G be a semi-simple algebraic group and let U be a maximal unipotent subgroup of G. Then U' is a finitely generated k-algebra.*

Sketch of proof. We may assume that G is simply connected. Let $B=TU$ be a Borel subgroup of G and let $\{\alpha_1,\ldots,\alpha_r\}$ be the corresponding fundamental system of roots. Let $\{\lambda_1,\ldots,\lambda_r\}$ be the set of highest weights of T so that $\langle\lambda_i,\alpha_j\rangle=\delta_{ij}$. For each $i=1,\ldots,r$, let $\rho_i: G\to GL(V_i)$ be a finite-dimensional, irreducible representation of G having highest weight λ_i and corresponding highest weight vector $v_i\in V_i$ [11; Theorem 39, p. 209]. Let $\rho=\rho_1\oplus\ldots\oplus\rho_r$ and $v=v_1\oplus\ldots\oplus v_r$. In [1; p. 247], it was shown that (a) $U=\{g\in G\colon \rho(g)v=v\}$ and (b) the irreducible components of $\mathrm{cl}(G\cdot v)-G\cdot v$ are orbits of vectors having the form $(0,0,\ldots,0,v_{m+1},\ldots,v_r)$ after a rearrangement of $1,\ldots,r$. Let S be the stabilizer in G of this vector. We need to show that $\dim S\geqq\dim U+2$.

If $m\geqq 2$, then S contains U and $T_1=\{t\in T\colon \lambda_i(t)=1 \text{ for } i=m+1,\ldots,r\}$. Since $\dim T_1\geqq 2$, we have $\dim S\geqq\dim T_1\cdot U\geqq\dim U+2$. If $m=1$, then S contains U and

$$T_1=\{t\in T\colon \lambda_i(t)=1 \text{ for } i=2,\ldots,r\}.$$

Let $\alpha=\alpha_1$ and let G_α be the three dimensional, connected, semi-simple subgroup of G determined by α [11; Corollary 6, p. 46]. A Borel subgroup B_α of G_α is in S. Since G_α/B_α is complete, G_α is in S and $\dim S\geqq\dim U+2$.

Note 1. Different proofs of this theorem, valid for characteristic 0, may be found in [2] and [12]. It follows from [12] that the integer m, above, is 1.

(2.2) **Theorem.** *Let* $\mathrm{char}\, k=0$. *Let G be a semi-simple algebraic group and let U_1 be the unipotent radical of a parabolic subgroup of G. Then U_1' is a finitely generated k-algebra.*

Proof. Let P be a parabolic subgroup of G and let $P=H\cdot U_1$ where H is reductive and U_1 is the unipotent radical of P. There is a maximal unipotent subgroup U of G so that $U=(U\cap H)U_1$. Let $f_1,\ldots,f_r\in k[G]$ be chosen so that $U'=k[f_1,\ldots,f_r]$ (by (2.1)).

The algebra U_1' is stable under H (Note 1 in (1.1)). There is a finite-dimensional, H-stable subspace W_1 of $k[G]$ which contains each f_i [4; Proposition, p. 62]. Let $W=W_1\cap U_1'$ and let $\{\phi_1,\ldots,\phi_m\}$ be a basis for W.

Since H acts on U_1' in a completely reducible manner, we can break U_1' up into irreducible subspaces. Let V be one such and suppose that it has highest weight vector v_0 with respect to $U\cap H$. Now $v_0\in U'$ and $V=kH\cdot v_0$ so $V\subset H\cdot U'\subset k[\phi_1,\ldots,\phi_m]\subset U_1'$. Hence, $U_1'=k[\phi_1,\ldots,\phi_m]$.

Note 2. In the proof above, we may choose each ϕ_i in $H\cdot U'$ so $U_1'=k[\phi_1,\ldots,\phi_m]=H\cdot U'$. This suggests the definition in (3.2).

3. The Algebra R

(3.1) **Terminology.** In the rest of this paper, we shall adapt the following notation. Let G be a simply-connected, semi-simple algebraic group. Let $B=TU$ be a Borel subgroup of G and let P be a parabolic subgroup of G which contains B. Then $P=HSU_1$ where H is connected and semi-simple, S is a torus, HS is reductive, and U_1 is the unipotent radical of P [4; p. 185]. There is a maximal unipotent subgroup U_H of H, normalized by T, and a maximal torus T_H in H such that $U=U_HU_1$ and $T=T_HS$. We shall denote by s_H that element in the normalizer of T_H in H such that $s_HU_Hs_H^{-1}=U_H^-$ [4; pp. 160–161].

(3.2) **Definition of R.** *The proof given in* (2.2) *suggests the following definition. Let R be the intersection of all those k-algebras R_1 in $k[G]$ such that*

(i) $h\cdot R_1=R_1$ *for all* $h\in H$;

(ii) R_1 *contains* U'.

Theorem. *The k-algebra R is finitely generated and is contained in U_1'.*

Proof. Since H normalizes U_1, we have $h\cdot U_1'=U_1'$ for all $h\in H$ (Note 1 in (1.1)). Hence, R is contained in U_1'. Next, let us suppose that $U'=k[f_1,\ldots,f_m]$. There is a finite-dimensional H-invariant vector space W_1 in $k[G]$ which contains each f_i [4; Proposition, p. 62]. Let W be the subspace of W_1 spanned by all the $h\cdot f_i$, $h\in H$. Let us change notation and let $\{f_1,\ldots,f_n\}$ be a basis for W. Then it is easy to see that $R=k[f_1,\ldots,f_n]$.

4. The Variety Y

(4.1) **Left Translations.** The group G acts on itself through left and right translation. It will be important to distinguish these actions, so, we shall write $g*g_1=gg_1$ for left translation. This gives rise to an action on $k[G]$ by $(g*f)(g_1)=f(g^{-1}g_1)$. The following properties will be important later. (Their easy proofs are omitted.)

P1) Left and right translation commute on $k[G]$, i.e.,

$$h \cdot (g * f) = g * (h \cdot f) \quad \text{for all } g, h \in G, f \in k[G].$$

P2) If H is any subgroup of G and $g \in G$, then $g * H' = H'$.

P3) Let R be as in (3.2) and $g \in G$. Then $g * R = R$.

(4.2) **Actions on Y.** Let Y be the affine variety such that $k[Y] = R$. We saw in P3), (4.1), that $g * R = R$ for all $g \in G$. Let us choose linearly independent functions $f_1, \ldots, f_n \in R$ so that

(a) $R = k[f_1, \ldots, f_n]$;

(b) the vector space spanned by $f_1, \ldots, f_n$ is $(G, *)$ stable.

Then there are functions $P_{ij} \in k[G]$, $1 \leqq i, j \leqq n$, such that

$$g * f_i = \sum_{j=1}^{n} P_{ij}(g) f_j$$

for all $g \in G$ [7; Lemma 3.1, p. 45]. We define a map $G \times Y \to Y$, $(g, y) \to g * y$, by

$$f_i(g * y) = (g^{-1} * f_i)(y) = \sum_{j=1}^{n} P_{ij}(g^{-1}) f_j(y).$$

This gives a (rational) action of G on Y.

Similarly, $h \cdot R = R$ for all $h \in H$ and we may define an action of H on Y. We shall denote this by $(h, y) \to h \cdot y$.

The inclusion π° of R into $k[G]$ gives rise to a morphism π: $G \to Y$ which is equivariant with respect to both $(G, *)$ and $(H, \cdot)$. We put $\pi(e) = y$.

P1) $\pi(g) = g * y$ and $(\pi^\circ f)(g) = f(g * y)$ for all $f \in R$, $g \in G$.

P2) For all $h \in H$, $g \in G$ we have $h \cdot (g * y) = g h^{-1} * y$.

Proof. We have for each $f \in R$,

$$\begin{aligned} f(h \cdot g * y) &= (h^{-1} \cdot f)(g * y) = \pi^\circ(h^{-1} \cdot f)(g) \\ &= (h^{-1} \cdot \pi^\circ f)(g) = (\pi^\circ f)(g h^{-1}) = f(g h^{-1} * y). \end{aligned}$$

P3) $U_1 = G_y = \{g \in G\colon g * y = y\}$.

Proof. Let $h \in H$, $g \in G_y$, and $f \in U'$. Then

$$(h \cdot f)(g * y) = (h \cdot f)(y)$$

and, by P2),

$$f(gh * y) = f(h * y) \quad \text{or} \quad (\pi^\circ f)(gh) = (\pi^\circ f)(h).$$

But U' separates points in G/U according to Note 2 in (1.1). Hence, we see that

$$(*) \qquad h^{-1} g h \in U \quad \text{for all } h \in H.$$

Putting $h = e$, we see that G_y is contained in U.

Next, let $g \in G_y$, $f \in R$, and $t \in T$. Then,

$$\begin{aligned} f(t^{-1}gt * y) &= (\pi^\circ f)(t^{-1}gt) = \pi^\circ(t * t \cdot f)(g) \\ &= \pi^\circ(t * t \cdot f)(e) \quad (\text{since } g \in G_y) \\ &= (\pi^\circ f)(t^{-1}et) = f(y). \end{aligned}$$

It follows from this calculation that G_y is normalized by T. Then G_y is generated by the U_α which it contains [4; Proposition, p. 170].

Now U_1 is contained in G_y. For R is contained in U_1' and, so, $u \cdot \pi^\circ f = \pi^\circ f$ for all $u \in U_1$, $f \in R$. Hence, $(\pi^\circ f)(eu) = (\pi^\circ f)(e)$ and, by P1), $f(u * y) = f(y)$. If there is a U_α in G_y which is not in U_1, then U_α is in U_H and $s_H^{-1} U_\alpha s_H$ is in U^-. This shows that U_α cannot be in G_y (using (*)).

(4.3) **Theorem.** *The orbit $H \cdot y$ is closed in Y.*

Proof. Let us assume that the theorem is false and let $S_1 = \mathrm{cl}(H \cdot y) - H \cdot y$. Then, there is a one-parameter subgroup λ of H so that

$$\mathrm{limit}_{t \to 0} \lambda(t)^{-1} \cdot y$$

exists and is contained in S_1 [5; Theorem 1.4, p. 302]. Then

$$\mathrm{limit}_{t \to 0} (h \cdot f)(\lambda(t)^{-1} \cdot y)$$

exists for all $h \in H$ and $f \in U'$. By P2), (4.2), we have

(*) $$\mathrm{limit}_{t \to 0} f(\lambda(t) h * y)$$

exists.

Let us return now to the notation and terminology used in (2.1). Let $Z = \mathrm{cl}(Gv)$. The inclusion of $k[Z]$ into U' and, then, R, gives rise to a morphism $\psi\colon Y \to Z$ such that $\psi(g * y) = gv$ for all g in G. From (*) we see that

$$\mathrm{limit}_{t \to 0} f(\lambda(t) h * y)$$

exists for all $f \in k[Z]$ or that

$$\mathrm{limit}_{t \to 0} f(\lambda(t) hv)$$

exists for all $h \in H$, $f \in k[Z]$. Hence,

(**) $$\mathrm{limit}_{t \to 0} \lambda(t) hv \quad \text{and} \quad \mathrm{limit}_{t \to 0} h^{-1} \lambda(t) hv$$

exist. We may now assume that $\lambda(t)$ is in T_H. Then taking $h = e$ and remembering the construction of v, we have $\langle \lambda_i, \lambda \rangle \geqq 0$ for $i = 1, \ldots, r$.

Let us assume that $\{\alpha_1, \ldots, \alpha_l\}$ is the subset of $\{\alpha_1, \ldots, \alpha_r\}$ defining H. Taking $h = s_H$ in (**) gives $\langle s_H \lambda_i, \lambda \rangle \geqq 0$ for each $i = 1, \ldots, l$. But there is an integer j, $1 \leqq j \leqq l$, so that $s_H \cdot \lambda_i | T_H = -\lambda_j | T_H$. Hence, $\langle -\lambda_j, \lambda \rangle \geqq 0$ and $\langle \lambda_i, \lambda \rangle = 0$ for $j, i = 1, \ldots, l$. These equations show that λ is trivial, contradicting our opening assumption.

5. Proof of Theorem

(5.1) **The variety X.** Let v be as in (2.1) and let X be the normalization of $\mathrm{cl}(Gv)$. We may assume that G acts on X via $(g,x)\to gx$ and that there is an $x\in X$ (lying over v) so that $X=\mathrm{cl}(Gx)$. The proof in (2.1) shows that $\dim(X-Gx)\leqq\dim X-2$. Furthermore, there are points $x_1,\ldots,x_n$ in X so that $X-Gx$ is the union of the orbits Gx_i.

The inclusion from $U'=k[X]$ into R gives rise to a morphism $\phi\colon Y\to X$ such that $\phi(g*y)=gx$ for all $g\in G$.

Theorem. $\phi(Y-G*y)\subset X-Gx$.

Proof. The set $F=Y-G*y$ is $(G,*)$ stable and if $gx\in\phi(F)$ for some $g\in G$, then $\phi(F)$ contains Gx. (We now seek a contradiction.) Now F is also $(H,\cdot)$ stable (by P2) in (4.2)). Hence, the ideal I in R corresponding to F is invariant under U_H; this implies that there is a non-zero $f\in I$ which is fixed by U_H. But f is in R and R is contained in U_1' so that f is fixed by U. In this case, we may consider f to be a function on X. Then $f(F)=0$ implies that $f(\phi(F))=0$ and $f(Gx)=0$. This proves that $f=0$, a contradiction.

(5.2) **Theorem.** *Let F be an irreducible, $(H,\cdot)$ stable closed subset in Y and let P be its corresponding ideal in R. Let $H\cdot(P\cap U')$ be the smallest $(H,\cdot)$ stable ideal in R which contains $P\cap U'$ and let J be the radical of $H\cdot(P\cap U')$. Then $J=P$.*

Proof. First, we note that J is contained in P and that J is $(H,\cdot)$ stable since $H\cdot(P\cap U')$ is. Next, we consider R/J and let us assume that $P/J\neq 0$. Then there is an element $f+J$ in P/J which is non-zero and is U_H invariant. According to the theorem in (1.3), there is an element $F\in R$, fixed by U_H, so that $F+J=f^t+J$ for some positive integer t. But F is in U_1' so $F\in P\cap U'$ and $f+J=J$, a contradiction.

Corollary. *Let F_1 and F_2 be $(H,\cdot)$ stable irreducible, closed subsets of Y. If*

$$\mathrm{cl}(\phi(F_1))\supset\mathrm{cl}(\phi(F_2)),\quad\text{then}\quad F_1\supset F_2.$$

Proof. Let P_1 (resp. P_2) be the prime ideal in R vanishing on F_1 (resp. F_2). Now $P_1\cap U'$ vanishes on $\phi(F_1)$ and, so, by assumption, $P_1\cap U'\subset P_2$. Then $H\cdot(P_1\cap U')\subset P_2$ and, by the Theorem, $P_1\subset P_2$.

(5.3) *The set $Y-G*y$.* Let $F_1,\ldots,F_r$ be the irreducible components of $Y-G*y$. Since both H and G are connected, each F_i is stable under $(H,\cdot)$ and $(G,*)$. Now, $\phi(F_i)$ is irreducible and a union of $(G,*)$ orbits so there is an x_i with $\mathrm{cl}(\phi(F_i))=\mathrm{cl}(Gx_i)$. Let $y_i\in F_i$ be any point such that $\phi(y_i)=x_i$.

The group $G\times H$ acts on Y via $(g,h)y=g*h\cdot y$. Let Z be the closure of the $G\times H$ orbit of y_i. Since F_i is stable with respect to both $(G,*)$ and $(H,\cdot)$, we have $Z\subset F_i$. On the other hand, $\phi(g*y_i)=gx_i$ so $\mathrm{cl}(\phi(Z))\supset\mathrm{cl}(Gx_i)=\mathrm{cl}(\phi(F_i))$. By the corollary in (5.2), we have $Z\supset F_i$.

Theorem. *The set $Y-G*y$ is the union of finitely many orbits of $G\times H$.*

(5.4) **Proof of Theorem in Introduction.** Let F be an irreducible component of $Y-G*y$. We shall assume that $\dim F=\dim Y-1$ and seek a contradiction.

Since H is connected, F is $(H, \cdot)$ stable. Furthermore, $\phi(F) \subset X - Gx$ by the theorem, (5.1). The $(H, \cdot)$ stable, closed sets F and $H \cdot y$ are disjoint. Hence, there is an $f \in R^H$ such that $f(F) = 0$ and $f(H \cdot y) = 1$. Now f is in U_1' and is fixed by U_H, so f is in U' and we may view f as a function on X.

The function f vanishes on $\phi(F)$. But $\phi(F)$, being in $X - Gx$, has codimension $\geqq 2$ in X. Hence, there is an irreducible set D in X, having codimension 1, on which f vanishes and which contains $\phi(F)$. Again, since $X - Gx$ has codimension $\geqq 2$ in X, D must contain points in Gx and, in fact, $D \cap Gx$ must be dense in D.

Let $F_1, \ldots, F_r$ be the irreducible components of $\{y \in Y: f(y) = 0\}$ with, say, $F = F_1$. Since $f \in R^H$, each F_i is $(H, \cdot)$ stable. Since $\phi(G * y) = Gx$, there is an i, $2 \leqq i \leqq r$, so that $\mathrm{cl}(\phi(F_i)) = D$. Then

$$\mathrm{cl}(\phi(F_i)) = D \supset \mathrm{cl}(\phi(F))$$

and, by the corollary in (5.2), $F_i \supset F_1$ which contradicts our choice of $F_1, \ldots, F_r$.

References

1. Grosshans, F.: Observable groups and Hilbert's fourteenth problem. Amer. J. Math. **95**, 229–253 (1973)
2. Hadžiev, Dž.: Some questions in the theory of vector invariants. Math. USSR Sbornik **1**, 383–396 (1967)
3. Hochschild, G., Mostow, G.D.: Unipotent groups in invariant theory. Proc. Natl. Acad. Sci. USA **70**, 646–648 (1973)
4. Humphreys, J.E.: Linear algebraic groups. New York: Springer 1975
5. Kempf, G.R.: Instability in invariant theory. Ann. of Math. **108**, 299–316 (1978)
6. Nagata, M.: Lectures on the Fourteenth Problem of Hilbert. Bombay: Tata Institute 1965
7. Newstead, P.E.: Introduction to moduli problems and orbit spaces. Bombay: Tata Institute 1978
8. Pommerening, K.: Invarianten unipotenter Gruppen. Math. Z. **176**, 359–374 (1981)
9. Pommerening, K.: Ordered sets with the standardizing property and straightening laws for algebras of invariants, to appear
10. Sampson, J.H., Washnitzer, G.: A Künneth formula for coherent algebraic sheaves. Ill. J. Math. **3**, 389–402 (1959)
11. Steinberg, R.: Lectures on Chevalley groups. Mimeographed lecture notes. New Haven: Yale University Math. Dept. 1968
12. Vinberg, B., Popov, V.L.: On a class of quasi-homogeneous affine varieties. Math. USSR Izv. **6**, 743–758 (1972)

Oblatum 1-VII-1982 & 19-I-1983

Invent. math. 73, 11-32 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

On Siegel's Lemma

E. Bombieri[1] and J. Vaaler[2]

[1] The Institute for Advanced Study, School of Mathematics, Princeton, NJ 08540, USA
[2] The University of Texas, Department of Mathematics, Austin, TX 78712, USA

Introduction

I.1. A basic technique in transcendental number theory and diophantine approximation requires the construction of auxiliary polynomials, not identically 0, with suitable properties such as the vanishing at certain points. In a few instances the auxiliary polynomials can be explicitly found but in the general case we must rely on Thue's idea of proving the existence of the required polynomial in an indirect way. In his well-known paper of 1909 ([10], pp. 288–289) Thue expressed the properties to be satisfied by means of linear conditions on the coefficients of the polynomial. By regarding the coefficients as unknowns Thue was led to solve in rational integers a set of M linear equations in N unknowns, with $N>M$. Now Thue was able to prove the existence of a non-trivial small solution by using Dirichlet's box principle.

Siegel ([7], Bd. I, p. 213, Hilfssatz) was the first to state this principle formally. He proved that the linear system

$$\begin{aligned} &a_{11}x_1+\ldots+a_{1N}x_N=0\\ &\cdots\cdots\cdots\cdots\cdots\\ &a_{M1}x_1+\ldots+a_{MN}x_N=0 \end{aligned}$$

where the a_{mn} are rational integers, not all 0, bounded by B and where $N>M$ has a solution $x_1,\ldots,x_N$ with the x_n's rational integers, not all 0, bounded by $1+(NB)^{M/(N-M)}$. Since then, all results of this type have been referred to in the recent literature as Siegel's lemma.

For more refined applications the coefficients a_{mn} may be taken as integers in a number field K and the unknowns x_n are restricted to integers of a subfield k of K; if $[K:k]=r$ we need $N>Mr$, since expressing the elements of K in terms of a basis over k replaces each equation having coefficients in K with r equations having coefficients in k, thus bringing the number of equations to Mr. In terms of a suitable notion of height, say $h(x)$, we get solutions satisfying bounds of the type

$$\max_n h(x_n)\leqq C_1(C_2 N\max_{mn}h(a_{mn}))^{Mr/(N-Mr)}.$$

Offprint requests to: E. Bombieri

I.2. Although in most cases Siegel's original estimate and even cruder ones suffice for applications, there are situations in which it is essential to obtain bounds which are as sharp as possible. This appears in the work of Thue and it is explicitly mentioned in Siegel ([7], p. 213). More recently Bombieri [1] obtained new results on effective approximations to algebraic numbers by using accurate bounds obtained with Siegel's lemma. In a successive paper, Bombieri and Mueller [2] were able to deal with effective approximations to some numbers of type $\sqrt[4]{a/b}$, by improving the estimates obtained by Siegel's lemma; it should be noted that the earlier general results in [1] were insufficient to deal with numbers of type $\sqrt[4]{a/b}$.

The method of [2] was rather ad hoc and the question remained to find a general formulation of Siegel's lemma which would cover the improvement obtained in those special cases. This is what we do in this paper.

Over the rational integers we formulate Siegel's lemma as follows.

Theorem 1. *Let* $\sum_{n=1}^{N} a_{mn}x_n=0$, $m=1,2,\ldots,M$ *be a linear system of* M *linearly independent equations in* $N>M$ *unknowns, with rational integer coefficients* a_{mn}. *Then there is a non-trivial solution in rational integers* x_n *with*

$$\max_n |x_n| \leqq (D^{-1}\sqrt{|\det(AA^T)|})^{1/(N-M)}, \tag{1.1}$$

where A *denotes the* $M\times N$ *matrix* $A=(a_{mn})$, A^T *is the transpose and where* D *is the greatest common divisor of the determinants of all* $M\times M$ *minors of* A.

Our results are stronger than the statement of Theorem 1. In some instances it is useful to control not only one solution to our linear system but rather a basis of solutions. We prove

Theorem 2. *With the same hypotheses as Theorem* 1 *there are* $N-M$ *linearly independent integral solutions* $\mathbf{x}_l=(x_{1l},\ldots,x_{Nl})$, $l=1,\ldots,N-M$, *with*

$$\prod_{l=1}^{N-M} \max_n |x_{nl}| \leqq D^{-1}\sqrt{|\det(AA^T)|}. \tag{1.2}$$

Theorems 1 and 2 provide us with an invariant formulation of Siegel's lemma, since the quantity $D^{-1}\sqrt{|\det(AA^T)|}$ remains unchanged if we multiply A on the left by any non-singular $M\times M$ matrix with integral entries. In fact there is no reason to limit ourselves to the case of integral coefficients in view of the homogeneous nature of the problem. It will turn out that the bound $D^{-1}\sqrt{|\det(AA^T)|}$ is nothing else than an absolute height on the Grassmannian parametrizing our linear systems. With that interpretation, our theorems go through essentially unchanged in the most general case.

I.3. As one may expect, our proofs are based on the powerful techniques of Geometry of Numbers rather than the simpler but less accurate Dirichlet box principle. However two difficulties appear and they will provide the motivation for working in the context of adèles, where these difficulties will be automatically resolved.

Let us consider the simplest situation, namely the case in which we want to solve the system

$$a_{11}x_1+\ldots+a_{1N}x_N=0$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$a_{M1}x_1+\ldots+a_{MN}x_N=0$$

in integers, where the a_{mn} are also integers. We think of $x_1, \ldots, x_N$ as forming a lattice $\Lambda\subset\mathbb{R}^N$ and we are interested in finding a lattice point with $|x_n|\leqq t$, $n=1, \ldots, N$ on the linear space T defined by the above linear system. Now $\Gamma=\Lambda\cap T$ is a lattice in T, so we want to find a point of Γ in the convex region $Q_t=\{|x_n|<t,\ n=1, \ldots, N\}\cap T$. According to Minkowski's first theorem it suffices to establish

$$\mathrm{vol}(Q_t)\geqq 2^{\dim T}\det(\Gamma),$$

and hence we need

a) a lower bound for $\mathrm{vol}(Q_t)$,

b) an upper bound for $\det(\Gamma)$.

Problem a) has been solved in an optimal way by Vaaler [11] with his cube slicing inequality; the solution of b) requires a description of the lattice Γ in T, and that was done by H.J.S. Smith [9] in 1861 by reduction of the system to a new one in triangular form; see Skolem [8], p. 5. As already noted in Skolem's monograph, if we try a similar procedure but working instead with the integers in a number field we run into difficulties because the reduction to triangular form is not always possible, unless we are in a principal ideal domain. A second difficulty which appears to be even more substantial arises if we try to deal with a basis of solutions rather than just one solution. Clearly this requires the use of the much deeper Minkowski's second theorem. An essential ingredient of all known proofs of Minkowski's second theorem deals with a lattice $\Lambda=\mathbb{Z}^N$ and a sublattice M and provides bases of M and Λ such that the basis of M can be expressed in triangular form in terms of a basis of Λ. If we replace $\mathbb{Z}^N$ by $(O_k)^N$, where O_k denotes the integers of the number field k, the above result holds only if O_k is a principal ideal domain, and otherwise the proof of Minkowski's second theorem cannot be completed in any obvious way.

A simple way for resolving this difficulty consists in restoring unique factorization by localizing on a sufficiently large set of primes, i.e. replacing O_k by $S^{-1}O_k$ for a suitable multiplicative set S. Now we have to restore the fact that we want $S^{-1}O_k$ to be discrete, and this is achieved by considering the diagonal embedding of $S^{-1}O_k$ not just in the usual product of real and complex spaces as is the case for O_k, but in a product with p-adic factors as well. But then, if we have to localize with respect to many places, why shouldn't we localize with respect to all places of k? This means that we should work with the adèles of k and it also means that we need Geometry of Numbers over the adèles.

We have been unable to locate in the literature an explicit formulation and proof of the first and second theorem of Minkowski over the adèles. Thus one

of the main results (Theorem 3) of this paper consists in a proof of Minkowski's second theorem in an adelic context. The application to Siegel's lemma will then follow more or less automatically, once we have Vaaler's cube slicing inequality over local fields to control the volume of the restriction of a cube to a linear space and once we know how to control the lattice $\Lambda \cap T$. This is done by appealing to the Brill-Gordan theorem on duality on Grassmann co-ordinates.

The content of our paper is as follows.

In Sect. II we introduce absolute values, heights and basic facts about adèles.

In Sect. III we prove the adelic Minkowski second theorem by adapting a theorem of Davenport and Estermann's proof to our adelic context.

In Sect. IV we prove the cube slicing inequality over the adèles.

In Sect. V we apply the Geometry of Numbers to the Siegel lemma and formulate some special but significant corollaries.

Acknowledgement. We wish to thank here R. Langlands and A. Borel for some helpful conversations.

Heights and Measures

II.1. In this section we fix the notation and give some definitions for the objects we will work with throughout the remainder of the paper.

Let k be an algebraic number field. For every place v of k we write k_v for the completion of k at v. We then write $d=[k:\mathbb{Q}]$ for the degree of k over $\mathbb{Q}$ and $d_v=[k_v:\mathbb{Q}_v]$ for the local degrees. If v is a finite place of k which lies over the prime number p, we write $v|p$. If v is an infinite place of k we write $v|\infty$.

For every place v of k we normalize an absolute value $|\ |_v$ as follows:

(i) if $v|p$ then $|p|_v=p^{-d_v/d}$,

(ii) if $v|\infty$ and $x\in k_v$ then $|x|_v=|x|^{d_v/d}$, where $|\ |$ is the usual Euclidean absolute value on $\mathbb{R}$ or $\mathbb{C}$.

In view of our normalizations we have the product formula

$$\prod_v |x|_v=1 \tag{2.1}$$

whenever $x\in k$ and $x\neq 0$. It will also be convenient to employ a second normalized absolute value which we define by

$$\|x\|_v=|x|_v^{d/d_v}. \tag{2.2}$$

Thus $\|\ \|_v$ extends the usual p-adic absolute value if $v|p$ and the Euclidean absolute value if $v|\infty$.

Let

$$\mathbf{x}=\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

denote a column vector in $(k)^N$ and let us define

$$|\mathbf{x}|_v = \max_n |x_n|_v; \tag{2.3}$$

then the (homogeneous) height of the vector $\mathbf{x}$ is

$$h(\mathbf{x}) = \prod_v |\mathbf{x}|_v. \tag{2.4}$$

A basic property of the height is that it depends only on the algebraic vector $\mathbf{x}$ and not on the field k containing its co-ordinates (see Weil [13]). Also if $a \in k$, $a \neq 0$ the product formula yields

$$h(a\mathbf{x}) = h(\mathbf{x})$$

and thus $h(\mathbf{x})$ is an absolute height on the projective space $\mathbb{P}_k^{N-1}$.

Let $X = (x_{mn})$ be an $M \times N$ matrix with entries in k and $\operatorname{rank}(X) = M < N$. If $J \subseteq \{1, 2, \ldots, N\}$ is a subset with $|J| = M$ elements we write

$$X_J = (x_{mn}), \qquad m = 1, 2, \ldots, M, \quad n \in J$$

for the corresponding submatrix. For each place v of k we define the local height $H_v(X)$ as follows:

(i) $H_v(X) = \max_{|J|=M} |\det X_J|_v$ if $v \nmid \infty$,

(ii) $H_v(X) = |\det X X^T|_v^{\frac{1}{2}}$ if v is real, where X^T is the transpose of X,

(iii) $H_v(X) = |\det X X^*|_v^{\frac{1}{2}}$ if v is complex, where X^* is the complex conjugate transpose of X.

For $v | \infty$ we also have

(iv) $H_v(X) = (\sum_{|J|=M} \|\det X_J\|_v^2)^{\frac{d_v}{2d}}$ by the Cauchy-Binet formula.

If we form a product over all valuations we obtain a global height

$$H(X) = \prod_v H_v(X).$$

Because of our normalizations, $H(X)$ depends only on the entries in the matrix X and not on the field containing them. In other words, $H(X)$ does not change if k is replaced by a different number field containing the entries x_{mn}. By (i) and (iv) we see that $H(X)$ depends only on the Grassmann co-ordinates $(\det X_J)$ of X and thus $H(X)$ is an intrinsic height on a Grassmann manifold. This makes it clear that if C is a nonsingular $M \times M$ matrix over k then

$$H(CX) = \{\prod_v |\det C|_v\} H(X) = H(X). \tag{2.5}$$

Now suppose that $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ is partitioned into blocks consisting of an $M_1 \times N$ matrix X_1 and an $M_2 \times N$ matrix X_2. For each place v we have

$$H_v(X) \leqq H_v(X_1) H_v(X_2). \tag{2.6}$$

If v is finite this follows from Laplace's expansion and the fact that $|\ |_v$ is ultrametric. If v is complex (or real) then

$$|\det XX^*|_v = \left|\det \begin{pmatrix} X_1X_1^* & X_1X_2^* \\ X_2X_1^* & X_2X_2^* \end{pmatrix}\right|_v \leqq |\det(X_1X_1^*)|_v\,|\det(X_2X_2^*)|_v$$

by an inequality of Fischer [5]. Of course (2.6) also holds for the global height H and extends by induction to matrices X which are partitioned similarly into any finite number of blocks.

We extend our definitions of H_v and H to $M \times N$ matrices X with $\operatorname{rank}(X) = N < M$ by operating with X^T.

II.2. We denote by O_k the ring of integers of k and if v is a finite place of k we write O_v for the maximal compact subring of k_v, that is,

$$O_v = \{x \in k_v : |x|_v \leqq 1\}.$$

Let $k_{\mathbb{A}}$ be the adèle ring of k; elements of $k_{\mathbb{A}}$ will be written as $x = (x_v)$ where x_v is the v-component of x, for all places v of k. All the properties of $k_{\mathbb{A}}$ which we will need can be found in Weil [14], Chap. IV.

Since the additive group of $k_{\mathbb{A}}$ is locally compact we can determine a Haar measure on $k_{\mathbb{A}}$ which is unique up to a multiplicative constant. We do this as follows:

(i) If $v|p$ we let β_v denote Haar measure on k_v normalized so that

$$\beta_v(O_v) = |\mathscr{D}_v|_v^{d/2},$$

where $\mathscr{D}_v$ is the local different of k at v.

(ii) If $v|\infty$ and $k_v = \mathbb{R}$ we let β_v denote the ordinary Lebesgue measure on $\mathbb{R}$.

(iii) If $v|\infty$ and $k_v = \mathbb{C}$ we let β_v denote the Lebesgue measure on the complex plane multiplied by 2.

Now the product measure

$$\beta = \prod_v \beta_v$$

is the required Haar measure on $k_{\mathbb{A}}$ (to be precise, β determines a Haar measure on all open subgroups $\prod_{v \in P} k_v \times \prod_{v \notin P} O_v$ where P is any finite set of places containing all infinite places, and the Haar measure on $k_{\mathbb{A}}$ is the unique measure which agrees with the product measure on this family of subgroups).

We consider k as a discrete subgroup of $k_{\mathbb{A}}$ by means of the usual diagonal embedding and we denote by φ the canonical homomorphism $\varphi: k_{\mathbb{A}} \to k_{\mathbb{A}}/k$ of $k_{\mathbb{A}}$ onto the compact group $k_{\mathbb{A}}/k$. The Haar measure induced by φ on $k_{\mathbb{A}}/k$ will also be denoted by β and in view of our normalizations we have

$$\beta(k_{\mathbb{A}}/k) = 1. \tag{2.7}$$

Let

$$G = G_1 \times G_2 \times \ldots \times G_L \tag{2.8}$$

be a product of locally compact groups in which each factor G_l is either $k_{\mathbb{A}}$ or $k_{\mathbb{A}}/k$. The Haar measure β on each factor determines a unique product measure on G. We write V for this product measure and note that V is then a Haar measure on G. Of course there are many different groups G that can be formed in this way. However, there will be no confusion if we use V to denote the corresponding Haar measure on each of them.

Let G be a group of the form (2.8) and suppose that the factor G_l is $k_{\mathbb{A}}$. We define a homomorphism φ_l on G by

$$\varphi_l(g_1, \ldots, g_l, \ldots, g_L) = (g_1, \ldots, \varphi(g_l), \ldots, g_L).$$

If each of the factors $G_1, G_2, \ldots, G_l$ is $k_{\mathbb{A}}$ we let Φ_l denote the homomorphism $\varphi_1 \circ \varphi_2 \circ \ldots \circ \varphi_l$. Since each mapping φ_l or Φ_l is a homomorphism, the Haar measure of a measurable set $X \subseteq G$ is preserved whenever the homomorphism restricted to X is injective. Thus we have

$$V(X) = V(\Phi_l(X))$$

if $X \subseteq G$ is measurable and Φ_l restricted to X is an injection (and similarly for φ_l).

Let $x = (x_v)$ be an element of $k_{\mathbb{A}}$ and let α be a real number. We introduce a scalar multiplication by defining αx to be the point $y = (y_v)$ in $k_{\mathbb{A}}$ determined by

$$\begin{aligned} y_v &= \alpha x_v \quad \text{if } v \mid \infty, \\ y_v &= x_v \quad \text{if } v \nmid \infty. \end{aligned}$$

If L is a positive integer we shall denote elements of $(k_{\mathbb{A}})^L$ by column vectors $\mathbf{x}$ and we extend our notion of scalar multiplication to vectors $\mathbf{x}$ by acting componentwise as

$$\alpha \mathbf{x} = \begin{pmatrix} \alpha x_1 \\ \alpha x_2 \\ \vdots \\ \alpha x_L \end{pmatrix}.$$

If $X \subseteq (k_{\mathbb{A}})^L$ then $\alpha X \subseteq (k_{\mathbb{A}})^L$ is obtained by applying scalar multiplication by α to each $\mathbf{x}$ in X. Clearly we have

$$V(\alpha X) = |\alpha|^{dL} V(X) \tag{2.9}$$

for each measurable subset X in $(k_{\mathbb{A}})^L$, where $|\alpha|$ is the Euclidean absolute value of the real number α.

Successive Minima

III.1. Throughout this section L will denote a fixed positive integer. We write $E = (k)^L$ and $E_v = (k_v)^L$ for the L-dimensional vector spaces over k and k_v respectively. We also write $E_{\mathbb{A}} = (k_{\mathbb{A}})^L$ for the L-fold product of adèles.

For each finite place v of k let M_v be a k_v-lattice in E_v, that is, a compact and open O_v-module in E_v. We assume that for almost all v, $M_v=(O_v)^L$. For each infinite place v of k let S_v be a nonempty, open, convex, symmetric subset of E_v. By symmetric we mean that $S_v=-S_v$. To avoid some minor complications it will be convenient to assume that each S_v is also bounded. We then define

$$\mathscr{S}=\prod_{v|\infty} S_v \times \prod_{v\nmid\infty} M_v. \tag{3.1}$$

It follows that $\mathscr{S}\subseteq E_{\mathbf{A}}$, $\mathscr{S}$ is an open neighborhood of 0 and the closure of $\mathscr{S}$ is compact. As E is a discrete subgroup of $E_{\mathbf{A}}$ it is clear that $\mathscr{S}\cap E$ is finite.

We are now able to define the successive minima for $\mathscr{S}$ with respect to the subgroup E. We do this as follows. For each integer l, $1\leqq l\leqq L$, let

$$\lambda_l=\inf\{\lambda>0\colon (\lambda\mathscr{S})\cap E \text{ contains } l \text{ linearly independent vectors}\}.$$

It is obvious that $0\leqq\lambda_1\leqq\lambda_2\leqq\ldots\leqq\lambda_L\leqq\infty$. In fact the set

$$M=\bigcap_{v\nmid\infty}(E\cap M_v)$$

is a k-lattice, that is, an O_k-module in E containing a basis for E over k ([14], Theorem 2, p. 84). If we write $E_\infty=\prod_{v|\infty}(k_v)^L$ then M can be viewed as the projection of

$$(E_\infty\times\prod_{v\nmid\infty}M_v)\cap E$$

into the first factor E_∞. Thus M is a discrete subgroup of E_∞. Since each S_v is open and bounded it follows that $M\cap\prod_{v|\infty}(\lambda S_v)$ is $\{0\}$ if $\lambda>0$ is sufficiently small and contains L linearly independent vectors if λ is sufficiently large. In this way we see that the successive minima λ_l satisfy $0<\lambda_1\leqq\lambda_2\leqq\ldots\leqq\lambda_L<\infty$. Indeed, it is clear that the numbers λ_l may also be defined by

$$\lambda_l=\inf\{\lambda>0\colon M\cap\prod_{v|\infty}(\lambda S_v) \text{ contains } l \text{ linearly independent vectors}\}. \tag{3.2}$$

Our version of the upper bound in Minkowski's second theorem takes the following form.

Theorem 3. *The successive minima $\lambda_1,\lambda_2,\ldots,\lambda_L$ satisfy the inequality*

$$(\lambda_1\lambda_2\ldots\lambda_L)^d V(\mathscr{S})\leqq 2^{dL}. \tag{3.3}$$

We prove Theorem 3 by first obtaining an adelic version of a result of Davenport [3] and Estermann [4]. As before M_v is a k_v-lattice for each finite place v of k and $M_v=(O_v)^L$ for almost all v. If v is an infinite place of k we let T_v denote an open, nonempty, convex subset of E_v. We obtain an open subset $\mathscr{T}\subseteq E_{\mathbf{A}}$ by defining

$$\mathscr{T}=\prod_{v|\infty}T_v\times\prod_{v\nmid\infty}M_v. \tag{3.4}$$

Next we suppose that $0<\mu_1\leqq\mu_2\leqq\ldots\leqq\mu_L<\infty$ satisfy the following condition: if $\mathbf{x}$ and $\mathbf{y}$ are vectors in $\mu_l\mathscr{T}$ with $\mathbf{x}-\mathbf{y}$ in E, then $x_n=y_n$ for $n=l, l+1, \ldots, L$.

Theorem 4. *Let $\mathscr{T}$ and $\mu_1, \mu_2, \ldots, \mu_L$ be as above, then*

$$(\mu_1\mu_2\ldots\mu_L)^d V(\mathscr{T})\leqq 1. \tag{3.5}$$

Before proving these theorems we make two remarks. Since the successive minima λ_l can be defined by (3.2) and $\prod_{v|\infty} S_v$ is a convex, symmetric subset of a Euclidean space, it is tempting to try to prove (3.3) from strictly Euclidean considerations. Now, however, the usual $\mathbb{Z}$-module has been replaced by the O_k-module M. As we noted in I.3 this leads to immediate difficulties. With each λ_l we may associate, in an obvious way, a vector $\mathbf{u}_l$ in M such that $\mathbf{u}_1, \ldots, \mathbf{u}_l$ are linearly independent and contained in $\prod_{v|\infty}(\lambda S_v)$ for all $\lambda>\lambda_l$. We would then like to determine a basis $\mathbf{b}_1, \ldots, \mathbf{b}_L$ for M as an O_k-module such that $\mathbf{u}_l$ is a linear combination of $\mathbf{b}_1, \ldots, \mathbf{b}_l$ for each l. In general such a basis may not exist if O_k is not a principal ideal domain. On the other hand, when we work over the adèles, the corresponding vectors $\mathbf{u}_l$ are selected from E. As E is a vector space over the field k, the problem of a suitable basis for E has a trivial solution.

We also note that our formulation of Minkowski's second theorem over the adèles is of some independent interest. We certainly do not believe that our application of this inequality to Siegel's lemma exhausts its usefulness.

III.2. In our proof of Theorem 4 we require the following inequality.

Lemma 5. *If $\mathscr{T}\subseteq E_{\mathbf{A}}$ has the form* (3.4) *and $\mu\geqq 1$, then*

$$V(\Phi_l\{\mu\mathscr{T}\})\geqq\mu^{d(L-l)}V(\Phi_l\mathscr{T}) \tag{3.6}$$

for each integer l, $1\leqq l\leqq L$.

Proof. First suppose that $l=L$ and let $\mathbf{y}\in\mathscr{T}$. Since $0\in\mathscr{T}-\mathbf{y}$ it follows from the convexity of each T_v, $v|\infty$, that

$$\mu(\mathscr{T}-\mathbf{y})\supseteq(\mathscr{T}-\mathbf{y}).$$

Therefore we have

$$\Phi_L\{\mu\mathscr{T}-\mu\mathbf{y}\}\supseteq\Phi_L\{\mathscr{T}-\mathbf{y}\}$$

and so

$$V(\Phi_L\{\mu\mathscr{T}-\mu\mathbf{y}\})\geqq V(\Phi_L\{\mathscr{T}-\mathbf{y}\}).$$

As V is translation invariant and Φ_L is a homomorphism,

$$V(\Phi_L\{\mu\mathscr{T}-\mu\mathbf{y}\})=V(\Phi_L\{\mu\mathscr{T}\})$$

and similarly

$$V(\Phi_L\{\mathscr{T}-\mathbf{y}\})=V(\Phi_L\mathscr{T}).$$

This established (3.6) when $l=L$.

Next we suppose that $1 \leqq l \leqq L-1$. We may identify $E_{\mathbf{A}} = (k_{\mathbf{A}})^L$ in an obvious way with

$$(k_{\mathbf{A}})^l \times (k_{\mathbf{A}})^{L-l} = \{(\mathbf{x}, \mathbf{y}) : \mathbf{x} \in (k_{\mathbf{A}})^l \quad \text{and} \quad \mathbf{y} \in (k_{\mathbf{A}})^{L-l}\}.$$

In this identification the first component of $(\mathbf{x}, \mathbf{y})$ represents the first l coordinates of a vector $\mathbf{z} \in E_{\mathbf{A}}$. Then for a fixed vector $\mathbf{y} \in (k_{\mathbf{A}})^{L-l}$ let $\mathscr{T}(\mathbf{y}) \subseteq (k_{\mathbf{A}})^l$ be defined by

$$\mathscr{T}(\mathbf{y}) = \{\mathbf{x} \in (k_{\mathbf{A}})^l : (\mathbf{x}, \mathbf{y}) \in \mathscr{T}\}.$$

It follows that

$$\Phi_l\{\mathscr{T}(\mathbf{y})\} = \{\omega \in (k_{\mathbf{A}}/k)^l : (\omega, \mathbf{y}) \in \Phi_l \mathscr{T}\}.$$

If we replace $\mathscr{T}$ by $\mu\mathscr{T}$ then

$$\begin{aligned} V(\Phi_l(\mu\mathscr{T})) &= \int_{\mathbf{y} \in (k_{\mathbf{A}})^{L-l}} \int_{\substack{\omega \in (k_{\mathbf{A}}/k)^l \\ (\omega, \mathbf{y}) \in \Phi_l(\mu\mathscr{T})}} dV(\omega)\, dV(\mathbf{y}) \\ &= \int_{\mathbf{y} \in (k_{\mathbf{A}})^{L-l}} V(\Phi_l\{(\mu\mathscr{T})(\mathbf{y})\})\, dV(\mathbf{y}). \end{aligned}$$

We make the change of variable $\mathbf{y} \to \mu\mathbf{y}$ in the last integral, so that

$$V(\Phi_l(\mu\mathscr{T})) = \mu^{d(L-l)} \int_{\mathbf{y} \in (k_{\mathbf{A}})^{L-l}} V(\Phi_l\{(\mu\mathscr{T})(\mu\mathbf{y})\})\, dV(\mathbf{y}). \tag{3.7}$$

Now we observe that

$$(\mu\mathscr{T})(\mu\mathbf{y}) = \{\mathbf{x} \in (k_{\mathbf{A}})^l : (\mathbf{x}, \mu\mathbf{y}) \in \mu\mathscr{T}\} = \mu(\mathscr{T}(\mathbf{y})). \tag{3.8}$$

Since $\mu \geqq 1$ we have

$$V(\Phi_l\{\mu(\mathscr{T}(\mathbf{y}))\}) \geqq V(\Phi_l\{\mathscr{T}(\mathbf{y})\}) \tag{3.9}$$

as before using the facts that V is translation invariant, Φ_l is a homomorphism and $\mathscr{T}(\mathbf{y}) \subseteq (k_{\mathbf{A}})^l$. By combining (3.7), (3.8) and (3.9) we obtain

$$\begin{aligned} V(\Phi_l(\mu\mathscr{T})) &\geqq \mu^{d(L-l)} \int_{\mathbf{y} \in (k_{\mathbf{A}})^{L-l}} V(\Phi_l(\mathscr{T}(\mathbf{y})))\, dV(\mathbf{y}) \\ &= \mu^{d(L-l)}\, V(\Phi_l \mathscr{T}). \end{aligned}$$

This completes our proof.

Proof of Theorem 4. For each integer l, $1 \leqq l \leqq L-1$ we apply Lemma 5 to the set $\mu_l\mathscr{T}$ with $\mu = (\mu_{l+1}/\mu_l)$. We find that

$$V(\Phi_l\{\mu_{l+1}\mathscr{T}\}) \geqq (\mu_{l+1}/\mu_l)^{d(L-l)}\, V(\Phi_l\{\mu_l\mathscr{T}\}). \tag{3.10}$$

Next we claim that

$$\varphi_{l+1} : \Phi_l\{\mu_{l+1}\mathscr{T}\} \to (k_{\mathbf{A}}/k)^{l+1} \times (k_{\mathbf{A}})^{L-l-1} \tag{3.11}$$

is injective. To verify this let $\mathbf{x}$ and $\mathbf{y}$ be distinct points in $\Phi_l\{\mu_{l+1}\mathscr{T}\}$. Then there are distinct points $\mathbf{x}'$ and $\mathbf{y}'$ in $\mu_{l+1}\mathscr{T}$ such that $\Phi_l(\mathbf{x}') = \mathbf{x}$ and $\Phi_l(\mathbf{y}') = \mathbf{y}$.

Now the equation $\varphi_{l+1}(\mathbf{x})=\varphi_{l+1}(\mathbf{y})$ implies that

$$x_n=y_n \quad \text{if } n\neq l+1 \tag{3.12}$$

and also implies that

$$\Phi_{l+1}(\mathbf{x}')=\Phi_{l+1}(\mathbf{y}'). \tag{3.13}$$

But (3.13) shows that $\mathbf{x}'-\mathbf{y}'\in E$ and so by the hypotheses on $\mu_1, \mu_2, \ldots, \mu_L$ we must have

$$x'_n=y'_n \quad \text{if } n=l+1, l+2, \ldots, L.$$

Since Φ_l is the identity map on the $(l+1)$st co-ordinate we find that

$$x_n=y_n \quad \text{if } n=l+1. \tag{3.14}$$

Of course (3.12) and (3.14) are inconsistent with the fact that $\mathbf{x}$ and $\mathbf{y}$ are distinct points in $\Phi_l\{\mu_{l+1}\mathscr{T}\}$. Thus our claim that (3.11) is injective has been established. As V is a Haar measure this shows that

$$V(\Phi_l\{\mu_{l+1}\mathscr{T}\})=V(\Phi_{l+1}\{\mu_{l+1}\mathscr{T}\}). \tag{3.15}$$

By combining (3.10) and (3.15) we have

$$V(\Phi_{l+1}\{\mu_{l+1}\mathscr{T}\})\geqq(\mu_{l+1}/\mu_l)^{d(L-l)}\,V(\Phi_l\{\mu_l\mathscr{T}\})$$

for $l=1, 2, \ldots, L-1$. Next we multiply these inequalities together to obtain

$$V(\Phi_L\{\mu_L\mathscr{T}\})\geqq\prod_{l=1}^{L-1}(\mu_{l+1}/\mu_l)^{d(L-l)}\,V(\Phi_1\{\mu_1\mathscr{T}\}).$$

Finally we note that

$$V(\Phi_1\{\mu_1\mathscr{T}\})=V(\mu_1\mathscr{T})=(\mu_1)^{dL}\,V(\mathscr{T}),$$

since $\Phi_1=\varphi_1$ applied to $\mu_1\mathscr{T}$ is obviously injective, and also

$$V(\Phi_L\{\mu_L\mathscr{T}\})\leqq V((k_{\mathbf{A}}/k)^L)=1.$$

Thus we have

$$1\geqq\prod_{l=1}^{L-1}(\mu_{l+1}/\mu_l)^{d(L-l)}(\mu_1)^{dL}\,V(\mathscr{T})$$

which is equivalent to (3.5).

III.3. To deduce Theorem 3 from Theorem 4 we argue as follows. With each successive minima λ_l we may associate a vector $\mathbf{u}_l$ in E so that $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L\}$ are linearly independent over k and for every l, $1\leqq l\leqq L$, and $\lambda>\lambda_l$, we have

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_l\}\subseteq(\lambda\mathscr{S})\cap E.$$

Now let $U=(\mathbf{u}_1\ldots\mathbf{u}_L)$ be the $L\times L$ matrix with columns $\mathbf{u}_1, \ldots, \mathbf{u}_L$. It follows that $\mathbf{x}\to U\mathbf{x}$ is an automorphism of $E_{\mathbf{A}}$. Since $\det(U)$ is in k, the modulus of

this automorphism is 1. That is, the automorphism preserves the Haar measure V. The sets $U^{-1}S_v$, $v|\infty$, and $U^{-1}M_v$, $v\nmid\infty$, have exactly the same properties as S_v and M_v. Thus the successive minima for $U^{-1}\mathscr{S}$ may be defined as before and are clearly equal to the minima $\lambda_1, \lambda_2, \ldots, \lambda_L$ of $\mathscr{S}$. Now, however, the vectors associated with the successive minima of $U^{-1}\mathscr{S}$ may be taken as $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_L$, where

$$\mathbf{e}_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \quad \mathbf{e}_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \ldots \\ 0 \end{pmatrix}, \quad \ldots .$$

Since $\mathscr{S}$ and $U^{-1}\mathscr{S}$ have the same Haar measure we may assume without loss of generality that $\mathbf{u}_l = \mathbf{e}_l$ to begin with.

Next we wish to apply Theorem 4 with $T_v = S_v$ for each infinite place v and $\mu_l = \frac{1}{2}\lambda_l$ for $l = 1, 2, \ldots, L$. We must check that the hypotheses of Theorem 4 are satisfied and it clearly suffices to consider $l = 1$ and those values of l for which $\lambda_{l-1} < \lambda_l$. Therefore we suppose that $\mathbf{x}$ and $\mathbf{y}$ are distinct points in $\mu_l\mathscr{S}$ with $\mathbf{x} - \mathbf{y} \in E$. Since each S_v, $v|\infty$, is convex and symmetric we have

$$\mathbf{x} - \mathbf{y} \in 2\mu_l\mathscr{S} = \lambda_l\mathscr{S}.$$

Also, each of the vectors $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{l-1}$ and $\mathbf{x}-\mathbf{y}$ is in $(\lambda_l - \delta)\mathscr{S}$ for some $\delta > 0$, since $\lambda_{l-1} < \lambda_l$ and $\mathscr{S}$ is open. It follows that

$$\{\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_{l-1}, \mathbf{x}-\mathbf{y}\}$$

cannot be linearly independent over k. As $\mathbf{e}_1, \ldots, \mathbf{e}_{l-1}$ are obviously linearly independent we must have

$$\mathbf{x} - \mathbf{y} = \sum_{j=1}^{l-1} \alpha_j \mathbf{e}_j$$

with $\alpha_j \in k$, $j = 1, 2, \ldots, l-1$. In other words, $x_n = y_n$ for $n = l, l+1, \ldots, L$.

We now apply Theorem 4 to obtain

$$(\mu_1 \mu_2 \ldots \mu_L)^d V(\mathscr{S}) \leqq 1.$$

Since $\mu_l = \frac{1}{2}\lambda_l$, this also proves Theorem 3.

III.4. The classical form of Minkowski's second theorem includes a lower bound for the product of the successive minima. This is also true in the context of adèles provided that we modify our hypotheses on the sets S_v for complex v. Specifically, we must assume that

$$S_v = \alpha S_v \quad \text{for complex } v \text{ and all complex numbers } \alpha \text{ with } |\alpha| = 1. \quad (3.16)$$

For real v we continue to assume that S_v is symmetric in the sense that $S_v = -S_v$, but for complex v our previous requirement that S_v be symmetric will now be replaced by (3.16).

Theorem 6. *Let $\mathscr{S}$ be as in Theorem 3 but with the additional requirement that (3.16) holds. Then the successive minima $\lambda_1, \lambda_2, \ldots, \lambda_L$ satisfy the inequality*

$$\frac{2^{dL}\pi^{sL}}{(L!)^r(2L!)^s|\Delta_k|^{\frac{1}{2}L}} \leqq (\lambda_1\lambda_2\ldots\lambda_L)^d V(\mathscr{S}), \tag{3.17}$$

where r is the number of real places, s is the number of complex places and Δ_k is the discriminant of k.

Proof. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$ and U be as in our proof of Theorem 3. For complex v the set $U^{-1}S_v$ also satisfies (3.16). Thus we may assume, as before, that $\mathbf{u}_l = \mathbf{e}_l$ for $1 \leqq l \leqq L$.

For each infinite place v we define

$$S'_v = \left\{ \mathbf{t} \in E_v : \sum_{l=1}^{L} \lambda_l \|t_l\|_v < 1 \right\}.$$

Since $\lambda \mathbf{e}_l \in S_v$ if $\lambda > \lambda_l$, it follows easily from the convexity and symmetry of S_v that $S'_v \subseteq S_v$. If v is complex the condition of symmetry that we require is exactly (3.16). When we compute the Haar measure of S'_v we find that

$$\beta_v^L(S'_v) = \frac{2^L}{(L!)(\lambda_1\lambda_2\ldots\lambda_L)} \qquad \text{if } v \text{ is real,}$$

$$\beta_v^L(S'_v) = \frac{(4\pi)^L}{(2L!)(\lambda_1\lambda_2\ldots\lambda_L)^2} \qquad \text{if } v \text{ is complex.}$$

If v is a finite place we know that $\mathbf{e}_l \in M_v$ for each l. As M_v is an O_v-module it follows that $(O_v)^L \subseteq M_v$. Of course the Haar measure of $(O_v)^L$ is $|\mathscr{D}_v|_v^{dL/2}$.

Let $\mathscr{S}' \subseteq E_{\mathbb{A}}$ be defined by

$$\mathscr{S}' = \prod_{v|\infty} S'_v \times \prod_{v \nmid \infty} (O_v)^L.$$

Then the Haar measure of $\mathscr{S}'$ is given by

$$V(\mathscr{S}') = \frac{2^{dL}\pi^{sL}}{(L!)^r(2L!)^s(\lambda_1\lambda_2\ldots\lambda_L)^d} \prod_{v \nmid \infty} |\mathscr{D}_v|_v^{dL/2}. \tag{3.18}$$

As $\mathscr{S}' \subseteq \mathscr{S}$ we have the inequality $V(\mathscr{S}') \leqq V(\mathscr{S})$. This observation together with (3.18) and

$$\prod_{v \nmid \infty} |\mathscr{D}_v|_v^{-d} = |\Delta_k|$$

establish the lower bound (3.17).

Cube Slicing

IV.1. Let N be a positive integer. For each place v of the number field k let C_v be the v-adic cube in $(k_v)^N$ defined as follows:

$$C_v=\{\mathbf{x}\in(k_v)^N:\ \max_{1\leqq n\leqq N}\|x_n\|_v<\tfrac{1}{2}\}\quad\text{if } v \text{ is real,}$$

$$C_v=\{\mathbf{x}\in(k_v)^N:\ \max_{1\leqq n\leqq N}\|x_n\|_v<(2\pi)^{-\frac{1}{2}}\}\quad\text{if } v \text{ is complex,}$$

$$C_v=\{\mathbf{x}\in(k_v)^N:\ \max_{1\leqq n\leqq N}\|x_n\|_v\leqq 1\}\quad\text{if } v\nmid\infty.$$

Obviously $C=\prod_v C_v$ is an open subset of $(k_{\mathbb{A}})^N$ having a compact closure. Let χ_C be the characteristic function of C in $(k_{\mathbb{A}})^N$ and χ_{C_v} the characteristic function of C_v in $(k_v)^N$.

Next we suppose that Y is an $N\times L$ matrix with entries in k and $\operatorname{rank}(Y)=L<N$. If $I\subseteq\{1,2,\ldots,N\}$ is a subset with $|I|=L$ elements we write

$$Y_I=(y_{nl}),\qquad n\in I,\ l=1,2,\ldots,L$$

for the corresponding submatrix. We recall that the height of the matrix Y is given by

$$H(Y)=\prod_{v\mid\infty}\Big(\sum_{|I|=L}\|\det Y_I\|_v^2\Big)^{\frac{d_v}{2d}}\cdot\Big(\prod_{v\nmid\infty}\max_{|I|=L}|\det Y_I|_v\Big). \tag{4.1}$$

Eventually we wish to apply Theorem 3 with

$$\mathscr{S}=\prod_v\{\mathbf{x}\in(k_v)^L:\ Y\mathbf{x}\in C_v\} \tag{4.2}$$

so as to obtain an upper bound for the product of the successive minima of $\mathscr{S}$. To accomplish this we require the following cube slicing inequality.

Theorem 7. *If Y is an $N\times L$ matrix over k with $\operatorname{rank}(Y)=L<N$, then the set $\mathscr{S}$ defined by* (4.2) *is a subset of $E_{\mathbb{A}}=(k_{\mathbb{A}})^L$ and has the form* (3.1). *Moreover, we have the inequality*

$$H(Y)^{-d}|\Delta_k|^{-\frac{1}{2}L}\leqq\int_{E_{\mathbb{A}}}\chi_C(Y\mathbf{x})\,dV(\mathbf{x}), \tag{4.3}$$

where Δ_k is the discriminant of k, and therefore

$$H(Y)^{-d}|\Delta_k|^{-\frac{1}{2}L}\leqq V(\mathscr{S}). \tag{4.4}$$

Proof. We begin by giving a lower bound for each of the local volumes

$$\int_{E_v}\chi_{C_v}(Y\mathbf{x})\,d\beta_v^L(\mathbf{x}),$$

where $E_v=(k_v)^L$ and β_v^L is the L-fold product of the Haar measures β_v. If v is real we have

$$\int_{E_v}\chi_{C_v}(Y\mathbf{x})\,d\beta_v^L(\mathbf{x})\geqq\|\det Y^TY\|_v^{-\frac{1}{2}}=\Big(\sum_{|I|=L}\|\det Y_I\|_v^2\Big)^{-\frac{1}{2}} \tag{4.5}$$

by Theorem 1 of Vaaler [11]. If v is complex this becomes

$$\int_{E_v}\chi_{C_v}(Y\mathbf{x})\,d\beta_v^L(\mathbf{x})\geqq\|\det Y^*Y\|_v^{-1}=\Big(\sum_{|I|=L}\|\det Y_I\|_v^2\Big)^{-1}. \tag{4.6}$$

(In (4.5) we have written Y^T for the transpose of the matrix Y and in (4.6) the complex conjugate transpose of Y is denoted by Y^*.) To verify the inequality (4.6) for complex v write $Y=A+iB$ and $\mathbf{x}=\mathbf{s}+i\mathbf{t}$. Then

$$\int_{E_v} \chi_{C_v}(Y\mathbf{x})\, d\beta_v^L(\mathbf{x}) = 2^L \int_{\mathbb{R}^L}\int_{\mathbb{R}^L} \chi_{C_v'}\left(\begin{pmatrix} A & -B \\ B & A \end{pmatrix}\begin{pmatrix}\mathbf{s}\\ \mathbf{t}\end{pmatrix}\right) d\mathbf{s}\, d\mathbf{t} \tag{4.7}$$

where

$$C_v' = \{\boldsymbol{\xi}\in\mathbb{R}^{2L}: (\xi_j)^2+(\xi_{L+j})^2 < (2\pi)^{-1} \quad \text{for } j=1,2,\ldots,L\}.$$

We may now apply Theorem 1 of [11] to the right-hand side of (4.7). We find that the left-hand side of (4.7) is

$$\begin{aligned} &\geqq \left\| \det \begin{pmatrix} A & -B \\ B & A \end{pmatrix}^T \begin{pmatrix} A & -B \\ B & A \end{pmatrix} \right\|_v^{-\frac{1}{2}} \\ &= \|\det (A+iB)^*(A+iB)\|_v^{-1} \\ &= \|\det Y^* Y\|_v^{-1}. \end{aligned}$$

Next we consider the case $v\nmid\infty$. Clearly there exists $J\subseteq\{1,2,\ldots,N\}$ such that $|J|=L$ and

$$\max_{|I|=L} \|\det Y_I\|_v = \|\det Y_J\|_v.$$

Since $\operatorname{rank}(Y)=L$ we also know that $\|\det Y_J\|_v \neq 0$. By making the change of variables $\mathbf{x}\to Y_J^{-1}\mathbf{x}$ we obtain

$$\begin{aligned} \int_{E_v} \chi_{C_v}(Y\mathbf{x})\, d\beta_v^L(\mathbf{x}) &= \operatorname{mod}_{k_v}(\det Y_J^{-1}) \int_{E_v} \chi_{C_v}(Y Y_J^{-1}\mathbf{x})\, d\beta_v^L(\mathbf{x}) \\ &= \|\det Y_J\|_v^{-d_v} \int_{E_v} \chi_{C_v}(W\mathbf{x})\, d\beta_v^L(\mathbf{x}), \end{aligned} \tag{4.8}$$

where $W=YY_J^{-1}$. By permuting the co-ordinates of $\mathbf{x}$ if necessary, we may assume without loss of generality that $J=\{1,2,\ldots,L\}$. Then the submatrix W_J is an $L\times L$ identity matrix and

$$\|\det W_I\|_v \leqq 1$$

holds for every $L\times L$ submatrix W_I. In particular, if we form a submatrix W_I with

$$I=\{1,2,\ldots,l-1,l+1,\ldots,L \text{ and } L+j\},$$

where $1\leqq j\leqq N-L$, we find that

$$\|\det W_I\|_v = \|w_{L+j,l}\|_v \leqq 1.$$

Since $\|\ \|_v$ is ultrametric it follows that $W\mathbf{x}\in C_v$ if and only if $\|x_l\|_v\leqq 1$ for each l. Thus we have

$$\int_{E_v} \chi_{C_v}(W\mathbf{x})\, d\beta_v^L(\mathbf{x}) = \beta_v^L\{(O_v)^L\} = |\mathscr{D}_v|_v^{dL/2}. \tag{4.9}$$

If v is an infinite place then obviously the set

$$S_v=\{\mathbf{x}\in E_v\colon Y\mathbf{x}\in C_v\}$$

is nonempty, open, convex, symmetric and bounded. For finite v it is also clear that

$$M_v=\{\mathbf{x}\in E_v\colon Y\mathbf{x}\in C_v\}$$

is a compact and open O_v-module. For almost all finite v we have

$$\|y_{nl}\|_v=1 \quad \text{for each non-zero entry } y_{nl} \text{ of } Y, \tag{4.10}$$

and

$$\|\det Y_J^{-1}\|_v=1. \tag{4.11}$$

If (4.10) holds then $(O_v)^L\subseteq M_v$ while (4.8), (4.9) and (4.11) imply that $\beta_v^L\{(O_v)^L\}=\beta_v^L(M_v)$. Thus $M_v=(O_v)^L$ for almost all finite v. This shows that $\mathscr{S}$ is a subset of $E_{\mathbf{A}}$ and has the form (3.1). Since $\mathscr{S}$ has a compact closure we have

$$\int_{E_{\mathbf{A}}}\chi_C(Y\mathbf{x})\,dV(\mathbf{x})=\prod_v\int_{E_v}\chi_{C_v}(Y\mathbf{x})\,d\beta_v^L(\mathbf{x}). \tag{4.12}$$

We now use (2.2), our estimates for the factors on the right of (4.12) and

$$\prod_{v\nmid\infty}|\mathscr{D}_v|_v^{-d}=|\Delta_k|$$

to establish the inequality (4.3). This completes our proof.

Linear Forms

V.1. We are now in a position to apply our results on successive minima and cube slicing over the adèles to a discussion of linear forms.

Theorem 8. *Let Y be an $N\times L$ matrix over k with $\operatorname{rank}(Y)=L<N$. Then there exist L linearly independent vectors $\mathbf{u}_1,\mathbf{u}_2,\ldots,\mathbf{u}_L$ in $E=(k)^L$ such that $Y\mathbf{u}_l\in(O_k)^N$ for each l and*

$$\prod_{l=1}^{L}h(Y\mathbf{u}_l)\leqq\left\{\left(\frac{2}{\pi}\right)^s\sqrt{|\Delta_k|}\right\}^{\frac{L}{d}}H(Y), \tag{5.1}$$

where s is the number of complex places of k.

Proof. Let $\mathscr{S}$ be defined by (4.2), let $\lambda>0$ and suppose that $\mathbf{u}$ is a nonzero vector in $(\lambda\mathscr{S})\cap E$. From the definition of $\lambda\mathscr{S}$ we have

$$\max_{1\leqq n\leqq N}\|(Y\mathbf{u})_n\|_v<\tfrac{1}{2}\lambda \quad \text{if } v \text{ is real,} \tag{5.2}$$

$$\max_{1\leqq n\leqq N}\|(Y\mathbf{u})_n\|_v<(2\pi)^{-\frac{1}{2}}\lambda \quad \text{if } v \text{ is complex,} \tag{5.3}$$

$$\max_{1\leqq n\leqq N}\|(Y\mathbf{u})_n\|_v\leqq 1 \quad \text{if } v \text{ is finite.} \tag{5.4}$$

Since $\mathbf{u}\in E$, $Y\mathbf{u}\in(k)^N$ and so the bound (5.4) shows that in fact $Y\mathbf{u}\in(O_k)^N$. By taking a product over all valuations v we find that

$$\prod_v \{\max_{1\leqq n\leqq N} \|(Y\mathbf{u})_n\|_v^{d_v/d}\} = h(Y\mathbf{u}) < (\tfrac{1}{2}\lambda)\left(\frac{2}{\pi}\right)^{s/d}, \tag{5.5}$$

where s is the number of complex valuations.

Let $0<\lambda_1\leqq\lambda_2\leqq\ldots\leqq\lambda_L<\infty$ be the successive minima of $\mathscr{S}$. As before there exists a corresponding set $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$ of linearly independent vectors such that $\mathbf{u}_l\in(\lambda\mathscr{S})\cap E$ for $\lambda>\lambda_l$. From our previous observations we see that $Y\mathbf{u}_l\in(O_k)^N$ and

$$h(Y\mathbf{u}_l)\leqq(\tfrac{1}{2}\lambda_l)\left(\frac{2}{\pi}\right)^{s/d} \tag{5.6}$$

for each l, $1\leqq l\leqq L$. The three inequalities (3.3), (4.4) and (5.6) can now be combined to prove that (5.1) holds.

The bound in Theorem 8 is essentially an analogue for algebraic number fields of Theorem 2 in [11]. In classical versions of Minkowski's theorem on linear forms the matrix Y is square. Here and in [11] the number of linear forms, that is, the number of rows in the matrix Y, is greater than the number of variables. It is this feature that requires a cube slicing inequality to estimate the appropriate volume.

V.2. We consider the system

$$\sum_{n=1}^{N} a_{mn}x_n = 0, \qquad m=1,2,\ldots,M, \tag{5.7}$$

of M linear equations in N variables. We write $A=(a_{mn})$ for the $M\times N$ matrix over k and assume that $\operatorname{rank}(A)=M<N$. Our first version of Siegel's lemma is the following result.

Theorem 9. *Let A be an $M\times N$ matrix over k with $\operatorname{rank}(A)=M<N$. Then there exist $N-M$ linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-M}$ in $(O_k)^N$ which satisfy*

$$A\mathbf{x}_l=\mathbf{0}, \qquad l=1,2,\ldots,N-M, \tag{5.8}$$

and

$$\prod_{l=1}^{N-M} h(\mathbf{x}_l)\leqq\left\{\left(\frac{2}{\pi}\right)^s\sqrt{|\Delta_k|}\right\}^{\frac{N-M}{d}} H(A), \tag{5.9}$$

where s is the number of complex places of k.

Proof. Since $\operatorname{rank}(A)=M$, the solution space for the system (5.7) has dimension $N-M=L$. Let $\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_L$ be a basis for this solution space and write

$$Y=(\mathbf{y}_1\cdots\mathbf{y}_L)$$

for the corresponding $N\times L$ matrix over k. Let $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$ be the vectors determined by applying Theorem 8 to Y. Obviously the vectors $\mathbf{x}_l=Y\mathbf{u}_l$,

$1 \leqq l \leqq L$, are linearly independent and satisfy (5.8). Moreover, we have

$$\prod_{l=1}^{L} h(\mathbf{x}_l) \leqq \left\{ \left(\frac{2}{\pi}\right)^s \sqrt{|\Delta_k|} \right\}^{\frac{L}{d}} H(Y).$$

By the theorem of P. Gordan [6] there is a constant $\gamma \in k$, $\gamma \neq 0$, such that for every subset $I \subseteq \{1, 2, \ldots, N\}$, $|I| = L$,

$$\det Y_I = \gamma(-1)^{\varepsilon(J)} \det A_J,$$

where J is the complement of I and $\varepsilon(J) = \sum_{j \in J} j$. Using the product formula (2.1) we obtain

$$H(Y) = \prod_v |\gamma|_v H(A) = H(A),$$

which completes the proof.

V.3. If $\mathbf{x}$ is a vector in $(k)^N$ we define the *inhomogeneous* height of $\mathbf{x}$ by

$$h_1(\mathbf{x}) = \prod_v \max\{1, |\mathbf{x}|_v\}.$$

Sometimes one may require the results of Theorem 8 or Theorem 9 with h_1 in place of h. Our proofs are easily modified to deal with this situation.

Corollary 10. *Let the hypotheses of Theorem* 8 *hold. Then the vectors* $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_L$ *satisfy*

$$\prod_{l=1}^{L} h_1(Y\mathbf{u}_l) \leqq |\Delta_k|^{\frac{L}{2d}} H(Y). \tag{5.10}$$

Proof. Let $\mathscr{S}$ be defined by (4.2), let $\lambda > 0$ and suppose that $\mathbf{u}$ is a nonzero vector in $(\lambda\mathscr{S}) \cap E$. As in the proof of Theorem 8 we find that (5.5) holds. Now at least one co-ordinate of the vector $Y\mathbf{u}$ is different from zero. Hence (5.5) and the product formula (2.1) imply that

$$1 < (\tfrac{1}{2}\lambda)^{s/d} \left(\frac{2}{\pi}\right)^{s/d}. \tag{5.11}$$

Using (5.11), an easy calculation will establish the bound

$$h_1(Y\mathbf{u}) = \prod_v \max\{1, |Y\mathbf{u}|_v\} < (\tfrac{1}{2}\lambda). \tag{5.12}$$

If we complete the proof of Theorem 8 using (5.12) rather than (5.5) we obtain our corollary.

In our proof of Theorem 9 the bound (5.10) could be used in place of (5.1). The statement of the following corollary is then obvious.

Corollary 11. *Let the hypotheses of Theorem* 9 *hold. Then the vectors* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-M}$ *satisfying* (5.8) *also satisfy*

$$\prod_{l=1}^{N-M} h_1(\mathbf{x}_l) \leqq |\Delta_k|^{\frac{N-M}{2d}} H(A).$$

V.4. Let K be a finite extension of the number field k with $[K:k]=r\geqq 2$. We again consider the system (5.7) of M linear equations in N variables. Now, however, we assume that $A=(a_{mn})$ is an $M\times N$ matrix over K and we wish to determine a basis of solutions in $(k)^N$.

Let F be a number field such that $k\subseteq K\subseteq F$, F is a Galois extension of k with Galois group $G(F/k)$, and F is also a Galois extension of K with Galois group $G(F/K)$. Obviously $G(F/K)$ is a subgroup of $G(F/k)$ having index r. Let $\sigma_1,\sigma_2,\ldots,\sigma_r$ be elements of $G(F/k)$ which form a set of distinct representatives of the cosets of $G(F/K)$ in $G(F/k)$. For each σ_i we write $\sigma_i(A)=(\sigma_i(a_{mn}))$ for the corresponding $M\times N$ matrix. We then assemble these matrices as blocks of the $Mr\times N$ matrix

$$\mathscr{A}=\begin{pmatrix}\sigma_1(A)\\ \sigma_2(A)\\ \cdots\\ \sigma_r(A)\end{pmatrix}. \tag{5.13}$$

Our second version of Siegel's Lemma takes the following form.

Theorem 12. *Let $\mathscr{A}$ be defined by* (5.13) *and suppose that* $\operatorname{rank}(\mathscr{A})=Mr<N$. *Then there exist $N-Mr$ linearly independent vectors $\mathbf{x}_1,\mathbf{x}_2,\ldots,\mathbf{x}_{N-Mr}$ in $(O_k)^N$ which satisfy*

$$A\mathbf{x}_l=\mathbf{0},\qquad l=1,2,\ldots,N-Mr, \tag{5.14}$$

and

$$\prod_{l=1}^{N-Mr}h(\mathbf{x}_l)\leqq\left\{\left(\frac{2}{\pi}\right)^s|\Delta_k|^{\frac{1}{2}}\right\}^{\frac{N-Mr}{d}}H(\mathscr{A}), \tag{5.15}$$

where s is the number of complex places of k.

Proof. Let $\omega_1,\omega_2,\ldots,\omega_r$ be a basis for K over k. Each entry a_{mn} in the matrix A can be written in the form

$$a_{mn}=\sum_{j=1}^{r}\omega_j b_{mn}(j) \tag{5.16}$$

with each $b_{mn}(j)$ in k. Since

$$\sum_{n=1}^{N}a_{mn}x_n=\sum_{j=1}^{r}\omega_j\left\{\sum_{n=1}^{N}b_{mn}(j)x_n\right\}$$

we find that a vector $\mathbf{x}$ in $(k)^N$ is a solution to (5.7) if and only if $\mathbf{x}$ is a solution to the system

$$\sum_{n=1}^{N}b_{mn}(j)x_n=0,\qquad m=1,2,\ldots,M,\ j=1,2,\ldots,r, \tag{5.17}$$

of Mr linear equations in N variables. In view of this equivalence we define $M\times N$ matrices $B(j)$ for each j, $1\leqq j\leqq r$, by

$$B(j)=(b_{mn}(j))_{\substack{m=1,2,\ldots,M,\\ n=1,2,\ldots,N.}}$$

We assemble these matrices in blocks so as to define the $Mr \times N$ matrix

$$\mathscr{B} = \begin{pmatrix} B(1) \\ B(2) \\ \cdots \\ B(r) \end{pmatrix}.$$

We also define an $Mr \times Mr$ matrix Ω by

$$\Omega = (\sigma_i(\omega_j) I_M)_{\substack{i=1,2,\ldots,r, \\ j=1,2,\ldots,r,}}$$

where I_M is an $M \times M$ identity matrix. Here our notation indicates that Ω is organized in blocks of $M \times M$ matrices and $\sigma_i(\omega_j) I_M$ is the (i,j)-block.

If we apply the automorphism σ_i to each side of (5.16) we obtain a system of identities equivalent to the matrix equation

$$\Omega \mathscr{B} = \mathscr{A}. \tag{5.18}$$

As Ω is clearly nonsingular we must have $\operatorname{rank}(\mathscr{B}) = \operatorname{rank}(\mathscr{A}) = Mr < N$. Therefore we can apply Theorem 9 to the system (5.17) determined by $\mathscr{B}$. It follows that there exist $N - Mr$ linearly independent vectors $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-Mr}$ in $(O_k)^N$ satisfying (5.14) and

$$\prod_{l=1}^{N-Mr} h(\mathbf{x}_l) \leqq \left\{ \left(\frac{2}{\pi}\right)^s |\Delta_k|^{\frac{1}{2}} \right\}^{\frac{N-Mr}{d}} H(\mathscr{B}).$$

Finally we note that $H(\mathscr{B}) = H(\mathscr{A})$ as a consequence of (5.18) and the identity (2.5).

In our bound (5.15) the height $H(\mathscr{A})$ must be computed with respect to the Galois extension F. For some applications it may be more convenient to have a bound depending on the height of A or even on the height of the rows of A. We write $A^{(m)}$, $m = 1, 2, \ldots, M$ for the m-th row of A.

Corollary 13. *Let $\mathscr{A}$ satisfy the hypotheses of Theorem* 12. *Then*

$$H(\mathscr{A}) \leqq H(A)^r \leqq \prod_{m=1}^{M} \{H(A^{(m)})\}^r \tag{5.19}$$

and so these bounds may be used on the right-hand side of (5.15).

Proof. By applying (2.6) we find that

$$\begin{aligned} H(\mathscr{A}) &\leqq \prod_{v} \prod_{i=1}^{r} H_v(\sigma_i(A)) \\ &= H(A)^r \\ &\leqq \prod_{m=1}^{M} \{H(A^{(m)})\}^r, \end{aligned}$$

which proves the Corollary.

We note that by using the height $H(A^{(m)})$ of the rows of A in the bound (5.15) we still obtain a sharper (and more general) result than that achieved by Vaaler and van der Poorten [12]. The method of [12] also uses Geometry of Numbers and cube slicing, but only in a Euclidean space.

Finally, we make two remarks concerning ways in which Theorem 12 could be modified. It may happen that the matrix A can be divided into blocks,

$$A=\begin{pmatrix} A_1 \\ A_2 \\ \cdots \\ A_L \end{pmatrix},$$

in such a way that A_l is an $M_l \times N$ matrix over the number field K_l with $[K_l:k]=r_l$, $l=1,2,\ldots,L$. Now we let F be a number field which is a Galois extension of k and a Galois extension of each K_l, $l=1,2,\ldots,L$. For each A_l we can define an $M_l r_l \times N$ matrix $\mathscr{A}_l$ as in (5.13), and we can assemble these matrices into a $\left(\sum_{l=1}^{L} M_l r_l\right) \times N$ matrix $\mathscr{Z}$. With only obvious modifications in our proof of Theorem 12, we obtain the following more general form of Siegel's Lemma.

Theorem 14. *Suppose that* $t=\operatorname{rank}(\mathscr{Z})=\sum_{l=1}^{L} M_l r_l < N$. *Then there exist* $N-t$ *linearly independent vectors* $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{N-t}$ *in* $(O_k)^N$ *such that*

$$A\mathbf{x}_l=\mathbf{0}, \qquad l=1,2,\ldots,N-t$$

and

$$\prod_{l=1}^{N-t} h(\mathbf{x}_l) \leqq \left\{\left(\frac{2}{\pi}\right)^s |\Delta_k|^{\frac{1}{2}}\right\}^{\frac{N-t}{d}} H(\mathscr{Z}), \tag{5.20}$$

where s *is the number of complex places of* k. *Moreover, on the right-hand side of* (5.20) *we can use*

$$H(\mathscr{Z}) \leqq \prod_{l=1}^{L} \{H(A_l)\}^{r_l}.$$

If $\mathscr{A}$ is defined by (5.13) and $\operatorname{rank}(\mathscr{A}) < Mr$ then Theorem 12 can not be applied. However, we can still establish (5.18) and deduce that $\operatorname{rank}(\mathscr{B}) = \operatorname{rank}(\mathscr{A})$. If $\mathscr{B}'$ is an $M' \times N$ submatrix of $\mathscr{B}$ such that $\operatorname{rank}(\mathscr{B}') = \operatorname{rank}(\mathscr{B}) = M'$ then we may apply Theorem 9 to obtain a bound involving $H(\mathscr{B}')$. Another approach if $\operatorname{rank}(\mathscr{A}) < Mr$ is to divide the matrix A into blocks so that Theorem 14 applies. Of course the idiosyncrasies of a particular application may suggest other ways in which Theorem 12 can be usefully modified.

References

1. Bombieri, E.: On the Thue-Siegel theorem. Acta Mathematica **148**, 255–296 (1982)

2. Bombieri, E., Mueller, J.: On effective measures of irrationality for $\sqrt[r]{\frac{a}{b}}$ and related numbers. J. reine angew. Math. (to appear)

3. Davenport, H.: Minkowski's Inequality for the minima associated with a convex body. Q. J. Math. **10**, 119–121 (1939) (= *Coll. Works*, I, 88–90)
4. Estermann, T.: Note on a theorem of Minkowski. J. London Math. Soc. **21**, 179–182 (1946)
5. Fischer, E.: Über den Hadamardschen Determinantensatz. Arch. Math. (Basel) **13**, 32–40 (1908)
6. Gordan, P.: Über den größten gemeinsamen Factor. Math. Annalen **7**, 443–448 (1873)
7. Siegel, C.L.: Über einige Anwendungen diophantischer Approximationen. Abh. der Preuß. Akad. der Wissenschaften. Phys.-math. Kl. 1929, Nr. 1 (= *Ges. Abh.*, I, 209–266)
8. Skolem, T.: Diophantische Gleichungen. Ergebnisse der Mathematik, Vol. 5. Berlin: Springer 1938
9. Smith, H.J.S.: On systems of linear indeterminate equations and congruences. Phil. Trans. Roy. Soc. London **151**, 293–326 (1861) (= *Coll. Math. Papers*, I, 367–409)
10. Thue, A.: Über Annäherungswerte algebraischer Zahlen. J. reine angew. Math. **135**, 284–305 (1909)
11. Vaaler, J.D.: A Geometric Inequality with Applications to Linear Forms. Pacific J. Math. **83**, 543–553 (1979)
12. Vaaler, J.D., van der Poorten, A.J.: Bounds for solutions of systems of linear equations. Bull. Australian Math. Soc. **25**, 125–132 (1982)
13. Weil, A.: Arithmetic on algebraic varieties. Annals of Math. **53**, 412–444 (1951) (= *Coll. Papers*, I, 454–486)
14. Weil, A.: Basic number theory. Berlin-Heidelberg-New York: Springer 1973

Oblatum 16-II-1983

Invent. math. 73, 33-49 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

The Birkhoff-Lewis Fixed Point Theorem and a Conjecture of V.I. Arnold

C.C. Conley[1] and E. Zehnder[2]

[1] Department of Mathematics, University of Wisconsin-Madison, 480 Lincoln Drive, Madison, WI 53706, USA

[2] Ruhr-Universität, Bochum, Abteilung Mathematik, Universitätsstr. 150 D-4630 Bochum, Federal Republic of Germany

1. Introduction, Results

The aim of this note is to prove two fixed point theorems for symplectic maps which are generated by timedependent globally Hamiltonian vectorfields, which however are not assumed to be close to the identity map. In the case that ψ is a symplectic diffeomorphism on the torus $T^{2n}=\mathbb{R}^{2n}/\mathbb{Z}^{2n}$ this requires that $\psi=\varphi^1$, where φ^t is the flow satisfying

$$\frac{d}{dt}\varphi^t(x)=J\nabla h(t,\varphi^t(x)), \qquad \varphi^0=\mathrm{id},$$

the function $h\in C^2(\mathbb{R}\times\mathbb{R}^{2n})$ being periodic in all the variables of period 1. The skew-symmetric matrix $J\in\mathfrak{L}(\mathbb{R}^{2n})$ defined by $J=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}$, 1 being the unit matrix on $\mathbb{R}^n$, is the standard symplectic structure on $\mathbb{R}^{2n}$. Clearly, a periodic solution on T^{2n} having period 1 of the Hamiltonian equation

$$\dot{x}=J\nabla h(t,x), \qquad x\in\mathbb{R}^{2n}, \tag{1}$$

gives rise to a fixed point of ψ, and the problem is to find 1-periodic solutions of the Hamiltonian system, which is periodic in time of period 1. The first result is as follows.

Theorem 1. *The Hamiltonian vectorfield* (1) *on* T^{2n}, *with the function* $h(t,x)\in C^2(\mathbb{R}\times\mathbb{R}^{2n})$ *being periodic of period* 1 *possesses at least* $2n+1$ *periodic solutions of period* 1.

The periodic solutions found by the theorem are contractible loops on T^{2n}, i.e. are given as periodic functions on $\mathbb{R}^{2n}$. One expects more periodic solutions, if all the periodic solutions are known to be nondegenerate. Here we call a 1-periodic solution nondegenerate, if it has no Floquet-multiplier equal to 1. Recall that $\lambda\in\mathbb{C}$ is a Floquet-multiplier of a periodic solution $x(t)=x(t+1)$, if λ

is an eigenvalue of $d\varphi^1(x(0))$, where φ^t is the flow of the corresponding timedependent vectorfield. Indeed, the following statement holds true:

Theorem 2. *Assume that all the periodic solutions having period* 1 *of the system* (1) *are nondegenerate, then there are at least* 2^{2n} *of them.*

From these existence statements for periodic solutions one deduces immediately the following Corollary for the symplectic map ψ. We call a fixed point $x=\psi(x)$ nondegenerate if 1 is not an eigenvalue of $d\psi(x)$.

Corollary 1. *Every symplectic* C^1*-diffeomorphism* ψ *on the torus* $T^{2n}=\mathbb{R}^{2n}/Z^{2n}$, *which is generated by a globally Hamiltonian vectorfield, possesses at least* $2n+1$ *fixed points. If, moreover, all the fixed points of* ψ *are nondegenerate, then there are at least* 2^{2n} *of them.*

The symplectic diffeomorphism which meet the assumption of the Corollary can be characterized as follows. If (M,ω) is any compact, symplectic and smooth manifold, we denote by $\mathrm{Diff}^\infty(M,\omega)$ the topological group of symplectic C^∞-diffeomorphisms ψ, i.e. $\psi^*\omega=\omega$. Let $\mathrm{Diff}_0^\infty(M,\omega)$ be the identity component in $\mathrm{Diff}^\infty(M,\omega)$, which can be shown to be the identity component by smooth arcs in $\mathrm{Diff}^\infty(M,\omega)$. It has been proved by A. Banyaga [4], that the commutator-subgroup of $\mathrm{Diff}_0^\infty(M,\omega)$ consists precisely of those symplectic diffeomorphisms, which are generated by globally Hamiltonian vectorfields on M, and hence agrees with the subgroup of symplectic diffeomorphisms having vanishing so-called Calabi-invariant.

As a special case we consider a measure preserving diffeomorphism of T^2, which is homologeous to the identity map on T^2 and hence is, on the covering space $\mathbb{R}^2$, of the form

$$\psi\colon\ x\mapsto x+f(x),\qquad x\in\mathbb{R}^2, \tag{2}$$

with f being periodic. As observed by V.I. Arnold, see [2], this map ψ is generated by a globally Hamiltonian vectorfield on T^2 if and only if the meanvalue of f over the torus vanishes, i.e. $[f]=0$. This fact will be proved in the appendix. We therefore conclude from Theorem 1 the following result, which was conjectured by V.I. Arnold in [2] and [3].

Corollary 2. *Every measure preserving* C^1*-diffeomorphism of* T^2 *which is of the form* (2) *with* $[f]=0$ *has at least* 3 *fixed points.*

The condition $[f]=0$ is clearly necessary in order to guarantee a fixed point, as the translation map $x\mapsto x+c$ shows, which has no fixed points on T^2, if $c\notin Z^2$.

We point out that it is not assumed that the symplectic maps considered are C^1-close to the identity map. Indeed under this additional assumption the above fixed points can easily been found as critical points of a so-called generating function, which is defined on the torus. The idea of relating fixed points of symplectic maps to critical points of a related function defined on the corresponding manifold goes back to H. Poincaré [11]. It has been exploited by A. Banyaga [5], J. Moser [8] and A. Weinstein [12] in order to guarantee fixed points for symplectic maps, which are however assumed to be C^1-close to the identity map.

The second result is related to the Birkhoff-Lewis fixed point theorem, for which we refer to J. Moser [9]. The problem can be reduced to an exact symplectic diffeomorphism on $T^n \times D$, $D \subset \mathbb{R}^n$ being a disc, which is assumed to be close to an integrable map. Instead we would like to replace such a smallness condition by a condition at the boundary $T^n \times \partial D$ only. To be precise we consider on the symplectic manifold $T^n \times \mathbb{R}^n$ the time-dependent globally Hamiltonian vectorfield given by

$$h(t, x, y) \in C^2(R \times T^n \times \mathbb{R}^n),$$

and periodic in t and x of period 1. We suppose that

$$h(t, x, y) = \tfrac{1}{2}\langle y, by\rangle + \langle a, y\rangle \tag{3}$$

if $|y| \geqq C > 0$, where $b \in \mathfrak{L}(\mathbb{R}^n)$ is a time-independent, symmetric and non-singular matrix, and where $a \in \mathbb{R}^n$ are constants. Set

$$T^n \times D = \{(x, y) \in T^n \times \mathbb{R}^n \mid |y| < C\}.$$

Under these assumptions, the following statement holds true.

Theorem 3. *The Hamiltonian vectorfield on $T^n \times \mathbb{R}^n$ admits at least $n+1$ periodic solutions, of period 1, which are contained in $T^n \times D$.*

Again the periodic solutions found are special: their projections onto T^n are contractible loops on T^n. The symplectic diffeomorphism generated by the above time-dependent Hamiltonian vectorfield admits then at least $n+1$ fixed points in $T^n \times D$.

2. Idea of the Proof

The proof of these theorems is based on a variational principle for which the periodic orbits are critical points. This variational problem differs from that customarily used in mechanics, which in the example of the geodesic flow on a manifold M is the length integral or the energy integral. In contrast, the variational principle used here is defined in the loop space over a symplectic space, in the above example, over T^*M. In our problem of the torus $T^{2n} = \mathbb{R}^{2n}/Z^{2n}$ we consider, on the covering space $\mathbb{R}^{2n}$, the action functional, defined on periodic functions $x(0) = x(1)$:

$$f(x) = \int_0^1 \{\tfrac{1}{2}\langle \dot{x}, Jx\rangle - h(t, x(t))\}\, dt,$$

whose Euler-equations are indeed the Hamiltonian Eqs. (1). This functional is neither bounded from above nor from below. That it still can be used effectively for existence proofs was first shown by P. Rabinowitz and subsequently used by many authors. Since h is periodic, following the ideas of [1], it will be shown by a Lyapunov-Schmidt reduction that the required critical points of f are in one-to-one correspondence with the critical points of a function g, which

approximates f but is defined on the finite dimensional manifold $M=T^{2n}\times\mathbb{R}^N\times\mathbb{R}^N$ for some large N. The critical points of g are then found as the rest points of the gradient flow ∇g on M. From the fact, that h and its derivatives are uniformly bounded it follows, that the set of bounded solutions of this gradient flow is compact and contained in the compact set $B:=T^{2n}\times D\times D$, where D is a disc in $\mathbb{R}^N$. Moreover $B^-:=T^{2n}\times\partial D\times D$ is the exit set and $B^+=T^{2n}\times D\times\partial D$ is the entrance set, so that B is an isolating block in the sense of [6]. The proof now follows from two general statements for flows, which are not necessarily gradient flows. First consider any continuous flow which admits the above very special isolating block B, with exit set B^- and entrance set B^+. Then the invariant set S of the flow contained in B carries cohomology which it obtains from the torus T^{2n}. In fact Theorem 4 states that

$$l(S)\geqq l(B)=l(T^{2n})=2n+1,$$

where $l(X)$ denotes the cup long of a compact space X. The second statement concerns Morse-decompositions. If $\{M_1,\ldots,M_k\}$ is an ordered Morse-decomposition of a compact, isolated invariant set S of a continuous flow, then

$$l(S)\leqq\sum_{j=1}^{k} l(M_j),$$

by Theorem 5. If, in addition, the flow on S is gradientlike with finitely many rest points then these rest points consititute a Morse-decomposition of S. In this case $l(M_j)=1$ and we obtain the estimate

$$l(S)\leqq\sum_{j=1}^{k} 1=\#\{\text{rest points}\}.$$

We conclude that the gradient flow ∇g possesses at least $2n+1$ rest points and theorem 1 follows. Theorem 3 is proved similarly. Theorem 2 is a simple application of the Morse-theory as developed in [6] and [7] to the compact set S of all bounded solutions of ∇g on M.

3. The Variational Principle and the Reduction

We shall look for special periodic solutions of the Hamiltonian system (1) on $T^{2n}=\mathbb{R}^{2n}/Z^{2n}$, namely for those whose orbits are contractible. In the covering space $\mathbb{R}^{2n}$ of the torus these solutions are described by periodic functions $t\mapsto x(t)\in\mathbb{R}^{2n}$, $x(0)=x(1)$. The required periodic solutions are the critical points of the functional

$$f(x):=\int_0^1\{\tfrac{1}{2}\langle\dot{x},Jx\rangle-h(t,x(t))\}\,dt, \tag{4}$$

defined on the space of periodic curves in $\mathbb{R}^{2n}$, i.e. $x(0)=x(1)$. Indeed one verifies immediately that

$$\nabla f(x)=-J\dot{x}-\nabla h(t,x). \tag{5}$$

To be precise we introduce the Hilbert space $H=L_2((0,1);\mathbb{R}^{2n})$. Define in H the linear operator A: $\mathrm{dom}(A)\subset H\to H$ by setting

$$\mathrm{dom}(A)=\{u\in H^1([0,1];\mathbb{R}^{2n})\,|\,u(0)=u(1)\}$$

and $\mathrm{Au}=-J\dot{u}$ if $u\in\mathrm{dom}(A)$. The space H^1 is the Sobolev space of absolutely continuous functions whose first derivative is in L^2. The continuous operator $F: H\to H$ is defined by $F(u)(t):=\nabla h(t,u(t))$, $u\in H$. Its potential $\Phi(u)$ is given by $\Phi(u):=\int_0^1 h(t,u(t))\,dt$, so that $F(u)=\nabla\Phi(u)$.

Since $J^2=-1$ we can write the equation (1) in the form $-J\dot{x}=\nabla h(t,x)$ and one sees that every solution $u\in\mathrm{dom}(A)$ of the equation

$$\mathrm{Au}=F(u) \tag{6}$$

defines (by periodic continuation) a classical 1-periodic solution of (1). Conversely, every 1-periodic solution on T^{2n} of (1), which is contractible on T^{2n} defines (by restriction) a solution u of the Eq. (6). With these notations the functional f defined by (4) becomes

$$f(u)=\tfrac{1}{2}\langle \mathrm{Au},u\rangle-\Phi(u), \tag{7}$$

for $u\in\mathrm{dom}(A)$. We look for critical points of f.

Since h is periodic, there is a constant $\alpha>0$ such that

$$|h''(t,x)|\leqq\alpha \tag{8}$$

for all $(t,x)\in\mathbb{R}\times\mathbb{R}^{2n}$, where $'$ stands for the derivative in the x-variable. We shall use this estimate in order to reduce the problem of finding critical points of the functional f on $\mathrm{dom}(A)$ to the problem of finding critical points of a related functional, which is defined on a finite dimensional subspace of the Hilbert space H.

First observe that the operator A is selfadjoint, $A^*=A$. It has closed range and a compact resolvent. The spectrum of A, $\sigma(A)$, is a pure point spectrum and $\sigma(A)=2\pi\mathbb{Z}$. Every eigenvalue $\lambda\in\sigma(A)$ has multiplicity $2n$ and the eigenspace $E(\lambda):=\ker(\lambda-A)$ is spanned by the orthogonal basis given by the loops:

$$t\mapsto e^{t\lambda J}e_k=(\cos\lambda t)\,e_k+(\sin\lambda t)\,Je_k,$$

$k=1,2,\ldots,2n$, where $\{e_k\,|\,1\leqq k\leqq 2n\}$ is the standard basis in $\mathbb{R}^{2n}$. In particular $\ker(A)=\mathbb{R}^{2n}$, the kernel of A consist precisely of the constant loops in $\mathbb{R}^{2n}$. Denoting by $\{E_\lambda\,|\,\lambda\in\mathbb{R}\}$ the spectral resolution of A we define the orthogonal projection $P\in\mathfrak{L}(H)$ by

$$P=\int_{-\beta}^{\beta} dE_\lambda,\quad\text{with}\quad \beta\geqq 2\alpha,\quad(\alpha\text{ as in (8)}),$$

where $\beta\notin 2\pi\mathbb{Z}$. Let $P^\perp=1-P$ and set $Z=P(H)$ and $Y=P^\perp(H)$. Then $H=Z\oplus Y$ and $\dim Z<\infty$. With these notations the equation $\mathrm{Au}-F(u)=0$, for $u\in\mathrm{dom}(A)$

is equivalent to the pair of equations

$$\begin{aligned} APu - PF(u) &= 0 \\ AP^{\perp}u - P^{\perp}F(u) &= 0. \end{aligned} \tag{9}$$

Now writing $u = Pu + P^{\perp}u = z + y \in Z \oplus Y$ we shall solve, for fixed $z \in Z$, the second equation of (9) which becomes $Ay - P^{\perp}F(z+y) = 0$. With $A_0 := A|Y$ this equation is equivalent to

$$y = A_0^{-1} P^{\perp} F(z+y). \tag{10}$$

Observe that $|A_0^{-1}| \leqq \beta^{-1}$ and $|P^{\perp}| = 1$. Also, from (8) we conclude that $|F(u) - F(v)| \leqq \alpha |u - v|$ for all $u, v \in H$. Consequently, in view of $\beta \geqq 2\alpha$, the right hand side of (10) is a contraction operator in H having contraction constant 1/2. We conclude, for fixed $z \in Z$, that the equation (10) has a unique solution $y = v(z) \in Y$. Since $(A_0^{-1} y)(t) = \int_0^t J y(s)\, ds$, we have $A_0^{-1}(Y) \subset H^1$ and therefore $v(z) \in \mathrm{dom}(A)$. Moreover, the map $z \mapsto v(z)$ from Z into Y is Lipschitz-continuous. In fact we have

$$|v(z_1) - v(z_2)| \leqq \tfrac{1}{2}\{|z_1 - z_2| + |v(z_1) - v(z_2)|\}.$$

Setting

$$u(z) = z + v(z)$$

we now have to solve the first equation of (9), namely $Az - PF(u(z)) = 0$, which in view of (10) is equivalent to the equation $\mathrm{A}u(z) - F(u(z)) = 0$. One verifies readily that

$$\nabla g(z) = Az - PF(u(z)) \quad \text{with} \quad g(z) := f(u(z)). \tag{11}$$

It remains to find critical points of the function g, which is defined on the finite dimensional space Z.

The following observation is crucial. Since h is periodic we conclude by uniqueness that $v(z+j) = v(z)$ for every $j \in \mathbb{Z}^{2n}$ and for every $z \in Z$. Therefore $u(z+j) = u(z) + j$ and consequently

$$\nabla g(z+j) = \nabla g(z), \qquad j \in \mathbb{Z}^{2n}, \tag{12}$$

for all $z \in Z$. If $z \in Z$, we set $z = x + \xi$, with $x = [z]$ being the mean value of z. Hence $x \in \mathrm{Ker}(A)$ and $\xi \in \mathrm{Ker}(A)^{\perp} \cap Z$. Writing $z = (x, \xi)$ we conclude from (12) that $\nabla g(z) = \nabla g(x, \xi)$ is a vectorfield on $(x, \xi) \in T^{2n} \times \mathbb{R}^{2M}$, where $T^{2n} = \mathbb{R}^{2n}/\mathbb{Z}^{2n}$. Summarizing we have proved

Lemma 1. *The rest points of the Lipschitz-continuous vectorfield $\nabla g(z) = \nabla g(x, \xi)$ on $T^{2n} \times \mathbb{R}^{2M}$ are in one-to-one correspondence with those periodic solutions having period 1 of the Hamiltonian Eq. (1) on T^{2n} which are contractible.*

In order to find the rest points of ∇g we study the gradient flow $\dfrac{d}{ds} z = \nabla g(z)$ on $(x, \xi) \in T^{2n} \times \mathbb{R}^{2M}$, which, explicitely, is given by

$$\frac{d}{ds}x = -Q_0 F(u(z)) \tag{13}$$
$$\frac{d}{ds}\xi = A\xi - QF(u(z)),$$

where Q_0 is the orthogonal projection onto the constants, i.e. onto the kernel of the operator A, and where Q is the orthogonal projection onto the complement of $\ker(A)$ in Z. We denote the splitting $\xi = (\xi_+, \xi_-) \in \mathbb{R}^M \times \mathbb{R}^M$ of QAQ into the positive and negative part by:

$$A\xi = \begin{pmatrix} A_+ & 0 \\ 0 & A_- \end{pmatrix} \begin{pmatrix} \xi_+ \\ \xi_- \end{pmatrix} \tag{14}$$
$$\langle A_+ \xi_+, \xi_+ \rangle \geqq 2\pi |\xi_+|^2$$
$$\langle A_- \xi_-, \xi_- \rangle \leqq -2\pi |\xi_-|^2.$$

Since h is uniformly bounded, there is a constant $K > 0$ such that $|F(u(z))| \leqq K$ for every $z \in Z$, and we conclude from (13) and (14), that with $\varepsilon = (2\pi - 1) K^2 > 0$ one has

$$\frac{d}{ds}|\xi_+|^2 \geqq \varepsilon \qquad \text{if } |\xi_+| \geqq K \tag{15}$$
$$\frac{d}{ds}|\xi_-|^2 \leqq -\varepsilon \qquad \text{if } |\xi_-| \geqq K.$$

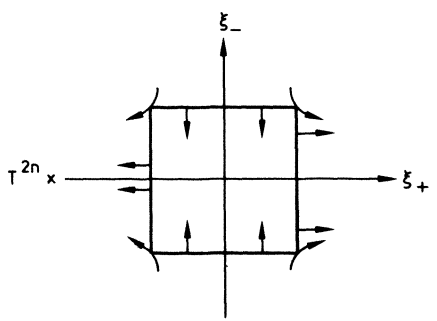
In fact,

$$\frac{d}{ds}\frac{1}{2}|\xi_+|^2 = \langle \xi_+, A_+ \xi_+ - Q_+ F(u(z)) \rangle = \langle \xi_+, A_+ \xi_+ \rangle - \langle \xi_+, Q_+ F(u(z)) \rangle$$
$$\geqq 2\pi |\xi_+|^2 - |\xi_+| \cdot K \geqq \varepsilon.$$

Similarly,

$$\frac{d}{ds}\frac{1}{2}|\xi_-|^2 \leqq -2\pi |\xi_-|^2 + |\xi_-| \cdot K \leqq -\varepsilon,$$

and (15) follows. Clearly all the rest points of ∇g are contained in the compact set $B = T^{2n} \times D_1 \times D_2$, where D_1 and D_2 are the discs of radius K, $D_1 = \{\xi_+ \in \mathbb{R}^M \mid |\xi_+| \leqq K\}$ and $D_2 = \{\xi_- \in \mathbb{R}^M \mid |\xi_-| \leqq K\}$. Moreover $B^- := T^{2n} \times \partial D_1 \times D_2$ is the exit set of B and $B^+ = T^{2n} \times D_1 \times \partial D_2$ is the entrance set, so that B is an isolating block in the sense of [6].



Summarising we have proved:

Lemma 2. *The compact set* $B=T^{2n}\times D_1\times D_2$ *is an isolating block for the flow of* ∇g *with exit set* $B^-=T^{2n}\times\partial D_1\times D_2$ *and with entrance set* $B^+=T^{2n}\times D_1\times\partial D_2$.

We shall prove that every continuous flow which admits the above very special isolating block (B,B^-,B^+) contains at least $2n+1$ rest points in B, provided the flow on the invariant set in B is gradient like.

4. Two Statements for Flows

We consider, more generally, on $M=T^m\times\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}$ with coordinates (x,y_1,y_2) a continuous flow, which is defined in an open neighborhood of the compact set $B=T^m\times D_1\times D_2$, where D_1 and D_2 are discs, $D_1=\{y_1\in\mathbb{R}^{N_1}\,|\,|y_1|\leqq K_1\}$ and $D_2=\{y_2\in\mathbb{R}^{N_2}\,|\,|y_2|\leqq K_2\}$. If $\gamma\in M$ and if $\varphi^t(\gamma)$ is the orbit of the flow through $\gamma=\varphi^0(\gamma)$ we shall write $\varphi^t(\gamma)=\gamma\cdot t$. For an interval $J\subset\mathbb{R}$ we set $\gamma\cdot J=\{\gamma\cdot t\,|\,t\in J\}$. The invariant set contained in B is defined to be

$$S=\{\gamma\in B\,|\,\gamma\cdot\mathbb{R}\subset B\}.$$

Recall the

Definition. Let $\check{H}^*(X)$ be the Alexander cohomology of a compact topological space X with real coefficients. Then the cup long of X is defined as

$$l(X)=1+\sup\{k\in\mathbb{N}\,|\,\exists\alpha_1,\ldots,\alpha_k\in H^*(X)\setminus 1 \quad\text{with}\quad \alpha_1\cup\alpha_2\cup\ldots\cup\alpha_k\neq 0\},$$

and $l(X)=1$ if no such class exists.

Theorem 4. *Assume a continuous flow on* $T^m\times\mathbb{R}^{N_1}\times\mathbb{R}^{N_2}$ *admits the isolating block* $B=T^m\times D_1\times D_2$ *with exit set* $B^-:=T^m\times\partial D_1\times D_2$ *and with entrance set* $B^+:=T^m\times D_1\times\partial D_2$. *Let* S *be the invariant set in* B, *then:*

$$l(S)\geqq l(B)=l(T^m)=m+1.$$

We need a Lemma and define by means of the flow the compact sets

$$\begin{aligned} A^+&:=\{\gamma\in B\,|\,\gamma\cdot\mathbb{R}^+\subset B\}\\ A^-&:=\{\gamma\in B\,|\,\gamma\cdot\mathbb{R}^-\subset B\}. \end{aligned}\tag{16}$$

It then follows, that

Lemma 3 (Wazewski's Principle). *Assume* B *is an isolating block with exit set* B^- *and entrance set* B^+. *Then* B^+ *is a strong deformation retract of* $B\setminus A^-$ *and* B^- *is a strong deformation retract of* $B\setminus A^+$.

Proof. The proof follows immediately from the definition of an isolating block. In fact, in order to prove that B^- is a strong deformation retract of $B\setminus A^+$ define the continuous function $\tau^+: B\setminus A^+\to\mathbb{R}$ by setting $\tau^+(x)=\sup\{t\,|\,x\cdot[0,t]$

$\subset B\}$. Then $\tau^+(x)=0$ if and only if $x\in B^-$. The deformation retraction F: $(B\setminus A^+)\times[0,1]\to B\setminus A^+$ is simply given by $F(x,s)=x\cdot\{s\tau^+(x)\}$. The other part of the lemma is proved similarly. □

Proof of Theorem 4. Note that $\check{H}^*(B)=\check{H}^*(T^m)$, since $B=T^m\times D_1\times D_2$. Let $\check{x}_1,\ldots,\check{x}_m$ be the classes in $\check{H}^*(B)$ which correspond to the de Rham classes dx_i on T^m. Then $\check{\vartheta}:=\check{x}_1\cup\check{x}_2\cup\ldots\cup\check{x}_m\neq 0$ in $\check{H}^*(B)$ since $dx_1\wedge\ldots\wedge dx_m$ is a volume form on T^m. We shall show that $\check{\vartheta}$ maps nontrivially to $\check{H}^*(S)$ under the inclusion induced map i^*: $\check{H}^*(B)\to\check{H}^*(S)$. The statement $l(S)\geqq m+1$ then follows.

To see this consider the following two diagrams, in which H^* and H_* denote the singular cohomology and homology respectively with real coefficients. Observe that $\partial B=B^+\cup B^-$, and

$$(B\setminus A^-)\cup(B\setminus A^+)=B\setminus(A^-\cap A^+)=B\setminus S.$$

Observe also, that the compact set S is contained in the interior of B.

$$\begin{array}{ccccccccc}
H^*(B,B^+) & \otimes & H^*(B,B^-) & \xrightarrow{\cup} & H^*(B,\partial B) & \cong & H_*(B,\partial B) & \xrightarrow{\check{D}} & \check{H}^*(B) \\
\uparrow{\scriptstyle i_1^*} & & \uparrow{\scriptstyle i_2^*} & & \uparrow{\scriptstyle j^*} & & \downarrow{\scriptstyle j_*} & & \downarrow{\scriptstyle i^*} \\
H^*(B,B\setminus A^-) & \otimes & H^*(B,B\setminus A^+) & \xrightarrow{\cup} & H^*(B,B\setminus S) & \cong & H_*(B,B\setminus S) & \xrightarrow{\check{D}} & \check{H}^*(S).
\end{array}$$

The vertical maps are inclusion induced. The maps i_1^* and i_2^* are isomorphisms, by Lemma 3. The isomorphism $\check{D}$ is the Alexander duality map: $\check{D}$: $H_j(B,\partial B)\to\check{H}^{d-j}(B)$, with $d=m+N_1+N_2$.

We shall find cohomology in $H^*(S)$ by means of the following argument. Let $\alpha\in H^*(B,B^+)$ and $\beta\in H^*(B,B^-)$ such that $\alpha\cup\beta\neq 0$ in $H^*(B,\partial B)$. Then $\alpha=i_1^*(\alpha^*)$ for some $\alpha^*\in H^*(B,B\setminus A^-)$ and $\beta=i_2^*(\beta^*)$ for some β^* in $H^*(B,B\setminus A^+)$, since i_1^* and i_2^* are isomorphisms. Hence $\alpha\cup\beta=j^*(\alpha^*\cup\beta^*)$. If now $\alpha\cup\beta\in H^j(B,\partial B)$ then, going to the dual spaces (in the sense of vector spaces), there is a $\mu\in H_j(B,\partial B)$ such that $j_*(\mu)=:\mu^*\neq 0$ in $H_j(B,B\setminus S)$. Consequently, the Alexander dual $\check{D}(\mu)\in\check{H}^{d-j}(B)$ is maped onto $i^*(\check{D}(\mu))=\check{D}(\mu^*)\neq 0$ in $\check{H}^{d-j}(S)$. If, in particular, $j=N_1+N_2$, then $d-j=m$ and $\check{D}(\mu)\in\check{H}^m(B)\cong\check{H}^m(T^m)$ which is generated by $\check{\vartheta}=\check{x}_1\cup\ldots\cup\check{x}_m$. Consequently $i^*(\check{\vartheta})=i^*(\check{x}_1\cup\ldots\cup\check{x}_m)\neq 0$ in $\check{H}^m(S)$ and so $l(S)\geqq l(T^m)=m+1$ as claimed in the theorem.

In order to carry out this argument concretely we first recall that

$$H^*(B,B^+)\cong H^*(T^m)\otimes H^*(D_2,\partial D_2)\otimes H^*(D_1) \quad \text{and}$$
$$H^*(B,B^-)\cong H^*(T^m)\otimes H^*(D_2)\otimes H^*(D_1,\partial D_1),$$

also

$$\begin{aligned}H^*(B,\partial B)&\cong H^*(T^m)\otimes H^*(D_2,\partial D_2)\otimes H^*(D_1,\partial D_1)\\&\cong H^*(T^m)\otimes H^*(D_1\times D_2,(D_1\times\partial D_2)\cup(\partial D_1\times D_2)).\end{aligned}$$

We choose now α to be the image $\hat{\xi}$ in $H^*(B,B^+)$ of the generator ξ of $H^{N_1}(D_2,\partial D_2)\cong H^{N_1}(S^{N_1},*)$, and we choose β to be the image $\hat{\eta}$ in $H^*(B,B^-)$ of the generator η of $H^{N_2}(D_1,\partial D_1)\cong H^{N_2}(S^{N_2},*)$. Then $\alpha\cup\beta$ is equal to the image

$\hat{\zeta}$ in $H^*(B, \partial B)$ of the generator ζ of $H^{N_1+N_2}((D_1, \partial D_1) \times (D_2, \partial D_2))$, and so $\alpha \cup \beta \neq 0$ in $H^{N_1+N_2}(B, \partial B)$, so the theorem follows, by the above argument. □

Definition. Let S be a compact invariant set of a continuous flow. A Morse decomposition of S is a finite collection $\{M_p\}_{p \in P}$ of disjoint, compact and invariant subsets of S, which can be ordered, say $(M_1, M_2, \ldots, M_k)$, $k = |P|$, so that the following property holds true. If

$$\gamma \in S \setminus \bigcup_{p \in P} M_p,$$

then there is a pair of indices $i < j$ such that the positive $(t \to +\infty)$ and negative $(t \to -\infty)$ limit sets $\omega(\gamma)$ and $\omega^*(\gamma)$ of γ satisfy:

$$\omega(\gamma) \subset M_i \quad \text{and} \quad \omega^*(\gamma) \subset M_j.$$

Theorem 5. *Let S be any compact invariant set of a continuous flow, with Morse-decomposition $\{M_p\}_{p \in P}$. Then*

$$l(S) \leqq \sum_{p \in P} l(M_p).$$

In particular, if $|P| < l(S)$, then some M_p has non-trivial Alexander cohomology (so contains a continuum of points).

Postponing the proof of this theorem we first derive the corollary:

Corollary. *Let S be a compact invariant set of a continuous flow. Assume, in addition, that S is gradient like (i.e. there exists a continuous real valued function on S which is strictly decreasing on non-constant orbits). Then S contains at least $l(S)$ rest points.*

In fact, assume there are only finitely many rest points in S. Then they form a Morse-decomposition of S, since S is assumed to be gradient like. As none of the sets of this decomposition has nontrivial cohomology, hence $l(M_p) = 1$, there must be at least $l(S)$ sets in the decomposition, hence $l(S)$ rest points, proving the corollary. It remains to prove Theorem 5.

Proof of Theorem 5. First observe that any decomposition of S can be obtained by first decomposing it into two sets, then decomposing one of these and continuing until the decomposition is reached. Therefore one needs only prove the theorem for decompositions into two sets. Thus let (M_1, M_2) be an ordered Morse-decomposition of S. From the definition we conclude that there is a compact neighborhood S_1 of M_1 in S and a compact neighborhood S_2 of M_2 in S with $S_1 \cup S_2 = S$ and such that

$$M_1 = \bigcap_{t>0} S_1 \cdot t \quad \text{and} \quad M_2 = \bigcap_{t>0} S_2 \cdot (-t).$$

Consequently, by the continuity property of the Alexander cohomology, $\check{H}^*(S_1) = \check{H}^*(M_1)$ and $\check{H}^*(S_2) = \check{H}^*(M_2)$, and it is therefore sufficient to prove that $l(S_1) + l(S_2) \geqq l(S)$ for $S_1 \cup S_2 = S$. This will follow from the following general observation:

Lemma 4. *Let $S_1 \cup S_2 \subset S$ be three compact sets. Denote by $i_1: S_1 \to S$, $i_2: S_2 \to S$ and $i: S_1 \cup S_2 \to S$ the inclusion maps. Let $\alpha, \beta \in \check{H}^*(S)$. Then $i_1^* \alpha = 0$ and $i_2^* \beta = 0$ imply $i^*(\alpha \cup \beta) = 0$.*

Proof. Consider the following diagram

$$\begin{array}{ccccc}
\check{H}^*(S, S_1) & \otimes & \check{H}^*(S, S_2) & \xrightarrow{\cup} & \check{H}^*(S, S_1 \cup S_2) \\
\downarrow{\scriptstyle j_1^*} & & \downarrow{\scriptstyle j_2^*} & & \downarrow{\scriptstyle j^*} \\
\check{H}^*(S) & \otimes & \check{H}^*(S) & \xrightarrow{\cup} & \check{H}^*(S) \\
\downarrow{\scriptstyle i_1^*} & & \downarrow{\scriptstyle i_2^*} & & \downarrow{\scriptstyle i^*} \\
\check{H}^*(S_1) & \otimes & \check{H}^*(S_2) & \xrightarrow{\cup} & \check{H}^*(S_1 \cup S_2).
\end{array}$$

The vertical sequences are exact. If $\alpha \in \check{H}^*(S)$ satisfies $i_1^*(\alpha) = 0$ then there is an $\hat{\alpha} \in \check{H}^*(S, S_1)$ with $j_1^*(\hat{\alpha}) = \alpha$. Similarly, if $i_2^*(\beta) = 0$, then $j_2^*(\hat{\beta}) = \hat{\alpha}$ for some $\hat{\beta} \in \check{H}^*(S, S_2)$. Since $j^*(\hat{\alpha} \cup \hat{\beta}) = j_1^*(\hat{\alpha}) \cup j_2^*(\hat{\beta}) = \alpha \cup \beta$ we conclude, by exactness, $i^*(\alpha \cup \beta) = i^* \circ j^*(\hat{\alpha} \cup \hat{\beta}) = 0$, as advertized. □

Going back to the proof of Theorem 5 we let $\alpha_1, \alpha_2, \ldots, \alpha_l$ be in $H^*(S)$ such that $\alpha_1 \cup \alpha_2 \cup \ldots \cup \alpha_l \neq 0$. Let the α's be ordered so that $\alpha_1 \cup \ldots \cup \alpha_r$ is the longest product not in the kernel of i_1^*. Therefore $l(S_1) \geqq r+1$ and $i_1^*(\alpha_1 \cup \ldots \cup \alpha_r \cup \alpha_{r+1}) = 0$. Since $S = S_1 \cup S_2$ it follows from Lemma 4, that $i_2^*(a_{r+2} \cup \ldots \cup \alpha_l) \neq 0$ and therefore $l(S_2) \geqq l - (r+1) + 1 = l - r$. Hence $l(S_1) + l(S_2) \geqq l+1$. We have shown that if S admits a nontrivial product with l factors, i.e. if $l(S) \geqq l+1$, then $l(S_1) + l(S_2) \geqq l+1$, so the statement of the theorem follows. □

5. Proof of Theorem 1

By Lemma 1 the 1-periodic solutions, which are contractible on T^{2n}, are in one-to-one correspondence with the rest points of the gradient flow of ∇g on $T^{2n} \times \mathbb{R}^{2M}$. If S is the invariant set of this flow in the block B of Lemma 2 we conclude, by the Corollary of Theorem 5, together with Theorem 4, that S contains at least $2n+1$ rest points. □

6. Proof of Theorem 3

The proof proceeds along the same lines as that of Theorem 1 and we sketch the essential points. On the manifold $M = T^n \times \mathbb{R}^n$ the Hamiltonian vectorfield is given by the Hamiltonian function $h(t, x, y)$, on the covering space $\mathbb{R}^n \times \mathbb{R}^n$. The function h is periodic of period 1 in t and in the x-variables. We look again for special periodic solutions which are, on $\mathbb{R}^n \times \mathbb{R}^n$, given by $(x(t), y(t))$, where both $x(t)$ and $y(t)$ are periodic of period 1. Hence, on T^n, the loop

$t \mapsto x(t)$ is contractible. These periodic solutions are the critical points of f as defined in (4), with x replaced this time by (x, y). We write

$$h(t,x,y) = \tfrac{1}{2}\langle y, by\rangle + \{h(t,x,y) - \tfrac{1}{2}\langle y, by\rangle\}$$
$$= h_1(y) + h_0(t,x,y).$$

It follows from the assumption of Theorem 3 that $|\nabla h_0| \leqq K$. We proceed now as in the proof of Theorem 1. The Eq. (6) looks as follows:

$$Au = F_1(u) + F_0(u) = F(u),$$

where $F_1(u)(t) = \nabla h_1(u(t))$ and $F_0(u)(t) = \nabla h_0(t, u(t))$. Moreover $|F_0(u)| \leqq K$ for all $u \in H$. The operator $F_1(u)$ is a bounded linear operator of the Hilbertspace H:

$$F_1(u) := Bu, \qquad Bu(t) = \begin{pmatrix} 0 & 0 \\ 0 & b \end{pmatrix} u(t).$$

Since $|h''(t,x,y)| \leqq \alpha$ for some $\alpha > 0$, the sought periodic solutions are again found as the critical points of a function g defined on a finite dimensional space Z. We find

$$\nabla g(z) = Az - Bz - PF_0(u(z)).$$

Here we have used, that B commutes with the projection P of the Hilbert space H onto Z. The vectorfield $\nabla g(z)$ is this time a vectorfield on $T^n \times \mathbb{R}^n \times \mathbb{R}^{2M}$. More specifically we can use as coordinates in Z the Fourier coefficients up to order $N = [\beta]$, with β as in the proof of Theorem 1. If $z(t) = (x(t), y(t))$ we thus have

$$x(t) = x_0 + \sum_{n=1}^{N} (\alpha_n \varphi_n(t) + \beta_n \psi_n(t))$$
$$y(t) = y_0 + \sum_{n=1}^{N} (a_n \varphi_n(t) + b_n \psi_n(t)),$$

where $\varphi_n(t) = \sin(2\pi n t)$ and $\psi_n(t) = \cos(2\pi n t)$. The meanvalue $x_0 = [x(t)]$ is the variable on the torus T^n. In these coordinates, the gradient equation $\frac{d}{ds} z = \nabla g(z)$ becomes, if we omit the nonlinear term $PF_0(u(z))$, which is uniformly bounded:

$$\frac{d}{ds} x_0 = 0$$

$$\frac{d}{ds} y_0 = -b y_0$$

$$\frac{d}{ds}\begin{pmatrix} \alpha_n \\ \beta_n \\ a_n \\ b_n \end{pmatrix} = \left(\begin{array}{cc|cc} 0 & & 2\pi n J & \\ & & & \\ \hline & & -b & 0 \\ & & 0 & -b \end{array}\right) \begin{pmatrix} \alpha_n \\ \beta_n \\ a_n \\ b_n \end{pmatrix};$$

$$2\pi n \quad J^T$$

$1 \leqq n \leqq N$. Since, by assumption, $\det b \neq 0$, we see from this representation, that the vectorfield transversal to the torus is hyperbolic, if we omit the nonlinear terms. Since the omitted terms are uniformly bounded, we can construct therefore, as in Lemma 2, a special isolating block B for the gradient flow $\nabla g(z)$ which is of the form: T^n times hyperbolic i.e. $B = T^n \times D_1 \times D_2$ with exit set $B^- = T^n \times \partial D_1 \times D_2$ and entrance set $B^1 = T^n \times D_1 \times \partial D_2$. Application of the Corollary to Theorem 5 together with Theorem 4 yields $(n+1)$ critical points of g, which give rise to $(n+1)$ periodic solutions. We claim that the periodic solutions found are contained in the region $T^n \times D \subset T^n \times \mathbb{R}^n$, where D is as in the assumption of Theorem 3. In fact, on $T^n \times (\mathbb{R}^n \setminus D)$ the Hamiltonian system is integrable;

$$\dot{x} = \frac{\partial}{\partial y} h(t, x, y) = by + a$$

$$\dot{y} = -\frac{\partial}{\partial x} h(t, x, y) = 0.$$

Hence the tori $T^n \times \{y\}$ are invariant under the flow, and the restriction of the flow onto a torus $T^n \times \{y\}$ is in fact linear: φ^t: $(x, y) \mapsto (x + t(by + a), y)$. In particular, the periodic solutions are not described by periodic functions on $\mathbb{R}^n \times (\mathbb{R}^n \setminus D)$ and hence do not count. Therefore the periodic solutions found above must lie in $T^n \times D$. The proof of Theorem 3 is finished. □

7. Proof of Theorem 2

We shall make use of the Morse-theory for flows as represented in [7]. In order to briefly outline the result we need, we consider a continuous flow on a locally compact and metric space X. A compact and invariant subset $S \subset X$ is called isolated, if it admits a compact neighborhood N such that S is the maximal invariant subset which is contained in N. With an isolated invariant set S a pair (N_1, N_2) of compact spaces can be associated, where $N_2 \subset N_1$ is roughly the "exit set" of N_1 and where $S \subset \operatorname{int}(N_1 \setminus N_2)$ is the invariant set contained in N_1. The homotopy type of the pointed space $(N_1/N_2, *)$ then does not depend on the particular choice of the "index-pair" (N_1, N_2) for S, and is called the index of S. It is denoted by $h(S) = [(N_1/N_2, *)]$. The algebraic invariants of $h(S)$ are defined to be

$$p(t, h(S)) := \sum_{j \geqq 0} t^j \dim \check{H}^j(N_1, N_2),$$

where (N_1, N_2) is any index-pair for S. Let now $\{M_1, \ldots, M_k\}$ be an ordered Morse-decomposition of S. Then the relation between the algebraic invariants of $h(M_k)$ and those of $h(S)$ is described by the following Morse-inequalities (see [7], Theorem 3.3):

$$\sum_{j=1}^{k} p(t, h(M_j)) = p(t, h(S)) + (1 + t)\, Q(t), \tag{17}$$

where Q is a formal power series with non-negative integer coefficients.

We shall apply this equation to the flow of ∇g on $T^{2n}\times\mathbb{R}^N\times\mathbb{R}^N$, with S being the set of bounded solutions. Then S is compact, since by (13), $|\nabla g(z)|\geqq\varepsilon$ for all $z=(x,\xi)\in T^{2n}\times\mathbb{R}^{2N}$ with $|\xi|\geqq K$. In particular S is contained in the interior of the compact set $B:=T^{2n}\times D_1\times D_2$, if we choose the radii of the discs sufficiently big. Moreover, by Lemma 2, the compact pair (B,B^-), with the exit set $B^-:=T^{2n}\times\partial D_1\times D_2$, is an index pair for S in the sense of ([7], Definition 3.4). Therefore $h(S)=[(B/B^-,*)]$ and the algebraic invariants of $h(S)$ are easily computed. Namely

$$\check{H}^*(h(S))\cong\check{H}(B/B^-,*)\cong\check{H}^*(B,B^-)\cong\check{H}^*(T^{2n}\times D_1,T^{2n}\times\partial D_1)$$

which, by the Künneth-formula, is isomorphic to $\check{H}^*(T^{2n})\otimes\check{H}^*(D_1,\partial D_1)$. As $\check{H}^*(D_1,\partial D_1)\cong\check{H}^*(S^N,*)$, S^N being a sphere of dimension N, we conclude $\check{H}^j(h(S))\cong\check{H}^{j-N}(T^{2n})$. Consequently, the algebraic invariants of $h(S)$ are given by

$$p(t,h(S))=\sum_{j=0}^{2n}\binom{2n}{j}t^{N+j}, \tag{18}$$

since $\dim\check{H}^j(T^{2n})=\binom{2n}{j}$. Recall Lemma 1 and assume that all the periodic solutions having period 1 are nondegenerate. In this case it can be shown ([7], Lemma 2.6) that the function g is a Morse-function, hence has only non-degenerate critical points. Their number is finite, since the critical points are contained in the compact set B. Therefore the critical points $\{z_j\}=M_j$ can be labeled so that they form an ordered Morse-decomposition of S. It is easily seen ([7], Sect. 3.6), that $h(\{z_j\})=[(S^{d_j},*)]$, where d_j is the Morse-index of the critical point z_j. Consequently $p(t,h(\{z_j\})=t^{d_j}$. Hence, by (18), the Morse-inequalities look as follows:

$$\sum_{j=1}^{k}t^{d_j}=\sum_{j=0}^{2n}\binom{2n}{j}t^{j+N}+(1+t)\,Q(t). \tag{19}$$

Since the polynomial Q has nonnegative integer coefficients we conclude in particular that indeed $k\geqq\sum_{j=0}^{2n}\binom{2n}{j}=2^{2n}$, as we wanted to prove. □

8. Appendix

We prove the following statement which we found in V.I. Arnold's book [2], Appendix 9, without proof however. The idea of the proof was suggested to us by J. Mather.

Theorem 6. *If ψ is a symplectic C^∞-diffeomorphism of $T^2=\mathbb{R}^2/\mathbb{Z}^2$ then the following statements are equivalent.*

(i) *ψ is, on $\mathbb{R}^2$, of the form $x\mapsto x+f(x)$ with f being periodic and $[f]=0$.*

(ii) *ψ is generated by a globally Hamiltonian vectorfield on T^2.*

(iii) *ψ belongs to the commutator subgroup of $\mathrm{Diff}_0^\infty(T^2,\omega)$, the identity component of the group $\mathrm{Diff}^\infty(T^2,\omega)$ of symplectic diffeomorphisms of T^2.*

The statement (ii)⇔(iii) is a special case of a result due to A. Banyaga [4]. In order to prove (i)⇔(ii) we begin with a Lemma, which is due to J. Moser.

Let M be a compact symplectic manifold. Fixing a symplectic form ω_0 on M we consider the set $\Omega = \{\text{symplectic form } \omega \text{ on } M \mid \int_c \omega = \int_c \omega_0 \text{ for all 2-cycles } c \text{ on } M\}$.

Lemma 5. *Let* $s \mapsto \omega_s \in \Omega$, $s \in [0,1]$ *be a closed curve which is contractible to* ω_0 *in* Ω, i.e. *there exist* $\omega_{st} \in \Omega$, $s, t \in [0,1]$ *with* $\omega_{s0} = \omega_0$, $\omega_{s1} = \omega_s$ *for* $s \in [0,1]$, *and* $\omega_{0t} = \omega_{1t} = \omega_0$ *for* $t \in [0,1]$. *Then there exists a closed curve* $s \mapsto \varphi_s \in \mathrm{Diff}^\infty(M)$ *satisfying*

$$\varphi_s^* \omega_s = \omega_0 \quad \textit{and} \quad \varphi_0 = \varphi_1 = \mathrm{id}.$$

Remark. For $\dim M = 2$ every closed curve ω_s meets the assumption of the Lemma. In fact, since Ω is convex we can set $\omega_{st} = t\omega_s + (1-t)\omega_0 \in \Omega$. For $\dim M > 2$ the assumption is met, if the component of Ω containing ω_0 is simply connected. When this is the case is not known to us.

Proof. We follow [10]. By the Hodge decomposition theorem with respect to a given metric on M we have the representation $\frac{d}{dt}\omega_{st} = d\alpha_{st} + h_{st}$. The 2-form h_{st} is harmonic. We shall require that $\alpha_{st} = \delta\beta_{st}$ so that the choice of α_{st} is unique. Since by assumption the periods of ω_{st} are independent of t, those of $\frac{d}{dt}\omega_{st}$ are zero and so $h_{st} = 0$. Thus

$$\frac{d}{dt}\omega_{st} = d\alpha_{st}, \tag{20}$$

where, by the above normalization, α_{st} is unique. As $\omega_{1t} = \omega_{0t} = \omega_0$ we therefore have $\alpha_{0t} = \alpha_{1t} = 0$. Let V_{st} be the unique vectorfield satisfying

$$\omega_{st}(V_{st}, \cdot) = -\alpha_{st}, \tag{21}$$

and let φ_{st} be the flow

$$\frac{d}{dt}\varphi_{st} = V_{st} \circ \varphi_{st}, \qquad \varphi_{s0} = \mathrm{id}. \tag{22}$$

Since $d\omega_{st} = 0$ one finds with (20) and (21)

$$\begin{aligned}\frac{d}{dt}(\varphi_{st}^* \omega_{st}) &= \varphi_{st}^* \left\{ \frac{d}{dt}\omega_{st} + d(\omega_{st}(V_{st}, \cdot)) \right\} \\ &= \varphi_{st}^* \left\{ \frac{d}{dt}\omega_{st} - d\alpha_{st} \right\} = 0.\end{aligned}$$

Hence $\varphi_{st}^* \omega_{st} = \omega_{s0} = \omega_0$. Also, from $V_{0t} = V_{1t} = 0$, we conclude $\varphi_{1t} = \varphi_{0t} = \mathrm{id}$. Therefore $\varphi_s = \varphi_{s1}$ is the desired loop of diffeomorphisms. □

Consider now $M = T^{2n}$ and let ω_0 be the standard symplectic form on T^{2n}. Define the subgroups A, A_1, A_2 of $\mathrm{Diff}^\infty(M)$ as follows

$$\begin{aligned}A &= \{\psi \in \mathrm{Diff}^\infty(M) \mid \psi \text{ homotopic to id}\} \\ A_1 &= \{\psi \in A \mid \psi^* \omega_0 = \omega_0\} \\ A_2 &= \{\psi \in \mathrm{Diff}^\infty(M) \mid \psi^* \omega_0 = \omega_0 \text{ and, on } R^{2n},\ \psi(x) = x + p(x) \\ &\qquad\qquad \text{with } [p] = 0\}.\end{aligned}$$

Then clearly $A \supset A_1 \supset A_2$.

Lemma 6. *If $M=T^2$, then A_1 and also A_2 are connected by smooth arcs.*

Proof. It is a nontrivial fact, that, since $\dim T^2=2$, the group of diffeomorphisms of T^2 which are homotopic to the identity is equal to the one component of $\mathrm{Diff}^\infty(T^2)$, which is connected by smooth arcs, see C.J. Earle and J. Eells [13]. If $\psi\in A_1$ we take a smooth arc $\chi_s\in A$ with $\chi_0=\mathrm{id}$ and $\chi_1=\psi$ and set $\omega_s=\chi_s^*\omega_0\in\Omega$ so that $\omega_s=\omega_0$ for $s=0,1$. By Lemma 5 there is φ_s with $\varphi_s^*\omega_s=\omega_0$ and $\varphi_0=\varphi_1=\mathrm{id}$. Therefore $\psi_s=\varphi_s\circ\chi_s\in A_1$ and $\psi_0=\mathrm{id}$, $\psi_1=\psi$.

If, moreover $\psi\in A_2$, then ψ_s: $x\mapsto x+a_s+p_s(x)$, $a_s\in\mathbb{R}^2$ and $a_0=a_1=0$, $[p_s]=0$. We set τ_s: $x\mapsto x-a_s$, then $\tilde{\psi}_s=\tau_s\circ\psi_s\in A_2$ and $\tilde{\psi}_0=\mathrm{id}$, $\tilde{\psi}_1=\psi_1=\psi$. □

In view of this lemma, the statement (i)$\Leftrightarrow$(ii) is an immediate consequence of the following simple

Lemma 7. *A smooth arc $\psi_t\in A_1$ with $\psi_0=\mathrm{id}$ is the flow of a globally Hamiltonian vectorfield on T^{2n} if and only if $\psi_t\in A_2$.*

Proof. Let $\psi_t\in A_2$, then ψ_t: $x\mapsto x+p_t(x)$ with $[p_t]=0$. Since $\psi_t^*\omega_0=\omega_0$ we conclude that the vectorfield $\left(\frac{d}{dt}\psi_t\right)\circ\psi_t^{-1}$ is of the form:

$$\left(\frac{d}{dt}\psi_t\right)\circ\psi_t^{-1}=J\nabla h.$$

$\nabla h(t,x)$ is periodic in x, so that $h(t,x)=\langle x,c(t)\rangle+h_1(t,x)$, with h_1 periodic in x. But $[\dot{p}_t]=0$, hence $[\dot{p}_t\circ\psi_t^{-1}]=[\dot{p}_t]=0$ and consequently $c(t)=0$, so that indeed $h=h_1$ is periodic in x. Conversely, if ψ_t is the flow of the globally Hamiltonian vectorfield $J\nabla h$ on T^{2n}, then $\psi_t^*\omega_0=\omega_0$ and $\psi_t\in A_2$. □

We would like to thank A. Dold, M. Herman, J. Mather, and J. Moser for suggestions and helpful discussions. We thank the Forschungsinstitut für Mathematik ETH Zürich for its hospitality. For the second author the visit in Zürich was made possible by a grant (Akademie-Stipendium) of the Stiftung Volkswagenwerk.

Note Added in Proof

We point out that M. Chaperon used the method introduced above to estimate the number of intersection points of Lagrangian submanifolds. He proved, for example, the following statement for an n-dimensional torus $M=T^n$. If $j: I\times M\to T^*M$ is an exact Lagrangian isotopy such that $j_0(M)$ is the zero section Σ of $T^*M\to M$, then $\Sigma\cap j_1(M)$ contains at least $n+1$ points.

References

1. Amann, H., Zehnder, E.: Nontrivial solutions for a class of nonresonance problems and applications to nonlinear differential equations. Annali Sc. Norm. Sup. Pisa, Serie IV. Vol. **7**, 539–603 (1980)
2. Arnold, V.I.: Mathematical methods of classical mechanics. (Appendix 9). Berlin-Heidelberg-New York: Springer 1978
3. Arnold, V.I.: Proceedings of symposia in pure mathematics. Vol. XXVIII A.M.S., p. 66, 1976
4. Banyaga, A.: Sur la structure du groupe des difféomorphismes qui préservent une forme symplectique. Comment. Math. Helvetici **53**, 174-227 (1978)
5. Banyaga, A.: On fixed points of symplectic maps. Preprint
6. Conley, C.C.: Isolated invariant sets and the Morse index. CBMS, Regional Conf. Series in Math., vol. 38 (1978)
7. Conley, C.C., Zehnder, E.: Morse type index theory for flows and periodic solutions for Hamiltonian equations. To appear in Comm. Pure and Appl. Math.

8. Moser, J.: A fixed point theorem in symplectic geometry. Acta Math. **141**, 17–34 (1978)
9. Moser, J.: Proof of a generalized form of a fixed point theorem due to G.D. Birkhoff. Lecture Notes in Mathematics, Vol. 597: Geometry and Topology, pp. 464–494. Berlin-Heidelberg-New York: Springer 1977
10. Moser, J.: On the volume elements on a manifold. Transactions Amer. Math. Soc. **120**, 286–294 (1965)
11. Poincaré, H.: Méthodes nouvelles de la mécanique célèste. Vol. 3, chap. 28. Paris: Gauthier Villars 1899
12. Weinstein, A.: Lectures on symplectic manifolds. CBMS, Regional conf. series in Math., vol. 29 (1977)
13. Earle, C.J., Eells, J.: A fibre bundle description of Teichmüller theory. J. Diff. Geometry **3**, 19–43 (1969)

Oblatum 15-III-1983

Invent. math. 73, 51-69 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Analyticity of CR Equivalences Between Some Real Hypersurfaces in $\mathbb{C}^n$ with Degenerate Levi Forms

C.K. Han

Department of Mathematics, Tulane University, New Orleans, LA 70118, USA

§1. Introduction and Statement of Main Results

This paper is concerned with real analyticity of CR equivalences and its application to the problem of continuation of a biholomorphic mapping across a real analytic hypersurface in $\mathbb{C}^n$.

Let M_1 and M_2 be real analytic hypersurfaces in $\mathbb{C}^n$ of real codimension 1 and $F: M_1 \to M_2$ be a CR equivalence. It is well known that if these hypersurfaces are of non-degenerate Levi form and if $F \in C^3$, then $F \in C^\omega$ ([8, 10]). In this paper we study two cases of degenerate hypersurfaces.

To state the main results, we fix some notation: Let $r \in C^\omega$ be a local defining function of a hypersurface M at a point $0 \in \mathbb{C}^n$. By a holomorphic change of coordinates we can get local coordinates $(z_1, \ldots, z_n)$ so that

$$r(z_1, \bar{z}_1, \ldots, z_n, \bar{z}_n) = z_n + \bar{z}_n + \phi(z_1, \bar{z}_1, \ldots, z_n, \bar{z}_n), \tag{1}$$

where $\phi(0) = 0$ and the Taylor series expansion has no purely holomorphic terms (therefore, no purely anti-holomorphic terms, since ϕ is real-valued).

For any $(n-1)$ tuple of nonnegative integers $\alpha = (a_1, \ldots, a_{n-1})$ and $\beta = (b_1, \ldots, b_{n-1})$, we let

$$\frac{\partial^{|\alpha|+|\beta|} r}{\partial \bar{z}^\alpha \partial z^\beta} = \left(\frac{\partial}{\partial \bar{z}_1}\right)^{a_1} \cdots \left(\frac{\partial}{\partial \bar{z}_{n-1}}\right)^{a_{n-1}} \left(\frac{\partial}{\partial z_1}\right)^{b_1} \cdots \left(\frac{\partial}{\partial z_{n-1}}\right)^{b_{n-1}} r.$$

Let $F: M_1 \to M_2$ be a C^1 diffeomorphism and let $F = (f_1, \ldots, f_n)$ be its coordinatewise expression. Then F is a CR equivalence if and only if each f_j $(j = 1, \ldots, n)$ satisfies the tangential Cauchy-Riemann equations. E. Cartan solved in [9] the equivalence problem of CR structure for C^ω strongly pseudoconvex hypersurfaces in $\mathbb{C}^2$, namely the problem of finding local differential invariants of a C^ω strongly pseudoconvex hypersurface in $\mathbb{C}^2$ such that two such hypersurfaces M_1 and M_2 are CR equivalent if and only if the corresponding invariants agree. A basic idea in [9] is to construct, by differentiating the tangential Cauchy-Riemann equations repeatedly and making use of nondegeneracy

conditions, a complete system of real analytic total differential equations whose integral manifolds correspond to CR equivalences. This is the so-called method of prolongation developed by E. Cartan (cf. [8–10, 17, 21, 22, 27]). Since an integral manifold of a system of C^ω total differential equations is C^ω real analyticity of a CR equivalence follows as a corollary.

By applying the same method to some degenerate hypersurfaces, we can prove the following

Theorem 1. *Let M_1 and M_2 be germs of C^ω real hypersurfaces in $\mathbb{C}^n$ at O_1 and O_2, respectively, and $F: M_1 \to M_2$ be a CR equivalence. Suppose that there exists a local defining function r of M_2 and a coordinate system $(z_1, \ldots, z_n)$ near O_2 normalized as in* (1) *such that for each $j=1, \ldots, n-1$, there exists an $(n-1)$ tuple of nonnegative integers*

$$\alpha_j=(a_j^1, \ldots, a_j^{n-1})$$

so that

$$\det\left(\frac{\partial^{|\alpha_j|+1} r}{\partial \bar{z}^{\alpha_j} \partial z_i}\right)_{i,j=1,\ldots,n-1} \neq 0.$$

Let $m=\max\{|\alpha_j|; j=1, \ldots, n-1\}$. Then $F \in C^{m^2+m+1}$ implies that $F \in C^\omega$.

Observe that Theorem 1 with $m=1$ is the well known nondegenerate case. As another case, if M_1 and M_2 are regular in the sense of Tanaka [27], we can construct a linear elliptic system of partial differential equations with analytic coefficients of which a CR equivalence between M_1 and M_2 is a solution. Analyticity of solutions follows from the theory of linear p.d.e. (cf. [18, 23]). For any real hypersurface M in $\mathbb{C}^n$ we define $H(M) \equiv T(M) \cap JT(M)$, where J is the almost complex structure tensor.

Theorem 2. *Let M_i $(i=1,2)$ be C^ω real hypersurfaces in $\mathbb{C}^n$. Suppose that for each point of M_i there exists a global C^ω infinitesimal CR automorphism which is transversal to $H(M_i)$ on a neighborhood of the point and the set of nondegenerate points is dense in M_i. Then $d_i=$dimension of the vector space of C^∞ infinitesimal CR automorphism on M_i is finite. Furthermore, if there exists a CR equivalence $F \in C^m$, where $m=\max(d_1, d_2)+5$, of M_1 onto M_2, then $d_1=d_2$ and FC^ω.*

This study was motivated by the problem of boundary regularity of biholomorphic mappings. There has recently been much progress in this old problem of function theory after the appearance of the fundamental paper [15] (1974) of C. Fefferman (for example, [1–5, 12–14, 16, 24–26]). In particular, the following theorem is one of the generalizations of [15].

Theorem (S. Bell and E. Ligocka [5]). *If D_1 and D_2 are bounded pseudoconvex domains with real analytic boundaries and Φ is a biholomorphic mapping of D_1 onto D_2, then Φ is C^∞ up to the boundary and thus induces a C^∞ CR equivalence between the boundaries.*

Analyticity of a CR equivalence becomes of interest in this context, for if the induced CR equivalence between the boundaries is real analytic, then Φ

continues holomorphically across the boundary due to the following well known theorem (cf. [7, 8]).

Theorem. *Suppose that M is a nonsingular C^ω real hypersurface in $\mathbb{C}^n$. If a complex valued function $f \in C^\omega$ on M satisfies the tangential Cauchy-Riemann equations, then f extends to a unique holomorphic function of a neighborhood of M.*

By combining the above two theorems and our Theorem 1, we get

Corollary 1. *Suppose that D_1 and D_2 are bounded pseudoconvex domains in $\mathbb{C}^n$ with real analytic boundaries and Φ is a biholomorphic mapping of D_1 onto D_2. Suppose that ∂D_2 satisfies the conditions of Theorem 1 at a point $O_2 \in \partial D_2$. Let $O_1 = \Phi^{-1}(O_2) \in \partial D_1$. Then there exists a neighborhood U of O_1 such that Φ extends to a biholomorphic mapping of $D_1 \cup U$.*

Holomorphic continuation of a biholomorphic mapping across a real analytic boundary has been studied recently by different approaches (cf. [2–4, 14, 25, 26]). In particular, the following theorem is proved by means of a reflexion principle.

Theorem (K. Diederich and S. Webster, [14]). *Suppose that D_1 and D_2 are bounded domains with real analytic boundaries and $\Phi: D_1 \to D_2$ is a biholomorphic mapping. Then there exists an integer k and a real analytic set $A \subsetneq \partial D_1$, $\dim A \leqq 2n-3$, such that if Φ and Φ^{-1} are C^k up to the boundary, then Φ continues holomorphically past $\partial D \setminus A$.*

They showed that outside a real analytic set A a higher order nondegeneracy condition similar to our hypothesis of Theorem 1 holds. In order for a C^k (for some sufficiently large k) CR equivalence to be analytic, there must be some nondegeneracy conditions (see Examples 1 and 2 of § 2). However, none of the conditions of Theorem 1 or [14] is necessary (see Example 3 of § 2). It seems to the author that at the points where these generalized nondegeneracy conditions fail, a new coordinate system may be used so that a similar computation as in the proof of Theorem 1 (§ 2) is possible. Thus, we are led to the following

Conjecture. *Suppose that M_1 and M_2 are germs of C^ω pseudoconvex hypersurfaces in $\mathbb{C}^n$ and $F: M_1 \to M_2$ is a CR equivalence. If there is no complex subvariety contained in M_1* (cf. [6, 11, 19, 20]), *then there exists an integer k such that $F \in C^k$ implies that $F \in C^\omega$.*

Now, if D is a bounded domain in $\mathbb{C}^n$ with real analytic boundary, then the Levi determinant of ∂D does not vanish identically and therefore the set of nondegenerate points is dense in ∂D. Thus we have the following corollary to Theorem 2.

Corollary 2. *Let D_1 and D_2 be bounded pseudoconvex domains in $\mathbb{C}^n$ with real analytic boundary and Φ; $D_1 \to D_2$ be a biholomorphic mapping. Suppose that for each point $p \in \partial D_i$ $(i=1,2)$ there exists a global C^ω infinitesimal CR automorphism which is transversal to the maximal complex subspace of $T_p(\partial D_i)$. Then Φ extends to a biholomorphic mapping of a neighborhood of $\bar{D}_1$.*

A bounded pseudoconvex Reinhardt domain in $\mathbb{C}^n$ with a C^ω boundary (Example 4 of § 3) satisfies the conditions of Corollary 2. For complete Reinhardt domains, however, there is a more general result:

Theorem (S. Bell [2]). *A proper holomorphic mapping between bounded complete Reinhardt domains extends holomorphically past the boundary.*

This is a revised version of a part of the author's Ph.D. dissertation written at the University of Michigan. The author would like to thank his thesis supervisor, Professor Daniel M. Burns, for suggesting the problem and for all his help. The author also thanks Professor M. Kuranishi for valuable discussions. The method of computation in § 2 was largely inspired from a note [17] by H. Goldschmidt and H. Rossi. The author thanks both of them.

§ 2. Hypersurfaces with a Generalized Nondegeneracy Condition

In this section we will prove Theorem 1 and give some examples. First of all, in order for a sufficiently smooth CR equivalence between two C^ω hypersurfaces to be real analytic, some nondegeneracy conditions on the hypersurfaces are necessary. Consider the following two examples.

Example 1 (Levi flat case). Let

$$M=\{(z_1, z_2)\in\mathbb{C}^2: \operatorname{Im} z_2=0\}$$

and let $h: R^1\to R^1$ be any C^k diffeomorphism. Then the mapping

$$(x_1, y_1, x_2, 0)\mapsto(x_1, y_1, h(x_2), 0)$$

is a C^k CR equivalence of M onto itself.

Example 2 (a hypersurface containing complex subvarieties). Let S be a C^ω hypersurface in $\mathbb{C}^{n-1}$ and $M=S\times\mathbb{C}$. Let $f\in C^k$ be a CR function on S, then the mapping

$$(p, z)\mapsto(p, f(p)\cdot z)$$

is a C^k CR equivalence locally where $f\neq 0$ of M into itself.

We fix notation: Let h_1 and h_2 be germs of C^k functions at a point 0 of a real analytic manifold M. By $h_1\underset{q}{\sim}h_2$ we mean that h_1 and h_2 belong to the same class of q-jet, namely, all the partial derivatives of h_1 at 0 up to order q are equal to those of h_2 ($q<k$). Let $x=(x_1, \ldots, x_d)$, where d is the dimension of M, be real analytic coordinates on M. For two C^k vector fields

$$X=\sum_i a_i\frac{\partial}{\partial x_i} \quad\text{and}\quad Y=\sum_i b_i\frac{\partial}{\partial x_i}, \quad \text{define}$$

$$X\underset{q}{\sim}Y \quad \text{if } a_i\underset{q}{\sim}b_i \quad \text{for each } i=1, \ldots, d.$$

Note that if $X_1\underset{q}{\sim}Y_1$ and $X_2\underset{q}{\sim}Y_2$, then $[X_1, X_2]\underset{q-1}{\sim}[Y_1, Y_2]$. Let a and b be germs of real analytic functions at $0\in M$. By $((a, b))$ we denote an element of the

ideal generated by a and b of the ring of germs of real analytic functions of M at 0. We denote by $C^\omega(x, u_1, u_2, \ldots)$ the set of real analytic functions of variables $x, u_1, u_2, \ldots$ where x is a real analytic coordinate of M and $u_1, u_2, \ldots$ are functions on M. The subset $C_0^\omega(x, u_1, u_2, \ldots)$ of $C^\omega(x, u_1, u_2, \ldots)$ consists of those vanishing at $(0, u_1(0), u_2(0), \ldots)$. An element of $C^\omega(x, u, u, \ldots)$ ($C_0^\omega(x, u_1, u_2, \ldots)$, resp.) will be denoted by $\varepsilon(x, u_1, u_2, \ldots)$ ($\varepsilon_0(x, u_1, u_2, \ldots)$, resp.).

Now let M be a C^ω hyperspace in $\mathbb{C}^n$. We shall use the standard notations: $H(M)$ is the bundle of the maximal complex subspaces of tangent spaces of M and its complexification $H_{\mathbb{C}}(M)$ has a decomposition

$$H_{\mathbb{C}}(M) = H^{1,0}(M) + H^{0,1}(M) \quad \text{etc.}$$

Proof of Theorem 1. Write $F = (f_1, \ldots, f_{n-1}, g)$ coordinatewise with respect to the normalized coordinate $(z_1, \ldots, z_n)$ as in (1). Let $\{\bar{Z}_1, \ldots, \bar{Z}_{n-1}\}$ be a set of independent C^ω sections of $H^{0,1}(M_1)$. Since F is a CR equivalence, we have

$$r \circ F = 0, \tag{2.1}$$

and

$$\bar{Z}_j f_i = 0 \tag{2.2}$$

$$\bar{Z}_j g = 0, \quad i, j = 1, \ldots, n-1$$

(tangential Cauchy-Riemann equations).

By using (2.2) and repeated application of the differential operators $\bar{Z}$'s and their conjugate Z's to (2.1), we will get all the partial derivatives of $f_1, \ldots, f_{n-1}$ and g of order m^2+m+1 as real analytic functions of the partial derivatives of lower order. This can be regarded as a system of C^ω total differential equations and analyticity of F follows as one of the applications (see [21]).

For any $h \in C^\infty(M_1)$ and a multi-index $A = (a_1, \ldots, a_{n-1})$, we express $Z_{n-1}^{a_{n-1}} \ldots Z_1^{a_1} h$ as $Z^A h$.

Lemma 1. *We can choose a real analytic basis $\{Z_1, \ldots, Z_{n-1}\}$ of $H^{1,0}(M_1)$ and a real analytic vector field T which is transversal to $H(M_1)$ having the following properties:*

i) $Z_i f_j(O_1) = \delta_{ij}$ (*Kronecker delta*)

ii) *For any multi-index A with $2 \leqq |A| \leqq m$*

$$Z^A f_j(O_1) = 0, \quad j = 1, \ldots, n-1.$$

iii) *Let $[Z_i, \bar{Z}_j] \equiv \sqrt{-1}\rho_{ij} T \pmod{Z, \bar{Z}\text{'s}}$. For each $i = 1, \ldots, n = 1$, there exists an integer $j_i \in \{1, \ldots, n-1\}$ and a multi-index A_i with $|A_i| \leqq m-1$ such that $\bar{Z}^{A_i} \rho_{ij_i} \neq 0$, and for any multi-index B with $|B| < |A_i|$, $\bar{Z}^B \rho_{ij}(O_1) = 0$ for all $j = 1, \ldots, n-1$.*

Proof of Lemma 1. On M_2 we choose a basis $\{L_1, \ldots, L_{n-1}\}$ of $H^{1,0}(M_2)$ and a real vector field N which is transversal to $H(M_2)$ as follows:

$$L_i = \frac{\partial}{\partial z_i} - \frac{r_i}{r_n} \frac{\partial}{\partial z_n}, \tag{2.3}$$

$$N=\tfrac{1}{2}J(\operatorname{grad} r)$$
$$=\sqrt{-1}\left\{\sum_{i=1}^{n-1}\left(r_{\bar{i}}\frac{\partial}{\partial z_i}-r_i\frac{\partial}{\partial \bar{z}_i}\right)+\left(r_{\bar{n}}\frac{\partial}{\partial z_n}-r_n\frac{\partial}{\partial \bar{z}_n}\right)\right\} \tag{2.4}$$

where $r_i=\dfrac{\partial r}{\partial z_i}$ etc.

Set

$$[L_i\bar{L}_j]=\sqrt{-1}R_{i\bar{j}}N+\sum_{t=1}^{n-1}a_{i\bar{j}}^{t}L_t+\sum_{t=1}^{n-1}b_{i\bar{j}}^{t}\bar{L}_t. \tag{2.5}$$

By substituting (2.3) and (2.4) in (2.5), we get

$$R_{i\bar{j}}=-\frac{r_{i\bar{j}}}{r_{\bar{n}}r_n}+((r_i,r_{\bar{j}})), \tag{2.6}$$

$$a_{i\bar{j}}^{t}=-\sqrt{-1}R_{i\bar{j}}\cdot r_{\bar{t}}, \tag{2.7}$$

$$b_{i\bar{j}}^{t}=\sqrt{-1}R_{i\bar{j}}\cdot r_t. \tag{2.8}$$

Now for each $i=1,\ldots,n-1$, let $\beta_i=(b_1^i,\ldots,b_{n-1}^i)$ be a multi-index of shortest length such that $\dfrac{\partial^{|\beta_i|+1}}{\partial\bar{z}^{\beta_i}\partial z_i}\neq 0$. By the hypothesis of the theorem such β_i exists and $|\beta_i|\leqq m$. Let t be the first integer with $b_t^i\neq 0$. Then define j_i and $A_i=(a_1,\ldots,a_{n-1})$ by

$$j_i=t \quad\text{and}\quad a_k=b_k^i\ (k\neq t), \qquad a_t=b_t^i-1.$$

Then

$$\frac{\partial^{|A_i|+1+1}r}{\partial\bar{z}^{A_i}\partial\bar{z}_{j_i}\partial\bar{z}_{j_i}\partial z_i}=\frac{\partial^{|\beta_i|+1}r}{\partial\bar{z}^{\beta_i}\partial z_i}\neq 0.$$

From (2.6) and (2.3), we see that

$$\bar{L}^{A_i}R_{i\bar{j}_i}=-\frac{r_{i\bar{j}}\bar{A}_i}{r_{\bar{n}}r_n}+((r_{i\bar{B}},r_{\bar{C}}\colon B \text{ sub-index of } A_i)).$$

(By a sub-index of a multi-index $A=(a_1,\ldots,a_{n-1})$, we mean a multiindex $A'=(a_1',\ldots,a_{n-1}')$ with $a_i'\leqq a_i$ for each $i=1,\ldots,n-1$.) Since the Taylor series of r has no pure antiholomorphic terms, $r_{\bar{C}}(O_2)=0$ and from the choice of A_i, $r_{i\bar{B}}(O_2)=0$. Therefore,

$$\bar{L}^{A_i}R_{i\bar{j}_i}(O_2)=-\frac{r_{i\bar{j}_i}\bar{A}_i}{r_n r_{\bar{n}}}(O_2)\neq 0. \tag{2.9}$$

Now choose a C^ω section Z_i of $H^{1,0}(M_1)$ for each $i=1,\ldots,n-1$ and a C^ω real vector field T so that

$$Z_i\underset{m}{\sim}F_*^{-1}(L_i) \quad\text{and}\quad T\underset{m}{\sim}F_*^{-1}(N). \tag{2.10}$$

Then we have

$$[Z_i, \bar{Z}_j]_{m-1}\sim[F_*^{-1}(L_i), F_*^{-1}(\bar{L}_j)]$$
$$=F_*^{-1}([L_i, \bar{L}_j])$$
Substitute (2.5)
$$=\sqrt{-1}(R_{i\bar{j}}\circ F)F_*^{-1}(N)+\sum_t(a_{i\bar{j}}^t\circ F)F_*^{-1}(L_t)+\sum_t(b_{i\bar{j}}^t\circ F)F_*^{-1}(\bar{L}_t) \quad (2.11)$$

Set

$$[Z_i, \bar{Z}_j]=\sqrt{-1}\rho_{i\bar{j}}T+\sum_{t=1}^{n-1}\alpha_{i\bar{j}}^t Z_t+\sum_{t=1}^{n-1}\beta_{i\bar{j}}^t\bar{Z}_t. \quad (2.12)$$

Since Z_i's and T are nonvanishing, from (2.10)–(2.12) we have

$$\rho_{i\bar{j}}\ {}_{m-1}\sim R_{i\bar{j}}\circ F. \quad (2.13)$$

Therefore,

$$\bar{Z}^{A_i}\rho_{i\bar{j}_i}(O_1)=(F_*^{-1}\bar{L})^{A_i}(R_{i\bar{j}_i}\circ F)(O_1)$$
$$=\bar{L}^{A_i}R_{ij_i}(O_2)$$
$$\neq 0 \quad \text{by (2.9).}$$

Thus $\{Z_1, \ldots, Z_{n-1}, T\}$ has the property (iii). It is easy to show that this set also satisfies (i) and (ii).

f_i $(i=1, \ldots, n-1)$ will be written as f when we do not need to specify. We write (f, g) for $(f_1, \ldots, f_{n-1}, g)$ and z for $(z_1, \ldots, z_n)$. Now rearrange the Taylor series expansion at O_2 of r as

$$r=z_n+\bar{z}_n+\sum_{i=1}^{n}z_i\phi^i(\bar{z})+\sum_{i\leqq j\leqq n}z_iz_j\phi^{ij}(\bar{z})$$
$$+\sum_{i\leqq j\leqq k\leqq n}z_iz_jz_k\phi^{ijk}(\bar{z})$$
$$+\text{----------}, \quad (2.14)$$

where all ϕ's are vanishing at O_2 and purely antiholomorphic. Then

$$r\circ F=g+\bar{g}+\left\{\sum_{i=1}^{n-1}f_i(\phi^i\circ F)+g(\phi^n\circ F)\right\}$$
$$+\left\{\sum_{i\leqq j\leqq n-1}f_if_j(\phi^{ij}\circ F)+\sum_{i=1}^{n-1}f_ig(\phi^{in}\circ F)+g^2(\phi^{nn}\circ F)\right\}$$
$$+\text{----------}. \quad (2.15)$$

Not we will get all the derivatives of f and g of the form $T^tZ^A\bar{Z}^Bf$ and $T^tZ^A\bar{Z}^Bg$, where A and B are multiindices with $|A|+|B|+t=m^2+m+1$, as real analytic functions of derivatives of f and g of order $\leqq m^2+m$.

If $|B|\neq 0$, then by (2.2) $T^tZ^A\bar{Z}^Bf=T^tZ^A\bar{Z}^Bg=0$, therefore it suffices to get T^tZ^Af and T^tZ^Ag $(t+|A|=m^2+m+1)$ in terms of lower order derivatives, which will be done in four steps, after fixing some notation.

For any nonnegative integer u, let

$$C_u \equiv C^\omega(x, \bar{f}_i, \bar{g}, g, T^t Z^A f_i : i=1, \ldots, n-1, t+|A| \leqq u).$$

A subset $C_{u,v}$ of C_u is defined by

$$C_{u,v} \equiv C^\omega(x, \bar{f}_i, \bar{g}, \bar{g}, T^t Z^A f_i : i=1, \ldots, n-1, t+|A| \leqq u, t \leqq v)$$

if $u \geqq v$ and $C_{u,v} = C_{u,u}$ if $u < v$. $\bar{C}_u$ and $\bar{C}_{u,v}$ are the complex conjugates of C_u and $C_{u,v}$ respectively, namely

$$\bar{C}_u = C^\omega(x, f_i, g, \bar{g}, T^t \bar{Z}^A \bar{f}_i : i=1, \ldots, n-1, t+|A| \leqq u) \text{ etc.}$$

By ε_u ($\varepsilon_{u,v}, \bar{\varepsilon}_u, \bar{\varepsilon}_{u,v}$, resp.) we mean an element of C_u ($C_{u,v}, \bar{C}_u, \bar{C}_{u,v}$, resp.).

The four steps are then

Step 1. *$Z^A f \in \bar{C}_m$ for any multi-index A.*

Step 2. *$C_{u,m} \subset \bar{C}_{m^2+m}$, or equivalently*

$\bar{C}_{u,m} \subset C_{m^2+m}$ for any nonnegative integer u.

Step 3. *If $t+|A| \geqq m+1$, $T^t Z^A f = \varepsilon_{t+|A|-1} + \bar{\varepsilon}_{u,m}$ for some sufficiently large integer u depending on t and $|A|$.*

Step 4. *$T^t Z^A g \in C^\omega(x, T^s Z^B f (s+|B| \leqq m^2+m), T^s \bar{Z}^B \bar{f} (s+|B| \leqq m^2+m), T^s g (s \leqq m^2+m), T^s \bar{g} (s \leqq m^2+m))$, for any t and A with $t+|A| = m^2+m+1$.*

By combining Step 2 and Step 3, we get

$$T^t Z^A f (t+|A| = m^2+m+1) \in C_{m^2+m}. \tag{2.16}$$

(2.16) together with Step 4 gives all derivatives of f and g or order m^2+m+1 as real analytic functions of derivatives of order $\leqq m^2+m$.

Proof of Step 1. For each $j=1, \ldots, n-1$, let α_j be the multiindex as in the hypothesis. Apply $\bar{Z}^{\alpha_j}$ to (2.15). Since $\bar{Z}^{\alpha_j} f = \bar{Z}^{\alpha_j} g = 0$ by (2.2), we have

$$\begin{aligned} 0 = \bar{Z}^{\alpha_j} \bar{g} &+ \left\{ \sum_{i=1}^{n-1} f_i \bar{Z}^{\alpha_j} (\phi^i \circ F) + g \bar{Z}^{\alpha_j} (\phi^n \circ F) \right\} \\ &+ \left\{ \sum_{p \leqq q \leqq n-1} f_p f_q \bar{Z}^{\alpha_j} (\phi^{pq} \circ F) + \sum_{p=1}^{n-1} f_p g \bar{Z}^{\alpha_j} (\phi^{pn} \circ F) + g^2 \bar{Z}^{\alpha_j} (\phi^{nn} \circ F) \right\} \\ &+ \ldots . \end{aligned} \tag{2.17}$$

Consider the coefficient of f_i. Since $|\alpha_j| \leqq m$,

$$\begin{aligned} \bar{Z}^{\alpha_j} (\phi^i \circ F)(O_1) &= (F_*^{-1} \bar{L})^{\alpha_j} (\phi^i \circ F)(O_1) \\ &\text{(where } F_*^{-1} \bar{L} \equiv (F_*^{-1} \bar{L}_1, \ldots, F_*^{-1} \bar{L}_{n-1})) \\ &= \bar{L}^{\alpha_j} \phi^i (O_2) \\ &= \frac{\partial^{|\alpha_j|+1} r}{\partial \bar{z}^{\alpha_j} \partial z_i} (O_2) \equiv \lambda_{ij}. \end{aligned}$$

Since $\det(\lambda_{i\bar{j}}) \neq 0$, we can solve (2.17) for f_i, to get

$$f_i = \psi_i + g\,\psi_i^n + \Big\{ \sum_{p \leqq q \leqq n-1} f_p f_q \psi_i^{pq} + \sum_{p=1}^{n-1} f_p\, g\, \psi_i^{pn} + g^2 \psi_i^{nn} \Big\} + \ldots, \quad (i = 1, \ldots, n-1) \tag{2.18}$$

where all the ψ's are in $C^\omega(x, \bar{g}, \bar{Z}^B \bar{f}: |B| \leqq m)$.

Let A be any multi-index. Apply Z^A to (2.18). To compute RHS we need the following

Lemma 2. a) *For any nonnegative integers u and k $(u \geqq k)$,*

$$\bar{Z}_j^k Z_i^u f = (-\sqrt{-1})^k (\rho_{i\bar{j}})^k \frac{u!}{(u-k)!} T^k Z_i^{u-k} f + \sum_{t=1}^{k-1} \gamma_t T^t Z_i^{u-t} f + \varepsilon_{u-1}, \quad \text{where } \gamma_t \in C^\omega(x).$$

b) *For any multi-indices B and C,*

$$Z^C \bar{Z}^B \bar{f} \in C^\omega(x, T^t \bar{Z}^D \bar{f} (t + |D| \leqq |B|,\ t \leqq |C|)),$$

or equivalently,

$$\bar{Z}^C Z^B f \in C^\omega(x, T^t Z^D f (t + |D| \leqq |B|,\ t \leqq |C|)).$$

Proof of Lemma 2. If $k = u = 1$,

$$\bar{Z}_j Z_i f = \{ Z_i Z_j - [Z_i, Z_j] \} f$$

substitute (2.12) for $[Z_i, \bar{Z}_j]$, then by (2.2)

$$= -\sqrt{-1}\,\rho_{i\bar{j}}\, Tf - \sum_{t-1}^{n-1} \alpha_{i\bar{j}}^t Z_t f.$$

Then use induction on k, u and the size of B and C.

By Lemma 2, $Z^A \bar{Z}^B \bar{f} \in \bar{C}_{|B|}$, therefore $Z^A(\psi\text{'s}) \in \bar{C}_m$. Thus Z^A applied to (2.18) becomes

$$Z^A f_i = \mu_i + \sum_{B \prec A} Z^B g\, \mu_{iB}^n + \Big\{ \sum_{\substack{p \leqq q \leqq n-1 \\ B \preceq A}} Z^B (f_p f_q)\, \mu_{iB}^{pq} + \sum_{\substack{p=1,\ldots,n-1 \\ B \prec A}} Z^B (f_p g)\, g\, \mu_{iB}^{pn} + \sum_{B \prec A} Z^B (g^2)\, \mu_{iB}^{nn} \Big\} + \ldots, \quad \text{where by } B \prec A$$

we denote that B is a subindex of A and all μ's are in $\bar{C}_m$. To compute $Z^B g$ of RHS apply Z^B to (2.15). Then we get $Z^B g \in C^\omega(x, \bar{f}, Z^C f, Z^D g: C \prec B,\ D \precneqq B)$. Therefore, by induction

$$Z^B g \in C^\omega(x, \bar{f}, \bar{g}, Z^C f, g: C \prec B). \tag{2.20}$$

We also observe that

$$Z^B(f_j f_k \dots f_t) = (Z^B f_j)(f_k \dots f_t) + \dots + (f_j f_k \dots)(Z^B f_t) + \varepsilon(Z^C f\colon |C| < |B|).$$

Substitute the above and (2.20) in (2.19) and use induction on $|A|$ to get

$$Z^{\tilde{A}} f_i \in C^\omega(x, g, \bar{g}, T^t \bar{Z}^B \bar{f}\colon t + |B| \leqq m), \quad \text{for any multi-index } \tilde{A} \tag{2.21}$$

which proves Step 1.

Lemma 3. *Let Z be any C^ω section of $H^{1,0}(M_1)$ and u be any integer $\geqq m$. Then*

$$Z\bar{C}_u \subset \bar{C}_u, \quad \text{or equivalently, } \bar{Z}C_u \subset C_u.$$

Proof of Lemma 3. If suffices to show that for any $i = 1, \dots, n-1$, $Z_i \bar{C}_u \subset \bar{C}_u$. Since $\bar{C}_u = C^\omega(x, f, g, \bar{g}, T^t \bar{Z}^A \bar{f}\,(t + |A| \leqq u))$, it suffices to show that $Z_i f$, $Z_i g$ and $Z_i T^t \bar{Z}^A \bar{f}$ are elements of $\bar{C}_u$. But $Z_i f \in \bar{C}_u$ by Step 1 and that $Z_i g \in \bar{C}_u$ follows from (2.20) and Step 1.

Now consider $Z_i(T^t \bar{Z}^A \bar{f})$, where $t + |A| \leqq u$.

$$Z_i(T^t \bar{Z}^A)\bar{f} = (T^t \bar{Z}^A) Z_j \bar{f} + \sum_{s,B} \gamma_{s,B} T^s \bar{Z}^B \bar{f} \quad (s + |B| \leqq t + |A|),$$

$$(\text{each } \gamma_{s,B} \in C^\omega(x)),$$

where the second term of RHS consists of those terms arising from permuting Z_i with T's and $\bar{Z}$'s. Since $Z_i \bar{f} = 0$ by (2.2) and $t + |A| \leqq u$, we get

$$Z_i(T^t \bar{Z}^A \bar{f}) \in \bar{C}_u.$$

Proof of Step 2. To prove that $C_{u,m} \subset \bar{C}_{m^2+m}$, it suffices to show that for any t and any multi-index A

$$T^t Z^A f \in \bar{C}_{m+tm}. \tag{2.22}$$

Use induction on t. (2.22) holds true for $t = 0$ by Step 1. Assume (2.22) for t and apply Z_1 to (2.22). Then $Z_1 T^t f \in \bar{C}_{m+tm}$ by Lemma 3. Apply $\bar{Z}_{j_1}$ to the above, where j_1 is as in Lemma 1. Then $\bar{Z}_{j_1} Z_1 T^t f \in \bar{C}_{m+tm+1}$. On the other hand,

$$\begin{aligned}
\bar{Z}_{j_1} Z_1 T^t f &= (Z_1 \bar{Z}_{j_1} - [Z_1, \bar{Z}_{j_1}])\, T^t f \\
&\quad \text{substitute (2.12) for } [Z_1, \bar{Z}_{j_1}] \\
&= Z_1 \bar{Z}_{j_1} T^t f - \left(\sqrt{-1}\,\rho_{1\bar{j}_1} T + \sum_{k=1}^{n-1} \alpha^k_{1\bar{j}_1} Z_k + \sum_{k=1}^{n-1} \beta^k_{1\bar{j}_1} \bar{Z}_k\right) T^t f \\
&= -\sqrt{-1}\,\rho_{1\bar{j}_1} T^{t+1} f + \varepsilon(x, T^s Z^B f\colon s \leqq t)
\end{aligned}$$

because $\bar{Z}_i T^t f = T^t \bar{Z}_i f + \varepsilon(x, T^s Z^B f\colon s \leqq t)$ for the same reason as in the proof of Lemma 3) and $T^t \bar{Z}_i f = 0$ by (2.2). By the induction hypothesis $\varepsilon(x, T^s Z^B f\colon s \leqq t) \in \bar{C}_{m+tm}$, therefore we have

$$\rho_{1\bar{j}_1} T^{t+1} f \in \bar{C}_{m+tm+1}. \tag{2.23}$$

Apply $\bar{Z}^{A_1}$ to (2.23), where A_1 is a multi-index as in Lemma 1, then

$$\bar{Z}^{A_1}(\rho_{1\bar{j}_1} T^{t+1} f) \in \bar{C}_{m+(t+1)m} \quad \text{because } |A_1| \leqq m-1. \tag{2.24}$$

$$\text{LHS of (2.24)} = (\bar{Z}^{A_1} \rho_{1\bar{j}_1} + \varepsilon_0(x))\, T^{t+1} f + \varepsilon(x, T^s Z^B f: s \leqq t)$$

and from the choice of A_1 in Lemma 1, $\bar{Z}^{A_1} \rho_{1\bar{j}_1} \neq 0$. Furthermore, by the induction hypothesis, each $T^s Z^B f (s \leqq t) \in \bar{C}_{m+tm}$. Therefore,

$$T^{t+1} f \in \bar{C}_{m+(t+1)m}. \tag{2.25}$$

Now let A be any multi-index and apply Z^A to (2.25). Then by Lemma 3, $Z^A T^{t+1} f \in \bar{C}_{m+(t+1)m}$. On the other hand,

$$Z^A T^{t+1} f = T^{t+1} Z^A f + \varepsilon(x, T^s Z^B f: s + |B| \leqq t + |A|, s \leqq t+1)$$

and by the induction hypothesis

$$T^s Z^B f (s \leqq t) \in \bar{C}_{m+tm}.$$

Therefore,

$$T^{t+1} Z^A f \in \bar{C}_{m+(t+1)m},$$

which proves (2.22).

Proof of Step 3. Suppose first that Z^A involves only one Z, namely, $Z^A = Z_i^s$. Without loss of generality assume that $i=1$. We will show that for any integer $v>0$

$$T^t Z_1^s f (t+s=v) = \varepsilon_{v-1} + \bar{\varepsilon}_{u_1, m} \tag{2.26}$$

for some sufficiently large integer u_1. By Step 1 we have

$$Z_1^v f \in \bar{C}_m. \tag{2.27}$$

Let k be any nonnegative integer $\leqq v$. Apply $\bar{Z}_{j_1}$ to (2.27) k times, then

$$Z_{j_1}^k Z_1^v f \in C_{m+k, m}.$$

On the other hand, by Lemma 2

$$\bar{Z}_{j_1}^k Z_1^v f = \sum_{j=1}^{k-1} \gamma_j T^j Z_1^{v-j} f + (-\sqrt{-1})^k \frac{v!}{(v-k)!} (\rho_{1\bar{j}_1}^k)^k T^k Z_1^{v-k} f + \varepsilon_{v-1},$$

where $\gamma_j \in C^\omega(x)$. Therefore,

$$(\rho_{1\bar{j}_1})^k T^k Z_1^{v-k} f = \sum_{j=1}^{k-1} \beta_j T^j Z_1^{v-j} f + \varepsilon_{v-1} + \bar{\varepsilon}_{m+k, m}, \tag{2.28}$$

where each $\beta_j \in C^\omega(x)$. Multiply (2.28) by $(\rho_{1\bar{j}_1})^{1+2+\ldots+(k-1)}$ to get

$$(\rho_{1\bar{j}_1})^{1+2+\ldots+k} T^k Z_1^{v-k} f = \sum_{j=1}^{k-1} \beta_1 (\rho_{1\bar{j}_1})^{1+2+\ldots+(k-1)} T^j Z_1^{v-j} f + \varepsilon_{v-1} + \bar{\varepsilon}_{m+k, m}. \tag{2.29}$$

By a repeated use of (2.28) for each $j=1,\ldots,k-1$,

$$(\rho_{1\bar{j}_1})^{1+\ldots+j}\, T^j Z_1^{\nu-j} f=\varepsilon_{\nu-1}+\bar{\varepsilon}_{m+j,m}.$$

Therefore, (2.29) becomes

$$(\rho_{1\bar{j}_1})^{1+\ldots+k} T^k Z_1^{\nu-k} f=\varepsilon_{\nu-1}+\bar{\varepsilon}_{m+k,m}. \tag{2.30}$$

Apply $\bar{Z}^{A_1}$ to (2.30) $(1+\ldots+k)$ times. Since $\bar{Z}^{A_1}\varepsilon_{\nu-1}\in C_{\nu-1}$ by Lemma 3 and $\bar{Z}^{A_1}\bar{\varepsilon}_{m+k,m}\in\bar{C}_{m+k+|A_1|,m}$ we have

$$\{(\bar{Z}^{A_1}\rho_{1\bar{j}_1})^{1+\ldots+k}+\varepsilon_0(x)\}\; T^k Z_1^{\nu-k} f+\sum_{j=1}^{\nu-k} a_j T^{k+j} Z_1^{\nu-k-j} f=\varepsilon_{\nu-1}+\bar{\varepsilon}_{u_1,m} \tag{2.31}$$

for some sufficiently large integer u_1, where each $a_j\in C_0^\omega(x)$.

Observe that the coefficient of $T^k Z_1^{\nu-k} f$ in (2.31) is nonzero. Solve the system of analytic equations (2.31) with $k=1,\ldots,\nu$ for $T^k Z_1^{\nu-k} f$, to get (2.26). To complete the proof we use induction on the number of nonzero integers in the multi-index A.

Proof of Step 4. Let t be a nonnegative integer and A be a multi-index such that $t+|A|=m^2+m+1$. If $|A|\neq 0$, by applying T^t to (2.20) we have

$$T^t Z^A g\in C(x, T^s Z^C f, T^s \bar{f}, T^s g, T^s \bar{g}: s\leqq t,\ s+|C|\leqq m^2+m+1). \tag{2.32}$$

Therefore, it suffices to find $T^{m^2+m+1} g$ as a real analytic function of lower order terms. Consider (2.32) with $Z^A=Z_1$, namely,

$$T^{m^2+m} Z_1 g\in C^\omega(x, T^s Z^B f, T^s \bar{f}, T^s g, T^s \bar{g}: s\leqq m^2+m,\ s+|B|\leqq m^2+m+1). \tag{2.32.1}$$

Apply $\bar{Z}_{j_1}$ to (2.32.1), where j_1 is as in Lemma 1. Then for the same reason as in the proof of Lemma 3, LHS of (2.32.1) becomes

$$-\sqrt{-1}\,\rho_{1\bar{j}_1} T^{m^2+m+1} g+\varepsilon(x, T^s Z^B g: s+|B|\leqq m^2+m+1,\ s\leqq m^2+m)$$

and by Lemma 2 RHS of (2.32.1) becomes

$$\begin{aligned}&\varepsilon(x, T^s Z^B f(s+|B|\leqq m^2+m+1),\quad T^s \bar{Z}^B \bar{f}(s\leqq m^2+m),\\ &\quad T^s Z^B g(s+|B|\leqq m^2+m+1),\quad T^s \bar{Z}^B g(s\leqq m^2+m)).\end{aligned}$$

By equating two sides and using (2.32) Step 2 and Step 3, we get

$$\rho_{1\bar{j}_1} T^{m^2+m+1} g\in C^\omega(x, T^s Z^B f, T^s \bar{Z}^B \bar{f}, T^s g, T^s \bar{g}: s+|B|\leqq m^2+m+1,\ s\leqq m^2+m). \tag{2.33}$$

Apply $\bar{Z}^{A_1}$ to (2.33), where A_1 is as in Lemma 1. Then LHS of (2.33) becomes

$$(\bar{Z}^{A_1}\rho_{1\bar{j}_1}+\varepsilon_0(x))\, T^{m^2+m+1} g+\varepsilon(x, T^s Z^B g: s+|B|\leqq m^2+m+1,\ s\leqq m^2+m)$$

and RHS of (2.33) becomes by Lemma 2

$$\begin{aligned}&C^\omega(x, T^s Z^B f(s+|B|\leqq m^2+m+1),\ T^s \bar{Z}^B \bar{f}(s\leqq m^2+m),\\ &\quad T^s Z^B g(s+|B|\leqq m^2+m),\ T^s \bar{Z}^B \bar{g}(s\leqq m^2+m)),\end{aligned}$$

which reduces to analytic functions of derivatives of order $\leqq m^2+m$ by (2.32), Step 2 and Step 3.

Since $\bar{Z}^{A_1}\rho_{1\bar{J}_1}\neq 0$, we get

$$T^{m^2+m+1}g\in C^\omega(x, T^sZ^Bf(s+|B|\leqq m^2+m),\ T^s\bar{Z}^B\bar{f}(s+|B|\leqq m^2+m),\\ T^sg(s\leqq m^2+m),\ T^s\bar{g}(s\leqq m^2+m)), \tag{2.34}$$

which proves Step 4, and concludes the proof of Theorem 1.

Example 3. Let M be a hypersurface in $\mathbb{C}^2$ defined by $r=0$, where

$$r=z_2+\bar{z}_2+h(z_2,\bar{z}_2)\{|z_1|^4+\varepsilon(\bar{z}_1z_1^3+z_1\bar{z}_1^3)\}, \tag{2.35}$$

where ε is a real number and $h\in C^\omega$ is a real valued function such that $h(0)=1$. Observe that M satisfies the hypothesis of Theorem 1 if and only if $\varepsilon\neq 0$ and M is pseudoconvex if $|\varepsilon|$ is sufficiently small. Consider the case $\varepsilon=0$. Let F be a CR equivalence of C^ω hypersurface M_1 onto another hypersurface M_2 defined by (2.35) with $\varepsilon=0$. Write $F=(f,g)$ coordinatewise and let $\tilde{f}=f^2$. Then a similar argument as in the proof of Theorem 1 can be applied to $\tilde{f}$ and g. In this case, one will be able to express all the partial derivatives of $\tilde{f}$ and g of order 31 as C^ω functions of derivatives of lower order by using the fact that

$$\bar{C}_2\subset C_{2+2^2}\subset\bar{C}_{2+2^2+2^3}\subset C_{2+2^2+2^3+2^4}=C_{30}.$$

However, actual computation is very complicated. If h is constant, M satisfies the hypotheses of Theorem 2.

§ 3. Regular Hypersurfaces

In this section we will prove Theorem 2 and give an example. A C^ω real hypersurface M in $\mathbb{C}^n$ is said to be regular at $0\in M$ if on a neighborhood of 0 in M there exists a C^ω infinitesimal CR automorphism K which is transversal to $H(M)$ (cf. [27]).

Lemma 4. (Proposition 4 of [27]). *Suppose that M is a C^ω real hypersurface in $\mathbb{C}^n$ which is regular at $0\in M$. Then there exist local coordinates $(z_1,\ldots,z_n)$ near 0 so that M is locally the graph*

$$2\operatorname{Re}z_n=\psi(z_1,\bar{z}_1,\ldots,z_{n-1},\bar{z}_{n-1})\quad \text{for some } \psi\in C^\omega$$

such that $\dfrac{\partial\psi}{\partial z_i}(0)=0$ $(i=1,\ldots,n-1)$ *and*

$$K=\frac{\partial}{\partial y_n},\quad \text{where } z_n=x_n+\sqrt{-1}\,y_n.$$

Lemma 5. *Suppose that M is a C^ω real hypersurface in $\mathbb{C}^n$ with non-degenerate Levi form at $P\in M$. Then the vector space of germs at P of C^4 infinitesimal CR automorphisms of M is finite dimensional* (cf. [8, 17]).

Now let X be a C^∞ infinitesimal CR automorphism defined on an open set $U \subset M$ and Y be a C^∞ section of $H(U)$. Then the defining property of X implies that $[X, JY] = J[X, Y]$, where J is the almost complex structure of $\mathbb{C}^n$ and $[\,,\,]$ is the Lie bracket. Let $Z = \frac{1}{2}(Y - \sqrt{-1}\,JY)$. Then

$$\begin{aligned}[X, Z] &\equiv \tfrac{1}{2}\{[X, Y] - \sqrt{-1}\,[X, JY]\} \\ &= \tfrac{1}{2}\{[X, Y] - \sqrt{-1}\,J[X, Y]\},\end{aligned}$$

therefore, $[X, Y]$ is a C^∞ vector field of type $(1,0)$. Thus for a given basis $\{Z_j\}_{j=1,\ldots,n-1}$ of $H^{1,0}(U)$, we have

$$[X, Z_k] = \sum_{j=1}^{n-1} \alpha_k^j Z_j, \alpha_k^j \in C^\infty(U), \qquad k = 1, \ldots, n-1. \tag{3.0}$$

Proof of Theorem 2. By Lemma 5, d_i = finite. Let $O_1 \in M_1$ be an arbitrary point and let $O_2 = F(O_1)$. For each $i = 1, 2$ let K_i be an infinitesimal CR automorphism on M_i as in the hypothesis. Let $(z_1, \ldots, z_n)$ and $(z'_1, \ldots, z'_n)$ be local coordinates at O_1 and O_2 respectively, as in Lemma 4. Define

$$L_j = \frac{\partial}{\partial z_j} + \frac{\partial \psi}{\partial z_j}\frac{\partial}{\partial z_n}, \qquad j = 1, \ldots, n-1.$$

Then L_j is a linearly independent set of holomorphic vector fields tangent to M_1. We use $(z_1, \ldots, z_{n-1}, \operatorname{Im} z_n)$ as local coordinates of M_1 near O_1, more precisely, let $\pi: M \to \mathbb{C}^{n-1} \times \mathbb{R}$ be the mapping

$$(z_1, \ldots, z_{n-1}, \psi(z_1, \bar{z}_1, \ldots, z_{n-1}, \bar{z}_{n-1}) + \sqrt{-1}\,y_n) \xrightarrow{\pi} (z_1, \ldots, z_{n-1}, y_n)$$

and let

$$\zeta_j = z_j \circ \pi, \qquad j = 1, \ldots, n-1$$

and

$$t = y_n \circ \pi, \qquad \text{then } (\zeta_1, \ldots, \zeta_{n-1}, t)$$

is a local coordinate system of M_1 near O_1. The coordinate fields are

$$\frac{\partial}{\partial \zeta_j} = \frac{\partial}{\partial z_j} + \frac{1}{2}\frac{\partial \psi}{\partial z_j}\frac{\partial}{\partial x_n}$$

and

$$\frac{\partial}{\partial t} = \frac{\partial}{\partial y_n} = K_1.$$

Then L_j in terms of the coordinate fields is

$$L_j = \frac{\partial}{\partial \zeta_j} - \frac{1}{2}\sqrt{-1}\frac{\partial \psi}{\partial z_j}\frac{\partial}{\partial t}, \qquad \text{on } M_1. \tag{3.1}$$

We put primes on the corresponding notions of M_2. Now write $F = (f_1, \ldots, f_{n-1}, g)$ coordinatewise with respect to the coordinates $(\zeta'_1, \ldots, \zeta'_{n-1}, t')$ of M_2 near O_2. Then

$$F_*(L_k)=\sum_{j=1}^{n-1}\left\{(L_k f_j)\circ F^{-1}\frac{\partial}{\partial\zeta'_j}+(L_k\bar{f}_j)\circ F^{-1}\frac{\partial}{\partial\bar{\zeta}'_j}\right\}+(L_k g)\circ F^{-1}\frac{\partial}{\partial t'}. \tag{3.2}$$

Since F is a CR equivalence,

$$F_*(L_k)=\sum_{j=1}^{n-1}\gamma_k^j L'_j, \quad \text{each } \gamma\in C^{m-1}. \tag{3.3}$$

Substitute (3.1) for L'_j and (3.2) for $F_*(L_k)$ to get

$$\begin{aligned}&\sum_{j=1}^{n-1}\left\{(L_k f_j)\circ F^{-1}\frac{\partial}{\partial\zeta'_j}+(L_k\bar{f}_j)\circ F^{-1}\frac{\partial}{\partial\bar{\zeta}'_j}\right\}+(L_k g)\circ F^{-1}\frac{\partial}{\partial t'}\\&\quad=\sum_{j=1}^{n-1}\gamma_k^j\left(\frac{\partial}{\partial\zeta'_j}-\frac{1}{2}\sqrt{-1}\frac{\partial\psi'}{\partial z'_j}\frac{\partial}{\partial t'}\right).\end{aligned} \tag{3.4}$$

By equating the corresponding components in (3.4), we get

$$(L_k f_j)\circ F^{-1}=\gamma_k^j, \tag{3.5}$$

$$(L_k\bar{f}_j)\circ F^{-1}=0, \tag{3.6}$$

$$(L_k g)\circ F^{-1}=-\tfrac{1}{2}\sqrt{-1}\sum_{j=1}^{n-1}\gamma_k^j\frac{\partial\psi'}{\partial z'_j}. \tag{3.7}$$

Now let $\{X_1=K_1,X_2,\ldots,X_{d_1}\}$ be a basis of C^∞ infinitesimal CR automorphisms of M_1 such that each $X_j(j=2,\ldots,d_1)$ is tangent to $H(M_1)$ at O_1. Since F is a C^m CR equivalence, $\{F_*(X_j):j=1,\ldots,d_1\}$ is a set of independent C^{m-1} infinitesimal CR automorphisms of M_2. By Lemma 6 at the end of the proof each F_*X_j is real analytic. Therefore, $d_1\leqq d_2$. Same argument applied to F^{-1} shows that $d_2\leqq d_1$. Thus, $d_1=d_2=d$.

Now we will use $\{K_2,F_*(X_2),F_*(X_3),\ldots,F_*(X_d)\}$ as a basis of C^∞ infinitesimal CR automorphisms of M_2. Since $F_*(K_1)$ is a C^ω infinitesimal CR automorphism of M_2, express this in terms of the basis chosen above:

$$\begin{aligned}F_*(K_1)&=c_1K_2+\sum_{j=2}^{d}c_jF_*(X_j)\\&=c_1\frac{\partial}{\partial t'}+\sum_{\substack{j=2,\ldots,d\\k=1,\ldots,n-1}}c_j\left\{(X_jf_k)\circ F^{-1}\frac{\partial}{\partial\zeta'_k}+(X_j\bar{f}_k)\circ F^{-1}\frac{\partial}{\partial\bar{\zeta}'_k}+(X_jg)\circ F^{-1}\frac{\partial}{\partial t'}\right\}\end{aligned} \tag{3.8}$$

where each c_j is a constant. On the other hand

$$F_*(K_1)=\sum_{j=1}^{n-1}(K_1f_j)\circ F^{-1}\frac{\partial}{\partial\zeta'_j}+\sum_{j=1}^{n-1}(K_1\bar{f}_j)\circ F^{-1}\frac{\partial}{\partial\bar{\zeta}'_j}+K_1g\circ F^{-1}\frac{\partial}{\partial t'}. \tag{3.9}$$

Equate the corresponding components of (3.8) and (3.9) to get

$$K_1f_j=\left(\sum_{i=2}^{d}c_iX_i\right)f_j, \quad j=1,\ldots,n-1 \tag{3.10}$$

and

$$K_1 g = c_1 + \sum_{i=2}^{d} (c_i X_i) g \tag{3.11}$$

(3.6) and (3.10) give a system of equations for each f_j, $j=1,\ldots,n-1$:

$$\begin{aligned} &\bar{L}_k f_j = 0, \quad k=1,\ldots,n-1, \\ &K_1 f_j = \left(\sum_{i=2}^{d} c_i X_i\right) f_j. \end{aligned} \tag{3.12}$$

Now apply τL_k ($\tau \gg 0$) to the first equation of (3.12) and apply K_1 to the second equation of (3.12) to get

$$\begin{aligned} &\tau L_k \bar{L}_k f_j = 0 \quad k=1,\ldots,n-1 \\ &K_1^2 f_j = \sum_{i=2}^{d} c_i K_1 (X_i f_j). \end{aligned} \tag{3.13}$$

By Lemma 6 below, X_k ($k=2,\ldots,d$) is C^ω and therefore, all the coefficients of (3.13) are C^ω. Let $\xi=(\xi_1,\ldots,\xi_{2n-1})$ be an element of $T^*_{O_1}(M_1)$ expressed with respect to the local coordinates $(\zeta_1,\ldots,\zeta_{n-1},t)$. The principal symbol of (3.13) is

$$P_2(O_1,\xi) = \tfrac{1}{4}\tau \sum_{k=1}^{2(n-1)} \xi_k^2 + \xi_{2n-1}^2 + \sum_{q=1}^{2n-2} a_q(O_1)\,\xi_{2n-1}\,\xi_q,$$

where each a_q is a real analytic function. If we take $\tau>0$ sufficiently large, then $P_2(O_1,\xi)\neq 0$ for any $\xi\neq 0$. This proves that each f_j ($j=1,\ldots,n-1$) is real analytic. To show analyticity of g, use (3.7) and (3.11): Substitute (3.5) for γ_k^j in (3.7) to get

$$L_k g = -\frac{\sqrt{-1}}{2} \sum_{j=1}^{n-1} (L_k f_j) \left(\frac{\partial \psi'}{\partial z_j'} \circ F\right). \tag{3.14}$$

But each $\dfrac{\partial \psi'}{\partial z_j'} \circ F$ ($j=1,\ldots,n-1$) is C^ω because $\dfrac{\partial \psi'}{\partial z_j'}$ is independent of y_n' and each f_j is C^ω. Now (3.14) and (3.11) give a system of equations for g:

$$\begin{aligned} &L_k g = C^\omega \quad \text{function} \\ &K_1 g = c_1 + \left(\sum_{j=2}^{d} c_j X_j\right) g. \end{aligned} \tag{3.15}$$

Again by Lemma 6, each X_j is C^ω and repeat the same argument as that below (3.13) to conclude that g is C^ω. This completes the proof of Theorem 2 up to the following lemma.

Lemma 6. *Let M be a C^ω real hypersurface in $\mathbb{C}^n$. Suppose that there exists a C^ω infinitesimal CR automorphism K which is transversal to $H(M)$ and that the set of nondegenerate points is dense in M. Then $\delta \equiv$ dimension of the vector space*

of the C^4 infinitesimal CR automorphisms is finite. Furthermore, any $C^{\delta+4}$ infinitesimal CR automorphism X is real analytic.

Proof of Lemma 6. By Lemma 5, δ is finite. On an open set $U \subset M$ choose a C^ω basis $\{Z_j\}_{j=1,\ldots,n-1}$ of $H^{1,0}(U)$. Let

$$X = \sum_{j=1}^{n-1} u_j Z_j + \sum_{j=1}^{n-1} \bar{u}_j \bar{Z}_j + hK, \quad h\text{: real} \tag{3.16}$$

and let

$$[Z_i, Z_j] = \sum_{k=1}^{n-1} a_j^k Z_k, \tag{3.17}$$

$$[Z_i, \bar{Z}_j] = \rho_{ij} K + \sum_{k=1}^{n-1} b_{jj}^k Z_k + \sum_{k=1}^{n-1} c_{ij}^k \bar{Z}_k \tag{3.18}$$

and

$$[K, Z_j] = \sum_{k=1}^{n-1} \beta_j^k Z_k. \tag{3.19}$$

Now substitute (3.16) for X in (3.0). Then by (3.17)–(3.19) we get

$$\begin{aligned}&(-Z_j u_k + u_i a_{ij}^k - \bar{u}_i b_{ji}^k + h\beta_j^k) Z_k + (-Z_j \bar{u}_k - \bar{u}_i c_{ji}^k) Z_k \\ &+ (-Z_h h - \bar{u}_i \rho_{ji}) K = \alpha_j^k Z_k \quad \text{(summation convention).}\end{aligned} \tag{3.20}$$

By comparing the corresponding components we get

$$\bar{Z}_j u_k + u_i \bar{c}_{ji}^k = 0 \tag{3.21}$$

and

$$Z_j h + \bar{u}_i \rho_{ji} = 0 \quad \text{(summation convention).} \tag{3.22}$$

Now consider a sequence of infinitesimal CR automorphisms

$$\{X_k\}_{k=0,1,\ldots,} \quad \text{where } X_0 = X,\ X_i = [K, X_{i-1}],\ i \geqq 1.$$

Since the Lie algebra of C^4 infinitesimal CR automorphisms is of dimension δ, there exists an integer $N \leqq \delta$ such that X_N is a linear combination of $\{X_0, \ldots, X_{N-1}\}$. Set

$$X_N = \sum_{k=1}^{N-1} c_k X_k, \quad \text{where } c_k\text{'s are constants.} \tag{3.23}$$

We introduce a notation: By $\langle K^s u_1 \rangle_{s<3}$ for instance, we denote a linear combination of the form $a_0 u_1 + a_1 K u_1 + a_2 K^2 u_1$, where each a_j is a known C^ω function. Then

$$X_1 = [K, X]$$

substitute (3.16) for X, substitute (3.19) for $[K, Z_j]$ and simplify

$$\begin{aligned}&= (K u_j + u_i \beta_i^j) Z_j + (K \bar{u}_j + \bar{u}_i \bar{\beta}_i^j) \bar{Z}_j + (Kh) K \\ &\quad \text{(summation convention).}\end{aligned}$$

By induction on k, we get

$$X_k=\sum_{j=1}^{n-1}\{K^k u_j+\langle K^s u_i\rangle_{s<k,\,i=1,\ldots,n-1}\}Z_j$$
$$+\sum_{j=1}^{n-1}\{K^k \bar{u}_j+\langle K^s \bar{u}_i\rangle_{s<k,\,i=1,\ldots,n-1}\}\bar{Z}_j$$
$$+(K^k h)K.$$

Now substitute (3.24) in (3.23) to get

$$\{K^N u_j+\langle K^s u_i\rangle_{s<N,\,i=1,\ldots,n-1}\}Z_j$$
$$+\{K^N \bar{u}_j+\langle K^s \bar{u}_i\rangle_{s<N,\,i=1,\ldots,n-1}\}\bar{Z}_j$$
$$+\{K^N h+\langle K^{N-1}h,\ldots,h\rangle\}K=0,\quad\text{(summation convention).}\tag{3.25}$$

Thus we get

$$K^N u_j=\langle K^s u_i\rangle_{s<N,\,i=1,\ldots,n-1},\tag{3.26}$$
$$K^N h=\langle K^{N-1}h,\ldots,h\rangle.\tag{3.27}$$

Note that $N\geqq 1$. Using the same argument as below (3.13), we can construct from (3.21), (3.22), (3.26) and (3.27) a system of elliptic linear partial differential equations of order N if N is even or $N+1$ if N is odd. Therefore, u_j $(j=1,\ldots,n-1)$ and h are real analytic.

Example 4. Let D be a Reinhardt domain in $\mathbb{C}^n$. If ∂D is real analytic, it is locally defined by

$$r(z_1,\bar{z}_1,\ldots,z_n,\bar{z}_n)=\psi(|z_1|^2,\ldots,|z_n|^2)=0$$

with $\operatorname{grad} r\neq 0$ on ∂D. At each point of ∂D there are infinitesimal CR automorphisms

$$\frac{\partial}{\partial\theta_\alpha}\equiv\sqrt{-1}\left(z_\alpha\frac{\partial}{\partial z_\alpha}-\bar{z}_\alpha\frac{\partial}{\partial\bar{z}_\alpha}\right),\qquad \alpha=1,\ldots,n.$$

Since $\operatorname{grad} r=2\sum_{\beta=1}^{n}\psi_\beta\left(z_\beta\frac{\partial}{\partial z_\beta}+\bar{z}_\beta\frac{\partial}{\partial\bar{z}_\beta}\right)\neq 0$ on ∂D and

$$\sqrt{-1}\,\partial r\left(\frac{\partial}{\partial\theta_\alpha}\right)=\sqrt{-1}\left(\sum_{\beta=1}^{n}\psi_\beta\bar{z}_\beta\,dz_\beta\right)\left(\frac{\partial}{\partial\theta_\alpha}\right)$$
$$=-\psi_\alpha|z_\alpha|^2,\qquad \alpha=1,\ldots,n,$$

we see that for some α, $\sqrt{-1}\,\partial r\left(\frac{\partial}{\partial\theta_\alpha}\right)\neq 0$, i.e., $\frac{\partial}{\partial\theta_\alpha}$ is transversal to $H(M)$ $=\operatorname{Ker}(\sqrt{-1}\,\partial r)$. Therefore ∂D is a regular hypersurface.

Reference

1. Bedford, E., Fornaess, J.: Biholomorphic maps of weakly pseudoconvex domains. Duke Math. J. **45**, 711–719 (1980)
2. Bell, S.: The Bergman kernel function and proper holomorphic mappings. preprint (1981)
3. Bell, S.: Analytic hypo-ellipticity of the $\bar{\partial}$-Neumann problem and extendability of holomorphic mappings. Acta Math. **147**, 109–116 (1981)
4. Bell, S., Boas, H.: Regularity of the Bergman projection in weakly pseudoconvex domains. Math. Ann. **257**, 23–30 (1981)
5. Bell, S., Ligocka, E.: A simplification and extension of Fefferman's theorem on biholomorphic mappings. Invent. Math. **57**, 283–289 (1980)
6. Bloom, T., Graham, I.: A geometric characterization of points of type m on real submanifolds of $\mathbb{C}^n$. J. Diff. Geom. **12**, 171–182 (1977)
7. Burns, D.: Pseudoconformal geometry. Lecture Note, University of Michigan 1980
8. Burns, D., Shnider, S.: Real hypersurfaces in complex manifolds. Proc. Symp. Pure Math. **30**, 141–167 (1976)
9. Cartan, E.: Sur la géométrie pseudo-conforme des hypersurfaces de deux variables complexes, I, II. Oeuvres II, **2**, 1231–1304 (1952), Oeuvres III, **2**, 1217–1238 (1952)
10. Chern, S.S., Moser, J.K.: Real hypersurfaces in complex manifolds. Acta Math. **133**, 219–271 (1974)
11. D'Angelo, J.: Real hypersurfaces, orders of contact and applications. preprint (1981)
12. Diederich, K., Fornaess, J.: Biholomorphic mappings between certain real analytic domains in $\mathbb{C}^2$. Math. Ann. **245**, 255–272 (1979)
13. Diederich, K., Fornaess, J.: Pseudoconvex domains with real analytic boundary. Ann. of Math. **107**, 371–384 (1978)
14. Diederich, K., Webster, S.: A reflection principle for degenerate real hypersurfaces. Duke Math. J. **47**, (No. 4) 835–843 (1980)
15. Fefferman, C.: The Bergman kernel and biholomorphic mappings of pseudoconvex domains. Invent. Math. **26**, 1–65 (1974)
16. Fornaess, J.: Biholomorphic mappings between pseudoconvex domains. Pac. J. Math. **74**, 63–65 (1978)
17. Goldschmidt, H., Rossi, H.: Notes on E. Cartan's computation, personal note
18. John, F.: Plane waves and spherical means applied to partial differential equations. New York: Interscience Publishers 1955
19. Kohn, J.J.: Boundary behavior of $\bar{\partial}$ on weakly pseudoconvex manifolds of dimension two. J. Diff. Geom. **6**, 523–542 (1972)
20. Kohn, J.J.: Subellipticity of the $\bar{\partial}$-Neumann problem on pseudoconvex domains: sufficient conditions. Acta Math. **142**, 79–122 (1979)
21. Kuranishi, M.: On E. Cartan's prolongation theorem of exterior differential systems. Amer. J. Math. **79**, 1–47 (1957)
22. Kuranishi, M.: Lectures on involutive systems of partial differential equations. Publ. Soc. Math. Sao Paulo 1967
23. Morrey, C.B., Nirenberg, L.: On the analyticity of the solutions of linear systems of partial differential equations. Comm. Pure Appl. Math. **10**, 271–290 (1957)
24. Nirenberg, L., Webster, S., Yang, P.: Local boundary regularity of holomorphic mappings. Comm. Pure Appl. Math. **33**, (No. 3) 305–338 (1980)
25. Pincuk, S.: On the analytic continuation of holomorphic mappings. Math. USSR Sbornik **27**, (No. 3) 375–392 (1975)
26. Pincuk, S.: On holomorphic mappings of real analytic hypersurfaces. Math. USSR Sbornik **34**, (No. 4) 503–519 (1978)
27. Tanaka, N.: On the pseudoconformal geometry of hypersurfaces of the space of n complex variables. J. Math. Soc. Japan **14**, 397–429 (1962)

Oblatum 11-XI-1982

Invent. math. 73, 71-106 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Einstein Metrics and Complex Structures

N. Koiso*

Department of Mathematics, Osaka University, Toyonaka, Osaka 560, Japan

0. Main Results

The main purpose of this paper is to point out some similarities between Einstein metrics and complex structures. The results may be regarded as generalizations of some facts which hold on 2-dimensional manifolds. First we consider an Einstein metric and a complex structure. Recall a classical result of Newlander and Nirenberg.

0.1. **Proposition.** *Any integrable almost complex structure (i.e., an almost complex structure whose Nijenhuis torsion tensor vanishes) defines a complex structure.*

Of course, we cannot say that an Einstein metric is "holomorphic". The following result is the best we can expect in general.

0.2. **Proposition** ([12, Theorem 5.2]). *Any C^r-Einstein metric ($r \geqq 2$) is real analytic with respect to some real analytic structure compatible with the original C^r-differentiable structure.*

On a two or three dimensional manifold, this fact is obvious because in these cases Einstein metrics have constant sectional curvature. Moreover, we know the following

0.3. **Fact.** *Any (Einstein) metric on an orientable 2-dimensional manifold is Kählerian, that is, there exists a complex structure such that the metric becomes a compatible Kähler metric.*

And conversely,

0.4. **Fact.** *Any complex structure on a 2-dimensional manifold admits a compatible Kähler-Einstein metric.*

---

* This research was done when the author was supported by the Sakkokai Foundation and invited to Ecole Polytechnique

Fact 0.4 was generalized to higher dimensions for some compact manifolds by Calabi, Aubin and Yau.

0.5. **Proposition** ([3], [4, Théorème 4], [31, p. 364 Theorem 2]). *Let J be a complex structure with negative or vanishing first Chern class. If J admits a compatible Kähler metric, then it also admits a compatible Kähler-Einstein metric.*

This result suggests some strong relationship between Einstein metrics and complex structures. But it seems that Fact 0.3 is regarded as that which holds characteristically in the 2-dimensional case. In fact, on an odd dimensional manifold, we can expect nothing more than Proposition 0.2. However, Hitchin obtained the following

0.6. **Proposition** ([18, Remark 2.2]). *Any Einstein metric on the K3-surface is Kählerian.*

Here, *the K3-surface* is a 4-dimensional C^∞-manifold which is defined by the equation: $\sum_{i=1}^{4}(z^i)^4=0$ in $P^3(\mathbf{C})$. We will give a weak generalization of this result.

0.7. **Theorem** (Theorem 10.5). *Let (J, g) be a Kähler-Einstein structure on a compact manifold M. Assume that the first Chern class is non-positive and that the local deformation space of the complex structure J coincides with an open set of the cohomology group $H^1(M, \Theta)$ with coefficient in the sheaf Θ of germs of holomorphic vector fields. Then any Einstein metric g_1 on M sufficiently close to the metric g is Kählerian.*

It is an interesting problem: *Are the assumptions for the local deformation space and that g_1 is close to g really necessary?* We may say that Theorem 0.7 is a kind of converse to Proposition 0.5. Next, we consider families of structures.

0.8. **Proposition** ([25]). *The space of all complex structures on a manifold locally forms a complex analytic set.*

For the meaning of this proposition, see Definition 2.7 of the local pre-moduli space of Einstein metrics. We will give a corresponding result for families of Einstein metrics.

0.9. **Theorem** (Theorem 3.1). *The local pre-moduli space of Einstein metrics forms a real analytic set.*

As a corollary, we will get

0.10. **Theorem** (Theorem 3.2). *Let g be an Einstein metric. The three notions: to be non-deformable, to be formally non-deformable and to be rigid are equivalent.*

By analogy with the construction of complex analytic structure on the space of complex structures in Proposition 0.8, we can construct a canonical riemannian metric on a family of riemannian metrics. Combining Propositions 0.5, 0.8 and Theorem 0.7, we have families of Kähler-Einstein structures which has a complex structure and a riemannian metric.

0.11. **Theorem** (Theorem 12.3). *Let* (J_t, g_t) *be a normal and stable family* (Definition 12.1) *of Kähler-Einstein structures. Then the canonical riemannian metric is a Kähler metric compatible with the complex structure introduced in Proposition* 0.8.

In the 2-dimensional case, the family is called *the Teichmüller space* and the canonical riemannian metric is known as *the Peterson-Weil metric.* We will also see that the canonical riemannian metric on a local pre-moduli space of Einstein metrics is real analytic with respect to the real analytic structure introduced in Theorem 0.9. So we can expect that this riemannian metric satisfies some elliptic equation. For example, we may ask, "*Is the canonical riemannian metric an Einstein metric?*" This problem is rather difficult except on a family of flat riemannian metrics. Even in the 2-dimensional case, it remains open.

An important difference between the theory of deformations of Einstein metrics and that of complex structures lies in the obstruction for integrability of infinitesimal deformations.

0.12. **Proposition** ([19, p. 452 Theorem]). *Let* J *be a complex structure on* M. *If* $H^2(M, \Theta)=0$, *then for any infinitesimal deformation* I *of* J, *there exists an actual deformation of* J *whose infinitesimal deformation coincides with* I.

For the theory of complex deformations, we may say that the space $H^2(M, \Theta)$ is *the obstruction space.* For the theory of Einstein deformations, we will give a similar but negative result (Theorem 0.13). It seems to be impossible to say that if some space defined solely in terms of a given Einstein metric g vanishes, then, for any infinitesimal deformation h of g, there exists an Einstein deformation of g whose infinitesimal deformation coincides with h.

0.13. **Theorem** (Proposition 5.4). *The obstruction space for the space* EEID *of all essential infinitesimal Einstein deformations is the space* EEID *itself.*

Table of Contents

1. Deformations and Infinitesimal Deformations of an Einstein Metric (Fundamental Definitions) 74
2. Moduli Space of Einstein Metrics (Definition) 76
3. Real Analyticity of the Local Pre-Moduli Space of Einstein Metrics (Theorems 0.9, 0.10) . 78
4. The Canonical Riemannian Metric on the Local Pre-Moduli Space of Einstein Metrics (Definition) 80
5. The Obstruction Space for the Space EEID (Theorem 0.13) 85
6. Deformations and Infinitesimal Deformations of a Complex Structure (Definitions: Different Way from that of Kodaira and Spencer) 87
7. The Space EEID on a Kähler-Einstein Manifold 89
8. The Space CEID on a Kähler-Einstein Manifold 91
9. Kähler Relation between the Spaces EEID and CEID (in 7, 8, 9, we treat the Space of Infinitesimal Deformations) 92
10. Einstein Metrics and Complex Structures (Theorem 0.7) 95
11. The Complex Structure on a Family of Complex Structures (Proposition 0.8) 99
12. The Metric on a Family of Kähler-Einstein Structures (Theorem 0.11) 101
13. Appendix (Proof of Theorem 0.9) 104

The author would like to express his sincere gratitude to Professors M. Berger, A.L. Besse, J.-P. Bourguignon, J.L. Kazdan, J. Lafontaine, D. Meyer, and P. Pansu for many useful discussions and their encouragements.

1. Deformations and Infinitesimal Deformations of an Einstein Metric

First, we introduce some fundamental definitions and facts due to [8]. Throughout this paper, M *denotes* a compact connected n-dimensional C^∞-manifold without boundary. All objects are assumed to belong to the real C^∞-category, unless otherwise stated. By *a family of geometric structures on* M we mean a family of structures which depends smoothly on t, where t runs through an open set of a Euclidean space $\mathbf{R}^k$ or more generally a manifold. This open set is called *the parameter space.*

1.1. *Definition.* Let g be an Einstein metric on M with volume 1. A family g_t of Einstein metrics on M with volume 1 such that $g_0=g$ is called *an Einstein deformation of* g.

Note that any family g_t of Einstein metrics on M can be reduced to an Einstein deformation by multiplying it by some function of t.

1.2. *Definition.* Let g_t be an Einstein deformation of an Einstein metric g on M. If there is a family γ_t of diffeomorphisms of M such that $g_t=\gamma_t^* g$, then g_t is said to be *trivial.*

Let g_t be an Einstein deformation of g with parameter space P. Then each g_t satisfies *the Einstein equation:*

$$E(g_t)\equiv r_t-(\int_{g_t} u_t/n)\, g_t=0, \tag{1.2.1}$$

where r_t is *the Ricci tensor*, u_t *the scalar curvature* and *the operator* $\int_g$ is defined by

$$\int_g f=\int_M f v_g, \tag{1.2.2}$$

v_g being *the volume element defined by* g. If we differentiate the Eq. (1.2.1) with respect to t, then we get a second order linear differential equation:

$$E'_g(h)\equiv(1/2)(\bar{\Delta}+2L-2\delta^*\delta-\mathrm{Hess}\,\mathrm{tr})\,h=0 \tag{1.2.3}$$

for $h=v[g_t]$, $v\in T_0P$, where the operators are defined by

$$(\bar{\Delta}\psi)_{ij}=-D^m D_m\psi_{ij}, \tag{1.2.4}$$

$$(L\psi)_{ij}=R_i{}^k{}_j{}^m\psi_{km}, \tag{1.2.5}$$

$$(\delta\psi)_i=-D^m\psi_{mi} \tag{1.2.6}$$

for bilinear forms ψ (which need not be symmetric), and

$$(\delta^*\xi)_{ij}=(1/2)(D_i\xi_j+D_j\xi_i) \tag{1.2.7}$$

for 1-forms ξ, D being *the covariant derivative* and the sign convention of the curvature tensor R is taken so that $R_{ijij} \leqq 0$ for the standard sphere.

1.3. *Definition.* Let g be an Einstein metric on M with volume 1. A symmetric bilinear form h is called *an infinitesimal Einstein deformation of* g if h satisfies the Eq. (1.2.3) and the following equation (to preserve the volume):

$$\int_g \operatorname{tr} h = 0. \tag{1.3.1}$$

The space of all infinitesimal Einstein deformations of g is *denoted by* EID(g) or simply EID.

Remark that the equation

$$L_X g = 2\delta_g^* X, \tag{1.3.2}$$

where L_X denotes the Lie derivative, holds under the canonical identification between 1-forms and vector fields given by the metric g.

1.4. *Definition.* Let g be an Einstein metric on M with volume 1. An infinitesimal Einstein deformation of g of the form $L_X g$ is said to be *trivial.* The space of all trivial infinitesimal Einstein deformations is *denoted by* ETID(g) or simply ETID. An infinitesimal Einstein deformation h of g is said to be *essential* if h is orthogonal to the space ETID with respect to the global inner product defined by g. The space of all essential infinitesimal Einstein deformations of g is *denoted by* EEID(g) or simply EEID. (See Definition 2.3.)

The three equations in the following Lemma may be regarded as defining the space EEID.

1.5. **Lemma** ([8, Lemma 7.1, (7.1)]). *Let g be an Einstein metric on M with volume* 1. *A symmetric bilinear form h is an element of* EEID(g) *if and only if h satisfies the following equations.*

$$(\bar{\Delta} + 2L)h = 0, \tag{1.5.1}$$

$$\delta h = 0, \tag{1.5.2}$$

$$\operatorname{tr} h = 0. \tag{1.5.3}$$

In particular, the space EEID *is finite dimensional.*

1.6. *Remark.* By definition, the isometry group $I(g)$ of g acts on the space EEID(g). This action induces an action of the space $K(g)$ of all Killing vector fields.

1.7. *Example.* Let (T^n, g) be a flat torus. Then the equations in Lemma 1.5 reduce to $Dh = 0$ and $\operatorname{tr} h = 0$. Thus $\dim \mathrm{EEID}(g) = n(n+1)/2 - 1$. For any infinitesimal Einstein deformation h, there is an Einstein deformation g_t such that $g_0' = h$. All Einstein deformations g_t of g are families of flat metrics ([8, Addition 8.1], [7, Proposition 3.2]).

1.8. *Example.* Let g be an Einstein metric on M whose sectional curvature ranges in $(3n/(7n-4), 1]$ ($(1/4, 1]$ for $n=4$). Then g has constant sectional

curvature ([6, Théorème 2], [15, Theorem 1]). It implies that if g has positive constant sectional curvature, then there is no non-trivial Einstein deformation of g. On the other hand, $\mathrm{EEID}(g)=0$ ([8, Corollary 7.3, Lemma 7.4]).

1.9. *Example.* Let (M, g) be a locally symmetric Einstein manifold of noncompact type without 2-dimensional factor. Then $\mathrm{EEID}(g)=0$ and there is no non-trivial Einstein deformation of g ([23, Corollary 3.5]).

1.10. *Example.* Let (M, g) be a simply connected irreducible symmetric space of compact type. Then the space $\mathrm{EEID}(g)$ vanishes except for the following types: $SU(k+1)$ $(k\geqq 2)$, $SU(k)/SO(k)$ $(k\geqq 2)$, $SU(2k)/Sp(k)$ $(k\geqq 3)$, $U(p+q)/U(p)\times U(q)$ $(p\geqq q\geqq 2)$ and E_6/F_4 ([24, Theorem 5.7]).

2. Moduli Spaces of Einstein Metrics

We recall Ebin's slice theorem. In this section we set $N=[n/2]+1$. Remark that by Sobolev's embedding theorem (c.f. [16]) H^s-differentiability implies C^{s-N}-differentiability for $s\geqq N$. So, for $s\geqq N$, the space of all H^s-riemannian metrics on M makes sense, which we denote by $\mathscr{M}^s$. And the group of all H^{s+1}-diffeomorphisms of M also makes sense, which we denote by $\mathscr{D}^{s+1}$. They become Hilbert manifolds ([29]). For $g\in\mathscr{M}^\infty$, we denote by $I(g)$ the isometry group of (M, g).

2.1. **Lemma** ([13, Theorem 8.20]). *Let $g\in\mathscr{M}^\infty$. If $s\geqq N+2$, then there exist a submanifold $\mathscr{S}^s_g$ of $\mathscr{M}^s$ and a local cross section $\chi^{s+1}: I(g)\backslash\mathscr{D}^{s+1}\to\mathscr{D}^{s+1}$ defined on an open neighbourhood $\mathscr{U}^{s+1}$ of the coset $I(g)$ with the following properties.*

(2.1.1) *If $\gamma\in I(g)$, then $\gamma^*(\mathscr{S}^s_g)=\mathscr{S}^s_g$.*

(2.1.2) *Let $\gamma\in\mathscr{D}^{s+1}$. If $\gamma^*(\mathscr{S}^s_g)\cap\mathscr{S}^s_g\neq\phi$, then $\gamma\in I(g)$.*

(2.1.3) *The map $F^s: \mathscr{S}^s_g\times\mathscr{U}^{s+1}\to\mathscr{M}^s$ defined by $F^s(g_1, u)=\chi^{s+1}(u)^* g_1$ is a homeomorphism onto an open neighbourhood $\mathscr{V}^s$ of g in $\mathscr{M}^s$.*

By the property (2.1.3), any $g_1\in\mathscr{V}^s$ is isometric with some $g_2\in\mathscr{S}^s_g$. And by the property (2.1.2), two H^s-riemannian metrics g_1 and g_2 in $\mathscr{S}^s_g$ are H^{s+1}-isometric if and only if they are isometric under an isometry $\gamma\in I(g)$. This means that the quotient space $\mathscr{M}^s/\mathscr{D}^{s+1}$ is locally identified with the quotient space $\mathscr{S}^s_g/\mathrm{I}(g)$. The following Lemma is the infinitesimal version of slice theorem.

2.2. **Lemma** ([13, Proposition 8.8]). *Let $g\in\mathscr{M}^\infty$. Then the following orthogonal decomposition holds.*

$$H^s(S^2M)=\delta_g^*(H^{s+1}(S^1M))\oplus\mathrm{Ker}\,\delta_g\cap H^s(S^2M), \tag{2.2.1}$$

where S^pM denotes the symmetric p-tensor bundle over M. The spaces $H^s(S^2M)$, $\delta_g^*(H^{s+1}(S^1M))$ and $\mathrm{Ker}\,\delta_g$ are the tangent spaces at g respectively of $\mathscr{M}^s$, $(\mathscr{D}^{s+1})^*g$ and $\mathscr{S}^s_g$.

2.3. *Definition.* Let $g\in\mathscr{M}^\infty$. Let ψ be a symmetric bilinear form on M. We decompose ψ into $\delta_g^*\xi+h$; $\delta_g h=0$. The symmetric bilinear form h (resp. $\delta_g^*\xi$) is called *the essential part* (resp. *the trivial part*) of ψ.

Lemma 2.1 is adapted to the ILH-category (c.f. [28]).

2.4. **Lemma** ([23, Theorem 2.2]). *In* Lemma 2.1, *the spaces* $\mathscr{S}_g^s$, $\mathscr{U}^{s+1}$, $\mathscr{V}^s$ *and the map* χ^{s+1} *can be taken so that* $\mathscr{S}_g^s = \mathscr{S}_g^{N+2} \cap \mathscr{M}^s$, $\mathscr{U}^{s+1} = \mathscr{U}^{N+3} \cap (I(g) \backslash \mathscr{D}^{s+1})$, $\mathscr{V}^s = \mathscr{V}^{N+2} \cap \mathscr{M}^s$ *and that* $\chi^{s+1} = \chi^{N+3} | \mathscr{U}^{s+1}$ *for* $s \geqq N+2$. *For any integers* $s \geqq N+2$ *and* $k \geqq 0$, *the mappings*

$$F_s^{s+k} \equiv F^{s+k} : \mathscr{S}_g^{s+k} \times \mathscr{U}^{s+k+1} \to \mathscr{V}^s \tag{2.4.1}$$

$$p_s^{s+k} \times q_s^{s+k} \equiv (F^{s+k})^{-1} : \mathscr{V}^{s+k} \to \mathscr{S}_g^s \times \mathscr{U}^{s+1} \tag{2.4.2}$$

are C^k*-differentiable.*

This means that, if we treat only C^∞-riemannian metrics, then the map F may be regarded as a C^∞-diffeomorphism.

In the following, we treat two C^∞-structures on M. We denote by C_∞ the original C^∞-structure on M and by C_r the C^r-structure induced by C_∞. If we denote by C'_∞ another C^∞-structure on M, then C'_r will denote the C^r-structure induced by C'_∞.

2.5 **Lemma.** *Let* $s \geqq N+2$ *and* $g \in \mathscr{M}^\infty$. *If* $g_1 \in \mathscr{S}_g^s$ *is* C^∞*-differentiable with respect to some* C^∞*-structure* C'_∞ *on* M *such that* $C'_{N+1} = C_{N+1}$, *then* g_1 *is* C^∞*-differentiable with respect to the original* C^∞*-structure* C_∞, *i.e.,* $g_1 \in \mathscr{S}_g^\infty$.

Proof. It is known by a theorem of Whitney (c.f. [17, p. 51 Theorem 2.9]) that if $C'_{N+1} = C_{N+1}$, then there exists a C^{N+1}-diffeomorphism γ of M such that $\gamma^* C'_\infty = C_\infty$. It means that $\gamma^* g_1 \in \mathscr{M}^\infty$. On the other hand, $\gamma^* g_1$ is an H^s-riemannian metric with respect to $\gamma^* C_\infty$. Let $\{x^i\}$ (resp. $\{\bar{x}^i\}$) be a local C^∞-coordinate with respect to C_∞ (resp. $\gamma^* C_\infty$) and $\{\Gamma^i{}_{jk}\}$ (resp. $\{\bar{\Gamma}^i{}_{jk}\}$) the Christoffel symbols of $\gamma^* g_1$ with respect to $\{x^i\}$ (resp. $\{\bar{x}^i\}$). Then the functions $\bar{x}^i$ are C^{N+1}-differentiable and $\bar{\Gamma}^i{}_{jk}$ are H^s-differentiable. By the transformation rule for the Christoffel symbols:

$$\frac{\partial^2 \bar{x}^k}{\partial x^i \partial x^j} = \Gamma^m{}_{ij} \frac{\partial \bar{x}^k}{\partial x^m} - \bar{\Gamma}^k{}_{lm} \frac{\partial \bar{x}^l}{\partial x^i} \frac{\partial \bar{x}^m}{\partial x^j}, \tag{2.5.1}$$

we see the following. If $\bar{x}^i$ are H^t-differentiable for $N+1 \leqq t \leqq s$, then the right hand side is H^{t-1}-differentiable by the composition law ([29, Theorem 11.3]) and so $\bar{x}^i$ are H^{t+1}-differentiable. Thus, by induction, we see that $\bar{x}^i$ are H^{s+1}-differentiable, which means that $\gamma \in \mathscr{D}^{s+1}$.

Now we approximate $\gamma \in \mathscr{D}^{s+1}$ by a sequence $\{\gamma_i\}$ of C^∞-diffeomorphisms. Then the sequence $\{\gamma_i^{-1*} \gamma^* g_1\}$ of C^∞-riemannian metrics converges to g_1 in $\mathscr{M}^s$, which implies that there is a C^∞-riemannian metric in $\mathscr{V}^s$ (defined in Lemma 2.1) which is H^{s+1}-isometric with g_1. Applying Lemma 2.4, we obtain a riemannian metric $g_2 \in \mathscr{S}_g^\infty$ which is also H^{s+1}-isometric with g_1. Denote by $\eta \in \mathscr{D}^{s+1}$ this isometry, i.e., $\eta^* g_2 = g_1$. Then the property (2.1.2) implies that $\eta \in I(g)$, in particular η is in $\mathscr{D}^\infty$ and so $g_1 \in \mathscr{M}^\infty$. Q.E.D.

2.6. **Lemma.** *Let* $s \geqq n+2$ *and* $g \in \mathscr{M}^\infty$. *Then all* H^s*-Einstein metrics in* $\mathscr{S}_g^s$ *are in* $\mathscr{S}_g^\infty$.

Proof. Let $g_1 \in \mathscr{S}_g^s$ be an H^s-Einstein metric. By Sobolev's embedding theorem, g_1 is C^{s-N}-differentiable. Then the result follows from Proposition 0.2 and Lemma 2.5. Q.E.D.

2.7. *Notation. Denote by* $\mathscr{M}_1^s$ the space of all H^s-riemannian metrics on M with volume 1.

2.8. *Definition.* Let $g \in \mathscr{M}_1^\infty$ be an Einstein metric. The space of all Einstein metrics in $\mathscr{S}_g^\infty \cap \mathscr{M}_1^\infty$ is called *the local pre-moduli space of Einstein metrics around g on M*, and denoted by ELPM(g) or simply ELPM.

2.9. *Remark.* By definition, the isometry group $I(g)$ acts on the space ELPM(g). This action induces an action of the space $K(g)$ on ELPM(g). See Remark 1.6.

3. Real Analyticity of the Local Pre-Moduli Space of Einstein Metrics

We know that "the local pre-moduli space of complex structures" forms a complex analytic set (Proposition 0.8). Concerning the moduli space of Einstein metrics, we have the following

3.1. **Theorem.** *Let $g \in \mathscr{M}_1^\infty$ be an Einstein metric. If $s \geqq n+2$, then there is an open neighbourhood $\mathscr{U}^s$ of g in $\mathscr{S}_g^s \cap \mathscr{M}_1^s$ such that the space* ELPM(g)$\cap \mathscr{U}^s$ *forms a real analytic set in a finite dimensional real analytic submanifold Z^s of $\mathscr{U}^s$ whose tangent space $T_g Z^s$ at g coincides with the space* EEID(g).

This theorem is proved by using the real analytic implicit function theorem in Banach space. See Appendix. As a corollary, we get the following

3.2. **Theorem.** *Let $g \in \mathscr{M}_1^\infty$ be an Einstein metric. The following four conditions are equivalent.*

(3.2.1) *The Einstein metric g is deformable, that is, there is a non-trivial Einstein deformation of g.*

(3.2.2) *There is a continuous one-parameter family g_t of Einstein metrics on M such that $g_0 = g$ and that g_t are not homothetic with g for all $t \neq 0$.*

(3.2.3) *The Einstein metric g is not rigid, that is, for any open neighbourhood $\mathscr{V}$ of g in $\mathscr{M}^s$, there is an Einstein metric $g_1 \in \mathscr{V}$ which is not homothetic with g.*

(3.2.4) *The Einstein metric g is formally deformable, that is, there is a formal power series*

$$g(t) = \sum_{i=0}^{\infty} \frac{1}{i!} h^{(i)} t^i, \qquad h^{(0)} = g \tag{3.2.5}$$

with coefficients in $C^\infty(S^2 M)$ which satisfies the formal equation $E(g(t)) = 0$ but is not in the formal orbit space $(\mathscr{D}^{s+1})^* g$.

Proof. Assume that condition (3.2.1) holds. We may assume that the parameter space of g_t is one-dimensional and that g_1 is not homothetic with g. [13, Theorem 8.10] says that the orbit space $(\mathscr{D}^{s+1})^* g$ is closed in $\mathscr{M}^s$. Therefore there is a maximal real number $0 \leqq t_0 < 1$ such that g_t are isometric with g for

all $0 \leqq t \leqq t_0$. Let $\gamma \in \mathscr{D}^{s+1}$ be an isometry between g_{t_0} and g, i.e., $\gamma^* g_{t_0} = g$. Since the metrics g_{t_0} and g are C^∞, so is γ. Thus we obtain a new family $\bar{g}_t = \gamma^* g_{t-t_0}$ such that $\bar{g}_0 = g$ and that for all sufficiently small $t > 0$ $\bar{g}_t$ is not isometric with g, i.e., condition (3.2.2) holds. Obviously, condition (3.2.2) implies condition (3.2.3). Assume that g is not rigid. By Lemmas 2.1 and 2.6, it means that the point g is not isolated in the space ELMP(g). Then by Theorem 3.1, there is a non-trivial real analytic curve in the local pre-moduli space, which implies that g is formally deformable. Finally, assume that g is formally deformable and let the formal power series (3.2.5) is a non-trivial formal deformation. Set

$$g_{(k)t} = \sum_{i=0}^{k} \frac{1}{i!} h^{(i)} t^i. \tag{3.2.6}$$

Then each $g_{(k)t}$ is a C^∞-curve in $\mathscr{M}^\infty$, so we obtain a curve $\tilde{g}_{(k)t}$ in the space $\mathscr{S}^s_g$ by Lemmas 2.1 and 2.4. Let $h^{(i)}_{(k)}$ be the i-th derivative of $\tilde{g}_{(k)t}$ at $t=0$. Then we see that

$$h^{(i)}_{(k)} = h^{(i)}_{(j)} \quad \text{for } i \leqq \min\{k, j\}, \tag{3.2.7}$$

and so the formal curve $\tilde{g}(t)$ defined by

$$\tilde{g}(t) = \sum_{i=0}^{\infty} \frac{1}{i!} h^{(i)}_{(i)} t^i \tag{3.2.8}$$

is a non-trivial curve in the space ELPM(g). But [2, Theorem 1.2] says that any formal curve

$$c(t) = \sum_{i=0}^{\infty} \frac{1}{i!} c^{(i)} t^i \tag{3.2.9}$$

in an analytic set can be approximated by convergent curves, that is, for any positive integer k, there exists a convergent curve

$$c_{(k)t} = \sum_{i=0}^{\infty} \frac{1}{i!} c^{(i)}_{(k)} t^i \tag{3.2.10}$$

in the real analytic set such that $c^{(i)}_{(k)} = c^{(i)}$ for $i \leqq k$. Therefore we obtain a non-trivial convergent curve in the space ELPM(g), i.e., g is deformable. Q.E.D.

3.3. *Remark.* The fact that the formal deformability implies the deformability may be used to construct some new Einstein metrics. But up to now we have no application of this relation.

3.4. *Remark.* In the category of deformations of complex structures, the rigidity does not imply the non-deformability. For counter examples, see [27, pp. 23–26].

3.5. **Corollary.** *Let $g \in \mathscr{M}^\infty_1$ be an Einstein metric. If each essential infinitesimal Einstein deformation h of g is integrable, that is, if there exists an Einstein deformation g_t of g such that $g'_0 = h$, then the space* ELPM(g) *forms a submanifold of $\mathscr{M}^s$ around g whose tangent space T_g(ELPM(g)) coincides with the space* EEID(g).

Proof. It is sufficient to prove that if f is a real analytic function defined on an open neighbourhood of the origin 0 in a Euclidean space $\mathbf{R}^k$ such that $f(0)=0$ and that for any $v\in\mathbf{R}^k$ there exists a curve $c_v(t)$ in $\mathbf{R}^k$ which satisfies $c_v(0)=0$, $c_v'(0)=v$ and $f(c_v(t))=0$, then f is identically zero. We prove it by induction. First, we have

$$0=(f\circ c_v)_0'=f_0'(c_v'(0))=f_0'(v). \tag{3.5.1}$$

Therefore $f_0'=0$. Assume that $f_0^{(i)}=0$ for $i\leqq r$. Then we see

$$0=\left(\frac{d}{dt}\right)_0^{r+1}(f\circ c_v)(t)=f_0^{(r+1)}(c_v'(0),\ldots,c_v'(0)), \tag{3.5.2}$$

and so $f_0^{(r+1)}=0$. Q.E.D.

3.6. *Example.* If EEID(g)=0, then ELPM(g) reduces to one point. See Examples in Sect. 1.

3.7. *Example.* Let g be the symmetric Einstein metric on $P^{2m}(\mathbf{C})\times S^2$. Then EEID($g$)$\neq$0 but ELPM($g$) reduces to one point ([24, Theorem 5.7]).

4. The Canonical Riemannian Metric on the Space ELPM(g)

If the local pre-moduli space ELPM of Einstein metrics around an Einstein metric g forms a submanifold of $\mathscr{M}^s$, then it has the induced riemannian metric. But this metric may depend on the origin g (c.f. Lemma 4.7). So we define another riemannian metric.

4.1. *Definition.* A family g_t of riemannian metrics on M with parameter space P is said to be *effectively parametrized* if we have

$$v[g_t]\notin \operatorname{Im}\delta_{g_t}^* \tag{4.1.1}$$

for all $t\in P$ and non-zero $v\in T_tP$. A family g_t is said to be *normal* if the dimension of the space $K(g_t)$ of all Killing vector fields is constant for $t\in P$.

Let g_t be a family of riemannian metrics on M with parameter space P. For each $t\in P$ and $v\in T_tP$, we have the decomposition (2.2.1):

$$v[g_t]=h+L_Xg_t;\quad \delta_{g_t}h=0. \tag{4.1.2}$$

We set

$$v^H=v-X, \tag{4.1.3}$$

which is regarded as a vector field along the map: $M\to M\times P$. If we denote by $[v^H, g_t]$ the symmetric bilinear form $v[g_t]-L_Xg_t$, then

$$h=[v^H, g_t]\in\operatorname{Ker}\delta_{g_t}. \tag{4.1.4}$$

We define a positive semi-definite inner product (,) on T_tP by setting

$$(v, w)=\langle[v^H, g_t], [w^H, g_t]\rangle_{g_t}, \tag{4.1.5}$$

where $\langle\ ,\ \rangle$ denotes the global inner product. Note that if $K(g_t)\neq 0$ then the vector field X which satisfies Eq. (4.1.2) is not unique, but $[v^H, g_t]$ is well-defined and so inner product (4.1.5) is well-defined.

4.2. **Lemma.** *Let g_t be an effectively parametrized normal family of riemannian metrics on M with parameter space P. Then the inner product* (4.1.5) *defines a riemannian metric on P.*

Proof. Since inner product (4.1.5) is positive definite on each T_tP by assumption, it is sufficient to show that the essential part h in decomposition (4.1.2) depends C^∞-ly on t and v. This follows directly from the next lemma. We will also see that the vector v^H can be taken so that it depends C^∞-ly on t and v. Q.E.D.

4.3. **Lemma.** *Let ω_t be a family of volume elements on M, E_t, F_t families of vector bundles over M with fiber metrics g_t^E, g_t^F and Q_t: $C^\infty(E_t)\to C^\infty(F_t)$ a family of differential operators of order k with injective symbol. Assume that ω_t, E_t, F_t, g_t^E, g_t^F and Q_t depends C^∞-ly (resp. real analytically) on t. That is, there are bundle isomorphisms e_t: $E_0\to E_t$ and f_t: $F_0\to F_t$ such that the coefficients of $e_t^* g_t^E$, $f_t^* g_t^F$ and $(f_t^{-1})_*\circ Q_t\circ(e_t)_*$ depend C^∞-ly (resp. real analytically) on t. Then the dimension of the space $\operatorname{Ker} Q_t$ is upper semicontinuous. If the dimension of the space $\operatorname{Ker} Q_t$ is constant, then the decompositions*

$$H^s(E_t)=Q_t^*(H^{s+k}(F_t))\oplus\operatorname{Ker} Q_t, \tag{4.3.1}$$

$$H^s(F_t)=Q_t(H^{s+k}(E_t))\oplus\operatorname{Ker} Q_t^* \tag{4.3.2}$$

depend C^∞-ly (resp. real analytically) on t, where Q_t^* is the formal adjoint operator of Q_t with respect to g_t^E, g_t^F and ω_t. Moreover the isomorphisms

$$Q_t^*+1\colon Q_t(H^{s+2k}(E_t))\oplus\operatorname{Ker} Q_t\to H^s(E_t), \tag{4.3.3}$$

$$Q_t+1\colon Q_t^*(H^{s+2k}(F_t))\oplus\operatorname{Ker} Q_t^*\to H^s(F_t) \tag{4.3.4}$$

also depend C^∞-ly (resp. real analytically) on t.

Proof. We may assume that the vector bundles E_t and F_t do not depend on t. The decomposition (4.3.2) for each t is due to [13, Theorem 8.5]. If we remark that $Q_t^*Q_t$ is an elliptic operator for each t, then the other isomorphisms for each t are easy to check. By Remark 13.8, the families of operators Q_t etc. are C^∞-(resp. real analytic) curves in the Banach spaces $L(H^{s+k}(E), H^s(F))$ etc. of all continuous linear operators for sufficiently large s. First we consider the map

$$\text{projection}\circ Q_t\colon Q_0^*(H^{s+k}(F))\oplus\operatorname{Ker} Q_0\to Q_0(H^s(F)). \tag{4.3.5}$$

For $t=0$, the restriction of this map on $Q_0^*(H^{s+k}(F))$ is an isomorphism. Hence by the implicit function theorem, there is a unique homomorphism ψ_t: $\operatorname{Ker} Q_0\to Q_0^*(H^{s+k}(F))$ which depends C^∞-ly (resp. real analytically) on t such that

$$Q_t(\psi_t(z)+z)\in\operatorname{Ker} Q_0^* \quad \text{for } z\in\operatorname{Ker} Q_0. \tag{4.3.6}$$

Let $x \in \operatorname{Ker} Q_t$ and decompose it as

$$x = u_x + z_x; \quad u_x \in \operatorname{Im} Q_0^*, \quad z_x \in \operatorname{Ker} Q_0. \tag{4.3.7}$$

Then we see that

$$Q_t(u_x + z_x) = 0. \tag{4.3.8}$$

Since such ψ_t is unique, we get $\psi_t(z_x) = u_x$. In particular, if $z_x = 0$, then $x = 0$. Thus we have an injection: $\operatorname{Ker} Q_t \to \operatorname{Ker} Q_0$, from which the upper semicontinuity follows. Assume that the dimension of $\operatorname{Ker} Q_t$ is constant. Then we see that for any $z \in \operatorname{Ker} Q_0$, there exists $x \in \operatorname{Ker} Q_t$ such that $z = z_x$. But then

$$Q_t(\psi_t(z) + z) = Q_t(x) = 0. \tag{4.3.9}$$

Thus if we set $a_t = 1 + \psi_t$, then a_t gives an isomorphism: $\operatorname{Ker} Q_0 \to \operatorname{Ker} Q_t$.

Next consider the map

$$\begin{aligned} Q_t^* + a_t \colon\ & H^s(F) \oplus \operatorname{Ker} Q_0 = \operatorname{Ker} Q_0^* \oplus Q_0(H^{s+k}(E)) \oplus \operatorname{Ker} Q_0 \\ & \to H^{s-k}(E) = Q_t^*(H^s(F)) \oplus \operatorname{Ker} Q_t. \end{aligned} \tag{4.3.10}$$

For $t = 0$, the restriction of this map to $Q_0(H^{s+k}(E)) \oplus \operatorname{Ker} Q_0$ is an isomorphism. Therefore by the implicit function theorem there exists a homomorphism

$$\psi_t = \psi_t^I + \psi_t^K \colon \operatorname{Ker} Q_0^* \to Q_0(H^{s+k}(E)) \oplus \operatorname{Ker} Q_0 \tag{4.3.11}$$

which depends C^∞-ly (resp. real analytically) on t such that

$$Q_t^*(x + \psi_t^I(x)) + a_t \psi_t^K(x) = 0 \quad \text{for } x \in \operatorname{Ker} Q_0^*. \tag{4.3.12}$$

But this implies that $Q_t^*(x + \psi_t^I(x)) = a_t \psi_t^K(x) = 0$. Thus if we set $b_t = 1 + \psi_t^I$, this gives an isomorphism: $\operatorname{Ker} Q_0^* \to \operatorname{Ker} Q_t^*$.

Finally we consider the maps

$$Q_t^* + a_t \colon Q_0(H^{s+2k}(E)) \oplus \operatorname{Ker} Q_0 \to H^s(E), \tag{4.3.13}$$

$$Q_t + b_t \colon Q_0^*(H^{s+2k}(F)) \oplus \operatorname{Ker} Q_0^* \to H^s(F). \tag{4.3.14}$$

They are isomorphisms for $t = 0$, so we have the inverse maps ψ_t and ϕ_t, i.e.,

$$(Q_t^* + a_t)\psi_t = \mathrm{id}_{H^s(E)}, \tag{4.3.15}$$

$$(Q_t + b_t)\phi_t = \mathrm{id}_{H^s(F)}, \tag{4.3.16}$$

which depend C^∞-ly (resp. real analytically) on t. Then the map $Q_t^*\psi_t + a_t\psi_t$ gives the decomposition (4.3.1) and the map $Q_t\phi_t + b_t\phi_t$ gives the decomposition (4.3.2). Then the spaces $Q_t(H^{s+k}(E))$ and $Q_t^*(H^{s+k}(F))$ depend C^∞-ly (resp. real analytically) on t, thus also depend isomorphisms (4.3.3) and (4.3.4). Q.E.D.

4.4. *Remark.* This Lemma in the C^∞-category is essentially done in [22, Theorem 5]. In their proof they used potential theory.

4.5. *Definition.* Let g_t be an effectively parametrized normal family with parameter space P. The riemannian metric on P defined in Lemma 4.2 is called *the canonical riemannian metric on* P and *denoted by* g^P.

4.6. *Definition.* Let g_t (resp. $\bar{g}_s$) be a family of riemannian metrics on M with parameter space P (resp. $\bar{P}$). They are said to be *equivalent* if there are a diffeomorphism $\psi: P \to \bar{P}$ and a family γ_t of diffeomorphisms of M with parameter space P such that

$$\gamma_t^* \bar{g}_{\psi(t)} = g_t. \tag{4.6.1}$$

4.7. **Lemma.** *Let g_t be an effectively parametrized normal family of riemannian metrics on M with parameter space P and $\bar{g}_t$ an equivalent family. Then the family $\bar{g}_t$ is also effectively parametrized and normal, and ψ becomes an isometry.*

Proof. Obviously, $\bar{g}_t$ is normal. We may assume that the parameter space of $\bar{g}_t$ is also P. Then if we differentiate Eq. (4.6.1), we get

$$[v, g_t] = \gamma_t^*[[v, \gamma_t] \circ \gamma_t^{-1}, \bar{g}_t] + \gamma_t^*[v, \bar{g}_t]. \tag{4.7.1}$$

Therefore if h is the essential part of $[v, g_t]$ with respect to g_t, then ${\gamma_t^{-1}}^* h$ is the essential part of $[v, \bar{g}_t]$ with respect to ${\gamma_t^{-1}}^* g_t = \bar{g}_t$. This implies that $\bar{g}_t$ is effectively parametrized, and we see that

$$\begin{aligned}(v_1, v_2) &= \langle h_1, h_2 \rangle_{g_t} \\ &= \langle {\gamma_t^{-1}}^* h_1, {\gamma_t^{-1}}^* h_2 \rangle_{{\gamma_t^{-1}}^* g_t},\end{aligned} \tag{4.7.2}$$

where h_i is the essential part of $[v_i, g_t]$. Q.E.D.

4.8. *Remark.* This Lemma means that the canonical riemannian metric is a well-defined notion on a family of riemannian metrics as a fiber structure: $M \times P \to P$ (c.f. [21, Definition 1.1]).

If the space ELPM becomes a submanifold of $\mathscr{S}_g^s$, it may be regarded as a family of riemannian metrics on M with parameter space ELPM.

4.9. *Definition.* Let $g \in \mathscr{M}_1^\infty$ be an Einstein metric. The space ELPM(g) is said to be *normal* (resp. *effectively parametrized*) if it forms a submanifold of $\mathscr{M}^s$ around g and if it is normal (resp. effectively parametrized) as a family of riemannian metrics.

4.10. **Lemma.** *Let $g_0 \in \mathscr{M}_1^\infty$ be an Einstein metric. Assume that the space* ELPM(g_0) *becomes a submanifold of $\mathscr{M}^s$ around g_0. Then the following conditions are equivalent.*

(4.10.1) ELPM(g_0) *is effectively parametrized.*

(4.10.2) ELPM(g_0) *is normal.*

(4.10.3) $K(g) = K(g_0)$ *for* $g \in$ ELPM(g_0).

(4.10.4) $K(g_0)$ *acts trivially on* ELPM(g_0).

(4.10.5) $K(g_0)$ *acts trivially on* T_{g_0}(ELPM(g_0)), *where* $K(g)$ *denotes the space of all Killing vector fields on* (M, g).

Proof. We will show the implications: $1\to 3$, $2\to 3\to 4$ and $5\to 4\to 1$. Combining the obvious implications: $3\to 2$ and $4\to 5$, we get the equivalence.

$1\to 3$: Assume that the space ELPM is effectively parametrized. Let $X\in K(g_0)$. Then, since $K(g_0)$ acts on ELPM, we have $L_X g\in T_g(\mathrm{ELPM})$, therefore $L_X g=0$, i.e., $X\in K(g)$. But here by the property (2.1.2), $I(g)\subset I(g_0)$ for $g\in\mathscr{S}^s_{g_0}$. Thus $K(g)=K(g_0)$.

$2\to 3$: Assume that the space ELPM is normal. Then the inclusion $I(g)\subset I(g_0)$ implies that $K(g)=K(g_0)$.

$3\to 4$: Assume that $K(g)=K(g_0)$ for $g\in$ELPM. Let $X\in K(g_0)$. Then $L_X g=0$, which implies that $K(g_0)$ acts trivially on ELPM.

$5\to 4$: Assume that $K(g_0)$ acts trivially on $T_{g_0}(\mathrm{ELPM})$. Let $X\in K(g_0)$ and $\gamma(t)$ the one-parameter group of diffeomorphisms generated by X. Then $\gamma(t)$ acts trivially on $T_{g_0}(\mathrm{ELPM})$. But here, the manifold ELPM is an $I(g_0)$-invariant submanifold of $\mathscr{M}^s$, so the induced riemannian metric on ELPM is $I(g_0)$-invariant. Thus the triviality of the action of $\gamma(t)$ on $T_{g_0}(\mathrm{ELPM})$ extends to that on the space ELPM, which implies that $K(g_0)$ acts trivially on ELPM.

$4\to 1$: Assume that $K(g_0)$ acts trivially on the space ELPM. Let $g\in$ELPM and X a vector field on M such that $L_X g\in T_g(\mathrm{ELPM})$. Denote by $\gamma(t)$ the one-parameter group of diffeomorphisms generated by X. Then in the situation of Lemmas 2.1 and 2.4, we have

$$\gamma(t)^* g=F^{s-1}_{s-2}(p^s_{s-1}(\gamma(t)^* g), q^s_{s-1}(\gamma(t)^* g)). \tag{4.10.6}$$

Therefore, by the property (2.1.2) we see that

$$\chi^s(q^s_{s-1}(\gamma(t)^* g))\circ\gamma(t)^{-1}\in I(g_0), \tag{4.10.7}$$

which implies that

$$(\chi^s)'\circ(q^s_{s-1})'_g(L_X g)-X\in K(g_0). \tag{4.10.8}$$

But here, since $L_X g\in T_g\mathscr{S}^s_{g_0}$, we have

$$(q^s_{s-1})'_g(L_X g)=0. \tag{4.10.9}$$

So $X\in K(g_0)$. Therefore, by assumption we see that $L_X g=0$. Q.E.D.

4.11. **Corollary.** *Let g be an Einstein metric on M with non-positive Ricci curvature and with volume 1. If (M, g) has no local flat factor and if the space* ELPM(g) *forms a submanifold of $\mathscr{M}^s$, then the canonical riemannian metric on the space* ELPM(g) *is well defined.*

Proof. By a well-known theorem of Bochner, the assumption implies that $K(g)$ vanishes. Q.E.D.

4.12. **Corollary.** *Let g be a flat riemannian metric on M with volume 1. Then the canonical riemannian metric on the space* ELPM(g) *is well defined.*

Proof. Let $h\in$EEID. Then h is parallel and so we can easily construct an Einstein deformation g_t such that $g'_0=h$. Therefore, by Corollary 3.5, the space ELPM becomes a submanifold of $\mathscr{M}^s$ whose tangent space at g coincides with

EEID. Moreover, if X is a Killing vector field of (M, g), then X is parallel and so $L_X h=0$. Q.E.D.

4.13. **Theorem.** *Let $g_0 \in \mathcal{M}_1^\infty$ be an Einstein metric. If the space* $\mathrm{ELPM}(g_0)$ *is normal (or equivalently if* $\mathrm{ELPM}(g_0)$ *forms a manifold and if one of the conditions in* Lemma 4.10 *holds), then the canonical riemannian metric on* $\mathrm{ELPM}(g_0)$ *is real analytic.*

Proof. It is sufficient to prove that the essential part h of ψ depends real analytically on g and $\psi \in T_g(\mathrm{ELPM})$. This follows from Theorem 3.1 and Lemma 4.3. Q.E.D.

4.14. *Example.* Let (T^n, g_0) be a flat torus of volume 1. Let EPM be the set of all T^n-invariant riemannian metrics on T^n with volume 1. Then the space $\mathrm{ELPM}(g_0)$ is an open neighbourhood of g_0 in EPM. In this case, the canonical riemannian metric on $\mathrm{ELPM}(g_0)$ coincides with the induced metric from $\mathcal{M}^s$. The space $\mathrm{ELPM}(g_0)$ is then isometric with an open set of the irreducible symmetric space $SL(n, \mathbf{R})/SO(n, \mathbf{R})$ (=EPM).

5. The Obstruction Space for the Space EEID

In this section, we treat only deformations of riemannian metrics. But to understand what we do, it would be better to recall Proposition 0.12, which says that the space $H^2(M, \Theta)$ is the "obstruction" for the integrability (see Corollary 3.5) of infinitesimal complex deformations (see Definition 6.2). Bourguignon posed the following

5.1. *Question. Is there a space which plays the role of an obstruction space for the integrability of infinitesimal Einstein deformation?*

We will give a negative answer to this question. Consider an equation $F(g)=0$ for riemannian metrics. (For example, Einstein's equation $E(g)=0$.) Assume that there is a linear operator B_g such that the equation

$$B_g(F(g))=0 \tag{5.1.1}$$

becomes an identity. For example, if $F=E$, we have such an operator, namely the Bianchi identity operator:

$$B_g(\psi)=\delta_g\psi+(1/2)\,d\mathrm{tr}_g\psi. \tag{5.1.2}$$

Let g_t be a 1-parameter family of riemannian metrics on M, and set $h_i = \left(\frac{d}{dt}\right)_0^i g_t$. Then we have the expression

$$\left(\frac{d}{dt}\right)_0^r F(g_t)=P_{g_0}^r(h_1, \ldots, h_{r-1})+F'_{g_0}(h_r). \tag{5.1.3}$$

If we take the r-th derivative of identity (5.1.1), we get the identity

$$\sum_{i=1}^{r} \binom{r}{i} \left[\left(\frac{d}{dt}\right)_0^i B_{g_t}\right]\left[\left(\frac{d}{dt}\right)_0^{r-i} F(g_t)\right]=0. \tag{5.1.4}$$

Now, we assume that

$$\left(\frac{d}{dt}\right)_0^i F(g_t)=0 \quad \text{for } 0\leqq i\leqq r-1. \tag{5.1.5}$$

Then equality 5.1.4 becomes

$$B_{g_0}(P^r_{g_0}(h_1,\ldots,h_{r-1})+F'_{g_0}(h_r))=0. \tag{5.1.6}$$

Remark that this equality holds for any h_r. So we have, since B_{g_0} is linear,

$$B_{g_0}\circ F'_{g_0}=0, \tag{5.1.7}$$

$$B_{g_0}(P^r_{g_0}(h_1,\ldots,h_{r-1}))=0. \tag{5.1.8}$$

We want to solve the equation

$$F'_{g_0}(h_r)=-P^r_{g_0}(h_1,\ldots,h_{r-1}) \tag{5.1.9}$$

for h_r. There is a solution h_r if and only if $P^r_{g_0}(h_1,\ldots,h_{r-1})\in\operatorname{Im}F'_{g_0}$. But here we have Eq. (5.1.8), so if $\operatorname{Ker}B_{g_0}\subset\operatorname{Im}F'_{g_0}$, then there exists a solution of (5.1.9). In general, by equality (5.1.7), we have

$$\operatorname{Ker}B_{g_0}\supset\operatorname{Im}F'_{g_0}. \tag{5.1.10}$$

5.2. *Definition.* Let g be a riemannian metric on M such that $F(g)=0$. The space $\operatorname{Ker}B_g/\operatorname{Im}F'_g$ is called *the obstruction space.* If $\operatorname{Im}F'_g$ is closed, then the space $\operatorname{Ker}B_g\cap(\operatorname{Im}F'_g)^\perp$ also is called *the obstruction space.*

In fact, we saw that if the obstruction space vanishes, then all infinitesimal deformations are formally integrable (see Corollary 3.5). In the case of Einstein's equation $E(g)=0$, it implies the following

5.3. **Lemma.** *Let g be an Einstein metric on M with volume 1. If the obstruction space* $\operatorname{Ker}B_g\cap(\operatorname{Im}E'_g)^\perp$ *vanishes, then the space* ELPM *forms a submanifold of M^s whose tangent space T_g(ELPM) coincides with the space* EEID.

Proof. By Corollary 3.5, it is sufficient to prove that for any $h\in$EEID, there is an Einstein deformation g_t of g such that $g'_0=h$. But it is already shown in the above discussion that there exists a formal power series

$$g(t)=\sum_{i=0}^{\infty}\frac{1}{i!}h_it^i \tag{5.3.1}$$

which satisfies the formal equation $E(g(t))=0$ such that $h_0=g$ and $h_1=h$. Then the result follows by the same argument as in the proof of Theorem 3.2. Q.E.D.

But unfortunately, we have the next proposition which says that we cannot apply this lemma.

5.4. **Proposition** ([24, Proposition 3.2]). *Let g be an Einstein metric on M with volume 1. Then the obstruction space* $\operatorname{Ker}B_g\cap(\operatorname{Im}E'_g)^\perp$ *coincides with the space* EEID *itself.*

5.5. *Remark.* We see that the definition of the obstruction space depends on the choice of the identity. But it seems to the author that for Einstein's equation, there is no other effective identity than the Bianchi identity. On the other hand, if an equation $F(g)=0$ is sufficiently nice, for example if the space of all "essential infinitesimal deformations" is finite dimensional, Lemma 5.3 remains true for this equation. If a solution g of the equation has "an essential infinitesimal deformation" and if "the obstruction space" vanishes, then we get a manifold as "the local pre-moduli space" for the equation. To analyse the real analytic set ELPM as a subset of such a manifold, we pose the following

5.6. *Problem. Find such an equation* $F(g)=0$. *We claim also that Einstein metrics* g *satisfy* $F(g)=0$.

6. Deformations and Infinitesimal Deformations of a Complex Structure

We introduce the notion of complex deformations, which was developed by [21, 22]. We introduce it in an exactly similar way as in Sect. 1 to analyse deformations of Kähler-Einstein structures.

6.1. *Definition.* Let J be a complex structure on M. A family J_t of complex structures on M such that $J_0=J$ is called *a complex deformation of* J. A complex deformation J_t of J is said to be *trivial* if the complex structure J_t is isomorphic with J for each t.

A tensor field J of type $(1,1)$ is a complex structure if and only if the following equations are satisfied.

$$J^2=-\mathrm{id}_{TM}, \tag{6.1.1}$$

$$N(J)=0, \tag{6.1.2}$$

where N denotes *the Nijenhuis torsion tensor*, defined by

$$N(J)(X,Y)=[X,Y]-[JX,JY]+J[JX,Y]+J[X,JY]. \tag{6.1.3}$$

6.2. *Definition.* Let J be a complex structure on M. A tensor field I of type $(1,1)$ is called *an infinitesimal complex deformation of* J if the following linear equations are satisfied.

$$IJ+JI=0, \tag{6.2.1}$$

$$N_J'(I)=0. \tag{6.2.2}$$

The space of all infinitesimal complex deformations of J is *denoted by* CID(J) or simply by CID.

6.3. *Definition.* Let J be a complex structure and g a riemannian metric on M. An infinitesimal complex deformation of the form L_XJ is said to be *trivial.* The space of all trivial infinitesimal complex deformations of J is *denoted by* CTID(J) or simply by CTID. An infinitesimal complex deformation I of J is said to be *essential with respect to* g if I is orthogonal to the space CTID(J)

with respect to the global inner product defined by g. The space of all essential infinitesimal complex deformations of J with respect to g is denoted by CEID(J, g) or simply CEID(J), CEID.

6.4. **Lemma.** *Let J be a complex structure on M. A (real) tensor field I of type $(1,1)$ is an infinitesimal complex deformation of J if and only if I satisfies the following two equations*

$$I^{\alpha}{}_{\beta}=0, \tag{6.4.1}$$

$$\partial_{\alpha}I^{\bar{\gamma}}{}_{\beta}-\partial_{\beta}I^{\bar{\gamma}}{}_{\alpha}=0. \tag{6.4.2}$$

If g is a Kähler metric on M, then Eq. (6.4.2) becomes

$$D_{\alpha}I^{\bar{\gamma}}{}_{\beta}-D_{\beta}I^{\bar{\gamma}}{}_{\alpha}=0, \tag{6.4.3}$$

and an element $I\in$CID is in the space CEID(J, g) *if and only if I satisfies the equation*

$$D^{\gamma}I^{\bar{\alpha}}{}_{\gamma}=0. \tag{6.4.4}$$

Proof. Equations (6.4.1), (6.4.2) and (6.4.3) are given by a straightforward tensor computations. Equation (6.4.4) follows from the equation

$$(L_X J)(\partial_{\alpha})=-2\sqrt{-1}(D_{\alpha}X^{\bar{\gamma}})\,\bar{\partial}_{\gamma} \text{ for vector fields } X. \tag{6.4.5}$$

Q.E.D.

6.5. **Lemma.** *Let J be a complex structure and g a riemannian metric on M. Then the space* CID(J) *admits an orthogonal decomposition*

$$\mathrm{CID}(J)=\mathrm{CEID}(J, g)\oplus\mathrm{CTID}(J) \tag{6.5.1}$$

with respect to the global inner product defined by g.

Proof. It suffices to show that the space CTID is a closed subspace, which is a direct consequence of decomposition (4.3.2) (for $t=0$). Q.E.D.

Also for the space CEID, a similar result as Lemma 1.5 holds. See Lemma 8.1. We defined the space CEID in analogy with the space EEID. But we have another space of infinitesimal deformations of complex structures, defined in [20, Sect. 5]. This space is the cohomology group $H^1(M, \Theta)$ with coefficients in the sheaf Θ of germs of holomorphic vector fields.

6.6. **Lemma.** *Let J be a complex structure and g a riemannian metric on M. Then the space* CEID(J, g) *and $H^1(M, \Theta)$ are canonically isomorphic.*

Proof. We use the ordinary notations for chain complexes and denote by X the sheaf of germs of smooth vector fields. First we construct a linear map: CID$\to H^1(M, \Theta)$. Let $I\in$CID. By Eq. 6.4.2 and by Dolbeault's lemma, there is a local complex vector field η such that $I^{\bar{\gamma}}{}_{\beta}=\partial_{\beta}\eta^{\bar{\gamma}}$. Taking the real part ξ of the anti-holomorphic part of η, we see that $L_{\xi}J=I$ by formula 6.4.5, i.e., there exists $\{\xi_{\alpha}\}\in C^0(M, X)$ such that $L_{\xi_{\alpha}}J=I$. If we define $\{\phi_{\alpha\beta}\}\in Z^1(M, X)$ by $\phi_{\alpha\beta}=\xi_{\alpha}-\xi_{\beta}$, then we see that $\{\phi_{\alpha\beta}\}\in Z^1(M, \Theta)$, and so it defines an element in

$H^1(M, \Theta)$. We easily see that this correspondence: $\mathrm{CID} \to H^1(M, \Theta)$ is well-defined, and its kernel coincides with the space CTID. So this map defines an injection: $\mathrm{CEID} \to H^1(M, \Theta)$. Next we show that this map is surjective. Let $\{\phi_{\alpha\beta}\} \in Z^1(M, \Theta)$. Since $Z^1(M, \Theta) \subset Z^1(M, X) = B^1(M, X)$, there exists $\{\eta_\alpha\} \in C^0(M, X)$ such that $\phi_{\alpha\beta} = \eta_\alpha - \eta_\beta$. Then $L_{\eta_\alpha} J$ defines an infinitesimal complex deformation. It is easy to see that I is mapped to $\{\phi_{\alpha\beta}\} \in H^1(M, \Theta)$ by the above correspondence: $\mathrm{CID} \to H^1(M, \Theta)$. Q.E.D.

Let J be a complex structure on M. If $I \in \mathrm{CID}$, then by Lemma 2.4, we see that $JI \in \mathrm{CID}$, i.e., the space CID becomes a complex vector space. If $\{\phi_{\alpha\beta}\} \in Z^1(M, \Theta)$, then $\{J\phi_{\alpha\beta}\} \in Z^1(M, \Theta)$. Using the operator J on $Z^1(M, \Theta)$, the space $H^1(M, \Theta)$ also becomes a complex vector space. Moreover, if g is a compatible Kähler metric on M, then by Eq. (6.4.3) and (6.4.4), the space $\mathrm{CEID}(J, g)$ becomes a subspace of the complex vector space $\mathrm{CID}(J)$. We easily see the following

6.7. **Lemma.** *Let (J, g) be a Kähler structure on M. Then the isomorphism defined in* Lemma 6.6 *is an isomorphism of complex vector spaces.*

7. The Space EEID on a Kähler-Einstein Manifold

First, we prepare the following

7.1. **Lemma.** *Let (J, g) be a Kähler-Einstein structure on M. Then the following formulae hold.*

$$(7.1.1)\quad D^\gamma D_\gamma \psi_{\alpha_1 \ldots \alpha_k \bar{\beta}_1 \ldots \bar{\beta}_l} = D^{\bar{\gamma}} D_{\bar{\gamma}} \psi_{\alpha_1 \ldots \alpha_k \bar{\beta}_1 \ldots \bar{\beta}_l} + (k-l)\, e \psi_{\alpha_1 \ldots \alpha_k \bar{\beta}_1 \ldots \bar{\beta}_l},$$

$$(7.1.2)\quad -D^\alpha (D_\alpha \psi_{\beta\gamma} - D_\beta \psi_{\alpha\gamma}) = \tfrac{1}{2}(\bar{\Delta} + 2L)\psi_{\beta\gamma} + D_\beta D^\alpha \psi_{\alpha\gamma},$$

$$(7.1.3)\quad -D^\alpha (D_\alpha \psi_{\beta\bar{\gamma}} - D_\beta \psi_{\alpha\bar{\gamma}}) = \tfrac{1}{2}(\bar{\Delta} + 2L + 2e)\psi_{\beta\bar{\gamma}} + D_\beta D^\alpha \psi_{\alpha\bar{\gamma}},$$

$$(7.1.4)\quad \langle D_\alpha \psi_{\beta\gamma} - D_\beta \psi_{\alpha\gamma}, D_{\bar{\alpha}} \psi_{\bar{\beta}\bar{\gamma}} - D_{\bar{\beta}} \psi_{\bar{\alpha}\bar{\gamma}} \rangle = \langle (\bar{\Delta} + 2L)\psi_{\beta\gamma}, \psi_{\bar{\beta}\bar{\gamma}} \rangle - 2\langle D^\alpha \psi_{\alpha\gamma}, D^{\bar{\beta}} \psi_{\bar{\beta}\bar{\gamma}} \rangle,$$

$$(7.1.5)\quad \langle D_\alpha \psi_{\beta\bar{\gamma}} - D_\beta \psi_{\alpha\bar{\gamma}}, D_{\bar{\alpha}} \psi_{\bar{\beta}\gamma} - D_{\bar{\beta}} \psi_{\bar{\alpha}\gamma} \rangle = \langle (\bar{\Delta} + 2L + 2e)\psi_{\beta\bar{\gamma}}, \psi_{\bar{\beta}\gamma} \rangle - 2\langle D^\alpha \psi_{\alpha\bar{\gamma}}, D^{\bar{\beta}} \psi_{\bar{\beta}\gamma} \rangle,$$

where e is the constant Ricci curvature and $\langle\ ,\ \rangle$ denotes the global inner product defined by g. If ϕ is an anti-symmetric 2-tensor, then

$$(7.1.6)\quad L\phi_{\alpha\beta} = 0.$$

Proof. Straightforward tensor calculation. C.f. [11, Sect. 6]. Q.E.D.

Now let (J, g) be a Kähler-Einstein structure on M with volume 1 and let $h \in \mathrm{EEID}$. We decompose h into its hermitian part h_H and its anti-hermitian part h_A:

$$(7.1.7)\quad h_H(JX, JY) = h_H(X, Y),$$

$$(7.1.8)\quad h_A(JX, JY) = -h_A(X, Y).$$

Then from equation 1.5.1: $(\bar{\Delta}+2L)h=0$ and equality (7.1.4), h_A satisfies Eq. (1.5.2): $\delta h_A=0$. Obviously h_A satisfies Eqs. (1.5.1) and (1.5.3): $\operatorname{tr} h_A=0$, therefore $h_A\in\mathrm{EEID}$ and so $h_H\in\mathrm{EEID}$. Q.E.D.

7.2. *Notation.* Let (J, g) be a Kähler-Einstein structure on M with volume 1. The space of all hermitian (resp. anti-hermitian) essential infinitesimal Einstein deformations is *denoted by* EEID_H (resp. EEID_A).

7.3. **Proposition.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. Then the decomposition:*

$$\mathrm{EEID}=\mathrm{EEID}_H\oplus\mathrm{EEID}_A \tag{7.3.1}$$

holds. An anti-hermitian symmetric 2-tensor field h is an element of EEID_A if and only if the following equations are satisfied.

$$D_\alpha h_{\beta\gamma}-D_\beta h_{\alpha\gamma}=0, \tag{7.3.2}$$

$$D^\alpha h_{\alpha\beta}=0. \tag{7.3.3}$$

Moreover, if $e<0$, then $\mathrm{EEID}_H=0$. If $e=0$, then a hermitian symmetric 2-tensor field h is an element of EEID_H if and only if the following equations are satisfied.

$$D_\alpha h_{\beta\bar{\gamma}}-D_\beta h_{\alpha\bar{\gamma}}=0, \tag{7.3.4}$$

$$h_\alpha{}^\alpha=0. \tag{7.3.5}$$

Proof. Let $h\in\mathrm{EEID}_A$. Then from equality (7.1.4) and Eq. (1.5.1): $(\bar{\Delta}+2L)h=0$, Eqs. (7.3.2) and (7.3.3) follow. Conversely, if an anti-hermitian symmetric 2-tensor field h satisfies Eqs. (7.3.2) and (7.3.3), then by equality (7.1.2), Eq. (1.5.1) holds. This implies that $h\in\mathrm{EEID}$.

Assume that $e\leqq 0$ and let $h\in\mathrm{EEID}_H$. Then, from equality (7.1.5) and Eqs. (1.5.1) and (1.5.2): $\delta h=0$ it follows that, if $e<0$, $h=0$ and, if $e=0$, h satisfies Eq. (7.3.4). Equation (7.3.5) is equivalent to Eq. (1.5.3): $\operatorname{tr} h=0$. Conversely, assume that $e=0$ and that a hermitian symmetric 2-tensor field h satisfies Eqs. (7.3.4) and (7.3.5). Then

$$D^{\bar{\alpha}} h_{\bar{\alpha}\gamma}=D^{\bar{\alpha}} h_{\gamma\bar{\alpha}}=D_\gamma h^{\bar{\alpha}}{}_{\bar{\alpha}}=0, \tag{7.3.6}$$

i.e., Eq. (1.5.2) holds. Then from equality (7.1.3) Eq. (1.5.1) follows, which implies that $h\in\mathrm{EEID}$. Q.E.D.

7.4. **Proposition.** *Let (J, g) be a Kähler-Einstein structure on M with $e=0$ and with volume 1. Then*

$$\mathrm{EEID}_H\cong H^{1,1}(M,\mathbf{R})/\mathbf{R}\cdot\omega, \tag{7.4.1}$$

where ω denotes the Kähler form.

Proof. For $h\in\mathrm{EEID}_H$, we define a real 2-form ψ by $\psi=hJ$, i.e.,

$$\psi_{\alpha\bar{\beta}}=-\sqrt{-1}\,h_{\alpha\bar{\beta}}. \tag{7.4.2}$$

Then from Eq. (1.5.2), $\delta\psi=0$ follows, and from Eq. (7.3.4) it follows that $d\psi=0$. It is easy to see that this correspondence gives the isomorphism (7.4.1). Q.E.D.

8. The Space CEID on a Kähler-Einstein Manifold

Let (J, g) be a Kähler structure on M and let $I\in\mathrm{CEID}$. A tensor of type $(1, 1)$ is identified with a tensor of type $(0, 2)$ via the usual correspondence defined by the metric tensor. Thus, by Eq. (6.4.1), the tensor field I is identified with an anti-hermitian 2-tensor field, which is denoted by $I_{\alpha\beta}$ or $I_{\bar{\alpha}\bar{\beta}}$.

8.1. **Lemma.** *Let (J, g) be a Kähler-Einstein structure on M. An anti-hermitian 2-tensor field I is an element of* CEID *if and only if it satisfies the following equation*

$$(\bar{\Delta}+2L)I=0. \tag{8.1.1}$$

Proof. If $I\in\mathrm{CEID}$, then from Eq. (6.4.3) and (6.4.4):

$$D_\alpha I^{\bar{\gamma}}{}_{\beta}-D_\beta I^{\bar{\gamma}}{}_{\alpha}=0, \qquad D^\gamma I^{\bar{\alpha}}{}_{\gamma}=0,$$

and formula (7.1.2), Eq. (8.1.1) follows. If Eq. (8.1.1) holds, then from formula (7.1.4), we see that Eqs. (6.4.3) and (6.4.4) follow. Q.E.D.

Let I_S (resp. I_A) be the symmetric part (resp. antisymmetric part) of $I\in\mathrm{CEID}$. Then by Lemma 8.1, we see that I_S and $I_A\in\mathrm{CEID}$. *Denote by* CEID_S (resp. CEID_A) the space of all symmetric (resp. antisymmetric) elements of CEID.

8.2. **Proposition.** *Let (J, g) be a Kähler-Einstein structure on M. Then the decomposition*

$$\mathrm{CEID}=\mathrm{CEID}_S\oplus\mathrm{CEID}_A \tag{8.2.1}$$

holds. An antisymmetric anti-hermitian 2-tensor field I belongs to CEID_A *if and only if I is parallel, i.e.,*

$$DI=0. \tag{8.2.2}$$

In particular, we have the isomorphism:

$$\mathrm{CEID}_A\cong H^{2,0}(M, \mathbf{C}). \tag{8.2.3}$$

Proof. Let I be an antisymmetric anti-hermitian 2-tensor field. Then, by formula (7.1.6), Eq. (8.1.1) is equivalent to the equation $\bar{\Delta}I=0$, which is also equivalent to Eq. (8.2.2). The last isomorphism then follows from the equivalence between the properties of being parallel and being harmonic. Q.E.D.

8.3. **Proposition.** *Let (J, g) be a Kähler-Einstein structure on M with $e\neq 0$. Then* $\mathrm{CEID}_A=0$.

Proof. If $e \neq 0$, then by formula (7.1.1), there is no non-zero parallel holomorphic 2-tensor field. So using isomorphism (8.2.3), $\mathrm{CEID}_A = 0$.

9. Kähler Relation Between the Space EID and CID

Let (J, g) be a Kähler structure on M and (J_t, g_t) a one-parameter family of Kähler structure such that $(J_0, g_0) = (J, g)$. Then the following equations are satisfied.

$$(g_t(X, J_t Y) + g_t(Y, J_t X))' = 0, \tag{9.1.1}$$

$$d\omega_t' = 0, \tag{9.1.2}$$

where ω_t is the Kähler form defined by $\omega_t(X, Y) = g_t(X, J_t Y)$. Set $g_0' = h$, $J_0' = I$ and $\omega_0' = \phi$. Then we see that ϕ is given by the following equations

$$\phi_{\alpha\beta} = \sqrt{-1}\, h_{\alpha\beta} + I_{\alpha\beta}, \tag{9.1.3}$$

$$\phi_{\alpha\bar{\beta}} = -\sqrt{-1}\, h_{\alpha\bar{\beta}} + I_{\alpha\bar{\beta}}. \tag{9.1.4}$$

From Eq. (9.1.1), we see that

$$2\sqrt{-1}\, h_{\alpha\beta} + (I_{\alpha\beta} + I_{\beta\alpha}) = 0. \tag{9.1.5}$$

Combining with Eq. (9.1.3), we get

$$\phi_{\alpha\beta} = (1/2)(I_{\alpha\beta} - I_{\beta\alpha}). \tag{9.1.6}$$

Therefore the following relation always holds.

$$\begin{aligned} 2(d\phi)_{\alpha\beta\gamma} &= D_\alpha(I_{\beta\gamma} - I_{\gamma\beta}) + \text{alternating terms} \\ &= (D_\alpha I_{\beta\gamma} - D_\gamma I_{\beta\alpha}) + \text{alternating terms} \\ &= 0, \end{aligned} \tag{9.1.7}$$

where the last equality follows from Eq. (6.4.3): $D_\alpha I^{\bar{\gamma}}{}_\beta - D_\beta I^{\bar{\gamma}}{}_\alpha = 0$. From Eqs. (6.4.1): $I^\alpha{}_\beta = 0$ and (9.4.1), we see that

$$\phi_{\alpha\bar{\beta}} = -\sqrt{-1}\, h_{\alpha\bar{\beta}}. \tag{9.1.8}$$

Combining with Eq. (9.1.6), we get

$$(d\phi)_{\alpha\beta\bar{\gamma}} = -\sqrt{-1}\, D_\alpha h_{\beta\bar{\gamma}} + \sqrt{-1}\, D_\beta h_{\alpha\bar{\gamma}} + (1/2) D_{\bar{\gamma}}(I_{\alpha\beta} - I_{\beta\alpha}). \tag{9.1.9}$$

Thus we may set the following

9.2. *Definition.* Let (J, g) be a Kähler structure on M. A symmetric 2-tensor field h and an infinitesimal complex deformation I are said to be *Kähler related* if they satisfy the following equations

$$2\sqrt{-1}\, h_{\alpha\beta} + (I_{\alpha\beta} + I_{\beta\alpha}) = 0, \tag{9.2.1}$$

$$2\sqrt{-1}(D_\alpha h_{\beta\bar{\gamma}} - D_\beta h_{\alpha\bar{\gamma}}) = D_{\bar{\gamma}}(I_{\alpha\beta} - I_{\beta\alpha}). \tag{9.2.2}$$

9.3. **Lemma.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. Then the space EEID_A and the space CEID_S are isomorphic under a canonical correspondence: $I \to h$ defined by*

$$h_{\alpha\beta} = \sqrt{-1}\, I_{\alpha\beta}. \tag{9.3.1}$$

Moreover, this isomorphism is equivalent with the Kähler relation.

Proof. The first half is obvious by Lemma 6.4 and Proposition 7.3. Under the assumption that h is anti-hermitian and that I is symmetric, Eq. (9.2.2) always holds and Eq. (9.3.1) is equivalent with Eq. (9.2.1). Q.E.D.

9.4. **Corollary.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. Then,*

(9.4.1) *if* $e<0$, $\dim \mathrm{EEID} = 2 \dim_{\mathbf{C}} H^1(M, \Theta)$,

(9.4.2) *if* $e=0$, $\dim \mathrm{EEID} = (\dim H^{1,1}(M, \mathbf{R}) - 1) + 2(\dim_{\mathbf{C}} H^1(M, \Theta) - \dim_{\mathbf{C}} H^{2,0}(M, \mathbf{C}))$,

(9.4.3) *if* $e>0$, $\dim \mathrm{EEID} \geqq 2 \dim_{\mathbf{C}} H^1(M, \Theta)$.

Proof. Combination of Lemma 6.6, Propositions 7.3, 7.4, 8.2 and Lemma 9.3. Q.E.D.

9.5. *Remark.* Formula (9.4.2) on the $K3$-surface is obtained in [10, p. 174 Theorem].

9.6. *Remark.* In general, we cannot replace the inequality sign in (9.4.3) by an equality sign. For example, on $P^1(\mathbf{C}) \times P^{2m}(\mathbf{C})$, $H^1(M, \Theta) = 0$ ([9, Theorem VII, Corollary]) but $\mathrm{EEID} \neq 0$ (Example 3.17).

9.7. **Lemma.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. If $e>0$, then for any non-zero element $h \in \mathrm{EEID}_H$, there is no element $I \in \mathrm{CID}$ which is Kähler related with h.*

Proof. Assume that I is Kähler related with h. Then by Eq. (9.2.1), I is anti-symmetric, and so from Eq. (9.2.2) it follows that

$$\sqrt{-1}(D_\alpha h_{\beta\bar{\gamma}} - D_\beta h_{\alpha\bar{\gamma}}) = D_{\bar{\gamma}} I_{\alpha\beta}. \tag{9.7.1}$$

But on the other hand, by Eq. (1.5.2): $\delta h = 0$, we know that

$$D^{\bar{\gamma}} D_\alpha h_{\beta\bar{\gamma}} = D_\alpha D^{\bar{\gamma}} h_{\beta\bar{\gamma}} = 0. \tag{9.7.2}$$

Therefore $D^{\bar{\gamma}} D_{\bar{\gamma}} I_{\alpha\beta} = 0$, which implies that

$$D_{\bar{\gamma}} I_{\alpha\beta} = 0. \tag{9.7.3}$$

Combining Eqs. (9.7.1), (1.5.1): $(\bar{\Delta} + 2L) h = 0$, (1.5.2): $\delta h = 0$ and (7.1.3), we see that $e h = 0$. Q.E.D.

9.8. **Lemma.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. The space of all infinitesimal complex deformations $I \in \mathrm{CID}$ which are Kähler related*

with $0 \in \mathrm{EID}$ *coincides with the space* CEID_A. *The space of all infinitesimal Einstein deformations* $h \in \mathrm{EID}$ *which are Kähler related with* $0 \in \mathrm{CID}$ *vanishes if* $e<0$, *coincides with the space* EEID_H *if* $e=0$ *and coincides with the space* $\{L_X g;$ X *is a holomorphic vector field*$\}$ *if* $e>0$.

Proof. If $I \in \mathrm{CEID}_A$, then I is parallel by Proposition 8.2 and so I is Kähler related with $0 \in \mathrm{EID}$. Conversely if $I \in \mathrm{CID}$ is Kähler related with $0 \in \mathrm{EID}$, then I is antisymmetric and $D_{\bar{\gamma}} I_{\alpha\beta}=0$, which implies that Eq. (6.4.4) holds. Therefore $I \in \mathrm{CEID}_A$.

If $e=0$ and if $h \in \mathrm{EEID}_H$, then by Eq. (7.3.4), h is Kähler related with $0 \in \mathrm{CID}$. If $e>0$ and if X is a holomorphic vector field on M, then it is easy to see that the infinitesimal Einstein deformation $L_X g$ is Kähler related with $L_X J = 0 \in \mathrm{CID}$. Conversely, assume that $h \in \mathrm{EID}$ is Kähler related with $0 \in \mathrm{CID}$. From decomposition (2.2.1), there exist an element $\psi \in \mathrm{EEID}$ and a vector field X on M such that $h=\psi+L_X g$. Then ψ is Kähler related with $-L_X J \in \mathrm{CID}$. By Lemma 9.3, there is an element $I_1 \in \mathrm{CEID}_S$ which is Kähler related with the anti-hermitian part ψ_A of ψ, so the hermitian part ψ_H of ψ is Kähler related with $-(I_1+L_X J)$. If $e<0$, then $\psi_H=0$ by Proposition 7.3. If $e>0$, then $\psi_H=0$ by Lemma 9.7. If $e=0$, then we have seen that ψ_H is Kähler related with $0 \in \mathrm{CID}$. Thus, in any case, $I_1+L_X J$ is Kähler related with $0 \in \mathrm{EID}$. But we have seen that then $I_1+L_X J \in \mathrm{CEID}_A$, which implies that $L_X J=0$ and $I_1=0$ by decompositions (6.5.1) and (7.3.1). Therefore $\psi_A=0$. If $e<0$, then we have seen that $\psi=\psi_H=0$ and so $h=0$ since there is no non-zero holomorphic vector field. If $e=0$, then $h=\psi \in \mathrm{EEID}_H$ since all holomorphic vector fields are Killing vector fields. If $e>0$, then we have seen that $\psi=\psi_H=0$ and so $h=L_X g$ where X is a holomorphic vector field. Q.E.D.

9.9. **Proposition.** *Let* (J, g) *be a Kähler-Einstein structure on* M *with volume* 1. *Let* $h_1 \in \mathrm{EID}$ *and* $I_1 \in \mathrm{CID}$. *We decompose them by decompositions* (2.2.1), (6.5.1), (7.3.1) *and* (8.2.1) *as*

$$h_1=h+L_X g, \quad h=h_A+h_H; \quad h_A \in \mathrm{EEID}_A,\ h_H \in \mathrm{EEID}_H, \tag{9.9.1}$$

$$I_1=I+L_Y J, \quad I=I_S+I_A; \quad I_S \in \mathrm{CEID}_S,\ I_A \in \mathrm{CEID}_A. \tag{9.9.2}$$

Then h_1 *and* I_1 *are Kähler related if and only if condition* (9.9.3), *or equivalently one of conditions* (9.9.4), (9.9.5) *and* (9.9.6), *holds.*

(9.9.3) $X-Y$ *is a holomorphic vector field,* h_H *is Kähler related with* $0 \in \mathrm{CID}$ *and* h_A *is Kähler related with* I_S.

(9.9.4) $e<0$, $X=Y$ *and* $h_{\alpha\beta}=\sqrt{-1}\, I_{\alpha\beta}$.

(9.9.5) $e=0$, $X-Y$ *is a Killing vector field and* $2h_{\alpha\beta}=\sqrt{-1}(I_{\alpha\beta}+I_{\beta\alpha})$.

(9.9.6) $e>0$, $X-Y$ *is a holomorphic vector field,* $h_{\alpha\bar{\beta}}=0$ *and* $h_{\alpha\beta}=\sqrt{-1}\, I_{\alpha\beta}$.

Proof. Assume condition (9.6.3). Then $L_X g$ and $L_Y J$ are Kähler related and h is Kähler related with I_S. If $e \neq 0$, then $I_A=0$ by Proposition 8.3. Even if $e=0$, I_A is Kähler related with $0 \in \mathrm{EID}$ by Lemma 9.8. Therefore h_1 is Kähler related

with I_1. Conversely, assume that h_1 and I_1 are Kähler related. Then since $L_X g$ is Kähler related with $L_X J$, h is Kähler related with $I+L_{(Y-X)}J$. By Lemma 9.3, there is an element $I_2 \in \mathrm{CEID}_S$ which is Kähler related with h_A. Then h_H is Kähler related with $I-I_2+L_{(Y-X)}J$. But here, if $e<0$ then $h_H=0$ by Lemma 7.3, if $e=0$ then h_H is Kähler related with $0\in\mathrm{CID}$ by Lemma 9.8, and if $e>0$ then $h_H=0$ by Lemma 9.7. Thus in any case h_H is Kähler related with $0\in\mathrm{CID}$ and so $I-I_2+L_{(Y-X)}J$ is Kähler related with $0\in\mathrm{EID}$. Then by Lemma 9.8, $I-I_2+L_{(Y-X)}J\in\mathrm{CEID}_A$. This implies that $I-I_2\in\mathrm{CEID}_A$ and $L_{(Y-X)}J=0$, i.e., $Y-X$ is a holomorphic vector field and $I_2=I_S$. But by definition of I_2, I_2 is Kähler related with h_A.

To prove the equivalence between conditions (9.9.3) and (9.9.4), (9.9.5), (9.9.6), it is sufficient to see the following. If $e<0$, then there is no non-zero holomorphic vector field. If $e=0$, then all holomorphic vector fields are Killing vector fields. If $e\leqq 0$, then h_H is always Kähler related with $0\in\mathrm{CID}$ by Lemmas 7.3 and 9.8. If $e>0$, then h_H is Kähler related with $0\in\mathrm{CID}$ only if $h_H=0$ by Lemma 9.7. If $e\neq 0$, then $I_S=I$ by Proposition 8.3. Combining these informations with Lemma 9.3 gives the equivalence. Q.E.D.

10. Einstein Metrics and Complex Structures

We expect that in the situation of Proposition 0.5, if we deform the complex structure J, then the Einstein metric g depends C^∞-ly on J. The following result justifies this observation also valid for the case of positive Chern class.

10.1. **Proposition.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. If the constant Ricci curvature $e>0$, then we assume that there is no non-zero holomorphic vector field. Let J_t be a one-parameter complex deformation of J. Then there exists an Einstein deformation g_t (defined for small t) of g such that each metric g_t is a Kähler metric compatible with J_t. Moreover, if we have an infinitesimal Einstein deformation $h\in\mathrm{EID}$ which is Kähler related with $I=J_0'$, then we can choose g_t so that $g_0'=h$.*

Proof. Recall the formula

$$\rho=\sqrt{-1}\,\partial\bar{\partial}\log|g| \tag{10.1.1}$$

for a Kähler structure, where ρ is the Ricci form defined by $\rho_{ij}=r_{ik}J^k{}_j$, ∂ and $\bar{\partial}$ the ordinary differential operators defined by J and $|g|=\det(g_{\alpha\bar{\beta}})$ for a complex coordinate system. By [22, Theorem 15], there is a one-parameter family $\hat{g}_t$ of riemannian metrics on M such that each $\hat{g}_t$ is a Kähler metric compatible with J_t. First we assume that $e\neq 0$. Set

$$\bar{g}_t=e^{-1}\hat{r}_t, \tag{10.1.2}$$

where $\hat{r}_t$ is the Ricci tensor of $\hat{g}_t$. Then the function f_t defined by

$$f_t=\log(|\bar{g}_t|\,|\hat{g}_t|^{-1}) \tag{10.1.3}$$

is well-defined and, by formula (10.1.1), satisfies the equation

(10.1.4) $$\bar{\rho}_t - e\,\bar{\omega}_t = \sqrt{-1}\,\partial_t \bar{\partial}_t f_t,$$

where $\bar{\omega}_t$ is the Kähler form of $(J_t, \bar{g}_t)$. We consider the equation

(10.1.5) $$\log(|\bar{\omega}_t + \sqrt{-1}\,\partial_t \bar{\partial}_t \psi|_t\, |\bar{\omega}_t|_t^{-1}) + f_t - e\,\psi = 0$$

for a function ψ. By formula (10.1.1), we can check that the Kähler metric defined by the Kähler form

(10.1.6) $$\omega_t = \bar{\omega}_t + \sqrt{-1}\,\partial_t \bar{\partial}_t \psi$$

is an Einstein metric (c.f. [5, Eq. I, II$^{\pm}$]), Consider the map: $\mathbf{R} \times H^{s+2}(M) \to H^s(M)$ $(s \geqq N+2 = [n/2]+3)$ defined by

(10.1.7) $$(t, \psi) \mapsto \log(|\bar{\omega}_t + \sqrt{-1}\,\partial_t \bar{\partial}_t \psi|_t\, |\bar{\omega}_t|_t^{-1}) + f_t - e\,\psi.$$

Then the derivative of this map with respect to ψ at $(0, 0)$ is given by

(10.1.8) $$\psi' \mapsto (1/2)\,\Delta\psi' - e\,\psi',$$

which is an isomorphism from $H^{s+2}(M)$ onto $H^s(M)$. In fact if $2e$ is an eigenvalue of Δ with ψ' as a corresponding eigenfunction, then $Jd\psi'$ becomes a holomorphic vector field ([26, pp. 134–136, 147]). Thus the implicit function theorem implies that a solution $\psi_t \in H^s(M)$ of (10.1.5) exists and depends C^∞-ly on t. By changing constant factor if necessary, we obtain an H^s-Einstein deformation of g defined by Eq. (10.1.6). Then $g_0' \in \mathrm{EID}$ is Kähler related with I, so $g_0' - h$ is Kähler related with $0 \in \mathrm{CID}$. This implies that $g_0' = h$ by Lemma 9.8.

If $e = 0$, the proof is less simple. By assumption, $\hat{g}_0'$ and h are Kähler related with the same I, so $\hat{g}_0' - h$ is Kähler related with $0 \in \mathrm{CID}$. Then from Eqs. (9.2.1) and (9.2.2), it follows that the tensor field $\hat{\psi} = (\hat{g}_0' - h)J$ is a closed hermitian form. [20, Theorem 4.2] says that the dimension of the space $H^{1;1}_t(M, \mathbf{R})$ defined by $(J_t, \hat{g}_t)$ is constant for t. Hence, by Lemma 4.3, we obtain a one-parameter family $\hat{\psi}_{Ht}$ of 2-forms such that $\hat{\psi}_{H0} = \hat{\psi}$ and that each $\hat{\psi}_{Ht}$ is in $H^{1;1}_t(M, \mathbf{R})$. We set

(10.1.9) $$\bar{g}_t = \hat{g}_t + t\,\phi_t,$$

where ϕ_t is defined by $\phi_t = \hat{\psi}_{Ht} J_t$. Then we see that $\bar{g}_t$ is a Kähler metric compatible with J_t and that $\bar{\omega}_0' - hJ$ is cohomologous to 0. Now we define a function f_t by

(10.1.10) $$\bar{\rho}_t = \sqrt{-1}\,\partial_t \bar{\partial}_t f_t.$$

Since $\bar{\rho}_t$ is cohomologous to 0, such an f_t exists and is unique up to constant for each t. Then, by Lemma 4.3, we see that such a function f_t can be taken to depend C^∞-ly on t and so that $f_0 = 0$ (which is obvious when $e \neq 0$). We replace the map (10.1.7) by the map: $\mathbf{R} \times \mathrm{Ker}(\int_g | H^{s+2}(M)) \times \mathbf{R} \to H^s(M)$ defined by

(10.1.11) $$(t, \psi, c) \mapsto \log(|\bar{\omega}_t + \sqrt{-1}\,\partial_t \bar{\partial}_t \psi|_t\, |\bar{\omega}_t|_t^{-1}) + f_t + c.$$

The derivative of this map with respect to ψ and c at $(0, 0, 0)$ is given by

$$(\psi', c') \mapsto (1/2)\,\Delta\psi' + c', \tag{10.1.12}$$

which is an isomorphism from $\operatorname{Ker} \int_g \times \mathbf{R}$ onto $H^s(M)$. Thus the implicit function theorem implies that the solution $\psi_t \in \operatorname{Ker} \int_g$ exists, depends C^∞-ly on t and gives an Einstein metric defined by Eq. (10.1.6). Moreover, $\omega_0' - \bar{\omega}_0'$ and $\bar{\omega}_0' - hJ$ are cohomologous to 0, and so is $\omega_0' - hJ$. But, by Lemma 9.8, we see that $g_0' - h \in \mathrm{EEID}_H$ and so $\omega_0' - hJ$ is harmonic (see Proposition 7.4). Thus $g_0' = h$. By changing the constant factor if necessary, we obtain an H^s-Einstein deformation of g.

Finally, we show the smoothness. Since each g_t is a C^2-Kähler-Einstein metric, g_t is C^∞ by [12, Theorem 6.1] (c.f. Proposition 0.2). Since the solution g_t is uniquely constructed in the above proof, we can repeat this proof, i.e., we can apply the implicit function theorem, for any $r \geqq s$ at each t. This means that g_t is a C^∞-curve in $\mathscr{M}^r$ for all $r \geqq s$. Q.E.D.

10.2. *Definition.* A complex structure J on M is said to *belong to a non-singular complete family of complex structures* if there is a family J_t of complex structures on M with parameter space P such that $J_0 = J$ and that the map: $T_0 P \to \mathrm{CID}$ defined by

$$v \mapsto [v, J_t] \tag{10.2.1}$$

is an isomorphism onto CEID.

10.3. *Remark.* By Lemma 6.6, it is obvious that the above definition does not depend on the choice of the riemannian metric with respect to which the space CEID is defined.

10.4. *Remark.* Recall Proposition 0.10. That is, if $H^2(M, \Theta)$ vanishes, then J belongs to a non-singular complete family of complex structures.

10.5. **Theorem.** *Let (J, g) be a Kähler-Einstein structure on M with volume 1. Assume that the complex structure J belongs to a non-singular complete family of complex structures. Moreover, if $e > 0$, we assume that $\mathrm{EEID}_H(J, g)$ vanishes and that there is no non-zero holomorphic vector field. Then any Einstein metric g_1 on M sufficiently close to g is Kählerian, that is, there is a complex structure J_1 on M such that g_1 becomes a Kähler metric compatible with J_1. Moreover, such J_1 can be taken to depend C^∞-ly on g_1.*

10.6. *Remark.* If the manifold M is of dimension 2 or the $K3$-surface, then the condition that g_1 is close to g is not necessary. See Proposition 0.3. Remark also that in these two cases the assumption for the original Kähler-Einstein structure (J, g) is satisfied.

Proof. In the following, we omit the suffix s which means H^s since all objects are C^∞ and the mappings p, q and χ in Lemma 2.4 may be considered to be C^∞. By Proposition 10.1, we have a C^∞-map ε: $P \times \mathrm{EEID}_H \to \mathscr{M}_1$ such that each $(t, h) \in P \times \mathrm{EEID}_H$ corresponds to a Kähler-Einstein metric compatible

with J_t. Moreover, by Proposition 7.3 and Lemma 6.8, the image of the differential ε'_0 at the origin coincides with the space EEID. We take the composition $p\circ\varepsilon$. Then the image of the differential $(p\circ\varepsilon)'_0$ at the origin coincides with the space EEID. Owing to Theorem 3.1, it means that the local pre-moduli space ELPM(g) locally becomes a submanifold of $\mathscr{S}_g$ whose tangent space at g coincides with EEID and that $p\circ\varepsilon$ is a local submersion from $P\times\mathrm{EEID}_H$ onto ELPM defined on an open neighbourhood of the origin. Therefore there is a cross section ψ: ELPM$\to P\times\mathrm{EEID}_H$. Now we may assume that the metric g_1 has volume 1. Since $p(g_1)\in$ELPM, we can define $\psi p(g_1)$ and the equation $p\varepsilon\psi p(g_1)=p(g_1)$ holds. Therefore we see that

$$\varepsilon\psi p(g_1)=(\chi q(\varepsilon\psi p(g_1)))^* p(g_1). \tag{10.6.1}$$

Here, $\varepsilon\psi p(g_1)$ is a Kähler metric compatible with the complex structure $\psi p(g_1)$. Thus $p(g_1)$ is a Kähler metric compatible with the complex structure $(\chi q(\varepsilon\psi p(g_1)))^{-1}*(\psi p(g_1))$ and so $g_1=(\chi q(g_1))^* p(g_1)$ is a Kähler metric compatible with

$$(\chi q(g_1))^*(\chi q(\varepsilon\psi p(g_1)))^{-1}*(\psi p(g_1)). \quad \text{Q.E.D.}$$

10.7. **Corollary.** *Let (J,g) be a Kähler-Einstein structure on M with volume 1 whose complex structure J is in a non-singular complete family of complex structures. Assume that $c_1<0$ or that $c_1=0$ and the second Betti number $b_2=1$. Then there is a local one-to-one correspondence between complex structures J_1 on M and compatible Kähler-Einstein metrics g_1. In particular, the space* ELPM(g) *may be regarded as a family of complex structures on M. Moreover the spaces* $\mathrm{EEID}_H(J_1,g_1)$ *and* $\mathrm{CEID}_A(J_1,g_1)$ *vanish for all pairs (J_1,g_1).*

Proof. By Remark 10.4, we can apply Theorem 10.5. If $\mathrm{EEID}_H(J_1,g_1)=0$ and $\mathrm{CEID}_A(J_1,g_1)=0$, then by Proposition 9.9, we see that the correspondence between complex structures and Einstein metrics becomes a local diffeomorphism around J_1. The vanishings of the spaces EEID_H and CEID_A follow from Proposition 7.4 and 8.2. Q.E.D.

10.8. *Example.* As a particular case, we can apply Theorem 10.5 to complex hypersurfaces of a complex projective space. Let $V_{m,d}$ be the set of all homogeneous polynomials f on $\mathbf{C}^{m+2}$ such that f defines a non-singular irreducible hypersurface in $P^{m+1}(\mathbf{C})$. The complex automorphism group $SL(m+2,\mathbf{C})$ of $P^{m+1}(\mathbf{C})$ acts canonically on $V_{m,d}$. Let $H_{m,d}$ be the quotient space $V_{m,d}/SL(m+2,\mathbf{C})$. The space $H_{m,d}$ may be regarded as a set of complex hypersurfaces of $P^{m+1}(\mathbf{C})$. Let $(M,J)\in H_{m,d}$. If $d\geqq m+2$, $m\geqq 2$ and $(m,d)\neq(2,4)$, then the first Chern class c_1 of (M,J) is negative or vanishes. Therefore we can apply Proposition 0.5. Let $MH_{m,d}$ be the set of all Kähler-Einstein structures obtained by Proposition 0.5. [21, Theorem 14.1] says that, under the same assumption for m and d, for any $(M,J)\in H_{m,d}$ the complex structure J is in a non-singular complete family of complex structures, and this family may be regarded as an open neighbourhood of J in $H_{m,d}$. Thus we can apply Theorem 10.5. Let $(M,J,g)\in MH_{m,d}$. If g_1 is an Einstein metric on M which is sufficiently close to g, then there exists a complex structure J_1 on M such that (M,J_1,g_1) is isomorphic to certain $(M,J_2,g_2)\in MH_{m,d}$ as Kähler manifolds. If

$d>m+2$, then the first Chern class of J is negative. Hence we see, by Corollary 9.4, that

$$\dim \mathrm{EEID}(g)=2\left\{\binom{m+d+1}{d}-(m+2)^2\right\}. \tag{10.8.1}$$

11. The Complex Structure on a Family of Complex Structures

In this section, we recall some results obtained by Kodaira and Spencer. The idea of our proofs is similar with that of [21] except for notations.

11.1. *Definition.* A family J_t of complex structures on M with parameter space P is said to be *normal* if the dimension of the space $H^0(M, \Theta(J_t))$ is constant, and said to be *stable* if the linear map: $T_tP \to H^1(M, \Theta(J_t))$ defined by

$$v \longmapsto [v, J_t] \tag{11.1.1}$$

is an isomorphism onto a complex subspace of $H^1(M, \Theta(J_t))$ for each $t \in P$.

Let (J_t, g_t) be a family of Kähler structures on M with parameter space P. For $t \in P$ and $v \in T_tP$, we can define an element $I \in \mathrm{CEID}(J_t, g_t)$ by

$$[v, J_t]=I+[X, J_t] \tag{11.1.2}$$

for some vector field X on M. We set

$$v^H=v-X. \tag{11.1.3}$$

If the family J_t is stable, then there is a unique vector $w \in T_tP$ such that $[w^H, J_t] = J_t[v^H, J_t]$. We define a complex structure J^P on T_tP by

$$[(J^P v)^H, J_t]=J_t[v^H, J_t]. \tag{11.1.4}$$

11.2. **Proposition** (c.f. [21, Proposition 11.1]). *Let (J_t, g_t) be a family of Kähler structures on M. Assume that the family J_t is normal and stable. Then the operator J^P depends C^∞-ly on $t \in P$ and becomes a complex structure on P.*

Proof. Since the family J_t is normal, we can apply Lemma 4.3 and see that X can be taken to depend C^∞-ly on t and $v \in T_tP$. Therefore, if v is a vector field on P, then we may assume that v^H is also a C^∞-vector field, which implies that J^P is a C^∞-tensor field on P. Now, we show that this almost complex structure J^P on P is integrable. For $v, w \in T_tP$, we set

$$A(v, w)=[v^H, w^H]-[v, w]^H. \tag{11.2.1}$$

It is a well-defined vector field on M for each pair $v, w \in T_tP$. I.e., $A(v, w)$ does not depend on the extension of v and w. Moreover we set

$$N^A(v, w)=A(v, w)-A(J^P v, J^P w)+J_t A(J^P v, w)+J_t A(v, J^P w) \tag{11.2.2}$$

and denote by N^P the Nijenhuis torsion tensor of J^P. We see that

(11.2.3) $[N^P(v, w)^H, J_t]$

$$
\begin{aligned}
&= [[v, w]^H, J_t] - [[J^P v, J^P w]^H, J_t] + [J^P[J^P v, w]^H, J_t] + [J^P[v, J^P w]^H, J_t] \\
&= [[v^H, w^H], J_t] - [A(v, w), J_t] - [[J^P v^H, J^P w^H], J_t] + [A(J^P v, J^P w), J_t] \\
&\quad + J_t[[J^P v^H, w^H], J_t] - J_t[A(J^P v, w), J_t] + J_t[[v^H, J^P w^H], J_t] - J_t[A(v, J^P w), J_t] \\
&= -[N^A(v, w), J_t] \\
&\quad + [v^H, [w^H, J_t]] - [w^H, [v^H, J_t]] - [J^P v^H, [J^P w^H, J_t]] + [J^P w^H, [J^P v^H, J_t]] \\
&\quad + J_t[J^P v^H, [w^H, J_t]] - J_t[w^H, [J^P v^H, J_t]] + J_t[v^H, [J^P w^H, J_t]] - J_t[J^P w^H, [v^H, J_t]] \\
&= -[N^A(v, w), J_t] + [v^H, [w^H, J_t]] - [w^H, [v^H, J_t]] \\
&\quad - [J^P v^H, J_t]\,[w^H, J_t] - J_t[J^P v^H, [w^H, J_t]] + [J^P w^H, J_t]\,[v^H, J_t] + J_t[J^P w^H, [v^H, J_t]] \\
&\quad + J_t[J^P v^H, [w^H, J_t]] - J_t[w^H, J_t]\,[v^H, J_t] + [w^H, [v^H, J_t]] \\
&\quad + J_t[v^H, J_t]\,[w^H, J_t] - [v^H, [w^H, J_t]] - J_t[J^P w^H, [v^H, J_t]] \\
&= -[N^A(v, w), J_t].
\end{aligned}
$$

Since the family J_t is stable, it means that $N^P(v, w)=0$. Q.E.D.

We have seen also the following

11.3. **Corollary.** *In the situation of* Proposition 11.2, $N^A(v, w)$ *is a holomorphic vector field for* J_t *for each* $v, w \in T_t P$.

Next, if $H^0(M, \Theta(J_t))=0$, then we can define an almost complex structure J^T on $T=M\times P$ by

$$J^T v^H = (J^P v)^H \quad \text{for } v \in TP, \tag{11.3.1}$$

$$J^T X = J_t X \quad \text{for } X \in TM, \tag{11.3.2}$$

since the vector field introduced in Eq. (11.1.2) is unique, and so the vector v^H is well-defined for each $v \in T_t P$.

11.4. **Proposition** (c.f. [21, Proposition 18.3, Theorem 18.4]). *Let* (J_t, g_t) *be a family of Kähler structures on* M *with parameter space* P. *If the family* J_t *is normal and stable and if* $H^0(M, \Theta(J_0))=0$, *then the almost complex structure* J^T *on* T *defined above is integrable.*

Proof. Denote by N^T the Nijenhuis tensor of J^T. Then $N^T(X, Y)=0$ for X, $Y \in TM$. For $v \in TP$ and $X \in TM$, we see

(11.4.1)

$$
\begin{aligned}
N^T(v^H, X) &= [v^H, X] - [J^P v^H, J_t X] + J_t[J^P v^H, X] + J_t[v^H, J_t X] \\
&= [v^H, X] - [J^P v^H, J_t]\, X - J_t[J^P v^H, X] + J_t[J^P v^H, X] + J_t[v^H, J_t]\, X - [v^H, X] \\
&= 0.
\end{aligned}
$$

For $v, w \in TP$, we see

$$(11.4.2)\quad \begin{aligned} N^T(v^H, w^H) &= [v^H, w^H] - [J^P v^H, J^P w^H] + J^T[J^P v^H, w^H] + J^T[v^H, J^P w^H] \\ &= A(v, w) + [v, w]^H - A(J^P v, J^P w) - [J^P v, J^P w]^H \\ &\quad + J_t A(J^P v, w) + J^P[J^P v, w]^H + J_t A(v, J^P w) + J^P[v, J^P w]^H \\ &= N^A(v, w) + N^P(v, w)^H. \end{aligned}$$

But here $N^P(v, w)=0$ by Proposition 11.2 and $N^A(v, w)=0$ by assumption and Corollary 11.3. Q.E.D.

12. The Canonical Riemannian Metric on a Family of Kähler-Einstein Structures

Let (J_t, g_t) be a family of Kähler-Einstein structures on M. If the family J_t is normal and stable and if the family g_t is normal and effectively parametrized, then the parameter space P can be endowed with the canonical complex structure and the canonical riemannian metric.

12.1. *Definition.* A family (J_t, g_t) of Kähler-Einstein structures on M with volume 1 is said to be *normal* if the family J_t is normal, and said to be *stable* if the family J_t is stable and if the spaces $\mathrm{CEID}_A(J_t, g_t)$ and $\mathrm{EEID}_H(J_t, g_t)$ vanish for all $t \in P$.

12.2. **Lemma.** *If a family (J_t, g_t) of Kähler-Einstein structures on M is normal and stable, then the family g_t is normal and effectively parametrized. In particular, the parameter space P can be endowed with the canonical riemannian metric.*

Proof. Let (J, g) be a Kähler-Einstein structure on M. We know that if $e<0$ then $\mathrm{Ker}\,\delta_g^*=0$, if $e=0$ then $\mathrm{Ker}\,\delta_g^*=H^0(M, \Theta)$ and if $e>0$ then $\mathrm{Ker}\,\delta_g^* + J(\mathrm{Ker}\,\delta_g^*)=H^0(M, \Theta)$. Therefore the normality of the family J_t implies the normality of the family g_t. Let $v \in T_t P$ and assume that $[v, g_t] \in \mathrm{Im}\,\delta_{g_t}^*$. I.e., there is a vector field X on M such that $[v, g_t]=[X, g_t]$. Then the assumption and Proposition 9.9 imply that $[v, J_t]$ is decomposed into $I_S+[Y, J_t]$, where $I_S \in \mathrm{CEID}(J_t, g_t)$ is Kähler related to $0 \in \mathrm{EEID}(g_t)$. But then Lemma 9.3 says that $I_S=0$, which contradicts the stability of the family J_t. Q.E.D.

12.3. **Theorem.** *Let (J_t, g_t) be a normal and stable family of Kähler-Einstein structures on M. Then the canonical riemannian metric g^P on the parameter space P is a Kähler metric compatible with the complex structure J^P on P.*

Proof. For $v \in T_t P$, let v^H be the vector field defined by Eq. (4.1.3). Since $[v^H, J_t] \in \mathrm{CID}(J_t)$ is Kähler related with $[v^H, g_t] \in \mathrm{EEID}_A(J_t, g_t)$, we see, by Proposition 9.9, that $[v^H, J_t] \in \mathrm{CEID}_S(J_t, g_t)$. Therefore, the notation v^H does not contradict that induced by Eq. 11.1.3. Moreover, by Lemma 9.3,

$$(12.3.1)\quad [v^H, \omega_t] = [v^H, g_t J_t] = [v^H, g_t]\, J_t + g_t[v^H, J_t] = 0.$$

and so

$$g^P(v,w)=\int_M([v^H,g_t],[w^H,g_t])\,v_{g_t}$$
$$=\int_M(g_t[v^H,J_t]\,J_t,g_t[w^H,J_t]\,J_t)\,v_{g_t}$$
$$=\int_M([v^H,J_t],[w^H,J_t])\,v_{g_t}. \tag{12.3.2}$$

Therefore,

$$g^P(J^Pv,J^Pw)=\int_M([J^Pv^H,J_t],[J^Pw^H,J_t])\,v_{g_t}$$
$$=\int_M(J_t[v^H,J_t],J_t[w^H,J_t])\,v_{g_t}$$
$$=g^P(v,w), \tag{12.3.3}$$

i.e., g^P is a hermitian metric. Now we calculate the exterior derivative $d\omega^P$ of the Kähler form ω^P of g^P.

$$\omega^P(v,w)=g^P(v,J^Pw)$$
$$=\int_M([v^H,J_t],[J^Pw^H,J_t])\,v_{g_t}$$
$$=\int_M([v^H,J_t],J_t[w^H,J_t])\,v_{g_t}. \tag{12.3.4}$$

Then assuming that $[v,w]=[w,z]=[z,v]=0$, we see

$$(d\omega^P)(v,w,z)$$
$$=v\int_M([w^H,J_t],J_t[z^H,J_t])\,v_{g_t}$$
$$+\text{alternating terms}. \tag{12.3.5}$$

Here,

$$\begin{aligned}v\{([w^H,J_t],J_t[z^H,J_t])\,v_{g_t}\}&\\
=&-[v^H,g_t]^{ij}(g_t)_{km}[w^H,J_t]^k{}_i(J_t[z^H,J_t])^m{}_j v_{g_t}\\
&+(g_t)^{ij}[v^H,g_t]_{km}[w^H,J_t]^k{}_i(J_t[z^H,J_t])^m{}_j v_{g_t}\\
&+([v^H,[w^H,J_t]],J_t[z^H,J_t])\,v_{g_t}\\
&+([w^H,J_t],[v^H,J_t]\,[z^H,J_t])\,v_{g_t}\\
&+([w^H,J_t],J_t[v^H,[z^H,J_t]])\,v_{g_t}\\
&+([w^H,J_t],J_t[z^H,J_t])\,(1/2)\,[v^H,g_t]^m{}_m v_{g_t},\end{aligned} \tag{12.3.6}$$

and

$$[v^H,g_t]^{ij}(g_t)_{km}[w^H,J_t]^k{}_i(J_t[z^H,J_t])^m{}_j$$
$$=2\,\mathrm{Re}\{[v^H,g_t]^{\alpha\beta}(g_t)_{\bar\gamma\delta}[w^H,J_t]^{\bar\gamma}{}_\alpha(J_t)^\delta{}_\varepsilon[z^H,J_t]^\varepsilon{}_\beta\}=0, \tag{12.3.7}$$

$$(g_t)^{ij}[v^H,g_t]_{km}[w^H,J_t]^k{}_i(J_t[z^H,J_t])^m{}_j=0, \tag{12.3.8}$$

$$([w^H,J_t],[v^H,J_t]\,[z^H,J_t])=2\,\mathrm{Re}\{[w^H,J_t]_\alpha{}^\beta[v^H,J_t]^\alpha{}_{\bar\gamma}[z^H,J_t]^{\bar\gamma}{}_\beta\}=0, \tag{12.3.9}$$

$$[v^H,g_t]^m{}_m=0. \tag{12.3.10}$$

Therefore,

$$(12.3.11)\qquad \begin{aligned}(d\omega^P)(v,w,z) &= \int_M ([v^H,[w^H,J_t]],J_t[z^H,J_t])\,v_{g_t} \\ &\quad -\int_M ([v^H,[z^H,J_t]],J_t[w^H,J_t])\,v_{g_t} \\ &\quad +\text{alternating terms} \\ &= \int_M ([A(v,w),J_t],[J^P z^H,J_t])\,v_{g_t} \\ &\quad +\text{alternating terms.}\end{aligned}$$

But here $[J^P z^H, J_t]\in \mathrm{CEID}(J_t, g_t)$ and $A(v, w)$ is a vector field on M. Thus

$$(12.3.12)\qquad d\omega^P=0. \quad \text{Q.E.D.}$$

12.4. **Corollary.** *In the situation of* Corollary 10.7, *the space* ELPM(g) *regarded as a family* (J_t, g_t) *of Kähler-Einstein structures on* M *becomes a Kähler manifold. Moreover, the complex structure is real analytic with respect to the real analytic structure of* ELPM(g).

Proof. By Theorem 4.13, the canonical riemannian metric, which is a Kähler metric, is real analytic. Q.E.D.

Assume that $H^0(M, \Theta(J_t))=0$. In Proposition 11.4 we defined a complex structure J^T on $T=M\times P$. Here we define a riemannian metric g^T on T by

$$(12.4.1)\qquad g^T(v^H, w^H)=g^P(v, w) \quad \text{for } v, w\in TP,$$

$$(12.4.2)\qquad g^T(X, Y)=g_t(X, Y) \quad \text{for } X, Y\in TM,$$

$$(12.4.3)\qquad g^T(v^H, X)=0 \quad \text{for } v\in TP,\ X\in TM.$$

Obviously, the metric g^T is a hermitian metric.

12.5. **Proposition.** *Let* (J_t, g_t) *be a normal and stable family of Kähler-Einstein structures on* M. *Assume that* $H^0(M, \Theta(J_t))=0$. *Then the hermitian metric* g^T *is a Kähler metric if and only if* $A=0$, *where* A *is defined by* Eq. (11.2.1).

Proof. Denote by ω^T the Kähler form of g^T. First we see that

$$(12.5.1)\qquad (d\omega^T)(X, Y, Z)=0 \quad \text{for } X, Y, Z\in TM$$

by definition. Next, for $v\in TP$, we extend X and $Y\in TM$ so that $[X, Y]=0$, $[v^H, X]=0$ and $[v^H, Y]=0$. Then

$$(12.5.2)\qquad \begin{aligned}(d\omega^T)(v^H, X, Y) &= v^H(\omega^T(X, Y))+X(\omega^T(Y, v^H))+Y(\omega^T(v^H, X)) \\ &= v^H(\omega_t(X, Y))=[v^H, \omega_t](X, Y) \\ &= 0. \quad (12.3.1)\end{aligned}$$

Moreover, for $v, w, z\in TP$,

$$(12.5.3)\qquad \begin{aligned}(d\omega^T)(v^H, w^H, z^H) &= v^H(\omega^T(w^H, z^H)) - \omega^T([v^H, w^H], z^H)\\ &\quad + \text{alternating terms}\\ &= v(\omega^P(w, z)) - \omega^P([v, w], z)\\ &\quad + \text{alternating terms}\\ &= (d\omega^P)(v, w, z) = 0. \quad (12.3.12)\end{aligned}$$

Finally,

$$(12.5.4)\quad \begin{aligned}(d\omega^T)(v^H, w^H, X) &= v^H(\omega^T(w^H, X)) + w^H(\omega^T(X, v^H)) + X(\omega^T(v^H, w^H))\\ &\quad - \omega^T([v^H, w^H], X) - \omega^T([w^H, X], v^H) - \omega^T([X, v^H], w^H)\\ &= -\omega_t(A(v, w), X). \quad \text{Q.E.D.}\end{aligned}$$

It seems to the author that the condition $A=0$ is rather strong and probably does not occur except on the one dimensional complex torus.

12.6. *Example.* In the situation of Example 10.8, if $d>m+2$, then the space $MH_{m,d}$ is identified with the space $H_{m,d}$. They canonically become Kähler manifolds.

13. Appendix - Proof of Theorem 3.1

To prove Theorem 3.1, we recall some basic definitions and facts in the theory of real analytic objects in Banach spaces, for which we refer to [14, Chap. IV]. This category is effectively used for the Plateau problem (c.f. [30]).

13.1. *Definition.* Let V and W be Banach spaces and U an open set of V. A mapping f: $U \to W$ is said to be *real analytic* if for each point $x \in U$ f can be represented by a convergent power series around x.

13.2. *Definition.* Let V and W be complex Banach spaces and U an open set of V. A mapping f: $U \to W$ is said to be *holomorphic* if f is of class C^1 and the derivative f'_x at each point $x \in U$ commutes with the almost complex structures.

13.3 **Lemma** ([14, p. 134, Theorem 3.7]). *Let V and W be complex Banach spaces and U an open set of V. A holomorphic mapping f: $U \to W$ is real analytic.*

13.4 **Lemma** ([1, Theorem 5.7], c.f. [14, p. 144, Theorem 3.11]). *Let V and W be Banach spaces and $V^{\mathbf{C}}$, $W^{\mathbf{C}}$ their complexifications. Let U be an open set of V and f: $U \to W$ a real analytic map. Then there exists an open set $U^{\mathbf{C}}$ of $V^{\mathbf{C}}$ which contains U such that f can be extended to a holomorphic map $f^{\mathbf{C}}$: $U^{\mathbf{C}} \to W^{\mathbf{C}}$.*

The most important fact is the following

13.5. **Lemma** ([14, p. 145, Theorem 3.12]). *In the real analytic category in Banach spaces, the implicit function theorem holds.*

So we see as a corollary the following

13.6. **Lemma.** *Let V and W be Hilbert spaces and f a real analytic mapping from V to W defined on an open neighbourhood of the origin $0 \in V$. Assume that $f(0)=0$ and that the image of the differential f'_0 at 0 is closed in W. Then there is an open neighbourhood U of $0 \in V$ such that the set $f^{-1}(0) \cap U$ is a real analytic set in a real analytic submanifold Z of U whose tangent space T_0Z coincides with $\mathrm{Ker} f'_0$.*

Proof. Let $p: W \to \mathrm{Im} f_0'$ be a projection map and set $q = \mathrm{id}_W - p$. Applying Lemma 13.5 to the map $p \circ f$, we see that there is an open neighbourhood U of $0 \in V$ such that the set $(p \circ f)^{-1}(0) \cap U$ forms a real analytic submanifold of U. If we set $Z = (p \circ f)^{-1}(0) \cap U$, then $T_0 Z = \mathrm{Ker} f_0'$ and $f^{-1}(0) \cap U = (q \circ f | Z)^{-1}(0)$. Q.E.D.

To work in this category, the following Lemma is basic.

13.7. **Lemma.** *Let E and F be vector bundles over M and $E^{\mathbf{C}}$, $F^{\mathbf{C}}$ their complexifications. Let f be a C^∞-cross section of E and $\psi: E \to F$ a fiber preserving C^∞-map defined on an open set of E which contains the image of f. Assume that ψ has an extension to a fiber preserving map $\psi^{\mathbf{C}}: E^{\mathbf{C}} \to F^{\mathbf{C}}$ defined on an open set of $E^{\mathbf{C}}$ such that the restriction $\psi^{\mathbf{C}}_x$ to each fiber $E^{\mathbf{C}}_x$ is holomorphic. Then the map $\Psi: H^s(E) \to H^s(F)$ defined by*

$$\Psi(u) = \psi \circ u, \tag{13.7.1}$$

defined on an open neighbourhood of f, is real analytic provided that $s > [n/2] + 1$.

Proof. By Lemma 13.3, it is sufficient to prove that the map $\Psi^{\mathbf{C}}: H^s(E^{\mathbf{C}}) \to H^s(F^{\mathbf{C}})$ defined by

$$\Psi^{\mathbf{C}}(u) = \psi^{\mathbf{C}} \circ u, \tag{13.7.2}$$

which is an extension of Ψ, is holomorphic. But in fact $\Psi^{\mathbf{C}}$ is C^∞ ([29, Theorem 11.3]), and for each $x \in M$ we have

$$\lim_{z \to 0} \frac{1}{z} \{\Psi^{\mathbf{C}}(u + zv)(x) - \Psi^{\mathbf{C}}(u)(x)\} = (\psi^{\mathbf{C}}_x)'_{u(x)}(v(x)), \tag{13.7.3}$$

where z denotes complex number. Thus $\Psi^{\mathbf{C}}$ is holomorphic. Q.E.D.

13.8. *Remark.* This Lemma, together with the observation that the differentiation is linear, says that ordinary tensor calculus operations on a compact C^∞-manifold are real analytic with respect to some suitable H^s-topology.

Now we come back to our space $\mathscr{M}^s$.

13.9 **Lemma.** *Ebin's slice $\mathscr{S}^s_g$ is a real analytic submanifold of $\mathscr{M}^s$.*

Proof. The definition of Ebin's slice $\mathscr{S}^s_g$ reduces as follows. Let V be a finite dimensional vector space and $S^2 V$ (resp. $S^2_+ V$) the space of all symmetric (resp. positive definite symmetric) bilinear forms on V. If we fix an inner product $g_0 \in S^2_+ V$, we can define a riemannian metric on $S^2_+ V$ by

$$(\psi, \phi)_g = \mathrm{Tr}(g^{-1} \psi g^{-1} \phi) \det(g_0^{-1} g)^{1/2} \tag{13.9.1}$$

for $g \in S^2_+ V$ and $\psi, \phi \in S^2 V$. This riemannian metric depends on g_0, but only up to constant factor. Therefore the exponential map exp does not depend on g_0. Coming back to the manifold M, we define an exponential map Exp on $\mathscr{M}^s$ by

$$(\mathrm{Exp}_g h)_x = \exp_{g_x} h_x \quad \text{for } x \in M, \tag{13.9.2}$$

where $h \in H^s(S^2 M)$. The slice $\mathscr{S}^s_g$ is defined as $\mathrm{Exp}_g(U)$, where U is an open neighbourhood of the origin in $\mathrm{Ker}\, \delta_g$. But by Lemma 3.12, the map Exp_g is real analytic, hence $\mathscr{S}^s_g$ is a real analytic submanifold of $\mathscr{M}^s$. Q.E.D.

Proof of Theorem 3.1. By Lemma 2.6, we see that

$$\mathrm{ELPM}(g) = (E | \mathscr{S}^s_g \cap \mathscr{M}^s_1)^{-1}(0), \tag{13.9.3}$$

where E is defined by Eq. (1.2.1) and regarded as a real analytic map from $\mathscr{M}^s$ into $H^{s-2}(S^2 M)$. By [24, Proposition 3.2], the image of the differential $(E | \mathscr{S}^s_g \cap \mathscr{M}^s_1)'_g$ at g is closed in $H^{s-2}(S^2 M)$. Therefore the proof reduces to Lemma 13.6. Q.E.D.

13.10. *Remark.* In the proof of Theorem 3.1, we did not effectively use the fact that any Einstein metric is real analytic (Proposition 0.2).

References

1. Alexiewicz, A., Orlicz, W.: Analytic operators in real Banach spaces. Studia Math. **14**, 57–78 (1954)
2. Artin, M.: On the solutions of analytic equations. Invent. math. **5**, 277–291 (1968)
3. Aubin, T.: Equations du type Monge-Ampère sur les variétés kählériennes compactes. C. R. Acad. Sc. Paris **283**, 119–121 (1976)
4. Aubin, T.: Equations du type Monge-Ampère sur les variétés kählériennes compactes. Bull. Sc. math. **102**, 63–95 (1978)
5. Bérard Bergery, L.: Enoncé des théorèmes et mise en équation. Astérisque **58**, 77–102 (1978)
6. Berger, M.: Sur quelques variétés d'Einstein compactes. Annali di Math. Pura e Appl. **53**, 89–96 (1961)
7. Berger, M.: Quelques formules de variation pour une structure riemannienne. Ann. scient. Ec. Norm. Sup. Paris **3**, 285–294 (1970)
8. Berger, M., Ebin, D.G.: Some decompositions of the space of symmetric tensors on a riemannian manifold. J. Differential Geometry **3**, 379–392 (1969)
9. Bott, R.: Homogeneous vector bundles. Ann. of Math. **66**, 203–248 (1957)
10. Bourguignon, J.-P.: Les surface K3 (Géométrie riemannienne en dimension 4, Exposé N° VIII) Séminaire Arthur Besse, CEDIC-NATHAN, Paris 1981, pp. 157–177
11. Calabi, E., Vessentini, E.: On compact locally symmetric Kähler manifolds. Ann. of Math. **71**, 472–507 (1960)
12. DeTurck, D.M., Kazdan, J.L.: Some regularity theorems in riemannian geometry. Ann. Scient. Ec. Norm. Sup. Paris **14**, 249–260 (1981)
13. Ebin, D.G.: Espace des métriques riemanniennes et mouvement des fluides via les variétés d'applications. Centre de Math. de l'Ecole Polytechnique, Paris 1972
14. Fučík, S., Nečas, J., Souček, J., Souček, V.: Spectral Analysis of Nonlinear Operators. Lecture Notes in Math., vol. 346. Berlin-Heidelberg-New York: Springer 1973
15. Fujitani, T.: Compact suitable pinched Einstein manifolds. Bulletin of Faculty of Liberal Arts, Nagasaki Univ. Natural Science **19**, 1–5 (1979)
16. Helffer, B.: Rappels sur les espaces fonctionnels utiles aux équations aux dérivées partielles. Astérisque **58**, 9–22 (1978)
17. Hirsch, M.W.: Differential topology. GTM 33. New York: Springer 1976
18. Hitchin, N.: Compact four-dimensional Einstein manifolds. J. Differential Geometry **9**, 435–441 (1974)
19. Kodaira, K., Nirenberg, L., Spencer, D.C.: On the existence of deformations of complex analytic structures. Ann. of Math. **68**, 450–459 (1958)
20. Kodaira, K., Spencer, D.C.: On the variation of almost-complex structure. Algebraic Geometry and Topology (Fox R.H. ed.), Princeton 1957, pp. 139–150
21. Kodaira, K., Spencer, D.C.: On deformations of complex analytic structures I–II. Ann. of Math. **67**, 328–466 (1958)
22. Kodaira, K., Spencer, D.C.: On deformations of complex analytic structures III. Stability theorems for complex structures. Ann of Math. **71**, 43–76 (1960)
23. Koiso, N.: Non-deformability of Einstein metrics. Osaka J. Math. **15**, 419–433 (1978)
24. Koiso, N.: Rigidity and infinitesimal deformability of Einstein metrics. Osaka. J. Math. **19**, 643–668 (1982)
25. Kuranishi, M.: New proof for the existence of locally complete families of complex structures. Proc. of the Conf. on Complex Analysis in Minneapolis. Berlin-Heidelberg-New York: Springer 1964, pp. 142–154
26. Lichnerowicz, A.: Géométrie des groupes de transformations. Paris: Dunod 1958
27. Morrow, J., Kodaira, K.: Complex manifolds. Holt, Rinehart and Winston, Inc., USA 1971
28. Omori, H.: On the group of diffeomorphisms on a compact manifold. Global Analysis, Proc. Sympos. Pure Math. **15**, (1968); Amer. Math. Soc., pp. 167–183
29. Palais, R.S.: Foundations of Global non-linear analysis. New York: Benjamin 1968
30. Tomi, F.: On the finite solvability of Plateau's Problem. Lectures Notes in Math., vol. 597, 679–695. Berlin-Heidelberg-New York: Springer 1977
31. Yau, S.T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I. Comm. Pure and Appl. Math. **31**, 339–411 (1978)

Oblatum 25-X-1982

Invent. math. 73, 107–115 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

A Summation Formula Related to the Binary Digits

J. Coquet

Département de Mathématique, Université de Valenciennes,
F-59326 Valenciennes Cedex, France

1. Introduction

In a paper which appeared in 1969, D.J. Newman [4] proved the preponderance, among the multiples of 3, of numbers which have an even sum of binary digits over those which have an odd sum.

If $s(n)$ denotes the sum of binary digits of n, we put:

$$S(N)=\sum_{n<N}(-1)^{s(3n)}$$

Newman obtained the inequalities:

$$\frac{3^\alpha}{20}<S(N)N^{-\alpha}<5.3^\alpha \quad \text{with } \alpha=\frac{\text{Log } 3}{\text{Log } 4}$$

Our purpose is to prove in a different way the more precise:

Theorem. 1) $S(N)=\frac{\eta}{3}+N^\alpha F\left(\frac{\text{Log } N}{\text{Log } 4}\right)$ *where* F *is a continuous nowhere differentiable function of period* 1 *and where*

$$\eta=\begin{cases}0 & \text{if } N \text{ is even}\\ (-1)^{s(3N-3)} & \text{if } N \text{ is odd}\end{cases}$$

2) $\limsup_{N\to\infty} S(N)\cdot N^{-\alpha}=\text{Sup } F=\frac{55}{3}\left(\frac{3}{65}\right)^\alpha=1.601958421\ldots$

$\liminf_{N\to\infty} S(N)\cdot N^{-\alpha}=\text{Inf } F=\frac{2\sqrt{3}}{3}=1.154700538\ldots$

Moreover, $\text{Sup } F=\lim_{N\in\mathscr{A}} S(N)\cdot N^{-\alpha}$ *where* $\mathscr{A}$ *is the sequence* $k\mapsto\frac{65\cdot 4^k+1}{3}$. *And* $\text{Inf } F=\lim_{N\in\mathscr{B}} S(N)\cdot N^{-\alpha}$ *where* $\mathscr{B}$ *is the sequence* $k\mapsto 2\cdot 4^k$.

In addition, we enlighten the link between the above function F and a classical fractal object.

The first part of this theorem can be compared to a result of Delange [1] concerning the summation of the sum of digits. As indicated in paragraph VII, our method could be used to give an alternative proof of Harborth's result [3] on the summation function of $2^{s(n)}$. An excellent bibliography about related topics is contained in a paper of Stolarsky [5].

2. Construction of F

$$S(N)=\tfrac{1}{3}\sum_{n<3N}(-1)^{s(n)}+\tfrac{2}{3}\mathscr{R}e\sum_{n<3N}(-1)^{s(n)}j^n \tag{1}$$

and since $\sum_{n<2K}(-1)^{s(n)}=0$, we get $\sum_{n<3N}(-1)^{s(n)}=\begin{cases}0 & \text{if } N \text{ is even}\\ (-1)^{s(3N-3)} & \text{if } N \text{ is odd.}\end{cases}$

On the other hand, the sequence $g\colon n\mapsto(-1)^{s(n)}j^n$ is a 4-multiplicative one [2], in other words, $g(0)=1$ and if

$$n=\varepsilon_k 4^k+\ldots+\varepsilon_r 4^r+\ldots+\varepsilon_0, \qquad \varepsilon_r\in\{0,1,2,3\},$$

$$g(n)=\prod_{r=0}^{k} g(\varepsilon_r 4^r). \tag{2}$$

Thus, g is completely determined by the relations:

$$g(4^r)=-j \qquad g(2\cdot 4^r)=-j^2 \qquad g(3\cdot 4^r)=1 \tag{3}$$

From (2) and (3), we derive by induction on r:

$$\sum_{n<4^r}g(n)=3^r \qquad \sum_{n<2\cdot 4^r}g(n)=(1-j)3^r \qquad \sum_{n<3\cdot 4^r}g(n)=2\cdot 3^r \tag{4}$$

which explain the choice of basis 4 instead of basis 2.

Now, let $T=\varepsilon_0+\varepsilon_1 4+\ldots+\varepsilon_k 4^k$ and put $T_p=\varepsilon_p\cdot 4^p+\ldots+\varepsilon_k\cdot 4^k$.

$$\begin{aligned}\sum_{n<T}g(n)&=\sum_{n<T_k}g(n)+\sum_{0\le p<k}\ \sum_{T_{p+1}\le n<T_p}g(n)\\ &=\sum_{n<\varepsilon_k 4_k}g(n)+\sum_{0\le p<k}g(T_{p-1})\sum_{m<\varepsilon_p 4^p}g(m) \qquad \text{from (2)}\\ &=\sum_{n<\varepsilon_k 4^k}g(n)+\sum_{0\le p<k}g(\varepsilon_k 4^k)\ldots g(\varepsilon_{p+1}4^{p+1})\sum_{n<\varepsilon_p 4^p}g(n).\end{aligned}$$

From now on, it is more convenient to write:

$$T=e_0 4^k+\ldots+e_{k-r}4^r+\ldots+e_k \qquad \text{with } e_0\neq 0.$$

With the notations $c(0)=0$, $c(1)=1$, $c(2)=1-j$, $c(3)=2$,

$$\alpha(0)=\beta(0)=0, \qquad \alpha(i)=\text{Card}\,\{\rho<i;\ e_\rho=1\},$$

$$\beta(i)=\text{Card}\,\{\rho<i;\ e_\rho=2\},$$

we deduce from (2), (3), (4), (5) the relation:

$$\sum_{n<T} g(n)=\sum_{r\leqq k}(-j)^{\alpha(k-r)}(-j^2)^{\beta(k-r)}c(e_{k-r})3^r.$$

Putting $x=T\cdot 4^{-k}=\sum_{a=0}^{\infty} e_a 4^{-a}\in[1,4]$, we obtain:

$$\sum_{n<T} g(n)=3^k\psi(x) \quad \text{with } \psi(x)=\sum_{a=0}^{\infty}(-j)^{\alpha(a)-\beta(a)}c(e_a)\cdot 3^{-a}.$$

Joined to (1), this relation gives with $3N$ in place of T:

$$S(N)=\frac{\eta}{3}+\frac{2}{3}3^k\mathscr{R}e\psi(x)=\frac{\eta}{3}+\frac{2}{3}\left(\frac{3N}{x}\right)^{\alpha}\mathscr{R}e\psi(x)=\frac{\eta}{3}+2\cdot 3^{\alpha-1}N^{\alpha}\theta(x)$$

with $\theta(x)=x^{-\alpha}\Phi(x)$ and $\Phi(x)=\mathscr{R}e\psi(x)$.

Since $\dfrac{\operatorname{Log} x}{\operatorname{Log} 4}=\left\{\dfrac{\operatorname{Log} 3N}{\operatorname{Log} 4}\right\}$ where $\{\cdot\}$ denotes the fractional part, we get

$$F(t)=2\cdot 3^{\alpha-1}\theta(4^{\left\{t+\frac{\operatorname{Log} 3}{\operatorname{Log} 4}\right\}}).$$

F has period 1. Making use of the table in the following paragraph, we easily verify that:

$$1=\theta(1)=\lim_{x\to 1^+}\theta(x)=\lim_{x\to 4^-}\theta(x).$$

3. Continuity and Non-Differentiability

We have to show that ϕ is continuous and nowhere differentiable in [1, 4].

3.1 A Simple Writing of ϕ.

$\phi(x)=\sum_{a=0}^{\infty} d(a)3^{-a}$ where the numbers $d(a)$ are defined in the following table.

Table

$\alpha(a)-\beta(a)\to$ mod 6	0	1	2	3	4	5
$e_a=0$	0	0	0	0	0	0
$e_a=1$	1	1/2	−1/2	−1	−1/2	1/2
$e_a=2$	3/2	0	−3/2	−3/2	0	3/2
$e_a=3$	2	1	−1	−2	−1	1

3.2 ϕ is Continuous

x and y denote two real numbers satisfying $1\leqq x<y<x+4^{-u}\leqq 4$.

1) If x and y belong to the same interval $[t\cdot 4^{-u}, (t+1)4^{-u}]$,

$$|\phi(x)-\phi(t\cdot 4^{-u})|\leqq 2\sum_{a>u} 3^{-a}=3^{-u}$$

and the analogous inequality for y give:

$$|\phi(x)-\phi(y)|\leqq 23^{-u}.$$

2) On the contrary, assume that

$$(t-1)4^{-u}<x<t4^{-u}\leqq y<(t+1)4^{-u}.$$

The inequality $|\phi(y)-\phi(t\cdot 4^{-u})|\leqq 3^{-u}$ still holds. We write $t4^{-u}=\sum_{a\leqq r} e_a 4^{-a}$ with $e_r\neq 0$ and $r\leqq u$, thus

$$(t-1)4^{-u}=\sum_{a<r} e_a 4^{-a}+(e_r-1)4^{-r}+\sum_{r<a\leqq u} 3\cdot 4^{-a}.$$

With the coefficients $d(a)$ related to the expansion of x, we have:

$$\begin{aligned}\phi(x)&=\phi((t-1)4^{-u})+\sum_{a>u} d(a)3^{-a}\\ &=\sum_{a<r} d(a)3^{-a}+d(r)3^{-r}+d(r+1)\sum_{r<a\leqq u} 3^{-a}+\sum_{a>u} d(a)3^{-a}\end{aligned} \tag{6}$$

On the other hand, $d'(r)$ being determined from the expansion of $t\cdot 4^{-u}$,

$$\phi(t\cdot 4^{-u})=\sum_{a<r} d(a)3^{-a}+d'(r)3^{-r}. \tag{7}$$

With the aid of the table, we verify (18 cases) that:

$$d'(r)=d(r)+d(r-1)\sum_{b>0} 3^{-b}=d(r)+\tfrac{1}{2}d(r+1). \tag{8}$$

(6), (7) and (8) give:

$$|\phi(x)-\phi(t\cdot 4^{-u})|\leqq\sum_{a>u}|d(a)-d(r+1)|3^{-a}\leqq 3^{-u}.$$

We still get $|\phi(x)-\phi(y)|\leqq 23^{-u}$ and ϕ is continuous.

3.3 ϕ is Nowhere Differentiable

We only prove the non-differentiability at $x=\sum_{a=0}^{\infty} e_a 4^{-a}\in]1,4[$ and omit the cases $x=1$ and $x=4$ which are obvious.

Set $x_k=\sum_{a\leqq k} e_a 4^{-a}$ and $y_k=x_k+4^{-k}$. If $e_k\in\{0,1,2\}$, the table gives $|\phi(y_k)-\phi(x_k)|\geqq\frac{1}{2}3^{-k}$. If $e_k=3$, put $c=\inf\{a\in\mathbb{N};\ e_{a+1}=\ldots=e_k=3\}$. Then,

$$\phi(x_k)=\sum_{a<c} d(a)3^{-a}+d(c)3^{-c}+d(c+1)\sum_{c<a\leqq k} 3^{-a}$$

and

$$\phi(y_k)=\sum_{a<c} d(a)3^{-a}+d'(c)3^{-c}$$

where $d'(c)$ is relative to the expansion of y_k.

As in (8), $d'(c)=d(c)+\frac{1}{2}d(c+1)$ and we obtain:

$$|\phi(x_k)-\phi(y_k)|=|d(c+1)\sum_{a>k} 3^{-a}|\geqq\tfrac{1}{2}\cdot 3^{-k}$$

because $|d(c+1)|\geqq 1$ as can be seen in the last lign of the Table.

We deduce that $\lim_{k\to\infty}\left|\dfrac{\phi(y_k)-\phi(x_k)}{y_k-x_k}\right|=+\infty$ and the proof is done.

4. ψ As a Fractal Object

Let

$$x=\sum_{a=0}^{\infty} e_a 4^{-a}\in[0,4[, \qquad x_k=\sum_{a\leqq k} e_a 4^{-a}, \qquad y_k=x_k+4^{-k}$$

From

$$\psi(x_{k+1})=\psi(x_k)+(-j)^{\alpha(k+1)-\beta(k+1)}c(e_{k+1})3^{-k-1}$$

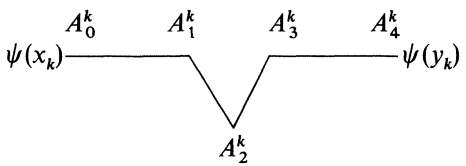
and

$$\psi(y_k)=\psi(x_k)+(-j)^{\alpha(k+1)-\beta(k+1)}3^{-k},$$

we derive:

$$\frac{\psi(x_{k+1})-\psi(x_k)}{\psi(y_k)-\psi(x_k)}=\frac{c(e_{k+1})}{3}.$$

Therefore, the sequence $k\mapsto\psi(x_k)$ which converges to $\psi(x)$ is built on the classical fractal scheme



where, if $e_{k+1}=j$, $\psi(x_{k+1})=A_j^k$ and $\psi(y_{k+1})=A_{j+1}^k$.

5. Lower Bound for θ and F

We prove that $\operatorname{Inf}\theta=\theta\left(\frac{3}{2}\right)=\left(\frac{2}{3}\right)^{\alpha}$ so that $\operatorname{Inf}F=\dfrac{2\sqrt{3}}{3}$ is the limit of $S(N)\cdot N^{-\alpha}$ when N increases through the sequence of numbers $2\cdot 4^k$.

5.1 *Out of* $\left]\frac{3}{2},\frac{7}{4}\right[$

A look at the table gives the following estimates.

1) If $1 \leqq x < \frac{3}{2}$, $\phi(x) \geqq 1$ thus $\theta(x) > \theta\left(\frac{3}{2}\right)$

2) If $\frac{7}{2} \leqq x < 2$, $\phi(x) \geqq \frac{4}{3}$ thus $\theta(x) \geqq \frac{4}{3 \cdot 2^{\alpha}} = \frac{4\sqrt{3}}{9} > \theta\left(\frac{3}{2}\right)$

3) If $2 \leqq x < \frac{5}{2}$, $\phi(x) \geqq \frac{3}{2}$ thus $\theta(x) \geqq \frac{3}{2}\left(\frac{2}{5}\right)^{\alpha} > \theta\left(\frac{3}{2}\right)$

4) If $\frac{5}{2} \leqq x < 3$, $\phi(x) \geqq \frac{11}{6}$ thus $\theta(x) \geqq \left(\frac{1}{3}\right)^{\alpha} \frac{11}{6} > \theta\left(\frac{3}{2}\right)$

5) If $3 \leqq x < \frac{7}{2}$, $\phi(x) \geqq 2$ thus $\theta(x) \geqq 2\left(\frac{2}{7}\right)^{\alpha} > \theta\left(\frac{3}{2}\right)$

6) If $\frac{7}{2} \leqq x < 4$, $\phi(x) \geqq \frac{5}{2}$ thus $\theta(x) \geqq \frac{5}{6} > \theta\left(\frac{3}{2}\right)$.

5.2 In the Interval $\left]\frac{3}{2}, \frac{7}{4}\right[$

1) If $\frac{3}{2} + \frac{1}{4^r} \leqq x < \frac{3}{2} + \frac{2}{4^r}$, $r \geqq 2$, we get from the table:

$$\phi(x) \geqq 1 + \frac{1}{3^r} \text{ thus } \theta(x) \geqq \left(1 + \frac{1}{3^r}\right)\left(\frac{3}{2} + \frac{2}{4^r}\right)^{-\alpha} > \theta\left(\frac{3}{2}\right).$$

2) If $\frac{3}{2} + \frac{2}{4^r} \leqq x < \frac{3}{2} + \frac{3}{4^r}$, $r \geqq 2$, we get

$$\theta(x) \geqq \left(1 + \frac{1}{2 \cdot 3^{r-1}}\right)\left(\frac{3}{2} + \frac{3}{4^r}\right)^{-\alpha} > \theta\left(\frac{3}{2}\right).$$

3) If $\frac{3}{2} + \frac{3}{4^r} \leqq x < \frac{3}{2} + \frac{1}{4^{r-1}}$, $r \geqq 2$, we get

$$\theta(x) \geqq \left(1 + \frac{2}{3^r}\right)\left(\frac{3}{2} + \frac{1}{4^{r-1}}\right)^{-\alpha} > \theta\left(\frac{3}{2}\right).$$

Finally, $\theta(x) > \theta\left(\frac{3}{2}\right)$ for all $x \in \left]\frac{3}{2}, \frac{7}{4}\right[$ since

$$\left]\frac{3}{2}, \frac{7}{4}\right[= \bigcup_{r \geqq 2} \left[\frac{3}{2} + \frac{1}{4^r}, \frac{3}{2} + \frac{1}{4^{r-1}}\right[.$$

6. Upper Bound for θ and F

We prove that $\operatorname{Sup} \theta = \theta\left(\frac{65}{64}\right) = \frac{55}{2(65)^{\alpha}}$ so that $\operatorname{Sup} F = \frac{55}{3}\left(\frac{3}{65}\right)^{\alpha}$ is the limit of $S(N)N^{-\alpha}$ when N runs through the sequence $k \mapsto \frac{1}{3}(65 \cdot 4^k + 1)$.

6.1 $\theta(x)<1$ *in the Interval* $\left[\frac{5}{4}, \frac{15}{4}\right[$

Making use of subintervals of length $\frac{1}{4}$ or $\frac{1}{16}$, this inequality is easily verified with the aid of the table.

6.2 $\theta(x)<1$ *in the Interval* $\left[\frac{15}{4}, 4\right[$

Indeed, $\left[\frac{15}{4}, 4\right[= \bigcup_{r\geqq 1} \left[4-\frac{1}{4^r}, 4-\frac{1}{4^{r+1}}\right[$ and

1) if $4-\frac{1}{4^r}\leqq x<4-\frac{1}{4^r}+\frac{1}{4^{r+1}}$, $\phi(x)\leqq 3-\frac{2}{3^{r+1}}$ hence

$$\theta(x)\leqq\left(1-\frac{2}{3^{r+2}}\right)\left(1-\frac{1}{4^{r+1}}\right)^{-\alpha}<1 \quad \text{for } r\geqq 1,$$

2) if $4-\frac{1}{4^r}+\frac{1}{4^{r+1}}\leqq x<4-\frac{1}{4^r}+\frac{2}{4^{r+1}}$, $\phi(x)\leqq 3-\frac{1}{2\cdot 3^r}$ hence

$$\theta(x)\leqq\left(1-\frac{1}{2\cdot 3^{r+1}}\right)\left(1-\frac{3}{4^{r+2}}\right)^{-\alpha}<1 \quad \text{for } r\geqq 1,$$

3) if $4-\frac{1}{4^r}+\frac{2}{4^{r+1}}\leqq x<4-\frac{1}{4^{r+1}}$, $\phi(x)\leqq 3-\frac{1}{3^{r+1}}$ hence

$$\theta(x)\leqq\left(1-\frac{1}{3^{r+2}}\right)\left(1-\frac{2}{4^{r+2}}\right)^{-\alpha}<1 \quad \text{for } r\geqq 1.$$

6.3 In the Interval $\left]\frac{17}{16}, \frac{5}{4}\right[$

In this interval $\theta(x)<\theta\left(\frac{17}{16}\right)=\frac{19}{2(17)^{\alpha}}=1.006041522\ldots$ Indeed,

1) if $\frac{17}{16}\leqq x<\frac{18}{16}$, $\phi(x)\leqq\phi\left(\frac{17}{16}\right)$ hence $\theta(x)<\theta\left(\frac{17}{16}\right)$,

2) if $\frac{18}{16}\leqq x<\frac{5}{4}$, $\theta(x)<1$ as can be seen from the table, using the partition points $\frac{37}{32}, \frac{75}{64}, \frac{19}{16}, \frac{77}{64}, \frac{78}{64}, \frac{79}{64}$.

6.4 In the Interval $\left]\frac{65}{64}, \frac{17}{16}\right[$

In this interval, $\theta(x)<\theta\left(\frac{65}{64}\right)=\frac{55}{2(65)^{\alpha}}=1.006080775\ldots$ Indeed,

1) if $x\in\left]\frac{65}{64}, \frac{66}{64}\right[$, $\phi(x)\leqq\phi\left(\frac{65}{64}\right)$ thus $\theta(x)<\theta\left(\frac{65}{64}\right)$,

2) if $x\in\left]\frac{66}{64}, \frac{67}{64}\right[$, $\theta(x)<\theta\left(\frac{65}{64}\right)$, an inequality which is proved making use of the partition points $\frac{265}{256}, \frac{266}{256}, \frac{267}{256}$.

3) if $x\in\left]\frac{67}{64}, \frac{17}{16}\right[$, $\theta(x)<\theta\left(\frac{17}{16}\right)$ since for all $k\geqq 3$, we have: for all

$$x\in\left[\frac{17}{16}-\frac{1}{4^k}, \frac{17}{16}-\frac{1}{4^k}+\frac{1}{4^{k+1}}\right[, \quad \phi(x)\leqq\frac{19}{18}-\frac{1}{3^{k+1}} \tag{9}$$

hence

$$\theta(x)\leqq\theta\left(\frac{17}{16}\right)\left(1-\frac{18}{19\cdot 3^{k+1}}\right)\left(1-\frac{16}{17\cdot 4^k}\right)^{-\alpha}<\theta\left(\frac{17}{16}\right),$$

for all

$$x\in\left[\frac{17}{16}-\frac{1}{4^k}+\frac{1}{4^{k+1}}, \frac{17}{16}-\frac{1}{4^k}+\frac{2}{4^{k+1}}\right[, \quad \phi(x)\leqq\phi\left(\frac{17}{16}-\frac{1}{4^k}+\frac{1}{4^{k+1}}\right) \tag{10}$$

hence a fortiori

$$\theta(x)\leqq\theta\left(\frac{17}{16}-\frac{1}{4^k}+\frac{1}{4^{k+1}}\right)<\theta\left(\frac{17}{16}\right),$$

for all

$$x\in\left[\frac{17}{16}-\frac{1}{4^k}+\frac{2}{4^{k+1}}, \frac{17}{16}-\frac{1}{4^{k+1}}\right[, \quad \phi(x)\leqq\frac{19}{18}=\frac{1}{2\cdot 3^{k+1}} \tag{11}$$

therefore

$$\theta(x)\leqq\theta\left(\frac{17}{16}\right)\left(1-\frac{1}{19\cdot 3^{k-1}}\right)\left(1-\frac{2}{17\cdot 4^{k-1}}\right)^{-\alpha}<\theta\left(\frac{17}{16}\right).$$

6.5 In the Interval $\left[1, \frac{65}{64}\right[$

As in paragraph 6.4, we get

$$\theta(x)<\theta\left(\frac{65}{64}\right) \qquad \text{for } x\in\left[\frac{257}{256}, \frac{65}{64}\right[,$$

$$\theta(x)<\theta\left(\frac{257}{256}\right)<\theta\left(\frac{65}{64}\right) \qquad \text{for } x\in\left[\frac{1025}{1024}, \frac{257}{256}\right[.$$

Moreover,

$$\theta(x) \leqq \phi(x) \leqq \phi\left(\frac{1025}{1024}\right) = \frac{487}{486} < \theta\left(\frac{65}{64}\right) \quad \text{for } x \in \left[1, \frac{1025}{1024}\right[.$$

7. Remark

In [3], Harborth studies the behavior of

$$Q(N) = N^{-\beta} \sum_{n<N} 2^{s(n)} \quad \text{where } \beta = \frac{\operatorname{Log} 3}{\operatorname{Log} 2}.$$

He shows that $\limsup_{N\to\infty} Q(N) = 1$ and $\liminf_{N\to\infty} Q(N) = 0.812556\ldots$ (the exact value is not known).

As in paragraph 2, we can prove that $\limsup Q$ and $\liminf Q$ are respectively the upper and lower bounds of the function θ: $[1, 2] \to \mathbb{R}$ defined as follows: if

$$x = \sum_{a=0}^{\infty} e_a \cdot 2^{-a}, \quad e_0 = 1, \quad e_a \in \{0, 1\},$$

$$\theta(x) = x^{-\beta} \cdot \sum_{a=0}^{\infty} \delta(a) 3^{-a} \quad \text{where } \delta(a) = e_a \prod_{k<a} (1 + e_k).$$

References

1. Delange, H.: Sur la fonction sommatoire de la fonction "somme des chiffres". Enseignement Mathématique (2) **21**, 31-47 (1975)
2. Delange, H.: Sur les fonctions q-additives ou q-multiplicatives. Acta Arithmetica **21**, 285-298 (1972)
3. Harborth, H.: Number of odd binomial coefficients. Proceedings of A.M.S. **62**, 19-22 (1977)
4. Newman, D.J.: On the number of binary digits in a multiple of three. Proceedings of A.M.S. **21**, 719-721 (1969)
5. Stolarsky, K.B.: Power and exponential sums of digital sums... SIAM J. Appl. Math. **32**, 717-730 (1977)

Oblatum 30-XI-1982 & 10-I-1983

Invent. math. 73, 117-138 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

On the Representation of Integers by Binary Cubic Forms of Positive Discriminant

J.-H. Evertse

University of Leiden, Department of Mathematics, Wassenaarseweg 80, Postbus 9512, NL-2300 RA Leiden, The Netherlands

§ 1. Introduction

Let $F(x, y)$ be an irreducible binary cubic form with integral coefficients and negative discriminant. Delone [3] and Nagell [8] showed that the equation

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

has at most five solutions in integers x, y. This can not be improved for if $F(x, y)$ equals $x^3-x y^2+y^3$ then F has discriminant -23 and (1) is satisfied by $(1, 0)$, $(0, 1)$, $(-1, 1)$, $(1, 1)$, $(4, -3)$. Delone and Nagell proved their result by considering units in the algebraic number field $L=\mathbb{Q}(\varepsilon)$, where ε is the real root of $F(x, 1)=0$. In both the proofs of Delone and Nagell the fact that the group of units in the ring of integers of K is generated by one fundamental unit is essential.

Now suppose that $F(x, y)$ satisfies the same conditions as above except that its discriminant is positive. Let $L=\mathbb{Q}(\varepsilon)$ where ε is some root of $F(x, 1)=0$. We cannot apply the methods of Delone and Nagell since L has a system of two fundamental units. However, it is possible to reduce (1) to a set of exponential equations to which a local method of Skolem can be applied (cf. [12], [13], [14], [6], [1], [7] Ch. 23). In this way, (1) was solved for the forms $F(x, y)=x^3-3 x y^2+y^3$ of discriminant 81 and $F(x, y)=x^3+x^2 y-2 x y^2-y^3$ of discriminant 49 by respectively Ljunggren [6] and Baulin [2]. In the first case, the six solutions of (1) are $(1,0)$, $(0,1)$, $(-1,-1)$, $(1,3)$, $(-3,-2)$, $(2,-1)$ whereas in the second case (1) is only satisfied by the nine pairs $(1,0)$, $(0,-1)$, $(-1,1)$, $(-1,-1)$, $(2,-1)$, $(-1,2)$, $(5,4)$, $(4,-9)$, $(-9,5)$. Note that in the first case $L=\mathbb{Q}(e^{2\pi i/9}+e^{-2\pi i/9})$ and in the second case, $L=\mathbb{Q}(e^{2\pi i/7}+e^{-2\pi i/7})$. In the proofs of Ljunggren and Baulin, use is made of the explicit values of some fundamental units in a quadratic extension of L. So it does not seem easy to derive general results on (1) by their method.

It is possible to derive more general, though ineffective, results on (1) by means of an approximation method in which hypergeometric functions are

used. In this way, Siegel ([9], [11]) showed that the number of solutions of the inequality

$$|F(x,y)| \leqq k \tag{2}$$

in integers x, y with $(x,y)=1$, $y>0$ or $(x,y)=(1,0)$, where k is a positive integer, is at most 18 if the discriminant of F is sufficiently large compared with k. In a student paper from 1949, A.E. Gel'man showed, by refining Siegel's estimates, that 18 can be replaced by 10. (For a proof of this, we refer to [4], Chap. 5). In particular this implies that (1) has at most ten solutions if the discriminant of F is large enough. We shall give a uniform upper bound for the number of solutions of (1).

Theorem 1. *Let F be a binary cubic form with integral coefficients and non-zero discriminant. Then the equation*

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

has at most twelve *solutions in integers x, y.*

As far as I know, no cubic forms F are known for which (1) has more than nine solutions and it is likely that our result can be improved.

The method which we shall apply in the proof of Theorem 1 allows us also to prove a result on the number of solutions of (2). Before we can state it we have to introduce the notion of a reduced cubic form. For any cubic form $F(x,y)$ its *quadratic covariant* $H(x,y)$ is defined by

$$H(x,y)=-\frac{1}{4}\left(\frac{\partial^2 F}{\partial x^2}\cdot\frac{\partial^2 F}{\partial y^2}-\left(\frac{\partial^2 F}{\partial x\,\partial y}\right)^2\right)=Ax^2+Bxy+Cy^2$$

say, where A, B, C are certain polynomials in the coefficients of F. It is well-known that if F has discriminant D, then H has discriminant $B^2-4AC=-3D$ and that H is positive definite if F has positive discriminant. Furthermore, H satisfies the covariance property, i.e. if a_{11}, a_{12}, a_{21}, a_{22} are real (or complex) numbers with $a_{11}a_{22}-a_{12}a_{21}=1$, then $H(a_{11}x+a_{12}y, a_{21}x+a_{22}y)$ is the quadratic covariant of $F(a_{11}x+a_{12}y, a_{21}x+a_{22}y)$.

Definition. A binary cubic form of positive discriminant is called *reduced* if its quadratic covariant H is reduced, i.e. if $C \geqq A \geqq |B|$.

Since every positive definite quadratic form is equivalent to a reduced quadratic form, the covariance property yields that every cubic form of positive discriminant is equivalent to a reduced cubic form.

Theorem 2. *Let F be a reduced, irreducible, binary cubic form of positive discriminant with integral coefficients and let k be a positive integer. Then the inequality*

$$|F(x,y)| \leqq k \tag{2}$$

has at most nine solutions in integers x, y with $(x,y)=1$, $y \geqq 12^{1/4}k^{3/2}$.

In fact, both Theorem 1 and Theorem 2 are consequences of

Theorem 3. *Let F be an irreducible, binary cubic form with integral coefficients, discriminant D with $D>0$ and quadratic covariant H. Let k be a positive integer.*

Then the number of solutions of the inequality.

$$|F(x,y)| \leqq k \tag{2}$$

in integers x, y *with*

$$H(x,y) \geqq \tfrac{3}{2}(3D)^{1/2} k^3, \quad (x,y)=1, \quad y>0 \quad \text{or} \quad (x,y)=(1,0) \tag{3}$$

is at most 9.

In the proof of Theorem 3 we have used ideas from [4], Chap. 5, [9] and [10].

§ 2. Proofs of Theorem 1 and Theorem 2

In this section we shall derive Theorem 1 and Theorem 2 from Theorem 3. In the sequel, let F be a binary cubic form with integral coefficients, discriminant D and quadratic covariant H. Theorem 2 is an immediate consequence of Theorem 3 and the lemma below.

Lemma 1. *If* $D>0$ *and if* F *is reduced and irreducible then*

$$H(x,y) \geqq \tfrac{3}{4} D^{1/2} y^2 \quad \text{for } x, y \in \mathbb{Z}, \tag{4}$$

$$H(x,y) \geqq \tfrac{3}{2}(3D)^{1/2} y^2 \quad \text{for } x, y \in \mathbb{Z} \text{ with } |x| \geqq |2y|. \tag{5}$$

Proof. We may assume that $y \neq 0$. Put $H(x,y) = Ax^2 + Bxy + Cy^2 = y^2 f(z)$, where $z = x/y$ and $f(z) = Az^2 + Bz + C$. Note that for $z = -B/2A$, $f(z)$ assumes a minimum on the reals which is equal to $3D/4A$. Furthermore F, whence H, is reduced, hence

$$A^2 \leqq AC \leqq \tfrac{1}{3}(4AC - B^2) = D.$$

This implies that $f(z) \geqq 3D/4D^{1/2} = 3D^{1/2}/4$ which is equivalent to (4).

We are now going to prove (5). Since $f(z)$ assumes its minimum in a point with absolute value not exceeding 1/2, it follows that for $|z| \geqq 2$,

$$\begin{aligned} f(z) &\geqq \min(f(2), f(-2)) = 4A - 2|B| + C \\ &= 4A + C - 2(4AC - 3D)^{1/2} =: g(A, C). \end{aligned}$$

We shall minimalize $g(A, C)$ on the (A, C)-plane. Note that (A, C) belongs to the area

$$\{(A, C) \in \mathbb{R}^2 : 1 \leqq A \leqq C,\ A^2 \geqq 4AC - 3D \geqq 0\},$$

since H is reduced. On this area, $g(A, C)$ is minimal in the point $((3D)^{1/2}/3, 5(3D)^{1/2}/6)$. This minimum equals $3(3D)^{1/2}/2$, which proves our lemma completely (cf. Fig. 1). □

We shall now prove Theorem 1. Suppose $D \neq 0$. Firstly we assume that F is reducible. Then there are integers a_1, a_2, b_1, b_2, b_3 such that

$$F(x,y) = (a_1 x + a_2 y)(b_1 x^2 + b_2 xy + b_3 y^2).$$

Hence solutions of (1) must satisfy

$$a_1 x + a_2 y = \pm 1, \quad b_1 x^2 + b_2 xy + b_3 y^2 = \pm 1.$$

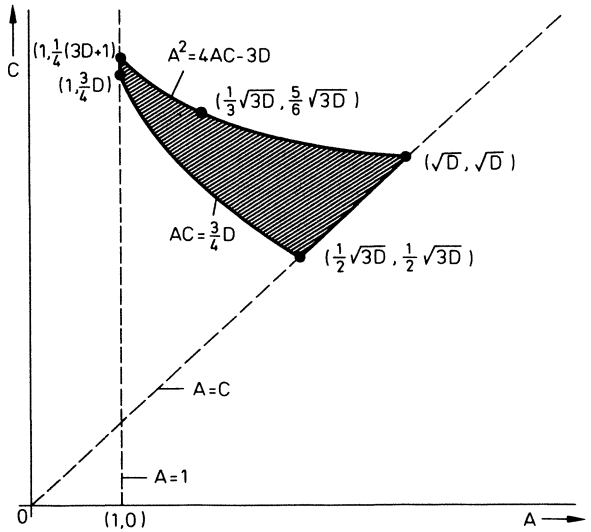


Fig. 1

Therefore,

$$a_2^2 b_1 x^2 + a_2 b_2 x(a_1 x \mp 1) + b_3(a_1 x \mp 1)^2 = \pm a_2^2.$$

Since $D \neq 0$, the latter equation has at most two solutions in x for each choice of the sign of $a_1 x + a_2 y$. Hence (1) has at most four solutions.

We shall now assume that F is irreducible. We shall also assume that $D > 0$, that F is reduced and that its coefficients of x^3 is positive. By the result of Delone and Nagell and by the fact that the number of solutions of (1) does not change when F is replaced by an equivalent form, these are no restrictions.

It follows easily from Theorem 3, that (1) has at most nine solutions in integers x, y for which $H(x,y) \geqq 3(3D)^{1/2}/2$. Note that by Lemma 1, $H(x,y) \geqq 3(3D)^{1/2}/2$ if $|xy| \geqq 2$. Hence it suffices to show that (1) has at most three solutions with $|xy| \leqq 1$.

First of all, there can be at most three solutions with $|y| = 1$, $|x| \leqq 1$. For such solutions belong to the set of six pairs (x,y) with $x \in \{-1,0,1\}$, $y \in \{-1,1\}$ and for each solution (x,y) belonging to this set, the pair $(-x,-y)$, which is clearly no solution of (1), also belongs to it. Hence we may restrict ourselves to the case that $(1,0)$ is a solution of (1).

We may assume without restriction that the number of solutions with $y=1$, $|x| \leqq 1$ is not less than the number of solutions with $y=-1$, $|x| \leqq 1$ and moreover, that there are at least two solutions with $y=1$, $|x| \leqq 1$. If there are three such solutions, we have

$$F(x,y) = x(x+y)(x-y) + y^3 = x^3 - xy^2 + y^3,$$

hence F has discriminant $-23 < 0$. If we have two such solutions, then

$$F(x,y) = (x - py)^2(x - qy) + y^3$$

for some $p, q \in \{-1, 0, 1\}$, since by our assumption that $F(1,0)=1$, $F(x,1)=1$ can not have any non-integral, rational solution. If three are also solutions with $y=-1$, then

$$(x+p)^2(x+q)=2,$$

hence $x+q=2$, $x+p=1$, $q=p+1$. But then F is equivalent to the form $x^2(x-y)+y^3=x^3-x^2y+y^3$ of negative discriminant -23. This implies that (1) has at most two solutions with $|y|=1$, $|x| \leqq 1$ if $F(1,0)=1$ which completes the proof of Theorem 1. □

Note that there are infinitely many reduced forms F for which (1) has three solutions with $H(x,y)<3(3D)^{1/2}/2$. For take

$$F(x,y)=x^3+ax^2y-(a+3)xy^2+y^3,$$

where a is some fixed integer. Then F has quadratic covariant $(a^2+3a+9)(x^2-xy+y^2)$ and discriminant $(a^2+3a+9)^2$. It is easy to check that F is irreducible and that $(1,0)$, $(0,1)$, $(-1,-1)$ are solutions of (1). For these solutions one has $H(x,y)=D^{1/2}$.

Note that (1) has at most one solution with $H(x,y)<\frac{1}{2}(3D)^{1/2}$ if F is irreducible and of positive discriminant. For by the covariance property of H we may assume that F is reduced. Hence, by Lemma 1, $H(x,y) \geqq 3(3D)^{1/2}/2$ if $|xy| \geqq 2$. Furthermore, $H(\pm 1,0)=A$, $H(0,\pm 1)=C$, $H(\pm 1;\pm 1) \geqq A-|B|+C$ and

$$A-|B|+C \geqq C \geqq (AC)^{1/2}=\tfrac{1}{2}(3D+B^2)^{1/2} \geqq \tfrac{1}{2}(3D)^{1/2}.$$

If D exceeds some absolute constant, D_0 say, one can prove, similar to Theorem 3, that Theorem 3 with $k=1$ holds even if the lower bound $(3/2)(3D)^{1/2}$ in (3) is replaced by $\frac{1}{2}(3D)^{1/2}$. Hence for $D \geqq D_0$, (1) has at most ten solutions. In fact, one can show that D_0 can be taken as 5×10^{10} but we shall not work this out here.

§ 3. Preliminaries to the Proof of Theorem 3

In the sequel we shall assume that $F(x,y)$ is an irreducible, binary cubic form of positive discriminant D, $F(x,y)=ax^3+bx^2y+cxy^2+dy^3$ say, where a, b, c, are rational integers. The quadratic covariant and the cubic covariant of F are defined by

$$H(x,y)=-\frac{1}{4}\left(\frac{\partial^2 F}{\partial x^2}\frac{\partial^2 F}{\partial y^2}-\left(\frac{\partial^2 F}{\partial x \partial y}\right)^2\right), \qquad G(x,y)=\frac{\partial F}{\partial x}\frac{\partial H}{\partial y}-\frac{\partial F}{\partial y}\frac{\partial H}{\partial x}$$

respectively. Here

$$\begin{aligned} H(x,y)&=Ax^2+Bxy+Cy^2, \\ G(x,y)&=a'x^3+3b'x^2y+3c'xy^2+d'y^3, \end{aligned} \tag{6}$$

where A, B, C, a', b', c', d' are rational integers given by

$$A=b^2-3ac, \quad B=bc-9ad, \quad C=c^2-3bd,$$
$$a'=9abc-2b^3-27a^2d, \quad b'=6ac^2-b^2c-9abd,$$
$$c'=9acd+bc^2-6b^2d, \quad d'=27ad^2+2c^3-9bcd.$$

We have the following identity: (cf. [7], Chap. 24)

$$4H(x,y)^3=G(x,y)^2+27DF(x,y)^2. \tag{7}$$

Let $\sqrt{-3D}$ be a fixed choice of the square-root of $-3D$. Let $M=\mathbb{Q}(\sqrt{-3D})$ and let $\mathcal{O}$ be the ring of integers in M. We define a subset $\mathcal{O}_0$ of $\mathcal{O}$ by

$$\mathcal{O}_0=\left\{\frac{m+n\sqrt{-3D}}{2}\in\mathcal{O}: m,n\in\mathbb{Z}\right\}.$$

It is easy to check that $\mathcal{O}_0$ is a subring of $\mathcal{O}$, on noting that $D\equiv B^2 \pmod 4$, i.e. $D\equiv 0,1 \pmod 4$ and that $(m+n\sqrt{-3D})/2\in\mathcal{O}$ if and only if $m\equiv nD \pmod 2$. For any complex number λ we denote its complex conjugate by $\bar{\lambda}$. We have clearly

$$\lambda\in\mathcal{O}_0 \text{ if and only if } \lambda\in\mathcal{O},\ \lambda-\bar{\lambda}\in\mathbb{Z}\cdot\sqrt{-3D}. \tag{8}$$

This implies that

$$|\lambda|\geqq|\operatorname{Im}\lambda|\geqq\tfrac{1}{2}\sqrt{3D} \quad \text{if } \lambda\in\mathcal{O}_0-\mathbb{Z}. \tag{9}$$

Put

$$U(x,y)=\frac{G(x,y)+3\sqrt{-3D}\,F(x,y)}{2},$$
$$V(x,y)=\frac{G(x,y)-3\sqrt{-3D}\,F(x,y)}{2}. \tag{10}$$

Then $U(x,y)$, $V(x,y)$ are cubic forms with coefficients belonging to M such that corresponding coefficients of U and V are complex conjugates. Since F must be also irreducible over M, $U(x,y)$ and $V(x,y)$ do not have factors in common. Moreover, by (7), $U(x,y)V(x,y)=H(x,y)^3$, hence $U(x,y)$, $V(x,y)$ must be cubes of linear forms, $\xi(x,y)$, $\eta(x,y)$ respectively say, of which the corresponding coefficients are complex conjugates. On noting that $\xi(x,y)\eta(x,y)$ must be a quadratic form which is a cube root of $H(x,y)^3$ and for which the coefficient of x^3 is a positive, real number, we have that

$$\xi(x,y)^3-\eta(x,y)^3=3\sqrt{-3D}\,F(x,y),$$
$$\xi(x,y)^3+\eta(x,y)^3=G(x,y), \tag{11}$$
$$\xi(x,y)\eta(x,y)=H(x,y).$$

These identities imply that for any pair of rational integers x_0, y_0, the numbers $\xi(x_0,y_0), \eta(x_0,y_0)$ are complex conjugate algebraic integers and, since H has

discriminant $-3D$,

$$\frac{\xi(x,y)}{\xi(1,0)}\in M[x,y], \quad \frac{\eta(x,y)}{\eta(1,0)}\in M[x,y]. \tag{12}$$

A pair of linear forms ξ, η with properties (11), (12) is called a *pair of resolvent forms*[1].

Lemma 2. (i) $\xi(x,y)^3\in\mathcal{O}_0[x,y]$, $\eta(x,y)^3\in\mathcal{O}_0[x,y]$.

(ii) *If* (x_1,y_1), (x_2,y_2) *are two pairs of rational integers then*

$$\xi(x_1,y_1)\eta(x_2,y_2)\in\mathcal{O}_0,$$
$$\xi(x_1,y_1)^2\xi(x_2,y_2)\in\mathcal{O}_0, \eta(x_1,y_1)^2\eta(x_2,y_2)\in\mathcal{O}_0.$$

Proof. Let $\xi(x,y)=\varepsilon_1 x+\varepsilon_2 y$. Clearly, $\varepsilon_1,\varepsilon_2$ are algebraic integers and so are $\varepsilon_1^3,\varepsilon_1^2\varepsilon_2,\varepsilon_1\varepsilon_2^2,\varepsilon_2^3$. Note that by (6) and (10),

$$\begin{aligned}\xi(x,y)^3&=\varepsilon_1^3x^3+3\varepsilon_1^2\varepsilon_2x^2y+3\varepsilon_1\varepsilon_2^2xy^2+\varepsilon_2^3y^3=U(x,y)\\&=\frac{a'+3a\sqrt{-3D}}{2}x^3+3\frac{b'+b\sqrt{-3D}}{2}x^2y+3\frac{c'+c\sqrt{-3D}}{2}xy^2\\&\quad+\frac{d'+3d\sqrt{-3D}}{2}y^3.\end{aligned}$$

Hence ε_1^3, $\varepsilon_1^2\varepsilon_2$, $\varepsilon_1\varepsilon_2^2$, ε_2^3 belong to $\mathcal{O}_0$. This clearly implies that $\xi(x,y)^3\in\mathcal{O}_0[x,y]$ and since the coefficients of $\eta(x,y)^3$ are complex conjugate to the corresponding coefficients of $\xi(x,y)^3$ that $\eta(x,y)^3\in\mathcal{O}_0[x,y]$. Furthermore this implies that $\xi(x_1,y_1)^2\xi(x_2,y_2)$ belongs to $\mathcal{O}_0$ since this number can be written as a linear combination with rational integral coefficients in $\varepsilon_1^3,\varepsilon_1^2\varepsilon_2,\varepsilon_1\varepsilon_2^2,\varepsilon_2^3$. But then, $\eta(x_1,y_1)^2\eta(x_2,y_2)=\bar{\xi}(x_1,y_1)^2\bar{\xi}(x_2,y_2)$ also belongs to $\mathcal{O}_0$. Finally we note that the linear transformation $(x,y)\to(\xi,\eta)$ has determinant $\pm\sqrt{-3D}$, since the product of ξ and η is a quadratic form of discriminant $-3D$. Hence

$$\begin{aligned}&\xi(x_1,y_1)\eta(x_2,y_2)-\bar{\xi}(x_1,y_1)\bar{\eta}(x_2,y_2)\\&\quad=\xi(x_1,y_1)\eta(x_2,y_2)-\xi(x_2,y_2)\eta(x_1,y_1)=\pm\sqrt{-3D}(x_1y_2-x_2y_1).\end{aligned} \tag{13}$$

Therefore, by the fact that $\xi(x_1,y_1)\eta(x_2,y_2)$ is an algebraic integer, that $\xi(x_1,y_1)/\xi(x_2,y_2)$ belongs to M by (12), that $\xi(x_2,y_2)\eta(x_2,y_2)$ is a rational integer and by (8), $\xi(x_1,y_1)\eta(x_2,y_2)$ belongs to $\mathcal{O}_0$. □

Note that there are exactly three pairs of resolvent forms, for if (ξ,η) is one pair, then $(\rho\xi,\rho^2\eta)$ and $(\rho^2\xi,\rho\eta)$ are the other two pairs, where ρ is a primitive third root of unity. Let (ξ_1,η_1), (ξ_2,η_2), (ξ_3,η_3) be the three pairs of re-

[1] One can show that for $x_0,y_0\in\mathbb{Z}$, $\xi(x_0,y_0)$, $\eta(x_0,y_0)$ are Lagrange resolvents, i.e. they can be written as $\omega+\rho\omega'+\rho^2\omega''$, where ω,ω',ω'' are conjugate cubic numbers and where ρ is a third root of unity (cf. [4], p. 167). We shall not use this fact

solvent forms. We say that a pair of rational integers (x, y) is *related* to a pair of resolvent forms (ξ, η) if

$$\left|1-\frac{\eta(x,y)}{\xi(x,y)}\right| = \min_{1\leqq i\leqq 3}\left|1-\frac{\eta_i(x,y)}{\xi_i(x,y)}\right|.$$

Our aim is to show that there are at most three solutions of (2) satisfying (3) which are related to a given pair of resolvent forms (ξ, η). This will clearly imply Theorem 3.

§ 4. Construction of the Auxiliary Polynomials

We shall construct some auxiliary polynomials which will be used in the proof of Theorem 3. We shall use results from the theory of hypergeometric functions (cf. [5]). Our method will be similar to that of Siegel in [10].

A hypergeometric function is a power series of the following type

$$F(\alpha,\beta,\gamma,z)=1+\frac{\alpha\cdot\beta}{1\cdot\gamma}z+\frac{\alpha(\alpha+1)\beta(\beta+1)}{1\cdot 2\cdot\gamma(\gamma+1)}z^2 + \frac{\alpha(\alpha+1)(\alpha+2)\beta(\beta+1)(\beta+2)}{1\cdot 2\cdot 3\cdot\gamma(\gamma+1)(\gamma+2)}z^3+\ldots. \tag{14}$$

Here z is a complex variable and α, β, γ are complex constants. If α or β is a non-positive integer and m is the smallest integer such that $\alpha(\alpha+1)\ldots(\alpha+m)\beta(\beta+1)\ldots(\beta+m)=0$ we agree that (14) truncates after the term with z^m. Furthermore, if γ is a non-positive integer, we assume that at least one of α, β is also a non-positive integer, not smaller than γ.

It is easy to check that (14) converges for $|z|<1$. If α, β, γ are reals with $\gamma>\alpha+\beta$ and $\gamma, \gamma-\alpha, \gamma-\beta$ no non-positive integers, we have by a result of Gauss that the series converges for $z=1$ and that

$$F(\alpha,\beta,\gamma,1)=\frac{\Gamma(\gamma)\Gamma(\gamma-\alpha-\beta)}{\Gamma(\gamma-\alpha)\Gamma(\gamma-\beta)} \tag{15}$$

(where Γ denotes the Γ-function, as usual). (14) is a solution of the linear differential equation

$$z(z-1)y''+((\alpha+\beta+1)z-\gamma)y'+\alpha\beta y=0 \tag{16}$$

in complex functions $y=y(z)$ which are analytic around $z=0$, for by an easy computation we have that a power-series $y(z)=\sum_{m=0}^{\infty} a_m z^m$ is a solution of (16) if and only if

$$(n+1)(\gamma+n)a_{n+1}=(\alpha+n)(\beta+n)a_n \quad \text{for } n=0,1,2,\ldots. \tag{17}$$

If γ is a positive integer, then $a_1, a_2, \ldots$ are completely determined by a_0, hence the solutions of (16) which are holomorphic around $z=0$ form a one-dimensional complex vector space. If $\gamma=-m$ for some non-negative integer m, then a_0 and a_{m+1} may vary independently of each other. Furthermore,

$a_1, \ldots, a_m, a_{m+2}, a_{m+3}, \ldots$ are uniquely determined by a_0 and a_{m+1}. Hence the solution space of (16) is two-dimensional.

For every polynomial we define a binary form by attaching an asterisk. The binary form corresponding to $f(z)=a_n z^n+a_{n-1} z^{n-1}+\ldots+a_1 z+a_0$ is given by

$$f^*(x, y)=a_0 x^n+a_1 x^{n-1} y+\ldots+a_{n-1} x y^{n-1}+a_n y^n (=x^n f(y/x) \text{ for } x \neq 0).$$

Lemma 3. *Let r, g be integers with $r \geqq 1$, $g \in \{0, 1\}$. Put*

$$\begin{aligned} A_{r,g}(z) &= \sum_{m=0}^{r} \binom{r-g+1/3}{m} \binom{2r-g-m}{r-g} (-z)^m, \\ B_{r,g}(z) &= \sum_{m=0}^{r-g} \binom{r-1/3}{m} \binom{2r-g-m}{r} (-z)^m. \end{aligned} \tag{18}$$

(i) *There exists a power series $F_{r,g}(z)$ such that for all complex numbers z with $|z|<1$*

$$A_{r,g}(z)-(1-z)^{1/3} B_{r,g}(z)=z^{2r+1-g} F_{r,g}(z) \tag{19}$$

(where $(1-z)^{1/3}=\sum_{m=0}^{\infty} \binom{1/3}{m} (-z)^m$) and

$$|F_{r,g}(z)| \leqq \frac{\binom{r-g+1/3}{r+1-g} \binom{r-1/3}{r}}{\binom{2r+1-g}{r}} (1-|z|)^{-\frac{1}{2}(2r+1-g)}. \tag{20}$$

(ii) *For all complex numbers z with $|1-z| \leqq 1$ we have*

$$|A_{r,g}(z)| \leqq \binom{2r-g}{r}. \tag{21}$$

(iii) *For all complex numbers z with $z \neq 0$ and for $h \in \{0, 1\}$ we have*

$$A_{r,0}(z) B_{r+h,1}(z) \neq A_{r+h,1}(z) B_{r,0}(z). \tag{22}$$

(iv) *For all complex numbers z we have*

$$A^*_{r,0}(z, z-\bar{z})=\overline{B^*_{r,0}(z, z-\bar{z})}. \tag{23}$$

(v) *For all $\lambda \in \mathcal{O}_0, f \in \mathbb{Z}$ we have*

$$A^*_{r,g}(\lambda, 3\sqrt{-3D} f) \in \mathcal{O}_0, \quad B^*_{r,g}(\lambda, 3\sqrt{-3D} f) \in \mathcal{O}_0. \tag{24}$$

Proof. Put

$$\begin{aligned} C_{r,g}(z) &= \sum_{m=0}^{r} \binom{r-1/3}{r-m} \binom{r-g+1/3}{m} z^m, \\ D_{r,g}(z) &= \sum_{m=0}^{r-g} \binom{r-1/3}{m} \binom{r-g+1/3}{r-g-m} z^m. \end{aligned}$$

A straightforward computation shows that

$$\begin{aligned} A_{r,g}(z) &= \binom{2r-g}{r} F(-\tfrac{1}{3}-r+g, -r, -2r+g, z), \\ B_{r,g}(z) &= \binom{2r-g}{r-g} F(\tfrac{1}{3}-r, -r+g, -2r+g, z), \\ C_{r,g}(z) &= \binom{r-1/3}{r} F(-\tfrac{1}{3}-r+g, -r, \tfrac{2}{3}, z), \\ D_{r,g}(z) &= \binom{r-g+1/3}{r-g} F(\tfrac{1}{3}-r, -r+g, \tfrac{4}{3}, z). \end{aligned} \tag{25}$$

These identities imply that

$$C_{r,g}(z) = A_{r,g}(1-z), \qquad D_{r,g}(z) = B_{r,g}(1-z). \tag{26}$$

We shall derive only the first of these identities; the other follows in a similar way. From (25) one deduces easily that both $A_{r,g}(1-z)$ and $C_{r,g}(z)$ satisfy (16) with $\alpha = -1/3-r+g, \beta = -r, \gamma = 2/3$. Since γ is not a non-positive integer, the remark made after formula (17) gives that $A_{r,g}(1-z)$ and $C_{r,g}(z)$ are linearly dependent. It follows from the identity in power series

$$(1+z)^{2r+g} = (1+z)^{r-1/3}(1+z)^{r-g+1/3},$$

on equating the coefficients of z^r, that

$$C_{r,g}(1) = \sum_{m=0}^{r} \binom{r-1/3}{r-m}\binom{r-g+1/3}{m} = \binom{2r-g}{r} = A_{r,g}(0).$$

Hence $C_{r,g}(z) = A_{r,g}(1-z)$. As a consequence of (26) and (25) we have

$$A_{r,g}(1) = \binom{r-1/3}{r}, \qquad B_{r,g}(1) = \binom{r-g+1/3}{r-g}. \tag{27}$$

It is not difficult to derive (21) and (23) from (26). Firstly, we note that $C_{r,g}(z)$ has positive coefficients. Hence for all complex numbers z with $|1-z| \leqq 1$,

$$|A_{r,g}(z)| = |C_{r,g}(1-z)| \leqq C_{r,g}(1) = A_{r,g}(0) = \binom{2r-g}{r}.$$

Secondly, we note that (23) holds for $z=0$ and that for $z \neq 0$, $D_{r,0}(z) = z^r C_{r,0}(z^{-1})$. Hence for $z \neq 0$,

$$\begin{aligned} A^*_{r,0}(z, z-\bar{z}) &= z^r A_{r,0}\left(1-\frac{\bar{z}}{z}\right) = z^r C_{r,0}\left(\frac{\bar{z}}{z}\right) \\ &= \bar{z}^r D_{r,0}\left(\frac{z}{\bar{z}}\right) = \bar{z}^r B_{r,0}\left(1-\frac{z}{\bar{z}}\right) \\ &= B^*_{r,0}(\bar{z}, \bar{z}-z) = \overline{B^*_{r,0}(z, z-\bar{z})}. \end{aligned}$$

We shall now prove (19). It is easy to check that for $|z|<1$ the functions $A_{r,g}(z)$, $(1-z)^{1/3}B_{r,g}(z)$ and $z^{2r+1-g}G_{r,g}(z)$, where

$$G_{r,g}(z)=F(r+1-g, r+2/3, 2r+2-g, z)$$

satisfy (16) with $\alpha=-1/3-r+g$, $\beta=-r$, $\gamma=-2r+g$. Hence these functions are linearly dependent, i.e. there are constants c_1, c_2, c_3, not all zero, such that

$$c_1 A_{r,g}(z)+c_2(1-z)^{1/3}B_{r,g}(z)+c_3 z^{2r+1-g}G_{r,g}(z)=0. \tag{28}$$

By inserting $z=0$ in (28) and using that $A_{r,g}(0)=B_{r,g}(0)$, we obtain that $c_1+c_2=0$. Clearly c_1 is non-zero. Hence we may assume that $c_1=1$, $c_2=-1$. By inserting $z=1$ in (28) we infer that $c_3=-A_{r,g}(1)/G_{r,g}(1)$. Thus we obtain (19) with

$$F_{r,g}(z)=A_{r,g}(1)\,G_{r,g}(1)^{-1}\,G_{r,g}(z).$$

We shall now estimate $|F_{r,g}(z)|$ from above. By (27) and (15) we have that

$$\begin{aligned}\frac{A_{r,g}(1)}{G_{r,g}(1)}&=\binom{r-1/3}{r}\frac{\Gamma(r+1)\Gamma(r+4/3-g)}{\Gamma(2r+2-g)\Gamma(1/3)}\\&=\binom{r-1/3}{r}\frac{(r+1/3-g)(r-2/3-g)\ldots 1/3}{(2r+1-g)(2r-g)\ldots(r+1)}\\&=\frac{\binom{r-1/3}{r}\binom{r-g+1/3}{r+1-g}}{\binom{2r+1-g}{r}}.\end{aligned}$$

We shall now show that $G_{r,g}(z)=\sum_{m=0}^{\infty} g_m z^m$, where

$$0\leqq g_m\leqq\binom{-r-1/2+g/2}{m}(-1)^m \quad \text{for } m=0,1,2,\ldots. \tag{29}$$

This suffices to prove (20). For then

$$|G_{r,g}(z)|\leqq\sum_{m=0}^{r}\binom{-r-1/2+g/2}{m}(-|z|)^m=(1-|z|)^{-\frac{1}{2}(2r+1-g)}.$$

Formula (29) follows by induction on m. For $g_0=\binom{-r-1/2+g/2}{0}(-1)^0=1$ and for $m\geqq 0$

$$\begin{aligned}\frac{g_{m+1}}{g_m}&=\frac{(r+1-g+m)(r+2/3+m)}{(m+1)(2r+2-g+m)}\leqq\frac{(r+1/2-g/2+m)}{m+1}\\&=\frac{(-1)^{m+1}\binom{-r-1/2+g/2}{m+1}}{(-1)^m\binom{-r-1/2+g/2}{m}}.\end{aligned}$$

We shall deduce (22) from (19) and (27). Note that by (19) with $g=0$ and by (19) with $r+h$ instead of r and $g=1$,

$$A_{r,0}(z)B_{r+h,1}(z)-A_{r+h,1}(z)B_{r,0}(z)=z^{2r+h}P_{r,h}(z) \tag{30}$$

for some power-series $P_{r,h}(z)$. However, the left-hand side of (30) is a polynomial of degree at most $2r+h$. Hence $P_{r,h}(z)$ is a constant. By inserting $z=1$ in (30) and using (27), it follows that this constant is non-zero. Hence the left-hand side of (30) equals 0 if and only if $z=0$ which proves (22).

Finally we shall prove (24). It clearly suffices to prove that $\binom{a/3}{n}(3\sqrt{-3D})^n \in \mathcal{O}_0$ for each pair of integers a, n with $n \geqq 0$. This is obvious for $n=0$ and for $n \geqq 1$ it suffices to show that $\binom{a/3}{n}3^{(3n-1)/2}$ is an algebraic integer. For this implies that for even n with $n \geqq 2$

$$\binom{a/3}{n}(3\sqrt{-3D})^n=\binom{a/3}{n}(-3^3D)^{n/2}=\binom{a/3}{n}3^{(3n-1)/2}\sqrt{3}(-D)^{n/2}$$

is a rational integer, whence belongs to $\mathcal{O}_0$ and that for odd n with $n \geqq 1$

$$\binom{a/3}{n}(3\sqrt{-3D})^n=\binom{a/3}{n}3^{(3n-1)/2}(-D)^{(n-1)/2}\sqrt{-3D}$$

also belongs to $\mathcal{O}_0$. We shall now show that $u(n):=\binom{a/3}{n}3^{3n-1}$ is a p-adic integer for each prime p and for each integer n with $n \geqq 1$. First of all, if $p \neq 3$, then $a/3$, whence $u(n)$ is a p-adic integer. Moreover, since

$$u(n)=\left(\frac{a}{3}\left(\frac{a}{3}-1\right)\ldots\left(\frac{a}{3}-n+1\right)\Big/n!\right)^2 3^{3n-1}=(a(a-3)(a-6)\ldots(a-3(n-1))^2\frac{3^{n-1}}{(n!)^2}$$

and the number of times that 3 divides $(n!)^2$ is equal to

$$2\sum_{j=1}^{\infty}\left[\frac{n}{3^j}\right]<2\sum_{j=1}^{\infty}\frac{n}{3^j}=n,$$

the number $u(n)$ is a 3-adic integer. This completes the proof of Lemma 3. □

§ 5. Proof of Theorem 3

Let $F(x,y)$ be an irreducible, binary cubic form with rational integral coefficients and positive discriminant. Let (ξ,η) be a pair of resolvent forms as constructed in § 3. We have to show that at most three solutions of (2) satisfying (3) are related to (ξ,η) i.e. that there are at most three pairs of rational

integers (x, y) related to (ξ, η) with

$$|\xi(x,y)^3 - \eta(x,y)^3| \leqq 3k\Delta^{1/2}, \quad |\xi(x,y)| \geqq \tfrac{1}{2}\sqrt{6}\,k^{3/2}\Delta^{1/4}, \tag{31}$$
$$(x,y)=1,\ y>0 \quad \text{or} \quad (x,y)=(1,0),$$

where $\Delta = 3D$. It is known that 49, 81, 148 are the three smallest positive integers which are discriminant of an irreducible cubic form and that for these values of D only one equivalence class of cubic forms exists.[2] Hence by the results of Ljunggren and Baulin we may assume that $D \geqq 148$ if $k=1$ and $D \geqq 49$ if $k \geqq 2$, i.e. $\Delta \geqq 444$ if $k=1$ and $\Delta \geqq 147$ if $k \geqq 2$ and we shall do so in the sequel. For each solution of (31) we have, by the fact that $\Delta^{1/4} k^{7/2} \geqq 444^{1/4} > 2^{3/2} 3^{-1/2}$,

$$\left|1 - \frac{\eta(x,y)^3}{\xi(x,y)^3}\right| \leqq \frac{3k\Delta^{1/2}}{|\xi(x,y)|^3} \leqq \frac{2^{3/2}}{3^{1/2}\Delta^{1/4}k^{7/2}} < 1. \tag{32}$$

This fact will be used in the proof of the lemma below.

Lemma 4. (i) *If* (x, y) *is a solution of* (31) *related to* (ξ, η) *then*

$$\left|1 - \frac{\eta(x,y)}{\xi(x,y)}\right| < \frac{\pi}{3}\frac{k\Delta^{1/2}}{|\xi(x,y)|^3}. \tag{33}$$

(ii) *If* $(x', y'), (x'', y'')$ *are distinct solutions of* (31) *related to* (ξ, η) *and if*

$$|\xi(x'', y'')| \geqq |\xi(x', y')|,$$

then

$$|\xi(x'', y'')| > \frac{0.948}{k}|\xi(x', y')|^2. \tag{34}$$

Proof. (i) I shall use an argument which was suggested to me by F. Beukers. By (32) and by the fact that $|\eta(x,y)^3/\xi(x,y)^3| = 1$ we have $|\arg(\eta(x,y)^3/\xi(x,y)^3)| < \pi/3$. Since (x, y) is related to (ξ, η), this implies by (23) that

$$\begin{aligned}\left|1 - \frac{\eta(x,y)}{\xi(x,y)}\right| &\leqq \left|\arg\left(\frac{\eta(x,y)}{\xi(x,y)}\right)\right| \\ &= \tfrac{1}{3}\left|\arg\left(\frac{\eta(x,y)^3}{\xi(x,y)^3}\right)\right| < \frac{1}{3}\cdot\frac{\pi}{3}\left|1 - \frac{\eta(x,y)^3}{\xi(x,y)^3}\right| \\ &\leqq \frac{1}{3}\cdot\frac{\pi}{3}\frac{3k\Delta^{1/2}}{|\xi(x,y)|^3} = \frac{\pi}{3}\frac{k\Delta^{1/2}}{|\xi(x,y)|^3}.\end{aligned}$$

---

[2] In [4] on p. 159 a table is given of rings of discriminant D with $0 < D \leqq 1296$ with a unit element which are contained in rings of integers of cubic fields. This table can also be considered as a table of equivalence classes of cubic forms of positive discriminant for Delone and Faddeev showed in Chap. 2 of [4] that there exists a bijective mapping of equivalence classes of irreducible cubic forms onto the rings described above which preserves the discriminant. This table shows that the class of discriminant 49 is represented by $x^3 - x^2y - 2xy^2 + y^3$, the class of discriminant 81 by $x^3 - 3xy^2 - y^3$, that of 148 by $x^3 - x^2y - 3xy^2 + y^3$ and that there are no other positive discriminants below 149

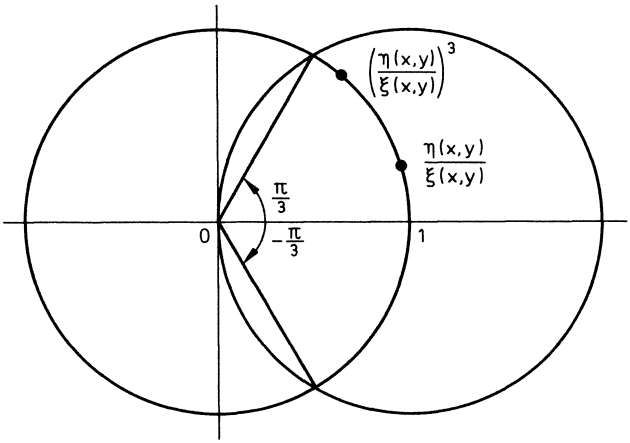


Fig. 2

(ii) Put $\xi'=\xi(x', y')$, $\eta'=\eta(x', y')$, $\xi''=\xi(x'', y'')$, $\eta''=\eta(x'', y'')$. We have by Lemma 2 that $\xi'\eta''\in\mathcal{O}_0$. Hence, by (13) and (33),

$$\Delta^{1/2}\leqq|\xi'\eta''-\xi''\eta'|=|\xi'|\cdot|\xi''|\left|\frac{\eta'}{\xi'}-\frac{\eta''}{\xi''}\right|$$

$$\leqq|\xi'|\cdot|\xi''|\left(\left|1-\frac{\eta'}{\xi'}\right|+\left|1-\frac{\eta''}{\xi''}\right|\right)<|\xi'|\cdot|\xi''|\cdot\frac{\pi}{3}k\Delta^{1/2}(|\xi'|^{-3}+|\xi''|^{-3}).$$

This implies that

$$|\xi'|^3+|\xi''|^3>\frac{3}{\pi k}(|\xi'|\cdot|\xi''|)^2.$$

Put $p=k|\xi''|\cdot|\xi'|^{-2}$, $h(z)=z^3-3z^2/\pi+(k/|\xi'|)^3$. Then $h(p)>0$. Note that $h(z)$ assumes a local maximum on the reals in $z=0$ and a local minimum in $z=2/\pi$. Clearly, $h(0)>0$. By (31) and the fact that $k^{1/2}\Delta^{1/4}\geqq 444^{1/4}$, we have

$$|\xi'|\geqq 999^{1/4}k. \tag{35}$$

Hence

$$h\left(\frac{2}{\pi}\right)=\left(\frac{2}{\pi}\right)^3-\frac{3}{\pi}\left(\frac{2}{\pi}\right)^2+\left(\frac{k}{|\xi'|}\right)^2<0.$$

Therefore, $h(z)$ has two positive zeros, p_1, p_2 say, with $p_1<p_2$, such that $h(z)$ is positive for $0<z<p_1$ and $z>p_2$ and negative for $p_1<z<p_2$. It is impossible that $p\leqq p_1$. For since $|\xi''|\geqq|\xi'|$ we have $p\geqq k/|\xi'|$ and by (35),

$$h\left(\frac{k}{|\xi'|}\right)=\frac{3}{\pi}\left(\frac{k}{|\xi'|}\right)^3\left(\frac{2\pi}{3}-\frac{|\xi'|}{k}\right)<0.$$

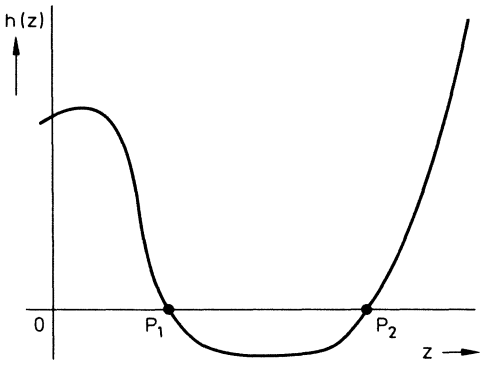


Fig. 3

Hence $p > p_2$. Furthermore we have $p_2 > 0.948$, since by (35),

$$h(0.948) = 0.948^3 - \frac{3}{\pi} \times 0.948^2 + \left(\frac{k}{|\xi'|}\right)^3 < 0$$

This proves Lemma 4 completely. □

We shall now assume that at least four solutions of (31) are related to (ξ, η), (x_{-1}, y_{-1}), (x_0, y_0), (x_1, y_1), (x_2, y_2) say, ordered such that $|\xi(x_{i+1}, y_{i+1})| \geqq |\xi(x_i, y_i)|$ for $i \in \{-1, 0, 1\}$ and we shall derive a contradiction from that. Put $\xi_i = \xi(x_i, y_i)$, $\eta_i = \eta(x_i, y_i)$ for $i \in \{1, 2\}$, $z_1 = 1 - (\eta_1/\xi_1)^3$. On applying (34) twice, we have by (31),

$$|\xi_1| \geqq \left(\frac{0.948}{k}\right)^3 |\xi(x_{-1}, y_{-1})|^4 \geqq \frac{0.948^3}{k^3} (\tfrac{1}{2}\sqrt{6}\, k^{3/2} \Delta^{1/4})^4 > 1.91 k^3 \Delta. \tag{36}$$

Furthermore, on applying (34) once again,

$$|\xi_2| \geqq \frac{0.948}{k} |\xi_1|^2. \tag{37}$$

Note that by (36), (32) and the fact that $k^8 \Delta^{5/2} \geqq 444^{5/2}$,

$$0 < |z_1| \leqq \frac{3k\Delta^{1/2}}{1.91^3 k^9 \Delta^3} = 3 \times 1.91^{-3} k^{-8} \Delta^{-5/2} \leqq 3 \times 1.91^{-3} \times 444^{-5/2} < 10^{-6}. \tag{38}$$

Note that for any complex number z with $|z| < 10^{-6}$ and for any primitive third root of unity ρ we have

$$|1 - (1-z)^{1/3}| < |1 - \rho(1-z)^{1/3}| \left(\text{where } (1-z)^{1/3} = \sum_{m=0}^{\infty} \binom{1/3}{m} (-z)^m\right).$$

Hence, by the fact that (x_1, y_1) is related to (ξ, η),

$$\frac{\eta_1}{\xi_1}=(1-z_1)^{1/3}. \tag{39}$$

We shall use (38) and (39) in the proof of the lemma below. r, g are integers with $r\geqq 1$, $g\in\{0,1\}$, and $A_{r,g}(z)$, $B_{r,g}(z)$ are the polynomials constructed in Lemma 3.

Lemma 5. *Put*

$$\Sigma_{r,g}=\frac{\eta_2}{\xi_2}A_{r,g}(z_1)-\frac{\eta_1}{\xi_1}B_{r,g}(z_1)$$

$$\sigma(r,g)=2^{-2g/3}\binom{2r}{r}\pi,$$

$$\tau(r,g)=2^{g/3}\tfrac{3}{2}(3.1)^{2r+1-g}\binom{r-g+1/3}{r+1-g}\binom{r-1/3}{r}\Big/\binom{2r+1-g}{r},$$

$$\Xi_{r,g}=\sigma(r,g)\,k\Delta^{g/3}|\xi_1|^{3r+1-g}|\xi_2|^{-2},$$

$$\Pi_{r,g}=\tau(r,g)\,k^{2r+1-g}\Delta^{r-g/6}|\xi_2||\xi_1|^{-3r-2(1-g)}.$$

If $\Sigma_{r,g}\neq 0$, *then*

$$1<\tfrac{1}{3}\Xi_{r,g}+\tfrac{2}{3}\Pi_{r,g}. \tag{40}$$

Proof. We shall consider the number

$$\Lambda_{r,g}=\xi_1^{3r+1-g}\xi_2\Sigma_{r,g}=\xi_1^{1-g}\eta_2A^*_{r,g}(\xi_1^3,\xi_1^3-\eta_1^3)-\xi_1^{2g}\xi_2\eta_1B^*_{r,g}(\xi_1^3,\xi_1^3-\eta_1^3).$$

We suppose that $\Sigma_{r,g}$, whence $\Lambda_{r,g}$ is non-zero.

First of all, we shall estimate $|\Lambda_{r,g}|$ from below. If $g=0$ then we have, by Lemma 2, (11), (23), (24) and (8) that $\Lambda_{r,0}\in\mathbb{Z}\cdot\sqrt{-3D}$, hence

$$|\Lambda_{r,0}|\geqq\Delta^{1/2}. \tag{41}$$

If $g=1$ then we have, by Lemma 2 and (24), that $\Lambda_{r,1}^3\in O_0$. Furthermore, $\Lambda_{r,1}^3\notin\mathbb{Z}$. For suppose otherwise. Then we have for some third root of unity ρ, that $\Lambda_{r,1}=\rho\overline{\Lambda_{r,1}}$. Hence by Lemma 2,

$$\Sigma_{r,1}=\xi_1^{-3r}\xi_2^{-1}\rho\overline{\Lambda_{r,1}}=\xi_1^{-3r}\xi_2^{-1}\rho\eta_1^{3r}\eta_2\left(\frac{\xi_2}{\eta_2}A_{r,1}\left(1-\frac{\xi_1^3}{\eta_1^3}\right)-\frac{\xi_1}{\eta_1}B_{r,1}\left(1-\frac{\xi_1^3}{\eta_1^3}\right)\right)$$

$$=\rho\frac{\eta_1^{3r}}{\xi_1^{3r}}\left(A_{r,1}\left(1-\frac{\xi_1^3}{\eta_1^3}\right)-\frac{\xi_1\eta_2}{\xi_2\eta_1}B_{r,1}\left(1-\frac{\xi_1^3}{\eta_1^3}\right)\right)\in M(\rho).$$

Hence $\Sigma_{r,1}$ has degree at most 2 over M. However, by (12) and Lemma 2, there exists a non-zero element θ of M such that $\Sigma_{r,1}=(\eta(1,0)/\xi(1,0))\,\theta$. Choose α such that $\xi(\alpha,1)=\eta(\alpha,1)$. Then, by (12),

$$\alpha\in M(\eta(1,0)/\xi(1,0))=M(\Sigma_{r,1})$$

and by (11), $F(\alpha,1)=0$. Therefore, by the irreducibility of F, $\Sigma_{r,1}$ has degree 3 over M. Thus obtain a contradiction. Hence $\Lambda_{r,1}^3$ can not belong to $\mathbb{Z}$. This implies by (9) that

$$|\Lambda_{r,1}|\geqq(\tfrac{1}{2}\Delta^{1/2})^{1/3}.$$

Together with (41) this yields that for positive integers r and for $g\in\{0,1\}$,

$$|\Lambda_{r,g}|\geqq 2^{-g/3}\Delta^{1/2-g/3}. \tag{42}$$

We shall now estimate $|\Lambda_{r,g}|$ from above. By (38), (39), (19), $|1-z_1|=1$, (21), (32), (20), (38), (33) we have

$$\begin{aligned}
|\Lambda_{r,g}|&=|\xi_1|^{3r+1-g}|\xi_2|\left|\left(\frac{\eta_2}{\xi_2}-1\right)A_{r,g}(z_1)+z_1^{2r+1-g}F_{r,g}(z_1)\right|\\
&<|\xi_1|^{3r+1-g}|\xi_2|\left(\binom{2r-g}{r}\frac{\pi}{3}\frac{k\Delta^{1/2}}{|\xi_2|^3}\right.\\
&\quad\left.+\frac{\binom{r-g+1/3}{r+1-g}\binom{r-1/3}{r}}{\binom{2r+1-g}{r}}\left((1-10^{-6})^{1/2}\frac{3k\Delta^{1/2}}{|\xi_1|^3}\right)^{2r+1-g}\right)\\
&<|\xi_1|^{3r+1-g}||\xi_2|\left(\binom{2r-g}{r}\frac{\pi}{3}\frac{k\Delta^{1/2}}{|\xi_2|^3}\right.\\
&\quad\left.+\frac{\binom{r-g+1/3}{r+1-g}\binom{r-1/3}{r}}{\binom{2r+1-g}{r}}(3.1)^{2r+1-g}\frac{(k\Delta^{1/2})^{2r+1-g}}{|\xi_1|^{3(2r+1-g)}}\right)\\
&=2^{-g/3}\Delta^{1/2-g/3}(\tfrac{1}{3}\Xi_{r,g}+\tfrac{2}{3}\Pi_{r,g}).
\end{aligned}$$

Together with (42) this proves our lemma. □

Lemma 6. *$\Sigma_{r,g}\neq 0$ for $(r,g)=(1,1)$, $(1,0)$, $(2,0)$, $(3,0)$.*

Proof. First of all we shall show that $\Sigma_{1,1}\neq 0$. Suppose $\Sigma_{1,1}=0$. Then $\eta_2\xi_2^{-1}A_{1,1}(z_1)=\eta_1\xi_1^{-1}B_{1,1}(z_1)$, i.e.

$$\frac{\eta_2}{\xi_2}(1-\tfrac{1}{3}z_1)=\frac{\eta_1}{\xi_1}.$$

Therefore,

$$\begin{aligned}
1&=\left|\frac{\eta_1}{\xi_1}\right|^2=\left|\frac{\eta_2}{\xi_2}\right|^2\left|1-\frac{1}{3}\left(1-\frac{\eta_1^3}{\xi_1^3}\right)\right|^2=\left|\frac{2}{3}+\frac{1}{3}\frac{\eta_1^3}{\xi_1^3}\right|^2\\
&=\left(\frac{2}{3}+\frac{1}{3}\frac{\eta_1^3}{\xi_1^3}\right)\left(\frac{2}{3}+\frac{1}{3}\frac{\xi_1^3}{\eta_1^3}\right)=\frac{5}{9}+\frac{2}{9}\left(\frac{\xi_1^3}{\eta_1^3}+\frac{\eta_1^3}{\xi_1^3}\right).
\end{aligned}$$

This implies that $\xi_1^3/\eta_1^3+\eta_1^3/\xi_1^3=2$, where $\xi_1^3/\eta_1^3=1$, which is impossible by (11) and the irreducibility of F.

We now assume that $\Sigma_{r,0}=0$ for some $r\in\{1,2,3\}$. Let $q^*(r)$ be the smallest positive rational integer such that both $q^*(r)A_{r,0}(z)$ and $q^*(r)B_{r,0}(z)$ have rational integral coefficients. Put $u=\xi_1^3$, $v=\xi_1^3-\eta_1^3$, $\bar{A}_r(z)=q^*(r)A_{r,0}(z)$, $\bar{B}_r(z)=q^*(r)B_{r,0}(z)$, and similarly for $\bar{A}_r^*(x,y)$, $\bar{B}_r^*(x,y)$. Then, by our assumption,

$$\frac{\eta_2^3}{\xi_2^3}=\frac{(u-v)(\bar{B}_r^*(u,v))^3}{u(\bar{A}^*(u,v))^3}.$$

Let $\mathfrak{a}_r$ be the integral ideal in M generated by $u\bar{A}_r^*(u,v)^3$ and $(u-v)\bar{B}_r^*(u,v)^3$. Then the ideal generated by $u(\bar{A}_r^*(u,v))^3-(u-v)(\bar{B}_r^*(u,v))^3$ divides $(\xi_2^3-\eta_2^3)\cdot\mathfrak{a}_r$. Hence by (31)

$$\begin{aligned}|u|^{3r+1}|\bar{A}_r(z_1)^3-(1-z_1)\bar{B}_r(z_1)^3|&=|u(\bar{A}_r^*(u,v))^3-(u-v)(\bar{B}_r^*(u,v))^3|\\&\leqq N(\mathfrak{a}_r)^{1/2}|\xi_2^3-\eta_2^3|\leqq N(\mathfrak{a}_r)^{1/2}\cdot 3k\Delta^{1/2}\end{aligned}\tag{43}$$

where $N(\mathfrak{a}_r)$ denotes the absolute norm of the ideal $\mathfrak{a}_r$. Here we used, that the norm of an element of an imaginary quadratic field is equal to the square of its absolute value. From (19) we infer that there exists a polynomial $K_r(z)$, with rational integral coefficients, such that

$$\bar{A}_r(z)^3-(1-z)\bar{B}_r(z)^3=z^{2r+1}K_r(z).$$

Hence, by (43),

$$|z_1|^{2r+1}|K_r(z_1)|\leqq N(\mathfrak{a}_r)^{1/2}|u|^{-3r-1}3k\Delta^{1/2}=|z_1|^{3r+1}N(\mathfrak{a}_r)^{1/2}|v|^{-3r-1}3k\Delta^{1/2}.$$

Therefore, by (31),

$$1\leqq\frac{N(\mathfrak{a}_r)^{1/2}|v|^{-3r-1}|z_1|^r\cdot 3k\Delta^{1/2}}{|K_r(z_1)|}\leqq\frac{N(\mathfrak{a}_r)^{1/2}|v|^{-3r-1}}{|K_r(z_1)|}(3k\Delta^{1/2})^{r+1}|\xi_1|^{-3r}.$$

This implies, by the fact that $k^2\Delta\geqq 444$,

$$|\xi_1|\leqq\left(\frac{N(\mathfrak{a}_r)^{1/2}|v|^{-3r-1}}{|K_r(z_1)|}\right)^{1/3r}3^{(r+1)/3r}(444)^{(1-5r)/6r}\cdot k^3\Delta.\tag{44}$$

A straightforward computation shows that, by (18),

$$\begin{aligned}\bar{A}_1(z)&=3-2z,\\\bar{A}_2(z)&=54-63z+14z^2,\\\bar{A}_3(z)&=81-135z+63z^2-7z^3,\\\bar{B}_1(z)&=3-z,\\\bar{B}_2(z)&=54-45z+5z^2,\\\bar{B}_3(z)&=81-108z+36z^2-2z^3,\\K_1(z)&=2-z,\\K_2(z)&=756-756z+125z^2,\\K_3(z)&=162-243z+97z^2-8z^3.\end{aligned}$$

Hence, by (38)

$$|K_1(z_1)|>1, \quad |K_2(z_1)|>755, \quad |K_3(z_1)|>161. \tag{45}$$

We shall now estimate $N(\mathfrak{a}_r)$ from above. We choose a finite extension M' of M in which the ideal generated by u and v is principal, with generator w say. We put $u'=u/w$, $v'=v/w$ and we denote the extension of $\mathfrak{a}_r$ to M' by $\mathfrak{a}_r'$. The ideal in M' generated by $\alpha_1, \alpha_2, \ldots, \alpha_n$ is denoted by $(\alpha_1, \alpha_2, \ldots, \alpha_n)$. Put $\mathfrak{r}_r = (\bar{A}_r^*(u', v'), \bar{B}_r^*(u', v'))$. Note that $A_r^*(x, x-y) = \bar{B}_r^*(y, y-x)$. This can be verified for $r \in \{1, 2, 3\}$ by straightforward computation but follows also from the proof of (23). Hence

$$\begin{aligned}
\mathfrak{a}_r' &= w^{3r+1}(u'\bar{A}_r^*(u', v')^3, (u'-v')\bar{B}_r^*(u', v')^3) \\
&\supset w^{3r+1}(u', u'-v')(u', \bar{B}_r^*(u', v')^3)(u'-v', \bar{A}_r^*(u', v')^3)\,\mathfrak{r}_r^3 \\
&\supset w^{3r+1}(u', \bar{B}_r^*(0, v')^3)(u'-v', \bar{A}_r^*(v', v')^3)\,\mathfrak{r}_r^3 \\
&\supset w^{3r+1}(u', B_r^*(0, v')^3)(u'-v', \bar{B}_r^*(0, v')^3)\,\mathfrak{r}_r^3 \supset w^{3r+1}\,\bar{B}_r^*(0, 1)^3\,\mathfrak{r}_r^3.
\end{aligned} \tag{46}$$

We have the following identities:

$$\begin{aligned}
&\bar{A}_1^*(x, y) - \bar{B}_1^*(x, y) = -y, \\
&(3x-y)\bar{A}_2^*(x, y) - (3x-2y)\bar{B}_2^*(x, y) = -4y^3, \\
&(27x-5y)\bar{A}_2^*(x, y) - (27x-14y)\bar{B}_2^*(x, y) = -72xy^2, \\
&\bar{A}_3^*(x, y) - \bar{B}_3^*(x, y) = (-27x^2+27xy-5y^2)y, \\
&(3x-y)\bar{A}_3^*(x, y) - (3x-2y)\bar{B}_3^*(x, y) = (-6x+3y)y^3.
\end{aligned}$$

These imply

$$\begin{aligned}
&\mathfrak{r}_1 \supset (3u'-v', v') \supset (v'), \\
&\mathfrak{r}_2 \supset (72u'v'^2, 4v'^3) \supset (72u', 4v')v'^2 \supset 72(v')^2, \\
&\mathfrak{r}_3 \supset (6u'-3v')v'^3, (27u'^2-27u'v'+5v'^2)v') \\
&\quad \supset 3v'^3(2u'-v', 27u'^2-27u'v'+5v'^2) \\
&\quad = 3v'^3(2u'-v', 27u'^2-27u'(2u')+5(2u')^2) \\
&\quad = 3v'^3(2u'-v', 7u'^2) \supset 21(v')^3.
\end{aligned}$$

Hence by (46),

$$\begin{aligned}
&\mathfrak{a}_1' \supset (w^4 v'^3) \supset (v)^4, \quad \mathfrak{a}_2' \supset (360^3 w^7 v'^6) \supset 360^3 (v)^7, \\
&\mathfrak{a}_3' \supset (42^3 w^{10} v'^9) \supset 42^3 (v)^{10}.
\end{aligned}$$

Therefore,

$$N(\mathfrak{a}_1)^{1/2}|v|^{-4} \leqq 1, \quad N(\mathfrak{a}_2)^{1/2}|v|^{-7} \leqq 360^3, \quad N(\mathfrak{a}_3)^{1/2}|v|^{-10} \leqq 42^3.$$

In combination with (44) and (45) this yields that

$$\begin{aligned}
&|\xi_1| < 0.04\,k^3\,\Delta \quad \text{if } r=1, \quad |\xi_1| < 0.12\,k^3\,\Delta \quad \text{if } r=2, \\
&|\xi_1| < 0.03\,k^3\,\Delta \quad \text{if } r=3.
\end{aligned}$$

But these inequalities are contradictory to (36) which proves our lemma. □

Lemma 7. *There are integers ℓ_1, ℓ_2 with $1 \leqq \ell_1 \leqq \ell_2 \leqq \ell_1 + 1$ such that*

$$\begin{aligned} \sigma(\ell_1, 0)|\xi_1|^{3\ell_1+1} \leqq k^{-1}|\xi_2|^2 < \sigma(\ell_1+1, 0)|\xi_1|^{3\ell_1+4}, \\ \sigma(\ell_2, 1)|\xi_1|^{3\ell_2} \leqq k^{-1}\Delta^{-1/3}|\xi_2|^2 < \sigma(\ell_2+1, 1)|\xi_1|^{3\ell_2+3}. \end{aligned} \tag{47}$$

Proof. For the sake of completeness we put $\sigma(0,0) = \sigma(0,1) = 0$. Then the sequences $\sigma(\ell, 0)|\xi_1|^{3\ell+1}$ and $\sigma(\ell, 1)|\xi_1|^{3\ell}$ increase monotonically to infinity and their terms with $\ell = 0$ are equal to 0. Hence there exist non-negative integers ℓ_1, ℓ_2 satisfying (47). Firstly we shall show that $\ell_1 \geqq 1$. We assume the contrary, i.e.

$$|\xi_2| < (\sigma(1,0)k)^{1/2}|\xi_1|^2 = (2\pi k)^{1/2}|\xi_1|^2. \tag{48}$$

Note that by Lemma 5 and Lemma 6, with $r = g = 1$, $3 < \Xi_{1,1} + 2\Pi_{1,1}$. On the other side, by (37), $\Delta \geqq 147$, and (36),

$$\begin{aligned} \Xi_{1,1} &= \sigma(1,1)k\Delta^{1/3}|\xi_1|^3|\xi_2|^{-2} \leqq \sigma(1,1)k^3\Delta^{1/3}0.948^{-2}|\xi_1|^{-1} \\ &\leqq \sigma(1,1)0.948^{-2} \times 147^{-2/3}k^3\Delta|\xi_1'|^{-1} < 1. \end{aligned}$$

Furthermore, by (48), (36) and the fact that $k^{1/2}\Delta^{1/6} \geqq 444^{1/6}$,

$$\begin{aligned} \Pi_{1,1} &= \tau(1,1)k^2\Delta^{5/6}|\xi_2||\xi_1|^{-3} \leqq \tau(1,1)(2\pi)^{1/2}k^{5/2}\Delta^{5/6}|\xi_1|^{-1} \\ &\leqq \tau(1,1)(2\pi)^{1/2}444^{-1/6}k^3\Delta|\xi_1|^{-1} < 1.84k^3\Delta|\xi_1| < 1. \end{aligned}$$

This contradiction shows that $\ell_1 \geqq 1$.

We shall now show that $\ell_1 \leqq \ell_2 \leqq \ell_1 + 1$. Suppose $\ell_1 \geqq \ell_2 + 1$. Then

$$\binom{2\ell_2+2}{\ell_2+1}\pi|\xi_1|^{3\ell_2+4} \leqq k^{-1}|\xi_2|^2 < 2^{-2/3}\Delta^{1/3}\binom{2\ell_2+2}{\ell_2+1}\pi|\xi_1|^{3\ell_2+3},$$

hence

$$|\xi_1| < \left(\frac{\Delta}{4}\right)^{1/3},$$

which is clearly impossible. If $\ell_2 \geqq \ell_1 + 1$, then

$$\begin{aligned} \binom{2\ell_1+2}{\ell_1+1}\pi|\xi_1|^{3\ell_1+6} &\leqq 2^{-2/3}\pi\binom{2\ell_1+4}{\ell_1+2}|\xi_1|^{3\ell_1+6} \\ &\leqq k^{-1}\Delta^{-1/3}\pi\binom{2\ell_1+2}{\ell_1+1}|\xi_1|^{3\ell_1+4}, \end{aligned}$$

hence

$$|\xi_1| < \Delta^{-1/6} < 1,$$

which is also impossible. This completes the proof of Lemma 7. □

We shall now complete the proof of Theorem 3. Let ℓ_1, ℓ_2 be the integers defined in Lemma 7. We choose $r = \ell_1$, $g = 0$ if $\Sigma_{\ell_1,0} \neq 0$, but $r = \ell_2$, $g = 1$ otherwise. Then we have by (22) and Lemma 7 that $\Sigma_{r,g} \neq 0$. Note that by Lemma 6, $g = 1$ implies that $r \geqq 4$. Furthermore, by (47) and our definitions of r and g,

$$\sigma(r,g)|\xi_1|^{3r+1-g} \leqq k^{-1}\Delta^{-g/3}|\xi_2|^2 < \sigma(r+1,g)|\xi_1|^{3r+4-g}. \tag{49}$$

The left-hand side inequality clearly implies that $\Xi_{r,g} \leqq 1$. We shall now show that $\Pi_{r,g} \leqq 1$. Then we have a contradiction with Lemma 5 which shows that (31) can not have four solutions being related to (ξ, η), i.e. that Theorem 3 holds. Note that, by the facts that $k^2 \Delta \geqq 444$, $r - 3g > 0$, and by the right-hand side inequality of (49),

$$\begin{aligned}
\Pi_{r,g} &= \tau(r,g)\, k^{2r+1-g}\, \Delta^{r-g/6}\, |\xi_2|\, |\xi_1|^{-3r-2(1-g)} \\
&< (\sigma(r+1,g)\, \tau(r,g)^2\, k^{4r+3-2g}\, \Delta^{2r}\, |\xi_1|^{3(g-r)})^{1/2} \\
&\leqq ((\sigma(r+1,g)\, \tau(r,g)^2 (k^2 \Delta)^{3g-r} (k^3 \Delta\, |\xi_1|^{-1})^{3(g-r)})^{1/2} \\
&\leqq (\Omega_{r,g}\, k^3 \Delta\, |\xi_1|^{-1})^{3(r-g)/2},
\end{aligned}$$

where

$$\begin{aligned}
\Omega_{r,g} &= (\sigma(r+1,g)\, \tau(r,g)^2\, 444^{3g-r})^{1/3(r-g)} \\
&= \left(\binom{2r+2}{r+1} \left(\frac{3}{2} \binom{r-g+1/3}{r+1-g} \binom{r-1/3}{r} \binom{2r+1-g}{r} \right)^2 \cdot \right. \\
&\qquad \left. \cdot\, \pi (3 \cdot 1)^{2(2r+1-g)}\, 444^{3g-r} \right)^{1/3(r-g)}.
\end{aligned}$$

By (36) it suffices to show that $\Omega_{r,g} \leqq 1.91$. But this is easy, since for each pair of integers (s,h) with $s \geqq 1, h \in \{0,1\}$,

$$\begin{aligned}
\frac{\Omega_{s+1,h}^{3(s+1-h)}}{\Omega_{s,h}^{3(s-h)}} &= \frac{(2s+4)(2s+3)}{(s+2)(s+2)} \\
&\quad \cdot \left(\frac{(s+1-h+1/3)(s+1-1/3)}{(s+2-h)(s+1)} \cdot \frac{(2s+3-h)(2s+2-h)}{(s+2-h)(s+1)} \right)^2 (3 \cdot 1)^4\, 444^{-1} \\
&< 4 \times (3 \cdot 1)^4 \times 444^{-1} \left(\frac{(s+4/3-h)(s-2/3)}{(2s+3-h)(2s+2-h)} \right)^2 \\
&< 4 \times (3 \cdot 1)^4 \times 444^{-1} < 1,
\end{aligned}$$

whereas

$$\Omega_{1,0} = \left(\binom{4}{2} \left(\frac{3}{2} \binom{4/3}{2} \binom{2/3}{1} \binom{3}{1} \right)^2 \pi \cdot (3 \cdot 1)^6\, 444^{-1} \right)^{1/3} < 0.6,$$

$$\Omega_{4,1} = \left(\binom{10}{5} \left(\frac{3}{2} \binom{10/3}{4} \binom{11/3}{4} \binom{8}{4} \right)^2 \pi \cdot (3 \cdot 1)^{16}\, 444^{-1} \right)^{1/9} < 1.85.$$

This proves Theorem 3 completely. □

References

1. Avanesov, È.T.: The representation of numbers by binary cubic forms of positive discriminant (Russian). Acta Arith. **14**, 13-25 (1967/68)
2. Baulin, V.I.: On an indeterminate equation of the third degree with least positive discriminant (Russian). Tul'sk Gos. Ped. Inst. U'čen Zap. Fiz. Math. Nauk. Vip. **7**, 138-170 (1960)
3. Delauney, B.: Über die Darstellung der Zahlen durch die binären kubischen Formen von negativer Diskriminante. Math. Z. **31**, 1-26 (1930)

4. Delone, B.N., Faddeev, D.K.: The theory of irrationalities of the third degree. Am. Math. Soc. Transl. of Math. Monographs **10**, (Providence, USA, 1964)
5. Klein, F.: Vorlesungen über die hypergeometrische Funktion. Grundl. der Math. Wiss. XXXIX. Berlin-Heidelberg-New York: Springer 1933
6. Ljunggren, W.: Einige Bemerkungen über die Darstellung ganzer Zahlen durch binäre kubische Formen mit positiver Diskriminante. Acta Math. **75**, 1–21 (1942)
7. Mordell, L.J.: Diophantine equations. London-New York: Academic Press 1969
8. Nagell, T.: Darstellung ganzer Zahlen durch binäre kubische Formen mit negativer Diskriminante. Math. Z. **28**, 10–29 (1928)
9. Siegel, C.L.: Über einige Anwendungen diophantischer Approximationen. Abh. preuss. Akad. Wiss. Phys. Math. Kl. (1929), Nr. 1. Gesammelte Abhandlungen I, pp. 209-266. Berlin-Heidelberg-New York: Springer 1966
10. Siegel, C.L.: Die Gleichung $ax^n - by^n = c$. Math. Ann. **114**, 57–68 (1937); Ges. Abh. II, pp. 8–20. Berlin-Heidelberg-New York: 1966
11. Siegel, C.L.: Einige Erläuterungen zu Thues Untersuchungen über Annäherungswerte algebraischer Zahlen und diophantische Gleichungen. Nachr. Akad. Wiss. Göttingen, II, Math. Phys. Kl. (1970, Nr. 8. Ges. Abh. IV, pp. 140–166. Berlin-Heidelberg-New York: Springer 1979
12. Skolem, Th.: Einige Sätze über gewisse Reihenentwicklungen und exponentiale Beziehungen mit Anwendung auf diophantische Gleichungen. Oslo Vid. Akad. Skrifter **I**, (Nr. 6) 1–61 (1933)
13. Skolem, Th.: Ein Verfahren zur Behandlung gewisser exponentialer Gleichungen und diophantischer Gleichungen. 8^{de} Skand. Mat. Kongr. Förh. Stockh 1934, 163–168
14. Skolem, Th.: Einige Sätze über p-adische Potenzreihen mit Anwendung auf gewisse exponentielle Gleichungen. Math. Ann. **111**, 399–424 (1935)

Oblatum 3-XII-1982

Invent. math. 73, 139–150 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Every K3 Surface is Kähler*

Y.-T. Siu

Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

A K3 surface means a simply connected compact complex manifold of complex dimension two whose canonical line bundle is trivial. In this paper it will be proved that every K3 surface is Kähler. This is the only case yet unconfirmed of the conjecture of Kodaira that every compact complex surface with even first Betti number is Kähler [7, p. 85]. Thus the result of this paper completes the confirmation of Kodaira's conjecture.

In [18] Todorov presented a proof that every K3 surface is Kähler. There are serious difficulties with his proof. These difficulties will be fully explained later in this introduction. The main purpose of publishing this paper is to clear up once for all the confusion caused by Todorov's proof by explaining the difficulties in his proof and presenting a correct proof of the Kähler character of K3 surfaces.

Yau [19] obtained, as a special case of his proof of the Calabi conjecture, a unique Kähler-Einstein metric in every Kähler class of a K3 surface. Yau's result immediately opened the way to the following approach of proving the Kähler character of K3 surfaces. Take an arbitrary K3 surface M which we hope to prove to be Kähler and deform it to obtain a holomorphic family of K3 surfaces M_s, $s\in U$, where U is an open ball in $\mathbb{C}^{20}$ centered at 0, such that $M_0=M$ and the family is complete and effectively parametrized at $s=0$. There is a dense subset A of U such that M_s is projective algebraic for $s\in A$. One hopes to be able to find a sequence $\{s_\nu\}_{\nu=1}^{\infty}$ in A approaching 0 and to find for each ν a suitable Kähler class $b_{s_\nu}\in H^{1,1}_{\mathbb{R}}(M_{s_\nu})$ so that the unique Kähler-Einstein metric g_{s_ν} of M_{s_ν} in the Kähler class b_{s_ν} converges to a Kähler metric g_0 of M_0 as $\nu\to\infty$. Here $H^{1,1}_{\mathbb{R}}(M_s)$ means the set of all elements of $H^2(M_s,\mathbb{R})$ whose cup product with the cohomology class of a nowhere zero holomorphic 2-form on M_s is zero. The first major difficulty in this approach is how to choose the sequence $\{s_\nu\}$ and the Kähler classes b_{s_ν} so that b_{s_ν} converges to a suitable class in $H^{1,1}_{\mathbb{R}}(M_0)$. For want of a better name we call this the topological difficulty. The topological difficulty corresponds to finding a class in

* Research partially supported by an NSF Grant

$H^{1,1}_{\mathbb{R}}(M_0)$ which can qualify as a candidate for the class of a Kähler form on M_0 in view of the stability of Kähler structures proved by Kodaira-Spencer [8, p. 75, Th. 15]. The natural way to get a candidate is to first choose one component from the two components of the positive cone of $H^{1,1}_{\mathbb{R}}(M_0)$ and then choose in that component a class whose value at the homology class of any effective divisor of self-intersection -2 in M_0 is positive. The difficulty is to prove that for sufficiently large ν and for one of the two choices of components, a class in $H^{1,1}_{\mathbb{R}}(M_{s_\nu})$ sufficiently close to this class is a Kähler class. The second major difficulty is to prove the convergence of g_{s_ν} to a Kähler metric g_0 of M_0. We call the second difficulty the analytical difficulty.

Prior to Yau's proof of the Calabi conjecture, Hitchin observed that from a Kähler-Einstein metric on a K3 surface one can construct a family of complex structures parametrized by the two-sphere [5, p. 436, Remark 2 and p. 439, lines 12–18] (see also Sommese's paper [16, p. 201, Prop. II and p. 213, Remark III F]). After Yau's proof of the Calabi conjecture, Todorov [18] used this family (together with Kulikov's result on the period map of algebraic K3 surfaces [9]) to obtain the surjectivity of the period map for Kähler K3 surfaces. Todorov's surjectivity result made possible a second approach. This second approach is to imitate the proof of the Torelli theorem for Kähler K3 surfaces by Burns-Rapoport [2] to try to prove that a K3 surface which has the same period as a Kähler K3 surface must be biholomorphic to it. We describe below this second approach.

One chooses one component from the two components of the positive cone of $H^{1,1}_{\mathbb{R}}(M_0)$. Then one takes a Kähler K3 surface N_0 with an isomorphism σ': $H^2(N_0, \mathbb{Z}) \to H^2(M_0, \mathbb{Z})$ compatible with the cup products such that the isomorphism σ: $H^2(N_0, \mathbb{C}) \to H^2(M_0, \mathbb{C})$ induced by σ' sends the period of N_0 to the period of M_0 and the image of the Kähler class of N_0 under σ is in the chosen component of the positive cone of $H^{1,1}_{\mathbb{R}}(M_0)$ and has positive value at the homology class of any effective divisor of self-intersection -2 in M_0. The existence of N_0 follows from Todorov's surjectivity result of the period map of Kähler K3 surfaces (see [10] and [18, Th. 3]). One deforms N_0 to get an effectively parametrized complete holomorphic family of Kähler K3 surfaces N_s, $s \in U'$. Without loss of generality one can assume that $U = U'$ and the periods of N_s and M_s are the same for $s \in U$ when compared by maps induced by σ. By the weak Torelli theorem for Kähler K3 surfaces [11, p. 145], for every $s \in A$ there is a biholomorphism f_s: $N_s \to M_s$. One hopes that for one of the two choices of the components of the positive cone of $H^{1,1}_{\mathbb{R}}(M_0)$, f_{s_ν} converges to a biholomorphism between N_0 and M_0 for some sequence $\{s_\nu\}$ in A approaching 0. The biholomorphism f_s induces an isomorphism τ_s: $H^2(N_s, \mathbb{C}) \to H^2(M_s, \mathbb{C})$ which in turn induces an isomorphism σ_s: $H^2(N_0, \mathbb{C}) \to H^2(M_0, \mathbb{C})$ via topological trivializations of the families $\{N_s\}_{s \in U}$ and $\{M_s\}_{s \in U}$. The first major difficulty is to prove the convergence of the maps σ_{s_ν} to an isomorphism between $H^2(N_0, \mathbb{C})$ and $H^2(M_0, \mathbb{C})$. In view of the (strong) Torelli theorem for algebraic or Kähler K3 surfaces [2, 11, 14], this difficulty is the same as the topological difficulty in the first approach. Note that one does not know that $\sigma_s = \sigma$ for $s \in A$ though one expects it to be the case. The second major difficulty is to prove the convergence of the maps f_{s_ν} to a biholomorphism. Again one difficulty is topologi-

cal and one is analytical. These two difficulties are parallel to the two important steps in the proof of the Torelli theorem for Kähler K3 surfaces [2, 11]. The first difficulty corresponds to getting the topological information needed to prove the finiteness of the volume of the graph of f_{s_ν} and the second difficulty corresponds to applying Bishop's theorem [1] on the convergence of a sequence of subvarieties of uniformly bounded volume. In our case the removal of the topological difficulty does not imply directly the uniform boundedness of the volume of the graph of f_{s_ν} because of the lack of Kähler forms on M_{s_ν} which converge to a uniformly positive definite form on M_0 as $\nu \to \infty$.

The topological difficulties in both approaches are the same, whereas the analytical difficulty in the second approach is easier to handle than in the first, because the convergence of holomorphic objects is easier to prove than the convergence of real-analytic objects.

What Todorov used in his proof of the Kähler character of K3 surfaces [18, §4] is essentially equivalent to the second approach, though on the surface it looks different from and more complicated than the above description. In the description above N_0 is produced by invoking the surjectivity of the period map and N_s is obtained by deforming N_0, whereas in his proof Todorov reproduced once more the Hitchin construction to obtain Kähler K3 surfaces corresponding the surfaces N_s described above.

In Todorov's proof there are two problems corresponding to the two major difficulties mentioned above. To describe the two problems, we use the notations of [18]. In [18, p. 264, (3.11)] one does not know that q is a Kähler class of X though $g(q)$ is a Kähler class of X for some element g of the group G of reflections introduced in [18, p. 261, (3.5.2)]. Thus one does not know that y belongs to $\mathbb{P}_x^1(q)$. This is the topological difficulty. He claimed that $w_{Y_s}(2,0)$ is harmonic with respect to the Kähler-Einstein metric of X, which is in general not true, because a biholomorphism between two K3 complex structures on a common differentiable manifold need not induce the identity map on the common underlying differentiable manifold. Thus one has to worry about the convergence of biholomorphisms, which is the analytical difficulty.

In this paper we show how the two major difficulties in the second approach can be overcome. Both difficulties are overcome by constructing on any given K3 surface a closed 2-form whose (1, 1)-component is everywhere positive definite. Such a 2-form is constructed by using the separation theorem for locally convex linear topological spaces, a technique first introduced by Sullivan [17].

I would like to thank S.-T. Yau for useful discussions.

1. Closed Forms with Positive (1, 1)-Component

We prove in this section the existence on any K3 surface a closed 2-form with positive definite (1, 1)-component.

(1.1) Let $\mathscr{H}_{(s)}$ be the Sobolev space of functions (or distributions) whose derivatives of order $\leqq s$ are square integrable. Sobolev spaces $\mathscr{H}_{(s)}$ can be defined

over Euclidean spaces for all real values of s by using Fourier transform and can be defined over compact manifolds by using partitions of unity.

Let M be a K3 surface. For any integer k let A_k be the set of all global real 2-forms (or 2-currents) on M whose coefficients are in the Sobolev space $\mathscr{H}_{(k)}$ of M. The Sobolev space A_k is a real Hilbert space with respect to a norm $\|\cdot\|_k$. (The norm $\|\cdot\|_k$ depends on the partition of unity and the choice of local coordinates. A different partition of unity and another choice of local coordinates give another equivalent norm.)

For 2-forms (or 2-currents) ξ, η on M define

$$(\xi, \eta) = \int_M \xi \wedge \eta$$

if the integral exists. Take a Riemannian metric on M and let $*$ be its Hodge star operator. The two Hilbert spaces A_k and A_{-k} are the dual of each other under the pairing which sends $\xi \in A_k$ and $\eta \in A_{-k}$ to $(\xi, *\eta)$ in the sense that for $\xi \in A_k$ the norm of ξ defined by $\sup |(\xi, *\eta)|$ over $\|\eta\|_{-k} \leqq 1$ is equivalent to $\|\xi\|_k$. Since $*$ defines an isomorphism of the Hilbert space A_{-k} to itself, it follows that the two Hilbert spaces A_k and A_{-k} are the dual of each other under the pairing which sends $\xi \in A_k$ and $\eta \in A_{-k}$ to (ξ, η). Hence the weak topology of the Hilbert space A_{-k} coincides with the weakest topology of A_{-k} in which the map $\eta \mapsto (\xi, \eta)$ is continuous for every $\xi \in A_k$.

For $R > 0$ let $B_{-k}(R)$ be the set of all $\eta \in A_{-k}$ with $\|\eta\|_{-k} \leqq R$. Then $B_{-k}(R)$ is compact in the weak topology of A_{-k} [3, p. 425, V. 4.7].

(1.2) We cover the K3 surface M by a finite number of coordinate charts U_ν with coordinates z_1^ν, z_2^ν $(1 \leqq \nu \leqq \ell)$. Let U_ν' be a relatively compact open subset of U_ν $(1 \leqq \nu \leqq \ell)$ such that $\{U_\nu'\}_{1 \leqq \nu \leqq \ell}$ covers M. Let S be the set of all $\lambda = (\lambda_1, \lambda_2) \in \mathbb{C}^2$ with $|\lambda_1|^2 + |\lambda_2|^2 = 1$. For $1 \leqq \nu \leqq \ell$, $P \in U_\nu$, and $\lambda \in S$, we denote by $T_{\nu, P, \lambda}$ the $(1,1)$-current on M whose value $T_{\nu, P, \lambda}(\phi)$ at every real smooth $(1,1)$-form

$$\phi = \sqrt{-1} \sum_{i,j=1}^{2} \phi_{i\bar{j}} dz_i^\nu \wedge \overline{dz_j^\nu}$$

on M is

$$\sum_{i,j=1}^{2} \phi_{i\bar{j}}(P) \lambda_i \overline{\lambda_j}.$$

For a nonnegative integer s let C_s be the Banach space of all s times continuously differentiable real 2-forms on M. By the Sobolev lemma [13, p. 78, Th. 3.5.1], for $k \geqq 3+s$ the Hilbert space A_k is contained in C_s and the inclusion map $A_k \hookrightarrow C_s$ is continuous with respect the Hilbert space topology of A_k and the Banach space topology of C_s. Hence for $k \geqq 3$ we have $T_{\nu, P, \lambda} \in A_{-k}$ and, moreover, the supremum Λ_1 of $\|T_{\nu, P, \lambda}\|_{-k}$ for $1 \leqq \nu \leqq \ell$, $P \in U_\nu'$, and $\lambda \in S$ is finite. We now fix k to be an integer $\geqq 4$.

(1.3) Choose a smooth positive definite $(1,1)$-form ω on M. Let Λ_2 be the infimum of $T_{\nu, P, \lambda}(\omega)$ for $1 \leqq \nu \leqq \ell$, $P \in U_\nu'$, and $\lambda \in S$. Then Λ_2 is positive. Let $\mathscr{P}$ be the set of all elements of A_k which are positive semidefinite $(1,1)$-froms on M. Let E be the set of all $\eta \in A_{-k}$ such that

(i) η is of type $(1,1)$,

(ii) $\|\eta\|_{-k} \leqq \Lambda_1$,

(iii) the value $\eta(\omega)$ of the $(1,1)$-current η at ω is $\geqq \Lambda_2$, and

(iv) $\eta(\xi) \geqq 0$ for every $\xi \in \mathscr{P}$.

Then every $T_{v,P,\lambda}$ is in E. Clearly E is a convex subset of A_{-k}. Since $B_{-k}(\Lambda_1)$ is compact in the weak topology of A_{-k} and since ω belongs to A_k, it follows that E is compact in the weak topology of A_{-k}.

Let F be the set of all elements of A_{-k} which are exact as 2-currents on M. Clearly F is a convex subset of A_{-k}. Moreover, F is closed in the weak topology of A_{-k}. For, if $\eta_i \in F$ approaches η in the weak topology of A_{-k}, then for every smooth closed real 2-form ψ on M, $\eta(\psi) = \lim \eta_i(\psi) = 0$.

(1.4) We claim that E and F are disjoint. To prove the claim, we assume that E and F contain a common element η and we want to derive a contradiction. Since $\eta \in E$, $\eta(\xi) \geqq 0$ for $\xi \in \mathscr{P}$. Hence η is a positive $(1,1)$-current on M. Since $\eta \in F$, it follows that $\eta = d\gamma$ for some real 1-current γ on M. We can write $\gamma = \beta + \overline{\beta}$, where β is a $(1,0)$-current on M. Since η is of type $(1,1)$, it follows that $\eta = \bar{\partial}\beta + \partial\overline{\beta}$ and $\bar{\partial}\overline{\beta} = 0$. From $H^1(M, \mathscr{O}_M) = 0$ we conclude that $\overline{\beta} = \bar{\partial}\sigma$ for some $(0,0)$-current σ on M. Hence $\eta = \sqrt{-1}\,\partial\bar{\partial}\tau$, where $\tau = \sqrt{-1}\,(\bar{\sigma} - \sigma)$. The positivity of the $(1,1)$-current η on M implies that τ is a plurisubharmonic function on M. By the compactness of M, τ must be constant. Thus η vanishes, which contradicts $\eta(\omega) \geqq \Lambda_2$. The claim is proved.

(1.5) By applying the separation theorem [3, p. 417, Th. 10] to the compact convex set E and the closed convex set F in the locally convex linear topological space A_{-k} endowed with its weak topology, we obtain a real linear functional f on A_{-k} such that

(i) f is continuous in the weak topology of A_{-k} and

(ii) $\sup_F f \leqq c < c' \leqq \inf_E f$ for some real numbers c and c'.

Since f is continuous in the weak topology of A_{-k}, it follows from [3, p. 421, V. 3.9] that there exists $\xi \in A_k$ such that $f(\eta) = (\xi, \eta)$ for every $\eta \in A_{-k}$. Since $\eta \in F$ implies $\mathbb{R}\eta \in F$ and since $\sup_F f \leqq c$, it follows that f is identically zero on F. This implies that $(\xi, d\zeta) = 0$ for every smooth real 1-form ζ on M. Hence ξ is d-closed. Let

$$\sum_{i,j=1}^{2} \xi_{i\bar{j}} dz_i^v \wedge \overline{dz_j^v}$$

be the $(1,1)$-component of ξ. From $T_{v,P,\lambda} \in E$ it follows that $(\xi, T_{v,P,\lambda}) \geqq c'$. Hence

$$\sum_{i,j=1}^{2} \xi_{i\bar{j}}(P)\, \lambda_i \overline{\lambda_j} \geqq c'$$

and the $(1,1)$-component of ξ is positive definite at every point of M. We have thus obtained a proof of the following.

(1.6) **Proposition.** *Let M be a* K3 *surface and k be an integer $\geqq 1$. Then there exists a real C^k d-closed 2-form on M whose $(1,1)$-component is positive definite at every point of M.*

(1.7) *Remark.* The method of using the separation theorem to construct forms was first introduced by Sullivan [17]. It was recently used by Harvey-Lawson [4] to obtain characterizations of Kähler manifolds and to get a new proof of Miyaoka's result [12].

§ 2. Overcoming the Analytic Difficulty

We overcome the analytical difficulty by using the closed 2-form with positive $(1,1)$-component constructed in §1 to refine the proof of the Main Lemma in the paper of Burns-Rapoport [2].

(2.1) **Proposition.** *Let $\pi\colon \mathscr{M} \to U$ and $\pi'\colon \mathscr{M}' \to U$ be two holomorphic families of* K3 *surfaces with a complex manifold U as the parameter space so that both are diffeomorphically identified with a trivial family $U \times X \to U$. Let τ be an isomorphism of $H^2(X, \mathbb{C})$ which is compatible with the quadratic form defined by the cup product. Let $M_s = \pi^{-1}(s)$ and $M'_s = \pi'^{-1}(s)$ for $s \in U$. Let s_0 be a point of U and let A be a subset of U such that s_0 is an accumulation point of A. Assume the following two conditions.*

(i) *M_{s_0} is Kähler.*

(ii) *For $s \in A$ the two* K3 *surfaces M_s and M'_s are biholomorphic under a map f_s which induces τ on $H^2(X, \mathbb{C})$.*

Then one concludes that the two K3 *surfaces M_{s_0} and M'_{s_0} are biholomorphic.*

(2.2) *Remark.* The assumption that f_s induces the same isomorphism τ of $H^2(X, \mathbb{C})$ for all $s \in A$ eliminates the topological difficulty.

We will devote the rest of this section to prove Prop. (2.1).

(2.3) Fix a positive integer k. By Prop. (1.6) there exists a real C^k d-closed 2-form θ on the underlying differentiable structure X such that the $(1,1)$-component $\omega'(s_0)$ of θ with respect to the complex structure of M'_{s_0} is positive definite at every point of M'_{s_0}. By continuity arguments there exists an open neighborhood W of s_0 in U such that for $s \in W$ the $(1,1)$-component $\omega'(s)$ of θ with respect to the complex structure of M'_s is positive definite at every point of M'_s.

Since M_{s_0} is assumed Kähler, (after shrinking W if necessary) we have for every $s \in W$ a Kähler form $\omega(s)$ on M_s which depends smoothly on s. Let η be a positive definite $(1,1)$-form on W. The collection of $(1,1)$-forms $\omega(s)$ on M_s and $(1,1)$-forms $\omega'(s)$ on M'_s for $s \in W$, together with η, define a Hermitian metric H on $\mathscr{M} \times_W \mathscr{M}'$. Let θ be the $(1,1)$-form on $\mathscr{M} \times_W \mathscr{M}'$ associated to the Hermitian metric H. Then the pullback of θ to the submanifold $M_s \times M'_s$ of $\mathscr{M} \times_W \mathscr{M}'$ is equal to $\omega(s) + \omega'(s)$, where for notational simplicity we use $\omega(s)$, $\omega'(s)$ to denote also their pullbacks under the projections from $M_s \times M'_s$ to M_s and M'_s respectively.

(2.4) For $s \in W \cap A$ let $\Gamma_s \subset M_s \times M'_s$ be the graph of the holomorphic map $f_s\colon M_s \to M'_s$. We want to compute the volume $\mathrm{Vol}(\Gamma_s)$ of Γ_s with respect to the Hermitian metric H of $\mathscr{M} \times_W \mathscr{M}'$ and show that it is bounded in s as s ap-

proaches s_0 so that we can apply Bishop's theorem to conclude the convergence of the subvariety Γ_s in $\mathscr{M} \times_W \mathscr{M}'$ as s approaches s_0.

Since $\mathrm{Vol}(\Gamma_s) = \frac{1}{2} \int_{\Gamma_s} \theta^2$ and since the pullback of η to Γ_s is zero, we have $\mathrm{Vol}(\Gamma_s) = I_s + II_s + III_s$, where

$$I_s = \tfrac{1}{2} \int_{M_s} \omega(s)^2$$

$$II_s = \tfrac{1}{2} \int_{M'_s} \omega'(s)^2$$

$$III_s = \int_{\Gamma_s} \omega(s) \wedge \omega'(s).$$

Since I_s and II_s are smooth in s, clearly they are both bounded in s as s approaches s_0. The only difficult term to handle is III_s. We can rewrite III_s as

$$\int_{M_s} u(s) \wedge f_s^* \omega'(s)$$

which is equal to

$$\int_{M_s} \omega(s) \wedge f_s^* \theta$$

because $\omega(s)$ is of type $(1,1)$ on M_s and the holomorphicity of f_s: $M_s \to M'_s$ implies that $f_s^* \omega'(s)$ is the $(1,1)$-component of $f_s^* \theta$ with respect to the complex structure of M_s. Let ψ be a real smooth d-closed 2-form on X whose cohomology class is mapped by τ to the cohomology class of θ. Since f_s induces the map τ on $H^2(X, \mathbb{C})$ and since both $\omega(s)$ and θ are d-closed 2-forms on X, it follows that

$$\int_{M_s} \omega(s) \wedge f_s^* \theta = \int_X \omega(s) \wedge \psi.$$

From the smooth dependence of $\omega(s)$ on s we conclude that III_s is bounded in s as s approaches s_0.

(2.5) For a subvariety Z of pure codimension in a complex manifold G, we denote by $[Z]$ the current on G defined by Z. If Z is contained in a complex submanifold G' of G, $[Z]$ is used also to denote the current on G' defined by Z and it will be clear from the context which meaning $[Z]$ takes. Now we invoke Bishop's theorem [1] and conclude that for some subsequence $\{s_\nu\} \subset A$ converging to s_0 the current $[\Gamma_{s_\nu}]$ on $\mathscr{M} \times_W \mathscr{M}'$ converges weakly to a current on $\mathscr{M} \times_W \mathscr{M}'$ of the form $\sum_{i=1}^{k} m_i [\Gamma^i]$, where m_i is a positive integer and Γ^i is an irreducible subvariety of complex dimension 2 in $M_{s_0} \times M'_{s_0}$.

(2.6) For any closed 4-current Ξ on $M_s \times M'_s$ define a linear map Ξ_*: $H^*(M_s, \mathbb{C}) \to H^*(M'_s, \mathbb{C})$ of cohomology rings as follows. A cohomology class defined by a closed p-form α on M_s is mapped by Ξ_* to the cohomology class defined by the closed p-current $(pr_2)_*(\Xi \wedge (pr_1)^* \alpha)$ on M'_s, where pr_1 and pr_2 are respectively the projections of $M_s \times M'_s$ onto the first and second factors and $(pr_2)_*$ and $(pr_1)^*$ mean respectively the corresponding pushforward and

pullback maps. By reversing the rôles of M_s and M'_s, we define analogously a linear map Ξ^*: $H^*(M'_s, \mathbb{C}) \to H^*(M_s, \mathbb{C})$.

The map $[\Gamma_s]_*$ defined by the 4-current $[\Gamma_s]$ in $M_s \times M'_s$ clearly agrees with the map from $H^*(M_s, \mathbb{C})$ to $H^*(M'_s, \mathbb{C})$ defined by f_s. Since f_s defines an isomorphism on $H^*(X, \mathbb{Z})$, by passing to limit along the subsequence $\{s_\nu\}$ we conclude that $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)_*$ is an isomorphism of $H^*(X, \mathbb{C})$. Likewise $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)^*$ is also an isomorphism of $H^*(X, \mathbb{C})$.

Let ϕ be a nowhere zero holomorphic 2-form on M_{s_0}. Since $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)_*$ is an isomorphism of $H^*(X, \mathbb{C})$, it follows that the 2-current

$$(pr_2)_* \left(\sum_{i=1}^{k} m_i[\Gamma^i] \wedge (pr_1)^* \phi\right)$$

on M'_{s_0} (which is automatically a holomorphic 2-form on M'_{s_0}) cannot be zero. Hence there must be some Γ^j which is projected both onto M_{s_0} and onto M'_{s_0}. There can only be one such Γ^j and, moreover, $m_j = 1$ and its projection maps onto M_{s_0} and M'_{s_0} are both of degree one, because both $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)_*$ and $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)^*$ must leave fixed the class in $H^0(X, \mathbb{C})$ which is defined by the function on X with constant value 1. This particular Γ^j must be projected biholomorphically onto both M_{s_0} and M'_{s_0}. For, if ϕ' is a nowhere zero holomorphic 2-form on M'_{s_0}, then the following two holomorphic 2-forms

$$(pr_1)_*([\Gamma^j] \wedge (pr_2)^* \phi') = (pr_1)_* \left(\sum_{i=1}^{k} m_i[\Gamma^i] \wedge (pr_2)^* \phi'\right)$$

$$(pr_2)_*([\Gamma^j] \wedge (pr_1)^* \phi) = (pr_2)_* \left(\sum_{i=1}^{k} m_i[\Gamma^i] \wedge (pr_1)^* \phi\right)$$

on M_{s_0} and M'_{s_0} respectively cannot be identically zero due to the fact that both $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)^*$ and $\left(\sum_{i=1}^{k} m_i[\Gamma^i]\right)_*$ are isomorphisms of $H^*(X, \mathbb{C})$ and therefore both holomorphic 2-forms are nowhere zero. This can happen only when Γ^j are projected biholomorphically onto both M_{s_0} and M'_{s_0}. Hence M_{s_0} and M'_{s_0} are biholomorphic. This concludes the proof of Prop. (2.1).

(2.7) *Remark.* The arguments in (2.6) are essentially the same as those on pp. 248–250 of the paper of Burns-Rapoport [2].

3. Overcoming the Topological Difficulty

First we introduce some notations. Let M be a K3 surface with a nowhere zero holomorphic 2-form ϕ. For a d-closed 2-form α on M we denote by $[\alpha]$ the cohomology class defined by α. Denote by $H^{1,1}_{\mathbb{R}}(M)$ the set of all

$a \in H^2(M, \mathbb{R})$ such that a has zero cup product with $[\phi]$ (and therefore also with $[\bar{\phi}]$ due to the real character of a). For $a \in H^2(M, \mathbb{R})$ we denote by $a^{\#}_M \in H^{1,1}_{\mathbb{R}}(M)$ the orthogonal projection of a onto $H^{1,1}_{\mathbb{R}}(M)$ with respect to the quadratic form defined by the cup product. That is, $a = c[\phi] + a^{\#}_M + \bar{c}[\bar{\phi}]$ for some complex number c. By the positive cone of $H^{1,1}_{\mathbb{R}}(M)$ we mean the set of all elements of $H^{1,1}_{\mathbb{R}}(M)$ whose cup product with itself is positive. The positive cone has two components.

(3.1) **Proposition.** *Let M be a* K3 *surface and θ be a continuous d-closed real 2-form on M whose $(1,1)$-component is positive definite everywhere on M. Then $[\theta]^{\#}_M$ belongs to the positive cone of $H^{1,1}_{\mathbb{R}}(M)$ and its value at the homology class of any effective divisor of M is positive. Moreover, if M is Kähler, then $[\theta]^{\#}_M$ lies in the same component of the positive cone of $H^{1,1}_{\mathbb{R}}(M)$ as any Kähler class of M.*

Proof. Let ϕ be a holomorphic 2-form on M with $\int_M \phi \wedge \bar{\phi} = 1$. Let $Q(\cdot, \cdot)$ be the quadratic form on $H^2(M, \mathbb{C})$ defined by the cup-product. Let $\theta = \alpha + \omega + \bar{\alpha}$ so that α is a $(2,0)$-form on M and ω is a $(1,1)$-form on M. (Note that in general α and ω are not closed.) By the definition of $[\theta]^{\#}_M$ we have $[\theta] = c[\phi] + [\theta]^{\#}_M + \bar{c}[\bar{\phi}]$ for some complex number c.

Let D be an effective divisor of M and let d be its cohomology class. Since d is orthogonal to both $[\phi]$ and $[\bar{\phi}]$ with respect to the quadratic form Q, it follows that

$$Q([\theta]^{\#}_M, d) = Q([\theta], d) = \int_D \theta = \int_D \alpha + \omega + \bar{\alpha} = \int_D \omega > 0,$$

because $\int_D \alpha$ and $\int_D \bar{\alpha}$ vanish due to type considerations and ω is positive definite at every point of M.

We now prove that $[\theta]^{\#}_M$ belongs to the positive cone of $H^{1,1}_{\mathbb{R}}(M)$. From

$$\begin{aligned} Q([\theta], [\theta]) &= Q(c[\phi] + [\theta]^{\#}_M + \bar{c}[\bar{\phi}], c[\phi] + [\theta]^{\#}_M + \bar{c}[\bar{\phi}]) \\ &= 2c\bar{c}\,Q([\phi], [\bar{\phi}]) + Q([\theta]^{\#}_M, [\theta]^{\#}_M) \end{aligned}$$

it follows that

$$\begin{aligned} Q([\theta]^{\#}_M, [\theta]^{\#}_M) &= -2c\bar{c}\,Q([\phi], [\bar{\phi}]) + Q([\theta], [\theta]) \\ &= -2c\bar{c} + Q([\theta], [\theta]) \\ &= -2c\bar{c} + \int_M \theta \wedge \theta \\ &= -2c\bar{c} + \int_M (\alpha + \omega + \bar{\alpha}) \wedge (\alpha + \omega + \bar{\alpha}) \\ &= -2c\bar{c} + 2\int_M \alpha \wedge \bar{\alpha} + \int_M \omega \wedge \omega \\ &> 2(-c\bar{c} + \int_M \alpha \wedge \bar{\alpha}), \end{aligned}$$

because ω is positive definite at every point of M. To finish the proof that $[\theta]^{\#}_M$ belongs to the positive cone of $H^{1,1}_{\mathbb{R}}(M)$, it suffices to show that $\int_M \alpha \wedge \bar{\alpha} \geqq c\bar{c}$. It

follows from $Q([\phi],[\bar{\phi}])=1$ and $Q([\theta]_M^{\#},[\bar{\phi}])=Q([\bar{\phi}],[\bar{\phi}])=0$ that

$$c=Q([\theta],[\bar{\phi}])=\int_M \theta\wedge\bar{\phi}=\int_M \alpha\wedge\bar{\phi}.$$

For any choice of complex numbers λ, μ,

$$\int_M (\lambda\alpha+\mu\phi)\wedge\overline{(\lambda\alpha+\mu\phi)}\geqq 0$$

because $\lambda a+\mu\phi$ is a $(2,0)$-form on M. Hence the quadratic form

$$\lambda\bar{\lambda}\int_M \alpha\wedge\bar{\alpha}+\lambda\bar{\mu}\int_M \alpha\wedge\bar{\phi}+\bar{\lambda}\mu\int_M \bar{\alpha}\wedge\phi+\mu\bar{\mu}\int_M \phi\wedge\bar{\phi}$$

in λ and μ is positive semidefinite. It follows that

$$c\bar{c}=|\int_M \alpha\wedge\bar{\phi}|^2\leqq(\int_M \alpha\wedge\bar{\alpha})(\int_M \phi\wedge\bar{\phi})=\int_M \alpha\wedge\bar{\alpha}.$$

It remains to prove that, if M is Kähler, then $[\theta]_M^{\#}$ lies in the same component of the positive cone of $H_{\mathbb{R}}^{1,1}(M)$ as any Kähler class of M. Assume that M is Kähler and ξ is a Kähler form of M. For any t in the unit interval $[0,1]$, $t\xi+(1-t)\theta$ is a continuous d-closed real 2-form on M whose $(1,1)$-component is positive definite everywhere on M. By the preceding argument $t[\xi]+(1-t)[\theta]_M^{\#}=[t\xi+(1-t)\theta]_M^{\#}$ belongs to the positive cone of $H_{\mathbb{R}}^{1,1}(M)$ for every t in $[0,1]$. Hence $[\theta]_M^{\#}$ lies in the same component of the positive cone of $H_{\mathbb{R}}^{1,1}(M)$ as $[\xi]$. Q.E.D.

(3.2) We need the following refined version of Todorov's result on the surjectivity of the period map for Kähler K3 surfaces. This refined version was mentioned without proof in Todorov's paper [18, Th. 3]. A complete proof was given by Looijenga [10]. Let X be a differentiable manifold which is diffeomorphic to a K3 surface. Let a be a nonzero element of $H^2(X,\mathbb{C})$ such that the cup product of a with itself is zero and the cup product of a with $\bar{a}$ is positive. Let Δ be the set of all elements c of $H^2(X,\mathbb{Z})$ such that the cup product of c with a is zero and the cup product of c with itself is equal to -2. Let b be an element of $H^2(X,\mathbb{R})$ such that the cup product of b with a is zero, the cup product of b with itself is positive, and the cup product of b with every c in Δ is nonzero. Then there exists an isomorphism τ of $H^2(X,\mathbb{Z})$ compatible with the cup product and there exists a Kähler K3 structure on X with a Kähler form ω and a nowhere zero holomorphic 2-form ϕ such that the isomorphism of $H^2(X,\mathbb{C})$ induced by τ maps $[\phi]$ to a and maps $[\omega]$ to b.

Actually for the purpose of this paper what we need is only the special case with the additional assumption that a is known to be the cohomology class of a 2-form on X which is holomorphic with respect to some K3 structure M on X and $b=[\theta]_M^{\#}$ for some continuous d-closed real 2-form θ on X whose $(1,1)$-component with respect to the complex structure M is everywhere positive definite. In view of Prop. (3.1) this special case follows quite easily from the Hitchin construction and the ampleness criterion of Nakai-Moisezon.

We are now ready to prove the main theorem of this paper.

(3.3) **Theorem.** *Every* K3 *surface is Kähler.*

Proof. Let M be a K3 surface. Let X be the underlying differentiable manifold. Let ϕ be a nowhere zero holomorphic 2-form on M. By Prop. (1.6) there exists a continuous d-closed real 2-form θ on X whose $(1,1)$-component with respect to the complex structure M is positive definite everywhere. By (3.2) there exists an isomorphism τ of $H^2(X,\mathbb{Z})$ and there exists a Kähler K3 structure N on X with a Kähler form ω and a nowhere zero holomorphic 2-form ψ such that the isomorphism $\tilde{\tau}$ of $H^2(X,\mathbb{C})$ induced by τ maps $[\psi]$ to $[\phi]$ and maps $[\omega]$ to $[\theta]^{\#}_M$ (because every class in $H^2(X,\mathbb{Z})\cap H^{1,1}_{\mathbb{R}}(M)$ whose cup product with itself is -2 comes either from an effective divisor on M or from the negative of an effective divisor on M, and by Prop. (3.1) the value of $[\theta]^{\#}_M$ on the homology class of an effective divisor of M is positive).

By the local Torelli theorem of Andreotti-Weil [6, p. 783, Th. 17], every K3 surface can be deformed to give a complete holomorphic family of K3 surfaces effectively parametrized by their periods. Hence from M and N we obtain two holomorphic families $\{M_s\}_{s\in U}$, $\{N_s\}_{s\in U}$ of K3 structures on X, where U is an open neighborhood of the origin 0 of $\mathbb{C}^{20}$, such that

i) $M_0 = M$, $N_0 = N$,

ii) the two holomorphic families are both complete and effectively parametrized at 0,

iii) for every $s\in U$ the isomorphism $\tilde{\tau}$ of $H^2(X,\mathbb{C})$ maps the cohomology class of a holomorphic 2-form on N_s to the cohomology class of a holomorphic 2-form on M_s.

By continuity arguments there exists an open neighborhood W of 0 in U such that for $s\in W$ the $(1,1)$-component of θ (respectively ω) with respect to the complex structure M_s (respectively N_s) is everywhere positive definite.

There exists a dense subset A of W such that for $s\in A$ the K3 surfaces M_s and N_s are both projective algebraic (because a K3 surface Y is projective algebraic if and only if there exists an element in $H^2(Y,\mathbb{Z})\cap H^{1,1}_{\mathbb{R}}(Y)$ whose cup product with itself is positive). It follows from Prop. (3.1) that for $s\in A$ the class $[\theta]^{\#}_{M_s}$ (respectively $[\omega]^{\#}_{N_s}$) lies in the same component of the positive cone of $H^{1,1}_{\mathbb{R}}(M_s)$ (respectively $H^{1,1}_{\mathbb{R}}(N_s)$) as any Kähler class of M_s (respectively N_s) and the value of $[\theta]^{\#}_{M_s}$ (respectively $[\omega]^{\#}_{N_s}$) at the homology class of any effective divisor of M_s (respectively N_s) is positive. Since for any K3 surface Y any element in $H^2(Y,\mathbb{Z})\cap H^{1,1}_{\mathbb{R}}(Y)$ whose cup product with itself is -2 is always the cohomology class of either an effective divisor on Y or the negative of an effective divisor on Y and since for $s\in A$ the isomorphism $\tilde{\tau}$ of $H^2(X,\mathbb{C})$ maps $[\omega]^{\#}_{N_s}$ to $[\theta]^{\#}_{M_s}$, it follows from the Torelli theorem for Kähler (or algebraic K3 surfaces [2, p. 235, Th. 1] (or [14]) that for $s\in A$ there exists a biholomorphic map from N_s to M_s which induces the isomorphism τ on $H^2(X,\mathbb{Z})$. By Prop. (2.1) $M = M_0$ is biholomorphic to N_0 and is therefore Kähler. Q.E.D.

I would like to take this opportunity to correct the following errors in [15]. On line 9 of p. 313, $\langle \Delta\rangle^{\perp}$ should read $\bigcup_{x\in\Delta} x^{\perp}$. On p. 318, line 3 from the bottom, $\langle E_2, q\rangle$ should read $\langle E_2, q'\rangle$. Moreover, on p. 318, line 4 from the bottom, one has to use [2, p. 241, Prop. 2.3] and replace Z by $\bigcup\{x^{\perp} \mid x\in\Delta\cap E^{\perp},\ E\in W\}$, where W is some neighborhood of E_1 in Ω.

References

1. Bishop, E.: Conditions for the analyticity of certain sets. Michigan Math. J. **11**, 289–304 (1964)
2. Burns, D., Rapoport, M.: On the Torelli problem for Kähler K3 surfaces. Ann. Scient. Éc. Norm. Sup. **8**, 235–274 (1975)
3. Dunford, N., Schwartz, J.T.: Linear Operators, Vol. I. New York: Interscience 1958
4. Harvey, F.R., Lawson, B.: An intrinsic characterization of Kähler manifolds.
5. Hitchin, N.: Compact four-dimensional Einstein manifolds. J. Diff. Geom. **9**, 435–441 (1974)
6. Kodaira, K.: On the structure of compact complex analytic surfaces, I. Amer. J. Math. **86**, 751–798 (1964)
7. Kodaira, K., Morrow, J.: Complex Manifolds. New York: Holt, Rinehart & Winston 1971
8. Kodaira, K., Spencer, D.C.: On deformations of complex analytic structures, III. Stability theorems for complex structures. Ann. of Math. **71**, 43–76 (1960)
9. Kulikov, V.: Epimorphicity of the period mapping for surfaces of type K3. Uspehi Mat. Nauk **32**, (No. 4) 257–258 (1977)
10. Looijenga, E.: A Torelli theorem for Kähler-Einstein K3 surfaces. Lecture Notes in Math. Vol. 892, pp. 107–112. Berlin-Heidelberg-New York: Springer 1980
11. Looijenga, E., Peters, C.: Torelli theorems for Kähler K3 surfaces. Compositio Math. **42**, 145–186 (1981)
12. Miyaoka, Y.: Kähler metrics on elliptic surfaces. Proc. Japan Acad. **50**, 533–536 (1974)
13. Morrey, C.B., Jr.: Multiple integrals in the calculus of variations. New York: Springer 1966
14. Piatetskii-Shapiro, I.I., Shafarevitch, I.R.: A Torelli theorem for algebraic surfaces of type K3. Izv. Akad. Nauk SSSR **35**, 547–572 (1971); (English Translation) Math. USSR Izvestija **5**, 547–588 (1971)
15. Siu, Y.-T.: A simple proof of the surjectivity of the period map of K3 surfaces. Manuscripta Math. **35**, 311–321 (1981)
16. Sommese, A.J.: Quaternionic manifolds. Math. Ann. **212**, 191–214 (1975)
17. Sullivan, D.: Cycles for the dynamical study of foliated manifolds and complex manifolds. Invent. Math. **36**, 225–255 (1976)
18. Todorov, A.N.: Applications of the Kähler-Einstein-Calabi-Yau metric to moduli of K3 surfaces. Invent. Math. **61**, 251–265 (1980)
19. Yau, S.-T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation, I. Comm. Pure Applied Math. **31**, 339–411 (1978)

Oblatum 10-I-1983

Invent. math. 73, 151-156 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Theta Series of Ternary and Quaternary Quadratic Forms

J.S. Hsia* and D.C. Hung

Department of Mathematics, The Ohio State University, 231 W. 18th Avenue, Columbus, OH 43210, USA

In the theory of integral quadratic forms or in the allied subject of modular forms very little seems to be known about the linear independence of theta series associated to positive definite forms. For a system of inequivalent positive binary quadratic forms with a fixed fundamental discriminant, Hecke [4] showed that their theta series are independent. However, for positive forms in three or more variables, even in a fixed genus, there does not appear to be any non-trivial general result of a similar nature in that certain explicitly described forms in the genus yield linearly independent theta series. We consider here even positive definite ternary quadratic forms of discriminant $2p$ and quaternary quadratic forms of discriminant p where p is a prime congruent to 1 mod 4. For a fixed such p these forms belong to a single genus $G := G(3, 2p)$, resp. $G(4, p)$. Let G' denote the subset consisting of those forms in G which represent 2. In [6] Kitaoka observed that the number of form-classes in $G'(4, p)$ is one-half the dimension of the space $\mathcal{M}\left(2, \Gamma_0(p), \left(\frac{*}{p}\right)\right)$ of automorphic forms of weight 2 and character $\left(\frac{*}{p}\right)$ with respect to the group $\Gamma_0(p)$, and asked whether the theta series of f together with its primitive adjoint (i.e., reciprocal) form constitute an explicit basis for $\mathcal{M}\left(2, \Gamma_0(p), \left(\frac{*}{p}\right)\right)$ where f ranges over the classes in $G'(4, p)$. This question had also entered in some of the recent works of Eichler, e.g., [2, 3]. Recently, Kitaoka has kindly communicated to one of us that, with the aid of a computer, he and Nashiro have found a counter-example at $p = 389$. For this discriminant, the class number of all the even quaternary lattices is 18 with 17 classes belonging to $G'(4, 389)$. There is a single linear relation involving 10 of the 17 classes, and if we replace one class by the 18th class, together with all their adjoints, we obtain an explicit basis for $\mathcal{M}\left(2, \Gamma_0(389), \left(\frac{*}{389}\right)\right)$. However, although in general the linear indepen-

* Research supported in part by NSF Grants MCS 80-02985 and MCS 82-02065.

Offprint requests to: J.S. Hsia

dence of theta series coming from all the classes in $G'(4,p)$ is false, certain explicit subsets of these series are found to be independent (Theorem 2). Also, in the case of ternary quadratic forms the theta series associated to all the classes in $G'(3,2p)$ are linearly independent (Theorem 3)[1]. Moreover, with regard to Siegel modular forms we show that the theta series of degree two arising from all the classes of forms in $G'(4,p)$ are independent (Theorem 1) even though the corresponding degree one statement does not hold.

While it may be more natural to prove linear independence results of theta series by analytic methods this does not seem to have been done. On the other hand, arithmetic methods have been used recently (see [8, 1, 5]) to reprove, and indeed to extend, representation-theoretic results of positive definite quadratic forms which previously had yielded only to analytic arguments ([12, 10] e.g.). Our proof here is again purely arithmetic and uses only results from quadratic forms theory. Although this method should be applicable to other genera as well as to theta series over number fields, it is still essential that one requires the technical features of the forms in these genera.

We wish to thank the referee for his careful reading of the original version and for his many valuable suggestions which resulted in a considerable streamlining of the proof of Theorem 1 which is the focal point of the whole argument.

§1. Roots Systems and Unit Groups

Throughout this paper we shall adopt the geometric language of lattices instead of forms. Thus, by a lattice over $\mathbb{Z}$ (resp. $\mathbb{Z}_p$) we mean a free abelian group (resp. free $\mathbb{Z}_p$-module) together with a $\mathbb{Z}$- (resp. $\mathbb{Z}_p$-) values non-degenerate symmetric bilinear function $B(x,y)$. An even positive definite integral quadratic form in n variables with discriminant D is then a $\mathbb{Z}$-lattice of rank n having a basis $\{u_1,\ldots,u_n\}$ with respect to which the matrix $(B(u_i,u_j))$ is integral, has even diagonal entries, and whose determinant is D.

Within the genus $G(4,p)$ of positive even quaternary lattices of discriminant p it is convenient to partition the classes according to their roots systems. A minimal vector x in a lattice is a vector of length $Q(x)=B(x,x)=2$. The roots system R_L of a lattice L is the set of all minimal vectors of L. Using the structural results for lattices in $G(4,p)$ as given in [6] it is readily shown that the only possible roots systems are the following types: ϕ, A_1, $A_1\oplus A_1$, A_2, $A_1\oplus A_2$, A_3, A_4 according to the number $a_L(2)$ of minimal vectors in L being resp. 0, 2, 4, 6, 8, 12, 20. The latter case can only occur at $p=5$, and since the class number of $G(4,5)$ is unity we shall henceforth ignore this case. Since the orthogonal transformations of a lattice L in $G(4,p)$ are generated by the symmetries of L and ± 1 the unit groups $O(L)$ can be easily determined. We just state the results for later use and skip the details.

Proposition 1.1. *Let $L\in G(4,p)$ with $p>5$. Then, the order of the unit group $|O(L)|$ is 2, 4, 8, 12, 24, 48 according to roots system R_L type being ϕ, A_1, $2A_1$, A_2, $A_1\oplus A_2$, A_3 respectively.*

[1] Kitaoka has independently proved Theorem 3 and part (a) of Theorem 2

Similarly, for ternary lattices in $G(3,2p)$ we have:

Proposition 1.2. *Let $L\in G(3,2p)$. Then, the roots system R_L can only be ϕ, A_1, A_2. The unit group $O(L)$ is generated by symmetries of L and ± 1, and its order is respectively* 2, 4, 12.

Remark 1.3. Theorem 3 in §3 remains valid for positive even ternary lattices of discriminant $2p$ where $p\equiv 3 \pmod 4$. While the unit groups in these cases are still generated by integral symmetries and ± 1, the lattices, however, may now be decomposable. The possible roots systems are: ϕ, A_1, $A_1\oplus A_1$, A_2 and $A_1\oplus A_2$ with unit groups having orders: 2, 4, 8, 12 and 24 respectively, and the latter occurs only at $p=3$.

§2. Theta Series of Degree Two

The aim of this section is to prove that although the ordinary theta series attached to quaternary lattices from $G'(4,p)$ are not linearly independent (see example in introduction), their counterparts in degree two theta series do form an independent set. The proof is divided into several steps.

1) Let $L_1,\ldots,L_h$ be a full set of non-isometric lattices in $G'(4,p)$. Fix minimal vectors e_i in L_i and denote by K_i the orthogonal complement of e_i in L_i. So, K_i has discriminant $2p$. By local integral theory (Chap. IX, [9]), choose vectors $x_{j,2}\in(K_i)_2$, $x_{j,p}\in(K_i)_p$, $(j=1,2)$ so that $d(x_{1,2},x_{2,2})\in\mathbb{Z}_2^\times$ and $d(x_{1,p},x_{2,p})\in\mathbb{Z}_p^\times$. By Lemma 1.6, [5], there exist global vectors x_1, $x_2\in K_i$ that approximate the local vectors at the respective primes, and $d(x_1,x_2)=q_i$ for some prime q_i. Set $J_i=\mathbb{Z}x_1+\mathbb{Z}x_2$.

2) Fix i and let $L=L_i$, $e=e_i$, $J=J_i$ and $q=q_i$. If $\langle f\rangle$ denotes the orthogonal complement of $\langle e\rangle\perp J$ in L, then $Q(f)=2pq$. Put $\Lambda=\langle e\rangle\perp J\perp\langle f\rangle\subset L$. There exist precisely two lattices in $G(4,p)$ which contain Λ, and the two are interchanged by the symmetry S_f. To see this, suppose L' is any lattice in $G(4,p)$ containing Λ. Then, $L_r=L'_r$ for $r\neq 2,q$. At $r=2$, we have $L_2=J_2\perp M$, $L'_2=J_2\perp M'$ for some 2-adic unimodular lattices M, M', both containing e. From a direct calculation, one sees that $M=M'$. (Equivalently, $\langle e\rangle\perp J_2$ is a 2-adic "characteristic sublattice" of L_2 in the sense of [7].) Now, look at $r=q$. $\Lambda_q=\langle e\rangle\perp\langle u\rangle\perp\langle v\rangle\perp\langle f\rangle$ where $Q(u)\in\mathbb{Z}_q^\times$, $Q(v)\in q\mathbb{Z}_q^\times$. Since $L'_q\supset\Lambda_q$, $L'_q=\Lambda_q+\left\langle\frac{1}{q}(v+af)\right\rangle$ with $a\in\mathbb{Z}_q$, and $\frac{Q(v)}{q}+\frac{a^2Q(f)}{q}\equiv 0\pmod q$. If $a \bmod q$ is the solution corresponding to L_q, then the only other one is $-a$, which corresponds to $S_f(L_q)$.

3) If $\phi\colon J_i\to L_j$ is an isometric embedding such that L_j contains a minimal vector e orthogonal to $\phi(J_i)$, then ϕ can be extended to an isometry $L_i\xrightarrow{\sim}L_j$. In particular, $i=j$. For, we may assume ϕ to be the inclusion map and $e=e_i$ by Witt's theorem. Now, both L_i and L_j contain a quaternary sublattice $\langle e_i\rangle\perp J_i\perp\langle f\rangle$ of discriminant $4q_i^2p$. By step 2) L_i is isometric to L_j.

4) If 2^{n_j} denotes the exact power of 2 in $|O(L_j)|$, then Proposition 1.1 gives $n_j=2$ for L_j being of roots system type A_1 or A_2; $=3$ for $2A_1$ or $A_1\oplus A_2$; $=4$

for A_3. Clearly, $O(L_j)$ acts on the set of isometric embeddings of J_i into L_j. We wish to measure the sizes of the $O(L_j)$-orbits. Let $\phi: J_i \to L_j$, $i \neq j$, and H_ϕ be the stabilizer of ϕ. If $\sigma \in O(L_j)$ is an element of order 2 then $\sigma = -1_W \perp 1_U$. Since $O(L_j)$ is generated by symmetries and ± 1, it follows readily that either U or W is spanned by pairwise orthogonal minimal vectors. But for $\sigma \in H_\phi$, $U \supset \mathbb{Q}(\phi(J_i))$ and so cannot be spanned in this way. Thus, it must be W, contradicting step 3). The same argument works for the case when $i=j$ and ϕ does not lie in the orbit of inclusion. When ϕ does lie in the orbit of inclusion, $S_{e_i} \in H_\phi$ and is the only element of order 2. Summarizing, we have: the number of elements in an $O(L_j)$-orbit is divisible by 2^{n_j} except for the single case for the orbit of the inclusion map when $i=j$ in which case this number is divisible exactly by 2^{n_j-1}.

Suppose a_{ij} is the number of isometric embeddings of J_i into L_j. We therefore have:

$$a_{ij} \equiv 0 \bmod 2^{n_j} \quad \text{if } i \neq j,$$

$$a_{jj} \equiv 0 \bmod 2^{n_j-1}, \quad a_{jj} \not\equiv 0 \bmod 2^{n_j}.$$

5) Since theta series (of degree two) of positive quadratic forms are integral Siegel modular forms, their linear independence over $\mathbb{Z}$ already implies linear independence over $\mathbb{C}$. Suppose there is a non-trivial integral linear relation $\sum c_j \theta_{L_j}^{(2)}(Z) = 0$. We may assume that the greatest common divisor of the coefficients is unity. If evaluated at the binary lattice J_i we have: $\sum_j c_j a_{ij} = 0$. Therefore, upon considering this equation modulo various congruences we obtain, using step 4), the following:

$$\begin{aligned}
\bmod 2^2\text{:}\ & c_j \equiv 0 \bmod 2 \quad \text{for } n_j = 2,\\
\bmod 2^3\text{:}\ & c_j \equiv 0 \bmod 2 \quad \text{for } n_j = 3,\\
& c_j \equiv 0 \bmod 4 \quad \text{for } n_j = 2,\\
\bmod 2^4\text{:}\ & c_j \equiv 0 \bmod 2 \quad \text{for } n_j = 4.
\end{aligned}$$

This is a contradiction, and completes the proof of the announced result.

Theorem 1. *The theta series $\theta_L^{(2)}(Z)$ of degree two for lattices L coming from the classes of even quaternary lattices of discriminant p and representing 2 are linearly independent.*

§3. Theta Series of Degree One

The objective of this section is to present two linear independence results for the theta series attached to certain positive definite even ternary and quaternary quadratic forms. Since our method is similar to the proof of Theorem 1, we shall merely state the necessary changes instead of giving detailed proofs.

Theorem 2. *The theta series $\theta_L(z)$ for lattices L coming from the classes in $G'(4,p)$ which have roots system of type (a) $R_L \supset 2A_1$ or (b) $R_L \supset A_2$ are linearly independent.*

Theorem 3. *The theta series $\theta_L(z)$ for lattices L coming from the classes in $G'(3,2p)$ are linearly independent.*

Remarks on the Proofs

(i) In Theorem 2 the vectors e_i are replaced by the roots system R_i of type $2A_1$ or A_2 respectively, J_i now has rank 1 and discriminant $2q_i$. In steps 2) and 3) above there will be four lattices in $G(4,p)$ containing $R_i \perp J_i \perp \langle f \rangle$ which are transitively permuted by the group generated by S_f and an automorphism of R_i not belonging to the Weyl group. For part (a), the Fourier coefficients in step 4) now read: $a_{ij} \equiv 0 \bmod 2^{n_j-1}$ for $i \neq j$, $a_{jj} \equiv 0 \bmod 2^{n_j-2}$ but $a_{jj} \not\equiv 0 \bmod 2^{n_j-1}$, and the congruences in step 5) are taken $\bmod 2^2$ and $\bmod 2^3$. As for part (b), congruences mod 3 are used.

(ii) In Theorem 3 when the binary K_i is non-ambigous there will be "twins" L_i, L_i', both containing K_i. Since these twins are not isometric, step 3) will no longer apply, and the same argument would yield only roughly half the number of classes in $G'(3,2p)$ with linearly independent theta series. To get all the classes to be independent, an appeal to the dual lattices $L_i^\#$ seems inevitable. But, the usual transformation formula shows that the linear independence for the L_i's or the $L_i^\#$'s are equivalent. See Chap. VII, [11]. Now, in steps 2) and 3) L_i becomes $L_i^\#$ which has scale $\frac{1}{2p}\mathbb{Z}$ and discriminant $\frac{1}{2p}$, e_i remains the same, K_i is binary with scale $\frac{1}{2p}\mathbb{Z}$ and discriminant $\frac{1}{p}$, J_i is unary of discriminant $\frac{q_i}{2p}$ with q_i prime, and both $L_i^\#$ and $L_j^\#$ contain a ternary sublattice $\langle e_i \rangle \perp J_i \perp \langle f \rangle$ of discriminant $\frac{2q_i^2}{p}$, and the rest proceeds as before.

References

1. Cassels, J.W.S.: Rational quadratic forms. London: Academic Press 1978
2. Eichler, M.: Über ambige definite quaternäre quadratische Formen von Primzahlendiskriminante. Comme. Pure & Appl. Math. **29**, 623-647 (1976)
3. Eichler, M.: Theta functions over $\mathbb{Q}$ and over $\mathbb{Q}(\sqrt{q})$. Modular functions of one variable VI. Lectures notes in math., vol. 627. 197-225. Berlin-Heidelberg-New York: Springer 1977
4. Hecke, E.: Mathematische Werke. Göttingen: Vandenhoeck & Ruprecht 1970
5. Hsia, J.S., Kitaoka, Y., Kneser, M.: Representations of positive definite quadratic forms. J. reine angew. Math. **301**, 132-141 (1978)
6. Kitaoka, Y.: Quaternary even positive definite quadratic forms of prime discriminant. Nagoya Math. J. **52**, 147-161 (1973)

7. Kitaoka, Y.: Representations of quadratic forms and their application to Selberg's zeta functions. Nagoya Math. J. **63**, 153–162 (1976)
8. Kneser, M.: Quadratische Formen. Vorlesungsausarbeitung. Göttingen (1973/4)
9. O'Meara, O.T.: Introduction to quadratic forms. Berlin-Heidelberg-New York: Springer 1963
10. Raghavan, S.: Modular forms of degree n and representations by quadratic forms. Ann. of Math. **70**, 446–477 (1959)
11. Serre, J-P.: A Course in arithmetic. Berlin-Heidelberg-New York: Springer 1973
12. Tartakovskij, V.: Die Gesamtheit der Zahlen, die durch eine positive quadratische Form $F(x_1, \ldots, x_s)$ $(s \geqq 4)$ darstellbar sind. Izv. Akad. Nauk SSSR **7**, 111–122, 165–196 (1929)

Oblatum 30-VII-1982

Invent. math. 73, 157–164 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Sur l'anneau canonique de certaines variétés de dimension 3

X. Benveniste

Centre de Mathématiques, Ecole Polytechnique, F-91128 Palaiseau, France

Introduction

Le but de cet article est de montrer la finitude de l'anneau canonique de certaines variétés de dimension 3. Cet article doit beaucoup à une prépublication de Y. Kawamata, qui par ailleurs publiera une autre version du résultat que nous annonçons ici.

Nous tenons à remercier E. Viehweg pour ses encouragements, ses remarques et pour nous avoir indiqué les généralisations possibles du résultat. Remercions aussi A. Beauville pour ses précieux conseils.

Ici nous donnerons une version du résultat dans laquelle figure la notion de décomposition de Zariski d'un diviseur effectif sur une variété de dimension 3 projective et lisse sur $\mathbb{C}$. Nous faisons cela car, sous cette forme, la finitude de l'anneau canonique d'une variété de type général, de dimension 3, lisse et projective sur $\mathbb{C}$, est équivalente à des conditions numériques assez simples sur la décomposition de Zariski du diviseur canonique d'un modèle birationnel de celle-ci. Elle a aussi l'avantage de montrer que si une variété de type général de dimension 3 admet un modèle birationnel dont les singularités sont canoniques au sens de M. Reid [7] (et [8]), et dont le faisceau dualisant est de degré positif sur toutes les courbes de ce modèle, l'anneau canonique d'une telle variété est de type fini sur $\mathbb{C}$. On peut espérer que les méthodes de S. Mori [6] permettront de prouver l'existence d'un tel modèle.

En reprenant les notations de [1] le résultat que nous démontrons est:

Théorème 0-1. *Soit X une variété lisse et projective de dimension 3 sur $\mathbb{C}$; soit K le diviseur canonique de X. Supposons que X soit de type général et que la classe k de K dans $NS(X)\otimes_{\mathbb{Z}}\mathbb{Q}$ admette une décomposition de Zariski relativement à la famille finite $(w_i)_{i\in L}$ des classes des composantes fixes des systèmes linéaires pluricanoniques*

$$k=u+\sum_{i\in L}\tau_i w_i$$

vérifiant les conditions suivantes:

(i) $u \in NS(X) \otimes \mathbb{Q}$;

(ii) pour tout $i \in L$, $w_i \cdot u \cdot u = 0$ *et* $\tau_i > 0$.

Alors l'anneau canonique de X est de type fini sur $\mathbb{C}$.

Nous obtenons le

Corollaire 0-1. *Soit X une variété lisse et projective de dimension* 3 *sur* $\mathbb{C}$. *Supposons X de type général et soit K le diviseur canonique de X. Supposons que K soit numériquement positif. Alors l'anneau canonique est de type fini sur* $\mathbb{C}$.

Corollaire 0-2. *Soit X une variété de type général de dimension* 3 *lisse et projective sur* $\mathbb{C}$. *Soit K son diviseur canonique; les conditions suivantes sont équivalentes:*

(i) *L'anneau* $\bigoplus_{n \geqq 0} H^0(X, \mathcal{O}_X(nK))$ *est de type fini sur* $\mathbb{C}$.

(ii) *X admet un modèle birationnel à singularités canoniques au sens de* [7] *dont le faisceau dualisant est de degré positif sur toutes les courbes de ce modèle.*

(iii) *Il existe une variété Y et un morphisme* $\varepsilon: Y \to X$ *composé d'une suite finite d'éclatements de sous-variétés lisses de dimension* $\leqq 1$ *tels que la classe* k_Y *du diviseur canonique de Y dans* $NS(Y)$ *vérifie les conditions suivantes:*

k_Y *admet une décomposition de Zariski relativement à la famille finie* $(v_i)_{i \in I}$ *des classes des composantes fixes des systèmes linéaires pluricanoniques et l'on a:*

a) $k_Y = u + \sum_{i \in I} \rho_i v_i$

b) $u \in NS(X) \otimes \mathbb{Q}$

c) *pour tout* $i \in I$ *tel que* $\rho_i > 0$, $w_i \cdot u \cdot u = 0$.

Dans tout ce qui suit, nous noterons pour tout réel x respectivement $[x]$ et $\{x\}$ la partie entière et le plus petit entier $\geqq x$. Le symbole $\equiv$ désigne l'équivalence rationnelle entre diviseurs.

1) Notion de décomposition de Zariski en dimension 3

Reprenons la définition de [1]; dans tout ce qui suit X est une variété lisse et projective de dimension 3 sur $\mathbb{C}$.

Définition 1-1. *On dit qu'une classe* $d \in NS(X) \otimes_{\mathbb{Z}} \mathbb{Q}$ *est numériquement positive si pour toute courbe irréductible A on a* $d \cdot A \geqq 0$.

Nous désignerons par $NS^+_{\mathbb{R}}$ l'enveloppe convexe dans $NS(X) \otimes_{\mathbb{Z}} \mathbb{R}$ de l'ensemble des classes de diviseurs effectifs dans $NS(X)$.

Définition 1-2. *Soit* $(v_i)_{i \in I}$ *une famille finie de classes de diviseurs irréductibles dans* $NS(X)$. *On dit qu'un élément* $x \in NS^+_{\mathbb{R}}$ *admet une décomposition de Zariski relativement à la famille* $(v_i)_{i \in I}$, *s'il existe* $u \in NS^+_{\mathbb{R}}$ *et des réels* $\rho_i \geqq 0$ *pour* $i \in I$ *tels que:*

(i) $x = u + \sum_{i \in I} \rho_i v_i$

(ii) *u est numériquement positif*

(iii) *pour chaque* $w \in \sum_{i \in I} \mathbb{R}_+ \rho_i v_i - \{0\}$, *la classe* $u+w$ *n'est pas numériquement positive.*

2) Quelques lemmes de préparation

Dans tout ce qui suit X désigne une variété de type général de dimension 3 projective et lisse sur $\mathbb{C}$ vérifiant les hypothèses du théorème 0-1.

Nous adopterons les notations suivantes de Fujita [2]: Pour un diviseur effectif D sur X, soit $B(D)$ le lieu fixe de $|D|$. Notons aussi $SB(D) = \bigcap_{n \geqq 0} B(nD)$ le lieu fixe stable de D.

Nous allons montrer la

Proposition 2-1. *Il existe un modèle birationnel* Y *de* X, *tel que si* K_Y *est le diviseur canonique de* Y *les conditions suivantes soient remplies:*

(i) *Il existe un diviseur* P *numériquement positif vérifiant* $P^3 > 0$, *un entier* $d > 0$, *une famille finie* $(E_i)_{i \in I}$ *de diviseurs lisses à croisements normaux, et une famille de rationnels positifs* $(\rho_i)_{i \in I}$ *tels que dans* $NS(Y) \otimes_{\mathbb{Z}} \mathbb{Q}$ *on ait:*

$$K_Y = P/d + \sum_{i \in I} \rho_i E_i$$

et que pour tout $n \geqq 1$: $H^0(Y, \mathcal{O}_Y(nP)) \xrightarrow{\sim} H^0(Y, \mathcal{O}_Y(ndK_Y))$.

(ii) *Il existe un entier* $m > 0$, *une famille d'entiers positifs* $(u_i)_{i \in I}$ *et un diviseur* M *tels que:*

1) $mP \equiv M + \sum_{i \in I} u_i E_i$,

2) *le système linéaire* $|M|$ *est sans point base,*

3) *pour chaque* $i \in I$ *tel que* $u_i > 0$, *on a* $E_i \subset SB(P)$.

(iii) *Il existe un entier* $b > 0$ *et une famille d'entiers* $> 0 (e_i)_{i \in I}$ *tels que pour toute famille* $(\varepsilon_i)_{i \in I} \in \{0, 1\}^I$, *le diviseur*

$$bmP - \sum_{i \in I} (e_i + \varepsilon_i) E_i \quad \textit{soit ample.}$$

Démonstration. Pour chaque $i \in L$, soit W_i un diviseur dont la classe est celle de w_i. Soit d un entier > 0 tel que pour chaque $i \in L$, $d\tau_i$ soit entier. Posons

$$U \equiv dK - \sum_{i \in L} d\tau_i W_i$$

et remarquons que la classe de U est celle de du dans $NS(X) \otimes_{\mathbb{Z}} \mathbb{Q}$.

Il est facile de voir que la condition (ii) du théorème 0-1 implique que pour $m \geqq 1$ on a: $B(mdK) = B(mU) \cup \bigcup_{i \in L} W_i$; de plus les parties mobiles de $|mdK|$ et de $|mU|$ sont les mêmes.

Soit $(W_i)_{i \in J_1}$ la famille des composantes irréductibles de dimension 2 de $SB(U)$; soit m un entier, H un diviseur ample, $(W_i)_{i \in J_2}$ une famille finie de diviseurs irréductibles sur X, $(h_i)_{i \in J_2}$ une famille d'entiers positifs tels que:

(i) $B(mU)=SB(U)$

(ii) $mU\equiv H+\sum_{i\in J_2} h_i W_i$

D'après les résultats de Hironaka [4], il existe une variété lisse Y, un morphisme birationnel $\varepsilon: Y\to X$, composé d'une suite finie d'éclatements de sous-variétés lisses de dimension $\leqq 1$ tel que la famille $(E_i)_{i\in I}$ des transformés stricts des W_i pour $i\in L\cup J_1\cup J_2$ et des diviseurs exceptionnels de ε soit à croisements normaux et que la partie mobile de $|m\varepsilon^*(U)|$ soit sans point base. Soit M un élément irréductible de cette partie mobile.

Il existe donc deux familles d'entiers positifs $(u_i)_{i\in I}$ et $(d_i)_{i\in I}$ telles que:

$$m\varepsilon^*(U)\equiv M+\sum_{i\in I} u_i E_i$$

$$m\varepsilon^*(U)\equiv \varepsilon^*(H)+\sum_{i\in I} d_i E_i.$$

Posons $P=\varepsilon^*(U)$. Soit K_Y le diviseur canonique de Y; il existe une famille de rationnels positifs $(\rho_i)_{i\in I}$ telle que dans $NS(Y)\otimes_{\mathbf{Z}}\mathbf{Q}$ on ait:

$$K_Y=P/d+\sum_{i\in I}\rho_i E_i.$$

Soit I' le sous-ensemble de I formé des i tels que E_i soit le transformé strict d'un diviseur exceptionnel de ε. D'après les résultats donnés dans [3] sur les éclatements, il existe un entier $b'>0$ et une famille d'entiers strictement positifs $(b'_i)_{i\in I'}$ tels que $b'\varepsilon^*(H)-\sum_{i\in I'} b'_i E_i$ soit ample. D'après les résultats de Serre sur les faisceaux amples (cf. par exemple [3]), il existe un entier $b>0$ et une famille d'entiers strictement positifs $(t'_i)_{i\in I}$ tels que pour toute famille $(\varepsilon_i)_{i\in I}\in\{0,1\}^I$ le diviseur

$$b\varepsilon^*(H)-\sum_{i\in I}(t'_i+\varepsilon_i)E_i \quad \text{soit ample.}$$

Pour avoir la proposition 2-1, il suffit de poser pour chaque $i\in I$, $e_i=b\,d_i+t'_i$.

3) La démonstration du thèoréme 0-1

La démonstration du théorème 0-1 s'appuie sur le théorème d'annulation dû à Kawamata et à Viehweg ([5] et [9]). Nous allons citer celui-ci sous une forme qui nous sera utile.

Théorème 3-1 (Kawamata-Viehweg). *Soient V une variété lisse et projective sur $\mathbf{C}$ de dimension n, $(E_i)_{i\in A}$ une famille finie de diviseurs lisses à croisements normaux, $(a_i)_{i\in A}$ une famille de rationnels positifs strictement inférieurs à 1, et D une classe de* $\mathrm{Pic}(V)\otimes_{\mathbf{Z}}\mathbf{Q}$ *numériquement positive (au sens de la définition* 1-1*) vérifiant $D^n>0$. Soit L un élément de* $\mathrm{Pic}(V)$ *dont la classe dans* $\mathrm{Pic}(V)\otimes_{\mathbf{Z}}\mathbf{Q}$ *s'écrit:*

$$L=D+\sum_{i\in I} a_i E_i.$$

Alors pour tout $p \leqq n-1$, *on a:*

$$H^p(V, \mathcal{O}_V(-L)) = 0.$$

Soit Y la variété introduite dans la proposition 2-1. A présent nous utiliserons les notations de la proposition 2-1 sans y faire référence. Nous appliquerons le théorème 3-1 sur Y pour montrer la

Proposition 3-1. *Supposons que* $SB(P)$, *soit non vide; alors il existe une famille d'entiers positifs* $(m_i)_{i\in I}$, *un entier* $n_0 > 0$, *un élément* $0 \in I$ *tels que pour* $n \geqq n_0$, *le diviseur* $D_n \equiv nP + \sum_{i \neq 0} m_i E_i$ *posséde les propriétés suivantes:*

(i) E_0 *est partie fixe de* $|D_n|$,

(ii) $H^p(Y, \mathcal{O}_Y(D_n)) = H^p(Y, \mathcal{O}_Y(D_n - E_0)) = 0$ *pour* $p > 0$,

(iii) *Soit* K_0 *le diviseur canonique de* E_0. *Il existe une famille* $(c_i)_{i\in I-\{0\}}$ *de rationnels positifs strictement inférieurs à* 1, *un diviseur ample* H, *un entier* $a > 0$ *tels que dans* $NS(E_0) \otimes_{\mathbb{Z}} \mathbb{Q}$ *on ait:*

$$D_n|_{E_0} = K_0 + (H/a + \sum_{i \neq 0} c_i E_i + (n - n_0) P)|_{E_0}.$$

Démonstration: Nous allons supposer $SB(P)$ non vide et procéder en trois étapes:

a) *Quelques calculs arithmétiques.* Soit r un entier > 0; considérons:

$$a(r) = \max\{(ru_i + e_i + 1)/(\rho_i + 1),\ u \in I\}.$$

Soit $E(r)$ l'ensemble des $i \in I$ tels que:

$$a(r) = (ru_i + e_i + 1)/(\rho_i + 1).$$

Il est clair qu'il existe un entier r_0 tel que pour $r \geqq r_0$, les conditions suivantes soient remplies:

(i) $E(r) = E(r_0)$

(ii) pour tout $i \in E(r)$ on a $u_i \neq 0$,

(iii) $1/a(r) < 1/2$.

Soit $r \geqq r_0$ fixé dorénavant et faisons la

Remarque: Si $u_i \neq 0$, *on a* $E_i \subset SB(P)$.

Soit $L(r)$ l'ensemble des $i \in I - E(r)$ tels que

$$(ru_i + e_i + 1)/a(r) - \rho_i \in \mathbb{Z};$$

remarquons que si $i \in E(r)$

$$(ru_i + e_i + 1)/a(r) - \rho_i = 1 \in \mathbb{Z}.$$

Soit $0 \in E(r)$; posons:

1) $s_0 = e_0 + 1$

2) $s_i = e_i$ pour $i \in L(r) \cup E(r) - \{0\}$

3) $s_i = e_i + 1$ pour $i \in I - L(r) \cup E(r)$.

On obtient ainsi les assertions suivantes:

(i) $a(r)=(ru_0+s_0)/(\rho_0+1)$

(ii) pour $i\in I-\{0\}$, on a $(ru_i+s_i)/(\rho_i+1)<a(r)$, et $(ru_i+s_i)/a(r)-\rho_i\notin\mathbb{Z}$.

On en déduit le

Lemme 3-1. *Il existe un rationnel $\varepsilon>0$ et une famille d'entiers positifs $(m_i)_{i\in I}$ tels que:*

(i) $m_0=0$, $(ru_0+s_0)/a(r)-\rho_0=1$, $[(ru_0+s_0)/(a(r)+\varepsilon)-\rho_0]=0$

(ii) *pour tout $i\in I-\{0\}$ on a:*

$$[(ru_i+s_i)/a(r)-\rho_i]=[(ru_i+s_i)/(a(r)+\varepsilon)-\rho_i]=-m_i \quad \text{et} \quad m_i\leqq\{\rho_i\}.$$

b) *La démonstration des assertions* (i) *et* (ii): Posons maintenant $H_0\equiv bmP-\sum_{i\in I}s_iE_i$; rappelons que d'après la proposition 2-1 ce diviseur est ample. Soit n_0 un entier $>m(r+b)/a(r)+1/d$ et $n\geqq n_0$; dans $\mathrm{Pic}(Y)\otimes_{\mathbb{Z}}\mathbb{Q}$ on peut écrire:

$$nP+\sum_{i\neq 0}m_iE_i-K_Y-E_0=(n-m(r+b)/a(r)-1/d)\,P$$

$$+(H_0+rM)/a(r)+\sum_{i\neq 0}((ru_i+s_i)/a(r)+m_i-\rho_i)\,E_i.$$

Si ε est le rationnel introduit au lemme 3-1 on a aussi:

$$nP+\sum_{i\neq 0}m_iE_i-K_Y=(n-m(r+b)/(a(r)+\varepsilon)-1/d)\,P$$

$$+(H_0+rM)/(a(r)+\varepsilon)+\sum_{i\in I}((ru_i+s_i)/(a(r)+\varepsilon)+m_i-\rho_i)\,E_i$$

Remarquons que si l'on pose pour $i\in I$ et $t\in\{0,\varepsilon\}$:

$$c_i(t)=(ru_i+s_i)/(a(r)+t)-\rho_i+m_i$$

on a d'après le lemme 3-1 pour $i\in I-\{0\}$, $0\leqq c_i(t)<1$, et $0\leqq c_0(\varepsilon)<1$. Désormais on posera $c_i=c_i(0)$ pour $i\in I-\{0\}$ et pour $n\geqq n_0$:

$$D_n\equiv nP+\sum_{i\neq 0}m_iE_i.$$

Alors E_0 est partie fixe de $|D_n|$ puisque pour $i\neq 0$, $m_i\leqq\{\rho_i\}$; en utilisant le théorème 3-1 on a pour tout $p\geqq 1$:

$$H^p(Y,\mathcal{O}_Y(D_n-E_0))=H^p(Y,\mathcal{O}_Y(D_n))=0.$$

Cela montre les assertions (i) et (ii) de la proposition 3-1.

c) *La démonstration de l'assertion* (iii). Remarquons que si K_0 est le diviseur canonique de E_0 on a:

$$K_0\equiv(K_Y+E_0)|_{E_0}.$$

Par suite on peut écrire dans $\mathrm{Pic}(E_0)\otimes_{\mathbb{Z}}\mathbb{Q}$:

$$D_n|_{E_0}=K_0+\{(n-m(r+b)/a(r)-1/d)\,P+(H_0+rM)/a(r)+\sum_{i\neq 0}c_iE_i\}|_{E_0}.$$

Pour terminer la démonstration il suffit de prendre pour a un entier tel que

$$a\{(n_0-m(r+b)/a(r)-1/d)\,P+(H_0+rM)/a(r)\}$$

soit un diviseur ample et de prendre pour H ce diviseur. □

Montrons maintenant comment la proposition 3-1 implique le théorème 0-1.

D'après les résultats de Zariski [10] si $SB(P)$ est vide, l'anneau $\bigoplus_{n\geqq 0} H^0(Y,\mathcal{O}_Y(nP))$ est de type fini sur $\mathbb{C}$; en utilisant toujours les résultats de [10] et le fait que les parties mobiles de $|mP|$ et de $|mdK_Y|$ sont les mêmes pour $m\geqq 1$, on en déduit que l'anneau $\bigoplus_{n\geqq 0} H^0(Y,\mathcal{O}_Y(nK_Y))$ est de type fini sur $\mathbb{C}$. Nous raisonnerons par l'absurde et supposerons que $SB(P)$ est non vide. Nous aurons besoin du

Lemme 3-2. *Soient S une surface lisse et projective sur $\mathbb{C}$, K son diviseur canonique, D et L deux diviseurs sur S et n un entier; on a:*

$$\chi(\mathcal{O}_S(K+D+nL))=\chi(\mathcal{O}_S)+n^2L^2/2+nL\cdot(K+2D)/2+D\cdot(K+D)/2.$$

C'est le théorème de Riemann-Roch pour les surfaces.

Posons $R=P|_{E_0}$ et adoptons les notations de la proposition 3-1. Soit $n\geqq n_0$; on a la suite exacte:

$$0\to\mathcal{O}_Y(D_n-E_0)\to\mathcal{O}_Y(D_n)\to\mathcal{O}_{E_0}(D_n)\to 0.$$

Les assertions (i) et (ii) impliquent que pour tout $p\geqq 0$ on a:

$$H^p(E_0,\mathcal{O}_{E_0}(D_n))=0.$$

Posons $G\equiv n_0R-K_0+\sum_{i\neq 0}m_iE_i|_{E_0}$. Remarquons que d'après l'assertion (iii) de la proposition 3-1 on a:

$$D_n|_{E_0}\equiv K_0+G+(n-n_0)\,R$$

et on a dans $\mathrm{Pic}(E_0)\otimes_{\mathbb{Z}}\mathbb{Q}$:

$$G=(H/a+\sum_{i\neq 0}c_iE_i)|_{E_0} \tag{1}$$

où les c_i sont des rationnels positifs strictement inférieurs à 1. Le lemme 3-2 implique:

$$R^2=R\cdot(K_0+2G)=0.$$

Comme R est numériquement positif et que par définition:

$$K_0+G\equiv n_0R+\sum_{i\neq 0}m_iE_i|_{E_0}$$

on déduit:

$$R \cdot G = R \cdot H = 0$$

en utilisant l'égalité (1). Cela implique que R est numériquement trivial sur E_0 et le théorème 3-1 montre que:

$$H^p(E_0, \mathcal{O}_{E_0}(K_0 + G - n_0 R)) = 0$$

pour $p \in \{1, 2\}$. Pour la même raison on a aussi:

$$\chi(\mathcal{O}_{E_0}(K_0 + G - n_0 R)) = \chi(\mathcal{O}_{E_0}(K_0 + G + (n - n_0) R)) = 0$$

et donc:

$$H^0(E_0, \mathcal{O}_{E_0}(K_0 + G - n_0 R)) = 0$$

mais le système linéaire $|K_0 + G - n_0 R|$ contient le diviseur effectif $\sum_{i \neq 0} m_i E_i|_{E_0}$ ce qui est absurde. □

References

1. Benveniste, X.: Sur la décomposition de Zariski en dimension 3. Note aux C.R.A.S. t 295, serie I, pp. 107-110 (1982)
2. Fujita, T.: Semi positive line bundles. A paraître
3. Hartshorne, R.: Algebraic Geometry. G.T.M. vol. 52. Berlin-Heidelberg-New York: Springer 1977
4. Hironaka, H.: Resolution of an Algebraic variety over a field of characteristic zero. (I, II). Annals of Math. **79**, (n° 1) 109-326 (1964)
5. Kawamata, Y.: A generalisation of Kodaira Ramanujam's vanishing Theorem. Math. Ann. **261**, 43-46 (1982)
6. Mori, S.: Threefolds whose canonical bundle is not numerically effective. Lecture Notes in Math., vol. 947. Algebraic threefolds, pp. 155-189. Berlin-Heidelberg-New York: Springer 1982
7. Reid, M.: Canonical 3-folds. Journées de Geométrie Algébrique. Angers 1979, pp. 273-309. Sijthoff et Noordhoff 1980
8. Reid, M.: Minimal models of canonical 3-folds. A paraître au Symposia in Math. Editeurs S. Iitaka et H. Morikawa
9. Viehweg, E.: Vanishing theorem. Journal fur die reine und angewandte Mathematik Band **335** (1982)
10. Zariski, O.: The theorem of Riemann-Roch for high multiples of an effective divisor on an algebraic surface. Ann. of Math. **76**, 560-615 (1962)

Oblatum 7-I-1983

Invent. math. 73, 165–188 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Differential Invariance of Multiplicity on Analytic Varieties

Y.-N. Gau[1] and J. Lipman[2]*

[1] Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA
[2] Department of Mathematics, Purdue University, West Lafayette, IN 47907, USA

Contents

Introduction 165
A. Preliminaries 168
B. The Homology Tangent Cone 172
C. Multiplicity of the Homology Tangent Cone 177
D. The e_i are Differential Invariants 179
E. Appendix: Algebraic Definition of the e_i 183

Introduction

A (complex-)*analytic subvariety* A of an open set $U \subseteq \mathbb{C}^n$ is a (closed) subset of U whose intersection with some open neighborhood of each point of U is the set of common zeros of a finite number of functions defined and holomorphic in that neighborhood. The *smooth points* of A are those points around which A is a complex manifold, so the local geometry of A is interesting only near a non-smooth, or *singular*, point P. A basic measure of singularity is the *multiplicity* of A at P, which is defined to be the local covering number of a generic complex projection from A to $\mathbb{C}^d$ ($d = \dim A$) (cf. (B.2)). This is the same as the multiplicity of the local ring of germs of holomorphic functions on A at P (cf. [22, Chap. VIII, § 10], [18, Chap. III] for the definition of algebraic multiplicity). The question considered here is "*How basic an invariant is multiplicity?*" i.e. how does it behave under topological or differentiable deformation of the pair (A, P)?

Our main result is the following (O = origin).

Theorem. *Let $A \ni O$ (resp. $B \ni O$) be an analytic subvariety of the open set $U \subseteq \mathbb{C}^n$ (resp. $V \subseteq \mathbb{C}^n$). Suppose there is a homeomorphism α from U to V such that $\alpha(A) = B$, $\alpha(O) = O$ and both α and α^{-1} (as functions from $\mathbb{R}^{2n}$ to $\mathbb{R}^{2n}$) have a derivative (= linear approximation [19, p. 16]) at O. Then A and B have the same multiplicity at O.*

* Research supported by NSF Grant MCS-7903444

In short, "weakly" diffeomorphic germs have the same multiplicity.

The above assertion is wrong if we drop the differentiability assumption. It is known, for example, that if A and B are any two curves in $\mathbb{C}^3$ irreducible at O then a (non-differentiable) homeomorphism α exists. More concretely, let $u: \mathbb{C}^2 \to \mathbb{C}$ be a continuous map such for $t \in \mathbb{C}$,

$$u(t^2, t^3) = t \qquad (|t| \leqq 1)$$
$$u(t^2, t^3) = t/|t| \qquad (|t| > 1)$$

(u exists by the Tietze extension theorem). Let $\alpha: \mathbb{C}^3 \to \mathbb{C}^3$ be the homeomorphism given by

$$\alpha(x, y, z) = (x - u(y + x^2, z + x^3), y + x^2, z + x^3).$$

Let $U = D^3$, where D is the open unit disc $|t| < 1$ in $\mathbb{C}$, and let $V = \alpha(U)$. Then α maps the manifold $U \cap (y = z = 0)$ onto the curve $\{(0, x^2, x^3) | x \in D\}$ which has a cusp (of multiplicity 2) at O.

However, Zariski asked [Bull. AMS 77 (1971), p. 483], whether for *hypersurfaces* ($\dim A = \dim B = n - 1$) multiplicity is indeed invariant under embedded homeomorphisms. This is still open.

Known positive results (special cases) for Zariski's question are:

1. for plane curves (i.e. $n = 2$) [21].
2. when one of A, B is smooth at O (follows from [1, p. 114, Th. 3] and [16, p. 261, Prop.]).
3. when the homeomorphism and its inverse both have a derivative at the origin [7].

Our result generalizes Ephraim's by removing the "hypersurface" condition.

We sketch the proof of the Theorem. We may assume that A, B are irreducible (B.2.1). The idea is to use the (Zariski) tangent cones of A and B as intermediaries. The *tangent cone* $C(A, O)$ of an analytic variety $A \ni O$ at O is the set of all limits of secants from O through points on A approaching O (cf. (B.1) for the precise definition).

Now the derivative of α at O takes $C(A, O)$ onto $C(B, O)$, hence (cf. (A.8)) takes each irreducible component A_i of $C(A, O)$ to an irreducible component B_i of $C(B, O)$. The derivative being real linear, we can deduce from a result of Ephraim [8, p. 310, Th. 4.6] that A_i *and* B_i *have the same multiplicity at* O (cf. (B.2.3)).

The multiplicity of A at O is, in general, not equal to the multiplicity of $C(A, O)$ which is the sum of the multiplicities of the A_i (e.g. a cusp). However (cf. Sect. C) it is equal to a linear combination of the multiplicities of the A_i, with coefficients e_i definable through a set $S \subset \mathbb{R} \times \mathbb{C}^n$ which is used by Draper [5, p. 197] in his analytic intersection theory. (In fact, the idea of this proof has its origin in Draper's paper.) S is a geometric avatar of the "Rees ring" often used in algebraic multiplicity theory (cf. (E.4)).

The set $S = S_A \subset \mathbb{R} \times \mathbb{C}^n$ has a natural projection onto $\mathbb{R}$ with $1 \times A$ and $0 \times C(A, O)$ as fibres (over 1 and 0 resp.). Thus $S \cap (0 \times \mathbb{C}^n) = \bigcup_i (0 \times A_i)$. The coefficient e_i is then defined by the relation $[S] \cdot [0 \times \mathbb{C}^n] = \sum e_i [0 \times A_i]$ where [ ]

denotes "fundamental class in Borel-Moore homology" and "$\cdot$" is the homology intersection product (cf. (B.4.2)). If we let f_i be the corresponding coefficients for $C(B,O)$, then it follows from the existence of a fibre-preserving homeomorphism between S_A and S_B that $e_i = f_i$ (cf. Sect. D). Thus with mult = multiplicity)

$$\text{mult of } A = \sum e_i(\text{mult of } A_i) = \sum f_i(\text{mult of } B_i) = \text{mult of } B.$$

In Sect. E, we give an algebraic characterization (not used elsewhere) of the e_i: e_i is the length of the artin local ring G_{p_i}, where G is the graded ring $\bigoplus_{q \geqq 0} \mathfrak{m}^q/\mathfrak{m}^{q+1}$ ($\mathfrak{m}$ = maximal ideal of the local ring $\mathcal{O}_{A,O}$ of germs of analytic functions) - so that $C(A,O) = \mathrm{Spec}(G)$ - and p_i is the minimal prime ideal in G corresponding to A_i.

The tools used are the Borel-Moore homology [4] and its intersection theory developed in [3]. Hence all homology groups in this paper are Borel-Moore homology groups. In this regard, we would like to thank Professor Reinhard Schultz for helpful discussions.

We conclude this introduction with remarks on the notion of "differential equivalence".

Let (A,O), (B,O) be two germs of analytic spaces (cf. [11, p. 150] for the definition). We say that (A,O) and (B,O) are *differentially equivalent* if one of the following three equivalent conditions holds. (For simplicity we write $\mathbb{C}^k$, $\mathbb{R}^k$ for suitable open sets $(\ni O)$ in $\mathbb{C}^k$, $\mathbb{R}^k$, and write A, B for suitable representatives of (A,O), (B,O) respectively):

1. *There exist analytic embeddings* $(A,O) \to (\mathbb{C}^n, O)$, $(B,O) \to (\mathbb{C}^n, O)$ *and a homeomorphism* α *with the same properties as in the Theorem.*

2. *There exist analytic embeddings* $i: (A,O) \to (\mathbb{C}^p, O)$, $j: (B,O) \to (\mathbb{C}^q, O)$ *and a homeomorphism* $h: A \to B$ *such that* i *and* $j \circ h$ *are "diffeo-equivalent" embeddings, i.e. there exist a pair of nonnegative integers* (r,s) *and a homeomorphism* $H: \mathbb{R}^{2p+r} \to \mathbb{R}^{2q+s}$ *such that both* H *and* H^{-1} *have a derivative at* $O \times O$ *and* $H \circ g_1 \circ i = g_2 \circ j \circ h$, *where* $g_1: \mathbb{R}^{2p} \to \mathbb{R}^{2p} \times \mathbb{R}^r$ (resp. $g_2: \mathbb{R}^{2q} \to \mathbb{R}^{2q} \times \mathbb{R}^s$) *is the inclusion* $x \to x \times O$ [12, p. 128].

A function $f: X \to \mathbb{R}^n$ (where (X,O) is an analytic germ) is *differentiable at* O if for some embedding $(X,O) \to (\mathbb{C}^N, O)$, f is the restriction (to X) of a function from $\mathbb{C}^N$ to $\mathbb{R}^n$ having a derivative at O. A map between analytic germs $f: (X,O) \to (Y,O)$ is *differentiable* if its composition with some embedding $(Y,O) \to (\mathbb{C}^M, O)$ is differentiable at O. (This property does not depend on the choice of the embeddings.) Condition 3 is:

3. *There exist differentiable maps* $f: (A,O) \to (B,O)$ *and* $g: (B,O) \to (A,O)$ *such that* $f \circ g$ *and* $g \circ f$ *are identity maps.*

[The implications $1 \Rightarrow 2 \Rightarrow 3$ are trivial. For $3 \Rightarrow 1$, choose embeddings $A \to \mathbb{C}^N$, $B \to \mathbb{C}^M$; extend f (resp. g) to a map $F: \mathbb{C}^N \to \mathbb{C}^M$ (resp. $G: \mathbb{C}^M \to \mathbb{C}^N$) differentiable at O; and define a homeomorphism $\alpha: \mathbb{C}^N \times \mathbb{C}^M \to \mathbb{C}^N \times \mathbb{C}^M$ by

$$\alpha(x,y) = (x - G(y + F(x)),\ y + F(x)).]$$

The Theorem then states that *differentially equivalent germs have the same multiplicity.*

Example. In [2, § 4], Bloom shows that the curves $A \subseteq \mathbb{C}^2$, $B \subseteq \mathbb{C}^3$ given by

$$A = \{t^3, t^5) | t \in \mathbb{C}\},$$
$$B = \{(t^3, t^5, t^7) | t \in \mathbb{C}\}$$

are differentially equivalent at their respective origins (in fact t^7 is a C^1 function everywhere on A). The multiplicities at O are of course the same (namely 3). Incidentally, the Hilbert-Samuel polynomials are *different*, namely $3n-3$ for A at O and $3n-2$ for B at O.

A. Preliminaries

This section gives definitions, fixes notations and lists some properties of homology groups and intersection products. At the end we prove a useful Lemma (A.8). The main references are [4], [3].

(A.1) $\mathbb{Z}$, $\mathbb{Z}^+$, $\mathbb{R}$, $\mathbb{C}$ denote as usual the integers, strictly positive integers, real numbers, complex numbers respectively. All spaces will be subsets of Euclidean spaces with the usual metric topology and hence will be paracompact. Unless otherwise indicated, *analytic* means *complex analytic* and *component* means *irreducible component* (of an analytic variety) (cf. [11] for these and other terminologies in Analytic Geometry).

(A.2) Recall that a *family of supports* Φ on a topological space X is a family of closed subsets of X having the following properties: If $A, B \in \Phi$, then $A \cup B \in \Phi$; if $A \in \Phi$ and if B is closed in A, then $B \in \Phi$.

We write $H_i^{\Phi}(X)$ (resp. $H^i_{\Phi}(X)$) to denote the i-th homology (resp. cohomology) group of X with coefficients in the constant sheaf $\mathbb{Z}$ and supports in Φ, as defined in [4] (resp. [9]). We omit i if its value is clear from the context.

In the preceding notations, we omit Φ if Φ is the family of all closed subsets of X, and we replace Φ by F (a closed subset of X) if Φ is the family of all closed subsets of F.

(A.2.1) In case F, $X \subset \mathbb{R}^n$ and $F \cap X$ is closed in X we write, by abuse of notation, $H^F(X)$ to mean $H^{F \cap X}(X)$.

(A.3) Let $f: X \to Y$ be a continuous map and Φ (resp. Ψ) a family of supports on X (resp. Y). If for every $\phi \in \Phi$, $f(\phi) \in \Psi$ (in symbol: $f(\Phi) \subset \Psi$) and the restriction of f to ϕ is proper, then f induces a homomorphism $f^{\Phi,\Psi}$ or f_* from $H_i^{\Phi}(X)$ into $H_i^{\Psi}(Y)$ (cf. [4, p. 143, Sect. 3] for the definition of f_* and its properties). As for *cohomology*, if $f^{-1}(\Psi) \subset \Phi$, then f induces a homomorphism $f^*: H^i_{\Psi}(Y) \to H^i_{\Phi}(X)$ [9, p. 199, 4.16].

If $\Phi \subset \Phi'$ are families of supports on X, the map $\mathrm{id}^{\Phi,\Phi'}: H_i^{\Phi}(X) \to H_i^{\Phi'}(X)$ is called "*enlargement of supports*". (It can be checked that this is equivalent to the definition given in [3, p. 464].)

For U open in X, there is a *restriction map* $j^{X,U}: H_i^{\Phi}(X) \to H_i^{\Phi \cap U}(U)$ where $\Phi \cap U = \{\phi \cap U | \phi \in \Phi\}$ is a family of supports on U [3, p. 464].

(A.3.1) *f_* is functorial:*

If $g: Y \to Z$ is continuous and $g^{\Psi,\Theta}$ is defined, then

$$(g \circ f)^{\Phi,\Theta} = g^{\Psi,\Theta} \circ f^{\Phi,\Psi}.$$

(A.3.2) *f_* is compatible with restriction maps:*

Let V be open in Y and $U = f^{-1}(V)$ (open in X). If $f^{\Phi,\Psi}$ is defined then so is $f^{\Phi\cap U,\Psi\cap V}: H^{\Phi\cap U}(U) \to H^{\Psi\cap V}(V)$, and we have

$$f^{\Phi\cap U,\Psi\cap V} \circ j^{X,U} = j^{Y,V} \circ f^{\Phi,\Psi}.$$

(A.3.3) *Inclusion maps* $i_{F,X}$ [4, p. 145, Th. 3.5(b)]:

$i_{F,X} = i^{\Phi\cap F,\Phi}: H^{\Phi\cap F}(F) \to H^{\Phi}(X)$ is the homomorphism induced by the inclusion $i: F \to X$ (A.3). We shall use $i_{F,X}$, $i^{\Phi\cap F,\Phi}$ interchangeably. If Φ is a family of supports on F, then $i_{F,X}$ is an isomorphism. In particular ($\Phi = E$, a closed subset of F; and $\Phi = F$)

$$H^E(F) \xrightarrow{\sim} H^E(X), \quad \text{and} \quad H(F) \xrightarrow{\sim} H^F(X).$$

(A.3.4) *Homology exact sequence* [4, p. 146, Th. 3.8]:

Let U be open in X and $F = X - U$. Then there exists an exact sequence:

$$\ldots \to H_q(F) \xrightarrow[i_{F,X}]{} H_q(X) \xrightarrow[j^{X,U}]{} H_q(U) \to H_{q-1}(F) \to \ldots$$

(A.4) *Fundamental classes:*

See [3, p. 469, 2.2] for the definition of fundamental classes. For a space X, $[X]$ will always denote a fundamental class of X (if it exists).

(A.4.1) For a d-dimensional analytic variety A, $H_k(A) = 0$ for $k > 2d$, and $H_{2d}(A) \simeq \Pi_\lambda \mathbb{Z}_\lambda$, one $\mathbb{Z}_\lambda = \mathbb{Z}$ for each irreducible component A_λ of A [3, pp. 476, 482]. In this case we write $[A] \in H_{2d}(A)$ for *the* fundamental class, i.e. the one which induces the canonical orientation at all smooth points. When A is irreducible, $[A]$ generates $H_{2d}(A) \simeq \mathbb{Z}$.

(A.4.2) Let M be a real n-dimensional oriented connected topological manifold. Then $H_{n+1}(M) = 0$ and $H_n(M) \simeq \mathbb{Z}$ [3, p. 467, 1.11]. (In particular, for a connected open subset U of $\mathbb{R}^n$, we have $H_n(U) \simeq \mathbb{Z}$.) We write $[M]$ for the generator of $H_n(M)$ which corresponds to the orientation.

(A.4.3) Let X be a space with fundamental class $[X]$. Then the *restriction* $j^{X,U}[X]$ of $[X]$ to an open subset U, is a fundamental class of U. If X is an analytic variety, then $j^{X,U}[X]$ is *the* fundamental class $[U]$. In particular, if M denotes the set of smooth points of an analytic variety A, then the restriction of $[A]$ to M gives the fundamental class $[M]$ [3, 3.2 and 2.3].

If X is closed in some space Y, we let $[X]^Y = i_{X,Y}[X]$ where $i_{X,Y}: H(X) \xrightarrow{\sim} H^X(Y)$. Note that if Y is closed in Z, then by (A.3.1)

$$[X]^Z = i_{X,Z}[X] = i_{Y,Z} \circ i_{X,Y}[X] = i_{Y,Z}[X]^Y.$$

(A.4.4) As in (A.2.1), $[F]^X$ will mean $[F \cap X]^X$ (notations as in (A.2.1)).

(A.5) *Poincaré duality:*

We shall write "$\cap$" for cap product [3, p. 504, Appendix]. (N.B. the cap product $H_i^{\Phi}(X)\otimes H_{\Psi}^{j}(X)\longrightarrow H_{i-j}^{\Phi\cap\Psi}(X)$ is defined only if Φ and $\Phi\cap\Psi$ are paracompactifying. This is no restriction for us since we consider only metric spaces.)

For an n-dimensional oriented connected manifold X there is an isomorphism $\Delta: H_i^{\Phi}(X)\simeq H_{\Phi}^{n-i}(X)$ which is compatible with restrictions to open subsets [3, p. 467, 1.11]. Moreover, the isomorphism is characterized by $\Delta^{-1}\xi = [X]\cap\xi$ for $\xi\in H_{\Phi}^{n-i}(X)$ [3, p. 510, Th. 7.9 and p. 467-8, 1.11].

(A.6) *Intersection Product:*

Let X be an n-dimensional oriented manifold. For $a\in H_i^{\Phi}(X)$, $b\in H_j^{\Psi}(X)$ their intersection product, denoted by $a\cdot b$, is defined to be

$$a\cdot b = a\cap\Delta b\in H_{i+j-n}^{\Phi\cap\Psi}(X).$$

(A.6.1) The intersection product is bilinear, associative, anticommutative and compatible with both *enlargement of supports* and *restriction to open subsets.* (The last assertion follows from the definition and the corresponding properties for cap product and Δ, cf. (A.5) and [3, p. 565, Th. 7.2].)

(A.6.2) If $i+j=n$ and $\Phi\cap\Psi=$ a point Q, then $a\cdot b\in H_0^Q(X)\cong H_0(Q)\cong\mathbb{Z}$. (The last isomorphism comes out in a straightforward way from the definition of H_0.) We shall call the absolute value of $a\cdot b$ the *intersection number* of a and b, and denote it by $a\circ b$.

(A.6.3) *Criterion of unit multiplicity* [3, p. 484, 4.8].

Let A, B be two analytic subvarieties of an open subset U of $\mathbb{C}^n$. Suppose A (resp. B) is pure i-(resp. j-)dimensional and $A\cap B=C$, where C is irreducible and $\dim C=i+j-n$ (i.e. A, B *intersect properly at* C). Assume moreover that C contains a point which is smooth in both A and B, and at which A intersects B transversally. Then the conclusion is $[A]^U\cdot[B]^U=[C]^U$. (The special case when A, B are transversal complex linear spaces is useful for our purpose.)

(A.6.4) Let $V\subset W$ be connected oriented open subsets of $\mathbb{R}^s$ and $\mathbb{R}^t$ respectively such that V is closed in W. Let $i: V\to W$ be the inclusion map. Let $Y\subset W$ be either an analytic subvariety of W or the intersection of W with an oriented real-linear space, and let $[Y]$ be its fundamental class ((A.4.1), (A.4.2)).

Suppose that V is transversal to Y at all points of $V\cap Y_R$ where $Y_R\subset Y$ is an open subset consisting of smooth points. Assume also that $\dim_{\mathbb{R}}(Y-Y_R) < \dim_{\mathbb{R}} Y$ and that $\dim_{\mathbb{R}}(Y-Y_R)\cap V < s-t+\dim_{\mathbb{R}} Y$.

Then $X=Y\cap V$ *has a unique fundamental class* $[X]\in H(X)$ *such that*

$$\Delta i_{X,V}[X] = i^*\Delta i_{Y,W}[Y], \quad \text{where } i^*: H_Y(W)\to H_X(V);$$

or equivalently, by (A.4.3)

$$\Delta[X]^V = i^*\Delta[Y]^W.$$

Proof. This is a special case of [3, p. 475, 2.15].

(A.7) Let U (resp. U') be a connected oriented open set in $\mathbb{R}^n$ (resp. $\mathbb{R}^{n'}$), Φ, Ψ (resp. Φ', Ψ') families of supports on U (resp. U') and let $f: U\to U'$ be such

that $f^{\Phi,\Phi'}, f^{\Phi\cap\Psi,\Phi'\cap\Psi'}$ and $f_{\Psi',\Psi}$ are defined, i.e. $f(\Phi)\subset\Phi', f|_{\phi}$ is proper for $\phi\in\Phi$; $f(\Phi\cap\Psi)\subset\Phi'\cap\Psi'$; $f^{-1}(\Psi')\subset\Psi$. Under these conditions we have the *projection formula* [3, p. 466, 1.9]): in $H^{\Psi}_{s-t}\cap^{\Psi'}(Y)$,

$$f_*(a\cap f^* b)=f_* a\cap b \qquad [a\in H^{\Phi}_s(X); b\in H^{t}_{\Psi'}(Y)].$$

(A.7.1) *Variant of projection formula:*
Let $\xi\in H^{\Phi}_i(U)$, $\eta\in H^{\Psi}_j(U)$. If $\eta'\in H^{\Psi'}_{j+n'-n}(U')$ is such that $\Delta\eta=f^{n-j}_{\Psi',\Psi}\Delta\eta'$, then

$$f^{\Phi\cap\Psi,\Phi'\cap\Psi'}_{i+j-n}(\xi\cdot\eta)=(f^{\Phi,\Phi'}_i\xi)\cdot\eta'.$$

Remark. In singular homology η is called the *transfer* $f_!\eta'$ of η', cf. [6, p. 310].

Proof. By Definition (A.6) and the hypothesis we have:

$$f_{i+j-n}(\xi\cdot\eta)=f_{i+j-n}(\xi\cap\Delta\eta)=f_{i+j-n}(\xi\cap f^{n-j}\Delta\eta').$$

By the projection formula the last term is:

$$f_i\xi\cap\Delta\eta'=f_i\xi\cdot\eta'. \quad \text{q.e.d.}$$

(A.7.2) We apply (A.7.1) to the situation in (A.6.4), i.e. take $U=V$, $U'=W$, $\Psi=X$, $\Psi'=Y$, $f=i$: $V\to W$ and Φ,Φ' such that $i^{\Phi,\Phi'}$ is defined ($\Rightarrow i^{\Phi\cap\Psi,\Phi'\cap\Psi'}$ is defined). By (A.6.4), $\eta=[X]^V$ and $\eta'=[Y]^W$ satisfy the condition in (A.7.1), hence:

$$i^{\Phi\cap X,\Phi'\cap Y}(\xi\cdot[X]^V)=i^{\Phi,\Phi'}(\xi)\cdot[Y]^W \qquad \text{for all } \xi\in H^{\Phi}(V).$$

In particular, letting $\Phi=\Phi'=V$ and $\xi=V$ (A.4.2), we have

$$i^{X,X}([V]\cdot[X]^V)=(i^{V,V}[V])\cdot[Y]^W=[V]^W\cdot[Y]^W.$$

But $[V]\cdot[X]^V\underset{\text{def}}{=}[V]\cap\Delta[X]^V=\Delta^{-1}\Delta[X]^V=[X]^V$ by (A.5). Thus by (A.6.3):

$$[X]^W\underset{(A.4.3)}{=}i^{X,X}[X]^V=[V]^W\cdot[Y]^W.$$

(A.7.3) *Topological Invariance of Intersection*

Let U, U' be as in (A.7). Assume f: $U\to U'$ is a homeomorphism with $\Phi'=f(\Phi)$, $\Psi'=f(\Psi)$. Write $f'_*=f^{\Phi,\Phi'}_i$, $f_*=f^{\Psi,\Psi'}_j$ and $\tilde{f}_*=f^{\Phi\cap\Psi,\Phi'\cap\Psi'}_{i+j-n}$. Then for $\xi'\in H^{\Phi}_i(U)$, $\xi\in H^{\Psi}_j(U)$ we have: $\tilde{f}_*(\xi'\cdot\xi)=\pm f'_*\xi'\cdot f_*\xi$.

Proof. By (A.7.1) it suffices to check:

$$\Delta\xi=\pm f^*\Delta f_*\xi.$$

Using (A.5) and (A.7) we have

$$f_*\Delta^{-1}(f^*\Delta\eta)=f_*([U]\cap f^*\Delta\eta)=f_*[U]\cap\Delta\eta \qquad \text{for } \eta\in H^{\Phi'}(U').$$

Since f_* is an isomorphism, $f_*[U]=\pm[U']$ by (A.4.2). By (A.5) again, we conclude that

$$f_*\Delta^{-1}f^*\Delta\eta=\pm\Delta^{-1}\Delta\eta=\pm\eta.$$

In particular, if $\eta = f_* \xi$ we get (since f_* is an isomorphism)

$$\Delta^{-1} f^* \Delta f_* \xi = \pm \xi$$

$$\text{i.e.} \quad f^* \Delta f_* = \pm \Delta \xi. \quad \text{q.e.d.}$$

(A.8) Lemma. *Let $f: A \to B$ be a homeomorphism of analytic spaces. Then for every component A_* of A, $f(A_*)$ is a component of B with the same dimension as A_*.*

Proof. Let $d = \dim A_*$ and $M = A_* - \mathrm{Sing}(A)$ be the set of all smooth points of A which lie in A_*. M is a connected manifold which is open in A and with closure (in A) $\overline{M} = A_*$. Hence $f(M)$ is an open connected subset of B which is locally homeomorphic to $\mathbb{R}^{2d}$ and, regarding $f(M)$ as an open subvariety of B, $f(M) - \mathrm{Sing}(f(M))$ is a pure d-dimensional complex manifold. It follows that $f(M)$ is also pure d-dimensional.

Now homology groups being topological invariants (A.3.1), we have $H_{2d}(f(M)) \simeq H_{2d}(M)$ which by (A.4.2) is isomorphic to $\mathbb{Z}$. By [3, p. 482, Lemma 4.3] $f(M)$ is irreducible, so is contained in one component of B, say B_*. This implies $f(A_*) = f(\overline{M}) \subset \overline{f(M)} \subset \bar{B}_* = B_*$. The same argument applied to f^{-1} and B_* gives that $f^{-1}(B_*)$ is contained in a component of A, say A_{**}.

Together we have: $A_* = f^{-1} f(A_*) \subseteq f^{-1}(B_*) \subseteq A_{**}$. But components are maximal irreducible analytic subspaces, hence $A_* = A_{**}$. In particular: $f(A_*) = B_*$. Also $\overline{f(M)} = B_*$, therefore B_* is d-dimensional. q.e.d.

B. The Homology Tangent Cone

(B.1) As mentioned in the introduction, the basic idea of the proof is to relate the analytic variety A to its (Zariski) tangent cone through a map $S \to \mathbb{R}$ which has A and the tangent cone as two of its fibres. Let $d = \dim A$.

Definition [20, p. 210, Def. 1G]: Let A be an analytic variety and $A \ni O$. The *tangent cone* of A at O, denoted by $C(A, O)$, is the set of all vectors v with the property that there are sequences of points $q_i \in A$ and of *complex* numbers a_i such that $q_i \to O$ and $a_i q_i \to v$, i.e. the direction of v (if $v \neq O$) is a limit of directions of secants from O to points of A converging to O.

(B.1.1) We may restrict the a_i in the above definition to be real and positive; this does not change $C(A, O)$ [20, p. 218, Remark 3E].

(B.1.2) If A is pure d-dimensional, then $C(A, O)$ is a pure d-dimensional algebraic variety which is a *cone* (i.e. $t \in \mathbb{C}$, $x \in C(A, O)$ implies $t x \in C(A, O)$) [20, p. 211, Lemma 1L and p. 214, Th. 2E].

(B.1.3) With α as in the Theorem, the derivative of α at O [19, p. 16], denoted by α_0, is a real-linear map inducing a homeomorphism between $C(A, O)$ and $C(B, O)$. The proof uses (B.1.1) but is otherwise straightforward, cf. [7, p. 802].

(B.2) Definition. The *multiplicity* of a d-dimensional analytic variety $A \subset \mathbb{C}^n$ at a point $O \in A$, *denoted by* $\mu(A, O)$, is defined to be the local covering number of

the projection of a small neighborhood around O in A to a generic d-dimensional complex linear subspace (cf. [20, p. 233, Def. 7J]).

In other words, for "most" linear projections $\pi: \mathbb{C}^n \to \mathbb{C}^d$, π induces a $\mu(A,O)$-sheeted unramified covering $V - \pi^{-1}(D) \to \pi(V) - D$ where $V \subset A$ is some neighborhood of O and D is a nowhere dense analytic subvariety of the open set $\pi(V) \subset \mathbb{C}^d$, and where $\mu(A,O)$ does not depend on π.

Remark 1. $\mu(A,O)$ can be shown to be equal to the algebraic multiplicity of the analytic local ring of A at O (i.e. the local ring of germs at O of analytic functions on A) [5, p. 198, Th. 6.5], [17, p. 123, (A.15)]. Also we can replace "generic projection" in the above definition by "*transversal projection* π" which means that O is an isolated point of $\pi^{-1}\pi(O) \cap C(A,O)$ (cf. [5, p. 196, Th. 6.3], [20, p. 234, Th. 7P]).

Remark 2. If $A^1, A^2, \ldots$ are those components of A which have dimension d ($d = \dim A$) and pass through O, then $\mu(A,O) = \sum_i \mu(A^i, O)$. To see this, just use any projection which is sufficiently generic for A and for all the A_i (for example, any projection which is transversal for A (Remark 1) and hence for the A_i). An algebraic proof can also be given by using [18, p. 76, (23.5)].

(B.2.1) *To prove the Theorem, we can and shall assume that A and B are irreducible.* Let A^i be as in Remark 2. By (A.8) α maps each A^i onto some component of B, call it B^i. Also $\dim B = d$ and the B^i are all the irreducible d-dimensional components of B containing O. By Remark 2, it suffices to prove the Theorem for $\alpha: A^i \to B^i$.

(B.2.2) It is clear from the definition that $\mu(A,O) = \mu(A \cap U, O)$ where $U \subset \mathbb{C}^n$ is any open set in $\mathbb{C}^n$ containing O. Algebraically this follows from the fact that A and $A \cap U$ have the same local ring at O.

For simplicity, we omit from now on the reference point O. Thus $\mu(A) = \mu(A,O)$, $C(A) = C(A,O)$ etc.

(B.2.3) Let A^* be a component of $C(A)$. By (B.1.3) $\alpha_0(A^*)$ is an irreducible component, say B^*, of $C(B)$.

Claim. $\mu(A^*) = \mu(B^*)$.

Proof. We use a structure theorem proved by R. Ephraim [8, p. 310, Th. 4.6]: Any germ of analytic variety has a *decomposition*, i.e. it is complex-analytically isomorphic to a (finite) product of analytic germs each of which can't be further decomposed. If X, Y are two irreducible analytic germs, and X is C^∞-isomorphic to Y, then the decomposition of X and Y are the same up to permutations and complex conjugations. To be precise, if $X = X_1 \times \ldots \times X_k$, $Y = Y_1 \times \ldots \times Y_h$ are decompositions of X and Y (into indecomposables), then $k = h$ and, after a permutation of $Y_1 \ldots Y_k$, either $X_i \simeq Y_i$ or $X_i \simeq \overline{Y_i}$. ("$\simeq$" means complex-analytically isomorphic, $\overline{Y_i}$ = complex conjugate of Y_i.)

Applying the above to $\alpha_0: A^* \to B^*$, we have $A^* = A_1^* \times \ldots \times A_k^*$, $B^* = B_1^* \times \ldots \times B_k^*$ and A_i^* is analytically isomorphic either to B_i^* or to the complex conjugate of B_i^*. In either case, A_i^* and B_i^* have isomorphic (as $\mathbb{R}$-algebras) ana-

lytic local rings at the origin [8, p. 301, Remark 2.4], hence have the same multiplicity (B.2, Remark 1). But arguing as in [18, (42.6)], we see that

$$\mu(A^*)=\prod_i \mu(A_i^*), \qquad \mu(B^*)=\prod_i \mu(B_i^*).$$

Hence $\mu(A^*)=\mu(B^*)$. q.e.d.

Remark. The assertions $\mu(\bar{A})=\mu(A)$ and $\mu(A_1\times A_2)=\mu(A_1)\mu(A_2)$ in the above proof can also be proved topologically by observing that if π (resp. π_1, π_2) is transversal to A (resp. A_1, A_2) then the conjugate of π (resp. $\pi_1\times\pi_2$) is transversal to $\bar{A}$ (resp. $A_1\times A_2$). The results then follow from (B.2, Remark 1).

(B.3) As in (B.2.1), A is an irreducible analytic subvariety of the open set $U\subseteq\mathbb{C}^n$, and $O\in A$. We now define the sets S, $\bar{S}$ from A (resp. T, $\bar{T}$ from B) which provide the foundations for the proof of the Theorem. Let $\mathbb{C}^*=\mathbb{C}-\{0\}$, $\mathbb{R}^*=\mathbb{R}-\{0\}$. Consider the map ψ from $\mathbb{C}\times\mathbb{C}^n$ into itself defined by:

$$\psi(t, z_1, \ldots, z_n)=(t, t z_1, \ldots, t z_n).$$

Then $\mathscr{U}=\psi^{-1}(\mathbb{C}\times U)$ is an open set in $\mathbb{C}\times\mathbb{C}^n$. (The corresponding set $\psi^{-1}(\mathbb{C}\times V)$ for V will be denoted by $\mathscr{V}$.) Since $\psi^{-1}(t\times U)=\left(t\times\frac{1}{t}U\right)$ for $t\neq 0$, we see that $\mathscr{U}=\bigcup_{t\in\mathbb{C}^*}\left(t\times\frac{1}{t}U\right)\cup(0\times\mathbb{C}^n)$.

To define S (an analytic subvariety of $\mathscr{U}$) we consider $\psi^{-1}(\mathbb{C}\times A)$. ψ being analytic, $\psi^{-1}(\mathbb{C}\times A)$ is an analytic subvariety of $\mathscr{U}$, of dimension $<n+1$, and one of its components is $0\times\mathbb{C}^n$.

(B.3.1) **Definition.** S is the union of the other components.

It is clear that $S-(0\times\mathbb{C}^n)=\psi^{-1}(\mathbb{C}^*\times A)$. Also note that ψ: $\mathscr{U}-(0\times\mathbb{C}^n)\to\mathbb{C}^*\times U$ being an analytic isomorphism, $((t, z)\to(t, (1/t)z)$ is the inverse), the smooth points of $S-(0\times\mathbb{C}^n)$ are mapped (by ψ) to the smooth points of $\mathbb{C}^*\times A$, i.e. the points in $\mathbb{C}^*\times A^-$ where A^- denotes the set of smooth points in A. Now using the fact that the components of an analytic variety W are the closures (in W) of the connected components of the smooth part of W [20, p. 74, Th. 1G], we can conclude the following (B.3.2) and (B.3.3):

(B.3.2) $S-(0\times\mathbb{C}^n)=\psi^{-1}(\mathbb{C}^*\times A)$ is an irreducible analytic variety.

(B.3.3) $S\cap(0\times\mathbb{C}^n)=0\times C(A)$ (in particular, $0\times C(A)$ is closed in S).

Proofs. It is obvious that, as sets, $\psi^{-1}(\mathbb{C}\times A)=(0\times\mathbb{C}^n)\cup(S-(0\times\mathbb{C}^n))$.

Since A is assumed to be irreducible, A^-, and hence $\psi^{-1}(\mathbb{C}^*\times A^-)$, is connected; (B.3.2) follows. By [20, p. 79, Th. 2J] and (B.3.1), $(0\times\mathbb{C}^n)-S$ and $\psi^{-1}(\mathbb{C}^*\times A^-)$ are all the connected components of the smooth part of $\psi^{-1}(\mathbb{C}\times A)$. It follows that S is *irreducible* and is the closure (in $\psi^{-1}(\mathbb{C}\times A)$) of $\psi^{-1}(\mathbb{C}^*\times A^-)$. Since

$$S\supset S-(0\times\mathbb{C}^n)=\psi^{-1}(\mathbb{C}^*\times A)\supset\psi^{-1}(\mathbb{C}^*\times A^-),$$

S is also the closure of $S-(0\times\mathbb{C}^n)=\bigcup_{t\neq 0}\left(t,\frac{1}{t}A\right)$. From the definition of the tangent cone (B.1), we see then that $S\cap(0\times\mathbb{C}^n)=0\times C(A)$.

(B.3.4) Next we consider a real section $\bar{S}$ of S. This $\bar{S}$ will play an essential role in section D (cf. (D.1)).

Definition. $\bar{\mathscr{U}}=\mathscr{U}\cap(\mathbb{R}\times\mathbb{C}^n)$, $\bar{S}=S\cap(\mathbb{R}\times\mathbb{C}^n)=S\cap\bar{\mathscr{U}}$.

It is readily seen that $\bar{\mathscr{U}}=\bigcup_{t\in\mathbb{R}^*}\left(t\times\frac{1}{t}U\right)\cup(0\times\mathbb{C}^n)$ is a connected open set in $\mathbb{R}\times\mathbb{C}^n$, $\bar{\mathscr{U}}$ is closed in $\mathscr{U}$, $\bar{S}$ is a real-analytic variety which is closed in $\bar{\mathscr{U}}$, $\bar{S}\cap(t\times\mathbb{C}^n)=t\times\frac{1}{t}A$ $(t\in\mathbb{R}^*)$ and $\bar{S}\cap(0\times\mathbb{C}^n)=0\times C(A)$. The last statement follows from (B.3.3).

Fix an orientation in $\mathbb{R}\times\mathbb{C}^n$. Consider the inclusion $i:\bar{\mathscr{U}}\to\mathscr{U}$ and the analytic variety $S\subset\mathscr{U}$. Note that $\dim_{\mathbb{R}}\mathscr{U}=2n+2$, $\dim_{\mathbb{R}}\bar{\mathscr{U}}=2n+1$, $\dim_{\mathbb{R}}S=2d+2$, $\dim_{\mathbb{R}}\bar{S}=2d+1$. It can be checked easily that $\bar{\mathscr{U}}$ is transversal to S at smooth points of S of the form (t,x) with $t\in\mathbb{R}^*$ (cf. (C.2.2, Remark)). Also $\dim_{\mathbb{R}}(S-\psi^{-1}(\mathbb{C}^*\times A^-))\cap\bar{\mathscr{U}}=2d$ (B.3.3). Thus we can apply (A.6.4) (with $Y=S$, $Y_{\mathbb{R}}=\psi^{-1}(\mathbb{C}^*\times A^-)$) to conclude that $\bar{S}=S\cap\bar{\mathscr{U}}$ has a fundamental class $[\bar{S}]\in H_{2d+1}(\bar{S})$ and, by (A.7.2) that

$$[\bar{S}]^{\mathscr{U}}=[\bar{\mathscr{U}}]^{\mathscr{U}}\cdot[\bar{S}]^{\mathscr{U}}. \tag{B.3.5}$$

(B.4) Through S (or $\bar{S}$) we can deform A to its tangent cone $C(A)$. However, to avoid loss of information we need something like the "algebraic" tangent cone $\sum e_iA_i$, $e_i\in\mathbb{Z}^+$, where A_i are the components of $C(A)$ (cf. [5, p. 198]). Actually, we work with the "homology tangent cone" $C_H(A)=C_H(A)^{\mathscr{U}}$ defined as the homology class:

$$C_H(A)=[S]^{\mathscr{U}}\cdot[0\times\mathbb{C}^n]^{\mathscr{U}}\in H_{2d}^{0\times C(A)}(\mathscr{U}).$$

Consider $\bigoplus_i H_{2d}(0\times A_i)\longrightarrow H_{2d}(0\times C(A))\xrightarrow{\sim}H_{2d}^{0\times C(A)}(\mathscr{U})$. The first map is an isomorphism induced by the direct sum of inclusion maps: $\bigoplus i^{0\times A_i,\,0\times C(A)}$ [cf. 3, p. 482, Lemma 4.3]. The second map is also an isomorphism by (A.3.3).

(B.4.1) Let $[\![0\times A_i]\!]^{\mathscr{U}}$ denote the image of $[0\times A_i]\in H(0\times A_i)$ in $H_{2d}^{0\times C(A)}(\mathscr{U})$. By (A.3.1), (A.3), (A.4.3) we see:

$$[\![0\times A_i]\!]^{\mathscr{U}}=\mathrm{id}^{0\times A_i,\,0\times C(A)}\circ i_{0\times A_i,\mathscr{U}}[0\times A_i]=\mathrm{id}^{0\times A_i,\,0\times C(A)}[0\times A_i]^{\mathscr{U}}.$$

(B.4.2) **Definition.** e_i is the integer uniquely determined by

$$C_H(A)=[S]^{\mathscr{U}}\cdot[0\times\mathbb{C}^n]^{\mathscr{U}}=\sum e_i[\![0\times A_i]\!]^{\mathscr{U}}.$$

Thus e_i is the *intersection number* (in $\mathscr{U}$) [3, p. 483]

$$e_i=i(S\cdot(0\times\mathbb{C}^n),A_i).$$

e_i does not change when U is replaced by a smaller open set (A.6.1).

(B.4.3) Similarly, from $B \subset V$ we can define $\mathscr{V}$, $\overline{\mathscr{V}}$, T, $\bar{T}$, and integers f_i are then defined by

$$C_H(B) = [T]^{\mathscr{V}} \cdot [0 \times \mathbb{C}^n]^{\mathscr{V}} = \sum f_i [\![0 \times B_i]\!]^{\mathscr{V}}.$$

(B.4.4) It is a fact that all the e_i (and f_i) are *positive* (cf. [3, p. 488, 4.16]). This also follows from the fact that e_i and f_i can be defined algebraically (E.1.1).

(B.5) The integers e_i can also be derived from $\bar{S} \subset \overline{\mathscr{U}}$ (resp. f_i from $\bar{T} \subset \overline{\mathscr{V}}$).

Let $[\![0 \times A_i]\!]^{\mathscr{U}}$ be the natural image of $[0 \times A_i]$ in $H^{0 \times C(A)}_{2d}(\overline{\mathscr{U}})$. Then by (A.4.3) and definitions:

(B.5.1) $$[\![0 \times A_i]\!]^{\mathscr{U}} = i^{0 \times C(A),\, 0 \times C(A)} [\![0 \times A_i]\!]^{\overline{\mathscr{U}}},$$

where $i: \overline{\mathscr{U}} \to \mathscr{U}$ is the inclusion.

(B.5.2) **Lemma.**

$$[\bar{S}]^{\mathscr{U}} \cdot [0 \times \mathbb{C}^n]^{\mathscr{U}} = \sum e_i [\![0 \times A_i]\!]^{\mathscr{U}} \in H^{0 \times C(A)}(\overline{\mathscr{U}}).$$

Proof. We choose an orientation on $\mathbb{R} \times \mathbb{C}^n$ as before (B.3.4), and on $i\mathbb{R} \times \mathbb{C}^n$ in such a way that $[\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}} \cdot [i\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}} = [0 \times \mathbb{C}^n]^{\mathscr{U}}$ (cf. (A.6.3)). (We are using the convention $[X]^{\mathscr{U}} = [X \cap \mathscr{U}]^{\mathscr{U}}$ (A.2.1); and $i\mathbb{R}$ stands for the pure imaginary numbers in $\mathbb{C}$.) Then the lemma results from (B.4.2), (B.5.1) and the following two assertions:

(1) $$i^{0 \times C(A),\, 0 \times C(A)}([\bar{S}]^{\overline{\mathscr{U}}} \cdot [0 \times \mathbb{C}^n]^{\overline{\mathscr{U}}}) = [\bar{S}]^{\mathscr{U}} \cdot [i\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}}.$$

(2) $$[\bar{S}]^{\mathscr{U}} \cdot [i\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}} = [S]^{\mathscr{U}} \cdot [0 \times \mathbb{C}^n]^{\mathscr{U}} \in H^{0 \times C(A)}(\mathscr{U}).$$

(1) follows immediately from (A.7.2) (let $V = \overline{\mathscr{U}}$, $W = \mathscr{U}$, $X = (0 \times \mathbb{C}^n) \cap \mathscr{U}$, $Y = (i\mathbb{R} \times \mathbb{C}^n) \cap \mathscr{U}$, $\Phi = \Phi' = \bar{S}$ and $\xi = [\bar{S}]^{\overline{\mathscr{U}}}$; by (A.4.3) $i[\bar{S}]^{\overline{\mathscr{U}}} = [\bar{S}]^{\mathscr{U}}$).

(2) follows from (B.3.5) and associativity of intersection (A.6.1):

$$\begin{aligned}[\bar{S}]^{\mathscr{U}} \cdot [i\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}} &= ([S]^{\mathscr{U}} \cdot [\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}}) \cdot [i\mathbb{R} \times \mathbb{C}^n]^{\mathscr{U}} \\ &= [S]^{\mathscr{U}} \cdot [0 \times \mathbb{C}^n]^{\mathscr{U}}\end{aligned}$$ q.e.d.

Similarly:

(B.5.3) $$[\bar{T}]^{\mathscr{V}} \cdot [0 \times \mathbb{C}^n]^{\mathscr{V}} = \sum f_i [\![0 \times B_i]\!]^{\mathscr{V}}.$$

(B.6) We can now outline the approach we shall take.

By (B.2.1) and (B.1.2) $C(A)$ is pure d-dimensional; let $A_1, \ldots, A_k$ be its components. Then $B_1 = \alpha_0(A_1), \ldots, B_k = \alpha_0(A_k)$ are the components of $C(B, O)$ (by (B.1.3), (A.8)).

Let $C_H(A) = \sum e_i [\![0 \times A_i]\!]^{\mathscr{U}}$ (resp. $C_H(B) = \sum f_i [\![0 \times A_i]\!]^{\mathscr{V}}$) be the homology tangent cone of A (resp. B).

In view of (B.2.3) ($\mu(A_i) = \mu(B_i)$), the theorem results from the following assertions:

(1) $$\mu(A) = \sum_i e_i \mu(A_i), \qquad \mu(B) = \sum f_i \mu(B_i).$$

(2) $$e_i = f_i.$$

These are proved in sections C and D respectively.

C. Multiplicity of the Homology Tangent Cone

In this section we prove assertion (1) of (B.6). Notation remains as in (B.3).

(C.1) Let $\pi: \mathbb{C}^n \to \mathbb{C}^d$ ($d = \dim A$) be a complex linear projection ($\pi^2 = \pi$) such that $\pi^{-1}(O) \cap C(A) = O$. (It is easy to see that such π do exist.) We write $L = \pi^{-1}(O) \cap U$.

The idea of the proof (quite similar to that of [5, Thm. 6.4]) is to interpret multiplicity as the intersection number $[A]^U \circ [L]^U$ between A and $\pi^{-1}(O)$. It is easy to see that $[A]^U \circ [L]^U = [1 \times A]^{\mathscr{U}} \circ [\mathbb{C} \times L]^{\mathscr{U}}$ which in turn can be shown to be the intersection number $[S]^{\mathscr{U}} \circ [1 \times L]^{\mathscr{U}}$ (using the associativity of intersection and the fact that S and $1 \times U$ intersect transversally). Similarly $\sum e_i \mu(A_i) = \sum e_i [A_i]^U \circ [L]^U$ can be shown to be the intersection number $[S]^{\mathscr{U}} \circ [0 \times L]^{\mathscr{U}}$. To complete the proof we need only show that $[S]^{\mathscr{U}} \circ [1 \times L]^{\mathscr{U}} = [S]^{\mathscr{U}} \circ [0 \times L]^{\mathscr{U}}$. This is done by showing that both numbers are equal to the value of $[S]^{\mathscr{U}} \cdot [\mathbb{C} \times L]^{\mathscr{U}} \in H_2^{\mathbb{C} \times O}(\mathscr{U}) \cong \mathbb{Z}$. Details are given in (C.2), (C.3).

(C.2) By shrinking U if necessary (this is O.K. by (B.2.2), (B.4.2), (B.2.1)) we can assume that $L \cap A = O$ (because near the origin A is "close" to $C(A)$, cf. [20, p. 215, 2.10]); and that $(A \cap U, \pi|_{A \cap U})$ is a branched covering over an open connected subset $W = \pi(A \cap U)$ of $\mathbb{C}^d$, i.e. $\pi|_{A \cap U}$ is a topological covering outside an analytic variety D in W [10, p. 101, Th. 16]. By (B.2, Remark 1) $\mu(A)$ = covering number of $\pi = \#(A \cap U \cap \pi^{-1}(a))$ for $a \in W - D$. Then for $a \in W - D$, $\pi^{-1}(a)$ intersects A transversally at smooth points of A [10, p. 106, Th. 19]. Moreover, joining a to O by an arc γ in W, we have that $\pi^{-1}(\gamma) \cap A$ is compact (because π is proper). Hence, by [3, p. 485, 4.10], $\#(A \cap \pi^{-1}(a)) = [A]^U \circ [L]^U$. Thus

$$\mu(A) = [A]^U \circ [L]^U. \tag{C.2.1}$$

Next we apply (A.7.2) with $V = U$, $W = \mathscr{U}$, $Y = (\mathbb{C} \times L) \cap \mathscr{U}$, $X = Y \cap U = 1 \times L$, $\Phi = \Phi' = A$ and $\xi = [A]^U$. (Here we identify U with $1 \times U$ and A with $1 \times A$). Since $i^{A \cap 1 \times L, A \cap \mathbb{C} \times L} (= i^{1 \times O, 1 \times O})$ and $i^{A,A}$ are isomorphisms, (A.7.2) gives:

$$[A]^U \circ [L]^U = [1 \times A]^{\mathscr{U}} \circ [\mathbb{C} \times L]^{\mathscr{U}}.$$

(C.2.2) Together with (C.2.1) we have: $\mu(A) = [1 \times A]^{\mathscr{U}} \circ [\mathbb{C} \times L]^{\mathscr{U}}$. Similarly,

$$\mu(A_i) = [0 \times A_i]^{\mathscr{U}} \circ [\mathbb{C} \times L]^{\mathscr{U}}.$$

Remark. Recall that $S \cap (1 \times U) = 1 \times A$. In fact the intersection is transversal. Let $s = (1, z_0)$ be a smooth point of S. By (B.3.1) $\phi: (t, z) \to \left(t, \frac{1}{t} z\right)$ induces an isomorphism $d\phi$ between $\mathbb{C} \times T_{z_0}(A)$, (the tangent space of $\mathbb{C} \times A$ at s) and $T_s S$ (the tangent space of S at $(1, z_0)$). Straightforward computations show that $d\phi(1, O)_s = (1, -z_0)_s$. (Here $O \in \mathbb{C}^n$). Hence $T_s S$ and $T_s(1 \times U) = (0 \times \mathbb{C}^n)_s$ generate $(\mathbb{C} \times \mathbb{C}^n)_s$ i.e. S intersects $1 \times U$ transversally at smooth points of S which lie in $1 \times U$.

By the above remark and the criterion of unit multiplicity (A.6.3) we can conclude

$$[S]^{\mathscr{U}} \cdot [1 \times U]^{\mathscr{U}} = [1 \times A]^{\mathscr{U}}.$$

And the Eq. (C.2.2) becomes, by associativity (A.6.1),

$$\mu(A) = [S]^{\mathcal{U}} \circ ([1 \times U]^{\mathcal{U}} \cdot [\mathbb{C} \times L]^{\mathcal{U}}).$$

But $1 \times U$ is transversal to $\mathbb{C} \times L$, hence, by (A.6.3),

(C.2.3) $$\mu(A) = [S]^{\mathcal{U}} \circ [1 \times L]^{\mathcal{U}}.$$

Similar arguments also give

$$\begin{aligned}[S]^{\mathcal{U}} \circ [0 \times L]^{\mathcal{U}} &= [S]^{\mathcal{U}} \circ ([0 \times \mathbb{C}^n]^{\mathcal{U}} \cdot [\mathbb{C} \times L]^{\mathcal{U}}) \\ &= ([S]^{\mathcal{U}} \cdot [0 \times \mathbb{C}^n]^{\mathcal{U}}) \circ [\mathbb{C} \times L]^{\mathcal{U}} \\ &= \sum e_i [\![0 \times A_i]\!]^{\mathcal{U}} \circ [\mathbb{C} \times L]^{\mathcal{U}} \\ &= \sum e_i ([\![0 \times A_i]\!]^{\mathcal{U}} \circ [\mathbb{C} \times L]^{\mathcal{U}}).\end{aligned}$$

(We have used (A.6.3), (A.6.1), (B.4.2), (B.4.4) here.)

In view of the definition of $[\![0 \times A_i]\!]^{\mathcal{U}}$ (B.4.1), by the compatibility between intersections and enlargement of supports (A.6.1) we have

$$[\![0 \times A_i]\!]^{\mathcal{U}} \circ [\mathbb{C} \times L]^{\mathcal{U}} = [0 \times A_i]^{\mathcal{U}} \circ [\mathbb{C} \times L]^{\mathcal{U}},$$

which in turn is equal to $\mu(A_i)$ by (C.2.2).

Altogether we conclude

(C.2.4) $$[S]^{\mathcal{U}} \circ [0 \times L]^{\mathcal{U}} = \sum e_i \mu(A_i).$$

(C.3) In view of (C.2.3) and (C.2.4), to complete the proof we are left with proving

$$[S]^{\mathcal{U}} \circ [1 \times L]^{\mathcal{U}} = [S]^{\mathcal{U}} \circ [0 \times L]^{\mathcal{U}}.$$

Proof. By associativity (A.6.1), for any $t \in \mathbb{C}$ we have

(C.3.1) $$[S]^{\mathcal{U}} \cdot ([\mathbb{C} \times L]^{\mathcal{U}} \cdot [t \times \mathbb{C}^n]^{\mathcal{U}}) = ([S]^{\mathcal{U}} \cdot [\mathbb{C} \times L]^{\mathcal{U}}) \cdot [t \times \mathbb{C}^n]^{\mathcal{U}}.$$

By the criterion of unit multiplicity (A.6.3), the left side of (C.3.1) is $[S]^{\mathcal{U}} \cdot [t \times L]^{\mathcal{U}}$. On the other hand, since $L \cap (1/t) A = O$ and $L \cap C(A) = O$, we see that $S \cap (\mathbb{C} \times L) = \mathbb{C} \times O$, and (since $H_2^{\mathbb{C} \times O}(\mathcal{U}) \simeq \mathbb{Z}$ (A.3.3), (A.4.2)) therefore $[S]^{\mathcal{U}} \cdot [\mathbb{C} \times L]^{\mathcal{U}} = \lambda [\mathbb{C} \times O]^{\mathcal{U}}$ for some $\lambda \in \mathbb{Z}$; hence, by (A.6.3) again, the right side of (C.3.1) is $\lambda [t \times O]^{\mathcal{U}}$. Thus:

$$[S]^{\mathcal{U}} \circ [t \times L]^{\mathcal{U}} = |\lambda| \quad \text{for all } t \in \mathbb{C}.$$

In particular $[S]^{\mathcal{U}} \circ [1 \times L]^{\mathcal{U}} = [S]^{\mathcal{U}} \circ [0 \times L]^{\mathcal{U}}$. q.e.d.

Remark (not needed elsewhere). If we define the *local degree* of $\pi: A \to \mathbb{C}^d$ (denoted by $\deg \pi|_A$) to be $|\pi_*[A]|$ where $\pi_*: H_{2d}(A) \to H_{2d}(\mathbb{C}^d) \simeq \mathbb{Z}$, then it can be shown (using the projection formula (A.7)) that

$$\pm \deg \pi|_A = [A]^U \circ [L]^U.$$

(In fact we need only that π is continuous and $\pi^{-1}(O) \cap C(A)$ has an isolated point at O.)

Moreover, if π is real-linear (or even "close to" real linear, e.g. differentiable) then the argument in this section gives

$$\deg \pi|_A = \sum \pm e_i \deg \pi|_{A_i}.$$

D. The e_i are Differential Invariants

Consider the situation in Section B (especially (B.6)), i.e. we have homeomorphisms $\alpha: (U, A, O) \to (V, B, O)$, $\alpha_0: (C(A), A_i) \to (C(B), B_i)$ and sets $\mathscr{U}, S, \bar{\mathscr{U}}, \bar{S}$ (resp. $\mathscr{V}, T, \bar{\mathscr{V}}, \bar{T}$) which are derived from (U, A) (resp. (V, B)). Here α_0 is induced by a *real-linear* map $\alpha_0: \mathbb{C}^n \to \mathbb{C}^n$ (the derivative at O of α). Note that $\dim_{\mathbb{R}} U = 2n$, $\dim_{\mathbb{R}} \mathscr{U} = 2n+2$, $\dim_{\mathbb{R}} \bar{\mathscr{U}} = 2n+1$, $\dim_{\mathbb{R}} A = \dim_{\mathbb{R}} A_i = 2d$, $\dim_{\mathbb{R}} S = 2d+2$, $\dim_{\mathbb{R}} \bar{S} = 2d+1$. Also recall that the natural numbers e_i (resp. f_i) are uniquely determined by the equation (B.5.2) (resp. (B.5.3)):

$$[\bar{S}]^{\bar{\mathscr{U}}} \cdot [0 \times \mathbb{C}^n]^{\bar{\mathscr{U}}} = \sum e_i [\![0 \times A_i]\!]^{\bar{\mathscr{U}}}$$
$$(\text{resp. } [\bar{T}]^{\bar{\mathscr{V}}} \cdot [0 \times \mathbb{C}^n]^{\bar{\mathscr{V}}} = \sum f_i [\![0 \times B_i]\!]^{\bar{\mathscr{V}}}).$$

(D.1) We show in this section that $e_i = f_i$.

The approach is the following. First we construct a homeomorphism h from $(\bar{\mathscr{U}}, \bar{S})$ to $(\bar{\mathscr{V}}, \bar{T})$ which respects the fibres, i.e. maps $\bar{\mathscr{U}} \cap (t \times \mathbb{C}^n)$ to $\bar{\mathscr{V}} \cap (t \times \mathbb{C}^n)$ for all $t \in \mathbb{R}$. (We can not do so for $(\mathscr{U}, S)$ and $(\mathscr{V}, T)$, and this is the sole reason for introducing $\bar{\mathscr{U}}, \bar{\mathscr{V}}$). Since homeomorphism preserves intersections (A.7.3), we have:

$$\begin{aligned} h_*[\bar{S}]^{\bar{\mathscr{U}}} \cdot h_*[0 \times \mathbb{C}^n]^{\bar{\mathscr{U}}} &= \pm h_*([\bar{S}]^{\bar{\mathscr{U}}} \cdot [0 \times \mathbb{C}^n]^{\bar{\mathscr{U}}}) \\ &= \pm \sum e_i h_* [\![0 \times A_i]\!]^{\bar{\mathscr{U}}}. \end{aligned}$$

On the other hand, we have (by definition)

$$[\bar{T}]^{\bar{\mathscr{V}}} \cdot [0 \times \mathbb{C}^n]^{\bar{\mathscr{V}}} = \sum f_i [\![0 \times B_i]\!]^{\bar{\mathscr{V}}}.$$

Now it is easy to see that: $h_*[0 \times \mathbb{C}^n]^{\bar{\mathscr{U}}} = \pm [0 \times \mathbb{C}^n]^{\bar{\mathscr{V}}}$, $h_*[\![0 \times A_i]\!]^{\bar{\mathscr{U}}} = \pm [\![0 \times B_i]\!]^{\bar{\mathscr{V}}}$ (cf. (D.3.2)). In (D.4) we show: $h_*[\bar{S}]^{\bar{\mathscr{U}}} = \pm [\bar{T}]^{\bar{\mathscr{V}}}$. It follows that $e_i = \pm f_i$. But the e_i and f_i are all positive (B.4.4), hence $e_i = f_i$.

Remark. It follows that the "$\pm$" in $h_*[\![0 \times A_i]\!]^{\bar{\mathscr{U}}} = \pm [\![0 \times B_i]\!]^{\bar{\mathscr{V}}}$ is independent of i.

(D.2) We now carry out the above scheme.

Definition. For $(t, x) \in \mathscr{U}$,

$$h(t, x) = \begin{cases} (t, t^{-1} \alpha(t x)) & t \neq 0 \\ (0, \alpha_0(x)) & t = 0. \end{cases}$$

Notation. For simplicity we also write h for restrictions of h to subsets of $\mathscr{U}$; and h_* will always mean $h_i^{\Phi, h(\Phi)}(X)$ for some subset X of $\mathscr{U}$ and family of supports Φ which will be clear from the context.

We check that h has the desired properties.

(D.2.1) *h maps $(\mathscr{U}, S)$ into $(\mathscr{V}, T)$ and $(\bar{\mathscr{U}}, \bar{S})$ into $(\bar{\mathscr{V}}, \bar{T})$.*

Proof. The first assertion follows easily from definitions and from $S \cap (0 \times \mathbb{C}^n) = 0 \times C(A)$ (B.3.3) and $\alpha_0(C(A)) = C(B)$ (B.1.3). The proof of the second assertion is similar.

(D.2.2) *The maps in* (D.2.1) *are bijective.* In fact the inverse is defined by

$$h^{-1}(t,x) = \begin{cases} (t, t^{-1}\alpha^{-1}(t\,x)) & t \neq 0 \\ (0, \alpha_0^{-1}(x)) & t = 0. \end{cases}$$

(D.2.3) *$h: \bar{\mathscr{U}} \to \bar{\mathscr{V}}$ is a homeomorphism* (hence so is $h: \bar{S} \to \bar{T}$).

Proof. By (D.2.2) it suffices to show that h is continuous (then by symmetry h^{-1} is continuous too). It is clear that h is continuous at all points $(t,x) \in \bar{\mathscr{U}}$ with $t \in \mathbb{R}^*$, so we consider points $(0,x) \in \bar{\mathscr{U}}$, i.e. $x \in \mathbb{C}^n$. Let $(t,y) \in \bar{\mathscr{U}}$ be a nearby point.

Case 1. If $t = 0$, then

$$\begin{aligned} h(0,y) - h(0,x) &= (0, \alpha_0(y)) - (0, \alpha_0(x)) \\ &= (0, \alpha_0(y-x)). \end{aligned}$$

Case 2. If $t \in \mathbb{R}^*$, then

$$\begin{aligned} h(t,y) - h(0,x) &= (t, t^{-1}\alpha(t\,y)) - (0, \alpha_0(x)) \\ &= (t, t^{-1}\alpha(t\,y) - \alpha_0(x)). \end{aligned}$$

We estimate the second coordinate:

$$\begin{aligned} \left| \frac{1}{t}\alpha(t\,y) - \alpha_0(x) \right| &\leqq \left| \frac{1}{t}\alpha(t\,y) - \frac{1}{t}\alpha_0(t\,y) \right| + \left| \frac{1}{t}\alpha_0(t\,y) - \alpha_0(x) \right| \\ &= \left| \frac{\alpha(t\,y) - \alpha(0) - \alpha_0(t\,y)}{t} \right| + |\alpha_0(y) - \alpha_0(x)| \\ &= |y - x + x| \frac{|\alpha(t\,y) - \alpha(0) - \alpha_0(t\,y)|}{|t\,y|} + |\alpha_0(y-x)|. \end{aligned}$$

The first term tends to 0 with t and $|y-x|$ (by the definition of derivative). Hence, in either case, $h(t,y) - h(0,x)$ can be made arbitrarily small if (t,y) is close enough to $(0,x)$, i.e. h is continuous at $(0,x)$.

(D.2.4) The same proof breaks down when we attempt to prove the continuity of $h: \mathscr{U} \to \mathscr{V}$ (because α_0 is only *real*-linear). However, h clearly induces a homeomorphism from $\mathscr{U} - (0 \times \mathbb{C}^n)$ to $\mathscr{V} - (0 \times \mathbb{C}^n)$; and this fact will be needed in (D.4.1).

(D.3) In this subsection (D.3) everything happens in $\bar{\mathscr{U}} = \mathscr{U} \cap (\mathbb{R} \times \mathbb{C}^n)$.

With the homeomorphism h defined above we can conclude, using (A.7.3) and (B.5.2), that

(D.3.1)
$$\begin{aligned} h_*[\bar{S}]^{\mathscr{U}} \cdot h_*[0\times\mathbb{C}^n]^{\mathscr{U}} &= \pm h_*([\bar{S}]^{\mathscr{U}} \cdot [0\times\mathbb{C}^n]^{\mathscr{U}}) \\ &= \pm h_*(\sum e_i [\![0\times A_i]\!]^{\mathscr{U}}) \\ &= \pm \sum e_i h_* [\![0\times A_i]\!]^{\mathscr{U}}), \end{aligned}$$

where $h_*[\![0\times A_i]\!]^{\mathscr{U}} \in H_{2d}^{0\times C(B)}(\overline{\mathscr{V}})$.

(D.3.2) *Remark.* It is clear that $h_*[0\times\mathbb{C}^n]^{\mathscr{U}} = \pm[0\times\mathbb{C}^n]^{\mathscr{V}}$ (h_* induces an isomorphism between $H_{2n}^{0\times\mathbb{C}^n}(\overline{\mathscr{U}}) \simeq \mathbb{Z}$ and $H_{2n}^{0\times\mathbb{C}^n}(\overline{\mathscr{V}}) \simeq \mathbb{Z}$ (A.3.3), (A.4.1)).

We show that $h_*[\![0\times A_i]\!]^{\mathscr{U}} = \pm[\![0\times B_i]\!]^{\mathscr{V}}$. Recall the Definition (B.5)

$$[\![0\times A_i]\!]^{\mathscr{U}} = i^{0\times A_i,\, 0\times C(A)}[0\times A_i],$$
$$[\![0\times B_i]\!]^{\mathscr{V}} = i^{0\times B_i,\, 0\times C(B)}[0\times B_i].$$

Since $h = h|_{0\times A_i} = 0\times(\alpha_0|_{0\times A_i})$ induces an isomorphism $\mathbb{Z} \simeq H_{2d}(0\times A_i) \to H_{2d}(0\times B_i) \simeq \mathbb{Z}$ (A.4.1), we have $h_*[0\times A_i] = \pm[0\times B_i]$. The result then follows from the commutativity of the following diagram (A.3.1):

$$\begin{array}{ccc} H(0\times A_i) & \xrightarrow{i^{0\times A_i,\, 0\times C(A)}} & H^{0\times C(A)}(\overline{\mathscr{U}}) \\ \downarrow{\scriptstyle h_*} & & \downarrow{\scriptstyle h_*} \\ H(0\times B_i) & \xrightarrow[i^{0\times B_i,\, 0\times C(B)}]{} & H^{0\times C(B)}(\overline{\mathscr{V}}) \end{array}$$

(D.4) As in (D.1), it remains to show that $h_*[\bar{S}]^{\mathscr{U}} = \pm[\bar{T}]^{\mathscr{V}}$; and for this *it suffices to prove:* $h_*[\bar{S}] = \pm[\bar{T}]$ (because then using $h_* \circ i_{S,\overline{\mathscr{U}}} = i_{\bar{T},\overline{\mathscr{V}}} \circ h_*$ (A.3.1) we have

$$h_*[\bar{S}]^{\mathscr{U}} = h_* i_{S,\overline{\mathscr{U}}}[\bar{S}] = i_{\bar{T},\overline{\mathscr{V}}} h_*[\bar{S}] = \pm i_{\bar{T},\overline{\mathscr{V}}}[\bar{T}] = \pm[\bar{T}]^{\mathscr{V}}).$$

Idea of Proof. By definition $[\bar{S}]$ and $[\bar{T}]$ come from intersections that occur in $\mathscr{U}$ and $\mathscr{V}$ respectively. However h is continuous only on $\mathscr{U} - (0\times\mathbb{C}^n)$ (D.2.4). Hence we restrict the intersection (between S and $\mathbb{R}\times\mathbb{C}^n$) to $\mathscr{U} - (0\times\mathbb{C}^n)$ in order to obtain

$$h_*[\bar{S} - (0\times\mathbb{C}^n)] = \pm[\bar{T} - (0\times\mathbb{C}^n)].$$

This will then imply: $h_*[\bar{S}] = \pm[\bar{T}]$. We now give the details.

Notation. "Superscript *" means "leave out the part in $0\times\mathbb{C}^n$". Thus $\mathscr{U}^* = \mathscr{U} - (0\times\mathbb{C}^n)$, $\overline{\mathscr{U}}^* = \overline{\mathscr{U}} - (0\times\mathbb{C}^n)$, $S^* = S - (0\times\mathbb{C}^n)$ etc.

(D.4.1) *Remark 1.* If X is a closed subset of $\mathscr{U}$ having a fundamental class $[X]$, we write $[X^*] = j^{X,X^*}[X]$, cf. (A.3). By (A.4.3), $[X^*]$ is a fundamental class of X^*. It follows from (A.3.2) that

$$\begin{aligned} j^{\mathscr{U},\mathscr{U}^*}[X]^{\mathscr{U}} &= j^{\mathscr{U},\mathscr{U}^*} i_{X,\mathscr{U}}[X] = i_{X^*,\mathscr{U}^*} j^{X,X^*}[X] \\ &= i_{X^*,\mathscr{U}^*}[X^*] = [X^*]^{\mathscr{U}^*}. \end{aligned}$$

In particular, the restriction of $[S]^{\mathscr{U}}$ (resp. $[\bar{S}]^{\mathscr{U}}$, $[\overline{\mathscr{U}}]^{\mathscr{U}}$) to $\mathscr{U}^*$ is $[S^*]^{\mathscr{U}^*}$ (resp. $[\bar{S}^*]^{\mathscr{U}^*}$, $[\overline{\mathscr{U}}^*]^{\mathscr{U}^*}$). (N.B.: $\bar{S}^*$ and $\overline{\mathscr{U}}^*$ are not connected.)

Remark 2. Since $[\bar{S}]^{\mathscr{U}} = [\bar{\mathscr{U}}]^{\mathscr{U}} \cdot [S]^{\mathscr{U}}$ (B.3.5), it follows from the above remark and the compatibility of intersection and restriction (A.6.1) that

$$[\bar{S}^*]^{\mathscr{U}^*} = [\bar{\mathscr{U}}^*]^{\mathscr{U}^*} \cdot [S^*]^{\mathscr{U}^*}.$$

Similarly

$$[\bar{T}^*]^{\mathscr{V}^*} = [\bar{\mathscr{V}}^*]^{\mathscr{V}^*} \cdot [T^*]^{\mathscr{V}}.$$

(D.4.2) *Claim.* $h_*[\bar{S}^*]^{\mathscr{U}^*} = \pm[\bar{T}^*]^{\mathscr{V}^*}$.

Proof. In view of (D.4.1, Remark 2) the assertion will result from (A.7.3) if we can show

(a) $h_*[S^*]^{\mathscr{U}^*} = \pm[T^*]^{\mathscr{V}^*}$ and
(b) $h_*[\bar{\mathscr{U}}^*]^{\mathscr{U}^*} = \pm[\bar{\mathscr{V}}^*]^{\mathscr{V}^*}$.

Proof of (a). Recall that S^* (resp. T^*) is an irreducible analytic variety (B.3.2). By (A.4.1), $H^{S^*}_{2d+2}(\mathscr{U}^*) \simeq H_{2d+2}(S^*) \simeq \mathbb{Z}$ (resp. $H^{T^*}_{2d+2}(\mathscr{V}^*) \simeq \mathbb{Z}$) and has the fundamental class $[S^*]^{\mathscr{U}^*}$ (resp. $[T^*]^{\mathscr{V}^*}$) as a generator. Since h_* is an isomorphism, (a) is proved.

Proof of (b). By definition we have (cf. (D.4.1, Remark 1))

$$[\bar{\mathscr{U}}^*]^{\mathscr{U}^*} = i_{\bar{\mathscr{U}}^*, \mathscr{U}^*} \circ j^{\bar{\mathscr{U}}, \bar{\mathscr{U}}^*}[\bar{\mathscr{U}}],$$
$$[\bar{\mathscr{V}}^*]^{\mathscr{V}^*} = i_{\bar{\mathscr{V}}^*, \mathscr{V}^*} \circ j^{\bar{\mathscr{V}}, \bar{\mathscr{V}}^*}[\bar{\mathscr{V}}].$$

By (A.4.2), $[\bar{\mathscr{U}}]$ (resp. $[\bar{\mathscr{V}}]$) generates $H_{2n+1}(\bar{\mathscr{U}}) \simeq \mathbb{Z}$ (resp. $H_{2n+1}(\bar{\mathscr{V}}) \simeq \mathbb{Z}$). Thus $h_*[\bar{\mathscr{U}}] = \pm[\bar{\mathscr{V}}]$ (h_* is an isomorphism). By (A.3.2), (A.3.1),

$$h_*[\bar{\mathscr{U}}^*]^{\mathscr{U}^*} = \pm[\bar{\mathscr{V}}^*]^{\mathscr{V}^*}. \quad \text{q.e.d.}$$

(D.4.3) We are now ready to prove $h_*[\bar{S}] = \pm[\bar{T}]$ (D.4). Let $i_S = i_{\bar{S}^*, \mathscr{U}^*}$, $i_T = i_{\bar{T}^*, \mathscr{V}^*}$, $j_S = j^{\bar{S}, \bar{S}^*}$, $j_T = j^{\bar{T}, \bar{T}^*}$. Consider the following commutative diagram (A.3.1), (A.3.2):

$$\begin{array}{ccccc}
[\bar{S}] \in H_{2d+1}(\bar{S}) & \xrightarrow{j_S} & H_{2d+1}(\bar{S}^*) & \xrightarrow{i_S} & H^{\bar{S}^*}_{2d+1}(\mathscr{U}^*) \\
\downarrow ? \quad \downarrow h_* & & \downarrow h_* & & \downarrow h_* \\
\pm[T] \in H_{2d+1}(\bar{T}) & \xrightarrow{j_T} & H_{2d+1}(\bar{T}^*) & \xrightarrow{i_T} & H^{\bar{T}^*}_{2d+1}(\mathscr{V}^*)
\end{array}$$

As in the proof of (b) (D.4.2) we have (by definition)

$$i_S j_S[\bar{S}] = [\bar{S}^*]^{\mathscr{U}^*},$$

and similarly

(D.4.4) $$i_T j_T[\bar{T}] = [\bar{T}^*]^{\mathscr{V}^*}.$$

By the commutativity of the above diagram and (D.4.2) we see that

$$i_T j_T h_*[\bar{S}] = h_* i_S j_S[\bar{S}] = h_*[\bar{S}^*]^{\mathscr{U}^*} = \pm[\bar{T}^*]^{\mathscr{V}^*}.$$

By (D.4.4) it remains to show that $i_T \circ j_T$ is injective. But $i_T: H(\bar{T}^*) \to H^{\bar{T}^*}(\mathscr{V}^*)$ is an isomorphism (A.3.3), and the fact that j_T is injective can be seen from the following exact sequence (A.3.4) (keeping in mind $\bar{T} - \bar{T}^* = \bar{T} \cap (0 \times \mathbb{C}^n) = 0 \times C(B)$ (B.3.3)):

$$0 = H_{2d+1}(0 \times C(B)) \to H_{2d+1}(\bar{T}) \xrightarrow{j_T} H_{2d+1}(\bar{T}^*),$$

where $\dim_{\mathbb{R}} C(B) = 2d$, so by (A.4.1), $H_{2d+1} C(B) = 0$.

This completes the proof of (D.1), hence of the Theorem.

E. Appendix: Algebraic Definition of the e_i

(E.1) In this section we prove the following algebraic characterization of e_i. (It is an adaptation of folklore, cf. e.g. [14].)

Proposition. *Let $O \in A$, an analytic subvariety of $U \subseteq \mathbb{C}^n$ all of whose components have the same dimension, and let $\mathcal{O}_{A,O}$ be the local ring of germs at O of analytic functions on A. Let $m = m_{A,O}$ be the maximal ideal of $\mathcal{O}_{A,O}$, and let G be the graded ring $\bigoplus_{q \geq 0} m^q/m^{q+1}$. Then there is a natural one-one correspondence $A_i \leftrightarrow p_i$ between components of the tangent cone $C(A) = C(A, O)$ and minimal prime ideals in G, and e_i is the length of the artin local ring G_{p_i}.*

Before discussing the proof we note:

(E.1.1) **Corollary.** $e_i > 0$.

(E.1.2) From [18, (23.5)], applied to the localization of G at its maximal ideal $\bigoplus_{q>0} m^q/m^{q+1}$, we now obtain the main result of Sect. C:

$$\mu(A) = \sum_i e_i \mu(A_i).$$

However the proof in Sect. C is more in keeping with the spirit of the rest of the paper; and it gives a more general result - cf. the remark at the end of Sect. C.

(E.2) We begin the proof of the Proposition with some preliminary observations.

Let $f = f(z_1, \ldots, z_n)$ be a convergent power series, and write

$$f = f_d + f_{d+1} + \ldots + f_m + \ldots$$

where $f_m = f_m(z_1, \ldots, z_n)$ is a homogeneous polynomial of degree m, and $f_d \neq 0$.

Recall [20, p. 221, Theorem 4D], that $C(A)$ is the algebraic subvariety of $\mathbb{C}^n$ defined by the vanishing of all the initial forms f_d as f runs through the power series vanishing in a neighborhood N of O on A (here both d and N depend on f). Accordingly, we consider $C(A)$ to be the (not necessarily reduced!) affine algebraic variety with coordinate ring $\mathbb{C}[z_1, \ldots, z_n]/\{f_d(z_1, \ldots, z_n)\}$ (factoring out all f_d as above), which is nothing but the graded ring G in the Proposition. So the minimal primes of G correspond naturally to the components of $C(A)$. ("Components" are the same in the algebraic or analytic categories, since irreducible algebraic varieties have connected smooth part, so are irreducible as analytic varieties.) These minimal primes are homogeneous, so each component of $C(A)$ is mapped into itself by multiplication by any complex number, and hence passes through O. Let us show that the map from the set of irreducible components of the *germ* of $C(A)$ at O to the set of algebraic components of $C(A)$, taking an irreducible germ to the unique algebraic component containing it, is *bijective* (whence each algebraic component is irreducible at 0).

Algebraically this map is realized by considering the algebraic local ring R of $C(A)$ at O (i.e. R is the localization of G at the maximal ideal $\bigoplus_{q>0} m^q/m^{q+1}$) as a subring of the analytic local ring

$$\tilde{R} = \mathbb{C}\langle z_1, \ldots, z_n \rangle / \{f_d(z_1, \ldots, z_n)\}$$

(where $\mathbb{C}\langle\ldots\rangle$ denotes "convergent power series") and associating to each minimal prime ideal $\tilde{p}$ of $\tilde{R}$ its contraction $p=\tilde{p}\cap R$. Now R and $\tilde{R}$ clearly have the same completion, say $\hat{R}$. $\hat{R}/p\hat{R}$ is the completion of R/p, and if $\hat{m}$ is the maximal ideal of $\hat{R}/p\hat{R}$, then the graded ring $\bigoplus_{q\geq 0}\hat{m}^q/\hat{m}^{q+1}$ is isomorphic to the graded *integral domain* $G/p\cap G$; it follows that $\hat{R}/p\hat{R}$ is also an integral domain, i.e. *$p\hat{R}$ is a prime ideal.* So

$$p\hat{R}=[(p\tilde{R})\hat{R}]\cap\tilde{R}=p\hat{R}\cap\tilde{R}$$

is a prime ideal contained in $\tilde{p}$, and since $\tilde{p}$ is minimal, therefore $p\tilde{R}=\tilde{p}$, so that our map is injective. Also, since *every* minimal prime in R is the contraction of a prime ideal in $\hat{R}$, our map is surjective (this is also geometrically clear), hence bijective.

In summary, we have:

Lemma. *Let A_i be an algebraic component of $C(A)$, and let p_i be the corresponding minimal prime ideal in G. Then A_i is analytically irreducible at 0, and if $\tilde{p}_i$ is the corresponding minimal prime ideal in the analytic local ring $\tilde{R}$, then $p_i=\tilde{p}_i\cap G$, $\tilde{p}_i=p_i\tilde{R}$, and the artin local rings G_{p_i}, $\tilde{R}_{\tilde{p}_i}$ have the same length.*

Proof. Everything has been proved except the last assertion, which follows from $\tilde{p}_i=p_i\tilde{R}$ and the flatness of $\tilde{R}_{\tilde{p}_i}$ over G_{p_i}. (Note that $\tilde{R}$ is flat over R since $\hat{R}$ is faithfully flat over both $\tilde{R}$ and R.)

(E.3) Lemma. *Let $S\subseteq\mathbb{C}\times\mathbb{C}^n$ be the analytic set defined in* (B.3), *and let $t:\mathbb{C}\times\mathbb{C}^n\to\mathbb{C}$ be the projection. Let $\tilde{R}$ be as in* (E.2) *above. Then*

$$\tilde{R}\cong\mathcal{O}_{S,O}/t\mathcal{O}_{S,O}.$$

After proving this Lemma, and in view of the preceding Lemma (E.2), we can get the Proposition from the following fact, which expresses the *equality of algebraic and topological intersection numbers.* (This fact is "well-known", cf. [3, p. 488, 4.16] and for zero-dimensional intersections [5, Theorem 6.5] or [17, p. 118, (A.8)]; but we don't know of an adequate reference, so a proof will be given.)

Fact. *Let L, S be pure-dimensional closed analytic subvarieties of an open set $U\subseteq\mathbb{C}^m$, and let B be an irreducible component of $L\cap S$ such that*

$$\operatorname{codim} B=\operatorname{codim} L+\operatorname{codim} S$$

(where "codim" denotes "codimension in $\mathbb{C}^m$"). The topological intersection number $i(L.S, B)$ is defined as in [3, p. 482, 4.4]. *Assume that B contains a point where L is smooth, and that $O\in B$. Let B^* be a component of the germ of B at O, set $R=\mathcal{O}_{S,O}$, and let p be the prime ideal in R consisting of germs of functions vanishing on B^*. Let I be the $\mathcal{O}_{\mathbb{C}^m,O}$-ideal consisting of germs of functions vanishing on L, so that (by standard local analytic geometry) IR_p is primary for the maximal ideal of R_p. Then the multiplicity of the ideal IR_p is $i(L.S, B)$.*

In (E.4) below we will prove the Lemma, and in (E.5) we will prove the Fact.

Finally, to get the Proposition, apply the Fact with L given by $t=0$ (notation as in Lemma), so that $I=t\mathcal{O}_{S,O}$ and $L\cap S=0\times C(A)$; and with B the component $0\times A_i$ of $L\cap S$. Then

$$\begin{aligned}
e_i &= i(L.S, B) && \text{(Definition (B.4.2))}\\
&= \text{multiplicity of } t(\mathcal{O}_{S,O})_p && \text{(by Fact)}\\
&= \text{length of } (\mathcal{O}_{S,O})_p/(t)\\
&= \text{length of } \tilde{R}_{\tilde{p}_i} && \text{(by Lemma)}\\
&= \text{length of } G_{p_i} && \text{(cf. (E.2)).}
\end{aligned}$$

(Note: from the proof of the Lemma, it will be clear that if $p\subset\mathcal{O}_{S,O}$ is the ideal of germs vanishing on $0\times A_i$, then $p/(t)=\tilde{p}_i$.)

(E.4) (Proof of Lemma). As before, represent a convergent power series f as a sum of homogeneous polynomials:

$$f=f_d+f_{d+1}+f_{d+2}+\ldots \qquad (f_d\neq 0)$$

and set

$$f^*(t, z_1, \ldots, z_n) = t^{-d} f(t z_1, \ldots, t z_n) = f_d + t f_{d+1} + t^2 f_{d+2} + \ldots.$$

We call f_d the "initial form" of f.

Let $I \subseteq \mathcal{O}_{\mathbb{C}^n, O}$ be the ideal of germs of functions vanishing on A, and let $F \subseteq I$ be a finite set of generators of I whose initial forms generate the kernel of the graded $\mathbb{C}$-algebra homomorphism $\mathbb{C}[Z_1, \ldots, Z_n] \to G$ taking the indeterminate Z_i to the natural image of the corresponding coordinate function z_i in $\mathfrak{m}/\mathfrak{m}^2 \subseteq G$. We will show that *the ideal of germs in $\mathcal{O}_{\mathbb{C}^{n+1}, O}$ vanishing on S is generated by $\{f^* | f \in F\}$*; the Lemma is then obvious.

With notations as in (B.3), let $N \subseteq U$ be a neighborhood of O such that all the $f \in F$ converge in N, and such that

$$(x_1, \ldots, x_n) \in A \cap N \Leftrightarrow f(x_1, \ldots, x_n) = 0 \quad \forall f \in F.$$

Then if $x \ \mathbb{C} - \{0\}$, we see that

$$\begin{aligned} &(x, x_1, \ldots, x_n) \in S \cap \psi^{-1}(\mathbb{C} \times N) \\ &\quad \Leftrightarrow (x x_1, \ldots, x x_n) \in A \cap N \\ &\quad \Leftrightarrow f(x x_1, \ldots, x x_n) = 0 \qquad \forall f \in F \\ &\quad \Leftrightarrow f^*(x, x_1, \ldots, x_n) = 0 \qquad \forall f \in F; \end{aligned}$$

while if $x = 0$, then

$$\begin{aligned} &(x, x_1, \ldots, x_n) \in S \cap \psi^{-1}(\mathbb{C} \times N) \\ &\quad \Leftrightarrow (x_1, \ldots, x_n) \in C(A) \qquad \text{(B.3.3)} \\ &\quad \Leftrightarrow f_d(x_1, \ldots, x_n) = 0 \qquad \forall f \in F \ (d \text{ depends on } f) \\ &\quad \Leftrightarrow f^*(x, x_1, \ldots, x_n) = 0 \qquad \forall f \in F. \end{aligned}$$

Thus the vanishing of the $f^*(f \in F)$ defines the germ of S at O *set-theoretically*, and it remains to be shown that the ideal I^* generated by the f^* is a *radical ideal*, i.e. is equal to its own radical. Roughly speaking, this is true because I^* is a "proper transform" of the radical ideal $I\mathbb{C}\langle t, z, \ldots, z_n\rangle$ under blowing up. Here are the technical details.

Let H_0 be the convergent power series ring $\mathbb{C}\langle t, y_1, \ldots, y_n\rangle$, and let $I_y \subseteq H_0$ be the ideal generated by $\{f(y_1, \ldots, y_n) | f \in F\}$, so that $I_y = \sqrt{I_y}$. Define a $\mathbb{C}$-algebra homomorphism $\gamma_0 : H_0 \to \mathbb{C}\langle t, z_1, \ldots, z_n\rangle$ by

$$\gamma_0(y_i) = t z_i \quad (1 \leqq i \leqq n)$$
$$\gamma_0(t) = t.$$

Set

$$H = H_0[t^{-1} y_1, \ldots, t^{-1} y_n].$$

Then $y_i \in tH$ $(1 \leqq i \leqq n)$, and so for $f = f_d + \ldots$ as above, since $f \in (y_1, \ldots, y_n)^d H_0$, therefore

$$t^{-d} f(y_1, \ldots, y_d) \in H.$$

The ideal J in H generated by all such $t^{-d} f$ $(f \in F)$ is the "proper transform" of I_y, i.e.

$$J = (I_y H_0[t^{-1}]) \cap H$$

(cf. [13, p. 216, Lemma 6]). Clearly $J = \sqrt{J}$.

Now γ_0 extends to a map

$$\gamma : H \to \mathbb{C}\langle t, z_1, \ldots, z_n\rangle$$

with

$$\gamma(t^{-1} y_i) = z_i \quad 1 \leqq i \leqq n,$$

and the ideal I^* is the extension of J:

$$I^* = \gamma(J)\, \mathbb{C}\langle t, z_1, \ldots, z_n\rangle.$$

To show that $I^* = \sqrt{I^*}$, let M be the maximal ideal

$$M = (t, t^{-1} y_1, \ldots, t^{-1} y_n) H.$$

Then γ extends to a local homomorphism of $(n+1)$-dimensional regular local rings

$$\gamma_M : H_M \to \mathbb{C}\langle t, z_1, \ldots, z_n \rangle.$$

Since $M\mathbb{C}\langle t, z_1, \ldots, z_n \rangle$ is the unique maximal ideal, we see that γ_M extends to an *isomorphism of completions*

$$\hat{\gamma}_M : (H_M)\hat{} \simeq \mathbb{C}[[t, z_1, \ldots, z_n]].$$

Hence

$$I^* = (\hat{\gamma}_M(J)\, \mathbb{C}[[\cdot]]) \cap \mathbb{C}\langle \cdot \rangle;$$

and since H_M is excellent and $J = \sqrt{J}$, therefore $\hat{\gamma}_M(J)\, \mathbb{C}[[\cdot]]$ is a radical ideal, so the same is true of I^*. Q.E.D.

(E.5) (Proof of Fact). Suppose first that O is smooth both on L and on $L \cap S$. Then we can carry out, as follows, the standard reduction to the case $B = \{O\}$:

There is a coordinate system $(t_1, t_2, \ldots, t_m)$ in a neighborhood $V \subseteq U$ of O in $\mathbb{C}^m$ such that $L \cap V$ is given (ideal theoretically) by

$$t_1 = t_2 = \ldots = t_r = 0 \quad (r = \text{codim. } L)$$

and $L \cap S \cap V = B \cap V$ is given (set-theoretically) by

$$t_1 = \ldots = t_r = t_{r+1} = \ldots = t_{r+e} = 0 \quad (e = \text{codim. } S).$$

By [3, p. 483, Proposition 4.5]

$$i(L.S, B) = i((L \cap V).(S \cap V), B \cap V);$$

so we may assume that $V = U$.

Algebraically, we have $I = (t_1, \ldots, t_r)\, \mathcal{O}_{\mathbb{C}^m, O}$, $p = \sqrt{IR}$, and R/p is a regular local ring of dimension $m - e - r$, with maximal ideal $(t_{r+e+1}, \ldots, t_m)(R/p)$. Algebraic associativity for multiplicity (denoted "μ") [18, p. 81, (24.7)] gives

$$\begin{aligned} \mu((t_1, \ldots, t_r, t_{r+e+1}, \ldots, t_m) R) &= \mu((t_1, \ldots, t_r) R_p)\, \mu((t_{r+e+1}, \ldots, t_m)(R/p)) \\ &= \mu(IR_p). \end{aligned}$$

If M is the submanifold of U given by $t_{r+e+1} = \ldots = t_m = 0$, then topological associativity [3, p. 484, 4.7] gives

$$i(M.L, M \cap L)\, i((M \cap L).S, O) = i(L.S, B)\, i(M.B, O)$$

that is (cf. (A.6.3))

$$i((M \cap L).S, O) = i(L.S, B).$$

So it suffices to show that

$$\mu((t_1, \ldots, t_r, t_{r+e+1}, \ldots, t_m) R) = i((M \cap L.S, O).$$

Thus we may indeed assume that $B = \{O\}$; and this case can be settled by [3, p. 485, 4.10] and the references mentioned in (E.3) above.

To reduce the general case to the preceding one, note that the points of B which are not smooth on both L and $L \cap S$ form an analytic subvariety of B, which is nowhere dense in B (since B is irreducible and contains, by assumption, a smooth point of L). So we can choose $s \in B$ at which L and $L \cap S$ are smooth, and set $R' = \mathcal{O}_{S,s}$. Let $\mathcal{I}$ (resp. $\mathcal{P}$) be the $\mathcal{O}_{\mathbb{C}^m}$-ideal of functions vanishing on L (resp. B). Then $p' = \mathcal{P}R'$ is a prime ideal in R' (since B is smooth at s), and it will suffice to show that the ideals $IR_{p'}$, $\mathcal{I}R'_{p'}$ have the same multiplicity. (The multiplicity of $\mathcal{I}R'_{p'}$ is $i(L.S, B)$ by the preceding case.)

In fact we will show more, namely that the Hilbert-Samuel functions

$$\lambda(n) = \text{length of the } R_p\text{-module } I^n R_p / I^{n+1} R_p$$

and

$$\lambda'(n) = \text{length of the } R'_{p'}\text{-module } \mathscr{I}^n R'_{p'} / \mathscr{I}^{n+1} R'_{p'}$$

coincide.

For this purpose, observe that if $\bar{I} = IR$ and

$$E^{a,n} = [p^a(\bar{I}^n / I^{n+1})] / [p^{a+1}(\bar{I}^n / \bar{I}^{n+1})]$$

then

$$\lambda(n) = \sum_{a=0}^{\infty} \dim_{k(p)}(E^{a,n} \otimes_R R_p) = (\text{say}) \sum_{a=0}^{\infty} e^{a,n}$$

where "$\dim_{k(p)}$" denotes vector-space dimension over the field $k(p) = R_p / pR_p$. (Note that $e^{a,n} = 0$ if $p^a R_p \subseteq IR_p$, so the sum makes sense). Now if $\mathscr{I} = \mathscr{I}\mathcal{O}_S$ and $\mathscr{E}^{a,n}$ is the coherent $\mathcal{O}_B$-module

$$\mathscr{E}^{a,n} = [\mathscr{P}^a(\bar{\mathscr{I}}^n / \bar{\mathscr{I}}^{n+1})] / [\mathscr{P}^{a+1}(\bar{\mathscr{I}}^n / \bar{\mathscr{I}}^{n+1})]$$

then $E^{a,n}$ is the stalk $\mathscr{E}^{a,n}_0$ and from [15, p. 20–10, Prop. 6] we see that outside a nowhere dense subvariety of B, $\mathscr{E}^{a,n}$ is locally free of rank $e^{a,n}$. Thus for fixed a, n, $e^{a,n}$ depends only on $\mathscr{P}$ and $\bar{\mathscr{I}}$; and a similar argument with s in place of O gives

$$\lambda'(n) = \sum_{a=0}^{\infty} e^{a,n}.$$

References

1. A'Campo, N.: Le nombre de Lefschetz d'une monodromie. Proc. Kon. Ned. Akad Serie A **76**, 113–118 (1973)
2. Bloom, T.: C^1 functions on a complex analytic variety. Duke Math J. **36**, 283–296 (1969)
3. Borel, A., Haefliger, A.: La class d'homologie fondamentale d'un espace analytique. Bull. Soc. Math. France **89**, 461–513 (1961)
4. Borel, A., Moore, J.C.: Homology theory for locally compact spaces. Mich. Math. J. **7**, 137–159 (1960)
5. Draper, R.N.: Intersection theory in analytic geometry. Math. Ann. **180**, 175–204 (1969)
6. Dold, A.: Lectures on algebraic topology. Berlin-Heidelberg-New York: Springer 1972
7. Ephraim, R.: C^1 preservation of multiplicity. Duke Math. J. **43**, 797–803 (1976)
8. Ephraim, R.: The cartesian product structure and C^∞ equivalences of singularities. Trans. Amer. Math. Soc. **224**, 299–311 (1976)
9. Godement, R.: Topologie algébrique et thèorie des faisceaux. Paris: Hermann 1958
10. Gunning, R.C.: Lectures on complex analytic varieties: The local parametrization theorem. Math. Notes, Princeton, N.J.: Princeton Univ. Press, 1970
11. Gunning, R.C., Rossi, H.: Analytic functions of several complex variables. Englewood Cliffs, N.J.: Prentice-Hall, 1965
12. Hironaka, H.: Normal cones in analytic Whitney stratifications. Publ. Math. I.H.E.S. **36**, 127–138 (1969)
13. Hironaka, H.: Resolution of singularities of an algebraic variety over a field of characteristic zero. Annals of Math. **79**, 109–326 (1964)
14. Hénaut, A.: Cycles et Cone Tangent de Zariski, Fonctions de Plusieurs Variables Complexes IV, Séminaire François Norguet 1977–79. Lecture Notes in Mathematics, vol. 807, pp. 145–166. Berlin-Heidelberg-New York: Springer 1980
15. Houzel, C.: Géométrie analytique locale, exposé 20 in Séminaire Henri Cartan 1960/61, Familles d'espaces complexes et fondements de la géométrie analytique, Secrétariat mathématique, Paris, 1962

16. Lê, Dũng Tráng: Calcul du nombre de cycles évanouissants d'une hypersurface complexe, Ann. Inst. Fourier (Grenoble) **23**, 261–270 (1973)
17. Mumford, D.: Algebraic geometry I: complex projective varieties. Berlin-Heidelberg-New York: Springer 1976
18. Nagata, M.: Local rings. New York: John Wiley and Sons 1962
19. Spivak, M.: Calculus on manifolds. New York: Benjamin 1965
20. Whitney, H.: Complex analytic varieties. Addison-Wesley, 1972
21. Zariski, O.: On the topology of algebroid singularities. Amer. J. Math. **54**, 453–465 (1932)
22. Zariski, O., Samuel, P.: Commutative Algebra. Vol. 2. Princeton, N.J.: Van Nostrand 1960

Oblatum 7-IX-1981

Invent. math. 73, 189–239 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

The Lefschetz Number of an Involution on the Space of Harmonic Cusp Forms of SL_3

R. Lee[1] and J. Schwermer[2]

[1] Department of Mathematics, Yale University, New Haven, CT 06520, USA
[2] Mathematisches Institut der Universität, Wegelerstraße 10, D-5300 Bonn 1, Bundesrepublik Deutschland

Contents

§0. Introduction 189
§1. Elementary Properties of $\Gamma_Q(m)$ 194
§2. Cohomology of $\Gamma_Q(m)$ at Infinity 196
§3. The Lefschetz Number at Infinity 200
§4. Lefschetz Number and Ramified Covering Spaces 208
§5. The Euler Characteristic of the Fixpoint Set F^σ 209
§6. Fixpoint Sets of Other Involutions 221
§7. Order Four Elements 225
§8. Fixpoint Set $F(\gamma_0)$ 229
§9. Fixpoint Set $F(\gamma_1)$ 231
References 238

§ 0. Introduction

The analysis of involutions (or more general, of finite group actions) plays an important role in the study of the arithmetic of modular curves and in the classical theory of modular forms. This appears already in the work of Fricke [7] or Hecke [12].

In this paper we deal with an arithmetic torsionfree subgroup Γ of $SL_3(\mathbb{Q})$ and study an involution σ on the arithmetic quotient X/Γ of the associated symmetric space $X = SO(3)\backslash SL_3(\mathbb{R})$ by Γ. Using a method of Rohlfs [19] and our previous work [16] an analysis of the Lefschetz-number of the induced map σ^*: $H^*(X/\Gamma, \mathbb{C}) \to H^*(X/\Gamma; \mathbb{C})$ in cohomology allows us to obtain results on the cusp cohomology $H^*_{\text{cusp}}(\Gamma, \mathbb{C})$ of Γ; by definition (cf. [2] or [3], 5.4), this is a natural subspace of the cohomology $H^*(X/\Gamma, \mathbb{C})$ and can be characterized as the space of harmonic cuspidal differential forms on X/Γ. We will prove the following main result.

Offprint requests to: J. Schwermer

Theorem. *Let* $\Gamma(m)=\{A\in SL_3(\mathbb{Z})\mid A\equiv \mathrm{Id} \bmod m\}$ *be the full congruence subgroup of level* m, *where* m *is a prime with* $m\equiv 3 \bmod 8$ *and* $m\equiv -1 \bmod 3$. *Then we have the following lower estimate for the dimension of the cusp cohomology of* $\Gamma(m)$

$$\dim_{\mathbb{C}} H^q_{\mathrm{cusp}}(\Gamma(m);\mathbb{C})\geqq m(m+1)$$

for $q=2, 3$. (*Note that* $H^q_{\mathrm{cusp}}(\Gamma(m);\mathbb{C})=0$ *for* $q\neq 2, 3$ *as pointed out in* [16], § 1.)

As an easy consequence one has

Corollary. *Let* Γ *be a subgroup of finite index in* $SL_3(\mathbb{Z})$, *and let* n *be a natural number. Then there exists a torsionfree subgroup* Γ' *of finite index in* Γ *with*

$$\dim_{\mathbb{C}} H^q_{\mathrm{cusp}}(\Gamma';\mathbb{C})\geqq n, \quad q=2, 3.$$

Remarks. (1) The interested reader may find analogous results for other groups than SL_3 or the classical case SL_2 (cf. [25]) in [6, 8, 9, 10, 14, 15, 17, 30].

(2) For $\Gamma_0(p)=\{A\in SL_3(\mathbb{Z})\mid A=(a_{ij}),\ a_{21}\equiv a_{31}\equiv 0(p)\}$ and $p=53, 61, 79, 89$ Ash has found in [1] by machine computation that $\dim H^2_{\mathrm{cusp}}(\Gamma_0(p),\mathbb{C})=2$.

We indicate now briefly the idea of the proof and the outline of the paper.

0.1. We derived in [16], 1.8 for a torsionfree subgroup Γ of finite index in $SL_3(\mathbb{Z})$ a short exact sequence ($q>0$)

$$0\to H^q_{\mathrm{cusp}}(\Gamma;\mathbb{C})\to H^q(\bar{X}/\Gamma,\mathbb{C})\to H^q_{\mathrm{inf}}(\bar{X}/\Gamma;\mathbb{C})\to 0 \tag{1}$$

where we have identified the cohomology of the 5-dimensional complete Riemannian manifold X/Γ with the cohomology of its compactification due to Borel and Serre [4]; by definition, the cohomology at infinity $H^*_{\mathrm{inf}}(\bar{X}/\Gamma,\mathbb{C})$ is the image of the natural restriction map of the cohomology of $\bar{X}/\Gamma$ to the cohomology $H^*(\partial(\bar{X}/\Gamma),\mathbb{C})$ of its boundary. Note that the cusp cohomology vanishes for $q\neq 2, 3$, and that $H^2_{\mathrm{cusp}}(\Gamma,\mathbb{C})\cong H^3_{\mathrm{cusp}}(\Gamma;\mathbb{C})$ (cf. [16], proof of 2.5 resp. 1.8(4)).

We consider an involution σ on $SL_3(\mathbb{R})$ such that Γ is σ-stable and the induced involution σ on X/Γ reverses the orientation. The Lefschetznumber of σ and Γ is then defined by

$$L(\sigma,\Gamma)=\sum_{q=0}^{\infty}(-1)^q \operatorname{trace}(\sigma^q:\ H^q(X/\Gamma,\mathbb{C})\to H^q(X/\Gamma,\mathbb{C})). \tag{2}$$

Since σ^* respects the terms in the short exact sequence (1) we get accordingly

$$L_{\mathrm{cusp}}(\sigma,\Gamma)=L(\sigma,\Gamma)-L_{\mathrm{inf}}(\sigma,\Gamma).$$

The isomorphism $H^2_{\mathrm{cusp}}(\Gamma,\mathbb{C})\cong H^3_{\mathrm{cusp}}(\Gamma,\mathbb{C})$ is given by the *-operator which depends on an orientation of X, and since σ is orientation reversing we have $\operatorname{tr}(\sigma^2_{\mathrm{cusp}})=-\operatorname{tr}(\sigma^3_{\mathrm{cusp}})$ in obvious notation. We obtain the formula

$$2\operatorname{trace}(\sigma^2_{\mathrm{cusp}})=L_{\mathrm{cusp}}(\sigma,\Gamma)=L(\sigma,\Gamma)-L_{\mathrm{inf}}(\sigma,\Gamma). \tag{3}$$

If we can prove now that the difference of the Lefschetz number $L(\sigma, \Gamma)$ and $L_{\inf}(\sigma, \Gamma)$ is non-zero for appropriately chosen $\Gamma \supset \Gamma(m)$ and involution σ we have as an immediate consequence our main result, namely the *non-vanishing of the cusp cohomology* $H^*_{\text{cusp}}(\Gamma(m); \mathbb{C})$. The actual estimate for $\dim H^q_{\text{cusp}}(\Gamma(m); \mathbb{C})$ is then only a matter of the following easy argument. Since $H^q(SL_3(\mathbb{Z}); \mathbb{C}) = H^q(\Gamma(m), \mathbb{C})^{SL_3(\mathbb{Z}/m\mathbb{Z})} = 0$ by Soulé [27] the trivial module doesn't occur in a decomposition of $H^q_{\text{cusp}}(\Gamma(m); \mathbb{C})$ into irreducible representations with respect to the natural action of $SL_3(\mathbb{Z}/m\mathbb{Z})$. It follows that the dimension of $H^q_{\text{cusp}}(\Gamma(m); \mathbb{C})$ if it doesn't vanish has as a lower bound the smallest possible dimension of a non-trivial irreducible representation of $SL_3(\mathbb{Z}/m\mathbb{Z})$. A short inspection of the character tables for $SL_3(\mathbb{Z}/m\mathbb{Z})$ [26] gives us then the estimate as asserted in the theorem.

Our computation of the total Lefschetz number is based on a method of Rohlfs developed in [19] in the case $\Gamma = \Gamma(m)$ and σ the Cartan involution τ on $SL_3(\mathbb{R})$. Unfortunately we cannot directly follow this approach because there are some difficulties in computing the Lefschetz number $L_{\inf}(\tau, \Gamma(m))$ at infinity. These are related to some arithmetic questions concerning special values of L-functions ([20, 21], § 8 for more details). Instead of the full congruence subgroup $\Gamma(m)$ we consider an intermediate group $\Gamma_Q(m) \supset \Gamma(m)$, for which the computation of the Lefschetz number at infinity is feasible. On the other hand, this group is close enough to $\Gamma(m)$ such that Rohlfs' method can be applied in a slightly modified way in order to compute its total Lefschetz number. These computations imply then by (3) and $H^*_{\text{cusp}}(\Gamma_Q(m), \mathbb{C}) = H^*_{\text{cusp}}(\Gamma(m), \mathbb{C})^{\Gamma_Q(m)/\Gamma(m)}$ the non-vanishing of $H^*_{\text{cusp}}(\Gamma(m), \mathbb{C})$.

For the proof of the corollary we put $\Gamma' = \Gamma \cap \Gamma(m)$ where m is chosen in such a way that $m(m+1) \geqq n$.

0.2. We begin in § 1 by considering a specific congruence subgroup Γ of $SL_3(\mathbb{Z})$ and specific involution σ. We fix a prime m; let Q denote the symmetric matrix

$$Q = \begin{pmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{1}$$

and $PO(Q)$ denote the projective orthogonal group in ${}_fSL = SL_3(\mathbb{Z}/m\mathbb{Z})$ consisting of elements which preserve Q up to a unit in $\mathbb{Z}/m\mathbb{Z}$. Our congruence subgroup $\Gamma = \Gamma_Q(m)$ is then defined as the subgroup of $SL_3(\mathbb{Z})$ of all matrices A such that A modulo m belongs to $PO(Q)$. We consider the 'twisted' Cartan involution $\sigma: SL_3(\mathbb{R}) \to SL_3(\mathbb{R})$ defined by

$$A \mapsto \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix} (A^t)^{-1} \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix}^{-1}. \tag{2}$$

By technical reasons we assume that m is a prime with $m \equiv 3 \bmod 8$ and $m \equiv -1 \bmod 3$. Under this assumption we show (in 1.2) that $\Gamma_Q(m)$ is a semidirect product

$$\Gamma_Q(m) = \Gamma_D(m) \rtimes \mathcal{O}, \tag{3}$$

where $\Gamma_D(m)$ is a torsionfree subgroup of $SL_3(\mathbb{Z})$ and $\mathcal{O}=\langle s,t|s^4=1,t^2=1,\ tst=s^{-1}\rangle$ the dihedral group of order 8.

0.3. Cohomology of $\Gamma_Q(m)$ at Infinity. In order to determine the cohomology of $\Gamma_Q(m)$ at infinity we note that $\Gamma_Q(m)\supset\Gamma(m)$. The cohomology of $\Gamma_Q(m)$ at infinity is then naturally given as $H^*_{\inf}(\bar{X}/\Gamma(m),\mathbb{C})^{\Gamma_Q(m)/\Gamma(m)}=H^*_{\inf}(\bar{X}/\Gamma_Q(m),\mathbb{C})$. Using our Theorem 2.4 in [16], where we studied $H^*_{\inf}(\bar{X}/\Gamma(m),\mathbb{C})$ as an $SL_3(\mathbb{Z}/m\mathbb{Z})$-module we compute $H^*_{\inf}(\bar{X}/\Gamma_Q(m),\mathbb{C})$ in §2. Let $\oplus H^3_{\text{cusp}}(e'(P),\mathbb{C})$ denote the cusp cohomology in degree 3 of all faces in $\partial(\bar{X}/\Gamma(m))$ corresponding to maximal parabolic $\mathbb{Q}$-subgroups of $SL_3(\mathbb{R})$; then $H^q_{\inf}(\bar{X}/\Gamma_Q(m),\mathbb{C})$ vanishes (resp. is equal to $\mathbb{C}$) for $q=1,2$ (resp. $q=0$) and in degree 3 we get

$$H^3_{\inf}(\bar{X}/\Gamma_Q(m),\mathbb{C})=[\widetilde{St}]^{PO(Q)}\oplus[\oplus H^3_{\text{cusp}}(e'(P),\mathbb{C})]^{PO(Q)} \tag{1}$$

where $[\widetilde{St}]^{PO(Q)}$ denotes the $PO(Q)$-invariant elements in the generalized Steinberg-module (defined in 2.1 (6)) for $SL_3(\mathbb{Z}/m\mathbb{Z})$. There is no contribution to the Lefschetznumber at infinity by the last term in (1). To deal with the other term we note that as $SL_3(\mathbb{Z}/m\mathbb{Z})$-module the generalized Steinberg module $\widetilde{St}$ can be identified with the cohomology $H^1(T/\Gamma(m),\mathbb{C})$ in degree 1 of the Tits building associated to $SL_3(\mathbb{Q})$ modulo the action of $\Gamma(m)$. Then a careful analysis of the induced action of σ on $[H^1(T/\Gamma(m),\mathbb{C})]^{PO(Q)}$ yields a contribution by trace $(\sigma,[\widetilde{St}]^{PO(Q)})=1-\left(\frac{m+1}{2}\right)$ (cf. 3.5), so we obtain finally for the Lefschetznumber of σ and $\Gamma_Q(m)$ at infinity

$$L_{\inf}(\sigma,X/\Gamma_Q(m))=\frac{m+1}{2}. \tag{2}$$

0.4. Using the description $\Gamma_Q(m)=\Gamma_D(m)\rtimes\mathcal{O}$ of $\Gamma_Q(m)$ as a semidirect product (cf. 0.2 (3)), the quotient space $X/\Gamma_Q(m)$ can be thought of as the quotient of $X/\Gamma_D(m)$ under the natural action of $\mathcal{O}$. Since $\Gamma_D(m)$ is torsionfree this allows to overcome the difficulty that $\Gamma_Q(m)$ has torsion elements. Let $F^{\sigma\gamma}$ denote the fixpoint set of the operation $x\mapsto x^\sigma\cdot\gamma$, $x\in X/\Gamma_D(m)$, $\gamma\in\mathcal{O}$. Then we have as a consequence of the Lefschetzfixpoint formula the following formula (cf. 4.2)

$$\begin{aligned}L(\sigma,\Gamma_Q(m))&=\frac{1}{|\mathcal{O}|}\sum_{\gamma\in\mathcal{O}}L(\sigma\gamma,\Gamma_D(m))\\&=\frac{1}{|\mathcal{O}|}\sum_{\gamma\in\mathcal{O}}\chi(F^{\sigma\gamma}),\end{aligned}$$

where $\chi(F^{\sigma\gamma})$, $\gamma\in\mathcal{O}$, denotes the Euler characteristic of $F^{\sigma\gamma}$.

0.5. The rest of the paper is then devoted to study the eight fixpointsets $F^{\sigma\gamma}$, $\gamma\in\mathcal{O}$, in $X/\Gamma_D(m)$ and to compute their Euler characteristics. Here we follow an approach of Rohlfs [19]. Studying the Lefschetznumber of the full congruence subgroup $\Delta(m)$ of $SL_n(\mathbb{Z})$ and the Cartan involution τ of $SL_n(\mathbb{R})$ he computed the Euler characteristic of the fixpoint set F of τ acting on the associated

locally symmetric space $Y/\Delta(m)$. This work is based on the parametrization of the connected components of F by the non-abelian cohomology set $H^1(\mathfrak{g}, \Delta(m))$, $\mathfrak{g}=\{1, \tau\}$, which can be interpreted as a set of certain equivalence classes of unimodular bilinear forms. The final formula for $L(\tau, \Delta(m))$ is then a matter of determining $H^1(\mathfrak{g}, \Delta(m))$, a certain local-global principle, computation of local volumes and the Minkowski-Siegel formula.

However, this method cannot be applied directly to our situation, because the operations $\sigma\gamma$ we are dealing with are different from the Cartan involution and $\Gamma_D(m)$ is not a full congruence subgroup. Moreover, apart from involutions there are two operations $\sigma\gamma$ of order four to consider. In § 5 we explain how we have to modify Rohlfs approach. As an example, we carry out in detail the case F^σ which contains all the essential features in the computations of the Euler characteristics. By the same method the Euler characteristic of the fixpoint set of the remaining five other involutions is computed in § 6 indicating briefly any necessary modifications. The study of the two operations of order four is done in § 8–10. We list the results (for notation we refer to the description of $\mathcal{O}$ in 0.2).

$$\chi(F^\sigma)=\chi(F^{\sigma s^2}) =\frac{(m^2-1)}{12}, \tag{1}$$

$$\chi(F^{\sigma t})=\chi(F^{\sigma t s^2})= -\frac{(m^2-1)}{12}, \tag{2}$$

$$\chi(F^{\sigma t s})=\chi(F^{\sigma t s^3})= -\frac{(m+1)}{6}, \tag{3}$$

$$\chi(F^{\sigma s})=\chi(F^{\sigma s^3}) = -\frac{m(m+1)}{6}. \tag{4}$$

Using (1)–(4) and 0.4 (1) we obtain for the total Lefschetz number of σ and $\Gamma_Q(m)$ the formula

$$L(\sigma, \Gamma_Q(m))= -\frac{(m+1)^2}{24}. \tag{5}$$

Comparing (5) with the Lefschetz number at infinity we get

$$L_{\text{cusp}}(\sigma, \Gamma_Q(m))= -\frac{(m+1)^2}{24} -\frac{(m+1)}{2}. \tag{6}$$

This proves the non-vanishing of $H^*_{\text{cusp}}(\Gamma_Q(m), \mathbb{C})$ as desired.

Both authors are very grateful to J. Rohlfs for some helpful comments on a first draft of this paper. We thank also G. Harder for stimulating discussions and remarks. The first (resp. second) named author would like to thank the Max Planck Institut für Mathematik, Bonn (resp. Dept. of Mathematics, Yale University) for its hospitality and support during certain periods of this work.

0.6. Notation. Beside the usual conventions we fix the following ones:

(1) If M is a set and f: $M \to M$ a map, we denote the fixpointset of f by $M^f=\{m\in M \mid f(m)=m\}$.

(2) By $M_n(R)$ we denote the R-algebra of $n \times n$-matrices with entries in the given ring R with 1. The transpose of an element A in $M_n(R)$ is written as A^t or ${}^tA$. The group of elements in $M_n(R)$ with in R invertible determinant (resp. determinant 1) id denoted by $GL_n(R)$ (resp. $SL_n(R)$). For a given element b in $GL_n(R)$ with $b = b^t$ we define the group of isometries with determinant 1 of the bilinear form b as $SO(b)(R) = \{A \in SL_n(R) \mid A^t b^{-1} A = b^{-1}\}$. If we consider the identity $b = E$ we use $SO_n(R) = SO(E)(R)$.

(3) Let G be a finite group and $H \subset G$ a subgroup of G. If (ϕ, V) is a finite dimensional complex representation of H, we denote by $\mathrm{Ind}_H^G[\phi] := V \otimes_{\mathbb{C}[H]} \mathbb{C}[G]$ the representation of G induced by σ. The restriction of a representation (ψ, W) of G on the subgroup H is denoted by $\mathrm{Res}_H \psi$.

(4) Let $\mathfrak{g}$ be a finite group acting from the left by group automorphisms on a given group Γ. For a given element σ in $\mathfrak{g}$ we denote the operation by $g \mapsto \sigma(g)$ or $g \mapsto {}^{\sigma}g$, $g \in \Gamma$. In this situation the first non-abelian cohomology set $H^1(\mathfrak{g}, \Gamma)$ is defined (cf. [24], I). If $\gamma = (\gamma_\sigma)$, $\sigma \in \mathfrak{g}$, is a cocycle for $H^1(\mathfrak{g}, \Gamma)$ then there is a new operation of $\mathfrak{g}$ on Γ twisted by γ; it is given by $g \mapsto \gamma_\sigma \sigma(g) \gamma_\sigma^{-1}$, $g \in \Gamma$. We denote Γ with this twisted operation by ${}^{\gamma}\Gamma$.

(5) Sometimes we denote $\mathbb{Z}/m\mathbb{Z}$ by $\mathbb{Z}/m$.

§1. Elementary Properties of $\Gamma_Q(m)$

1.1. We begin by considering the full congruence subgroup $\Gamma(m) = \{A \in SL_3(\mathbb{Z}) \mid A \equiv E \bmod m\}$ in $SL_3(\mathbb{Z})$ of level m, $m \in N$ as in our previous paper [16]. Throughout this paper, we need by technical reasons which will become clear lateron the following assumptions on m:

(1) m is a prime, (2) $m \equiv 3 \bmod 8$, (3) $m \equiv -1 \bmod 3$.

Condition (2) is not substantially used until it becomes crucial in 9.4. The weaker version $m \equiv 3 \bmod 4$ would be sufficient for most considerations.

Because of (1), the quotient $\mathbb{Z}/m\mathbb{Z}$ forms a field, and the finite group ${}_fSL := SL_3(\mathbb{Z})/\Gamma(m)$ is the special linear group of degree 3 over this field. Let Q denote the symmetric matrix

$$Q = \begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & 0 \end{pmatrix}$$

over $\mathbb{Z}/m\mathbb{Z}$, and let $PO(Q)$ denote the projective orthogonal group in ${}_fSL$ which preserves Q up to a scalar, i.e.

$$PO(Q) := \{A \in SL_3(\mathbb{Z}/m\mathbb{Z}) \mid A Q {}^tA = \alpha Q, \alpha \in \mathbb{Z}/m\mathbb{Z}, \alpha \neq 0\}.$$

Define $\Gamma_Q(m)$ to be the congruence subgroup in $SL_3(\mathbb{Z})$ consisting of matrices A such that A modulo m belongs to the subgroup $PO(Q)$ of ${}_fSL$. This

group $\Gamma_Q(m)$ contains elements of finite order. If s and t denote the matrices

$$s=\begin{pmatrix} 0 & 1 & \\ -1 & 0 & \\ & & 1 \end{pmatrix}, \quad t=\begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix},$$

then s and t generate a dihedral group $\mathcal{O}$ of order 8 in $\Gamma_Q(m)$; the relations are given by $s^4=1$, $tst=s^{-1}$, $t^2=1$. From the condition (2), it follows that (-1) is not a square modulo m, and so the group of units $(\mathbb{Z}/m\mathbb{Z})^*$ decomposes into a direct product

$$(\mathbb{Z}/m\mathbb{Z})^* \simeq \mathbb{Z}_{\frac{m-1}{2}} \times \mathbb{Z}_2$$

of two cyclic groups. There is a projection

$$\varepsilon\colon (\mathbb{Z}/m\mathbb{Z})^* \to \mathbb{Z}/2\mathbb{Z}$$

of the group of units $(\mathbb{Z}/m\mathbb{Z})^*$ onto its 2-primary component. We have in the projective orthogonal group $PO(Q)$ the subgroup

$$D=\left\{ \begin{pmatrix} a_{11} & & a_{13} \\ & a_{22} & a_{23} \\ & & a_{33} \end{pmatrix} \in PO(Q) \,\middle|\, \begin{matrix} a_{12}=a_{21}=a_{31}=a_{32}=0 \\ \varepsilon(a_{11})=\varepsilon(a_{22})=\varepsilon(a_{33})=1 \end{matrix} \right\}.$$

By means of D, we define $\Gamma_D(m)$ to be the subgroup in $\Gamma_Q(m)$ consisting of matrices A such that A modulo m is in D.

1.2. Proposition. *The congruence subgroup $\Gamma_Q(m)$ is a semi-direct product of $\Gamma_D(m)$ and $\mathcal{O}$, i.e.*

$$\Gamma_Q(m)=\Gamma_D(m)\rtimes\mathcal{O}.$$

The subgroup $\Gamma_D(m)$ is torsion free. If $\sigma\colon SL_3(\mathbb{Q})\to SL_3(\mathbb{Q})$ is the involution given by the formula:

$$\sigma(A)=\begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix} {}^tA^{-1} \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix}^{-1},$$

then both $\Gamma_Q(m)$ and $\Gamma_D(m)$ are σ-stable.

For an invertible matrix

$$A=\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in {}_fSL$$

to lie in $PO(Q)$, it must preserve the null space $\langle Q\rangle^\perp=\{(0,0,z)\,|\,z\in\mathbb{Z}/m\mathbb{Z}\}$ of Q. This is the same as the requirement $a_{31}=a_{32}=0$. As for the upper $2-$ by -2

block $\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$, it must satisfy the equation

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix} = \alpha \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

A direct computation shows that either one of the following two conditions must be fulfilled

$$a_{12} = a_{21} = 0, \quad \text{or} \quad a_{11} = a_{22} = 0.$$

Hence the projective orthogonal group $PO(Q)$ consists of the following two types of matrices:

$$\begin{pmatrix} a_{11} & & a_{13} \\ & a_{22} & a_{23} \\ & & \dfrac{1}{a_{11}a_{22}} \end{pmatrix}, \quad \begin{pmatrix} & 1 & \\ 1 & & \\ & & -1 \end{pmatrix} \cdot \begin{pmatrix} a_{11} & & a_{13} \\ & a_{22} & a_{23} \\ & & \dfrac{1}{a_{11}a_{22}} \end{pmatrix}.$$

From this it is not difficult to show that $PO(Q)$ is a semi-direct product of D and $\mathcal{O}$, $PO(Q) = D \rtimes \mathcal{O}$, and from the definition of $\Gamma_Q(m)$ there is a similar semi-direct product decomposition, $\Gamma_Q(m) = \Gamma_D(m) \rtimes \mathcal{O}$. This proves the first statement in 1.2.

As is well known ([18, 27, 28]), the group $SL_3(\mathbb{Z})$ has only 2-torsion and 3-torsion, and the full congruence subgroup $\Gamma(m)$ is torsion free for $m \geqq 3$. Since we have a group extension

$$1 \to \Gamma(m) \to \Gamma_D(m) \to D \to 1,$$

a torsion element in $\Gamma_D(m)$ must project onto a non-trivial element in D. On the other hand, our assumptions (2), (3) on m imply that the finite group D does not have $2-$ or $3-$ torsion. Hence $\Gamma_D(m)$ must be torsion free.

Let A be an element in $\Gamma_D(m)$. Its inverse $A^{-1} = (b_{ij})$, $i, j = 1, 2, 3$, must be an element of the same type i.e. $b_{12} \equiv b_{21} \equiv b_{31} \equiv b_{32} \equiv 0 \bmod m$ resp. $\varepsilon(b_{11} \bmod m) = \varepsilon(b_{22} \bmod m) = \varepsilon(b_{33} \bmod m) = 1$. Under the action of σ, it is sent to the matrix

$$\sigma(A) = \begin{pmatrix} b_{11} & b_{12} & b_{31}/m \\ b_{21} & b_{22} & b_{32}/m \\ b_{13}m & b_{23}m & b_{33} \end{pmatrix}. \tag{1}$$

Clearly, this is an element in $\Gamma_D(m)$. This proves that $\Gamma_D(m)$ is σ-stable. Note that s and t are invariant under the action of σ and so the dihedral group $\mathcal{O}$ is fixed by σ. Since $\Gamma_Q(m)$ is a semi-direct product of $\Gamma_D(m)$ and $\mathcal{O}$, it also must be stable under the action of σ.

§ 2. Cohomology of $\Gamma_Q(m)$ at Infinity

We recall briefly some of the results in our previous paper [16] and adopt for the most part the notation there.

2.1. Let Γ be a torsion free arithmetic subgroup of $SL_3(\mathbb{Q})$, and denote by X the symmetric space $X = SO(3)\backslash SL_3(\mathbb{R})$. Then there is a short exact sequence (cf. 1.8 [16])

$$0 \to H^q_{\text{cusp}}(\Gamma, X; \mathbb{C}) \to H^q(\Gamma, X; \mathbb{C}) \to H^q_{\text{inf}}(\bar{X}/\Gamma; \mathbb{C}) \to 0, \quad q > 0 \tag{1}$$

where $H^q_{\text{cusp}}(\Gamma, X; \mathbb{C})$ stands for the cusp cohomology of Γ identified with the space of harmonic cusp forms with respect to Γ on X and where the cohomology at infinity is, by definition, the image of the natural restriction map of the cohomology $H^*(\bar{X}/\Gamma; \mathbb{C}) = H^*(\Gamma, \bar{X}; \mathbb{C})$ of the Borel-Serre compactification $\bar{X}/\Gamma$ (cf. [2, 4]) to the cohomology $H^*(\partial(\bar{X}/\Gamma), \mathbb{C})$ of its boundary $\partial(\bar{X}/\Gamma)$. If Γ' is now an arithmetic subgroup of $SL_3(\mathbb{Q})$ (possibly with torsion) and contains Γ as a normal subgroup, then Γ'/Γ acts on the Eilenberg-MacLane cohomology spaces $H^*(\Gamma; \mathbb{C}) = H^*(\Gamma, X; \mathbb{C})$ and we have (cf. [5], VIII, 2) $H^*(\Gamma'; \mathbb{C}) = H^*(\Gamma; \mathbb{C})^{\Gamma'/\Gamma}$. Since taking invariants under a finite group is an exact functor in characteristic zero, there is also an exact short sequence for the group Γ'

$$0 \to H^q_{\text{cusp}}(\Gamma', X; \mathbb{C}) \to H^q(\Gamma', X; \mathbb{C}) \to H^q_{\text{inf}}(\bar{X}/\Gamma'; \mathbb{C}) \to 0, \quad q > 0 \tag{2}$$

where $H^q_{\text{inf}}(\bar{X}/\Gamma'; \mathbb{C}) := H^q_{\text{inf}}(\bar{X}/\Gamma; \mathbb{C})^{\Gamma'/\Gamma}$. This applies in particular to the situation we are considering now where we assume $\Gamma = \Gamma(m)$ and $\Gamma' = \Gamma_Q(m)$ as in § 1.

In Theorem 2.4 in [16], we gave a complete description of the cohomology $H^*(\partial(\bar{X}/\Gamma(m)); \mathbb{C})$ of the boundary $\partial(\bar{X}/\Gamma(m))$ as a module over ${}_fSL := SL_3(\mathbb{Z})/\Gamma(m) = SL_3(\mathbb{Z}/m\mathbb{Z})$. In order to recall the result we have to fix some notation. Let P_i, $i = 1, 2$, resp. P_0 be the standard maximal (resp. minimal) parabolic $\mathbb{Q}$-subgroups of $SL_3(\mathbb{R})$ with its usual Langlands decomposition $P_i = A_i \cdot {}^0M_i \cdot N_i$, $i = 0, 1, 2$, as in 2.1 [16], and denote the quotients by

$${}_fP_i = (P_i \cap SL_3(\mathbb{Z}))/P_i \cap \Gamma(m), \tag{3}$$

resp.

$${}_fM_i = ({}^0M_i \cap SL_3(\mathbb{Z}))/{}^0M_i \cap \Gamma(m). \tag{4}$$

Then we proved that the cohomology $H^*_{\text{inf}}(\bar{X}/\Gamma(m); \mathbb{C})$ as a representation space of ${}_fSL$ can be described by the following formula:

$$[H^q_{\text{inf}}(\bar{X}/\Gamma(m); \mathbb{C})] = \begin{cases} \mathbb{C}, & q = 0 \\ 0, & q = 1,\ q \geqq 4 \\ 2[\text{Ind}_{{}_fP_1}^{{}_fSL}[\varepsilon_1]] \oplus [\text{Ind}_{{}_fP_1}^{{}_fSL}[H^1_{\text{cusp}}(\Gamma(2, m); E)]], & q = 2 \\ \bigoplus_{i=1}^{2} [\text{Ind}_{{}_fP_i}^{{}_fSL}[H^1_{\text{cusp}}(\Gamma(2, m); \mathbb{C})]] \oplus [\widetilde{St}], & q = 3 \end{cases} \tag{5}$$

where the action of ${}_fP_i$ on $H^1_{\text{cusp}}(\Gamma(2, m); _)$ is the pullback of the natural action of ${}_fM_i \cong SL_2^{\pm}(\mathbb{Z}/m\mathbb{Z})$ on the cusp cohomology $H^1_{\text{cusp}}(\Gamma(2, m); _)$ of the full congruence subgroup of level m of $SL_2(\mathbb{Z})$ with coefficients in $\mathbb{C}$ resp. the 2-dimensional representation E of $SL_2(\mathbb{R})$; where $\widetilde{St}$ denotes the generalized Steinberg-module of ${}_fSL$ defined by the short exact sequence

$$0 \to \mathbb{C} \to \bigoplus_{i=1}^{2} \text{Ind}_{{}_fP_i}^{{}_fSL}[\mathbb{C}] \to \text{Ind}_{{}_fP_0}^{{}_fSL}[\mathbb{C}] \to \widetilde{St} \to 0 \tag{6}$$

and ε_i denotes the pullback to ${}_fP_i$ of the one-dimensional signum-representation of ${}_fM_i \cong SL_2^{\pm}(\mathbb{Z}/m\mathbb{Z})$. We remark here that the term $H^1_{\text{cusp}}(\Gamma(2,m))$ comes from the cohomology of the base space $SO(3)\cap {}^0M_i \backslash {}^0M_i/{}^0M_i \cap \Gamma(m) \cong H/\Gamma(2,m)$ (H = upper half plane) in the natural fibration of the face $e'(P_i)$ in $\partial(\bar{X}/\Gamma(m))$ associated to P_i (cf. 3.1 [16]). To simplify our notations we use the symbols ψ_i, ϕ_i, resp. Ψ_i, Φ_i, $i=1,2$, to denote the representations (resp. representation spaces) of ${}_fP_i$ on the cusp cohomology:

$$\Psi_i := H^1_{\text{cusp}}(\Gamma(2,m); E), \qquad \Phi_i := H^1_{\text{cusp}}(\Gamma(2,m); \mathbb{C}), \qquad i=1,2. \tag{7}$$

The object of this section is to establish the following

2.2. Proposition. *The cohomology of $\Gamma_Q(m)$ at infinity is given by*

$$H^q_{\text{inf}}(\bar{X}/\Gamma_Q(m); \mathbb{C}) = \begin{cases} \mathbb{C}, & q=0 \\ 0, & q\neq 0,3 \\ \bigoplus_{i=1}^{2} [\text{Ind}_{{}_fP_i}^{{}_fSL}[\phi_i]]^{PO(Q)} \oplus [\widetilde{St}]^{PO(Q)}, & q=3. \end{cases}$$

Comparing the formulas for the cohomology $H^q_{\text{inf}}(\bar{X}/\Gamma(m); \mathbb{C})$ with the above it is clear that we only have to show that the invariant subspaces

$$H^0(PO(Q), \text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i]) = [\text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i]]^{PO(Q)}$$
$$H^0(PO(Q), \text{Ind}_{{}_fP_i}^{{}_fSL}[\varepsilon_i]) = [\text{Ind}_{{}_fP_i}^{{}_fSL}[\varepsilon_i]]^{PO(Q)}$$

are trivial.

We begin with the first case. Choose a set T_i of representatives for the double cosets of ${}_fSL \bmod (PO(Q), {}_fP_i)$. For $t\in T_i$ we denote by I_t the subgroup $PO(Q)\cap t^{-1}({}_fP_i)\,t$ of $PO(Q)$. If we put

$$\psi_{i,t}(p) := \psi_i(tpt^{-1}), \qquad p\in I_t$$

we get a homomorphism $\psi_{i,t}\colon I_t \to GL(\Psi_i)$ and therefore a representation $(\psi_{i,t}, \Psi_{i,t})$ of I_t. Up to an isomorphism this representation does not depend on the choice of t. The restriction of the representation $\text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i]$ on the subgroup $PO(Q)$ is then given by (cf. [22], 7.3):

$$\text{Res}_{PO(Q)} \text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i] = \bigoplus_{t\in T_i} \text{Ind}_{I_t}^{PO(Q)}[\psi_{i,t}].$$

This implies

$$\begin{aligned} H^0(PO(Q), \text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i]) &\cong \text{Hom}_{PO(Q)}(\mathbb{C}, \text{Ind}_{{}_fP_i}^{{}_fSL}[\psi_i]) \\ &\cong \bigoplus_{t\in T_i} \text{Hom}_{PO(Q)}(\mathbb{C}, \text{Ind}_{I_t}^{PO(Q)}[\psi_{i,t}]) \\ &\cong \bigoplus_{t\in T_i} \text{Hom}_{I_t}(\mathbb{C}, \Psi_{i,t}) \end{aligned} \tag{1}$$

where the last isomorphism is due to Frobenius reciprocity.

Since the representations $\mathrm{Ind}_{{}_fP_1}^{{}_fSL}[\psi_1]$, $\mathrm{Ind}_{{}_fP_2}^{{}_fSL}[\psi_2]$ are isomorphic representations of ${}_fSL$, it is enough to consider only one of these two. For definiteness, we consider $i=2$, and choose

$$ {}_fP_2 = \left\{ \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & a_{32} & a_{33} \end{pmatrix} \in {}_fSL \,\middle|\, a_{33} = \pm 1 \right\}. $$

Note that this is a different choice of representative for the parabolic subgroup of type (P_2) from the choice in [16], but this causes no difficulties. The advantage is that the left cosets ${}_fSL/{}_fP_2$ can be identified with column vectors

$$ \xi = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \qquad x, y, z \in \mathbb{Z}/m\mathbb{Z}. $$

Two such columns $\xi = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$, $\xi' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix}$, give the same coset if and only if they differ by a multiple of ± 1, i.e.

$$ x = \pm x', \qquad y = \pm y', \qquad z = \pm z'. $$

To describe the double coset space $PO(Q)\backslash {}_fSL/{}_fP_2$, we consider the action of the projective orthogonal group $PO(Q)$ on this space of column vectors. It is elementary to verify that there are three orbits for this action given by

$$ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \ (2), \qquad \begin{pmatrix} -1 \\ 0 \\ 0 \end{pmatrix} \ (3), \qquad \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} \ (4). $$

As a consequence there are only three distinct elements in the double coset space $PO(Q)\backslash {}_fSL/{}_fP_2$ represented by

$$ t_1 = \begin{pmatrix} 1 & & \\ & 1 & \\ & & 1 \end{pmatrix} \ (5), \qquad t_2 = \begin{pmatrix} & & -1 \\ & 1 & \\ 1 & & \end{pmatrix} \ (6), \qquad t_3 = \begin{pmatrix} -1 & 1 & -1 \\ & 0 & -1 \\ & 1 & 0 \end{pmatrix} \ (7). $$

In the case of (5), (6), the diagonal matrices

$$ d_1 = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 \end{pmatrix} \ (8), \qquad d_2 = \begin{pmatrix} 1 & & \\ & -1 & \\ & & -1 \end{pmatrix} \ (9), $$

lie respectively in the subgroups I_{t_1} and I_{t_2}. On one hand they operate trivially on the one dimensional representation $\mathbb{C}$ of I_{t_j}, $j=1,2$, and on the other hand

their conjugate

$$t_1 d_1 t_1^{-1} = t_2 d_2 t_2^{-1} = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 \end{pmatrix}$$

operates as -1 on the space Ψ_{2,t_j}. Hence, in formula (1), there is no intertwining operator from the first two orbits. As for the orbit (7), the element

$$\begin{pmatrix} & 1 & \\ 1 & & \\ & & -1 \end{pmatrix}$$

lies in the subgroup I_{t_3} whereas its conjugation by t_3 is of the form

$$\begin{pmatrix} -1 & & \\ & -1 & \\ 1 & -1 & 1 \end{pmatrix}.$$

The first element operates as $+1$ on the trivial representation and the second element operates as -1 on the space Ψ_{2,t_3}. Thus there is no intertwining between the representations $\mathrm{Ind}_{PO(Q)}^{fSL}[\mathbb{C}]$ and $\mathrm{Ind}_{fP_2}^{fSL}[\Psi_2]$ i.e. $H^0(PO(Q), \mathrm{Ind}_{fP_2}^{fSL}[\Psi_2]) = 0$.

It remains to prove that the cohomology $H^0(PO(Q), \mathrm{Ind}_{fP_i}^{fSL}[\varepsilon_i])$ is also trivial. We repeat a similar argument as before, using the fact that the element d_j, $j \neq i$, $j = 1, 2$ operates as -1 on the space ε_i (see [16], 2.2.9). The detailed argument is left to the reader.

§ 3. The Lefschetz Number at Infinity

3.1. The object of this section is to evaluate the Lefschetz number $L_{\mathrm{inf}}(\sigma, X/\Gamma_Q(m))$. By 2.2, $H^3_{\mathrm{inf}}(X/\Gamma_Q(m))$ consists of the two terms

$$H^0(PO(Q); \widetilde{St}) \quad (1), \qquad \text{and} \qquad \bigoplus_{i=1}^{2} H^0(PO(Q); \mathrm{Ind}_{fP_i}^{fSL}[\phi_i]) \quad (2),$$

where $\widetilde{St}$ resp. ϕ_i as defined in 2.1 (6) resp. (7). However, for the generalized Steinberg module $\widetilde{St}$ there is a geometric definition in terms of the Tits building of $SL_3(\mathbb{Q})$. We consider the set of all proper, parabolic $\mathbb{Q}$-subgroups of $SL_3(\mathbb{Q})$ partially ordered by inclusion. The geometric realization $T = T(3)$ of the partially ordered set is a simplicial complex known as the Tits building of $SL_3(\mathbb{Q})$ (see [4], § 8.4). The group $SL_3(\mathbb{Q})$ operates simplicially on this space. It follows that the quotient space $T(3)/\Gamma(m)$ of the building $T(3)$ modulo the action of the principal congruence subgroup $\Gamma(m)$ is a CW-complex, and the finite group ${}_fSL \cong SL_3(\mathbb{Z})/\Gamma(m)$ operates on this complex.

3.2. Proposition. *As ${}_fSL$-modules, the cohomology spaces of $T(3)/\Gamma(m)$ are given by*

$$H^q(T(3)/\Gamma(m), \mathbb{C}) = \begin{cases} \mathbb{C} & \text{if } q=0 \\ \widetilde{St} & \text{if } q=1. \\ 0 & \text{otherwise} \end{cases}$$

The proof is similar to the proof of Proposition 3.7 in [16]. There is a one-to-one correspondence between the set of vertices in $T/\Gamma(m)$ and the set of $\Gamma(m)$-conjugacy classes of minimal parabolic $\mathbb{Q}$-subgroups. Since the last space can be identified with the coset space ${}_fSL/{}_fP_0$ (cf. [16], § 3), the 0-cochains form a ${}_fSL$-representation space isomorphic to the induced representation $\mathrm{Ind}_{{}_fP_0}^{{}_fSL}[\mathbb{C}]$. Similarly the 1-cochains are in one-to-one correspondence with the $\Gamma(m)$-conjugacy classes of maximal $\mathbb{Q}$-parabolic subgroups in $SL_3(\mathbb{Q})$. This gives a formula for the cochain complexes of $T/\Gamma(m)$:

$$C^1(T/\Gamma(m); \mathbb{C}) \cong \bigoplus_{i=1}^{2} \mathrm{Ind}_{{}_fP_i}^{{}_fSL}[\mathbb{C}], \quad \text{resp. } C^0(T/\Gamma(m); \mathbb{C}) \cong \mathrm{Ind}_{{}_fP_0}^{{}_fSL}[\mathbb{C}].$$

As all the cells in $T(3)/\Gamma(m)$ are of dimension less than one, we have

$$H^q(T/\Gamma(m); \mathbb{C}) = 0, \qquad q>1,$$

and for the first two cohomology groups there is an exact sequence:

$$0 \to H^0(T/\Gamma(m); \mathbb{C}) \to \mathrm{Ind}_{{}_fP_0}^{{}_fSL}[\mathbb{C}] \to \bigoplus_{i=1}^{2} \mathrm{Ind}_{{}_fP_i}^{{}_fSL}[\mathbb{C}] \to H^1(T/\Gamma(m); \mathbb{C}) \to 0.$$

It is easy to verify that $T(3)$ is connected, and so the first term in the above exact sequence is the trivial one-dimensional representation of ${}_fSL$, i.e. $H^0(T/\Gamma(m), \mathbb{C}) \cong \mathbb{C}$. Comparing this with the previous definition of the generalized Steinberg representation, it follows that the last term $H^1(T/\Gamma(m); \mathbb{C})$ coincides with $\widetilde{St}$.

3.3. There is another definition of the Tits building $T(3)$ in terms of rational flags. As in the above paragraph, there are two $SL_3(\mathbb{Q})$-conjugacy classes of maximal parabolic groups represented by the standard maximal parabolic $\mathbb{Q}$-subgroups P_i, $i=1,2$ of $SL_3(\mathbb{Q})$. Those in the first conjugacy class (P_1) are in one-to-one correspondence with rational lines in $\mathbb{Q}^3$. For given any line in $\mathbb{Q}^3$, the isotropy subgroup in $SL_3(\mathbb{Q})$ which keeps this line fixed is such a maximal parabolic group of type (P_1). On the other hand, any line determines up to sign ± 1 an integral unimodular row vector in $\mathbb{Z}^3$,

$$\pm L = \pm(a_1, a_2, a_3), \qquad a_i \in \mathbb{Z}, \qquad \text{g.c.d. } (a_1, a_2, a_3) = 1. \tag{1}$$

In a similar fashion, those in the second conjugacy class (P_2) are in one-to-one correspondence with rational planes in $\mathbb{Q}^3$ or equivalently with an integral unimodular column vectors

$$\pm H = \pm \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

As for 1-simplices, they can be represented by flags (L, H) consisting of a line L, a plane H, and L lying in H. In the language of row and column vectors, the last condition means that the dot product $L \cdot H$ is zero, i.e.

$$L \cdot H = a_1 x_1 + a_2 x_2 + a_3 x_3 = 0 \quad \text{for } L = \pm(a_1 a_2 a_3), \quad H = \pm \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}.$$

Note that we identify two pairs (L, H), (L', H') if their components differ only by a sign ± 1,

$$(L, H) = (\pm L, H) = (L, \pm H) = (\pm L, \pm H).$$

Let A be an element in the principal congruence subgroup $\Gamma(m)$, i.e. $A \equiv I \bmod m$. Then the action of A on the simplices of T is given by the formulas:

$$\{\pm L\} \quad \mapsto \{\pm L \cdot A\} \tag{2}$$

$$\{\pm H\} \quad \mapsto \{\pm A^{-1} \cdot H\} \tag{3}$$

$$\{\pm L, \pm H\} \mapsto \{\pm L \cdot A, \pm A^{-1} \cdot H\}. \tag{4}$$

Here the symbols $L \cdot A$, $A^{-1} \cdot H$ denote the matrix multiplication of A (resp. A^{-1}) on the row vectors L (resp. on the column vectors H).

The quotient space $T/\Gamma(m)$ admits a similar interpretation. Its vertices can be represented by row or column vectors in $(\mathbb{Z}/m\mathbb{Z})^3$

$$\pm \bar{L} = \pm(\bar{a}_1, \bar{a}_2, \bar{a}_3), \quad \pm \bar{H} = \pm \begin{pmatrix} \bar{x}_1 \\ \bar{x}_2 \\ \bar{x}_3 \end{pmatrix}, \quad \begin{matrix} \bar{a}_i \in \mathbb{Z}/m\mathbb{Z} \\ \bar{x}_i \in \mathbb{Z}/m\mathbb{Z} \end{matrix}.$$

Its 1-simplices are pairs of column and row vectors $\{\pm \bar{L}, \pm \bar{H}\}$ such that

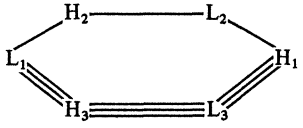
$$\bar{L} \cdot \bar{H} = \bar{a}_1 \bar{x}_1 + \bar{a}_2 \bar{x}_2 + \bar{a}_3 \bar{x}_3 = 0.$$

The action of the finite group $PO(Q)$ on this complex is defined by the same formulas as in (2), (3) and (4). Since the orbit of this action is canonically isomorphic to $T/\Gamma_Q(m)$, we have

$$H^1(T/\Gamma_Q(m)) \cong H^0(PO(Q), H^1(T/\Gamma(m))) \cong H^0(PO(Q), St).$$

3.4. Proposition. *The $PO(Q)$-invariant elements in the generalized Steinberg representation $\widetilde{St}$ form a vector space $H^0(PO(Q), \widetilde{St})$ of dimension $\frac{3}{2}(m-3)+1$.*

We will show that the quotient space $T/\Gamma_Q(m)$ is a 1-dimensional simplicial complex consisting of 6 vertices and $\frac{3}{2}(m-1)+3$ edges as indicated in the following figure:



(1)

Here the 6 vertices can be represented by the orbits of 3 row vectors and 3 column vectors:

$$L_1 = \pm(0,0,1), \quad L_2 = \pm(-1,0,0), \quad L_3 = \pm(1,-1,0),$$

$$H_1 = \pm\begin{pmatrix}0\\0\\1\end{pmatrix}, \quad H_2 = \pm\begin{pmatrix}-1\\0\\0\end{pmatrix}, \quad H_3 = \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

The symbol $\cdot\!-\!\cdot$ denotes an edge, and the symbol $\cdot\equiv\cdot$ denotes $\dfrac{m-1}{2}$ edges connecting two vertices at the end. A direct computation of the Euler characteristic of the building $T/\Gamma_Q(m)$ shows that the cohomology $H^1(T/\Gamma_Q(m))$ has dimension $\frac{3}{2}(m-3)+1$ as claimed.

To prove (1), we begin by considering those vertices of $T/\Gamma(m)$ which correspond to non-zero column vectors

$$H = \pm\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}, \quad x_i \in \mathbb{Z}/m\mathbb{Z}.$$

As in the proof of Proposition 2.2, there are three different orbits of these vertices under the action of $PO(Q)$:

$$H_1 = \pm\begin{pmatrix}0\\0\\1\end{pmatrix}, \quad H_2 = \pm\begin{pmatrix}-1\\0\\0\end{pmatrix}, \quad H_3 = \pm\begin{pmatrix}1\\1\\0\end{pmatrix},$$

each of which represents a unique vertex in $T/\Gamma_Q(m)$.

For row vectors $L = \pm(a_1, a_2, a_3)$, the situation is similar. In fact, there is the involution

$$\sigma\colon T/\Gamma(m) \to T/\Gamma(m)$$

induced by the automorphism in 1.2, which interchanges the role of column and row vectors

$$\sigma(a_1,a_2,a_3) = \begin{pmatrix}a_1\\a_2\\a_3 m\end{pmatrix}, \quad \sigma\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix} = (x_1 m, x_2 m, x_3).$$

Hence we have three other vertices in $T/\Gamma(m)$ defined by

$$L_1 = \sigma(H_1) = \pm(0,0,1), \quad L_2 = \sigma(H_2) = \pm(-1,0,0), \quad L_3 = \sigma(H_3) = \pm(1,1,0).$$

Note that the element $s^3 = \begin{pmatrix}0 & -1 & \\ 1 & 0 & \\ & & 1\end{pmatrix}$ sends the vector $(1,1,0)$ to $(1,-1,0)$, and so the two vectors L_3, L_3 represent the same point in $T/\Gamma_Q(m)$.

It remains to describe the 1-simplices in $T/\Gamma_Q(m)$. For this we consider the set of all 1-simplices in $T/\Gamma_Q(m)$ with the vertex $\begin{pmatrix}0\\0\\1\end{pmatrix}$ as one of its endpoints. In the ramified covering space $T/\Gamma(m)$, this is covered by a corresponding set $\mathrm{Star}(H_1)$ which contains all 1-simplices in $T/\Gamma(m)$ with the vertex $\begin{pmatrix}0\\0\\1\end{pmatrix}$ as its endpoint. Let $I(H_1)$ denote the isotropy subgroup in $PO(Q)$ which keeps this vertex H_1 fixed. Then $I(H_1)$ operates on $\mathrm{Star}(H_1)$ and its orbit space is the same as the star of $\begin{pmatrix}0\\0\\1\end{pmatrix}$ in $T/\Gamma(m)$. A direct computation shows that $\mathrm{Star}(H_1)$ consists of the following 1-simplices:

$$\left\{(a_1, a_2, 0), \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}, \qquad a_i \in \mathbb{Z}/m\mathbb{Z},$$

and the isotropy subgroup $I(H_1)$ consists of the following matrices:

$$\begin{pmatrix}a_{11} & & \\ & a_{22} & \\ & & \pm 1\end{pmatrix}, \quad \begin{pmatrix}0 & 1 & \\ 1 & 0 & \\ & & -1\end{pmatrix} \quad \begin{pmatrix}a_{11} & & \\ & a_{22} & \\ & & \pm 1\end{pmatrix}.$$

Under the action of $I(H_1)$, two simplices

$$\left\{(a_1, a_2, 0), \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}, \qquad \left\{(a_1', a_2', 0), \begin{pmatrix}0\\0\\1\end{pmatrix}\right\}$$

lie in the same orbit if and only if $a_1 a_2 = \pm a_1' a_2'$. As a result, there are $\dfrac{m+1}{2}$ orbits denoted by

$$\left\{(-1, 0, 0), \begin{pmatrix}0\\0\\1\end{pmatrix}\right\} \quad (2), \qquad \left\{(1, \lambda, 0), \begin{pmatrix}0\\0\\1\end{pmatrix}\right\} \lambda \neq 0,\ \varepsilon(\lambda) = 1. \quad (3)$$

The first orbit (2) represents a single line connecting the vertices L_2 and H_1, and the second group of orbits in (3) represents $\dfrac{m-1}{2}$ different lines connecting the vertices L_3, H_1 in $T/\Gamma_Q(m)$.

The involution σ takes the above simplices to the simplex connecting L_1, H_2 and the $\dfrac{m-1}{2}$ simplices connecting L_1, H_3.

To complete the diagram (1), we have to examine 1-simplices connecting L_2, H_2 and L_3, H_3. We only discuss the second case L_3, H_3 and leave the first case to the reader. In the space $T/\Gamma(m)$ the star of the vertex $\begin{pmatrix}1\\1\\0\end{pmatrix}$ consists of the following 1-simplices

$$\left\{(a_1, -a_1, a_3), \begin{pmatrix}1\\1\\0\end{pmatrix}\right\}, \tag{4}$$

and as for the isotropy subgroup it consists of the following matrices

$$\begin{pmatrix} 0 & \pm 1 & a_{13} \\ \pm 1 & 0 & a_{23} \\ & & -1 \end{pmatrix}, \quad \begin{pmatrix} \pm 1 & & a_{13} \\ & \pm 1 & a_{23} \\ & & 1 \end{pmatrix}.$$

We can get rid of the entry a_3 in (4) by the action of the matrix

$$\begin{pmatrix} 1 & & -a_3/a_1 \\ & 1 & 0 \\ & & 1 \end{pmatrix}$$

and so it is enough to consider the 1-simplices $\left\{(a_1, -a_1, 0), \begin{pmatrix}1\\1\\0\end{pmatrix}\right\}$. Under the action of $\begin{pmatrix} \pm 1 & & \\ & \pm 1 & \\ & & 1 \end{pmatrix}, \begin{pmatrix} 0 & \pm 1 & \\ \pm 1 & 0 & \\ & & -1 \end{pmatrix}$, there are a total of $\frac{m-1}{2}$ orbits as in (1). This completes the proof.

3.5. Proposition. *Let $\tilde{St}_Q$ denote the vector space of $PO(Q)$-invariant elements in $\tilde{St}$, i.e. $\tilde{St}_Q \cong H^0(PO(Q), \tilde{St})$. Then there is an induced action by σ on this space $\tilde{St}_Q$ and its trace is given by the formula:*

$$\operatorname{trace}(\sigma, \tilde{St}_Q) = 1 - \left(\frac{m+1}{2}\right).$$

As in the previous proof, there is an induced action σ on the space $T/\Gamma_Q(m)$, $\sigma\colon T/\Gamma_Q(m) \to T/\Gamma_Q(m)$, and so an action on its cohomology $H^1(T/\Gamma_Q(m)) \cong H^0(PO(Q), \tilde{St}) \cong \tilde{St}_Q$. We claim that there are exactly $\frac{m+1}{2}$ fixpoints of σ on $T/\Gamma_Q(m)$. The above trace formula follows immediately from this claim by applying the Lefschetz fixpoint formula. In the proof of 3.4(1), we showed that σ interchanges the vertices L_1, L_2, L_3 with the corresponding vertices H_1, H_2, H_3. Thus a fixpoint can occur only at the midpoint of the lines connecting L_2, H_2 and L_3, H_3. On one hand, there must be a fixpoint at the middle of the line segment $\{L_2, H_2\}$ because σ interchanges its two endpoints.

On the other hand, the 1-simplex $l_a = \left\{(a, -a, 0); \begin{pmatrix}1\\1\\0\end{pmatrix}\right\}$ is mapped onto the simplex $\sigma(l_a) = \left\{(1,1,0); \begin{pmatrix}a\\-a\\0\end{pmatrix}\right\}$. This is $\Gamma_Q(m)$-equivalent to l_a because

$$(1,1,0)\begin{pmatrix}a & & \\ & -a & \\ & & -1\end{pmatrix} = (a,-a,0), \quad \text{resp.} \begin{pmatrix}\frac{1}{a} & & \\ & \frac{-1}{a} & \\ & & -1\end{pmatrix}\begin{pmatrix}a\\-a\\0\end{pmatrix} = \begin{pmatrix}1\\1\\0\end{pmatrix}.$$

Thus for every line connecting L_3, H_3 in $T/\Gamma_Q(m)$, there is a fixpoint at the middle.

3.6. Proposition. *The Lefschetz number* $L_{\inf}(\sigma, X/\Gamma_Q(m))$ *of* $X/\Gamma_Q(m)$ *at infinity is* $\frac{m+1}{2}$.

In order to prove this, we show that the two direct summands $H^0(PO(Q), \mathrm{Ind}_{{}_fP_1}^{{}_fSL}[\phi_1])$ resp. $H^0(PO(Q), \mathrm{Ind}_{{}_fP_2}^{{}_fSL}[\phi_2])$ in $H^3_{\inf}(X/\Gamma_Q(m))$ are interchanged by the involution σ. Then the trace of σ on $H^3_{\inf}(X/\Gamma_Q(m))$ is given by

$$\begin{aligned}&\mathrm{trace}(\sigma, H^3_{\inf}(X/\Gamma_Q(m)))\\ &\quad = \mathrm{trace}(\sigma, \widetilde{St}_Q) \oplus \mathrm{trace}(\sigma, H^0(PO(Q), \mathrm{Ind}_{{}_fP_1}^{{}_fSL}[\phi_1]) \oplus H^0(PO(Q), \mathrm{Ind}_{{}_fP_2}^{{}_fSL}[\phi_2])\\ &\quad = \mathrm{trace}(\sigma, \widetilde{St}_Q) = 1 - \left(\frac{m+1}{2}\right),\end{aligned}$$

and the Lefschetz number at infinity is given by

$$\begin{aligned}L_{\inf}(\sigma, X/\Gamma_Q(m)) &= -\mathrm{trace}(\sigma, H^3_{\inf}(X/\Gamma_Q(m))) + \mathrm{trace}(\sigma, H^0_{\inf}(X/\Gamma_Q(m)))\\ &= \left(\frac{m+1}{2}\right).\end{aligned}$$

Replacing ψ_i by ϕ_i, $i=1,2$, the argument in Proposition 2.2 yields a direct sum decomposition

$$H^0(PO(Q), \mathrm{Ind}_{{}_fP_i}^{{}_fSL}[\phi_i]) = \bigoplus_{t \in PO(Q)\backslash {}_fSL/{}_fP_i} H^0(J_t, \phi_{i,t})$$

where t runs through all the double cosets in $PO(Q)\backslash {}_fSL/{}_fP_i$ and $J_t = t^{-1}PO(Q)t \cap {}_fP_i$. We fix $i=2$. As in 2.2(5), 2.2(6), 2.2(7), there are three double cosets t_1, t_2, t_3 in $PO(Q)\backslash {}_fSL/{}_fP_2$, and the corresponding subgroups in

${}_fP_2$ are

$$J_{t_1}=PO(Q)\cap {}_fP_2=\left\{\begin{pmatrix} a_{11} & & \\ & a_{22} & \\ & & \pm 1\end{pmatrix},\begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & 1\end{pmatrix}\cdot\begin{pmatrix} a_{11} & & \\ & a_{22} & \\ & & \pm 1\end{pmatrix}\middle| a_{ij}\in\mathbb{Z}/m\mathbb{Z}\right\},$$

$$J_{t_2}=t_2^{-1}PO(Q)\,t_2\cap {}_fP_2=\left\{\begin{pmatrix} a_{11} & & \\ a_{21} & a_{22} & \\ a_{31} & & \pm 1\end{pmatrix}\middle| a_{11}a_{22}=\pm 1, a_{ij}\in\mathbb{Z}/m\mathbb{Z}\right\},$$

$$J_{t_3}=t_3^{-1}PO(Q)\,t_3\cap {}_fP_2=\left\{\begin{pmatrix} \pm 1 & a_{12} & \\ & -1 & \\ \pm 1 & a_{32} & \pm 1\end{pmatrix},\begin{pmatrix} \pm 1 & a_{12} & \\ & 1 & \\ & a_{32} & \pm 1\end{pmatrix}\middle| a_{ij}\in\mathbb{Z}/m\mathbb{Z}\right\}.$$

Let K_1, K_2, K_3 denote respectively the image of $J_{t_1}, J_{t_2}, J_{t_3}$ in $SL_2^{\pm}(\mathbb{Z}/m\mathbb{Z})$ under the natural projection ${}_fP_2\to {}_fM_2\subset SL_2^{\pm}(\mathbb{Z}/m\mathbb{Z})$. In other words, K_i is obtained by taking the upper 2-by-2 block of matrices in J_{t_i}, i.e.

$$K_1=\left\{\begin{pmatrix} a_{11} & \\ & a_{22}\end{pmatrix},\begin{pmatrix} 0 & 1\\ 1 & 0\end{pmatrix}\cdot\begin{pmatrix} a_{11} & \\ & a_{22}\end{pmatrix}\middle| a_{ij}\in\mathbb{Z}/m\mathbb{Z}, a_{11}a_{22}=\pm 1\right\};$$

$$K_2=\left\{\begin{pmatrix} a_{11} & \\ a_{21} & a_{22}\end{pmatrix}\middle|\begin{matrix} a_{ij}\in\mathbb{Z}/m\mathbb{Z},\\ a_{11}a_{22}=\pm 1\end{matrix}\right\};$$

$$K_3=\left\{\begin{pmatrix} \pm 1 & a_{12}\\ & -1\end{pmatrix},\begin{pmatrix} \pm 1 & a_{12}\\ & 1\end{pmatrix}\middle| a_{12}\in\mathbb{Z}/m\mathbb{Z}\right\}.$$

Corresponding to each K_i, there is the congruence subgroup Γ_{K_i} in $SL_2^{\pm}(\mathbb{Z})$ defined by the group extension: $1\to\Gamma(2,m)\to\Gamma_{K_i}\to K_i\to 1$. If we recall that the representation ϕ_i of ${}_fP_i$ on $H^1_{\text{cusp}}(\Gamma(2,m))$ was given as the pullback of the natural action of $SL_2^{\pm}(\mathbb{Z}/m\mathbb{Z})$ on this space, then the $PO(Q)$-invariant subspace in $\text{Ind}_{{}_fP_2}^{{}_fSL}[\phi_2]$ can be expressed as a sum of cusp cohomology groups of these congruence subgroups:

$$\begin{aligned} H^0(PO(Q),\text{Ind}_{{}_fP_2}^{{}_fSL}[\phi_2]) &= \bigoplus_{i=1}^{3} H^0(J_{t_i},\Phi_2)\\ &= \bigoplus_{i=1}^{3} H^0(K_i, H^1_{\text{cusp}}(\Gamma(2,m))) = \bigoplus_{i=1}^{3} H^1_{\text{cusp}}(\Gamma(2,m))^{K_i}\\ &= \bigoplus_{i=1}^{3} H^1_{\text{cusp}}(\Gamma_{K_i}).\end{aligned}$$

In the same manner for ${}_fP_1$, we define the congruence subgroups $\Gamma_{K_i'}$ in $SL_2^{\pm}(\mathbb{Z})$, and the cohomology $H^0(PO(Q),\text{Ind}_{{}_fP_1}^{{}_fSL}[\phi_1])$ can be expressed in terms of the cusp cohomology of $\Gamma_{K_i'}$. Therefore we get

$$\bigoplus_{i=1}^{2} H^0(PO(Q),\text{Ind}_{{}_fP_i}^{{}_fSL}[\phi_i]) = \bigoplus_{i=1}^{3} H^1_{\text{cusp}}(\Gamma_{K_1})\oplus\bigoplus_{i=1}^{3} H^1_{\text{cusp}}(\Gamma_{K_1'}).$$

As in (3.4), the complex $T/\Gamma_Q(m)$ consists of 6 vertices $L_1, L_2, L_3, H_1, H_2, H_3$. Each of these vertices gives rise to a boundary component X_{L_i}, X_{H_i}, and the fundamental groups of these components are respectively the extension of Γ_{K_i}, $\Gamma_{K_i'}$ by a free abelian group of rank 2

$$0 \to \mathbb{Z}^2 \to \pi_1(X_{H_i}) \to \Gamma_{K_i} \to 0, \qquad 0 \to \mathbb{Z}^2 \to \pi_1(X_{L_i}) \to \Gamma_{K_i'} \to 0.$$

From a spectral sequence arguement, the direct sums $\bigoplus H^1_{\text{cusp}}(\Gamma_{K_i})$, resp. $\bigoplus H^1_{\text{cusp}}(\Gamma_{K_i'})$ are isomorphic to cohomology $\bigoplus H^1(X_{H_i})$ resp. $\bigoplus H^1(X_{L_i})$ $i=1,2,3$. Since σ interchanges the components X_{H_i} with X_{L_i}, its induced action on the cohomology also interchanges these two direct summands $\bigoplus_{i=1}^{2} H^0(PO(Q), \operatorname{Ind}_{fP_i}^{rSL}[\phi_i])$. This proves 3.6.

§ 4. Lefschetznumber and Ramified Covering Spaces

4.1. In this section, we set up some of the preliminary steps in the computation of the Lefschetznumber $L(\sigma, X/\Gamma_Q(m))$. As mentioned before, $\Gamma_Q(m)$ is a semi-direct product of the torsionfree group $\Gamma_D(m)$ and a dihedral group $\mathcal{O}$ of order 8 in $\Gamma_Q(m)$. So, the quotient space $X/\Gamma_Q(m)$ can be thought of as the quotient of $X/\Gamma_D(m)$ under the natural action of $\mathcal{O}$. In particular, we have by 1.2 a commutative diagram

$$\begin{array}{ccc} X/\Gamma_D(m) & \xrightarrow{\sigma} & X/\Gamma_D(m) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ X/\Gamma_Q(m) & \xrightarrow{\sigma} & X/\Gamma_Q(m), \end{array}$$

where π denotes the natural projection and σ the automorphism induced by the involution σ. For an element γ in $\mathcal{O}$ we denote by $\sigma\gamma$ the automorphism $x \mapsto x^\sigma \cdot \gamma$ of $X/\Gamma_D(m)$.

4.2. Proposition. *Let $F^{\sigma\gamma}$ denote the fixpoint set of $\sigma\gamma$, $\gamma \in \mathcal{O}$, in $X/\Gamma_D(m)$; then we have for the Lefschetznumber the following formula*

$$\begin{aligned} L(\sigma, X/\Gamma_Q(m)) &= \frac{1}{|\mathcal{O}|} \sum_{\gamma \in \mathcal{O}} L(\sigma\gamma, X/\Gamma_D(m)) \\ &= \frac{1}{|\mathcal{O}|} \sum_{\gamma \in \mathcal{O}} \chi(F^{\sigma\gamma}), \end{aligned} \tag{1}$$

where $\chi(F^{\sigma\gamma})$ denotes the Euler characteristic of $F^{\sigma\gamma}$.

The second equality is given by the classical Lefschetz fixpoint formula which says (by a modified version of the formulation in [19], 1.9) for each $\gamma \in \mathcal{O}$, $L(\sigma\gamma, X/\Gamma_D(m)) = \chi(F^{\sigma\gamma})$. The following lemma implies then (1). We thank J.

Rohlfs for this general argument which simplifies our original approach; there we proved an analogue for the Euler characteristics involved here, precisely the formula: $\chi((X/\Gamma_Q(m))^\sigma)=\frac{1}{|\mathcal{O}|}\sum\chi(F^{\sigma\gamma})$.

4.3. Lemma. *Let Γ_0 be a normal subgroup of finite index in $\Gamma=\Gamma_Q(m)$, such that Γ_0 is stable under σ. Then we have*

$$L(\sigma,\Gamma)=\frac{1}{[\Gamma:\Gamma_0]}\sum_{\gamma\in\Gamma/\Gamma_0}L(\sigma\gamma,\Gamma_0).$$

Since $\gamma_0\in\Gamma_0$ operates by conjugation on Γ_0 and induces the trivial map on $H^*(\Gamma_0,\mathbb{C})$ the Lefschetznumbers $L(\sigma\gamma,\Gamma_0)$ are well defined. Then Γ/Γ_0 acts on $H^*(\Gamma_0,\mathbb{C})$ and we have $H^*(\Gamma,\mathbb{C})=H^*(\Gamma_0,\mathbb{C})^{\Gamma/\Gamma_0}$ as follows e.g. from the Hochschild-Serre spectral sequence. Let $p^*\colon H^*(\Gamma_0,\mathbb{C})\to H^*(\Gamma_0,\mathbb{C})^{\Gamma/\Gamma_0}$ be the projection operator defined by

$$p^*(a)=\frac{1}{[\Gamma:\Gamma_0]}\sum_{\gamma\in\Gamma/\Gamma_0}\gamma^*(a).$$

Then we have

$$\begin{aligned}
L(\sigma,\Gamma)&=\sum_{i=0}^{\infty}(-1)^i\operatorname{tr}\sigma^i|_{H^i(\Gamma,\mathbb{C})}\\
&=\sum_{i=0}^{\infty}(-1)^i\operatorname{tr}\sigma^i|_{p^i(H^i(\Gamma_0,\mathbb{C}))}\\
&=\sum_{i=0}^{\infty}(-1)^i\operatorname{tr}(\sigma^i\circ p^i)|_{H^i(\Gamma_0,\mathbb{C})}\\
&=\sum_{i=0}^{\infty}\frac{1}{[\Gamma:\Gamma_0]}(-1)^i\sum_{\gamma\in\Gamma/\Gamma_0}\operatorname{tr}(\sigma^i\circ\gamma^i)|_{H^i(\Gamma_0,\mathbb{C})}\\
&=\frac{1}{[\Gamma:\Gamma_0]}\sum_{\gamma\in\Gamma/\Gamma_0}L(\sigma\gamma,\Gamma_0)
\end{aligned}$$

as claimed.

One sees that the argument is of a general nature.

§ 5. The Euler Characteristic of the Fixpointset F^σ

The basic idea underlying our computations of the various Lefschetznumbers $L(\sigma\gamma,\Gamma_D(m))$, $\gamma\in\mathcal{O}$, (resp. Euler characteristics) in 4.2 is due to J. Rohlfs. In his paper [19] he considers the Cartan involution τ of $SL_n(\mathbb{R})$ given by $A\mapsto(A^t)^{-1}$ and the cohomology of a torsionfree τ-stable congruence subgroup $\Gamma(m)\subset SL_n(\mathbb{Z})$. Since our computational work can be viewed a "variation of his theme" we have to assume some familiarity with his approach in computing the Lefschetz number $L(\tau,\Gamma(m))$. In this section we will deal with the fixpoint set F^σ and compute its Euler characteristic. We will precisely explain in this

case in which way we have to modify and to improve Rohlfs' methods. In order to leave lateron computational details to the reader (cf. §6) this is worked out quite carefully.

The final result of this section will be

5.1. Proposition. *The Euler characteristic of the fixpointset F^σ of the involution σ on $X/\Gamma_D(m)$* (cf. §1) *is given by*

$$\chi(F^\sigma)=(1/12)(m^2-1).$$

First of all, we note that the involution σ is defined by "twisting" the Cartan involution τ by means of the matrix $M=\begin{pmatrix}1 & & \\ & 1 & \\ & & m\end{pmatrix}$, i.e. $\sigma(A)=M\tau(A)M^{-1}$, $A\in SL_3(\mathbb{R})$. One observes that M is not unimodular. However we will modify Rohlfs' method in a slightly more general setting.

5.2. Let σ be an involution of $SL_n(\mathbb{R})$ defined by

$$\sigma(A)=M\tau(A)M^{-1}, \qquad A\in SL_n(\mathbb{R}), \tag{1}$$

where τ is the Cartan involution and M is an integral symmetric matrix of determinant $\det M=m$. The last condition is still unnecessarily restrictive, but is introduced to simplify our notations lateron. We consider a torsionfree, σ-stable, arithmetic subgroup of $SL_n(\mathbb{Z})$. The involution induced on the associated symmetric space $X=SO_n(\mathbb{R})\backslash SL_n(\mathbb{R})$ resp. the arithmetic quotient X/Γ is again denoted by σ. Then the connected components of the fixpoint set $(X/\Gamma)^\sigma$ are parametrized by the nonabelian cohomology set $H^1(\mathfrak{g},\Gamma)$ with $\mathfrak{g}=\{1,\sigma\}$. (For a definition of $H^1(\mathfrak{g},\Gamma)$ we refer to [24], Chap. I, 5.) Although this observation is an analogue of the result 1.3 in [19] we explain this basic parametrization in here for the convenience of the reader.

Consider a cocycle $b=(1,b)$ for $H^1(\mathfrak{g},\Gamma)$. We will denote the class determined by b with the same letter. By definition $b=(1,b)$, $b\in\Gamma$, is a cocycle if $b\cdot\sigma(b)=1$. Two cocycles $(1,b)$ resp. $(1,d)$ are equivalent if there exists a $c\in\Gamma$ such that $d=c^{-1}\cdot b\cdot\sigma(c)$. To the given cocycle $b=(1,b)$ for $H^1(\mathfrak{g},\Gamma)$ we associate by $\bar\sigma(A)=b\cdot\sigma(A)\cdot b^{-1}$ a new operation $\bar\sigma$ on $SL_n(\mathbb{R})$ and on Γ. Analogously we define $\bar\sigma(x)=\sigma(x)\cdot b^{-1}$ for $x\in X$. We see that $\bar\sigma(x\cdot g)=\bar\sigma(x)\cdot\bar\sigma(g)$ holds for $x\in X$ and $g\in SL_n(\mathbb{R})$. Therefore this twisted operation induces an operation $\bar\sigma$ on X/Γ, and this operation coincides with the operation σ on X/Γ. We denote by $X(b)$ the set of points in X fixed under the twisted involution $\bar\sigma$ and by $\Gamma(b)$ the set of elements in Γ fixed under $\bar\sigma$. Since Γ is torsionfree the natural map $X(b)/\Gamma(b)\to X/\Gamma$ is injective, and its image

$$F(b):=\mathrm{Im}(X(b)/\Gamma(b)\to X/\Gamma) \tag{2}$$

lies in the fixpointset $(X/\Gamma)^\sigma$. Note that $F(b)$ depends only on the cohomology class in $H^1(\mathfrak{g},\Gamma)$ represented by the cocycle b. Since the twisted operation acts as an isometry on X we see that $F(b)$ is non-empty.

On the other hand, if $\bar{x}\in(X/\Gamma)^\sigma$ is represented by $x\in X$ there exists an uniquely determined element b in Γ such that $\sigma(x)=x\cdot b$. Since σ is an involution we have $x=\sigma(\sigma(x))=\sigma(x\cdot b)=\sigma(x)\cdot\sigma(b)=x\cdot b\cdot\sigma(b)$. It follows that $b\cdot\sigma(b)=1$, and $(1,b)$ is a cocycle for $H^1(\mathfrak{g},\Gamma)$. For another representative $x'=x\cdot c$, $c\in\Gamma$, of $\bar{x}\in(X/\Gamma)^\sigma$ the associated cocycle is $b'=c^{-1}\cdot b\cdot\sigma(c)$. Thus every point in the fixpointset $(X/\Gamma)^\sigma$ determines uniquely a class in $H^1(\mathfrak{g},\Gamma)$, and we have a disjoint union of connected non-empty sets

$$(X/\Gamma)^\sigma=\coprod_{b\in H^1(\mathfrak{g},\Gamma)} F(b). \tag{3}$$

5.3. By definition, a cocycle $b=(1,b)$, $b\in\Gamma$, for $H^1(\mathfrak{g},\Gamma)$ satisfies $b\cdot\sigma(b)=1$ i.e.

$$b\cdot M\cdot\tau(b)\,M^{-1}=1. \tag{1}$$

Two cocycles b, d represent the same class if there exists a c in Γ with

$$d=c^{-1}\cdot b\cdot\sigma(c)=c^{-1}\cdot b\cdot M\cdot\tau(c)\cdot M^{-1}. \tag{2}$$

If we put $B=b\cdot M$ resp. $D=d\cdot M$ then conditions (1) and (2) are written as

$$B \text{ is symmetric, i.e. } B={}^tB \tag{3}$$

$$B \text{ and } D \text{ are } \Gamma\text{-congruent, i.e. } B=cD\cdot{}^tc \text{ for some } c\in\Gamma. \tag{4}$$

It follows that $H^1(\mathfrak{g},\Gamma)$ can be interpreted as the set of Γ-congruence classes of symmetric matrices of the form $b\cdot M$, $b\in\Gamma$.

5.4. In order to compute $H^1(\mathfrak{g},\Gamma)$ we describe it in terms of local data. A typical place of $\mathbb{Q}$ will be denoted by v, and, as usual, the ring of adeles of $\mathbb{Q}$ is denoted by A.

We suppose that $\Gamma=\bigcap_v(\Gamma_v\cap SL_n(\mathbb{Q}))$, where Γ_v is an open compact subgroup of $SL_n(\mathbb{Q}_v)$ for all finite places v of $\mathbb{Q}$ ($\Gamma_v=SL_n(\mathbb{Z}_v)$ for almost all v) and $\Gamma_\infty=SL_n(\mathbb{R})$. Note that by the strong approximation property of $SL_n/\mathbb{Q}$ the closure of Γ in $SL_n(\mathbb{Q}_v)$ is Γ_v for all finite places v. Since Γ is σ-stable by assumption (cf. 5.2) all Γ_v are σ-stable and the inclusion $\Gamma\to\Gamma_v$ induces a map $h_v\colon H^1(\mathfrak{g},\Gamma)\to H^1(\mathfrak{g},\Gamma_v)$. We have $H^1(\mathfrak{g},\Gamma_v)=\{1\}$ for almost all v (cf. [19], 3.1); therefore $\prod_v h_v=h$ induces a map

$$h\colon H^1(\mathfrak{g},\Gamma)\to\coprod_v H^1(\mathfrak{g},\Gamma_v). \tag{1}$$

On the other hand, $H^1(\mathfrak{g},\Gamma_v)$ admits an interpretation as Γ_v-congruence classes of symmetric matrices just as before in 5.3; so for b_v in $H^1(\mathfrak{g},\Gamma_v)$ the Hasse-Witt invariant $c_v(b_v)\in\{\pm1\}$ is defined. For $\coprod\{b_v\}\in\coprod_v H^1(\mathfrak{g},\Gamma_v)$ we can form the product

$$c\colon\coprod_v H^1(\mathfrak{g},\Gamma_v)\to\{\pm1\},\qquad\prod_v b_v\mapsto\prod_v c_v(b_v) \tag{2}$$

which makes sense because almost all factors of the infinite product are 1.

We denote by $SO(M)(A)$ the group of A-valued points of the special orthogonal group $SO(M)/\mathbb{Q}$ corresponding to M. Then there is an exact sequence of pointed sets

$$1 \to SO(M)(\mathbb{Q}) \backslash SO(M)(A)/\Pi \Gamma_v^{\sigma} \xrightarrow{\partial} H^1(\mathfrak{g}, \Gamma) \xrightarrow{h} \coprod_v H^1(\mathfrak{g}, \Gamma_v) \xrightarrow{c} \{\pm 1\} \to 1. \tag{3}$$

Here Γ_v^{σ} denotes the subgroup of elements in Γ_v fixed under σ. We refer to Proposition 2.2 in [19] for a proof, where one can also find a description of the injective map ∂.

Remark. This sequence determines the image of h and the fiber of h through the distinguished element of $H^1(\mathfrak{g}, \Gamma)$. If $b=(1, b)$ is a cocycle for $H^1(\mathfrak{g}, \Gamma)$ then the fiber of h through the element represented by b is obtained by twisting (cf. [24], I, 5.4). One has $h^{-1}(h(b)) \cong SO(B)(\mathbb{Q}) \backslash SO(B)(A)/\Pi \Gamma_v(B)$ where $\Gamma_v(B)$ is the subgroup in Γ_v fixed under the 'twisted' involution $g \to b\sigma(g)b^{-1}$ (cf. 0.6 (4)).

5.5. Let B be a symmetric matrix of the form $b \cdot M$. By 5.3 this represents a cohomology class in $H^1(\mathfrak{g}, \Gamma)$, and by 5.2 there is a fixpoint component $F(b)$. There is a biinvariant measure ω_v on $SO(B)(\mathbb{Q}_v)$ for every place v of $\mathbb{Q}$ such that

$\omega = \Pi \omega_v$ is the Tamagawa measure on $SO(B)(A)$ (1)

$(-1)^{rs} \binom{l}{s} \operatorname{vol}_{\omega_\infty}(SO_n(\mathbb{R}))^{-1}$. ω_∞ is the Euler-Poincaré measure (in the sense of Serre [23] § 3) on $SO(B)(\mathbb{R})$ where $(r, 2s)$ is the signature of B and $l = \left[\frac{n}{2}\right]$ (2)

(see [11], [19] §4 resp. 5.6 in this paper). Let $\Gamma(B)$ denote the subgroup in Γ fixed under the 'twisted' involution $g \mapsto b \cdot \sigma(g) \cdot b^{-1}$. By 5.2 there is a natural identification between the fixpoint component $F(b)$ and the manifold $SO(B)(\mathbb{R})/\Gamma(B)$. Hence by the above condition (2), we have

$$\chi(F(b)) = (-1)^{r \cdot s} \binom{l}{s} \frac{1}{\operatorname{vol}_{\omega_\infty}(SO_n(\mathbb{R}))} \int_{SO(B)(\mathbb{R})/\Gamma(B)} \omega_\infty. \tag{3}$$

As described in 5.4, there is an isomorphism between the inverse image $h^{-1}h(b)$ and the double coset space $SO(B)(\mathbb{Q}) \backslash SO(B)(A)/\Pi \Gamma_v(B)$. It turns out to be easier and sufficient to compute the sum $\sum_{c \in h^{-1}h(b)} \chi(F(c))$ of all Euler characteristics together. The double coset space $SO(B)(\mathbb{Q}) \backslash SO(B)(A)/\prod_v \Gamma_v(B)$ is finite, and we choose a set of representatives $x_1, \ldots, x_{n(B)}$,

$$SO(B)(A) = \bigcup_{i=1}^{n(B)} \prod \Gamma_v(B) \cdot x_i \cdot SO(B)(\mathbb{Q}).$$

By strong approximation, each element x_i can be written as a product $x_i = g_i u_i$ where $g_i \in SL_n(\mathbb{Q})$, $u_i \in \prod_v \Gamma_v$. From the definition of the map ∂ in 5.4 (3), it is

easy to see that the element $c_i = g_i^{-1} \cdot B \cdot g_i^{\sigma}$ represents a cohomology class in $H^1(\mathfrak{g}, \Gamma)$, and as $i = 1, \ldots, n(B)$, these elements run through all classes in $h^{-1}h(B)$. In addition, the fixpoint component $F(c_i)$ can be identified with $F(c_i) \cong SO(B)(\mathbb{R})/g_i \Gamma g_i^{-1} \cap SO(B)(\mathbb{R})$, and the computation reduces to the problem of calculating the sum

$$\sum_{i=1}^{n(B)} \mathrm{vol}(SO(B)(\mathbb{R})/g_i \Gamma g_i^{-1} \cap SO(B)(\mathbb{R})).$$

In 5.1, 5.2 of [19], Rohlfs showed that there is a sequence of further reductions in writing the above sum; one gets

$$\sum_{i=1}^{n(B)} \mathrm{vol}(SO(B)(\mathbb{R})/g_i \Gamma g_i^{-1} \cap SO(B)(\mathbb{R}))$$

$$= \frac{1}{\prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(B))} \cdot \mathrm{vol}(SO(B)(A)/SO(B)(\mathbb{Q})).$$

By a result of T. Tamagawa (see [29]) we have $\mathrm{vol}(SO(B)(A)/SO(B)(\mathbb{Q})) = 2$, and so

$$\sum_{c \in h^{-1}h(b)} \chi(F(c)) = 2(-1)^{rs} \binom{l}{s} \mathrm{vol}(SO_n(\mathbb{R}))^{-1} \prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(B))^{-1}. \tag{4}$$

5.6. For our applications, we need an explicit description of invariant measures ω_v on $SO(B)(\mathbb{Q}_v)$ which satisfy the conditions (1) and (2) in 5.5. Let

$$so(B)(\mathbb{Q}) = \{A \in M_n(\mathbb{Q}) \mid A + B \cdot {}^tA \cdot B^{-1} = 0\}$$

be the rational Lie algebra of the special orthogonal group $SO(B)(\mathbb{Q})$. Over this algebra, there is a bilinear pairing e defined by the formula:

$$e(X, Y) = -\mathrm{Tr}(X \cdot Y)/2, \qquad X, Y \in so(B)(\mathbb{Q}).$$

Since e differs from the Killing form only by a constant factor $-2(n-2)$, this is a non-degenerate symmetric pairing invariant under the adjoint action of $SO(B)(\mathbb{Q})$, $n > 2$. Define a top dimensional alternating form on the vector space $so(B)(\mathbb{Q})$ by means of the formula:

$$\omega(x_1, \ldots, x_d) = \mathrm{sign}(x_1, \ldots, x_d)(|\det(e(x_i, x_j))_{i,j=1,\ldots,d}|)^{1/2}. \tag{1}$$

Here $d = \dfrac{n(n-1)}{2}$, the function $\mathrm{sign}(x_1, \ldots, x_d) = 1$ if the sequence $\{x_1, \ldots, x_d\}$ of vectors forms a basis system compatible with respect to a fixed orientation, and $\mathrm{sign}(x_1, \ldots, x_d) = -1$ otherwise. However, unlike the situation when B is non-singular, $\det B = 1$, the above alternating form is not necessarily defined over the rationals. To simplify our discussion, we restrict ourselves to only those non-degenerate symmetric forms B such that ω is rational.

Assumption A. *The number* $|\det(e(x_i, x_j))_{i,j=1,\ldots,d}|$ *is a square for any set of linear independent vectors* $x_1, \ldots, x_d$ *in* $so(B)(\mathbb{Q})$.

Example. Let M be a n-by-n symmetric matrix of the form

$$M=\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & m \end{pmatrix}.$$

Then the following matrices $\{U_{i,j},\ U_{i,n}\}_{1\leqq i<j<n}$ form a basis system for $so(M)(\mathbb{Q})$:

$$U_{i,j}=i\begin{pmatrix} & & & \\ & & 1 & \\ & & & \\ j & -1 & & \end{pmatrix}\quad 1\leqq i<j<n,\qquad U_{i,n}=\begin{pmatrix} & & \\ & & 1 \\ & & \\ & -m & \end{pmatrix}\quad 1\leqq i<n.$$

A straightforward computation shows that

$$\begin{aligned} e(U_{i,j}, U_{k,l})&=0 \quad \text{if } (i,j)\neq(k,l),\\ e(U_{i,j}, U_{i,j})&=-1 \quad \text{if } 1\leqq i<j<n,\\ e(U_{i,n}, U_{i,n})&=-m \quad \text{if } 1\leqq i<n \end{aligned}$$

and $|\det(e(U_{i,j}, U_{k,l})_{|1\leqq i<j,\,1\leqq k<l|}=m^{(n-1)}$. Hence our assumption A is satisfied if and only if n is odd.

For simplicity we assume from now on that $n=3$.

By left translation, we extend ω to a biinvariant form ω of highest degree on $SO(B)(\mathbb{Q})$. Since ω is rational, the method of Tamagawa (see [29]) allows us to define a biinvariant measure ω_v on each of the locally compact groups $SO(B)(\mathbb{Q}_v)$ such that condition 5.5 (1) resp. 5.5 (2) is satisfied by [29] resp. 4.5 in [19].

At a finite prime $v\neq\infty$, we compute the volume $\mathrm{vol}_{\omega_v}(\Gamma_v(B))$ by considering a sufficiently deep congruence subgroup $\Gamma_v(B)(j)$, $j\gg 0$ so that the exponential map becomes bijective, $\exp: so(B_v)(j)\xrightarrow{\simeq}\Gamma_v(B)(j)$, and its differential D exp becomes the identity map (cf. [19], § 4). Then

$$\begin{aligned} \mathrm{vol}(\Gamma_v(B)(j))&=\int_{\Gamma_v(B)(j)}\omega_v=\int_{so(B_v)(j)}\omega\\ &=\|\det(e(x_i,x_j))_{i,j=1,\ldots,d}\|_v^{1/2} \end{aligned}\tag{2}$$

where $x_1,\ldots,x_d$ forms a $\mathbb{Z}_v$-basis for $so(B_v)(j)$. To recover the volume of $\Gamma_v(B)$, we use the formula:

$$\mathrm{vol}(\Gamma_v(B))=[\Gamma_v(B):\Gamma_v(B)(j)]\,\mathrm{vol}(\Gamma_v(B)(j)),\tag{3}$$

where the notation $[\Gamma_v(B):\Gamma_v(B)(j)]$ stands for the index of the subgroup $\Gamma_v(\mathrm{B})(\mathrm{j})$ in $\Gamma_v(\mathrm{B})$.

Note that the space $so(B)(\mathbb{Q}_v)$ is obtained by taking the tensor product (extension of scalars) $so(B)(\mathbb{Q}_v)\simeq so(B)(\mathbb{Q})\underset{\mathbb{Z}}{\otimes}\mathbb{Z}_v$. Our assumption A implies

that for any linear independent elements $x_1, \ldots, x_d$ the expression $\|\det(e(x_i, x_j))_{i,j=1,\ldots,d}\|_v$ is a square and $\mathrm{vol}(\Gamma_v(B))$ is a rational number. Conversely, if we can prove that the expression $\|\det(e(x_i, x_j)_{i,j=1,\ldots,d})\|_v$ is a square at every local place v, then for a set of linear independent elements $y_1, \ldots, y_d$ in $so(B)(\mathbb{Q})$ the term $|\det(e(y_i, y_j)_{i,j=1,\ldots,d}|$ is also a square. This means that our assumption (A) is equivalent to the condition: *local volume* $\mathrm{vol}_v(\Gamma_v(B))$ *is a well defined rational number for every prime* v.

It is worthwhile to remark two techniques in computing local volumes. First, if two symmetric matrices B_1 and B_2 are Γ_v-congruent to each other, $B_2 = cB_1\,{}^t c$, $c \in \Gamma_v$ then they have the same volume. For there exists an isomorphism of Lie algebras Ad_c: $so(B_1) \xrightarrow{\sim} so(B_2)$ given by $A \mapsto cAc^{-1}$. This isomorphism is an isometry with respect to the inner product defined by e, and also the congruence subspaces $so(B_1)(j) \xrightarrow{\sim} so(B_2)(j)$. Hence we have the same volumes,

$$\begin{aligned} \mathrm{vol}(\Gamma_v(B_1)(j)) &= \|\det(e(x_i, x_j)_{i,j=1,\ldots,d})\|_v^{1/2} \\ &= \|\det(e(Ad_c x_i, Ad_c x_j)_{i,j=1,\ldots,d})\|_v^{1/2} \\ &= \mathrm{vol}(\Gamma_v(B_2)(j)). \end{aligned} \tag{4}$$

Another technique in computing the local volume is to change a symmetric matrix B by a non-zero element α in $\mathbb{Z}_v$. Note that the symmetric matrices B, αB have the same group of automorphisms, $SO(B) = SO(\alpha B)$, the same Lie algebras $so(B) = so(\alpha B)$, the same congruence subgroup $\Gamma_v(B) = \Gamma_v(\alpha B)$, and the same family of congruence subspaces $so(B_v)(j) = so(\alpha B_v)(j)$. Hence they have the same volume

$$\mathrm{vol}(\Gamma_v(B)) = \mathrm{vol}(\Gamma_v(\alpha B)).$$

5.7. Proof of Proposition 5.1. By an abuse of notation, we denote by Γ the congruence subgroup $\Gamma_D(m)$, $\Gamma = \Gamma_D(m)$, in this proof. From 5.2, we have the decomposition of the fixpoint set $F^\sigma = \coprod_{\{b\} \in H^1(\mathfrak{g}, \Gamma)} F(b)$. The cohomology $H^1(\mathfrak{g}, \Gamma)$ is the same as Γ-congruence classes of symmetric matrices $B = b \cdot M$, $b \in \Gamma$. Note in order for a symmetric matrix to be of this form, it must satisfy the conditions:

$$B \equiv \begin{pmatrix} b_{11} & 0 & 0 \\ 0 & b_{22} & 0 \\ 0 & 0 & 0 \end{pmatrix} \bmod m, \tag{1}$$

$$\varepsilon(b_{11}) = \varepsilon(b_{22}) = 1, \tag{2}$$

$$\det B = m. \tag{3}$$

Thus, to study the cohomology $H^1(\mathfrak{g}, \Gamma)$, we have to study Γ-congruence classes of 3-by-3 symmetric matrices satisfying conditions (1)–(3).

For $v \neq m$, we let Γ_v be the group $SL_3(\mathbb{Z}_v)$, and for $v = m$ we let $\Gamma_{(m)}$ be the subgroup in $SL_3(\mathbb{Z}_{(m)})$ consisting of matrices

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}, \tag{4}$$

$$a_{12} \equiv a_{21} \equiv a_{31} \equiv a_{32} \equiv 0 \bmod m, \tag{5}$$

$$\varepsilon(a_{11}) = \varepsilon(a_{22}) = \varepsilon(a_{33}) = 1. \tag{6}$$

We now compute the local cohomology groups $H^1(\mathfrak{g}, \Gamma_v)$ in the sequence 5.4 (3).

Lemma (5.7.1). *Let Γ_v be defined as above. Then the local cohomology sets $H^1(\mathfrak{g}, \Gamma_v)$ are given by the following formulas:*

(i) $H^1(\mathfrak{g}, \Gamma_\infty) = \{M, \mu \cdot M\}$,

(ii) For $v \neq 2, \infty$, $H^1(\mathfrak{g}, \Gamma_v) = \{M\}$,

(iii) $H^1(\mathfrak{g}, \Gamma_{(2)}) = \{M, \mu \cdot M\}$

where $M = \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix}$, $\mu = \begin{pmatrix} -1 & & \\ & -1 & \\ & & 1 \end{pmatrix}$, $\mu \cdot M = \begin{pmatrix} -1 & & \\ & -1 & \\ & & m \end{pmatrix}$.

The cohomology group $H^1(\mathfrak{g}, \Gamma_v)$ admits an interpretation as set of Γ_v-congruence classes of symmetric matrices as in (5.4).

For $v = \infty$, $\Gamma_v = SL_3(\mathbb{R})$, the computation reduces then to the classification of such symmetric matrices. Because of the restriction on the determinant, there are two alternatives: positive definite, or indefinite of signature $(1, 2)$. Hence $H^1(\mathfrak{g}, SL_3(\mathbb{R}))$ consists of two elements, and the symmetric matrices M, $\mu \cdot M$ are their representatives. This proves (i).

For $v \neq 2, m, \infty$, we have $\Gamma_v = SL_3(\mathbb{Z}_v)$. As wellknown, two integral v-adic forms of unit determinant are $SL_3(\mathbb{Z}_v)$-congruent if they have the same number of variables and the same determinant invariant (cf. [13] p. 92). In our situation, the determinants of our symmetric matrices are all equal to m and so they must be Γ_v-congruent to each other,

$$H^1(\mathfrak{g}, SL_3(\mathbb{Z}_v)) = \{M\} \quad \text{for } v \neq 2, m, \infty.$$

To prove (ii), it remains to consider the prime $v = m$. As before, a cohomology class in $H^1(\mathfrak{g}, \Gamma_{(m)})$ is represented by symmetric matrices:

$$S = \begin{pmatrix} S_{11} & S_{12} & S_{13} \\ S_{12} & S_{22} & S_{23} \\ S_{13} & S_{23} & S_{33} \end{pmatrix}, \qquad S_{12} \equiv S_{13} \equiv S_{23} \equiv 0 \bmod m,$$

$$\varepsilon(S_{11}) = \varepsilon(S_{22}) = 1, \qquad \det S = m.$$

To prove that S is $\Gamma_{(m)}$-congruent to M, we proceed to simplify S by a sequence of elementary row-column operations. First, S_{11} is a unit in $\mathbb{Z}_{(m)}$ and we want to kill off the off-diagonal elements in the first row and the first column, let

$$C = \begin{pmatrix} 1 & \dfrac{-S_{12}}{S_{11}} & \dfrac{-S_{13}}{S_{11}} \\ & 1 & \\ & & 1 \end{pmatrix}.$$

Since $\frac{S_{12}}{S_{11}} \equiv \frac{S_{13}}{S_{11}} \equiv 0 \bmod m$, this is an element in $\Gamma_{(m)}$. By applying C to S on the right and ${}^tC$ on the left, we obtain a symmetric matrix $S^1 = {}^tC \cdot S \cdot C$ satisfying the above cocycle conditions, but with the entries $S'_{12} = S'_{13} = 0$

$$ {}^tC \cdot S \cdot C = \begin{pmatrix} S_{11} & 0 & 0 \\ 0 & S'_{22} & S'_{23} \\ 0 & S'_{23} & S'_{33} \end{pmatrix}. $$

Repeating the same procedure to kill off S'_{23}, we set

$$ D = \begin{pmatrix} 1 & & \\ & 1 & \frac{-S'_{23}}{S'_{22}} \\ & & 1 \end{pmatrix} \qquad \frac{S'_{23}}{S'_{22}} \equiv 0 \bmod m $$

in $\Gamma_{(m)}$ and consider the symmetric matrix $S'' = {}^tDS'D$. Then

$$ S'' = \begin{pmatrix} S_{11} & 0 & 0 \\ 0 & S'_{22} & 0 \\ 0 & 0 & S''_{33} \end{pmatrix} \quad \text{with } S''_{33} = \frac{m}{S_{11}S'_{22}}. $$

Under our assumption $m \equiv 3 \bmod 4$, the condition $\varepsilon(S_{11}) = \varepsilon(S'_{22}) = 1$ implies that they lie in the odd exponent subgroup $\mathbb{Z}\Big/\frac{m-1}{2}$ in $(\mathbb{Z}/m)^* \cong \mathbb{Z}/m - 1 \cong \mathbb{Z}/2 \times \mathbb{Z}\Big/\frac{m-1}{2}$ (cf. §1). Hence they can be written as squares of some elements in $\mathbb{Z}\Big/\frac{m-1}{2}$. By Hensel's lemma, there exist elements a, b in $\mathbb{Z}_{(m)}$ such that

$$ S_{11} = a^2, \quad S'_{22} = b^2, \quad \varepsilon(a) = \varepsilon(b) = 1. $$

Then S'' is $\Gamma_{(m)}$-congruent to the symmetric matrix

$$ {}^tA \cdot S'' \cdot A = \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix} \quad \text{with } A = \begin{pmatrix} \frac{1}{a} & & \\ & \frac{1}{b} & \\ & & ab \end{pmatrix}. $$

Thus we have $S \sim S' \sim S'' \sim M$ and this completes the proof of (ii).

It remains to investigate the situation when $v = 2$. From the theory of quadratic forms, two properly primitive forms of unit determinants in $\mathbb{Z}_{(2)}$ are congruent if they are congruent modulo 4 and their determinant invariants are the same modulo 8 (cf. [13], p. 92). By properly primitive we mean some of diagonal entries S_{ii} is odd. For ternary forms, it is easy to see that if every diagonal entry is even, then its determinant is divisible by 2. Since the de-

terminant of our symmetric matrix is the odd prime m, we only have to consider properly primitive forms. Again from the theory of quadratic forms, we can always diagonalize with respect to $\Gamma_{(2)}$ such ternary forms (see [13], p. 85). In terms of matrices, this means

$$S \sim \begin{pmatrix} a & & \\ & b & \\ & & c \end{pmatrix}.$$

Congruent modulo 4, the diagonal entries can be either $+1$ or -1, and their product $abc \equiv m \equiv -1 \bmod 4$. Up to permutation of diagonal entries, there are only two possibilities:

$$S \equiv \begin{pmatrix} 1 & & \\ & 1 & \\ & & -1 \end{pmatrix} \bmod 4, \quad \text{or } S \equiv \begin{pmatrix} -1 & & \\ & -1 & \\ & & -1 \end{pmatrix} \bmod 4.$$

Since the determinant is fixed modulo 8, this implies that $H^1(\mathfrak{g}, \Gamma_{(2)})$ can have at most two distinct elements. On the other hand, a direct computation of Hasse-Witt invariants shows that the two elements M, $\mu \cdot M$ belong to different congruence classes; one has $c_2(M)=1$ resp. $c_2(\mu \cdot M)=-1$. The proof of the lemma is complete.

5.8. Continuation of the Proof of 5.1. Having determined the local cohomology groups we now determine the image of the map h in the exact sequence 5.4 (3)

$$1 \to SO(M)(\mathbb{Q}) \backslash SO(M)(A) / \prod_v \Gamma_v^{\sigma} \to H^1(\mathfrak{g}, \Gamma) \xrightarrow{h} \coprod_v H^1(\mathfrak{g}, \Gamma_v) \xrightarrow{c} \{\pm 1\} \to 1.$$

Lemma. *With the notation as introduced in the previous lemma, the image of h consists of two elements: $x=(x_v)$, $y=(y_v)$ where*

(i) $x_v = y_v = M$, *for all* $v \neq 2, \infty$,

(ii) $x_\infty = x_2 = M$,

(iii) $y_\infty = y_2 = \mu \cdot M$.

In fact let $\{B_1\}=b_1$ denote the class in $H^1(\mathfrak{g}, \Gamma)$ represented by M, and let $\{B_2\}=b_2$ denote the class represented by $\mu \cdot M$. Then $h(b_1)=x$, $h(b_2)=y$.

For the proof, we consider the map 5.4 (2) $c: \coprod_v H^1(\mathfrak{g}, \Gamma_v) \to \{\pm 1\}$ defined by the product of Hasse-Witt invariants. A straightforward computation shows that the symmetric matrices M, $\mu \cdot M$ take the following values:

$$c_v(\mu \cdot M) = c_v(M) = 1, \quad \text{for } v \neq 2, m, \infty, \tag{1}$$

$$c_\infty(M) = 1, \quad c_\infty(\mu \cdot M) = -1, \tag{2}$$

$$c_2(M) = -1, \quad c_2(\mu \cdot M) = 1, \tag{3}$$

$$c_m(M) = c_m(\mu \cdot M) = -1. \tag{4}$$

There are 4 classes of elements in $\coprod_v H^1(\mathfrak{g}, \Gamma_v)$, and the above formulas completely determine the value of c on these four classes. The two elements x, y are the only elements in the kernel of c, and so by exactness they are the only elements in the image of h.

5.9. Let b_1, b_2 be as introduced in 5.8. Then, from the decomposition of fixpoint set F^σ, it follows that

$$\chi(F^\sigma)=\sum_{c\in h^{-1}h(b_1)}\chi(F(c))+\sum_{c\in h^{-1}h(b_2)}\chi(F(c)).$$

As described in 5.5, each of these two sums can be expressed as an infinite product, and we obtain

$$\chi(F^\sigma)=\frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))}\left(\frac{1}{\prod_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_1)}-\frac{1}{\prod_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_2)}\right). \tag{1}$$

Therefore, it remains to compute the local volume factor $\mathrm{vol}\,\Gamma_v(B_i)$ at each finite prime $v\neq\infty$, $i=1,2$.

For each finite prime v we denote as in [19], §3, by E resp. L the symmetric bilinear forms on $\mathbb{Z}_v^3$ given for $x=(x_j)$, $y=(y_j)\in\mathbb{Z}_v^3$ by $E(x,y)=\sum x_jy_j$ resp. $L(x,y)=-x_1y_1-x_2y_2+x_3y_3$. The symmetric matrices representing E and L with respect to the standard basis are denoted with the same letters; they denote also the cohomology classes represented by E, L considered as cocycles.

At a prime $v\neq 2$, m, the symmetric matrices $B_{i,v}$ $(i=1,2)$ are now Γ_v-congruent to the symmetric matrices L, E just defined. By the remark in the last paragraph of 5.6, we have

$$\mathrm{vol}\,\Gamma_v(B_1)=\mathrm{vol}\,\Gamma_v(B_2)=\mathrm{vol}\,\Gamma_v(L)=\mathrm{vol}\,\Gamma_v(E), \tag{2}$$

and this value is computed in 4.7 in [19].

At the prime $v=2$, the symmetric matrices B_1, B_2 have determinant different from $\det L=\det E=1$, and so they are not congruent to L or E. However, after multiplying B_1, B_2 by the scalar m, we have

$$\det(m\cdot B_i)\equiv m^3\det(B_i)\equiv m^4\equiv 1 \bmod 8,$$

$$m\cdot B_1\equiv\begin{pmatrix} m & & \\ & m & \\ & & m^2\end{pmatrix}\equiv\begin{pmatrix} -1 & & \\ & -1 & \\ & & 1\end{pmatrix}\bmod 4,$$

$$m\cdot B_2\equiv\begin{pmatrix} -m & & \\ & -m & \\ & & m^2\end{pmatrix}\equiv\begin{pmatrix} 1 & & \\ & 1 & \\ & & 1\end{pmatrix}\bmod 4.$$

Hence by our criterion on properly primitive forms, mB_1, mB_2 are congruent respectively to L and E,

$$mB_1\underset{(2)}{\sim}L,\qquad mB_2\underset{(2)}{\sim}E.$$

Thus we have

$$\mathrm{vol}(\Gamma_2(B_1)) = \mathrm{vol}(\Gamma_2(m \cdot B_1)) = \mathrm{vol}(\Gamma_2(L)), \tag{3}$$

$$\mathrm{vol}(\Gamma_2(B_2)) = \mathrm{vol}(\Gamma_2(m \cdot B_2)) = \mathrm{vol}(\Gamma_2(E)). \tag{4}$$

At the prime $v = m$, we choose the following basis for $so(M)(\mathbb{Z}_{(m)})$

$$U_{12} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad U_{23} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -m & 0 \end{pmatrix}, \quad U_{13} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -m & 0 & 0 \end{pmatrix}.$$

Then the elements $m^j U_{12}$, $m^j U_{23}$, $m^j U_{13}$ generate the congruence subspace $so(M)(j)$. In the first example in 5.6 we computed the inner products

$$e(U_{i,j}, U_{k,l}) = 0 \quad \text{for } (i,j) \neq (k,l), \qquad e(U_{12}, U_{12}) = 1,$$
$$e(U_{23}, U_{23}) = m, \qquad e(U_{13}, U_{13}) = m,$$

and by the integration formula in 5.6(2), we have

$$\mathrm{vol}(\Gamma_m(M)(j)) = m^{-3j-1}.$$

Since $\Gamma_m(M)(j)$ is a subgroup of index m^{3j-1} in $\Gamma_m(M)$, it follows that

$$\mathrm{vol}(\Gamma_m(M)) = m^{3j-1} \cdot \mathrm{vol}(\Gamma_m(M)(j)) = m^{-2}.$$

Comparing this with the local volumes $\mathrm{vol}(\Gamma_{(m)}(E))$ resp. $\mathrm{vol}(\Gamma_{(m)}(L))$ computed in 4.7 [19], there is an increase in our expression by a factor m,

$$\mathrm{vol}(\Gamma_m(B_i)) = m \,\mathrm{vol}(\Gamma_m(E)) = m \,\mathrm{vol}(\Gamma_m(L)). \tag{5}$$

Putting the formulas (1)–(5) together, we obtain for the Euler characteristic of F^σ

$$\begin{aligned} \chi(F^\sigma) &= \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))} \left(\frac{1}{\prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(B_1))} - \frac{1}{\prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(B_2))} \right) \\ &= \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))} \left(\frac{1}{m \cdot \prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(L))} - \frac{1}{m \cdot \prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(E))} \right). \end{aligned} \tag{6}$$

We compare this formula now with the final steps of Rohlfs in computing the Euler characteristic $\chi(F^\tau)$ of the fixpointset of the Cartan involution τ acting on $X/\Gamma(m)$. In 5.3 in [19] the formula

$$\chi(F^\tau) = \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))} \left(\frac{1}{\prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(E))} - \frac{1}{\prod_{v \neq \infty} \mathrm{vol}(\Gamma_v(L))} \right) \tag{7}$$

is established. Using the computations of the local factors one gets then as a final result

$$\begin{aligned}\chi(F^{\tau}) &= m(m^2-1)\cdot\zeta(-1) \\ &= -\frac{m(m^2-1)}{12}\end{aligned} \tag{8}$$

where one uses the fact that the Riemann zeta function $\zeta(s)$ takes the value $\zeta(-1)=-1/12$ at the point -1. Comparing (7) with formula (6) we see that $\chi(F^{\sigma})$ differs from $\chi(F^{\tau})$ by a factor $-\frac{1}{m}$; so we get

$$\chi(F^{\sigma}) = -\frac{1}{m}\chi(F^{\tau}) = \frac{m^2-1}{12}. \tag{9}$$

This completes the proof of Proposition 5.1.

§ 6. Fixpoint Sets of Other Involutions

6.1. In this section we apply Rohlfs' method to compute the Euler characteristic of five fixpoint sets $F^{\sigma\gamma}$, where $\gamma = s^2, t, ts^2, ts, ts^3$ in $\mathcal{O}$ (cf. § 1). By definition (cf. § 4) these five involutions are given explicitly in terms of matrices by

$$\sigma\gamma(x) = \gamma M\cdot\tau(x)\cdot(\gamma M)^{-1}$$

where γM is one of the following matrices.

$$s^2M = \begin{pmatrix} -1 & & \\ & -1 & \\ & & m \end{pmatrix}, \quad tM = \begin{pmatrix} 1 & & \\ & -1 & \\ & & -m \end{pmatrix}, \quad ts^2M = \begin{pmatrix} -1 & & \\ & 1 & \\ & & -m \end{pmatrix},$$

$$tsM = \begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & -m \end{pmatrix}, \quad ts^3M = \begin{pmatrix} 0 & -1 & \\ -1 & 0 & \\ & & -m \end{pmatrix}$$

and τ the Cartan involution. Since the case F^{σ} considered in § 5 contains all the essential features in the computations of Euler characteristics, we indicate in the following only briefly any necessary modifications in the analysis of the exact sequence analogue to 5.4(3) and the volume computations. The final result is

6.2. Proposition. *With the notation as defined above, we have*

(i) $\chi(F^{\sigma}) = \chi(F^{\sigma s^2}) = \dfrac{(m^2-1)}{12}$,

(ii) $\chi(F^{\sigma t}) = \chi(F^{\sigma ts^2}) = -\dfrac{(m^2-1)}{12}$,

(iii) $\chi(F^{\sigma ts}) = \chi(F^{\sigma ts^3}) = -\dfrac{(m+1)}{6}$.

Note that the six fixpoint sets are divided into three families: $\{F^{\sigma}, F^{\sigma s^2}\}$, $\{F^{\sigma t}, F^{\sigma t s^2}\}$, $\{F^{\sigma t s}, F^{\sigma t s^3}\}$. Two members of the same family differ only by the factor s^2.

Proof. In §5 we computed the Euler characteristic $\chi(F^{\sigma})$ in (i) above. To compute $\chi(F^{\sigma s^2})$, we observe that the cohomology set $H^1(\mathfrak{g}, \Gamma)$, $\mathfrak{g}=\{1, \sigma s^2\}$, $\Gamma=\Gamma_D(m)$ is the same as the Γ-congruence classes of symmetric matrices S such that

$$S \equiv \begin{pmatrix} s_{11} & & \\ & s_{22} & \\ & & 0 \end{pmatrix} \bmod m \quad (1), \qquad \varepsilon(s_{11})=\varepsilon(s_{22})=-1 \quad (2), \qquad \det S = m. \quad (3)$$

The difference between this and the situation in the previous section lies in the sign appearing in condition (2) (cf. 5.7). However, the computation of the cohomology set $H^1(\mathfrak{g}, \Gamma_v)$ at places v different from m remain unchanged, i.e. using the notation of the lemma in 5.7 we have

$$H^1(\mathfrak{g}, \Gamma_v)=\{M\}, \qquad v \neq 2, m, \infty, \tag{4}$$

$$H^1(\mathfrak{g}, \Gamma_{\infty})=\{M, \mu \cdot M\}, \tag{5}$$

$$H^1(\mathfrak{g}, \Gamma_{(2)})=\{M, \mu \cdot M\}. \tag{6}$$

At the prime m, the cohomology still consists of one element but with a different representative

$$H^1(\mathfrak{g}, \Gamma_{(m)})=\{\mu \cdot M\}. \tag{7}$$

It is not difficult to verify that the map h: $H^1(\mathfrak{g}, \Gamma) \to \coprod_v H^1(\mathfrak{g}, \Gamma_v)$ occuring in the analogue of the exact sequence 5.4(3), has again exactly two elements $h(\bar{b}_i)$, $i=1, 2$, in its image (cf. 5.8). The two cohomology classes $\bar{b}_1, \bar{b}_2$ can be represented by symmetric matrices $\bar{B}_1$, $\bar{B}_2$ with the following properties:

$$\det(\bar{B}_1)=\det(\bar{B}_2)=m, \quad (8) \qquad c_v(\bar{B}_1)=c_v(\bar{B}_2)=1, \qquad v \neq 2, m, \infty, \tag{9}$$

$$c_m(\bar{B}_1)=c_m(\bar{B}_2)=-1, \quad (10) \qquad c_2(\bar{B}_1)=1, \qquad c_2(\bar{B}_2)=-1, \tag{11}$$

$$c_{\infty}(\bar{B}_1)=-1, \qquad c_{\infty}(\bar{B}_2)=1. \tag{12}$$

Comparing $\bar{B}_1$, $\bar{B}_2$ with the symmetric matrices B_1, B_2 in 5.8, 5.9, it can be shown that at every place v the symmetric matrices $\bar{B}_1$, $\bar{B}_2$ are $SL_3(\mathbb{Z}_v)$-*congruent* to the corresponding matrices B_1, B_2. From the discussion of Tamagawa measure in 5.6, and formula 5.6(4) it follows that they give the same local volume factors at every finite place $v \neq m$

$$\operatorname{vol}(\Gamma_v(\bar{B}_i))=\operatorname{vol}(\Gamma_v(B_i)) \qquad i=1, 2.$$

At $v=m$, we let $U'_{12}=U_{12} \cdot \mu$, $U'_{13}=U_{13} \cdot \mu$, $U'_{23}=U_{23} \cdot \mu$ where U_{ij} is defined as in 5.9. Then $\{U'_{i,j}\}$ forms a $\mathbb{Z}_{(m)}$-basis for the space $SO(\bar{B}_i)(\mathbb{Z}_m)$, and a direct computation shows that

$$\|\det(e(U'_{i,j}, U'_{k,l}))_{1 \leqq i<j \leqq 3,\, 1 \leqq k<l \leqq 3}\| = \|\det(e(U_{i,j}, U_{k,l}))_{1 \leqq i<j \leqq 3,\, 1 \leqq k<l \leqq 3}\|.$$

It follows that the local volumes for a deep enough congruence subgroup are the same

$$\mathrm{vol}(\Gamma_m(\bar{B}_i)(j)) = \mathrm{vol}(\Gamma_m(B_i)(j)) \qquad i=1,2;\ j \gg 0.$$

Since for the index $[\Gamma_m(\bar{B}_i)\colon \Gamma_m(\bar{B}_i)(j)] = [\Gamma_m(B_i)\colon \Gamma_m(B_i)(j)]$, the above equation remains valid for $\Gamma_m(\bar{B}_i)$ and $\Gamma_m(B_i)$, i.e. we get

$$\mathrm{vol}(\Gamma_m(\bar{B}_i)) = \mathrm{vol}(\Gamma_m(B_i)).$$

Hence we have

$$\chi(F^{\sigma s^2}) = \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))}\left[\frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(\bar{B}_1)} - \frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(\bar{B}_2)}\right]$$

$$= \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))}\left[\frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_1)} - \frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_2)}\right] = \chi(F^{\sigma}).$$

This proves formula (i) in Proposition 6.2.

6.3. Proof of 6.2 (ii). Next we come to the involution σt and its fixpoint set $F^{\sigma t}$. The cohomology $H^1(\mathfrak{g}, \Gamma_D(m))$, $\mathfrak{g}=\{1, \sigma t\}$, is the same as $\Gamma_D(m)$-congruence classes of symmetric matrices S such that

$$S \equiv \begin{pmatrix} s_{11} & & \\ & s_{22} & \\ & & 0 \end{pmatrix} \bmod m, \quad (1) \qquad \varepsilon(s_{11})=1,\ \varepsilon(s_{22})=-1, \quad (2) \qquad \det S = m. \quad (3)$$

We proceed in the same manner as before: the computation of the local cohomology sets $H^1(\mathfrak{g}, \Gamma_v)$ gives the same results as in 6.2. The image of $h\colon H^1(\mathfrak{g}, \Gamma) \to \coprod H^1(\mathfrak{g}, \Gamma_v)$ consists of two elements $h(D_1)$, $h(D_2)$; the representatives D_1, D_2 have the following properties which differ from the properties 6.2(8)–(12) in the previous consideration:

$$\det(D_1) = \det(D_2) = m, \quad (4) \qquad c_v(D_1) = c_v(D_2) = 1, \qquad v \neq 2, \infty, m, \qquad (5)$$

$$c_m(D_1) = c_m(D_2) = 1, \quad (6) \qquad c_2(D_1) = 1, \quad c_2(D_2) = -1, \qquad (7)$$

$$c_\infty(D_1) = 1, \quad c_\infty(D_2) = -1. \qquad (8)$$

Condition (5) follows from the well known fact that an integral symmetric matrix with unit determinant in $\mathbb{Z}_v(\det \not\equiv 0 \bmod v)$ has Hasse-Witt invariant one. From conditions (1), (2), (3), it is easy to calculate the Hasse-Witt invariant directly from a formula for ternary forms (cf. [13] p. 35), and the answer appears as in (6). Since the product $\prod\limits_v c_v = 1$, the Hasse-Witt invariants at prime 2 and at the infinite prime must have the sign as in (7) and (8).

For $v \neq m$, it is easy to see that D_i and B_i are Γ_v-congruent to each other, and so they contribute the same volume factor $\mathrm{vol}(\Gamma_v(D_i)) = \mathrm{vol}(\Gamma_v(B_i))$, $i=1,2$, by 5.6(4).

At $v=m$, we let $U''_{12} = U_{12}\cdot t$, $U''_{13} = U_{13}\cdot t$, $U''_{23} = U_{23}\cdot t$ where $U_{i,j}$ is defined as in 5.9. Then $U''_{i,j}$ span the space $SO(D_i)(\mathbb{Z}_m)$, and a direct computation shows

$$\|\det(e(U''_{i,j}, U''_{k,l}))_{1\leq i<j\leq 3,\, 1\leq k<l\leq 3}\| = \|\det(e(U_{i,j}, U_{k,l}))_{1\leq i<j\leq 3,\, 1\leq k<l\leq 3}\|.$$

It follows that the local volume factor remains the same at prime m, $\mathrm{vol}(\Gamma_m(D_i)) = \mathrm{vol}(\Gamma_m(B_i))$. However, the forms have different signatures, and so they appear with different signs in the Euler characteristic formula. Thus

$$\chi(F^{\sigma t}) = \frac{2}{\mathrm{vol}(SO_3(\mathbb{R}))}\left[\frac{-1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(D_1)} + \frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(D_2)}\right]$$

$$= \frac{-2}{\mathrm{vol}(SO_3(\mathbb{R}))}\left[\frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_1)} - \frac{1}{\prod\limits_{v\neq\infty}\mathrm{vol}\,\Gamma_v(B_2)}\right] = -\chi(F^{\sigma}).$$

For the involution $\sigma t s^2$, its relation with $F^{\sigma t}$ is similar to that between F^{σ} and $F^{\sigma s^2}$ in 6.2. Carrying through the same arguement as in the previous section, we have

$$\chi(F^{\sigma t}) = \chi(F^{\sigma t s^2}) = -\frac{(m^2-1)}{12}.$$

6.4. This brings us to the last family $F^{\sigma t s}$ and $F^{\sigma t s^3}$. Again we treat only the case $F^{\sigma t s}$ because the other case $F^{\sigma t s^3}$ is identical. For the involution $\sigma t s$, the cohomology $H^1(\mathfrak{g}, \Gamma_D(m))$, $\mathfrak{g} = \{1, \sigma t s\}$ is the same as $\Gamma_D(m)$-congruence classes of integral, 3-by-3, symmetric matrices S such that

$$S \equiv \begin{pmatrix} 0 & s_{12} & \\ s_{12} & 0 & \\ & & 0 \end{pmatrix} \bmod m, \quad (1) \qquad \varepsilon(s_{12}) = 1, \quad (2) \qquad \det S = m. \quad (3)$$

Note that the above conditions imply for the Hasse-Witt invariant $c_m(S) = 1$. If we proceed as before, then there are two symmetric matrices G_1, G_2 such that the image of h: $H^1(\mathfrak{g}, \Gamma) \to \coprod H^1(\mathfrak{g}, \Gamma_v)$ consists of two elements $h(G_1)$, $h(G_2)$ and the local Hasse-Witt invariants of G_1, G_2 are the same as the ones of D_1, D_2 in 6.3. At places v different from m, $v \neq m$, the local volume factors remain unchanged; i.e. at the prime m, the symmetric matrices G_1, G_2 are congruent to the matrix $G = \begin{pmatrix} 0 & 1 & \\ 1 & 0 & \\ & & m \end{pmatrix}$. We put

$$Y_{12} = U_{12} \cdot st = \begin{pmatrix} 1 & & \\ & -1 & \\ & & 0 \end{pmatrix}, \qquad Y_{13} = U_{13} \cdot st = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & -m & 0 \end{pmatrix},$$

$$Y_{23} = U_{23} \cdot st = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ -m & 0 & 0 \end{pmatrix}.$$

We multiply these elements by m^j and consider the congruence subspace generated by these vectors $m^j \cdot Y_{12}$, $m^j Y_{13}$, $m^j Y_{23}$. The integration formula in 5.6(2) gives us the value

$$\mathrm{vol}(\Gamma_m(G)(j)) = m^{-3j-1}.$$

On the other hand, $\Gamma_m(G)(j)$ is a normal subgroup in $\Gamma_m(G)$ of index

$$[\Gamma_m(G): \Gamma_m(G)(j)] = \frac{(m-1)}{2} m^{3j-1}.$$

Comparing this with the previous situation (cf. 5.9 resp. 6.3), there is an increase by a factor $\left(\frac{m-1}{2}\right)$. The reason for this increase is because the $\left(\frac{m-1}{2}\right)$ diagonal matrices

$$\begin{pmatrix} a_{11} & & \\ & a_{11}^{-1} & \\ & & 1 \end{pmatrix}, \quad \varepsilon(a_{11}) = 1$$

are contained in $\Gamma_m(G)$ but not in $\Gamma_m(D)$. This accounts for the increase of volume by $\frac{m-1}{2}$, and a decrease in Euler characteristic by the same factor, so we have

$$\chi(F^{\sigma ts}) = \chi(F^{\sigma t}) \Big/ \frac{m-1}{2} = -\frac{m+1}{6}.$$

Since $\chi(F^{\sigma ts^3}) = \chi(F^{\sigma ts})$, this proves formula (iii) of 6.2 and completes the proof of our proposition.

§ 7. Order Four Elements

7.1. We now investigate the fixpoint sets $F^{\sigma s}$, $F^{\sigma s^3}$ of order 4 elements σs, σs^3. Since the square of σs, σs^3 is the same as the action by the element μ, $(\sigma s)^2 = (\sigma s^3)^2 = s^2 = \mu$, the fixpoint sets of σs, σs^3 must lie in the fixpoint set F^μ of the involution μ induced by μ. (For μ we refer to the lemma in 5.7.) In fact, if we consider the restriction of σs, σs^3 over F^μ, then they become order two and the fixpoint sets of these actions $\sigma s | F^\mu$, $\sigma s^3 | F^\mu$ coincide with $F^{\sigma s}$, $F^{\sigma s^3}$

$$F^{\sigma s} = \mathrm{Fix}(\sigma s | F^\mu, F^\mu) \quad (1) \qquad \text{resp.} \quad F^{\sigma s^3} = \mathrm{Fix}(\sigma s^3 | F^\mu, F^\mu). \quad (2)$$

On the other hand, since σs and σs^3 differ only by the factor μ, $\sigma s^3 = (\sigma s) \cdot \mu$, the restriction of σs, σs^3 over F^μ are the same and so they have the same fixpoint set $F^{\sigma s}$, $F^{\sigma s^3}$,

$$F^{\sigma s} = F^{\sigma s^3}. \quad (3)$$

Therefore it is enough to study only the case $F^{\sigma s}$.

This is achieved in three steps. First in this section we prove that F^μ consists of two connected components $F(\gamma_0)$, $F(\gamma_1)$ each of which is invariant under σs. Then in the next section we study the action σs on the first component $F(\gamma_0)$, and finally in the last section we study the action σs on the remaining fixpoint component $F(\gamma_1)$.

7.2. There is a major difference between the involution μ and those involutions τ, σ, ... studied in the previous sections. In the present situation, the involution μ is an inner automorphism of $SL_3(\mathbb{Z})$ whereas before the involutions are outer automorphisms. Nevertheless, the connected components of F^μ are in one-to-one correspondence with the non-abelian cohomology set $H^1(\mathfrak{g}, \Gamma_D(m))$, $\mathfrak{g}=\{1,\mu\}$ (see 5.2).

To study this cohomology $H^1(\mathfrak{g}, \Gamma_D(m))$, $\mathfrak{g}=\{1,\mu\}$, we consider the cohomology $H^1(\mathfrak{g}, SL_3(\mathbb{Z}))$ of $\mathfrak{g}$ with coefficients in the entire special linear group $SL_3(\mathbb{Z})$. A cocycle in the cohomology set $H^1(\mathfrak{g}, SL_3(\mathbb{Z}))$ is by definition an element δ in $SL_3(\mathbb{Z})$ such that

$$\delta\cdot\mu\cdot\delta\cdot\mu^{-1}=1 \quad \text{or equivalently } (\delta\mu)^2=1. \tag{1}$$

Moreover, two cocycles δ and $\bar{\delta}$ represent the same cohomology class if and only if there exists an element β in $SL_3(\mathbb{Z})$ such that

$$\beta\cdot\delta\cdot\mu\beta^{-1}\mu^{-1}=\bar{\delta} \quad \text{or equivalently } \beta(\delta\mu)\beta^{-1}=(\bar{\delta}\mu). \tag{2}$$

If we let $\gamma=\delta\mu$, $\bar{\gamma}=\bar{\delta}\mu$, then condition (1) means they are order 2 elements in $SL_3(\mathbb{Z})$ and condition (2) means δ, $\bar{\delta}$ are conjugate to each other. In other words, the cohomology set $H^1(\mathfrak{g}, SL_3(\mathbb{Z}))$ is the same as the set of conjugacy classes of order two elements $\gamma=\delta\mu$ in $SL_3(\mathbb{Z})$.

7.3. Proposition. *The cohomology set $H^1(\mathfrak{g}, SL_3(\mathbb{Z}))$, $\mathfrak{g}=\{1,\mu\}$, consists of two elements. Using the identification as explained above, these two elements correspond to the following conjugacy classes of order two elements in $SL_3(\mathbb{Z})$, which are represented by*

$$\gamma_0=\begin{pmatrix}-1 & & \\ & -1 & \\ & & 1\end{pmatrix}, \quad \gamma_1=\begin{pmatrix}-1 & & \\ & -1 & 1\\ & & 1\end{pmatrix}.$$

The proof follows immediately from the classification of conjugacy classes of order 2 elements in $SL_3(\mathbb{Z})$ (see [27, 28]).

We now turn to the cohomology $H^1(\mathfrak{g}, \Gamma_D(m))$ of $\mathfrak{g}=\{1,\mu\}$ with coefficients in $\Gamma_D(m)$. From conditions 7.2(1) and (2), it follows that there is a one-to-one correspondence between $H^1(\mathfrak{g}, \Gamma_D(m))$ and the set of $\Gamma_D(m)$-conjugacy classes of order two elements of the form $\delta\cdot\mu$, $\delta\in\Gamma_D(m)$. Let $\langle\Gamma_D(m),\mu\rangle$ be the subgroup in $SL_3(\mathbb{Z})$ generated by the elements in $\Gamma_D(m)$ and μ. Note that this is a semi-direct product of $\Gamma_D(m)$ and the cyclic group $\langle\mu\rangle$ of order two,

$$\langle\Gamma_D(m),\mu\rangle\cong\Gamma_D(m)\ltimes\mathbb{Z}/2.$$

Since $\Gamma_D(m)$ has no torsion elements by 1.2, an element of order two in $\langle\Gamma_D(m),\mu\rangle$ must be of the form $\delta\mu$ for some δ in $\Gamma_D(m)$. Thus $H^1(\mathfrak{g}, \Gamma_D(m))$ is the same as the set of $\Gamma_D(m)$-conjugacy classes of order two elements in $\langle\Gamma_D(m),\mu\rangle$.

7.4. Proposition. *The cohomology $H^1(\mathfrak{g}, \Gamma_D(m))$, $\mathfrak{g}=\{1,\mu\}$, consists of two elements. Using the identification as explained above, these two elements correspond*

to the following conjugacy classes of order two elements

$$\gamma_0=\begin{pmatrix}-1 & & \\ & -1 & \\ & & 1\end{pmatrix}, \quad \gamma_1=\begin{pmatrix}-1 & & \\ & -1 & 1\\ & & 1\end{pmatrix}$$

in $\langle \Gamma_D(m), \mu\rangle$.

Proof. Consider the fibration:

$$1\to\Gamma_D(m)\to SL_3(\mathbb{Z})\to SL_3(\mathbb{Z})/\Gamma_D(m), \tag{1}$$

and the associated exact sequence of pointed cohomology sets (cf. [24], Prop. 36)

$$\begin{aligned}1\to H^0(\mathfrak{g}, \Gamma_D(m))\to H^0(\mathfrak{g}, SL_3(\mathbb{Z}))\to H^0(\mathfrak{g}, SL_3(\mathbb{Z})/\Gamma_D(m))\\ \to H^1(\mathfrak{g}, \Gamma_D(m))\xrightarrow{f} H^1(\mathfrak{g}, SL_3(\mathbb{Z})).\end{aligned} \tag{2}$$

Here f is induced by inclusion, $H^0(\mathfrak{g}, SL_3(\mathbb{Z})/\Gamma_D(m))$ and $H^0(\mathfrak{g}, SL_3(\mathbb{Z}))$ denote respectively the fixpoint set of μ in the space $SL_3(\mathbb{Z})/\Gamma_D(m)$ and in the group $SL_3(\mathbb{Z})$. From Proposition 7.3 the map f is surjective because the elements γ_0, γ_1 in 7.3 are elements in $\langle \Gamma_D(m), \mu\rangle$. The inverse image $f^{-1}f(\gamma_0)$ can be identified with the quotient space ([24], Cor. 1 to Prop. 36)

$$SL_3(\mathbb{Z})^\mu\backslash\{SL_3(\mathbb{Z})/\Gamma_D(m)\}^\mu.$$

Here $\{SL_3(\mathbb{Z})/\Gamma_D(m)\}^\mu$ is the subspace in the right coset space $SL_3(\mathbb{Z})/\Gamma_D(m)$ fixed by μ, the group $SL_3(\mathbb{Z})^\mu$ is the subgroup in $SL_3(\mathbb{Z})$ fixed by μ, and the symbol $SL_3(\mathbb{Z})^\mu\backslash\{SL_3(\mathbb{Z})/\Gamma_D(m)\}^\mu$ denotes the orbit space of $\{SL_3(\mathbb{Z})/\Gamma_D(m)\}^\mu$ under the left action of $SL_3(\mathbb{Z})^\mu$. There are natural isomorphisms (cf. § 1)

$$SL_3(\mathbb{Z})/\Gamma_D(m)\simeq SL_3(\mathbb{Z}/m)/D, \tag{3}$$

$$SL_3(\mathbb{Z})^\mu=\left\{\begin{pmatrix}a_{11} & a_{12} & a_{13}\\ a_{21} & a_{22} & a_{23}\\ a_{31} & a_{32} & a_{33}\end{pmatrix}\in SL_3(\mathbb{Z})\,\middle|\, a_{13}=a_{23}=a_{31}=a_{32}=0\right\}. \tag{4}$$

$$SL_3(\mathbb{Z}/m)^\mu=\left\{\begin{pmatrix}\bar a_{11} & \bar a_{12} & \bar a_{13}\\ \bar a_{21} & \bar a_{22} & \bar a_{23}\\ \bar a_{31} & \bar a_{32} & \bar a_{33}\end{pmatrix}\in SL_3(\mathbb{Z}/m)\,\middle|\, \bar a_{13}=\bar a_{23}=\bar a_{31}=\bar a_{32}=0\right\}. \tag{5}$$

It is not difficult to verify that if an element x in $SL_3(\mathbb{Z}/m)$ projects onto an element in the fixpoint set $\{SL_3(\mathbb{Z}/m)/D\}^\mu$, then x is an element in $SL_3(\mathbb{Z}/m)^\mu$. Since the modulo m reduction map

$$SL_3(Z)^\mu\to SL_3(\mathbb{Z}/m)^\mu$$

is onto, the quotient space $SL_3(\mathbb{Z})^\mu\backslash\{SL_3(\mathbb{Z})/\Gamma_D(m)\}^\mu$ has only one element.

The same argument can be carried out to show that the inverse image $f^{-1}f(\gamma_1)$ has only one element. Since this is very similar to the discussion above, we will not repeat the details in here. The proof of 7.4 follows.

7.5. Proposition. *The fixpoint set F^{μ} of μ in $X/\Gamma_D(m)$ consists of two connected components $F(\gamma_0)$ resp. $F(\gamma_1)$, $\gamma_0, \gamma_1 \in H^1(\mathfrak{g}, \Gamma_D(m))$, which are stable under the action of σ s.*

If $X(\gamma_i)$, $i=0,1$, denotes the fixpoint submanifold in X under the action of γ_i and $\Gamma_D(m)^{\gamma_i}$ the centralizer of γ_i in $\Gamma_D(m)$, then there are natural isomorphisms

$$F(\gamma_0)=X(\gamma_0)/\Gamma_D(m)^{\gamma_0}, \tag{1}$$

$$F(\gamma_1)=X(\gamma_1)/\Gamma_D(m)^{\gamma_1}. \tag{2}$$

The first statement follows immediately from 7.2 and the determination of $H^1(\mathfrak{g}, \Gamma_D(m))$ in 7.4.

For an order two element γ in $\langle \Gamma_D(m), \mu \rangle$, we let $X(\gamma)$ be the fixpoint submanifold in X under the action of γ. These submanifolds decompose into two families according to the $\Gamma_D(m)$-conjugacy classes $\{\gamma_0\}$, $\{\gamma_1\}$. Define X_0, X_1 to be the unions of these two families of submanifolds,

$$X_i=\bigcup_{\gamma \sim \gamma_i} X(\gamma)=\bigcup_{\beta \in \Gamma_D(m)} X(\beta^{-1}\gamma_i\beta), \qquad i=0,1. \tag{1}$$

For an element δ in $\Gamma_D(m)$, it is easy to see that the action of δ maps the submanifold $X(\gamma)$ into $X(\delta^{-1}\gamma\delta)$. Hence the group $\Gamma_D(m)$ operates on these two unions X_0, X_1 of spaces. We claim their orbit spaces coincide respectively with the fixpoint component $F(\gamma_0)$, $F(\gamma_1)$. Let x be an element of X which projects onto a fixpoint of μ in $X/\Gamma_D(m)$. Then we must have

$$x \cdot \mu = x \cdot \delta \qquad \text{or equivalently} \quad x = x \cdot \delta\mu$$

for some δ in $\Gamma_D(m)$. It follows that $\delta\mu$ is an element of finite order in $\langle \Gamma_D(m), \mu \rangle$ and x belongs to the component $X(\delta\mu)$. From the discussion in 7.2, it is clear that $\delta\mu$ is of order two, and so $X(\delta\mu)$ lies in one of the unions X_0 or X_1.

Next the subspaces $X(\gamma)$, where γ in $\langle \Gamma_D(m), \mu \rangle$ is an element of order two, are disjoint from each other. For if x were in the intersection of $X(\bar{\gamma}_2)$, $X(\gamma_2)$, $x \in X(\bar{\gamma}_2) \cap X(\gamma_2)$, then x would be fixed by the element $\bar{\gamma}_2\gamma_2^{-1}$. Since the last element lies in $\Gamma_D(m)$, this would contradict the fact that $\Gamma_D(m)$ is torsion free.

We can now write the sums in (1) as the disjoint unions:

$$X_i=\coprod_{\beta \in \Gamma_D(m)} X(\beta^{-1}\gamma_i\beta), \qquad i=0,1. \tag{2}$$

As indicated before, the action of an element δ, $\delta \in \Gamma_D(m)$, brings the fixpoint set $X(\gamma)$ to $X(\delta^{-1}\gamma\delta)$. Hence the action of $\Gamma_D(m)$ is transitive on the members of this disjoint union in X_0, X_1. If we fix one component $X(\gamma)$, then the element δ in $\Gamma_D(m)$ which keeps $X(\gamma)$ stable must satisfy the equation:

$$\delta^{-1}\gamma\delta=\gamma.$$

Hence it belongs to the centralizer of γ. This leads us to the equation:

$$X_i/\Gamma_D(m)=\coprod_{\beta \in \Gamma_D(m)} X(\beta^{-1}\gamma_i\beta)/\Gamma_D(m)=X(\gamma_i)/\Gamma_D(m)^{\gamma_i}, \qquad i=0,1.$$

This proves the proposition.

§ 8. Fixpoint Set $F(\gamma_0)$

8.1. In this section, we give a description of $F(\gamma_0)$ as a product of an arithmetic quotient of the upper half plane modulo a congruence subgroup $\Gamma_D(2,m)$ of $SL_2(\mathbb{Z})$ and $\mathbb{R}_+$. By means of this description, the Euler characteristic of the fixpoint set $\mathrm{Fix}(\sigma s, F(\gamma_0))$ is easily determined in Proposition 8.4.

Throughout the rest of this paper, we identify X with the symmetric space $SO(M)(\mathbb{R})\backslash SL_3(\mathbb{R})$. An element in X can be thought of as a positive-definite, symmetric, 3-by-3 matrix A with determinant m i.e.

$$A=\begin{pmatrix} s_{11} & s_{12} & s_{13} \\ s_{12} & s_{22} & s_{23} \\ s_{13} & s_{23} & s_{33} \end{pmatrix}, \quad (1) \qquad \begin{array}{l} s_{11}>0,\ \Delta=s_{11}s_{22}-s_{12}^2>0, \\ \det A=m. \end{array} \quad (2)$$

The group $SL_3(\mathbb{R})$ operates transitively on X by $A\mapsto {}^t\gamma\cdot A\cdot\gamma$, and the isotropy subgroup at the point M is the special orthogonal group $SO(M)(\mathbb{R})$. Note that if A is an element as in (1), then under the action of γ_0 it becomes

$${}^t\gamma_0\cdot A\cdot\gamma_0=\begin{pmatrix} s_{11} & s_{12} & -s_{13} \\ s_{12} & s_{22} & -s_{23} \\ -s_{13} & -s_{23} & s_{33} \end{pmatrix}.$$

Thus the fixpoint set $X(\gamma_0)$ coincides with the subspace in X consisting of positive definite, symmetric matrices A with the entries $s_{13}=s_{23}=0$, i.e.

$$A=\begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{12} & s_{22} & 0 \\ 0 & 0 & s_{33} \end{pmatrix}, \quad (3) \qquad \begin{array}{l} s_{11}>0,\ \Delta=\begin{vmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{vmatrix}>0, \\ \Delta s_{33}=m. \end{array} \quad (4)$$

As for the group $\Gamma_D(m)^{\gamma_0}$, it lies in $SL_3(\mathbb{Z})^\mu$ as a subgroup, and by 7.4(4) it consists of the following type of matrices:

$$\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & \pm 1 \end{pmatrix}.$$

However, because of the conditions

$$\varepsilon(a_{11})=\varepsilon(a_{22})=1, \qquad a_{12}\equiv a_{21}\equiv 0 \bmod m$$

on $\Gamma_D(m)$, the entry a_{33} at the lower corner must be 1 and so

$$\Gamma_D(m)^{\gamma_0}=\left\{\begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ 0 & 0 & 1 \end{pmatrix}\middle|\begin{array}{l}\varepsilon(a_{11})=\varepsilon(a_{22})=1, \\ a_{12}\equiv a_{21}\equiv 0 \bmod m.\end{array}\right\}. \quad (5)$$

We denote by $\Gamma_D(2,m)$ the congruence subgroup in $SL_2(\mathbb{Z})$ consisting of matrices:

$$\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \qquad \varepsilon(a_{11})=\varepsilon(a_{22})=1, \qquad a_{12}\equiv a_{21}\equiv 0 \bmod m.$$

Then from the above formula (5) it is clear that the group $\Gamma_D(m)^{\gamma_0}$ is isomorphic to this congruence subgroup $\Gamma_D(2, m)$ in $SL_2(\mathbb{Z})$.

8.2. Let Y be the symmetric space $SO_2(\mathbb{R})\backslash SL_2(\mathbb{R})$. Then there is a natural right action of $\Gamma_D(2, m)$ on Y. This quotient space Y can be identified with the space of positive definite, symmetric, 2-by-2 matrices R of determinant 1,

$$R = \begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}, \quad s_{11} > 0, \quad \det R = 1. \tag{1}$$

There is a diffeomorphism of $X(\gamma_0)$ onto the product space $Y \times \mathbb{R}_+$ defined by the equation:

$$X(\gamma_0) \xrightarrow{\simeq} Y \times \mathbb{R}_+, \quad \begin{pmatrix} & R & & 0 \\ & & & 0 \\ 0 & & 0 & s_{33} \end{pmatrix} \mapsto \left(\frac{R}{\sqrt{\Delta}}, s_{33} \right). \tag{2}$$

In fact, this is an equivariant diffeomorphism with respect to the action of $\Gamma_D(2, m)$ on the first factor Y in $Y \times \mathbb{R}_+$ and with respect to the action of $\Gamma_D(m)^{\gamma_0}$ on the space X. From Proposition 7.5 it follows that there exists a diffeomorphism of the fixpoint set $F(\gamma_0)$ and the product manifold $Y/\Gamma_D(2, m) \times \mathbb{R}_+$

$$F(\gamma_0) \cong X(\gamma_0)/\Gamma_D(m)^{\gamma_0} \cong Y/\Gamma_D(2, m) \times \mathbb{R}_+. \tag{3}$$

8.3. We now investigate the action of σs on the latter space. On the symmetric space X, the involution σ is given by the formula:

$$\sigma(A) = \begin{pmatrix} 1 & & \\ & 1 & \\ & & m \end{pmatrix} \cdot A^{-1} \cdot \begin{pmatrix} 1 & & \\ & 1 & \\ & & m^{-1} \end{pmatrix}. \tag{1}$$

Let A be a point in $X(\gamma_0)$, given as

$$A = \begin{pmatrix} s_{11} & s_{12} & 0 \\ s_{12} & s_{22} & 0 \\ 0 & 0 & s_{33} \end{pmatrix}.$$

Then under the action of σs it becomes

$$\sigma s(A) = \begin{pmatrix} & -1 & \\ 1 & & \\ & & m \end{pmatrix} \begin{pmatrix} \frac{s_{22}}{\Delta} & \frac{-s_{12}}{\Delta} & \\ \frac{-s_{12}}{\Delta} & \frac{s_{11}}{\Delta} & \\ & & \frac{1}{s_{33}} \end{pmatrix} \begin{pmatrix} & -1 & \\ 1 & & \\ & & m \end{pmatrix}$$

$$= \begin{pmatrix} \frac{s_{11}}{\Delta} & \frac{s_{12}}{\Delta} & \\ \frac{s_{12}}{\Delta} & \frac{s_{22}}{\Delta} & \\ & & \frac{m^2}{s_{33}} \end{pmatrix}.$$

Since $\sigma s(A)$ is again a point in $X(\gamma_0)$, the action of σs keeps the submanifold $X(\gamma_0)$ stable. If we identify $X(\gamma_0)$ with $Y \times \mathbb{R}_+$ by the diffeomorphism in 8.2(2), then σs induces the following action on $Y/\Gamma_D(2,m) \times R_+$:

$$\left(\frac{R}{\sqrt{\Delta}}, s_{33}\right) \mapsto \left(\frac{R}{\sqrt{\Delta}}, \frac{m^2}{s_{33}}\right). \tag{2}$$

In order for a point in $Y/\Gamma_D(2,m) \times \mathbb{R}_+$ to be a fixpoint, a necessary and sufficient condition is that it lies in the submanifold $Y/\Gamma_D(2,m) \times \{m\}$ where $s_{33} = m$.

8.4. Proposition. *With the notation as above, the fixpoint set of σs on $F(\gamma_0)$ is diffeomorphic to the surface $Y/\Gamma_D(2,m)$. The Euler characteristic of this fixpoint set is given by*

$$\chi(Y/\Gamma_D(2,m)) = -\frac{m(m+1)}{6}.$$

The first statement follows immediately from the discussion above. To compute the Euler characteristic, we note that we have

$$\chi(\Gamma_D(2,m)) = [SL_2(\mathbb{Z}):\Gamma_D(2,m)] \cdot \chi(SL_2(\mathbb{Z})) = -\frac{[SL_2(\mathbb{Z}):\Gamma_D(2,m)]}{12}$$
$$= -\frac{m(m+1)}{6}$$

where we use $\chi(SL_2(\mathbb{Z})) = -\frac{1}{12}$ (cf. [23], p. 100) and the notion of virtual Euler characteristic.

§ 9. Fixpoint Set $F(\gamma_1)$

9.1. We now turn to the other fixpoint component $F(\gamma_1)$ defined in Sect. 7.5. Recall that the universal covering space is the manifold $X(\gamma_1)$ with the covering transformation group $\Gamma_D(m)^{\gamma_1}$. Although γ_0, γ_1 represent two distinct conjugacy classes in $SL_3(\mathbb{Z})$, they are conjugate in $SL_3(\mathbb{Q})$. Let β be the matrix

$$\begin{pmatrix} -1 & & \\ & -1 & \frac{1}{2} \\ & & 1 \end{pmatrix}$$

in $SL_3(\mathbb{Q})$. Then conjugation by β brings γ_0 to γ_1, $\beta^{-1}\gamma_0\beta=\gamma_1$. Hence, from the discussion in § 7, the action of β on X translates the component $X(\gamma_0)$ to $X(\gamma_1)$. Conjugation by β brings $\Gamma_D(m)^{\gamma_0}$ into a subgroup in $SL_3(\mathbb{Q})^{\gamma_1}$ which is bigger than $\Gamma_D(m)^{\gamma_1}$. For given an element g in $\Gamma_D(m)^{\gamma_0}$,

$$g=\begin{pmatrix} a_{11} & a_{12} & \\ a_{21} & a_{22} & \\ & & 1 \end{pmatrix},$$

the conjugation by β brings this to the matrix $\beta^{-1}g\beta$

$$\beta^{-1}g\beta=\begin{pmatrix} a_{11} & a_{12} & \dfrac{-a_{12}}{2} \\ a_{21} & a_{22} & \dfrac{1-a_{22}}{2} \\ & & 1 \end{pmatrix}.$$

In order for $\beta^{-1}g\beta$ to be in $\Gamma_D(m)^{\gamma_1}$, a necessary and sufficient condition is that

$$a_{12}\equiv 0 \bmod 2, \quad (1) \qquad a_{22}\equiv 1 \bmod 2. \qquad (2)$$

However, because of the condition on determinant, the second condition (2) is an immediate consequence of the first condition (1).

9.2. Proposition. *Let $\Gamma_D^0(2,m)$ be the congruence subgroup in $SL_2(\mathbb{Z})$ defined by:*

$$\Gamma_D^0(2,m)=\left\{\begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}\in SL_2(\mathbb{Z})\,\middle|\,\begin{array}{l} a_{12}\equiv 0 \bmod 2 \\ a_{12}\equiv a_{21}\equiv 0 \bmod m \\ \varepsilon(a_{11})=\varepsilon(a_{22})=1 \end{array}\right\}.$$

Then $\Gamma_D^0(2,m)$ is a subgroup in $\Gamma_D(2,m)$, and β induces a natural isomorphism of $\Gamma_D^0(2,m)$ onto $\Gamma_D(m)^{\gamma_1}$.

The proof of the above proposition is straightforward, and we will omit it in here.

9.3. Recall that $X(\gamma_0)$ is diffeomorphic to the product space $Y\times\mathbb{R}_+$. Since $X(\gamma_1)$ is diffeomorphic to $X(\gamma_0)$ under the action of β, there is an equivariant diffeomorphism of $X(\gamma_1)$ onto $Y\times\mathbb{R}_+$ with respect to the action of $\Gamma_D(m)^{\gamma_1}$ on $X(\gamma_1)$ and the action of $\Gamma_D^0(2,m)$ on $Y\times\mathbb{R}_+$. We have

Proposition. *The fixpoint component $F(\gamma_1)$ is diffeomorphic to the product manifold*

$$Y/\Gamma_D^0(2,m)\times\mathbb{R}_+.$$

9.4. To describe the action of σs on $F(\gamma_1)$, we bring into play our initial assumption $m\equiv 3 \bmod 8$ (cf. 1.1 (2)).

Lemma. *There exists a matrix* α,

$$\alpha = \begin{pmatrix} a & b & 0 \\ c & d & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

in $\Gamma_D(m)$ *with the following properties:*

$$a^2 \equiv -2 \bmod m, \tag{1}$$

$$b \equiv c \equiv 0 \bmod m, \quad \varepsilon(a) = \varepsilon(d) = 1, \tag{2}$$

$$a \equiv d \equiv 0 \bmod 2. \tag{3}$$

First under our assumption $m \equiv 3 \bmod 8$, the integer -2 is a square modulo m. By the same argument there exists also an integer a' such that $(a')^2 \equiv -2 \bmod m$, $\varepsilon(a') = 1$. If we modify a' appropriately then this element satisfies our requirements on a.

Since a is a unit modulo m, there exists an element b, $b \equiv 0 \bmod m$ such that a and b are relatively prime. By Euclidean algorithm, there exist integers c, d such that $a\,d - b\,c = 1$, or in other words

$$\begin{pmatrix} a & b & \\ c & d & \\ & & 1 \end{pmatrix}$$

is in $SL_3(\mathbb{Z})$. Since $b \equiv 0 \bmod m$, the integer d must be a unit modulo m and $\varepsilon(d) = 1$. It remains to modify the above matrix so as to satisfy the conditions:

$$c \equiv 0 \bmod m, \tag{4}$$

$$d \equiv 0 \bmod 2. \tag{5}$$

If $c \not\equiv 0 \bmod m$, we multiply the above from the left by a matrix

$$\begin{pmatrix} 1 & & \\ c_1 & 1 & \\ & & 1 \end{pmatrix}$$

where $c_1 \equiv \dfrac{-c}{a} \bmod m$. The result is an element

$$\begin{pmatrix} 1 & & \\ c_1 & 1 & \\ & & 1 \end{pmatrix} \begin{pmatrix} a & b & \\ c & d & \\ & & 1 \end{pmatrix} = \begin{pmatrix} a & b & \\ c + a\,c_1 & d + b\,c_1 & \\ & & 1 \end{pmatrix}$$

in $\Gamma_D(m)$ and so condition (4) is satisfied.

On the other hand, if $d \not\equiv 0 \bmod 2$, we can right multiply our original matrix by the matrix

$$\begin{pmatrix} 1 & m & \\ & 1 & \\ & & 1 \end{pmatrix}$$

in $\Gamma_D(m)$, the resulting element

$$\begin{pmatrix} a & b+am & \\ c & d+cm & \\ & & 1 \end{pmatrix}$$

will remain in $\Gamma_D(m)$. However, since $c \equiv m \equiv 1 \bmod 2$, the entry $d+cm$ is divisible by 2. Hence we can make both of the conditions (4) and (5) satisfied, and the proof of the lemma follows.

9.5. Throughout the rest of our discussion, we fix once for all a set of integers a, b, c, d so that

$$\alpha = \begin{pmatrix} a & b & \\ c & d & \\ & & 1 \end{pmatrix}$$

satisfies all the requirements in the previous lemma. In addition, we also use the notation δ to denote the following matrix

$$\delta = \begin{pmatrix} 1 & & \\ & -2 & \\ & & -\frac{1}{2} \end{pmatrix}.$$

Lemma. *Let α, β, δ, σ, s be defined as above. Then there are two commutative diagrams:*

$$\begin{array}{ccccc} X(\gamma_0) & \xrightarrow{\sigma} & X(\gamma_0) & \xrightarrow{\cdot\delta} & X(\gamma_0) \\ \downarrow{\scriptstyle\beta} & & & & \downarrow{\scriptstyle\beta} \\ X(\gamma_1) & & & & X(\gamma_1) \\ \downarrow{\scriptstyle\pi} & & & & \downarrow{\scriptstyle\pi} \\ F(\gamma_1) & & \xrightarrow{\sigma} & & F(\gamma_1), \end{array} \tag{1}$$

$$\begin{array}{ccccc} X(\gamma_0) & \xrightarrow{\cdot s} & X(s^{-1}\gamma_0 s) & \xrightarrow{\cdot\alpha} & X(\gamma_0) \\ \downarrow{\scriptstyle\beta} & & & & \downarrow{\scriptstyle\beta} \\ X(\gamma_1) & & & & X(\gamma_1) \\ \downarrow{\scriptstyle\pi} & & & & \downarrow{\scriptstyle\pi} \\ F(\gamma_1) & & \xrightarrow{\cdot s} & & F(\gamma_1). \end{array} \tag{2}$$

where the vertical arrow π represents the covering map, and the symbols $\cdot\alpha$, $\cdot\beta$, $\cdot\delta$, $\cdot s$ denote the action induced by the elements α, β, δ, s.

Since the proof relies on checking a number of simple matrix identities, we will omit all the straightforward computations.

First we have the identities:

$$\sigma(\gamma_0)=\gamma_0, \quad \delta^{-1}\gamma_0\delta=\gamma_0, \quad \alpha^{-1}s^{-1}\gamma_0 s\alpha=\gamma_0,$$

which show that all the top horizontal arrows are well-defined mappings.

To prove the diagram (1) is commutative, we define the element

$$\bar{\delta}=\begin{pmatrix} 1 & & \\ & -2 & 1 \\ & m & \dfrac{-(m+1)}{2} \end{pmatrix}.$$

Then, using this element, we decompose the diagram in (1) into four squares I), II), III), IV):

$$\begin{array}{ccccc} X(\gamma_0) & \xrightarrow{\sigma} & X(\gamma_0) & \xrightarrow{\cdot\delta} & X(\gamma_0) \\ \downarrow{\scriptstyle\beta} & \text{I} & \downarrow{\scriptstyle\sigma(\beta)} & \text{III} & \downarrow{\scriptstyle\cdot\beta} \\ X(\gamma_1) & \xrightarrow{\sigma} & X(\sigma(\gamma_1)) & \xrightarrow{\cdot\delta} & X(\gamma_1) \\ \downarrow{\scriptstyle\pi} & \text{II} & \downarrow{\scriptstyle\pi} & \text{IV} & \downarrow{\scriptstyle\pi} \\ F(\gamma_1) & \xrightarrow{\sigma} & F(\gamma_1) & \longrightarrow & F(\gamma_1). \end{array} \tag{3}$$

It follows immediately from the definition that the two squares (I) and (II) on the left hand side are commutative. To prove the square (III) is commutative, we use the identities:

$$\bar{\delta}^{-1}\sigma(\gamma_1)\bar{\delta}=\gamma_1, \quad \sigma(\beta)\cdot\bar{\delta}\cdot\beta^{-1}=\delta.$$

By our assumption on m, the matrix $\bar{\delta}$ represents an element in $\Gamma_D(m)$. Hence it induces the identity map on the quotient space, and so the square (IV) is commutative.

An immediate consequence of the commutativity of the four squares is that the diagram (1) is also commutative.

Next we set

$$\bar{\alpha}=\begin{pmatrix} a & b & -\left(\dfrac{b+1}{2}\right) \\ c & d & -\dfrac{d}{2} \\ & & 1 \end{pmatrix}.$$

Then diagram (2) again decomposes into four squares:

$$\begin{array}{ccccc} X(\gamma_0) & \xrightarrow{\cdot s} & X(s^{-1}\gamma_0 s) & \xrightarrow{\cdot\alpha} & X(\gamma_0) \\ \downarrow{\scriptstyle \cdot\beta} & & \downarrow{\scriptstyle \cdot s^{-1}\beta s} & & \downarrow{\scriptstyle \cdot\beta} \\ X(\gamma_1) & \xrightarrow{\cdot s} & X(s^{-1}\gamma_1 s) & \xrightarrow{\bar{\alpha}} & X(\gamma_1) \\ \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} & & \downarrow{\scriptstyle \pi} \\ F(\gamma_1) & \xrightarrow{\cdot s} & F(\gamma_1) & \xrightarrow{\mathrm{id}} & F(\gamma_1). \end{array}$$

To prove that they are commutative, we use the identities

$$(s\bar{\alpha})^{-1}\cdot\gamma_1\cdot(s\bar{\alpha})=\gamma_1, \qquad s^{-1}\beta s\bar{\alpha}=\alpha\beta,$$

together with the fact that $\bar{\alpha}$ is an element in $\Gamma_D(m)$. This completes the proof of our lemma.

9.6. An immediate consequence of Lemma 9.5 is that the action σs on the manifold $F(\gamma_1)$ is induced by the composition $\alpha\circ s\circ\delta\circ\sigma$ on the covering space level:

$$\begin{array}{ccccccc} X(\gamma_0) & \xrightarrow{\sigma} & X(\gamma_0) & \xrightarrow{\cdot\delta} & X(\gamma_0) & \xrightarrow{\cdot s} & X(s^{-1}\gamma_0 s) \xrightarrow{\cdot\alpha} X(\gamma_0) \\ \downarrow{\scriptstyle \beta} & & & & & & \downarrow{\scriptstyle \beta} \\ X(\gamma_1) & & & & & & X(\gamma_1) \\ \downarrow & & & & & & \downarrow{\scriptstyle \pi} \\ F(\gamma_1) & & & & \xrightarrow{\sigma s} & & F(\gamma_1) \end{array} \tag{1}$$

Let A be an element in $X(\gamma_0)$,

$$A=\begin{pmatrix} s_{11} & s_{12} & \\ s_{12} & s_{22} & \\ & & s_{33} \end{pmatrix} \qquad \begin{array}{l} s_{11}>0,\ \Delta=s_{11}s_{22}-s_{12}^2>0, \\ \Delta\cdot s_{33}=m. \end{array} \tag{2}$$

Then under the composite map $\alpha\circ s\circ\delta\circ\sigma$, it becomes the element

$$\begin{pmatrix} & * & 0 \\ & & 0 \\ 0 & 0 & m^2/4s_{33} \end{pmatrix}$$

where

$$*={}^t\begin{pmatrix} -c & -d \\ a & b \end{pmatrix}\cdot\begin{pmatrix} 1 & \\ & -2 \end{pmatrix}\cdot\begin{pmatrix} s_{11} & s_{12} \\ s_{12} & s_{22} \end{pmatrix}^{-1}\cdot\begin{pmatrix} 1 & \\ & -2 \end{pmatrix}\cdot\begin{pmatrix} -c & -d \\ a & b \end{pmatrix}. \tag{3}$$

If we identify $F(\gamma_1)$ with the product space $Y/\Gamma_D^0(2m)\times\mathbb{R}_+$ as in 9.3, then the involution σs is a product involution. On the second factor $\mathbb{R}_+$, it is defined by the equation:

$$s_{33}\mapsto\frac{m^2}{4s_{33}}, \qquad s_{33}\in\mathbb{R}_+.$$

Hence, for a fixpoint (R, s_{33}) in $Y/\Gamma_D^0(2m)\times\mathbb{R}_+$, its second coordinate must satisfy the equation $s_{33}=\dfrac{m^2}{4s_{33}}$, or in other words, it must lie in the level

$$Y/\Gamma_D^0(2,m)\times\left\{\frac{m}{2}\right\}.$$

At this level, $s_{33}=\dfrac{m}{2}$, $\Delta=2$, the involution sends a point $\{R\}$,

$$R=\frac{1}{\sqrt{2}}\begin{pmatrix}s_{11} & s_{12}\\ s_{12} & s_{22}\end{pmatrix},$$

to the point $\{v(R)\}$ where

$$v(R)=\frac{1}{\sqrt{2}}\,{}^t\begin{pmatrix}-c & -d\\ a & b\end{pmatrix}\begin{pmatrix}1 & \\ & -2\end{pmatrix}\begin{pmatrix}s_{11} & s_{12}\\ s_{12} & s_{22}\end{pmatrix}^{-1}\begin{pmatrix}1 & \\ & -2\end{pmatrix}\begin{pmatrix}-c & -d\\ a & b\end{pmatrix}.$$

Comparing this with the formulas in 8.1, it is not difficult to see that this involution is induced by the "twisted" Cartan involution:

$$v\colon SL_2(\mathbb{Q})\to SL_2(\mathbb{Q}),$$

$$v(A)=\begin{pmatrix}-c & -d\\ a & b\end{pmatrix}\begin{pmatrix}1 & \\ & -2\end{pmatrix}\cdot\tau(A)\cdot\begin{pmatrix}1 & \\ & -2\end{pmatrix}^{-1}\begin{pmatrix}-c & -d\\ a & b\end{pmatrix}^{-1} \tag{4}$$

on the group level.

9.7. Proposition. *The fixpoint set of σs on $F(\gamma_1)$ has zero Euler characteristic,* i.e.

$$\chi(F(\gamma_1)^{\sigma s})=0.$$

To compute the Euler characteristic of the fixpoint set $F(\gamma_1)^{\sigma s}$, we observe that the action defined in the previous paragraph

$$R\mapsto v(R)$$

is orientation reversing. Their fixpoint set must consist of 1-dimensional submanifolds, and so they can either be circles or straight lines. Since the Euler characteristic of a circle is zero, we can omit it from our consideration.

To deal with the situation of the straight lines, we compactify the manifold by attaching a boundary circle to each cusp. If l is a fixed line, then its endpoints consist of two fixpoints lying on boundary circles. These circles must be stable under the action of σs. Conversely, if σs keeps a boundary circle invariant, then such circles must have two fixpoints and these two fixpoints must be the endpoints of some fixed lines. From this it follows that the Euler characteristic of $F(\gamma_1)^{\sigma s}$ is the same as the number of boundary components fixed by σs.

To determine the last number, we let $\Gamma_0(2,2)$ denote the Hecke congruence subgroup of level 2:

$$\Gamma_0(2,2)=\left\{ \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \in SL_2(\mathbb{Z}) \middle| a_{21} \equiv 0 \bmod 2 \right\}.$$

Now $\Gamma_D^0(2,m)$ is a subgroup in $\Gamma_0(2,2)$, and so there is a natural induced mapping

$$\Phi\colon\ Y/\Gamma_D^0(2,m) \to Y/\Gamma_0(2,2),$$

which maps the boundary cusps to the boundary cusps. It is easy to see that the involution v in 9.6(4) keeps the congruence subgroup $\Gamma_0(2,2)$ stable, and so it induces an involution on the quotient space $Y/\Gamma_0(2,2)$. Furthermore the map Φ is equivariant with respect to these two involutions.

On the other hand, in our definition of v, we use the conjugation by the matrix

$$\begin{pmatrix} -c & -d \\ a & b \end{pmatrix}.$$

Since this is an element in $\Gamma_0(2,2)$ it has no effect on the quotient space $Y/\Gamma_0(2,2)$, and so our involution is the same as

$$\tilde{v}\colon\ R \mapsto \frac{1}{\sqrt{2}} \begin{pmatrix} 1 & \\ & -2 \end{pmatrix} R^{-1} \begin{pmatrix} 1 & \\ & -2 \end{pmatrix}.$$

As well known, there are two cusps for $Y/\Gamma_0(2,2)$ which are represented by

$$\lim_{s_{11}\to 0} \begin{pmatrix} s_{11} & \\ & s_{11}^{-1} \end{pmatrix} \quad \text{resp.} \quad \lim_{s_{11}\to \infty} \begin{pmatrix} s_{11} & \\ & s_{11}^{-1} \end{pmatrix}.$$

Under the involution $\tilde{v}$, these two cusps are interchanged and so there is no fixpoint on the boundary of $Y/\Gamma_0(2,2)$. Since Φ is equivariant, it follows that the same is true for $Y/\Gamma_D^0(2,m)$, and so

$$\chi(F(\gamma_1)^{\sigma_S})=0.$$

This completes the proof of Proposition 9.7.

References

1. Ash, A.: Cuspidal cohomology of congruence subgroups of $SL(3,\mathbb{Z})$. Preprint 1980
2. Borel, A.: Cohomology of arithmetic groups. In: Proc. Int. Congress Math. Vancouver, vol. I, pp. 435–442, Vancouver 1974
3. Borel, A.: Stable real cohomology of arithmetic groups II. In: Hano, J., et al (eds.) Manifolds and Lie groups. Progress in Maths., vol. 14, pp. 21–55. Boston-Basel-Stuttgart: Birkhäuser 1981
4. Borel, A., Serre, J.P.: Corners and arithmetic groups. Comment. Math. Helvetici **48**, 436–491 (1973)
5. Borel, A., Wallach, N.: Continuous cohomology, discrete subgroups and representations of reductive groups. Annals of Math. Studies **94**, Princeton 1980

6. Cartan, H.: Fonctions automorphes et séries de Poincaré. J. Anal. Math. **6**, 169–175 (1958) (=Oeuvres, II, 731–737)
7. Fricke, R.: Die elliptischen Funktionen und ihre Anwendungen, 2. Teil: Die algebraischen Ausführungen. Leipzig-Berlin: Teubner 1922
8. Grunewald, F., Schwermer, J.: A non-vanishing theorem for the cuspidal cohomology of SL_2 over imaginary quadratic integers. Math. Annalen **258**, 183–200 (1981)
9. Harder, G.: On the cohomology of $SL_2(\mathcal{O})$. In: Gelfand, I.M. (ed.) Lie groups and their representations, pp. 139–150. London: Hilger 1975
10. Harder, G.: On the cohomology of discrete arithmetically defined groups. In: Proc. of the Int. Colloq. on Discrete subgroups of Lie groups and Appl. to Moduli (Bombay 1973), pp. 129–160. Oxford: University Press 1975
11. Harder, G.: A Gauss-Bonnet formula for discrete arithmetically defined groups. Ann. Scient. Ec. Norm. Sup. (4) **4**, 409–455 (1971)
12. Hecke, E: Über ein Fundamentalproblem aus der Theorie der elliptischen Modulfunktionen, Abh. math. Seminar der hamburg. Universität **6**, 235–257 (1928) (=Math. Werke, 525–547)
13. Jones, B.: The arithmetic theory of quadratic forms. Carus Math. Monographs 10. New York: MAA (John Wiley) 1950
14. Langlands, R.: Modular forms and l-adic representations, In: Modular functions of one variable II, Lect. Notes in Maths., vol. 349; pp. 361–500. Berlin-Heidelberg-New York: Springer 1973
15. Langlands, R.: Automorphic representations, Shimura varieties, and motives. Ein Märchen. In: Proc. Symp. Pure Maths., vol. XXXIII, part 2, pp. 205–246. Providence: AMS 1979
16. Lee, R., Schwermer, J.: Cohomology of arithmetic subgroups of SL_3 at infinity. J. f. d. Reine u. Angewandte Math. **330**, 100–131 (1982)
17. Millson, J.: Higher weight cycles. J. Fac. Science Univ. Tokyo **28**, 465–479 (1981)
18. Minkowski, H.: Gesammelte Abhandlungen. I. Leipzig-Berlin: Teubner 1911
19. Rohlfs, J.: The Lefschetznumber of an involution on the space of classes of positive definite quadratic forms. Comment. Math. Helvetici **56**, 272–296 (1981)
20. Schwermer, J.: Sur la cohomologie de $SL_n(\mathbb{Z})$ à l'infini et les séries d'Eisenstein. C. R. Acad. Sc. Paris **289**, 413–415 (1979)
21. Schwermer, J.: Kohomologie arithmetisch definierter Gruppen und Eisensteinreihen, Lect. Notes in Maths., vol. 988. Berlin-Heidelberg-New York: Springer 1983
22. Serre, J.-P.: Representations linéaires des groupes finis. Paris: Hermann 1967
23. Serre, J.-P.: Cohomologie des groupes discrets. In: Prospects in Mathematics. Annals of Math. Studies **70**, 77–169 Princeton: University Press 1971
24. Serre, J.-P.: Cohomologie Galoisienne. Lect. Notes in Maths. vol. 5. Berlin-Heidelberg-New York: Springer 1964
25. Shimura, G.: Introduction to the arithmetic theory of automorphic functions. Publ. Math. Soc. Japan vol. 11, Princeton University Press 1971
26. Simpson, W.A., Frame, J.S.: The character tables for $SL(3, q)$, $SU(3, q^2)$, $PSL(3, q)$, $PSU(3, q^2)$. Canadian J. Math. **25**, 486–494 (1973)
27. Soulé, C.: The cohomology of $SL_3(\mathbb{Z})$. Topology **17**, 1–22 (1978)
28. Tahara, K.: Finite subgroups of $GL_3(\mathbb{Z})$. Nagoya Math. J. **41**, 169–209 (1971)
29. Tamagawa, T.: Adèles. In: Proc. Symp. Pure Maths., vol IX, pp. 113–121. Providence: AMS 1966
30. Wallach, N.: Square integrable automorphic forms and cohomology of arithmetic quotients of $SU(p, q)$. Preprint 1981

Oblatum 18-X-1982

Inventiones math. 73, 241-245 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

On the *K*-Theory of Algebraically Closed Fields

A. Suslin

191011 USSR, Leningrad, Fontanaka 27, LOMI

Let F be an algebraically closed field. It was conjectured by Lichtenbaum (see [1], [5]) that $K_i(F)$ is divisible if $i \geqq 1$, the torsion being zero if i is even and isomorphic to $W_{(n)}$ (where W is the group of roots of unity in F^*) if $i=2n-1$. This conjecture is equivalent to the conjecture concerning the structure of the cohomology ring $H^*(GL(F), \mathbf{Z}/\ell)$ with ℓ prime different from $\operatorname{char} F$ (see [5]) and also to the conjecture concerning the structure of the K-theory ring $K_*(F, \mathbf{Z}/\ell)$. All these conjectures are known to be true when F is the algebraic closure of a finite field by the work of Quillen [3]. The main result of the present paper shows that, if $F_0 \subset F$ is an extension of algebraically closed fields, then the natural homomorphisms ${}_nK_*(F_0) \to {}_nK_*(F)$, $K_*(F_0)/n \to K_*(F)/n$, $K_*(F_0, \mathbf{Z}/n) \to K_*(F, \mathbf{Z}/n)$ are isomorphisms. This theorem solves affirmatively Lichtenbaum's conjecture for fields of positive characteristic; its proof is based on the rigidity property for $K_*(F, \mathbf{Z}/n)$ which, as *I* hope, may be useful also in dealing with the zero characteristic case. It was noticed by *C.* Soulé that the same proof may be used to show that K'-theory with coefficients of any variety defined over an algebraically closed field does not change under (algebraically closed) base change, so we prove below this more general statement. I am also much obliged to E. Friedlander for his stimulating interest in the problem and to O. Gabber, who has simplified certain proofs below. This work was done during my staying at the University Paris VII and I want to thank it for its hospitality.

Let O be a discrete valuation ring with associated valuation v, residue field F and quotient field E and let X be a scheme of finite type over O. We'll denote by $M(X)$ the category of coherent sheaves on X and by $M_0(X)$ its full subcategory consisting of sheaves with support in the closed fiber of $X \to \operatorname{Spec} O$. By the localization theorem (see [4]) the sequence $BQM_0(X) \to BQM(X) \to BQM(X_E)$ is a homotopy fibration and by the dévissage theorem [4] the natural imbedding $M(X_F) \hookrightarrow M_0(X)$ induces a homotopy equivalence $BQM(X_F) \tilde{\to} BQM_0(X)$. In particular we have an exact localization sequence:

$$\ldots \to K'_{i+1}(X_F, \mathbf{Z}/n) \to K'_{i+1}(X, \mathbf{Z}/n) \to K'_{i+1}(X_E, \mathbf{Z}/n) \xrightarrow{d} K'_i(X_F, \mathbf{Z}/n) \to \ldots$$

Denote by φ the following homomorphism:

$$K'_i(X_E, \mathbf{Z}/n)\otimes E^* = K'_i(X_E, \mathbf{Z}/n)\otimes K_1(E) \xrightarrow{\text{mult}} K'_{i+1}(X_E, \mathbf{Z}/n) \xrightarrow{d} K'_i(X_F, \mathbf{Z}/n).$$

We shall use the notation $\ell(a)$ to denote the element of $K_1(E)$ corresponding to $a\in E^*$ under the canonical identification $K_1(E)=E^*$. The well known properties of d with respect to products (see [2]) show that

$$\text{If } t\in K'_i(X_E, \mathbf{Z}/n) \text{ and } a\in O^*, \text{ then } \varphi(t\otimes a)=d(t)\cdot\ell(\bar{a}). \tag{1.1}$$

$$\text{If } t\in K'_i(X, \mathbf{Z}/n) \text{ and } a\in E^*, \text{ then } \varphi(t\otimes a)=(-1)^i\bar{t}\cdot d(\ell(a)) = (-1)^i v(a)\bar{t} \tag{1.2}$$

where $\bar{t}$ denotes the image of t under the homomorphism induced on the K'-groups by the (finite Tor-dimension) closed imbedding $X_F\hookrightarrow X$).

Suppose that O' is a finite normal O-algebra. The ring O' is semilocal and if p_j $(j=1,\ldots,k)$ are its maximal ideals then $O'_j=O'_{p_j}$ are discrete valuation rings. We denote by E' the quotient field of O', by F'_j the residue field of O'_j and by e_j the ramification index of O'_j over O. Finally we denote by $N_{E'/E}$ (resp. $N_{F'_j/F}$) the transfer (or direct image homomorphism) corresponding to the finite morphism $X_{E'}\to X_E$ (resp. $X_{F'_j}\to X_F$). From the commutative diagram of fibrations

$$\begin{array}{ccccc} BQM_0(X_{O'}) & \longrightarrow & BQM(X_{O'}) & \longrightarrow & BQM(X_{E'}) \\ \downarrow & & \downarrow & & \downarrow \\ BQM_0(X) & \longrightarrow & BQM(X) & \longrightarrow & BQM(X_E) \end{array}$$

we deduce immediately the following well known fact:

Lemma 1, *The following diagram commutes for any* $i\geqq 0$

$$\begin{array}{ccc} K'_{i+1}(X_{E'}, \mathbf{Z}/n) & \xrightarrow{N_{E'/E}} & K'_{i+1}(X_E, \mathbf{Z}/n) \\ \downarrow{\scriptstyle (d_j)} & & \downarrow{\scriptstyle d} \\ \coprod_{j=1}^{k} K'_i(X_{F'_j}, \mathbf{Z}/n) & \xrightarrow{\Sigma N_{F'_j/F}} & K'_i(X_F, \mathbf{Z}/n). \end{array}$$

Suppose now that the field F is algebraically closed. The formula (1.1) shows that $\varphi(K'_i(X_E, \mathbf{Z}/n)\otimes O^*)$ is in the image of $K'_{i-1}(X_E, \mathbf{Z}/n)\otimes F^*\to K'_i(X_F, \mathbf{Z}/n)$, which is zero since F^* is divisible. This shows that φ factors through $K'_i(X_E, \mathbf{Z}/n)\otimes(E^*/O^*)=K'_i(X_E, \mathbf{Z}/n)$. We denote the resulting homomorphism $K'_i(X_E, \mathbf{Z}/n)\to K'_i(X_F, \mathbf{Z}/n)$ by s and we call it the specialization

homomorphism. By the definition of s we have the following formulae:

$$\text{For any } t\in K_i'(X_E,\mathbf{Z}/n) \text{ and any } a\in E^* \quad d(t\cdot \ell(a))=v(a)\cdot s(t) \tag{1.3}$$

$$\text{For any } t\in K_i'(X,\mathbf{Z}/n) \quad s(t)=(-1)^i\cdot \bar{t}. \tag{1.4}$$

Lemma 2. *s vanishes on the image of $K_{i-1}'(X,\mathbf{Z}/n)\otimes E^*\to K_i'(X_E,\mathbf{Z}/n)$.*

Proof. Suppose, that $t\in K_{i-1}'(X,\mathbf{Z}/n)$, $a\in E^*$ and let π be any prime element of O. Then

$$\begin{aligned} s(t\cdot \ell(a)) &= d(t\cdot \ell(a)\cdot \ell(\pi)) \\ &= (-1)^{i-1}\,\bar{t}\cdot d(\ell(a)\cdot \ell(\pi))\in K_{i-1}'(X_F,\mathbf{Z}/n)\,K_1(F)=0. \end{aligned}$$

In the situation of Lemma 1 suppose that F is algebraically closed, then $F_j'=F$ and the discrete valuation rings O_j' define specialisations $s_j\colon K_i'(X_{E'},\mathbf{Z}/n)\to K_i'(X_F,\mathbf{Z}/n)$.

Lemma 3. $s N_{E'/E}=\sum_{j=1}^{k} e_j\cdot s_j$.

Proof. Let $\pi\in O$ be a prime element, then for any $t\in K_i'(X_{E'},\mathbf{Z}/n)$ we have (in view of the projection formula, Lemma 1 and (1.3)):

$$\begin{aligned} s(N_{E'/E}(t)) &= d(N_{E'/E}(t)\cdot \ell(\pi)) = d(N_{E'/E}(t\cdot \ell(\pi)) \\ &= \sum_{j=1}^{k} d_j(t\cdot \ell(\pi)) = \sum_{j=1}^{k} e_j s_j(t). \end{aligned}$$

Let now X be a variety over an algebraically closed field F and let C/F be a smooth curve with function field $E=F(C)$. If $c\in C$ is a closed point then we can consider the corresponding specialization $s\colon K_i'(X_E,\mathbf{Z}/n)\to K_i'(X,\mathbf{Z}/n)$. The restriction of s to $K_i'(X\times C,\mathbf{Z}/n)$ coincides (up to sign) with the natural homomorphism (which we denote by $t\mapsto t(c)$) induced by the morphism $X=c\hookrightarrow X\times C$.

Rigidity Theorem. *If $c_0,c_1\in C$ are two closed points and s_0,s_1 are the corresponding specializations, then $s_0=s_1$.*

Proof. Suppose first that $C=\mathbf{P}_F^1$ and denote by T the coordinate function on C. We have the direct sum decomposition:

$$K_i'(X_E,\mathbf{Z}/n)=K_i'(X,\mathbf{Z}/n)\oplus \coprod_{a\in F} K_{i-1}'(X,\mathbf{Z}/n)\cdot \ell(T-a).$$

Both s_0 and s_1 are equal to $(-1)^i$ on the first summand and are zero on the remaining summands by Lemma 2. In the general case we may clearly suppose that C is complete. Consider the homomorphism $q\colon K_i'(X_E,\mathbf{Z}/n)\otimes \mathrm{Div}^0(C)\to K_i'(X,\mathbf{Z}/n)$ given by $t\otimes\sum n_i c_i\mapsto\sum n_i s_i(t)$. If $f\colon C\to\mathbf{P}_F^1$ is a rational function then, in view of Lemma 3 and what was proved above, we have:

$$q(t\otimes(f))=s_0(N(t))-s_\infty(N(t))=O$$

(here N is taken with respect to the field imbedding $F(\mathbf{P}^1)\hookrightarrow F(C)$). This shows that q factors through $K_i'(X_E, \mathbf{Z}/n)\otimes \mathrm{Pic}^0(C)$ which is zero since $\mathrm{Pic}^0(C)$ is divisible.

Corollary 1. *Suppose that A is a connected smooth affine F-algebra and $h_0, h_1 : A\to F$ are any F-algebra homomorphisms. The induced homomorphisms $K_i'(X_A, \mathbf{Z}/n)\to K_i'(X, \mathbf{Z}/n)$ coincide.*

Proof. Let $Z=\mathrm{Spec}\, A$ and let z_0, z_1 be closed points on Z corresponding to h_0 and h_1. There exists an irreducible curve $C'\subset Z$ joining z_0 and z_1 (see [6] p. 56) and the statement follows from the rigidity theorem, applied to the normalization C of C'.

Let $F_0\subset F$ be an extension of algebraically closed fields and let X_0 be a variety over F_0. Denote $(X_0)_F$ by X.

Corollary 2. *If A is any smooth affine F_0-algebra without zero divisors and $h_0, h_1 : A\to F$ are any F_0-algebra homomorphisms, then the induced homomorphisms $K_i'((X_0)_A, \mathbf{Z}/n)\to K_i'(X, \mathbf{Z}/n)$ coincide.*

Main Theorem. *In the above notation the homomorphisms ${}_nK'_*(X_0)\to {}_nK'_*(X)$, $K'_*(X_0)/n\to K'_*(X)/n$, $K'_*(X_0, \mathbf{Z}/n)\to K'_*(X, \mathbf{Z}/n)$ are isomorphisms.*

Proof. Let R denote any of the functors under consideration. We can write F in the form $F=\varinjlim A$, where A runs through all smooth affine F_0-subalgebras of F. Since R respects direct limits we see that $R(X)=\varinjlim R((X_0)_A)$ and hence the element of $R(X_0)$ dying in $R(X)$ also dies in some $R((X_0)_A)$. Since F_0 is algebraically closed there exists some F_0-algebra homomorphism h: $A\to F_0$ and any such homomorphism induces a splitting $R((X_0)_A)\to R(X_0)$ for the map $R(X_0)\to R((X_0)_A)$. This shows that $R(X_0)\to R((X_0)_A)$ is split injective and hence $R(X_0)\to R(X)$ is injective.

Finally we see from Corollary 2 that the homomorphism $K_i'(X_0, \mathbf{Z}/n)\to K_i'(X, \mathbf{Z}/n)$ induced by the imbedding $A\hookrightarrow F$ coincides with the homomorphism induced by $A\xrightarrow{h} F_0\hookrightarrow F$, and hence its image is contained in the image $K_i'(X_0, \mathbf{Z}/n)\to K_i'(X, \mathbf{Z}/n)$. This proves surjectivity for K-groups with coefficients. The remaining isomorphisms follow from the snake lemma applied to the commutative diagram.

$$\begin{array}{ccccccc}
0\to K_i'(X_0)/n & \longrightarrow & K_i'(X_0, \mathbf{Z}/n) & \longrightarrow & {}_nK'_{i-1}(X_0)\to 0 \\
\downarrow & & \downarrow & & \downarrow \\
0\to K_i'(X)/n & \longrightarrow & K_i'(X, \mathbf{Z}/n) & \longrightarrow & {}_nK'_{i-1}(X)\to 0
\end{array}$$

Corollary 1. *With the same notation as above the imbedding $F_0\hookrightarrow F$ induces isomorphisms*

$$H_i(GL_n(F_0), \mathbf{Z}/\ell)\xrightarrow{\sim} H_i(GL_n(F), \mathbf{Z}/\ell) \quad \text{when } i\leqq n.$$

Proof. Consider the morphism $BSL(F_0)^+\to BSL(F)^+$. It follows from the main theorem that this morphism induces isomorphism of homotopy groups with

coefficients $\mathbf{Z}/\ell$ and since both spaces are simply connected we deduce from this (see [7]) that it induces also an isomorphism of homology groups with coefficients $\mathbf{Z}/\ell$. Thus $H_*(SL(F_0), \mathbf{Z}/\ell) \xrightarrow{\sim} H_*(SL(F), \mathbf{Z}/\ell)$. Comparing the Hochschild-Serre spectral sequences corresponding to the extensions

$$1 \to SL(F_0) \to GL(F_0) \to F_0^* \to 1$$
$$1 \to SL(F) \to GL(F) \to F^* \to 1$$

we deduce that $H_*(GL(F_0), \mathbf{Z}/\ell) \to H_*(GL(F), \mathbf{Z}/\ell)$ is an isomorphism. The result for GL_n follows from this and from the homology stability theorem for infinite fields ([8]).

Corollary 2. *If F is an algebraically closed field of positive characteristic p then $H_*(GL(F), \mathbf{Z}/p) = \mathbf{Z}/p$ and for $\ell \neq p$ the homology ring $H_*(GL(F), \mathbf{Z}/\ell)$ is a polynomial ring over $\mathbf{Z}/\ell$ in variables $c_i \in H_{2i}(GL(F), \mathbf{Z}/\ell)$.*

Proof. This statement is known for the algebraic closure F_0 of the prime subfield in F ([3]). The general case follows now from Corollary 1.

References

1. Gersten, S.M.: Problems about higher K-functors. Lecture Notes Math. vol. 341, pp. 43–57. Berlin-Heidelberg-New York: Springer 1973
2. Grayson, D.: Products in K-theory and intersecting algebraic cycles. Invent. Math. **47**, 71–84 (1978)
3. Quillen, D.: On the cohomology and K-theory of the general linear group over a finite field. Ann. of Math. **96**, 552–586 (1972)
4. Quillen, D.: Higher algebraic K-theory I. Lecture Notes Math. vol. 341, pp. 85–147. Berlin-Heidelberg-New York: Springer 1973
5. Quillen, D.: Higher algebraic K-theory. Proc. Vanc. Intern. Congr. Math. **1**, 171–177 (1974)
6. Mumford, D.: Abelian varieties. Oxford: University Press 1974
7. Neisendorfer, J.: Primary homotopy theory. Memoirs of the AMS, N 232, 1980
8. Suslin, A.: Homology of GL_n, characteristic classes and Milnor K-theory. LOMI preprint E-4-82

Oblatum February 9, 1983

Investiones math. 73, 247–250 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Homology of the Infinite Orthogonal and Symplectic Groups over Algebraically Closed Fields

An Appendix to the Paper of A. Suslin

M. Karoubi

UER de Mathématiques-Université de Paris VII, 2 Place Sussien, F-75251 Paris, France

As a consequence of Suslin's result and the fundamental theorem of Hermitian K-theory [3], we shall prove the following statement:

Theorem. *Let k be an algebraically closed field and k_0 the algebraic closure of the prime field. Then, the obvious homomorphisms*

$$O_{2n}(k_0)\to O_{2n}(k) \quad \text{and} \quad Sp_{2n}(k_0)\to Sp_{2n}(k)$$

induce isomorphisms

$$\varinjlim H_*(O_{2n}(k_0); \mathbf{Z}/q)\xrightarrow{\approx}\varinjlim H_*(O_{2n}(k); \mathbf{Z}/q)$$
$$\varinjlim H_*(Sp_{2n}(k_0); \mathbf{Z}/q)\xrightarrow{\approx}\varinjlim H_*(Sp_{2n}(k); \mathbf{Z}/q)$$

where q is arbitrary if $\mathrm{char}(k)\neq 2$ *and q is odd if* $\mathrm{char}(k)=2$.

Proof. Let F denote either k or k_0. Here $Sp_{2n}(F)$ is the usual symplectic group and $O_{2n}(F)$ the orthogonal group of F^{2n} provided with the "hyperbolic" quadratic form

$$(x_1, \ldots, x_n, y_1, \ldots, y_n)\mapsto \sum_{i=1}^{n} x_i y_i.$$

Therefore $O(F)=\varinjlim O_{2n}(F)$ and $Sp(F)=\varinjlim Sp_{2n}(F)$ are just the groups denoted by ${}_1O(F)$ and ${}_{-1}O(F)$ in [2]. If $\mathrm{char}(F)\neq 2$, since $\sqrt{-1}\in F$, $O(F)$ may be considered as the direct limit of the orthogonal groups of F^n provided with the quadratic form defined by the sum of the squares of the coordinates. Since homology commutes with direct limits, we have to prove isomorphisms

$$H_*(O(k); \mathbf{Z}/q)\approx H_*(O(k_0); \mathbf{Z}/q) \quad \text{and} \quad H_*(Sp(k); \mathbf{Z}/q)\approx H_*(Sp(k_0); \mathbf{Z}/q).$$

Let us recall now some basic definitions and results of [2, 3] adapted to this situation. The Hermitian K-theory spectrum ${}_\varepsilon\mathscr{L}(F)$, with $\varepsilon=\pm 1$, is defined as follows ([2], p. 212–218):

$${}_\varepsilon\mathscr{L}_n(F)=\Omega^n({}_\varepsilon L_0(F)\times B_\varepsilon O(F)^+) \qquad \text{for } n\geqq 0$$
$${}_\varepsilon\mathscr{L}_n(F)={}_\varepsilon L_0(S^{-n}F)\times B_\varepsilon O(S^{-n}F)^+ \qquad \text{for } n<0.$$

The Hermitian "L-groups" are given by the formulas

$$_\varepsilon L_n(F)=\pi_0(_\varepsilon\mathscr{L}_n(F)),\qquad _\varepsilon L_n(F;\mathbf{Z}/q)=\pi_0(_\varepsilon\mathscr{L}_n(F);\mathbf{Z}/q)=\pi_i(_\varepsilon\mathscr{L}_{n-i}(F);\mathbf{Z}/q)$$

(the last equality may be taken as a definition for $i\geqq 2$). For example $_{-1}L_0(F)\approx {}_1L_0(F)\approx\mathbf{Z}$ (this is just the Grothendieck group of the category of F-vector spaces provided with a non degenerate skew symmetric or quadratic form). We have also $_{-1}L_1(F)\approx H_1(Sp(F);\mathbf{Z})=O$ and

$$_1L_1(F)\approx H_1(O(F);\mathbf{Z})\approx\mathbf{Z}/2.$$

The Quillen K-theory spectrum $\mathscr{K}(F)$ is described in an analogous way ([2], p. 212–218):

$$\mathscr{K}_n(F)=\Omega^n(K_0(F)\times BGL(F)^+)\qquad\text{for } n\geqq 0,$$
$$\mathscr{K}_n(F)=K_0(S^{-n}F)\times BGL(S^{-n}F)^+\qquad\text{for } n<0.$$

Note that $K_n(F)=\pi_0(\mathscr{K}_n(F))$ and $K_n(F;\mathbf{Z}/q)\approx\pi_i(\mathscr{K}_{n-i}(F);\mathbf{Z}/q)$ for $i\geqq 2$. Note that $K_n(F)$ (resp. $_\varepsilon L_n(F)$) are for $n>0$ the homotopy groups of an infinite loop space which homology is precisely that of $GL(F)$ (resp. $_\varepsilon O(F)$). Note finally that for $n<0$, $K_n(F)=0$ ([4], p. 73), but $_\varepsilon L_n(F)\neq 0$ in general (compare with [3], p. 278).

Finally, there are maps of spectra

$$\mathscr{K}_n(F)\to{}_\varepsilon\mathscr{L}_n(F)\quad\text{and}\quad{}_\varepsilon\mathscr{L}_n(F)\to\mathscr{K}_n(F)$$

induced by the hyperbolic and the forgetful functor respectively. If we denote by $_\varepsilon\mathscr{U}_n(F)$ and $_\varepsilon\mathscr{V}_n(F)$ the homotopy fibers of these two maps, then the fundamental theorem of [3] implies the existence of an homotopy equivalence

$$_\varepsilon\mathscr{V}_n(F)\approx\Omega\,{}_{-\varepsilon}\mathscr{U}_n(F)$$

if $\mathrm{char}(F)\neq 2$ and

$$_\varepsilon\mathscr{V}_n(F)_{(2)}\sim\Omega\,{}_{-\varepsilon}\mathscr{U}_n(F)_{(2)}$$

if $\mathrm{char}(F)=2$ ([3], p. 260; cf. [2], p. 253 for the definition of $X_{(2)}$ when X is an H-space). Therefore, if we define $_\varepsilon V_n(F)$ (resp. $_\varepsilon U_n(F)$) as $\pi_0(_\varepsilon\mathscr{V}_n(F))$ (resp. $\pi_0(_\varepsilon\mathscr{U}_n(F))$, then we have an isomorphism $_\varepsilon V_n(F)\approx {}_{-\varepsilon}U_{n+1}(F)$ if $\mathrm{char}(F)\neq 2$ and $_\varepsilon V_n(F)\otimes\mathbf{Z}[\tfrac{1}{2}]\approx {}_{-\varepsilon}U_{n+1}(F)\otimes\mathbf{Z}[\tfrac{1}{2}]$ if $\mathrm{char}(F)=2$. In the same way as above we may introduce U and V groups with coefficients in $\mathbf{Z}/q$ and then we have four types of exact sequences:

$$K_{n+1}(F)\to{}_\varepsilon L_{n+1}(F)\to{}_\varepsilon U_n(F)\to K_n(F)\to{}_\varepsilon L_n(F)$$
$$_\varepsilon L_{n+1}(F)\to K_{n+1}(F)\to{}_\varepsilon V_n(F)\to{}_\varepsilon L_n(F)\to K_n(F)$$
$$\bar{K}_{n+1}(F)\to{}_\varepsilon\bar{L}_{n+1}(F)\to{}_\varepsilon\bar{U}_n(F)\to\bar{K}_n(F)\to{}_\varepsilon\bar{L}_n(F)$$
$$_\varepsilon\bar{L}_{n+1}(F)\to\bar{K}_{n+1}(F)\to{}_\varepsilon\bar{V}_n(F)\to{}_\varepsilon\bar{L}_n(F)\to\bar{K}_n(F)$$

where, for short, we denote by $\bar{K}$, $\bar{L}$, $\bar{U}$ and $\bar{V}$ the corresponding K, L, U and V groups with coefficients in $\mathbf{Z}/q$. Note that $_\varepsilon\bar{V}_n(F)\approx {}_{-\varepsilon}\bar{U}_{n+1}(F)$ (q is assumed odd if $\mathrm{char}(F)=2$). We are going to prove in three steps that $_\varepsilon\bar{L}_n(k)\approx{}_\varepsilon\bar{L}_n(k_0)$, i.e. $\pi_n(_\varepsilon\mathscr{L}(k);\mathbf{Z}/q)\approx\pi_n(_\varepsilon\mathscr{L}(k_0);\mathbf{Z}/q)$ for all values of n. The Hurewicz theorem mod. q will then imply the theorem of this appendix.

Step 1. *The groups* ${}_\varepsilon\bar{L}_n(k)$ *and* ${}_\varepsilon\bar{L}_n(k_0)$ *are isomorphic for* $n=0$ *and* 1. The exact sequence (E_n):

$$ {}_\varepsilon L_{n+1}(F) \xrightarrow{\times q} {}_\varepsilon L_{n+1}(F) \to {}_\varepsilon\bar{L}_{n+1}(F) \to {}_\varepsilon L_n(F) \xrightarrow{\times q} {}_\varepsilon L_n(F) $$

written for $n=0$, shows that ${}_\varepsilon\bar{L}_1(F) \approx \operatorname{Coker}\, {}_\varepsilon L_1(F) \xrightarrow{\times q} {}_\varepsilon L_1(F)$, therefore is 0 if q is odd or $\varepsilon=-1$ and is $\mathbf{Z}/2$ if q is even and $\varepsilon=1$. For the group ${}_\varepsilon\bar{L}_0(F)$ let us distinguish two cases according to the characteristic of F.

If $\operatorname{char}(F)\neq 2$, we write the two exact sequences

$$ {}_\varepsilon L_0(F) \longrightarrow K_0(F) \longrightarrow {}_\varepsilon V_{-1}(F) \longrightarrow {}_\varepsilon L_{-1}(F) \longrightarrow K_{-1}(F) $$

$$ \wr\wr $$

$$ K_1(F) \longrightarrow {}_{-\varepsilon}L_1(F) \longrightarrow {}_{-\varepsilon}U_0(F) \longrightarrow K_0(F) \longrightarrow {}_{-\varepsilon}L_0(F) $$

Let ${}_\varepsilon\alpha$ be the map ${}_\varepsilon L_0(F) \to K_0(F)$. Since ${}_{-1}L_1(F)=0$, the second line in the diagram above shows that $0={}_{-1}U_0(F) \approx {}_{+1}V_{-1}(F) \approx \operatorname{Coker}\, {}_1\alpha$. Since $K_{-1}(F)=0$, we also get ${}_1L_{-1}(F)=0$. In the same way, since ${}_{-1}V_{-1}(F) \approx {}_1U_0(F) \approx \mathbf{Z}/2$ and $\operatorname{Coker}\, {}_{-1}\alpha=\mathbf{Z}/2$, we have also ${}_{-1}L_{-1}(F)=0$. Therefore, the exact sequence (E_{-1}) above will imply ${}_\varepsilon\bar{L}_0(F) \approx \operatorname{Coker}\, {}_\varepsilon L_0(F) \xrightarrow{\times q} {}_\varepsilon L_0(F) \approx \mathbf{Z}/q$.

If $\operatorname{char}(F)=2$, let us denote $\tilde{K}$, $\tilde{L}$, $\tilde{U}$, $\tilde{V}$ the groups K, L, U and V tensorized by $\mathbf{Z}[\frac{1}{2}]$. Then the same arguments as above show that ${}_\varepsilon\tilde{L}_{-1}(F)=0$ since ${}_\varepsilon\tilde{V}_{-1}(F)={}_{-\varepsilon}\tilde{U}_0(F)=\operatorname{Coker}\, {}_\varepsilon\tilde{\alpha}=0$. Therefore ${}_\varepsilon\bar{L}_0(F)=\mathbf{Z}/q$ for q odd.

Step 2. *If* $\operatorname{char}(k)\neq 2$, ${}_\varepsilon L_n(k) \approx {}_\varepsilon L_n(k_0)$ *and* ${}_\varepsilon\bar{L}_n(k) \approx {}_\varepsilon\bar{L}_n(k_0)$ *for* $n<0$.
If $\operatorname{char}(k)=2$, ${}_\varepsilon\bar{L}_n(k) \approx {}_\varepsilon\bar{L}_n(k_0)$ *for* $n<0$.

This step is not really necessary for the proof of our main theorem. It is just added for the sake of completeness. If $\operatorname{char}(F)\neq 2$ we have already proved in Step 1 that ${}_\varepsilon L_{-1}(F)=0$. Now if $p>2$,

$$ {}_\varepsilon L_{-p}(F) \approx {}_\varepsilon V_{-p}(F) \approx {}_{-\varepsilon}U_{-p+1}(F) \approx {}_{-\varepsilon}L_{-p+2}(F), $$
$$ {}_\varepsilon L_{-2}(F) \approx {}_\varepsilon V_{-2}(F) \approx {}_{-\varepsilon}U_{-1}(F) \approx \operatorname{Coker}(K_0(F) \to {}_{-\varepsilon}L_0(F)). $$

Therefore ${}_\varepsilon L_{-p}(k) \approx {}_\varepsilon L_{-p}(k_0)$ by induction on p. The exact sequence (E_n), written for negative values of n, will then imply ${}_\varepsilon\bar{L}_{-p}(k) \approx {}_\varepsilon\bar{L}_{-p}(k_0)$. If $\operatorname{char}(k)=2$, the arguments are completely analogous when replacing the groups K, L, U and V by $\tilde{K}$, $\tilde{L}$, $\tilde{U}$ and $\tilde{V}$.

Step 3. *The isomorphism* ${}_\varepsilon\bar{L}_{n-1}(k) \approx {}_\varepsilon\bar{L}_{n-1}(k_0)$ *implies the isomorphism*

$$ {}_{-\varepsilon}\bar{L}_{n+1}(k) \approx {}_{-\varepsilon}\bar{L}_{n+1}(k_0). $$

This step, together with Step 1 will of course finish the proof of the theorem. Let us write the two exact sequences

$$ \begin{array}{ccccccccc}
{}_\varepsilon\bar{L}_n(k) & \longrightarrow & \bar{K}_n(k) & \longrightarrow & {}_\varepsilon\bar{V}_{n-1}(k) & \longrightarrow & {}_\varepsilon\bar{L}_{n-1}(k) & \longrightarrow & \bar{K}_{n-1}(k) \\
\uparrow & & \uparrow & & \uparrow & & \uparrow\approx & & \uparrow \\
{}_\varepsilon\bar{L}_n(k_0) & \longrightarrow & \bar{K}_n(k_0) & \longrightarrow & {}_\varepsilon\bar{V}_{n-1}(k_0) & \longrightarrow & {}_\varepsilon\bar{L}_{n-1}(k_0) & \longrightarrow & \bar{K}_{n-1}(k_0)
\end{array} $$

The main theorem of Suslin implies that the maps $K_n(k_0) \to K_n(k)$ and $K_{n-1}(k_0) \to K_{n-1}(k)$ are isomorphisms. If, at the beginning of his proof of the main theorem, we take for R the functor ${}_{\varepsilon}\bar{V}_{n-1}$, we see that the map ${}_{\varepsilon}\bar{V}_{n-1}(k_0) \to {}_{\varepsilon}\bar{V}_{n-1}(k)$ is injective. The surjectivity of this map is obvious by diagram chasing. Using the fundamental theorem of Hermitian K-theory, we deduce that ${}_{-\varepsilon}\bar{U}_n(k_0) \approx {}_{-\varepsilon}\bar{U}_n(k)$. We have now two other exact sequences

$$\begin{array}{ccccccccc}
{}_{-\varepsilon}\bar{U}_{n+1}(k) & \longrightarrow & \bar{K}_{n+1}(k) & \longrightarrow & {}_{-\varepsilon}\bar{L}_{n+1}(k) & \longrightarrow & {}_{-\varepsilon}\bar{U}_n(k) & \longrightarrow & \bar{K}_n(k) \\
\uparrow & & \uparrow & & \uparrow & & \uparrow \approx & & \uparrow \\
{}_{-\varepsilon}\bar{U}_{n+1}(k_0) & \longrightarrow & K_{n+1}(k_0) & \longrightarrow & {}_{-\varepsilon}\bar{L}_{n+1}(k_0) & \longrightarrow & {}_{-\varepsilon}\bar{U}_n(k_0) & \longrightarrow & \bar{K}_n(k_0)
\end{array}$$

and again ${}_{-\varepsilon}\bar{L}_{n+1}(k_0) \approx {}_{-\varepsilon}\bar{L}_{n+1}(k)$ by the same argument as above (take for R the functor ${}_{-\varepsilon}\bar{L}_{n+1}$).

Corollary. *Let us assume* $\mathrm{char}(k) > 0$. *Then:*

$$H_*(O(k); \mathbf{Z}/q) \approx H_*(Sp(k); \mathbf{Z}/q) \approx \mathbf{Z}/q[p_1, p_2, \ldots]$$

for q *odd and* $\mathrm{char}(k) \nmid q$, *where* p_i *are homology classes in dimension* $4i$.

$$H_*(O(k); \mathbf{Z}/q) = H_*(Sp(k); \mathbf{Z}/q) = \mathbf{Z}/q$$

if q *is odd and is a power of the characteristic.*

$$H_*(O(k); \mathbf{Z}/2) \approx \mathbf{Z}/2[w_1, w_2, \ldots]$$

where $w_i \in H_i(O(k); \mathbf{Z}/2)$ *if* $\mathrm{char}(k) \neq 2$.

$$H_*(Sp(k); \mathbf{Z}/2) \approx \mathbf{Z}/2)[p_1, p_2, \ldots]$$

where $p_i \in H_{4i}(Sp(k); \mathbf{Z}/2)$ *if* $\mathrm{char}(k) \neq 2$.

The corollary follows from the main theorem of this appendix and the computation of $H_*(O(k_0))$ and $H_*(Sp(k_0))$ made by E. Friedländer [1]. The work of K. Vogtman [5] on the stability of the homology of the orthogonal group enables us to write explicitly some homology groups $H_i(O_{2n}(k); \mathbf{Z}/q)$. More precisely, $H_i(O_{2n}(k); \mathbf{Z}/q) \approx H_i(O(k); \mathbf{Z}/q)$ if $n \geqq 3i$.

References

1. Friedländer, E.: Unstable K-theories of the algebraic closure of a finite field. Comment. Math. Helvetici **50**, 145–154 (1975)
2. Karoubi, M.: Théorie de Quillen et homologie du groupe orthogonal. Ann. of Math. **112**, 207–257 (1980)
3. Karoubi, M.: Le theorème fondamental de la K-théorie hermitienne. Ann. of Math. **112**, 259–282 (1980)
4. Karoubi, M.: La périodicité de Bott en K-théorie générale. Ann. Scient. Ec. Norm. Sup. 4^e série **4**, 63–95 (1971)
5. Vogtmann, K.: Spherical posets and homology stability for $O_{n,n}$. Topology **20**, 119–132 (1981)

Oblatum 9-II-1983

Invent. math. 73, 251-269 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

Growth Series of Finite Extensions of $\mathbb{Z}^n$ are Rational

M. Benson

Department of Mathematical Sciences, University of Minnesota, Duluth, MN 55812, USA

§ 1. Introduction

Let G be a group which is generated as a monoid by the finite subset S; that is, any element of G can be written as a word formed from S without using inverses. Let $\Sigma(S)$ be the set of all words from S. For any $\sigma \in \Sigma(S)$, let $\bar{\sigma}$ be the element of G that it represents. By assigning a positive integer weight $\omega(s)$ to each element $s \in S$, we can define the *weighted length* $\omega(\sigma)$ of any word $\sigma \in \Sigma(S)$ as follows: if $\sigma = s_1 s_2 \ldots s_k$, then $\omega(\sigma) = \omega(s_1) + \omega(s_2) + \ldots + \omega(s_k)$. Any element $g \in G$ is defined to have weighted length $\omega(g)$ equal to the minimum of the weighted lengths of all words in $\Sigma(S)$ representing g. Any word of length $\omega(g)$ in $\Sigma(S)$, which represents g, will be called a *minimal word* representing g. The most important case is when the weights are all 1. Then, ω is called the *standard length* function and is denoted by l_S.

The *growth series* $p^{\omega}_{G,S}(t)$ is defined by the formula

$$p^{\omega}_{G,S}(t) = \sum_{N=0}^{\infty} g^{\omega}_{G,S}(N) t^N, \tag{1.1}$$

where $g^{\omega}_{G,S}(N)$ is the number of elements of G having weighted length equal to N. Computations by L. Solomon [12], N. Bourbaki [3], J. Cannon [5], and P. Wagreich [16] have shown that $p^{\omega}_{G,S}(t)$ often can be expressed as a rational function of t, at least using their specific choices of generating sets. J-P. Serre [11] (p. 112) has asked about the location of the zeroes and poles of growth series, singling out the compact hyperbolic and Euclidean groups as being especially interesting. Recently, J. Cannon [4] has shown that the growth series associated with certain isometry groups of hyperbolic space are always rational functions, no matter what set of generators S is chosen.

The object of this paper is to show that growth series are rational in the Euclidean case, as well. More specifically, we will show:

(1.2) **Theorem.** *Suppose G contains a subgroup of finite index isomorphic to $\mathbb{Z}^n$. Then, for any finite generating set S and weighted length ω, $p^{\omega}_{G,S}(t)$ is a rational function of t with $t=1$ as a pole of order n. The other poles of $p^{\omega}_{G,S}(t)$ are roots of unity.*

Underlying the proof of this result is an algorithm that allows computation of the growth series of any such G, although in many cases, it is not particularly easy to carry out.

In the special case $G=\mathbb{Z}^n$ and $\omega=l_S$, we can show that $t=1$ is the only pole of $p_{\mathbb{Z}^n,S}(t)$. The proof for this special case was obtained independently of an earlier result of D. Klarner [7, 8].

Recently P. Wagreich has given a very short proof of Theorem (1.2) when G is a finitely generated Abelian group. Let F_i be the submodule of the group ring $k[G]$ (k any field) generated by the elements of G having weighted length $\leqq i$. He notes that $p^{\omega}_{G,S}(t)$ is just the Poincaré series of the graded, finite-type k-algebra $\bigoplus_{i=0}^{\infty} F_{i+1}/F_i$. Then the standard theory of Poincaré series applies, e.g. [1] (Theorem 11.1, p. 117).

Alternately one may study the cumulative growth function

$$\gamma^{\omega}_{G,S}(N)=\sum_{i=0}^{N} g^{\omega}_{G,S}(i).$$

In view of the remark of J. Cannon [4] (Lemma 8.4), Theorem (1.2) may be restated. It says that $\gamma^{\omega}_{G,S}(N)$ is *periodically polynomial* of degree n. In particular, when G is a finitely generated abelian group and $\omega=l_S$, then $\gamma^{\omega}_{G,S}(N)$ is *strictly polynomial* of degree n.

This agrees with results of H. Bass [2] (Theorem 2, p. 608), which show that if G contains a finitely generated nilpotent subgroup H of finite index in G which has lower central series:

$$\{1\}\subseteq H_{p+1}\subseteq H_p\subseteq\ldots\subseteq H_2\subseteq H_1=H;$$

then $\gamma^{\omega}_{G,S}(N)$ grows like a polynomial of degree $n=\sum_{i=1}^{p} i\cdot\operatorname{rank}(G_i/G_{i+1})$.

Other results concerning $\gamma_{G,S}(N)$ have been given by M. Gromov [6], J. Milnor [9, 10], J. Tits [15], and J. Wolf [17]. None of these papers consider the case when the weighting is other than the standard length but most of the results concerning polynomial growth will carry over once the following lemma is established:

(1.3) **Lemma.** *Suppose $\gamma^{\omega}_{G,S}(N)$ grows like a polynomial of degree n, then for any weighting ω' and set of generators S', $\gamma^{\omega'}_{G,S'}(N)$ grows like a polynomial of degree n as well.*

This lemma has been proved in [2] (Lemma 1, p. 604) for the case when $\omega=l_S$ and $\omega'=l_{S'}$. The proof given there easily extends to our case as well.

In the case $G=\mathbb{Z}^n$, Theorem (1.2) is proved by putting $\Sigma(S)$ in one-one correspondence with $\mathbb{Z}_+^{\operatorname{card}(S)}$, and selecting a subset V^{ε} of minimal words, one for each element of G. We show that V^{ε} has the structure of a finite disjoint union of products of intervals in $\mathbb{Z}_+$. The growth series $p_{\mathbb{Z}^n,S}(t)$ is then just the generating function counting the points of V along the hyperplane slices perpendicular to the *weight vector*. A simple formula for this generating function can be given, which expresses it as a rational function with a denominator of $\prod_{s\in S}(1-t^{\omega(s)})$.

The general case is more complicated. The first step is to replace S by a larger generating set $S' \supseteq S$ and ω by ω', so that $p^{\omega}_{G,S} = p^{\omega'}_{G,S'}$. With the new generating set we can show that every element of G can be represented by a minimal word of one of a finite number of word *patterns* $\pi_1, \pi_2, \ldots, \pi_M$. We divide G into a finite disjoint union $\bigcup_{i=1}^{M} G_{\pi_i}$, where each subset G_{π_i} consists of elements of G which can be represented by a minimal word having the pattern π_i. We show that each of these subsets G_{π_i} is in one-one correspondence with a subset of $\mathbb{Z}^m_+$ which is defined by linear inequalities and congruences. The growth function $p^{\omega}_{G,S}(t)$ in this case is the sum of the generating functions counting lattice points along hyperplane slices of the G_{π_i}. By a slicing procedure we can show that each of these generating functions is a rational function of t with cyclotomic factors in the denominator, so $p^{\omega}_{G,S}(t)$ is as well.

The author wishes to thank Phil Wagreich for his encouragement and advice while working on this problem.

§ 2. Patterns for Words in $\Sigma(S)$

Suppose that H is a subgroup of finite index in G isomorphic to $\mathbb{Z}^n$. It is well known that the *core* of H is a *normal* subgroup of finite index in G which is contained in H. It follows that it is also isomorphic to $\mathbb{Z}^n$.

From now on we consider $\mathbb{Z}^n$ as being a *normal* subgroup of G of finite index $d = [G:\mathbb{Z}^n]$. Let S be a finite subset which generates G as a monoid and

$$\begin{aligned} S \cap \mathbb{Z}^n &= \{\varkappa_1, \varkappa_2, \ldots, \varkappa_r\} \\ S \setminus \mathbb{Z}^n &= \{\gamma_1, \gamma_2, \ldots, \gamma_s\}. \end{aligned} \tag{2.1}$$

Finally let ω be a weighted length function as defined in § 1.

Any $\pi \in \Sigma(S \setminus \mathbb{Z}^n)$ will be called a *pattern* for $\Sigma(S)$. If $\sigma \in \Sigma(S)$, $\sigma = s_1 s_2 \ldots s_k$, then set $c(\sigma) = \operatorname{card}\{s_i | s_i \notin \mathbb{Z}^n\}$. Suppose that $s_{i_1}, s_{i_2}, \ldots, s_{i_{c(\sigma)}}$ are those s_i for which $s_i \notin \mathbb{Z}^n$. Then $\pi(\sigma) = s_{i_1} s_{i_2} \ldots s_{i_{c(\sigma)}}$ is the *pattern of* σ. Given a pattern π, the set of all words patterned after π will be denoted by $\Sigma(S)^{\pi}$. Clearly $\Sigma(S) = \coprod_{\pi} \Sigma(S)^{\pi}$.

Suppose $\pi = \gamma_{\iota_1} \gamma_{\iota_2} \ldots \gamma_{\iota_k}$ is some pattern. Any word in $\Sigma(S)^{\pi}$ can be rewritten into the form

$$w^{\pi} = \varkappa_1^{v_1} \varkappa_2^{v_2} \ldots \varkappa_r^{v_r} \gamma_{\tau_1} \varkappa_1^{v_{r+1}} \varkappa_2^{v_{r+2}} \ldots \varkappa_r^{v_{2r}} \gamma_{\iota_2} \ldots \gamma_{\iota_k} \varkappa_1^{v_{kr+1}} \ldots \varkappa_r^{v_{(k+1)r}} \tag{2.2}$$

by permuting adjacent commuting generators $\varkappa_1, \varkappa_2, \ldots, \varkappa_r$. Clearly this rearrangement changes neither the element of G that it represents nor its weighted length. Any word that is in the form (2.2) will be called *arranged*. Let $W^{\pi} \subset \Sigma(S)^{\pi}$ be the subset of arranged words. The elements of W^{π} are in one-one correspondence with vectors in $\mathbb{Z}^m_+$, that is, the $m = (k+1)r$ tuples with non-negative integer coordinates. Indeed we will freely identify any $w^{\pi} \in W^{\pi}$ in (2.2) with the corresponding vector in $\mathbb{Z}^m_+$, $(w_1, w_2, \ldots, w_m)$.

The weighted length of some $w^{\pi} = (w_1, w_2, \ldots, w_m) \in W^{\pi}$ is given by the formula

$$\omega(w^\pi)=\sum_{i=0}^{k}\sum_{j=1}^{r}\omega(\varkappa_j)v_{ir+j}+\sum_{j=1}^{k}\omega(\gamma_{i_i}) \tag{2.3}$$

If we let $A^\pi_{n+1}=(\omega(\varkappa_1),\omega(\varkappa_2),\ldots,\omega(\varkappa_r),\omega(\varkappa_1),\ldots)\in\mathbb{Z}^m_+$ and $B^\pi_{n+1}=\sum_{i=1}^{k}\omega(\gamma_{i_i})\in\mathbb{Z}_+$, then (2.3) can be written in the more compact form

$$\omega(w^\pi)=A^\pi_{n+1}\cdot w^\pi+B^\pi_{n+1} \tag{2.4}$$

§ 3. A Standard Form for Elements of G

Given any $w^\pi_1, w^\pi_2\in W^\pi$, we would like to check whether $\overline{w^\pi_1}=\overline{w^\pi_2}$. In order to do this we introduce a standard form for elements of G. Choose a set of coset representatives $g_1, g_2, \ldots, g_d$ so that $G=\mathbb{Z}^n g_1\cup\mathbb{Z}^n g_2\cup\ldots\cup\mathbb{Z}^n g_d$. Then each element of G may be expressed uniquely in the form $(c_1, c_2, \ldots, c_n)g_i$, for some $c_1, c_2, \ldots, c_n\in\mathbb{Z}$, and some $i=1,\ldots,d$.

In order to express $\overline{w^\pi}$ in terms of this standard form we need to introduce some structure constants for the group. Let

$$\begin{aligned}\varkappa_i&=(z_{i1}, z_{i2}, \ldots, z_{in}), \qquad i=1,2,\ldots,r\\ \gamma_k e_i\gamma_k^{-1}&=(g^k_{i1}, g^k_{i2}, \ldots, g^k_{in}), \qquad i=1,2,\ldots,n;\ k=1,2,\ldots,s\end{aligned} \tag{3.1}$$

where $e_1, e_2, \ldots, e_n$ are the standard coordinate vectors in $\mathbb{Z}^n$. Let $Z=(z_{ij})$ be the $r\times n$ matrix, and $G^k=(g^k_{ij})$ be the $n\times n$ matrices formed from these constants. Then we have the more general relationships

$$\begin{aligned}\varkappa_1^{u_1}\varkappa_2^{u_2}\ldots\varkappa_n^{u_n}&=(u_1,u_2,\ldots,u_n)Z\\ \gamma_k(u_1,u_2,\ldots,u_n)\gamma_k^{-1}&=(u_1,u_2,\ldots,u_n)G^k.\end{aligned} \tag{3.2}$$

We also need to express $\bar\pi$ in terms of this standard form, say

$$\bar\pi=(B^\pi_1, B^\pi_2, \ldots, B^\pi_n)g_\pi \tag{3.3}$$

where $g_\pi\in\{g_1, g_2, \ldots, g_d\}$. With this it is easy to see that any $w^\pi\in W^\pi$ in the form given in (2.2) represents

$$\begin{aligned}\overline{w^\pi}=\{&(w_1,\ldots,w_r)Z+(w_{r+1},\ldots,w_{2r})ZG^{i_1}+\ldots\\ &+(w_{kr+1},\ldots,w_{(k+1)r})ZG^{i_r}G^{i_{r-1}}\ldots G^{i_1}+(B^\pi_1,B^\pi_2,\ldots,B^\pi_n)\}g_\pi.\end{aligned} \tag{3.4}$$

Let A^π_i be the m-tuple formed by taking the i^{th} columns of the matrix formed by lining up Z, ZG^{i_1}, $ZG^{i_2}G^{i_1}$, …, $ZG^{i_r}G^{i_{r-1}}\ldots G^{i_1}$ one by one vertically. Then (3.4) may be written more compactly as

$$\overline{w^\pi}=\{(A^\pi_1\cdot w^\pi, A^\pi_2\cdot w^\pi, \ldots, A^\pi_n\cdot w^\pi)+(B^\pi_1, B^\pi_2, \ldots, B^\pi_n)\}g_\pi \tag{3.7}$$

In particular, this shows that all words in W^π represent elements of the same coset $\mathbb{Z}^n g_\pi$.

§ 4. The π-Minimal Words in $\Sigma(S)$

We are going to construct a subset $V^\pi \subseteq W^\pi$ satisfying the properties

(4.1)
a) Every v^π is minimal in the sense that if there exists $\sigma \in \Sigma(S)^\pi$ with $\overline{v^\pi} = \bar{\sigma}$, then $\omega(v^\pi) \leqq \omega(\sigma)$.
b) V^π is "complete" in the sense that for any $\sigma \in \Sigma(S)^\pi$ there exists $v^\pi \in V^\pi$ such that $\overline{v^\pi} = \bar{\sigma}$.
c) No two elements of V^π represent the same element of G.

In order to do this we introduce a well-ordering on W_c^π, the set of all w^π for which $\overline{w^\pi} = (c_1, c_2, \ldots, c_n) g_\pi$ where $c = (c_1, c_2, \ldots, c_n)$.

Suppose that the $\mathbb{R}$-rank of the set of vectors $A_1^\pi, \ldots, A_{n+1}^\pi$ introduced in § 2 and § 3 is p. Choose $m-p$ linearly independent vectors $A_{n+2}^\pi, \ldots, A_N^\pi$, $N = n+1+m-p$, from the set of standard coordinate vectors in $\mathbb{Z}^m$ so that together $A_1^\pi, \ldots, A_N^\pi$ have maximal $\mathbb{R}$-rank.

For any two $w_1^\pi, w_2^\pi \in W_c^\pi$, define $w_1^\pi >_\pi w_2^\pi$ if either $w_1^\pi = w_2^\pi$ or else for some i, $n+1 \leqq i \leqq N$,

$$A_{n+1}^\pi \cdot w_1^\pi = A_{n+1}^\pi \cdot w_2^\pi, \qquad A_{n+2}^\pi \cdot w_1^\pi = A_{n+2}^\pi \cdot w_2^\pi, \ldots, A_{i-1}^\pi \cdot w_1^\pi = A_{i-1}^\pi \cdot w_2^\pi \quad A_i^\pi \cdot w_1^\pi > A_i^\pi \cdot w_2^\pi. \tag{4.2}$$

Because $A_1^\pi, A_2^\pi, \ldots, A_N^\pi$ have maximal $\mathbb{R}$-rank, $>_\pi$ is a partial order. In fact it is a total order on W_c^π which is compatible with ω in the sense that if $w_1^\pi >_\pi w_2^\pi$, then $\omega(w_1^\pi) \geqq \omega(w_2^\pi)$.

(4.3) **Proposition.** *The ordering $>_\pi$ is a well-ordering of W_c^π which is compatible with ω.*

Proof. Suppose that $V \subseteq W_c^\pi$, nonempty, does not have a least element. We can construct a descending chain of distinct elements $v_1^\pi >_\pi v_2^\pi >_\pi \ldots$ in V. Then we have $A_{n+1}^\pi \cdot v_1^\pi \geqq A_{n+1}^\pi \cdot v_2^\pi \geqq \ldots$. But A_{n+1}^π and $v_1^\pi, v_2^\pi, \ldots$ have non-negative coordinates so their dot products are non-negative as well. Hence this chain of inequalities must eventually become stationary, say $A_{n+1}^\pi \cdot v_{i_1}^\pi = A_{n+1}^\pi \cdot v_{i_1+1}^\pi = \ldots$ From then on $A_{n+2}^\pi \cdot v_{i_1}^\pi \geqq A_{n+2}^\pi \cdot v_{i_1+1}^\pi \geqq \ldots$. For the same reasons these must also become equalities. At some point $A_i^\pi \cdot v_{i_{N-n}}^\pi = A_i^\pi \cdot v_{i_{N-n+1}}^\pi = \ldots$ for $i = 1, 2, \ldots, N$. But this implies $v_{i_{N-n}}^\pi = v_{i_{N-n+1}}^\pi = \ldots$, a contradiction. Hence every nonempty subset of W_c^π has a least element.

The set of all such least elements will be denoted by

$$V^\pi = \{v^\pi \mid v^\pi \in W_c^\pi \text{ for some } c \text{ and } v^\pi \text{ minimal with respect to } >_\pi \text{ in } W_c^\pi\} \tag{4.4}$$

Any element of V^π will be called a *π-minimal* word. It is clear that the set V^π fulfills properties (4.1) a)–c).

§ 5. Weak Hyperplane Orderings of $\mathbb{Z}^m$

Suppose now that $H_1, H_2, \ldots, H_m \in \mathbb{Z}^m$ are linearly independent vectors. The *weak ordering* of rank q obtained from the vectors $H_1, H_2, \ldots, H_m$ is defined as

follows: For any $v, w \in \mathbb{Z}^m$, $v \underset{H,q}{>} w$ if

$$\begin{aligned} &H_1 \cdot v \geqq H_1 \cdot w, \quad H_2 \cdot v \geqq H_2 \cdot w, \ldots, H_q \cdot v \geqq H_q \cdot w \\ &H_{q+1} \cdot v = H_{q+1} \cdot w, \quad H_{q+2} \cdot v = H_{q+2} \cdot w, \ldots, H_m \cdot v = H_m \cdot w. \end{aligned} \tag{5.1}$$

This is a partial order, but not every two elements are comparable. Let

$$\mathrm{Cut}_{H,q}(c_1, c_2, \ldots, c_m) = \{v \in \mathbb{Z}^m \mid v \underset{H,q}{>} (c_1, c_2, \ldots, c_m)\}. \tag{5.2}$$

Although $\underset{H,q}{>}$ does not well-order $\mathrm{Cut}_{H,q}(c_1, c_2, \ldots, c_m)$, any nonempty subset of it is bounded below by a finite number of minimal elements.

(5.3) **Proposition.** *Every nonempty subset* $W \subseteq \mathrm{Cut}_{H,q}(c_1, c_2, \ldots, c_m)$ *contains a finite subset* $\{w_1, w_2, \ldots, w_l\}$ *with the property that for any* $w \in W$, *there is an* i *for which* $w \underset{H,q}{>} w_i$.

Proof. We proceed by induction on q. If $q=0$, then W consists of only the one point $(c_1, c_2, \ldots, c_m)$, so in this case it is trivial. Now assume that it is true for $q=k$, and consider the case $q=k+1$. By using an argument like the one given in the proof of Proposition (4.3), we conclude that there must be a minimal element $w_0 \in W$ (although it need not be a least element). w_0 will bound all elements of W below except those in the set

$$\bigcup_{i=0}^{k+1} \bigcup_{j=c_i}^{H_i \cdot w_0 - 1} \{w \in W \mid H_i \cdot w = j\}.$$

Each $\{w \in W \mid H_i \cdot w = j\}$ is a subset of $\mathrm{Cut}_{H',k}(c_1, \ldots, c_{i-1}, c_{k+1}, c_{i+1}, \ldots, c_k, c_i, c_{k+2}, \ldots, c_m)$, where H' is the sequence of vectors $H_1, H_2, \ldots, H_{i-1}, H_{k+1}, H_{i+1}, \ldots, H_k, H_i, H_{k+2}, \ldots, H_m$. Since we have assumed that the result is true for $q=k$, the sets $\{w \in W \mid H_i \cdot w = j\}$ must each be either empty or else be bounded below by a finite number of minimal elements. The union of all of the minimal elements in each of these sets along with w_0 is a finite set of minimal elements which bounds W below.

From now on we will be interested in weak orderings of maximal rank, so we will write $\underset{H}{>}$ to denote the weak ordering of rank m obtained from the vectors $H_1, H_2, \ldots, H_m$.

§ 6. Fundamental π-Translations

Suppose $\tau \in \mathbb{Z}^m$ is a vector with the property that for some i, $n+1 \leqq i \leqq N$, we have

$$\begin{aligned} &A_1^\pi \cdot \tau = \ldots = A_n^\pi \cdot \tau = \ldots = A_{i-1}^\pi \cdot \tau = 0 \\ &A_i^\pi \cdot \tau > 0. \end{aligned} \tag{6.1}$$

Then we will call τ a *π-translation*. The set of all such τ will be denoted by $\mathscr{T}^\pi$. The reason for their name is that if $w^\pi \in W^\pi$ is a word representing c with pattern π, then $w^\pi + \tau$ also represents c with pattern π, whenever $\tau \in \mathscr{T}^\pi$.

Note that if $w_1^\pi, w_2^\pi \in W_c^\pi$, then $w_1^\pi \underset{\pi}{>} w_2^\pi$ is and only if $w_1^\pi - w_2^\pi \in \mathscr{T}^\pi$.

(6.2) *Remark.* If W_c^π is nonempty, then it may be decomposed as $W_c^\pi = w_c^\pi + \mathscr{T}^\pi$ where w_c^π is the π-minimal element of W_c^π.

We are now going to use the results of §5 to construct a finite set $T^\pi \subset \mathscr{T}^\pi$ of π-translations which we need in order to give an alternate construction of V^π.

For each $I \subseteq \{1, 2, \ldots, m\}$, let $H_1^I, H_2^I, \ldots, H_m^I$ be the sequence of vectors defined by

$$H_i^I = \begin{cases} e_i & \text{if } i \in I \\ -e_i & \text{if } i \notin I. \end{cases} \tag{6.3}$$

Weakly order $\mathscr{T}^\pi$ by $\underset{H^I}{>}$. Using Proposition (5.3), there exists a finite subset $T_I^\pi \subset \mathscr{T}^\pi$ which bounds $\mathscr{T}^\pi \cap \mathrm{Cut}_{H^I, m}(0, 0, \ldots, 0)$ below. Let $T^\pi = \bigcup_{I \subset \{1, 2, \ldots, m\}} T_I^\pi$. The elements of T^π will be called *fundamental π-translations.*

(6.4) **Proposition.** $V^\pi = \bigcap_{\tau \in T^\pi} (\mathbb{Z}_+^m \setminus (\tau + \mathbb{Z}_+^m))$.

Proof. Suppose first that $w \in V^\pi$, but $w \notin \bigcap_{\tau \in T^\pi} (\mathbb{Z}_+^m \setminus (\tau + \mathbb{Z}_+^m))$. Then for some $\tau \in T^\pi$ we have $w - \tau \in \mathbb{Z}_+^m$. This means that $w - \tau$ is a word representing the same element of G as w does and $w \underset{\pi}{>} w - \tau$. So w is not π-minimal, and $w \notin V^\pi$, a contradiction.

Now suppose $w \in \bigcap_{\tau \in T^\pi} (\mathbb{Z}_+^m \setminus (\tau + \mathbb{Z}_+^m))$ and $w \notin V^\pi$. Write $w = w_c^\pi + \tau$ for some $\tau \in \mathscr{T}^\pi$. Let I be the set of indices of the nonnegative coordinates of τ. Choose some $\tau_0 \in T_I^\pi$ for which $\tau \underset{H^I}{\geq} \tau_0$. Then $w - \tau_0 \in \mathbb{Z}_+^m$ (since if $i \in I$, then $w^{(i)} - \tau_0^{(i)} = w_c^{\pi(i)} + (\tau^{(i)} - \tau_0^{(i)}) \geqq w_c^{\pi(i)} + \tau^{(i)} = w^{(i)} \geqq 0$ and if $i \notin I$, then $w^{(i)} - \tau_0^{(i)} \geqq w^{(i)} \geqq 0$). Hence $w \in \tau_0 + \mathbb{Z}_+^m$, so $w \notin \mathbb{Z}_+^m \setminus (\tau_0 + \mathbb{Z}_+^m)$, a contradiction.

§ 7. Positive Interval Sets

A basic positive interval set will be any set of the form

$$I_j = [a_1^{(j)}, b_1^{(j)}) \times [a_2^{(j)}, b_2^{(j)}) \times \ldots \times [a_m^{(j)}, b_m^{(j)}) \cap \mathbb{Z}_+^m \tag{7.1}$$

where the $a_i^{(j)}$ are nonnegative integers, the $b_i^{(j)}$ are nonnegative integers or ∞, and $a_i^{(j)} \leqq b_i^{(j)}$. Any finite disjoint union of these will be called a *positive interval set.*

The *dimension* of any basic positive interval set I_j as in (7.1) is defined by

$$\dim I_j = \mathrm{card}\{i \mid b_i^{(j)} = \infty\}. \tag{7.2}$$

For any positive interval set represented as a finite union of basic positive interval sets, the dimension is the maximum of the dimensions of the basic positive interval sets in the union. It is easy to see that this definition is independent of the representation.

For any basic positive interval set I_j as in (7.1), we also define

$$c_i(I_j) = \begin{cases} a_j^{(i)} & b_j^{(i)} = \infty \\ b_j^{(i)} & \text{otherwise.} \end{cases} \tag{7.3}$$

In order to extend this to any positive interval set I, we first define $c_i(I, \mathscr{I})$ as the maximum of the $c_i(I_j)$ for all I_j appearing in the finite disjoint union $\mathscr{I}$ representing I. Then $c_i(I)$ is the minimum of the $c_i(I, \mathscr{I})$ over all $\mathscr{I}$ representing I.

The following proposition may be easily verified:

(7.4) **Proposition.** *If $I_1, I_2, \ldots, I_k$ are positive interval sets, then $I_1 \cap I_2 \cap \ldots \cap I_k$ is a positive interval set with*

$$c_i(I_1 \cap I_2 \cap \ldots \cap I_k) \leqq \max(c_i(I_1), c_i(I_2), \ldots, c_i(I_k)).$$

§8. V^π is a Positive Interval Set

This section is devoted to proving the following result

(8.1) **Proposition.** *V^π is a positive interval set with*

a) $c_i(V^\pi) \leqq \max_{\tau \in T^\pi} \tau_+^{(i)}$, where $\tau_+^{(i)} = \max(\tau^{(i)}, 0)$,

b) $\dim V^\pi \leqq n$.

Proof. Suppose $\tau \in \mathbb{Z}^m$, then $\mathbb{Z}_+^m \setminus (\tau + \mathbb{Z}_+^m)$ can be expressed as a disjoint union of basic positive interval sets

$$\begin{aligned} &[0, \tau_+^{(1)}) \times [0, \infty) \times \ldots \times [0, \infty) \cap \mathbb{Z}_+^m \\ &\cup [\tau_+^{(1)}, \infty) \times [0, \tau_+^{(2)}) \times \ldots \times [0, \infty) \cap \mathbb{Z}_+^m \\ &\qquad \ldots\ldots\ldots \\ &\cup [\tau_+^{(1)}, \infty) \times [\tau_+^{(2)}, \infty) \times \ldots \times [0, \tau_+^{(m)}) \cap \mathbb{Z}_+^m . \end{aligned} \tag{8.2}$$

Then $c_i(\mathbb{Z}_+^m \setminus (\tau + \mathbb{Z}_+^m)) \leqq \tau_+^{(i)}$. Combining Propositions (6.4) and (7.4) we see that V^π is a positive interval set with $c_i(V^\pi) \leqq \max_{\tau \in T^\pi} \tau_+^{(i)}$.

Now we show that $\dim V^\pi \leqq n$. For suppose not, then without loss of generality, assume that there exists a basic positive interval set $I_j \subseteq V^\pi$ in the form (7.1) with $b_1^{(j)} = b_2^{(j)} = \ldots = b_{n+1}^{(j)} = \infty$. The linear transformation $A^\pi = (A_1^\pi, A_2^\pi, \ldots, A_n^\pi)$ (see §3 for the construction of the A_j^π) must carry the $n+1$ standard coordinate vectors $e_1, e_2, \ldots, e_{n+1} \in \mathbb{Z}^m$ into the $n+1$ linearly independent vectors $A_1^\pi e_1, A^\pi e_2, \ldots, A^\pi e_{n+1}$. For some integers $c_1, c_2, \ldots, c_{n+1}$ (not all zero) $\sum_{i=1}^{n+1} c_i A^\pi e_i = 0$.

Write $(c_1, c_2, \ldots, c_{n+1}, 0, 0, \ldots, 0) = c_+ - c_-$, the difference of two distinct non-negative vectors c_+, $c_- \in \mathbb{Z}_+^m$. Then $A^\pi c_+ = A^\pi c_-$. If $a_+ = (a_1, a_2, \ldots, a_m) + c_+$, $a_- = (a_1, a_2, \ldots, a_m) + c_-$, then a_+, $a_- \in V^\pi$ and $A^\pi a_+ + B^\pi = A^\pi a_- + B^\pi$ where $B^\pi = (B_1^\pi, B_2^\pi, \ldots, B_n^\pi)$ (see §3 for the construction of the B_i^π). This means that a_+, a_- are distinct words in V^π representing the same element of G, a contradiction. Hence $\dim V^\pi \leqq n$.

§ 9. Generating Functions on Positive Interval Sets

We are going to find an expression for the generating function which counts the lattice points in a positive interval set I lying on slices perpendicular to the positive weight vector $\omega=(\omega_1, \omega_2, \ldots, \omega_m)$. It is defined by the formula

$$p_I^{\omega}(t)=\sum_{N=0}^{\infty} \operatorname{card}\{(x_1, x_2, \ldots, x_n)\in I \mid \omega_1 x_1+\ldots+\omega_m x_m=N\}\, t^N. \tag{9.1}$$

For a basic positive interval set I_j as defined in (7.1), this generating function can be factored as

$$p_{I_j}^{\omega}(t)=\prod_{i=1}^{m}\left(t^{\omega_i a_i^{(j)}}+\ldots+t^{\omega_i(b_i^{(j)}-1)}\right) \tag{9.2}$$

where it is understood that if $b_i^{(j)}=\infty$, then the i^{th} term in the product is an infinite sum. In this case term may be replaced by $t^{\omega_i a_i^{(j)}}/1-t^{\omega_i}$.

More generally if $I=\coprod_{j=1}^{M} I_j$ is a union of basic positive interval sets, then the generating function $p_I^{\omega}(t)$ is the sum of the generating functions on each of the I_j.

If we bring in the quantities $\dim I$, $c_i(I)$, $i=1, 2, \ldots, m$, introduced in § 7, we can get a useful description of the generating function.

(9.3) **Proposition.** *If I is a positive interval set and ω a positive weight vector, then*

$$p_I^{\omega}(t)=\frac{q_I^{\omega}(t)}{(1-t)^{\dim I}\prod_{i=0}^{m}(1+t+\ldots+t^{\omega_i-1})}$$

where $q_I(t)$ is a polynomial with integer coefficients of degree less than or equal to $\sum_{i=1}^{m}\omega_i c_i(I)$. $q_I(t)$ is relatively prime to $1-t$.

§ 10. The Case $G=\mathbb{Z}^n$

When $G=\mathbb{Z}^n$, there is only one pattern, the empty word ε. The elements of G are in one-one correspondence with those in V^{ε} and

$$g_{\mathbb{Z}^n, S}^{\omega}(N)=\operatorname{card}\{x_1, \ldots, x_m)\in V^{\varepsilon} \mid \omega(\mathfrak{z}_1)\, x_1+\ldots+\omega(\mathfrak{z}_m)\, x_m=N\}. \tag{10.1}$$

In this case we can easily show that $\dim V^{\varepsilon}=n$. For if $\dim V^{\varepsilon}<n$, then V would be contained in a finite union of vector subspaces of $\mathbb{R}^m$, each having dimension $<n$. The linear transformation A^{ε} (see § 8) must carry this union of subspaces onto $\mathbb{R}^n$, because the image of V^{ε} is $\mathbb{Z}^n$ which spans $\mathbb{R}^n$ as a $\mathbb{R}$-vector space. This means $\mathbb{R}^n$ is equal to a finite union of vector spaces of dimension $<n$. This contradicts the irreducibility of $\mathbb{R}^n$. So $\dim V^{\varepsilon}=n$.

Using this remark along with Proposition (9.3), we get

(10.2) **Theorem.** *There exists a polynomial* $q^{\omega}_{\mathbb{Z}^n,S}(t)$ *with integer coefficients and degree less than or equal to* $\sum_{i=1}^{m} \omega_i \max_{\tau \in T^{\pi}} \tau_+^{(i)}$ *which is relatively prime to* $1-t$ *such that*

$$p^{\omega}_{\mathbb{Z}^n,S}(t) = \frac{q^{\omega}_{\mathbb{Z}^n,S}(t)}{(1-t)^n \prod_{i=1}^{m} (1+t+\ldots+t^{\omega_i-1})}.$$

The following examples demonstrate how some growth series of $\mathbb{Z}^n$ may be calculated.

(10.3) *Example.* Suppose $G=\mathbb{Z}$, $S=\{\pm 1\}$, and $\omega=(\omega_1,\omega_2)$. In the notation of §3, $A_1^{\varepsilon}=(1,-1)$, and $A_2^{\varepsilon}=(\omega_1,\omega_2)$. The set of ε-translations is

$$\mathscr{T}^{\varepsilon}=\{(z_1,z_2)\in\mathbb{Z}^2 \mid z_1-z_2=0,\ \omega_1 z_1+\omega_2 z_2>0\}=\{(z,z)\mid z\in\mathbb{Z}_+^2\}.$$

The single minimal ε-translation is $(1,1)$. It follows that $V=\mathbb{Z}_+^2\setminus(\mathbb{Z}_+^2+(1,1))$ $=(([0,1)\times[0,\infty))\cup([1,\infty)\times[0,1)))\cap\mathbb{Z}_+^2$. Thus

$$p^{\omega}_{\mathbb{Z},S}(t)=(1+t^{\omega_2}+\ldots)+(t^{\omega_1}+t^{2\omega_1}+\ldots)=\frac{1+\ldots+t^{\omega_1+\omega_2-1}}{(1-t)(1+\ldots+t^{\omega_1-1})(1+\ldots+t^{\omega_2-1})}.$$

In the special case when $\omega=l_S$, then $P_{\mathbb{Z},S}(t)=\dfrac{1+t}{1-t}$.

(10.4) *Example.* Suppose $G=\mathbb{Z}$, $S=\{2,3,-1\}$, and $\omega=l_S$. In the case A_1^{ε} $=(2,3,-1)$, $A_2^{\varepsilon}=(1,1,1)$, and we choose $A_3^{\varepsilon}=(1,0,0)$. Then ε-translations are

$$\begin{aligned}\mathscr{T}^{t}=&\{(z_1,z_2,z_3)\in\mathbb{Z}^3 \mid 2z_1+3z_2-z_3=0,\ z_1+z_2+z_3>0\}\\ &\cup\{(z_1,z_2,z_3)\in\mathbb{Z}^3 \mid 2z_1+3z_2-z_3=0,\ z_1+z_2+z_3=0,\ z_1>0\}.\end{aligned}$$

The ε-minimal translations can be tabulated as follows

I	T_{ε}^I
$\{1\}$	$\{(3,-2,0)\}$
$\{1,3\}$	$\{(1,0,2),(2,-1,1)\}$
$\{2,3\}$	$\{(0,1,3),(-1,1,1)\}$
$\{1,2,3\}$	$\{(1,0,2),(0,1,3)\}$.

For other $I\subseteq\{1,2,3\}$, T_{ε}^I is empty. Therefore

$$\begin{aligned}V^{\varepsilon}=&(\mathbb{Z}_+^3\setminus(\mathbb{Z}_+^3+(3,-2,0)))\cap(\mathbb{Z}_+^3\setminus(\mathbb{Z}_+^3+(1,0,2)))\\ &\cap(\mathbb{Z}_+^3\setminus(\mathbb{Z}_+^3+(2,-1,1)))\cap(\mathbb{Z}_+^3\setminus(\mathbb{Z}_+^3+(0,1,3)))\\ &\cap(\mathbb{Z}_+^3\setminus(\mathbb{Z}_+^3+(-1,1,1)))\\ =&([0,1)\times[0,1)\times[0,\infty))\cup([1,2)\times[0,1)\times[0,2))\\ &\cup([2,3)\times[0,1)\times[0,1))\cup([0,3)\times[1,\infty)\times[0,1))\cap\mathbb{Z}_+^3\end{aligned}$$

and

$$p_{\mathbb{Z},S}(t)=\frac{1}{1-t}+t(1+t)+t^2+(1+t+t^2)\frac{t}{1-t}=\frac{1+2t+2t^2-t^3}{1-t}.$$

(10.6) *Example.* Suppose $G=\mathbb{Z}^2$, $S=\{(1,2),(2,3),(-1,-1),(0,-2)\}$ and $\omega=l_S$. Then $A_1^\varepsilon=(1,2,-1,0)$, $A_2^\varepsilon=(2,3,-1,-2)$, $A_3^\varepsilon=(1,1,1,1)$, and we choose a fourth linearly independent vector $A_4^\varepsilon=(0,0,0,1)$. The set of ε-translations takes the form $\mathscr{T}^\varepsilon=\{(-Z+4w,z-2w,z,w)\in\mathbb{Z}^4 \mid z+3w>0\}\cup\{(7w,-5w,-3w,w)\mid w>0\}$. Then minimal ε-translations are given in the table below:

I	T_ε^I
$\{2,3\}$	$\{(-1,1,1,0)\}$
$\{1,4\}$	$\{(4,-2,0,1)\}$
$\{2,3,4\}$	$\{(0,2,4,1),(-1,1,1,0)\}$
$\{1,3,4\}$	$\{(4,-2,0,1),(3,-1,1,1),(2,0,2,1)\}$
$\{1,2,3,4\}$	$\{(0,2,4,1),(2,0,2,1)\}$.

For other choices of $I\subset\{1,2,3,4\}$, $T_\varepsilon^I=\emptyset$. In order to avoid the tedious intersection process to find a disjoint union of positive interval sets, we use Corollary (11.3) to conclude that $p_{\mathbb{Z}^2,S}(t)$ has the form $\frac{q(t)}{(1-t)^2}$, for some polynomial $q(t)$. Since $c_1=4$, $c_2=2$, $c_3=4$, $c_4=1$, the degree of $q(t)$ is no more than 11. By using a computer to count group elements was found that

$$p_{\mathbb{Z}^2,S}(t)=1+4t+9t^2+16t^3+25t^4+33t^5+41t^6+49t^7$$
$$+57t^8+65t^9+73t^{10}+81t^{11}+\ldots.$$

Multiplying this by $(1-t)^2$ we get $q(t)=1+2t+2t^2+2t^3+2t^4+t^5$. As a result

$$p_{\mathbb{Z}^2,S}(t)=\frac{1+2t+2t^2+2t^3+2t^4+t^5}{(1-t)^2}.$$

§ 11. Elementary Patterns

In the general case $\mathbb{Z}^n\subsetneq G$, there are infinitely many patterns $\pi\in\Sigma(G\backslash S)$. Instead of working with the words formed from S we will construct a larger generating set S^* that contains S. Let $S^*=\{\overline{s_1\ldots s_j}\mid s_1,\ldots,s_j\in S,\ 0<j\leqq d\}$. To each $s^*\in S^*$ we assign the weight $\omega(s^*)$, inducing a weighted length function ω^* on the set of words $\Sigma(S^*)$ and the group G. For this reason $p_{G,S}^\omega(t)=p_{G,S^*}^{\omega^*}(t)$.

Any pattern π for $\Sigma(S^*)$ with $c(\pi)\leqq d$ will be called an *elementary pattern.* Clearly there are only finitely many elementary patterns. We will show that every $g\in G$ can be represented by a minimal word in $\Sigma(S^*)$ with an elementary pattern.

The following lemma is key to showing that a minimal word in $\Sigma(S^*)$ can be transformed into a new minimal word with an elementary pattern. In it we use the quantity $\mu_L(\sigma)=\omega(\pi(\sigma))\cdot L+c(\sigma)$ defined for any $\sigma\in\Sigma(S^*)$, L a positive integer.

(11.1) **Lemma.** *Suppose $s^*,s^{*\prime}\in S^*\backslash\mathbb{Z}^n$ and $s^*s^{*\prime}$ is a minimal word representing $g\in G$. Then $s^*s^{*\prime}$ may be replaced by a minimal word $\sigma\in\Sigma(S^*)$ also representing g for which $\mu_L(s^*s^{*\prime})>\mu_L(\sigma)$ for any $L\geqq l_S(g)$.*

Before proving this we make a basic remark that is a consequence of the "pigeon hole principle."

(11.2) *Remark.* Any product $g_1 g_2 \dots g_k$, $k \geqq d$ of elements of G must contain a subproduct $g_{e+1} g_{e+2} \dots g_f \in \mathbb{Z}^n$.

Proof of Lemma (11.1). Express s^*, $s^{*\prime}$ as minimal words in S, say $s^* = s_1 s_2 \dots s_j$, $s^{*\prime} = s_{j+1} \dots s_k$. If $k < d$ then $s^{*\prime\prime} = s_1 s_2 \dots s_k$ is in S^* and is a minimal word representing g, with $\mu_L(s^{*\prime\prime}) = \omega(g) L < \omega(g) L + 2 = \mu_L(s^* s^{*\prime})$.

Otherwise $k \geqq d$. In this case apply Remark (11.2) to $s_1 s_2 \dots s_d$. Some substring in S^*, say $s^{*\prime\prime} = s_{e+1} s_{e+2} \dots s_f$ must represent an element of $\mathbb{Z}^n$. The word $s_1 \dots s_e s^{*\prime\prime} s_{f+1} \dots s_k$ is a minimal word in $\Sigma(S^*)$ representing g with

$$\mu_L(s_1 \dots s_e s^{*\prime\prime} s_{f+1} \dots s_k) = (\omega(g) - \omega(s^{*\prime\prime})) L + k + e - f + 1 \leqq \omega(g) L < \omega(g) L + 2 = \mu_L(s^* s^{*\prime})$$

where we have used the inequalities $\omega(s^{*\prime\prime}) \geqq 1$, $k \leqq L$, and $f < e$.

We are now ready to show that any $g \in G$ is represented by a minimal word $\sigma \in \Sigma(S^*)$ with $c(\sigma) \leqq d$.

(11.3) **Proposition.** *Suppose $g \in G$, then there exists a minimal word $\sigma \in \Sigma(S^*)$ representing g with an elementary pattern.*

Proof. Choose $\sigma \in \Sigma(S^*)$, $\sigma = s_1^* s_2^* \dots s_k^*$, a minimal word representing g, for which $\mu_L(\sigma) = \min\{\mu_L(\sigma') | \sigma' \in \Sigma(S^*)$ is a minimal word and $\bar{\sigma}' = g\}$. Here $L = l_S(g)$. We are going to show that $c(\sigma) \leqq d$.

Suppose instead that $c(\sigma) > d$. Then some of the s_i^* will not be in $\mathbb{Z}^n$, say $s_{i_1}^*, s_{i_2}^*, \dots, s_{i_{c(\sigma)}}^*$. None of these may be adjacent in the word σ, because otherwise we could use Lemma (11.1) to replace the pair by another minimal word with a smaller value of μ_L. The result would be a new minimal word for g with a smaller μ_L value, a contradiction. We now apply Remark (11.2) to the product $(s_{i_1}^* s_{i_1+1}^* \dots s_{i_2-1}^*)(s_{i_2}^* s_{i_2+1}^* \dots s_{i_3-1}^*) \dots (s_{i_d}^* s_{i_d+1}^* \dots s_{i_{d+1}-1}^*)$. Some subproduct, say $(s_{i_e}^* \dots s_{i_{e+1}-1}^*) \dots (s_{i_f}^* \dots s_{i_{f+1}-1}^*)$ must be in $\mathbb{Z}^n$. But then we may permute the terms $s_{i_e}^* \dots s_{i_f}^*$ and $s_{i_f+1}^* \dots s_{i_{f+1}-1}^*$ since both of them are in $\mathbb{Z}^n$ and hence commute. The resulting word will still be a minimal word representing g and have the same μ_L value, but now $s_{i_f}^*$ and $s_{i_{f+1}}^*$ will be adjacent and neither are in $\mathbb{Z}^n$. According to the argument given above, this leads to a contradiction. Therefore $c(\sigma) \leqq d$.

§ 12. Construction of the U^π

Let P be the set of elementary patterns. Then P is a finite set of patterns with the property that every $g \in G$ is represented by a minimal word $\sigma \in \bigcup_{\pi \in P} V^\pi$. However it may happen that there exist words $\sigma_1 \in V^{\pi_1}$, $\sigma_2 \in V^{\pi_2}$, $(\pi_1 \neq \pi_2)$ for which $\bar{\sigma}_1 = \bar{\sigma}_2$.

We are now going to construct subsets $U^\pi \subseteq V^\pi$, so that

(12.1) a) Every $u^\pi \in U^\pi$ is minimal in the sense that if $\sigma \in \Sigma(S)$ is a word for which $\overline{u^\pi} = \bar{\sigma}$, then $\omega(u^\pi) \leqq \omega(\sigma)$.

b) Every $g \in G$ is represented by some $\sigma \in \bigcup_{\pi \in P} U^\pi$.

c) No two elements in $\bigcup_{\pi \in P} U^\pi$ represent the same element of G.

For this let P_i be the set of elementary patterns π for which $\bar{\pi} \in \mathbb{Z}^n g_i$. The elements of P_i will be denoted by $\pi_i^{(j)}$, $j = 1, 2, \ldots, \operatorname{card} P_i$. Also let $p: \mathbb{Z}^{n+1} \times \mathbb{Z}^{n+1} \to \mathbb{Z}^{n+1}$ be the projection onto the first factor and $\Delta, \Delta_+ \subseteq \mathbb{Z}^{n+1} \times \mathbb{Z}^{n+1}$ be the sets defined by

$$\begin{aligned} \Delta &= \{(v, w) \in \mathbb{Z}^{n+1} \times \mathbb{Z}^{n+1} \mid v_j = w_j \text{ for } j = 1, 2, \ldots, n \text{ and } v_{n+1} \geqq w_{n+1}\} \\ \Delta_+ &= \{(v, w) \in \mathbb{Z}^{n+1} \times \mathbb{Z}^{n+1} \mid v_j = w_j \text{ for } j = 1, 2, \ldots, n \text{ and } v_{n+1} > w_{n+1}\}. \end{aligned} \tag{12.2}$$

Using the affine transformation $\mathscr{E}^\pi$ given by

$$\mathscr{E}_w^\pi = (A_1^\pi \cdot w + B_1^\pi, A_2^\pi \cdot w + B_2^\pi, \ldots, A_{n+1}^\pi \cdot w + B_{n+1}^\pi) \tag{12.3}$$

where the A_j^π and B_j^π are the ones that were introduced in §2 and §3. We define the sets

$$\operatorname{Overlap}_{j,k}^i = ((\mathscr{E}^{\pi_i^{(j)}})^{-1}\, p(\mathscr{E}^{\pi_i^{(j)}}(V^{\pi_i^{(j)}}) \times \mathscr{E}^{\pi_i^{(k)}}(V^{\pi_i^{(k)}}) \cap \Delta_*)) \cap V^{\pi_i^{(j)}} \tag{12.4}$$

where $\Delta_* = \Delta_+$ for $j < k$ and $\Delta_* = \Delta$ otherwise. This set consists of the words in $V^{\pi_i^{(j)}}$ which could be replaced by shorter words in $V^{\pi_i^{(k)}}$ representing the same element of G. Note that $\operatorname{Overlap}_{j,j}^i = \emptyset$.

We can now define the sets U^π for any elementary pattern $\pi_i^{(j)}$ by the formula

$$U^{\pi_i^{(j)}} = V^{\pi_i^{(j)}} \setminus \bigcup_{k=1}^{\operatorname{card} P_i} \operatorname{Overlap}_{j,k}^i \tag{12.5}$$

It is easy to see that the $U^{\pi_i^{(j)}} \subseteq V^{\pi_i^{(j)}}$ satisfy the properties (12.1).

The U^π are no longer positive interval sets like the V^π were, however they are "polyhedral sets". The section which follows develops a theory of polyhedral sets.

§ 13. Polyhedral Sets

This section gives a self-contained theory of "polyhedral sets," including an algorithm to decompose them into "sliced sets." Once this decomposition is made, it is relatively easy to compute the generating functions associated with these sets. Closely related work may be found in [13] and [14].

Any subset $P \subseteq \mathbb{Z}^m$ which can be expressed in one of the forms

$$\{z \in \mathbb{Z}^m \mid a \cdot z = b\}, \{z \in \mathbb{Z}^m \mid a \cdot z > b\}, \quad \text{or} \quad \{z \in \mathbb{Z}^m \mid a \cdot z \equiv b (\operatorname{mod} c)\},$$

for some $a \in \mathbb{Z}^m$, $b, c \in \mathbb{Z}$, $c > 0$ will be called an *elementary region* in $\mathbb{Z}^m$.

A finite intersection of elementary regions is a *basic polyhedral set*, and

finite disjoint unions of basic polyhedrals sets will be called simply *polyhedral sets.*

Note that any finite intersection of basic polyhedral sets is again a basic polyhedral set. Also the complement of an elementary region in $\mathbb{Z}^m$ is a basic polyhedral set, since

$$\mathbb{Z}^m \setminus \{z \in \mathbb{Z}^m \mid a \cdot z = b\} = \{z \in \mathbb{Z}^m \mid a \cdot z > b\} \cup \{z \in \mathbb{Z}^m \mid (-a) \cdot z > -b\},$$

$$\mathbb{Z}^m \setminus \{z \in \mathbb{Z}^m \mid a \cdot z > b\} = \{z \in \mathbb{Z}^m \mid (-a) \cdot z > -b+1\},$$

$$\mathbb{Z}^m \setminus \{z \in \mathbb{Z}^m \mid a \cdot z \equiv b (\mathrm{mod}\, c)\} = \bigcup_{i=1}^{c-1} \{z \in \mathbb{Z}^m \mid a \cdot z \equiv b + i (\mathrm{mod}\, c)\}.$$

Note also that $\mathbb{Z}^m$ and $\emptyset$ are elementary regions. More generally it is easy to check that the following holds.

(13.1) **Proposition.** *Polyhedral sets in $\mathbb{Z}^m$ are closed under finite unions, finite intersections, and set differences.*

(13.2) *Remark.* Suppose $P_1 \subseteq \mathbb{Z}^{m_1}$, $P_2 \subseteq \mathbb{Z}^{m_2}$ are polyhedral sets, then so is $P_1 \times P_2 \subseteq \mathbb{Z}^{m_1+m_2}$. For it is easy to see that $P_1 \times \mathbb{Z}^{m_1}$ and $\mathbb{Z}^{m_2} \times P_2$ are polyhedral sets and $P_1 \times P_2 = P_1 \times \mathbb{Z}^m \cap \mathbb{Z}^m \times P_2$.

Any elementary region may be expressed in one of the following four forms:

(13.3)

$$\text{a)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k = \sum_{i=1}^{k-1} a_i z_i + b \right\},$$

$$\text{b)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k \geqq \sum_{i=1}^{k-1} a_i z_i + b \right\},$$

$$\text{c)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k \leqq \sum_{i=1}^{k-1} a_i z_i + b \right\},$$

$$\text{d)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k \equiv \sum_{i=1}^{k-1} a_i z_i + b \,(\mathrm{mod}\, c) \right\}$$

for some $a_i, b, c \in \mathbb{Z}$, $a_k, c > 0$. In this case $P = P' \times \mathbb{Z}^{m-k}$, where P' is a region in $\mathbb{Z}^k$. P will be called a *z_k-elementary region.*

A polyhedral set will be called a *z_k-slice* if it can be written in one of the forms

(13.4)

$$\text{a)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k = \sum_{j=1}^{k-1} a_j z_j + b \right\},$$

$$\text{b)} \quad \{ z \in \mathbb{Z}^m \mid z_k \equiv b (\mathrm{mod}\, c) \},$$

$$\text{c)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k \geqq \sum_{j=1}^{k-1} a_j z_j + b,\ a_k z_k \equiv \sum_{j=1}^{k-1} a_j z_j + b (\mathrm{mod}\, c),\ a_k | c \right\},$$

$$\text{d)} \quad \left\{ z \in \mathbb{Z}^m \mid a_k z_k \leqq \sum_{j=1}^{k-1} a_j z_j + b,\ a_k z_k \equiv \sum_{j=1}^{k-1} a_j z_j + b (\mathrm{mod}\, c),\ a_k | c \right\},$$

$$\text{e)} \quad \left\{ z \in \mathbb{Z}^m \mid \sum_{j=1}^{k-1} a'_j z_j + b' \leqq a_k z_k \leqq \sum_{j=1}^{k-1} a_j z_j + b,\ a_k | c \right.$$

$$\left. a_k z_k \equiv \sum_{j=1}^{k-1} a_j z_j + b (\mathrm{mod}\, c),\ a_k z_k \equiv \sum_{j=1}^{k-1} a_j z_j + b (\mathrm{mod}\, c) \right\}$$

for some a_i, a_i', b, $c \in \mathbb{Z}$, a_k, $c > 0$. Note that if $c = 1$, then the congruences will be satisfied trivially.

Suppose $R \subseteq \mathbb{Z}^m$, $S \subseteq \mathbb{Z}^{k-1}$ are polyhedral sets. A *z_k-separated intersection* $R \cap S \times \mathbb{Z}^{m-k+1}$ is one in which

i) R is a z_k-slice,

ii) if R is of types (13.3) a), c), or d), $S \subseteq \left\{ z \in \mathbb{Z}^{k-1} \mid \sum_{j=1}^{k-1} a_j z_j + b \equiv 0 (\mathrm{mod}\, a_k) \right\}$,

iii) if R is of type (13.3) e),

$$S \subseteq \left\{ z \in \mathbb{Z}^{k-1} \mid \sum_{j=1}^{k-1} a_j z_j + b \equiv 0 (\mathrm{mod}\, a_k), \sum_{j=1}^{k-1} a_j' z_j + b \equiv 0 (\mathrm{mod}\, a_k) \right\}.$$

Note that if $R \cap S \times \mathbb{Z}^{m-k+1}$ is z_k-separated, then $\pi_{k-1}(R \cap S \times \mathbb{Z}^{m-k+1}) = S$, where π_{k-1} is the projection on the first $k-1$ coordinates of $\mathbb{Z}^m$.

A *sliced set* $P \subseteq \mathbb{Z}^m$ is a polyhedral set which can be represented as $P = P_m \cap P_{m-1} \times \mathbb{Z} \cap \ldots \cap P_1 \times \mathbb{Z}^{m-1}$ where for each k, $1 < k \leqq m$, the intersection $P_k \cap (P_{k-1} \times \mathbb{Z} \cap \ldots \cap P_1 \times \mathbb{Z}^{k-1})$ is z_k-separated. Any such representation of P will be called a *sliced presentation* of P.

(13.5) *Reduction Lemma.* Let R, S be z_m-elementary regions in $\mathbb{Z}^m$, $m > 1$.

a) If R has type (13.3) a), then $R \cap S = R \cap Q \times \mathbb{Z}$, where $Q \subseteq \mathbb{Z}^{m-1}$ is an elementary region of the same type as S.

b) If both R and S are of type (13.3) b) or both of type (13.3) c), then $R \cap S = (R \cap Q \times \mathbb{Z}) \cup (S \cap (\mathbb{Z}^{m-1} \setminus Q) \times \mathbb{Z})$ where Q is an elementary region of type (13.3) b) or (13.3) c) in $\mathbb{Z}^{m-1}$.

c) If R is of type (13.3) d), then $R = \bigcup_{\text{finite}} R_i \cap S_i \times \mathbb{Z}$ where each $R_i \subseteq \mathbb{Z}^m$ is a z_m-slice of type (13.4) b) and each $S_i \subseteq \mathbb{Z}^{m-1}$ is an elementary region of type (13.3) d).

d) If R and S are both z_m-slices of type (13.4) b), then either $R \cap S$ can be expressed as a z_m-slice of type (13.4) b) or else it is empty.

e) If R is of type (13.3) b) or c) and S is a z_m-slice of type (13.4) b), then $R \cap S = \bigcup_{\text{finite}} R_i \cap S_i \times \mathbb{Z}$, a union of z_m-separated intersections, R_i being a z_m-slice of type (13.4) c) or d) depending on the original type of R.

Proof. The proofs of a) and b) are similar. In each case Q is obtained by eliminating z_m between the two defining relations. For instance if $R = \left\{ z \in \mathbb{Z}^m \mid a_m z_m = \sum_{k=1}^{m} a_k z_k + b \right\}$, $S = \left\{ z \in \mathbb{Z}^m \mid a_m' z_m \geqq \sum_{k=1}^{m} a_k' z_k + b' \right\}$, then let

$$Q = \left\{ z \in \mathbb{Z}^m \mid a_m' \left(\sum_{k=1}^{m-1} a_j z_j + b \right) \geqq a_m \left(\sum_{k=1}^{m-1} a_j' z_j + b' \right) \right\}.$$

For c) let $R = \left\{ z \in \mathbb{Z}^m \mid a_m z_m \equiv \sum_{k=1}^{m-1} a_k z_k + b (\mathrm{mod}\, c) \right\}$. Let $c' = c/gcd(a_m, c)$, $R_i = \{ z \in \mathbb{Z}^m \mid z_m \equiv i (\mathrm{mod}\, c') \}$, $S_i = \left\{ z \in \mathbb{Z}^{m-1} \mid \sum_{k=1}^{m-1} a_j z_j + b \equiv a_m i (\mathrm{mod}\, c) \right\}$. Then $R \cap S = \bigcup_{i=0}^{c'} R_i \cap S_i \times \mathbb{Z}$.

For d), let $R=\{z\in\mathbb{Z}^m \mid z_m\equiv b(\mathrm{mod}\, c)\}$. $S=\{z\in\mathbb{Z}^m \mid z_m\equiv b'(\mathrm{mod}\, c')\}$. Then $R\cap S=\{z\in\mathbb{Z}^m \mid (c+c')z_m\equiv cb'+c'b(\mathrm{mod}\, cc')\}$. If $gcd(c+c', cc')$ divides $cb'+c'b$ then this congruence is solvable and $R\cap S$ is a z_m-slice. Otherwise $R\cap S=\emptyset$.

For e) let

$$R=\left\{z\in\mathbb{Z}^m \mid a_m z_m \geqq \sum_{k=1}^{m-1} a_k z_k + b\right\}, \qquad S=\{z\in\mathbb{Z}^m \mid z_m\equiv c(\mathrm{mod}\, d)\}.$$

Then let

$$R_i=\left\{z\in\mathbb{Z}^m \mid a_m z_m > \sum_{k=1}^{m-1} a_k z_k + b + i,\ a_m z_m \equiv \sum_{k=1}^{m-1} a_k z_k + b + i(\mathrm{mod}\, a_m d)\right\}$$

and

$$S_i=\left\{z\in\mathbb{Z}^{m-1} \mid \sum_{k=1}^{m-1} a_k z_k + b + i \equiv a_m c(\mathrm{mod}\, a_m d)\right\}.$$

Then $R\cap S=\bigcup_{i=0}^{a_m d-1} R_i\cap S_i\times\mathbb{Z}$ proving e) for the case when R is of type (13.3) b). The case for regions of type (13.3) c) is similar.

(13.6) **Proposition.** *Every polyhedral set $P\subseteq\mathbb{Z}^m$ is a finite disjoint union of sliced sets in $\mathbb{Z}^m$.*

Proof. We use induction on m. If $m=1$, any basic polyhedral set is a z_1-slice so we are done. For $m>1$ we want to show that after a finite number of modification steps any representation $P=P_1\cup\ldots\cup P_s$ as a disjoint union of basic polyhedral sets $P_i=P_i^{(1)}\cap\ldots\cap P_i^{(r_i)}$ (the $P_i^{(j)}$ being elementary regions) will become a disjoint union of sliced representations. First, whenever one of the $P_i^{(j)}$ is of type (13.3) a), and some other z_m-elementary region $P_i^{(k)}$ appears in P_i, replace $P_i^{(j)}\cap P_i^{(k)}$ using Lemma (13.5) a). Keep performing these replacements until no more can be done. Next for any two $P_i^{(j)}$, $P_i^{(k)}$ which are both of type (13.3) b) or both of (13.3) c), replace $P_i^{(j)}\cap P_i^{(k)}$ using Lemma (13.5) b). This may increase the number of P_i but the process will eventually terminate. For any elementary regions $P_i^{(j)}$ of type (13.3) d) replace them according to Lemma (13.5) c). Then any two z_m-slices $P_i^{(j)}$, $P_i^{(k)}$ of type (13.4) b) may be combined into one by Lemma (13.5) d).

At this point each P_i will appear as one of the following forms: 1) $R_i\cap S_i\times\mathbb{Z}$ where R_i is a z_m-slice of type (13.4) a) and S_i a polyhedral set in $\mathbb{Z}^{m-1}$, 2) $R_i\cap S_i\times\mathbb{Z}$ where R_i is of type (13.4) b), 3) $R_i\cap R_i''\cap S_i\times\mathbb{Z}$ where R_i is a z_m-elementary region of type (13.3) b) or c) and R_i'' is a z_m-slice of type (13.4) b) (perhaps $\mathbb{Z}^m$), or 4) $R_i\cap R_i'\cap R_i''\cap S_i\times\mathbb{Z}$ where R_i, R_i' are z_m-elementary regions of type (13.3) b) (13.3) c), and R_i'' is a z_m-slice of type (13.4) b). By replacing S_i by $S_i\cap\left\{z\in\mathbb{Z}^{m-1} \mid \sum_{k=1}^{m-1} a_k z_k + b\equiv 0(\mathrm{mod}\, a_m)\right\}$, form 1) will become z_m-separated. Form 2) is already z_m-separated. For 3), use Lemma (13.5) e) to replace $R_i\cap R_i''$ by a disjoint union of z_m-separated intersections. Intersecting with $S_i\times\mathbb{Z}$ will not affect this. Finally for 4) we apply Lemma (13.5) e) twice; once for $R_i\cap R_i''$ and another time for $R_i'\cap R_i''$.

The result of this process will be a finite disjoint union of z_m-separated intersections $R_i\cap S_i\times\mathbb{Z}$. Now apply the induction hypothesis to each S_i and we get a representation of P as a finite disjoint union of sliced representations.

Suppose that $E: \mathbb{R}^m \to \mathbb{R}^n$ is an affine transformation. L is *integral* if $E(\mathbb{Z}^m) \subseteq \mathbb{Z}^n$. Equivalently, if E is given by $E(w) = Aw + B$, where A is a matrix, B a vector, then both A and B must have integer entries.

(13.7) **Proposition.** *Suppose $Q \subseteq \mathbb{Z}^n$ is a polyhedral set and $E: R^m \to R^n$ is an integral affine transformation. Then $E^{-1}Q \cap \mathbb{Z}^m$ is a polyhedral set.*

Proof. It is enough to show this for the case when Q is an elementary region. For example, if Q has the form $Q = \{z \in \mathbb{Z}^n \mid a \cdot z = b\}$, then $E^{-1}Q \cap \mathbb{Z}^m$ is $\{w \in \mathbb{Z}^m \mid aA \cdot w = b - a \cdot B\}$, which is again an elementary region. The pullbacks of the other types can be computed similarly.

(13.8) **Proposition.** *Suppose $P \subseteq \mathbb{Z}^m$ is a polyhedral set and $E: R^m \to R^n$ is an integral affine transformation, then $E(P) \cap \mathbb{Z}^n$ is a polyhedral set.*

Proof. Let $A = (A_1, A_2, \ldots, A_n)$. Factor $A = \tau \circ \pi \circ \iota$ as the composition of an injection followed by a projection, and finally a translation, where $\iota: \mathbb{Z}^m \to \mathbb{Z}^{m+n}$ $\iota(z_1, z_2, \ldots, z_n) = (A_1 \cdot z, \ldots, A_n \cdot z, z_1, z_2, \ldots, z_m)$, $\pi: \mathbb{Z}^{m+n} \to \mathbb{Z}^n$, $\pi(z_1, z_2, \ldots, z_{m+n})$ $= (z_1, \ldots, z_n)$, and $\tau: \mathbb{Z}^n \to \mathbb{Z}^n$, $\tau(z_1, z_2, \ldots, z_n) = (z_1 + B_1, \ldots, z_n + B_n)$. We only need to verify that Proposition (13.8) holds for ι, π, and τ separately.

The maps ι and τ are both injective, and so it is enough to simply show that they carry elementary regions into polyhedral sets. For example

$$\tau(\{z \in \mathbb{Z}^m \mid a \cdot z = b\}) = \{w \in \mathbb{Z}^n \mid a \cdot w = b + a \cdot B\}$$

and

$$\iota(\{z \in \mathbb{Z}^m \mid a \cdot z = b\}) = \{w \in \mathbb{Z}^{m+n} \mid w_n = A_n \cdot (w_{n+1}, \ldots, w_{n+m})\}.$$

We can show in a similar manner that ι and τ carry elementary regions of other types into polyhedral sets as well. This verifies Proposition (13.8) for ι and τ.

For the projection π, it is enough to show that $\pi: \mathbb{Z}^m \to \mathbb{Z}^{m-1}$, given by $\pi(z_1, z_2, \ldots, z_m) = (z_1, \ldots, z_{m-1})$ satisfies Proposition (13.8). We write any polyhedral set $P = P_1 \cup \ldots \cup P_r$, a finite disjoint union of sliced sets. In turn, express each $P_i = R_i^{(m)} \cap R_i^{(m-1)} \times \mathbb{Z} \cap \ldots \cap R_i^{(1)} \times \mathbb{Z}^{m-1}$ in a sliced presentation. Since the intersection $R_i^{(m)} \cap (R_i^{(m-1)} \times \mathbb{Z} \cap \ldots \cap R_i^{(1)} \times \mathbb{Z}^{m-1})$ is z_m-separated, $\pi(P_i) = R_i^{(m-1)}$ $\cap \ldots \cap R_i^{(1)} \times \mathbb{Z}^{m-2}$. Therefore, $\pi(P_i)$ is a polyhedral. But, then $\pi(P)$ is a polyhedral set by Proposition (13.1) b).

§ 14. Generating Functions on Positive Polyhedral Sets

Any $I \subseteq \mathbb{Z}_+^m$ which is a polyhedral set will be called a *positive polyhedral set.* Positive interval sets are examples of these. Evidently, a sliced presentation for a positive polyhedral set may contain only slices of the form (13.4) a), c), or e).

The generating function which counts the lattice points in I lying on slices perpendicular to the weight vector $(\omega_1, \ldots, \omega_m)$ can be defined by the formula (9.1).

(14.1) **Proposition.** *Suppose $I \subseteq \mathbb{Z}_+^m$ is a positive polyhedral set, then the generating function is rational with the form*

$$p_I^{\omega}(t)=\frac{q_I^{\omega}(t)}{\prod_{i=1}^{k}(1+\ldots+t^{v_i-1})(1-t)^n}$$

for some polynomial $q_I^{\omega}(t)$ *with integer coefficients and for some positive integers* $n, k, v_1, v_2, \ldots, v_k$.

Proof. We will prove more. Suppose that ω is a weight vector consisting of positive *rational* weights, then we will show that

$$p_I^{\omega}(t)=\sum_{x\in I} t^{\omega\cdot x}$$

has the above form for some polynomial $q_I^{\omega}(t)$ (in fractional powers of t) with integer coefficients and for some positive rational numbers v_i. When ω is integral then the v_i must be positive integers and $q_I^{\omega}(t)$ must really be a polynomial in t.

It is enough to prove this for the case when I is a positive sliced set. We will use induction on m. A simple computation settles the case for $m=1$. For $m>1$, let $I=I_m\cap I_{m-1}\times\mathbb{Z}\cap\ldots\cap I_1\times\mathbb{Z}^{m-1}$ be a sliced presentation. Let $I'=I_{m-1}\cap I_{m-2}\times\mathbb{Z}\cap\ldots\cap I_1\times\mathbb{Z}^{m-2}$, so that $I=I_m\cap I'\times\mathbb{Z}$. Depending on the type of I_m it is possible to derive the following formulas for $p_I^{\omega}(t)$ in terms of $p_{I'}^{\omega'}(t)$:

(1) type 13.3a): $p_I^{\omega}(t)=t^{\omega_m b/a_m}\,p_{I'}^{\omega'}(t)$,

(2) type 13.3c): $p_I^{\omega}(t)=\dfrac{t^{\omega_m b/a_m}\,p_{I'}^{\omega'}(t)}{1-t^{c\,\omega_m/a_m}}$,

(3) type 13.3e): $p_I^{\omega}(t)=\dfrac{t^{\omega_m b/a_m}\,p_{I'}^{\omega'}(t)-t^{\omega_m(b'+c)/a_m}\,p_{I'}^{\omega''}(t)}{1-t^{c\,\omega_m/a_m}}$.

In the formulas above, $\omega'=(\omega_1+a_1\omega_m/a_m,\ldots,\omega_{m-1}+a_{m-1}\omega_m/a_m)$ and $\omega''=(\omega_1+a_1'\omega_m/a_m,\ldots,\omega_{m-1}+a_{m-1}'\omega_m/a_m)$. In either case we may apply the induction hypothesis to $p_{I'}^{\omega'}(t)$ and $p_{I'}^{\omega''}(t)$. This implies that $p_I^{\omega}(t)$ has the specified form, proving Proposition (14.1).

§ 15. The General Case

The elements of G are in one-to-one correspondence with the elements of $\bigcup_{\pi\in P} U^{\pi}$, and, in fact,

$$g_{G,S}^{\omega}(N)=\sum_{\pi\in P}\operatorname{card}\{u\in U^{\pi}\mid A_{n+1}^{\pi}\cdot u+B_{n+1}^{\pi}=N\}. \tag{15.1}$$

In terms of the generating function, (15.1) translates to

$$p_{G,S}^{\omega}(t)=\sum_{\pi\in P} t^{B_{n+1}^{\pi}}\,p_{U^{\pi}}^{A_{n+1}^{\pi}}(t). \tag{15.2}$$

The construction of U^π in §12 produces polyhedral sets according to Propositions (8.1), (13.1), (13.7), and (13.8) and Remark (13.2). Since $\dim V^\pi \leqq n$, by Proposition (8.1), Remark (13.2) shows that $\dim U^\pi \leqq n$, as well. We may use (14.1) to get:

(15.3) **Corollary.** *If G is a group containing a subgroup of finite index isomorphic to $\mathbb{Z}^n$ and S is a finite subset generating G, then*

$$p^{\omega}_{G,S}(t) = \frac{q^{\omega}_{G,S}(t)}{\prod_{i=1}^{k} (1 + \ldots + t^{v_i - 1})(1-t)^n}$$

for some polynomial $q^{\omega}_{G,S}(t)$ with integer coefficients and for some positive integers $v_1, \ldots, v_k$.

Theorem (1.2) now follows from Corollary (15.3) and from the result of Bass [2] (Theorem 2, p. 608) because if $p^{\omega}_{G,S}(t)$ were to have a pole of order $< n$, then $\gamma^{\omega}_{G,S}(t)$ would be periodically polynomial of degree $< n$, contradicting Bass' result.

References

1. Atiyah, M.F., MacDonald, I.G.: Introduction to commutative algebra. Reading: Addison-Wesley 1969
2. Bass, H.: The degree of polynomial growth of finitely generated nilpotent groups. Proc. London Math. Soc. (3) **25**, 603–614 (1972)
3. Bourbaki, N.: Groupes et algèbres de Lie, Chapitres 4, 5, et 6. Paris: Hermann 1968
4. Cannon, J.W.: The growth of the closed surface groups and the compact hyperbolic Coxeter groups (unpublished)
5. Cannon, J.W.: The combinatorial structure of cocompact discrete hyperbolic groups (to appear in Geometriae Dedicata)
6. Gromov, M.: Groups of polynomial growth and expanding maps. Publ. Math. I.H.E.S. **53** (1981)
7. Klarner, D.A.: Mathematical crystal growth I. Discrete Applied Mathematics **3**, 47–52 (1981)
8. Klarner, D.A.: Mathematic crystal growth II. Discrete Applied Mathematics **3**, 113–117 (1981)
9. Milnor, J.: A note on curvature and fundamental group. J. Differential Geometry **2**, 1–7 (1968)
10. Milnor, J.: Growth of finitely generated solvable groups. J. Differential Geometry **2**, 447–448 (1968)
11. Serre, J-P.: Cohomologie des groupes discrets. In: Prospects in Mathematics, pp. 77–169. Princeton: Princeton University Press 1971
12. Solomon, L.: The orders of the finite Chevalley groups. J. of Algebra **3**, 376–393 (1966)
13. Stanley, R.P.: Linear diophantine equations and local cohomology. Invent. Math. **68**, 175–193 (1982)
14. Stanley, R.P.: Linear homogeneous diophantine equations and magic labellings of graphs **40**, 607–632 (1973)
15. Tits, J.: Groupes à croissance polynomiale. Séminaire Bourbaki **572**, 1–13 (1981)
16. Wagreich, P.: The growth function of a discrete group (to appear in the Proceedings of the Conference on Algebraic Varieties with Group Actions, Vancouver, 1982)
17. Wolf, J.A.: Growth of finitely generated solvable groups and curvatures of Riemannian manifolds. J. Differential Geometry **2**, 421–446 (1968)

Oblatum 8-XI-1982

Invent. math. 73, 271-285 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Flat Bundles and Residues for Foliations

J.L. Heitsch*

Institute for Advanced Study, University of Illinois at Chicago, Chicago, IL 60680, USA

§1. Introduction

In this note we compute the residues for a vector field preserving a foliation. We assume the vector field disappears only to first order along its singular set and that its singular set consists of closed separated leaves. Under these conditions, the residue is completely determined by the characteristic classes of the flat bundle ν_0, which is the normal bundle ν of the foliation restricted to the singular set, and the action of the vector field on ν_0.

For a foliation of even codimension q whose normal bundle has Euler class zero, we interpret the Pontrjagin ring of the normal bundle in dimension $2q$ as an obstruction to the existence of such a vector field. We show that if the Euler class of the normal bundle of a foliation is zero and it admits such a vector field, then many of its secondary characteristic classes must be also be zero. This is one of the few results relating the geometry of a foliation to its characteristic classes.

We will make extensive use of the material in [6]. It will be very helpful to the reader to be familiar with this paper.

§2. The Main Theorem

All objects are assumed to be smooth, i.e., C^∞. If τ is a bundle over a manifold M, the fiber of τ over $x \in M$ is denoted τ_x, and the space of smooth sections of τ is written $C^\infty(\tau)$. If $X \in C^\infty(\tau)$, its value at $x \in M$ is written $X(x)$. The tangent bundle of M is denoted TM. We write $GL_q^+\mathbb{R}$ for the group of $q \times q$ invertible real matrices with positive determinant and $gl_q\mathbb{R}$ for its Lie algebra. The special orthogonal group is denoted SO_q and its Lie algebra so_q.

In [6], a residue theorem for vector fields preserving a foliation was proven. We briefly recall this result.

* Partially supported by NSF Grant MSC 81-08814

Let τ be a codimension q foliation on a manifold M. We always assume the normal bundle ν of τ is orientable. A vector field $X \in C^\infty(TM)$ is a Γ vector field for τ if for all $Y \in C^\infty(\tau)$, $[X, Y] \in C^\infty(\tau)$. The singular set of X is

$$\mathscr{A}(X) = \{x \in M \mid X(x) \in \tau_x\}$$

and it is a union of leaves of τ. If $\mathscr{A}(X)$ is a finite union of closed separated leaves, we say X has isolated singularities.

The differential graded algebra WO_q is

$$WO_q = R_q[c_1, \ldots, c_q] \otimes \Lambda(\hat{c}_1, \hat{c}_3, \ldots, \hat{c}_{2s+1})$$

where degree $c_i = 2i$, degree $\hat{c}_i = 2i-1$, and $2s+1 = q$ or $q-1$. $\mathbb{R}_q[c_1, \ldots, c_q]$ is the polynomial algebra $\mathbb{R}[c_1, \ldots, c_q]$ truncated above degree $2q$, and Λ is the exterior algebra on the $\hat{c}_i$s. The differential is given by setting

$$d(c_i \otimes 1) = 0 \quad d(1 \otimes \hat{c}_i) = c_i \otimes 1$$

and extending as a derivation.

Associated to the foliation τ is a natural map

$$\beta_\tau : H^*(WO_q) \to H^*(M, \mathbb{R}).$$

For the construction of this map see [3]. A basis for $H^*(WO_q)$, due to J. Vey [5], is given by the classes

$$c_{2j_1} \ldots c_{2j_l} \qquad 2j_1 + \ldots + 2j_l \leqq q$$

and the classes

$$\begin{aligned} \hat{c}_{i_1} \ldots \hat{c}_{i_k} c_{j_1} \ldots c_{j_l} \qquad & i_1 < i_2 < \ldots \\ & j_1 \leqq j_2 \leqq \ldots \\ & i_1 \leqq \text{any odd } j \\ & i_1 + j_1 + \ldots + j_l \geqq q+1. \end{aligned}$$

Let $I_q(WO_q)$ be the ideal in WO_q generated by the elements of the form $c_{j_1} \ldots c_{j_l}$ where $j_1 + \ldots + j_l = q$.

Theorem 2.1 (3.11 of [6].) *Let τ be a codimension q foliation on M whose normal bundle ν is oriented. Let X be a Γ vector field for τ with isolated singularities and set $\mathscr{A}(X) = \bigcup_{i=1}^{r} N_i$. Let $\varphi \in I_q(WO_q)$ be an element of degree k. Then φ, τ and X determine a cohomology class*

$$\mathrm{Res}_\varphi(\tau, X, N_i) \in H^{k-q}(N_i; \mathbb{R}),$$

the residue of φ, τ, and X at N_i, such that

i) $\mathrm{Res}_\varphi(\tau, X, N_i)$ *depends only on the behavior of τ and X in a neighborhood of N_i.*

ii) $\sum_i j_* \mathrm{Res}_\varphi(\tau, X, N_i) = \beta_\tau(\varphi)$

where j_* is the composition

$$H^{k-q}(N_i; \mathbb{R}) \xrightarrow{t} H_c^k(D_i; \mathbb{R}) \xrightarrow{e_i^*} H^k(M; \mathbb{R})$$

of the Thom isomorphism and the map given by trivially extending a form with fiber compact support. D_i is an embedded disc bundle with zero section N_i and $H_c^k(D_i; \mathbb{R})$ denotes the cohomology of forms with fiber compact support.

In what follows we will be concerned with computing the residue for a single leaf $N \in \mathscr{A}(X)$. As the residues are local we may assume that M is an oriented R^q bundle over N, that τ is transverse to the fibers of M, and that the singular set of X is precisely the zero section N of M. We may assume that X is tangent to the fibers of M [6].

For any such vector field X, we denote by A the following matrix and call it the matrix associated to X. On a fiber $\mathbb{R}^q$ of M, with coordinates $x_1, \ldots, x_q$, we have

$$X|_{\mathbb{R}^q} = \sum_{i=1}^{q} f_i(x_1, \ldots, x_q)\, \partial/\partial x_i.$$

Set $A = (a_{ij})$ where $a_{ij} = \dfrac{\partial f_i}{\partial x_j}(0, \ldots, 0)$.

The matrix A is well defined up to conjugation by an invertible matrix. Thus if c_i is the ith Chern polynomial on $gl_q\mathbb{R}$, then $c_i(A)$ is well defined. We call X a linear Γ vector field if its associated matrix A is invertible.

Write $\tilde{N}$ for the universal cover of N and for any linear Γ vector field X, let X_A be the vector field on $\tilde{N} \times \mathbb{R}^q$ given by

$$X_A = \sum_{i,j=1}^{q} a_{ij} x_j\, \partial/\partial x_i.$$

Let $h: \pi_1 N \to GL_q^+\mathbb{R}$ be the linear holomony of M and set $M_0 = \tilde{N} \times_h \mathbb{R}^q$. Then X_A induces a vector field X_A on M_0 which is a Γ vector field for the natural flat foliation τ_0. $\mathscr{A}(X_A)$ is the zero section N of M_0.

Theorem 2.2 [7]. *For all $\varphi \in I_q(WO_q)$,*

$$\mathrm{Res}_\varphi(\tau, X, N) = \mathrm{Res}_\varphi(\tau_0, X_A, N).$$

That is, the residues of a linear Γ vector field are determined by the linear parts of τ and X on the singular set of X. We shall show that the residues in this case are in fact determined by the characteristic classes of the flat bundle M_0 and the matrix A. Note that M_0 is equivalent to ν_0, the normal bundle of τ restricted to N.

The flat $GL_q^+\mathbb{R}$ structure on M_0 determines a canonical homomorphism

$$\alpha: H^*(gl_q\mathbb{R}, SO_q) \to H^*(N; R)$$

which measures the incompatibility of the flat structure with an SO_q structure. Recall that for q even

$$H^*(gl_q\mathbb{R}, SO_q) = \Lambda(h_1, h_3, \ldots, h_{q-1}, \chi)$$

where χ is the Euler class; for q odd

$$H^*(gl_q\mathbb{R}, SO_q)=\Lambda(h_1, h_3, \ldots, h_q).$$

The main result of this paper is the following.

Theorem 2.3. *Let τ, X, N be as above where X is a linear Γ vector field for τ with associated matrix A. Let $\varphi\in I_q(WO_q)$ be of the form $\varphi=\hat{c}_{i_1}\ldots\hat{c}_{i_k}c_{j_1}\ldots c_{j_l}$. Then, for q odd*

$$\mathrm{Res}_\varphi(\tau, X, N)=0$$

and for q even

$$\mathrm{Res}_\varphi(\tau, X, N)=\frac{c_{j_1}\ldots c_{j_l}(A)}{c_q(A)}\cdot\alpha(h_{i_1}\ldots h_{i_k}\chi).$$

Note that it is permitted to take $\varphi=c_{j_1}\ldots c_{j_l}$ in the above formula.

Corollary 2.4. *Let τ be a codimension q foliation with oriented normal bundle ν. Let X be a linear Γ vector field for τ with isolated singularities. If the Euler class of $\nu|_{\mathscr{A}(X)}$ is zero in $H^q(\mathscr{A}(X))$, then for all $\varphi\in I_q(WO_q)$,*

$$\beta_\tau(\varphi)=0.$$

§3. Computation of the Residues

By Theorem 2.2 we are reduced to computing the residues in the following situation. Let $h\colon \pi_1 N\to GL_q^+\mathbb{R}$ be a homomorphism and denote the associated $\mathbb{R}^q$ bundle by $M=\tilde{N}\times_h\mathbb{R}^q$. For any $A=(a_{ij})$ in the centralizer of $h(\pi_1 N)$, the vector field $X_A=\sum_{i,j=1}^{q} a_{ij}x_j\partial/\partial x_i$ is invariant by h and so induces a Γ vector field X_A for the natural flat foliation τ on M. $\mathscr{A}(X_A)=N$, the zero section of M. Let $\varphi=\hat{c}_I c_J$ be as in Theorem 2.3. In [6], see also [10], it was shown that

$$\mathrm{Res}_\varphi(\tau, X_A, N)=\alpha(h_{i_1}\ldots h_{i_k})\,\mathrm{Res}_{c_J}(\tau, X_A, N).$$

Thus we may assume $\varphi=c_J$.

To compute Res_{c_J}, let ω_A be any one form on M with $\omega_A|_\tau=0$ and $\omega_A(X_A)=1$ off some disc subbundle D of M. The foliation τ on M is transverse to the fibers of M so we may identify the normal bundle of τ with $T\mathbb{R}^q$, the tangent bundle along the fiber of M. Define a connection θ on $T\mathbb{R}^q$ by requiring its covariant derivative ∇ to satisfy

$$\nabla_Y\,\partial/\partial x_i=\omega_A(Y)[X_A, \partial/\partial x_i].$$

Here $Y\in TM$ and $(x_1, \ldots, x_q)$ are coordinates on the fiber $\mathbb{R}^q$ of M coming from the natural coordinates on $\mathbb{R}^q$ in $\tilde{N}\times\mathbb{R}^q$. It is not difficult to show that θ is well defined and that it is a basic X_A connection on $T\mathbb{R}^q$ supported off D. See [6] pp. 439–440.

$\mathrm{Res}_{c_J}(\tau, X_A, N)$ is obtained by integrating the form $c_J(\Omega)$, Ω the curvature of θ, over the fiber of M. With respect to the local framing $\partial/\partial x_1, \ldots, \partial/\partial x_q$ of $T\mathbb{R}^q$, the local form of θ is

$$-\omega_A \cdot A$$

and so

$$\Omega = -d\omega_A \cdot A$$

and

$$c_J(\Omega) = c_J(A) \cdot (-d\omega_A)^q.$$

Denoting integration over the fiber of M by $\int$, we have

$$\mathrm{Res}_{c_J}(\tau, X_A, N) = c_J(A) \cdot [\int (-d\omega_A)^q]$$

where $[\gamma]$ means the cohomology class determined by γ.

Since the identity matrix I is always in the centralizer of h, the vector field $X_I = \sum_{i=1}^{q} x_i \partial/\partial x_i$ induces a Γ vector field X_I for τ on M with singular set N. Let ω_I be a one form as above for X_I.

Proposition 3.1.

$$[\int (d\omega_A)^q] = \frac{1}{\det A} [\int (d\omega_I)^q].$$

Corollary 3.2.

$$\mathrm{Res}_{c_J}(\tau, X_A, N) = \frac{c_J(A)}{c_q(A)} \mathrm{Res}_{c_q}(\tau, X_I, N).$$

This follows immediately from 3.1 and the definition of the c_j.

Proof of 3.1. Define a bundle map

$$\tilde{A}\colon T(\tilde{N} \times \mathbb{R}^q) \to T(\tilde{N} \times \mathbb{R}^q)$$

covering the identity as follows. Let (x, y) be coordinates on $T\mathbb{R}^q = \mathbb{R}^q \times \mathbb{R}^q$ where y is the fiber coordinate. Let $\tilde{A}$ be the diffeomorphism of $T\mathbb{R}^q$ given by

$$\tilde{A}(x, y) = (x, A^{-1} y).$$

Extend $\tilde{A}$ to $T(\tilde{N} \times \mathbb{R}^q)$ by making it the identity on $T\tilde{N}$. This diffeomorphism of $T(\tilde{N} \times \mathbb{R}^q)$ commutes with the action of $\pi_1 N$ and so induces a bundle map $\tilde{A}$ of TM covering the identity.

Denote the map which $\tilde{A}$ induces on the forms on M by $\tilde{A}^*$.

Lemma 3.3. *The form $\omega = \tilde{A}^* \omega_I$ on M satisfies*

i) $\omega|_\tau \equiv 0$

ii) $\omega(X_A) = 1$ *off a disc subbundle.*

Proof. As $\tilde{A}$ preserves τ, i) is immediate. Property ii) follows from the fact that $\tilde{A}(X_A) = X_I$.

Thus we may assume that $\omega_A = \tilde{A}^* \omega_I$. Let (n, x), $x = (x_1, \ldots, x_q)$ be local coordinates on M coming from local coordinates on $\tilde{N} \times \mathbb{R}^q$. The form ω_I is

given locally by

$$\omega_I = \sum_{i=1}^{q} f_i(n,x)\, dx_i$$

and

$$\tilde{A}^* dx_i = \sum_{k=1}^{q} b_{ik}\, dx_k$$

where $A^{-1} = (b_{ij})$. Thus

$$\omega_A = \sum_{i=1}^{q} f_i(n,x) \cdot \tilde{A}^* dx_i .$$

Since we want to compute $(d\omega_A)^q$ and each term of ω_A already contains a dx_k, we may take

$$d\omega_A = \sum_{i=1}^{q} d_N f_i \wedge \tilde{A}^* dx_i$$

where d_N is the exterior derivative on N. Then the equality

$$(d\omega_A)^q = \frac{1}{\det A} (d\omega_I)^q$$

follows immediately from the observation that

$$\prod_{i=1}^{q} \tilde{A}^* dx_i = \prod_{i=1}^{q} \left(\sum_{i=1}^{q} b_{ik}\, dx_k \right) = \det A^{-1} \cdot \prod_{i=1}^{q} dx_i .$$

To complete the proof of Theorem 2.3 we need only the following.

Theorem 3.4. *If $\chi(M)$ denotes the Euler class of the bundle M over N, then*

$$\mathrm{Res}_{c_q}(\tau, X_I, N) = \chi(M).$$

Proof. The cohomology class $\mathrm{Res}_{c_q}(\tau, X_I, N) \in H^q(N; \mathbb{R})$ is natural under pull backs and it is a differential invariant. Thus the theorem of van Est [13] implies that it is the image under α: $H^*(gl_q\mathbb{R}, SO_q) \to H^*(M, \mathbb{R})$ of a universal class

$$\mathrm{Res}_{c_q} = a_0 \chi + a_1 h_1 + a_3 h_3 + \ldots, \qquad a_j \in \mathbb{R}$$

where $a_0 = 0$ when q is odd.

Denote $GL_q^+ R$ by G and SO_q by K and let $\Gamma \subset G$ be a discrete cocompact subgroup which acts freely on G/K. Let Γ act on $G \times R^q$ on the left by

$$g_1(g,x) = (g_1 g, g_1 x).$$

Let K act on the right on G and form the flat bundle

$$M = (G/K) \times_\Gamma R^q$$

over

$$N = \Gamma \backslash G / K.$$

The map α in this case is well known to be injective. Thus we need only compute Res_{c_q} for this one example. On this account see also [8], [9].

To do so we view M not as a flat bundle but as the K bundle

$$M=(\Gamma\backslash G)\times_K \mathbb{R}^q.$$

Here Γ acts on the left by

$$g_1(g,x)=(g_1 g,x)$$

and K acts on the right by

$$(g,x)\cdot k=(gk,k^{-1}x).$$

Let τ be the foliation on $G\times\mathbb{R}^q$ whose leaves are the submanifolds

$$L_x=\{(g,g^{-1}x)|g\in G\},\qquad x\in\mathbb{R}^q.$$

If Y is a left invariant vector field on G and Y^* is the vector field it induces on $\mathbb{R}^q$, then the vector field $Y-Y^*$ on $G\times\mathbb{R}^q$ is tangent to τ. The action of K preserves each L_x while $g_1\in\Gamma$ takes L_x to $L_{g_1 x}$. Thus τ induces a foliation τ on M. The diffeomorphism of $G\times\mathbb{R}^q$ given by

$$(g,x)\to(g,gx)$$

induces a diffeomorphism from $(\Gamma\backslash G)\times_K\mathbb{R}^q$ to $(G/K)\times_\Gamma\mathbb{R}^q$ and it maps τ to the natural flat foliation.

Let ω_I be the one form on $\mathbb{R}^q$,

$$\omega_I=\sum_{i=1}^{q}x_i\,dx_i.$$

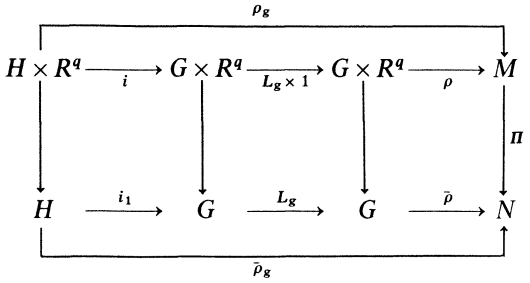
The action of K on $\mathbb{R}^q$ preserves ω_I so it induces a one form ω_I on M. We identify the normal bundle of τ with $T\mathbb{R}^q$, the tangent bundle along the fiber of M. Let θ be the unique basic connection on $T\mathbb{R}^q$ whose covariant derivative satisfies for all $Z\in T\mathbb{R}^q$

$$\nabla_Z\,\partial/\partial x_i=\omega_I(Z)\,[X_I,\partial/\partial x_i].$$

The vector fields $\partial/\partial x_i$ on M are induced from the projection $\rho\colon G\times\mathbb{R}^q\to M$. By Lemmas 5.7 and 4.15 of [6], θ is well defined and we may use it to compute Res_{c_q} provided that we integrate only over the disc subbundle D of M

$$D=\left\{\rho(g,x)\,\middle|\,\sum_{i=1}^{q}x_i^2=1\right\}.$$

The following allows us to choose nice local coordinates on M and N. Let H be the subgroup of G consisting of the upper triangular matrices and denote its Lie algebra by h. Let τ_H be the foliation on $H\times R^q$ spanned by the vector fields $Y-Y^*$, where $Y\in h$. Suppose $z\in N$ is the image of $g\in G$ under the natural map $\bar{\rho}\colon G\to N$. Consider the commutative diagram



Here i and i_1 are the injections, $L_g(g_1)=gg_1$, and all the vertical maps are the obvious projections. There is a neighborhood U of the identity e in H such that $\bar{\rho}_g|_U$ is a diffeomorphism onto the neighborhood $\bar{\rho}_g(U)$ of z in N. In addition, $\rho_g|_{U\times\mathbb{R}^q}$ is a bundle isomorphism onto $\Pi^{-1}(\rho_g(U))$ which maps τ_H to τ. See the proof of 5.1 in [6]. Thus we may use $U\times R^q$ as local coordinates on M, and the foliation τ in these local coordinates is just τ_H.

Let A_{ij}, $1\leqq i\leqq j\leqq q$, be the basis of h where A_{ij} has a 1 in the i,j spot and zeros elsewhere. Let $\bar{\omega}_{ij}$ be the left invariant one forms on H which are dual to the A_{ij}. Now

$$A^*_{ij}=x_j\,\partial/\partial x_i$$

and the tangent bundle to $U\times R^q$ is spanned by the vector fields $\partial/\partial x_i$ and the left invariant vector fields A_{ij}. For the local framing $\partial/\partial x_1,\ldots,\partial/\partial x_q$ of the normal bundle $T\mathbb{R}^q$, the covariant derivative ∇ of θ satisfies

$$\nabla_{\partial/\partial x_i}\partial/\partial x_k=-\omega_I(\partial/\partial x_i)\cdot\partial/\partial x_k,$$

and

$$\begin{aligned}\nabla_{A_{ij}}\partial/\partial x_k&=\nabla_{A_{ij}-A^*_{ij}}\partial/\partial x_k+\nabla_{A^*_{ij}}\partial/\partial x_k\\&=-[x_j\,\partial/\partial x_i,\partial/\partial x_k]-\omega_I(A^*_{ij})\,\partial/\partial x_k\\&=\delta^j_k\cdot\partial/\partial x_i-x_ix_j\cdot\partial/\partial x_k.\end{aligned}$$

Thus locally

$$\theta=-(\omega_I+\sum_{i\leqq j}x_ix_j\,\bar{\omega}_{ij})\cdot I+(\bar{\omega}_{ij})$$

i.e.,

$$\theta_{lk}=-\delta^k_l(\omega_I+\sum_{i\leqq j}x_ix_j\,\bar{\omega}_{ij})+\bar{\omega}_{lk}$$

where we set $\bar{\omega}_{lk}=0$ for $l>k$.

If we set $\bar{\omega}=(\bar{\omega}_{ij})$, the Maurier-Cartan equation for H is

$$d\bar{\omega}+\bar{\omega}\wedge\bar{\omega}=0.$$

Using $d\omega_I=0$ and $\theta\wedge\theta=\bar{\omega}\wedge\bar{\omega}$, we have that the curvature $\bar{\Omega}=d\theta+\theta\wedge\theta$ of θ is

$$\bar{\Omega}=-d(\sum_{i\leqq j}x_ix_j\,\bar{\omega}_{ij})\cdot I.$$

To find the local expression of Res_{c_q} at z in N, we integrate $c_q(\bar{\Omega})$ over the disc $D=\{(x_1,\ldots,x_q)|\sum x_i^2\leqq 1\}$. We may ignore terms in $\bar{\Omega}$ of the form $d\bar{\omega}_{ij}$, since any term of $c_q(\bar{\Omega})$ containing a $d\bar{\omega}_{ij}$ must be zero for dimensional reasons. Thus we may take for $\bar{\Omega}$,

$$\bar{\Omega}=-\Big(\sum_{i\leqq j} d(x_i x_j)\,\bar{\omega}_{ij}\Big)\cdot I.$$

Now consider the form Ω on $G\times R^q$

$$\Omega=-\left(\sum_{i,j=1}^{q} d(x_i x_j)\,\omega_{ij}\right)\cdot I$$

where the ω_{ij} are the canonical left invariant one forms on G. Then

$$i^*(\omega_{ij})=\bar{\omega}_{ij}$$

(recall $\bar{\omega}_{ij}=0$ for $i>j$) and so

$$i^*\Omega=\bar{\Omega}.$$

Thus $c_q(\bar{\Omega})=i^*c_q(\Omega)$ and

$$\int_D c_q(\bar{\Omega})=i_1^*\Big(\int_D c_q(\Omega)\Big).$$

That is, the local expression for Res_{c_q} at z in N with respect to the coordinates obtained from H is given by $(i_1^*\int_D c_q(\Omega))_e$. In the remainder of this section we compute $\int_D c_q(\Omega)$. In the next section we show that it is equal to $(\bar{\rho}\circ L_g)^*\chi$ where χ is a form on N representing the Euler class of the bundle M. But $(i_1^*(\bar{\rho}\circ L_g)^*\chi)_e$ is just the local expression for $\chi(M)$ at z in N with respect to the coordinates obtained from H. This proves the Theorem.

Set

$$\gamma=\sum_{i,j=1}^{q} d(x_i x_j)\,\omega_{ij},$$

then

$$c_q(\Omega)=\left(\frac{1}{2\pi i}\cdot\gamma\right)^q.$$

The form γ is left invariant on $G\times\mathbb{R}^q$ and so is determined by its restriction to gl_qR. Restricted to so_q it is zero. Thus it is completely determined by its restriction to the subspace of symmetric matrices s_q, which amounts to setting $\omega_{ij}=\omega_{ji}$.

Doing so we have

$$\gamma=2\sum_{i,j=1}^{q} x_j\,dx_i\,\omega_{ij}.$$

The form γ^q is now easily seen to be

$$\gamma^q=(-1)^a\,2^q q!\sum x_{j_1}\ldots x_{j_q}\,dx_1\ldots dx_q\,\omega_{1j_1}\ldots\omega_{qj_q}. \tag{3.5}$$

where $a=q(q-1)/2$ and the sum is over $j_1=1,\ldots,q;\ldots;j_q=1,\ldots,q$.

When we integrate γ^q over the disc D, each of the terms in (3.5) where $x_{j_1} \dots x_{j_q}$ is an odd function drops out. Thus we may take the sum in (3.5) to be over those $j_1, \dots, j_q$ where $x_{j_1} \dots x_{j_q}$ is an even function. When q is odd, every term $x_{j_1} \dots x_{j_q}$ in (3.5) is an odd function. Thus we have that for q odd, $\mathrm{Res}_{c_q}(\tau, X_I, N) = 0$.

We now assume that q is an even integer, $q = 2m$.

Lemma 3.6. *Let $a_1, \dots, a_{2m}$ be non-negative integers with $a_1 + \dots + a_{2m} = m$. Then*

$$\int_{\sum_{i=1}^{2m} x_i^2 \leqq 1} x_1^{2a_1} \dots x_{2m}^{2a_{2m}} \, dx_1 \dots dx_{2m}$$

$$= \prod_{i=1}^{2m} (1 \cdot 3 \cdot 5 \dots (2a_i - 1)) \int_{\sum_{i=1}^{2m} x_i^2 \leqq 1} x_1^2 \dots x_m^2 \, dx_1 \dots dx_{2m}$$

$$= \frac{\pi^m}{2^m (2m)!} \prod_{i=1}^{2m} (1 \cdot 3 \cdot 5 \dots (2a_i - 1)).$$

The proof is by two fairly easy inductions and is omitted. Summarizing we now have

Proposition 3.7.

i) *For q odd, $\mathrm{Res}_{c_q}(\tau, X_I, N) = 0$.*

ii) *for q even, $q = 2m$, the local expression for $\mathrm{Res}_{c_q}(\tau, X_I, N)$ at z in N is $(i_1^* \Phi)_e$ where Φ is a left invariant form on G which is completely determined by its restriction to s_q and*

$$\Phi|_{s_q} = \frac{1}{(2\pi)^m} \sum A_{j_1 \dots j_{2m}} \omega_{1 j_1} \dots \omega_{2m j_{2m}}. \tag{3.8}$$

The sum is over all $j_1, \dots, j_{2m}$ where $x_{j_1} \dots x_{j_{2m}}$ is an even function. The coefficient $A_{j_1 \dots j_{2m}}$ is given as follows. The j's come in groups of equal integers, say r groups which are $2a_1, \dots, 2a_r$ in number. Then

$$A_{j_1 \dots j_{2m}} = \prod_{i=1}^{r} (1 \cdot 3 \cdot 5 \dots (2a_i - 1)).$$

To obtain Φ one replaces ω_{ij} by $\frac{1}{2}(\omega_{ij} + \omega_{ji})$ in (3.8). Thus to obtain the local expression for $\mathrm{Res}_{c_q}(\tau, X_I, N)$ at z in N it is only necessary to replace ω_{ii} by $\bar{\omega}_{ii}$ and for $i \neq j$, ω_{ij} by $\frac{1}{2}\bar{\omega}_{ij}$ in (3.8) and adopt the convention $\bar{\omega}_{ij} = \bar{\omega}_{ji}$ for $i > j$.

§ 4. The Euler Class

In this section we calculate the Euler class of the bundle $M = \Gamma \backslash GL^+_{2m}\mathbb{R} \times_{SO_{2m}} \mathbb{R}^{2m}$. The principal SO_{2m} bundle associated to M is $\Gamma \backslash GL^+_{2m}\mathbb{R}$. Let s

$=s_{2m}$ be as above and write Π for the projection

$$\Pi\colon\ gl_{2m}R\to so_{2m}$$

along s. Denote by ω the canonical left invariant $gl_{2m}\mathbb{R}$ valued one form on $GL^+_{2m}\mathbb{R}$. Since s is invariant under the adjoint representation of so_{2m} on $gl_{2m}\mathbb{R}$, the so_{2m} valued one form

$$\theta=\Pi\circ\omega$$

is an SO_{2m} connection on $\Gamma\backslash GL^+_{2m}\mathbb{R}$, and it is given by

$$\theta_{ij}=\tfrac{1}{2}(\omega_{ij}-\omega_{ji}).$$

The Maurier-Cartan equation for $GL^+_{2m}\mathbb{R}$ is $d\omega+\omega\wedge\omega=0$, so the curvature $\Omega=d\theta+\theta\wedge\theta$ of θ is given by

$$\Omega_{ij}=-\tfrac{1}{4}\sum_{k=1}^{2m}(\omega_{ik}+\omega_{ki})\wedge(\omega_{kj}+\omega_{jk}).$$

Pulled back to $GL^+_{2m}\mathbb{R}$ this form is left invariant so $L_g^*\Omega=\Omega$. Thus the pull back to $GL^+_{2m}\mathbb{R}$ by $\bar{\rho}\circ L_g$ of a form χ on N representing the Euler class of the bundle M is given by

$$\chi(\Omega)=\frac{(-1)^m}{2^{2m}\pi^m m!}\sum\varepsilon_{i_1\dots i_{2m}}\Omega_{i_1i_2}\wedge\dots\wedge\Omega_{i_{2m-1}i_{2m}}$$

where the sum is over all permutations of $1,\dots,2m$ and $\varepsilon_{i_1\dots i_{2m}}$ is the sign of the permutation.

Now $i_Y\Omega_{ij}=0$ for all $Y\in so_{2m}$, so Ω is completely determined by its restriction to s, and

$$\Omega_{ij}|_s=-\sum_{k=1}^{2m}\omega_{ik}\wedge\omega_{jk}.$$

Thus $\chi(\Omega)$ is completely determined by its restriction to s, and

$$\chi(\Omega)|_s=\frac{1}{2^{2m}\pi^m m!}\sum\varepsilon_{i_1\dots i_{2m}}\sum_{\substack{k_1=1\\ \vdots\\ k_m=1}}^{2m}\omega_{i_1k_1}\omega_{i_2k_1}\dots\omega_{i_{2m-1}k_m}\omega_{i_{2m}k_m}. \tag{4.1}$$

By permuting the ω_{ij}'s we may write each

$$\varepsilon_{i_1\dots i_{2m}}\omega_{i_1k_1}\omega_{i_2k_1}\dots\omega_{i_{2m-1}k_m}\omega_{i_{2m}k_m}$$

as

$$\omega_{1j_1}\omega_{2j_2}\dots\omega_{2mj_{2m}}$$

where the indices $j_1,\dots,j_{2m}$ come in pairs $j_{i_1}=j_{i_2}$, $j_{i_3}=j_{i_4},\dots$. Thus

$$\chi(\Omega)|_s=\frac{1}{2^{2m}\pi^m m!}\sum B_{j_1\dots j_{2m}}\omega_{1j_1}\dots\omega_{2mj_{2m}}$$

where the sum is over all $j_1, \ldots, j_{2m}$ with $x_{j_1} \ldots x_{j_{2m}}$ an even function on R^{2m}. The coefficient $B_{j_1 \ldots j_{2m}}$ is given by

Lemma 4.2.

$$B_{j_1 \ldots j_{2m}} = 2^m m! A_{j_1 \ldots j_{2m}}.$$

Proof. We need to count how many times $\omega_{1j_1} \ldots \omega_{2mj_{2m}}$ occurs in (4.1). This term comes from a term of the form $\Omega_{i_1 i_2} \ldots \Omega_{i_{2m-1} i_{2m}}$. There are $2^m m!$ terms of this exact type in $\chi(\Omega)$, obtained by permuting the Ω_{ij}'s and by switching the i and j in a single Ω_{ij}. We now assume that $\omega_{1j_1} \ldots \omega_{2mj_{2m}}$ comes from a term of the form $\Omega_{i_1 i_2} \ldots \Omega_{i_{2m-1} i_{2m}}$ where $i_1 < i_3 < i_5 \ldots$ and each $i_{2k-1} < i_{2k}$. Suppose that the j's in $\omega_{1j_1} \ldots \omega_{2mj_{2m}}$ come in r distinct groups of equal integers which are $2a_1, \ldots, 2a_r$ in number. By symmetry we may assume that our term is of the form

$$\omega_{1j_1} \cdots \omega_{2a_1 j_1} \omega_{2a_1+1\, j_2} \cdots \omega_{2a_1+2a_2 j_2} \omega_{2a_1+2a_2+1\, j_3} \cdots.$$

We get one term of this form from each $\Omega_{i_1 i_2} \ldots \Omega_{i_{2m-1} i_{2m}}$ where

$i_1 < i_3 < \ldots < i_{2m-1}$, and $i_{2k-1} < i_{2k}$ and

$i_1, \ldots, i_{2a_1}$ is a permutation of $1, \ldots, 2a_1$

$i_{2a_1+1}, \ldots, i_{2a_1+2a_2}$ is a permutation of $2a_1+1, \ldots, 2a_1+2a_2$.

$\vdots$

It is easy to see that there are precisely $A_{j_1 \ldots j_{2m}}$ such terms.

Thus we have shown

Proposition 4.3. The Euler class of the bundle M is represented on N by a form whose pull back to $GL^+_{2m}\mathbb{R}$ is a left invariant form which is determined by its restriction to the subspace s. This restricted form is

$$\frac{1}{(2\pi)^m} \sum A_{j_1 \ldots j_{2m}} \omega_{1j_1} \cdots \omega_{2mj_{2m}}$$

where the sum is over all $j_1, \ldots, j_{2m}$ with $x_{j_1} \ldots x_{j_{2m}}$ an even function on $\mathbb{R}^{2m}$.

The fact that

$$\mathrm{Res}_{c_q}(\tau, X_I, N) = \chi(M)$$

is immediate from (4.3) and (3.7). This completes the proofs of Theorems 3.4 and 2.3.

§5. Applications

1. Using Theorem 2.3 one can now easily rederive many of the results of Sect. 5 of [6]. The interested reader is referred to that paper.

2. Let X be a linear Γ vector field for a codimension q foliation τ on M and suppose that X has isolated singularities. Assume the normal bundle ν of τ is oriented and for $N \in \mathscr{A}(X)$, denote by

$$\alpha_N \colon H^*(gl_q\mathbb{R}, SO_q) \to H^*(N, \mathbb{R})$$

the map determined by ν_0, the normal bundle of τ restricted to N. Let A_N be the matrix associated to X on N.

Let

$$j_N\colon H^*(N;\mathbb{R})\to H^{*+q}(M;\mathbb{R})$$

be the map given in Theorem 2.1.

For $\hat{c}_I c_J=\hat{c}_{i_1}\dots\hat{c}_{i_k}c_{j_1}\dots c_{j_l}\in I_q(WO_q)$, we have

Theorem 5.1.

$$\beta_\tau(\hat{c}_I c_J)=\sum_{N\in\mathscr{A}(X)}\frac{c_J(A_N)}{c_q(A_N)}j_N\circ\alpha_N(h_{i_1}\dots h_{i_k}\chi).$$

This follows immediately from Theorems 2.1 and 2.3. Note that $I=(i_1,\dots,i_k)$ may be vacuous.

Corollary 5.2. *If $\chi(\nu)=0$ and for some $\hat{c}_I c_J$ as in* 5.1, *$\beta_\tau(\hat{c}_I c_J)\neq 0$, then τ does not admit a linear Γ vector field with isolated singularities.*

We give two examples where 5.2 applies.

a) The foliation τ_λ of Example 5.1 of [6] has odd codimension and satisfies $\beta_{\tau_\lambda}(\hat{c}_1 c_1^{2m-1})\neq 0$. Thus τ_λ does not admit any linear Γ field with isolated singularities.

b) The foliation $\Pi^*(\tau_\lambda)$ of Theorem 6.3 of [6] has trivial normal bundle and satisfies $\beta_{\Pi^*(\tau_\lambda)}(\hat{c}_1 c_1^{2m-1})\neq 0$. Thus the foliation $\Pi^*(\tau_\lambda)$ with trivial normal bundle does not admit any linear Γ vector field with isolated singular set.

Corollary 5.3. *Let τ be a codimension q foliation on a compact manifold M with oriented normal bundle ν. Suppose all leaves of τ have subexponential growth. If τ admits a linear Γ vector field X with isolated singularities, then*

$$\beta_\tau(\hat{c}_I c_J)=0$$

for all $\hat{c}_I c_J\in I_q(WO_q)$.

Proof. Let $N\in\mathscr{A}(X)$ and note that N must be compact. The holonomy of N must have subexponential growth so the linear holonomy $h\colon \Pi_1(N)\to GL_q^+\mathbb{R}$ must also have subexponential growth. Let $\nu_0=\nu|_N$ and consider S_0, the sphere bundle associated to ν_0. $S_0=\tilde{N}\times_h S^{q-1}$, where $GL_q^+\mathbb{R}$ acts on S^{q-1} as the space of directed lines in R^q. Then S_0 has a natural foliation coming from the flat structure and all leaves of this foliation must also have subexponential growth. By [12], there is a class in $H_*(S_0)$ which is dual to the class of the fiber of S_0. Thus $\chi(S_0)=0$, but $\chi(\nu_0)=\chi(S_0)$.

In this regard see also [4] and [11], where Corollary 5.3 is proven in codimension 1 without requiring the existence of X (i.e., the Sullivan conjecture for the Godbillon-Vey invariant).

3. In [1], Bott showed that for any subbundle ν of TM of dimension q, the Pontrjagin ring Pont$^*(\nu)$ of ν in dimensions greater than $2q$ is an obstruction to ν being homotopic to the normal bundle of a codimension q foliation on M. For q even and $\chi(\nu)=0$, we may now interpret the Pontrjagin ring of ν in dimension $2q$ as an obstruction to ν being homotopic to the normal bundle of a foliation which admits a linear Γ vector field with isolated singularities. This follows directly from Theorem 5.1 and the fact that for $2j_1+\dots+2j_l=q$,

$$\beta_\tau(c_{2j_1}\dots c_{2j_l})=(-1)^{q/2}p_{j_1}\dots p_{j_l}(\nu),$$

where $p_i(\nu)$ is the ith Pontrjagin class of ν.

4. Denote by $\widehat{\mathrm{Diff}^+} S^{2m-1}$ the group of smooth orientation preserving diffeomorphisms of S^{2m-1} equipped with the discrete topology. Let $f: N \to B\widehat{\mathrm{Diff}^+} S^{2m-1}$ be a map and let M_0 be the associated flat S^{2m-1} bundle. For $\varphi \in H^k(WO_{2m-1})$, the natural flat foliation τ_0 of M_0 determines the class $\beta_{\tau_0}(\varphi) \in H^k(M_0; \mathbb{R})$ and so the class $\int \beta_{\tau_0}(\varphi) \in H^{k-2m+1}(N; \mathbb{R})$. This construction is natural so there is a universal map Φ and a commutative diagram

$$\begin{array}{ccc} H^k(WO_{2m-1}) & \xrightarrow{\Phi} & H^{k-2m+1}(B\widehat{\mathrm{Diff}^+} S^{2m-1}; \mathbb{R}) \\ & {\scriptstyle \int \beta \tau_0} \searrow \quad \swarrow {\scriptstyle f^*} & \\ & H^{k-2m+1}(N; \mathbb{R}). & \end{array}$$

Let $\widehat{GL^+_{2m}}\mathbb{R}$ be $GL^+_{2m}\mathbb{R}$ with the discrete topology. The usual action of $GL^+_{2m}\mathbb{R}$ on $\mathbb{R}^{2m}$ induces an action on S^{2m-1} and so also an inclusion

$$\psi: \widehat{GL^+_{2m}}\mathbb{R} \to \widehat{\mathrm{Diff}^+} S^{2m-1}.$$

This gives rise to

$$\psi^*: H^*(B\widehat{\mathrm{Diff}^+} S^{2m-1}; \mathbb{R}) \to H^*(B\widehat{GL^+_{2m}}\mathbb{R}; \mathbb{R})$$

and for any $g: N \to B\widehat{GL^+_{2m}}\mathbb{R}$, the diagram

$$\begin{array}{ccccc} H^*(WO_{2m-1}) & \xrightarrow{\Phi} & H^*(B\widehat{\mathrm{Diff}^+} S^{2m-1}; \mathbb{R}) & \xrightarrow{\psi^*} & H^*(B\widehat{GL^+_{2m}}\mathbb{R}, \mathbb{R}) \\ & {\scriptstyle \int \beta \tau_0} \searrow & \downarrow {\scriptstyle (g\Psi)^*} & \swarrow {\scriptstyle g^*} & \\ & & H^*(N; \mathbb{R}) & & \end{array}$$

commutes.

For $\hat{c}_I c_J = \hat{c}_{i_1} \ldots \hat{c}_{i_k} c_{j_1} \ldots c_{j_l}$ an element of the Vey basis of $H^*(WO_{2m-1})$, set

$$\delta \hat{c}_I c_J = \hat{c}_{i_2} \ldots \hat{c}_{i_k} c_{i_1} c_{j_1} \ldots c_{j_l}.$$

Theorem 4.2 of [6] says that for the flat $\mathbb{R}^{2m}$ bundle M determined by g,

$$\int \beta_{\tau_0}(\hat{c}_I c_J) = \mathrm{Res}_{\delta c_I c_J}(\tau, X_I, N).$$

Here τ is the natural flat foliation on M, and τ_0 is the foliation on the flat S^{2m-1} bindle over N determined by $\psi \circ g$. In particular, if $\hat{c}_i c_J \in H^{4m-1}(WO_{2m-1})$ we have

$$\int \beta_{\tau_0}(\hat{c}_i c_J) = \mathrm{Res}_{c_i c_J}(\tau, X_I, N) = \frac{c_i c_J(I)}{c_{2m}(I)} \cdot \chi(M).$$

This means that the classes $\Phi(\hat{c}_i c_J)$ and χ, the Euler class, in $H^{2m}(B\widehat{\mathrm{Diff}^+} S^{2m-1}; \mathbb{R})$ are mapped to linearly dependent classes by Ψ^*. However, these two classes are linearly independent in $H^{2m}(B\widehat{\mathrm{Diff}^+} S^{2m-1}; \mathbb{R})$. For $m=1$, this is a result of Thurston [2]. For $m>1$ it follows from example 5.1

and the remark on p. 453 of [6]. There the foliation τ_λ on the bundle $M_0 = (\Gamma \backslash G) \times_K S^{2m-1}$ is transverse to the fibers of M_0 so M_0 is a $\widehat{\mathrm{Diff}}^+ S^{2m-1}$ bundle. One calculates

$$\int \beta_{\tau_\lambda}(\hat{c}_i c_J) = \mathrm{Res}_{c_i c_J}(\tau, X_\lambda, N) = \frac{c_i c_J(A_\lambda)}{c_{2m}(A_\lambda)} \chi(M)$$

which varies with a variation in the λ's. Thus $\Phi(\hat{c}_i c_J)$ and χ are linearly independent in $H^*(B\widehat{\mathrm{Diff}}^+ S^{2m-1}; \mathbb{R})$.

References

1. Bott, R.: On a topological obstruction to integrability. Proceedings of Symposia in Pure Mathematics, Vol. 10, Amer. Math. Society, 1970, pp. 127–131
2. Bott, R.: On some formulas for the characteristic classes of group actions (Appendix by R. Brooks). In: Differential topology, foliations, and Gelfand Fuks cohomology. Lecture Notes Math. vol. 652. Berlin-Heidelberg-New York: Springer 1978
3. Bott, R., Haefliger, A.: On characteristic classes of Γ foliations. Bull. A.M.S. **78**, 1038–1044 (1972)
4. Duminy, G.: L'invariant de Godbillon-Vey d'un feuilletage se localise dans les feuilles ressort, preprint
5. Godbillon, C.: Cohomologie d'algebres de Lie de champs de vectures formels. Sem. Bourbaki 421.01–421.19
6. Heitsch, J.: Independent variation of secondary classes. Annals of Mathematics **108**, 421–460 (1978)
7. Heitsch, J.: Linearity and residues for foliations. Boletim da Sociedade Brasileira de Matematica **12**, 87–94 (1981)
8. Kamber, F., Tondeur, Ph.: Foliated bundles and characteristic classes. Lecture Notes in Math., vol. 493. Berlin-Heidelberg-New York: Springer 1975
9. Kamber, F., Tondeur, Ph.: On the linear independence of certain cohomology classes of $B\Gamma_q$. Adv. in Math. Suppl. Studies, vol. 5, pp. 213–263. New York: Academic Press 1979
10. Lazarov, C.: A permanence theorem for exotic classes. J. Diff. Geo. **14**, 475–486 (1979)
11. Mitzutani, T., Morita, S., Tsuboi, T.: The Godbillon-Vey class of codimension one foliations which are almost without holonomy. Annals of Mathematics **113**, 515–527 (1981)
12. Plante, J.: Foliations with measure preserving holonomy. Annals of Mathematics **102**, 327–361 (1975)
13. van Est, W.: Une application d'une methods de Cartan-Leray. Indag. Math. **17**, 542–544 (1955)

Oblatum 21-VII-1982 & 16-III-1983

Invent. math. 73, 287–321 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

An a Priori Estimate for the Oscillation of the Normal to a Hypersurface Whose First and Second Variation With Respect to an Elliptic Integrand is Controlled

W.K. Allard

Department of Mathematics, Duke University, Durham, NC 27706, USA

Table of Contents

Introduction 287
1. Preliminary Material and Basic Definitions and Hypotheses 288
2. Some Basic Estimates 302
3. The Main Lemma 311
4. The Main Theorem 328

Introduction

The object of this paper is to prove for any elliptic integrand the analogue for such an integrand of Theorem 1 of a recent paper [11] of Richard Schoen and Leon Simon. We now state the main implication of our results. Suppose Z is an open ball in $\mathbf{R}^{n+1}$ with center c and Φ is a positive parametric integrand of degree 1 and of class 3 on Z in the sense of [7, 5.1.1]; this amounts to specifying for each $z \in Z$ a non-negative function Φ_z on $\mathbf{R}^{n+1}$ such that (i) $\Phi_z(t\alpha) = t\Phi_z(\alpha)$ whenever $0 \leqq t < \infty$ and $\alpha \in \mathbf{R}^{n+1}$, (ii) $\{\alpha : \Phi_z(\alpha) < 1\}$ is a uniformly convex open set containing the origin, (iii) $\Phi_z(\alpha)$ is of class 3 in (z, α) on $Z \times (\mathbf{R}^{n+1} \sim \{0\})$. Suppose M is an oriented hypersurface of class 2 embedded in Z which is closed relative to Z and suppose ν is a unit normal field along M which is compatible with the orientation of M. Let υ_1 and υ_2 be such that

$$\left| \frac{d}{dt} \int \Phi_z(\nu_t(z))\, d\|M_t\|\, z \Big|_{t=0} \right| \leqq \upsilon_1 \int |\phi|\, d\|M_0\| \tag{1}$$

and

$$\left(\frac{d}{dt}\right)^2 \int \Phi_z(\nu_t(z))\, d\|M_t\|\, z \Big|_{t=0} \geqq -\upsilon_2 \int \phi^2\, d\|M_0\| \tag{2}$$

This research was partially supported by the National Science Foundation

whenever ϕ is a real valued continuously differentiable function on M with compact support and where, for sufficiently small t, $h_t(z)=z+t\,\phi(z)\,\nu(z)$ for $z\in M$, $M_t=h_t(M)$, $\nu_t(h_t(z))$ is normal to M_t at $h_t(z)$, $\nu_t(z)$ depends continuously on (z,t) and $\nu_0=\nu$. Suppose $c\in M$, u is a unit vector in $\mathbf{R}^{n+1}$ and ε is a positive real number; we shall prove the existence of a positive real number δ and a radius R not exceeding the distance from c to the complement of Z, depending only on Φ and not on M, such that

$$1-\nu(c)\cdot u^2\leqq\varepsilon$$

provided $\upsilon_1 R\leqq\delta$, $\upsilon_2 R^2\leqq\delta$, the n-area of the intersection of M with the ball of radius R centered at c does not exceed $1/\varepsilon$ and $|(z-c)\cdot u|\leqq\delta$ whenever $z\in M$ and $|z-c|\leqq R$. In fact, we prove somewhat more than this but the statement above is the main result.

In case Φ is the length integrand of degree one on Z and M is Φ minimizing in the sense of [7, 5.1.6] the conclusion of the above statement is due to De Giorgi [6]; in case Φ is an arbitrary elliptic integrand and M is Φ minimizing the conclusion of the above statement is due to Almgren [1]. In case Φ is, essentially, the length integrand of a Riemannian metric on Z the above statement is due to Schoen and Simon [11]. In case Φ is the length integrand of a Riemannian metric on Z the above statement is proved in [2] *with* the additional hypothesis that the n area of the intersection of M with the ball of radius of R with center c is near the area of an n disk with this radius but *without* the hypothesis $\upsilon_2 R^2\leqq\delta$. Whether the corresponding theorem is true for an arbitrary elliptic integrand is unknown. That a condition on the second variation is necessary in the absence of a stringent condition on area is illustrated by simple examples such as catenaries of revolution; see [5].

Theorem 1 of [11] along with other results is used there to obtain existence theorems along the lines of those obtained by Pitts [10]. We are working to obtain results along those lines for more general integrands but cannot as yet claim any substantial success. In carrying out this work we use the same general scheme that Schoen and Simon used. Just about everything needs to be reproved, however. The point of departure is the second variation formula in [4]. Of independent interest is our lower bound for mass ratios obtained in 2.2(9). We have been unable as yet to obtain an upper bound for mass ratios and that is why we cannot allow *a priori* a singular set of codimension 2 as Schoen and Simon did. I feel that persistence will be rewarded with success in this instance. The main motivation for allowing the singular set is to be able to apply Pitts' arguments; there remains a considerable amount of work to do before this becomes feasible for general integrands.

1. Preliminary Material and Basic Definitions and Hypotheses

We fix once and for all an integer $n\geqq 2$. The basic assumptions of the paper are set forth in 1.2 and 1.4. Unless we indicate otherwise, we use the notation

and terminology of [7]; those not familiar with such may consult the comprehensive glossaries of [7, pp. 669–671].

Throughout this paper,

$$\mathbf{c}_1, \ldots, \mathbf{c}_{13}$$

will be positive real numbers, depending only on n, whose definitions when not given explicitly will be clear from the context in which they first appear. Furthermore,

$$\zeta_1, \ldots, \zeta_{26}$$

will be real numbers *not less than* 1, depending only on n and the numbers

$$m, \Gamma_{0,0}, \Gamma_{0,1}, \Gamma_{0,2}, \Gamma_{0,3},$$

introduced in 1.2, whose definitions when not given explicitly will be clear from the context in which they first appear.

In 1.1 we give some notation. In 1.2 we introduce the integrand. In Sect. 1.3 we set up our calculus on hypersurfaces. In Sect. 1.4 we fix our hypersurface and make our variational assumptions. In Sect. 1.5 we recall the results of [4]. In Sect. 1.6–1.11 we give, in the order in which they will be used, some lemmas and other material of an elementary nature which will be used in later chapters but whose treatment there would be distracting.

In this work we will use frequently and without acknowledgement the Schwarz inequality as well as the inequality

$$|ab| \leqq \tau a^2 + \frac{1}{4\tau} b^2$$

which holds whenever $a, b \in \mathbf{R}$ and $0 < \tau < \infty$.

1.1. We let

$$\mathbf{R}^{n+1} \quad \text{and} \quad \mathbf{S}^n$$

be $(n+1)$ dimensional Euclidean space and the set of unit vectors in $\mathbf{R}^{n+1}$, respectively. We let

$$e_1, \ldots, e_{n+1} \quad \text{and} \quad e^1, \ldots, e^{n+1}$$

be the standard basic vectors of $\mathbf{R}^{n+1}$ and $\operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R})$, respectively. We let

$$\boldsymbol{\beta}: \mathbf{R}^{n+1} \to \operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R})$$

be such that $\boldsymbol{\beta}(u)(v) = u \cdot v$ whenever $u, v \in \mathbf{R}^{n+1}$. Whenever $0 \leqq p < \infty$ we let

$$\mathscr{H}^p$$

be p dimensional Hausdorff measure on $\mathbf{R}^{n+1}$.

We let

$$\mathscr{L}^n$$

be Lebesgue measure on $\mathbf{R}^n$. Whenever $a \in \mathbf{R}^n$ and $0 < r < \infty$ we let

$$\mathbf{U}(a, r) = \{x \in \mathbf{R}^n : |x - a| < r\}.$$

We let

$$\alpha(n) = \mathscr{L}^n[\mathbf{U}(0,1)].$$

Whenever the context makes it appropriate, we identify $(x, y) \in \mathbf{R}^n \times \mathbf{R}$ with $(x_1, \ldots, x_n, y) \in \mathbf{R}^{n+1}$ and we identify $x \in \mathbf{R}^n$ with $(x_1, \ldots, x_n, 0) \in \mathbf{R}^{n+1}$. We let

$$\mathbf{p}: \mathbf{R}^{n+1} \to \mathbf{R}^n \quad \text{and} \quad \mathbf{q}: \mathbf{R}^{n+1} \to \mathbf{R}$$

carry $z = (x, y) \in \mathbf{R}^{n+1}$ to $x \in \mathbf{R}^n$ and $y \in \mathbf{R}$, respectively. Whenever $a \in \mathbf{R}^n$ and $0 < r < \infty$ we let

$$\mathbf{C}(a, r) = \mathbf{p}^{-1}[\mathbf{U}(a, r)].$$

1.2. **The Integrand.** *Throughout this paper* (1)–(5) *below hold.*

(1) *Z is an open subset of $\mathbf{R}^{n+1}$.*

(2) *Φ is a parametric integrand of degree* 1 *and of class* 3 *on Z* in the sense of [7, 5.1.1]; thus

$$\Phi: Z \times \mathbf{R}^{n+1} \to \mathbf{R},$$

$\Phi z, t\alpha) = t\Phi(z, \alpha)$ whenever $(z, \alpha) \in Z \times \mathbf{R}^{n+1}$ and $0 \leqq t < \infty$, and $\Phi | Z \times (R^{n+1} \sim \{0\})$ is of class 3.

(3) $$m = \inf\{\Phi(z, \alpha): (z, \alpha) \in Z \times \mathbf{S}^n\} > 0;$$

thus Φ is *positive* in the sense of [7, 5.1.1].

(4) $$\langle (0, \beta) \odot (0, \beta), D^2 \Phi(z, \alpha) \rangle \geqq |\alpha \wedge \beta|^2$$

whenever $(z, \alpha) \in Z \times \mathbf{S}^n$ and $\beta \in \Lambda_n \mathbf{R}^{n+1}$; thus *$\Phi$ is elliptic with ellipticity bound* 1 *at each point of Z* in the sense of [7, 5.1.2].

(5) *For each ordered pair (k, l) of nonnegative integers such that $k + l \leqq 3$ we have*

$$\Gamma_{k,l} < \infty$$

where $\Gamma_{k,l}$ is the supremum of the set of numbers

$$|\langle (\beta_1, 0) \odot \ldots \odot (\beta_k, 0) \odot (0, \gamma_1) \odot \ldots \odot (0, \gamma_l), D^{k+l} \Phi(z, \alpha) \rangle|$$

corresponding to $(z, \alpha) \in Z \times S^n$ and $\beta_1, \ldots, \beta_k; \gamma_1, \ldots, \gamma_l \in \mathbf{S}^n$.

From the homogeneity condition that Φ satisfies one readily verifies that

(6) $$\langle (0, \beta)^k, D^k \Phi(z, t\alpha) \rangle = t^{1-k} \langle (0, \beta)^k, D^k \Phi(z, \alpha) \rangle,$$

(7) $$(0, \alpha) \lrcorner D^k \Phi(z, \alpha) = (2 - k) D^{k-1} \Phi(z, \alpha)$$

whenever $k \in \{1, 2, 3\}$, $\alpha, \beta \in \mathbf{R}^{n+1}$, $\alpha \neq 0$ and $0 < t < \infty$. We let

$$*\Phi: Z \times \Lambda_n \mathbf{R}^{n+1} \to \mathbf{R}$$

be such that $*\Phi(z, \alpha) = \Phi(z, *\alpha)$ whenever $(z, \alpha) \in Z \times \Lambda_n \mathbf{R}^{n+1}$, where $*$ is the Hodge star operator as in [7, 1.7.8].

According to [7, 1.5.3],

$$(u \cdot v)\,\eta = u \wedge (\eta \llcorner \boldsymbol{\beta}(v)) + (u \wedge \eta) \llcorner \boldsymbol{\beta}(v) \tag{8}$$

whenever $u, v \in \mathbf{R}^{n+1}$ and $\eta \in \Lambda_n \mathbf{R}^{n+1}$;
with the help of (8) we verify that

$$*[u \wedge (\eta \llcorner \boldsymbol{\beta}(v))] = (u \cdot v)\,v = (u \cdot v)\,v \tag{9}$$

whenever $u, v \in \mathbf{R}^{n+1}$, $\eta \in \Lambda_n \mathbf{R}^{n+1}$ *and* $v = *\eta$. With the help of (8) we infer that

$$\langle (t, u \wedge (\eta \llcorner \boldsymbol{\beta}(v)))^k, D^k(*\Phi)(z, v) \rangle = \langle (t, (u \cdot v)\,v - (u \cdot v)\,v)^k, D^k \Phi(z, v) \rangle \tag{10}$$

whenever $k \in \{1, 2, 3\}$; $t, u, v \in \mathbf{R}^{n+1}$, $(z, v) \in Z \times \Lambda_n \mathbf{R}^{n+1}$ *and* $v = *\eta \neq 0$.

For each $s \in \{1, -1\}$ *we let*

$$\Phi_s^{\S} : Z \times \mathbf{R}^n \to \mathbf{R}$$

be such that $\Phi_s^{\S}(x, y, \alpha) = \Phi((x, y), s(-\alpha, 1))$ *whenever* $(x, y, \alpha) \in Z \times \mathbf{R}^n$. Evidently,

$$\langle (u, v, \beta)^k, D\Phi_s^{\S}(x, y, \alpha) \rangle = \langle ((u, v), s(-\beta, 0))^k, D\Phi((x, y), s(-\alpha, 1)) \rangle \tag{11}$$

whenever $s \in \{1, -1\}$, $k \in \{1, 2, 3\}$, $(x, y, \alpha) \in Z \times \mathbf{R}^n$ *and* $(u, v, \beta) \in \mathbf{R}^n \times \mathbf{R}^n$.

We define

$$\Sigma, \Sigma_1 : Z \times \mathbf{S}^n \to \operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R}^{n+1}),$$
$$\Sigma_2, \Sigma_3 : Z \times \mathbf{S}^n \to \mathbf{R},$$
$$\Sigma_4, \Sigma_6 : Z \times S^n \to \mathbf{R}^{n+1},$$
$$\Sigma_5 : Z \times \mathbf{S}^n \to \operatorname{Hom}(\mathbf{R}^{n+1}, \operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R}^{n+1}))$$

at $(z, \alpha) \in Z \times \mathbf{S}^n$ *by requiring that*

$$\langle u, \Sigma(z, \alpha) \rangle \cdot v = \langle (0, u) \odot (0, v), D^2 \Phi(z, \alpha) \rangle,$$
$$\langle u, \Sigma_1(z, \alpha) \rangle \cdot v = \langle (0, u) \odot (v, 0), D^2 \Phi(z, \alpha) \rangle,$$
$$\Sigma_2(z, \alpha) = \langle (\alpha, 0), D\Phi(z, \alpha) \rangle,$$
$$\Sigma_3(z, \alpha) = \langle (\alpha, 0) \odot (\alpha, 0), D^2 \Phi(z, \alpha) \rangle,$$
$$\Sigma_4(z, \alpha) \cdot v = \sum_{i=1}^{n+1} \langle (0, e_i) \odot (e_i, 0) \odot (0, v), D^3 \Phi(z, \alpha) \rangle,$$
$$\langle u, \Sigma_5(z, \alpha) \rangle \cdot l = \sum_{i=1}^{n+1} \langle (0, e_i) \odot (0, l(e_i)) \odot (u, 0), D^3 \Phi(z, \alpha) \rangle,$$
$$\Sigma_6(z, \alpha) \cdot u = \sum_{i=1}^{n+1} \langle (0, e_i) \odot (e_i, 0) \odot (u, 0), D^3 \Phi(z, \alpha) \rangle,$$

whenever $u, v \in \mathbf{R}^{n+1}$ *and* $l \in \operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R}^{n+1})$.

1.3. **Submanifolds.** *We assume throughout this section that*
(1) *W is a finite dimensional inner product space.*
(2) *j and k are integers,* $0 \leqq j \leqq k$ *and* $1 \leqq k$.

(3) *M is an n dimensional submanifold of class k of $\mathbf{R}^{n+1}$ in the sense of* [7, 3.1.19]; that is, $M \subset \mathbf{R}^{n+1}$ and for each $c \in M$ there are U, V, ϕ, ψ such that U is an open subset of $\mathbf{R}^n$, V is an open subset of $\mathbf{R}^{n+1}$, $c \in V$, $\phi: U \to V$, ϕ is of class k, $\psi: V \to U$, ψ is of class k, $\psi \circ \phi$ is the identity map of U and $V \cap M = \text{range } \phi$.

(4) *v is such that $v: M \to \mathbf{S}^n$, v is continuous and $v(z) \in \text{Nor}(M, z)$ whenever $z \in M$.*

We adopt the definitions which follow throughout this paper. We let

$$\mathscr{M}^k$$

be the set of those ordered pairs (M, v) as in (3), (4) above. We let

$$\mathscr{E}^j(M, W) \quad \text{and} \quad \mathscr{L}(M, W)$$

be the vector spaces of those f such that $f \circ \phi$ is of class j and $f \circ \phi$ is locally Lipschitzian, respectively, whenever U is an open subset of R^n, $\phi: U \to M$ and ϕ is of class k. Whenever $f: M \to W$ we let the support of f or, more briefly,

$$\text{spt } f$$

equal the closure relative to M of $\{z \in M: f(z) \neq 0\}$. We let

$$\mathscr{E}^j_c(M, W) \quad \text{and} \quad \mathscr{L}_c(M, W)$$

be the vector spaces of those f such that $f \in \mathscr{E}^j(M, W)$ and $f \in \mathscr{L}(M, W)$, respectively, and spt f is compact. We let

$$\mathscr{E}^j(M), \ \mathscr{L}(M), \ \mathscr{E}^j_c(M), \ \mathscr{L}_c(M)$$

be, respectively, $\mathscr{E}^j(M, \mathbf{R})$, $\mathscr{L}(M, \mathbf{R})$, $\mathscr{E}^j_c(M, \mathbf{R})$, $\mathscr{L}_c(M, \mathbf{R})$. Whenever $f: M \to W$ and $c \in M$ we say *f is differentiable at c* if $f \circ \phi$ is differentiable at b whenever U is an open subset of $\mathbf{R}^n$, $\phi: U \to M$, ϕ is of class k, $b \in U$ and $\phi(b) = c$ in which case we let

$$Df(c) \in \text{Hom}(\mathbf{R}^{n+1}, W),$$

the *differential of f at c*, be determined by the requirement that $\langle v(c), Df(c) \rangle = 0$ and $Df(c) \circ D\phi(b) = D(f \circ \phi)(b)$ whenever ϕ is as above. The reason we arrange that the domain of $Df(c)$ is all of $\mathbf{R}^{n+1}$ and not just $\text{Tan}(M, c)$ is so that, in the event that f is differentiable at all points of M, we have that Df is, like f, a function on M with values in a finite dimensional inner product space. If $f: M \to \mathbf{R}$, $c \in M$ and f is differentiable at c we let

$$\nabla f(c) \in \text{Tan}(M, c),$$

the *gradient of f at c*, be such that $u \cdot \nabla f(c) = \langle u, Df(c) \rangle$ whenever $u \in \mathbf{R}^{n+1}$. We let

$$\|M\|$$

be the measure on $\mathbf{R}^{n+1}$ which assigns to each subset A of $\mathbf{R}^{n+1}$ the n dimensional Hausdorff measure of $A \cap M$.

Suppose $f \in \mathscr{L}(M, W)$. As is well known, f is differentiable at $\|M\|$ almost all points of $\mathbf{R}^{n+1}$ and Df is an $\|M\|$ measurable and locally $\|M\|$ essentially

bounded $\operatorname{Hom}(\mathbf{R}^{n+1}, W)$ valued function; furthermore, whenever G is an open subset of M, $\operatorname{spt} f \subset G$ and $0<\varepsilon<\infty$ there is $g \in \mathscr{C}^k(M, W)$ such that $\operatorname{spt} g \subset G$,

$$|f-g| \leqq \varepsilon \quad \text{and} \quad \int |Df-Dg|\, d\|M\| \leqq \varepsilon. \tag{5}$$

For the remainder of this section we assume that $k \geqq 2$. It is elementary that $v \in \mathscr{C}^{k-1}(M, \mathbf{R}^{n+1})$, that

$$Dv(z)^* = Dv(z) \quad \textit{whenever } z \in M \tag{6}$$

and that

$$\begin{aligned} &\langle u, \langle v, D(Df)(z)\rangle\rangle - \langle v, \langle u, D(Df)(z)\rangle\rangle \\ &\quad = \langle v \cdot \nu(z)\langle u, Dv(z)\rangle - u \cdot \nu(z)\langle v, Dv(z)\rangle, Df(z)\rangle \end{aligned} \tag{7}$$

whenever $f \in \mathscr{C}^2(M, W)$ *and* $u, v \in \mathbf{R}^{n+1}$.

We let

$$H = -\operatorname{trace} Dv \in \mathscr{C}^{k-2}(M) \tag{8}$$

and call H the *mean curvature of* (M, v).

Whenever $X \in \mathscr{C}^1(M, \mathbf{R}^{n+1})$ and $X(z) \in \operatorname{Tan}(M, z)$ whenever $z \in M$ we let

$$\operatorname{div} X = \operatorname{trace} DX \in \mathscr{C}^0(M);$$

as is well known,

$$\int \phi \operatorname{div} X\, d\|M\| = -\int \nabla\phi \cdot X\, d\|M\| \tag{9}$$

whenever $\phi \in \mathscr{C}^1(M)$ *and* $\operatorname{spt}\phi \cap \operatorname{spt} X$ *is compact.*

Whenever X is an $\mathbf{R}^{n+1}$-valued $\|M\|$ summable function, $X(z) \in \operatorname{Tan}(M, z)$ for $\|M\|$ almost all $z \in \mathbf{R}^{n+1}$ and f is an R-valued $\|M\|$ summable function we say

$$\operatorname{div} X = f \textit{ weakly}$$

provided

$$\int \phi f\, d\|M\| = -\int \nabla\phi \cdot X\, d\|M\| \quad \text{whenever } \phi \in \mathscr{C}^1_c(M) \tag{10}$$

and we say

$$\nabla f = X \textit{ weakly}$$

provided

$$\int W \cdot X\, d\|M\| = -\int f \operatorname{div} W\, d\|M\| \tag{11}$$

whenever $W \in \mathscr{C}^1_c(M, \mathbf{R}^{n+1})$ *and* $W(z) \in \operatorname{Tan}(M, z)$ *for* $z \in M$.

Finally, we record the following fundamental isoperimetric theorem.

(12) **Lemma.** *We have*

$$\left(\int |\phi|^{n/(n-1)} d\|M\|\right)^{(n-1)/n} \leqq \mathbf{c}_1 \int |\nabla\phi| + |\phi||H|\, d\|M\|$$

whenever $\phi \in \mathscr{L}_c(M)$;

here $c_1 = B(n+1)\, e\, n^{-1}\, \alpha(n)^{-1/m}$ *as in* [2, 7.1].

Proof. One modifies the proof of [2, 7.3] by replacing the last ten lines of the proof with the following: "By 4.10.2,

$$\int_0^\infty \|\delta V_r\|(U)\,dr \leqq \int |S[\operatorname{grad} f(x)]|\,dV(x,S) + \int_0^\infty \|\delta V\|\{x: r<f(x)\}\,dr;$$

furthermore,

$$\int_0^\infty \|\delta V\|\{x: r<f(x)\}\,dr = \int f\,d\|\delta V\|.$$

By 7.1 applied to V_r, with ϕ replaced by approximations to the characteristic function of $\{x: r<f(x)\}$,

$$\|V_r\|(U)^{1/\beta} \leqq \mathbf{c}_1 \|\delta V_r\|(U).$$

Therefore,

$$\int_0^\infty \|V_r\|(U)^{1/\beta}\,dr \leqq \mathbf{c}_1\left(\int |S[\operatorname{grad} f(x)]|\,dV(x,S) + \int f\,d\|\delta V\|\right)."$$

Remark. Another proof may be found in [9].

1.4. **The Hypersurface.** *Throughout this paper we assume that*

(1) $(M, v)\in\mathscr{M}^2$;

(2) $M\subset Z$ *and* M *is closed relative to* Z;

(3) $i: M\to Z$ *is the inclusion map of* M *into* Z *and* $\eta: M\to \Lambda_n\mathbf{R}^{n+1}$ *is such that* $v = *\eta$;

(4) $L = -Dv\cdot\Sigma(i, v) - \operatorname{trace}\Sigma_1(i, v)$;

(5) $0\leqq v_1<\infty$ *and*

$$|\mathscr{F}(\phi)| \leqq v_1 \int |\phi|\,d\|M\| \text{ whenever } \phi\in\mathscr{C}^1_1(M);$$

(6) $0\leqq v_2<\infty$ *and*

$$\mathscr{S}(\phi) \geqq -v_2\int \phi^2\,d\|M\| \text{ whenever } \Phi\in\mathscr{C}^1_c(M)$$

where, for each $\phi\in\mathscr{C}^1_c(M)$, *with* $h_t(z) = z + t\phi(z)$ *for* $z\in M$, *we set*

$$\mathscr{F}(\phi) = \frac{d}{dt}\int_{\operatorname{spt}\phi} *\Phi(h_t(z), \langle\eta(z), \Lambda_n Dh_t(z)\rangle)\,d\|M\|\big|_{t=0}$$

and

$$\mathscr{S}(\phi) = \left(\frac{d}{dt}\right)^2 \int_{\operatorname{spt}\phi} *\Phi(h_t(z), \langle\eta(z), \Lambda_n Dh_t(z)\rangle)\,d\|M\|\big|_{t=0}.$$

We let

$$\Delta = \max\{\Gamma^2_{1,0}, \Gamma^2_{1,1}, \Gamma^2_{1,2}, v_1^2, \Gamma_{2,0}, \Gamma_{2,1}, v_2\}. \tag{7}$$

1.5. **Incorporation of [4] Into the Present Context.** *The formulae which follow are fundamental for our work.*

(1) $\mathscr{F}(\phi) = -\int \phi L d\|M\|$;

(2) $\mathscr{S}(\phi) = \int \nabla\phi \cdot \langle \nabla\phi, \Sigma(i,v)\rangle - \phi^2 Dv^2 \cdot \Sigma(i,v)\, d\|M\|$
$-2\int \phi \langle \nabla\phi, \Sigma_1(i,v)\rangle \cdot v\, d\|M\|$
$+\int \phi^2 \{\Sigma_3(i,v) - Dv \cdot \Sigma_1(k,v) - H\Sigma_2(i,v)\}\, d\|M\| + \int \phi^2 LH\, d\|M\|$.

These formulae follow from [4]. Indeed, it is evident that the statements and proofs there are valid if M there is of class 3 and Υ there is of class 2. Furthermore, if M there is merely of class 2 the statements there are valid provided one takes divergences in the weak sense as in 1.3(10), and one may prove these statements by representing M locally as the graph in rectangular coordinates of a function of class 2, regularizing this function, applying the statement to the graph of the regularization and then letting the regularization parameter converge to the critical value. Thus [4, 3.11] holds with m, n, Υ, ξ there equal the present n, $n+1$, $*\Phi$, η. To see that this modification of [4, 3.11] gives (1) and (2) we first observe that

$$\langle v, \langle u, D\tau(z)\rangle\rangle = -v \cdot \langle u, Dv(z)\rangle\, v(z) - v \cdot v(z) \langle u, Dv(z)\rangle,$$
$$\langle u, \langle v, A(z)\rangle\rangle = -v \cdot v(z) \langle u, Dv(z)\rangle$$

whenever $z \in M$ *and* $u, v \in \mathbf{R}^{n+1}$

where on the left hand sides of these equations we are using the notation of [4]. Finally, we use these equations as well as 1.1 (1), (2) and 1.2 (7) to calculate

$$\langle Z \otimes \tau \otimes D\tau, s(i,\xi)\rangle = -\phi Dv \cdot \Sigma(i,v);$$
$$\langle Z \otimes \tau, s_1(i,\xi)\rangle - \langle Z, o_1(i,\xi)\rangle = -\phi \operatorname{trace} \Sigma_1(i,v);$$
$$\langle D_v Z \odot D_v Z, q(i,\xi)\rangle = \nabla\phi \cdot \langle \nabla\phi, \Sigma(i,v)\rangle;$$
$$\langle Z \otimes \langle Z, A\rangle \otimes D\tau, s(i,\xi)\rangle = -\phi^2 \langle v \otimes Dv \otimes D\tau, s(i,\xi)\rangle$$
$$= \phi^2 Dv^2 \cdot \Sigma(i,v);$$
$$\langle Z \otimes \langle Z, A\rangle, s_1(i,\xi)\rangle = -\phi^2 \langle v \otimes Dv, s_1(i,\xi)\rangle$$
$$= \phi^2 Dv \cdot \Sigma_1(i,v)$$
$$\langle \phi Z, o_1(i,\xi)\rangle = \phi \Sigma_2(i,v);$$
$$\langle v \odot v, q_2(i,\xi)\rangle = \phi^2 \Sigma_3(i,v);$$
$$\langle D_v Z \otimes Z, q_1(i,\xi)\rangle = \phi \langle (D\phi)\, v \otimes v, q_1(i,\xi)\rangle$$
$$= -\phi \langle \nabla\phi, \Sigma_1(i,v)\rangle \cdot v$$

where we have set $Z = \phi v$ and where, once again, we use the notation of [4] when appropriate.

1.6. The reader might first consider the case when B is the identity below.

(1) **Lemma.** *Suppose* A *and* B *are symmetric endomorphisms of* $\mathbf{R}^{n+1}$ *such that* $A \cdot B = 0$ *and*

$$\xi = \inf\{B(u) \cdot u : u \in \mathbf{S}^n\} > 0.$$

Then

$$\|A \circ B \circ A\| \leqq \max\left\{1 - \frac{\xi}{\operatorname{trace} B}, \frac{\|B\|}{\operatorname{trace} B}\right\} A^2 \cdot B.$$

Remark. Recall that

$$\|C\| = \sup\{|C(u)| : u \in \mathbf{S}^n\} \quad \text{and} \quad |C| = (\operatorname{trace} C^* \circ C)^{1/2}$$

for an endomorphism C of $\mathbf{R}^{n+1}$.

Proof. Suppose B is a symmetric endomorphism of $\mathbf{R}^{n+1}$, $\operatorname{trace} B = 1$, $\xi = \inf\{B(u) \cdot u : u \in \mathbf{S}^n\} > 0$ and

$$\sigma = \max\{1 - \xi, \|B\|\}.$$

Let

$$E = \{(X, x) : X \text{ is a symmetric endomorphism of } \mathbf{R}^{n+1} \text{ and } x \in \mathbf{R}^{n+1}\}$$

and let

$$f(X, x) = X \circ B \circ X(x) \cdot x \quad \text{for } (X, x) \in E.$$

Suppose (A, a) is a point at which f attains its maximum value on

$$C = \{(X, x) \in E : X \cdot B = 0, X^2 \cdot B = 1 \text{ and } |x| = 1\}.$$

We must show that $f(A, a) \leqq \sigma$.

Let

$$I(X, x) = X \cdot B, \quad K(X, x) = X^2 \cdot B, \quad L(X, x) = |x|^2$$

for $(X, x) \in E$; then there are $\iota, \kappa, \lambda \in \mathbf{R}$ such that

$$\operatorname{grad} f(A, a) = \iota \operatorname{grad} I(A, a) + \kappa \operatorname{grad} K(A, a) + \lambda \operatorname{grad} L(A, a).$$

Since

$$\operatorname{grad} f(A, a) = (\boldsymbol{\beta}(B \circ A(a))\, a + \boldsymbol{\beta}(a)\, B \circ A(a), 2A \circ B \circ A(a)),$$
$$\operatorname{grad} I(A, a) = (B, 0),$$
$$\operatorname{grad} K(A, a) = (A \circ B + B \circ A, 0),$$
$$\operatorname{grad} L(A, a) = (0, 2a)$$

we see that

$$\boldsymbol{\beta}(B \circ A(a))\, a + \boldsymbol{\beta}(a)\, B \circ A(a) = \iota B + \kappa(A \circ B + B \circ A), \tag{1}$$

$$A \circ B \circ A(a) = \lambda a. \tag{2}$$

Thus we must show that $\lambda \leqq \sigma$.

It is evident that

$$A \circ B \circ A \quad \text{is symmetric} \tag{3}$$

and that

$$A \circ B \circ A(u) \cdot u \geqq 0 \quad \text{whenever } u \in \mathbf{R}^{n+1}. \tag{4}$$

Since $\xi > 0$ we infer that $\ker A \circ B \circ A \subset \ker A$ from which it follows immediately that

$$\ker A \circ B \circ A = \ker A \tag{5}$$

which in turn implies

$$\operatorname{rng} A \circ B \circ A = \operatorname{rng} A. \tag{6}$$

Since $A^2 \cdot B = \operatorname{trace} A \circ B \circ A$ we have

$$\operatorname{trace} A \circ B \circ A = 1. \tag{7}$$

We now may conclude that

$$0 < \lambda < 1; \tag{8}$$

that $0 \leqq \lambda \leqq 1$ follows from (3), (4), (7); that $\lambda = 0$ is absurd follows from $f(X, x) \leqq f(A, a) = \lambda$ for $(X, x) \in C$. Were it the case that $\lambda = 1$ we would have rank $A \circ B \circ A = 1$ so rank A would equal 1 which is incompatible with $A \cdot B = 0$. Taking the trace of (1) we obtain

$$2B \circ A(a) \cdot a = \iota; \tag{9}$$

taking the inner product of A with (1) we obtain

$$\lambda = \kappa; \tag{10}$$

taking the inner product of (1) with $\boldsymbol{\beta}(a)\, a$ we obtain

$$2B \circ A(a) \cdot a = \iota B(a) + 2\kappa B \circ A(a) \cdot a. \tag{11}$$

In case $B \circ A(a) \cdot a \neq 0$ we infer from (9), (10), (11) that $1 = B(a) \cdot a + \lambda$ which gives $\lambda = 1 - B(a) \cdot a \leqq 1 - \xi \leqq \sigma$.

In case $B \circ A(a) \cdot a = 0$ we infer from (1), (9), (10) that

$$\boldsymbol{\beta}(B \circ A(a))\, a + \boldsymbol{\beta}(a)\, B \circ A(a) = \lambda (A \circ B + B \circ A). \tag{12}$$

Taking the inner product of (12) with $\beta(A(a))\, A(a)$ we obtain

$$2\lambda A(a) \cdot a = 2\lambda^2 A(a) \cdot a \tag{13}$$

which in view of (8) gives

$$A(a) \cdot a = 0. \tag{14}$$

Since $\lambda \neq 0$ we see that $A(a) \neq 0$ so that we may set $b = |A(a)|^{-1} A(a)$ and have $|a| = 1$, $a \cdot b = 0$, $|b| = 1$. We next apply (12) to $A(a)$ and obtain

$$\lambda a = \lambda (\lambda a + B \circ A(A(a))) \tag{15}$$

which gives

$$A \circ B \circ A(a) = \lambda a, \qquad A \circ B \circ A(b) = (1 - \lambda)\, b. \tag{16}$$

From (3), (4), (7) we infer rank $A \circ B \circ A = 2$ and then from (6) that rank $A = 2$ which forces

(17) $$A = \mu(\boldsymbol{\beta}(a)\, b + \boldsymbol{\beta}(b)\, a) \quad \text{for some } \mu \in \mathbf{R} \sim \{0\}.$$

This in turn implies $B(a) \cdot b = 0$ since $A \cdot B = 0$; setting $\delta = B(a) \cdot a$ and $\varepsilon = B(b) \cdot b$ we deduce that

(18) $$A \circ B \circ A(a) = \mu^2 \varepsilon a, \quad A \circ B \circ A(b) = \mu^2 \delta b.$$

Thus $\lambda = \mu^2 \varepsilon$ and $1 = \operatorname{trace} A \circ B \circ A = \mu^2(\varepsilon + \delta)$ so

$$\lambda = \frac{\varepsilon}{\delta + \varepsilon} \leqq \frac{\varepsilon}{\operatorname{trace} B} \leqq \frac{\|B\|}{\operatorname{trace} B} \leqq \|B\| \leqq \sigma.$$

(2) **Lemma.** *Suppose B is a symmetric endomorphism of $\mathbf{R}^{n+1}$ such that*

$$\inf\{B(u) \cdot u : u \in \mathbf{S}^n\} > 0.$$

Then

$$\|A \circ B \circ A\| \leqq \max\left\{1 - \frac{\|B\|}{\operatorname{trace} B}, \frac{\|B\|}{\operatorname{trace} B}\right\} A^2 \cdot B + 6(n+1)^{1/2} \|A\|\, |A \cdot B|$$

whenever A is a symmetric endomorphism of $\mathbf{R}^{n+1}$.

Proof. Suppose A, B, u are as above. Let $\varepsilon = A \cdot B / B \cdot B$, let $C = A - \varepsilon B$ and let

$$\sigma = \max\left\{1 - \frac{\|B\|}{\operatorname{trace} B}, \frac{\|B\|}{\operatorname{trace} B}\right\}.$$

Since $C \cdot B = 0$ we may use the previous Lemma to obtain

$$\begin{aligned}
&A(u) \cdot \langle A(u), B\rangle \\
&\quad = C(u) \cdot \langle C(u), B\rangle + 2\varepsilon B(u) \cdot \langle C(u), B\rangle + \varepsilon^2 B(u) \cdot \langle B(u), B\rangle \\
&\quad \leqq \sigma C^2 \cdot B + 2\varepsilon B(u) \cdot \langle C(u), B\rangle + \varepsilon^2 B(u) \cdot \langle B(u), B\rangle \\
&\quad = \sigma A^2 \cdot B - \varepsilon A \circ B \cdot B - \varepsilon B \circ A \cdot B + \varepsilon^2 B^2 \cdot B \\
&\qquad + 2\varepsilon B(u) \cdot \langle A(u), B\rangle - 2\varepsilon B(u) \cdot \langle \varepsilon B(u), B\rangle + \varepsilon^2 B(u) \cdot \langle B(u), B\rangle
\end{aligned}$$

1.7. **A Formula for First Variation.** *Suppose $g : Z \to \mathbf{R}^{n+1}$ is continuously differentiable with compact support. For each $t \in \mathbf{R}$ let*

$$h_t(z) = z + t g(z) \quad \text{for } z \in M.$$

Then

$$\begin{aligned}
&\frac{d}{dt} \int * \Phi(h_t(z), \langle \eta(z), \Lambda_n D h_t(z)\rangle)\, d\|M\|\, z\big|_{t=0} \\
&\quad = -\int Lg \cdot v\, d\|M\| \\
&\quad = \int \operatorname{trace} Dg\, \Phi(i, v) - \langle (g, \langle v, (Dg)^*\rangle), D\Phi(i, v)\rangle\, d\|M\|.
\end{aligned}$$

Do not confuse $Dg(z)$ **with** $D(g|M)(z)$**!**

The first equation above can be proved in exactly the same way [4, 3.11(1)] is proved. To prove that the first quantity above equals the third we use [4, 2.2(4)] and 1.2(10) to compute

$$\begin{aligned}
&\frac{d}{dt} *\Phi(h_t(z), \langle \eta(z), \Lambda_n Dh_t(z)\rangle)|_{t=0} \\
&\quad = \left\langle \left(g(z), \sum_{i=1}^{n+1} \langle e_i, Dg(z)\rangle \wedge (v(z) \llcorner e^i)\right), D(*\Phi)(z, v(z))\right\rangle \\
&\quad = \left\langle \left(0, \sum_{i=1}^{n+1} \langle e_i, Dg(z)\rangle \cdot e_i\, v(z)\right), D\Phi(z, v(z))\right\rangle \\
&\qquad - \left\langle \left(g(z), \sum_{i=1}^{n+1} \langle e_i, Dg(z)\rangle \cdot v(z)\, e_i\right), D\Phi(z, v(z))\right\rangle \\
&\quad = \operatorname{trace} Dg(z)\, \Phi(z, v(z)) - \langle (g(z), \langle v(z), (Dg(z))*\rangle), D\Phi(z, v(z))\rangle \qquad \text{for } z \in M.
\end{aligned}$$

1.8. **Some Simple Geometric Inequalities.** *Whenever* $u \in \mathbf{S}^n$ *and* $u \cdot e_{n+1} \neq 0$ *we let*

$$\imath(u) = -(u \cdot e_{n+1})^{-1}\, \mathbf{p}(u) \in \mathbf{R}^n$$

and note that $\imath(u)$ is the gradient of a linear function whose graph has u as a normal vector. The reader may verify the inequality

$$(1) \qquad (u \cdot e_{n+1})^2 (v \cdot e_{n+1})^2 |\imath(u) - \imath(v)|^2 \leqq 1 - u \cdot v^2 \leqq (u \cdot e_{n+1})^2 |\imath(u) - \imath(v)|^2$$

which holds for $u, v \in \mathbf{S}^{n+1}$ *with* $u \cdot e_{n+1} \neq 0$ *and* $v \cdot e_{n+1} \neq 0$. It is obvious that

$$(2) \qquad |\imath(u)|^2 = (u \cdot e_{n+1})^{-2} - 1 \quad \textit{for } u \in \mathbf{S}^n \quad \textit{with } u \cdot e_{n+1} \neq 0.$$

The reader may also verify the inequality

$$(3) \qquad (1 - u \cdot v^2)^{1/2} \leqq (1 - u \cdot w^2)^{1/2} + (1 - v \cdot w^2)^{1/2}$$

which holds *for* $u, v, w \in \mathbf{S}^n$.

In fact, it is useful to think about and calculate with $\boldsymbol{\beta}(u) u \in \operatorname{Hom}(\mathbf{R}^{n+1}, \mathbf{R}^{n+1})$ rather than $u \in \mathbf{S}^n$ in all of the above.

1.9. **Some Isoperimetric Lemmas.** Suppose $a \in \mathbf{R}^n$ and $0 < r < \infty$.

Lemma. *Suppose* A *is an open subset of* $\mathbf{U}(a, r)$ *and* B *is the boundary of* A *relative to* $\mathbf{U}(a, r)$. *Then*

$$(1) \qquad \min\{\mathscr{L}^n(A), \mathscr{L}^n(\mathbf{U}(a, r) \sim A)\} \leqq \mathbf{c}_2\, \mathscr{H}^{n-1}(B)^{n/(n-1)}.$$

Proof. See [7; 4.5.3, 4.5.11, 4.5.6].

It is the following corollary of this that we need.

Lemma. *Suppose* T *is a closed subset of* $\mathbf{U}(a, r)$ *and*

$$(2) \qquad \mathscr{H}^{n-1}(T) \leqq \mathbf{c}_3\, r^{n-1}.$$

Then

(3) there is a unique connected component C *of* $\mathbf{U}(a,r)\sim T$ *such that* $\mathscr{L}^n(C)>(1/2)\alpha(n)r^n$;

(4) $$\mathscr{L}^n(\mathbf{U}(a,r)\sim C)\leqq \mathbf{c}_2\mathscr{H}^{n-1}(T)^{n/(n-1)}.$$

Proof. Let $\mathscr{D}$ be the family of connected components of $\mathbf{U}(a,r)\sim T$. Note that $\mathscr{D}$ is a nonempty countable disjointed family of open subsets of $\mathbf{U}(a,r)$ and that the boundary of $\bigcup\mathscr{C}$ relative to $\mathbf{U}(a,r)$ is a subset of T whenever $\mathscr{C}\subset\mathscr{D}$. Were it the case that $\mathscr{L}^n(D)\leqq(1/3)\alpha(n)r^n$ for all D in $\mathscr{D}$ there would exist a subfamily $\mathscr{C}$ of $\mathscr{D}$ such that

$$(1/3)\alpha(n)r^n<\mathscr{L}^n(\textstyle\bigcup\mathscr{C})\leqq(2/3)\alpha(n)r^n.$$

But, according to (1),

$$\begin{aligned}(1/3)\alpha(n)r^n&\leqq\min\{\mathscr{L}^n(\textstyle\bigcup\mathscr{C}),\mathscr{L}^n(\mathbf{U}(a,r)\sim\bigcup\mathscr{C})\}\\&\leqq\mathbf{c}_2\mathscr{H}^n(T)^{n/(n-1)}\\&\leqq\mathbf{c}_2(\mathbf{c}_3r^{n-1})^{n/(n-1)}\\&<(1/3)\alpha(n)r^n,\end{aligned}$$

where we have set $\mathbf{c}_3=(4\mathbf{c}_2)^{-(n-1)/n}\alpha(n)$. So let $C\in\mathscr{D}$ be such that $\mathscr{L}^n(C)>(1/3)\alpha(n)r^n$. Using (1) once again we obtain

$$\min\{\mathscr{L}^n(C),\mathscr{L}^n(\mathbf{U}(a,r)\sim C)\}\leqq c_2\mathscr{H}^n(T)^{n/(n-1)}\leqq(1/3)\alpha(n)r^n$$

so that

$$\mathscr{L}^n(\mathbf{U}(a,r)\sim C)\leqq\mathbf{c}_2\mathscr{H}^n(T)^{n/(n-1)}\leqq(1/3)\alpha(n)r^n.$$

1.10. We will need the following Poincarè type inequality.

Lemma. *Suppose* $a\in\mathbf{R}^n$, $0<r<\infty$ *and* g *is an* $\mathbf{R}$*-valued* $\mathscr{L}^n$ *summable function on* $\mathbf{U}(a,r)$. *Suppose*

(1) V *is the infinum of the set of numbers* C *such that*

$$\int_{\mathbf{U}(a,r)} g\langle u,D\phi\rangle\,d\mathscr{L}^n\leqq C\sup\{|\phi(x)|:x\in\mathbf{U}(a,r)\}\,|u|$$

for each $u\in\mathbf{R}^n$ *and each* $\phi\in\mathscr{C}^1_c(\mathbf{U}(a,r))$.

(We call V *the total variation of the gradient of* g.*)* Then

(2) *there is* $t\in\mathbf{R}$ *such that*

$$\int_{\mathbf{U}(a,r)}|g-t|\,d\mathscr{L}^n\leqq\mathbf{c}_4rV.$$

Proof. In case $g\in\mathscr{C}^1(\mathbf{U}(a,r))$, $V=\int_{\mathbf{U}(a,r)}|\nabla g|\,d\mathscr{L}^n$ and (2) is well known; in the general case we regularize g as in [7, 4.1.18].

1.11. Suppose $a \in \mathbf{R}^n$ and $0 < r < \infty$. For each $\phi \in \mathscr{L}(\mathbf{U}(a, r))$ we let

$$\|\phi\|_0 = (r^{-n-2} \int_{\mathbf{U}(a,r)} \phi^2 \, d\mathscr{L}^n)^{1/2},$$
$$\|\phi\|_1 = (r^{-n} \int_{\mathbf{U}(a,r)} |\nabla \phi|^2 \, d\mathscr{L}^n)^{1/2}.$$

Lemmas of the type which follow are familiar to those who work on the theory of minimal surfaces.

Lemma. *Suppose*

(1) $0 < b \leqq B < \infty$, l *is a symmetric endomorphism of* $\mathbf{R}^{n+1}$ *and*

$$b|u| \leqq |l(u)| \leqq B(u) \quad \text{whenever } u \in \mathbf{R}^{n+1};$$

(2) $0 < \varepsilon < \infty$, $g \in \mathscr{L}(\mathbf{U}(a, r))$ *and*

$$Q(g, \phi) \leqq \varepsilon(\|\phi\|_0 + \|\phi\|_1) \quad \text{whenever } \phi \in \mathscr{C}^1_c(\mathbf{U}(a, r));$$

here we have set

$$Q(f, g) = r^{-n} \int_{\mathbf{U}(a,r)} \nabla f \cdot \langle \nabla g, l \rangle \, d\mathscr{L}^n$$

whenever $f, g \in \mathscr{L}(\mathbf{U}(a, r))$ *and the integral is defined.*

Then there is $u \in \mathbf{R}^n$ *such that*

$$|u| \leqq \mathbf{c}_6 \left(\frac{B}{b}\right)^{(\hat{n}+1)/2} \|g\|_1; \tag{3}$$

$$(\theta r)^{-n} \int_{\mathbf{U}(a,\theta r)} |\nabla g - u|^2 \, d\mathscr{L}^n \leqq \mathbf{c}_8 \left\{ \frac{1}{b} \left(\frac{B}{b}\right)^{1/2} \theta^{-n} \varepsilon \|g\|_1 + \left(\frac{B}{b}\right)^{\hat{n}+2} (\theta \|g\|_1)^2 \right\} \tag{4}$$

whenever $0 < \theta \leqq (1/2)$;

here $\hat{n}$ *is the smallest integer greater than* $n/2$.

Proof. It suffices to consider the case $a = 0$ and $r = 1$. Furthermore, we may assume g has a continuous extension $\bar{g}$ to the closure of $\mathbf{U}(0, 1)$; otherwise we could define $g_\lambda(x) = g(\lambda x)$ for $0 < \lambda < 1$ and $|x| < \lambda^{-1}$, apply our results to g_λ and let $\lambda \uparrow 1$.

Let h be the unique continuous function on the closure of $\mathbf{U}(0, 1)$ such that $h | \mathbf{U}(0, 1) \in \mathscr{C}^2(\mathbf{U}(0, 1))$ and h solves the Dirichlet problem

$$\operatorname{div} \langle \nabla h, l \rangle = 0 \quad \text{on } \mathbf{U}(0, 1),$$
$$h(x) = \bar{g}(x) \quad \text{if } x \in \mathbf{R}^n \text{ and } |x| = 1.$$

Then

$$\begin{aligned} b(\|h\|_1^2 + \|g - h\|_1^2) &\leqq Q(h, h) + Q(g - h, g - h) \\ &= Q(g, g) \\ &\leqq B \|g\|_1^2 \end{aligned}$$

and, by the Poincarè inequality (see [8, 3.2.1]),

$$\|g-h\|_0^2 \leqq \mathbf{c}_5 \|g-h\|_1^2.$$

Thus (2) implies

$$\begin{aligned} b\|g-h\|_1^2 &\leqq Q(g-h, g-h) \\ &= Q(g, g-h) \\ &\leqq \varepsilon(\|g-h\|_0 + \|g-h\|_1) \\ &\leqq (1+\mathbf{c}_5)\varepsilon \left(\frac{B}{b}\right)^{1/2} \|g\|_1; \end{aligned}$$

moreover, from standard elliptic estimates (see [7, 5.2.5] or [8]) we have

$$|\nabla h(0)| \leqq \mathbf{c}_6 \left(\frac{B}{b}\right)^{(\hat{n}+1)/2} \|g\|_1,$$

$$\|D\nabla h(0)\| \leqq \mathbf{c}_7 \left(\frac{B}{b}\right)^{(\hat{n}+2)/2} \|g\|_1 \quad \text{for } x \in \mathbf{U}(0, 1/2).$$

Setting $u = \nabla h(0)$ and taking $0 < \theta < 1/2$ we obtain

$$\int_{\mathbf{U}(0,\theta)} |\nabla h - u|^2 \, d\mathscr{L}^n \leqq \mathbf{c}_7^2 \left(\frac{B}{b}\right)^{\hat{n}+2} \|g\|_1^2 \int_{\mathbf{U}(0,\theta)} |x|^2 \, d\mathscr{L}^n.$$

Since

$$\int_{\mathbf{U}(0,\theta)} |x|^2 \, d\mathscr{L}^n = \frac{\boldsymbol{\alpha}(n)}{n+2} \theta^{n+2}$$

the proof is complete.

2. Some Basic Estimates

In 2.1 we use the main result of [4] to estimate the square integral of $\|Dv\|^2$. In 2.2 we use this estimate in conjunction with the isoperimetric estimate of 1.3(12) to obtain the lower bound 2.2(9) for mass ratios. In 2.3 we show that, for each $u \in \mathbf{S}^n$, $v \cdot u$ satisfies an elliptic constraint on M and use this to obtain in 2.4 the crucial estimate which is the analogue in the present context of [11; Section 2, Lemma 1]. Finally, in 2.5, we estimate the integral of $1 - u \cdot v^2$ in terms of the integral of the absolute value of the position vector dotted with u.

2.1. **Lemma.** *We have*

$$\text{(1)} \qquad \int \phi^2 \|Dv\|^2 \, d\|M\| \leqq \zeta_1 \int |\nabla \phi|^2 + \Delta \phi^2 \, d\|M\| \quad \text{whenever } \phi \in \mathscr{L}_c(M).$$

Proof. Suppose $\psi \in \mathscr{C}_c^1(M)$. For any τ with $0 < \tau < \infty$ we have

$$|\nabla \psi \cdot \langle \nabla \psi, \Sigma(i, v) \rangle| \leqq \Gamma_{0,2} |\nabla \psi|^2;$$

$$\psi^2 \|Dv\|^2 \leqq \psi^2 Dv^2 \cdot \Sigma(i, v);$$

$$2|\psi\langle\nabla\psi, \Sigma_1(i,v)\rangle \cdot v| \leqq |\nabla\psi|^2 + \Gamma_{1,1}^2\psi^2;$$
$$\psi^2|\Sigma_3(i,v)| \leqq \Gamma_{2,0}\psi^2;$$
$$\psi^2|Dv \cdot \Sigma_1(i,v)| \leqq \tau\psi^2\|Dv\|^2 + \frac{n}{4\tau}\psi^2\Gamma_{1,1}^2;$$
$$\psi^2|H\Sigma_2(i,v)| \leqq \tau\psi^2\|Dv\|^2 + \frac{n}{4\tau}\psi^2\Gamma_{1,0}^2;$$
$$\psi^2|LH| \leqq \tau\psi^2\|Dv\|^2 + \frac{n}{4\tau}v_1^2\psi^2.$$

We infer from 1.4 (5), (6) and 1.5 (2) that

$$(1-3\tau)\int\psi^2\|Dv\|^2 d\|M\| \leqq (\Gamma_{0,2}+1)\int|\nabla\psi|^2 d\|M\| + \left\{\Gamma_{1,1}^2 + \Gamma_{2,0} + \frac{n}{4\tau}(\Gamma_{1,1}^2 + \Gamma_{1,0}^2 + v_1^2) + v_2\right\}\int\psi^2 d\|M\|$$

whenever $0<\tau<\infty$. We set $\tau = 1/6$ to complete the proof.

2.2. **A Lower Bound for Mass Ratios.** We are referring here to the ratio of the area of M in a ball centered at a point of M to the n-th power of the radius of the ball. The estimate (9) below should be compared with [2, 5.1] and [7, 5.1.6]. See also [3].

(1) **Lemma.** *Suppose*

$$I: (0,\infty) \to [0,\infty), \qquad 0<C<\infty$$

and

$$I(np/(n-1)) \leqq (Cp)^{1/p} I(p-1)^{(p-1)/p} \qquad \textit{whenever } 1<p<\infty.$$

Then

$$\limsup_{p\to\infty} I(p) \leqq \mathbf{c}_9 C^{n/(q+n)} I(q)^{1/(q+n)} \qquad \textit{whenever } 0<q<\infty.$$

Proof. Let $\beta = n/(n-1)$. Suppose $0<q<\infty$. For each $\sigma = 0, 1, 2, \ldots$ we define q_σ by requiring that

$$q_0 = q \quad \text{and} \quad q_\sigma = \beta(q_{\sigma-1}+1) \quad \text{for } \sigma = 1, 2, \ldots.$$

We also set

$$r_\sigma = \frac{1}{q_\sigma+1} \quad \text{and} \quad s_\sigma = \frac{q_\sigma}{q_\sigma+1} \quad \text{for } \sigma = 0, 1, 2, \ldots.$$

Thus

(2) $$I(q_{\sigma+1}) \leqq C^{r_\sigma} r_\sigma^{-r_\sigma} I(q_\sigma)^{s_\sigma} \quad \text{for } \sigma = 0, 1, 2, \ldots.$$

Let us now set

$$t_{\rho,\sigma} = \prod_{\xi=\rho}^{\sigma} s_\xi \quad \text{for } \sigma = 0, 1, 2, \ldots \text{ and } \rho = 0, 1, \ldots, \sigma.$$

By a simple induction we infer from (2) that

$$I(q_{+1}) \leqq \left(\prod_{\rho=0}^{\sigma} r_\rho^{-\frac{t_{\rho,\sigma}}{q_\rho}} C^{\sum_{\rho=0}^{\sigma} \frac{t_{\rho,\sigma}}{q_\rho}} \right) I(q_0)^{t_{0,\sigma}} \quad \text{for } \sigma = 0, 1, 2, \ldots. \tag{3}$$

By a simple induction we obtain

$$\frac{q_\sigma + 1}{\beta^\sigma} = (q+n) + (1-n)\beta^{-\sigma} \quad \text{for } \sigma = 0, 1, 2, \ldots. \tag{4}$$

Furthermore,

$$\frac{t_{\rho,\sigma}}{q_\rho} = \frac{1}{q_\sigma + 1} \prod_{\xi=\rho+1}^{\sigma} \frac{q_\xi}{q_{\xi-1}+1} = \frac{\beta^\sigma}{q_\sigma + 1} \beta^{-\rho} \tag{5}$$

$$\text{for } \sigma = 0, 1, 2, \ldots \text{ and } \rho = 0, 1, \ldots, \sigma,$$

which in turn implies

$$\sum_{\rho=0}^{\sigma} \frac{t_{\rho,\sigma}}{q_\rho} = \frac{\beta^\sigma}{q_\sigma + 1} \frac{1 - \beta^{-\sigma-1}}{1 - \beta^{-1}} \quad \text{for } \sigma = 0, 1, 2, \ldots. \tag{6}$$

Now (4) and (5) imply

$$\lim_{\sigma \to \infty} t_{0,\sigma} = \frac{1}{q+n}$$

and (4) and (6) imply

$$\lim_{\sigma \to \infty} \sum_{\rho=0}^{\sigma} \frac{t_{\rho,\sigma}}{q_\rho} = \frac{n}{q+n}.$$

Finally, (4) and (5) imply

$$\begin{aligned} -\sum_{\rho=0}^{\sigma} \frac{t_{\rho,\sigma}}{q_\rho} \log r_\rho &= \sum_{\rho=0}^{\sigma} \frac{\beta^\sigma}{q_\sigma + 1} \beta^{-\rho} \log (q_\rho + 1) \\ &\leqq \frac{1}{q+1} \sum_{\rho=0}^{\sigma} \beta^{-\rho} \log \beta^\rho (q+n) \\ &= \frac{1}{q+1} \sum_{\rho=0}^{\sigma} \beta^{-\rho} \rho \log \beta + \frac{\log (q+n)}{q+1} \sum_{\rho=0}^{\sigma} \beta^{-\rho} \\ &= \frac{1}{q+1} n^2 \log \beta + \frac{\log (q+n)}{q+1} n \\ &\quad \text{for } \sigma = 0, 1, 2, \ldots \end{aligned}$$

so that our proof is complete since $(q+1)^{-1} \log (q+n)$ is decreasing on $0 < q < \infty$; in fact, we may take

$$\mathbf{c}_9 = \exp [(n/(n-1))^{n^2} n^n].$$

Lemma. *Suppose* $\phi \in \mathscr{L}_c(M)$ *and*

$$s = \sup\{|\phi(z)| : z \in M\}, \qquad l = \sup\{|\nabla\phi(z)| : z \in M\}.$$

Then

$$s^{n+1} \leqq \mathbf{c}_{10}[\zeta_1(l^2 + \Delta s^2)]^{n/2} \int |\phi| \, d\|M\|. \tag{7}$$

Proof. Suppose $1 < p < \infty$. From 1.3(12) we obtain

$$\left(\int |\phi|^{pn/(n-1)} d\|M\|\right)^{(n-1)/n} \leqq \mathbf{c}_1 \int |\nabla|\phi|^p| + |\phi|^p |H| \, d\|M\|.$$

It is evident that

$$\int |\nabla|\phi|^p| \, d\|M\| \leqq pl \int |\phi|^{p-1} d\|M\|;$$

moreover, the Schwarz Inequality and 2.1 imply

$$\begin{aligned}
&\int |\phi|^p |H| \, d\|M\| \\
&\quad \leqq n\left(\int |\phi|^{p-1} d\|M\|\right)^{1/2} \left(\int |\phi|^{p+1} \|Dv\|^2 d\|M\|\right)^{1/2} \\
&\quad \leqq n\left(\int |\phi|^{p-1} d\|M\|\right)^{1/2} \left(\zeta_1 \int |\nabla|\phi|^{(p+1)/2}|^2 + \Delta |\phi|^{p+1} d\|M\|\right)^{1/2} \\
&\quad \leqq n\left(\int |\phi|^{p-1} d\|M\|\right)^{1/2} \left(\zeta_1 \left(\left(\frac{p+1}{2}\right)^2 l^2 + \Delta s^2\right) \int |\phi|^{p-1} d\|M\|\right)^{1/2}.
\end{aligned}$$

Thus

$$\left(\int |\phi|^{pn/(n-1)} d\|M\|\right)^{(n-1)/n} \leqq pC \int |\phi|^{p-1} d\|M\|$$

where $C = n(\zeta_1(l^2 + \Delta s^2))^{1/2}$ so that our inequality follows from (1) with q there equal 1.

Theorem. *Suppose* $c \in M$, $0 < r < \infty$, $\operatorname{dist}(c, \mathbf{R}^{n+1} \sim Z) > r$ *and*

$$\Delta r^2 \leqq 1. \tag{8}$$

Then

$$r^n \leqq \zeta_2 \|M\|(\{z \in \mathbf{R}^{n+1} : |z - c| < r\}). \tag{9}$$

Proof. We let

$$\phi(z) = \begin{cases} r(1 - |z-c|/r) & \text{if } z \in M \text{ and } |z-c| \leqq r, \\ 0 & \text{if } z \in M \text{ and } |z-c| > r, \end{cases}$$

and apply (7).

2.3. **A Central Identity.** *Suppose* $u \in \mathbf{S}^n$. Let

$$\alpha, \beta \in \mathscr{C}^1(M) \quad \textit{and} \quad T \in \mathscr{C}^1(M, \mathbf{R}^{n+1})$$

be such that

$$\alpha(z) = v(z) \cdot u, \qquad \beta(z) = (1 - \alpha(z)^2)^{1/2}, \qquad T(z) = u - \alpha(z)\, v(z)$$

whenever $z \in M$.

Keeping in mind 1.3(6) we obtain

(1) $$\alpha^2+\beta^2=1;$$

(2) $$|T|=\beta$$

(3) $$\nabla\alpha=\langle u, Dv\rangle;$$

(4) $$|\nabla\alpha|\leqq\beta\|Dv\|;$$

(5) $$\beta\in\mathscr{L}(M);$$

(6) $$|\nabla\beta|\leqq|\alpha|\,\|Dv\| \qquad \|M\| \text{ almost everywhere.}$$

Then identity which follows is well-known in case Φ is the length integrand, that is when $\Phi(z,\alpha)=|\alpha|$ for $(z,\alpha)\in Z\times\mathbf{R}^{n+1}$.

Theorem.

$$\begin{aligned}\operatorname{div}(\langle\nabla\alpha, \Sigma(i,v)\rangle+LT)=&-\alpha Dv^2\cdot\Sigma(i,v)+L\operatorname{div}T\\&+\Sigma_4(i,v)\cdot\nabla\alpha-\langle u,\Sigma_5(i,v)\rangle\cdot Dv-\Sigma_6(i,v)\cdot u\\&-\operatorname{trace}\langle\langle u, Dv\rangle, \Sigma_5(i,v)\rangle.\end{aligned}$$

Proof. We will prove this under the assumption that M is of class 3; the general case will then follow by approximation.

Using 1.3(7), we calculate that

$$\begin{aligned}\langle v, D\nabla\alpha(z)\rangle=&\langle u,\langle v, D(Dv)(z)\rangle\rangle=\langle v,\langle u, D(Dv)(z)\rangle\rangle\\&+v\cdot\nu(z)\langle u, Dv(z)^2\rangle-u\cdot\nu(z)\langle v, Dv(z)^2\rangle\\&\text{for } v\in\mathbf{R}^{n+1} \quad\text{and}\quad z\in M.\end{aligned}$$

Now suppose $c\in M$, $(u_1,\ldots,u_{n+1})$ is an orthonormal basic sequence for $\mathbf{R}^{n+1}$ and $u_{n+1}=\nu(c)$. Since

$$\langle\nabla\alpha(z),\Sigma(z,\nu(z))\rangle\cdot u_i=\langle(0,\nabla\alpha(z))\odot(0,u_i), D^2\Phi(z,\nu(z))\rangle$$

for $z\in M$ and $i=1,\ldots,n+1$ we see that

$$\begin{aligned}&\operatorname{div}\langle\nabla\alpha,\Sigma(i,\nu)\rangle(c)\\&=\sum_{i=1}^{n}\langle(0,\langle u_i, D\nabla\alpha(c)\rangle\odot(0,u_i), D^2\Phi(c,\nu(c))\rangle\\&\quad+\sum_{i=1}^{n+1}\langle(0,\nabla\alpha(c))\odot(0,u_i)\odot(u_i,0), D^3\Phi(c,\nu(c))\rangle\\&\quad+\sum_{i=1}^{n}\langle(0,\nabla\alpha(c))\odot(0,u_i)\odot(0,\langle u_i, Dv(c)\rangle), D^3\Phi(c,\nu(c))\rangle\\&=\sum_{i=1}^{n}\langle(0,\langle u_i,\langle u, D(Dv)(c)\rangle\rangle)\odot(0,u_i), D^2\Phi(c,\nu(c))\rangle\\&\quad-\sum_{i=1}^{n}\langle(0,\alpha(c)\langle u_i, Dv(c)^2\rangle)\odot(0,u_i), D^2\Phi(c,\nu(c))\rangle\end{aligned}$$

$$+\Sigma_4(c, v(c))\cdot\nabla\alpha(c)$$
$$+\sum_{i=1}^{n}\langle(0,\langle u, Dv(c)\rangle)\odot(0,u_i)\odot(0,\langle u_i, Dv(c)\rangle), D^3\Phi(c, v(c))\rangle.$$

On the other hand, 1.4(4) says that

$$-L(z)=\sum_{i=1}^{n+1}\langle(0,u_1)\odot(0,\langle u_i, Dv(z)\rangle), D^2\Phi(z, v(z))\rangle$$
$$+\sum_{i=1}^{n+1}\langle(0,u_i)\odot(u_i,0), D^2\Phi(z, v(z))\rangle$$

for all $z\in M$. Differentiating at c in the direction u gives

$$-T(c)\cdot\nabla L(c)=\sum_{i=1}^{n+1}\langle(0,u_i)\odot(0,\langle u_i,\langle u, D(Dv)(c)\rangle\rangle), D^2\Phi(c, v(c))\rangle$$
$$+\sum_{i=1}^{n}\langle(0,u_i)\odot(0,\langle u_i, Dv(z)\rangle)\odot(u,0), D^3\Phi(c, v(c))\rangle$$
$$+\sum_{i=1}^{n}\langle(0,u_i)\odot(0,\langle u_i, Dv(c)\rangle)\odot(0,\langle u, Dv(c)\rangle), D^3\Phi(c, v(c))\rangle$$
$$+\sum_{i=1}^{n+1}\langle(0,u_i)\odot(u_i,0)\odot(u,\langle u, Dv(c)\rangle), D^3\Phi(c, v(c))\rangle$$
$$=\sum_{i=1}^{n}\langle(0,u_i)\odot(0,\langle u_i,\langle u, D(Dv)(c)\rangle\rangle), D^2\Phi(c, v(c))\rangle$$
$$+\langle u, \Sigma_5(c, v(c))\rangle\cdot Dv(c)$$
$$+\sum_{i=1}^{n}\langle(0,u_i)\odot(0,\langle u_i, Dv(c)\rangle)\odot(0,\langle u, Dv(c)\rangle), D^3\Phi(c, v(c))\rangle$$
$$+\Sigma_6(c, v(c))\cdot u+\operatorname{trace}\langle\langle u, Dv(c)\rangle, \Sigma_5(c, v(c))\rangle$$

where we have used 1.2 with k there equal 2.

2.4. **A Central Estimate.** Here we use more carefully the identity 1.5(2) which gave us 2.1 as well as 2.3 to obtain a sharper estimate. We will make use of 1.6. We let u, α, β, T be as in 2.3.

Theorem. *We have*

$$\int\phi^2\|Dv\|^2 d\|M\|\leqq\zeta_3\int|\nabla\phi|^2\beta^2+\Delta\phi^2 d\|M\|$$

whenever $\phi\in\mathscr{L}_c(M)$.

Proof. Suppose $\phi\in\mathscr{C}_c^1(M)$. Let

$$a_1=\int\phi^2\|Dv\|^2 d\|M\|,$$
$$a_2=\int\phi^2 Dv^2\cdot\Sigma(i, v)\, d\|M\|,$$
$$a_3=\int\phi^2\alpha^2 Dv^2\cdot\Sigma(i, v)\, d\|M\|,$$

$$a_4 = \int \phi^2 \, \| Dv \circ \Sigma(i, v) \circ Dv \| \, d \| M \|,$$
$$a_5 = \int \nabla \phi \, \beta \cdot \langle \nabla \phi \, \beta, \Sigma(i, v) \rangle \, d \| M \|,$$
$$a_6 = \int (1 - \alpha^2) \, |\nabla \phi|^2 \, d \| M \|,$$
$$a_7 = \int \phi^2 \, d \| M \|.$$

Let

$$\rho_1(\tau) = \left\{ \left(1 + \frac{1}{4\tau} + \frac{n^2}{4\tau} \right) v_1^2 + \frac{3(n+1)^2}{4\tau} \Gamma_{1,2}^2 + (n+1) \Gamma_{2,1} \right\},$$
$$\rho_2(\tau) = \frac{9(n+1)}{\tau} (v_1^2 + (n+1)^2 \Gamma_{1,1}^2),$$
$$\rho_3(\tau) = \left\{ \frac{n^2}{4\tau} \Gamma_{1,0}^2 + \left(1 + \frac{1}{\tau} + \frac{n^2}{4\tau} \right) \Gamma_{1,1}^2 + \Gamma_{2,0} + \frac{n^2}{4\tau} v_1^2 + v_2 \right\}$$

whenever $0 < \tau < \infty$.

Part One. Suppose ψ is a real valued function of class 2 whose domain is an open subset of $\mathbf{R}$ *containing* $[-1, 1]$. *We have*

$$\begin{aligned}
&\nabla \phi (\psi \circ \alpha) \cdot \langle \nabla \phi (\psi \circ \alpha), \Sigma(i, v) \rangle \\
&\quad = (\psi \circ \alpha)^2 \, \nabla \phi \cdot \langle \nabla \phi, \Sigma(i, v) \rangle \\
&\qquad + \operatorname{div} \phi^2 (\psi \psi') \circ \alpha (\langle \nabla \alpha, \Sigma(i, v) \rangle + LT) \\
&\qquad - \phi^2 (\psi \psi'') \circ \alpha \nabla \alpha \cdot \langle \nabla \alpha, \Sigma(i, v) \rangle \\
&\qquad - 2 \phi (\psi \psi') \circ \alpha L \nabla \phi \cdot T \\
&\qquad - \phi^2 (\psi \psi')' \circ \alpha L \nabla \alpha \cdot T \\
&\qquad - \phi^2 (\psi \psi') \circ \alpha \{ - \alpha Dv^2 \cdot \Sigma(i, v) - L \operatorname{div} T + \Sigma_4(i, v) \cdot \nabla \alpha \\
&\qquad - \langle u, \Sigma_5(i, v) \rangle \cdot Dv - \Sigma_6(i, v) \cdot u - \operatorname{trace} \langle \langle u, Dv \rangle, \Sigma_5(i, v) \rangle \}
\end{aligned}$$

in the weak sense.

Indeed,

$$\begin{aligned}
&\nabla \phi (\psi \circ \alpha) \cdot \langle \nabla \phi (\psi \circ \alpha), \Sigma(i, v) \rangle \\
&\quad = (\psi \circ \alpha)^2 \, \nabla \phi \cdot \langle \nabla \phi, \Sigma(i, v) \rangle + 2 \phi (\psi \circ \alpha) \nabla \phi \cdot \langle \nabla (\psi \circ \alpha), \Sigma(i, v) \rangle \\
&\qquad + \phi^2 \nabla (\psi \circ \alpha) \cdot \langle \nabla (\psi \circ \alpha), \Sigma(i, v) \rangle
\end{aligned}$$

and

$$\begin{aligned}
&2 \phi (\psi \circ \alpha) \nabla \phi \cdot \langle \nabla (\psi \circ \alpha), \Sigma(i, v) \rangle + \phi^2 \nabla (\psi \circ \alpha) \cdot \langle \nabla (\psi \circ \alpha), \Sigma(i, v) \rangle \\
&\quad = \operatorname{div} \phi^2 (\psi \psi') \circ \alpha (\langle \nabla \alpha, \Sigma(i, v) \rangle + LT) \\
&\qquad - \phi^2 (\psi \psi'') \circ \alpha \nabla \alpha \cdot \langle \nabla \alpha, \Sigma(i, v) \rangle \\
&\qquad - 2 \phi (\psi \psi') \circ \alpha L \nabla \phi \cdot T \\
&\qquad - \phi^2 (\psi \psi')' \circ \alpha L \nabla \alpha \cdot T \\
&\qquad - \phi^2 (\psi \psi') \circ \alpha \operatorname{div} (\langle \nabla \alpha, \Sigma(i, v) \rangle + LT)
\end{aligned}$$

in the weak sense, so the asserted identity follows from 2.3.

Part Two. We have

$$a_5 \leqq 5\tau a_1 + a_4 - a_3 + (\Gamma_{0,2}+1)\, a_6 + \rho_1(\tau)\, a_7$$

whenever $0<\tau<\infty$.

We prove this as follows. For each ε with $0<\varepsilon<\infty$ we let

$$\zeta_\varepsilon(t) = (1+\varepsilon-t^2)^{1/2} \quad \text{whenever} \quad -(1+\varepsilon)^{1/2} < t < (1+\varepsilon)^{1/2}.$$

Note that

$$\zeta_\varepsilon'(t) = -\frac{t}{\zeta_\varepsilon(t)} \quad \text{and} \quad \zeta_\varepsilon''(t) = -\frac{1+\varepsilon}{\zeta_\varepsilon(t)^3}$$

so that

$$\zeta_\varepsilon\, \xi_\varepsilon''(t) = -\frac{1+\varepsilon}{\zeta_\varepsilon^2(t)}, \qquad \zeta_\varepsilon \zeta_\xi'(t) = -t, \qquad (\zeta_\varepsilon \zeta_\varepsilon')'(t) = -1$$

whenever $-(1+\varepsilon)^{1/2} < t < (1+\varepsilon)^{1/2}$. So if $0<\varepsilon<\infty$, $\psi = \zeta_\varepsilon$ and $0<\tau<\infty$ we have

$$(\psi\circ\alpha)^2 \nabla\phi \cdot \langle \nabla\phi, \Sigma_1(i,v)\rangle \leqq (1+\varepsilon-\alpha^2)\, \Gamma_{0,2} |\nabla\phi|^2,$$

$$-\phi^2 (\psi\psi'')\circ\alpha \nabla\alpha \cdot \langle \nabla\alpha, \Sigma(i,v)\rangle \leqq (1+\varepsilon)\, \phi^2 \|Dv\circ\Sigma(i,v)\circ Dv\|,$$

$$-2\phi(\psi\psi')\circ\alpha L\nabla\phi \cdot T \leqq (1-\alpha^2)|\nabla\phi|^2 + v_1^2\phi^2,$$

$$-\phi^2(\psi\psi')'\circ\alpha L\nabla\alpha \cdot T \leqq \tau\phi^2 \|Dv\|^2 + \frac{1}{4\tau}\, v_1^2\, \phi^2,$$

$$\phi^2(\psi\psi')'\circ\alpha\alpha Dv^2 \cdot \Sigma(i,v) = -\phi^2\alpha^2 Dv^2 \cdot \Sigma(i,v),$$

$$\phi^2(\psi\psi')\circ\alpha L \operatorname{div} T \leqq \tau\phi^2\|Dv\|^2 + \frac{n^2}{4\tau}\, v_1^2\, \phi^2,$$

$$-\phi^2(\psi\psi')\circ\alpha\Sigma_4(i,v)\cdot\nabla\alpha \leqq \tau\phi^2\|Dv\|^2 + \frac{(n+1)^2}{4\tau}\, \Gamma_{1,2}^2\, \phi^2,$$

$$\phi^2(\psi\psi')\circ\alpha\langle u, \Sigma_5(i,v)\rangle \cdot Dv \leqq \tau\phi^2\|Dv\|^2 + \frac{(n+1)^2}{4\tau}\, \Gamma_{1,2}^2\, \phi^2,$$

$$\phi^2(\psi\psi')\circ\alpha\Sigma_6(i,v)\cdot u \leqq (n+1)\, \Gamma_{2,1}\, \phi^2,$$

$$\phi^2(\psi\psi')\circ\alpha \operatorname{trace}\langle\langle u, Dv\rangle, \Sigma_5(i,v)\rangle \leqq \tau\phi^2\|Dv\|^2 + \frac{(n+1)^2}{4\tau}\, \Gamma_{1,2}^2\, \phi^2.$$

We now integrate the identity in Part One with respect to $\|M\|$, use the inequalities listed immediately above and let $\varepsilon\downarrow 0$ to derive the asserted inequality.

Part Three. We have

$$a_4 \leqq (1-\gamma)\, a_2 + \tau a_1 + \rho_2(\tau)\, a_7 \qquad \textit{whenever} \quad 0<\tau<\infty$$

where

$$\gamma = \min\left\{\frac{1}{n+1}, \frac{n}{n+\Gamma_{0,2}}\right\}.$$

To prove this observe that

$$\max\left\{1-\frac{\|\Sigma(i,v)\|}{\operatorname{trace}\Sigma(i,v)}, \frac{\|\Sigma(i,v)\|}{\operatorname{trace}\Sigma(i,v)}\right\}\leqq\max\left\{\frac{n}{n+1}, \frac{\Gamma_{0,2}}{n+\Gamma_{0,2}}\right\}=1-\gamma$$

and that

$$\begin{aligned}&6(n+1)^{1/2}\|Dv\|\,|Dv\cdot\Sigma(i,v)|\\&\quad=6(n+1)^{1/2}\|Dv\|\,|-L-\operatorname{trace}\Sigma_1(i,v)|\\&\quad\leqq\tau\|Dv\|^2+\frac{9(n+1)}{\tau}\,v_1^2+\frac{9}{\tau}(n+1)^3\,\Gamma_{1,1}^2,\end{aligned}$$

and then use 1.6.

Part Four. We have

$$a_2-a_3\leqq a_5+a_6+4\tau a_1+\rho_3(\tau)\,a_7$$

whenever $0<\tau<\infty$.

To prove this, we observe that

$$\begin{aligned}&-2\phi\beta\langle\nabla\phi\beta,\Sigma_1(i,v)\rangle\cdot v\\&\quad\leqq2\phi\beta^2|\nabla\phi|\,\Gamma_{1,1}+2\phi^2\beta|\nabla\beta|\,\Gamma_{1,1}\\&\quad\leqq\beta^2|\nabla\phi|^2+\Gamma_{1,1}^2\,\phi^2+\tau\phi^2\|Dv\|^2+\frac{1}{\tau}\,\Gamma_{1,1}^2\,\phi^2,\\&\phi^2\beta^2\Sigma_3(i,v)\leqq\Gamma_{2,0}\phi^2,\\&\quad-\phi^2Dv\cdot\Sigma_1(i,v)\leqq\tau\phi^2\|Dv\|^2+\frac{n^2}{4\tau}\,\Gamma_{1,1}^2\,\phi^2,\\&\quad-\phi^2\beta^2H\Sigma_2(i,v)\leqq\tau\phi^2\|Dv\|^2+\frac{n^2}{4\tau}\,\Gamma_{1,0}^2\phi^2,\\&\quad\phi^2\beta^2LH\leqq\tau\phi^2\|Dv\|^2+\frac{n^2}{4\tau}\,v_1^2\,\phi^2,\\&\quad v_2\,\phi^2\beta\leqq v_2\,\phi^2\end{aligned}$$

$\|M\|$ essentially whenever $0<\tau<\infty$. We then invoke 1.5(1), (2) and 1.4(5), (6) with ϕ there replaced by $\phi\beta$.

Part Five. Using, in order, Parts Four, Two and Three we obtain

$$\begin{aligned}a_2-a_3&\leqq a_5+a_6+4\tau a_1+\rho_3(\tau)\,a_7\\&\leqq(5\tau a_1+a_4+a_3+(\Gamma_{0,2}+1)\,a_6+\rho_1(\tau)\,a_7)+a_6+4\tau a_1+\rho_3(\tau)\,a_7\\&\leqq5\tau a_1+((1-\gamma)\,a_2+\tau a_1+\rho_2(\tau)\,a_7)\\&\quad-a_3+(\Gamma_{0,2}+1)\,a_6+\rho_1(\tau)\,a_7+a_6+4\tau a_1+\rho_3(\tau)\,a_7;\end{aligned}$$

since $a_1\leqq a_2$ we infer that

$$(\gamma - 10\tau)a_1 \leqq (\Gamma_{0,2}+2)a_6 + (\rho_1(\tau)+\rho_2(\tau)+\rho_3(\tau))a_7.$$

So we set $\tau = \gamma/20$ to complete the proof.

2.5. There is no ζ_4.

3. The Main Lemma

3.1. Throughout This Chapter we Assume that (1)–(6) Below Hold.

(1) $0 < R < \infty$, $X = \mathbf{U}(0, 3R)$, *Y is a bounded open interval in* $\mathbf{R}$, $Z = X \times Y$;

(2) *M is closed relative to* $X \times \mathbf{R}$;

(3) β *is as in* 2.3 *with u there equal* e_{n+1};

(4) $B = \{\mathbf{p}(z) : z \in M \text{ and } \beta(z) = \sqrt{2}/2\}$;

(5) $N = \{z \in M \sim \mathbf{p}^{-1}(B) : \beta(z) < \sqrt{2}/2\}$;

(6) $\pi = \mathbf{p} | M$.

Using (1) and (2) we infer that B is closed relative to X, N is open relative to M, $\mathbf{p}|N$ is an open mapping and $\pi : M \to X$ is proper. Moreover, if C is an open connected subset of $X \sim B$ then $\mathbf{p}$ carries each connected component of $N \cap \mathbf{p}^{-1}(C)$ univalently onto C; in particular, the number of these components is finite. See the Proof of the Lemma in 3.3 if this is not clear.

3.2. A Basic Lemma About B.

Lemma. Suppose $a \in X$, $0 < r < \infty$, $|a| + r < 3R$,

(1) $$\varepsilon = r^{2-n} \int_{\mathbf{C}(a,r)} \|Dv\|^2 d\|M\| + r^{-n}\|M\|(\{z \in M \cap \mathbf{C}(a,r) : \beta(z) \geqq \sqrt{2}/4\});$$

(2) $$\mathbf{c}_{11}\varepsilon \leqq 1.$$

Then

(3) *there is an open connected subset C of* $\mathbf{U}(a,r) \sim B$ *such that*

(i) $\mathscr{L}^n(C) > (1/2)\alpha(n) r^n$;

(ii) $\mathscr{H}^{n-1}(E) \leqq 2\varepsilon r^{n-1}$ *where E is the boundary of C relative to* $\mathbf{U}(a,r)$;

(iii) $\mathscr{L}^n(\mathbf{U}(a,r) \sim C) \leqq \mathbf{c}_2 (2\varepsilon)^{n/(n-1)} r^n$;

(iv) $\|M\|[\mathbf{C}(a,r) \sim (\mathbf{p}^{-1}(C) \cap N)] \leqq \mathbf{c}_{12}[1 + \varepsilon^{1/(n-1)} r^{-n}\|M\|(\mathbf{C}(a,r))]\varepsilon r^n$;

(v) $\kappa \leqq 2\alpha(n)^{-1} r^{-n} \|M\|(\mathbf{C}(a,r))$

where κ *is the number of connected components of* $\mathbf{p}^{-1}(C) \cap N$;

(4) $$\kappa > 0$$

if, in addition to (2), *we have*

(i) $\Delta r^2 \leqq 1$, $a \in \operatorname{rng} \pi$ *and* $\zeta_2(2\sqrt{2}\mathbf{c}_4 + 1)\varepsilon < 1$.

Proof. For each $\tau\in(0,1/4)$ we let

$$S_\tau=\{z\in M\cap\mathbf{C}(a,r):|\beta(z)-\sqrt{2}/2|<\tau\},$$
$$T_\tau=\{z\in M\cap\mathbf{C}(a,r):|\beta(z)-\sqrt{2}/2|=\tau\}$$

and we let Γ_τ be the boundary of $\mathbf{p}(S_\tau)$ relative to $\mathbf{U}(a,r)$. From the coarea formula of [7, 3.2.22] we obtain

$$\begin{aligned}
&r^{1-n}\int_0^{1/4}\mathscr{H}^{n-1}(T_\tau)\,d\tau\\
&\quad=r^{n-1}\int_{\{z\in M\cap\mathbf{C}(a,r):\,|\beta(z)-\sqrt{2}/2|<1/4\}}|\nabla\beta|\,d\|M\|\,z\\
&\quad\leqq r^{1-n}(1/2)\,[r^{-1}\|M\|(\{z\in M\cap\mathbf{C}(a,r):\beta(z)\geqq\sqrt{2}/4\})+r\int_{\mathbf{C}(a,r)}\|Dv\|^2\,d\|M\|]\\
&\quad=(1/2)\,\varepsilon.
\end{aligned}$$

Hence we may choose $t\in(0,1/4)$ such that, with $T=T_\tau$ and $\Gamma=\Gamma_t$ we have

$$\mathscr{H}^{n-1}(\Gamma)\leqq\mathscr{H}^{n-1}(T)\leqq 2\varepsilon r^{n-1};\tag{5}$$

in fact, because $\sqrt{2}/2+1/4<1$, $\mathbf{p}(S_t)$ is open from which it is easy to see that $\Gamma_t\subset\mathbf{p}(T_t)$. Putting $S=S_t$ we have

$$\mathscr{L}^n[\mathbf{p}(S)]\leqq r^n\|M\|(\{z\in M\cap\mathbf{C}(a,r):\beta(z)\geqq\sqrt{2}/4\}).\tag{6}$$

Suppose

$$2\varepsilon\leqq\mathbf{c}_3,\qquad \mathbf{c}_2(2\varepsilon)^{n/(n-1)}<(1/2)\,\boldsymbol{\alpha}(n),\qquad \varepsilon\leqq(1/2)\,\boldsymbol{\alpha}(n).\tag{7}$$

Then by (5) and 1.9 there is a connected component C of $\mathbf{U}(a,r)\sim\Gamma$ such that

$$\mathscr{L}^n(\mathbf{U}(a,r)\sim C)\leqq\mathbf{c}_2(2\varepsilon)^{n/(n-1)}r^n<(1/2)\,\boldsymbol{\alpha}(n).\tag{8}$$

Were it the case that $C\cap\mathbf{p}(S)\neq\emptyset$ we would have $C\subset\mathbf{p}(S)$ but that is impossible since $\mathscr{L}^n(\mathbf{p}(S))\leqq(1/2)\,\boldsymbol{\alpha}(n)\,r^n$ by (6) and (8) holds. Thus $C\cap B=\emptyset$.

Let $U=\{z\in M\cap\mathbf{C}(a,r):\beta(z)<\sqrt{2}/2-t\}$ and for each $x\in\mathbf{U}(a,r)$ let

$$\#(x)=\operatorname{card}\{z\in U:\mathbf{p}(z)=x\}.$$

Since the total variation of the gradient of $\#$ on $\mathbf{U}(a,r)$ does not exceed $\mathscr{H}^{n-1}(T)$, we infer from 1.10 and (5) the existence of a real number p such that

$$\int_{\mathbf{U}(a,r)}|\#-p|\,d\mathscr{L}^n\leqq\mathbf{c}_4\,r(2\varepsilon r^{n-1}).$$

Let k be the number of connected components of $N\cap\mathbf{p}^{-1}(C)$. We have

$$\begin{aligned}
\int_C|\#-k|\,d\mathscr{L}^n&\leqq\|M\|(p^{-1}(C)\sim U)\\
&\leqq\|M\|(\{z\in\mathbf{p}^{-1}(C):\beta(z)\geqq\sqrt{2}/2-t\})\\
&\leqq\varepsilon r^n
\end{aligned}$$

so that

$$\begin{aligned}(1/2)\,\boldsymbol{\alpha}(n)\,r^n|p-k| &\leqq \int_C |p-k|\,d\mathscr{L}^n \\ &\leqq \int_C |\#-p|+|\#-k|\,d\mathscr{L}^n \\ &\leqq \mathbf{c}_4\, r(2\varepsilon r^{n-1})+\varepsilon r^n \\ &< (1/2)\,\boldsymbol{\alpha}(n)\,r^n\end{aligned}$$

which implies $p=k$ provided we assume

$$(2\mathbf{c}_4+1)\,\varepsilon < (1/2)\,\boldsymbol{\alpha}(n). \tag{9}$$

Thus

$$\begin{aligned}(10)\qquad &\|M\|\,[\mathbf{C}(a,r)\sim(p^{-1}(C)\cap N)] \\ &\quad\leqq \|M\|\,[(\mathbf{C}(a,r)\cap U)\sim \mathbf{p}^{-1}(C)]+\|M\|\,[C(a,r)\sim U] \\ &\quad\leqq \sqrt{2}\int_{\mathbf{U}(a,r)\sim C} \#\,d\mathscr{L}^n+\varepsilon r^n \\ &\quad\leqq \sqrt{2}\int_{\mathbf{U}(a,r)} |\#-k|\,d\mathscr{L}^n+\sqrt{2}k\mathscr{L}^n(\mathbf{U}(a,r)\sim C)+\varepsilon r^n \\ &\quad\leqq 2\sqrt{2}\mathbf{c}_4\,\varepsilon r^n+\sqrt{2}k\mathbf{c}_2(2\varepsilon)^{n/(n-1)}\,r^n+\varepsilon r^n.\end{aligned}$$

If κ is the number of connected components of $N\cap p^{-1}(C)$ we have

$$(1/2)\,k\boldsymbol{\alpha}(n)\,r^n \leqq (1/2)\,\kappa\boldsymbol{\alpha}(n)\,r^n \leqq \|M\|\,[N\cap p^{-1}(C)] \leqq \|M\|\,\mathbf{C}(a,r)$$

so (3)(iv), (v) hold.

Finally, if (4)(i) holds we have from 2.2(9) that

$$r^n \leqq \zeta_2\,\|M\|(\mathbf{C}(a,r)).$$

Were it the case that $N\cap p^{-1}(C)=\emptyset$ we would have from (10) that

$$\|M\|(\mathbf{C}(a,r)) \leqq (2\sqrt{2}\mathbf{c}_4+1)\,\varepsilon r^n$$

which would contradict (4)(i).

3.3. *The function γ we are about to define will be a basic tool in our work.* We define $\chi\colon [0,1]\to[0,\sqrt{2}/2]$ by letting

$$\chi(t)=\begin{cases} t & \text{if} \quad 0\leqq t\leqq \sqrt{2}/2, \\ 4(\sqrt{2}/2+1/4-t) & \text{if} \quad \sqrt{2}/2<t<\sqrt{2}/2+1/4, \\ 0 & \text{if } \sqrt{2}/2+1/4<t\leqq 1\end{cases}$$

Since the sets $M\cap\mathbf{p}^{-1}(\{x\})$ are compact for each $x\in X$ *we may define* $\gamma\colon X\to[0,\sqrt{2}/2]$ by setting

$$\gamma(x)=\max\,[\{0\}\cup\{\chi\circ\beta(z)\colon z\in M\cap\mathbf{p}^{-1}(\{x\})\}].$$

Lemma. *We have*

(1) $\gamma(x)=\sqrt{2}/2$ *whenever* $x\in B$;

(2) $\gamma\in\mathscr{L}(X)$;

(3) $\int\psi|\nabla\gamma|^2\,d\mathscr{L}^n\leqq 16\int\psi\circ\mathbf{p}\,\|Dv\|^2\,d\|M\|$

whenever ψ *is a nonnegative Borel function on* X.

Proof. (1) is obvious. To prove (2) and (3) we suppose $a\in X$ and $\sqrt{2}/2+1/4<\lambda<1$ and set $T=\{z\in M:\mathbf{p}(z)=a \text{ and } \beta(z)\leqq\lambda\}$. If $T=\emptyset$, it is easily seen that γ vanishes near a. Suppose $T\neq\emptyset$. Since T is compact and $M\in\mathscr{M}^2$, T is finite; moreover, there are a positive real number δ and, for each $z\in T$, a function $\Lambda_z: U(a,\delta)\to M$ of class 2 such that $\mathbf{U}(a,\delta)\subset X$, $\mathbf{p}\circ\Lambda_z(x)=x$ for $x\in\mathbf{U}(a,\delta)$ and

$$\{z\in M\cap\mathbf{C}(a,\delta):\beta(z)\leqq\sqrt{2}/2+1/4\}\subset\cup\{\operatorname{rng}\Lambda_z: z\in T\}.$$

This implies

$$\gamma|\mathbf{U}(a,\delta)=\max\{\chi\circ\beta\circ\Lambda_z: z\in T\}.$$

Thus local versions of (2) and (3) hold near each point of X so the proof is complete.

3.4. *For the remainder of this chapter we assume* (1), (2) *below hold and adopt the terminology introduced here.*

(1) W *is a connected component of* $\mathbf{U}(0,2R)\sim B$;

(2) $G=N\cap\mathbf{p}^{-1}(W)\neq\emptyset$.

Note that W is open, G is open relative to M and that $\mathbf{p}$ carries each connected component of G univalently onto W. *Hence the number*

$$\kappa$$

of these components is a positive integer. We let

$$\mathbf{K}=\{1,\ldots,\kappa\}$$

and note that *for each* $i\in\mathbf{K}$ *there are functions*

$$f_i: W\to Y,\qquad F_i: W\to M$$

of class 2 *uniquely determined by the conditions*

(7) $F_i(x)=(x,f_i(x))$ *for each* $i\in\mathbf{K}$ *and each* $x\in W$,

(8) $f_1<\ldots<f_\kappa$,

(9) $G=\bigcup_{i=1}^{\kappa}G_i$

where *we have set*

$$G_i=\operatorname{rng}F_i\quad\text{for each } i\in\mathbf{K};$$

for each $i \in \mathbf{K}$ *we let*

$$s_i = \begin{cases} 1 & \text{if } v \cdot e_{n+1} > 0 \quad \text{on } G_i, \\ -1 & \text{if } v \cdot e_{n+1} < 0 \quad \text{on } G_i; \end{cases}$$

we let

$$\mathscr{U}$$

be the set of κ-tuples $(u_1, \ldots, u_\kappa) \in (\mathbf{S}^n)^\kappa$ such that for each $i \in \mathbf{K}$ we have $u_i \cdot e_{n+1} > 0$ if $s_i = 1$ and $u_i \cdot e_{n+1} < 0$ if $s_i = -1$; *we let*

$$\boldsymbol{\delta}(u, v) = \max \{(1 - u_i \cdot v_i^2)^{1/2} : i \in \mathbf{K}\} \quad \text{for } u, v \in \mathscr{U}$$

from 1.8(3) we see that $\boldsymbol{\delta}$ *is a metric on* $\mathscr{U}$; *we let*

$$o \in \mathscr{U}$$

be such that $o_i = s_i e_{n+1}$ *for* $i \in \mathbf{K}$.

Whenever $a \in X$, $0 < r < \infty$, $|a| + r < 2R$, $0 < \lambda \leqq 1$ *and* $u \in \mathscr{U}$ *we set*

$$\begin{aligned}
A(a, r) &= r^{-n} \|M\| (\mathbf{C}(a, r)), \\
V(a, r, \lambda) &= r^{-n} \|M\| (\{z \in G \cap \mathbf{C}(a, r) : \beta(z) \geqq \lambda\}), \\
P(a, r) &= r^{-n} \|M\| (\mathbf{C}(a, r) \sim G), \\
D(a, r, u) &= \sum_{i=1}^{\kappa} r^{-n} \int_{\mathbf{C}(a, r) \cap G_i} 1 - u_i \cdot v^2 \, d\|M\|, \\
O(a, r, u) &= P(a, r) + D(a, r, u), \\
K(a, r) &= r^{2-n} \int \phi^2 \|Dv\|^2 \, d\|M\|, \\
K(a, r, \lambda) &= r^{2-n} \int_{\{z \in M : z \notin G \text{ or } \gamma \circ \pi(z) \geqq \lambda\}} \phi^2 \|Dv\|^2 \, d\|M\|;
\end{aligned}$$

here, in the definition of $K(a, r)$ and $K(a, r, \lambda)$, $\phi(z) = 1$ if $z \in \mathbf{C}(a, r/2)$ and $\phi(z) = (2/r)(r - |\mathbf{p}(z) - a|)$ if $z \in \mathbf{C}(a, r) \sim \mathbf{C}(a, r/2)$.

The reader may find it useful, as we do, to associate "A" with "area", "V" with "vertical", "P" with "peripheral", "D" with "Dirichlet", "O" with "oscillation" and "K" with "curvature". We will show that if $O(a, r, u)$ is sufficiently small and other reasonable conditions are met then $O(a, t, v)$ behaves like a positive power function of t as $t \downarrow 0$ for v chosen appropriately.

Whenever $0 < \lambda \leqq 1$ *and* $0 < \tau < \infty$ *we define*

$$\rho_{\lambda, \tau} : [0, 1] \to [0, 1], \quad \sigma_\lambda : [0, 1] \to [0, 1]$$

as follows:

$$\rho_{\lambda, \tau}(t) = \begin{cases} 0 & \text{if } 0 \leqq t < \lambda e^{-1/\tau}, \\ \tau \log(t \lambda^{-1} e^{1/\tau}) & \text{if } \lambda e^{-1/\tau} \leqq t < \lambda, \\ 1 & \text{if } \lambda \leqq t \leqq 1; \end{cases}$$

$$\sigma_\lambda(t) = \begin{cases} 1 & \text{if } 0 \leqq t < \lambda/2, \\ (2/\lambda)(\lambda - t) & \text{if } \lambda/2 \leqq t < \lambda, \\ 0 & \text{if } \lambda \leqq t \leqq 1. \end{cases}$$

Whenever $0<\lambda\leqq\sqrt{2}/2$ *we define*

$$\mu_\lambda=\sigma_\lambda\circ\gamma\in\mathscr{L}(X).$$

3.5. **Definition.** *We let*

$$\lambda_0=(\sqrt{2}/4)(\zeta_7\kappa)^{-1},\qquad \lambda_1=\lambda_0 e^{-1/\lambda_0}.$$

Note that $0<\lambda_0\leqq\sqrt{2}/4$, $0<\lambda_1\leqq\sqrt{2}/8$.

Lemma. *Suppose* $a\in X$, $0<r<\infty$, $|a|+r<2R$, $0<\lambda\leqq\sqrt{2}/2$, $0<\tau<\infty$ *and* $u\in\mathscr{U}$. *Then*

(1) $K(a,r,\lambda)\leqq\zeta_5[\tau^2\kappa K(a,r)+V(a,r,\lambda e^{-1/\tau})+P(a,r)+\Delta r^2 A(a,r)]$;

(2) $K(a,r)-K(a,r,\lambda)\leqq\zeta_6[\kappa K(a,r,\lambda)+D(a,r,u)+\Delta r^2 A(a,r)]$;

provided $0<\lambda\leqq\sqrt{2}/4$ *and* $\boldsymbol{\delta}(u,o)\leqq\lambda$;

(3) $K(a,r)\leqq\zeta_8\kappa[D(a,r,u)+V(a,r,\lambda_1)+\Delta r^2 A(a,r)]$

provided $\boldsymbol{\delta}(u,o)\leqq\lambda_0$.

Proof. Let ϕ be as in the definition of $K(a,r)$. To prove (1) we let $\psi\in\mathscr{L}(M)$ equal 1 on $M\sim G$ and equal $\rho_{\lambda,\tau}\circ\gamma\circ\pi$ on G. Using 2.4 with u there equal e_{n+1} we obtain

$$\begin{aligned}K(a,r,\lambda)&\leqq\int(\psi\phi)^2\|Dv\|^2\,d\|M\|\\&\leqq\zeta_3\int|\nabla(\psi\phi)|^2\beta^2+\Delta(\psi\phi)^2\,d\|M\|\\&\leqq\zeta_3\int 2|\nabla\phi|^2\psi^2\beta^2+\Delta(\psi\phi)^2\,d\|M\|\\&\quad+2\zeta_3\tau^2\int\limits_{\{z\in G:\,\gamma\circ\pi(z)>0\}}\phi^2(\gamma\circ\pi)^{-2}(|\nabla\gamma|\circ\pi)^2\beta^2\,d\|M\|.\end{aligned}$$

Evidently

$$\begin{aligned}&\int 2|\nabla\phi|^2\psi^2\beta^2+\Delta(\psi\phi)^2\,d\|M\|\\&\quad\leqq 8r^{n-2}[V(a,r,\lambda e^{-1/\tau})+P(a,r)+\Delta r^2 A(a,r)].\end{aligned}$$

For each $i\in\mathbf{K}$ we use 3.3(3) to estimate

$$\begin{aligned}&\int\limits_{\{z\in G_i:\,\gamma\circ\pi(z)>0\}}\phi^2(\gamma\circ\pi)^{-2}(|\nabla\gamma|\circ\pi)^2\beta^2\,d\|M\|\\&\quad\leqq\int\limits_{G_i}\phi^2(|\nabla\gamma|\circ\pi)^2\,d\|M\|\\&\quad\leqq\sqrt{2}\int\limits_{W}(\phi\circ F_i)^2|\nabla\gamma|^2\,d\mathscr{L}^n\\&\quad\leqq\sqrt{2}\,16\int\phi^2\|Dv\|^2\,d\|M\|.\end{aligned}$$

This proves (1).

To prove (2), we let $\psi=\sigma_{2\lambda}\circ\gamma\circ\pi|G\in\mathscr{L}_c(G)$ and suppose $i\in\mathbf{K}$. Using 2.4 with u there equal u_i and 1.8(3) we obtain

$$\int_{\{z\in G_i:\,\gamma\circ\pi(z)<\lambda\}} \phi^2\|Dv\|^2\,d\|M\|$$
$$\leqq \int_{G_i}(\phi\psi)^2\,d\|M\| \leqq \zeta_3\int_{G_i}|\nabla(\phi\psi)|^2(1-u_i\cdot v^2)+\Delta(\phi\psi)^2\,d\|M\|$$
$$\leqq 2\zeta_3\int_{G_i}|\nabla\phi|^2\,1-u_i\cdot v^2\,d\|M\|$$
$$+2\zeta_3\int_{\{z\in G_i:\,\lambda\leqq\gamma\circ\pi(z)\leqq 2\lambda\}}\phi^2\lambda^{-2}(|\nabla\gamma|\circ\pi)^2\,2(\beta^2+\lambda^2)\,d\|M\|$$
$$+\zeta_3\Delta\|M\|[\mathbf{C}(a,r)\cap G_i].$$

Using 3.3(3) we see that

$$\int_{\{z\in G_i:\,\lambda\leqq\gamma\circ\pi(z)\leqq 2\lambda\}}\phi^2\lambda^{-2}(|\nabla\gamma|\circ\pi)^2\,2(\beta^2+\lambda^2)\,d\|M\|$$
$$\leqq 10\int_{\{z\in G_i:\,\lambda\leqq\gamma\circ\pi(z)\}}\phi^2(|\nabla\gamma|\circ\pi)^2\,d\|M\|$$
$$\leqq 10\sqrt{2}\int_{\{x\in W:\,\lambda\leqq\gamma(x)\}}(\phi\circ F_i)^2\,|\nabla\gamma|^2\,d\mathscr{L}^n$$
$$\leqq 10\sqrt{2}\,16\int_{\{z\in M:\,\lambda\leqq\gamma\circ\pi(z)\}}\phi^2\|Dv\|^2\,d\|M\|.$$

Thus (2) is proved.

To prove (3), we use (1) twice and (2) once to estimate

$$K(a,r)=K(a,r,\lambda)+[K(a,r)-K(a,r,\lambda)]$$
$$\leqq\zeta_5[\tau^2\kappa K(a,r)+V(a,r,\lambda e^{-1/\tau})+P(a,r)+\Delta r^2A(a,r)]$$
$$+\zeta_6[\kappa K(a,r,\lambda)+D(a,r,u)+\Delta r^2A(a,r)]$$
$$\leqq\zeta_5[\tau^2\kappa K(a,r)+V(a,r,\lambda e^{-1/\tau})+P(a,r)+\Delta r^2A(a,r)]$$
$$+\zeta_6\kappa\zeta_5[\tau^2\kappa K(a,r)+V(a,r,\lambda e^{-1/\tau})+P(a,r)+\Delta r^2A(a,r)]$$
$$+\zeta_6[D(a,r,u)+\Delta r^2A(a,r)];$$

we then set $\tau=\lambda=\lambda_0$.

3.6. **Lemma.** *Suppose* $a\in X$, $0<r<\infty$, $|a|+2r<2R$ *and* $0<\lambda\leqq\sqrt{2}/2$. *Then*

$$P(a,r)+V(a,r,\lambda)$$
$$\leqq\mathbf{c}_{13}[P(a,2r)+\kappa(2/\lambda)\,V(a,2r,\lambda/2)+\kappa(2/\lambda)\,K(a,2r)]^{n/(n-1)}.$$

Proof. Let $\psi\in\mathscr{D}(M)$ equal 1 on $M\sim G$ and $1-\sigma_\lambda\circ\gamma\circ\pi$ on G and let ϕ be as in the definition of $K(a,2r)$. Using 1.3(12) we estimate

$$r^{n-1}[P(a,r)+V(a,r,\lambda)]^{(n-1)/n}$$
$$\leqq(\int(\phi\psi)^{n/(n-1)}\,d\|M\|)^{(n-1)/n}$$
$$\leqq\mathbf{c}_1(\int|\nabla(\phi\psi)|+|\phi\psi|\,|H|\,d\|M\|)$$
$$\leqq\mathbf{c}_1\int_{\{z\in M:\,z\notin G\text{ or }\gamma\circ\pi(z)\geqq\lambda/2\}}|\nabla\phi|+\phi|H|\,d\|M\|$$
$$+\mathbf{c}_1(2/\lambda)\int_{\{z\in G:\,\gamma\circ\pi(z)\geqq\lambda/2\}}\phi(|\nabla\phi|\circ\pi)\,d\|M\|.$$

We have

$$\int_{M\sim G} |\nabla\phi| + \phi|H|\, d\|M\|$$
$$\leqq (1/4)(2r)^n P(a,2r) + (1/2)(r^{-1}(2r)^n P(a,2r)$$
$$+ r n^2 \int \phi^2 \|Dv\|^2\, d\|M\|).$$

Using 3.3(3) we obtain

$$\int_{\{z\in G_i:\, \gamma\circ\pi(z)\geqq\lambda/2\}} \phi(|\nabla\gamma|\circ\pi)\, d\|M\|$$
$$\leqq \sqrt{2} \int_{\{x\in W:\, \gamma(x)\geqq\lambda/2\}} \phi\circ F_i |\nabla\gamma|\, d\|M\|$$
$$\leqq (\sqrt{2}/2)(r^{-1}\mathscr{L}^n\{x\in W\cap \mathbf{U}(a,2r):\, \gamma(x)\geqq\lambda/2\}$$
$$+ r\int_W (\phi\circ F_i)^2 |\nabla\gamma|^2\, d\mathscr{L}^n)$$
$$\leqq (\sqrt{2}/2)(r^{-1}(2r)^n V(a,2r,\lambda/2) + 16r\int\phi^2\|Dv\|^2\, d\|M\|)$$

for each $i\in\mathbf{K}$. We sum over $\mathbf{K}$ to complete the proof.

3.7. **Lemma.** *Suppose* $a\in X$, $0<r<\infty$, $|a|+r<2R$, $0<\lambda\leqq 1$, $u\in\mathscr{U}$ *and*

$$\delta(u,0)\leqq\lambda/2.$$

Then

$$V(a,r,\lambda)\leqq(2/\lambda)^2 D(a,r,u).$$

Proof. Suppose $i\in\mathbf{K}$. Using 1.8(3) we obtain

$$\int_{\{z\in G_i\cap\mathbf{C}(a,r):\, \beta(z)\geqq\lambda\}} 1-u_i\cdot v^2\, d\|M\|$$
$$\leqq \int_{\{z\in G_i\cap\mathbf{C}(a,r):\, \beta(z)\geqq\lambda\}} \tfrac{1}{2}\beta^2 - (1-u_i\cdot e_{n+1})^2\, d\|M\|$$
$$\geqq \frac{\lambda^2}{4}\|M\|(\{z\in G_i\cap\mathbf{C}(a,r):\, \beta(z)\geqq\lambda\}).$$

3.8. *Whenever* $i\in\mathbf{K}$, $w\in S^n$, $w\cdot e_{n+1}\neq 0$ *and* E *is a Borel subset of* W we may use 1.8(1) to estimate

(1) $$\int_E |\nabla f_i - \iota(w)|^2\, d\mathscr{L}^n \leqq \int_{G_i\cap\mathbf{p}^{-1}(E)} |\iota(v)-\iota(w)|^2\, d\|M\|$$
$$\leqq 2(w\cdot e_{n+1})^{-2} \int_{G_i\cap\mathbf{p}^{-1}(E)} 1-w\cdot v^2\, d\|M\|$$

and

(2) $$\int_{G_i\cap\mathbf{p}^{-1}(E)} 1-w\cdot v^2\, d\|M\| \leqq \int_{G_i\cap\mathbf{p}^{-1}(E)} (v\cdot e_{n+1})^2 |\iota(v)-\iota(w)|^2\, d\|M\|$$
$$\leqq \int_E |\nabla f_i - w|^2\, d\mathscr{L}^n.$$

3.9. *Throughout this section we suppose* $a \in X$, $0 < r < \infty$, $|a| + 2r < 2R$, $u \in \mathscr{U}$ *and we set*

$$\xi = O(a, 2r, u).$$

Lemma. *Suppose*

$$\delta(u, 0) \leqq \lambda_1/2; \tag{1}$$

$$\kappa \Delta r^2 \leqq \xi \leqq 1. \tag{2}$$

Then

$$K(a, 2r, u) \leqq \zeta_9 \kappa \lambda_1^{-2} \xi. \tag{3}$$

Proof. By 3.7, $V(a, 2r, \lambda_1) \leqq (\lambda_1/2)^{-1} D(a, 2r, u)$ and it is evident that $A(a, 2r) \leqq \sqrt{2}\, \kappa \alpha(n) + P(a, 2r)$ so (3) now follows from 3.5(3).

Lemma. *Suppose*

$$0 < \lambda \leqq \lambda_1 \quad \text{and} \quad \delta(u, 0) \leqq \lambda/4; \tag{4}$$

$$\kappa \Delta r^2 \leqq \xi \leqq 1. \tag{5}$$

Then

$$P(a, r) + V(a, r, \lambda) \leqq \zeta_{10} (\kappa^2/\lambda^3)^{n/(n-1)} \xi^{n/(n-1)}. \tag{6}$$

Proof. Use 3.6, 3.7 and (3).

Lemma. *Suppose*

$$0 < \lambda \leqq 2\lambda_1, \quad 0 < \tau < \infty \quad \text{and} \quad \delta(u, 0) \leqq \lambda e^{-1/\tau}/8; \tag{7}$$

$$0 \leqq \omega \leqq 1, \quad \xi \leqq 1 \quad \text{and} \quad \kappa \Delta r^2 \leqq \omega \xi. \tag{8}$$

Then

$$r^{-n} \mathscr{L}^n(\{x \in \mathbf{U}(a, r) : 0 < \mu_\lambda(x) < 1\}) \leqq \zeta_{10} (8\kappa^2/\lambda^3)^{n/(n-1)} \xi^{n/(n-1)}; \tag{9}$$

$$r^{2-n} \int_{\mathbf{U}(a, 2r)} |\nabla \mu_\lambda|^2 \, d\mathscr{L}^n \leqq \zeta_{11} \lambda^{-2} [\kappa(\kappa \lambda_1^{-2} \tau^2 + \omega) \xi + (\kappa^2 e^{3/\tau}/\lambda^3)^{n/(n-1)} \xi^{n/(n-1)}]. \tag{10}$$

Proof. To prove (9) we observe that

$$\begin{aligned} &r^{-n} \mathscr{L}^n(\{x \in \mathbf{U}(a, r) : 0 < \mu_\lambda(x) < 1\}) \\ &\quad \leqq r^{-n} \mathscr{L}^n(\{x \in \mathbf{U}(a, r) \cap W : \gamma \circ \pi(x) \geqq \lambda/2\}) \\ &\quad \leqq V(a, r, \lambda/2) \end{aligned}$$

and invoke (6).

To prove (10) we first use 3.3(3) to estimate

$$\begin{aligned} &r^{2-n} \int_{\mathbf{U}(a, r)} |\nabla \mu_\lambda|^2 \, d\mathscr{L}^n \\ &\quad = r^{2-n} \int_{\{x \in W \cap \mathbf{U}(a, r) : \lambda/2 < \gamma \circ \pi(x) < \lambda\}} (2/\lambda)^2 |\nabla \gamma|^2 \, d\mathscr{L}^n \\ &\quad \leqq (2/\lambda)^2 \, 16 K(a, 2r, \lambda/2). \end{aligned}$$

Second, we use 3.5(1) to estimate

$$\lambda^{-2} K(a, 2r, \lambda/2)$$
$$\leqq \zeta_5 \lambda^{-2}[\kappa \tau^2 K(a, 2r) + V(a, 2r, \lambda e^{-1/\tau}/2) + P(a, 2r) + \Delta r^2 A(a, 2r)].$$

Third, we use (3) to estimate

$$\kappa K(a, 2r) \leqq 2^n \kappa \zeta_9 \kappa \lambda_1^{-2} \xi.$$

Fourth, we use (6) to estimate

$$\lambda^{-2}[V(a, 2r, \lambda e^{-1/\tau}/2) + P(a, 2r)]$$
$$\leqq \lambda^{-2} \zeta_{10}(8 e^{3/\tau} \kappa^2/\lambda^3)^{n/(n-1)} \xi^{n/(n-1)}.$$

Finally we estimate

$$\lambda^{-2} \Delta r^2 A(a, 2r) \leqq \lambda^{-2} \omega \xi[\sqrt{2}\kappa + P(a, 2r)].$$

Lemma. *Suppose*

(11) $$\delta(u, 0) \leqq \lambda_1/2;$$

(12) $$\kappa \Delta r^2 \leqq \xi \leqq 1;$$

(13) $$\zeta_{12} \kappa \lambda_1^{-2} \xi \leqq 1.$$

Then

(14) *there is a unique connected component* Q *of* $\mathbf{U}(a, r) \sim B$ *such that* $\mathscr{L}^n(Q) > (1/2)\alpha(n) r^n$;

(15) $\mathscr{L}^n(\mathbf{U}(a, r) \sim Q) \leqq \zeta_{13}(\kappa \lambda_1^{-2} \xi)^{n/(n-1)} r^n$;

(16) *if* $Q \subset W$ *then there is for each* $i \in \mathbf{K}$ *a point*

$$b_i \in Y$$

such that,

$$|f_i(x) - b_i| \leqq \zeta_{14} \kappa r \quad \text{whenever } x \in Q.$$

Remark. When in the course of the proof we cite (13) we are in effect requiring ζ_{12} to be sufficiently large, or, what is the same thing, requiring $\kappa \lambda_1^{-2} \xi$ to be sufficiently small.

Proof. Part One. From (11), (12), 3.7 and (3) we obtain

(17)
$$\begin{aligned} \varepsilon &= r^{-n} \|M\|(\{z \in M \cap \mathbf{C}(a, r) : \beta(z) \geqq \sqrt{2}/4\}) \\ &\quad + r^{2-n} \int_{\mathbf{C}(a, r)} \|Dv\|^2 \, d\|M\| \\ &\leqq 2^n[P(a, 2r) + V(a, 2r, \sqrt{2}/4) + K(a, 2r)] \\ &\leqq 3 \cdot 2^n \zeta_9 \kappa \lambda_1^{-2} \xi \\ &= (1/2)\tilde{\xi} \end{aligned}$$

where we have set $\tilde{\xi} = 3 \cdot 2^{n+1} \zeta_9 \kappa \lambda_1^{-2} \xi$. From (13) we obtain

$$\xi \leqq \tilde{\xi} \leqq 1. \tag{18}$$

Using (17), (13) we apply 3.2 to obtain an open connected subset C of $\mathbf{U}(a,r) \sim B$ such that

$$\mathscr{L}^n(C) > (1/2)\alpha(n) r^n, \tag{19}$$

$$\mathscr{H}^{n-1}(E) \leqq 2^{-n} \tilde{\xi} r^{n-1}, \tag{20}$$

$$\mathscr{L}^n(\mathbf{U}(a,r) \sim C) \leqq \mathbf{c}_2 (2^{-n} \tilde{\xi})^{n/(n-1)} r^n < (1/2)\alpha(n) r^n, \tag{21}$$

where we have set E equal the boundary relative to $\mathbf{U}(a,r)$ of C. Let Q equal the connected component of $\mathbf{U}(a,r) \sim B$ which contains C.

From (19), (20) it is clear that (14), (15) hold.

Part Two. Suppose D is a connected component of $\{z \in M : |p(z) - a| \leqq r\}$. For each $j = 0, 1, 2, \ldots$ let $q_j = \min \mathbf{p}(D) + 2jr$. Let $J = \min\{j : q_j \notin \mathbf{q}(D)\}$. For each $j \in \{0, 1, \ldots, J-1\}$ choose $c_j \in D$ such that $q_j = \mathbf{q}(c_j)$. From (12) and 2.2(9) we infer that

$$\begin{aligned}
(\operatorname{diam} D) \zeta_2^{-1} r^n &\leqq (2Jr) \zeta_2^{-1} r^n \\
&\leqq 2r \sum_{j=0}^{J-1} \|M\| (\{z \in \mathbf{R}^{n+1} : |z - c| < r\}) \\
&\leqq 2r \|M\| (\mathbf{C}(a, 2r) \\
&\leqq 2r [\|M\| (\mathbf{C}(a,2r) \cap G) + \|M\| (\mathbf{C}(a,2r) \sim G)] \\
&\leqq 2r [\sqrt{2} \kappa \alpha(n) (2r)^n + O(a, 2r, u) (2r)^n].
\end{aligned}$$

Thus

$$\operatorname{diam} D \leqq \zeta_2 2^{n+1} (\sqrt{2} \kappa \alpha(n) + 1). \tag{22}$$

Thus (16) holds if for each $i \in \mathbf{K}$ we take b_i to be any point in the closure of $\mathbf{q} \circ F_i(Q)$.

The Main Lemma. Suppose

$$0 \leqq \omega \leqq 1, \quad 0 < \delta \leqq 1, \quad 0 < \lambda \leqq \lambda_1, \quad 0 < \tau \leqq \delta \lambda_1 \lambda; \tag{23}$$

$$\boldsymbol{\delta}(u, 0) \leqq \lambda e^{-1/\tau}/8; \tag{24}$$

$$\kappa \Delta r^2 \leqq \omega \xi; \tag{25}$$

$$\zeta_{12} \kappa \lambda_1^{-2} \xi \leqq 1; \tag{26}$$

$$\lambda^{-2} \omega \leqq \delta^2; \tag{27}$$

$$\lambda^{-2} (e^{3/\tau}/\lambda^3)^{n/(n-1)} \xi^{1/(n-1)} \leqq \delta^2; \tag{28}$$

$$0 < \theta \leqq 1/2. \tag{29}$$

Then (14)–(16) *hold and either* $Q \cap W = \emptyset$ *and*

$$O(a, \theta r, u) \leqq \zeta_{15} \theta^{-n} \delta^2 \xi \tag{30}$$

or $Q \subset W$ *and*

(31) *there is* $v \in \mathscr{U}$ *such that*

(i) $\boldsymbol{\delta}(u, v)^2 \leqq \zeta_{18} \kappa^5 \xi$,

(ii) $O(a, \theta r, v) \leqq \zeta_{20} \kappa^5 [\theta^{-n} \max\{\lambda, \delta\} + \theta^2] \xi$.

Proof. Part One. Note that (14)–(18) hold. Let Q be as (14). Either $Q \cap W = \emptyset$ or $Q \subset W$.

Suppose $Q \cap W = \emptyset$. Using (15) we obtain

$$\begin{aligned} D(a, \theta r, u) &\leqq \theta^{-n} D(a, r, u) \\ &\leqq \theta^{-n} \sqrt{2}\, \kappa\, \mathscr{L}^n(\mathbf{U}(a, r) \cap W)\, r^{-n} \\ &\leqq \theta^{-n} \sqrt{2}\, \kappa\, \zeta_{13} (\kappa \lambda_1^{-2} \xi)^{n/(n-1)} \end{aligned}$$

and using (6) we obtain

$$\begin{aligned} P(a, \theta r) &\leqq \theta^{-n} P(a, r) \\ &\leqq \theta^{-n} \zeta_{10} (\kappa/\lambda_1)^{2n/(n-1)}\, \xi^{n/(n-1)} \end{aligned}$$

so (30) follows from (28) since $\tau \leqq \lambda_1 \leqq e^{-\kappa}$.

Henceforth we assume $Q \subset W$.

Part Two. Linearization of the first variation. (Ultimately, we shall apply 1.11.) For each $i \in \mathbf{K}$ we let b_i be as in (16), we let

$$\alpha_i = \iota(u_i)$$

and we let

$$g_i(x) = f_i(x) - \alpha_i \cdot (x - a) - b_i \quad \text{for } x \in Q.$$

We fix $i \in \mathbf{K}$ for the remainder of Part Two.

We define $h_i \in \mathscr{L}(\mathbf{U}(a, r))$ by letting

$$h_i(x) = \begin{cases} \mu_\lambda(x)\, g_i(x) & \text{for } x \in Q, \\ 0 & \text{for } x \in \mathbf{U}(a, r) \sim Q. \end{cases}$$

We define $l_i \in \mathrm{Hom}(\mathbf{R}^n, \mathbf{R}^n)$ by requiring that

$$\langle \alpha, l_i \rangle \cdot \beta = \langle (0, 0, \alpha) \odot (0, 0, \beta), D^2 \Phi^{\S}_{s_i}(a, b_i, \alpha_i) \rangle$$

whenever $\alpha, \beta \in \mathbf{R}^n$. For each $x \in Q$ we let $\Omega_i(x) \in \odot^2 \mathbf{R}^n$ be such that

$$\langle \alpha^2, \Omega_i(x) \rangle = \int_0^1 \langle (0, 0, \alpha)^2, D^2 \Phi^{\S}_{s_i}(a, b_i, \alpha_i + t \nabla g_i(x)) - D^2 \Phi^{\S}_{s_i}(a, b_i, \alpha_i) \rangle\, dt$$

whenever $\alpha \in \mathbf{R}^n$.

Let j be the inclusion map of $\mathbf{U}(a, r)$ in $\mathbf{R}^n$. For each $\phi \in \mathscr{C}^1_c(U(a, r))$ we let

$$T^i(\phi) = \int_Q \langle (0, \mu_\lambda \phi, \nabla(\mu_\lambda \phi)), D\Phi^{\S}_{s_i}(j, f_i, \nabla f_i) \rangle\, d\mathscr{L}^n,$$

$$U^i(\phi) = \int_{\mathbf{U}(a, r)} \nabla \phi \cdot \langle \nabla h_i, l_i \rangle\, d\mathscr{L}^n;$$

we let

$$T_1^i(\phi)=\int_Q \langle (0,\mu_\lambda\phi,0), D\Phi_{s_i}^\S(j,f_i,\nabla f_i)\rangle\, d\mathscr{L}^n,$$

$$T_2^i(\phi)=\int_Q \langle (0,0,\nabla(\mu_\lambda\phi)), D\Phi_{s_i}^\S(j,f_i,\nabla f_i)-D\Phi_{s_i}^\S(a,b_i,\nabla f_i)\rangle\, d\mathscr{L}^n,$$

$$T_3^i(\phi)=\int_Q \langle \nabla(\mu_\lambda\phi)\odot\nabla g_i, \Omega_i\rangle\, d\mathscr{L}^n,$$

$$T_4^i(\phi)=\int_Q (0,0,\nabla(\mu_\lambda\phi))\, D\Phi_{s_i}^\S(a,b_i,\alpha_i)\rangle\, d\mathscr{L}^n,$$

$$T_5^i(\Phi)=\int_Q \langle (0,0,\nabla(\mu_\lambda\phi))\odot(0,0,\nabla g_i), D^2\Phi_{s_i}^\S(a,b_i,\alpha_i)\rangle\, d\mathscr{L}^n,$$

$$T_6^i(\phi)=\int_{G_i} L\phi\circ\mathbf{p}\, v\cdot e_{n+1}\, d\|M\|;$$

we let

$$U_1^i(\phi)=\int_Q g_i\nabla\phi\cdot\langle\nabla\mu_\lambda, l_i\rangle\, d\mathscr{L}^n,$$

$$U_2^i(\phi)=\int_Q \nabla(\mu_\lambda\phi)\cdot\langle\nabla g_i, l_i\rangle\, d\mathscr{L}^n,$$

$$U_3^i(\phi)=-\int_Q \phi\nabla\mu_\lambda\cdot\langle\nabla g_i, l_i\rangle\, d\mathscr{L}^n.$$

From Taylor's Theorem we see that

$$\begin{aligned}\langle\nabla(\mu_\lambda\phi)\odot\nabla g_u, \Omega_i\rangle &= \langle (0,0,\nabla(\mu_\lambda\phi)), D\Phi_{s_i}^\S(a,b_i,\nabla f_i)-D\Phi_{s_i}^\S(a,b_i,\alpha_i)\rangle\\ &\quad -\langle (0,0,\nabla(\mu_\lambda\phi))\odot(0,0,\nabla g_i), D^2\Phi_{s_i}^\S(a,b_i,\alpha_i)\rangle\end{aligned}$$

on Q from which we see that $T^i=\sum_{j=1}^5 T_j^i$; in fact, the corresponding equation holds for the integrands defining these quantities. Using 1.7 and 1.2(11) we see that $T^i=T_6^i$. It is obvious that $U^i=\sum_{j=1}^3 U_j^i$ and that $U_2^i=T_5^i$. Finally $T_4^i=0$. Thus

$$U^i=-(T_1^i+T_2^i+T_3^i)+T_6^i+U_1^i+U_3^i. \tag{32}$$

Part Three. Suppose $i\in\mathbf{K}$. We now estimate U^i.

We begin with some simple preliminary estimates. We have

$$\begin{aligned}1\leqq|(-\alpha_i,1)+(-t\nabla g_i,1)| &= |(1-t)(-\alpha_i,1)+t(-\nabla f_i,1)|\\ &\leqq\sqrt{2}\quad \text{on } Q \text{ for } 0\leqq t\leqq 1.\end{aligned} \tag{33}$$

We have

$$|\nabla g_i|\leqq 4\lambda\quad \text{on } \{x\in Q: \mu_\lambda(x)>0\}; \tag{34}$$

in fact, if $x\in Q$, $\mu_\lambda(x)>0$ and $w_j=v(F_j(x))$ for $j\in\mathbf{K}$ we have $\beta(w_j)<\lambda$ for $j\in\mathbf{K}$ so that 1.8(1) and (24) imply

$$\begin{aligned}|\nabla g_i(x)|^2 &\leqq (w_i \cdot e_{n+1})^{-2}(u_i \cdot e_{n+1})^{-2}(1-u_i \cdot w_i^2)\\ &\leqq 4\,\boldsymbol{\delta}(u,w)^2\\ &\leqq 8[\boldsymbol{\delta}(u,0)^2+\boldsymbol{\delta}(w,0)^2]\\ &\leqq 16\,\lambda^2.\end{aligned}$$

Using 3.8(1) we obtain

$$r^{-n}\int_Q |\nabla g_i|^2\, d\mathscr{L}^n \leqq 4D(a,r,u) \leqq 2^{n+2}\,\xi. \tag{35}$$

In making the estimates below we will use (33) and 1.2(11) repeatedly and without acknowledgement. Henceforth, we let $\|\cdot\|_1$ and $\|\cdot\|_2$ be as in 1.11 and we set

$$\Lambda = \mathrm{r}^{2-n}\int_{\mathbf{U}(a,r)} |\nabla \mu_\lambda|^2\, d\mathscr{L}^n.$$

Suppose $\phi \in \mathscr{C}^1_c(\mathbf{U}(a,r))$. Using (25) we obtain

$$|T_1^i(\phi)| \leqq \Gamma_{1,0}\int_{\mathbf{U}(a,r)} |\phi|\, d\mathscr{L}^n \leqq (\omega\,\xi)^{1/2}\,\|\phi\|_0\, r^n. \tag{36}$$

Using (16) and (25) we obtain

$$\begin{aligned}|T_2^i(\phi)| &\leqq \Gamma_{1,1}\,\zeta_{14}\,\kappa\, r\int_{\mathbf{U}(a,r)} |\nabla\phi| + |\phi|\,|\nabla\mu_\lambda|\, d\mathscr{L}^n\\ &\leqq \zeta_{14}\,\kappa(\omega\,\xi)^{1/2}[\boldsymbol{\alpha}(n)^{1/2}\,\|\phi\|_1 + \Lambda^{1/2}\,\|\phi\|_0]\, r^n.\end{aligned} \tag{37}$$

Using (34) and (35) we obtain

$$\begin{aligned}|T_3^i(\phi)| &\leqq \Gamma_{0,3}\int_{\mathbf{U}(a,r)} [|\nabla\phi| + |\phi|\,|\nabla\mu_\lambda|]\,|\nabla g_i|^2\, d\mathscr{L}^n\\ &\leqq \Gamma_{0,3}(4\lambda)\,\|\phi\|_1 (2^{n+2}\,\xi)^{1/2}\, r^n\\ &\quad + \Gamma_{0,3}(16\lambda^2)\,\|\phi\|_0\,\Lambda^{1/2}\, r^n.\end{aligned} \tag{38}$$

Using (25) we obtain

$$|T_6^i(\phi)| \leqq \upsilon_1 \int_{\mathbf{U}(a,r)} |\phi|\, d\mathscr{L}^n \leqq (\omega\,\xi)^{1/2}\,\|\phi\|_0. \tag{39}$$

Using (16) we obtain

$$\begin{aligned}|U_1^i(\phi)| &\leqq \zeta_{14}\,\kappa\, r\,\Gamma_{0,2}\int_{\mathbf{U}(a,r)} |\nabla\phi|\,|\nabla\mu_\lambda|\, d\mathscr{L}^n\\ &\leqq \zeta_{14}\,\kappa\,\Gamma_{0,2}\,\|\phi\|_1\,\Lambda^{1/2}\, r^n.\end{aligned} \tag{40}$$

Using (34) we obtain

$$\begin{aligned}|U_3^i(\phi)| &\leqq \Gamma_{0,2}(4\lambda)\int_{\mathbf{U}(a,r)} |\phi|\,|\nabla\mu_\lambda|\, d\mathscr{L}^n\\ &\leqq \Gamma_{0,2}(4\lambda)\,\|\phi\|_0\,\Lambda^{1/2}\, r^n.\end{aligned} \tag{41}$$

Using (35) and (16) we obtain

$$r^{-n}\int\limits_{\mathbf{U}(a,r)}|\nabla h_i|^2\,d\mathscr{L}^n\leqq 2r^{-n}\int\limits_{Q}|\nabla g_i|^2+g_i^2|\nabla\mu|^2\,d\mathscr{L}^n \tag{42}$$
$$\leqq 2[2^{n+2}\xi+(\zeta_{14}\kappa+1)^2\Lambda].$$

Using (10), (27) and (28) we infer that

$$\Lambda\leqq 3\zeta_{11}\kappa^3\delta^2\xi. \tag{43}$$

It follows from (27) and (36)–(43) that

$$|U^i(\phi)|\leqq\zeta_{16}\kappa^{5/2}\max\{\lambda,\delta\}\,\xi^{1/2}(\|\phi\|_0+\|\phi\|_1), \tag{44}$$

$$r^{-n}\int\limits_{\mathbf{U}(a,r)}|\nabla h_i|^2\,d\mathscr{L}^n\leqq\zeta_{17}\kappa^5\xi. \tag{45}$$

Finally, we have

$$\sqrt{2}/2|\alpha|\leqq|\langle\alpha,l_i\rangle|\leqq\Gamma_{0,2}|\alpha| \quad \textit{whenever}\ \ \alpha\in\mathbf{R}^n \tag{46}$$

Part Four. Conclusion. From (44)–(46) and 1.11 we secure

$$\tau_1,\ldots,\tau_\kappa\in\mathbf{R}^n$$

such that

$$|\tau_i|^2\leqq\zeta_{18}\kappa^5\xi, \tag{47}$$

$$(\theta r)^{-n}\int\limits_{\mathbf{U}(a,\theta r)}|\nabla h_i-\tau_i|^2\,d\mathscr{L}^n\leqq\zeta_{19}\kappa^5[\theta^{-n}\max\{\lambda,\delta\}+\theta^2\xi] \tag{48}$$

whenever $0<\theta\leqq 1/2$ *and* $i\in\mathbf{K}$.

For each $i\in\mathbf{K}$ we let $v_i\in\mathbf{S}^n$ be such that $v_i\cdot e_{n+1}=s_i$ and

$$\imath(v_i)=\imath(u_i)+\tau_i=\alpha_i+\tau_i.$$

From 1.8(1) we infer that

$$1-u_i\cdot v_i^2\leqq(u_i\cdot e_{n+1})^2\,|\imath(u_i)-\imath(v_i)|^2\leqq|\tau_i|^2$$

so that (31)(i) holds.

Suppose $0<\theta\leqq 1/2$. From 3.8(2) we obtain for each $i\in\mathbf{K}$ that

$$\begin{aligned}(\theta r)^{-n}&\int\limits_{G_i\cap\mathbf{p}^{-1}(Q)\cap\mathbf{C}(a,\theta r)}1-v_i\cdot\nu^2\,d\|M\|\\
&\leqq\int\limits_{Q\cap\mathbf{U}(a,\theta r)}|\nabla f_i-(\alpha_i+\tau_i)|^2\,d\mathscr{L}^n\\
&\leqq 3(\theta r)^{-n}\int\limits_{\mathbf{U}(a,\theta r)}|\nabla h_i-\tau_i|^2\,d\mathscr{L}^n\\
&\quad+3(\theta r)^{-n}\int\limits_{Q\cap\mathbf{U}(a,r)}g_i^2|\nabla\mu_\lambda|^2\,d\mathscr{L}^n\\
&\quad+3(\theta r)^{-n}\int\limits_{Q\cap\mathbf{U}(a,r)\cap\{x:\,0<\mu_\lambda(x)\}}(1-\mu_\lambda)^2|\nabla g_i|^2\,d\mathscr{L}^n;\end{aligned}$$

the second of the last three integrals does not exceed

$$3\theta^{-n}(\zeta_{14}\kappa+1)^2\Lambda$$

by (16); the last of the three integrals does not exceed

$$3\theta^{-n}(\zeta_{10}(8\kappa^2/\lambda^3)^{n/(n-1)})(16\lambda^2)$$

by (9) and (37). For each $i\in\mathbf{K}$ we obtain from (15) that

$$\begin{aligned}(\theta r)^{-n} &\int\limits_{[G_i\cap\mathbf{C}(a,r)]\sim\mathbf{p}^{-1}(Q)} 1-v_i\cdot\nu^2\, d\|M\| \\ &\leqq(\theta r)^{-n}\sqrt{2}\,\mathscr{L}^n(\mathbf{U}(a,r)\sim Q) \\ &\leqq(\theta r)^{-n}\sqrt{2}\,\zeta_{13}(\kappa\lambda_1^{-2}\xi)^{n/(n-1)}r^n.\end{aligned}$$

Finally, we obtain from (6) that

$$P(a,\theta r)\leqq\theta^{-n}P(a,r)\leqq\theta^{-n}\zeta_{10}(\kappa^2/\lambda^3)^{n/(n-1)}\xi^{n/(n-1)}.$$

Putting these estimates together, making sure to sum over $\mathbf{K}$ when appropriate and using (43) and (28) we obtain (31)(ii).

(49) **Definition.** We let

$$\begin{gathered}\theta_1=2^{-1}\zeta_{20}^{-1/2}\kappa^{-5/2}; \\ \delta_1=2^{-1/2}\zeta_{15}^{-1/2}\theta_1^{n/2}\kappa^{-5/2},\qquad \delta_2=4^{-1}\zeta_{20}^{-1}\kappa^{-5}\theta_1^n, \\ \delta_3=\min\{\delta_1,\delta_2\}; \\ \lambda_2=\min\{\lambda_1,\delta_2\}; \\ \tau_1=\delta_3\lambda_1\lambda_2; \\ \omega_1=\lambda_2^2\delta_3^2.\end{gathered}$$

Lemma. *Suppose*

(50) $$\boldsymbol{\delta}(u,\theta)\leqq\lambda_2 e^{-1/\tau_1}/8;$$

(51) $$\zeta_{12}\kappa\lambda_1^{-2}\xi\leqq 1;$$

(52) $$\lambda_2^{-2}(e^{3/\tau_1}/\lambda_2{}^3)^{n/(n-1)}\xi^{1/(n-1)}\leqq\delta_3^2.$$

Then there is $v\in\mathscr{U}$ such that

(53) $$\boldsymbol{\delta}(u,v)^2\leqq\zeta_{18}\kappa^5\xi$$

and

(54) $$O(a,\theta_1 r,v)\leqq\max\{\xi/2,(2\theta_1)^{-n}\omega_1^{-1}\kappa\Delta r^2\}.$$

Proof. In case $\kappa\Delta r^2>\omega_1\xi$, we have

$$O(a,\theta_1 r,u)\leqq(2\theta_1)^{-n}\xi$$
$$<(2\theta_1^{-n})\omega_1^{-1}\kappa\Delta r^2$$

so that (53), (54) hold with $v=u$. *In case* $\kappa\Delta r^2\leqq\omega_1\xi$, we let $\omega=\omega_1$, $\delta=\delta_3$, $\lambda=\lambda_2$ and $\tau=\tau_1$ and we observe that (23)–(29) hold. The existence of v as in (53) satisfying $O(a,\theta_1 r,v)\leqq\xi/2$ now follows from the Main Lemma.

Lemma. *Suppose*

$$\delta(u,\theta)\leqq(\zeta_{21}\exp(\exp\kappa))^{-1}; \tag{55}$$

$$\xi\leqq(\zeta_{21}\exp(\exp\kappa))^{-1}. \tag{56}$$

Then there is $v\in\mathscr{U}$ *such that*

$$\delta(u,v)^2\leqq\zeta_{18}\kappa^5\xi \tag{57}$$

and

$$O(a,\theta_1 r,v)\leqq\max\{\xi/2,\zeta_{22}(\exp\kappa)\Delta r^2\}. \tag{58}$$

Proof. This is a straightforward consequence of the previous Lemma.

3.10. **Lemma.** *Suppose*

$$a\in X,\quad 0<r<\infty\quad\text{and}\quad|a|+r<2R; \tag{59}$$

$$O(a,r,0)=(\zeta_{23}\exp(\exp\kappa))^{-1}; \tag{60}$$

$$\Delta r^2\leqq(\zeta_{23}\exp(\exp\kappa))^{-1}. \tag{61}$$

Then there is a sequence

$$v_0,v_1,v_2,\ldots\quad\text{in }\mathscr{U}$$

such that

$$\delta(v_j,v_{j-1})^2\leqq\zeta_{18}\kappa^5 O(a,(\theta_1/2)^j r,v_{j-1}); \tag{62}$$

and

$$O(a,(\theta_1/2)^j r,v_j)$$
$$\leqq\max\{O(a,(\theta_1/2)^{j-1}r,v_{j-1})/2,\zeta_{22}(\exp\kappa)\Delta(\theta_1/2)^{2(j-1)}r^2\} \tag{63}$$

whenever $j=1,2,\ldots$; *here* θ_1 *is as in* 3.9 (49).

Furthermore, if $a\in W$ *then*

$$|\nabla f_i(a)|\leqq\zeta_{24}(\exp\kappa)\max\{O(a,r,\theta),\Delta r^2\}. \tag{64}$$

Proof. The sequence $v_0,v_1,v_2,\ldots$ is constructed by letting $v_0=\theta$ and using the final Lemma in 3.9 inductively. The estimate (62) allows estimation of $\delta(\theta,\lim_{j\to\infty}v_j)$ so that (64) holds.

Remark. Other estimation on f_i, $i\in\mathbf{K}$, may be made; see [11, Chap. 4].

4. The Main Theorem

Recall our permanent assumptions of 1.2 and 1.4.

4.1. **The Main Theorem.** *Suppose*

(1) $0<R<\infty$, $X=\mathbf{U}(0,3R)$, *Y is a bounded interval in* $\mathbf{R}$ *and* $Z=X\times Y$;

(2) *M is closed relative to* $X\times\mathbf{R}$;

(3) $0\in M$;

(4) $\mu=R^{-n}\|M\|(\mathbf{C}(0,3R))$;

(5) $\Delta R^2\leqq(\zeta_{25}\exp(\exp\mu))^{-1}$

(6) $\rho=R^{-n}\int 1-\nu\cdot e_{n+1}^2\,d\|M\|\leqq(\zeta_{25}\exp(\exp\mu))^{-1}$

Then there are a positive integer κ *and for each* $i\in\{1,\ldots,\kappa\}$ *a function* $f_i\colon \mathbf{U}(0,R)\to Y$ *of class* 2 *such that*

$$M\cap\mathbf{C}(0,R)=\bigcup_{i=1}^{\kappa}\{(x,y)\in\mathbf{C}(0,R)\colon y=f_i(x)\}. \tag{7}$$

$$|\nabla f_i|\leqq\zeta_{26}(\exp\mu)\max\{\rho,\Delta R^2\}\quad \text{for } i\in\mathbf{K}. \tag{8}$$

Proof. Let X, β, B, N be as in 3.1 and let

$$\varepsilon_1=(2R)^{2-n}\int_{\mathbf{C}(0,2R)}\|Dv\|^2\,d\|M\|,$$

$$\varepsilon_2=(2R)^{-n}\|M\|(\{z\in M\cap\mathbf{C}(0,2R)\colon\beta(z)\geqq\sqrt{2}/4\}),$$

$$\varepsilon=\varepsilon_1+\varepsilon_2,$$

$$\varepsilon_3=\Delta R^2.$$

According to 2.4,

$$\begin{aligned}(2R)^{2-n}\varepsilon_1&\leqq\zeta_3\int_{\mathbf{C}(0,3R)}R^{-2}\beta^2+\Delta\,d\|M\|\\&\leqq\zeta_3(R^{-2+n}\rho+\varepsilon_3R^{-2+n}\mu);\end{aligned}$$

evidently,

$$(2R)^n\varepsilon_2\leqq 8R^n\rho.$$

Consequently,

$$\varepsilon=\varepsilon_1+\varepsilon_2\leqq 2^n\zeta_3(\rho+\mu\varepsilon_3)+2^{n+2}\rho. \tag{9}$$

Since $\Delta R^2\leqq 1$ by (5) we have from (3) and 2.2(9) that

$$\zeta_2\mu\geqq 1. \tag{10}$$

Thus (5), (6), (9), (10) imply $\mathbf{c}_{11}\varepsilon\leqq 1$ and $\zeta_2(2\sqrt{2}\,\mathbf{c}_4+1)\varepsilon<1$ so that we may use 3.2 to secure a connected component C of $\mathbf{U}(0,2R)$ such that

$$0<\kappa\leqq 2\alpha(n)^{-1}\mu \tag{11}$$

where κ is the number of connected components of $\mathbf{p}^{-1}(C)\cap N$ and

$$\|M\|[\mathbf{C}(0,2R)\sim(\mathbf{p}^{-1}(C)\cap N)] \leqq \mathbf{c}_{12}[1+\varepsilon^{1/(n-1)}\mu]\,\varepsilon(2R)^n. \tag{12}$$

We let W be the connected component of $\mathbf{U}(0,2R)\sim B$ containing C and we let $G=\mathbf{p}^{-1}(W)\cap N$; we adopt the terminology of 3.4. We have

$$P(a,R)\leqq R^{-n}\|M\|[\mathbf{C}(0,2R)\sim(\mathbf{p}^{-1}(C)\cap N)],$$
$$D(a,R,0)\leqq\rho.$$

whenever $a\in\mathbf{U}(0,R)$. Thus by (5), (6), (9), (11), (12) we have

$$O(a,R,0)\leqq(\exp(\exp c\,\zeta_{25}2^{-1}\alpha(n)\kappa))^{-1} \tag{13}$$

whenever $a\in\mathbf{U}(0,R)$ where c is a number depending only on n, ζ_3 and ζ_4. The Main Theorem now follows from 3.10.

4.2. *Remark.* Suppose R, X, Y, Z are as in 4.1. To prove the assertion made in the introduction it will suffice to show that

$$R^{-n}\int 1-\nu\cdot e_{n+1}^2\,d\|M\|$$

becomes small as

$$\Delta R^2 \quad\text{and}\quad R^{-n-1}\int|\mathbf{q}|\,d\|M\|$$

become small and μ stays bounded.

In order to show this is indeed the case we consider sequences

$$\Phi_\gamma, M_\gamma, \nu_\gamma, i_\gamma, \qquad \gamma=1,2,3,\ldots,$$

as in 1.2 and 1.4 satisfying

(1) $\sup\{\|M_\gamma\|(Z):\gamma=1,2,3,\ldots\}<\infty$;

(2) $\sup\{\|D^j\Phi_\gamma(z,\alpha)\|:\gamma=1,2,3,\ldots,j\in\{0,1,2\}\text{ and }(z,\alpha)\in Z\times\mathbf{S}^n\}<\infty$;

(3) $\lim\limits_{\gamma\to\infty}\sup\{|\langle(u,0),D\Phi_\gamma(z,\alpha)\rangle|:\gamma=1,2,3,\ldots\text{ and }(u,z,\alpha)\in\mathbf{S}^n\times Z\times\mathbf{S}^n\}=0$;

(4) $\lim\limits_{\gamma\to\infty}\int|\mathbf{q}|\,d\|M_\gamma\|=0$;

(5) $\lim\limits_{\gamma\to\infty}\int(\operatorname{trace}Dg)\,\Phi_\gamma(i_\gamma,\nu_\gamma)-\langle(g,\langle\nu_\gamma,Dg^*\rangle),D\Phi_\gamma(i_\gamma,\nu_\gamma)\rangle\,d\|M_\gamma\|=0$

whenever $g:Z\to\mathbf{R}^{n+1}$ is continuously differentiable with compact support.

We will show that

$$\lim_{\gamma\to\infty}\int 1-\nu_\gamma\cdot e_{n+1}^2\,d\|M_\gamma\|=0. \tag{6}$$

To prove (6), for each $\gamma=1,2,3,\ldots$ we let V_γ be the Radon measure on $Z\times\mathbf{S}^n$ such that

$$\int f\,dV_\gamma=\int f(z,\nu_\gamma(z))\,d\|M_\gamma\|\,z$$

whenever $f: Z \times \mathbf{S}^n \to \mathbf{R}$ is continuous with compact support and we let V be a Radon measure on $Z \times \mathbf{S}^n$ which is the weak limit of a subsequence of the sequence $V_1, V_2, V_3, \ldots$. In order to prove (6) it will suffice to show that

$$\int 1 - \alpha \cdot e_{n+1}^2 \, dV(z, \alpha) = 0. \tag{7}$$

Note that (4) implies

$$\int |z \cdot e_{n+1}| \, dV(z, \alpha) = 0. \tag{8}$$

Owing to (2) and (3) we may obtain a function $\psi: \mathbf{R}^{n+1} \to \mathbf{R}$ such that

(9) $\psi(t\alpha) = t\psi(\alpha)$ whenever $0 \leqq t < \infty$ and $\alpha \in \mathbf{R}^{n+1}$;

(10) $\psi | \mathbf{R}^{n+1} \sim \{\mathbf{0}\}$ is continuously differentiable;

(11) $\liminf_{\gamma \to \infty} \sup \{|\Phi_\gamma(z, \alpha) - \psi(\alpha)| : (z, \alpha) \in Z \times \mathbf{S}^n\} = 0$;

(12) $\liminf_{\gamma \to \infty} \sup \{\| \langle (u, \beta), D\Phi_\gamma(z, \alpha) \rangle - \langle \beta, D\psi(\alpha) \rangle \| :$

$$(u, \beta) \in \mathbf{S}^n \times \mathbf{S}^n \quad \text{and} \quad (z, \alpha) \in Z \times S^n\} = 0.$$

As a consequence of (5) and (12) we see that

$$\int \operatorname{trace} Dg(z) \psi(\alpha) - \langle \langle \alpha, Dg(z) \rangle^*, D\psi(\alpha) \rangle \, dV(z, \alpha) = 0 \tag{13}$$

whenever $g: Z \; R^{n+1}$ is continuously differentiable with compact support.

Suppose $\phi: Z \to \mathbf{R}$ is continuously differentiable with compact support. Suppose $v \in \mathbf{R}^{n+1}$ and $v \cdot e_{n+1} = 0$ and let $g(z) = \phi(z)(z \cdot e_{n+1}) v$ for $z \in Z$. We infer from (8) and (13) that

$$\int \phi(z) v \cdot \alpha \langle e_{n+1}, D\psi(\alpha) \rangle \, dV(z, \alpha) = 0.$$

Owing to the arbitrariness of ϕ we deduce that

$$\textit{either } \alpha \in \{e_{n+1}, -e_{n+1}\} \textit{ or } \langle e_{n+1}, D\psi(\alpha) \rangle = 0 \qquad \text{for } V \text{ almost all } (z, \alpha). \tag{14}$$

Now let $g(z) = \phi(z)(z \cdot e_{n+1}) e_{n+1}$ for $z \in Z$. We infer from (8) and (13) that

$$\int \phi(z) [\psi(\alpha) - \alpha \cdot e_{n+1} \langle e_{n+1}, D\psi(\alpha) \rangle] \, dV(z, \alpha) = 0.$$

Owing to the arbitrariness of ϕ we infer that

$$\psi(\alpha) = \alpha \cdot e_{n+1} \langle e_{n+1}, D\psi(\alpha) \rangle \quad \text{for } V \text{ almost all } (z, \alpha). \tag{15}$$

The conjunction of (14) and (15) gives (7).

References

1. Almgren, F.J., Jr.: Existence and regularity almost everywhere of solutions to elliptic variational problems among surfaces of varying topological type and singularity structure. Am. of Math. **87**, 321-391 (1968)

2. Allard, W.K.: On the first variation of a varifold. Ann. of Math. **95**, 417–491 (1972)
3. Allard, W.K.: A characterization of the area integrand. Symposia Mathematica Vol. **XIV**, 429–444 (1974)
4. Allard, W.K.: On the first and second variation of a parametric integral. Am. J. Math. (in press) (1983)
5. Bliss, G.A.: Calculus of variations. (Carus mathematical monographs, 1) La Salle, Illinois: Open Court Publishing Co., 1925
6. De Giorgi, E.: Frontiere orientate di misura minima. Seminario Mat. Scuola Norm. Sup. Pisa, 1961
7. Federer, H.: Geometric measure theory. Berlin-Heidelberg-New York: Springer 1969
8. Morrey, C.B., Jr.: Multiple integrals in the calculus of variations. Berlin-Heidelberg-New York: Springer 1966
9. Michael, J.H., Simon, L.: Sobolev and mean value inequalities on generalized submanifolds of $\mathbf{R}^n$. Comm. on Pure and Appl. Math. Vol. XXVI, 361–379 (1973)
10. Pitts, J.: Existence and regularity of minimal surfaces on Riemannian manifolds. preprint, 1979
11. Schoen, R., Simon, L.: Regularity of stable minimal hypersurfaces. Comm. on Pure and Appl. Math., Vol. XXXIV, 741–797 (1981)

Oblatum 21-X-1982

Invent. math. 73, 333-336 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

On Projective Varieties of Codimension 2

Z. Ran*

Mathematics Department, Brandeis University, Waltham, MA 02254/USA

One of the important developments of the early 1970's in algebraic geometry was the discovery by Barth [1], subsequently enlarged on by many others, that submanifolds X of projective space, whose codimension is small compared with their dimension, "look" homologically like complete intersections. This focused attention on the fact that no such X were known other than complete intersections, and led Hartshorne [4] to conjecture that X must in fact be a complete intersection if $2\operatorname{codim}(X)<\dim(X)$. Hartshorne and, independently, Barth and Van de Ven [3] proved that if the degree (hence also the codimension) of X are much smaller than its dimension, then X is a complete intersection. The purpose of this paper is to contribute to the program of proving that projective varieties of small codimension are complete intersections, by proving the following.

Theorem. *Let X be a locally complete intersection subvariety of codimension 2 in $\mathbb{P}^{m+2}$ over an algebraically closed field of arbitrary characteristic. Let N denote the normal bundle of X and d its degree. Assume that $\Lambda^2 N \cong \mathcal{O}_X(v)$, $v \in \mathbb{Z}$ and moreover that either*

(i) $v \geqq \frac{d}{m}+m$, or

(ii) $d \leqq m$.

Then X is a complete intersection.

As is well-known ([1], [6]), our initial hypothesis is satisfied as soon as $m \geqq 4$, at least if char $=0$ and X is smooth. As the dualizing sheaf $\omega_X \simeq \mathcal{O}_X(v-m-3)$, hypothesis (i) means that ω_X is "large" relative to d and m. Note that in general $v \leqq d/2+2$ whenever X is nondegenerate (this follows from Castelnuovo's bound on the genus of a curve in $\mathbb{P}^3$). To the best of my knowledge the only published result on this question, due to Barth and Van de Ven, [2], is that X is a complete intersection whenever $d \leqq \frac{m+7}{4}$.

* Partially supported by the NSF

Our method of proof is inspired by an often-quoted paper [8] of Severi, in addition to the cited work of Barth and Van de Ven, and is based on the following *Obvious Fact*: if X is contained in a hypersurface Y of degree k, then so too must be any $(k+1)$-secant line of X, hence no such line can pass through a generic point P of projective space. For X a surface in $\mathbb{P}^4$ and $k=2$, Severi (op.cit.) shows the converse. Here we proceed by first proving, for X as in the Theorem, the existence of $(k+1)$-secants through P for $k \leqq m$, unless X looks *numerically* like a complete intersection. On the other hand a result of Maruyama shows that X must be contained in a hypersurface of degree $\leqq m$. Putting the two together, we then show that X must be a complete intersection.

To begin the proof proper, first some preliminaries. As is well-known [4], there is a vector bundle E on $\mathbb{P}^{m+2}$ and an exact Koszul sequence.

$$0 \to \mathcal{O} \to E \to \mathcal{I}_X(v) \to 0. \tag{1}$$

Note that $v=c_1(E)$, $d=c_2(E)$. Put $e(t)=c_2(E(-t))=d-t(v-t)$. Let G denote the Grassmannian $G(1,\mathbb{P}^{m+2})$, $q: I \to G$ the tautological $\mathbb{P}^1$-bundle, $p: I \to \mathbb{P}^{m+2}$ the natural map. Now fix a generic point $P \in \mathbb{P}^{m+2}$ and put

$$\Sigma_k = \Sigma_{k,P} = \{L \in G: P \in L \text{ and length } (L \cap X) \geqq k\},$$

the set of k-secant lines through P. For existence of such, we have the following result, analogous to a result of Barth and Van de Ven [3] about existence of "jumping lines".

Proposition. *If* $e(0)\,e(1) \ldots e(k) \neq 0$ *and* $k \leqq m$, *then* $\Sigma_{k+1} \neq \emptyset$.

Proof.[1] By induction it clearly suffices to show that if $C \subset \Sigma_k$ is an irreducible curve with $C \cap \Sigma_{k+1} = \emptyset$ and $\min\{\text{length}(L \cap X): L \in C\} = k$, then $e(k)=0$. Let $\tilde{C} \to C$ be a desingularization, $\pi: S \to \tilde{C}$ the $\mathbb{P}^1$-bundle induced by q, and $\tilde{p}$: $S \to \mathbb{P}^{m+2}$ the natural map. Put $D=\tilde{p}^{-1}(X)$ (as a scheme). By assumption, the fibres of D over $\tilde{C}$ all have length k, i.e. D is flat over $\tilde{C}$ and hence Cohen-Macaulay, i.e. a divisor on S. Let F denote a fibre of π, Z the section of π contracted by $\tilde{p}$, and $H=\tilde{p}^{-1}$ (hyperplane). We clearly have a numerical equivalence

$$D \sim kH + lF$$

for some l. But since $D \cdot Z = H \cdot Z = 0$, it follows that $l=0$, i.e. $D \sim kH$. As the map $\mathcal{O} \to \tilde{p}^*(E)$ vanishes precisely on D, we have an exact sequence

$$0 \to \mathcal{O}(D) \to \tilde{p}^*(E) \to Q \to 0 \tag{2}$$

for some line-bundle Q. Computing Chern classes from (2) yields $c_1(Q) \sim (v-k)H$, and $c_2(\tilde{p}^*(E)) \sim k(v-k).\, H^2$, i.e. $d=k(v-k)$ or $e(k)=0$.

The Obvious Fact mentioned above now yields

---

[1] The methods of [7] also yield a proof of this result, showing moreover that if Σ_{k+1} has dimension $m-k$ then its degree is $e(0) \ldots e(k)/(k+1)!$, counting multiplicities.

Corollary. *If* $e(0)\dots e(k)\neq 0$ *and* $k\leqq m$, *then* X *is not contained in any hypersurface of degree* k.

We now complete the proof of the Theorem. Assume first, instead of (i), the weaker hypothesis

(i)′ $$v\geqq\frac{d}{\alpha}+\alpha \quad \text{for some } \alpha\in(0,m].$$

Note that (i)′ means $e(\alpha)\leqq 0$; in particular, e has nonnegative discriminant, i.e. $c_1^2(E)\geqq 4c_2(E)$. By a theorem of Maruyama ([5], 3.3; see also 9.7), it follows that the vector bundle E is unstable. This means that the smallest integer k such that

$$h^0(E(k-v))\neq 0$$

satisfies

$$k\leqq v/2.$$

As $h^0(E(k-v-1))=0$, a section of $E(k-v)$ must vanish only in codimension 2, whence

$$e(k)=e(v-k)=c_2(E(k-v))\geqq 0.$$

As $e(t)$ is decreasing for $t\leqq v/2$, it follows that $k\leqq\alpha$, hence $k\leqq m$.

Now tensoring (1) with $\mathcal{O}(k-v)$ and using $h^0(E(k-v))\neq 0$, we get a hypersurface of degree k containing X, whence by the Corollary, an integer $i\leqq k$ such that

$$e(i)=0.$$

$e(t)$ being decreasing for $t\leqq k$, we have $i=k$, i.e.

$$c_2(E(k-v))=0.$$

Thus a section of $E(k-v)$ vanishes nowhere, hence a Koszul sequence

$$0\to\mathcal{O}\to E(k-v)\to\mathcal{O}(2k-v)\to 0$$

shows that E is decomposable, i.e. X is a complete intersection.

To complete the proof it now suffices to prove X is a complete intersection if $d\leqq m$. But note that any $(d+1)$-secant line of X is contained in X, whence $\Sigma_{d+1}=\emptyset$. So by the Proposition, $e(k)=0$ for some $k\leqq m$. In particular, hypothesis (i)′ is satisfied, so we are done.

References

1. Barth, W.: Transplanting cohomology classes in complex-projective space. Amer. J. Math. **92**, 951-967 (1970)
2. Barth, W.: Submanifolds of low codimension in projective space. Proc. Int. Cong. Math., Vancouver, 1974

3. Barth, W., Van de Ven, A.: A decomposability criterion for algebraic 2-bundles in projective spaces. Invent. Math. **25**, 91–106 (1974)
4. Hartshorne, R.: Varieties of small codimension in projective space. Bull. AMS **80**, 1017–1032 (1974)
5. Hartshorne, R.: Stable reflexive sheaves. Math. Ann. **254**, 121–176 (1980)
6. Larsen, M.E.: On the topology of complex projective manifolds. Invent. Math. **19**, 251–260 (1973)
7. Ran, Z.: A united-set formula. Bull. (N.S.) AMS **8**, 329–332 (1983)
8. Severi, F.: Intorno ai punti doppi impropri di una superficie generale dello spazio a quattro dimensioni, e a suo punti tripli apparenti. Rend. Circ. Mat Palermo **15**, 33–51 (1901)

Oblatum 20-I-1983

Notes Added in Proof

1) the Proposition. An analogous method applies more generally to proof of the existence of multisecants under a suitable numerical hypothesis. However, even with no such hypotheses at all, one can still prove: if $X^m \subseteq \mathbb{P}^{m+r}$ and $(k-2)r \leqq m-1$, then X has at least $\infty^{2m-(k-2)(r-1)}$ k-secant lines, where each line contained in X counts ∞^k times. Details will be given elsewhere.

2) the Theorem. Ballico and Chiantini have now proved that X is a complete intersection whenever $v \leqq m$. Together with part (i) this implies that X is a complete intersection whenever $d \leqq m^2$.

Invent. math. 73, 337–347 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Arakelov's Theorem for Abelian Varieties

G. Faltings

Fachbereich Mathematik, Universität-Gesamthochschule Wuppertal, Gaußstr. 20, D-5600 Wuppertal 1, Federal Republic of Germany

§ 1. Introduction

We start with an algebraic curve B over an algebraically closed field k of characteristic zero, and let $S \subseteq B$ be a finite set of points. In [1] Arakelov has shown that there are only finitely many families of algebraic curves of genus $g > 1$ on B, with good reduction outside S, except for isotrivial families (isotrivial = becomes constant on a finite cover of B).

We want to consider the same question for principally polarized abelian varieties. Here the answer is more complicated:

There is a condition (*) such that the number of abelian varieties fulfilling (*) is finite, while any variety not fulfilling it can be deformed. The condition (*) says essentially that all endomorphisms of the cohomology of the abelian variety are endomorphisms of the abelian variety itself.

The method of proof consists of a combination of Arakelov's methods and Deligne's description of abelian varieties via Hodge-structures. In the next chapter we recall the necessary prerequisites, and after that we prove the theorem in two steps as in [1].

We first derive a boundedness-theorem and then we show that families fulfilling (*) cannot be deformed. From the form of the theorem it seems that it is difficult to take it over to characteristic $p > 0$ (see [7]). The author has learned about this subject from L. Szpiro, who kindly printed out to him that the following results are already known:

a) L. Moret-Bailly has proved a boundedness-theorem in any characteristic. The result is contained in his thesis. (It will appear in the proceedings of the "seminar on pencils of abelian varieties, Paris 1981/82", and in the Comptes-Rendues). Unfortunately this theorem is weaker in characteristic zero than ours.

b) L. Szpiro and L. Moret-Bailly have a very good theorem about boundedness for $S = \emptyset$ (any characteristic).

c) They have derived rigidity for relative dimension two. (To appear also in "seminar on pencils of abelian varieties"). In characteristic zero our results cover relative dimension up to three.

The results were found during a stay at the I.H.E.S. I have to thank P. Deligne for some help concerning the example of an abelian variety not fulfilling (*).

The referee has told me that the results of Zucker (Ann. of Math. **109**) should allow to treat more general Hodge-structures on $B-S$. As *I* am not an expert in this field *I* leave this to the reader. In any case *I* thank the referee for his suggestions.

§ 2. Notations

k always denotes an algebraically closed field of characteristic 0, and in the proofs we assume that $k=\mathbb{C}$, which we may do by the Lefschetz-principle. B denotes a connected complete smooth curve over k, and $S\subseteq B$ a finite set of points. We want to consider families

$$p: X \to B-S$$

of abelian varieties of relative dimension g over S. By [4] we know that giving such a family is the same as giving a polarizable Hodge-structure (or variation of Hodge-structure) of degree 1 on $B-S$, that is (except for the polarization) a locally constant sheaf $\mathbf{V}$ on $B-S$, locally isomorphic to $\mathbb{Z}^{2g}$, plus a subbundle

$$\mathscr{W}=\mathscr{W}_X\subseteq \mathbf{V}\otimes_{\mathbb{Z}}\mathcal{O}_B$$

of rank g, such that $\mathscr{W}$ and its complex conjugate span the fibre of $\mathbf{V}$ in each point. We furthermore assume that it has a principal polarization, which means a skew-symmetric form

$$\langle\ ,\ \rangle: \mathbf{V}\times\mathbf{V}\to\mathbb{Z},$$

which identifies $\mathbf{V}$ with its own dual, vanishes on $\mathscr{W}\times\mathscr{W}$, and which has the additional property that for $w\in\mathscr{W}$ and $w\neq 0$

$$-\frac{1}{2\pi i}\langle w,\bar{w}\rangle>0.$$

Definition. (See [4], §4.4.) A family

$$p: X\to B-S$$

satisfies (*) if any anti-symmetric endomorphism A of

$$\mathbf{V}=R^1p_*\mathbb{Z}$$

defines an endomorphism of X (i.e., is of type $(0,0)$). A is called anti-symmetric if

$$\langle Au, v\rangle=-\langle u, Av\rangle.$$

(We may use étale or de Rham cohomology to formulate that over base fields $k\neq\mathbb{C}$.)

In [4], Proposition 4.4.11, there are given conditions which force X to fulfill $(*)$.

We furthermore need some information about moduli-spaces. We fix a number $n \geqq 3$ and denote by $\mathscr{A} = \mathscr{A}_{n,g}$ the quotient of Siegel's upper half-plane $\mathbb{H}_g$ under the congruence-subgroup group Γ_n of level n of $Sp(2g, \mathbb{Z})$. It is a fine moduli-space for principally polarized abelian varieties with a level-n-structure. As before we have on it a principally polarized Hodge-structure $\mathbf{V}$, together with a subbundle $\mathscr{W}$ of $\mathbf{V} \otimes_{\mathbb{Z}} \mathcal{O}$. $\mathscr{W}$ can also be described as the vector-bundle associated with the natural g-dimensional representation of $U(g, \mathbb{C})$, the maximal compact subgroup of $Sp(2g, \mathbb{R})$. We denote its maximal exterior power by ω.

$$\omega = \Lambda^g \mathscr{W}.$$

We know [3] that $\mathscr{A}$ can be imbedded as an open subvariety into a compact variety $\mathscr{A}^*$, such that ω extends to an ample line-bundle on $\mathscr{A}^*$. There further exists another compactification $\bar{\mathscr{A}}$ of $\mathscr{A}$ which dominates $\mathscr{A}^*$, such that $\bar{\mathscr{A}}$ is nonsingular and that

$$D_\infty = \bar{\mathscr{A}} - \mathscr{A}$$

is a divisor with normal crossings (see [2]). We furthermore know that $\mathscr{W}$ extends as a bundle to $\bar{\mathscr{A}}$, such that the hermitian metric defined on it by the polarization of $\mathbf{V}$ has only logarithmic singularities at D_∞ (see [5])*. It is also wellknown that the extension of the second symmetric power is isomorphic to the sheaf of differential forms with logarithmic poles at D_∞:

$$\mathbf{S}^2(\mathscr{W}) \cong \Omega^1_{\bar{\mathscr{A}}}[D_\infty].$$

The latter fact can be seen as follows:

There is such an isomorphism over $\mathscr{A}$, and we only have to show that $\Omega^1_{\bar{\mathscr{A}}}[D_\infty]$ is Mumford's extension of $\Omega^1_{\mathscr{A}}$. But this follows from the fact that locally the compactification looks like an embedding

$$(\mathbb{C}^x)^a \times \mathbb{C}^b \subseteq \mathbb{C}^a \times \mathbb{C}^b,$$

and that the extended bundle is generated by sections invariant under $(\mathbb{C}^x)^a$. (See the construction in [5], p. 257).

If we have an abelian variety

$$p: X \to B - S$$

as in the beginning, and if we assume that X has a level-n-structure, we obtain a mapping

$$\phi: B \to \bar{\mathscr{A}}$$

with

$$\phi^{-1}(D_\infty) \subseteq S.$$

The pullback $\phi^*(\mathscr{W})$ on B is an extension of our previous bundle $\mathscr{W}$ (which was defined on $B - S$). It can be described as follows:

---

* This means that for a matrix $H = (h_{pq})$ giving the metric the functions $|h_{pq}|$, $|\det(H)|^{-1}$ grow at most like powers of $\log|1/f|$, f a local equation for D_∞

It is known that X extends to a stable family

$$\bar{p}: \bar{X} \to B$$

a semiabelian varieties over B. If

$$s: B \to \bar{X}$$

denotes its zero-section, then

$$\mathscr{W} = s^*(\Omega^1_{\bar{X}/B}).$$

We can show this as follows:

We may enlarge s, replace B by a covering B', or replace X by a larger abelian variety in which it is a factor. We may also change the polarization (since two of them are comparable), and finally we may assume that X is the Jacobian of a family of curves Y over $B-S$, which extends to a semi-stable family $\bar{Y}$ over B. But then some easy and rather explicit calculations show that the canonical metric on $s^*(\Omega^1_{\bar{X}/B})$ has a only logarithmic singularities, and we are done.

The referee has pointed out to me that in the following chapter we need the equality of Mumford's extension of ω to $\bar{\mathscr{A}}$ and of the pullback of the ample line-bundle ω on $\mathscr{A}^*$. This can be derived from the calculations in [2], Chap. IV. § 1, by rewriting them for differential forms with logarithmic singularities (instead of ordinary differential forms): We get that for $l \gg 0$ the global sections of $\omega^{\otimes l}$ on $\mathscr{A}^*$ generate $\omega^{\otimes l}$ on $\bar{\mathscr{A}}$.

§ 3. Boundedness

We want to prove the following

Theorem 1. *Let $\mathscr{L}$ be a line-bundle on $\bar{\mathscr{A}}$. If (B, S) is a curve plus a finite set, there exists a constant c (depending on $\mathscr{L}$, B and S), such that*

$$\deg(\phi^*(\mathscr{L})) \leqq c$$

for any morphism

$$\phi: B \to \bar{\mathscr{A}}$$

with

$$\phi^{-1}(D_\infty) \subseteq S.$$

(ϕ corresponds to an abelian variety over $B-S$ with a level-n-structure).

Proof. A moment's thought shows us that it is sufficient to prove the theorem for $\mathscr{L} = \omega = \Lambda^g \mathscr{W}$ (since $\phi(B) \not\subseteq D_\infty$). We shall prove that for any stable abelian variety X over $B-S$

$$\deg(\mathscr{W}_{\bar{X}}) = \deg(\Lambda^g \mathscr{W}_{\bar{X}})$$

is bounded, where $\bar{X}$ is the Néron-model of X over B, and $\mathscr{W}_{\bar{X}}$ is the extension of $\mathscr{W}_X$ described at the end of the previous chapter.

We proceed as follows:

By the Riemann-Roch theorem it is sufficient to bound the dimension of

$$\Gamma(B, \mathscr{W}_{\bar{X}} \otimes \Omega^1_B[S])$$

($\Omega^1_B[S]$ = differentials with simple poles in S). Any global section of $\mathscr{W}_{\bar{X}} \otimes \Omega^1_B[S]$ defines a holomorphic 2-form on X and therefore a class in

$$H^2(X, \mathbb{C}).$$

We note that the dimension of $H^2(X, \mathbb{C})$ is bounded (use $E_2^{p,q} = H^p(B-S, R^q p_* \mathbb{C}) \Rightarrow H^{p+q}(X, \mathbb{C})$), and so we are done if we prove that the mapping

$$\Gamma(B, \mathscr{W}_{\bar{X}} \otimes \Omega^1_B[S]) \to H^2(X, \mathbb{C})$$

is injective.

By Deligne's theorem ([4], Cor. 3.2.13) this would be true if our 2-forms would have only logarithmic poles at infinity, for some compactification of X. We avoid the construction of such a compactification if we note that for the proof of our claim we may enlarge S, replace B by a finite covering, or replace X by a bigger abelian variety in which it is a factor. We thus reduce to the case that X is the Jacobian of a semi-stable curve Y over $B-S$, with a regular semi-stable model $\bar{Y}$ over B. We furthermore may assume that the mapping from $\bar{Y}$ to B has a section. This gives us a commutative diagram

$$\begin{array}{ccc} Y & \longrightarrow & X \\ & \searrow \quad \swarrow & \\ & B-S, & \end{array}$$

and it is sufficient to show that the mapping

$$\Gamma(B, \mathscr{W}_{\bar{X}} \otimes \Omega^1_B[S]) \to H^2(Y, \mathbb{C})$$

is an injection.

Now for semi-stable curves explicit calculations are rather easy, and they show that the elements of $\Gamma(B, \mathscr{W}_{\bar{X}} \otimes \Omega^1_B[S])$ are precisely the global 2-forms on $\bar{Y}$ with logarithmic poles in the preimage of S. By Deligne's theorem we are ready.

Remark. The proof gives the bound (which is not the best):

$$\deg(\mathscr{W}) \leqq g(3 \cdot \text{genus}(B) + \text{order}(S) + 1).$$

Corollary. *The scheme*

$$\mathbf{Hom}((B, S), (\bar{\mathscr{A}}, D_\infty))$$

which classifies the morphisms

$$\phi: B \to \bar{\mathscr{A}}$$

with

$$\phi^{-1}(D_\infty) \subseteq S$$

is of finite type over k.

§ 4. Deformations

Let

$$\phi: B \to \bar{\mathscr{A}}$$

be a morphism as before, with

$$\phi^{-1}(D_\infty) \subseteq S.$$

We want to determine the tangent-space at ϕ of

$$\mathbf{Hom}((B, S), (\bar{\mathscr{A}}, D_\infty)).$$

By the usual calculations it is isomorphic to

$$\mathrm{Hom}_B(\phi^*(\Omega^1_{\mathscr{A}}[D_\infty]), \mathscr{O}_B) = \mathrm{Hom}_B(S^2(\mathscr{W}_{\bar{X}}), \mathscr{O}_B).$$

We know by general arguments that the canonical metric on $\mathscr{W}_{\bar{X}}$ has positive curvature. It is known that this tends to prevent the existence of linear forms on such a bundle. There are some inequalities which have to become equalities, and we leave it to the reader to verify the following

Lemma. *Suppose $\mathscr{E}$ is a bundle on B which has a hermitian metric with logarithmic singularities and non-negative curvature. If*

$$\phi: \mathscr{E} \to \mathscr{O}_B$$

is a nontrivial morphism let

$$\mathscr{F} = \mathrm{Kern}(\phi).$$

Then $\mathscr{E}$ is isomorphic to the orthogonal direct sum

$$\mathscr{E} \cong \mathscr{F} \oplus \mathscr{O}_B,$$

and ϕ corresponds to the projection onto the second direct summand.

Retourning to our problem we see that any nonzero element s of

$$\mathrm{Hom}_B(\mathbf{S}^2(\mathscr{W}_{\bar{X}}), \mathscr{O}_B)$$

gives rise to a section

$$t \in \Gamma(B, \mathbf{S}^2(\mathscr{W}_{\bar{X}}))$$

of constant norm 1, and such that $\langle h, \bar{t} \rangle$ is holomorphic for any local section h of $\mathbf{S}^2 \mathscr{W}_{\bar{X}}$.

Here $\langle\ ,\ \rangle$ denotes the scalarproduct on

$$\mathbf{S}^2(\mathbf{V}) \otimes_{\mathbf{Z}} \mathscr{O}_B \qquad (\mathbf{V} = R^1 p_* \mathbf{Z})$$

derived from the Hodge-structure on $\mathbf{S}^2(\mathbf{V})$.

We denote by

$$\nabla: \mathbf{S}^2(\mathbf{V}) \otimes_{\mathbf{Z}} \mathscr{O}_B \to \mathbf{S}^2(\mathbf{V}) \otimes_{\mathbf{Z}} \Omega^1_B$$

the connection (defined on $B - S$).

Then

$$\langle \nabla(t), \nabla(\bar{t}) \rangle = \partial\bar{\partial}\langle t, \bar{t} \rangle = 0$$

and for h a local section of $\mathbf{S}^2(\mathscr{W})$ we have

$$\langle \nabla(t), \bar{h} \rangle = \partial\langle t, \bar{h} \rangle = \overline{\bar{\partial}\langle h, \bar{t} \rangle} = 0.$$

By the fundamental properties of ∇ we known that $\nabla(t)$ has components of bidegree $(2,0)$ and $(1,1)$. The second equality says that it is of bidegree $(1,1)$, and by the first identity it vanishes. Thus t is parallel.

Finally some easy calculations show that there exists a (parallel) endomorphism A of $\mathbf{V}\otimes_{\mathbf{Z}}\mathbf{C}$, such that

i) A is pure of type $(-1,1)$
ii) A is anti-symmetric for $\langle\ ,\ \rangle$,

such that

$$s(h_1 \otimes h_2) = \langle h_1, A(h_2) \rangle$$

for local sections h_1, h_2 of $\mathscr{W}_{\bar{X}}$.

Conversely it is clear that any such A define an element s (using for example the results of Schmid [6] to treat the behaviour at S). We have thus nearly proved the

Theorem 2.

$$\mathbf{Hom}((B, S), (\bar{\mathscr{A}}, D_\infty))$$

is smooth.

Its tangent-space at a point corresponding to an abelian variety X over $B-S$ is isomorphic to the space of $\langle\ ,\ \rangle$-anti-symmetric endomorphisms of

$$\mathbf{V}\otimes_{\mathbf{Z}}\mathbf{C} = R^1 p(\mathbf{C})$$

of pure type $(-1,1)$.

Proof. It remains to show that any A as above defines deformations. A as well as its complex conjugate $\bar{A}$ are ∇-parallel, so that the eigenvalues of $\bar{A}A$ are constant on B. They are easily seen to be (positive) real numbers, and we shall define a deformation corresponding to A provided these eigenvalues are smaller than $+1$.

If these conditions are fulfilled we define a new Hodge-structure supported by $\mathbf{V}$, by choosing

$$\exp(A)(\mathscr{W}) = (\mathrm{Id} + A)(\mathscr{W}) \subseteq \mathbf{V}\otimes_{\mathbf{Z}}\mathscr{O}_B$$

as space of type $(1,0)$:

We only have to show that for h a local nonzero section of $\mathscr{W}$

$$\langle (\mathrm{Id}+A)(h), \overline{(\mathrm{Id}+A)(h)} \rangle = \langle (\mathrm{Id} - \bar{A}A)(h), \bar{h} \rangle$$

is positive.

This follows from our assumptions because the selfadjoint transformation

$$\mathrm{Id} - \bar{A}A$$

of $\mathscr{W}$ has positive eigenvalues.

We thus have defined a deformation of the Hodge-structure (and therefore of X) parametrized by a neighbourhood of the identity in our space of A's. This proves theorem 2.

Corollary. *There exist only finitely many families of principally polarized abelian varieties of dimension g over B, with good reduction outside S and satisfying* (∗). *If there is one such variety which does not fulfill* (∗), *there are infinitely many.*

Proof. We know that there is a finite Galois-covering B' of B such that any such family has a level n-structure after base-change to B'. We let S' be the preimage of S in B', and π the Galois-group of the covering B'/B. For any representation

$$\rho: \pi \to Sp(2g, \mathbb{Z}/n\mathbb{Z}) \subseteq \mathrm{Aut}(\bar{\mathscr{A}})$$

we obtain an action of π on

$$\mathbf{Hom}((B', S'), (\bar{\mathscr{A}}, D_\infty)),$$

and the set of abelian varieties over B injects into the disjoint union of the fixed-point sets of π, corresponding to all representations ρ.

From Theorem 2 it can be derived that these fixed-point-sets are manifolds at the points given by families

$$p: X \to B - S$$

over $B-S$, and that their tangent-spaces there are the symmetric endomorphisms (over B) of

$$R^1 p_*(\mathbb{C})$$

of type $(-1, 1)$.

They vanish precisely if X fulfills (∗).

Therefore the points corresponding to such X ly in the zero-dimensional components of the fixed-point-sets. As this are varieties of finite type over k (Theorem 1) their number is finite.

Conversely if X does not fulfill (∗), we have constructed deformations of X in the proof of Theorem 2.

Remark. The proof gives that only finitely many pairs $(\mathbf{V}, \langle\ ,\ \rangle)$ occur as cohomology of principally polarized abelian varieties over $B-S$. By an hyperplane section argument the same is true for any normal algebraic variety U.

§ 5. An Example

We want to show that there exist indeed nontrivial examples of abelian varieties which do not satisfy (∗) (constant ones of course fall into this category). We proceed by following the arguments in [4], §4.4. We restrict ourselves to the simplest case, but the method can be generalized.

Let K be a real quadratic number-field, D a quaternion-algebra over K which splits at one of the infinite places of K but not at the other. D has a canonical

involution denoted by

$$a \mapsto a^*,$$

such that for $a \in D$

$$N(a) = a\,a^*$$

and

$$T(a) = a + a^*$$

are in K.

We denote by $V_{\mathbb{Q}}$ the rational vectorspace (of dimension 8)

$$V_{\mathbb{Q}} = D.$$

D acts by multiplication on the left and on the right on V. For any element s of D with

$$s^* = -s$$

we obtain a new involution on D, denoted by

$$a^t = s\,a^*\,s^{-1}.$$

We further obtain a skew-symmetric form on $V_{\mathbb{Q}}$ by

$$\langle a, b \rangle = \mathrm{tr}_{K/\mathbb{Q}}(T(a\,s\,b^*)).$$

Then $\langle a\,c, b \rangle = \langle a, b\,c^t \rangle$, and $\langle\ ,\ \rangle$ is respected by the algebraic group G with

$$G(\mathbb{Q}) = \{a \in D \mid N(a) = 1\},$$

which acts by left-multiplication.

By assumption

$$D_{\mathbb{R}} = D \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbf{Q} \times M_2(\mathbb{R}) \qquad (\mathbf{Q} = \text{quaternions}),$$

$$G(\mathbb{R}) \cong SU(2, \mathbb{C}) \times SL(2, \mathbb{R})$$

and

$$V_{\mathbb{R}} = V \otimes_{\mathbb{Q}} \mathbb{R} \cong \mathbb{C}^2 \oplus (\mathbb{R}^2 \otimes \mathbb{R}^2).$$

The decomposition of $V_{\mathbb{R}}$ is written in such a way that $SU(2, \mathbb{C})$ acts on the first summand by its natural representation, $SL(2, \mathbb{R})$ on the first factor of the tensor-product, and $M_2(\mathbb{R})$ by right-multiplication on the second factor.

The bilinear form $\langle\ ,\ \rangle$ decomposes accordingly:

$$\langle\ ,\ \rangle = \langle\ ,\ \rangle_1 \oplus \langle\ ,\ \rangle_2 \otimes \langle\ ,\ \rangle_3.$$

Here $\langle\ ,\ \rangle_1$ (for suitable coordinates) is a real multiple of the skew-form

$$(u_1, u_2) \times (v_1, v_2) \mapsto \mathrm{Im}\,(\bar{u}_1\,v_1 + \bar{u}_2\,v_2)$$

on $\mathbb{C}^2$, $\langle\ ,\ \rangle_2$ the skew form

$$(u_1, u_2) \times (v_1, v_2) \mapsto u_1\,v_2 - u_2\,v_1$$

on $\mathbb{R}^2$, and $\langle\ ,\ \rangle_3$ a symmetric form on $\mathbb{R}^2$. We now choose s in such a way that $\langle\ ,\ \rangle_1$ is a positive multiple of the standard form listed above, and that $\langle\ ,\ \rangle_3$ is positive definite.

After extending scalars to the complex-numbers we obtain

$$V_{\mathbb{C}} = V_{\mathbb{Q}} \otimes_{\mathbb{Q}} \mathbb{C} = (\mathbb{C} \oplus \bar{\mathbb{C}}^2) \oplus (\mathbb{C}^2 \otimes \mathbb{R}^2)$$

where the first direct sum has been split such that

$$\bar{\mathbb{C}}^2 = \mathrm{Kern}\,(\mathbb{C}^2 \otimes_{\mathbb{R}} \mathbb{C} \mapsto \mathbb{C}^2).$$

For $0 \neq w \in \mathbb{C}^2$ we have $-\frac{1}{2\pi i}\langle w, \bar{w}\rangle > 0$, and thus we see that for a suitable one-dimensional complex subspace $L \subseteq \mathbb{C}^2$ we can define a polarized Hodge-structure on V by

$$V_{\mathbb{C}}^{1,0} = (\mathbb{C}^2 \oplus 0) \oplus (L \otimes \mathbb{R}^2).$$

The stabilizer in $G(\mathbb{R})$ of such a Hodge-structure is a maximal compact subgroup

$$M \cong SU(2, \mathbb{C}) \times U(1, \mathbb{C}),$$

and $G(\mathbb{R})$ acts transitively on these Hodge-structures which are parametrized by the upper half-plane

$$\mathbb{H} \cong G(\mathbb{R})/M.$$

Note that any $a \in D$ defines an endomorphism of $V_{\mathbb{Q}}$, but that not all such endomorphisms respect the Hodge-structure.

Otherwise the subspaces

$$\mathbb{C}^2 \oplus 0$$

and

$$0 \oplus \bar{\mathbb{C}}^2$$

of

$$\mathbb{C}^2 \oplus \bar{\mathbb{C}}^2$$

would be fixed by $SU(2, \mathbb{C})$ and $\mathbf{Q} \otimes_{\mathbb{R}} \mathbb{C}$, which cannot happen by a simple dimension-count. Unfortunately the antisymmetric a's still respect the Hodge-structure.

We thus take the direct sum of two copies of $V_{\mathbb{Q}}$: Here the algebra

$$A = M_2(D)$$

acts, with involution

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^* = \begin{pmatrix} a^* & c^* \\ b^* & d^* \end{pmatrix}.$$

A is generated by its antisymmetric elements, so some of them will not respect the Hodge-structure.

If we choose a lattice

$$V_{\mathbf{Z}} \subseteq V_{\mathbb{Q}}$$

such that $\langle\ ,\ \rangle$ takes integral values on $V_{\mathbf{Z}}$ and has determinant 1, we obtain a family of 8-dimensional abelian varieties over $\mathbb{H}$ which does not satisfy (*). If $\Gamma \subset G(\mathbb{Q})$ is a torsion-free arithmetic subgroup which stabilizes $V_{\mathbf{Z}}$ we can take

quotients and we obtain such a family over the compact Riemann surface

$$B = \Gamma \backslash \mathbb{H} = \Gamma \backslash G(\mathbb{R})/M.$$

As Γ is Zariski-dense in G it can be easily seen that this Hodge-structure does not have a fixed part on any finite covering of B.

References

1. Arakelov, S.: Families of curves with fixed degeneracy. Izv. Akad. Nauk. **35**, 1269–1293 (1971)
2. Ash, A., Mumford, D., Rapoport, M., Tai, Y.: Smooth compactification of Locally Symmetric Varieties. Math. Sci. Press, Brooklin (1975)
3. Baily, W.L., Borel, A.: Compactification of arithmetic quotients of bounded symmetric domains. Ann. of Math. **84**, 442–528 (1966)
4. Deligne, P.: Théorie de Hodge II. Publ. Math. **40**, 5–58 (1971)
5. Mumford, D.: Hirzebruch's proportionality theorem in the non-compact case. Invent. math. **42**, 239–272 (1977)
6. Schmid, W.: Variation of hodge structure: The singularities of the period mapping. Invent. math. **22**, 211–319 (1973)
7. Szpiro, L.: Sur le theorème de rigidité d'Arakelov et Parsin. Astérisque **64**, 169–202 (1979)

Oblatum 22-XII-1982 & 28-III-1983

Invent. math. 73, 349-366 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Endlichkeitssätze für abelsche Varietäten über Zahlkörpern

G. Faltings

Fachbereich Mathematik, Bergische Universitäts-Gesamthochschule Wuppertal, Gaußstr. 20, D-5600 Wuppertal 1, Bundesrepublik Deutschland

§1. Einleitung

Sei K eine endliche Erweiterung von $\mathbb{Q}$, A eine über K definierte abelsche Varietät, $\pi = \mathrm{Gal}(\bar{K}/K)$ die absolute Galois-Gruppe von K, l eine Primzahl. Dann operiert π auf dem Tate-Modul

$$T_l(A) = \varprojlim A[l^n](\bar{K}).$$

Ziel der vorliegenden Arbeit ist der Beweis der folgenden Resultate:

a) Die Darstellung von π auf $T_l(A) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ ist halbeinfach.

b) Die Abbildung

$$\mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Z}_l \to \mathrm{End}_\pi(T_l(A))$$

ist ein Isomorphismus.

c) Sei S eine endliche Menge von Stellen von K, $d > 0$. Dann gibt es nur endlich viele Isomorphieklassen d-fach polarisierter abelscher Varietäten über K, welche gute Reduktion außerhalb S haben.

a) und b) sind bekannt unter dem Namen Tate-Vermutung, c) als Shafarevich-Vermutung. Weiter weiß man ([9]), daß aus c) die Mordell-Vermutung folgt. Die Tate-Vermutung ist von Tate selbst für abelsche Varietäten über endlichen Körpern gezeigt worden. Zarhin verallgemeinerte dies auf Funktionenkörper ([15, 16]) über solchen, und unser Beweis ist eine Übertragung seiner Methoden auf den Zahlkörper-Fall. Das zur Übersetzung notwendige Wörterbuch hat Arakelov geliefert ([2]), und seine Methoden sind vom Verfasser ausgebaut worden ([5]). Kurz gesagt handelt es sich darum, „alles" mit hermiteschen Metriken zu versehen.

Der Beweis für c) erfolgt so, daß zunächst nur die Endlichkeit für Isogenie-Klassen gezeigt wird. Die grundlegende Idee dazu hat mir der Gutachter der „Inventiones" anläßlich der Veröffentlichung meiner Arbeit [6] mitgeteilt, und ich mußte sie nur noch von der Hodge-Theorie in die étale Kohomologie über-

setzen. Ich möchte daher dem mir persönlich unbekannten Gutachter an dieser Stelle für seine Anregung herzlich danken.

Der Rest des Beweises von c) benutzt eine Variante der bei a) und b) verwendeten Methoden.

Die Arbeit beginnt zunächst mit einigen technischen Details über Höhen. Die Komplikationen ergeben sich daraus, daß meines Wissens noch kein guter Modulraum semiabelscher Varietäten über $\mathbb{Z}$ existiert. (Mit ähnlichen Problemen hat sich auch L. Moret-Bailly zu kämpfen, der die Verhältnisse über Funktionenkörpern untersucht ([7]).) Danach werden die sehr schönen Ergebnisse von Tate ([13]) über p-divisible Gruppen benutzt. Der Schluß ist dann wieder etwas technischer.

Ich habe viel über das Thema von L. Szpiro gelernt, dem ich an dieser Stelle für seine Einführung in diesen Problemkreis danke. P. Deligne hat mich auf eine Unstimmigkeit in der ursprünglichen Fassung der Arbeit aufmerksam gemacht.

§ 2. Semiabelsche Varietäten

Definition. Sei S ein Schema (oder ein algebraic stack). Eine semiabelsche Varietät der relativen Dimension g über S ist eine glatte algebraische Gruppe $p\colon G \to S$, so daß die Fasern von p zusammenhängend von der Dimension g sind, und Erweiterungen einer abelschen Varietät durch einen Torus.

Beispiel. Sei $q\colon C \to S$ eine stabile Kurve vom Geschlecht g ([4]). Dann ist

$$J = \mathrm{Pic}^{\tau}(C/S) \to S$$

eine semiabelsche Varietät der relativen Dimension g.

Wir benötigen das folgende

Lemma 1. *Sei S normal, $U \subset S$ offen und dicht $p_1\colon A_1 \to S$ und $p_2\colon A_2 \to S$ zwei semiabelsche Varietäten, $\phi\colon A_1/U \to A_2/U$ ein über U definierter Homomorphismus algebraischer Gruppen. Dann läßt sich ϕ eindeutig auf ganz S fortsetzen.*

Beweis. Dies ist wohlbekannt, falls S Spektrum eines kompletten diskreten Bewertungsringes ist. Im allgemeinen reduziert man sofort auf S noethersch und exzellent, und bezeichnet mit

$$X \subseteq A_1 \times_S A_2$$

den Abschluß des Graphen von ϕ.

Nach Basiswechsel mit geeigneten Bewertungsringen sieht man, daß die Projektion $pr_1\colon X \to A_1$ eigentlich ist, und daß ihre Fasern nur einen Punkt besitzen. Da A_1 normal ist, muß pr_1 ein Isomorphismus sein, und X der Graph der eindeutig bestimmten Fortsetzung von ϕ. (Eindeutigkeit folgt zum Beispiel durch Betrachtung von Torsionspunkten, oder auf tausend andere Arten.)

Definition. Sei $p\colon A \to S$ eine semiabelsche Varietät der relativen Dimension g, $s\colon S \to A$ der Nullschnitt.

Setze:

$$\omega_{A/S} = s^*(\Omega^g_{A/S}),$$

$\omega_{A/S}$ ist ein Geradenbündel auf S.

Bemerkungen. a) Wenn p eigentlich ist, so ist $\omega_{A/S} \cong p_*(\Omega^g_{A/S})$.

b) $\omega_{A/S}$ kommutiert mit Basiswechsel.

c) Wenn $A = \mathrm{Pic}^\tau(C/S)$ mit einer stabilen Kurve $q\colon C \to S$, so ist $\omega_{A/S} \cong \Lambda^g q_*(\omega_{C/S})$, wobei $\omega_{C/S}$ den relativen dualisierenden Modul bezeichnet.

d) Wenn $S = \mathrm{Spec}(\mathbb{C})$ und p eigentlich ist (d.h., $A/\mathbb{C}$ ist eine komplexe abelsche Varietät), so besitzt

$$\omega_{A/S} \cong \Gamma(A, \Omega^g_{A/\mathbb{C}})$$

ein kanonisches hermitesches Skalarprodukt:

Wenn α, β holomorphe Differentialformen auf A sind, so setze

$$\langle \alpha, \beta \rangle = \left(\frac{i}{2}\right)^g \int_A \alpha \wedge \bar{\beta}.$$

Wir benötigen einige Tatsachen über die Modulräume stabiler Kurven und abelscher Varietäten. Dazu scheint die Sprache der „algebraic stacks" am angemessensten zu sein. Falls dem Leser diese Notation zu abstrakt erscheint, so möge er die folgenden Überlegungen durchführen:

Es geht eigentlich um Endlichkeitsaussagen.

Wenn $\mathfrak{S}$ einer der demnächst einzuführenden algebraic stacks ist, und S den zugehörigen groben Modulraum bezeichnet, so gibt es stets eine offene Überdeckung

$$S = \bigcup_{i=1}^{r} U_i$$

und endlich surjektive Abbildungen $V_i \to U_i$, so daß über V_i das „universelle Objekt zu $\mathfrak{S}$" existiert. Man kann dann alle Rechnungen in den V_i durchführen.

Nun zu den hier verwendeten algebraic stacks.

1. $\overline{\mathfrak{M}}_g$ klassifiziere die stabilen Kurven vom Geschlecht g ([4]). $\overline{\mathfrak{M}}_g$ ist eigentlich über $\mathrm{Spec}(\mathbb{Z})$, und der zugehörige grobe Modulraum heiße $\bar{M}_g$.

2. $\mathfrak{A}_g$ klassifiziere die prinzipal-polarisierten abelschen Varietäten der relativen Dimension g, und A_g sei der zugehörige Modulraum.

$\mathfrak{A}_g$ ist nicht eigentlich über $\mathrm{Spec}(\mathbb{Z})$, aber die folgenden Tatsachen sind bekannt:

a) Wenn

$$p\colon A \to \mathfrak{A}_g$$

die universelle abelsche Varietät über $\mathfrak{A}_g$ bezeichnet, so gibt es ein $r > 0$, für welches $(\omega_{A/\mathfrak{A}_g})^{\otimes r}$ ein sehr amples Geradenbündel auf $A_g/\mathbb{Q}$ definiert ([3]). Sei $\bar{A}_g/\mathbb{Q}$ der Zariski-Abschluß von $A_g/\mathbb{Q}$ in dem zugehörigen projektiven Raum $\mathbb{P}^N_{\mathbb{Q}}$, $\bar{A}_g/\mathbb{Z}$ der Zariski-Abschluß in $\mathbb{P}^N_{\mathbb{Z}}$, und $\mathscr{M}$ das Geradenbündel $\mathscr{O}(1)$ auf $\bar{A}_g/\mathbb{Z}$. ($\mathscr{M}$ setzt auf $\bar{A}_g/\mathbb{Q}$ $(\omega_{A/\mathfrak{A}_g})^{\otimes r}$ fort.)

b) Es gibt über $\mathbb{C}$ einen eigentlichen dominanten Morphismus

$$\phi\colon \mathfrak{N} \to \bar{A}_g/\mathbb{C},$$

so daß über $\mathfrak{N}$ eine semiabelsche Varietät existiert, welche die universelle abelsche Varietät über $\mathfrak{A}_g$ fortsetzt (siehe [8], §9). Außerdem ist bekannt, daß $\omega^{\otimes r}$ dieser semiabelschen Varietät isomorph zu $\phi^*(\mathscr{M})$ ist. (Hierzu muß man direkt rechnen, siehe meine Ausführungen dazu in [6], §2.)

Lemma 2. *Es gibt über* Spec($\mathbb{Z}$) *einen eigentlichen algebraic stack* $\mathfrak{Z}$, *eine offene Teilmenge* $\mathfrak{U} \subset \mathfrak{Z}$ *und einen eigentlichen Morphismus* $\psi\colon \mathfrak{U} \to \mathfrak{A}_g$, *welcher sich zu einem* $\bar{\psi}\colon \mathfrak{Z}/\mathbb{Q} \to \bar{A}_g/\mathbb{Q}$ *fortsetzt, so daß über* $\mathfrak{Z}$ *die folgenden Objekte existieren:*

a) *Eine stabile Kurve* $q\colon C \to \mathfrak{Z}$.

b) *Ein Untergeradenbündel* (= *lokal direkter Sumand*) $\mathscr{L} \subseteq \Lambda^g q_*(\omega_{C/\mathfrak{Z}})$.

c) *Über* $\mathfrak{U}$ *ein Paar von Gruppenhomomorphismen*

$$\alpha\colon \mathrm{Pic}^\tau(C/\mathfrak{Z}) \to \psi^*(A),$$
$$\beta\colon \psi^*(A) \to \mathrm{Pic}^\tau(C/\mathfrak{Z})$$

mit

$$\alpha \circ \beta = \textit{Multiplikation mit einem } d \in \mathbb{N},\ d \neq 0.$$

(*Dabei sei A wieder die universelle abelsche Varietät über* $\mathfrak{A}_g$.)

d) *Auf* $\mathfrak{Z} \otimes_{\mathbb{Z}} \mathbb{Q}$ *existiert ein Isomorphismus* $\mathscr{L}^{\otimes r} = \bar{\psi}^*(\mathscr{M})$, *und* $\mathscr{L}$ *ist über* $\mathfrak{U}/\mathbb{Q}$ *das Bild von*

$$\alpha^*\colon \psi^*(\omega_{A/\mathfrak{A}_g}) \to \Lambda^g q_*(\omega_{C/Z}).$$

Der daraus resultierende Isomorphismus (*über* $\mathfrak{U}/\mathbb{Q}$)

$$\psi^*(\omega_{A/\mathfrak{A}_g})^{\otimes r} \cong \psi^*(\mathscr{M})$$

ist ψ^**-Pullback des bei der Konstruktion von* $\mathscr{M}$ *angegebenen Isomorphismus über* A_g.

Beweis. In dem generischen Punkte von $\mathfrak{A}_g$ ist die zugehörige abelsche Varietät Quotient einer Jacobischen. Die zugehörige Kurve entspricht einer rationalen Abbildung von $\mathfrak{A}_g$ in $\overline{\mathfrak{M}}_{\tilde{g}}$, für ein $\tilde{g}$.

Wenn man den Graphen dieser Abbildung betrachtet, so erhält man (mit Hilfe einiger trivialer Zusatzüberlegungen) einen ersten Kandidaten $\mathfrak{Z}$, so daß schon a) und (nach Lemma 1) c) erfüllt sind. $\mathscr{L}$ ist dann über $\mathfrak{U} \otimes_{\mathbb{Z}} \mathbb{Q}$ durch d) schon festgelegt, und liefert eine rationale Abbildung von $\mathfrak{U} \otimes_{\mathbb{Z}} \mathbb{Q}$ in ein geeignetes projektives Bündel über $\mathfrak{Z}$. Man ersetzt $\mathfrak{Z}$ durch die Normalisierung des Abschlusses des zugehörigen Graphen, und dann sind auch b) und der zweite Teil von d) erfüllt. Zum Rest von d) ist zu vermerken, daß man den gesuchten Isomorphismus schon über $\mathfrak{U} \otimes_{\mathbb{Z}} \mathbb{Q}$ konstruiert hat, und man muß nur noch die Fortsetzbarkeit auf $\mathfrak{Z} \otimes_{\mathbb{Z}} \mathbb{Q}$ zeigen. Dazu kann man den Grundkörper von $\mathbb{Q}$ auf $\mathbb{C}$ erweitern, und es reicht, die Fortsetzbarkeit für ein $\tilde{\mathfrak{Z}}/\mathbb{C}$ zu zeigen, welches dominant und eigentlich über $\mathfrak{Z}$ liegt.

Mit Hilfe des weiter oben eingeführten $\phi\colon \mathfrak{N} \to \bar{A}_g/\mathbb{C}$ konstruiert man ein normales $\tilde{\mathfrak{Z}}$, so daß $\psi^*(A)$ sich auf $\tilde{\mathfrak{Z}}$ zu einer semiabelschen Varietät fortsetzt. Nach Lemma 1 kann man auch α und β fortsetzen, und diese liefern den gewünschten Isomorphismus über $\tilde{\mathfrak{Z}}$.

Korollar. *Es gibt eine natürliche Zahl* $e > 0$ *mit der folgenden Eigenschaft:*

Sei K ein Zahlkörper, R der Ring der ganzen Zahlen in K,

$$p\colon A \to \mathrm{Spec}(R)$$

eine semiabelsche Varietät, so daß die generische Faser A/K eigentlich über K ist, und eine prinzipale Polarisation bestitzt. Dem entspricht eine Abbildung ρ: $\mathrm{Spec}(K) \to A_g/\mathbb{Q}$, welche sich zu einem ρ: $\mathrm{Spec}(R) \to \bar{A}_g/\mathbb{Z}$ fortsetzt.

Nach Konstruktion besteht ein Isomorphismus

$$\rho^*(\mathscr{M}) \otimes_R K \cong (\omega_{A/R})^{\otimes r} \otimes_R K.$$

Unter Benutzung dieses Isomorphismus gilt:

$$e \cdot \rho^*(\mathscr{M}) \subseteq (\omega_{A/R})^{\otimes r} \subseteq e^{-1} \cdot \rho^*(\mathscr{M}) \left(\subseteq \rho^*(\mathscr{M}) \otimes_R K\right).$$

Beweis. Wir dürfen annehmen, daß $\bar{\psi}$: $\mathfrak{Z}/\mathbb{Q} \to \bar{A}_g/\mathbb{Q}$ sich fortsetzt zu einem eigentlichen $\bar{\psi}$: $\mathfrak{Z}/\mathbb{Z} \to \bar{A}_g/\mathbb{Z}$. Dann gibt es eine endliche Körpererweiterung $K' \supseteq K$ (mit ganzen Zahlen $R' \subseteq K'$), so daß sich ρ liften läßt zu

$$\tilde{\rho}\colon \mathrm{Spec}(R') \to \mathfrak{Z}.$$

Da über $\mathfrak{Z} \otimes_{\mathbb{Z}} \mathbb{Q}$ $\bar{\psi}^*(\mathscr{M})$ und $\mathscr{L}^{\otimes r}$ isomorph sind, gibt es ein $e_1 > 0$, *so daß über* $\mathfrak{Z}$

$$e_1 \cdot \mathscr{L}^{\otimes r} \subseteq \bar{\psi}^*(\mathscr{M}) \subseteq e_1^{-1} \cdot \mathscr{L}^{\otimes r}.$$

Es reicht, die Behauptung nach dem Basiswechsel zu R' zu zeigen, und wir müssen nur noch $\omega_{A/R'}$ und $\tilde{\rho}^*(\mathscr{L})$ vergleichen.

Durch Pullback erhält man eine stabile Kurve

$$q\colon C \to \mathrm{Spec}(R')$$

und

$$\alpha\colon \mathrm{Pic}^\tau(C/R') \to A/R', \qquad \beta\colon A/R' \to \mathrm{Pic}^\tau(C/R')$$

mit $\alpha \circ \beta = d \cdot \mathrm{id}$ (Benutze Lemma 1 über R'), so daß $\tilde{\rho}^*(\mathscr{L})$ das Unterbündel von $\Lambda^g q_*(\omega_{C/R'})$ ist, welches vom Bild von

$$\alpha^*\colon \omega_{A/R'} \to \Lambda^g q_*(\omega_{C/R'})$$

erzeugt wird. Daraus folgt unmittelbar, daß $d^g \cdot \bar{\rho}^*(\mathscr{L}) \subseteq \omega_{A/R'} \subseteq \bar{\rho}^*(\mathscr{L})$, und wir sind fertig.

§ 3. Höhen

Sei K wieder ein Zahlkörper, R der Ring der ganzen Zahlen in K. Analog zu [5] definieren wir ein metrisiertes Geradenbündel auf $\mathrm{Spec}(R)$ als einen projektiven R-Modul P vom Rang 1, zusammen mit Normen $\| \ \|_v$ auf $P \otimes_R K_v$ für alle unendlichen Stellen v von K. Dabei bezeichnet K_v die Komplettierung von K in v, und wir definieren ein $\varepsilon_v = 1$ oder 2, je nachdem ob $K_v \cong \mathbb{R}$ oder $K_v \cong \mathbb{C}$.

Der Grad des metrisierten Geradenbündels wird definiert als („#" = Ordnung)

$$\mathrm{Grad}(P, \| \ \|) = \log(\#(P/R \cdot p)) - \sum_v \varepsilon_v \cdot \log \|p\|_v.$$

Dabei ist p ein von Null verschiedenes Element von P, die Summe geht über alle unendlichen Stellen von K, und die rechte Seite ist selbstverständlich unabhängig von p.

Bemerkung. Der Begriff des metrischen Geradenbündels stammt von Arakelov ([2]). Der Grad von P hängt natürlich auch mit dem Volumen von P zusammen.

Uns interessieren besonders die metrisierten Geradenbündel $\omega_{A/R}$, wobei

$$p\colon A \to \mathrm{Spec}(R)$$

eine semiabelsche Varietät ist, mit eigentlicher generischer Faser A/K. Die Metriken an den unendlichen Stellen kommen vom schon erwähnten Skalarprodukt:

$$\|\alpha\|_v^2 = \left(\frac{i}{2}\right)^g \cdot \int_{A(\bar{K}_v)} \alpha \wedge \bar{\alpha}.$$

Definition. Die modultheoretische Höhe $h(A)$ ist

$$h(A) = \frac{1}{[K:\mathbb{Q}]} \mathrm{Grad}(\omega_{A/R}).$$

Man sieht sofort, daß $h(A)$ invariant ist gegenüber Erweiterungen des Grundkörpers. Der Name „Höhe" rechtfertigt sich wie folgt:

Im allgemeinen definiert man die Höhe eines Punktes $x \in \mathbb{P}^n(K)$, indem man x einen Morphismus $\rho\colon \mathrm{Spec}(R) \to \mathbb{P}^n_{\mathbb{Z}}$ zuordnet, das Bündel $\mathcal{O}(1)$ auf $\mathbb{P}^n_{\mathbb{C}}$ mit einer Metrik versieht, und dann als Höhe von x

$$\frac{1}{[K:\mathbb{Q}]} \cdot \mathrm{Grad}(\rho^* \mathcal{O}(1))$$

definiert.

Bei Veränderung der hermiteschen Metrik ändert sich die Höhenfunktion nur um einen beschränkten Betrag, und es ist bekannt, daß für jedes c nur endlich viele K-rationale Punkte des $\mathbb{P}^n$ Höhe $\leqq c$ haben. Entsprechende Überlegungen gelten für abgeschlossene Untervariatäten des $\mathbb{P}^n$. Angewandt auf unsere Situation bettet man wie bisher A_g mittels $\mathcal{M}$ in $\mathbb{P}^n_{\mathbb{Z}}$ ein. Außerdem hat man schon eine Metrik $\| \ \|$ auf dem von $\mathcal{M}$ auf $A_g(\mathbb{C})$ induzierten Bündel definiert. Wenn sich diese Metrik auf $\bar{A}_g(\mathbb{C})$ fortsetzen ließe, so könnte man sie zur Definition der Höhe verwenden, und aus dem Korollar zu Lemma 2 würde folgen, daß für eine semiabelsche Varietät A über R (wie oben), welche über K eine prinzipale Polarisation besitzt und damit ein $x \in A_g(K)$ definiert, $h(x)$ und $r \cdot h(A)$ sich nur um eine beschränkten Betrag unterscheiden.

Leider hat die Metrik $\| \ \|$ Singularitäten längs $\bar{A}_g(\mathbb{C}) - A_g(\mathbb{C})$, doch sind diese so mild, daß die fundamentale Endlichkeits-Eigenschaft der Höhe erhalten bleibt:

Definition. Sei $X/\mathbb{C}$ eine kompakte komplexe Varietät, $Y \subseteq X$ eine abgeschlossene Untervarietät, $\mathscr{M}$ eine Geradenbündel auf X, $\| \ \|$ eine hermitesche Metrik auf $\mathscr{M}|X-Y$. Die Metrik $\| \ \|$ hat logarithmische Singularitäten längs Y, wenn folgendes gilt: Es gibt eine eigentliche dominante Abbildung

$$\phi\colon \tilde{X} \to X,$$

so daß $\tilde{X}$ glatt und $\phi^{-1}(Y)$ ein Divisor mit normalen Überkreuzungen ist, und so daß für ein lokales Erzeugendes h von $\phi^*(\mathscr{M})$ und eine lokale Gleichung f von $\phi^{-1}(Y)$.

$$\operatorname{Sup}\{\|h\|, \|h\|^{-1}\} \leqq c_1 \cdot |(\log|f|)|^{c_2}$$

(mit Konstanten $c_1, c_2 > 0$) gilt.

Beispiel.

$$X = \bar{A}_g(\mathbb{C}), \qquad Y = \bar{A}_g(\mathbb{C}) - A_g(\mathbb{C}), \qquad \mathscr{M} \text{ und } \| \ \|$$

wie bisher.

Dies wurde zwar schon in [6] gezeigt (Ende von §2), doch folgt hier auf Wunsch der Referenten eine kurze Beweisskizze:

Allgemeiner gilt für ein glattes X, einen Divisor Y von X mit normalen Überkreuzungen und eine semiabelsche Varietät

$$p\colon A \to X,$$

so daß über $X-Y$ p eigentlich und A prinzipal polarisiert ist, daß die kanonische Metrik auf $\omega_{A/X}$ logarithmische Singularitäten längs Y hat.

Dazu betrachtet man statt $\omega_{A/X}$ $p_*(\Omega^1_{A/X})$ („logarithmische Singularitäten" läßt sich auch für Vektorbündel definieren), und mit den Methoden des §2 reduziert man das Problem auf den Fall, daß A Jacobische einer semi-stabilen Kurve $q\colon C \to X$ ist.

Wir behandeln kurz den Fall einer semi-stabilen Kurve über dem Einheitskreis $\mathbb{D}$. Der allgemeine Fall geht genauso. Wenn

$$q\colon C \to \mathbb{D} = \{t \mid |t| < 1\}$$

eine semi-stabile Kurve ist, mit guter Reduktion außerhalb 0, so besitzt C eine Überdeckung

$$C = \bigcup_{i=1}^{l} U_i,$$

so daß entweder

a) $U_i = \{(z,t) \mid |z| < 1, |t| < 1\}$, $q|U_i\colon U_i \to \mathbb{D}$ glatt, z liefert Koordinate auf allen Fasern

oder

b) $U_i = \{(z_1, z_2, t) \mid |z_1| < \varepsilon, |z_2| < \varepsilon, z_1 z_2 = t^m\}$, $q(z_1, z_2, t) = t$.

Wenn α ein lokaler Schnitt von $q_*(\omega_{C/\mathbb{D}})$ ist, so ist α auf den U_i von der Form

a) $\alpha = (\text{holomorph}) \cdot dz$ bzw.

b) $\alpha = (\text{holomorph}) \cdot \dfrac{dz_1}{z_1}$.

Eine explizite Rechnung ergibt, daß

$$\frac{i}{2}\int_{U_i\cap q^{-1}(t)} \alpha\wedge\bar{\alpha}$$

für $t\to 0$ entweder beschränkt bleibt, oder höchstens wie $|\log|t||$ wächst. Außerdem sieht man sofort, daß

$$\| \ \| \geqq (\text{pos. Konst.}) \ \| \ \|_1,$$

wobei $\| \ \|_1$ eine auf ganz X erklärte hermitesche Metrik auf $q_*(\omega_{C/X})$ bezeichnet.

Lemma 3. *Sei $X\subseteq \mathbb{P}^n_{\mathbb{Z}}$ Zariski-abgeschlossen, $Y\subseteq X$ abgeschlossen, $\| \ \|$ eine hermitesche Metrik auf $\mathcal{O}(1)|(X(\mathbb{C})-Y(\mathbb{C}))$, mit logarithmischen Singularitäten längs Y. Für einen Zahlkörper K und $x\in X(K)-Y(K)$ definiert man wie bisher $h(x)$. Dann gibt es für jedes c nur endlich viele $x\in X(K)-Y(K)$ mit $h(x)\leqq c$.*

Beweis. Sei $\| \ \|_1$ eine hermitesche Metrik für $\mathcal{O}(1)|X(\mathbb{C})$, h_1 die zugehörige Höhenfunktion, und man wähle ein $s>0$ und

$$f_1,\ldots,f_t\in\Gamma(X/\mathbb{Z},\mathcal{O}(s)),$$

deren gemeinsame Nullstellenmenge genau Y ist. Dann definiert $\| \ \|_1$ eine Metrik auf $\mathcal{O}(s)$ (welche auch $\| \ \|_1$ heiße), und aus den Voraussetzungen folgt sofort, daß Konstanten $c_1, c_2>0$ existieren

$$\text{mit } \log\left|\frac{\| \ \|}{\| \ \|_1}(z)\right| \leqq c_1+c_2\cdot\inf\{\log(|\log\| f_i(z)\|_1|)\}$$

für $z\in X(\mathbb{C})$.

Wenn $x\in X(K)-Y(K)$, entsprechend einem

$$\rho\colon \operatorname{Spec}(R)\to X,$$

so definieren die f_i Schnitte $\rho^*(f_i)$ von $\rho^*(\mathcal{O}(s))$, mit deren Hilfe man die Höhe $h_1(x)$ berechnen kann. Da $\|f_i(z)\|_1$ auf $X(\mathbb{C})$ nach oben beschränkt ist, erhält man sofort Konstanten $c_3, c_4>0$ mit

$$|h(x)-h_1(x)|\leqq c_3+c_4\cdot\log(h_1(x)).$$

Daraus folgt unmittelbar die Behauptung.

Wir können nun die Früchte unserer Bemühungen ernten. Das folgende Resultat ist schon fast bewiesen:

Satz 1. *Sei c gegeben. Dann gibt es nur endlich viele Isomorphieklassen von Paaren aus*

i) *einer semiabelschen Varietät der relativen Dimension g*

$$p\colon A\to\operatorname{Spec}(R)$$

mit eigentlicher generischer Faser A/K

ii) *einer prinzipalen Polarisation auf* A/K,

für welche

$$h(A) \leqq c$$

gilt.

Beweis. Nach dem Korollar zu Lemma 2 ist die Differenz zwischen $h(x)$ und $r \cdot h(A)$ beschränkt ($x \in A_g(K)$ gehört zu A). Nach Lemma 3 liefern die A mit $h(A) \leqq c$ somit nur endlich viele verschiedene $x \in A_g(K)$. Wir müssen uns nun noch überlegen, daß nur endlich viele K-Isomorphieklassen dieselbe Klasse über dem algebraischen Abschluß $\overline{K}$ induzieren können. Wir fixieren also eine solche Klasse über $\overline{K}$, und betrachten die entsprechenden A/K. Es ist bekannt, daß alle diese an denselben Stellen von K schlechte Reduktion haben. Aus dem Lemma 4 weiter unten folgt dann sofort die Existenz einer endlichen Erweiterung $K' \supseteq K$, welche für ein $n \geqq 3$ die Koordinaten der n-Teilungspunkte aller A/K enthält. Bekanntlich sind dann unsere A's schon über K' isomorph, und der Rest folgt aus allgemeinen Grundsätzen der Galois-Kohomologie.

Es bleibt nachzutragen das

Lemma 4. *Sei K ein Zahlkörper, S eine endliche Menge von Stellen von K. Dann gibt es nur endlich viele Körpererweiterungen $K' \supseteq K$ von vorgegebener Ordnung, welche außerhalb von S unverzweigt sind.*

Beweis. Bekannt (Hermite-Minkowski).

§ 4. Isogenien

Wir untersuchen das Verhalten von $h(A)$ unter Isogenien. Wie immer ist K ein Zahlkörper, $R \subset K$ der Ring der ganzen Zahlen.

Seien

$$p_1 \colon A_1 \to \mathrm{Spec}(R)$$

und

$$p_2 \colon A_2 \to \mathrm{Spec}(R)$$

semiabelsche Varietäten mit eigentlicher generischer Faser,

$$s \colon \mathrm{Spec}(R) \to A_1$$

der Nullschnitt, und $\phi \colon A_1 \to A_2$ eine Isogenie. (Nach Lemma 1 reicht es natürlich, wenn man ϕ über K definiert.)

Wir setzen $G = \mathrm{Ker}(\phi) \subseteq A_1$. Da ϕ automatisch flach ist, ist G ein quasiendliches flaches Gruppenschema über $\mathrm{Spec}(R)$.

ϕ induziert eine Injektion

$$\phi^* \colon \omega_{A_2/R} \to \omega_{A_1/R},$$

und man sieht sofort, daß

$$\#(\omega_{A_1/R}/\phi^*(\omega_{A_2/R})) = \#s^*(\Omega^1_{A_1/A_2}) = \#s^*(\Omega^1_{G/R}).$$

Da außerdem ϕ^* die Normen an den unendlichen Stellen um $(\mathrm{Grad}(\phi))^{1/2}$ verändert, folgt unmittelbar

Lemma 5.

$$h(A_2)=h(A_1)+\tfrac{1}{2}\log(\mathrm{Grad}(\phi))-\frac{1}{[K:\mathbb{Q}]}\cdot\log(\#s^*(\Omega^1_{G/R})).$$

Bemerkung. Wenn G von einer Zahl $n\in\mathbb{N}$ annulliert wird, so annulliert n auch $\Omega^1_{G/R}$. Daraus folgt, daß

$$\exp(2[K:\mathbb{Q}]\cdot(h(A_2)-h(A_1)))$$

eine rationale Zahl ist, in deren Zähler und Nenner nur Primteiler von $\mathrm{Grad}(\phi)$ auftauchen. Die Exponenten dieser Primteiler darin können durch ihre Exponenten in $\mathrm{Grad}(\phi)$ beschränkt werden.

Wir untersuchen nun das Verhalten der $h(A_n)$, falls $A_n=A/G_n$, wobei die G_n die Stufen einer l-divisiblen Gruppe $G\subseteq A[l^\infty]$ durchlaufen.

Satz 2. *Sei p: $A\to\mathrm{Spec}(R)$ eine semiabelsche Varietät mit eigentlicher generischer Faser, l eine Primzahl, und $G/K\subseteq A[l^\infty]/K$ eine l-divisible Untergruppe.*

Weiter sei G_n der Kern von l^n auf G, und A_n die semiabelsche Varietät $A_n=A/G_n$. Dann ist

$$h(A_n)=h(A).$$

Beweis. Seien $v_1,\ldots,v_r$ die über l liegenden Stellen von k, $K_i=K_{v_i}$ die entsprechenden lokalen Körper, $R_i\subseteq K_i$ die Bewertungsringe, $m_i=[K_i:\mathbb{Q}_l]$, so daß

$$m=[K:\mathbb{Q}]=\sum_{i=1}^{r}m_i.$$

Wir wählen ein festes i, und betrachten das formale Gruppenschema $\hat{A}$ über $Spf(R_i)$, die Komplettierung von A/R_i längs der Faser A_s über dem abgeschlossenen Punkt s von $\mathrm{Spec}(R_i)$.

A_s ist eine Erweiterung $0\to T_s\to A_s\to B_s\to 0$ mit T_s Torus, B_s eine abelsche Varietät.

Nach allgemeinen Grundsätzen kann man T_s liften zu einem Torus T über $\mathrm{Spec}(R_i)$, und $\hat{T}$ ist abgeschlossenes formales Unterschema von $\hat{A}$. (Morphismen von T_s in glatte Gruppenschemata können geliftet werden.)

Sei

$$\hat{H}_i=\hat{A}[l^\infty]$$

die assoziierte l-divisible Gruppe. $\hat{H}_i$ ist formale Komplettierung einer l-divisiblen Gruppe H_i über R_i, und H_i/K_i ist eine l-divisible Untergruppe von $A[l^\infty]/K_i$. Dasselbe gilt für $T[l^\infty]$, und zu diesen Untergruppen gehören $\mathbb{Z}_l$-Untergitter

$$T_l(T)\subseteq T_l(H_i)\subseteq T_l(A).$$

Lemma 6. *Sei $D_i=\mathrm{Gal}(\overline{K}_i/K_i)$ die absolute Galois-Gruppe von K_i, $I_i\subseteq D_i$ die Verzweigungsgruppe.*

Dann operiert I_i trivial auf $T_l(A)/T_l(H_i)$, und die induzierte Operation von $D_i/I_i\cong\hat{\mathbb{Z}}$ erfolgt über einen endlichen Quotienten von $\hat{\mathbb{Z}}$.

Beweis. Sei

$$\langle\ ,\ \rangle\colon\ T_l(A)\times T_l(A)\to\mathbb{Z}_l(1)=T_l(\mathbb{G}_m)$$

die von einer Polarisation von A/K induzierte symplektische Form. $\langle\ ,\ \rangle$ ist nicht ausgeartet, und bekanntlich (SGA VII, Exp. IX, § 7) ist

$$\langle T_l(T), T_l(H_i)\rangle = 0.$$

Aus Dimensionsgründen ist $T_l(H_i) = T_l(T)^{\perp}$, und wir erhalten eine Injektion

$$T_l(A)/T_l(H_i) \hookrightarrow \mathrm{Hom}_{\mathbb{Z}_l}(T_l(T), \mathbb{Z}_l(1)).$$

Diese Injektion ist D_i-linear, und D_i operiert in der verlangten Art und Weise auf $\mathrm{Hom}_{\mathbb{Z}_l}(T_l(T), \mathbb{Z}_l(1))$.

Nunmehr zurück zu unserem $G/K \subseteq A[l^{\infty}]/K$. Nach Basiserweiterung $K \subseteq K_i$ können wir den Durchschnitt $G_i = G \cap H_i$ bilden. Dies ist die maximale l-divisible Untergruppe von G_i/K_i, welche sich über R_i ausdehnen läßt, und es ist

$$\#(s^* \Omega^1_{A/A_n} \otimes_R R_i) = \# s^*(\Omega^1_{(G_i)_n/R_i}).$$

Nach [13], Proposition 2 kann man dies sofort ausrechnen: Sei d_i die Dimension der maximalen formalen Untergruppe von G_i. Dann ist

$$\# s^*(\Omega^1_{(G_i)_n/R_i}) = l^{n \cdot m_i \cdot d_i}.$$

Wenn C_i die Komplettierung des algebraischen Abschlusses von K_i bezeichnet, so ist weiter bekannt ([13], Theorem 3, Corollary 2), daß

$$T_l(G_i) \otimes_{\mathbb{Z}_l} C_i \cong C_i^{h_i - d_i} \oplus C_i^{d_i}(+1),$$

($h_i = $ Höhe(G_i), „$(+1)$" = Tate-Twist) als D_i-Moduln.

Zusammen mit Lemma 6 ergibt sich daraus, daß D_i auf

$$\Lambda^h T_l(G) \otimes_{\mathbb{Z}_l} C_i \subseteq \Lambda^h T_l(A) \otimes_{\mathbb{Z}_l} C_i$$
$$(h = \text{Höhe}(G))$$

wie auf

$$C_i(\chi_0^{d_i}) = C_i(d_i)$$

operiert ($\chi_0 =$ zyklotomischer Charakter).

Wir übertragen dies nun wieder auf den globalen Fall:

Es ist

$$\# s^*(\Omega^1_{A/A_n}) = l^{n \sum_{i=1}^{r} m_i d_i}, \qquad (l^n \cdot \Omega^1_{A/A_n} = 0)$$

und

$$h(A_n) - h(A) = n \cdot \log(l) \cdot \left(\frac{h}{2} - \sum_{i=1}^{r} \frac{m_i}{m} \cdot d_i\right).$$

Wir müssen also zeigen, daß $\sum_{i=1}^{r} m_i d_i = \frac{1}{2} m h$. Dazu betrachten wir die absolute Galois-Gruppe $\tilde{\pi} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q})$, und den $\tilde{\pi}$-Modul

$$\tilde{V} = \mathrm{Ind}_{\pi}^{\tilde{\pi}}(T_l(A)) \qquad (\pi = \mathrm{Gal}(\overline{K}/K)).$$

Dieser enthält den Untermodul

$$\tilde{W} = \mathrm{Ind}_{\pi}^{\tilde{\pi}}(T_l(G))$$

vom Rang mh, und $\tilde{\pi}$ operiert auf der Geraden

$$L = \Lambda^{mh}(\tilde{W}) \subseteq \Lambda^{mh}(\tilde{V})$$

via einen Charakter $\chi\colon \tilde{\pi} \to \mathbb{Z}_l^*$.

Aus der Klassenkörpertheorie folgt, daß χ von der Form $\chi = (l\text{-adische Potenz von } \chi_0) \cdot (\text{Charakter endlicher Ordnung})$ ist.

Die darin auftretende l-adische Potenz von χ_0 bestimmt sich wie folgt: Sei C die Komplettierung des algebraischen Abschlusses von $\mathbb{Q}_l$,

$$D \cong \mathrm{Gal}(\overline{\mathbb{Q}}_l/\mathbb{Q}_l) \subseteq \tilde{\pi}$$

die Zerlegungs-Gruppe von l. Dann gilt als D-Modul

$$L \otimes_{\mathbb{Z}_l} C \cong C\left(+\sum_{i=1}^{r} m_i d_i\right)$$

(dies folgt aus unseren vorherigen Berechnungen), und somit ist nach [13], Theorem 2

$$\chi \cdot \chi_0^{-\sum_{i=1}^{r} m_i d_i}$$

ein Charakter endlicher Ordnung auf D und auch auf $\tilde{\pi}$. Schließlich folgt aus dem schon von Weil bewiesenen Teil der Weil-Vermutungen mit einigen lokalen Überlegungen, daß $\chi(F_p)$ ($p = $ Primzahl, $F_p = $ Frobenius) für fast alle p eine algebraische Zahl ist, all deren Konjugierte Betrag $p^{\frac{mh}{2}}$ haben. Da $\chi_0(F_p) = p$, ist wie verlangt

$$\sum_{i=1}^{r} m_i d_i = \frac{mh}{2}.$$

§ 5. Endomorphismen

Sei K Zahlkörper, A/K eine abelsche Varietät der Dimension g, l eine Primzahl, $T_l = T_l(A)$ der Tate-Modul, auf dem $\pi = \mathrm{Gal}(\overline{K}/K)$ operiert.

Satz 3. *Die Operation von π auf $T_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ ist halbeinfach.*

Satz 4. *Die Abbildung*

$$\mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Z}_l \to \mathrm{End}_{\pi}(T_l)$$

ist ein Isomorphismus.

Beweis. Die beiden Sätze werden zusammen bewiesen. Es reicht bekanntlich, statt Satz 4 die etwas schwächere Aussage zu beweisen, daß die Abbildung

$$\mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Q}_l \to \mathrm{End}_{\pi}(T_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l)$$

bijektiv ist.

Man darf dann zum Beweis den Grundkörper erweitern, oder A durch eine isogene abelsche Varietät ersetzen. Wir können also annehmen, daß A/K prinzipal polarisiert ist, und daß A sich zu einer semiabelschen Varietät über $\mathrm{Spec}(R)$ fortsetzt. Dann besitzt T_l eine nichtausgeartete schiefsymmetrische Bilinearform. Sei

$$W \subseteq T_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$$

ein π-invarianter maximal isotroper Teilraum. Dem entspricht eine l-divisible Untergruppe $G \subseteq A[l^\infty]$, und die semiabelschen Varietäten $A_n = A/G_n$ tragen wieder prinzipale Polarisationen.

Nach Satz 2 ist $h(A_n) = h(A)$, und nach Satz 1 sind unendlich viele A_n's isomorph.

Wie in [16] folgt daraus, daß W Bild eines Idempotents aus $\mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Q}_l$ ist. Der Rest des Beweises geht genauso wie in [16], und sei daher nur skizziert:

Wähle $a, b, c, d \in \mathbb{Q}_l$ mit $a^2+b^2+c^2+d^2 = -1$. Setze

$$v = \begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix}$$

(entsprechend dem Quaternion $a+bi+cj+dk$), so daß $v \cdot {}^t v = -1$. Wenn W ein beliebiger π-invarianter Teilraum von $T_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ ist, so wendet man obige Überlegungen an auf den maximal isotropen Teilraum

$$W_1 = \{(x, vx) \mid x \in W^4\} \oplus \{(y, -vy) \mid y \in (W^\perp)^4\} \subseteq T_l(A)^8 \otimes_{\mathbb{Z}_l} \mathbb{Q}_l.$$

Korollar 1. *Seien A_1 und A_2 abelsche Varietäten über K. Dann ist*

$$\mathrm{Hom}_K(A_1, A_2) \otimes_{\mathbb{Z}} \mathbb{Z}_l \to \mathrm{Hom}_\pi(T_l(A_1), T_l(A_2))$$

ein Isomorphismus.

Beweis. Satz 4 für $A_1 \times A_2$.

Die L-Reihe eines A's ist bekanntlich definiert als

$$L(A,s) = \prod_v \frac{1}{\det(1-(N_v)^{-s} \cdot F_v \mid T_l(A))} = \prod_v L_v(A,s),$$

wobei das Produkt über fast alle Stellen von K zu erstrecken ist. Die lokalen L-Faktoren $L_v(A,s)$ sind unabhängig von l.

Korollar 2. *Seien A_1, A_2 wie im Korollar 1. Es sind äquivalent*

i) *A_1 und A_2 sind isogen.*
ii) *$T_l(A_1) \otimes_{\mathbb{Z}} \mathbb{Q}_l \cong T_l(A_2) \otimes_{\mathbb{Z}} \mathbb{Q}_l$ als π-Modul.*
iii) *$L_v(s, A_1) = L_v(s, A_2)$ für fast alle Stellen v von K.*
iv) *$L_v(s, A_1) = L_v(s, A_2)$ für alle v.*

Beweis. Die Äquivalenz von i) und ii) folgt aus Satz 4, die von ii) und iii) aus Satz 3 (+Čebotarev), und aus ii) folgt iv) folgt iii) ist trivial.

Korollar 3. *Sei A/K eine abelsche Varietät, $d>0$. Dann gibt es nur endlich viele Isomorphieklassen d-fach polarisierter abelscher Varietäten B/K, so daß für alle l $T_l(A)\cong T_l(B)$.*

Beweis. Die Voraussetzung besagt, daß zu jedem l eine Isogenie vom Grad prim zu l zwischen A und B existiert. Wir dürfen weiter den Grundkörper K zum Beweis erweitern, und dann annehmen, daß A und alle B's sich zu semiabelschen Varietäten über $\mathrm{Spec}(R)$ fortsetzen, und daß es für alle B's eine Isogenie vom Grad $\sqrt{d}$ mit einer prinzipal polarisierten abelschen Varietät gibt. Man kommt dann leicht auf folgende Voraussetzungen:

a) alle B's haben semistabile Reduktion,

b) alle B/K sind prinzipal polarisiert,

c) es gibt ein N, so daß für jede Primzahl l und alle B's Isogenien $\phi\colon A\to B$ existieren, für die die größte l-Potenz in $\mathrm{Grad}(\varphi)$ N teilt.

Die Bemerkung nach Lemma 5 zeigt dann, daß $\exp(2[K:\mathbb{Q}](h(B)-h(A))$ eine rationale Zahl ist, deren Zähler und Nenner durch eine geeignete Potenz von N abgeschätzt werden kann.

Also sind die $h(B)$ beschränkt, und man kann Satz 1 anwenden.

§ 6. Endlichkeitssätze

Satz 5. *Sei S eine endliche Menge von Stellen von K. Dann gibt es nur endlich viele Isogenie-Klassen abelscher Varietäten vorgegebener Dimension über K, welche gute Reduktion außerhalb S haben.*

Beweis. Sei A eine solche abelsche Varietät. Nach den Weil-Vermutungen gibt es für $v\notin S$ nur endlich viele Möglichkeiten für den lokalen L-Faktor $L_v(A,s)$. Wir werden endlich viele Stellen $v_1,\dots,v_r$ konstruieren, so daß zwei A's schon isogen sind, wenn sie an diesen Stellen denselben lokalen L-Faktor haben. Dazu wähle man eine Primzahl l. Nach Lemma 4 existiert eine endliche Galois-Erweiterung $K'\supseteq K$, welche alle außerhalb l und S unverzweigten Körpererweiterungen von K vom Grad $\leqq l^{8g^2}$ umfaßt ($g=\dim(A)$).

Sei $G=\mathrm{Gal}(K'/K)$; und man wähle $v_1,\dots,v_r$ so daß jede Konjugationsklasse in G das Bild eines Frobenius F_v für $v\in\{v_1,\dots,v_r\}$ enthält (Čebotarev). Dann erfüllen $v_1,\dots,v_r$ unsere Forderungen: Seien A_1,A_2 zwei abelsche Varietäten über K, welche denselben lokalen L-Faktor an $v_1,\dots,v_r$ haben.

Sei

$$M\subseteq \mathrm{End}_{\mathbb{Z}_l}(T_l(A_1))\times \mathrm{End}_{\mathbb{Z}_l}(T_l(A_2))$$

die $\mathbb{Z}_l$-Unteralgebra, die vom Bild von π erzeugt wird.

Dann ist M freier $\mathbb{Z}_l$-Modul vom Rang $\leqq 8g^2$, und M besitzt Darstellungen auf $T_l(A_1)$ und $T_l(A_2)$.

Wir müssen zeigen, daß für jedes $m\in M$

$$\mathrm{Spur}(m\,|\,T_l(A_1))=\mathrm{Spur}(m\,|\,T_l(A_2)).$$

Es reicht natürlich, dies für m aus einer $\mathbb{Z}_l$-Modulbasis von M zu zeigen, und nach Voraussetzung gilt die Gleichheit schon, wenn m Bild eines Elements aus

der Konjugationsklasse von F_v ist, für $v \in \{v_1, \ldots, v_r\}$. Wir zeigen, daß diese Bilder M über $\mathbb{Z}_l$ erzeugen. Nach Nakayama reicht es, wenn sie M/lM erzeugen. Dies gilt aber aus folgendem Grund:

Wir haben eine Darstellung

$$\rho\colon \pi \to (M/lM)^* = \text{Einheiten von } M/lM,$$

deren Bild M/lM erzeugt.

Da

$$\#(M/lM)^* \leqq l^{8g^2},$$

faktorisiert ρ über G, und $\rho(\pi)$ ist die Vereinigung der Bilder der Konjugationsklassen der F_v, $v \in \{v_1, \ldots, v_r\}$.

Satz 6 (Shafarevich-Vermutung). *Sei S eine endliche Menge von Stellen von K, $d > 0$. Dann gibt es nur endlich viele Isomorphie-Klassen d-fach polarisierter abelscher Varietäten über K von vorgegebener Dimension, welche außerhalb S gute Reduktion haben.*

Beweis. Nach Satz 5 nehmen wir an, daß die betrachteten abelschen Varietäten B/K alle isogen zu einem festen A/K sind. Wie im Beweis des Korollar 3 zu Satz 4 dürfen wir weiter voraussetzen, daß sich alle B's zu semiabelschen Varietäten über $\mathrm{Spec}(R)$ fortsetzen, und daß $d = 1$ ist. Wir wissen schon, daß

$$\exp(2[K:\mathbb{Q}](h(B) - h(A)))$$

eine rationale Zahl ist. Wir werden ein N konstruieren, so daß Zähler und Nenner dieser rationalen Zahl keine Primfaktoren $l > N$ besitzen, und daß die auftretenden l-Potenzen für Primzahlen $l \leqq N$ beschränkt sind.

Das letztere ist ganz einfach:

Wenn für zwei abelsche Varietäten B_1/K und B_2/K $T_l(B_1)$ und $T_l(B_2)$ als π-Moduln isomorph sind, so existiert nach Satz 4 eine Isogenie vom Grad prim zu l zwischen B_1 und B_2, und l tritt nicht in

$$\exp(2[K:\mathbb{Q}](h(B_1) - h(B_2)))$$

auf.

Es reicht also, wenn es nur endlich viele Isomorphie-Klassen π-invarianter Gitter in $T_l(A) \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ gibt. Dazu sei M_l die von π erzeugte $\mathbb{Z}_l$-Unteralgebra von $\mathrm{End}_{\mathbb{Z}_l}(T_l(A))$. Es folgt dann alles aus der Tatsache, daß $M_l \otimes_{\mathbb{Z}_l} \mathbb{Q}_l$ halbeinfach ist (Satz 3).

Wir kommen nun zur Wahl von N. Dazu sei n das Produkt der Primzahlen l, für welche entweder die Erweiterung $K \supseteq \mathbb{Q}$ in l verzweigt, oder A nicht gute Reduktion an allen Stellen der Charakteristik l hat.

Wähle eine Primzahl p, welche n nicht teilt. Sei wieder

$$\tilde{\pi} = \mathrm{Gal}(\overline{\mathbb{Q}}/\mathbb{Q}) \supseteq \pi = \mathrm{Gal}(\overline{K}/K),$$

und, für $0 \leqq h \leqq 2gm$ ($g = \dim(A)$, $m = [K:\mathbb{Q}]$), sei

$$P_h(T) = \det[T - F_p \mid \Lambda^h(\mathrm{Ind}_{\tilde{\pi}}^{\pi}(T_l(A)))].$$

Dabei ist l eine zu pn prime Primzahl, und F_p bezeichnet den Frobenius an der Stelle p.

Die $P_h(T)$ sind unabhängig von l, haben Koeffizienten in $\mathbb{Z}$, und ihre Nullstellen haben absoluten Betrag $p^{+\frac{h}{2}}$ (Weil-Vermutung oder besser -Satz).

Wir wählen nun $N \geqq 2$ so groß, daß keine Primzahl $l > N$ $P_h(\pm p^j)$ teilt, falls

$$0 \leqq h \leqq 2gm$$
$$0 \leqq j \leqq gm$$
$$j \neq \tfrac{1}{2}h.$$

Außerdem sei $N \geqq np$.

Wir werden zeigen, daß für jede Isogenie

$$\phi\colon B_1 \to B_2$$

von zu A isogenen abelschen Varietäten, deren Grad eine l-Potenz mit einer Primzahl $l > N$ ist, $h(B_1)$ und $h(B_2)$ übereinstimmen. Dies geht ähnlich wie beim Beweis des Satzes 2: Wir dürfen annehmen, daß l den Kern G von ϕ annulliert. Sei

$$V_l = T_l(B_1)/l \cdot T_l(B_1) \cong B_1[l](\overline{K}),$$
$$\tilde{V}_l = \mathrm{Ind}_{\pi}^{\tilde{\pi}}(V),$$
$$W_l = G(\overline{K}) \subseteq V_l$$
$$\tilde{W}_l = \mathrm{Ind}_{\pi}^{\tilde{\pi}}(W_l) \subseteq \tilde{V}_l.$$

Wenn ϕ die Ordnung l^h hat, so operiert $\tilde{\pi}$ auf

$$L = \Lambda^{mh}(\tilde{W}_l) \subseteq \Lambda^{mh}(\tilde{V}_l)$$

via einen Charakter $\chi\colon \tilde{\pi} \to (\mathbb{Z}/l\mathbb{Z})^*$.

Wenn $\varepsilon\colon \tilde{\pi} \to \{\pm 1\}$ den Charakter bezeichnet, mit dem $\tilde{\pi}$ auf $\Lambda^m \mathrm{Ind}_{\pi}^{\tilde{\pi}}(\mathbb{Z})$ operiert, so ist $\chi \cdot \varepsilon^h$ unverzweigt außerhalb l, denn die Trägheitsgruppen der Stellen v von K, welche l nicht teilen, operieren unipotent auf V_l (semistabile Reduktion). Nach der Klassenkörpertheorie ist $\chi \cdot \varepsilon^h$ eine Potenz des zyklotomischen Charakters χ_0. Diese Potenz läßt sich mit Hilfe von [10], Théorème 4.11 (statt der Tateschen Theorie [13]) wie folgt bestimmen:

Sei

$$l^d = \#s^*(\Omega^1_{G/R}),$$
$$0 \leqq d \leqq gm.$$

Dann ist (nach Raynaud) $\chi \cdot \varepsilon^h = \chi_0^{+d}$. Also ist

$$\chi_0^d(F_p) = \pm p^d$$

eine Nullstelle von $P_{mh}(T)$ modulo l, und nach Wahl von N muß $d = \dfrac{hm}{2}$ gelten.

Da wieder

$$h(B_2) - h(B_1) = \log(l)\left(\frac{h}{2} - \frac{d}{m}\right),$$

folgt unsere Behauptung, und es ergibt sich, daß die $h(B)$'s der betrachteten B's beschränkt sind. Damit folgt Satz 6 aus Satz 1.

Korollar 1. *Es gibt nur endlich viele Isomorphie-Klassen glatter Kurven X/K vom Geschlecht $g \geqq 2$, welche außerhalb S gute Reduktion haben.*

Beweis. Torelli.

Satz 7 (Mordell-Vermutung). *Sei X/K eine glatte Kurve vom Geschlecht $g \geqq 2$. Dann ist $X(K)$ endlich.*

Beweis. Dies steht in [9]: Nach eventueller Erweiterung des Grundkörpers gibt es eine unverzweigte Überlagerung vom Grad $m > 2$:

$$\phi\colon X_1 \to X.$$

Lemma 4 liefert einen endlichen Oberkörper $K_1 \supseteq K$, so daß für jedes $x \in X(K)$ $\phi^{-1}(x)$ aus m verschiedenen K_1-rationalen Punkten besteht. Man wähle einen davon aus, etwa $y \in p^{-1}(x)$.

Sei $D = \phi^{-1}(x) - \{y\}$, und A/K_1 die verallgemeinerte Jacobische zu dem Paar (X_1, D). Mit Hilfe von y konstruiert man eine Abbildung von $X_1 - D$ nach A.

Die Multiplikation mit 2 auf A induziert dann eine genaue über D verzweigte Überlagerung $Y(x) \to X_1$, wobei die Kurve $Y(x)$ nur an solchen Stellen v von K_1 schlechte Reduktion haben kann, für die eine der drei folgenden Bedingungen gilt:

a) v teilt 2.
b) X_1 hat schlechte Reduktion in v.
c) ϕ verzweigt in der Faser mod v.

Dies sind nur endlich viele Stellen, und es gibt somit nur endlich viele Möglichkeiten für $Y(x)$.

Dasselbe gilt für die Abbildung $Y(x) \to X_1 \to X$, welche genau über x verzweigt. Es folgt die Behauptung.

Bemerkungen. 1. Man erhält auf diesem Wege auch einen Beweis des Siegelschen Satzes (über ganze Punkte), welcher ohne diophantische Approximation auskommt.

2. Mit Hilfe der Methode aus [16] kann man aus Satz 6 folgern, daß für fast alle Primzahlen l die von π erzeugte Unteralgebra M_l von $\mathrm{End}_{\mathbb{Z}_l}(T_l(A))$ der volle Kommutator von $\mathrm{End}_K(A) \otimes_{\mathbb{Z}} \mathbb{Z}_l$ ist.

Literatur

1. Arakelov, S.: Families of curves with fixed degeneracies. Math. USSR Izvestija **5**, 1277–1302 (1971)
2. Arakelov, S.: An Intersection theory for divisors on an arithmetic surface. Math. USSR Izvestija **8**, 1167–1180 (1974)
3. Baily, W.L., Borel, A.: Compactification of arithmetic quotients of bounded symmetric domains. Ann. of Math. **84**, 442–528 (1966)

4. Deligne, P., Mumford, D.: The irreducibility of the space of curves of a given genus. Publ. math. I.H.E.S. **36**, 75–110 (1969)
5. Faltings, G.: Calculus on arithmetic surfaces. Eingereicht bei Ann. of Math.
6. Faltings, G.: Arakelov's theorem for abelian varieties. Invent. math. **73**, 337–347 (1983)
7. Moret-Bailly, L.: Variétés abéliennes polarisées sur les corps de fonctions. C.R. Acad. Sc. Paris **296**, 267–270 (1983)
8. Namikawa, Y.: Toroidal compactification of Siegel spaces. Lecture Notes in Mathematics, vol. 812. Berlin-Heidelberg-New York: Springer 1980
9. Parshin, A.N.: Algebraic curves over function fields I. Math. USSR Izvestija **2**, 1145–1170 (1968)
10. Raynaud, M.: Schémas en groupes de type $(p, \ldots, p)$. Bull. Soc. Math. France **102**, 241–280 (1974)
11. Szpiro, L.: Sur le théorème de rigidité de Parsin et Arakelov. Astérisque **64**, 169–202 (1979)
12. Szpiro, L.: Séminaire sur les pinceaux de courbes de genre au moins deux. Astérisque **86** (1981)
13. Tate, J.: p-divisible groups. Proceedings of a conference on local fields, Driebergen 1966, pp. 158–183. Berlin-Heidelberg-New York: Springer 1967
14. Tate, J.: Endomorphisms of abelian varieties over finite fields. Invent. math. **2**, 134–144 (1966)
15. Zarhin, J.G.: Isogenies of abelian varieties over fields of finite characteristics. Math. USSR Sbornik **24**, 451–461 (1974)
16. Zarhin, J.G.: A remark on endomorphisms of abelian varieties over function fields of finite characteristics. Math. USSR Izvestija **8**, 477–480 (1974)

Oblatum 8-VI & 10-VII-1983

Zusatz bei der Korrektur

Herr O. Gabber hat mir mitgeteilt, daß der Beweis von Satz 2 nicht ganz korrekt ist. Man erhält nur, daß die Folge $h(A_n)$ stationär wird. Dies reicht für unsere Zwecke.

Invent. math. 73, 367-381 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

A Decomposition Theorem for the Integral Homology of a Variety

J.B. Carrell[1] and R.M. Goresky[2]

University of British Columbia, Department of Mathematics, Vancouver, B.C. Canada V6T 1Y4

Introduction

In [2] it was shown (using the Weil conjectures) that the ℓ-adic Betti numbers of a nonsingular projective algebraic variety X (over an algebraically closed field) having a G_m action can be recovered from those of the components $X_1, \ldots, X_r$ of the fixed point set X^{G_m}. It follows that for a smooth complex projective variety X with algebraic $\mathbf{C}^*$ action, there exist canonical isomorphisms

$$\bigoplus_{j=1}^{r} H_{k-2m_j}(X_j; \mathbf{Q}) \cong \bigoplus_{j=1}^{r} H_{k-2m_j}(X_j^+; \mathbf{Q}) \cong H_k(X; \mathbf{Q}) \tag{1}$$

for $k = 0, 1, 2, \ldots$ although such isomorphisms were not constructed explicitly. Here, $X = \bigcup X_j^+$ is the Bialynicki-Birula plus decomposition of X (see [1]),

$$X_j^+ = \{x \in X \mid \lim_{\lambda \to 0} \lambda \cdot x \in X_j\}$$

and m_j is the fibre dimension of the bundle

$$P_j\colon\ X_j^+ \to X_j; \qquad P_j(x) = \lim_{\lambda \to 0} \lambda \cdot x.$$

This homology basis formula (1) was motivated by a theorem of Frankel [7]. In [5], also motivated by [7], it was shown that (1) is valid for integer coefficients if X is a compact Kaehler manifold with holomorphic $\mathbf{C}^*$ action having fixed points. In this paper we further extend (1) by allowing X to be a (singular) compact complex space admitting a so called good decomposition, which is a generalization of the Bialynicki-Birula decomposition of a smooth complex projective variety with algebraic $\mathbf{C}^*$ action.

[1] Partially supported by the Natural Sciences and Engineering Research Council of Canada

[2] Partially supported by the Natural Sciences and Engineering Research Council of Canada and the Sloan Foundation

Theorem 1. *Suppose* $X = \bigcup_{j=1}^{r} W_j$ *is a "good decomposition"* (§1) *of* X *into locally closed subsets* $W_1, \ldots, W_r$. *Then there are integers* m_j *such that*

$$H_k(X; \mathbf{Z}) \cong \bigoplus_{j=1}^{r} H_{k-2m_j}(W_j; \mathbf{Z}). \tag{2}$$

The isomorphism (2) is explicitly constructed on the chain level (for a certain complex of "transverse" chains) and this leads to insight into certain phenomena such as restriction of (2) to subvarieties of X.

Theorem 1 is a partial generalization to singular spaces (and to integer coefficients) of the "glueing principle" [3] which arises from the Weil conjectures: certain algebraic decompositions of X also decompose the homology and cohomology of X. Unfortunately, not every locally closed algebraic decomposition is "good", as we show in Example 5 (§3) where the homology basis formula (1) fails.

If X is nonsingular, the isomorphism in Theorem 1 preserves (p, q) types as in [5]. However, for singular X we do not know whether this isomorphism is compatible with the weight filtration.

An important source for examples of good decompositions is among the invariant subvarieties of compact Kaehler manifolds with $\mathbf{C}^*$ action. In Theorem 4 of §2 we state a sufficient condition that the plus decomposition of an invariant subvariety be good. It is an open question whether formula (2) holds for each $\overline{X_j^+}$ when X itself is smooth.

The plan of this paper is as follows: §1 concerns the definition of a good decomposition and the statements of the main results. In §2 we describe how a good decomposition can arise from a $\mathbf{C}^*$ action. §3 consists of examples. §4 contains the main technical tool of this paper (the 'plus construction'). The rest of the paper is devoted to the proofs of the theorems.

We would like to mention that Peter Orlik originally suggested to us that the homology basis formula might hold on singular varieties with $\mathbf{C}^*$ action. The final form of this paper has been greatly influenced by A. Bialynicki-Birula and we would like to thank him for several valuable discussions. We would also like to thank Andrew Sommese for his enlightening comments.

§ 1. Results on Good Decompositions

In this section, X will denote a not necessarily irreducible but connected compact complex space.

Definition. A *good decomposition* of X consists of a locally closed decomposition of X, say $X = \bigcup_{i=1}^{r} W_i$, where the subvarieties W_i, called *cells* of the decomposition, satisfy the following:

(1a) for each i, $1 \leqq i \leqq r$, there exists a holomorphic map p_i: $W_i \to X_i$ onto a compact complex space X_i making W_i a topologically locally trivial fibre bundle with affine fibres $\mathbf{C}^{m_i}$;

(1b) each p_i extends meromorphically to $\overline{W_i}$ in the sense that

$$\Gamma_i = \text{topological closure } \{(x, p_i(x)) \mid x \in W_i\} \subset X \times X_i$$

is a closed subvariety of $X \times X_i$ containing the graph of p_i as a Zariski open set.

(1c) if X_i is singular, then it admits an analytic Whitney stratification such that if A is a connected component of a stratum, then $g_i^{-1}(A)$ is irreducible, where $g_i\colon \Gamma_i \to X_i$ denotes the projection on the second factor; and

(1d) there exists a filtration of X by closed subvarieties

$$\emptyset = Z_0 \subset \ldots \subset Z_r = X$$

such that (for some renumbering of $X_1, \ldots, X_r$), $Z_i - Z_{i-1} = W_i$ for $1 \leqq i \leqq r$.

The technical condition (1c) allows us to define homomorphisms

$$\mu_j\colon H_{k-2m_j}(X_j; \mathbf{Z}) \to H_k(X; \mathbf{Z})$$

with the following useful property:

For almost every cycle z in X_j, $\mu_j(c\ell\, z)$ is (3)

is represented by $\overline{p_j^{-1}(z)}$ where $c\ell\, z$ denotes the homology class of z. The isomorphism (2) is the composition

$$\oplus H_{k-2m_j}(W_j; \mathbf{Z}) \to \oplus H_{k-2m_j}(X_j; \mathbf{Z}) \to H_k(X; \mathbf{Z})$$

defined by $\bigoplus_{j=1}^{r} \mu_j \circ p_{j*}$.

Remarks. (a) Suppose $\pi\colon E \to Y$ is a holomorphic affine bundle where Y is smooth and compact. If X is any compactification of E so that π extends meromorphically to X in the sense of (1b), then $X = (X - E) \cup E$ is a good decomposition of X and formula (2) decomposes a homology class in X into a component at ∞ plus a component along Y which comes from the Thom isomorphism. Unfortunately (1b) cannot be relaxed, for the homology basis formula fails for the decomposition of the elliptic Hopf surface $H = (\mathbf{C}^2 - (0,0))/\mathbf{Z}$, where $n \in \mathbf{Z}$ acts via $n \cdot (w,z) = (2^n w, 2^n z)$, as $T \cup (H - T)$ where T is the torus $T = \{(w,0)\} \subset H$. $H - T$ is made into the trivial line bundle on T via the map $\pi(w,z) = z$ but π does not extend meromorphically to H.

(b) Condition (1d) is necessary in order to avoid certain spaces such as the union of three $\mathbf{P}^1$'s joined together so as to form a ring. This space can be decomposed as $\mathbf{C} \cup \mathbf{C} \cup \mathbf{C}$ but (2) obviously fails although (1a, b, c) hold. Of course (1d) implies that W_1 is closed and $\overline{W_r} = X$. In the terminology of $\mathbf{C}^*$ actions, X_1 is called the *sink* of X and X_r is called the *source* of X (with respect to the plus decomposition).

(c) Formula (1) is also known to hold for algebraic $\mathbf{C}^*$ actions on a complete algebraic manifold X. Condition (1d) may fail however for such actions as was noticed first in [11] using torus embeddings (see also [12]) and later in [6] and [19] using quite different methods.

Theorem 1 has several interesting corollaries.

Corollary 1. *Let X have a good decomposition, let $\{Z_j\}$ denote a filtration satisfying* (1d), *and let Y be a closed subvariety of X such that, for some i, $Y - (Y \cap Z_i)$ is a union of certain cells W_j, say*

$$Y - (Y \cap Z_i) = W_{j_1} \cup \ldots \cup W_{j_s}$$

Then $Y = (Y \cap Z_i) \cup W_{j_1} \cup \ldots \cup W_{j_s}$ is a good decomposition of Y. In particular, if $J = \{j_1, \ldots, j_s\}$, then for every k,

$$H_k(Y;\mathbf{Z}) = H_k(Y \cap Z_i;\mathbf{Z}) \oplus [\bigoplus_{j \in J} H_{k-2m_j} X_j;\mathbf{Z})]$$

If Y is, in fact, a union of cells in X, e.g. if $Y = Z_i$ for some i, then $H.(Y;\mathbf{Z})$ is a direct summand in $H.(X;\mathbf{Z})$.

Proof. In fact, for the first part one only needs to show that the decomposition of Y is filterable (i.e. satisfies (1d)), and this follows from induction using the fact that X satisfies (1d). For the second part, just use Theorem 1. For the last assertion, use the fact that for every a, the diagram

$$\begin{array}{ccc} \bigoplus_{j \in J} H_{k-2m_j}(X_j;\mathbf{Z}) & \longrightarrow & H_k(Y;\mathbf{Z}) \\ \downarrow & & \downarrow \\ \bigoplus_{a=1}^{s} H_{k-2m_a}(X_a;\mathbf{Z}) & \longrightarrow & H_k(X;\mathbf{Z}) \end{array}$$

is commutative with respect to the natural maps, which follows from (3). □

There is a useful relative version of Theorem 1 for pairs (X, Y) where X has a good decomposition $X = \bigcup W_i$ and Y is a closed subvariety of X that is a union of certain W_i, say $Y = W_{j_1} \cup \ldots \cup W_{j_s}$ as in Corollary 1.

Corollary 2. *For each k,*

$$H_k(X, Y;\mathbf{Z}) \cong \bigoplus_{i \notin J} H_{k-2m_i}(X_i;\mathbf{Z}),$$

where $J = \{j_1, \ldots, j_s\}$.

The next theorem was proved in [5] for $\mathbf{C}^*$ actions. We state it here for completeness.

Theorem 2. *Let X be a compact Kaehler manifold with good decomposition $X = \bigcup_{i=1}^{r} W_i$ where all X_i are smooth. Let m_i denote fibre $\dim_{\mathbf{C}} p_i$. Then:*

(1) *for all p and q,*

$$H^p(X, \Omega^q) \cong \bigoplus_{i=1}^{r} H^{p-m_i}(X_i, \Omega^{q-m_i});$$

(2) *there exists an exact sequence*

$$0 \to K \to \mathrm{Pic}(X) \to \mathrm{Pic}(X_r) \to 0,$$

where X_r denotes the source of X and K denotes the $\mathbf{Z}$-module of divisors generated by the $\overline{W_i}$ of codimension one in X,

(3) $\mathrm{Index}(X) = \sum_{i=1}^{r} \mathrm{Index}(X_i)$,

and

(4) $\pi_1(X) = \pi_1(X_r)$.

One of the pleasant features of good decompositions is that they are well behaved with respect to certain birational transformations.

Theorem 3. *Let f: $X \to X'$ be a holomorphic map of compact complex spaces where X has a good decomposition $X = \bigcup W_i$, $1 \leqq i \leqq r$. Suppose Y is a closed subvariety of X so that $Y - (Y \cap W_1)$ is a union of cells of X, say $Y = (Y \cap W_1) \cup W_2 \cup \ldots \cup W_s$. Then if f: $X - Y \to X' - f(Y)$ is an isomorphism, the decomposition $X' = W_1' \cup W_2' \cup \ldots \cup W_{r-s}'$, where $W_1' = f(W_1 \cup Y)$ and $W_i' = f(W_{s+i})$ for $1 \leqq i \leqq r-s$, is a good decomposition of X'.*

Corollary 3. (i) *Let f: $X \to X'$ be a holomorphic surjective map blowing down a subvariety Y of the sink W_1 of X. Then X' has a good decomposition $W_1' \cup W_2' \cup \ldots \cup W_r'$ where each $W_i' = f(W_i)$.*

(ii) *If the subvariety of Y of X is a union of cells from X, say $Y = W_1 \cup \ldots \cup W_s$, then there exists a short exact sequence*

$$0 \to H_k(Y) \xrightarrow{\alpha} H_k(X) \oplus H_k(f(Y))$$

$$\xrightarrow{\beta} H_k(X') \to 0$$

where $\alpha(x) = (i_*(x), 0)$ and $\beta(x, y) = f_*(x) + j_*(y)$, i: $Y \to X$ and j: $f(Y) \to X'$ denoting the inclusions. Moreover, α and β split over **Z**.

This sequence bears some similarity to the exact homology sequence of a birational morphism [10].

§2. Good Decompositions Arising from C* Actions

In this chapter we assume X is a closed invariant subvariety of some compact Kaehler manifold Y with a holomorphic **C*** action which has a nonempty fixed point set.

Let $X_1, \ldots, X_r$ denote the components of the fixed point set of X. The *Bialynicki-Birula plus decomposition* of X is the union

$$X = \bigcup_{j=1}^{r} X_j^+ \quad \text{where } X_j^+ = \{x \in X \mid \lim_{\lambda \to 0} \lambda \cdot x \in X_j\}.$$

This decomposition exists by [18].

Definition. The **C*** action on X (which is induced from that on Y) is *good as* $\lambda \to 0$ if the associated Bialynicki-Birula plus decomposition is good.

Remark. Conditions (1b) and (1d) are always satisfied by the Bialynicki-Birula decomposition - see [4, 5] for the basic properties of this decomposition.

For good actions, $H_i(X_j) = H_i(X_j^+)$ so in this context, Theorem 1 becomes

Theorem 1'. *Suppose X has a good* **C*** *action with fixed point components $X_1, X_2, \ldots, X_r$. Then for all k,*

$$H_k(X) \cong \bigoplus_{j=1}^{r} H_{k-2m_j}(X_j)$$

where m_j is the fibre dimension of the affine space bundle $X_j^+ \to X_j$.

Definition. The $\mathbf{C}^*$ action on X is *singularity preserving* as $\lambda \to 0$ if there exists an equivariant Whitney stratification of X such that for every stratum A, and for every $x \in A$, the limit $x_0 = \lim_{\lambda \to 0} \lambda \cdot x$ is also in A.

The role of the singularity preserving condition is in the following lemma.

Lemma 1. *Suppose X admits a Whitney stratification which is singularity preserving as $\lambda \to 0$. Then condition* (1a) *is satisfied, i.e. p_j: $X_j^+ \to X_j$ is a locally trivial affine space bundle.*

Theorem 4. *Suppose X admits a singularity preserving analytic Whitney stratification. Suppose that for any stratum A of X and for any fixed point component X_j, either $A \cap X_j = \emptyset$ or $\overline{A} \cap \overline{X_j^+} = \overline{(A \cap X_j)^+}$. Then the nonempty $A \cap X_j$ are the strata of a Whitney stratification of $X^{\mathbf{C}^*}$ which makes the $\mathbf{C}^*$ action good.*

Of course, one can also consider actions that are good as $\lambda \to \infty$, i.e. those for which the minus decomposition is good. For singular X, an action that is good as $\lambda \to 0$ may or may not be good as $\lambda \to \infty$ - see Examples (1), (2), (4).

§3. Examples of Good Actions and Other Good Decompositions

Example (1). A good action with no minus isomorphism. Let E be the cone (in $\mathbf{P}^3$) with vertex $p = (0,0,0,1)$ over an elliptic curve $Y \subset \mathbf{P}^2$. E is invariant under the $\mathbf{C}^*$ action $\lambda \cdot (z_0, z_1, z_2, z_3) = (z_0, z_1, z_2, \lambda z_3)$ and $E^{\mathbf{C}^*} = Y \cup \{p\}$. E is the Thom space of a negative line bundle on Y and the plus isomorphism is the Thom isomorphism. In particular, $H_1(E) = 0$ which shows there is no minus isomorphism.

Example (2). A good action with a singular source. The $\mathbf{C}^*$ action on $\mathbf{C}^5$ which is given by

$$\lambda \cdot (z_1, z_2, z_3, z_4, z_5) = (z_1, z_2, \lambda z_3, z_4, z_5)$$

takes subspaces into subspaces and thus determines an action on the Grassman manifold $G_2(\mathbf{C}^5)$. The Schubert variety

$$X = \Omega(3,5) = \{P \in G_2(\mathbf{C}^5) \mid \dim_{\mathbf{C}}(P \cap \mathbf{C}^3) \geqq 1\}$$

is invariant. X may be stratified with one singular stratum

$$\Sigma = \{P \in G_2(\mathbf{C}^5) \mid P \subset \mathbf{C}^3\}$$

and the action is singularity preserving as $\lambda \to 0$. $X^{\mathbf{C}^*}$ has two components

$$X_1 = \{P \in X \mid P \subset \langle e_1, e_2, e_4, e_5 \rangle \text{ and } \dim_{\mathbf{C}}(P \cap \langle e_1, e_2 \rangle) \geqq 1\},$$
$$X_2 = \{P \in X \mid P \supset \langle e_3 \rangle \text{ and } \dim_{\mathbf{C}}(P \cap \langle e_1, e_2, e_4, e_5 \rangle) = 1\}$$

where e_i denote the standard basis vectors in $\mathbf{C}^5$ and $\langle\ \rangle$ denotes the span. Then $X_1 \cong \Omega(2,4)$, is singular, and has Betti numbers $1,0,1,0,2,0,1$. $X_2 \cong \mathbf{P}^3$

and has Betti numbers $1,0,1,0,1,0,1$. $\Sigma \cap X_1 \cong$ a point and $\Sigma \cap X_2 \cong \mathbf{P}^1$. $X_1^+ \to X_1$ is a bundle with fibre $\cong \mathbf{C}^2$ while $X_2^+ = X_2$. X_1 can be stratified with one singular stratum (consisting of the singular point) and this renders the action good as $\lambda \to 0$.

From Theorem 1, the Betti numbers of X can be read off from the following table:

dim	0	1	2	3	4	5	6	7	8	9	10
β_i	1	0	1	0	1+1	0	1+1	0	2	0	1

$X_2^- \to X_2$ is not a fibration and the homology basis formula fails if we use the minus homomorphism.

Example (3). The Springer fibre for $SL(n, \mathbf{C})$ ([17, 20]). Let $F(V)$ denote the flag manifold of $V = \mathbf{C}^n$ and let $u \in SL(n, \mathbf{C})$ be a fixed unipotent matrix. The variety X_u in $F(V)$ consisting of flags fixed by u admits a good decomposition into affines, by [17]. X_u also admits $\mathbf{C}^*$ actions induced by the one parameter subgroups of the centralizer of u. If the Young diagram associated to u as in [17] has r rows with λ_i cases in the i-th row and $\lambda_1 \geqq \lambda_2 \geqq \ldots \geqq \lambda_r$, then a $\mathbf{C}^*$ action on X_u is obtained as follows: fix integers $\alpha_1, \ldots, \alpha_r$ and for all i such that

$$\sum_{j=1}^{s-1} \lambda_j \leqq i \leqq \sum_{j=1}^{s} \lambda_j$$

set

$$\lambda \cdot e_i = \lambda^{\alpha_s} e_i \qquad (s = 1, \ldots, r).$$

This action defines a one parameter subgroup of the centralizer of u, hence a $\mathbf{C}^*$ action on X_u. Unfortunately, the affine decomposition of [17] is only the plus decomposition associated to a $\mathbf{C}^*$ action with isolated fixed points in special cases, as the following lemma shows.

Lemma. (a) *If $\alpha_1, \ldots, \alpha_r$ are all distinct, then $X_u^{\mathbf{C}^*}$ is an isolated set having $\binom{n}{\lambda_1, \ldots, \lambda_r}$ points.*

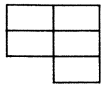
(b) *X_u has a good action with isolated fixed points precisely in the following cases:* (i) *$r = 1$ or $r = 2$, and* (ii) *r is arbitrary but $\lambda_2 = \lambda_3 \ldots = \lambda_r = 1$ and $\alpha_1 > \alpha_2 > \ldots > \alpha_r$. In the case $r = 1$ or $r = 2$, the action is good both as $\lambda \to 0$ and $\lambda \to \infty$. In case* (ii), *the action is only good as $\lambda \to 0$.*

(c) *If the diagram of u has more entries in the first column than the second, then any $\mathbf{C}^*$ action with $\alpha_1 = \alpha_2 = \ldots = \alpha_{r-1} > \alpha_r$ is good as $\lambda \to 0$ (but fails to have isolated fixed points).*

Outline of the Proof. The first part of part (a) follows from the fact that the maximal torus in the centralizer of u has the same fixed point set as the one parameter group being considered, so that $X_u^{\mathbf{C}^*}$ must be finite. To see the second part of (a) it is easiest to consider an example. Suppose that the diagram of u is  Then we may suppose that u has been chosen so that (u

$-I$): $e_2 \to e_1 \to 0$, $e_4 \to e_3 \to 0$. Hence the flags fixed by both u and $\mathbf{C}^*$ are of the form $(F_0, F_1, F_2, F_3, F_4)$ where F_i is generated by i of the vectors e_1, e_2, e_3, e_4 and $(u-I)F_i \subset F_{i-1}$. Therefore the fixed flags can be uniquely labelled by the permutations of 1, 2, 3, 4 where 1 preceeds 2 and 3 preceeds 4. The fixed flags are (1 2 3 4), (1 3 2 4), (1 3 4 2), (3 1 2 4), (3 1 4 2), (3 4 1 2). Obviously, in the general case, this procedure leads to exactly $\binom{n}{\lambda_1, \ldots, \lambda_r}$ fixed points.

The proofs of (b) and (c) of the lemma are done by induction, using the construction of Spaltenstein in [17, Lemma, p. 453] which is, in fact, $\mathbf{C}^*$ equivariant. As soon as the diagram for u contains the sub-diagram



(4)

any action on X_u with isolated fixed points fails to be good both as $\lambda \to 0$ and $\lambda \to \infty$. The action on X_u associated to (4) with $\alpha_1 > \alpha_2 > \alpha_3$ has a singular three dimensional plus cell and a singular two dimensional minus cell. In the case (b) where u has two Jordan blocks ($r=2$) or is of one hook type, the plus decomposition decomposes X into a union of affines (if the fixed points are isolated), giving us another proof of a special case of the result of Spaltenstein [17].

Example (4). Blowing down a good action. Let $\mathbf{F}_n$ denote the rational ruled surface $\mathbf{P}(\mathbf{C} \oplus \mathcal{O}(n))$, and let C denote the exceptional curve with $C^2 = -n$. $\mathbf{F}_n$ is a compactification of $\mathbf{C}^* \times \mathbf{C}^*$, i.e. a torus embedding. The map f: $\mathbf{F}_n \to \mathbf{F}'_n$ blowing C down to a point is equivariant with respect to any $\mathbf{C}^*$ action on $\mathbf{F}_n$ and by Theorem 3 any $\mathbf{C}^*$ action on $\mathbf{F}_n$ induces a good one on $\mathbf{F}'_n$. Notice that if $\mathbf{F}_n^{\mathbf{C}^*}$ is finite, then the action on $\mathbf{F}'_n$ is not singularity preserving either as $\lambda \to 0$ or $\lambda \to \infty$, although it is good in both directions.

The rest of this section is devoted to showing that some condition such as the graph condition (1c) is needed for Theorem 1 to hold.

Example (5). In this example we consider a variety X with a decomposition satisfying (1a), (1b), (1d) but not (1c) for which the homology basis formula is false. Let Y denote the nodal curve or "pinched torus" obtained by identifying two points of $\mathbf{P}^1$. Thus the singularity of Y is a normal crossing at the singular point y_0 of Y. Let X denote the space obtained from $Y \times \mathbf{P}^1$ by blowing up (y_0, ∞) where $\infty = [0,1]$. Let f: $X \to \mathbf{P}^1$ be the composition $X \to Y \times \mathbf{P}^1 \to \mathbf{P}^1$ and define a decomposition of X by

$$W_1 = f^{-1}(\infty)$$
$$W_2 = f^{-1}(\mathbf{P}^1 - \infty).$$

W_2 is the trivial line bundle on Y and $W_1 = Z_1$ consists of three copies of $\mathbf{P}^1$ joined at points so as to form a cycle (i.e. ring). Condition (1c) fails for the projection $W_2 \to Y$ for $g_2^{-1}(y_0)$ is a union of three $\mathbf{P}^1$'s meeting at a point. The homology basis formula fails because $H_3(W_1) \oplus H_1(W_2) = \mathbf{Z}$, while $H_3(X) = 0$. In fact the vanishing of $H_3(X)$ can most easily be seen by observing that X

admits a good action. Namely the action $\lambda\cdot(y,[z_0,z_1])=(y,[\lambda z_0,z_1])$ on $Y\times\mathbf{P}^1$ induces a good action on X whose source is the $\mathbf{P}^1\subset W_1$ corresponding to the desingularization of Y, whose sink is $Y\times[1,0]$, and whose remaining fixed component is an isolated fixed point whose plus cell is $\mathbf{C}$. The plus cell over the source is $\mathcal{O}(-2)$. Since $H^1(\mathbf{P}^1)=0$, $H_3(X)=0$ too.

§4. Preliminary Remarks on the Plus Construction

The main step in the proof of Theorem 1 is the construction of wrong way homomorphisms $H_k(X_j;\mathbf{Z})\to H_{k+2m_j}(\overline{W}_j;\mathbf{Z})$. The definition of these maps, when X_j is smooth, goes as follows ($\mathbf{Z}$ coefficients understood). We have the sequence

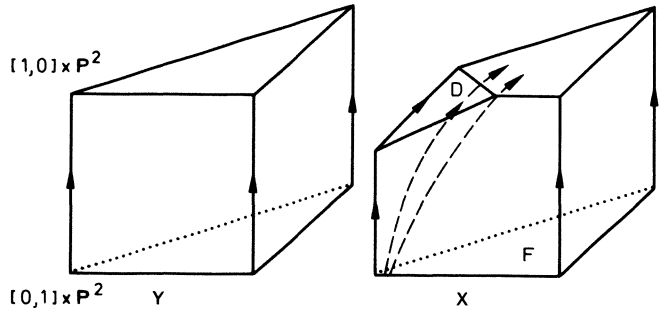
$$H_k(X_j)\xrightarrow{PD}H^{2b_j-k}(X_j)\to H^{2b_j-k}(\tilde{\Gamma}_j)$$

$$\xrightarrow{PD}H_{k+2m_j}(\tilde{\Gamma}_j)\to H_{k+2m_j}(\overline{W}_j)$$

where PD stands for Poincaré duality, $\tilde{\Gamma}_j$ denotes a resolution of singularities of Γ_j, and $b_j=\dim_{\mathbf{C}}X_j$. However, our aim is to describe these wrong way maps on the chain level even when X_j is singular. To do this, we describe a sub-complex $C_*^T(X_j)$ of the complex of singular chains on X_j which admits a wrong way chain map $C_i^T(X_j)\to C_{i+2m_j}(\Gamma_j)$. If X_j is smooth, then the first step will be to filter X_j by subvarieties $X_{jk}=\{x\in X_j|\dim_{\mathbf{C}}g_j^{-1}(x)\geqq k+m_j\}$. The smoothness implies that the codimension of X_{jk} in X_j is at least $k+1$. The chains that lift nicely under the plus construction to Γ_j are those "transverse" to the stratification $\{X_{jk}-X_{jk+1}\}$ of X_j.

There are two reasons such transverse cycles must be used. First, not all cycles on X_j lift to cycles on Γ_j (see Example (5)). The other reason is that homologous cycles in X_j may lift to nonhomologous cycles in $\overline{W}_j$. Such behavior is due to the existence of envelopes at infinity, as the following example shows.

Example (6). Let Y denote $\mathbf{P}^1\times\mathbf{P}^2$. Blow up the point $[1,0]\times[1,0,0]$ on Y to obtain X. The $\mathbf{C}^*$ action $(\lambda,[z_0,z_1],[y_0,y_1,y_2])\to([z_0,\lambda z_1],[y_0,y_1,y_2])$ on Y lifts to X, and the plus decomposition associated to this action on X is a good decomposition. Let $D\subset X$ denote the exceptional divisor. D is the "envelope" of limits of orbits. Take $F=[0,1]\times\mathbf{P}^2$. Then $D\subset\overline{F^+}-F^+$ but D is not contained in the fiberwise compactification of F^+. In fact D is isomorphic to a subvariety of the fiber of the projection $g\colon\Gamma\to F$ over $[0,1]\times[1,0,0]$ as in Example (5) where Γ is the closure of the graph defined in §1. The naive definition of the plus homomorphism is to lift the cycle ξ in F to $\overline{\xi^+}$ and show that this yields a cycle in X. To see why homologous cycles in F may lift to nonhomologous cycles, in X, let $X=[0,1]\times[1,0,0]$ and let y be any other point of F. Then $\overline{y^+}\sim\overline{x^+}+z$, where z is a nontrivial 2-cycle in D, so $\overline{x^+}$ and $\overline{y^+}$ are nonhomologous. Note that in the plus isomorphism for X, the homology class of z comes from the sink of X; in the diagram below, this class arises by



intersecting D and the top face of X. Instead, we will show in §5 how to lift transverse cycles (e.g. y) in F to X.

§5. The Plus Construction

Throughout this section, W_j will denote one cell of a good decomposition of X, and A a connected stratum of X_j satisfying (1c). Homology groups are understood to have integer coefficients. We will define the homomorphism $H_i(X_j)\to H_{i+2m}(X)$, where m is the complex dimension of the fibre of the affine space bundle p: $W_j\to X_j$. This construction associates to *almost* every oriented cycle $\xi\subset X_j$ the oriented cycle $\tilde{\xi}=\pi_* g^{-1}(\xi)$ where g: $\Gamma\to X_j$ and π: $\Gamma\to X$ are the projections and Γ denotes the graph Γ_j of (1b). The surprising fact is that $\tilde{\xi}$ is a cycle, i.e. that $\partial\tilde{\xi}=0$.

Let $\Gamma_A=g^{-1}(\bar{A})$. Since the projection $\Gamma_A\to\bar{A}$ is analytic, there is a proper subvariety $\Sigma_A\subset\bar{A}$ such that $g^{-1}(\bar{A}-\Sigma_A)\to\bar{A}-\Sigma_A$ is a topologically locally trivial fibre bundle whose fibre is an analytic compactification of $\mathbf{C}^m$. Define

$$A_0=\Sigma_A,$$
$$A_k=\{x\in\bar{A}\,|\dim_{\mathbf{C}} g^{-1}(x)\geqq m+k\}\quad\text{for } k\geqq 1.$$

Then we have the following key fact based on the graph condition.

Proposition. *For any $k\geqq 0$ the complex codimension of A_k in A is at least $k+1$.*

Proof. Since A is connected, Γ_A is irreducible. If the proposition were false for some k, then $g^{-1}(A_k)$ would be a proper subvariety of codimension 0 in Γ_A, which is impossible. □

Remark. Γ_A is, in fact, the closure of the graph of $p_j|p_j^{-1}(A)$ in $X\times X_j$.

Let $C_*(X_j)$ be the complex of subanalytic chains on X_j (see [9]). For each chain $\xi\in C_i(X_j)$, let $|\xi|$ denote the support of ξ.

Definition. A subanalytic chain $\xi\in C_i(X_j)$ is *transverse* if, for every stratum A of X_j and for every $k\geqq 0$,

$$\dim_{\mathbf{R}}(|\xi|\cap A_k)\leqq i-2k-2$$

and

$$\dim_{\mathbf{R}}(|\partial\xi|\cap A_k)\leqq i-2k-3$$

where

A_k is defined in §4.

The subcomplex of transverse chains will be denoted $C_*^T(X_j)$.

Lemma. *The inclusion $C_*^T(X_j)\to C_*(X_j)$ induces an isomorphism on homology.*

Proof. For any subanalytic cycle ξ there is a triangulation of X_j such that $|\xi|$ and each A_k are subcomplexes. By [14], ξ can be approximated by a P.L. cycle ξ' such that for each stratum A, $|\xi'|\cap A$ is in general position with respect to each $A_k\cap A$. Thus, ξ' is "transverse" and it may be further approximated by a nearby subanalytic cycle ξ'' which is still transverse. Therefore every homology class has a transverse representative. A relative version of this argument shows that any two transverse representations of a given homology class can be made homologous by a transverse chain. □

Definition. For any transverse chain $\xi\in C_i^T(X_j)$, let $g^*(\xi)\in C_{i+2m}(\Gamma)$ denote the chain whose support is $|g^*(\xi)|=g^{-1}(|\xi|)$ and whose orientation is the product of the orientation of ξ with the canonical orientation of the fibres of g.

Proposition. *If $\xi\in C_i^T(X_j)$ then $\partial g^*(\xi)=g^*(\partial\xi)$ so g^* induces a map on homology (which we also denote by g^*: $H_i(X_j)\to H_{i+2m}(\Gamma)$).*

Proof. We shall denote by Σ the set of nongeneric points in the image of g,

$$\Sigma=\bigcup_A \Sigma_A$$

where the union runs over all strata A of X_j. It follows that $g^{-1}(X_j-\Sigma)\to(X_j-\Sigma)$ is a topological fibre bundle with a compact oriented fiber of real dimension $2m$. But $g^{-1}(|\xi|)$ is an oriented subanalytic subset of Γ with real dimension $i+2m$ and

$$\dim_{\mathbf{R}} g^{-1}(|\xi|\cap\Sigma)\leqq i+2m-2,$$
$$\dim_{\mathbf{R}} g^{-1}(|\partial\xi|\cap\Sigma)\leqq i+2m-3$$

(which can be verified by counting $\dim g^{-1}(|\xi|\cap A_k)$. Therefore $\partial g^*(\xi)$ has no contribution from $g^{-1}(|\xi|\cap\Sigma)$. This implies $\partial g^*(\xi)=g^*(\partial\xi)$. □

Definition. The homomorphism

$$\mu_j\colon\ H_i(X_j)\to H_{i+2m}(X)$$

is the composition of the homomorphisms

$$g^*\colon\ H_i(X_j)\to H_{i+2m}(\Gamma)$$

and

$$H_{i+2m}(\Gamma)\to H_{i+2m}(X)$$

which is induced by projection to the second factor.

It is clear from the construction that $|\xi| = \text{closure } p^{-1}(|\xi|)$ whenever ξ is a transverse chain in F. It is also clear that the homomorphism μ_j factors through $H_{i+2m}(\overline{W_j})$.

Our construction can be used to produce wrong way homomorphisms in other circumstances. For example, if $g: Z \to W$ is a surjective morphism between normal compact algebraic varieties such that all the fibres of g have dimension k, then g induces $g^*: H_i(W) \to H_{i+2k}(Z)$. The appropriate dimension estimates follow from the isomorphism $H_i(X) \cong IH_i^{\bar{t}}(X)$ of [8].

§6. Proof of Theorem 1

The decomposition is filterable, i.e., there exists a filtration of X by closed invariant subvarieties,

$$\emptyset = Z_0 \subset Z_1 \subset \ldots \subset Z_r = X$$

such that $Z_i - Z_{i-1} = W_i$. Consider the following exact sequences, where α, β, γ are the plus homomorphisms of the previous section:

$$\begin{array}{ccccccccc}
H_{q+1}(Z_k, Z_{k-1}) & \xrightarrow{\partial} & H_q(Z_{k-1}) & \longrightarrow & H_q(Z_k) & \longrightarrow & H_q(Z_k, Z_{k-1}) & \xrightarrow{\partial} & \\
 & & \uparrow{\scriptstyle \alpha} & & \uparrow{\scriptstyle \beta} & \nwarrow{\scriptstyle \gamma} & \uparrow{\scriptstyle \delta} & & \\
0 & \longrightarrow & \bigoplus_{j<k} H_{q-2m_j}(X_j) & \longrightarrow & \bigoplus_{j\leqq k} H_{q-2m_j}(X_j) & \longrightarrow & H_{q-2m_k}(X_k) & \longrightarrow & 0
\end{array}$$

The decomposition $Z_{k-1} = W_1 \cup \ldots \cup W_{k-1}$ is good. Hence by induction on k, α is an isomorphism. We must show δ is an isomorphism. Now $Z_k - X_k$ deformation retracts to Z_{k-1} because:

(a) X_k has a tubular neighbourhood V in Z_k so $Z_k - X_k$ deforms to $Z_k - V$,

(b) one can use the affine bundle structure on W_k to deform $Z_k - V$ into some regular neighbourhood U of Z_{k-1} (by compactness), and

(c) U can be chosen to deformation retract to Z_{k-1}. Thus one sees that δ is the Thom isomorphism since

$$H_q(Z_k, Z_{k-1}) \cong H_q(Z_k, Z_k - X_k) \cong H_q(W_k, W_k - X_k)$$

by excision. The diagram commutes, so $\partial = 0$. Theorem 1 now follows from the five lemma and induction on k. □

To prove Corollary 2, consider the sequences

$$\begin{array}{ccccccccc}
0 & \longrightarrow & \bigoplus_{j\in J} H_{k-2m_j}(X_j; \mathbf{Z}) & \longrightarrow & \bigoplus_{j} H_{k-2m_j}(X_j; \mathbf{Z}) & \longrightarrow & \bigoplus_{j\notin J} H_{k-2m_j}(X_j; \mathbf{Z}) & \longrightarrow & 0 \\
 & & \downarrow{\scriptstyle \alpha} & & \downarrow{\scriptstyle \beta} & & \downarrow{\scriptstyle \gamma} & & \\
0 & \longrightarrow & H_k(Y; \mathbf{Z}) & \xrightarrow{i_*} & H_k(X; \mathbf{Z}) & \to & H_k(X, Y; \mathbf{Z}) & \longrightarrow & 0
\end{array}$$

where the maps α and β are the isomorphisms of Theorem 1 and the homomorphism γ is induced by an obvious relativization of the plus homomorphism. Now the top sequence is exact and from this and the fact that the bottom sequence contains part of the long homology sequence of the pair (X, Y), it follows that the bottom sequence is also exact, due to the fact that ι_*: $H_k(Y; \mathbf{Z}) \to H_k(X; \mathbf{Z})$ is injective for all k. Hence, by the five lemma, γ is an isomorphism. □

§ 7. Proof of Theorem 3

Note first that the decomposition $X' = f(Y \cup W_1) \cup f(W_{s+1}) \cup \ldots \cup f(W_r)$ consists of locally closed affine bundles provided we let $f(Y \cup W_1)$ be the sink of X'. Moreover, since $Y \cup W_1$ is closed and a union of plus cells in X, one can show directly that the decomposition of X' is filterable. Hence conditions (1a, b, d) are satisfied. To show (1c), let us suppose for simplicity that $W = W_t$, $W' = f(W_t)$ and $Z = X_t$. Also, let Γ and Γ' denote, respectively, the closures of the graphs of p: $W \to Z$ and p': $W' \to Z$. We must verify that $\Gamma' | A$ is irreducible for any connected component of a stratum for the Whitney stratification of Z satisfying (1c) for Γ. Now the point of condition (1c) is that for any such A,

$$\Gamma | \overline{A} = \overline{\text{graph } (p|A)} \subset X \times Z \tag{5}$$

For the right hand side of (5) is always an irreducible component of $\Gamma | \overline{A}$. Using the isomorphism f_*: graph $(p|A) \to$ graph $(p'|A)$ induced by f, we obtain a morphism f_* of $\Gamma | \overline{A}$ onto an irreducible variety containing $\Gamma' | A$. Hence $\Gamma' | A$ must also be irreducible. □

Corollary 3(i) follows immediately from Theorem 3. The proof of Corollary 3(ii) is identical with the proof of Corollary 2, so it will be omitted.

§ 8. Proof of Theorem 4

Suppose X has a Whitney stratification that is singularity preserving as $\lambda \to 0$, and let F denote a fixed point component. To show that for any $x \in X$, $p^{-1}(x)$ is equivariantly biholomorphic to some $\mathbf{C}^m$, let $x \in A \cap F$ for some stratum A of X. By the singularity preserving condition, it follows that $x^+ = \{y \in X: y_0 = x\} \subset A$. But as A is nonsingular, the result follows from [4]. The linearized $\mathbf{C}^*$ action on $T_x A$ gives rise to subspaces $(T_x A)^-$, $(T_x A)^0$, $(T_x A)^+$ generated by all vectors of negative, zero, or positive weight. These subspaces may be locally identified with $T_x(x^- \cap A)$, $T_x(F \cap A)$ and $T_x(x^+ \cap A)$ by exponentiation. Therefore $F \cap A$ is a manifold and the fibres of $p^{-1}(F \cap A) \to F \cap A$ form a smooth affine space bundle as in [4] or [1].

Now suppose $y \in \overline{A} \cap F$. We shall show $\dim_{\mathbf{C}}(y^+) = \dim_{\mathbf{C}}(x^+)$. By the upper semicontinuity principle applied to $\overline{A} \cap F$, we have $\dim_{\mathbf{C}} x^+ \leqq \dim_{\mathbf{C}} y^+$. To get

the reverse inequality, suppose B is the stratum containing y and choose a sequence $x_i \in A \cap F$ such that $x_i \to y$. By choosing a subsequence if necessary we may assume the $T_{x_i} A$ converge to some limiting plane τ and the subspaces $(T_{x_i} A)^-$, $(T_{x_i} A)^0$, and $(T_{x_i} A)^+$ converge to subspaces of τ. By continuity, τ is $\mathbf{C}^*$ invariant. Since the weights are integers, $\lim(T_{x_i} A)^0 \subset \tau^0$ and $\lim(T_{x_i} A)^{\mp} \subset \tau^{\mp}$. Thus these inclusions are equalities. Whitney's condition A says $T_y B \subset \tau$, so $(T_y B)^+ \subset \tau^+$. Hence

$$\dim_{\mathbf{C}} y^+ = \dim_{\mathbf{C}} (T_y B)^+ \leqq \dim_{\mathbf{C}} \tau^+ = \dim_{\mathbf{C}} (T_x A)^+ = \dim_{\mathbf{C}} x^+$$

which establishes the reverse inequality. In fact, we have shown that $(T_y B)^+ = \tau^+ = \lim(T_{x_i} A)^+$ so $F^+ \to F$ is (topologically) locally trivial near y. This completes the proof of Lemma 1. □

Let F be any fixed point component. We claim that F can be Whitney stratified with the manifolds $A^* = A \cap F$ where A runs over the strata of X. We check the Whitney conditions for a pair of such manifolds $(A^*, B^*) = (A \cap F, B \cap F)$ at a point $y \in B^*$. Let $x_i \in A^*$ and $y_i \in B^*$ converge to y and suppose the secant lines $\overrightarrow{x_i y_i}$ converge to some line ℓ and suppose the tangent planes $T_{x_i} A^*$ converge to some τ^*. We must show that $\ell \subset \tau^*$. We may suppose the $T_{x_i} A^*$ converge to some limiting plane τ as above, so $\tau^0 = \lim(T_{x_i} A)^0 = \lim T_{x_i} A^* = \tau^*$. Whitney's condition applied to X guarantees that $\ell \subset \tau$ and in fact $\ell \subset \tau^0$ because x_i and y_i are in F. Thus $\ell \subset \tau^*$.

Now suppose $\Gamma_{A \cap F}$ denotes the closure in $X \times F$ of $\{(a, p(a)) | a \in A \cap F^+\}$. We will verify (1c) for the above stratification of F. To do so, it suffices to show $g^{-1}(A \cap F) = \Gamma_{A \cap F}$ since $\Gamma_{A \cap F}$ is irreducible. Let $w \in g^{-1}(A \cap F)$. Clearly $w = (g(w), \pi(w))$ where $g(w) \in g^{-1}(A \cap F)$. Clearly $w = (g(w), \pi(w))$ where $g(w) \in A \cap F$ and $\pi(w) \in \overline{F^+}$. We claim that there exists a chain of orbits from $g(w)$ to $\pi(w)$ which is completely contained in $\overline{A} \cap \overline{F^+}$. To see this, note that w is a limit of points $(p(x_n), x_n)$ in Γ such that $g(w) = \lim p(x_n)$. We may, by [13, Corollary to Theorem 1.1] choose a subsequence so that $\lim \mathbf{C}^* \cdot x_n$ exists and is a chain $\overline{\mathcal{O}}_1 \cup \ldots \cup \overline{\mathcal{O}}_t$ of closures of orbits, i.e. $(\overline{\mathcal{O}}_i)_\infty = (\overline{\mathcal{O}}_{i+1})_0$ for $1 \leqq i \leqq t-1$. Clearly $(\overline{\mathcal{O}}_1)_0 = g(w)$ and $\pi(w)$ lies in some $\overline{\mathcal{O}}_\ell$. The singularity preserving assumption implies that each $\overline{\mathcal{O}}_k \subset \overline{A}$. In fact if $y \in \overline{A} \cap F$ then $\overline{y^+} \subset \overline{A}$. (For if B is the stratum of X containing y, then $y^+ \subset B \subset \overline{A}$ by the frontier condition). It follows by induction that $\overline{\mathcal{O}}_1 \cup \ldots \cup \overline{\mathcal{O}}_t \subset \overline{A}$. Hence $\pi(w) \in \overline{A} \cap \overline{F^+} = \overline{(A \cap F)^+}$ (since $A \cap F \neq \emptyset$). This means $w = (g(w), \pi(w)) \in \Gamma_{A \cap F}$ which proves that $g^{-1}(A \cap F) \subset \Gamma_{A \cap F}$. The reverse inclusion is clear. This completes the proof that the action is good and the proof of Theorem 4. □

References

1. Bialynicki-Birula, A.: Some theorems on actions of algebraic groups. Annals of Mathematics **98**, 480–497 (1973)
2. Bialynicki-Birula, A.: On fixed points of torus actions on projective varieties. Bull. de l'Acad. Polon des Sciences **32**, 1097–1101 (1974)
3. Bialynicki-Birula, A., Sommese, A.J.: Complete quotients of torus actions. Transactions of American Math. Society in press (1983)

4. Carrell, J.B., Sommese, A.J.: **C*** actions. Math. Scand. **43**, 49–59 (1978)
5. Carrell, J.B., Sommese, A.J.: Some topological aspects of **C*** actions on compact Kaehler manifolds. Comment. Math. Helvetici **54**, 567–582 (1979)
6. Carrell, J.B., Sommese, A.J.: Filtrations of complex manifolds with a **C*** action. Math. Scand. in press (1983)
7. Frankel, T.: Fixed points and torsion on Kaehler manifolds. Ann. of Math. **70**, 1–8 (1959)
8. Goresky, M., MacPherson, R.: Intersection homology theory. Topology **19**, 135–162 (1980)
9. Hardt, R.: Topological properties of subanalytic sets. Trans. Amer. Math. Soc. **211**, 57–70 (1975)
10. Hartshorne, R.: On the deRham cohomology of algebraic varieties. Publ. Math. I.H.E.S. n° 45, 1–99 (1975)
11. Jurkiewicz, J.: An example of algebraic torus action which determines the non-filterable decomposition, Bull. de l'Acad. Polon. des Sciences **24**, 667–674 (1976)
12. Konarski, J.: Some pathological actions of k^*. Lecture Notes in Mathematics, vol. 956. Berlin-Heidelberg-New York: Springer
13. Lieberman, D.: Compactness of the Chow scheme: applications to automorphisms and deformations of Kaehler manifolds. Seminaire Francois Norguet. Lecture Notes in Mathematics, vol. 670. Berlin-Heidelberg-New York: Springer
14. McCrory, C.: Stratified general position. Algebraic and geometric topology. Lecture Notes in Mathematics, vol. 664, pp. 142–146. Berlin-Heidelberg-New York: Springer
15. Mather, J.: Stratifications and mappings. Dynamical systems Peixote, M.M., (ed.). New York: Academic Press 1973
16. Milnor, J., Stasheff, J.: Characteristic classes. Annals of mathematics study no. 76. Princeton, N.J.: Princeton University Press 1974
17. Spaltenstein, N.: The fixed point set of a unipotent transformation on the flag manifold. Indag. Math. **38**, 542–456 (1976)
18. Sommese, A.J.: Extension theorems for reductive group actions on compact Kaehler manifolds. Math. Ann. **218**, 107–116 (1975)
19. Sommese, A.J.: Some examples of **C*** actions. Lectures Notes in Mathematics, vol. 956. Berlin-Heidelberg-New York: Springer
20. Vargas, J.A.: Fixed points under the action of unipotent elements of SL_n in the flag variety. Boletin de la Sociedad Matematica Mexicana **24**, 1–15 (1979)

Oblatum 10-XI-1982

Invent. math. 73, 383-411 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Automorphisms of Enriques Surfaces

W. Barth[1] and C. Peters[2]

[1] Mathematisches Institut der Universität, Bismarckstr. $1\frac{1}{2}$, D-8520 Erlangen, Federal Republic of Germany
[2] Mathematisch Instituut Rijksuniversiteit, Wassenaarseweg 80, NL-2300 RA Leiden, The Netherlands

0. Introduction

The aim of this note is to compute the group $\mathrm{Aut}(Y)$ of (biholomorphic) automorphisms for the general Enriques surface Y. The basic tool is the global Torelli theorem for projective K3-surfaces as it was given by Piatetski-Shapiro and Shafarevich [11] and refined by Burns and Rapaport [2]. The essential result is that - in contrast to the case of curves - $\mathrm{Aut}(Y)$ is big for general Y and small for special Y.

In this paper we consider the complex case only. Recall that an Enriques surface Y is a (projective) complex surface with universal double cover a K3-surface. One knows that $H^2(Y,\mathbb{Z})=\mathbb{Z}^{10}\oplus\mathbb{Z}_2$ and that the cup-product provides $H^2(Y,\mathbb{Z})/\mathrm{torsion}=\mathbb{Z}^{10}$ with the structure of a lattice M of signature (1, 9).

Theorem. *For a generic Enriques surface Y the representation of* $\mathrm{Aut}(Y)$ *on* $H^2(X,\mathbb{Z})$ *defines an isomorphism of* $\mathrm{Aut}(Y)$ *with the 2-congruence subgroup of* $O^\uparrow(M)$, *where* $O^\uparrow(M)$ *is the group of isometries of M not interchanging the two positive half-cones in* $M\otimes_{\mathbb{Z}}\mathbb{R}$, *or in other words, the reflection group of the lattice M.*

Here the notion "generic" needs some explanation. Horikawa [7, 8] defined a quasi-projective period domain D^0/Γ for Enriques surfaces. The assertion in the theorem holds for all surfaces Y with period point $\tau(Y)\in D^0/\Gamma$ in the complement of countably many images of 9-dimensional analytic varieties (recall $\dim D^0/\Gamma=10$). It was pointed out to us by Dolgachev that the theorem also follows from results of Nikulin [10], although it is not stated there explicitely.

For special Y the automorphism group can be quite different. We describe a 2-dimensional family of surfaces Y where $\mathrm{Aut}(Y)$ in general is $\mathbb{Z}_2\times D_\infty$, but for special cases $\mathbb{Z}_4\times D_\infty$ or D_4. Here $D_4(D_\infty)$ is the dihedral group $\mathbb{Z}_2\ltimes\mathbb{Z}_4(\mathbb{Z}_2\ltimes\mathbb{Z})$. The example of surfaces with finite group $\mathrm{Aut}(Y)$ was communicated to us by Dolgachev.

We apply the knowledge of $\mathrm{Aut}(Y)$ for generic Y to count the number of inequivalent realisations of Y as elliptic fibre space over $\mathbb{P}_1$, as double cover of

a quadri-nodal complete intersection of two quadrics in $\mathbb{P}_4$ (double plane realisation), as sextic surface in $\mathbb{P}_3$ passing doubly through the edges of a tetrahedron (Enriques-realisation), or as smooth surface in $\mathbb{P}_5$ of degree 10 (deformations of Reye-congruences). There are

$$\begin{aligned} 527 &= 17\cdot 31 && \text{realisations as elliptic fibration}\\ 67456 &= 2^7\cdot 17\cdot 31 && \text{double plane realisations}\\ 5396480 &= 2^{11}\cdot 5\cdot 17\cdot 31 && \text{Enriques-realisations}\\ 25903104 &= 2^{13}\cdot 3\cdot 17\cdot 31 && \text{realisations as 10th degree surface in } \mathbb{P}_5. \end{aligned}$$

We owe much to stimulating discussions on this subject with many other geometers, in particular to I. Dolgachev.

1. Some Lattices and Their Isometries

1.1. **Preliminaries.** A *lattice* is a free $\mathbb{Z}$-module of finite rank endowed with an integral quadratic form. $L\perp M$ denotes the *orthogonal direct sum* of two lattices L and M. $L^\vee=\mathrm{Hom}_{\mathbb{Z}}(L,\mathbb{Z})$ is the *dual* $\mathbb{Z}$*-module* (the canonical quadratic form on $L^\vee$ in general is not integral). The symmetric bilinear form on a lattice L, associated with the quadratic form, usually is denoted by $\langle\ ,\ \rangle$.

This form defines the *correllation* morphism

$$\varphi_L: L\to L^\vee, \qquad x\to\langle x,-\rangle.$$

If L is *nondegenerate*, φ_L is injective, and we may identify L with the submodule $\varphi_L(L)\subset L^\vee$. Then $L^\vee/L$ is a finite abelian group, trivial precisely when L is *unimodular*.

A submodule $M\subset L$ is called *primitive*, if L/M is free of torsion. In this case every integral functional on M extends to L, i.e., the restriction $L^\vee\to M^\vee$ is surjective.

If L is nondegenerate and M is primitive and nondegenerate, the composition $L\xrightarrow{\sim}L^\vee\to M^\vee\to M^\vee/M$ is surjective. It sends $x\in L$ to the $\varphi(M)$-equivalence class of $\langle x,-\rangle|M$. So its kernel is $M\perp M^\perp$ and we obtain an isomorphism $i_M: L/(M\perp M^\perp)\to M^\vee/M$. Interchanging the rôles of M and $M^\perp$ we obtain $i_{M^\perp}: L/(M\perp M^\perp)\xrightarrow{\sim}(M^\perp)^\vee/M^\perp$. Then we put $j_M=i_{M^\perp}\circ i_M^{-1}$: $M^\vee/M\xrightarrow{\sim}(M^\perp)^\vee/M^\perp$.

(1.1) **Lemma** [9, Prop. 1.1]. *Let L be a unimodular lattice, $M\subset L$ a nondegenerate primitive sublattice, and α: $M\to M$, β: $M^\perp\to M^\perp$ isometries. Then the isometry (α,β) of $M\perp M^\perp$ extends to L if and only if the automorphisms $\tilde\alpha$ on $M^\vee/M$ induced by α and $\tilde\beta$ on $(M^\perp)^\vee/M^\perp$ induced by β satisfy $j_M\circ\tilde\alpha=\tilde\beta\circ j_M$.*

(1.2) **Corollary.** *Let $M\subset L$ and $\alpha: M\to M$ be as above. If α extends to an isometry of L restricting to $\pm\mathrm{id}$ on $M^\perp$, then this extension is unique. Such an extension exists if and only if $\alpha^\vee: M^\vee\to M^\vee$ induces $\pm\mathrm{id}$ on $M^\vee/M$.*

We shall use the following *notation*: For $n\in\mathbb{N}$, by nL we denote the sublattice $\{n\cdot x: x\in L\}$ of L, whereas $L(n)$ is the $\mathbb{Z}$-module L endowed with the qua-

dratic form $x \to nx^2$. A *root* in the lattice L is an element w of square $w^2 = -2$. Any root w defines the *reflection* $s_w \in O(L)$, $s_w(x) = x + \langle x, w \rangle w$. Given a lattice L we put $L_{\mathbb{Q}} = L \otimes_{\mathbb{Z}} \mathbb{Q}$, $L_{\mathbb{R}} = L \otimes_{\mathbb{Z}} \mathbb{R}$, $L_{\mathbb{C}} = L \otimes_{\mathbb{Z}} \mathbb{C}$ and for any homomorphism $g: L \to L$ we denote by $g_{\mathbb{Q}}$, $g_{\mathbb{R}}$, resp. $g_{\mathbb{C}}$ the natural extension of g to these vector spaces.

1.2. **Application.** For applying 1.2 to the Picard lattice of K3 and Enriques surfaces we fix the following notation.

$$\mathbb{H} = \mathbb{Z}e_1 + \mathbb{Z}e_2 \qquad \text{with } e_1^2 = e_2^2 = 0,\ e_1 e_2 = 1 \text{ (hyperbolic plane)},$$
$$\mathbb{E} = E_8(-1) \qquad \text{(root lattice for Dynkin diagram } E_8),$$
$$L = \mathbb{H} \perp \mathbb{H} \perp \mathbb{H} \perp \mathbb{E} \perp \mathbb{E},$$
$$M = \mathbb{H} \perp \mathbb{E},$$
$$s: L \to L \text{ the involution } (h_1, h_2, h_3, e_1, e_2) \mapsto (-h_1, h_3, h_2, e_2, e_1),$$
$$L^+ = \{x \in L: s(x) = x\} = \{(0, h, h, e, e) \in L: h \in \mathbb{H}, e \in \mathbb{E}\},$$
$$L^- = \{x \in L: s(x) = -x\} = \{(h_1, h, -h, e, -e) \in L: h_1, h \in \mathbb{H}, e \in \mathbb{E}\}.$$

There are obvious isometries

$$\varepsilon^+: L^+ \to M(2) \qquad (0, h, h, e, e) \mapsto (h, e),$$
$$\varepsilon^-: L^- \to \mathbb{H} \perp M(2), \quad (h_1, h, -h, e, -e) \mapsto (h_1, h, e).$$

In particular this shows

$$(L^{\pm})^{\vee}/L^{\pm} = M/2M = (\mathbb{Z}_2)^{10}.$$

L^+ and L^- are primitive nondegenerate sublattices of L, one the orthogonal complement of the other. Using that all elements in $(L^{\pm})^{\vee}/L^{\pm}$ are induced by inner products with elements $(0, h, 0, e, 0)$ one traces the isomorphism $j: (L^-)^{\vee}/L^- \to (L^+)^{\vee}/L^+$ and finds: j is induced by the obvious isometry

$$\psi: L^- \to \mathbb{H} \perp L^+, \qquad (h_1, h, -h, e, -e) \mapsto (h_1, h, h, e, e).$$

We also put

$$\Gamma = \{g \in O(L): gs = sg\}.$$

For any $g \in \Gamma$ we have $g: L^{\pm} \to L^{\pm}$, and there are obvious restrictions $r^{\pm}: \Gamma \to O(L^{\pm})$.

(1.3) **Lemma.** *For $g \in O(L^{\pm})$ the following properties are equivalent:*

a) *there is a (unique) extension $\gamma \in \Gamma$ of g with $r^{\mp}(\gamma) = \mathrm{id}$.*

b) *g belongs to the 2-congruence subgroup of $O(L^{\pm})$.*

The proof follows from Corollary (1.2), because g induces the identity on $(L^{\pm})^{\vee}/L^{\pm}$ if and only if it belongs to the 2-congruence subgroup.

The quadratic form on M has signature (1, 9). So the set $\{x \in M_{\mathbb{R}}: x^2 > 0\}$ consists of two disjoint cones $\mathscr{C}_M$ and $-\mathscr{C}_M$. We put $O^{\uparrow}(M) = \{g \in O(M): g\mathscr{C}_M = \mathscr{C}_M\}$.

Then $O(M)$ is the direct product $O^{\uparrow}(M) \times \{\pm \mathrm{id}\}$.

1.3. **On the Root Lattice $\mathbb{E}$.** In this section we collect a few properties of $\mathbb{E}$ which are needed later. We use the description [1, p. 268] of $\mathbb{E}$. So $\mathbb{E} \subset \mathbb{R}^8$ (with the negative of the usual inner product) is the set of vectors $(x^1, \ldots, x^8)$ where either all x^i are integers or all x^i are half-integers, and $\sum x^i \in \mathbb{Z}$ is even. The 240 roots are $(0 \ldots 0, \pm 1_i, 0 \ldots 0, \pm 1_j, 0 \ldots 0)$, $1/2(\pm 1, \ldots, \pm 1)$.

(1.4) **Lemma.** *There are exactly* 135 *equivalence classes* mod $2\mathbb{E}$ *of vectors* $x \in \mathbb{E}$ *with* $x^2 = -4$.

Proof. An integral vector $x \in \mathbb{E}$ with $x^2 = -4$ is up to permutation of the coordinates of the form $\pm(2, 0, \ldots, 0)$ or $(\pm 1, \pm 1, \pm 1, \pm 1, 0, 0, 0, 0)$. Since $(0 \ldots 0, \pm 2_i, 0 \ldots 0, \pm 2_j, 0 \ldots 0) \in 2\mathbb{E}$, all vectors $(0 \ldots 0, \pm 2, 0 \ldots 0)$ are equivalent mod $2\mathbb{E}$. Of the second type there are $2^4 \cdot \binom{8}{4}$ vectors and mod $2\mathbb{E}$ each of them is equivalent with 2^4 ones.

So there are $\binom{8}{4} = 70$ inequivalent ones. Any half-integral vector is up to permutation of coordinates of the form

$$\pm\tfrac{1}{2}(3, -1, \ldots, -1), \qquad \pm\tfrac{1}{2}(3, 1, 1, -1, \ldots, -1),$$
$$\pm\tfrac{1}{2}(3, 1, 1, 1, 1, -1, -1, -1), \quad \pm\tfrac{1}{2}(3, 1, 1, 1, 1, 1, 1, -1).$$

Here all vectors of the first and of the last type are equivalent mod $2\mathbb{E}$. Vectors of the first, second, and third type are inequivalent mod $2\mathbb{E}$. There are $2.8 \cdot \binom{7}{2}$ vectors of the second type, each equivalent with 2.6 of them. So there are $\frac{1}{2}7.8 = 28$ inequivalent ones mod $2\mathbb{E}$. Of the third type there are $2.8 \cdot \binom{7}{3}$ vectors, each of them equivalent with 16 ones, so 35 inequivalent ones. Altogether we have $1 + 70 + 1 + 28 + 35 = 135$. □

(1.5) **Corollary.** *Choosing* 135 *representatives of the equivalence classes above, one from each of the* 120 *pairs* $\pm w$ *of roots, and* 0, *one obtains a system of representatives of* $\mathbb{E}$ mod $2\mathbb{E}$.

Proof. We only have to show that $w_1 - w_2 \in 2\mathbb{E}$ for two roots w_1, w_2 implies $w_2 = \pm w_1$. But if $w_1 - w_2 \in 2\mathbb{E}$, then $(w_1 - w_2)^2 = -4 - 2w_1 w_2$ is divisible by 8. Since $|w_1 w_2| \leqq 2$ this implies $w_1 w_2 = \pm 2$, i.e., $w_2 = \pm w_1$. □

We denote by $W = W(E_8)$ the *Weyl group*. Since the Dynkin diagram of E_8 admits no symmetries, W coincides with $O(\mathbb{E})$, see [1, p. 270]. W contains in particular

- all permutations of coordinates x^i
- simultaneous changes $x^i, x^j \mapsto -x^i, -x^j$ of the signs of two coordinates.

(1.6) **Lemma.** *W operates transitively on the set of all ordered* 8*-tuples of roots* $w_1, \ldots, w_8 \in \mathbb{E}$ *satisfying* $\langle w_i, w_j \rangle = -1$ *whenever* $i \neq j$.

Proof. $W(E_8)$ operates transitively on the roots, so we may assume

$$w_1 = \tfrac{1}{2}(1, \ldots, 1).$$

If w_i, $i \geqq 2$, is integral, then

$$w_i = (0 \ldots 0, 1, 0 \ldots 0, 1, 0 \ldots 0).$$

If w_i is not integral, say $w_i=\frac{1}{2}(1,1,1,1,1,1,-1,-1)$ we use the reflection s_w with $w=\frac{1}{2}(-1,1,1,1,1,-1,-1,-1)\perp w_1$ and transform w_i into an integral root. After permuting coordinates we have

$$w_2=(1,1,0\ldots 0).$$

Since $w_2\perp w$, by the same argument we may assume w_3 integral. Then after permutation

$$w_3=(1,0,1,0\ldots 0).$$

Again $w_3\perp w$ and we may arrange it that w_4 is integral, i.e.,

$$w_4=(1,0,0,\ldots 1\ldots 0) \quad \text{or} \quad w_4=(0,1,1,0\ldots 0).$$

In the second case we transform w_4 under s_u with

$$u=\tfrac{1}{2}(1,-1,-1,-1,-1,1,1,1)\perp w_1,w_2,w_3 \quad \text{into } \tfrac{1}{2}(1,1,1,-1,-1,1,1,1)$$

and then with

$$u=\tfrac{1}{2}(1,-1,-1,1,1,1,-1,-1)\perp w_1,w_2,w_3 \quad \text{into } (1,0,0,0,0,1,0,0).$$

So after permutation

$$w_4=(1,0,0,1,0,0,0,0).$$

Still $w_4\perp w$, hence we may assume w_5 integral, and after permutation

$$w_5=(1,0,0,0,1,0,0,0)\perp w$$
$$w_6=(1,0,0,0,0,1,0,0).$$

So, after permutation, we have

$$w_7=(1,0,0,0,0,0,1,0) \quad \text{or} \quad \tfrac{1}{2}(1,1,1,1,1,1,-1,-1).$$

In the first case necessarily $w_8=(1,0,\ldots,0,1)$ and in the second case there is no w_8 at all. □

1.4. **Reduction Modulo 2.** In this section we examine the reduction morphism $O(M)\to \mathrm{Aut}(M/2M)=GL(10,\mathbb{F}_2)$. By $\bar{x}\in M/2M$ we denote the class represented by $x\in M$. On $M/2M$ we have the bilinear form

$$\langle\bar{x},\bar{y}\rangle=\langle x,y\rangle \mod 2.$$

Since the form on M is even, by

$$q(\bar{x})=\tfrac{1}{2}x^2 \mod 2$$

one defines a nondegenerate quadratic form q on $M/2M$, i.e., a form satisfying

$$q(\bar{x}+\bar{y})=q(\bar{x})+q(\bar{y})+\langle\bar{x},\bar{y}\rangle.$$

On each $\mathbb{F}_2$-vector space of even dimension $2k$ there are - up to conjugation - exactly two such forms, q^+ and q^-, differing by their number $\nu = 2^{k-1}(2^k \pm 1)$ of zeros.

Using Lemma (1.4) we count the zeros of $q|\mathbb{E}/2\mathbb{E}$ and find

$$\nu_{\mathbb{E}} = 2^8 - 120 = 136 = 2^3(2^4+1).$$

We observe that the elements $\bar{x} \in \mathbb{E}/2\mathbb{E}$ with $q(\bar{x}) = 1$ are precisely the images of the roots.

Now $\mathbb{H}/2\mathbb{H} = \mathbb{F}_2^2$ and q has 3 zeros on it. The zeros of q on $M/2M$ are exactly the pairs $(\bar{h}, \bar{x})$, $h \in \mathbb{H}$, $x \in \mathbb{E}$ satisfying $q(\bar{h}) = (\bar{x})$. Their number is

$$3 \cdot \nu_E + 256 - \nu_E = 3 \cdot 136 + 120 = 528 = 2^4(2^5+1).$$

Hence q and $q|\mathbb{E}/2\mathbb{E}$ is the corresponding form q^+. Its group of automorphisms is denoted by $O^+(2k, \mathbb{F}_2)$. E.g. in the book [5] one finds (Chap. III, § 10)

i) $|O^+(2k, \mathbb{F}_2)| = 2^{1+k(k-1)} \cdot (2^k - 1) \cdot \sum_{j=1}^{k-1} (2^{2j} - 1)$,

ii) the group $O^+(2k, \mathbb{F}_2)$ is generated by transvections if $k \neq 2$; these are maps $\bar{x} \mapsto \bar{x} + \langle \bar{x}, \bar{a} \rangle \bar{a}$, $q^+(\bar{a}) = 1$.

iii) the group $O^+(2k, \mathbb{F}_2)$ contains a normal subgroup of index 2 consisting of all products of an even number of transvections. For $k \geqq 3$ this group is simple. In our cases, $k = 4$ and 5, we find

$$|O(\mathbb{E}/2\mathbb{E}, q)| = 2^{13} \cdot 3^5 \cdot 5^2 \cdot 7,$$
$$|O(M/2M, q)| = 2^{21} \cdot 3^5 \cdot 5^2 \cdot 7 \cdot 17 \cdot 31.$$

Now any root $w \in M$ reduces in $M/2M$ to an element $\bar{w}$ with $q(\bar{w}) = 1$, and the reflection s_w reduces to the transvection defined by $\bar{w}$. Conversely, if $\bar{a} \in M/2M$ with $q(\bar{a}) = 1$, then $\bar{a} = (\bar{h}, \bar{x})$, $h \in \mathbb{H}$, $x \in \mathbb{E}$, such that one of the following holds:

- either $q(\bar{h}) = 1$ and $q(\bar{x}) = 0$, i.e. $h \in \mathbb{H}$ is modulo $2\mathbb{H}$ equivalent with $h_1 + h_2$, $(h_1 + h_2)^2 = 2$, and $x \in \mathbb{E}$ to an element of square -4 (cf. 1.3). So $\bar{a}$ is the image of a root in M.
- or $q(\bar{h}) = 0$ and $q(\bar{x}) = 1$, i.e., $h \in \mathbb{H}$ is equivalent to $0, h_1$, or h_2 and $\bar{x} \in \mathbb{E}$ is equivalent to a root. In this case too, $\bar{a}$ is the image of a root in M.

This proves that all transvections are reductions mod 2 of reflections s_w and the reduction maps

$$W(E_8) \to O(\mathbb{E}/2\mathbb{E}, q) \qquad O(M) \to O(M/2M, q)$$

are surjective. Since $O(M) = O^\uparrow(M) \times \{\pm \mathrm{id}\}$, even $O^\uparrow(M) \to O(M/2M, q)$ is surjective. Recalling that

$$|W(E_8)| = 2^{14} \cdot 3^5 \cdot 5^2 \cdot 7$$

we find the following well known

(1.7) **Proposition.** a) *The 2-congruence subgroup of $W(E_8)$ is just $\{\pm \mathrm{id}\}$* (cf. [1] Exercise in Chap. 6, § 4).

b) *The 2-congruence subgroup in $O(M)$ has index $2^{21} \cdot 3^5 \cdot 5^2 \cdot 7 \cdot 17 \cdot 31$.*

1.5. **An Auxiliary Result.** Denote by K the lattice $\mathbb{H}\perp\mathbb{H}(2)$ and fix a basis $h_1, h_2\in\mathbb{H}$, $k_1, k_2\in\mathbb{H}(2)$ with $h_1^2=h_2^2=k_1^2=k_2^2=0$, $\langle h_1, h_2\rangle=1$, $\langle k_1, k_2\rangle=2$. Let $G\subset O(K)$ be the subgroup acting trivially on $K^\vee/K$. It contains all $g\in O(K)$ such that $g(k_i)-k_i\in 2K$ for $i=1,2$. In 4.5 we shall apply the following fact. (We are indebted to Y. Namikawa for pointing out to us an error in the first version of this lemma.)

(1.8) **Lemma.** *All vectors $x\in K$ of square $x^2=-4$, which are of the form*

$$x=2s_1h_1+2s_2h_2+t_1k_1+t_2k_2, \qquad s_i, t_i\in\mathbb{Z},$$

are under G conjugate with k_1+k_2.

Proof. Given x as above we put

$$x_1=2s_1h_1+t_1k_1, \qquad x_2=-2s_2h_2-t_2k_2,$$
$$e_1=t_2h_1-s_2k_1, \qquad e_2=-t_1h_2+s_2k_2.$$

They satisfy

$$x_1^2=x_2^2=e_1^2=e_2^2=\langle x_i, e_j\rangle=0$$

and since

$$x^2=4(2s_1s_2+t_1t_2)=-4$$

we have additionally

$$\langle x_1, x_2\rangle=2, \qquad \langle e_1, e_2\rangle=1.$$

So e_1, e_2, x_1, x_2 form a basis of K with e_1, e_2 generating a sublattice $\mathbb{H}$ and x_1, x_2 generating a sublattice $\mathbb{H}(2)$. Then there is some $g\in O(K)$ mapping

$$e_1\mapsto h_1, \qquad e_2\mapsto h_2, \qquad x_1\mapsto k_1, \qquad x_2\mapsto k_2,$$
$$x=x_1+x_2\mapsto k_1+k_2.$$

Since $t_1\cdot t_2=1-2s_1s_2$ is odd, the vectors

$$x_1-k_1=2s_1h_1+(t_1-1)k_1 \qquad x_2-k_2=-2s_2h_2-(t_2+1)k_2$$

belong to $2K$ and $g\in G$.

2. Periods of Enriques Surfaces

2.1. **Notation.** Let X be any complex projective surface. The cup-product form $\langle\ ,\ \rangle$ on $H^2(X,\mathbb{R})$ restricts to the subspace $H^{1,1}(X,\mathbb{R})=H^2(X,\mathbb{R})\cap H^{1,1}(X)$ as a form of signature $(1, h^{1,1}(X)-1)$. The set $\{x\in H^{1,1}(X,\mathbb{R})\colon \langle x,x\rangle>0\}$ consists of two disjoint connected cones. For two elements in the same connected component the cup-product is positive, while it is negative for two elements in different components. So only one of the cones, say $\mathscr{C}_X$, *contains classes of ample divisors.*

The inclusion $\mathbb{Z}\to\mathbb{R}$ induces a map $H^2(X,\mathbb{Z})\to H^2(X,\mathbb{R})$. Its image $H^2(X,\mathbb{Z})/\text{torsion}$ is denoted by $H^2(X,\mathbb{Z})_f$. Its elements are called the *integral*

points of $H^2(X,\mathbb{R})$. The cup-product provides $H^2(X,\mathbb{Z})_f$ with a quadratic form. The sublattice

$$S_X = H^{1,1}(X,\mathbb{R}) \cap H^2(X,\mathbb{Z})_f$$

is called the *algebraic lattice*. Its elements are precisely the cohomology classes d of divisors D on X. $T_X = S_X^{\perp} \subset H^2(X,\mathbb{Z})_f$ is called the *transcendental lattice*. A curve $D \subset X$ is called *nodal* or (-2)-*curve*, if it is smooth rational with $D^2 = -2$. A *nodal class* is the class $d \in S_X$ of such a curve. We put

$$\mathscr{C}_X^+ = \{x \in \mathscr{C}_X : \langle x, d\rangle > 0 \text{ for all nodal classes } d\}.$$

(2.1) **Lemma.** *If X is a* K3 *or Enriques surface, the set $\mathscr{C}_X^+ \cap H^2(X,\mathbb{Z})_f$ of integral points in $\mathscr{C}_X^+$ consists precisely of the classes of ample divisors.*

Proof. By the Nakai-Moishezon criterion a divisor D with $D^2 > 0$ is ample if and only if $D \cdot E > 0$ for all irreducible curves $E \subset X$. But for such a curve the adjunction formula shows $E^2 = -2$ or $E^2 \geqq 0$. In the second case its class e belongs to the closure of $\mathscr{C}_X$ and hence $\langle x, e\rangle > 0$ for all $x \in \mathscr{C}_X$. It follows that an integral point of $\mathscr{C}_X^+$ is the class of an ample divisor and conversely. □

Therefore, in the case of a K3 or Enriques surface X, $\mathscr{C}_X^+$ is called the *ample cone*.

In the remainder of this section X is a K3-surface.

(2.2) **Lemma.** *Let Δ^+ denote the set of all classes $d \in H^2(X,\mathbb{Z})$ of effective divisors satisfying $d^2 = -2$. Then*

$$\mathscr{C}_X^+ = \{x \in \mathscr{C}_X : \langle x, d\rangle > 0 \text{ for all } d \in \Delta^+\}.$$

Proof. Let $\mathscr{C}'$ denote the cone on the right-hand side. Since Δ^+ contains all nodal classes, obviously $\mathscr{C}' \subset \mathscr{C}_X^+$. Conversely, if $d \in \Delta^+$ and $\langle x, d\rangle \leqq 0$ for some $x \in \mathscr{C}_X^+$, then also $\langle y, d\rangle < 0$ for some integral point $y \in \mathscr{C}_X^+$. This contradicts (2.1). So $\langle x, d\rangle > 0$ for all $d \in \Delta^+$, $x \in \mathscr{C}_X^+$ and this shows $\mathscr{C}_X^+ \subset \mathscr{C}'$. □

(2.3) **Lemma.** *For an isometry g of $H^2(X,\mathbb{Z})_f$ the following properties are equivalent.*

i) $g_{\mathbb{R}} \mathscr{C}_X^+ \subset \mathscr{C}_X^+$.

ii) *g maps each class of an ample divisor to the class of an ample divisor.*

iii) *g maps the class of one ample divisor to the class of an ample divisor.*

iv) $g_{\mathbb{R}} \mathscr{C}_X = \mathscr{C}_X$ *and* $g\Delta^+ = \Delta^+$.

Proof. i) ⇒ ii) follows from (2.1), ii) ⇒ iii) is trivial. If iii) holds then of course $g\mathscr{C}_X = \mathscr{C}_X$. If $d \in \Delta^+$, then $(gd)^2 = -2$, so by Riemann-Roch either gd or $-gd$ is effective. But let $a \in H^2(X,\mathbb{Z})$ be an ample class with ga ample again. Then $\langle gd, ga\rangle = \langle d, a\rangle$ is positive and $-gd$ cannot be effective. This proves iv). The step iv) ⇒ i) follows from (2.2). □

We denote by $O^{\uparrow}(X) \subset O(H^2(X,\mathbb{Z}))$ the subgroup of isometries g with properties i)–iv).

2.2. **The Universal Covering of an Enriques Surface.** Let Y be an Enriques surface and $\pi : X \to Y$ its universal (double) covering. Let $\sigma \in \mathrm{Aut}(X)$ be the covering

involution. According to Horikawa [7, Theorem 5.1] there is an identification $H^2(X,\mathbb{Z})\xrightarrow{\sim} L$ such that σ^* acts on $H^2(X,\mathbb{Z})$ as the involution s from 1.2. The map $\pi^*: H^2(Y,\mathbb{Z})_f \to H^2(X,\mathbb{Z})$ is an isomorphism onto $L^+ \subset L$. In particular there is an isometry $H^2(Y,\mathbb{Z})_f \xrightarrow{\sim} M$. Such an identification $H^2(X,\mathbb{Z})\to L$ is called a *marking of the Enriques surface* Y. Let $H^{1,1}(X,\mathbb{R})^\sigma$ denote the vector subspace of σ^*-invariants.

(2.4) **Lemma.** *The map* $\pi^*: H^{1,1}(Y,\mathbb{R})\to H^{1,1}(X,\mathbb{R})$ *maps* $\mathscr{C}_Y^+$ *bijectively onto* $\mathscr{C}_X^+ \cap H^{1,1}(X,\mathbb{R})^\sigma$. *The integral points in* $\mathscr{C}_Y^+$ *correspond* 1 *to* 1 *under* π^* *to the classes of ample* σ*-invariant divisors on* X.

Proof. To test whether $x\in\mathscr{C}_X$ belongs to $\mathscr{C}_X^+$ we have to check $\langle x,d\rangle>0$ for nodal classes d. If $\sigma^* x=x$ then $\langle x,d\rangle=\frac{1}{2}\langle x,d+\sigma^* d\rangle$, and if $\langle d,\sigma^* d\rangle>0$, then $(d+\sigma^* d)^2=-4+2\langle d,\sigma^* d\rangle\geqq 0$, because this number is divisible by 4. Hence $d+\sigma^* d\in\bar{\mathscr{C}}_X$ and $\langle x,d\rangle>0$. So we have to check $\langle x,d\rangle>0$ only for nodal classes d with $\langle d,\sigma^* d\rangle=0$. If now $D\subset X$ is the (-2)-curve representing d, then $\langle d,\sigma^* d\rangle=0$ if and only if $D\cap\sigma^* D=\emptyset$, i.e., if and only if $\pi(D)=\pi(\sigma D)$ is a (-2)-curve on Y. Since every σ^*-invariant $x\in\mathscr{C}_X$ is of the form $\pi^* y$, $y\in\mathscr{C}_Y$, it follows that $H^{1,1}(X,\mathbb{R})^\sigma\cap\mathscr{C}_X^+=\pi^*\mathscr{C}_Y^+$. If $c\in\mathscr{C}_Y^+$ is an integral point, then $\pi^* c$ is the class of a σ-invariant divisor. Since we have proven $\pi^* c\in\mathscr{C}_X^+$, from Lemma (2.1) we obtain the ampleness of this divisor. □

2.3. **The Torelli Theorem for K3-Surfaces.** In this section we state the global Torelli theorem [11, p. 534], [2, Cor. 32] in the form we need it.

So let X be a projective K3-surface and ω_X a nonzero holomorphic 2-form on X. This ω_X spans $H^{0,2}(X)$ and is unique up to scalars. Using the Hodge decomposition we view $H^{0,2}(X)$ as a subspace of $H^2(X,\mathbb{C})$ and ω_X as a class in $H^2(X,\mathbb{C})$. Obviously

$$\langle\omega_X,\omega_X\rangle=0,\qquad\langle\omega_X,\bar{\omega}_X\rangle>0.$$

For $\operatorname{Re}\omega_X$ and $\operatorname{Im}\omega_X\in H^2(X,\mathbb{R})$ these relations are equivalent with

$$\langle\operatorname{Re}\omega_X,\operatorname{Re}\omega_X\rangle=\langle\operatorname{Im}\omega_X,\operatorname{Im}\omega_X\rangle>0,$$
$$\langle\operatorname{Re}\omega_X,\operatorname{Im}\omega_X\rangle=0.$$

So $\operatorname{Re}\omega_X$ and $\operatorname{Im}\omega_X$ span in $H^2(X,\mathbb{R})$ a two-dimensional subspace, on which the cup-product is positive definite.

Since $H^{1,1}(X)=\operatorname{Re}\omega_X^\perp\cap\operatorname{Im}\omega_X^\perp$, we have

$$S_X=H^2(X,\mathbb{Z})\cap\operatorname{Re}\omega_X^\perp\cap\operatorname{Im}\omega_X^\perp.$$

If $\rho=\operatorname{rank} S_X$ is the *Picard number*, we have signature $S_X=(1,\rho-1)$, signature $T_X=(2,20-\rho)$.

(2.5) **Theorem** (Global Torelli). *Let* g *be an isometry of* $H^2(X,\mathbb{Z})$. *Then* g *is induced by a unique automorphism of* X *if and only if* $g\in O^\uparrow(X)$ *and* $g_{\mathbb{C}}\omega_X=\lambda\omega_X$ *for some* $\lambda\in\mathbb{C}$.

(2.6) **Corollary.** *The representation of* $\operatorname{Aut}(X)$ *on* $H^2(X,\mathbb{Z})$ *is faithful and identifies* $\operatorname{Aut}(X)$ *with a subgroup of* $O^\uparrow(X)$.

2.4. **Periods of Enriques Surfaces.** We recall Horikawa's results [7, 8] on the moduli space of Enriques surfaces. Let $Y=X/\sigma$ be an Enriques surface and $\varphi: H^2(X,\mathbb{Z})\to L$ a marking of Y. Since on Y there are no holomorphic 2-forms, we have $\sigma^*\omega_X=-\omega_X$. So $\varphi_{\mathbb{C}}(\omega_X)$ defines a point $\tau(Y,\varphi)$ in the period domain

$$D:=\{\mathbb{C}\cdot\omega\in\mathbb{P}(L^-_{\mathbb{C}}):\ \langle\omega,\omega\rangle=0,\ \langle\omega,\bar{\omega}\rangle>0\}.$$

This D is the union of two copies of a bounded symmetric domain of type IV and dimension 10. The group Γ (or rather $r^-(\Gamma)$) acts on D in a properly discontinuous way. It contains an involution interchanging the two connected components of D [7, Lemma 8.1]. Since $r^-(\Gamma)$ is an arithmetic group [7], by Baily-Borel the analytic space D/Γ carries the structure of a quasi-projective variety. Since two markings for Y differ by an element in Γ, the Γ-equivalence class $\tau(Y)\in D/\Gamma$ of $\tau(Y,\varphi)$ is independent of the choice of φ. This point $\tau(Y)$ is called the *period point* of Y.

Horikawa proves:

(i) $\tau(Y_1)=\tau(Y_2)$ if and only if Y_1 is isomorphic with Y_2.

(ii) The points $\tau(Y,\varphi)$ belong to

$$D^0=D\setminus\bigcup_{d\text{ root in }L^-} d^\perp,$$

where $d^\perp=\{\mathbb{C}\cdot\omega\in\mathbb{P}(L^-_{\mathbb{C}}):\ \langle\omega,d\rangle=0\}$.

(iii) All points $\tau\in D^0$ are of the form $\tau(Y,\varphi)$ for some marked Enriques surface Y,φ.

(2.7) **Lemma.** *The union of all hyperplanes $d^\perp$, $d\in L^-$ a root, is locally finite in D. Hence $D\cap\bigcup_d d^\perp$ is an analytic set in D.*

Proof. If the union is not locally finite, there are infinitely many distinct roots $d_\nu\in L^-$ and points $\omega_\nu\in D\cap d_\nu^\perp$ such that $\omega=\lim_{\nu\to\infty}\omega_\nu\in D$. Since ω_ν converges to ω, the hyperplanes $(\mathrm{Re}\,\omega_\nu)^\perp$ and $(\mathrm{Im}\,\omega_\nu)^\perp$ as points in $\mathbb{P}((L^-_{\mathbb{R}})^\vee)$ converge to $(\mathrm{Re}\,\omega)^\perp$, resp. $(\mathrm{Im}\,\omega)^\perp$. The cup-product on L^- has signature (2, 10) with $\mathrm{Re}\,\omega$, $\mathrm{Im}\,\omega$ spanning a plane, on which this form is positive definite. So the cup-product is negative definite on $(\mathrm{Re}\,\omega)^\perp\cap(\mathrm{Im}\,\omega)^\perp$. In particular the vectors in $(\mathrm{Re}\,\omega)^\perp\cap(\mathrm{Im}\,\omega)^\perp$ of square -2 form a compact sphere, and there is a compact neighborhood of this sphere containing all vectors of square -2 in $(\mathrm{Re}\,\omega_\nu)^\perp\cap(\mathrm{Im}\,\omega_\nu)^\perp$ for all $\nu\in\mathbb{N}$. All the infinitely many roots d_ν would belong to this compact set, a contradiction. □

The analytic set $D\cap\bigcup d^\perp$ in D is Γ-invariant. So its image in D/Γ is analytic too. The Baily-Borel compactification $\overline{D/\Gamma}$ of D/Γ is obtained by attaching a curve [11, §4, Lemma 1]. By the extension theorem of Remmert-Stein [12, Satz 13] the analytic hypersurface in D/Γ extends to a hypersurface in the projective variety $\overline{D/\Gamma}$. It follows that Horikawa's period domain D^0/Γ is quasi-projective [8, Thm. 3.1].

2.5. **Nodal Curves on Enriques Surfaces.** If $C\subset Y$ is a nodal curve, then $\pi^*C\subset X$ decomposes as $B+\sigma B$ with a nodal curve B on X satisfying $B\cdot\sigma(B)=0$.

Conversely, given nodal curves $B, \sigma(B)$ on X with $B \cdot \sigma(B) = 0$, there is a nodal curve $C = \pi(B) \subset X$ such that $B + \sigma(B) = \sigma^* C$. So there is a 1 to 1 correspondence between nodal classes $c \in H^2(Y, \mathbb{Z})_f$ and pairs $b, \sigma^*(b)$ of nodal classes on X satisfying $\langle b, \sigma^*(b) \rangle = 0$. Fix a marking $H^2(X, \mathbb{Z}) = L$ as above. On L^- the linear forms $\langle b, - \rangle$ and $\langle sb, - \rangle$ differ only by the sign. For each root $c \in H^2(Y, \mathbb{Z})_f = M$ with $\pi^* c = b + \sigma(b)$ put $D_c = D \cap b^\perp$. Since D is not contained in any hyperplane and $b \notin L^+$, we have $D_c \neq D$. Since there are only countably many roots in L^+, the set

$$D_{\text{gen}} = D^0 \setminus \bigcup_{c \text{ root in } L^+} D_c$$

is still dense in D^0 and D_{gen}/Γ is dense in D^0/Γ. The period point $\tau(Y) \in D^0/\Gamma$ is contained in the image of $\bigcup D_c$ if and only if Y contains a (-2)-curve. So, if we understand by a "generic" Enriques surface Y a surface with $\tau(Y) \in D_{\text{gen}}/\Gamma$, then we have shown:

(2.8) **Proposition.** *The generic Enriques surface contains no* (-2)*-curve.*

Now for given Y we define the following sublattices $L_1, L_2, L_3, L_4 \subset L$: Let $M' \subset M$ be the smallest primitive sublattice containing all nodal classes and $L_1 = \pi^* M' \subset L^+$. Let $L_2 \subset L^-$ be the smallest primitive sublattice containing all the classes $d - s(d)$, where $d + s(d) = \pi^* c$, $c \in M$ a nodal class. We put $L_3 = L_2^\perp \cap S_X \cap L^-$ and $L_4 = T_X$. Finally we let $N \subset L^-$ be the smallest primitive sublattice containing L_3 and L_4. Since the form on $S_X \cap L^-$ is negative definite we have

$$L^-_{\mathbb{R}} = L_2 \otimes \mathbb{R} \perp \underbrace{L_3 \otimes \mathbb{R} \perp L_4 \otimes \mathbb{R}}_{N_{\mathbb{R}}} .$$

Notice that the sublattice $N \subset L^-$ determines L_2, L_2 determines M' and hence L_1. We call N the *nodal type* of the marked surface Y, φ. Proposition (2.8) means of course $L_1 = L_2 = 0$ for generic Y.

2.6. **Generic Enriques Surfaces of Fixed Nodal Type.** We fix a primitive sublattice $N \subset L^-$ and consider all marked Enriques surfaces Y, φ of fixed nodal type N. Their period points $\tau(Y, \varphi)$ belong to $D^0 \cap \mathbb{P}(N_{\mathbb{C}})$. If there is at least one surface of nodal type N, then $D^0 \cap \mathbb{P}(N_{\mathbb{C}})$ is a non-empty open set in a quadric of $\mathbb{P}(N_{\mathbb{C}})$. Put $n = \operatorname{rank} N$. If $n \geqq 3$, the union of countably many hyperplanes of $\mathbb{P}(N_{\mathbb{C}})$ intersects $D^0 \cap \mathbb{P}(N_{\mathbb{C}})$ in a set with dense complement. We apply this simultaneously to two different kinds of hyperplanes.

a) The hyperplanes $d^\perp \cap N_{\mathbb{C}}$ where $d \in L$, $d \notin N^\perp$, is a nodal class satisfying $\langle d, s(d) \rangle = 0$. The period points in the complement of these hyperplanes belong to surfaces of nodal type precisely equal to N (and not smaller).

b) The hyperplanes of $\mathbb{P}(N_{\mathbb{C}})$ defined over some algebraic number field $k \subset \mathbb{C}$. In particular we take k the extension of $\mathbb{Q}$ obtained by adjoining all primitive l-th roots of unity with Euler function $\varphi(l) \leqq n$. Since $n \leqq 12$, only the following values of l occur: $l = 1, \ldots, 16, 18, 20, \ldots, 32, 36, 42$.

(2.9) **Lemma.** *Let* Y, φ *be a marked Enriques surface of nodal type* N *with period point* $\tau(Y, \varphi) \in D^0 \cap \mathbb{P}(N_{\mathbb{C}})$ *not contained in any proper linear subspace of*

$\mathbb{P}(N_{\mathbb{C}})$ *defined over* k. *Let* $g \in \mathrm{Aut}(X)$ *be an automorphism commuting with the covering involution* σ. *Then* $g^* | N = \pm \mathrm{id}_N$.

Proof. Since g commutes with σ it induces an automorphism of Y, so g^* leaves invariant all sublattices L_i, $N \subset Y$. Let ω_X be a nonzero holomorphic 2-form on X. Then $g^* \omega_X = \lambda \omega_X$ with $\lambda \in \mathbb{C}$. Since $\langle g^* \omega_X, \overline{g^* \omega_X} \rangle = \langle \omega_X, \bar{\omega}_X \rangle$, obviously $|\lambda| = 1$. But since X is projective, from [13, p. 178/179] it follows that λ is a root of unity. Since λ is an eigenvalue for $g^* | N_{\mathbb{C}}$ and $n \leqq 12$, we have $\deg[\mathbb{Q}(\lambda): \mathbb{Q}] \leqq 12$. So $\lambda^l = 1$ with $\varphi(l) \leqq 12$ and $\lambda \in k$. g^* is defined over $\mathbb{Q}$, so the λ-eigenspace for $g^* | N \otimes \mathbb{C}$ is defined over k. But then the assumption on $\tau(Y, \varphi)$ implies that this eigenspace is all of $N_{\mathbb{C}}$, i.e., $g^* | N_{\mathbb{C}} = \lambda \cdot \mathrm{id}$. This is possible only if $\lambda = \pm 1$. □

If $n = \mathrm{rank}\ N = 2$ we may embed N in the euclidean plane $\mathbb{R}^2$ such that the form on N is induced by the usual inner product on $\mathbb{R}^2$. The two isotropic subspaces in $N_{\mathbb{C}} = \mathbb{C}^2$ are generated by the vectors $(1, \pm i)$. An orientation reversing orthogonal transformation of $\mathbb{R}^2$ would interchange these two subspaces. So any isometry g of N leaving both lines $\mathbb{C} \cdot (1, \pm i)$ invariant is of the form $g = \pm \mathrm{id}$, unless N is isometric with the period lattice in $\mathbb{C} = \mathbb{R}^2$ of an elliptic curve with $\mathbb{Z}_4$ or $\mathbb{Z}_6$-symmetry. In this case g may be of order 3, 4, or 6. So with these exceptions the analog of Lemma (2.9) for $n = 2$ also holds.

3. Generic Enriques Surfaces

3.1. **Automorphisms of Enriques Surfaces.** As above, fix a marking $H^2(X, \mathbb{Z}) = L$, $H^2(Y, \mathbb{Z})_f = M$ of the Enriques surface Y. Recall that $\Gamma \subset O(L)$ is the subgroup of isometries commuting with s and $O^\uparrow(X) \subset O(L)$ the subgroup of isometries g for which $g_{\mathbb{R}}$ leaves invariant $\mathscr{C}_X^+ \subset L_{\mathbb{R}}$. Let $O^\uparrow(Y)$ denote the group of isometries $g \in O(M)$ with $g_{\mathbb{R}}$ leaving invariant $\mathscr{C}_Y^+ \subset M_{\mathbb{R}}$. Each automorphism $\alpha \in \mathrm{Aut}(Y)$ lifts (in two ways) to an automorphism $\tilde{\alpha} \in \mathrm{Aut}(X)$ commuting with s. So if we put

$$\mathrm{Aut}(X, \sigma) = \{g \in \mathrm{Aut}(X): g\sigma = \sigma g\}$$

then $\mathrm{Aut}(Y) = \mathrm{Aut}(X, \sigma)/\{\mathrm{id}, \sigma\}$. Any $g \in \mathrm{Aut}(X)$ commuting with s leaves invariant the sublattices $L_1, \ldots, L_4 \subset L$ defined in 2.5. So we have canonical maps r_i: $\mathrm{Aut}(X, \sigma) \to O(L_i)$, $i = 1, \ldots, 4$, and an embedding (r^+, r_2, r_3, r_4): $\mathrm{Aut}(X, \sigma) \to O(L^+) \times O(L_2) \times O(L_3) \times O(L_4)$ with $r^- = (r_2, r_3, r_4)$.

(3.1) **Proposition.** *Under* r^+ *the kernel of* r^-: $\mathrm{Aut}(X, \sigma) \to O(L^-)$ *is identified with the 2-congruence subgroup of* $O^\uparrow(Y)$.

Proof. By Lemma (1.3) the kernel of $r^- = \Gamma \to O(L^-)$ is under r^+ identified with the 2-congruence subgroup of $O(L^+)$. It is clear that $r^+(\gamma)_{\mathbb{R}}$ leaves invariant the ample cone $\mathscr{C}_Y^+$ for every $\gamma \in \mathrm{Aut}(X, \sigma)$. So we only have to show the converse: if $\gamma \in \Gamma$ with $r^-(\gamma) = \pm \mathrm{id}$ and $r^+(\gamma)_{\mathbb{R}}$ leaving invariant $\mathscr{C}_Y^+$, then γ is induced by an automorphism of X. But take some arbitrary ample divisor C on Y with class $c \in C_Y^+ \cap H^2(Y, \mathbb{Z})_f$. Then $r^+(\gamma) c$ is the class of an ample divisor again. The class $\pi^* c \in L^+$ is the class of the ample divisor $\pi^* C$ and $\gamma(\pi^* c) = \pi^* r^+(\gamma) c$ too.

So $\gamma \in O^\uparrow(X)$ by Lemma (2.3). Since $r^-(\gamma) = \pm \mathrm{id}$, the extension $\gamma_{\mathbb{C}}$ leaves invariant the subspace $\mathbb{C}\omega_X \subset L^-_{\mathbb{C}}$. The assertion follows from Theorem (2.5). □

(3.2) **Proposition.** *The image of* r^-: $\mathrm{Aut}(X, \sigma) \to O(L^-)$ *is a finite group.*

Proof. The form on L^- has signature (2, 10). For each $\gamma \in \Gamma(Y)$ we have $\gamma_{\mathbb{C}}\omega_X = \lambda\omega_X$, $\lambda \in \mathbb{C}$. If $\omega_X = u + iv$, $u, v \in L^-_{\mathbb{R}}$, this shows that $\gamma_{\mathbb{R}}$ leaves invariant the 2-dimensional real subspace $\mathbb{R}u \oplus \mathbb{R}v \subset L^-_{\mathbb{R}}$, on which the form is positive definite. On $(\mathbb{R}u \oplus \mathbb{R}v)^\perp \subset L^-_{\mathbb{R}}$ the form is negative definite. So $r^-\gamma$ belongs to the compact group $O(\mathbb{R}u \oplus \mathbb{R}v) \times O((\mathbb{R}u \oplus \mathbb{R}v)^\perp)$. □

Combining propositions (3.1) and (3.2) we find

(3.3) **Theorem.** *For every Enriques surface* Y *the automorphism group* $\mathrm{Aut}(Y)$ *contains the 2-congruence subgroup of* $O^\uparrow(Y)$ *as normal subgroup of finite index.*

Next we consider generic Enriques surfaces and obtain the main result of this paper.

(3.4) **Theorem.** *Let* Y, φ *be a marked Enriques surface with period point* $\tau(Y, \varphi) \in D^0$ *not contained in any proper linear subspace of* $\mathbb{P}(L^-_{\mathbb{C}})$ *defined over the number field* k *from Lemma* (2.9). *Then the representation of* $\mathrm{Aut}(Y)$ *on* $H^2(Y, \mathbb{Z})_f = M$ *identifies* $\mathrm{Aut}(Y)$ *with the 2-congruence subgroup of* $O^\uparrow(M)$.

Proof. The assumption implies $\tau(Y, \varphi) \in D_{\mathrm{gen}}$. So $O^\uparrow(Y) = O^\uparrow(M)$. Also Lemma (2.9) shows that $r^-(\mathrm{Aut}(X, \sigma)) = \pm \mathrm{id}$, and the assertion follows from Proposition (3.1). □

Now if a generic surface Y is deformed into less generic ones, the following phenomena, working against each other, can happen.

- Y acquires nodal curves, so $O^\uparrow(Y)$ and $\mathrm{Aut}(Y)$ probably too, become smaller.
- Y acquires nodal curves and/or S_X becomes bigger, ω_X becomes more special, hence $r^-(\mathrm{Aut}(X, \sigma))$ and $\mathrm{Aut}(Y)$ probably too grow bigger.

We do not know, whether one can control these effects. In particular we do not know the Enriques surface with the "biggest" or the "smallest" automorphism group.

3.2. **Computation of Some Stabilizer Groups.** Denote by $G \subset O^\uparrow(M)$ the 2-congruence subgroup. It is the purpose of this section, to compute the stabilizer subgroups $G_c \subset G^\uparrow(M)_c$ for certain elements $c \in M$. An element $c \in M$ will be called

primitive, if $\mathbb{Z}c \subset M$ is a primitive sublattice

0-*class*, if $c^2 = 0$

forward pointing, if $c \in \bar{\mathscr{C}}_Y^+$ (here and in the sequel fix an isomorphism $M = H^2(Y, \mathbb{Z})_f$)

fwp, if c is forward pointing and primitive.

In particular we consider elements

$$c = e_1 + \ldots + e_n$$

where $e_1, \ldots, e_n$ are 0-classes satisfying

(3.5) $e_i \cdot e_j = 1$ for $i \neq j$.

(3.6) **Lemma.** i) *If* $n \geqq 2$, *then* (3.5) *implies that all* e_i *are primitive.*

ii) *If* $n \geqq 2$ *and one* e_i *is fwp, then so are all.*

iii) *If* $c = e_1 + \ldots + e_n = e'_1 + \ldots + e'_n$ *with* $e_1, \ldots, e_n$ *and* $e'_1, \ldots, e'_n$ *fwp* 0*-classes satisfying* (3.5), *then up to a permutation we have* $e_1 = e'_1, \ldots, e_n = e'_n$.

Proof. The assertions i) and ii) being obvious, we prove iii). We compute $c \cdot e'_1$ in two ways:

$$(e_1 + \ldots + e_n) e'_1 = \sum_1^n e_i e'_1 \quad \text{and} \quad (e'_1 + \ldots + e'_n) \cdot e'_1 = n - 1.$$

Since e_i and e'_1 both are *fwp*, we have $e_i \cdot e'_1 \geqq 0$ and $e_i \cdot e'_1 = 0$ only if $e_i = e'_1$. This shows $e_i = e'_1$ for some i. The assertion follows by induction on n. □

(3.7) **Proposition.** *For* $n = 1, 2, 3$ *or* 10 *the group* $O^\uparrow(M)$ *operates transitively on the set of ordered n-tuples* $e_1, \ldots, e_n \in M$ *consisting of fwp* 0*-classes satisfying* (3.5).

Proof. $n = 1$. Let $e_1 \in M$ be an arbitrary *fwp* 0-class. Since M is unimodular, there is some $c \in M$ with $e_1 \cdot c = 1$. Put $k = \frac{1}{2} c^2$ and $e_2 = c - k e_1$. Then $e_2^2 = 0$ and $e_1 \cdot e_2 = 1$. So e_1, e_2 are *fwp* generators of a sublattice $\mathbb{H} \subset M$. Since e_1 and e_2 may be permuted by some $g \in O^\uparrow(M)$, the assertion will follow from the case $n = 2$.

$n = 2$. It suffices to show that $O^\uparrow(\mathrm{M})$ operates transitively on the set of sublattices $\mathbb{H} \subset M$. But when an embedding $\mathbb{H} \subset M$ is given, we have $M = \mathbb{H} \perp \mathbb{H}^\perp$ with $\mathbb{H}^\perp$ unimodular, even, and negative definite, hence $\mathbb{H}^\perp = \mathbb{E}$. Then there is some $g \in O^\uparrow(M)$ mapping this decomposition $M = \mathbb{H} \perp \mathbb{E}$ into the standard one.

$n = 3$. From the case $n = 2$ it follows that we may assume e_1, e_2 to be the standard generators of $\mathbb{H}$ in the standard decomposition $M = \mathbb{H} \perp \mathbb{E}$. Then $e_3 = e_1 + e_2 + w$ with $w \in \mathbb{E}$ some root. The assertion follows from the well-known fact that the Weyl group $W(E_8)$ operates transitively on the roots of E_8.

$n = 10$. We take $e_1, e_2 \in \mathbb{H} \subset M$ as in the case $n = 3$ and for $i = 3, \ldots, 10$ we have $e_i = e_1 + e_2 + w_i$ with roots $w_i \in \mathbb{E}$ satisfying

$$w_i \cdot w_j = -1 \qquad \text{whenever } i \neq j.$$

The assertion follows from Lemma (1.6). □

Now we consider sums $c = e_1 + \ldots + e_n$ of *fwp* 0-classes $e_i \in M$ satisfying (3.5). For $n = 1, 2, 3$, and 10 we saw that $O^\uparrow(M)$ operates transitively on such c. It follows from Lemma (1.1) and Lemma (3.6) iii) that

$$O^\uparrow(M)_c = \mathfrak{S}_n \times O(e_1^\perp \cap \ldots \cap e_n^\perp),$$

$\mathfrak{S}_n$ the permutation group of degree n. We are interested in the number $N(n)$ = number of G-orbits of elements c.

Clearly the stabilizer subgroup G_c is $G \cap O^\uparrow(M)_c$, and since G is a normal subgroup in $O^\uparrow(M)$, all G-orbits are equivalent under $O^\uparrow(M)$. So the set of G-orbits is a homogeneous space under $O^\uparrow(M)/G = O^+(10, \mathbb{F}_2)$ and

$$N(n) = |O^+(10, \mathbb{F}_2)| / [O^\uparrow(M)_c : G_c].$$

The results are given in the following table

n	$O^\uparrow(M)_c$	G_c	$[O^\uparrow(M)_c : G_c]$	$N(n)$
1	$\mathbb{E} \times W_8$	$2\mathbb{E} \times \{\pm 1\}$	$2^{21} \cdot 3^5 \cdot 5^2 \cdot 7$	$17 \cdot 31 =$ 527
2	$\mathfrak{S}_2 \times W_8$	$\{\pm 1\}$	$2^{14} \cdot 3^5 \cdot 5^2 \cdot 7$	$2^7 \cdot 17 \cdot 31 =$ 67456
3	$\mathfrak{S}_3 \times W_7$	1	$2^{11} \cdot 3^5 \cdot 5 \cdot 7$	$2^{10} \cdot 5 \cdot 17 \cdot 31 =$ 2698240
10	$\mathfrak{S}_{10}$	1	$10! = 2^8 \cdot 3^4 \cdot 5^2 \cdot 7$	$2^{13} \cdot 3 \cdot 17 \cdot 31 =$ 12951552

Proofs. Recall (Sect. 1.4) that $|O^+(10, \mathbb{F}_2)| = 2^{21} \cdot 3^5 \cdot 5^2 \cdot 7 \cdot 17 \cdot 31$. To compute the stabilizer groups we use the standard decomposition $M = \mathbb{H} \perp \mathbb{E}$ with $\mathbb{H} = \mathbb{Z}e_1 \oplus \mathbb{Z}e_2$. Any $g \in O^\uparrow(M)$ has a matrix decomposition

$$g = \left(\begin{array}{c|c} g_H & g_{EH} \\ \hline g_{HE} & g_E \end{array}\right) \qquad \begin{array}{ll} g_H : \mathbb{H} \to \mathbb{H}, & g_E : \mathbb{E} \to \mathbb{E} \\ g_{EH} : \mathbb{E} \to \mathbb{H}, & g_{HE} : \mathbb{H} \to \mathbb{E}. \end{array}$$

$n = 1$. Assume $ge_1 = e_1$. Then

$$g = \left(\begin{array}{cc|c} 1 & t & f \\ 0 & 1 & 0 \\ \hline 0 & y & g_E \end{array}\right) \qquad \begin{array}{l} t \in \mathbb{Z}, \quad y \in \mathbb{E} \\ f : \mathbb{E} \to \mathbb{Z} \end{array}$$

and orthogonality of g is equivalent with

$$t = -\tfrac{1}{2} y^2, \quad g_E \in W_8, \quad f = -\langle g_E y, \rangle.$$

So $O^\uparrow(M)_{e_1}$ is $\mathbb{E} \times W_8$, the extended Weyl group, under the identification

$$\mathbb{E} \times W \ni (y, h) \mapsto \left(\begin{array}{cc|c} 1 & -\frac{1}{2} y^2 & -\langle hy, \rangle \\ 0 & 1 & 0 \\ \hline 0 & y & h \end{array}\right).$$

It is known (Sect. 1.4) that the 2-congruence subgroup of W_8 is just $\{\pm 1\}$. So $G_c = 2\mathbb{E} \times \{\pm 1\}$.

$n = 2$. If $c = e_1 + e_2$, we have $O^\uparrow(M)_c = \mathfrak{S}_2 \times W_8$ with $\mathfrak{S}_2$ permuting e_1 and e_2. Since a nontrivial permutation of e_i's cannot belong to G, we have $G_c = \{\pm 1\}$, the 2-congruence subgroup of W_8.

$n = 3$. If $c = e_1 + e_2 + e_3$ and $e_3 = e_1 + e_2 + w$ as above with some root $w \in \mathbb{E}$, then $\mathbb{Z}e_1^\perp \cap \mathbb{Z}e_2^\perp \cap \mathbb{Z}e_3^\perp$ is $w^\perp \subset \mathbb{E}$, the orthogonal complement of a root. This $w^\perp$ is isomorphic with the lattice $(-E_7)$. So $O^\uparrow(M)_c = \mathfrak{S}_3 \times W_7$ with $\mathfrak{S}_3$ permuting e_1, e_2, e_3 and W_7 the Weyl group of E_7. In particular G_c is trivial.

$n = 10$. Obvious. □

3.3. Representations of Generic Enriques Surfaces

a) *Elliptic Pencils.* In this section let Y be an Enriques surface, general in the sense that Theorem (3.4) applies. In particular, there are no (-2)-curves on Y.

We fix an isomorphism $H^2(Y,\mathbb{Z})_f = M$. It is classical that (because of the absence of (-2)-curves) each fwp 0-class $e \in M$ is effective. There are exactly two curves E and E' representing e. They are either nonsingular elliptic or irreducible rational with a node or a cusp, and adjoint in the sense that $K_Y = \mathcal{O}_Y(E - E')$. The linear system $|2E|$ is an elliptic pencil without base points. It provides Y with the structure of an elliptic fibre space over $\mathbb{P}_1$. This fibration contains exactly two multiple fibres, namely $2E$ and $2E'$. (The divisors E and E' are called the half-pencils of $|2E|$.) Each elliptic fibration over $\mathbb{P}_1$ is defined by such a linear system $|2E|$ with $e \in M$ some fwp 0-class. We shall not distinguish between two elliptic fibrations $Y \to \mathbb{P}_1$ differing by an automorphism of $\mathbb{P}_1$. So two elliptic fibre spaces $Y \to \mathbb{P}_1$ are isomorphic (as fibre spaces, modulo $\mathrm{Aut}(\mathbb{P}_1)$) if and only if they differ by an automorphism of X. So the different representations of X as elliptic fibrations over $\mathbb{P}_1$ correspond 1 to 1 with G-orbits of fwp 0-classes $e \in M$. From 6.5 we conclude

(3.8) **Theorem.** *For a general Enriques surface Y there are exactly $17 \cdot 31$ nonisomorphic elliptic fibre spaces over $\mathbb{P}_1$ with total space Y.*

b) *Double Plane Representations.* Now we consider pairs $e_1, e_2 \in M$ of fwp 0-classes with $e_1 \cdot e_2 = 1$. We let E_i, E_i' be the curves representing e_i, we put $D = 2(E_1^{(\prime)} + E_2^{(\prime)})$ and consider the linear system $|D|$. It is known [3, Theorem 6.1] that this linear system defines a 2 to 1 map of Y onto a "quartic Del Pezzo surface Q of Segre symbol (11)(11)1", i.e., a surface Q in $\mathbb{P}_4$ projectively equivalent to the complete intersection $z_0^2 = z_1 z_2 = z_3 z_4$ of two rank-3 quadrics. The map is ramified over the four nodes of Q and a complete intersection curve $B = Q \cap Z$ with some quadric $Z \subset \mathbb{P}_4$ not passing through any of the four nodes of Q. The absence of (-2)-curves on Y forces B to be nonsingular. The double cover $Y \to Q$ is related to Horikawa's representation of Y [7, 8] through the commutative diagram

$$\begin{array}{ccc} X & \longrightarrow & Y \\ \downarrow & & \downarrow \\ \mathbb{P}_1 \times \mathbb{P}_1 & \longrightarrow & Q \end{array}$$

where $\mathbb{P}_1 \times \mathbb{P}_1 \to Q$ is a double cover ramified over the four nodes of Q.

Up to an automorphism of $\mathbb{P}_4$ the map defined by $|D|$ is uniquely determined by the two classes $e_1, e_2 \in M$. We consider two double plane representations of Y as equivalent, if the pairs (Q, B) defining them differ by an automorphism of $\mathbb{P}_4$, i.e., if the classes D differ by an automorphism of Y. From Sect. 3.2 we obtain:

(3.9) **Theorem.** *For a general Enriques surface there are exactly $2^7 \cdot 17 \cdot 31$ inequivalent double plane representations.*

c) *Enriques Representations.* Let $e_1, e_2, e_3 \in M$ be a triplet of fwp 0-classes satisfying $e_i \cdot e_j = 1$ for $i \neq j$, let E_i, E_i' be the curves representing e_i, and put $D = E_1$

$+E_2+E_3$ (defined uniquely by $d=e_1+e_2+e_3$ up to the ambiguity between D and $D'=D+K_y$). It is known [3, Theorem 7.4] that $|D|$ defines a birational map of Y onto a sextic surface in $\mathbb{P}_3$ passing doubly through the edges of a tetrahedron. The image surfaces are projectively equivalent if and only if the linear systems $|D|$ differ by an automorphism g of Y. Using the double-plane representation it is easy to see that for general Y (i.e., general choice of the branch curve $B\subset Q$) there is no automorphism $g\in G$ leaving e_1, e_2 invariant and interchanging E_1 and E_1'. This shows that in general the systems $|D|$ and $|D'|$ have projectively inequivalent images. From Sect. 3.2 we conclude:

(3.10) **Theorem.** *For a general Enriques surface there are exactly* $2^{11}\cdot 5\cdot 17\cdot 31$ *inequivalent Enriques representations.*

d) *Representations as Surfaces of Degree* 10 *in* $\mathbb{P}_5$. Let $e_1,\ldots,e_{10}$ be *fwp* 0-classes satisfying $e_1\cdot e_j=1$ for $i\neq j$ and let E_i, E_i' be the curves representing e_i. We consider the linear system $|D|$ with class

$$d=\tfrac{1}{3}\sum_1^{10} e_i.$$

(Notice that because of the explicit form of the e_i given in Sect. 3.2 and Lemma (1.6) one easily checks that $\sum_1^{10} e_i$ in M is divisible by 3.) It is known [4, 3.2.1 iii) and 3.3.2] that there are (special) Enriques surfaces carrying such e_i with $|D|$ defining an embedding of Y in $\mathbb{P}_5$ of degree 10 ("Reye-congruences"). So for general Y, the system $|D|$ will also define such an embedding $Y\to\mathbb{P}_5$. As above one proves that $|D|$ and $|D'|$ define projectively inequivalent embeddings. So we conclude from Sect. 3.2

(3.11) **Theorem.** *For a general Enriques surface there are exactly* $2^{14}\cdot 3\cdot 17\cdot 31$ *embeddings in* $\mathbb{P}_5$, *defined by linear systems* $|D|=|\frac{1}{3}\sum E_i|$ *as above, as* 10*th degree surfaces which are projectively inequivalent.*

4. Examples of Enriques Surfaces with Small Automorphism Group

In this section we use the double plane presentation of Enriques surfaces to compute explicitly the automorphism groups for some examples. The observations that $\mathrm{Aut}(Y)$ is finite in case 3 below is due to Dolgachev [6].

4.1. **The Branch Curve.** Denote by Q the quadric $\mathbb{P}_1\times\mathbb{P}_1$ and let $((u_0:u_1),(v_0:v_1))$ be bihomogeneous coordinates on Q. By a line on Q we mean a smooth rational curve belonging to one of the two rulings on Q.

Take constants $a,b,c,d\in\mathbb{C}$ satisfying

$$a\neq 0,\quad c\neq 0\neq d,\quad c\neq d$$

and consider the curve $B\subset Q$ of bidegree (4, 4) with equation

$$(v_0^2-v_1^2)\{a(v_0^2-v_1^2)u_0^4+2b(v_0^2-v_1^2)u_0^2u_1^2+(cv_0^2-dv_1^2)u_1^4\}=0.$$

Then B splits as $B=N^+ + N^- + C$ with the two lines

$$N^{\pm}: v_0 \pm v_1 = 0$$

and C a curve of bidegree (4, 2). This curve C meets the line $N^{\pm}$ at

$$P^{\pm} = (1:0), (1: \mp 1) \in N^{\pm}$$

with multiplicity 4. Since $a \neq 0$, C is smooth in these points, so they are A_7-singularities on B.

It turns out that one has to distinguish between the following three cases:

Case 1 (general case): $b \neq 0$, $ac \neq b^2 \neq ad$.
Case 2 (symmetric case): $b = 0$.
Case 3 (special case): $ac = b^2$ or $ad = b^2$.

Here Case 3 leads to the surface first considered by Dolgachev [6].

(4.1) **Lemma.** *Each line* $(v_0 : v_1) = \text{const} \neq \pm 1$ *meets* C *at four distinct points, unless it is one of the two lines*

$$L^{\pm}: (v_0 : v_1) = (\sqrt{ad-b^2} : \pm\sqrt{ac-b^2}).$$

Case 1: *The two lines are different and* C *is smooth at the two distinct points of contact.*
Case 2: *The two lines are different and* C *meets them with multiplicity four at a smooth point of* C.
Case 3: *The two lines coincide and* C *has two ordinary nodes on this line.*

(4.2) **Corollary.** *Away from* $P^{\pm}$ *the curve* B *is smooth in case* 1 *and* 2, *and has two* A_1*-singularities on the line* $L^+ = L^-$ *in case* 3.

Proof of the Lemma. An arbitrary line L with equation $v_0 : v_1 = t_0 : t_1 \neq \pm 1$ intersects C at four distinct points unless $a(ct_0^2 - dt_1^2) = b^2(t_0^2 - t_1^2)$, i.e.

$$(ac-b^2)t_0^2 = (ad-b^2)t_1^2.$$

This condition determines the lines $L^{\pm}$. The restriction of C of $L^{\pm}$ has equation $(au_0^2 + bu_1^2)^2 = 0$, so the points of contact are

$$((u_0 : u_1), (v_0 : v_1)) = ((\sqrt{b} : \pm i\sqrt{a}), (\sqrt{ad-b^2} : \pm\sqrt{ac-b^2})).$$

In these points we differentiate the equation for C

$$\begin{aligned} \partial/\partial v_1 &= -2av_1 u_0^4 - 4bv_1 u_0^2 u_1^2 - 2dv_1 u_1^4 \\ &= -2v_1(ab^2 - 2ab^2 + da^2) \\ &= -2av_1(ad-b^2). \end{aligned}$$

So C is smooth here in case 1 and 2, but singular in case 3.

In case 3 let e.g. $ac=b^2$, hence $L^+=L^-$ is the line $v_1=0$. We use inhomogeneous coordinates $u=u_1/u_0$ and $v=v_1/v_0$ to form partial derivatives of the equation for C in the points $\left(\left(1:\pm i\sqrt{\frac{a}{b}}\right),(1:0)\right)$.

$$\partial^2/\partial u^2=4b+12cu^2=4\left(b-3c\frac{a}{b}\right)=-8b\neq 0,$$
$$\partial^2/\partial u\partial v=0,$$
$$\partial^2/\partial v^2=-2a-4bu^2-2du^4$$
$$=-2\left(a-2a+d\frac{a^2}{b^2}\right)=-2a(ad-b^2)\neq 0.$$

So the singularities are ordinary nodes. □

The equation for B is invariant under the group $\mathbb{Z}_2\times\mathbb{Z}_2$ generated by

$$\tau_1:((u_0:u_1),(v_0:v_1))\mapsto((u_0:-u_1),(v_0:v_1)),$$
$$\tau_2:((u_0:u_1),(v_0:v_1))\mapsto((u_0:u_1),(v_0:-v_1)).$$

We put

$$\tau=\tau_1\tau_2.$$

If $b=0$ the group $\mathbb{Z}_2\times\mathbb{Z}_2$ can be enlarged to $\mathbb{Z}_4\times\mathbb{Z}_2$ generated by

$$\rho:((u_0:u_1),(v_0:v_1))\mapsto((u_0:iu_1),(v_0:v_1))$$

and τ_2, i.e., $\rho^2=\tau_1$.

(4.3) **Lemma.** *In case* 1 *and* 3 *the group* $\mathbb{Z}_2\times\mathbb{Z}_2$ *generated by* τ_1,τ_2 *and in case* 2 *the group* $\mathbb{Z}_4\times\mathbb{Z}_2$ *generated by* ρ,τ_2 *is the full automorphism group of the pair* $B\subset Q$.

Proof. Any automorphism α of (B,Q) respects the pair of lines $N^{\pm}$, hence does not interchange u and v. Therefore $\alpha=\alpha_1\alpha_2$ with α_1 acting on u and α_2 on v. Additionally α respects the line $u_1=0$, the pair $L^{\pm}$ and the pair of lines $au_0^2+bu_1^2=0$.

Case 1 and 3. Here the equation $au_0^2+bu_1^2=0$ defines two distinct lines, interchanged by τ_1. This implies $\alpha_1=\mathrm{id}$ or $\alpha_1=\tau_1$. Now either α_2 or $\alpha_2\tau_2$ leaves invariant both the points $(1:\pm 1)$, so it is of the form

$$(v_0:v_1)\mapsto(t_0v_0+t_1v_1:t_1v_0+t_0v_1),\qquad t_0^2\neq t_1^2.$$

This substitution changes

$$v_0^2-v_1^2\mapsto(t_0^2-t_1^2)(v_0^2-v_1^2),$$
$$cv_0^2-dv_1^2\mapsto(ct_0^2-dt_1^2)v_0^2+2(c-d)t_0t_1v_0v_1+(ct_1^2-dt_0^2)v_1^2,$$

and the invariance of C under $\alpha_1\alpha_2$ implies first $t_0t_1=0$ and then $t_1=0$. So either $\alpha_2=\mathrm{id}$ or $\alpha_2=\tau_2$.

Case 2. Now the line $u_0=0$ is fixed under α, because on it C touches $L^{\pm}$. This implies

$$\alpha_1:(u_0:u_1)\mapsto(u_0:su_1),\qquad s\neq 0.$$

As in cases 1 and 3 either α_2 or $\alpha_2\tau_2$ is of the form

$$(v_0:v_1)\mapsto(t_0v_0+t_1v_1:t_1v_0+t_0v_1).$$

The invariance of the point pair $cv_0^2=dv_1^2$ implies $t_1=0$, so $\alpha_2=\mathrm{id}$ or $\alpha_2=\tau_2$. Considering the equation for C we find $s^4=1$, hence α_1 is a power of ρ. □

4.2. **The K3-Surface X.** Let $\overline{X}\to Q$ be the double covering branched over B and $q: X\to Q$ its minimal desingularisation. On X we have the following curves:

In all three cases:

$F_1^{\pm},\ldots,F_7^{\pm}$	(-2)-curves resolving the A_1-singularities over $P^{\pm}$
$F_8^{\pm}$	two (-2)-curves over the line $u_1=0$
$N^{\pm}$	two (-2)-curves in the branch locus
F	smooth elliptic curve over $u_0=0$

In case 1 and 2 additionally:

$L_1^{\pm}, L_2^{\pm}$	(-2)-curves over $L^{\pm}$ ($L_1^{\pm}$ and $L_2^{\pm}$ touch in case 2)
E, E'	smooth elliptic curves over $v_0=0$, $v_1=0$

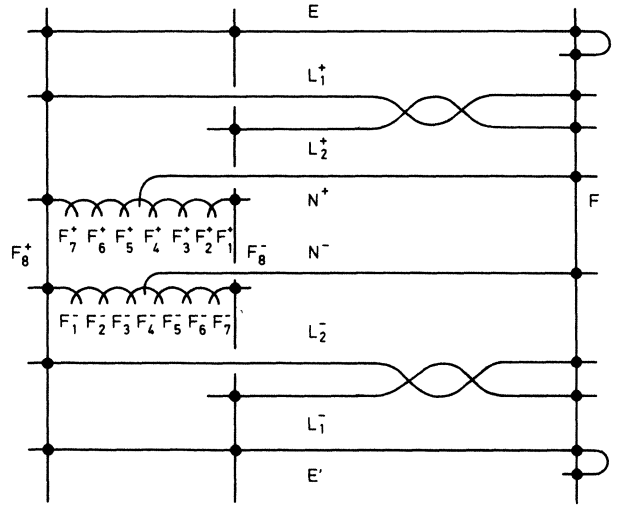
In case 3 additionally:

$E_1^{\pm}$	two (-2)-curves over the line $L^+=L^-$
$E_2^{\pm}$	(-2)-curves resolving the A_1-singularities of $\overline{X}$
E'	smooth elliptic curve over $v_0=0$, resp. $v_1=0$.

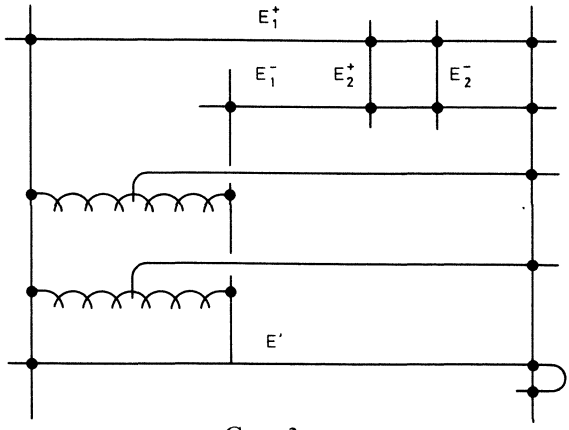
Let $\sigma_3\in\mathrm{Aut}(X)$ be the covering involution interchanging the two sheets of $q: X\to Q$. The automorphisms of Q from Lemma (4.3) lift to X in the following way:

σ_1 is an involution lifting τ_1 and having F, $F_2^{\pm}$, $F_4^{\pm}$, $F_6^{\pm}$, and $F_8^{\pm}$ as curves of fixed points.

σ_2 is an involution lifting τ_2 and having E (resp. $E_1^{\pm}$) and E' as curves of fixed points.



Case 1 (in case 2, $L_1^{\pm}$ and $L_2^{\pm}$ touch on F)



Case 3

σ_1 and σ_2 commute with σ_3, so the involutions $\sigma_1, \sigma_2, \sigma_3$ generate a subgroup $(\mathbb{Z}_2)^3 \subset \mathrm{Aut}(X)$. This group contains in particular $\sigma = \sigma_1\sigma_2\sigma_3$, the involution without fixed points.

In case 2 this group is enlarged by lifting ρ to $\tilde{\rho} \in \mathrm{Aut}(X)$ with $F_8^{\pm}$ being curves of fixed points for $\tilde{\rho}$. Then necessarily $F_2^{\pm}$, $F_4^{\pm}$ and $F_6^{\pm}$ are curves of fixed points, and $\tilde{\rho}|F$ is an involution with four isolated fixed points. Further $(\tilde{\rho})^2 = \sigma_1, \tilde{\rho}$ commutes with σ_2 and σ_3, and we have a subgroup $\mathbb{Z}_4 \times (\mathbb{Z}_2)^2$ in $\mathrm{Aut}(X)$.

Notice that the involutions σ_1 and $\sigma_1\sigma_3$ interchange E_2^+ and E_2^-. The involution $\sigma_2\sigma_3$ interchanges all other pairs of curves differing by a $\pm$-sign.

4.3. **The Elliptic Pencil $|F|$ on X.** The elliptic curve F in X is linearly equivalent with $\sum_1^8 (F_i^+ + F_i^-)$. We denote by $\phi: X \to \mathbb{P}_1$ the elliptic fibration defined by the pencil $|F|$. We know already the following sections for this pencil:

$$N^+, N^-, L_1^+, L_1^-, L_2^+, L_2^- \quad \text{(case 1 and 2)}$$
$$N^+, N^-, E_1^+, E_1^- \quad \text{(case 3)}.$$

We denote by $\mathfrak{S}$ the set of all sections and introduce on it the structure of an abelian group by distinguishing N^- as origin. For any of the curves $N^-, N^+, L_i^{\pm}, E_1^{\pm}$ we denote the corresponding group element by $\mathfrak{o}, \mathfrak{n}, \mathfrak{l}_i^{\pm}, \mathfrak{e}^{\pm}$.

With N^- as origin the 2-torsion elements on every elliptic curve in $|F|$ are the intersection points with the ramification divisor of q. Since C does not split, the only non-trivial 2-torsion element in $\mathfrak{S}$ is $\mathfrak{n}$. The involution $\sigma_2\sigma_3$ acts on $\mathfrak{S}$ as addition by $\mathfrak{n}$.

(4.4) **Proposition.** *The torsion subgroup $\mathfrak{S}_{\mathrm{tors}} \subset \mathfrak{S}$ is*

$$\mathbb{Z}_2 \text{ generated by } \mathfrak{n} \quad (\text{cases 1 and 2})$$
$$\mathbb{Z}_4 \text{ generated by } \mathfrak{e}^{\pm} \quad (\text{case 3}).$$

Proof. If $\mathfrak{s} \in \mathfrak{S}$ is any torsion element represented by the section $S \subset X$, then $qS \subset Q$ is a smooth rational curve of bidegree $(1, n)$. If qS intersects B in a smooth point, then necessarily $2\mathfrak{s}=0$, i.e., $S=N^{\pm}$. The only way to avoid meeting B in a smooth point (on N^+ or N^-) is $n=0$ and $qS=q(E_1^{\pm})$ in the case 3. So $S=E_1^+$ or E_1^-. Since $\mathfrak{S}$ contains only one 2-torsion element $\mathfrak{n} \neq 0$, necessarily $\mathfrak{S}_{\text{tors}} = \mathbb{Z}_4$ generated by $\mathfrak{e}^+$ or $\mathfrak{e}^-$ in this case. □

(4.5) **Corollary.** *In case* 1 *and* 2 *we have* $\operatorname{rank} \mathfrak{S} \geqq 1$.

In fact $\mathfrak{S}$ contains the non-torsion elements $\mathfrak{l}_i^{\pm}$. Notice that $\mathfrak{n} + \mathfrak{l}_i^+ = \mathfrak{l}_i^-$ and $-\mathfrak{l}_1^{\pm} = \mathfrak{l}_2^{\pm}$.

The automorphisms σ_1 (resp. $\tilde{\rho}$ in case 2) and $\sigma_3 \in \operatorname{Aut}(X)$ respect $|F|$ and leave N^- fixed.

(4.6) **Proposition.** *The subgroup of* $\operatorname{Aut}(X)$ *respecting* $|F|$ *and leaving* N^- *fixed is*

$$\mathbb{Z}_2 \times \mathbb{Z}_2 \text{ generated by } \sigma_1 \text{ and } \sigma_3 \quad (\text{case } 1 \text{ and } 3),$$
$$\mathbb{Z}_4 \times \mathbb{Z}_2 \text{ generated by } \tilde{\rho} \text{ and } \sigma_3 \quad (\text{case } 2).$$

Proof. Any $\alpha \in \operatorname{Aut}(X)$ respecting the pencil $|F|$ fixes the cycle $\sum_1^8 (F_i^+ + F_i^-)$. If $\alpha N^- = N^-$, then (after replacing α by $\alpha\sigma_3$) we may assume $\alpha F_i^+ = F_i^+$, $\alpha F_i^- = F_i^-$ for $i=1, \ldots, 8$. So α leaves invariant the $\tilde{E}_7$-fundamental cycle

$$Z = F_1^- + 2F_2^- + 3F_3^- + 4F_4^- + 3F_5^- + 2F_6^- + F_7^- + 2N^-$$

linearly equivalent with E (resp. $E_1^+ + E_1^- + E_2^+ + E_2^-$) and E'. This means that the map $X \to \mathbb{P}_1$ defined by the elliptic pencil $|E'|$ is α-equivariant too, hence α is induced by some symmetry of (B, Q). The assertion follows from Lemma (4.3). □

This group $\mathbb{Z}_2 \times \mathbb{Z}_2$ (resp. $\mathbb{Z}_4 \times \mathbb{Z}_2$) acts naturally on $\mathfrak{S}$. On the sections given above this action can be traced easily:

σ_3 being the covering involution induces $-\mathrm{id}$ on all elliptic curves in the pencil $|F|$, hence acts on $\mathfrak{S}$ as $-\mathrm{id}$.

σ_1 (resp. $\tilde{\rho}$) leaves invariant each of the sections $\mathfrak{n}$, $\mathfrak{l}_i^{\pm}$, $\mathfrak{e}^{\pm}$ so it acts trivially on $\mathfrak{S}_{\text{tors}}$ and on the rank-1 subgroup generated by the $\mathfrak{l}_i^{\pm}$.

Now let

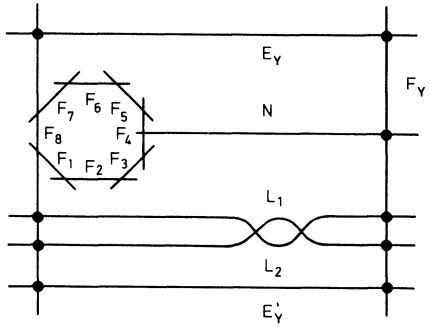
$$\mathfrak{R} = (\mathbb{Z}_2 \times \mathbb{Z}_2) \ltimes \mathfrak{S}, \qquad \text{resp. } (\mathbb{Z}_4 \times \mathbb{Z}_2) \ltimes \mathfrak{S}$$

be the semidirect product w.r. to this action.

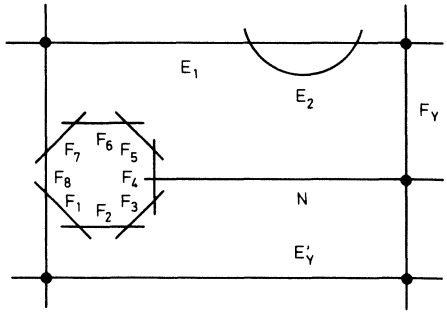
(4.7) **Proposition.** $\mathfrak{R} \subset \operatorname{Aut}(X) \subset O(L)$ *is just the stabilizer subgroup of the class* $f \in L$ *of* F.

Proof. Assume $\alpha f = f$ for some $\alpha \in \operatorname{Aut}(X)$. Then $\phi: X \to \mathbb{P}_1$ is α-equivariant. After replacing α by $\alpha \circ \{\text{translation by } -\alpha(\mathfrak{o}) \in \mathfrak{S}\}$ we may assume $\alpha N^- = N^-$. The assertion follows from Proposition (4.6) above. □

4.4. **The Enriques Surface** $Y = X/\sigma$. Since $\sigma = \sigma_1\sigma_2\sigma_3 \in \operatorname{Aut}(X)$ has no fixed points on X, the surface $Y = X/\sigma$ is an Enriques surface. As usual denote the projection by $\pi: X \to Y$. Under σ all the curves $E_i^{\pm}$, $F_i^{\pm}$, $L_i^{\pm}$, $N_i^{\pm}$ differing by a $\pm$sign are identified. So on Y we have the following curves:



Case 1 (in case 2, L_1 and L_2 touch on F_Y)



Case 3

First we determine divisors representing a $\mathbb{Z}$-basis of $H^2(Y,\mathbb{Z})_f = M$. Consider the cycle

$$Z = 2F_2 + 4F_3 + 6F_4 + 5F_5 + 4F_6 + 3F_7 + 2F_8 + 3N.$$

It is the fundamental cycle of an E_8-configuration, hence $Z^2 = -2$. We complete Z to $\tilde{E}_8$-configurations

$$\begin{aligned} Z_1 &= L_1 + Z, \quad Z_2 = L_2 + Z \qquad \text{(case 1 and 2)}, \\ Z_1 &= E_1 + Z \qquad\qquad\qquad\quad \text{(case 3)}. \end{aligned}$$

Then there are classes $h_1, h_2 \in M$ with $h_1^2 = h_2^2 = 0$ and

$$\begin{aligned} z_1 &= 2h_1, \quad z_2 = 2h_2 \qquad \text{(case 1 and 2)}, \\ z_1 &= 2h_1, \quad h_1 \cdot e_2 = 1 \qquad \text{(case 3)}. \end{aligned}$$

(Here as usual we denote the class in M represented by a cycle with the corresponding small letter.) Putting $h_2 = h_1 + e_2$ in case 3, we have in all cases

$$\begin{aligned} &h_1^2 = h_2^2 = 0, \quad h_1 \cdot h_2 = 1, \\ &h_i \cdot f_j = h_i \cdot n = 0, \qquad i = 1, 2,\ j = 2, \ldots, 8. \end{aligned}$$

This proves: $h_1, h_2, f_2, \ldots, f_8, n$ form a $\mathbb{Z}$-basis of M.

Recall that $\mathrm{Aut}(Y) = \mathrm{Aut}(X, \sigma)/\sigma$. Putting

$$\mathfrak{R}(\sigma) = \mathfrak{R} \cap \mathrm{Aut}(X, \sigma) = \{\alpha \in \mathfrak{R} : \alpha\sigma = \sigma\alpha\}$$

we have $\mathfrak{R}(\sigma)/\sigma$ as subgroup of $\mathrm{Aut}(Y)$. To describe this group more explicitely recall that $\sigma = \sigma_1 \cdot (\sigma_2 \sigma_3)$ and that $\sigma_2 \sigma_3$ acts on $\mathfrak{S}$ as translation by the unique 2-torsion element $\mathfrak{n}$. So $\sigma_2 \sigma_3$ commutes with all elements of $\mathfrak{R}$ and $\mathfrak{R}(\sigma)$ is the centralizer of σ_1. We observed that σ_1 centralizes $\mathfrak{S}_{\mathrm{tors}}$. If σ_1 commutes with translation by a non-torsion element $\mathfrak{s} \in \mathfrak{S}$, then necessarily $\sigma_1(\mathfrak{s}) = \mathfrak{s}$. So putting

$$\mathfrak{S}(\sigma) = \{\mathfrak{s} \in \mathfrak{S} : \sigma_1(\mathfrak{s}) = \mathfrak{s}\}$$

we have $\mathfrak{R}(\sigma) = (\mathbb{Z}_2 \times \mathbb{Z}_2) \ltimes \mathfrak{S}(\sigma)$, resp. in case 2 $\mathfrak{R}(\sigma) = (\mathbb{Z}_4 \times \mathbb{Z}_2) \ltimes \mathfrak{S}(\sigma)$. We put

$$s = \mathrm{rank}\, \mathfrak{S}(\sigma),$$

then $\mathfrak{S}(\sigma) = \mathfrak{S}_{\mathrm{tors}} \times \mathbb{Z}^s$. Notice that we do not yet know s, but in the cases 1 and 2 we have $\mathfrak{l}_i^{\pm} \in \mathfrak{S}(\sigma)$, hence $s \geqq 1$. So we obtain the following description of $\mathfrak{R}(\sigma)/\sigma \subset \mathrm{Aut}(Y)$.

Case 1: $\mathbb{Z}_2(\sigma_1) \times \mathbb{Z}_2(\sigma_2 \sigma_3) \times (\mathbb{Z}_2(\sigma_3) \ltimes \mathbb{Z}^s)/\sigma_1 \sigma_2 \sigma_3 = \mathbb{Z}_2 \times (\mathbb{Z}_2 \ltimes \mathbb{Z}^s)$.

Case 2: $\mathbb{Z}_2(\sigma_2 \sigma_3) \times ((\mathbb{Z}_4(\tilde{\rho}) \times \mathbb{Z}_2(\sigma_3)) \ltimes \mathbb{Z}^s)/\sigma_1 \sigma_2 \sigma_3 = (\mathbb{Z}_4 \times \mathbb{Z}_2) \ltimes \mathbb{Z}^s$.

Case 3: $\mathbb{Z}_2(\sigma_1) \times (\mathbb{Z}_2(\sigma_3) \ltimes (\mathbb{Z}_4(\mathfrak{e}^{\pm}) \times \mathbb{Z}^s))/\sigma_1 \sigma_2 \sigma_3 = \mathbb{Z}_2 \ltimes (\mathbb{Z}_4 \times \mathbb{Z}^s)$.

The aim of this section is to prove that there are no other automorphisms of Y. For $\alpha \in \mathrm{Aut}(X, \sigma)$ let us denote by $\alpha \bmod \sigma$ the induced automorphism of Y. The key observation is the following one.

(4.8) **Proposition.** a) *The involution* $\sigma_1 \bmod \sigma$ *acts trivially on* $H^2(Y, \mathbb{Z})$ *and generates the kernel of the representation of* $\mathrm{Aut}(Y)$ *on* $H^2(Y, \mathbb{Z})$.

b) *In case* 2 *the automorphism* $\tilde{\rho} \bmod \sigma$ *acts trivially on* $H^2(Y, \mathbb{Z})_f$ *and generates the kernel of the representation of* $\mathrm{Aut}(Y)$ *on* $H^2(Y, \mathbb{Z})_f$.

Proof. σ_1 (as well as $\tilde{\rho}$ in case 2) leaves invariant all the curves on X specified above, except for interchanging E_2^+ and E_2^-. This proves that $\sigma_1 \bmod \sigma$ (as well as $\tilde{\rho} \bmod \sigma$) acts trivially on the basis of $H^2(Y, \mathbb{Z})_f$ considered above.

To prove that $\sigma_1 \bmod \sigma$ acts trivially on $H^2(Y, \mathbb{Z})$ already (and $\tilde{\rho} \bmod \sigma$ does not do it) we observe that not only the classes $h_1, h_2, f_2, \ldots, f_8, n$, but also the curves $Z_1, Z_2, F_2, \ldots, F_8, N$ are left invariant under $\sigma_1 \bmod \sigma$, resp. $\tilde{\rho} \bmod \sigma$. So it suffices to consider the action on the two half-pencils in the linear systems $|Z_1|$ and $|Z_2|$.

Let us denote by $p_1', p_2' \in E_Y'$ the points where F_8, F_Y meet the smooth elliptic curve E_Y'. Then $|2p_1'| = |2p_2'|$ is the linear system cut out on E_Y' by both $|Z_1|$ and $|Z_2|$. The two half-pencils in $|Z_1|$ and $|Z_2|$ intersect E_Y' in the two other points p_3', p_4' with $\mathcal{O}_{E_Y'}(p_i') = \mathcal{O}_{E_Y'}(p_j')$, $i = 1, 2$, $j = 3, 4$. Now $\sigma_1 \bmod \sigma$ fixes the points p_3', p_4' and the corresponding half-pencils, whereas $\tilde{\rho}$ interchanges p_3' and p_4'.

Conversely, consider an arbitrary $\alpha \in \mathrm{Aut}(X, \sigma)$ with $\alpha \bmod \sigma$ acting trivially on $H^2(Y, \mathbb{Z})_f$. After replacing α by $\alpha\sigma$ we may assume $\alpha N^- = N^-$. By Proposi-

tion (4.6) this α is one of the following

$$\mathrm{id}_X, \tilde{\rho}, \sigma_1, \tilde{\rho}^3, \sigma_3, \tilde{\rho}\sigma_3, \sigma_1\sigma_3, \tilde{\rho}^3\sigma_3.$$

But the last four automorphisms in this list reverse the orientation in the cycle $\sum_1^8 (F_i^+ + F_i^-)$. This proves the assertion. □

(4.9 a) **Corollary.** $\sigma_1 \bmod \sigma$ *belongs to the center of* $\mathrm{Aut}(Y)$ *and its fixed point set is stable under each automorphism of* Y.

Now $\sigma_1 \bmod \sigma$ has the following set of fixed points

Case 1 and 2: $F_Y, F_2, F_4, F_6, F_8, \{p_3, p_4, p_3', p_4'\}\ p_i \in E_Y,\ p_i' \in E_Y'$,

Case 3: $F_Y, F_2, F_4, F_6, F_8, E_2, \{p_3', p_4'\}$.

It follows that any automorphism of Y is of the form $\alpha \bmod \sigma$ with $\alpha \in \mathfrak{N}(\sigma)$.

(4.9 b) **Corollary.** $\mathrm{Aut}(Y) = \mathfrak{N}(\sigma)/\sigma$.

It remains to determine the rank s of the abelian subgroup $\mathbb{Z}^s \subset \mathfrak{S}(\sigma)$.

(4.10) **Lemma.** a) *The subgroup of* $O(M)$ *leaving invariant* $f_1, \ldots, f_8$ *is* $D_\infty = \mathbb{Z}_2 \ltimes \mathbb{Z}$, *the infinite dihedral group.*

b) *The subgroup leaving in addition* e_2 *invariant is trivial.*

Proof. Any $\alpha \in O(M)$ leaving invariant $f_2, \ldots, f_8$ is determined by its action on $\{f_2, \ldots, f_8\}^\perp$. Now h_1, h_2 and $f = \sum_1^8 f_i$ belong to this orthogonal complement. Since

$$h_1^2 = h_2^2 = f^2 = 0, \quad h_1 \cdot h_2 = 1, \quad h_1 \cdot f = h_2 \cdot f = 2,$$

their 3×3 intersection matrix has determinant -8, hence equals $\det(f_i \cdot f_j)_{2 \leqq i, j \leqq 8}$. This shows that h_1, h_2, f are a $\mathbb{Z}$-basis of $\{f_2, \ldots, f_8\}^\perp$.

Now α acts as

$$h_1 \mapsto r_1 h_1 + r_2 h_2 + rf,$$
$$h_2 \mapsto s_1 h_1 + s_2 h_2 + sf,$$
$$f \mapsto f,$$

with $r, s, r_i, s_i \in \mathbb{Z}$ and determinant

$$r_1 s_2 - r_2 s_1 = \pm 1.$$

Denote by $\alpha_0 \in O(M)$ the permutation $h_1 \leftrightarrow h_2$. It generates a subgroup $\mathbb{Z}_2 \subset O(M)$ and modulo this subgroup we may assume the determinant above to be $+1$. Now orthogonality of α implies

$$2 = h_1 \cdot f = 2(r_1 + r_2) \Rightarrow r_2 = 1 - r_1,$$
$$2 = h_2 \cdot f = 2(s_1 + s_2) \Rightarrow s_2 = 1 - s_1$$

and from the determinant condition we conclude $r_1 = s_1 + 1$. Putting $s_1 = t \in \mathbb{Z}$ we find

$$h_1 \mapsto (t+1)h_1 - th_2 + rf,$$
$$h_2 \mapsto th_1 + (1-t)h_2 + sf$$

and orthogonality of α is equivalent with

$$0 = h_1^2 = -2t(t+1) + 4r \Rightarrow r = t(t+1)/2,$$
$$0 = h_2^2 = 2t(1-t) + 4s \Rightarrow s = t(t-1)/2.$$

So $\alpha = \alpha_t$ is an element of the group

$$\mathbb{Z} \ni t \mapsto \alpha_t = \begin{pmatrix} 1+t & -t & t(t+1)/2 \\ t & 1-t & t(t-1)/2 \\ 0 & 0 & 1 \end{pmatrix}.$$

On this group α_0 acts by $\alpha_t \to \alpha_{-t}$. So α_0 and $\mathbb{Z}$ generate D_∞.

b) We have $2E_2 = Z_2 - Z_1$ and hence $e_2 = h_2 - h_1$. So $\alpha_t(e_2) = e_2 - 2t(e_1 + e_2)$. So $\alpha_t(e_2) = e_2$ implies $t = 0$. Since $\alpha_0(e_2) = -e_2$, the assertion follows. □

(4.11) **Corollary.** *In case* 1 *or* 2 *we have* $s = 1$, *and* $s = 0$ *in case* 3.

Proof. If $\mathfrak{s} \in \mathbb{Z}^8 \subset \mathfrak{S}(\sigma)$ then addition by $8\mathfrak{s}$ leaves invariant $f_1, \ldots, f_8$ and in case 3 also e_2. So $s \leqq 1$, resp. $s = 0$ in case 3, by the lemma above. But $s \geqq 1$ in case 1 or 2 was observed already. □

The final result is the following.

(4.12) **Theorem.** *Let* Y *be an Enriques surface as considered above. Then* $\mathrm{Aut}(Y)$ *is isomorphic with*

$$\mathbb{Z}_2 \times D_\infty \quad \text{(case 1)},$$
$$\mathbb{Z}_4 \times D_\infty \quad \text{(case 2)},$$
$$D_4 \quad \text{(case 3)}.$$

4.5. **Invariants.** In this section we compute the Picard number $\rho(X)$ and the nodal type N for the general surfaces in our family. As above we use the decomposition $L = \mathbb{H} \perp \mathbb{H} \perp \mathbb{H} \perp \mathbb{E} \perp \mathbb{E}$. In 4.4 we observed that the nodal classes $f_2, \ldots, f_8, n \in M = \mathbb{H} \perp \mathbb{E}$ form a $\mathbb{Z}$-basis for $\mathbb{E}$. Their inverse images on X decompose into 16 nodal classes forming two E_8-configurations. In fact these two configurations are bases for the two factors $\mathbb{E} \subset L$, see [7, Sect. 5]. Since the algebraic lattice S_X contains $L^+ \subset L$, this proves the following.

(4.13) **Lemma.** *For all surfaces* X *considered here the algebraic lattice* $S_X \subset L$ *contains the* rank-18 *sublattice*

$$\mathbb{H}(2) \perp \mathbb{E} \perp \mathbb{E} = \{(0, h, h, x, y) : h \in \mathbb{H},\ x, y \in \mathbb{E}\}.$$

The orthogonal complement in L of this rank-18 lattice is

$$N_1 = \mathbb{H} \perp \mathbb{H}(2) = \{(h, h', -h', 0, 0) : h, h' \in \mathbb{H}\}.$$

In case 3 also the class $e_2=h_2-h_1$ is nodal. Then $\pi^* e_2=d+s(d)$ with $d\in\mathbb{H}\perp\mathbb{H}\perp\mathbb{H}\subset L$ a nodal class such that $\langle d, s(d)\rangle=0$ and

$$\begin{aligned} d+s(d)&=(0, h_2-h_1, h_2-h_1),\\ d-s(d)&=(h, h', -h')\in N_1,\\ 2d&=(h, h'+h_2-h_1, -h'+h_2-h_1). \end{aligned}$$

This shows $h\in 2\mathbb{H}$. By Corollary (1.2) and Lemma (1.8) it follows that there is an automorphism of N_1 extending to $\mathbb{H}\perp\mathbb{H}\perp\mathbb{H}$ as identity on $N_1^\perp$ mapping $d-s(d)$ to $(0, h_2-h_1, h_1-h_2)$. If we change the fixed marking of Y by this automorphism we have

$$d+s(d)=(0, h_2-h_1, h_2-h_1)\quad d-s(d)=(0, h_2-h_1, h_1-h_2)$$

and therefore

$$d=(0, h_2-h_1, 0),\qquad s(d)=(0, 0, h_2-h_1).$$

Let $N_3\subset N_1$ be the rank-3 sublattice orthogonal to d. Then we have shown the following.

(4.14) **Lemma.** *In case 3 for all surfaces X the algebraic lattice S_X contains the rank-19 lattice $N_3^\perp$.*

In other words: If $N\subset L$ is the nodal type of X, then always $N\subset N_1$ and in case 3 even $N\subset N_3$.

(4.5) **Proposition.** *For general X in our family we have $T_X=N=N_1$. For general X under case 3 we have $T_X=N=N_3$. In particular $\rho(X)=18$ for general X, and $\rho(X)=19$ for general X in case 3.*

Proof. The cup product on N_1 is non-degenerate. So there is no 3-dimensional isotropic linear subspace of $N_1\otimes\mathbb{C}$. If $T_X\neq N_1$ for all X, then the period point $\tau(Y,\varphi)$ would vary in a countable union of 1-dimensional quadrics. This is impossible, because we show that in D^0/Γ our family has an image containing a (local analytic) variety of dimension two. Similarly, in case 3, we show that the image contains a curve.

Case 1. Consider the map $\Phi_Y: Y\to\mathbb{P}_1$ given by the pencil $|2F_Y|$. We have $\Phi_Y\circ\pi=\Phi$, and $\Phi|E$ is the quotient map w.r. to $\mathbb{Z}_2(\sigma_1|E)\times\mathbb{Z}_2(\sigma_3|E)$. So $\Phi|E$ factors as $E\xrightarrow{q}\mathbb{P}_1(u_0: u_1)\to\mathbb{P}_1(w_0: w_1)$, $w_i=u_i^2$, and similarly for $\Phi|E'$. The four isolated fixed points $p_3, p_4\in E_Y$, $p_3', p_4'\in E_Y'$ and their images under Φ_Y (modulo the natural $\mathbb{C}^*$-action on $\mathbb{P}_1$) depend only on the isomorphism class of Y. These four image points are the four roots of the polynomial

$$\begin{aligned} &(aw_0^2+2bw_0w_1+cw_1^2)(aw_0^2+2bw_0w_1+dw_1^2)\\ &\quad=a^2\left(w_0^4+4\frac{b}{a}w_0^3w_1+\left(4\frac{b^2}{a^2}+\frac{c+d}{a}\right)w_0^2w_1^2+2\frac{b}{a}\left(\frac{c+d}{a}\right)s_0w_1^3+\frac{cd}{a^2}w_1^4\right). \end{aligned}$$

Since $\frac{b}{a}$, $\frac{c}{a}$, and $\frac{d}{a}$ vary independently, we see that these 4-tuples in $\mathbb{P}_1^4$ form a 3-dimensional set and dividing by the $\mathbb{C}^*$-action the dimension is two.

Case 3. It suffices to show that the elliptic curve F varies. It is ramified over the four points ± 1, $\pm\sqrt{\frac{c}{d}}$, or after multiplying with $\sqrt{d}$ over $\pm\sqrt{c}$, $\pm\sqrt{d}$. The cross-ratio

$$\frac{\sqrt{d}-\sqrt{c}}{\sqrt{d}+\sqrt{c}} : \frac{-\sqrt{d}-\sqrt{c}}{-\sqrt{d}+\sqrt{c}} = \frac{(\sqrt{d}-\sqrt{c})^2}{(\sqrt{d}+\sqrt{c})^2}$$

varies with c/d. □

The lattices $L_1, \ldots, L_4$ for general X then are as follows

$$\begin{aligned} &L_1 = L^+, \\ &L_2 = N_1^\perp \cap LL^-, \quad L_3 = 0, \quad L_4 = N_1 \qquad \text{in case 1,} \\ &L_2 = N_3^\perp \cap L^-, \quad L_3 = 0, \quad L_4 = N_3 \qquad \text{in case 3.} \end{aligned}$$

4.6 **The action of Aut(X, σ) on N.** Let us finish by giving a few properties of the representation of $\mathrm{Aut}(X, \sigma)$ on N_1 (cases 1 and 2) and N_3 (case 3). In the proof of (4.15) we showed that the period point $\tau(Y, \varphi)$ moves in an open set of $D^0 \cap \mathbb{P}(N_1)$ (resp. $D^0 \cap \mathbb{P}(N_3)$) if Y moves in our family (resp. in the part of the family under case 3). So the argument of (2.9) applies and shows that the image of

$$\begin{aligned} \mathbb{Z}_2(\sigma_1) \times \mathbb{Z}_2(\sigma_2\sigma_3) \times (\mathbb{Z}_2(\sigma_3) \ltimes \mathbb{Z}) &\to \mathrm{Aut}(N_1) \\ \mathbb{Z}_2(\sigma_1) \times (\mathbb{Z}_2(\sigma_3) \ltimes \mathbb{Z}_4(e^{\pm})) &\to \mathrm{Aut}(N_3) \end{aligned}$$

in general is the group $\{\pm \mathrm{id}\}$. But this assertion is invariant under deformations, so it holds for *all our surfaces Y.*

Now for $\tilde{\rho}$ the situation is quite different.

(4.16) **Proposition.** *We have* $(\tilde{\rho}|N_1)^2 = -\mathrm{id}$, *in particular the order of* $\tilde{\rho}$ *on* N *is four.*

Proof. It suffices to show that $\sigma_1 = \tilde{\rho}^2$ acts on N_1 as $-\mathrm{id}$. To do this, we use the Lefschetz fixed point formula. It reads

$$\mathrm{Trace}(\sigma_1 \text{ on } \oplus H^{2q}(X, \mathbb{R})) = e(\mathrm{Fix}(\sigma_1)),$$

where on the right-hand side we add over the Euler numbers of all components of the fixed point set of σ_1. Now σ_1 fixes point-wise the eight rational curves $F_2^\pm$, $F_4^\pm$, $F_6^\pm$, and $F_8^\pm$ as well as the smooth elliptic curve F. So $e(\mathrm{Fix}(\sigma_1)) = 16$ and $\mathrm{Trace}(\sigma_1|L_{\mathbb{R}}) = 14$. Since σ_1 leaves invariant all the curves $F_i^\pm$, $L_i^\pm$, $N^\pm$, from which a basis for the lattice $N_1^\perp$ can be chosen (cf. (4.13)), we have $(\sigma_1|N_1^\perp) = \mathrm{id}$ and $\mathrm{Trace}(\sigma_1|(N_1^\perp) \otimes \mathbb{R}) = 18$.

This implies

$$\mathrm{Trace}(\sigma_1|N_1 \otimes \mathbb{R}) = 14 - 18 = -4,$$

hence σ_1 acts on N_1 as $-\mathrm{id}$. □

References

1. Bourbaki, N.: Groupes et algèbres de Lie, Chap. 4, 5 et 6. Paris: Hermann 1968
2. Burns, D., Rapaport, M.: On the Torelli theorem for kählerian K3-surfaces. Ann. Sci. E.N.S. **4**^{e} (ser. 8) 235–274 (1975)
3. Cossec, F.: Projective models of Enriques surfaces. To appear in Math. Ann.
4. Cossec, F.: Reye congruences. To appear in Trans. A.M.S.
5. Dieudonné, J.: La géométrie des groupes classiques. Erg. d. M. N.F. 5. Berlin-Heidelberg-New York: Springer 1955
6. Dolgachev, I.: On automorphisms of Enriques surfaces. To appear Invent. Math. in press (1983)
7. Horikawa, E.: On the periods of Enriques surfaces I. Math. Ann. **234**, 73–108 (1978)
8. Horikawa, E.: On the periods of Enriques surfaces II. Math. Ann. **235**, 217–246 (1978)
9. Nikulin, V.V.: Finite automorphism groups of kähler K3-surfaces. Trans. Moscow Math. Soc. **38**, 71–135 (1980)
10. Nikulin, V.V.: On the quotient groups of the automorphism groups groups of hyperbolic forms by the subgroups generated by 2-reflections. Algebraic-geometrical applications. Modern Problems in Math., t. 18, 1–114, Moscow VINITI (1981) (To be translated by J. Sov. Math.)
11. Piatecky-Shapiro, I., Shafarevich, I.: A Torelli theorem for algebraic surfaces of type K3. Izv. A. N. SSSR, Math. **35**, 530–572 (1971) Transl. to English: Math. USSR-Izv. **5**, 547–588 (1971)
12. Remmert, R., Stein, K.: Über die wesentlichen Singularitäten analytischer Mengen. Math. Ann. **126**, 263–306 (1953)
13. Ueno, K.: Classification theory of algebraic varieties and and compact complex spaces. Lecture Notes in Mathematics, vol. 439. Berlin-Heidelberg-New York: Springer 1975

Oblatum 31-III-1983

Invent. math. 73, 413–418 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

On a New Characterization of Demuskin Groups

D. Dummit[1]* and J. Labute[2]**

[1] Department of Mathematics, Harvard University, Cambridge, MA 02138, USA
[2] Department of Mathematics, McGill University, Montreal, Quebec H3A 2K6, Canada

1. Statement of Results

Let G be a pro-p-group (p a prime). The cohomology groups $H^i(G) = H^i(G, \mathbf{Z}/p\mathbf{Z})$ are vector spaces over the field with p elements. The group G is a Demuskin group if the following three properties are satisfied: (i) $\dim H^1(G) < \infty$, (ii) $\dim H^2(G) = 1$, (iii) the cup product $H^1(G) \times H^1(G) \to H^2(G)$ is a non-degenerate bilinear form. Property (i) expresses the fact that G is a finitely generated pro-p-group with $n(G) = \dim H^1(G)$ as the cardinality of a minimal (topological) generating set (cf. [7, Chap. I, §4.2]). Property (ii) says that G is a one-relator pro-p-group (cf. [7, Chap. I, §4.3]). A canonical form for the relator defining a Demuskin group has been determined (cf. [4]). Examples of Demuskin groups include $\mathbf{Z}/2\mathbf{Z}$, the pro-p-completions of the fundamental groups of compact Riemann surfaces, and the Galois groups of the maximal p-extensions of local fields which contain the p-th roots of unity.

If G is a Demuskin group $\neq \mathbf{Z}/2\mathbf{Z}$ then G has cohomological dimension 2. Moreover, in this case every open subgroup of G is also a Demuskin group (cf. [8]) and, in particular, is a one-relator group. Using properties of the Euler-Poincaré characteristic of G, we also see that $n(U) - 2 = [G:U](n(G) - 2)$ for every open subgroup U of G. The purpose of this paper is to show that each of the last two properties characterizes Demuskin groups $\neq \mathbf{Z}/2\mathbf{Z}$. In fact we prove the following stronger result:

Theorem 1. *Let G be a finitely generated one-relator pro-p-group of rank $n(G) \neq 1$. Then the following statements are equivalent:*

Current addresses

[1] Department of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
[2] Department of Mathematics, Harvard University, Cambridge, MA 02138, USA

* Supported by a National Science Foundation Postdoctoral Research Fellowship
** Supported in part by a grant from the National Sciences and Engineering Research Council of Canada

(a) *The group G is a Demuskin group;*

(b) *Every open subgroup of G is a one-relator group;*

(b′) *Every open subgroup of G of index p is a one-relator group;*

(c) *If U is any open subgroup of G then $n(U)-2=[G:U](n(G)-2)$;*

(c′) *If U is any open subgroup of G of index p then $n(U)-2=[G:H](n(G)-2)$.*

This theorem is a immediate consequence of the following result:

Theorem 2. *Let G be a finitely generated one-relator pro-p-group of rank $n\neq 1$ which is not a Demuskin group. Then there exists an open subgroup U of G of index p such that $\dim H^2(U)>1$ and $n(U)-2>p(n-2)$.*

Corollary. *Let k be a field containing the p-th roots of unity. Suppose that $k^*/(k^*)^p$ has finite rank $\neq 1$ and that for every cyclic extension K of k of degree dividing p the subgroup of elements killed by p in the Brauer group of K is cyclic of order p. Then the Galois group of the maximal p-extension of k is either a Demuskin group or a free pro-p-group.*

The corollary follows from the fact that, for a field K containing the p-th roots of unity, the kernel of multiplication by p in the Brauer group of K contains $H^2(U)$, where U is the Galois group of the maximal p-extension of K.

It is well known (cf. [3, § 5.2] for example) that a finitely generated pro-p-group G is a free pro-p-group if and only if $n(U)-1=[G:U](n(G)-1)$ for every open subgroup U of G. Iwasawa observed that pro-p-groups satisfying $n(U)-c=[G:U](n(G)-c)$ for a constant positive integer c (and all open subgroups U of G) have interesting representation-theoretic properties (see [1, § 7] for one application in the case $c=1$) and raised the question of determining those groups with $c\neq 1$. Theorem 1 answers this question for 1-relator groups (in a strengthened form) in the case $c=2$. The referee has shown, using partial Euler-Poincaré characteristics (cf. [3, § 5.2]) that the cohomological dimension is less than or equal to 2 for a finitely generated pro-p-group G satisfying $\dim H^2(G)=c-1$ and $n(U)-c=[G:U](n(G)-c)$ for every open subgroup U of G. This yields an alternate proof of the fact that (c) implies (b) in Theorem 1. That (b) implies (a) can be seen directly by using results on the dualizing module of G as in [7, Chap. II, § 5.2].

The authors would like to thank the referee for his helpful comments in revising the paper.

2. The (x, τ, p)-Filtration of a Free Pro-p-group

In this section we outline the facts we need about filtrations. For more details, and proofs, the reader is referred to Lazard [6].

Let F be the free pro-p-group generated by $x_1, x_2, \ldots, x_n$. The completed group algebra $\Lambda=\mathbf{Z}_p[[F]]$ is isomorphic to the algebra of non-commutative formal power series in the indeterminates $X_1, X_2, \ldots, X_n$ with coefficients in the ring $\mathbf{Z}_p$ of p-adic integers, and the isomorphism can be chosen so that it sends x_i to $1+X_i$. We identify Λ with this algebra of formal power series.

Let $\tau_1, \tau_2, \ldots, \tau_n$ be n real numbers >0. We define a valuation w on Λ by setting

$$w(\sum_{i_1,\ldots,i_k} a_{i_1\ldots i_k} X_{i_1}\ldots X_{i_k}) = \operatorname*{Inf}_{i_1,\ldots,i_k}(b_{i_1\ldots i_k}),$$

where $b_{i_1\ldots i_k} = v(a_{i_1\ldots i_k}) + \tau_{i_1} + \ldots + \tau_{i_k}$ and v is the p-adic valuation of $\mathbf{Z}_p$ ($v(p) = 1$). For each real number $\alpha \geqq 0$, set

$$\Lambda_\alpha = \{u \in \Lambda : w(u) \geqq \alpha\}, \quad \Lambda_{\alpha+} = \{u \in \Lambda : w(u) > \alpha\}, \quad \operatorname{gr}_\alpha(\Lambda) = \Lambda_\alpha / \Lambda_{\alpha+}$$

and set $A = \operatorname{gr}(\Lambda) = \oplus \operatorname{gr}_\alpha(\Lambda)$ ($\alpha \geqq 0$). The valuation w induces the usual p-adic valuation on $\mathbf{Z}_p \subseteq \Lambda$. If π is the image of p in $\operatorname{gr}_1(\mathbf{Z}_p) = p\mathbf{Z}/p^2\mathbf{Z}$, then π is an indeterminate over $\mathbf{F}_p = \operatorname{gr}_0(\mathbf{Z}_p) = \mathbf{Z}/p\mathbf{Z}$ and $\operatorname{gr}(\mathbf{Z}_p) = \mathbf{F}_p[\pi]$. Hence $\operatorname{gr}(\Lambda)$ is naturally a graded $\mathbf{F}_p[\pi]$-algebra. If ξ_i denotes the image of X_i in $\operatorname{gr}_\alpha(\Lambda)$, where $\alpha = \tau_i$, then A is the graded algebra of non-commutative polynomials in the indeterminates $\xi_1, \ldots, \xi_n$ with coefficients in $\mathbf{F}_p[\pi]$, the grading being defined by $\deg(\pi) = 1$, $\deg(\xi_i) = \tau_i$. If L is the Lie subalgebra of A generated by the ξ_i, then, by the Birkhoff-Witt theorem, L is the free Lie algebra over $\mathbf{F}_p[\pi]$ on $\xi_1, \ldots, \xi_n$.

For $x \in F$, set $\omega(x) = w(x-1)$. The mapping $\omega: F \to \mathbf{R}_{>0} \cup \{+\infty\}$ satisfies (i) $\omega(xy^{-1}) \geqq \min(\omega(x), \omega(y))$, (ii) $\omega(x^{-1}y^{-1}xy) \geqq \omega(x) + \omega(y)$, and (iii) $\omega(x^p) \geqq \min(p\omega(x), \omega(x)+1)$. It defines the (x, τ, p)-filtration of F. For any $\alpha > 0$, set $F_\alpha = \{x \in F : \omega(x) \geqq \alpha\}$, $F_{\alpha+} = \{x \in F : \omega(x) > \alpha\}$. Then F_α and $F_{\alpha+}$ are closed normal subgroups of F and the quotient $\operatorname{gr}_\alpha(F) = F_\alpha / F_{\alpha+}$ is an elementary abelian p-group which we denote additively. Let $i_\alpha: F_\alpha \to \operatorname{gr}_\alpha(F)$ be the canonical surjection. If $x \in F_\alpha$, $x \notin F_{\alpha+}$, then $i_\alpha(x)$ is called the *initial form* of x.

If $\xi = i_\alpha(x)$ and $\eta = i_\beta(y)$, then the commutator $[x, y] = x^{-1}y^{-1}xy$ is an element of $F_{\alpha+\beta}$ and the element $i_{\alpha+\beta}([x, y])$ is independent of the choice of x and y. This element is denoted by $[\xi, \eta]$. This bracket is bilinear and satisfies the Jacobi identity. In particular, it defines a Lie algebra structure on $\operatorname{gr}(F) = \oplus \operatorname{gr}_\alpha(F)$ ($\alpha > 0$). Moreover, the mapping $x \mapsto x - 1$ of F into Λ induces an injective Lie algebra homomorphism $\Psi: \operatorname{gr}(F) \to A$, i.e., $\Psi([\xi, \eta]) = \Psi(\xi)\Psi(\eta) - \Psi(\eta)\Psi(\xi)$. We identify $\operatorname{gr}(F)$ with its image in A. To identify $\operatorname{gr}(F)$ as a (homogeneous) Lie subalgebra of A over $\mathbf{F}_p$ we have to introduce the operator $P: A_+ \to A_+$ (where $A_+ = \oplus A_\alpha$ ($\alpha > 0$), $A_\alpha =$ the homogeneous component of A of degree α). For a homogeneous element $\xi \in A_\alpha$, define

$$\begin{aligned} P\xi &= \xi^p && \text{if } \alpha < (p-1)^{-1}, \\ P\xi &= \xi^p + \pi\xi && \text{if } \alpha = (p-1)^{-1}, \\ P\xi &= \pi\xi && \text{if } \alpha > (p-1)^{-1}, \end{aligned}$$

and $P(\sum \xi_\alpha) = \sum P\xi_\alpha$ ($\xi_\alpha \in A_\alpha$). Note that $P: A_\alpha \to A_{\phi(\alpha)}$, where $\phi(\alpha) = \min(p\alpha, \alpha + 1)$.

If $x \in F_\alpha$, then $x^p \in F_{\phi(\alpha)}$ and the image of x^p in $\operatorname{gr}_{\phi(\alpha)}(F)$ is $P(i_\alpha(x))$. Hence $\operatorname{gr}(F)$ is stable under P and since $\operatorname{gr}(F) = \oplus P^i L_0$ ($i > 0$), where L_0 is the smallest (homogeneous) F_p Lie-subalgebra of A_+ containing $\xi_1, \ldots, \xi_n$, we see that $\operatorname{gr}(F)$ is the smallest $\mathbf{F}_p$ Lie-subalgebra of A_+ containing $\xi_1, \ldots, \xi_n$ and which is stable under P. The operator P has the following properties:

(1) For $\alpha \leqq (p-1)^{-1}$ and $\xi, \eta \in A_\alpha$, $P(\xi+\eta)=P(\xi)+P(\eta)+J(\xi,\eta)$, where $J(\xi,\eta)$ is the Lie polynomial $(\xi+\eta)^p-\xi^p-\eta^p$ in ξ,η with coefficients in $\mathbf{F}_p$ (cf. [2, V, § 5.2]);

(2) For $\alpha>(p-1)^{-1}$ and $\xi,\eta\in A_\alpha$, $P(\xi+\eta)=P(\xi)+P(\eta)$;

(3) For $\alpha<(p-1)^{-1}$ and $\xi\in A_\alpha$, $\eta\in A_\beta$, $[P\xi,\eta]=\mathrm{ad}(\xi)^p(\eta)$;

(4) For $\alpha=(p-1)^{-1}$ and $\xi\in A_\alpha$, $\eta\in A_\beta$, $[P\xi,\eta]=\mathrm{ad}(\xi)^p(\eta)+P[\xi,\eta]$;

(5) For $\alpha>(p-1)^{-1}$ and $\xi\in A_\alpha$, $\eta\in A_\beta$, $[P\xi,\eta]=P[\xi,\eta]$.

In particular, $\mathrm{gr}(F)$ is a mixed Lie algebra in the terminology of Lazard and is in fact free by the Birkhoff-Witt theorem for mixed Lie algebras. We shall not make use of this fact here.

2. Proof of Theorem 2

Let G be a finitely generated pro-p-group of rank $n>1$ with the property that every open subgroup of index dividing p is a one-relator group. In particular, G is a one-relator group and hence has the presentation $G=F/R$, where F is a free pro-p-group on generators $x_1,\ldots,x_n$ and R is the closed normal subgroup of F generated by some element $r\in F^p[F,F]$, $r\neq 1$. Here $[F,F]$ is the closed subgroup of F generated by the commutators $[x,y]=x^{-1}y^{-1}xy$ with $x,y\in F$. We have

$$G/[G,G]\cong(\mathbf{Z}_p/q\mathbf{Z}_p)\times(\mathbf{Z}_p)^{n-1},$$

where $q=p^f$, $1\leqq f\leqq\infty$ (by convention, $p^\infty=0$). We have $q=0$ if and only if $G/[G,G]$ is torsion free, and if $q\neq 0$ the integer q is the order of the cyclic torsion subgroup of $G/[G,G]$. If $q\neq 0$ there is an element $s\in F$, uniquely determined modulo $[F,F]$, such that $r\equiv s^q$ modulo $[F,F]$. In this case, since $s\notin F^p[F,F]$, we can assume that our basis $x_1,\ldots,x_n$ of F has been so chosen that $x_1=s$.

Lemma. *For a suitable choice of basis $x_1,\ldots,x_n$ of F, the relator r can be written in one of the following two forms:*

(1) $r\equiv x_1^q[x_1,x_2][x_3,x_4]\ldots[x_{m-1},x_m] \pmod T$,

(2) $r\equiv x_1^q[x_2,x_3][x_4,x_5]\ldots[x_{m-1},x_m] \pmod T$,

where $T=[F,F]^p[F,[F,F]]$ and $m\leqq n$.

Proof. We have $r=x_1^q u$, where

$$u\equiv\prod_{i<j}[x_i,x_j]^{a_{ij}} \pmod T,$$

with $a_{ij}\in F_p$. If there is a j with $a_{1j}\neq 0$, then changing $x_2,\ldots,x_n$ by a linear change of variables not involving x_1 we can write.

$$u\equiv[x_1,x_2]\prod_{3\leqq i<j\leqq n}[x_i,x_j]^{a_{ij}} \pmod T.$$

Repeating this procedure, we obtain (1). If $a_{ij}=0$ for all j then we obtain (2) by the same procedure.

Remark. Assume $q \neq 0$ and define the linear form b on $H^1(G)$ by $b(\phi) = \phi(s)$. Then, for $q \neq 2$, the forms (1) and (2) above are distinguished by $\mathrm{Ker}(b) \supseteq \mathrm{rad}\, H^1(G)$ and $\mathrm{Ker}(b) \not\supseteq \mathrm{rad}\, H^1(G)$ respectively, where $\mathrm{rad}\, H^1(G)$ is the radical of $H^1(G)$ with respect to the cup product pairing. If $q = 2$, the two forms are equivalent.

Now assume that G is not a Demuskin group and choose a presentation for G as in the lemma. The integer m there is the rank of the cup product pairing. If $m = n$ then r must have the form (2) with $q \neq 2$ and $n \geqq 3$. In this case, by renaming $x_1, \ldots, x_n$, we may take r in the form

$$(3)\quad r \equiv x_n^q [x_1, x_2][x_3, x_4] \ldots [x_{n-2}, x_{n-1}] \pmod{T}.$$

Choose $x_1, \ldots, x_n$ so that r is in the form (1) or (2) for $m < n$ and in the form (3) if $m = n$. Let N be the closed normal subgroup of F generated by the elements $x_1, x_2, \ldots, x_{n-1}, x_n^p$. The group N is a free pro-p-group with basis

$$x_n^p,\ \mathrm{ad}(x_n)^{j-1} x_i \qquad (1 \leqq i \leqq n-1,\ 1 \leqq j \leqq p),$$

where $\mathrm{ad}(x)y = [x, y]$. We have $F/N \cong \mathbf{Z}/p\mathbf{Z}$, generated by the image of x_n in F/N. Since $r \in N$ we have $R \subseteq N$ and R is generated as a (closed) normal subgroup of N by the elements $r_1, \ldots, r_p$, where $r_i = \mathrm{ad}(x_n)^{i-1} r$ (so, in particular, $\dim R/R^p[R, N] \leqq p$). Set $U = N/R$. Then U is the required open subgroup of G. To show this we distinguish two cases:

Case 1. *The relator is either in the form* (1) *or* (2) *with* $m < n$, *or in the form* (3) *with* $q \neq p$.

Case 2. *The relator* r *is in the form* (3) *with* $q = p \neq 2$ *and* $n \geqq 3$.

We deal with case 1 first. We first show that $\dim R/R^p[R, N] = p$. To do this we use the (x, τ, p)-filtration of F with $\tau_i = 1$ for $i \neq n$ and $\tau_n = 1/p$. Then $\mathrm{gr}(F) = \oplus\, \mathrm{gr}_\alpha(F)$ $(\alpha \in S)$, where S is the set of rational numbers > 0 with denominators dividing p. With respect to this filtration, the initial form ρ of r is not a polynomial in ξ_n in the free associative A. It follows easily that $\mathrm{ad}(\xi_n)^{i-1}\rho \neq 0$ for $i > 1$ by an elementary exercise in non-commutative polynomials. If ρ_i is the initial form of r_i, it follows that $\rho_i = \mathrm{ad}(\xi_n)^{i-1}\rho$ and hence that $\rho_1, \ldots, \rho_p$ are linearly independent over $\mathbf{F}_p$ since $\deg(\rho_i) = \deg(\rho) + (i-1)/p$. If we set $t = \deg(\rho)$ and $R_\alpha = R \cap F_\alpha$ then $R = R_t$ and $R^p[R, N] \subseteq R_{t+1}$ since $N \subseteq F_1$. Since $\dim R/R^p[R, N] \leqq p$ and

$$\dim R_{t+(i-1)/p}/R_{t+i/p} = 1$$

(generated by ρ_i), it follows that $R/R^p[R, N]$ is p-dimensional (generated by the images of $r_1, \ldots, r_p$).

The exact sequence $1 \to R \to N \to U \to 1$ gives rise to the exact sequence

$$0 \to H^1(U) \to H^1(N) \to H^1(R)^N \to H^2(U) \to 0,$$

where $H^1(R)^N = H^1(R/R^p[R, N])$. If $r \in N^p[N, N]$ then $H^1(U) \to H^1(N)$ is an isomorphism so that $H^2(U) \cong R/R^p[R, N]$ is of dimension p and

$$\dim H^1(U) = p(n-1) + 1 > p(n-2) + 2$$

which yields the required result. If $r \notin N^p[N,N]$, then the presentation above for U is not minimal and $t=\omega(r)=2-s/p$ with $1 \leqq s \leqq p-2$ since $1+2/p \leqq t \leqq 2-1/p$. In particular, $p \neq 2$. Since $\omega(r_i)=t+(i-1)/p$ and $N^p[N,N] \subseteq F_2$, it follows that the relators $r_1, \ldots, r_s$ are linearly independent mod $N^p[N,N]$ while the relators $r_{s+1}, \ldots, r_p$ lie in $N^p[N,N]$. It follows that $\dim H^1(N) - \dim H^1(U) = s$. Since

$$\dim H^2(U) = \dim H^1(R)^N - \dim H^1(N) + \dim H^1(U) = p - s > 1$$

and

$$\dim H^1(U) = \dim H^1(N) - s = p(n-1) + 1 - s > p(n-2) + 2,$$

the theorem follows in this case.

For case 2 we use the (x, τ, p)-filtration of F where $\tau_i = 1$ for all i. The initial form of r, with respect to this filtration, is

$$\pi \xi_n + [\xi_1, \xi_2] + \ldots + [\xi_{n-2}, \xi_{n-1}].$$

Let $M = R/R^p[R,R]$ and let $\Gamma = \mathbf{F}_p[[G]]$ be the completed algebra of G over $\mathbf{F}_p$. Then, since $\rho \notin \pi L$, the Γ-module M is free of rank 1 (cf. [5, Theorem 4']. If $\Delta = \mathbf{F}_p[[U]]$ then Γ is a free Δ-module of rank p and hence $M/IM \cong (\mathbf{F}_p)^p$, where I is the augmentation ideal of Δ. Since $\mathrm{Hom}(M/IM, \mathbf{F}_p) = H^1(R)^N$, it follows that $\dim H^1(R)^N = p$. Hence $R/R^p[R,N]$ is p-dimensional, generated by the images of $r_1, \ldots, r_p$. Let $u = x_n^{-p} r$. Then $\mathrm{ad}(x_n) r = \mathrm{ad}(x_n) u$. Let s be the smallest integer $\geqq 0$ such that $r_{s+1} \in N^p[N,N]$. Then $1 \leqq s \leqq p-2$ since $u \in N^p[N,N][F,[F,F]]$ and $r \notin N^p[N,N]$. Using the structure of $N/N^p[N,N]$ as an $F_p[F/N]$-module, we see that $r_1, \ldots, r_s$ are linearly independent mod $N^p[N,N]$ while the relators $r_{s+1}, \ldots, r_p \in N^p[N,N]$. It follows that $\dim H^1(N) - \dim H^1(U) = s$ and the theorem follows as in case 1.

References

1. Iwasawa, K.: Riemann-Huritz formula and p-Adic galois representations for number fields. Tohuko Math. J. (2nd series), **33**, 263–288 (1981)
2. Jacobson, N.: Lie algebras. New York: Interscience 1962
3. Koch, H.: Galoissche Theorie der p-Erweiterungen. Berlin-Heidelberg-New York: Springer 1962
4. Labute, J.P.: Classification of Demuskin Groups. Can. J. Math. **19**, 106–132 (1966)
5. Labute, J.P.: Algèbres de Lie et Pro-p-groupes définis par une seule relation. Invent. Math. **4**, 142–158 (1967)
6. Lazard, M.: Groupes Analytiques p-Adiques. Publications Mathematiques, No. 26, Inst. Hautes Etudes Scient., Paris, 1965
7. Serre, J-P.: Cohomologie Galoisienne. Lecture Notes in Mathematics, Vol. 5. Berlin-Heidelberg-New York: Springer 1965
8. Serre, J-P.: Structures de certains pro-p-groupes. Sem. Bourbaki, t. 15, No. 252, pp. 1–11 (1962–63)

Oblatum 2-IX-1982 & 7-II-1983

Invent. math. 73, 419-436 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

Classification C^∞ des èquations de pfaff complétement intégrables a singularité isolée

R. Moussu

Laboratoire de Topologie - ERA 945, Département de Mathématiques, Université de Dijon, B.P. 138, F-21004 Dijon Cedex, France

0. Introduction

Soit $I(n)$ l'ensemble des germes en $0\in\mathbb{R}^n$, $n\geqq 3$ de 1-formes, complètement *intégrables*, de classe C^∞, dont 0 est une *singularité algébriquement isolée*; c'est à dire:

$$\omega=\sum_{i=1}^{n} a_i dx_i, \qquad \omega\wedge d\omega=0,$$

les $a_i\in\mathscr{E}(n)$, l'anneau des germes en 0 de fonctions de classe C^∞ sur $\mathbb{R}^n$, et le quotient de $\mathscr{E}(n)$ par l'idéal qu'ils engendrent est un $\mathbb{R}$-espace vectoriel de dimension finie.

Cet article a essentiellement deux buts. Le premier est de montrer (théorème 1) que sous certaines conditions topologiques un élément $\omega\in I(n)$ possède une *intégrale première* $f\in\mathscr{E}(n)$, c'est à dire:

$$\omega=g\,df \quad \text{avec } g\in\mathscr{E}(n). \tag{0.1}$$

Le second est de caractériser algébriquement la relation de C^∞-équivalence sur $I(n)$ (théorème 2): $\omega,\omega'\in I(n)$ *sont* C^∞-*équivalentes* si

$$\omega\wedge\Phi^*(\omega')=0 \quad \text{avec } \Phi\in G(n) \tag{0.2}$$

le groupe des germes de C^∞-difféomorphismes de $(\mathbb{R}^n,0)$ qui fixent 0.

La recherche de critères algébriques ou topologiques d'existence d'intégrale première est un problème classique. Son histoire commence à la fin du 19ème siècle avec certains travaux de Darboux et Poincaré. Dans le cadre analytique réel ou complexe, il est résolu par G. Reeb [15], lorsque $J^1\omega$ est à singularité isoleé, puis il est repris dans [13, 14] et il est complètement résolu par B. Malgrange [7]: tout $\omega\in I(n)$ analytique possède une intégrale première analytique.

Un tel résultat n'est pas toujours vrai dans le cadre C^∞ comme le montre l'exemple suivant: soit ψ une application C^∞ de $\mathbb{R}$ dans $\mathbb{R}$ nulle sur $]-\infty,0]$. Alors le germe en $0\in\mathbb{R}^3$ de

$$\omega = dQ + \frac{\psi \circ Q(x_1\,dx_2 - x_2\,dx_1)}{x_1^2 + x_2^2} \quad \text{avec } Q = x_1^2 + x_2^2 - x_3^2$$

est un élément de $I(3)$ qui possède trois types de variétés intégrales: les surfaces $Q=t$ avec $t<0$, le cône $Q=0$ qui est une *séparatrice*; les autres coupent transversalement $\mathbb{R}^2 \times \{0\}$ suivant les courbes intégrales de l'équation différentielle $\omega/\mathbb{R}^2 \times \{0\}$. Cette équation possède une application premier retour de Poincaré $h \in G(1)$.

Si h est l'application identique de $\mathbb{R}$, les variétés intégrales de ω sont fermées et ω possède une intégrale première ([13], théorème 1) et réciproquement.

Si ω', h' sont obtenus de la même façon à partir d'un $\psi' \in \mathscr{E}(1)$, alors ω et ω' sont C^∞-équivalentes si et seulement si h et h' sont C^∞-conjugués (théorème 2).

Cet exemple est un résumé de toute la suite. Nous verrons avec la proposition 1, dont la démonstration technique est rejetée dans le chapitre 3, que les éléments hyperboliques de $I(n)$ possèdent toujours des séparatrices. Dans le chapitre 1 nous montrerons qu'un $\omega \in I(n)$ possède une intégrale première lorsque ses variétés intégrales sont fermées. Enfin, dans le chapitre 2, le concept d'*holonomie évanescente*, qui généralise celui d'application premier retour nous permettra de caractériser la C^∞-équivalence.

1. Existence d'intégrales premières

Notons $J_0^\infty \alpha$ le développement de Taylor en 0 d'un germe α en $0 \in \mathbb{R}^n$ d'une fonction, d'une forme, d'un difféomorphisme C^∞ et posons

$$\hat{\mathscr{E}}(n) = J_0^\infty \mathscr{E}(n), \qquad \hat{G}(n) = J_0^\infty G(n).$$

Une *intégrale première formelle* de $\omega \in I(n)$ est un $\hat{f} \in \hat{\mathscr{E}}(n)$ tel que

$$J_0^\infty \omega = \hat{g}\,d\hat{f} \quad \text{avec } \hat{g} \in \hat{\mathscr{E}}(n).$$

Théorème formel [7, 11, 13]. *Tout élément $\omega \in I(n)$ possède des intégrales premières formelles. Si $\hat{f}$ est l'une d'elles les autres lui sont L-équivalentes (au sens de Mather), c'est à dire du type $\hat{\varphi} \circ \hat{f}$ avec $\hat{\varphi} \in \hat{G}(1)$.*

Soit $f \in \mathscr{E}(n)$ un prolongement de Borel d'une intégrale première formelle $\hat{f}$ de $\omega \in I(n)$. Nous dirons que ω est *elliptique* si le germe de $f^{-1}(0)$ en $0 \in \mathbb{R}^n$ est réduit au point 0 et nous dirons que ω est *hyperbolique* dans le cas contraire. Cette définition ne souffre d'aucune ambiguité; en effet, un prolongement de Borel f' d'une autre intégrale première formelle $\hat{f}' = \hat{\varphi} \circ \hat{f}$, $\hat{\varphi} \in \hat{G}(1)$, de ω est $A - C^\infty$-équivalent à f, i.e.

$$f' = \varphi \circ f \circ \Phi \quad \text{avec } \Phi \in G(n), \quad \varphi \in G(1), \quad J_0^\infty \varphi = \hat{\varphi}, \tag{1.1}$$

d'après [10, 8], f et f' étant des germes de codimension finie.

Supposons que ω hyperbolique possède une intégrale première $f \in \mathscr{E}(n)$. Soit U un voisinage de 0 tel que 0 soit le seul point singulier de $\omega_U = g\,df$, un représentant de ω. Les variétés intégrales de $\omega_U/U - \{0\}$ sont les hypersurfaces

de niveau $f/U-\{0\}$; ce sont des fermés de U, excepté un nombre fini d'entre elles, les composantes connexes de $f^{-1}(0)\cap(U-\{0\})$. Nous dirons que $\omega\in I(n)$ est à feuilles fermées s'il existe un représentant ω_U de ω tel que ses variétés intégrales possèdent cette propriété.

Théorème 1. *(Existence d'intégrales premières C^∞). Un germe $\omega\in I(n)$ possède une intégrale première $f\in\mathscr{E}(n)$ si et seulement s'il est elliptique ou hyperbolique à feuilles fermées.*

Ainsi, comme dans le cadre holomorphe (voir [11]), il existe un critère topologique d'existence d'intégrale première C^∞ pour les éléments $I(n)$. La clef de la démonstration de ce théorème est la *«désingularisation»* des éléments de $I(n)$. Afin de pouvoir l'énoncer correctement, précisons quelques notations et rappelons le théorème de Milnor [12] et un très joli résultat de A.C. Roche [17].

Soient M une variété C^∞, K et K' des fermés de M. Nous notons $\Lambda^\cdot(M)$ (resp. $\mathscr{I}^\infty(K)$) l'algèbre des formes différentielles (resp. l'anneau de fonctions infiniment plates sur K) de classe C^∞ sur M et nous noterons $\Lambda^\cdot(X)_{K'}$ (resp. $\mathscr{I}^\infty(K)_{K'}$) leurs germes le long de K'. Lorsque $M=\mathbb{R}^n$, $K=K'=\{0\}$, nous écrivons de façon classique:

$$\Lambda^\cdot(n)=\Lambda^\cdot(\mathbb{R}^n)_0, \qquad \mathscr{I}^\infty(0)_0=\mathscr{M}^\infty(n).$$

Théorème de Milnor [12]. *Soit $P\in\mathbb{R}[x_1,x_2,\ldots,x_n]$ dont 0 est une singularité algébriquement isolée. Pour $\varepsilon>0$ assez petit, il existe $\eta>0$ tel que:*

i) *la sphère $\varepsilon S^{n-1}=\{x\in\mathbb{R}^n/\|x\|=\varepsilon\}$ est transverse aux hypersurfaces $P^{-1}(t)$ pour $|t|\leqq\eta$.*

ii) *$K=P^{-1}(0)\cap\varepsilon B$, où εB est la boule de bord εS^{n-1}, est homéomorphe à un cône de sommet 0 et de base $K\cap\varepsilon S^{n-1}$.*

iii) *$|P|/U-K$ où $U=|P|^{-1}(]0,\eta[)\cap\varepsilon B$ est un C^∞-fibré de base $]0,\eta[$, de fibre $Y=|P|^{-1}(\eta)\cap\in B$.*

Théorème de Roche [17, 16]. *Le fibré $|P|: U-K\to]0,\eta[$ possède une trivialisation:*

$$T: \Lambda^\cdot(-Y_0)\to U-K \quad \text{où} \quad X=[0,\eta[\times Y,\ Y_0=\{0\}\times Y \tag{1.2}$$

telle que l'isomorphisme T^* de $\Lambda^\cdot(U-K)$ sur $\Lambda^\cdot(x-Y_0)$ induit un isomorphisme

$$\bar{T}^*: \mathscr{I}^\infty(K)\Lambda^\cdot(U)\to\mathscr{I}^\infty(Y_0)\Lambda^\cdot(X).$$

La proposition suivante, qui pour un ω elliptique n'est rien d'autre que le théorème formel, montre qu'un $\omega\in I(n)$ hyperbolique possède des *séparatrices algébriques*, c'est à dire des variétés intégrales singulières algébriques, $K_1, K_2, \ldots, K_r$, qui sont les adhérences des composantes connexes de $P^{-1}(0)-\{0\}\cap U$. Elle joue un rôle analogue à celui des mises sous formes normales des champs de vecteurs complexes de $\mathbb{C}^2$ dans [9] et [11] comme nous le verrons dans le chapitre 2.

Proposition 1. *Soit ω un élément quelconque de $I(n)$. Il existe des coordonnées $(x_1,x_2,\ldots,x_n)$, un polynome P dans ces variables et une fonction g de classe C^∞*

sur $U=|P|^{-1}([0,\eta[)\cap\varepsilon B$ *tels que, pour* ε, η *assez petits,* ω *possède un représentant* ω_U *du type suivant:*

$$\omega_U=g(dP+\pi),\qquad \pi\in\mathscr{I}^\infty(K)\cdot\Lambda^*(U). \tag{1.3}$$

Notons (t,y) les points de X. La proposition 1 et le théorème de Roche implique la:

Désingularisation. *Soit* ω_U *un représentant de* $\omega\in I(n)$ *du type* (1.3). *Il existe une* (*unique*) 1*-forme* Ω, *intégrable de classe* C^∞ *sur* X *telle que*

$$\begin{aligned}&\Omega=dt+\alpha,\qquad \alpha\in\mathscr{I}^\infty(Y_0)\,\Lambda^*(X)\\ &\Omega/X-Y_0=T^*(dP+\pi)=T^*\left(\frac{\omega_U}{g}\right).\end{aligned} \tag{1.4}$$

Le feuilletage $\mathscr{F}_\Omega$ défini par Ω est C^∞ tangent au bord Y_0 de X transverse au facteur $[0,\eta[$ pour ε,η assez petits. T étant un C^∞-difféomorphisme, le lecteur peut visualiser le feuilletage $\mathscr{F}_{\omega_U}$ défini par ω_U de la façon suivante; il est constitué de deux parties distinctes: la réunion de ses séparatrices, le cone K, est son squelette et ses autres feuilles (lisses) sont des «semi-revêtements» de la fibre de Milnor Y qui s'enroulent autour de K. Ce spiralement est quantifié par la notion d'holonomie définie ci-dessous.

Notons $Y_1\ Y_2,\ldots,Y_q$ les composantes connexes de Y_0, $\mathrm{Diff}^\infty(\mathbb{R}^+,0)$ le groupe des germes de C^∞-difféomorphismes de $(\mathbb{R}^+,0)$ infiniment tangents à $1_{\mathbb{R}^+}$. Soit

$$h_k(\Omega)\colon \pi_1(Y_k)\to\mathrm{Diff}^\infty(\mathbb{R}^+,0),\qquad k=1,2,\ldots,q \tag{1.5}_k$$

l'homomorphisme d'holonomie de la feuille Y_k de $\mathscr{F}_\Omega$ et soit

$$h(\Omega)\colon \pi_1(Y)\to(\mathrm{Diff}^\infty(\mathbb{R}^+,0))^q \tag{1.5}$$

le produit cartésien des $h_k(\Omega)$ que nous appellerons représentation d'holonomie de Ω. Deux tels morphismes produits h^1, h^2:

$$h^i=\prod_k h_k^i\colon G=\prod_k G_k\to(\mathrm{Diff}^\infty(\mathbb{R}^+,0))^q,\qquad i=1,2$$

sont dits C^∞*-équivalents* s'il existe $(\varphi_k)\in(\mathrm{Diff}^\infty(\mathbb{R}^+,0))^q$ tel que pour chaque k

$$h_k^1(g)=\varphi_k\circ h_k^2(g)\circ\varphi_k^{-1}\qquad \text{si } g\in G_k. \tag{1.6}$$

Ils sont dits C^∞*-conjugués* s'il existe $\varphi\in G(1)$ tel que pour chaque k

$$h_k^1(g)=\varphi\circ h_k^2(g)\circ\varphi^{-1} \tag{1.7}$$

nous écrivons cette relation $h^1=\bar\varphi(h^2)$.

Rappelons un résultat bien connu sur les germes de feuilletages:

Lemme 1 [15, 4]. *Soient* Ω_i, pour $i=1,2$ *des* 1*-formes intégrables du type* (1.4):

$$\Omega_i=dt+\alpha_i,\qquad \alpha_i\in\mathscr{I}^\infty(Y_0)\,\Lambda^*(X).$$

Les représentations d'holonomie $h(\Omega_1)$ et $h(\Omega_2)$ de Y_0 sont C^∞-équivalentes si et seulement s'il existe un germe de C^∞-difféomorphisme $\tilde{\Phi}$ du germe de X le long de Y_0

$$\tilde{\Phi}: (t, y) \to (t + A(t, y), y), \qquad A \in \mathscr{I}^\infty(Y_0)$$

qui applique le germe de $\mathscr{F}_{\Omega_1}$ le long de Y_0 sur le germe de $\mathscr{F}_{\Omega_2}$ le long de Y_0.

Démonstration du théorème 1. Lorsque ω est elliptique, la fibre de Milnor Y de P est connexe et simplement connexe (homéomorphe à S^{n-1}) et lorsque ω est hyperbolique, les feuilles de $\mathscr{F}_{\omega_U}$ distinctes des séparatrices étant fermées, les feuilles de $\mathscr{F}_\Omega$ sont aussi fermées dans X. Dans les deux cas $\mathscr{F}_\Omega$ n'a pas d'holonomie, c'est à dire l'image de $h(\Omega)$ est réduite à $(1_{\mathbb{R}^+})^q$. D'après le lemme 1, il existe $A \in \mathscr{I}^\infty(Y_0)$ tel que

$$\Omega \wedge d(t + A(t, y)) = 0.$$

Les feuilles de $\mathscr{F}_\Omega$ sont les composantes connexes des hypersurfaces de niveau de $t + A$ et celles de $\mathscr{F}_{\omega_U}/U - K$ sont celles de $P + a$ avec

$$a = A \circ T^{-1}/U - K.$$

D'après le théorème de Roche, a s'étend à U en une fonction infiniment plate sur K, notée encore a telle que

$$\omega_U \wedge d(P + a) = 0.$$

Le théoréme de division de G. De Rham [1, 13] implique que le germe de $P + a$ en $0 \in \mathbb{R}^n$ est une intégrale première C^∞ de ω.

Remarque 1. Soit $\omega \in I(n)$ qui possède une intégrale première C^∞, f. Si la fibre de Milnor de f est connexe, c'est toujours le cas si ω est elliptique, l'ensemble des intégrales premières C^∞ de ω est constitué des $f' \in \mathscr{E}(n)$ L-équivalent à f (voir [13]), i.e.:

$$f' = \varphi \circ f, \qquad \varphi \in G(1).$$

Comme le montre l'exemple suivant, ceci n'est pas toujours vrai. Soient

$$f = x_1^2 + x_2^2 - x_3^2,$$

$$f' = \begin{cases} f & \text{si } x_3 \geqq 0 \quad \text{ou } f \geqq 0 \\ f + \psi \circ f & \text{si } x_3 \leqq 0 \quad \text{et } f \leqq 0 \end{cases}$$

où $\psi \in \mathscr{M}^\infty(1)$, $\psi(t) \neq 0$ si $t < 0$ et $\psi(t) \equiv 0$ si $t \geqq 0$.

Il est clair que $df \wedge df' = 0$ et f, f' ne sont pas L-équivalentes. Le lemme suivant montre comment f et f' sont reliées et il nous sera très utile pour préciser la proposition 1 dans le chapitre suivant.

Lemme 2. *Soient $f \in \mathscr{E}(n)$ un germe de codimension finie et soit $f' \in \mathscr{E}(n)$ tel que en tout point $a \in f^{-1}(0)$*

$$J_a^\infty df \wedge df' = 0.$$

i) *Il existe $\varphi \in \mathscr{E}(1)$ tel que $f' - \varphi \circ f \in \mathscr{I}^\infty(f^{-1}(0))$.*

ii) *Si, de plus, f' est de codimension finie, il existe $\Phi\in G(n)$ infiniment tangent à $1_{\mathbb{R}^n}$ le long de $f^{-1}(0)$ tel que:*

$$f'=\varphi\circ f\circ\Phi.$$

Démonstration. Notons encore f,f' des représentants des germes f,f' sur un voisinage U du type

$$U=|f|^{-1}([0,\eta[)\cap\varepsilon B$$

tel que 0 soit le seul point singulier de f et

$$J_a^\infty df\wedge df'=0 \quad \text{si } a\in K=f^{-1}(0)\cap U.$$

Notons $K_k, k=1,2,\ldots,p$, les composantes connexes de $K-\{0\}$. Puisque f est une submersion en tout point $a\in K_1$, il existe une carte (z,t) sur un voisinage v de a telle que

$$f(z,t)=t.$$

Soient $\bar f,\bar f'$ les champs tayloriens définis par f et f' le long de K. Dans la carte (z,t) on a

$$\bar f'(z,t)=\sum_{n\geqq 0} t^n f_n'(z) \quad \text{et } dt\wedge d\bar f'=0.$$

Les fonction (C^∞) $f_n'(z)$ sont des constantes α_n. Il existe un unique élément

$$\hat\varphi_v=\sum_{n\geqq 0}\alpha_n t^n\in\hat{\mathscr{E}}(1)$$

tel que

$$\bar f'/v=\hat\varphi_v\circ\bar f/v.$$

De cette unicité, on déduit que si v' est un voisinage du même type qui rencontre v, $\hat\varphi_v$ et $\hat\varphi_{v'}$ sont identiques. Un argument élémentaire de connexité permet d'affirmer

$$\bar f'/K_k=\hat\varphi_k\circ\bar f/K_k, \quad k=1,2,\ldots,p,$$

où les $\varphi_k\in\hat{\mathscr{E}}(1)$. La continuité en $0\in K$ des champs tayloriens $\bar f',\bar f$ implique que les $\hat\varphi_k$ sont tous égaux et ainsi

$$\bar f'=\hat\varphi\circ\bar f, \quad \hat\varphi\in\hat{\mathscr{E}}(1).$$

Désignons par $\varphi\in\mathscr{E}(1)$ un prolongement de Borel de $\hat\varphi$, alors $f'-\varphi\circ f$ est un élément de $\mathscr{I}^\infty(K)$.

Si f' est de codimension finie, φ est un élément de $G(1)$. Un argument «à la Mather» (voir par exemple [8] page 30) permet de montrer l'existence de fonctions $X_i(s,x)$ définies C^∞ sur un voisinage de $[0,1]\times\{0\}$ dans $\mathbb{R}\times U$, infiniment plates le long de $[0,1]\times K$ telles que

$$X(\varphi\circ f+t(f'-\varphi\circ f))=0 \quad \text{avec } X=\frac{\partial}{\partial s}+\Sigma X_i(s,x)\frac{\partial}{\partial x_i}.$$

Par intégration de X jusqu'au temps 1 on obtient un difféomorphisme infiniment tangent à l'identité le long de K; par construction, si Φ est son germe en 0 on a:

$$f'=\varphi\circ f\circ\Phi.$$

2. Holonomie et C^∞-équivalence

Dans tout ce paragraphe ω désigne un élément hyperbolique de $I(n)$ et nous notons encore ω un représentant de ω du type (1.3); c'est à dire

$$\begin{aligned} &\omega = g(df+\pi), \qquad \pi \in \mathscr{I}^\infty(K)\,\Lambda^\cdot(U) \\ &K = f^{-1}(0)\cap \varepsilon B, \qquad U = |f|^{-1}([0,\eta[)\cap \varepsilon B \end{aligned} \tag{1.3}$$

où f, g sont des fonctions C^∞ sur U, 0 étant le seul point singulier de f et g ne s'annulant pas.

Si $T: X - Y_0 \to U - K$ est une trivialisation (de Roche) du fibré de Milnor associé à $|f|$ on écrit ((1.4))

$$\Omega = dt + \alpha, \qquad \text{avec } \alpha \in \mathscr{I}^\infty(Y_0)\,\Lambda^\cdot(X) \tag{1.4}$$

la désingularisation correspondante de ω et

$$h(\Omega): \pi_1(Y) \to (\mathrm{Diff}^\infty(\mathbb{R}^+, 0))^q \tag{1.5}$$

sa représentation d'holonomie. Ce concept est un peu ambigu et avant d'énoncer le théorème 2 nous allons le préciser avec les quatre assertions suivantes.

Assertion 1. *La classe de C^∞-équivalence* (1.6) *de $h(\Omega)$, $[h(\Omega)]$ est indépendante de la trivialisation T. Elle ne dépend que de ω et f et nous la noterons*

$$[h(\Omega)] = h(\omega, f). \tag{2.1}$$

Démonstration. Soient Y_k, $k = 1, 2, \ldots, q$ les composantes connexes de Y et soient $X_k = [0, \eta[\times Y_k$ les composantes connexes correspondantes de X. Précisions tout d'abord la définition de $h(\Omega)$. C'est le produit cartésien des homomorphismes d'holonomie

$$h_k(\Omega): \pi_1(Y_k, a_k) \to \mathrm{Diff}(\mathbb{R}^+, 0)$$

que nous évaluons sur des petites courbes transverses τ_k à $\Omega_k = \Omega / X_k$ telles que

$$\tau_k: [0, \eta[\to X_k, \qquad \tau_k(0) = a_k, \qquad |f(T(\tau_k(t)))| = t.$$

Puisque Ω est C^∞-tangent à dt le long $0 \times Y$, la classe de C^∞-équivalence de $h(\Omega)$ est indépendante des τ_k ainsi choisis.

Soit T' une deuxième trivialisation de Roche de $|f|$. Par construction (voir [17]) on peut chosir, pour chaque k, τ_k telle que $T'^{-1} \circ T \circ \tau_k /]0, \eta[$ se prolonge en une courbe τ'_k du même type que τ_k, transverse à Ω' la désingularisation de ω par T':

$$\tau'_k: [0, \eta[\to X_k, \qquad \tau'_k(0) = a'_k, \qquad |f(T'(\tau'_k(t)))| = t.$$

Les homomorphismes d'holonomie $h_k(\Omega)$, $h_k(\Omega')$ sont égaux puisque

$$\Omega' / X - Y_0 = (T'^{-1} \circ T)^*(\Omega / X - Y_0);$$

ce qui démontre l'assertion.

Assertion 2. *Soit Φ un élément de $G(n)$. Alors*

$$\Phi^*(\omega)=(g\circ\Phi)(d(f\circ\Phi)+\Phi^*(\pi))$$

est une écriture du type (1.3) *pour $\Phi^*(\omega)$ et*

$$h(\Phi^*(\omega),f\circ\Phi)=h(\omega,f). \tag{2.2}$$

Démonstration. La 1-forme $\Phi^*(\pi)$ étant infiniment plate le long de $\Phi^{-1}(K)$, la première partie de l'assertion est évidente. D'autre part, $\Phi^{-1}\circ T$ étant une trivialisation du fibré de Milnor de $|f\circ\Phi|,\omega$ et $\Phi^*(\omega)$ possède les mêmes désingularisations, d'où (2.2).

Assertions 3. *Soit φ un élément de $G(1)$. Alors*

$$\omega=g_1(d(\varphi^{-1}\circ f)+g_1^{-1}\cdot\pi) \quad \text{avec } g_1=\varphi'\circ\varphi^{-1}\circ f$$

est encore une écriture de type (1.3) *de ω. La C^∞-conjugaison définie par φ (voir* (1.7)) *échange $h(\omega,f)$ et $h(\omega,\varphi^{-1}\circ f)$ et nous écrivons:*

$$\bar{\varphi}(h(\omega,\varphi^{-1}\circ f))=h(\omega,f). \tag{2.3}$$

Démonstration. La première partie de l'assertion est évidente. Soit $(\tilde{\varphi})$ le difféomorphisme de X:

$$(\tilde{\varphi})\colon (t,y)\to(\varphi(t),y)).$$

Alors $T\circ\tilde{\varphi}/X-Y_0$ est une trivialisation du fibré de Milnor de $|\varphi^{-1}\circ f|$. Il lui correspond une nouvelle désingularisation Ω' de ω (associée à la nouvelle écriture (1.3)) telle que

$$(\tilde{\varphi})^*(\Omega)\wedge\Omega'=0.$$

Les représentations d'holonomie $h(\Omega)$ et $h(\Omega')$ sont C^∞-conjuguées par φ (au sens de 1.7) et en passant aux classes de C^∞-équivalence on obtient (2.3).

Assertion 4. *Soit $g'(df'+\pi')$ un autre représentant de ω du type* (1.3). *Il existe $\varphi\in G(1)$, $\Phi\in G(n)$ C^∞-tangent à $1_{\mathbb{R}^n}$ le long de K tels que*

$$f=\varphi\circ f'\circ\Phi \quad \textit{et alors } h(\omega,f')=\bar{\varphi}(h(\omega),f). \tag{2.4}$$

Démonstration. Puisque K a un sens intrinsèque (c'est la réunion des séparatrices de ω) on a:

$$K=f^{-1}(0)=f'^{-1}(0) \quad \text{et } J^\infty_a df\wedge df'=0 \quad \text{si } a\in K.$$

D'après le lemme (2.2), ceci démontre la première partie de (2.4). Des deux assertions précédentes on déduit:

$$h(\omega,f')=\bar{\varphi}((h(\omega),f\circ\Phi))=\bar{\varphi}(h((\varphi^{-1})^*(\omega),f)).$$

La trivialisation T de $|f|$ permet de construire une désingularisation Ω' de $(\Phi^{-1})^*(\omega)$ pour l'écriture de type (1.3):

$$(\Phi^{-1})^*(\omega)=g''(df+\pi'') \quad \text{avec } g''=\varphi'(f)\cdot g'\circ\Phi^{-1}.$$

D'après le théorème de Roche, le difféomorphisme Φ, qui est infiniment tangent à $1_{\mathbb{R}^n}$ le long de K, se relève dans X en un difféomorphisme $\tilde{\Phi}$ infiniment tangent à 1_X le long de Y_0 tel que:

$$\tilde{\Phi}^*(\Omega) \wedge \Omega' = 0.$$

Les représentations d'holonomie de $h(\Omega)$ et $h(\Omega')$ sont C^∞-équivalentes, d'après le lemme 1, et par conséquent:

$$h((\Phi^{-1})^*(\omega), f) = h(\omega, f).$$

Théorème 2. *Soient $g(df+\pi)$ et $g'(df'+\pi')$ des représentants de type* (1.3) *de ω et $\omega' \in I(n)$. Alors, ω, ω' sont C^∞-équivalents si et seulement si les deux conditions suivantes sont satisfaites:*

i) *f et f' sont A-équivalents, i.e. il existe $\varphi \in G(1)$ et $\Phi \in G(n)$ tels que*

$$f = \varphi \circ f' \circ \Phi$$

ii) *les classes de C^∞-équivalence $h(\omega, f)$, $h(\omega', f')$ sont C^∞-conjuguées par φ, i.e.*

$$h(\omega, f) = \bar{\varphi}\, h(\omega', f').$$

Démonstration. Remarquons tout d'abord que i) est vrai si et seulement si ω et ω' sont formellement conjugués: il est clair que si f et f' sont A-équivalents:

$$J^\infty \omega \wedge \Phi^*(\omega') = 0.$$

Réciproquement, soit $\Phi \in G(n)$ tel que cette relation soit vraie. Alors on a

$$J^\infty df \wedge f' \circ \Phi = 0$$

et comme nous l'avons vu dans (1.1) il existe $\varphi \in G(1)$, $\Phi_1 \in G(n)$ tels que

$$f = \varphi \circ (f' \circ \Phi) \circ \Phi_1.$$

Supposons ω, ω' C^∞-équivalents et soit $\Phi \in G(n)$ tel que $\Phi^*(\omega') \wedge \omega = 0$. D'après le théorème de G. De Rahm,

$$\Phi^*(\omega') = g' \circ \Phi(d(f' \circ \Phi) + \Phi^*(\pi'))$$

est une nouvelle écriture du type (1.3) de ω (Assertion 2). L'assertion 4 montre que i) et ii) sont alors vrais.

Si ω, ω' vérifient i) et ii), d'après les assertions 2 et 3

$$h(\omega, f) = \bar{\varphi}(h(\omega', f)) = h(\Phi^*(\omega'), f).$$

Soient Ω, Ω' les désingularisations de ω, $\Phi^*(\omega')$ correspondants à la trivialisation T de $|f|$. Leurs représentations d'holonomie $h(\Omega)$, $h(\Omega')$ sont C^∞-equivalentes (égalité précédente). Il existe, d'après le lemme 1 un C^∞-difféomorphisme $\tilde{\Phi}_1$ de X, infiniment tangent à 1_X le long de Y_0 tel que

$$\tilde{\Phi}_1^*(\Omega') \wedge \Omega = 0,$$

et d'après le théorème de Roche il existe $\Phi_1 \in G(n)$ tel que

$$\Phi_1^*(\Phi^*(\omega')) \wedge \omega = 0.$$

Assertion 5. *Soient $f \in \mathscr{E}(n)$ un germe de codimension finie, Y la fibre de Milnor de $|f|$ et soit*

$$h_f : \pi_1(Y) \to (\mathrm{Diff}(\mathbb{R}^+, 0))^q$$

un homomorphisme (au sens de 1.5*). Il existe $\omega \in I(n)$ tel que*

$$h(\omega, f) = [h_f] \quad \text{où } \omega = df + \pi, \quad \pi \in \mathscr{I}^\infty(f^{-1}(0)) \cdot \Lambda^*(n).$$

Démonstration. Soit $T : X - Y_0 \to U - K$ une trivialisation (de Roche) du fibré de Milnor d'un représentant noté encore f du germe f (notations de (1.4)). Il existe un feuilletage $\tilde{\mathscr{F}}$ de X, infiniment tangent à Y_0 tel que le produit cartésien des représentations d'holonomie des composantes connexes de Y_0 soit exactement h_f (voir par exemple [15]). Le champ de $(n-1)$-plans tangents à $\tilde{\mathscr{F}}$ est formé des noyaux d'une 1-forme du type:

$$\Omega = dt + \alpha, \quad \text{avec } \alpha \in \mathscr{I}^\infty(Y_0) \cdot \Lambda^0(X).$$

D'après le théorème de Roche il existe $\pi \in \mathscr{I}^\infty(K)\,\Lambda^*(U)$ tel que

$$\omega = df + \pi \quad \text{et} \quad T^*(\omega) = \Omega / X - Y_0.$$

Ceci démontre l'assertion.

Le corollaire suivant est un résumé des cinq assertions et du théorème 2.

Corollaire. *Soit $J(n)$ l'ensemble des paires (f, h_f) définies dans l'assertion* 5*) et soit $\bar{J}(n)$ son quotient par la relation*

$$(f, h_f) \sim (f', h'_{f'}) \Leftrightarrow \begin{cases} f = \varphi \circ f' \circ \Phi, & \varphi \in G(1),\ \Phi \in G(n) \\ \bar{\varphi}[h_f] = [h_{f'}]. \end{cases}$$

L'application F de $I(n)$ dans $\bar{J}(n)$ définie par:

$$F : \omega = g(df + \pi) \to [(f, h(\omega, f))] \quad \text{où } \pi \in \mathscr{I}^\infty(f^{-1}(0)\,\Lambda^*(n)$$

est bien définie, compatible avec la relation de C^∞-équivalent sur $I(n)$; elle induit une bijection $\bar{F}$ du quotient $\bar{I}(n)$ de $I(n)$ par cette relation dans $\bar{J}(n)$.

Ce corollaire et le théorème 2 sont à rapprocher de résultats récents de J. Martinet et J.P. Ramis [9] sur les germes en $0 \in \mathbb{C}^2$ d'équations de Pfaff holomorphes résonnantes:

$$\omega = p\,x\,dy + q\,y(1 + a(x, y))\,dx = 0.$$

En effet, dans ce cas, comme dans le cadre $I(n)$, on montre que les classes d'équivalences de telles équations sont caractérisées par une paire du type (formel, germe de difféomorphisme). Cependant, il subsiste une différence fondamentale entre les deux: dans le cas des équations résonnantes «l'holonomie» est portée par les séparatrices alors que pour les $\omega \in I(n)$ l'holonomie est

«portée» par la fibre de Milnor d'une fibration C^∞-tangente à ω le long de ses séparatrices. Néanmoins, ce deuxième point de vue qui apparait déjà dans le cadre holomorphe chez Ju.S. Il'Iasenko, permet peutêtre de mieux comprendre géométriquement certains résultats de [9]; par exemple la relation entre les holonomies des deux variétés invariantes, ...

3. Démonstration de la proposition 1

Dans toute la suite ω désigne un représentant d'un élément hyperbolique de $I(n)$ que l'on peut écrire, d'après le théorème formel,

$$\omega = dP + \pi$$

où P est une fonction polynomiale de $x=(x_1, x_2, \ldots, x_n)$ et π une 1-forme, infiniment plate en 0 de classe C^∞ sur un ouvert U.

La démonstration de la proposition 1 repose sur le théorème de désingularisation d'Hironaka: il existe une variété analytique $\tilde{U}$ et une surjection propre analytique θ: $\tilde{U} \to U$ telle que si

$$D_0 = \theta^{-1}(0), \qquad D = \theta^{-1}(P_{(0)}^{-1}), \qquad \tilde{P} = P \circ \theta:$$

i) $U - D_0$ est dense dans U et $\theta / U - D_0$ est un isomorphisme sur $U - \{0\}$,
ii) l'adhérence $\tilde{K}$ de $D - D_0$ est une sous-variété lisse,
iii) D est à croisements normaux.

Plus précisément, en un point $a \in D_0$, il existe une carte centrée en ce point que nous notons

$$\begin{aligned} (x,t) &= (x, t_1, \ldots, t_{n-1}) \quad \text{si } a \in \tilde{K} \quad \text{et alors } \tilde{P}(x,t) = x\, t^r \\ t &= (t_1, \ldots, t_n) \quad \text{si } a \notin \tilde{K} \quad \text{et alors } |\tilde{P}(t)| = t^r \\ r &= (r_1, r_2, \ldots, r_p, 0, 0, \ldots, 0) \equiv (r_1, r_2, \ldots, r_p). \end{aligned} \tag{3.1}$$

Nous allons tout d'abord montrer que la proposition 1 se déduit du lemme suivant:

Lemme 3 *(existence d'integrale première locale). En tout point $a \in D_0$, il existe un germe f_a de fonction C^∞ tel que:*

$$\theta^*(\omega)_a \wedge df_a = 0 \quad \text{et} \quad h_a = (f_a - \tilde{P}) \in \mathscr{I}^\infty(D_0)_a. \tag{3.2}$$

Une généralisation du théorème de division de De Rham par un germe de 1-forme qui n'est plus à singularité isolée due à J.C. Tougeron [21] (voir encore [3]) permet d'écrire

$$\theta^*(\omega)_a = g_a\, df_a \quad \text{et} \quad (g_a - 1) \in \mathscr{I}^\infty(D_0)_a. \tag{3.2'}$$

Nous dirons encore que f_a est une *intégrale première* C^∞ de $\theta^*(\omega)_a$. Nous noterons f_V, g_V des représents de f_a, g_a sur une carte de (3.1), centrée en a et de domaine de définition V tels que:

$$\theta^*(\omega)/V = g_V\, df_V, \quad \text{avec}$$
$$h_V = (f_V - \tilde{P}) \quad \text{et} \quad (g_V - 1) \in \mathscr{I}^\infty(D_0 \cap V). \tag{3.2}''$$

Lorsque $a \in \tilde{K}$ nous avons ainsi:

$$f_V = x\, t^r + h_V = (x + h'_V)\, t^r \quad \text{avec } h'_V \in \mathscr{I}^\infty(D_0 \cap V)$$

et lorsque $a \notin \tilde{K}$, nous écrirons:

$$f_V = \varepsilon\, t^r + h_V = \varepsilon(1 + h'_V)\, t^r \quad \text{avec } h'_V \in \mathscr{I}^\infty(D_0 \cap V) \quad \text{et} \quad \varepsilon = \pm 1.$$

Démonstration de la proposition 1. Soit $\{V_j\}_{j=1,2,\ldots,m}$ un recouvrement de D_0 par des ouverts V_j de (3.2)'' associés à des $f_j = f_{V_j}$ qui vérifient:

$$\begin{aligned} &V_j = |f_j|^{-1}([0, \eta[) \cap \varepsilon B(a) \\ &V_j \cap \tilde{K} \neq \emptyset \quad \text{si } 1 \leqq j \leqq l \\ &V_j \cap \tilde{K} = \emptyset \quad \text{si } l+1 \leqq j \leqq m. \end{aligned} \tag{3.3}$$

La 1-forme $\theta^*(\omega)$, ou plus exactement les formes $\tilde{\omega}_j$ définies sur les V_j, par

$$\tilde{\omega}_j = \frac{\theta^*(\omega)}{t^{r-1}}, \quad r - 1 = (r_1 - 1, \ldots, r_p - 1)$$

définissent un feuilletage $\mathscr{F}_{\tilde{\omega}}$ de $\tilde{U}_0 = \bigcup_{j=1}^{m} V_j$ qui a pour lieu singulier $D_0 \cap \tilde{K}$ et dont D_0 est une réunion de *séparatrices*, i.e de feuilles lisses, propres (fermées dans U_0) qui rencontrent le lieu singulier de $\mathscr{F}_{\tilde{\omega}}$.

a) *Existence de séparatrices distinctes de* D_0. Soit F_1 la composante connexe de $\tilde{K} \cap D_0$ qui coupe V_1; elle est contenue dans une feuille singulière L_1 de $\mathscr{F}_{\tilde{\omega}}$ qui, d'après (3.2)'' et (3.3), est propre. En effet, c'est la réunion des

$$\overline{f_j^{-1}(0) - D_0} = (x + h'_j)^{-1}(0)$$

pour les j tels que V_j rencontrent F_1. Elle coupe D_0 suivant F_1 et elle est infiniment tangente à $\tilde{K}$ le long de F_1. C'est à dire:

$$D' = \bigcup_{j=1}^{m} f_j^{-1}(0) = D_0 \cup L_1 \cup L_2 \ldots \cup L_q$$

est l'ensemble des séparatrices de $\mathscr{F}_{\tilde{\omega}}$; ces séparatrices sont lisses et à croisements normaux comme D, infiniment tangentes à D le long de D_0.

b) Equation globale de D'. Soient j, j' tels que $V_j \cap V_{j'} \neq \emptyset$, W_k une composante connexe de $V_j \cap V_{j'} - D_0$ et soient $f_j, f_{j'}$ les intégrales premières correspondantes de (3.2)'. Elles sont reliées par:

$$df_j \wedge df_{j'} = 0, \quad f_j - f_{j'} \in \mathscr{I}^\infty(D_0 \cap V_j \cap V_{j'}). \tag{3.4}$$

Leurs restrictions à W_k ont les mêmes hypersurfaces de niveau; chacune d'elles est connexe et c'est d'après (3.3) l'intersection d'une hypersurface de niveau de

f_j ou $f_{j'}$ avec W_k. Il existe un C^∞-difféomorphisme φ_k entre les images de W_k par $f_j, f_{j'}$ tel que

$$f_j/W_k = \varphi_k \circ f_{j'}/W_k.$$

Le même argument que dans [13] montre que φ_k se prolonge en un C^∞-difféomorphisme $\bar{\varphi}_k$ entre deux voisinages de $0 \in \mathbb{R}$, infiniment tangent à $1_{\mathbb{R}}$ en 0 d'après (3.4). On en déduit par le même argument que dans le lemme 2:

$$f_j = f_{j'} + h_{i,j}, \qquad h_{i,j} \in \mathscr{I}^\infty(D' \cap V_j \cap V_{j'}).$$

Il existe un unique champ taylorien sur D' ([6, 20]) qui en un point $b \in D' \cap V_j$ a pour valeur $J_b^\infty f_j$. Soit $\tilde{P}'$ un prolongement de Whitney de ce champ. Il est (comme les f_j) infiniment tangent à $\tilde{P}$ le long de D_0:

$$\begin{aligned} \tilde{P}' &= \tilde{P} + \tilde{H}, \qquad \tilde{H} \in \mathscr{I}^\infty(D_0) \\ f_j &= \tilde{P}' + h_j'', \qquad h_j'' \in \mathscr{I}^\infty(D'). \end{aligned} \tag{3.5}$$

Alors, (3.2)'' s'écrit:

$$\theta^*(\omega)/V_j = g_j\, d\tilde{P}' + g_j\, dh_j'';$$

en utilisant une partition de l'unité, on obtient:

$$\theta^*(\omega) = \tilde{g}\, d\tilde{P}' + \tilde{\pi}', \qquad (1-\tilde{g}) \in \mathscr{I}^\infty(D_0), \qquad \tilde{\pi}' \in \mathscr{I}^\infty(D')\, \Lambda^\cdot(\tilde{U}_0).$$

Soient U_0, K' les images de $\tilde{U}_0$ et D' par θ. Il existe, d'après [17] (ou même seulement [20]), des fonctions C^∞ sur U_0, g, P', et une 1-forme $\pi' \in \mathscr{I}^\infty(K') \cdot \Lambda^\cdot(U_0)$ tels que

$$\begin{aligned} &\tilde{g} = g \circ \theta, \qquad \tilde{P}' = P' \circ \theta, \qquad \tilde{\pi}' = \theta^*(\pi'), \\ &\omega = g\, dP' + \pi', \end{aligned}$$

et ceci démontre la proposition 1 puisque

$$J_0^\infty P' = P \quad \text{et} \quad K' = P'^{-1}(0) \cap U_0.$$

Il existe certainement une démonstration du lemme 3 utilisant les arguments de Cerveau-Lins [3]; c'est à dire, au voisinage de chaque $a \in D_0$, on construit une action locale de $\mathbb{R}^{n-1}$ dont les orbites sont les variétés intégrales de $\theta^*(\omega)$ et ensuite avec les techniques de Chaperon [2] on linéarise cette action. Mais dans le cas particulier qui nous intéresse, il est plus simple de construire directement les intégrales premières locales.

La restriction de $\theta^*(\omega)$ à un voisinage V de $a \in D_0$ s'écrit dans une carte (3.1)

$$\theta^*(\omega) = \begin{cases} d(x\, t^r) + \theta^*(\pi) & \text{si } a \in \tilde{K} \\ \varepsilon\, d(t^r) + \theta^*(\pi) & \text{si } a \notin \tilde{K}, \qquad \varepsilon = \pm 1 \end{cases}$$

avec $r = (r_1, r_2, \ldots, r_p, 0, \ldots, 0)$, $r_k \neq 0$ et $\theta^*(\pi) = \tilde{\pi} \in \mathscr{I}^\infty(D_0)\, \Lambda(\tilde{U})$. Démontrons tout d'abord que $\theta^*(\omega)$ est fibrée sur sa restriction au sous-espace linéaire $\bigcap_{i>p} \{t_i = 0\}$; plus précisément, montrons:

Assertion 1. *Il existe une carte notée encore* (x,t) *si* $a\in\tilde{K}$, (t) *si* $a\notin\tilde{K}$ *infiniment tangente à la carte* (3.1) *le long de* D_0 *telle que*

$$\begin{aligned}\theta^*(\omega)&=d(x\,t^r)+c\,dx+\sum_{i=1}^{p}c_i\,dt_i, \qquad \text{si } a\in K\\ \theta^*(\omega)&=\varepsilon\,dt^r+\sum_{i=1}^{p}c_i\,dt_i, \qquad \text{si } a\notin\tilde{K},\quad \varepsilon=\pm 1.\end{aligned} \tag{3.6}$$

où $c,c_i\in\mathscr{I}^\infty(D_0)$ *sont des fonctions de* $(t_1,\dots,t_p)$ *si* $a\notin\tilde{K}$ *et de* $(x,t_1,t_2,\dots,t_p)$ *si* $a\in\tilde{K}$.

Démonstration lorsque $a\in\tilde{K}$. Pour $k=p+1,\dots,n-1$, il existe $d_k\in\mathscr{I}^\infty(D_0)$ tel que le champ de vecteurs

$$X_k=\frac{\partial}{\partial t_k}-d_k\frac{\partial}{\partial x}$$

annule $\theta^*(\omega)$. Puisque $\theta^*(\omega)$ est intégrable le crochet de deux tels champs

$$[X_k,X_l]=\left(-\frac{\partial d_l}{\partial t_k}+\frac{\partial d_k}{\partial t_l}\right)\frac{\partial}{\partial x}$$

annule aussi $\theta^*(\omega)$. Il est donc nul. Les coordonnées cherchées s'obtiennent à partir de l'action de $(\mathbb{R}^{n-p-1},0)$ correspondant à ces champs (voir [3] chapitre VI, [5]).

Lorsque $a\notin\tilde{K}$, l'assertion se démontre par les mêmes arguments. La différence essentielle entre les cas $a\in\tilde{K}$ et $a\notin\tilde{K}$ est: lorsque $a\notin\tilde{K}$ les plans de coordonnées $t_i=0$, $i=1,2,\dots,p$ sont des séparatrices de $\theta^*(\omega)$, par contre si $a\in\tilde{K}$ les $t_i=0$ sont encore des séparatrices mais $x=0$ n'en est pas nécessairement une. L'assertion suivante a pour but de mettre fin à ce privilège.

Assertion 2. *Si* $a\in\tilde{K}$, *il existe une carte notée encore* (x,t) *infiniment tangente à la carte* (3.1) *le long de* D_0 *telle que*

$$\theta^*(\omega)=d(x\,t^r)+c\,dx+\sum_{i=1}^{p}c_i\,dt_i$$

où c *et les* c_i *sont des fonctions de* $(x,t_1,t_2,\dots,t_p)$ *infiniment plates le long des plans* $x=0$ *et* $t_i=0$.

Démonstration. Soit X l'unique champ de vecteurs C^∞ du type

$$X=-|r|(x+d)\frac{\partial}{\partial x}+\sum_{i=1}^{p}t_i\frac{\partial}{\partial t_i}, \qquad d\in\mathscr{I}^\infty(D_0\cap V)$$

qui annule $\theta^*(\omega)$. La démonstration classique [16, 17, 2] du théorème de linéarisation d'un champ de vecteurs qui est une perturbation plate en 0 de sa partie linéaire permet de construire un difféomorphisme H du type

$$H:(x,t)\to(x+f,t_1+f_1,\dots,t_p+f_p),$$

où f et les $f_i \in \mathscr{I}^\infty(D_0 \cap V)$, qui linéarise X. En désignant encore par (x,t) la carte $(x,t) \circ H$ on peut donc supposer que le champ linéaire

$$X_0 = -|r| \frac{\partial}{\partial x} + \sum_{i=1}^{p} t_i \frac{\partial}{\partial t_i}$$

annule $\theta^*(\omega)$. Nous allons montrer que dans ces nouvelles coordonnées $\theta^*(\omega)$ possède les propriétés de l'assertion. Les fonctions c et c_i de l'écriture (3.6) de $\theta^*(\omega)$ appartenant à $\mathscr{I}^\infty(D_0 \cap V)$, sont divisibles par t^r. Leurs quotients c' et c'_i par t^r sont encore des éléments de $\mathscr{I}^\infty(D_0 \cap V)$. Soient alors

$$\bar{\omega} = \frac{\theta^*(\omega)}{t^r(1+c')} = dx + x\frac{dt^r}{t^r} + \sum_{i=1}^{p} c'_i dt_i,$$

$$\hat{\omega} = dx + x\frac{dt^r}{t^r} + \pi_0 + x\pi_1 + \ldots + x^k \pi_k + \ldots$$

le développement de taylor de $\bar{\omega}$ suivant les puissances de x; les π_k s'écrivant

$$\pi_k = \sum_{i=1}^{p} c'_{k,i} dt_i$$

où les $c'_{k,i}$ sont des fonctions de $(t_1, t_2, \ldots, t_p)$, infiniment plates le long des hyperplans $t_i = 0$. D'autre part, la condition X_0 annule $\theta^*(\omega)$, entraine

$$\pi_k \left(\sum_{i=1}^{p} t_i \frac{\partial}{\partial t_i} \right) = 0. \tag{3.7}$$

En écrivant dans la condition d'intégralité

$$(x^{-1}\hat{\omega}) \wedge d(x^{-1}\hat{\omega}) = 0$$

la nullité du terme en $x^{-2} dx \wedge \ldots$ on obtient

$$0 = d\pi_0 + \frac{dt^r}{t^r} \wedge \pi_0;$$

c'est à dire que:

$$t^r \pi_0 = df_0.$$

D'après (3.7), f_0 est constante sur les droites passant par $0 \in \mathbb{R}^p$. C'est une fonction constante et $\pi_0 = 0$.

Supposons $\pi_0 = \pi_1 = \ldots = \pi_{n-1} = 0$; c'est à dire:

$$x^{-1}\omega = \frac{dx}{x} + \frac{dt^r}{t^r} + x^{n-1}\pi_n + x^n \pi_{n+1} + \ldots.$$

En écrivant la nullité du terme en $x^{n-2} dx \wedge \ldots$ dans la condition d'intégrabilité de $x^{-1}\hat{\omega}$, on obtient:

$$d\pi_n - (n-1)\frac{dt^r}{t^r} \wedge \pi_n = 0;$$

la 1-forme $\frac{\pi_n}{(t^r)^{n-1}}$ est fermée. Les coefficients de π_n étant plats sur les hyperplans $t_i=0$,

$$\frac{\pi_n}{(t^r)^{n-1}}=df_n$$

est la différentielle d'une fonction C^∞ en t. Comme précédemment, (3.7) permet d'affirmer que π_n est identiquement nul; ce qui démontre l'assertion 2.

Démonstration de la proposition 1. Il suffit de démontrer, après les deux assertions, qu'un germe en $0\in\mathbb{R}^p$ de 1-forme intégrable

$$\begin{aligned}&\omega_p=dt^r+\sum_{i=1}^{p}a_i(t)\,dt_i\\&a_i\in\mathscr{I}^\infty(D_p),\qquad D_p=\bigcup_{i=1}^{p}\{t_i=0\}\end{aligned}\tag{3.8}$$

possède une intégrale première f_p du type

$$f_p(t)=t^r(1+h_p),\qquad h_p\in\mathscr{I}^\infty(D_p).\tag{3.9}$$

Cette démonstration se fait par récurrence sur p. Lorsque $p=1$, la primitive k_1 de $r_1 a_1$ est un élément de $\mathscr{I}^\infty(D_1)$ et

$$\omega_1=d(t_1^{r_1}+k_1)=d(t_1^{r_1}(1+h_1)).$$

Supposons (3.9) vrai pour des ω_k du type (3.8) pour $k=1,2,\ldots,p-1$, et montrons que (3.9) est vrai pour ω_p. Soit $X=X_1+X_\infty$ un germe en $0\in\mathbb{R}^p$ de champ de vecteurs qui annule ω_p, dont la partie linéaire est

$$X_1=\sum_{i=1}^{p-1}t_i\frac{\partial}{\partial t_i}-\frac{|r'|}{r_p}t_p\frac{\partial}{\partial t_p},\qquad r'=(r_1,\ldots,r_{p-1}),$$

les coefficients de X_∞ étant des éléments de $\mathscr{I}^\infty(D_p)$. Le théorème de linéarisation de Sternberg [19, 18, 2], ou plutôt sa démonstration permet de trouver de nouvelles coordonnées notées encore $t=(t_1,t_2,\ldots,t_p)$, infiniment tangentes le long de D_p aux anciennes dans lesquelles nous avons

$$\begin{aligned}&X=X_1=\sum_{i=1}^{p-1}t_i\frac{\partial}{\partial t_i}-\frac{|r'|}{r_p}t_p\frac{\partial}{\partial t_p},\\&\omega_p=dt^r+\sum_{i=1}^{p}a_i(t)\,dt_i,\\&\omega_p(X_1)=0\quad\text{et}\quad a_i(t)\in\mathscr{I}^\infty(D_p).\end{aligned}$$

Modulo une homothétie de centre $0\in\mathbb{R}^p$, on peut supposer que ω_p possède un représentant noté encore ω_p sur un voisinage V du polycube $|t_i|\leqq 1$.

Notons $t'=(t_1,t_2,\ldots,t_{p-1})$, J le plongement de $\mathbb{R}^{p-1}$ dans $\mathbb{R}^p$:

$$J(t')=(t',1)\quad\text{et}\quad\omega_{p-1}=J^*(\omega_p).$$

Il existe d'après l'hypothèse de récurrence $h_{p-1} \in \mathscr{I}^\infty(D_{p-1})$ tel que

$$f_{p-1}(t') = t'^{r'}(1 + h_{p-1}(t'))$$

soit une intégrale première de ω_{p-1}. Notons encore f_{p-1} un représentant de ce germe défini sur un voisinage V_{p-1} de $0 \in \mathbb{R}^{p-1}$ et tel que

$$df_{p-1} \wedge J^*(\omega_p) = 0.$$

Soit V_p^+ l'ensemble des points $t = (t', t_p)$ du polycube $|t_i| \leqq 1$ tel que la courbe intégrale

$$t(s) = (e^s t', e^{-\frac{|r'|}{r_p} s} t_p)$$

du champ $X = X_1$ coupe $V_{p-1} \times \{1\}$ en un point $(t'_1, 1)$. L'application

$$f_p^+ : V_p^+ \to \mathbb{R}$$

définie par $f_p^+(t) = f_{p-1}(t'_1)$ est C^∞, constante sur les variétés intégrales de ω_p / V_p^+ et elle s'écrit

$$f_p^+(t', t_p) = t'^{r'} \cdot t_p^{r_p}(1 + h_{p-1}(t' \cdot t_p^{\frac{r_p}{|r'|}})).$$

Puisque $h_{p-1} \in \mathscr{I}^\infty(D_{p-1})$ elle s'écrit encore

$$f_p^+(t) = t^r(1 + h_p^+(t)), \qquad h_p^+ \in \mathscr{I}^\infty(D_p).$$

On construit de la même façon

$$f_p^-(t) = t^r(1 + h_p^-(t)), \qquad h_p \in \mathscr{I}^\infty(D_p)$$

et alors

$$f_p(t) = \begin{cases} f_p^+(t) & \text{si } t_p > 0 \\ 0 & \text{si } t_p = 0 \\ f_p^-(t) & \text{si } t_p < 0 \end{cases}$$

est l'intégrale première cherchée.

Bibliographie

1. De Rahm, G.: Sur la division des formes et des courants par une forme linéaire. Commentari Math. Hel. **28**, 346–352 (1954)
2. Chaperon, M.: Sur la linéarisation C^∞ des germes hyperboliques d'actions: Cours de l'E.N.S. A paraitre dans Astérisque 1982
3. Cerveau, D., Lins, A.: Forme tangente à des actions commutatives. A paraitre (1980)
4. Haefliger, A.: Variétés feuilletées. Ann. Ec. Norm. Sup. de Pise Série 3, **6**, 367–397 (1962)
5. Kupka, I.: Singularities of integrable structurally stable Pfaffian forms. Proc. Natl. Acad. Sc. **52**, 1431–1432 (1964)
6. Malgrange, B.: Ideals of differentiable functions. Oxford: Oxford University Press 1968
7. Malgrange, B.: Frobenius avec singularités 1-codimension un. Publ. Math. IHES **46**, 163–173 (1976)
8. Martinet, J.: Singularités des fonctions et applications différentiables. Monographias de Mat. DAPUC/R.J.N° 1 (1977)

9. Martinet, J., Ramis, J.P.: Classification analytique des équations différentielles non linéaires résonnantes du premier ordre. A paraitre aux publ. Math. IHES (1983)
10. Mather, J.: III: Finitely determined map germs. Publ. IHES **35**, 127–156 (1968)
11. Mattei, J.F., Moussu, R.: Holonomie et intégrale première. Ann. Sc. Ec. Norm. Sup. 4ème série *t.* **13**, 459–523 (1980)
12. Milnor, J.: Singular points of complex hypersurfaces. Princeton (1968)
13. Moussu, R.: Sur l'existence d'intégrales premières pour un germe de forme de Pfaff. Ann. Inst. Fourier **26**, (fasc. 2) 171–220 (1976)
14. Moussu, R.: Existence d'intégrales premières pour un germe de forme de Pfaff C^∞-non dégénérée. Bol. Soc. Bras. Mat. **72**, 111–120 (1976)
15. Reeb, G.: Sur certaines propriétés topologiques des variétés feuilletées. Thèse. Hermann 1952
16. Roche, C.A.: Cohomologie relative dans le domaine réel. Thèse 3ème cycle. Grenoble (1982)
17. Roche, C.A.: Real relative cohomology of finite codimension germs.. Bull Am. Math. Soc. **7**, (n° 3) 596–598 (1982)
18. Roussarie, R.: Modèles locaux de champs et de formes. Astérisque N° 30 (1975)
19. Sternberg, S.: Local contractions and a theorem of Poincaré. Am. J. of Math. **79**, 809–824 (1957)
20. Tougeron, J.C.: Idéaux de fonction différentiables. Erg. Math. Berlin-Heidelberg-New York: Springer 1972
21. Tougeron, J.C.: Sur le complexe de Koszul (preprint)

Oblatum 15-III-1983

Invent. math. 73, 437–443 (1983)

Inventiones mathematicae
© Springer-Verlag 1983

An Obstruction to the Existence of Einstein Kähler Metrics

A. Futaki*

Department of Mathematics, College of Arts and Sciences, Chiba University, Yayoicho, Chiba, 260 Japan

Introduction

The purpose of this paper is to present an obstruction to the existence of Einstein Kähler metrics on compact complex manifolds of positive first Chern class.

Let M be a compact complex manifold of positive first Chern class. We put $c_1^+(M)$ to be the set consisting of all positive $(1,1)$-forms representing $c_1(M)$. Let $\omega=(i/2\pi)\sum_{\alpha,\beta} g_{\alpha\bar{\beta}}\,dz^\alpha\wedge d\bar{z}^\beta$ be an element of $c_1^+(M)$. Then $\sum_{\alpha,\beta} g_{\alpha\bar{\beta}}\,dz^\alpha d\bar{z}^\beta$ defines a Kähler metric on M and ω is its Kähler form. The Ricci form

$$\gamma_\omega=-(i/2\pi)\,\partial\bar{\partial}\log\det(g_{\alpha\bar{\beta}})$$

also represents $c_1(M)$. Therefore we can find a real valued smooth function F such that

$$\gamma_\omega-\omega=(i/2\pi)\,\partial\bar{\partial}F.$$

By definition ω is *Einstein* if and only if F is a constant function. Let $\mathfrak{h}(M)$ be the complex Lie algebra of all holomorphic vector fields on M. We define a linear function f of $\mathfrak{h}(M)$ into $\mathbb{C}$ by

$$f(X)=\int_M XF\omega^m,$$

where $X\in\mathfrak{h}(M)$ and $m=\dim M$. Our main result is stated as follows.

Theorem. *The function f does not depend on the choice of $\omega\in c_1^+(M)$. Therefore the number $\delta_M=\dim\mathfrak{h}(M)/\mathfrak{h}_0(M)$ depends only on the complex structure of M where $\mathfrak{h}_0(M)$ is the kernel of f. If M admits an Einstein Kähler metric then $\delta_M=0$.*

* Partially supported by the Grant-in-Aid for Scientific Research A-57740012 from the Ministry of Education

Up to now there have been several examples of compact complex manifolds of $c_1(M)>0$ which do not admit any Einstein Kähler metrics ([18, 9, 16, 17, 7]). In these examples the Lie algebras of holomorphic vector fields are not reductive. Then the nonexistence of Einstein Kähler metrics follows from Matsushima's theorem [15] which says that if M is a compact Einstein Kähler manifold then $\mathfrak{h}(M)$ is reductive (see also [13] in which Lichnerowicz generalized it to the effect that the same conclusion holds for a compact Kähler manifold of constant scalar curvature). Our theorem stated above seems dissimilar to Matsushima's theorem. In fact we can find a compact complex manifold of $c_1(M)>0$ such that $\mathfrak{h}(M)$ is reductive and that $\delta_M=1$ (Proposition 3.1).

In Sect. 2 we shall show that f is invariant under the group of all holomorphic transformations of M. It follows that the derived algebra of $\mathfrak{h}(M)$ is contained in $\mathfrak{h}_0(M)$. If $\mathfrak{h}(M)$ is semisimple then $\delta_M=0$. If $\delta_M=1$ then $\mathfrak{h}(M)$ contains a hyperplane invariant under the group of all holomorphic transformations of M.

In Sect. 5 we shall remark that our result is an answer to the question of Kazdan [10]. In their work [11] Kazdan and Warner found an integrability condition for the existence of a solution to the differential equation $2-\Delta u=he^u$ on S^2 which can be considered as a 1-dimensional complex Monge-Ampère equation. Our Theorem is a complex version of their result.

§ 1. Proof of the Main Theorem

The space $c_1^+(M)$ of allowable Kähler forms is contractible (e.g., by linear interpolation). Let t be a real variable in [0, 1], and ω_t a differentiable family of Kähler forms parametrized by t, f_t the corresponding family of linear functions of $\mathfrak{h}(M)$ into $\mathbb{C}$. Since the forms ω_t belong to a fixed cohomology class, there exists a corresponding family of real valued functions ϕ_t on M, unique up to any family of M-constant functions of t, such that

$$\frac{\partial\omega}{\partial t}=\frac{i}{2\pi}\partial\bar{\partial}\phi_t.$$

From this, one obtains in succession

$$\frac{\partial(\omega^m)}{\partial t}=m\frac{\partial\omega}{\partial t}\wedge\omega^{(m-1)}=\Delta\phi\,\omega^m;$$

$$\frac{\partial\gamma_\omega}{\partial t}=-\frac{i}{2\pi}\partial\bar{\partial}\Delta\phi_t;\qquad\frac{\partial}{\partial t}(\partial\bar{\partial}F_t)=-\partial\bar{\partial}(\Delta\phi+\phi)$$

whence we may choose F_t so that

$$\frac{\partial F}{\partial t}=-\Delta\phi-\phi.$$

Given $X \in \mathfrak{h}(M)$, consider the function $f_t(X)$ of t: the theorem is proved by showing that the derivative of this function vanishes identically. In fact

$$\frac{df_t(X)}{dt} = \int_M \left(X\left(\frac{\partial F}{\partial t}\right) + (XF)\,\Delta\phi \right) \omega^m. \tag{1}$$

We develop the integrand as follows

$$\begin{aligned} X\left(\frac{\partial F}{\partial t}\right) + (XF)\,\Delta\phi &= X(-\Delta\phi - \phi) + (XF)\,\Delta\phi \\ &= X^\alpha(-g^{\lambda\bar{\mu}}\phi_{\lambda\bar{\mu}\alpha} - \phi_\alpha + F_\alpha g^{\lambda\bar{\mu}}\phi_{\lambda\bar{\mu}}) \\ &= X^\alpha(-g^{\lambda\bar{\mu}}(\phi_{\lambda\alpha\bar{\mu}} - \phi_\beta R^\beta_{\lambda\alpha\bar{\mu}} - F_\alpha \phi_{\lambda\bar{\mu}}) - \phi_\alpha) \\ &= g^{\lambda\bar{\mu}} X^\alpha(-\phi_{\lambda\alpha\bar{\mu}} + \phi_\lambda S_{\alpha\bar{\mu}} + \phi_{\lambda\bar{\mu}} F_\alpha - \phi_\lambda g_{\alpha\bar{\mu}}) \\ &= g^{\lambda\bar{\mu}}(-X^\alpha \phi_{\lambda\alpha} + X^\alpha \phi_\lambda F_\alpha)_{\bar{\mu}} \end{aligned}$$

where R is the full curvature tensor and S is the Ricci tensor. Since the integrand in (1) can be thus expressed in a divergence form, the integral vanishes identically, proving the theorem.

§ 2. Properties of δ_M

We denote by $\mathrm{Aut}(M)$ the group of all holomorphic transformations of M. In this section we shall prove

Theorem 2.1. *The function f is invariant under* $\mathrm{Aut}(M)$.

Proof. Let a be any element of $\mathrm{Aut}(M)$ and h_t a one parameter subgroup of $\mathrm{Aut}(M)$. Then one finds

$$\begin{aligned} \int_M ((a \circ h_t \circ a^{-1})^* F)\,\omega^m &= \int_M F((a \circ h_t \circ a^{-1})^{-1})^* \omega^m \\ &= \int_M F((a \circ h_t)^{-1})^* (a^* \omega)^m \\ &= \int_M ((a \circ h_t)^* F)\,(a^* \omega)^m \\ &= \int_M (h_t^*(a^*(F))\,(a^* \omega)^m. \end{aligned} \tag{2}$$

If h_t is generated by $X \in \mathfrak{h}(M)$, differentiating (2) with respect to t one obtains

$$\int_M ((a d_a X) F)\,\omega^m = \int_M (X(a^* F))\,(a^* \omega)^m.$$

Since $\gamma_{a^*\omega} - a^*\omega = (i/2\pi)\,\partial\bar{\partial}(a^* F)$ and f is independent of the choice of a Kähler form, we have

$$f(a\,d_a X)) = f(X).$$

This proves the theorem.

Corollary 2.2. *The derived algebra of* $\mathfrak{h}(M)$ *is contained in* $\mathfrak{h}_0(M)$. *In particular* f *is a Lie algebra homomorphism.*

Corollary 2.3. *If* $\mathfrak{h}(M)$ *is semisimple then* $\delta_M=0$. *If* $\delta_M=1$ *then* $\mathfrak{h}(M)$ *contains a hyperplane invariant under* $\mathrm{Aut}(M)$.

§ 3. An Example

To begin with we recall about the hyperplane bundle H_m over the complex projective space $\mathbb{P}^m(\mathbb{C})$ of dimension m. There are three equivalent definitions. The first one is to be the dual of the universal (natural or tautological) bundle. So $c_1(H_m)$ is the positive generator of $H^2(\mathbb{P}^m(\mathbb{C});\mathbb{Z})$. The second one is to be the associated line bundle of the divisor defined by a hyperplane in $\mathbb{P}^m(\mathbb{C})$ (see [6]). Let $[z_0:z_1:\ldots:z_m]$ be the homogeneous coordinates of $\mathbb{P}^m(\mathbb{C})$ and $s(\alpha)_i=z_i/z_\alpha$, $0\leqq i\leqq m$, $i\neq\alpha$, be the inhomogeneous coordinates on $U_\alpha=\{z_\alpha\neq 0\}$. We choose defining functions $\{f_\alpha\}$ of the divisor $\{z_0=0\}$ as $f_\alpha=s(\alpha)_0$ for $\alpha\neq 0$ and $f_0=1$. Then the transition functions $\{g_{\alpha\beta}\}$ of H_m are given by $g_{\alpha\beta}=f_\alpha/f_\beta=z_\alpha/z_\beta$ on $U_\alpha\cap U_\beta$. We put, on U_α, $h_\alpha=(1+\sum_{i\neq\alpha}|s(\alpha)_i|^2)^{-1}$. Then $h=\{h_\alpha\}$ defines a fiber metric of H_m. Note that $(m+1)$ times the Chern form of H_m with respect to h defines an Einstein Kähler form of $\mathbb{P}^m(\mathbb{C})$, corresponding to the Fubini-Study metric. The third definition is to be a homogeneous line bundle (see [3]). Let $GL(1,m,\mathbb{C})$ be the subgroup of the general linear group $GL(m+1,\mathbb{C})$ consisting of all matrices whose $(k,1)$-elements are 0 except $(1,1)$-element. There is a holomorphic representation ρ of $GL(1,m,\mathbb{C})$ into $\mathbb{C}^*$ given by $\rho(A)=a_{11}^{-1}$ where $A\in GL(1,m,\mathbb{C})$ and a_{11} is its $(1,1)$-element. Then $\mathbb{P}^m(\mathbb{C})$ is the homogeneous space $GL(m+1,\mathbb{C})/GL(1,m,\mathbb{C})$ and H_m is defined to be the homogeneous line bundle $GL(m+1,\mathbb{C})\times_\rho\mathbb{C}$. Note that $GL(m+1,\mathbb{C})$ acts on H_m and the center of $GL(m+1,\mathbb{C})$ acts on H_m as scalar multiplications of the fiber.

Now consider the vector bundle $E=\pi_1^*H_1\oplus\pi_2^*H_2$ over $\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C})$ where π_i is the projection of $\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C})$ onto $\mathbb{P}^i(\mathbb{C})$. We take M to be the total space of the projective bundle $\mathbb{P}(E)$.

Proposition 3.1. The first Chern class of M is positive and $\mathfrak{h}(M)$ is reductive but $\delta_M=1$.

Proof. From the third definition of the hyperplane bundle, we see that $P(GL(2,\mathbb{C})\times GL(3,\mathbb{C}))$, the factor group of $GL(2,\mathbb{C})\times GL(3,\mathbb{C})$ by the center of $GL(5,\mathbb{C})$, acts on M effectively. This action induces the usual action of $PGL(2,\mathbb{C})\times PGL(3,\mathbb{C})$ on $\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C})$. On the other hand there is a homomorphism Π of $\mathfrak{h}(M)$ into $\mathfrak{h}(\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C}))$ (see [2]). From the observation above Π is surjective. By the arguments of [9] the kernel of Π is isomorphic to $H^0(\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C}),\mathrm{End}(E))/\mathbb{C}\cong\mathbb{C}$ where $\mathrm{End}(E)$ is the endomorphism bundle of E. Hence $\mathfrak{h}(M)$ is isomorphic to the Lie algebra of $P(GL(2,\mathbb{C})\times GL(3,\mathbb{C}))$, which is reductive.

Next we see $c_1(M)>0$ (see [18, 8] for the technique used here). Let $p: E-\{0\}\to\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C})$ be the projection where $\{0\}$ denotes the zero section.

We choose a Hermitian metric of E. For $\varepsilon>0$ we define a form $\tilde{\omega}_\varepsilon$ on $E-\{0\}$ by

$$\tilde{\omega}_\varepsilon = p^*\eta + \varepsilon i\partial\bar{\partial}\log\|v\|^2$$

where η is a Kähler form on $\mathbb{P}^1(\mathbb{C})\times\mathbb{P}^2(\mathbb{C})$, $v\in E-\{0\}$. This is positive if ε is small enough and projects to a closed form ω_ε on M. Let $s=z_1/z_0$ (resp. $t_1=w_1/w_0$, $t_2=w_2/w_0$) be the inhomogeneous coordinate on $U_0=\{z_0\neq 0\}$ (resp. $V_0=\{w_0\neq 0\}$). The restriction of $M=\mathbb{P}(E)$ to $U_0\times V_0$ is isomorphic to the product bundle $U_0\times V_0\times\mathbb{P}^1(\mathbb{C})$. Let $\xi=\xi_1/\xi_0$ be the inhomogeneous coordinate of the fiber. Then we may put

$$\omega_\varepsilon = \frac{i}{2\pi}\partial\bar{\partial}\log(1+|s|^2)^2(1+|t|^2)^3 + \varepsilon i\partial\bar{\partial}\log\left(\frac{1}{1+|s|^2}+\frac{|\xi|^2}{1+|t|^2}\right),$$

where $|t|^2=|t_1|^2+|t_2|^2$. Then

$$\omega_\varepsilon^4 = \left(\frac{3}{\pi}\right)^3 \varepsilon i^4 \frac{ds\wedge d\bar{s}\wedge dt_1\wedge d\bar{t}_1\wedge dt_2\wedge d\bar{t}_2\wedge d\xi\wedge d\bar{\xi}}{(1+|s|^2)(1+|t|^2)^2\,T^2} + o(\varepsilon)$$

where $T=1+|t|^2+(1+|s|^2)|\xi|^2$. Though ω_ε does not necessarily belong to $c_1(M)$, the Ricci form of ω_ε, denoted by $\gamma_{\omega_\varepsilon}$, belongs to $c_1(M)$ for each ε. Putting

$$\begin{aligned}\omega &= \lim_{\varepsilon\to 0}\gamma_{\omega_\varepsilon}\\ &= \frac{i}{2\pi}\partial\bar{\partial}\log(1+|s|^2)(1+|t|^2)^2\,T^2\end{aligned}$$

we obtain a positive form in $c_1(M)$, proving the claim.

Finally we see $\delta_M=1$. A straightforward computation shows

$$\omega^4 = \left(\frac{i}{\pi}\right)^4 \frac{8T^3+2T(1+|s|^2)(1+|t|^2)|\xi|^2+(1+|s|^2)(1+|t|^2)^2|\xi|^2}{T^5(1+|s|^2)(1+|t|^2)^2} \cdot ds\wedge d\bar{s}\wedge dt_1\wedge d\bar{t}_1\wedge dt_2\wedge d\bar{t}_2\wedge d\xi\wedge d\bar{\xi}$$

and

$$F = -\log\left(\frac{1}{T}\right)^3 [8T^3+2T(1+|s|^2)(1+|t|^2)|\xi|^2 + (1+|s|^2)(1+|t|^2)^2|\xi|^2].$$

As is seen easily $\xi\dfrac{\partial}{\partial\xi}$ defines a global vector field on M. One obtains

$$\xi\frac{\partial F}{\partial\xi}\omega^4 = \frac{2T(1+|s|^2)|\xi|^4-3(1+|t|^2)^2|\xi|^2}{T^6(1+|t|^2)} \times\left(\frac{i}{\pi}\right)^4 ds\wedge d\bar{s}\wedge dt_1\wedge d\bar{t}_1\wedge dt_2\wedge d\bar{t}_2\wedge d\xi\wedge d\bar{\xi},$$

$$f\left(\xi\frac{\partial}{\partial\xi}\right)=C\int_0^\infty\int_0^\infty\int_0^\infty\frac{\xi^2[2(1+t^2+(1+s^2)\xi^2)(1+s^2)\xi^2-3(1+t^2)^2]}{(1+t^2+(1+s^2)\xi^2)^6(1+t^2)}st^3\xi\,ds\,dt\,d\xi$$

$$=\frac{C}{120}\int_0^\infty\int_0^\infty\frac{st^3\,ds\,dt}{(1+s^2)(1+t^2)^3}>0$$

where $C=(16/\pi)^4\,\mathrm{vol}(S^3)\,\mathrm{vol}(S^1)^2$. This completes the proof.

§4. Conjecture

We consider a functional $\mathscr{S}c$ of $c_1^+(M)$ into the real numbers $\mathbb{R}$ defined by

$$\mathscr{S}c(\omega)=\int_M|\sigma(\omega)|^2\,\omega^m$$

where $\sigma(\omega)=g^{\alpha\bar\beta}S_{\alpha\bar\beta}$ is the scalar curvature of $\omega=(i/2\pi)g_{\alpha\bar\beta}\,dz^\alpha\wedge d\bar z^\beta$, S being the Ricci tensor. We can see that $\omega\in c_1^+(M)$ is a critical point of $\mathscr{S}c$ if and only if $V''V''\sigma(\omega)=0$ where V'' denotes the $(0,1)$-part of the covariant differentiation with respect to ω, or equivalently if $g^{\alpha\bar\beta}\sigma_{\bar\beta}\partial/\partial z^\alpha$ is a holomorphic vector field on M (see [4] for the proof). Assume that one can find a critical point $\omega\in c_1^+(M)$. Noting that $\gamma_\omega-\omega=(i/2\pi)\partial\bar\partial F$ implies $\sigma=m+\Delta F$, one finds

$$f\left(\sigma^\alpha\frac{\partial}{\partial z^\alpha}\right)=\int_M(\Delta F)^\alpha F_\alpha\,\omega^m=-\int_M(\Delta F)^2\,\omega^m.$$

If $\delta_M=0$ then F is a constant function and ω is an Einstein Kähler form. So we conjecture that if $\delta_M=0$ then M will admit an Einstein Kähler metric.

§5. Remark

In this section we explain that our main theorem is a complex version of the result of Kazdan and Warner in Sect. 8 of [11].

Let (M,g_0) be a compact Einstein Kähler manifold such that the Kähler form ω_0 of g_0 represents $c_1(M)$. We consider the problem;

Given a real valued smooth function F, does there exists a Kähler metric g such that the Kähler form ω of g represents $c_1(M)$ and that $\gamma_\omega-\omega=(i/2\pi)\partial\bar\partial F$?

This problem is equivalent to solving the following nonlinear partial differential equation of Monge-Ampere type;

$$\det(g_{0\alpha\bar\beta}-\phi_{\alpha\bar\beta})/\det(g_{0\alpha\bar\beta})=\exp(\phi-F).\tag{3}$$

(M,g_0) being as above, let Λ be the vector space over $\mathbb{C}$ consisting of all complex valued smooth functions satisfying $\Delta u+u=0$. Then Λ is isomorphic to $\mathfrak{h}(M)$ through the correspondence $u\mapsto g_0^{\alpha\bar\beta}u_{\bar\beta}\partial/\partial z^\alpha$ (see [14]).

Proposition 5.1. Notations being as above if $F\in\Lambda-\{0\}$ then the differential equation (3) has no solution.

Proof. Since M admits an Einstein metric $f(X)=0$ for all $X\in\mathfrak{h}(M)$. Suppose ϕ was a solution of (3). If we put

$$\omega=\frac{i}{2\pi}(g_{0\alpha\bar{\beta}}-\phi_{\alpha\bar{\beta}})\,dz^{\alpha}\wedge d\bar{z}^{\beta}$$

then

$$f(X)=\int_M g_0^{\alpha\bar{\beta}}u_{\bar{\beta}}F_{\alpha}e^{-F+\phi}\omega_0^m \tag{4}$$

where $u\in\Lambda$, $X=g_0^{\alpha\bar{\beta}}u_{\bar{\beta}}\partial/\partial z^{\alpha}$. If $u=F$ then $f(X)\neq 0$ and this is a contradiction, and the proposition is proved.

If we put $K=e^{-F}$, the right hand side of (4) is equal to $-\int_M u^{\alpha}K_{\alpha}e^{\phi}\omega_0^m$ which is the left hand side of (8.10) in [11] multiplied by -1.

Acknowledgement. The author is grateful to the referee for leading to numerous improvements of the paper. In addition the referee found that, for a manifold M obtained from $\mathbb{P}^3(\mathbb{C})$ by blowing up a quadratic curve in a hyperplane and a point not lying on the hyperplane, $c_1(M)$ is positive and $\mathfrak{h}(M)$ is reductive but $\delta_M=1$. Since there is no such compact Kähler surface, the referee's example has the lowest dimension among all manifolds with those properties.

References

1. Aubin, T.: Equations du type Monge-Ampère sur les variétés kählériennes compactes. C.R. Acad. Sci. Paris **283**, 119–121 (1976)
2. Blanchard, M.A.: Sur les variétés analytiques complexes. Ann. Sci. Ecole Norm. Sup. **73**, 157–202 (1956)
3. Bott, R.: Homogeneous vector bundles. Ann. of Math. **66**, 203–248 (1957)
4. Bourguignon, J.P.: Sur la deuxième conjecture de Calabi, Première classe de Chern et courbure de Ricci: preuve de la conjecture de Calabi. Soc. Math. de France, Astérisque **58**, 135–147 (1978)
5. Calabi, E.: Extremal Kähler metrics, Seminar on differential geometry. Ann. of Math. Studies, No. 102, pp. 259–290. Princeton Univ. Press, 1982
6. Griffiths, P., Harris, J.: Principles of algebraic geometry. New York: Wiley-Interscience 1978
7. Hano, J-I.: Examples of projective manifolds not admitting Kähler metric with constant scalar curvature, preprint
8. Hitchin, N.: On the curvature of rational surfaces. Proc. Symp. Pure Math. **27**, 65–80 (1975)
9. Ishikawa, K., Sakane, Y.: On complex projective bundles over a Kähler C-space. Osaka J. Math. **16**, 121–132 (1979)
10. Kazdan, J.L.: Gaussian and scalar curvature, an update, Seminar on differential geometry. Ann. of Math. Studies, No. 102, pp. 185–191. Princeton Univ. Press 1982
11. Kazdan, J.L., Warner, F.W.: Curvature functions for compact 2-manifolds. Ann. of Math. **99**, 14–47 (1974)
12. Kobayashi, S.: Transformation groups in differential geometry. Berlin-Heidelberg-New York: Springer 1972
13. Lichnerowicz, A.: Sur les transformations analytiques des variétés kählériennes. C.R. Acad. Sci. Paris **244**, 3011–3014 (1957
14. Lichnerowicz, A.: Géometrie des groupes de transformations. Paris: Dunod 1958
15. Matsushima, Y.: Sur la structure du groupe d'homéomorphismes analytiques d'une certaine variété kaehlérienne. Nagoya Math. J. **11**, 145–150 (1957)
16. Sakane, Y.: On compact Einstein Kähler manifolds with abundant holomorphic transformations, Manifolds and Lie groups, papers in honor of Matsushima. Progress in Math., pp. 337–358. Boston: Birkhäuser 1981
17. Sakane, Y.: On nonsingular hyperplane sections of a Hermitian symmetric space, preprint
18. Yau, S.-T.: On the curvature of compact Hermitian manifolds. Invent. Math. **25**, 213–239 (1974)
19. Yau, S.-T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I. Comm. Pure Appl. Math. **31**, 339–411 (1978)

Oblatum 10-I & 4-IV-1983

Invent. math. 73, 445-465 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

On the Poles of a Local Zeta Function for Curves

D. Meuser

Department of Mathematics, Boston University, Boston, MA 02215, USA

In this paper we determine the poles of a zeta function associated to a curve defined over a nonarchimedean local field K of characteristic zero.

Let $|\ |_K$ denote the usual absolute value on K and $|dx\,dy|_K$ the usual Haar measure on K^2. Let $f(x,y)$ denote a polynomial in two variables. Let $f(x,y)=0$ define a curve whose only singularity is at $(0,0)$. We also assume that this curve is absolutely analytically irreducible, i.e., irreducible in $\bar{K}[[x,y]]$, the ring of formal power series in two variables with coefficients in the algebraic closure of K.

The zeta function is defined by

$$Z_\Phi(s)=\int_{K^2}\Phi(x,y)\,|f(x,y)|_K^s\,|dx\,dy|_K,$$

where $\Phi(x,y)$ is a Schwartz-Bruhat function on K^2. This is a holomorphic function for $\mathrm{Re}(s)>0$. It has a meromorphic continuation to the entire complex plane with possible poles occurring among the values $s=-n_i/N_i$ where the (N_i,n_i) are the "numerical data" given by Hironaka's theorem on the resolution of singularities. This is shown in [4] where it is also shown that the above function is rational in q^{-s}. The possible poles of a similar function in the archimedean case were given by Bernstein-Gel'fand in [1]. We show that among the above possible poles many do not actually occur and that exactly one simple pole comes from each "characteristic exponent" of $f(x,y)$. Characteristic exponents are discussed by Zariski in [6]. These characteristic exponents were first considered in connection with the above zeta function by Igusa in [3], where the pole closest to the origin was determined.

I wish to thank Leon Strauss for referring me to his thesis [5], where he began the investigation of this question.

1. Characteristic Exponents and the Resolution Process

By interchanging x and y if necessary we shall assume that the tangent cone to $f=0$ is not $x=0$. If necessary, we may also replace y by $y+dx$ for some d in order to assume that $y=0$ is the tangent cone to $f=0$. One can then expand y

in a Puiseux series, i.e. a power series in $x^{1/v}$ where v is the multiplicity of the tangent $y=0$, and the coefficients of the series are in K. We rewrite this series as

$$y(x)=\sum_{i=1}^{k_0} a_{0,i}x^i+\sum_{i=0}^{k_1} a_{1,i}x^{(u_1+i)/v_1}+\ldots+\sum_{i=0}^{\infty} a_{g,i}x^{(u_g+i)/v_1 v_2\ldots v_g}, \tag{1}$$

where $v=v_1 v_2\ldots v_g$, the exponents are strictly increasing, μ_i and v_i are relatively prime for $1\leqq i\leqq g$, and $v_1, v_2, \ldots, v_g\geqq 2$. It is possible for any of the $a_{j,i}$ to be zero if $i\geqq 1$, but $a_{j,0}\neq 0$ for $1\leqq j\leqq g$. Expansions of this type are discussed in [6]. Based upon our assumptions on the tangent cone we also have $\mu_1/v_1>1$. The g exponents

$$\mu_1/v_1, \mu_2/v_1 v_2, \ldots, \mu_g/v_1 v_2\ldots v_g$$

are called the "characteristic exponents" of the curve.

We recall the main ideas of quadratic transformations. Let $X=K^2$. The quadratic transform Y and the quadratic transformation h with center $(0,0)$ is defined as

$$\begin{array}{l} X\times Z\supset Y=\{(x_0,y_0,z_1,z_2)\mid x_0 z_2-y_0 z_1=0\} \\ \quad\quad\quad\

where $N_I \geqq 1$ for all I and C' is the final strict transform. The preimage of any 2-form without zeroes or poles is given by

$$\sum_I (n_I - 1) E_I$$

where $n_I \geqq 2$ for every I.

The ordered pairs (N_I, n_I) will be referred to as the numerical data for the I^{th} exceptional curve. We shall review the algebraic notation we shall be using, as developed in [3], for expressing the numerical data in terms of the above continued fraction expansions.

Let $p_n = p_n(k_0, k_1, \ldots, k_n)$ denote a polynomial in $n+1$ variables $k_0, k_1, \ldots, k_n$ with integer coefficients defined inductively as follows: it represents 0, 1, respectively, for $n = -2, -1$ and

$$p_n = k_0 p_{n-1}(k_1, \ldots, k_n) + p_{n-2}(k_2, \ldots, k_n)$$

for every $n \geqq 0$. We shall drop the n from p_n, as this is clear from the way it is written. Fix a positive integer t and consider the $t+1$ variables $k_0, \ldots, k_t$.

We define

$$a(s) = \begin{cases} p(k_0, \ldots, k_s)\, p(k_1, \ldots, k_t) & s \text{ even} \\ p(k_1, \ldots, k_s)\, p(k_0, \ldots, k_t) & s \text{ odd.} \end{cases}$$

For $s = -2, -1$ we put $a(-2) = a(-1) = 0$. Define $b(s)$ for $-2 \leqq s \leqq t$ as

$$b(s) = a(s) + p(k_{s+2}, \ldots, k_t).$$

Define $c(s)$ for $s \geqq 0$ as

$$c(s) = p(k_0, \ldots, k_s) + p(k_1, \ldots, k_s)$$

and put $c(-2) = c(-1) = 1$.

For $0 \leqq s \leqq t$, $a(s)$, $b(s)$, and $c(s)$ satisfy the following recursion formulas

$$\begin{aligned} a(s) &= k_s b(s-1) + a(s-2) \\ b(s) &= k_s a(s-1) + b(s-2) \\ c(s) &= k_s c(s-1) + c(s-2). \end{aligned} \tag{4}$$

Let $P(-2) = 1$, $P(-1) = 0$, $P(0) = 1$, and $P(s) = p(k_1, \ldots, k_s)$ for $1 \leqq s \leqq t$. We then have a recursion formula for $P(s)$.

$$P(s) = k_s P(s-1) + P(s-2) \tag{5}$$

for $0 \leqq s \leqq t$.

We can now explicitly describe the resolution process. Let $a_1(s)$, $b_1(s)$, $c_1(s)$ and $P_1(s)$ denote the $a(s)$, $b(s)$, $c(s)$ and $P(s)$ for the sequence $k_{10}, k_{11}, \ldots, k_{1,t_1}$. We pass from (x, y) to (x', y') by

$$x' = x \qquad y' = y - \sum_{i=1}^{k_{10}} a_{0,i} x^i. \tag{6}$$

Let

$$I = k_{10} + k_{11} + \ldots + k_{1,s-1} + j$$

where $1 \leqq j \leqq k_{1,s}$, $0 \leqq s \leqq t$. Put $x = y_{-1}$, $y = x_{-1}$ and consider the sequence of transformations

$$x_{i-1} = x_i^{k_{1i}} y_i \qquad y_{i-1} = x_i$$

for $0 \leqq i < s$; then put

$$x_{s-1} = x_I^j y_I \qquad y_{s-1} = x_I.$$

Then after I quadratic transformations the transform of f is

$$f(x, y) = (x_I^{a(s-2)+jb(s-1)} y_I^{a(s-1)})^{\nu_2 \dots \nu_g} f_I(x_I, y_I)$$

and the preimage of $dx\,dy$ is

$$\pm x_I^{c(s-2)+jc(s-1)-1} y_I^{c(s-1)-1} dx_I\,dy_I$$

in the affine plane that still requires desingularization, where $x_I = 0$ is a local equation for the I^{th} exceptional curve. In other words we get the following expressions for the numerical data (N_I, n_I) which appears as Lemma 1 in [3]:

$$N_I = (a_1(s-2) + j\,b_1(s-1))\,\nu_2 \dots \nu_g$$
$$n_I = c_1(s-2) + j\,c_1(s-1)$$

for $I = k_{10} + k_{11} + \dots + k_{1,s-1} + j$, where $1 \leqq j \leqq k_{1,s}$, $0 \leqq s \leqq t$.

Let $I_1 = k_{10} + k_{11} + \dots + k_{1,t_1}$. When $I = I_1$ the singularity of the strict transform is no longer at the origin. The Puiseux series of the strict transform is

$$y_{I_1} = \sum_{i=0}^{k_{2,0}} \alpha_{0,i} x_{I_1}^i + \alpha_{1,0} x_{I_1}^{\mu_2/\nu_2 - \nu_1} + \dots$$

where $\alpha_{0,0} \neq 0$. We pass from (x_{I_1}, y_{I_1}) to $(\tilde{x}_{I_1}, \tilde{y}_{I_1})$ by

$$\tilde{x}_{I_1} = x_{I_1} \qquad \tilde{y}_{I_1} = y_{I_1} - \sum_{i=0}^{k_{2,0}} \alpha_{0,i} x_{I_1}^i.$$

We get

$$f(x, y) = \tilde{x}_{I_1}^{\mu_1 \nu_1 \dots \nu_g} f_{I_1}(\tilde{x}_{I_1}, \tilde{y}_{I_1})$$

and the preimage of the usual Haar measure $|dx\,dy|_K$ on K^2 is

$$\tilde{x}_{I_1}^{\mu_1 + \nu_1 - 1} |d\tilde{x}_{I_1}\,d\tilde{y}_{I_1}|.$$

We interchange the $\tilde{x}_{I_1}$ and $\tilde{y}_{I_1}$ coordinates and then apply the sequence of quadratic transformations corresponding to the continued fraction expansion of $\mu_2/\nu_2 - \mu_1$. By continuing this process we get the following expressions for the numerical data associated to the I^{th} exceptional curve. This is stated as Lemma 2 in [3]. If $I_i = k_{10} + k_{11} + \dots + k_{i,t_i}$, then if

$$I = I_{i-1} + k_{i,0} + k_{i,1} + \dots + k_{i,s-1} + j \tag{8}$$

where $1 \leqq j \leqq k_{i,s}$ and $0 \leqq s \leqq t_i$, we have

$$\begin{aligned} N_I &= (P_i(s-2) + j\,P_i(s-1))\,N_{I_{i-1}} + (a_i(s-2) + j\,b_i(s-1))\,\nu_{i+1} \dots \nu_g \\ n_I &= (P_i(s-2) + j\,P_i(s-1))(n_{I_{i-1}} - 1) + c_i(s-2) + j\,c_i(s-1), \end{aligned} \tag{9}$$

where $a_i(s)$, $b_i(s)$, $c_i(s)$ and $P_i(s)$ denote the $a(s)$, $b(s)$, $c(s)$ and $P(s)$ for the sequence k_{i0}, $k_{i1} \dots k_{i,t_i}$. This expression remains valid for $I < I_1$ by letting $I_0 = 0$, $N_0 = 0$, $n_0 = 1$.

We shall also refer to the following expressions for the numerical data of the exceptional curves $I_1, I_2, \dots, I_g$ in terms of the characteristic exponents, which is also given in [3]

$$\begin{aligned} n_{I_1} &= \mu_1 + \nu_1 \\ N_{I_1} &= \mu_1 \nu_1 \nu_2 \dots \nu_g \end{aligned} \tag{10}$$

and

$$\begin{aligned} n_{I_i} &= n_{I_{i-1}} \nu_i + \mu_i - \mu_{i-1} \nu_i \\ N_{I_i} &= N_{I_{i-1}} \nu_i + \mu_i \nu_i \nu_{i+1} \dots \nu_g - \mu_{i-1} \nu_i^2 \nu_{i+1} \dots \nu_g. \end{aligned} \tag{11}$$

2. Relations on the Numerical Data of Intersecting Exceptional Curves

Fix $1 \leqq i \leqq g$ and consider the I-th exceptional curve E_I for $I_{i-1} < I < I_i$. We wish to obtain information on the numerical data of intersecting exceptional curves. E_I will intersect one other exceptional curve E_{i_3} with $I < i_3$, or E_I will intersect two other exceptional curves E_{i_1} and E_{i_3}, with $i_1 < I < i_3$. We denote the numerical data corresponding to the exceptional curves E_{i_1}, E_I, E_{i_3}, by (N_1, n_1), (N_2, n_2), and (N_3, n_3) respectively. If E_{i_1} does not occur we assign trivial numerical data $n_1 = 1$ and $N_1 = 0$. The numerical data then satisfy the following algebraic expression:

Lemma 1. *For $I_{i-1} < I < I_i$, and referring to the above notation we have*

$$\frac{n_2}{N_2} = \frac{n_1 + n_3}{N_1 + N_3}.$$

Proof. We index our exceptional curves by s and j as in (8) and refer to the numerical data given in (9). We shall change notation slightly by denoting $k_{i,s}$ by k_s for $0 \leqq s \leqq t_i$, and $a(s)$, $b(s)$, $c(s)$, $P(s)$ for $a_i(s)$, $b_i(s)$, $c_i(s)$ and $P_i(s)$. In terms of the index notation one can describe the intersections of the exceptional curves. We consider each case separately.

(a) $0 \leqq s \leqq t$ and $1 < j < k_s$. Then $E_{s,j}$ intersects $E_{s,j-1}$ and $E_{s,j+1}$. The identity is immediate from the insertion of the expressions for the numerical data.

(b) $s = 0$, $j = 1$, and $k_0 = 1$. Then $E_{0,1}$ intersects $E_{2,1}$ and the exceptional curve $E_{I_{i-1}}$.

We can verify our formula by showing

$$\begin{aligned} c(-2) + c(-1) &= K(c(0) + c(1) + 1), \\ a(-2) + b(-1) &= K(a(0) + b(1)) \end{aligned}$$

and

$$P(-2) + P(-1) = K(P(0) + P(1) + 1)$$

for some constant K. From the definitions of $a(s)$, $b(s)$, $c(s)$ and $P(s)$ and the recursion formulas in (4) and (5) we determine the following values:

$$\begin{aligned}
c(0) &= k_0\, c(-1) + c(-2) = 2,\\
c(1) &= k_1\, c(0) + c(-1) = 2k_1 + 1,\\
b(-1) &= p(k_1, \ldots, k_t),\\
a(0) &= k_0\, b(-1) + a(-2) = k_0\, p(k_1, \ldots, k_t) = p(k_1, \ldots, k_t),\\
b(1) &= k_1\, a(0) + b(-1) = (k_1 + 1)\, p(k_1, \ldots, k_t),\\
P(1) &= k_1\, P(0) + P(-1) = k_1.
\end{aligned}$$

We then verify that the above equations are true, with $K = k_1 + 2$.

(c) $s = 0$, $j = 1$ and $k_0 > 1$. Then $E_{0,1}$ intersects $E_{0,2}$ and $E_{I_{i-1}}$. The identity is immediate from the insertion of the expressions for the numerical data.

(d) $s = 1$, $j = 1$ and $k_1 = 1$. Then $E_{1,1}$ intersects $E_{3,1}$.

We can verify our formula by showing

$$\begin{aligned}
c(-1) + c(0) &= K(c(1) + c(2) + 1),\\
a(-1) + b(0) &= K(a(1) + b(2))
\end{aligned}$$

and

$$P(-1) + P(0) = K(P(1) + P(2))$$

for some constant K. We have

$$\begin{aligned}
c(0) &= k_0 + 1,\\
c(1) &= k_0 + 2,\\
c(2) &= k_2\, c(1) + c(0) = k_2\, k_0 + 2k_2 + k_0 + 1,\\
a(1) &= k_1\, b(0) + a(-1) = b(0),\\
b(0) &= k_0\, a(-1) + b(-2) = b(-2) = a(-2) + p(k_0, \ldots, k_t) = p(k_0, \ldots, k_t),\\
b(2) &= k_2\, a(1) + b(0) = (k_2 + 1)\, p(k_0, \ldots, k_t),\\
P(1) &= 1,\\
P(2) &= k_2 + 1.
\end{aligned}$$

We then verify that the above equations are true, with $K = k_2 + 2$.

(e) $s = 1$, $j = 1$ and $k_1 > 1$. Then $E_{1,1}$ intersects $E_{1,2}$.

We can verify our formula by showing

$$\begin{aligned}
c(-1) + c(0) &= K(c(-1) + 2c(0) + 1),\\
a(-1) + b(0) &= K(a(-1) + 2b(0))
\end{aligned}$$

and

$$P(-1) + P(0) = K(P(-1) + 2P(0)) \quad \text{for some constant } K.$$

We then verify that these equations are true after inserting the expressions for $P(-1)$, $P(0)$, $c(-1)$, $c(0)$, $a(-1)$ and $b(0)$; with $K = 2$.

(f) $1<s<(t-1)$, $j=1$ and $k_s=1$. Then $E_{s,1}$ intersects $E_{s-2,k_{s-2}}$ and $E_{s+2,1}$.
We can verify our formula by showing that

$$c(s-2)+c(s-1)=K(c(s-4)+k_{s-2}c(s-3)+c(s)+c(s+1)),$$
$$a(s-2)+b(s-1)=K(a(s-4)+k_{s-2}b(s-3)+a(s)+b(s+1)),$$

and

$$P(s-2)+P(s-1)=K(P(s-4)+k_{s-2}P(s-3)+P(s)+P(s+1)),$$

for some constant K. Using the recursion formulas for $c(s-2)$, $a(s-2)$, and $P(s-2)$ on the right hand side it remains to show

$$c(s-2)+c(s-1)=K(c(s-2)+c(s)+c(s+1)),$$
$$a(s-2)+b(s-1)=K(a(s-2)+a(s)+b(s+1)),$$

and

$$P(s-2)+P(s-1)=K(P(s-2)+P(s)+P(s+1)).$$

After using the recursion formulas for $c(s)$, $c(s+1)$, $b(s+1)$, $a(s)$, $P(s)$ and $P(s+1)$, we conclude that the above equations are true, with $K=k_{s+1}+2$.

(g) $1<s<t$, $j=1$ and $k_s>1$. Then $E_{s,1}$ intersects $E_{s-2,k_{s-2}}$ and $E_{s,2}$.
We can verify our formula by showing

$$c(s-2)+c(s-1)=K(c(s-4)+k_{s-2}c(s-3)+c(s-2)+2c(s-1)),$$
$$a(s-2)+b(s-1)=K(a(s-4)+k_{s-2}b(s-3)+a(s-2)+2b(s-1))$$

and

$$P(s-2)+P(s-1)=K(P(s-4)+k_{s-2}P(s-3)+P(s-2)+2P(s-1))$$

for some constant K. Using the recursion formulas for $a(s-2)$, $c(s-2)$, and $P(s-2)$ on the right hand side we conclude that the above equations are true, with $K=2$.

(h) $s<(t-1)$, $j=k_s$, and $k_s>1$. Then E_{s,k_s} intersects E_{s,k_s-1} and $E_{s+2,1}$.
We can verify our formula by showing

$$c(s-2)+k_s c(s-1)=K(c(s-2)+(k_s-1)c(s-1)+c(s)+c(s+1)),$$
$$a(s-2)+k_s b(s-1)=K(a(s-2)+(k_s-1)b(s-1)+a(s)+b(s+1))$$

and

$$P(s-2)+k_s P(s-1)=K(P(s-2)+(k_s-1)P(s-1)+P(s)+P(s+1))$$

for some constant K. Using the recursion formulas for $c(s+1)$, $b(s+1)$, $c(s)$, $a(s)$, $P(s+1)$, and $P(s)$ on the right hand side we see that the above equations are true, with $K=k_{s+1}+2$.

(i) $s=t-1$, $j=k_{t-1}$, and $k_{t-1}>1$. Then $E_{t-1,k_{t-1}}$ intersects $E_{t-1,k_{t-1}-1}$ and E_{t,k_t}.
We can verify our formula by showing

$$c(t-3)+k_{t-1}c(t-2)=K(c(t-3)+(k_{t-1}-1)c(t-2)+c(t-2)+k_t c(t-1)),$$
$$a(t-3)+k_{t-1}b(t-2)=K(a(t-3)+(k_{t-1}-1)b(t-2)+a(t-2)+k_t b(t-1))$$

and

$$P(t-3)+k_{t-1}P(t-2)=K(P(t-3)+(k_{t-1}-1)P(t-2)+P(t-2)+k_tP(t-1))$$

for some constant K. Using the recursion formulas for $c(t-1)$, $a(t-1)$, and $P(t-1)$, the first and third equalities are immediate with $K=k_t+1$. So it remains to show

$$a(t-1)=(k_t+1)(a(t-1)-b(t-2)+a(t-2)+k_tb(t-1)).$$

Using $b(t-2)=a(t-2)+k_t$ and $b(t-1)=a(t-1)+1$ we see that this is true.

(j) $s=t-1$, $j=k_{t-1}$ and $k_{t-1}=1$. Then $E_{t-1,1}$ intersects $E_{t-3,k_{t-3}}$ and E_{t,k_t}. We can verify our formula by showing

$$c(t-3)+c(t-2)=K(c(t-5)+k_{t-3}c(t-4)+c(t-2)+k_tc(t-1)),$$
$$a(t-3)+b(t-2)=K(a(t-5)+k_{t-3}b(t-4)+a(t-2)+k_tb(t-1))$$

and

$$P(t-3)+P(t-2)=K(P(t-5)+k_{t-3}P(t-4)+P(t-2)+k_tP(t-1)),$$

for some constant K. Using the recursion formulas for $c(t-3)$, $a(t-3)$, and $P(t-3)$ on the right it remains to show

$$c(t-3)+c(t-2)=K(c(t-3)+c(t-2)+k_tc(t-1)),$$
$$a(t-3)+b(t-2)=K(a(t-3)+a(t-2)+k_tb(t-1))$$

and

$$P(t-3)+P(t-2)=K(P(t-3)+P(t-2)+k_tP(t-1))$$

for some constant K. Using the recursion formulas for $c(t-1)$, $b(t-1)$ and $P(t-1)$ the first and third equalities are immediate with $K=k_t+1$. The second one follows from $a(t-2)=b(t-2)-k_t$ and $b(t-3)=a(t-3)+k_t$.

Recalling that if $i>1$ we can have $k_0=0$, we also have the cases where we consider the possible intersections of $E_{2,1}$. In particular if $k_2>1$ then $E_{2,1}$ intersects $E_{I_{i-1}}$ and $E_{2,2}$, and if $k_2=1$ we have that $E_{2,1}$ intersects $E_{I_{i-1}}$ and $E_{4,1}$. Each of these cases can be verified in a similar manner to cases (b) and (c) above. q.e.d.

Now consider an I_i-th exceptional curve for $1\leqq i\leqq g$, which we shall denote by E_{I_i}. If $i\neq g$, E_{I_i} will intersect three exceptional curves E_{i_1}, E_{i_2}, and E_{i_3} with $i_1<i_2<I_i<i_3$. If $i=g$, E_{I_g} will intersect two exceptional curves E_{i_1} and E_{i_2}, and the final strict transform. Denote the numerical data for the exceptional curve E_{i_k} by (N_k,n_k) for $k=1,2,3$.

We then have the following inequalities:

Lemma 2. *For* $1\leqq i\leqq g$; *and* $k=1,2,3$ *we have*

$$n_kN_{I_i}-n_{I_i}N_k\neq\pm N_{I_i}.$$

Proof. We again index our exceptional curves by s and j as in (8) and refer to the numerical data given in (9). We also use the same abbreviated notation as in Lemma 1. Then the E_{i_1}, E_{i_2} and E_{I_i} curves are denoted by $E_{t-1,k_{t-1}}$, E_{t,k_t-1}, and E_{t,k_t} respectively.

Consider the expression $n_1 N_{I_t} - n_{I_t} N_1$. By using the formulas in (9) along with the recursion formulas, we have

$$n_1 = P(t-1)(n_{I_{t-1}} - 1) + c(t-1),$$
$$N_1 = P(t-1) N_{I_{t-1}} + a(t-1) v,$$

where $v = v_{i+1} \ldots v_g$. Recalling that $a(t) = b(t)$ we have

$$n_{I_t} = P(t)(n_{I_{t-1}} - 1) + c(t),$$
$$N_{I_t} = P(t) N_{I_{t-1}} + b(t) v.$$

Then

$$n_1 N_{I_t} - n_{I_t} N_1 = \det \begin{bmatrix} P(t) N_{I_{t-1}} + b(t) v & P(t-1) N_{I_{t-1}} + a(t-1) v \\ P(t)(n_{I_{t-1}} - 1) + c(t) & P(t-1)(n_{I_{t-1}} - 1) + c(t-1) \end{bmatrix}.$$

Applying the formulas (4) and (5) to the first column we see that this is equal to

$$\det \begin{bmatrix} P(t-2) N_{I_{t-1}} + b(t-2) v & P(t-1) N_{I_{t-1}} + a(t-1) v \\ P(t-2)(n_{I_{t-1}} - 1) + c(t-2) & P(t-1)(n_{I_{t-1}} - 1) + c(t-1) \end{bmatrix}.$$

We can then apply the formulas in (4) and (5) to the second column, etc., until we get

$$\det \begin{bmatrix} P(-2) N_{I_{t-1}} + b(-2) v & P(-1) N_{I_{t-1}} + a(-1) v \\ P(-2)(n_{I_{t-1}} - 1) + c(-2) & P(-1)(n_{I_{t-1}} - 1) + c(-1) \end{bmatrix}$$

if t even, and

$$\det \begin{bmatrix} P(-1) N_{I_{t-1}} + b(-1) v & P(-2) N_{I_{t-1}} + a(-2) v \\ P(-1)(n_{I_{t-1}} - 1) + c(-1) & P(-2)(n_{I_{t-1}} - 1) + c(-2) \end{bmatrix}$$

if t odd.

We recall that $P(-2) = 1$, $P(-1) = 0$, $a(-1) = a(-2) = 0$, and $c(-1) = c(-2) = 1$. Thus we are reduced to showing

$$N_{I_{t-1}} + b(-2) v \neq \pm N_{I_t} \tag{12}$$

if t even, and

$$n_{I_{t-1}} b(-1) v - N_{I_{t-1}} \neq \pm N_{I_t} \tag{13}$$

if t odd.

Now consider the expression $n_2 N_{I_t} - n_{I_t} N_2$. We have

$$n_2 = (P(t) - P(t-1))(n_{I_{t-1}} - 1) + c(t) - c(t-1),$$
$$N_2 = (P(t) - P(t-1)) N_{I_{t-1}} + (a(t) - b(t-1)) v.$$

Using the expressions for n_{I_t} and N_{I_t} above we get that

$$n_2 N_{I_t} - n_{I_t} N_2 = -\det \begin{bmatrix} P(t) N_{I_{t-1}} + a(t) v & P(t-1) N_{I_{t-1}} + b(t-1) v \\ P(t)(n_{I_{t-1}} - 1) + c(t) & P(t-1)(n_{I_{t-1}} - 1) + c(t-1) \end{bmatrix}.$$

By an argument similar to the above we have that this is equal to

$$N_{I_{i-1}} - n_{I_{i-1}} b(-1) v$$

if t even, and

$$-N_{I_{i-1}} - b(-2) v$$

if t odd. Thus we must show the inequalities (12) and (13) given above.

Consider the inequality (12). We clearly cannot have $N_{I_{i-1}} + b(-2) v = -N_{I_i}$. So using the expression for N_{I_i} again we must show

$$(P(t)-1) N_{I_{i-1}} + (b(t) - b(-2)) v \neq 0.$$

But referring to the original definitions of $P(s)$ and $b(s)$ we have

$$P(t) = p(k_1, \ldots, k_t) \geqq 1$$

and

$$b(t) - b(-2) = p(k_0, \ldots, k_t)(p(k_1, \ldots, k_t) - 1) > 1$$

since $k_t \geqq 2$. This establishes (12).

Now consider (13). We shall show that for $i \geqq 2$, $n_{I_{i-1}} b(-1) v - N_{I_{i-1}} < 0$. We observe that $b_i(-1) = v_i$, as one can see by observing that the multiplicity of the strict transform at the I_{i-1} stage is $v_i v_{i+1} \ldots v_g$ and by considering the expression for $s=0$ in (9). Thus we are reduced to showing

$$n_{I_{i-1}} v_i v_{i+1} \ldots v_g < N_{I_{i-1}}.$$

We proceed by induction. We refer to the expressions of the numerical data in terms of the characteristic exponents given in (10) and (11). If $i=2$ we must show

$$(\mu_1 + v_1) v_2 v_3 \ldots v_g < \mu_1 v_1 v_2 \ldots v_g.$$

But $v_1 < \mu_1$ implies $\mu_1 + v_1 < 2\mu_1$ and $2\mu_1 \leqq \mu_1 v_1$ since $v_1 \geqq 2$. So assume our desired expression is true for i, and consider $n_{I_i} v_{i+1} v_{i+2} \ldots v_g$. By inserting the expression for n_{I_i} we have that

$$\begin{aligned} n_{I_i} v_{i+1} \ldots v_g &= n_{I_{i-1}} v_i v_{i+1} \ldots v_g + \mu_i v_{i+1} \ldots v_g - \mu_{i-1} v_i v_{i+1} \ldots v_g \\ &< N_{I_{i-1}} + \mu_i v_{i+1} \ldots v_g - \mu_{i-1} v_i v_{i+1} \ldots v_g \end{aligned}$$

by our inductive hypothesis. But by using the expression for N_{I_i} along with the fact that $v_i \geqq 2$, we see that this is less than N_{I_i} as desired. So $n_{I_{i-1}} b(-1) v - N_{I_{i-1}} \neq N_{I_i}$ for $i \geqq 2$. The $i=1$ case is true by considering the expression for N_1 in terms of the characteristic exponents. We clearly cannot have $n_{I_{i-1}} b(-1) v - N_{I_{i-1}} = -N_{I_i}$ as this would imply $n_{I_{i-1}} b(-1) v < 0$. Thus we have established (13) and the desired inequalities for $k=1,2$.

We now consider the expression $n_3 N_{I_i} - n_{I_i} N_3$. Let $k_{i+1,0}$ denote the first term in the continued fraction expansion of $\mu_{i+1}/v_{i+1} - \mu_i$. We first consider the case where $k_{i+1,0} \neq 0$. Then E_{I_i} intersects $E^*_{0,1}$ where $E^*_{0,1}$ denotes the $s=0$, $j=1$ exceptional curve for the above continued fraction expansion. Referring to the expressions for the numerical data we have

$$N_3 = N_{I_i} + v_{i+1} \dots v_g,$$
$$n_3 = n_{I_i} + 1.$$

Inserting these in our expression we conclude that $n_3 N_{I_i} - n_{I_i} N_3 \neq N_{I_i}$.

Now consider the case where $k_{i+1,0} = 0$. Then E_{I_i} intersects $E^*_{2,1}$ where $E^*_{2,1}$ denotes the $s=2$, $j=1$ exceptional curve for the continued fraction expansion above. We then have

$$N_3 = (P(0) + P(1)) N_{I_i} + (a(0) + b(1)) v_{i+2} \dots v_g,$$
$$n_3 = (P(0) + P(1))(n_{I_i} - 1) + (c(0) + c(1)).$$

We have $P(0) = 1$, $P(1) = k_{i+1,1}$, $a(0) = 0$, $b(1) = b(-1)$, $c(0) = 1$, $c(1) = k_{i+1,1}$. Inserting these in our expression we again conclude $n_3 N_{I_i} - n_{I_i} N_3 \neq N_{I_i}$.

We need not consider the case $n_3 N_{I_i} - n_{I_i} N_3 = -N_{I_i}$ since from results in [3] which we shall refer to again below we have $n_3/N_3 > n_{I_i}/N_{I_i}$. q.e.d.

In [3] Igusa shows that $n_I/N_I > n_{I_1}/N_{I_1}$ for $I < I_1$ and $n_I/N_I > n_{I_i}/N_{I_i}$ for $I > I_i$ and $1 \leqq i < g$. Using this fact we can obtain the following information on the ratios n_I/N_I for intersecting exceptional curves. Let $I_{i-1} < I < I_i$, and let E_I intersect E_{i_1}, and E_{i_3} with $i_1 < I < i_3$, where we allow for the possibility that E_{i_1} does not occur. As usual, let the numerical data for the exceptional curves E_{i_1}, E_I, and E_{i_3} be denoted by (N_1, n_1), (N_2, n_2), and (N_3, n_3) respectively. As usual if E_{i_1} does not occur we assign trivial numerical data $N_1 = 0$, $n_1 = 1$.

Lemma 3. *n_2/N_2 cannot equal n_1/N_1 or n_3/N_3.*

Proof. We first suppose $0 < I < I_1$. Suppose n_2/N_2 is equal to either n_1/N_1 or n_3/N_3. Then by Lemma 1 we would have $n_2/N_2 = n_3/N_3$. Now repeat the argument with the exceptional curve E_{i_3} if $i_3 < I_1$. By repeating this argument we eventually get $n_2/N_2 = n_{I_1}/N_{I_1}$, a contradiction to the fact that $n_2/N_2 > n_{I_1}/N_{I_1}$.

Now let $I_i \leqq I$ for $i \geqq 1$. If E_I intersects only E_{i_3} then Lemma 1 applies with $N_1 = 0$, $n_1 = 1$. Then clearly we cannot have $n_2/N_2 = n_3/N_3$. So suppose E_I intersects E_{i_1} and E_{i_3} and n_2/N_2 is equal to either n_1/N_1 or n_3/N_3. Then by Lemma 1 we would have n_2/N_2 equal to n_1/N_1. Now repeat the argument with the exceptional curve E_{i_1} if $i_1 > I_i$ and E_{i_1} intersects two exceptional curves. By repeating this we eventually get that $n_2/N_2 = n_{I_i}/N_{I_i}$ or that $n_2/N_2 = n_k/N_k$ where (N_k, n_k) is the numerical data for an exceptional curve E_k that only intersects one other exceptional curve E_l with $k < l$. The first case contradicts the fact that $n_2/N_2 > n_{I_i}/N_{I_i}$. In the second case our process gives $n_2/N_2 = n_k/N_k = n_l/N_l$. However, applying Lemma 1 to E_k gives $n_k/N_k = (1 + n_l)/N_l$, a contradiction. q.e.d.

3. On the Vanishing of Possible Poles of $Z(s)$

Let R denote the ring of integers of K, P the unique maximal ideal of R, and $q = \mathrm{card}(R/P)$. In the definition of the zeta function, since Φ is a locally constant function with compact support, and since the possible poles are determined by

the singularity at the origin, we shall assume that Φ is the characteristic function of $R \times R$. We shall also assume that f has coefficients in R.

We shall make some observations about the resolution process. We recall that by a combination of interchanging x and y, and replacing y by $y+dx$ for some d in R we may assume that $y=0$ is the tangent cone to $f=0$. We observe that these changes do not affect the domain of integration $R \times R$. We next observe the following properties of the Puiseux series for $f=0$.

Lemma 4. *In the Puiseux series expansion for* $f(x, y)$ *as given in* (1), *we have* $a_{0,i} \in R$ *for* $1 \leqq i \leqq k_0$.

Proof. Let $a_{0,d}$ be the first non-zero term in the Puiseux expansion. Then we can write

$$f(x, y) = c \prod_{j=0}^{\nu-1} (y - (a_{0,d} x^d + P^{(j)}(x^{1/\nu})))$$

where $P^{(j)}(x^{1/\nu})$ is the power series obtained from $y(x) - a_{0,d} x^d$ by replacing $x^{1/\nu}$ by $\zeta^j x^{1/\nu}$ where ζ is a ν-th root of unity. The coefficient of y^k is then c times

$$(-1)^{\nu-k} \sum_{0 \leqq j_1, \ldots, j_{(\nu-k)} \leqq \nu-1} \prod_{l=1}^{\nu-k} (a_{0,d} x^d + P^{(j_l)}(x^{1/\nu}))$$

and each term in the above summation has the form

$$(a_{0,d} x^d)^{\nu-k} + \sum_{i=1}^{\nu-k} (a_{0,d} x^d)^{\nu-k-i} \sum_{0 \leqq j_1, \ldots, j_i \leqq \nu-1} \prod_{l=1}^{i} P^{(j_l)}(x^{1/\nu}).$$

So we can write the coefficient of y^k as $c(a_k(x) + b_k(x^{1/\nu}))$ where

$$a_k(x) = (-1)^{\nu-k} \binom{\nu}{k} (a_{0,d} x^d)^{\nu-k}$$

and $b_k(x^{1/\nu})$ is the appropriate power series in $x^{1/\nu}$ combined from terms of the form

$$(a_{0,d} x^d)^{\nu-k-i} \prod_{l=1}^{i} P^{(j_l)}(x^{1/\nu})$$

for $i \geqq 1$.

As usual, for any power series in $x^{1/\nu}$ we define ord $a(x^{1/\nu})$ to be the lowest fractional power of x that occurs. We have that ord $a_k(x) = d(\nu-k)$. On the other hand since ord $P^{(j)}(x^{1/\nu}) > d$ for all j we have ord $b_k(x^{1/\nu}) > d(\nu-k)$. Thus

$$f(x, y) = c(y - a_{0,d} x^d)^\nu + Q(x, y)$$

for some polynomial $Q(x, y)$. Hence $a_{0,d} \in R$.

We now change coordinates by letting $\bar{y} = y - a_{0,d} x^d$. Then $f(x, \bar{y})$ has coefficients in R and the Puiseux series expansion for $f(x, \bar{y})$ will have $a_{0,d+i} x^{d+i}$ as its first non-zero term if $a_{0,d+i} x^{d+i}$ is the next non-zero term in the Puiseux series expansion for $f(x, y)$. Repeating the argument proves the lemma. q.e.d.

Thus Lemma 4 shows that the initial change of coordinates given in (6) does not affect the domain of integration $R \times R$. At the first quadratic transformation h of the resolution process we pull back the domain of integration $R \times R$ to the two affine planes with coordinates (x_1, y_1) and (x_1', y_1') as given in (2) and (3). Then we observe that $h^{-1}(R \times R)$ can be covered disjointly by $R \times R$ in one affine plane and $R \times P$ in the other affine plane. We shall take the $R \times P$ cover to occur in the affine plane which contains no singularity of $h^{-1}(f^{-1}(0))$. We then continue the process. After I_1 quadratic transformations the change of coordinates given in (7) is described by the Puiseux expansion of the strict transform of f at this stage, hence by Lemma 4 this does not affect our domain of integration $R \times R$. The process continues in a similar manner until our curve is desingularized.

If $Z_{i1}(s)$ and $Z_{i2}(s)$ denote the two pullbacks of the integral for $Z(s)$ at the i-th stage of the resolution process, where $Z_{i1}(s)$ is the integral over $R \times P$ and $Z_{i2}(s)$ is the integral over $R \times R$, we have that $Z(s)$ is equal to

$$Z_{1,1}(s) + Z_{2,1}(s) + \ldots + Z_{I_g,1}(s) + Z_{I_g,2}(s).$$

The form of $Z_{I,1}(s)$ is

$$\int_P \int_R |x|^{N_I s + n_I - 1} |y|^{N_J s + n_J - 1} |f_I(x,y)|^s |dx|\,|dy|$$

where (N_I, n_I) is the numerical data for the exceptional curve E_I; (N_J, n_J) denotes the numerical data for the exceptional curve E_J if E_J intersects E_I with $J < I$, and f_I is the strict transform at the I-th stage. Since $f_I(x,y) \neq 0$ in a neighborhood of the origin, and

$$\int_{Pe} |x|^{N_I s + n_I - 1} |dx| = \frac{(1 - q^{-1}) q^{-e(N_I s + n_I)}}{(1 - q^{-(N_I s + n_I)})}$$

we have that the possible poles of $Z(s)$ lie at the values $s = -n_I/N_I$ where (N_I, n_I) is the numerical data for an exceptional curve. There is also a possible pole at $s = -1$ coming from the strict transform of f.

We are now ready to prove the first theorem.

Theorem 1. *Let $I_{i-1} < I < I_i$. Suppose $n_I/N_I \neq 1$ and $n_I/N_I \neq n_{I_l}/N_{I_l}$ for any $1 \leqq l \leqq g$. Then $Z(s)$ does not have a pole at $s = -n_I/N_I$.*

Proof. In order to investigate this pole we examine the pullback of the integral for $Z(s)$ in a collection of disjoint neighborhoods that cover any exceptional curve that gives a possible contribution to this pole.

Denote the I-th exceptional curve by E_{i_2} and as before we have that E_{i_2} intersects the exceptional curves E_{i_1} and E_{i_3} with $i_1 < i_2 < i_3$. Denote the numerical data for E_{i_k} by (N_k, n_k) where we recall that $(N_1, n_1) = (0, 1)$ if E_{i_1} doesn't exist.

We have that $Z_{i_2,1}(s)$ and $Z_{i_3,1}(s)$ will contribute to the pole at $-n_2/N_2$. We have

$$Z_{i_2,1}(s) = \int_P \int_R |x_2|^{N_2 s + n_2 - 1} |y_2|^{N_1 s + n_1 - 1} |f_2(x_2, y_2)|^s |dx_2|\,|dy_2|$$

and

$$Z_{i_3,1}(s)=\int_P\int_R |x_3|^{N_3 s+n_3-1}|y_3|^{N_2 s+n_2-1}|f_3(x_3,y_3)|^s|dx_3||dy_3|$$

where f_2 and f_3 denote the strict transforms at the i_2-th and i_3-th stages respectively.

We now examine f_3 and f_2 more closely. At the (i_2-1)th stage of the resolution process we consider the pullback of f in the affine plane that still requires desingularization. By interchanging coordinates and replacing y_1 by y_1+dx_1 for some constant $d\in R$ if necessary, we may assume that $f(x,y)=x_1^{N_1}f_1(x_1,y_1)$ with

$$f_1(x_1,y_1)=c\,y_1^n+\sum_{i=1}^{n} y_1^{n-i}P_i(x_1)+\sum_{k=0}^{\infty} y_1^{n+k}R_k(x_1),$$

where $\operatorname{ord}_{x_1}P_i>i$ and $\operatorname{ord}_{x_1}R_0>0$.

If we now apply the quadratic transformation $x_1=x_2y_2$, $y_1=x_2$, we obtain

$$f(x,y)=x_2^{N_2}y_2^{N_1}f_2(x_2,y_2)$$

where

$$f_2(x_2,y_2)=c+\sum_{i=1}^{n} P_i'(x_2,y_2)+\sum_{k=0}^{\infty} x_2^k R_k(x_2y_2),$$

and $\operatorname{ord}_{x_2}P_i'>0$. Similarly, if at the (i_2-1)th stage we apply the quadratic transformation $x_1=x_2'$, $y_1=x_2'y_2'$, we obtain in the other affine plane

$$f(x,y)=x_2'^{N_2}f_2'(x_2',y_2')$$

where

$$f_2'(x_2',y_2')=\left(c\,y_2'^n+\sum_{i=1}^{n} y_2'^{n-i}P_i'(x_2')+\sum_{k=0}^{\infty} x_2'^k\,y_2'^{n+k}R_k(x_2')\right)$$

and $\operatorname{ord}_{x_2'}P_i'>0$.

To reach the i_3-th stage we have a sequence of transformations resulting in the coordinate change

$$x_2'=x_3^l\,y_3,$$
$$y_2'=x_3$$

for some $l\geqq 1$. I.e. at the i_3-th stage we have in the desingularized affine plane

$$f(x,y)=x_3^{N_3}y_3^{N_2}f_3(x_3,y_3)$$

where $f_3(x_3,y_3)$ is obtained by factoring out powers of x_3 from $f_2'(x_3^l\,y_3,x_3)$.

We have

$$f_2'(x_3^l\,y_3,x_3)=\left(c\,x_3^n+\sum_{i=1}^{n} x_3^{(n-i)}\,P_i'(x_3^l\,y_3)+\sum_{k=0}^{\infty} x_3^{lk+n+k}\,y_3^k\,R_k(x_3^l\,y_3)\right).$$

Since f_3 must have a non-zero constant term and we have that $\operatorname{ord}_{x_2'}P_i'>0$ and $\operatorname{ord}_{x_2'}R_0>0$, f_3 is obtained from this by factoring out x_3^n. Thus f_3 has the form $c+y_3\,g(x_3,y_3)$. Referring back to f_2 we see it has the form $c+x_2\,h(x_2,y_2)$. Thus if $\operatorname{ord}c=e$ we have that $|f_2|$ is constant on $P^{e+1}\times P$ and $|f_3|$ is constant on $R\times P^{e+1}$.

We now split the domain of integration for $Z_{i_2,1}(s)$ into $P^{e+1}\times P$ and $(R-P^{e+1})\times P$. We first consider the domain of integration $(R-P^{e+1})\times P$. We can write this as a disjoint union of U_i where each U_i is a coset of the form $U_i=(a_i+P^{e+1})\times P$. We have that $|x_2|$ is constant on each U_i. We consider separately the cases where $f_2\neq 0$ on U_i and $f_2=0$ on U_i.

If $f_2\neq 0$ on U_i we can write U_i as a disjoint union of cosets of $P^k\times P^k$ for k sufficiently large so that $|x_2|$ and $|f_2|$ are constant on each coset. Then the integrals over these cosets will contribute terms of the form

$$\frac{P(q^{-s})}{1-q^{-(N_1s+n_1)}} \tag{14}$$

for some polynomial P in q^{-s}.

Now consider the case where $f_2=0$ on U_i. Since f_2 is non-singular we must have $\partial f_2/\partial x_2(x_2,y_2)\neq 0$ or $\partial f_2/\partial y_2(x_2,y_2)\neq 0$ at every point (x_2,y_2) in U_i. Refine U_i into smaller cosets U_{ij} so that on each U_{ij} one of $|\partial f_2/\partial x_2|$ or $|\partial f_2/\partial y_2|$ is a non-zero constant. Suppose $|\partial f_2/\partial x_2|=q^{-k}$ on U_{ij}. Write U_{ij} as a disjoint union of cosets of $P^l\times P^l$ for $l>k$. Let D be one of these cosets. Then the change of coordinates given by $\tilde{x}_2=f_2(x_2,y_2)$, $\tilde{y}_2=y_2$ maps D homeomorphically to its image, cf. [4], Lemma 5. Then the integrals over $f(D)$ will contribute terms of the form

$$\frac{P(q^{-s})}{(1-q^{-(N_1s+n_1)})(1-q^{-(s+1)})} \tag{15}$$

for some polynomial $P(q^{-s})$. Since by our assumptions $n_2/N_2\neq 1$ and n_2/N_2 cannot equal n_1/N_1 by Lemma 2 we see that the terms in (14) and (15) do not contribute to the residue at the pole $s=-n_2/N_2$. By splitting the domain of integration for $Z_{i_3,1}$ into $R\times P^{e+1}$ and $R\times(P-P^{e+1})$ and applying a similar argument we see that the integral over $R\times(P-P^{e+1})$ does not contribute to the residue of the pole at $s=-n_2/N_2$.

Therefore the contributions to the residue come from the integrals

$$\int_P\int_{P^{e+1}} |x_2|^{N_2s+n_2-1}|y_2|^{N_1s+n_1-1}|f_2(x_2,y_2)|^2\,|dx_2|\,|dy_2|$$

and

$$\int_{P^{e+1}}\int_R |x_3|^{N_3s+n_3-1}|y_3|^{N_2s+n_2-1}|f_3(x_3,y_3)|^s\,|dx_3|\,|dy_3|.$$

Evaluating these integrals gives us

$$\frac{(1-q^{-1})^2q^{-es}q^{-(N_2s+n_2)(e+1)}q^{-(N_1s+n_1)}}{(1-q^{-(N_1s+n_1)})(1-q^{-(N_2s+n_2)})}+\frac{(1-q^{-1})^2q^{-es}q^{-(N_2s+n_2)(e+1)}}{(1-q^{-(N_2s+n_2)})(1-q^{-(N_3s+n_3)})}.$$

Thus $Z(s)$ has at most a simple pole at $-n_2/N_2$, and the residue is a constant multiple of

$$\frac{q^{-(-N_1n_2/N_2+n_1)}}{1-q^{-(-N_1n_2/N_2+n_1)}}+\frac{1}{1-q^{-(-N_3n_2/N_2+n_3)}}=\frac{1}{1-q^{(n_2N_1-n_1N_2)/N_2}}+\frac{1}{1-q^{(n_2N_3-n_3N_2)/N_2}}-1.$$

The residue will be zero $\Leftrightarrow$

$$n_1 N_2 - n_2 N_1 = n_2 N_3 - n_3 N_2.$$

Rewriting the above expression we see that the residue will be zero if and only if

$$\frac{n_2}{N_2} = \frac{n_1 + n_3}{N_1 + N_3}.$$

Since this is true for $I_{k-1} < I < I_k$; by Lemma 1 we see that the above integrals do not contribute to a possible pole at $s = -n_2/N_2$. A similar argument applied to any exceptional curve whose quotient of numerical data is equal to n_2/N_2 shows that no pole occurs at $s = -n_I/N_I$ in our original notation. q.e.d.

4. The Existence of Poles of $Z(s)$

As is shown in [3], there is a pole at $s = -n_{I_1}/N_{I_1}$, which is the pole closest to the origin. We now proceed to show that poles must also occur at the values $s = -n_{I_i}/N_{I_i}$ for $1 < i \leqq g$.

Theorem 2. *$Z(s)$ has simple poles at the values $s = -n_{I_i}/N_{I_i}$, for $1 \leqq i \leqq g$.*

Proof. We first consider the case $i < g$. Denote the I_i exceptional curve by E_{I_i} and as before we have that E_{I_i} intersects three exceptional curves E_{i_1}, E_{i_2}, and E_{i_3} with $i_1 < i_2 < I_i < i_3$. Denote the numerical data for E_{i_k} by (N_k, n_k). The contribution to the pole at $s = -n_{I_i}/N_{I_i}$ will come from $Z_{I_i,1}(s)$ and $Z_{I_{i_3},1}(s)$.

We examine the pullback of f at the $(I_i - 1)^{\text{st}}$ stage of the resolution process in the affine plane that still requires desingularization. At this stage we have that $f(x, y) = x_0^{N_2} y_0^{N_1} \cdot [(a x_0 + b y_0)^n + Q(x_0, y_0)]$, where $Q(x_0, y_0)$ consists of terms of degree $> n$. If $ab^{-1} \in R$ we replace y_0 by $y_0 - ab^{-1} x_0$. If $ab^{-1} \notin R$ we interchange coordinates and replace y_0 by $y_0 - a^{-1} b x_0$. Then we are either in the situation where $f(x, y) = x_0^{N_2} (y_0 + d x_0)^{N_1} \cdot \tilde{f}_0(x_0, y_0)$ or $f(x, y) = x_0^{N_1} (y_0 + d x_0)^{N_2} \cdot \tilde{f}_0(x_0, y_0)$ where

$$\tilde{f}_0(x_0, y_0) = c\, y_0^n + \sum_{i=1}^{n} y_0^{n-i} P_i(x_0) + \sum_{k=0}^{\infty} y_0^{n+k} R_k(x_0),$$

where $\operatorname{ord}_{x_0} P_i > i$ and $\operatorname{ord}_{x_0} R_0 > 0$. The proof of both cases is similar, so suppose we are in the first case.

Applying the quadratic transformation $x_0 = x_1 y_1$, $y_0 = x_1$; gives

$$f(x, y) = x_1^{N_{I_i}} y_1^{N_2} (1 + d y_1)^{N_1} f_1(x_1, y_1)$$

where

$$f_1(x_1, y_1) = c + \sum_{i=1}^{n} P_i'(x_1, y_1) + \sum_{k=0}^{\infty} x_1^k R_k(x_1 y_1)$$

and $\operatorname{ord}_{x_1} P_i' > 0$. Applying the quadratic transformation $x_0 = x_2$, $y_0 = x_2 y_2$; gives

$$f(x, y) = x_2^{N_{I_i}} (y_2 + d)^{N_1} f_2(x_2, y_2)$$

where

$$f_2(x_2, y_2) = c\, y_2^n + \sum_{i=1}^{n} y_2^{n-i} P_i'(x_2) + \sum_{k=0}^{\infty} x_2^k y_2^{n+k} R_k(x_2)$$

and $\operatorname{ord}_{x_2} P_i' > 0$. This is the form of the pullback of f at the I_i^{th} stage in the affine plane that still requires desingularization, so after applying a sequence of transformations resulting in the coordinate change $x_2 = x_3^l y_3$, $y_2 = x_3$ we obtain

$$f(x, y) = x_3^{N_3} y_3^{N_{I_i}} (x_3 + d)^{N_1} f_3(x_3, y_3)$$

where f_3 is obtained by factoring out powers of x_3 from $f_2'(x_3^l y_3, x_3)$. We have

$$f_2'(x_3^l y_3, x_3) = c\, x_3^n + \sum_{i=1}^{n} x_3^{n-i} P_i'(x_3^l y_3) + \sum_{k=0}^{\infty} x_3^{lk+n+k} y_3^k R_k(x_3^l y_3).$$

Since f_3 must have a non-zero constant term and we have that $\operatorname{ord}_{x_2} P_i' > 0$ and $\operatorname{ord}_{x_2} R_0 > 0$, f_3 is obtained from this by factoring out x_3^n and f_3 has the form $c + y_3 g(x_3, y_3)$.

So $Z_{I_i,1}(s)$ and $Z_{i_3,1}(s)$ have the form

$$Z_{I_i,1}(s) = \int_P \int_R |x_1|^{N_{I_i} s + n_{I_i} - 1} |y_1|^{N_2 s + n_2 - 1} |1 + d y_1|^{N_1 s + n_1 - 1} |f_1(x_1, y_1)|^s |dx_1|\, |dy_1|$$

and

$$Z_{i_3,1}(s) = \int_P \int_R |x_3|^{N_3 s + n_3 - 1} |y_3|^{N_{I_i} s + n_{I_i} - 1} |x_3 + d|^{N_1 s + n_1 - 1} |f_3(x_3, y_3)|^s |dx_3|\, |dy_3|.$$

First consider $Z_{I_i,1}(s)$. Since f_1 has the form $c + x_1 h(x_1, y_1)$ we see that $|f_1| = q^{-l}$ on $P^{l+1} \times P$ where $\operatorname{ord} c = l$. We split the domain of integration into $P^{l+1} \times P$ and $(R - P^{l+1}) \times P$. First consider the domain of integration $(R - P^{l+1}) \times P$. We have $|1 + d y_1| = 1$ on this domain. The strict transform $f_1(x_1, y_1)$ is non-singular hence by applying techniques similar to those used in Theorem 1 we see that this domain of integration does not contribute to the pole at $s = -n_{I_i}/N_{I_i}$. So we need only consider the domain $P^{l+1} \times P$. The integral over this domain is

$$\frac{(1 - q^{-1})^2 q^{-ls} q^{-(N_{I_i} s + n_{I_i})(l+1)} q^{-(N_2 s + n_2)}}{(1 - q^{-(N_{I_i} s + n_{I_i})})(1 - q^{-(N_2 s + n_2)})},$$

so the contribution to the residue is

$$\frac{(1 - q^{-1})^2 q^{-ls} q^{-(N_2 s + n_2)}}{(1 - q^{-(N_2 s + n_2)})}, \tag{16}$$

where $s = -n_{I_i}/N_{I_i}$.

Now consider $Z_{i_3,1}(s)$. Since f_3 has the form $c + y_3 g(x_3, y_3)$, we see that $|f_3| = q^{-l}$ on $R \times P^{l+1}$. So we split the domain of integration into $R \times P^{l+1}$ and $R \times (P - P^{l+1})$. We also write $d = \pi^k u$ where $\operatorname{ord} \pi = 1$ and $u \in U = R - P$. In each of these cases we will further split up the domain of integration by writing R as the disjoint union of P^{k+1}, $R - P^k$, $-d + P^{k+1}$, and $\pi^k u_i + P^{k+1}$ for $1 \leqq i \leqq q-2$, where the u_i are a set of coset representatives modulo P in $U - (-u + P)$.

First consider $R\times(P-P^{l+1})$. Write $R-P^k$ as the disjoint union of cosets modulo P^{k+1}. Then on domains of integration of the form $(a+P^{k+1})\times(P-P^{l+1})$ where $a\not\equiv -d \bmod P^{k+1}$ we have that $|x_3+d|$ is constant. We can now use the fact that f_3 is nonsingular and apply techniques similar to those used above and in Theorem 1 to show that these domains do not contribute to the residue. On $(-d+P^{k+1})\times(P-P^{l+1})$ we have that $|x_3|$ is constant, so after replacing x_3 by $-d+x_3$ we can use the same techniques to show that this domain does not contribute to the residue.

Now consider the domain $R\times P^{l+1}$, and its further decomposition into the subdomains described above. In the domain $P^{k+1}\times P^{l+1}$, we have $|x_3+d|=q^{-k}$, so the contribution to the residue from the integral over this domain is

$$\frac{(1-q^{-1})^2 q^{-ls} q^{-k(N_3 s+n_3-1)} q^{-(N_3 s+n_3)(k+1)}}{(1-q^{-(N_3 s+n_3)})} \tag{17}$$

where $s=-n_{I_i}/N_{I_i}$.

In the domain $(-d+P^{k+1})\times P^{l+1}$, we change coordinates by replacing x_3 by $-d+x_3$, and since $|-d+x_3|=q^{-k}$, this integral becomes

$$q^{-ls} q^{-k(N_3 s+n_3)} \int_{P^{l+1}} \int_{P^{k+1}} |y_3|^{N_{I_i} s+n_{I_i}-1} |x_3|^{N_1 s+n_1-1} |dx_3||dy_3|$$

which gives a contribution to the residue of

$$\frac{(1-q^{-1})^2 q^{-ls} q^{-k(N_3 s+n_3)} q^{-(N_1 s+n_1)(k+1)}}{(1-q^{-(N_1 s+n_1)})}, \tag{18}$$

where $s=-n_{I_i}/N_{I_i}$.

In the domains of the form $(\pi^k u_i+P^{k+1})\times P^l$ we have $|x_3+d|=q^{-k}$, and $|x_3|=q^{-k}$. Changing coordinates by replacing x_3 by $\pi^k u_i+x_3$ we get integrals of the form

$$q^{-ls} q^{-k(N_1 s+n_1)} q^{-k(N_3 s+n_3-1)} \int_{P^{l+1}} \int_{P^{k+1}} |y_3|^{N_{I_i} s+n_{I_i}-1} |dx_3||dy_3|.$$

There are $q-2$ integrals of this form so by evaluating the residue contribution from each of them we get a combined residue contribution of

$$(1-q^{-1}) q^{-ls}(q-2) q^{-k((N_1+N_3)s+n_1+n_3)} q^{-1}. \tag{19}$$

On $(R-P^k)\times P^l$ we have $|x_3+d|=|x_3|$, so we get an integral of the form

$$\int_{P^l} \int_{R-P^k} |x_3|^{(N_1+N_3)s+(n_1+n_3)-2} |y_3|^{N_I s+n_I-1} |dx_3||dy_3|$$

which gives a residue contribution of

$$(1-q^{-1})^2 \frac{1-q^{-((N_1+N_3)s+n_1+n_3-1)k}}{1-q^{-((N_1+N_3)s+n_1+n_3-1)}}. \tag{20}$$

Combining the contributions to the residue from (16), (17), (18), (19), and (20) gives $(1-q^{-1})^2 q^{-ls}$ times

$$q^{(\alpha_4-1)k}\frac{q^{\alpha_1}}{1-q^{\alpha_1}}+\frac{q^{\alpha_2}}{1-q^{\alpha_2}}+q^{\alpha_4 k}\frac{q^{\alpha_3}}{1-q^{\alpha_3}}+q^{\alpha_4 k}(q-2)\frac{q^{-1}}{1-q^{-1}}+\frac{1-q^{\alpha_4 k}}{1-q^{\alpha_4}}$$

where $\alpha_j=-(N_j s+n_j)$ for $1\leqq j\leqq 3$, $\alpha_4=\alpha_1+\alpha_3+1$, and $s=-n_{I_i}/N_{I_i}$.

If $q=p^f$, then we consider the above expression in the fully ramified extension of Q_p obtained by adjoining π where $\pi^{N_{I_i}}=p$. We then consider the expansion of the above expression in integral powers of π. Let $\gamma_j=N_{I_i}\alpha_j$ for $1\leqq j\leqq 4$. We have

$$\frac{q^{\alpha_j}}{1-q^{\alpha_j}}=\pi^{\gamma_j f}+\pi^{2\gamma_j f}+\ldots$$

if $\alpha_j>0$,

$$\frac{q^{\alpha_j}}{1-q^{\alpha_j}}=-1-\pi^{-\gamma_j f}-\pi^{-2\gamma_j f}-\ldots$$

if $\alpha_j<0$, and

$$(q-2)\frac{q^{-1}}{1-q^{-1}}=2+\pi^{N_{I_i}f}+\pi^{2N_{I_i}f}+\ldots.$$

Since $n_3/N_3>n_{I_i}/N_{I_i}$ we have $\alpha_3<0$. By referring to the proof of Lemma 2 we also see that we have either $\alpha_1<0$ and $\alpha_2>0$ or $\alpha_1>0$ and $\alpha_2<0$. Thus the only way the expression for the residue could possibly be zero is if the integers $\{|\gamma_1|,|\gamma_2|,-\gamma_3,N_{I_i}\}$ are equal in pairs. But none of $|\gamma_1|,|\gamma_2|$ or $-\gamma_3$ can equal N_{I_i} by Lemma 2. Hence the residue is non-zero at $s=-n_{I_i}/N_{I_i}$ for $1\leqq i\leqq g$.

We now consider the case $i=g$. Denoting the I_g exceptional curve by E_{I_g} we have that E_{I_g} intersects two exceptional curves E_{i_1} and E_{i_2} with $i_1<i_2<I_g$; and the final strict transform. As before let the numerical data for E_{i_k} be denoted by (N_k,n_k) where $k=1,2$. The contribution to the pole at $s=-n_{I_g}/N_{I_g}$ will come from $Z_{I_g,1}(s)$ and $Z_{I_g,2}(s)$. By an argument entirely similar to the $i\neq g$ case we have that

$$Z_{I_g,1}(s)=\int_P\int_R|x_1|^{N_{I_g}s+n_{I_g}-1}|y_1|^{N_2 s+n_2-1}|1+dy_1|^{N_1 s+n_1-1}|f_1(x_1,y_1)|^s|dx_1||dy_1|$$

and

$$Z_{I_g,2}(s)=\int_R\int_R|x_2|^{N_{I_g}s+n_{I_g}-1}|y_2+d|^{N_1 s+n_1-1}|f_2(x_2,y_2)|^s|dx_2||dy_2|,$$

where $f_1(x_1,y_1)=c+P'(x_1,y_1)$ with $\operatorname{ord}_{x_1}P'\geqq 1$, and $f_2(x_2,y_2)=cy_2+P''(x_2,y_2)$ with $\operatorname{ord}_{x_2}P''\geqq 1$.

First consider $Z_{I_g,1}(s)$. By an argument entirely similar to the case $i\neq g$ we have that this gives a contribution to the residue of

$$\frac{(1-q^{-1})^2 q^{-ls}q^{-(N_2 s+n_2)}}{(1-q^{-(N_2 s+n_2)})},\tag{21}$$

where $s=-n_{I_g}/N_{I_g}$, and $\operatorname{ord}c=l$.

Now consider $Z_{I_g,2}(s)$. We change coordinates by replacing y_2 by y_2-d, so our integral has the form

$$\int_R \int_R |x_2|^{N_{I_g} s + n_{I_g} - 1} |y_2|^{N_1 s + n_1 - 1} |b + c y_2 + Q(x_2, y_2)|^s |dx_2| |dy_2|$$

where $b = -cd$ and $\operatorname{ord}_{x_2} Q(x_2, y_2) \geqq 1$. Let $\operatorname{ord} c = l$ and $\operatorname{ord} d = k$. We can reduce the domain of integration to $P^m \times R$ where $m = k + l + 1$ without affecting the residue. Writing $d = \pi^k u$ we further split this domain of integration into $P^m \times P^{k+1}$, $P^m \times (R - P^k)$, $P^m \times (-d + P^{k+1})$, and $\coprod_{i=1}^{q-2} P^m \times (\pi^k u_i + P^{k+1})$ where the u_i are a set of coset representatives modulo P in $U - (-u + P)$.

On $P^m \times P^{k+1}$ we have $|b + c y_2 + Q(x_2, y_2)| = |b| = q^{-(m-1)}$. So the integral over $P^m \times P^{k+1}$ contributes

$$(1 - q^{-1})^2 q^{-(m-1)s} \frac{q^{-(N_1 s + n_1)(k+1)}}{(1 - q^{-(N_1 s + n_1)})} \tag{22}$$

to the residue. On $P^m \times (R - P^k)$ we have $|b + c y_2 + Q(x_2, y_2)| = |c y_2| = q^{-l} |y_2|$. So the integral over this domain gives a residue contribution of

$$(1 - q^{-1})^2 q^{-ls} \frac{1 - q^{-((N_1+1)s + n_1)k}}{1 - q^{-((N_1+1)s + n_1)}}. \tag{23}$$

On the domains $P^m \times (\pi^k u_i + P^{k+1})$ we have $|b + c y_2 + Q(x_2, y_2)| = q^{-(m-1)}$ and $|y_2| = q^{-k}$. Since there are $q - 2$ of these domains, they give a combined residue contribution of

$$(q-2) q^{-(m-1)s} q^{-k(N_1 s + n_1)} q^{-1} (1 - q^{-1}). \tag{24}$$

It remains to consider the domain $P^m \times (-d + P^{k+1})$. We have $|y_2| = q^{-k}$, and by replacing y_2 by $d + y_2$ we have

$$q^{-k(N_1 s + n_1)} \int_{P^{k+1}} \int_{P^m} |x_2|^{N_{I_g} s + n_{I_g} - 1} |c y_2 + Q'(x_2, y_2)|^s |dx_2| |dy_2|$$

where $\operatorname{ord}_{x_2} Q_2' \geqq 1$. Let $I(q^{-s})$ denote the above integral. We are interested in calculating

$$\lim_{s \to -n_{I_g}/N_{I_g}} (1 - q^{-(N_{I_g} s + n_{I_g})}) I(q^{-s}).$$

Let

$$I_n(q^{-s}) = \int_{P^{k+1} - P^{k+n-1}} \int_{P^m} |x_2|^{N_{I_g} s + n_{I_g} - 1} |c y_2 + Q'(x_2, y_2)|^s |dx_2| |dy_2|.$$

Since $(1 - q^{-(N_{I_g} s + n_{I_g})}) I_n(q^{-s})$ converges uniformly to $(1 - q^{-(N_{I_g} s + n_{I_g})}) I(q^{-s})$ as $n \to \infty$ for s in a neighborhood of $-n_{I_g}/N_{I_g}$; we have that the contribution to the residue is equal to

$$\lim_{n \to \infty} \lim_{s \to -n_{I_g}/N_{I_g}} (1 - q^{-(N_{I_g} s + n_{I_g})}) I_n(q^{-s}).$$

By reasoning similar to before the contribution to the residue at $s = -n_{I_g}/N_{I_g}$ of $I_n(q^{-s})$ is equal to the contribution from

$$\int_{P^{k+1} - P^{k+n-1}} \int_{P^{k+n}} |x_2|^{N_{I_g} s + n_{I_g} - 1} |c y_2 + Q'(x_2, y_2)|^s |dx_2| |dy_2|.$$

But $|cy_2+Q'(x_2,y_2)|=q^{-l}|y_2|$ on this domain of integration, so this integral gives a contribution to the residue of

$$(1-q^{-1})^2 q^{-ls}\frac{q^{-(s+1)(k+1)}-q^{-(s+1)(k+n-1)}}{(1-q^{-(s+1)})}.$$

Taking the limit as $k\to\infty$ we get a contribution to the residue of

$$(1-q^{-1})^2 q^{-ls}\frac{q^{-(s+1)(k+1)}}{1-q^{-(s+1)}}. \tag{25}$$

Combining the contributions to the residue from (21), (22), (23), (24), and (25) gives $(1-q^{-1})^2$

$$q^{-(m-1)s}q^{\alpha_1 k}\frac{q^{\alpha_1}}{1-q^{\alpha_1}}+q^{-ls}\frac{q^{\alpha_2}}{1-q^{\alpha_2}}+q^{-ls}q^{\alpha_3 k}\frac{q^{\alpha_3}}{1-q^{\alpha_3}}$$
$$+q^{-(m-1)s}q^{\alpha_1 k}(q-2)\frac{q^{-1}}{1-q^{-1}}+q^{-ls}\frac{1-q^{(\alpha_1+\alpha_3-1)k}}{1-q^{(\alpha_1+\alpha_3-1)}}$$

where $\alpha_j=-(N_j s+n_j)$ for $j=1,2$; $\alpha_3=-(s+1)$, and $s=-n_{I_g}/N_{I_g}$.

Again consider the above expression in the fully ramified extension of Q_p obtained by adjoining π where $\pi^{N_{I_g}}=p$. Let $\gamma_j=N_{I_g}\alpha_j$ for $1\leqq j\leqq 3$. We clearly have $\alpha_3<0$ and again by referring to the proof of Lemma 2 we have either $\alpha_1<0$ and $\alpha_2>0$ or $\alpha_1>0$ and $\alpha_2<0$. Thus the only way the expression could be zero is if the integers $\{|\gamma_1|,|\gamma_2|,-\gamma_3,N_{I_g}\}$ are equal in pairs. But none of $|\gamma_1|,|\gamma_2|$, or $-\gamma_3$ can equal N_{I_g} by Lemma 2. Hence the residue at $s=-n_{I_g}/N_{I_g}$ is non-zero. q.e.d.

We also mention that at each stage of the resolution process we have introduced the possibility of a pole at $s=-1$ corresponding to the strict transform, whose existence we shall not consider.

References

1. Bernstein, I.N., Gel'fand, S.I.: Meromorphic property of the function P^{λ}. Functional Anal. Appl. **3**, 68-69 (1969)
2. Hironaka, H.: Introduction to the Theory of Infinitely Near Singular Points, Memorias de Matematica del Istituto "Jorge Juan", 28, Madrid, 1974
3. Igusa, J.-I.: On the first terms of certain asymptotic expansions. Complex analysis and algebraic geometry. (Baily, W.L., Jr., Shioda, T., eds.) Cambridge University Press, 1977, pp. 357-368
4. Igusa, J.-I.: Some observations in higher degree characters. Am. J. Math. **99**, 393-417 (1977)
5. Strauss, L.: On the evaluation of certain integrals through desingularization. Ph.D. thesis; Johns Hopkins University, 1978
6. Zariski, O.: Algebraic Surfaces. Ergeb. der Math., Springer (1932); Chelsea (1948)
7. Zariski, O.: Studies in equisingularity III, Saturation of local rings and equisingularity. Amer. J. Math. **90**, 961-1023 (1968)

Oblatum 22-V-1982 & 12-IV-1983

Invent. math. 73, 467-489 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

Calderón-Zygmund Kernels Carried by Linear Subspaces of Homogeneous Nilpotent Lie Algebras

D. Müller

Fakultät für Mathematik, Universität Bielefeld, Universitätsstr. 25
D-4800 Bielefeld, Federal Republic of Germany

Introduction

In recent years Calderón-Zygmund kernels on nilpotent Lie groups and the corresponding convolution operators have been studied by several authors (see e.g. [4], [13], [15]). In [13] it was shown that the classical theory of Calderón-Zygmund kernels on $\mathbb{R}^n$ has very natural extensions to kernels on nilpotent Lie groups. However, if one wants to study more singular convolution operators, for example convolution with distributions which are extensions of Calderón-Zygmund kernels on linear subspaces, then many of the corresponding questions are open. In [3] D. Geller and E.M. Stein initiated research in this direction and showed that certain homogeneous distribution of degree $-2n$, carried by subspaces complementary to the center of the $(2n+1)$-dimensional Heisenberg group H_n, when extended to H_n, yield bounded convolution operators on $L^p(H_n)$ for $1<p<\infty$.

In this paper we will prove some quite general results about the L^2-boundedness of convolution operators on nilpotent Lie groups which arise from certain classes of Calderón-Zygmund kernels on linear subspaces, supplementary to the center, of the corresponding Lie algebra. The article is organized as follows:

In Sect. 1 we introduce some basic notations, and in Sect. 2 we define two classes (H1) and (H2) of Calderón-Zygmund kernels with weak homogeneity properties and study some properties of these kernels. In Sect. 3 we prove our first main result (Theorem 3.1) which essentially states that a Calderón-Zygmund kernel of class (H1) carried by a linear subspace which is supplementary to the center, and which contains a nontrivial central part, yields a bounded convolution operator on L^2, provided that certain means of the kernel over all affine subspaces which are translates of the central part are zero. Section 4 is a short review of results of Moore and Wolf about square integrable representations of nilpotent Lie groups and concludes with an estimate for twisted convolutions. In the following section we prove another estimate for twisted convolutions with compactly supported distributions on nilpotent Lie groups,

which is based on results of H. Leptin [7]. Finally, in Sect. 6 we present our second main result (Theorem 6.1) which states the following: If k is a Calderón-Zygmund kernel of class (H2) on a subspace supplementary to the center of the Lie algebra (of a homogeneous nilpotent Lie group), and if the group is "completely square integrable", then k yields a bounded convolution operator on L^2. This result generalizes the L^2-part of the result of Geller and Stein.

We would like to thank J. Cygan, A. Hulanicki, H. Leptin and L. Richardson for helpful conversations about this subject.

1. The General Set-Up

Let $\mathfrak{n}$ be a nilpotent Lie algebra with a one-parameter family of dilations $D=\{D_r\}_{r>0}$, satisfying the following conditions:

i) D_r is an automorphism of $\mathfrak{n}$ for any $r>0$,

ii) $D_r D_{r'}=D_{rr'}$ for any $r, r'>0$,

iii) D_r is diagonalizable for any $r>0$, and its eigenvalues are $r^{\alpha_1}, r^{\alpha_2}, \ldots, r^{\alpha_k}, r$, where

$$0<\alpha_1<\alpha_2<\ldots<\alpha_k<\alpha_{k+1}=1.$$

If we denote by $\mathfrak{n}_{\alpha_j}$ the eigenspace associated to the eigenvalues r^{α_j}, then the family $\{\mathfrak{n}_{\alpha_j}\}$ yields a gradation of $\mathfrak{n}$, since

$$[\mathfrak{n}_{\alpha_j}, \mathfrak{n}_{\alpha_l}] \subseteq \mathfrak{n}_{\alpha_j+\alpha_l} \tag{1}$$

(with the convention that $\mathfrak{n}_\beta=0$ if β is not one of the α_j).

If $\mathfrak{v}$ is a D-invariant linear subspace of $\mathfrak{n}$, then of course $\mathfrak{v}$ decomposes into eigenspaces $\mathfrak{v}_{\alpha_j}$, i.e.

$$\mathfrak{v}=\sum_j^{\oplus} \mathfrak{v}_{\alpha_j}, \quad \text{with } \mathfrak{v}_{\alpha_j}=\mathfrak{v}\cap\mathfrak{n}_{\alpha_j}, \tag{2}$$

where the summation is understood to be over all indices j with $\mathfrak{v}\cap\mathfrak{n}_{\alpha_j}\neq 0$.

Let $\mathfrak{z}$ denote the center of $\mathfrak{n}$, and let $\mathfrak{w}$ be a fixed D-invariant linear subspace of $\mathfrak{n}$ supplementary to $\mathfrak{z}$. Then, since $\mathfrak{z}$ is of course D-invariant,

$$\mathfrak{z}=\sum_j^{\oplus} \mathfrak{z}_{\alpha_j} \quad \text{and} \quad \mathfrak{w}=\sum_j^{\oplus} \mathfrak{w}_{\alpha_j}.$$

If we set $\mathfrak{z}_0=\mathfrak{w}\cap\mathfrak{z}$, and if we choose a linear subspace $\mathfrak{n}_0$ of $\mathfrak{w}$ which is D-invariant and complementary to $\mathfrak{z}_0$, and a D-invariant subspace $\tilde{\mathfrak{z}}$ of $\mathfrak{z}$ which is complementary to $\mathfrak{z}_0$, then we have

$$\mathfrak{w}=\mathfrak{n}_0\oplus\mathfrak{z}_0 \tag{3}$$

and

$$\mathfrak{n}=\mathfrak{w}\oplus\tilde{\mathfrak{z}}, \quad \mathfrak{z}=\mathfrak{z}_0\oplus\tilde{\mathfrak{z}}. \tag{4}$$

Let further N denote the connected, simply connected Lie group having $\mathfrak{n}$ as its Lie algebra. We will use the elements of $\mathfrak{n}$ as parameters of N via the exponential mapping, i.e. if $x, y \in \mathfrak{n}$, xy denotes the element of $\mathfrak{n}$ such that $\exp_N xy = \exp_N x \exp_N y$. In particular, $x^{-1} = -x$. In the same way as we identify N with $\mathfrak{n}$, we will also identify the center $Z = \exp_N \mathfrak{z}$ of N with $\mathfrak{z}$ and $W = \exp \omega$ with ω. Lebesgue measure on $\mathfrak{n}$ (resp. $\mathfrak{z}$) is the same as the Haar measure for the group N (resp. Z).

Now, since $\mathfrak{n}_0 \simeq \mathfrak{n}/\mathfrak{z}$ as linear spaces, we can regard $\mathfrak{n}_0$ as a Lie algebra. Since $\mathfrak{n} = \mathfrak{n}_0 \oplus \mathfrak{z}$, we will often denote an element x of $\mathfrak{n}$ as a pair $x = (z, t)$ with $z \in \mathfrak{n}_0$, $t \in \mathfrak{z}$. Multiplication in $\mathfrak{n}$ may then be written

$$(z, s)(w, t) = (z w, s + t + P(z, w)) \quad \text{if } z, w \in \mathfrak{n}_0, s, t \in \mathfrak{z}, \tag{5}$$

where P is a polynomial cocycle from $\mathfrak{n}_0 \times \mathfrak{n}_0$ to $\mathfrak{z}$. To be more specific, P has the following properties:

$$P(-z, -w) = -P(w, z), \tag{6}$$

$$P(k_1 z, k_2 z) = 0 \quad \text{for any } k_1, k_2 \in \mathbb{R}, \tag{7}$$

$$P(z, w) + P(zw, y) = P(z, wy) + P(w, y), \tag{8}$$

$$P(D_r z, D_r w) = D_r P(z, w) \quad \text{for any } r > 0. \tag{9}$$

For $j = 1, \ldots, k+1$, we fix a norm $|\cdot|_{\alpha_j}$ on $\mathfrak{n}_{\alpha_j}$ and define a homogeneous gauge $|\cdot|$ on $\mathfrak{n}$ as

$$|x| = \sum_{j=1}^{k+1} |x_j|_{\alpha_j}^{1/\alpha_j}, \quad \text{if } x = \sum_j x_j \quad \text{with} \quad x_j \in \mathfrak{n}_{\alpha_j}.$$

If again $\mathfrak{v}$ denotes a D-invariant linear subspace of $\mathfrak{n}$, we also denote by $|\cdot|$ the restriction of $|\cdot|$ to $\mathfrak{v}$, and define a dual gauge, again denoted by $|\cdot|$, on $\mathfrak{v}^*$ as follows:

Let D_r act on $\mathfrak{v}^*$ by the adjoint action. Then $\mathfrak{v}^*$ also decomposes into eigenspaces $\mathfrak{v}_{\alpha_j}^*$, i.e.

$$\mathfrak{v}^* = \bigoplus_j \mathfrak{v}_{\alpha_j}^*.$$

Since obviously $\langle \lambda_i, x_j \rangle = 0$ for $\lambda_i \in \mathfrak{v}_{\alpha_i}^*$, $x_j \in \mathfrak{v}_{\alpha_j}$ and $i \neq j$, $\mathfrak{v}_{\alpha_j}^*$ may be canonically identified with the dual space of $\mathfrak{v}_{\alpha_j}$. If $|\cdot|_{\alpha_j}^*$ denotes the dual norm on $\mathfrak{v}_{\alpha_j}^*$, i.e.

$$|\lambda_j|_{\alpha_j}^* = \sup_{\substack{x \in \mathfrak{v}_{\alpha_j} \\ |x|_{\alpha_j} \leq 1}} |\langle \lambda_j, x \rangle|,$$

then we set

$$|\lambda| = \sum_j |\lambda_j|_{\alpha_j}^{*\,1/\alpha_j}, \quad \text{if } \lambda = \sum_j \lambda_j \quad \text{with} \quad \lambda_j \in \mathfrak{v}_{\alpha_j}^*.$$

Then it is clear that

$$|D_r x| = r|x| \quad \text{and} \quad |D_r \lambda| = r|\lambda| \tag{10}$$

for any $x \in \mathfrak{n}$ and any $\lambda \in \mathfrak{v}^*$. Moreover, it is not hard to see that there exist constants $\gamma_1 > 0$

and $\gamma > 0$ such that:

$$\gamma_1(|z|+|t|) \leqq |(z,t)| \leqq \gamma(|z|+|t|), \quad \text{if } z \in \mathfrak{n}_0,\ t \in \mathfrak{x}; \tag{11}$$

$$|x+y| \leqq \gamma(|x|+|y|); \tag{12}$$

$$|xy| \leqq \gamma(|x|+|y|); \tag{13}$$

any $\lambda \in \mathfrak{v}^*$ there exists an $x \in \mathfrak{v}$, such that

$$\langle \lambda, x \rangle = 1 \quad \text{and} \quad |x| \leqq \gamma |\lambda|^{-1}. \tag{14}$$

Now, let f be an integrable function on $\mathfrak{n}$. For $z \in \mathfrak{n}_0$ and $\lambda \in \mathfrak{x}^*$ let

$$f_\lambda(z) = \int_{\mathfrak{x}} f(z,t)\, e^{-i\langle \lambda, t\rangle}\, dt. \tag{15}$$

If we extend this notion in the usual way also to square-integrable functions, then Plancherel's theorem yields

$$|f|^2_{L^2(\mathfrak{n})} = \int_{\mathfrak{x}^*} |f_\lambda|^2_{L^2(\mathfrak{n}_0)}\, d\lambda \tag{16}$$

for a suitable normalization of the Lebesgue measure $d\lambda$ on $\mathfrak{x}^*$. Moreover, if f and g are integrable functions on $\mathfrak{n}$, then a simple computation (compare for example [13]) shows that

$$(f * g)_\lambda(z) = \int_{\mathfrak{n}_0} f_\lambda(z w^{-1})\, g_\lambda(w)\, e^{i\langle \lambda, P(z,-w)\rangle}\, dw. \tag{17}$$

For $\lambda \in \mathfrak{x}^*$ we denote by ω_λ the cocycle

$$\omega_\lambda(z,w) = e^{i\langle \lambda, P(z,w)\rangle}.$$

If we define for integrable functions φ and ψ on $\mathfrak{n}_0$ as in [13] the λ-twisted convolution of ψ and φ as

$$\varphi *_\lambda \psi(z) = \int_{\mathfrak{n}_0} \varphi(z w^{-1})\, \psi(w)\, \omega_\lambda(z,-w)\, dw, \tag{18}$$

then (17) reads

$$(f * g)_\lambda = f_\lambda *_\lambda g_\lambda. \tag{19}$$

We also adapt the definition of a Calderón-Zygmund kernel from [13]:

Definition. A Calderón-Zygmund kernel on $\mathfrak{w}$ is a tempered distribution k on $\mathfrak{w}$ which is a locally integrable function away from the origin and satisfies the following conditions for some positive constant $A(k)$:

$$|\hat{k}|_{L^\infty(\mathfrak{w}^*)} \leqq A(k), \tag{20}$$

$$\int_{|x| > \beta|y|} |k(x-y) - k(x)|\, dx \leqq A(k) \quad \text{for every } y \neq 0, \tag{21}$$

where dx denotes Lebesgue measure on $\mathfrak{w}$, and where β is a fixed constant with $\beta > 2\gamma$ for the constant γ which appears in (11) to (14).

In the following, let k be a Calderón-Zygmund kernel on $\mathfrak{w}$. We extend k to a tempered distribution K on $\mathfrak{n}$ by

$$\langle K, \varphi \rangle = \langle k, \varphi|_{\omega} \rangle, \quad \text{if } \varphi \in \mathcal{S}(n), \tag{22}$$

i.e., if δ denotes the point measure at 0 on $\tilde{z}$, then $K = k \otimes \delta$, if we write $n = \omega \oplus \tilde{z}$.

In the following, we will study conditions on k such that K yields a bounded convolution operator on $L^2(n)$. For example, in case that $y = n$, it has been shown by F. Ricci in [13] that $K = k$ always yields a bounded convolution operator on $L^2(n)$. In [3], D. Geller and E.M. Stein have shown that in case of the $(2n+1)$-dimensional Heisenberg group N and ω complementary to the center each kernel k on ω which is homogeneous of degree $-n$ (with respect to our notation) and which is smooth outside the origin, yields a bounded operator by convolution with K. They even showed more, namely that the corresponding operator is bounded on $L^p(n)$ for $1 < p < \infty$. The advantage of our method is that it also works for more general groups, but on the other hand it seems not to be clear how one could also get results for $L^p(n)$, $p \neq 2$.

With view to the results of Ricci in [13], we will restrict ourselves to the case $\omega \subsetneq n$, although our methods would also give a partially shorter proof of the results in [13].

Remark. Throughout the paper, for convenience we will frequently denote constants by C or C', even if sometimes different constants are meant.

2. Two Classes of Calderón-Zygmund Kernels

It will turn out that for our studies two special classes of Calderón-Zygmund kernels will be wellsuited.

We introduce the following weak homogeneity properties (let v be again a D-invariant subspace of n):

Definition. i) A Calderón-Zygmund kernel k on v is said to have property (H1), if there exists a constant $C_1(k)$ such that

$$\int_{R < |x| < 2R} |k(x)|\, dx \leqq C_1(k) \tag{H1}$$

for any $R > 0$.

ii) A Calderón-Zygmund kernel k on v is said to have property (H2), if there exists a constant $C_2(k)$ such that

$$\int_{|x| > R} |k(x)|^2\, dx \leqq C_2(k)^2\, R^{-q}, \tag{H2}$$

where $q = \Sigma\, \alpha_j \dim v_{\alpha_j}$ denotes the homogeneous dimension of v.

Remarks. 1. (H2) implies (H1), because, if (H2) holds, then

$$\begin{aligned} \int_{R < |x| < 2R} |k(x)|\, dx &\leqq \Big(\int_{|x| > R} |k(x)|^2\, dx \Big)^{1/2} \Big(\int_{R < |x| < 2R} dx \Big)^{1/2} \\ &\leqq C_2(k)\, R^{-q/2} \big(R^q \int_{1 < |x| < 2} dx\big)^{1/2} \\ &= C_2(k) \Big(\int_{1 < |x| < 2} dx \Big)^{1/2}. \end{aligned}$$

2. (H1) is equivalent to (H1)′: For any constants σ_1, σ_2 with $0<\sigma_1<\sigma_2$ there exists a constant $C'>0$ such that

$$\int\limits_{\sigma_1 R<|x|<\sigma_2 R} |k(x)|\,dx \leqq C' A_1(k)$$

for any $R>0$.

Of course we may assume that $\sigma_1=1$ and $\sigma_2>1$. If then $l\in\mathbb{N}$ is such that $2^l \geqq \sigma_2$, then obviously (H1) yields

$$\int\limits_{R<|x|<\sigma_2 R} |k(x)|\,dx \leqq \sum_{j=1}^{l} \int\limits_{2^{j-1}R<|x|<2^j R} |k(x)|\,dx \leqq l\, C_1(k).$$

3. Let ω be as in the preceding section. One can show that there exists a bounded Radon measure $d\sigma$ on the "unit sphere" $\Sigma=\{x\in\omega: |x|=1\}$ in ω such that for any $f\in L^1(\omega)$

$$\int\limits_{\omega} f(x)\,dx = \int\limits_0^{\infty} \int\limits_{\Sigma} f(D_r\sigma)\, r^{m-1}\, d\sigma\, dr, \tag{23}$$

(compare [11]), where m denotes the homogeneous dimension of ω.

Now, let k be a homogeneous kernel of degree $-m$ on ω, and let us assume that k can be written as

$$k(x)=\frac{\Omega(x')}{|x|^m},$$

where $x'=D_{|x|^{-1}}x$, if $x\neq 0$, and where Ω is an L^∞-function on Σ which satisfies

$$\int\limits_{\Sigma} \Omega(\sigma)\,d\sigma=0 \tag{24}$$

and the Dini condition

$$\int\limits_0^1 \frac{\omega(t)}{t}\,dt<\infty. \tag{25}$$

Here $\omega(t)$ is the modulus of continuity,

$$\omega(t)=\sup\{|\Omega(\sigma_1)-\Omega(\sigma_2)|: \sigma_1,\sigma_2\in\Sigma, |\sigma_1-\sigma_2|<t\}.$$

Then it is known (see [4], [14]), that k is a Calderón-Zygmund kernel on ω. Moreover, applying (23) we see that k also satisfies condition (H2).

4. If a Calderón-Zygmund kernel on v has property (H1) (resp. (H2)), we write

$$A_1(k)=A(k)+C_1(k) \quad \text{(resp. } A_2(k)=A(k)+C_2(k)\text{)}. \tag{26}$$

Now we will prove some properties of such Calderón-Zygmund kernels.

Let k be a Calderón-Zygmund kernel on $\omega=\nu_0\oplus\varkappa_0$. For $\lambda_0\in\varkappa_0^*$, $z\in\nu_0$, we denote by $k_{\lambda_0}(z)$ the partial Fourier transform of k in direction of $\varkappa_0$, i.e., since $\hat{k}$ is in $L^\infty(\omega^*)$, for a.e. $\lambda_0\in\varkappa_0$ we have $\hat{k}_{\lambda_0}(\xi)=\hat{k}(\xi,\lambda_0)$, $\xi\in\nu_0^*$. Formally we may write

$$k_{\lambda_0}(z)=\int_{\varkappa_0} k(z,t)\,e^{-i\langle\lambda_0,t\rangle}\,dt. \tag{27}$$

The following lemma is an adaption of a result of F. Ricci in [13], Lemma 2, to our situation:

Lemma 2.1. *Let γ and γ_1 be as in* (11)–(14). *For almost every $\lambda_0\in\varkappa_0^*$, k_{λ_0} is a Calderón-Zygmund kernel on $\varkappa_0$, with the constant β in* (21) *replaced by $\beta'=\beta\gamma\gamma_1^{-1}$, and with $A(k_{\lambda_0})\leqq A(k)$. Moreover,*

$$\int_{|z|>\beta'|\lambda_0|^{-1}} |k_{\lambda_0}(z)|\,dz\leqq 2A(k).$$

We will omit the proof, because it is essentially the same given in [13].

Lemma 2.2. *Let k be a Calderón-Zygmund kernel on ω with property* (H1). *Let further $\chi\in C_0^\infty(\omega)$ with $0\leqq\chi(x)\leqq 1$, and $\chi(x)=1$ if $|x|\leqq 1$, and let $\chi_R(x)=\chi(D_{R^{-1}}x)$.*

Then for any $R>0$, $k\chi_R$ is a Calderón-Zygmund kernel with $A(k\chi_R)\leqq CA_1(k)$.

Proof. First, we have $|\widehat{k\chi_R}|_{L^\infty}\leqq|\hat{k}|_{L^\infty}|\hat{\chi}_R|_{L^1}$, and $|\hat{\chi}_R|_{L^1}=|\hat{\chi}|_{L^1}$, hence $|\widehat{k\chi_R}|_{L^\infty}\leqq|\hat{\chi}|_{L^1}A(k)$.

Now let $w\neq 0$. Then

$$\begin{aligned}
&\int_{|z|>\beta|w|}|(k\chi_R)(z-w)-(k\chi_R)(z)|\,dz\\
&\quad\leqq\int_{|z|>\beta|w|}|k(z-w)-k(z)|\,\chi_R(z-w)\,dz\\
&\quad+\int_{|z|>\beta|w|}|k(z)|\,|\chi_R(z-w)-\chi_R(z)|\,dz\\
&\quad\leqq A(k)+\int_{|z|>\beta|w|}|k(z)|\,|\chi_R(z-w)-\chi_R(z)|\,dz.
\end{aligned}$$

Assume that $\chi(x)=0$ if $|x|\geqq\rho>1$:

We distinguish two cases:

1. $2\beta\gamma|w|<R$: If then $|z|\geqq(\rho\gamma+(2\beta)^{-1})R$, then we have $|z|\geqq\rho R$, hence $\chi_R(z)=0$, and

$$|z-w|\geqq\gamma^{-1}|z|-|w|)\geqq(\rho+(2\beta\gamma)^{-1})R-(2\beta\gamma)^{-1}R=\rho R,$$

hence also $\chi_R(z-w)=0$.

Moreover, if $|z|\leqq(2\gamma)^{-1}R$, then $|z|\leqq R$, hence $\chi_R(z)=1$, and

$$|z-w|\leqq\gamma(|z|+|w|\leqq R/2+R/2\beta\leqq R,$$

since $\beta>2\gamma\geqq 2$, hence also $\chi_R(z-w)=1$. Therefore in this case we have

$$\begin{aligned}
&\int_{|z|>\beta|w|}|k(z)|\,|\chi_R(z-w)-\chi_R(z)|\,dz\\
&\quad\leqq\int_{(\rho\gamma+\frac{1}{2\beta})R>|z|>\frac{1}{2\gamma}R}|k(z)|\,dz\leqq CA_1(k)\quad\text{by (H1)}'.
\end{aligned}$$

2. $2\beta\gamma|w| \geqq R$: If now $|z| > (2\rho+1)\beta\gamma^2|w|$, then $|z| > \rho R$, and also

$$|z-w| \geqq \gamma^{-1}|z| - |w| \geqq ((2\rho+1)\beta\gamma - 1)|w| \geqq 2\rho\beta\gamma|w| \geqq \rho R,$$

hence $\chi_R(z) = \chi_R(z-w) = 0$. Therefore

$$\int\limits_{|z|>\beta|w|} |k(z)|\,|\chi_R(z-w) - \chi_R(z)|\,dz$$
$$\leqq \int\limits_{(2\rho+1)\beta\gamma^2|w|>|z|>\beta|w|} |k(z)|\,dz \leqq C' A_1(k) \quad \text{by (H1)'}.$$

Hence, $A(k\chi_R) \leqq A(k) + (C + C')A_1(k)$.

To formulate the following proposition, we first introduce a second homogeneous gauge on $\varkappa_0$ (compare with [11], Proposition 1); we assume here $\varkappa_0 \neq 0$:

Let $\|\ \|$ denote a Euclidean norm (or any "smooth" norm) on $\varkappa_0$. Then we define

$$|t|' = r, \quad \text{if } \|D_{r^{-1}} t\| = 1, \quad \text{for } t \in \varkappa_0. \tag{28}$$

It is not hard to show that $t \to |t|'$ is everywhere continuous and C^∞ on $\varkappa_0 \setminus \{0\}$ (which is in general not true for $|\ |$), and that $|\ |$ and $|\ |'$ are equivalent.

Proposition 2.3. *Let k be a Calderón-Zygmund kernel on ω having property* (H1), *and assume that for a.e. $z \in \varkappa_0$ and every a, b with $0 < a < b$,*

$$\int\limits_{a<|t|'<b} k(z,t)\,dt = 0. \tag{29}$$

Then k_{λ_0} is in $L^1(\varkappa_0)$ for a.e. $\lambda_0 \in \varkappa_0^*$, and

$$|k_{\lambda_0}|_{L^1(\varkappa_0)} \leqq C A_1(k),$$

independently of λ_0.

Remark. If for example $R_1(x) = \Gamma\left(\frac{n+1}{2}\right)\pi^{-\frac{n+1}{2}} \cdot \frac{x_1}{\|x\|^{n+1}}$ denotes a Riesz-kernel on $\mathbb{R}^n$, then Proposition 2.3 shows that $(R_1)^1_\lambda$ has uniformly bounded L^1-norm, if we denote by $(R_1)^1_\lambda$ the partial Fourier transform in direction of x_1, evaluated at λ. But, one can show easily that $(R_1)^j_\lambda$ is not even integrable, if $(R_1)^j_\lambda$ denotes the evaluation at λ of the partial Fourier transform of R_1 in direction of x_j, $j \neq 1$.

Before we prove Proposition 2.3, we state another lemma which is proved by the same method as Lemma 3 in [13]:

Lemma 2.4. *For each $\alpha > 0$ there exists a constant $C_\alpha > 0$, such that for any $R > 0$ and any Calderón-Zygmund kernel k on ω with support contained in the "ball" $B_R = \{x \in \omega : |x| \leqq R\}$ the following estimate holds:*

$$\int\limits_\omega |k(x)|\,|x|^\alpha\,dx \leqq C_\alpha A(k) R^\alpha. \tag{30}$$

Proof. Let $x_0 \in \omega$, $x_0 \neq 0$, let $c>1$ and $x_n = D_{c^n} x_0$. For $r>0$, let $B_{x_n} = \{x \in \omega : |x-x_n| < r\}$. If $y_n = x_n - x_{n-1}$, it is clear that

$$\int_{B_{x_{n-1}}} |k(x)|\,dx \leqq \int_{B_{x_n}} |k(x-y_n)-k(x)|\,dx + \int_{B_{x_n}} |k(x)|\,dx. \tag{31}$$

If $x \in B_{x_n}$, then, by (12), $|x| \geqq \gamma^{-1}|x_n| - r$, i.e.

$$|x| \geqq \gamma^{-1} c^n |x_0| - r. \tag{32}$$

Furthermore, if $x_0 = \sum_j z_j$ with $z_j \in \omega_{\alpha_j}$, then

$$\begin{aligned} |y_n| &= |D_{C^n} x_0 - D_{C^{n-1}} x_0| \\ &= \sum_j |c^{n\alpha_j} z_j - c^{(n-1)\alpha_j} z_j|_{\alpha_j}^{1/\alpha_j} \\ &= c^{n-1} \sum_j |c^{\alpha_j} - 1|^{1/\alpha_j} |z_j|_{\alpha_j}^{1/\alpha_j} \\ &\leqq c^{n-1}(c-1)|x_0|. \end{aligned} \tag{33}$$

The middle integral in (30) can easily be estimated if $|x| > \beta |y_n|$ for $x \in B_{x_n}$. For this, by (32), (33), it suffices that $\gamma^{-1} c^n |x_0| - r > \beta c^{n-1}(c-1)|x_0|$, i.e.

$$r < c^{n-1}(\gamma^{-1} c - \beta(c-1))|x_0|. \tag{34}$$

Next we choose c small enough but greater than 1 such that $0<\mu<1$, where $\mu = \gamma^{-1} c - \beta(c-1)$. Then with $r = \mu |x_0|$ we have (34) for all x_0 and all $n \geqq 1$. Let us assume that c and r are chosen in such a way. Iterating (31) we get

$$\int_{B_{x_0}} |k(x)|\,dx \leqq n A(k) + \int_{B_{x_n}} |k(x)|\,dx,$$

since, by (21), $\int_{B_{x_j}} |k(x-y_j)-k(x)|\,dx \leqq A(k)$.

Now, if $\gamma^{-1} c^n |x_0| - r > R$, then, by (32), $|x| > R$ for any $x \in B_{x_n}$, hence $k(x)=0$ for $x \in B_{x_n}$. But, this happens if

$$c^n > \gamma(R\,|x_0|^{-1} + \mu). \tag{35}$$

For such an n, we have

$$\int_{B_{x_0}} |k(x)|\,dx \leqq n A(k). \tag{36}$$

Now, if $|x_0|$ is of order $2^{-j}R$, then the right hand side of (35) is of order 2^j, so we can choose an integer n in (35) which is of order j.

Let $\Gamma_j = \{x \in \omega : 2^{-j-1} R \leqq |x| < 2^{-j} R\}$. Using dilations, one can show that there is a certain number N, such that each Γ_j may be covered by N "balls" of the form $B_m^j = \{x \in \omega : |x - x_m^j| < \mu |x_m^j|\}$ with $x_m^j \in \Gamma_j$. Hence, by (36),

$$\int_{\Gamma_j} |k(x)|\,dx \leqq \sum_{m=1}^{N} \int_{B_m^j} |k(x)|\,dx \leqq Cj A(k). \tag{37}$$

Take now $\alpha>0$. We have, since $\operatorname{supp} k\subset B_R$,

$$\begin{aligned}\int_{\omega}|k(x)|\,|x|^{\alpha}\,dx&=\sum_{j\geqq 0}\int_{\Gamma_j}|k(x)|\,|x|^{\alpha}\,dx\\&\leqq\sum_{j\geqq 0}2^{-j\alpha}R^{\alpha}\int_{\Gamma_j}|k(x)|\,dx\\&\leqq CA(k)R^{\alpha}\sum_{j\geqq 0}j2^{-j\alpha}\\&=C_{\alpha}A(k)R^{\alpha}.\end{aligned}$$

Let us now prove Proposition 2.3.

Let $\chi_1\in C_0^{\infty}(\varkappa_0)$ such that $0\leqq\chi_1\leqq 1$, $\chi_1(z)=1$ if $|z|\leqq\rho_1$ with $\rho_1>0$, let $\chi_2^+\in C_0^{\infty}(\mathbb{R}^+)$ such that $0\leqq\chi_2^+\leqq 1$, $\chi_2^+(r)=1$ if $0\leqq r\leqq\rho_2$, with $\rho_2>0$, and let $\chi_2(t)=\chi_2^+(|t|')$ for $t\in\varkappa_0$. Write $\chi(z,t)=\chi_1(z)\,\chi_2(t)$ for $(z,t)\in\omega$. Then, obviously $\chi\in C_0^{\infty}(\omega)$, and, since $|\ |$ and $|\ |'$ are equivalent on $\varkappa_0$, we can choose ρ_1 and ρ_2 big enough such that $\chi(x)=1$ if $|x|\leqq 1$, which we will suppose from now on.

We write $\chi_R=\chi\circ D_{R^{-1}}$ for $R>0$, and decompose k as

$$k=k'+k'',$$

where

$$k'=k\chi_{\beta\gamma|\lambda_0|^{-1}}. \tag{38}$$

By Lemma 2.2, k' is a Calderón-Zygmund kernel with

$$A(k')\leqq CA_1(k), \tag{39}$$

and

$$\operatorname{supp} k'\subset B_{\rho\beta\gamma|\lambda_0|^{-1}}, \tag{40}$$

if $\rho>1$ is chosen such that $\operatorname{supp}\chi\subset B_{\rho}$.

Now, after remark 3 there exists a bounded Radon measure $d\mu(t)$ on $\Sigma'=\{t\in z_0\colon |t|'=1\}$, such that for any $f\in L^1(\varkappa_0)$

$$\int_{\varkappa_0}f(t)\,dt=\int_0^{\infty}\int_{\Sigma'}f(D_r t)\,r^{p-1}\,d\mu(t)\,dr, \tag{41}$$

where $p=\Sigma\alpha_j\dim(\varkappa_0\cap\nu_{\alpha_j})$ denotes the homogeneous dimension of $\varkappa_0$.

Hence, for a.e. $z\in\nu_0$

$$\begin{aligned}&\int_{\varkappa_0}k'(z,t)\,dt\\&\quad=\chi_1(D_{1/(\beta\gamma|\lambda_0|^{-1})}z)\int_0^{\infty}\chi_2^+((\beta\gamma|\lambda_0|^{-1})^{-1}r)\,k(z,D_r t)\,r^{p-1}\,d\mu(t)\,dr.\end{aligned}$$

But, (29) implies that for any $\psi^+\in C_0^{\infty}(\mathbb{R}^+)$ with $0\notin\operatorname{supp}\psi^+$

$$0=\int_0^{\infty}\int_{\Sigma'}k(z,D_r t)\,\psi^+(r)\,r^{p-1}\,d\mu(t)\,dr,$$

and therefore, for a.e. $z\in\nu_0$

$$\int_{\varkappa_0}k'(z,t)\,dt=0.$$

So we get

$$k'_{\lambda_0}(z)=\int_{\varkappa_0} k'(z,t)(e^{-i\langle\lambda_0,t\rangle}-1)\,dt. \tag{42}$$

Now, if $t=\sum_j t_j$ with $t_j\in\varkappa_0\cap\mathcal{N}_{\alpha_j}$, and $\lambda_0=\sum_j\lambda_{0,j}$ with $\lambda_{0,j}\in\varkappa^*_{0,\alpha_j}$, then

$$\begin{aligned}|e^{-i\langle\lambda_0,t\rangle}-1|&\leqq|\langle\lambda_0,t\rangle|\\&=|\sum_j\langle\lambda_{0,j},t_j\rangle|\leqq\sum_j|\lambda_{0,j}|^*_{\alpha_j}|t_j|_{\alpha_j}=\sum_j|\lambda_{0,j}|^{\alpha_j}|t_j|^{\alpha_j},\end{aligned}$$

which, together with (42), yields

$$\begin{aligned}\int_{\varkappa_0}|k'_{\lambda_0}(z)|\,dz&\leqq\sum_j|\lambda_{0,j}|^{\alpha_j}\iint|k'(z,t)|\,|t_j|^{\alpha_j}\,dt\,dz\\&\leqq\sum_j|\lambda_{0,j}|^{\alpha_j}\int|k'(x)|\,|x|^{\alpha_j}\,dx.\end{aligned}\tag{43}$$

By Lemma 2.4 and (39), (40),

$$\int|k'(x)|\,|x|^{\alpha_j}\,dx\leqq C_{\alpha_j}A_1(k)(\beta\gamma\rho|\lambda_0|^{-1})^{\alpha_j}\leqq C'A_1(k)|\lambda_0|^{-\alpha_j},$$

hence

$$\int_{\varkappa_0}|k'_{\lambda_0}(z)|\,dz\leqq C'A_1(k)\sum_j|\lambda_{0,j}|^{\alpha_j}|\lambda_0|^{-\alpha_j}\leqq CA_1(k). \tag{44}$$

Next we choose $s\in\varkappa_0$ such that $\langle\lambda_0,s\rangle=1$ and $|s|\leqq\gamma|\lambda_0|^{-1}$ (compare (14)), and define

$$H(z,t)=k''(z,t-s)-k''(z,t).$$

We claim that H is in fact an integrable function on ω:

Since by (39) k'' is a Calderón-Zygmund kernel with $A(k'')\leqq CA_1(k)$, we have

$$\iint_{|(z,t)|>\beta|s|}|H(z,t)|\,dz\,dt\leqq CA_1(k). \tag{45}$$

Suppose now $|(z,t)|\leqq\beta|s|$. Then $|(z,t)|\leqq\beta\gamma|\lambda_0|^{-1}$, hence, by (38), $k''(z,t)=0$, i.e. $H(z,t)=k''(z,t-s)$. Therefore,

$$\begin{aligned}\iint_{|(z,t)|\leqq\beta|s|}|H(z,t)|\,dz\,dt&=\iint_{|(z,t)|\leqq\beta|s|}|k''(z,t-s)|\,dz\,dt\\&\leqq\int_{|z|\leqq\beta|s|}\int_{|t+s|\leqq\beta|s|}|k''(z,t)|\,dz\,dt.\end{aligned}$$

But, if $|z|\leqq\beta|s|$ and $|t+s|\leqq\beta|s|$, then

$$|(z,t)|\leqq\gamma(|z|+|t+s-s|)\leqq\gamma(|z|+\gamma|t+s|+\gamma|s|)\leqq\tau|\lambda_0|^{-1}$$

with

$$\tau=\gamma^2(\beta+\gamma+\beta\gamma).$$

Moreover, $k''(z,t)=0$, if $|(z,t)|\leqq\beta\gamma|\lambda_0|^{-1}$. It follows that

$$\iint\limits_{|(z,t)|\leqq\beta|s|}|H(z,t)\,dz\,dt\leqq\int\limits_{\beta\gamma|\lambda_0|^{-1}<|x|<t|\lambda_0|^{-1}}|k(x)|\,dx\leqq CA_1(k) \tag{46}$$

by (H1)′.

So, from (45) and (46) we derive

$$|H|_{L^1(\omega)}\leqq CA_1(k). \tag{47}$$

An argument analogues to the one used in the proof of Lemma 2 in [13] shows that

$$k''_{\lambda_0}(z)=(e^{-i\langle\lambda_0,s\rangle}-1)^{-1}\int\limits_{\varkappa_0}H(z,t)\,e^{-i\langle\lambda_0,t\rangle}\,dt.$$

But, since $\langle\lambda_0,s\rangle=1$, this, together with (47), immediately gives us that

$$\int\limits_{\varkappa_0}|k''_{\lambda_0}(z)|\,dz\leqq|(e^{-i}-1)^{-1}|\,|H|_{L^1(\omega)}\leqq C'A_1(k). \tag{48}$$

Together with (44), this finishes the proof.

Proposition 2.5. *If k is a Calderón-Zygmund kernel on ω having property* (H2), *then for almost every $\lambda_0\in\varkappa_0^*$ k_{λ_0} is a Calderón-Zygmund kernel having property* (H2) *on $\varkappa_0$, and there exists some constant $C>0$, independent of k and λ_0, such that $A_2(k_{\lambda_0})\leqq CA_2(k)$.*

Proof. By Lemma 2.1, k_{λ_0} is a Calderón-Zygmund kernel such that $A(k_{\lambda_0})\leqq A(k)$ for a.e. $\lambda_0\in\varkappa_0^*$. So, it remains to prove property (H2) for k_{λ_0}.

Let $R>0$. By Minkowsky's inequality we have

$$\Big(\int\limits_{|z|>R}|k_{\lambda_0}(z)|^2\,dz\Big)^{1/2}\leqq\int\limits_{\varkappa_0}\Big(\int\limits_{|z|>R}|k(z,t)|^2\,dz\Big)^{1/2}dt. \tag{49}$$

One should realize here that we assumed that k_{λ_0} agrees with the vector-valued integral $\int k(\cdot,t)e^{-i\langle\lambda_0,t\rangle}\,dt$ as an element of the Hilbert space $L^2(\varkappa_0,d\mu_R)$, where $d\mu_R$ is the measure on $\varkappa_0$ which is defined by $\int\varphi\,d\mu_R=\int\limits_{|z|>R}\varphi(z)\,dz$. But, since the right hand side of the inequality (49) will turn out to be finite, one can show that in the fact our assumption is true for almost every $\lambda_0\in\varkappa_0^*$.

Let

$$h(t)=\Big(\int\limits_{|z|>R}|k(z,t)|^2\,dz\Big)^{1/2}.$$

After Fubini's theorem, h is Lebesgue measurable. Let m denote the homogeneous dimension of ω. From (H_2) we conclude that for any $\varepsilon>0$

$$\begin{aligned}\int\limits_{|t|>\varepsilon}h^2(t)\,dt&=\int\limits_{|t|>\varepsilon}\int\limits_{|z|>R}|k(z,t)|^2\,dz\,dt\\&\leqq\int\limits_{|(z,t)|>\gamma_1(\varepsilon+R)}|k(z,t)|^2\,dz\,dt\\&\leqq CA_2(k)^2(\varepsilon+R)^{-m},\end{aligned} \tag{50}$$

since after (11) $|z|>R$ and $|t|>\varepsilon$ imply

$$|(z,t)|\geqq\gamma_1(|z|+|t|)\geqq\gamma_1(\varepsilon+R).$$

So, for $\varepsilon>0$ and $n\in\mathbb{N}_0$, by Schwartz' inequality

$$\int\limits_{2^n\varepsilon<|t|<2^{n+1}\varepsilon} h(t)\,dt\leqq\Big(\int\limits_{2^n\varepsilon<|t|<2^{n+1}\varepsilon} h^2(t)\,dt\Big)^{1/2}\Big(\int\limits_{2^n\varepsilon<|t|<2^{n+1}\varepsilon} dt\Big)^{1/2}$$

$$\leqq C^{1/2}A_2(k)(R+2^n\varepsilon)^{-\frac{m}{2}}\Big(\int\limits_{1<|t|<2} dt(2^n\varepsilon)^p\Big)^{1/2},$$

where $p=\sum\limits_j \alpha_j\dim(\varkappa_0\cap\nu_{\alpha_j})$ denotes the homogeneous dimension of $\varkappa_0$. So, for some new constant $C>0$ we have

$$\int\limits_{2^n\varepsilon<|t|<2^{n+1}\varepsilon} h(t)\,dt\leqq CA_2(k)\frac{(2^n\varepsilon)^{p/2}}{(R+2^n\varepsilon)^{m/2}},$$

hence

$$\int\limits_{\varepsilon<|t|} h(t)\,dt=\sum_{n\geqq 0}\int\limits_{2^n\varepsilon<|t|<2^{n+1}\varepsilon} h(t)\,dt$$

$$\leqq CA_2(k)\sum_{n\geqq 0}\frac{(2^n\varepsilon)^{p/2}}{(R+2^n\varepsilon)^{m/2}}. \tag{51}$$

Let $\sigma=\log_2\dfrac{R}{\varepsilon}$. Then

$$\sum_{n\geqq 0}\frac{(2^n\varepsilon)^{p/2}}{(R+2^n\varepsilon)^{m/2}}$$

$$\leqq\sum_{n\leqq\sigma}\frac{(2^n\varepsilon)^{p/2}}{R^{m/2}}+\sum_{n>\sigma}\frac{(2^n\varepsilon)^{p/2}}{(2^n\varepsilon)^{m/2}}$$

$$\leqq R^{-m/2}\varepsilon^{p/2}\frac{2^{(\sigma+1)p/2}-1}{2^{p/2}-1}+\varepsilon^{(p-m)/2}2^{\sigma(p-m)/2}\frac{1}{1-2^{(p-m)/2}}$$

$$\leqq C'R^{\frac{p-m}{2}},$$

where C' is independent of R and ε.

Hence, by (51),

$$\int\limits_{\varkappa_0} h(t)\,dt\leqq CC'A_2(k)R^{\frac{p-m}{2}},$$

and so, by (49),

$$\int\limits_{|z|>R}|k_{\lambda_0}(z)|^2\,dz\leqq\Big(\int\limits_{\varkappa 0} h(t)dt\Big)^2\leqq C''A_2(k)^2R^{-(m-p)}. \tag{52}$$

But, $m-p$ is just the homogeneous dimension of ν_0. So, (52) proves that $A_2(k_{\lambda_0})\leqq CA_2(k)$ for some $C>0$.

3. The case where $\varkappa_0 = \omega \cap \varkappa \neq 0$

In case that $\varkappa_0 \neq 0$, we have the following application of Proposition 2.3:

Theorem 3.1. *Let k be a Calderón-Zygmund kernel on ω having property* (H1), *and assume that for a.e. $z \in \nu_0$ and every a, b with $0 < a < b$,*

$$\int_{a<|t|'<b} k(z,t)\,dt = 0. \tag{29}$$

Let K denote the extension of k to ν as described in (22), *i.e. $K = k \otimes \delta$. Then the operator T: $C_0^\infty(N) \to C^\infty(N)$ given by $Tf = f * K$ extends to a bounded operator from $L^2(N)$ to $L^2(N)$, and $\|T\| \leqq CA_1(k)$.*

Proof. By (16) we have for $f \in C_0^\infty(\nu)$

$$|Tf|^2_{L^2(\nu)} = \int_{\varkappa^*} |(Tf)_\lambda|^2_{L^2(\nu_0)}\,d\lambda, \tag{53}$$

and (19) yields

$$(Tf)_\lambda = f_\lambda *_\lambda K_\lambda. \tag{54}$$

Now, if we write $\lambda = (\lambda_0, \lambda_1)$ with $\lambda_0 \in \varkappa_0^*$, $\lambda_1 \in \tilde{\varkappa}^*$ (compare (4)), then for $\xi \in \nu_0^*$

$$\hat{K}_\lambda(\xi) = \hat{K}(\xi, \lambda_0, \lambda_1) = \hat{k}(\xi, \lambda_0) = \hat{k}_{\lambda_0}(\xi),$$

hence

$$K_\lambda = k_{\lambda_0}. \tag{55}$$

But, after Proposition 2.3, $|k_{\lambda_0}|_{L_1(\nu_0)} \leqq CA_1(k)$, and so

$$\begin{aligned} &|f_\lambda *_\lambda K_\lambda|_{L^2(\nu_0)} \\ &\quad = |f_\lambda *_\lambda k_{\lambda_0}|_{L^2(\nu_0)} \leqq \| |f_\lambda| * |k_{\lambda_0}| \|_{L^2(\nu_0)} \\ &\quad \leqq |k_{\lambda_0}|_{L^1(\nu_0)} |f_\lambda|_{L^2(\nu_0)} \leqq CA_1(k) |f_\lambda|_{L^2(\nu_0)}, \end{aligned}$$

which finally yields

$$|Tf|^2_{L^2(\nu)} \leqq \int C^2 A_1(k)^2 |f_\lambda|^2_{L^2(\nu_0)}\,d\lambda = C^2 A_1^2(k) |f|^2_{L^2(\nu)},$$

i.e. $\|T\| \leqq CA_1(k)$.

Remarks. In [13] it has been proved that in case $\omega = \nu$ we can dispense with property (H1) and (29) in Theorem 3.1. Moreover, in the next sections we will show that for a certain class of groups we can also omit property (29), if we replace property (H1) by (H2). So, it is not clear whether property (29) is superfluous at all. Moreover, it is an open question whether the operator T is bounded in general if $\omega \cap \varkappa = 0$. However, for "completely square integrable" groups this is true, as will also be shown in the following sections.

4. Nilpotent Lie groups Having Square Integrable Irreducible Unitary Representations

Let us first give a short summary of some well-known facts about square integrable unitary representations of nilpotent Lie groups, which all can be

found in [10]. For convenience, we will state these results only for the special groups considered above and we will use the same notation:

Let π be a unitary factor representation of N on a Hilbert space $\mathfrak{H}_\pi$. Then there exists a unitary character χ of Z with $\pi(t)=\chi(t)\cdot\mathbb{1}$ for $t\in Z$. Thus the absolute values of the coefficients $c_{\xi,\eta}(x)=\langle\pi(x)\xi,\eta\rangle$ for $\xi,\eta\in\mathfrak{H}_\pi$ of π may be considered as functions on N/Z. So, if we again parametrize N by $\mathfrak{n}$, and identify $\mathfrak{n}/\mathfrak{z}$ with $\mathfrak{n}_0$, then we say that π is *square integrable*, if for some nonzero coefficient $c_{\xi,\eta}$

$$\int_{\mathfrak{n}_0} |c_{\xi,\eta}(z)|^2\,dz<\infty. \tag{56}$$

It is known (see [10], Theorem B) that if π is irreducible and square integrable, then there exists a constant $d(\pi)>0$, called the *formal degree* of π, such that

$$\int_{\mathfrak{n}_0} c_{\xi_1,\eta_1}(z)\overline{c_{\xi_2,\eta_2}}(z)\,dz=d(\pi)^{-1}\langle\xi_1,\xi_2\rangle\overline{\langle\eta_1,\eta_2\rangle} \tag{57}$$

for any $\xi_1, \xi_2, \eta_1, \eta_2\in\mathfrak{H}_\pi$.

One should realize that $d(\pi)$ of course depends on the choice of the Haar measure dz on $\mathfrak{n}_0$.

According to the Kirillov theory (see [6]), the irreducible unitary representation of N are parametrized by the space of N-orbits in the dual space $\mathfrak{n}^*$ of $\mathfrak{n}$ under the coadjoint action. We shall write π_λ for the representation corresponding to an orbit $\mathfrak{o}$ containing $\lambda\in\mathfrak{n}^*$, and by b_λ we denote the skew symmetric bilinear form

$$b_\lambda(x,y)=\langle\lambda,[x,y]\rangle$$

on $\mathfrak{n}$. Since the singular subspace of b_λ clearly contains $\mathfrak{z}$, b_λ may be viewed as a skew two-form on $\mathfrak{n}/\mathfrak{z}\simeq\mathfrak{n}_0$. Finally, by χ_λ we will denote the character $\chi_\lambda(t)=\exp(i\langle\lambda,\log t\rangle)$ on Z, and the representation which is induced by χ_λ on N will be written as $\mathrm{ind}_Z^N\chi_\lambda$. If $\mathfrak{z}^\perp$ denotes the annihilator of $\mathfrak{z}$ in $\mathfrak{n}^*$, then Theorem 1 in [10] reads as follows:

Theorem A (Moore, Wolf). *For a linear functional λ on $\mathfrak{n}$ with orbit $\mathfrak{o}$ and corresponding representation $\pi=\pi_\lambda$ the following conditions are equivalent:*

i) *π is square integrable.*

ii) *$\mathrm{ind}_Z^N\chi_\lambda$ is a factor representation.*

iii) *$\mathfrak{o}=\lambda+\mathfrak{z}^\perp$.*

iv) *b_λ is nondegenerate on $\mathfrak{n}_0$.*

Now, we choose a fixed volume element α on $\mathfrak{n}_0$, and take the Haar measure dz on $\mathfrak{n}_0$ to be the measure defined by α. We define the Pfaffian $P(\lambda)$ of b_λ as follows:

Let r be the dimension of $\mathfrak{n}_0$. If r is odd, then let $P(\lambda)=0$, and if $r=2l$ is even, then we define $P(\lambda)$ by $b_\lambda^l=P(\lambda)\alpha$, where b_λ^l denotes the l-th exterior power of the restriction of b_λ to $\mathfrak{n}_0$. So, b_λ is nondegenerate on $\mathfrak{n}_0$ if and only if $P(\lambda)\neq 0$. Moreover, from Theorem A it follows easily that $P(\lambda)$ in fact depends only on the restriction of λ to $\mathfrak{z}$. Therefore, we may consider P as a function on $\mathfrak{z}^*$, and the theorem shows that the square integrable irreducible

representations of N are in one-to-one correspondence to the set of functionals $\lambda \in \mathfrak{z}^*$ with $P(\lambda) \neq 0$.

It is not surprising that P is homogeneous with respect to the dilations D_r. In fact, since D_r is an automorphism of $\mathfrak{n}$, one easily shows that $D_r b_\lambda = b_{D_r \lambda}$, and so

$$P(D_r\lambda)\alpha = b^l_{D_r\lambda} = D_r(b^l_\lambda) = P(\lambda)D_r\alpha = (\det D_r|_{\mathfrak{n}_0})P(\lambda)\alpha,$$

where D_r also denotes the canonical extension of D_r to the exterior algebra of $\mathfrak{n}_0$. Since

$$\det D_r|_{\mathfrak{n}_0} = r^q,$$

where q denotes the homogeneous dimension of $\mathfrak{n}_0$, we get

$$P(D_r\lambda) = r^q P(\lambda). \tag{58}$$

The Pfaffian on $\mathfrak{z}^*$ and the formal degree of square integrable representations are related as follows:

Theorem B (Moore, Wolf). *If for $\lambda \in \mathfrak{z}^*$ with $P(\lambda) \neq 0$ the corresponding irreducible representation of N is denoted by π_λ, then the formal degree of π_λ is $d(\pi_\lambda) = P((2\pi)^{-1}\lambda)$.*

The factor $(2\pi)^{-1}$ here stems from another normalization of the Fourier transform than in [10].

From these results, we draw the following conclusion:

Corollary 4.1. *Let $\lambda \in \mathfrak{z}^*$ with $P(\lambda) \neq 0$. Then, for any f and g in $L^2(\mathfrak{n}_0)$,*

$$|f *_\lambda g|_{L^2(\mathfrak{n}_0)} \leqq |P((2\pi)^{-1}\lambda)|^{-1/2} |f|_{L^2(\mathfrak{n}_0)} |g|_{L^2(\mathfrak{n}_0)}. \tag{59}$$

Proof. Since N is of type I, Theorem A shows that $\sigma = \operatorname{ind}_Z^N \chi_\lambda$ is unitary equivalent to a multiple of π_λ. This, together with (57), easily implies that

$$\int_{\mathfrak{n}_0} |\langle \sigma(z)\xi, \eta\rangle|^2 \, dz \leqq d(\pi_\lambda)^{-1} |\xi|^2 |\eta|^2 \tag{60}$$

for any $\xi, \eta \in \mathfrak{H}_\sigma$.

But, if f and g are in $L^2(\mathfrak{n}_0)$, then the functions $\xi(z,t) = f(z)\bar{\chi}_\lambda(t)$ and $\eta(z,t) = g(z)\bar{\chi}_\lambda(t)$ are in $\mathfrak{H}_\sigma$ by the definition of induced representations, and $|\xi|_{\mathfrak{H}_\sigma} = |f|_{L^2(\mathfrak{n}_0)}$, $|\eta|_{\mathfrak{H}_\sigma} |g|_{L^2(\mathfrak{n}_0)}$. Moreover, an easy computation, using (5)-(8), shows that for $z \in \mathfrak{n}_0$

$$\langle \sigma(z)\xi, \eta\rangle = \langle \xi, \sigma(z^{-1})\eta\rangle = \int_{\mathfrak{n}_0} \bar{g}(zw^{-1}) f(w^{-1}) e^{i\langle \lambda, P(z,-w)\rangle} \, dw,$$

since, by the definition of induced representations,

$$(\sigma(z^{-1})\eta)(w,t) = \eta((z,0)(w,t)).$$

So, if we write $\check{f}(w) = f(w^{-1})$, then, by (18),

$$\langle \sigma(z)\xi, \eta\rangle = (\bar{g} *_\lambda \check{f})(z),$$

and (60) reads as

$$|\bar{g} *_\lambda \check{f}|_{L^2(\mathfrak{n}_0)} \leqq d(\pi_\lambda)^{-1/2} |f|_{L^2(\mathfrak{n}_0)} |g|_{L^2(\mathfrak{n}_0)}.$$

But, since $|\check{f}|_{L^2(\pi_0)}=|f|_{L^2(\pi_0)}$ and $|\bar{g}|_{L^2(\pi_0)}=|g|_{L^2(\pi_0)}$, this together with Theorem B implies (59).

5. An Estimate for Twisted Convolutions

Let N be a nilpotent Lie group as above, but not necessarily having dilations. Let $X_1,\dots,X_n$ be a Jordan-Hölder basis for the Lie algebra $\mathfrak{n}$ of N, such that $[X,X_j]$ is always contained in the span $\mathfrak{g}_{j-1}$ of $X_1,\dots,X_{j-1}$ for any $X\in\mathfrak{n}$, and such that $\mathfrak{n}=\mathfrak{g}_r$ for some r.

We consider $X\in\mathfrak{n}$ as a left-invariant differential operator on N, and define a left invariant differential operator D on N as

$$D=(X_1+1)(X_2+1)\dots(X_n+1). \tag{61}$$

Similarly, if $G_j=\exp_N\mathfrak{g}_j$ denotes the closed subgroup of N corresponding to $\mathfrak{g}_j$, and if we consider X_i for $i\leqq j$ also as a left-invariant differential operator on G_j, we define D_j on G_j to be $D_j=(X_1+1)(X_2+1)\dots(X_j+1)$.

In the following, we will use a well-known method to construct a fundamental solution for D (compare for example to the proof of Lemma 3.1 in [1]).

Let $h(t)=e^{-t}$, if $t\geqq 0$, and $h(t)=0$ for $t<0$. Then h is a fundamental solution for the differential operator $\frac{d}{dt}+1$ on $\mathbb{R}$. We will now define inductively fundamental solutions ω_j on G_j for D_j.

First, of course $\omega_1(\exp tX_1)=h(t)$ is a fundamental solution for $D_1=X_1+1$. So, let us assume that we already defined a fundamental solution ω_j for D_j on G_j. If we identify $\exp\mathbb{R}X_{j+1}$ with $\mathbb{R}$, and if we let $\mathbb{R}$ operate on G_j by $x^t=\exp(-tX_{j+1})X\exp(tX_{j+1})$, then we may write G_{j+1} as $\mathbb{R}XG_j$, with multiplication $(t,x)(s,y)=(t+s,x^s y)$. We define ω_{j+1} by

$$\omega_{j+1}(t,x)=h(t)\,\omega_j(x^{-t}).$$

Then

$$\begin{aligned}(X_{j+1}+1)\,\omega_{j+1}(t,x)&=\int_{\mathbb{R}}\omega_{j+1}((t,x)(-s,e))(X_{j+1}+1)\,\delta(s)\,ds\\&=\int_{\mathbb{R}}h(t-s)\,\omega_j(x^{-t})\left(\frac{d}{ds}+1\right)\delta(s)\,ds\\&=\left(\frac{d}{dt}+1\right)h(t)\,\omega_j(x^{-t})=\delta(t)\,\omega_j(x^{-t})=\delta(t)\,\omega_j(x),\end{aligned}$$

where e denotes the neutral element in G_j, δ the Dirac distribution on $\mathbb{R}\simeq\exp\mathbb{R}X_{j+1}$, and where the integrals have to be interpreted as applications of distributions. So, if we apply our induction hypothesis, then

$$D_{j+1}\,\omega_{j+1}(t,x)=D_j(\delta(t)\,\omega_j(x))=\delta(t)\,\delta(x).$$

Thus $\omega=\omega_n$ is a fundamental solution for D.

Moreover, it is not hard to see that there exist polynomial functions $p_j(t_n, t_{n-1}, \ldots, t_{j+1})$ for $j=1,\ldots,n-2$, not even depending on $t_1,\ldots,t_r$ (since $\varkappa$ is spanned by $X_1,\ldots,X_r$), such that

$$\begin{aligned}\omega(\exp t_n X_n \exp t_{n-1} X_{n-1} \ldots \exp t_1 X_1) \\ = h(t_n)\, h(t_{n-1}) \prod_{j=1}^{n-2} h(t_j + p_j(t_n, \ldots, t_{j+1})),\end{aligned} \tag{62}$$

see e.g. [12] as general reference.

Since we may choose the Haar measure dx on N such that

$$\int_N \varphi(x)\,dx = \int_{\mathbb{R}^n} \varphi(\exp t_n X_n \ldots \exp t_1 X_1)\,dt_1 \ldots dt_n,$$

(62) shows that

$$|\omega|_{L^p(N)} = |h|^n_{L^p(\mathbb{R})} \tag{63}$$

for any $p\in[1,\infty)$.

We again write (compare (15))

$$\begin{aligned}&\omega_\lambda(z_n, \ldots, z_{r+1}) \\ &= \int \omega(\exp(z_n X_n + \ldots + z_1 X_1)) e^{-i\langle \lambda, z_r X_r + \ldots + z_1 X_1\rangle}\, dz_r \ldots dz_1\end{aligned}$$

for $\lambda\in\varkappa^*$.

Since there exist polynomial functions $q'_j(z_n, z_{n-1}, \ldots, z_{j+1})$ for $j=1,\ldots,n-2$, also not depending on $z_1,\ldots,z_r$, such that

$$\begin{aligned}&\exp(z_n X_n + \ldots + z_1 X_1) = \exp(z_n X_n) \exp(z_{n-1} X_{n-1})\ x \\ &\exp((z_{n-2} + q'_{n-2}(z_n, z_{n-1})) X_{n-2}) \ldots \exp((z_1 + q'_1(z_n, \ldots, z_2)) X_1),\end{aligned}$$

(62) shows that

$$\omega(\exp(z_n X_n + \ldots + z_1 X_1)) = h(z_n)\, h(z_{n-1}) \prod_{j=1}^{n-2} h(z_j + q_j(z_n, \ldots, z_{j+1})),$$

where the q_j are polynomial functions having the same properties as the functions p_j or q'_j. So we have

$$\begin{aligned}\omega_\lambda(z_n, \ldots, z_{r+1}) = h(z_n)\, h(z_{n-1}) \ldots h(z_{r+1} + q_{r+1}(z_n, \ldots, z_{r+2})) \\ \cdot \prod_{j=1}^{r} \hat{h}(\langle \lambda, X_j\rangle) e^{i\langle \lambda, X_j\rangle q_j(z_n, \ldots, z_{r+1})};\end{aligned}$$

realize here that for $j=1,\ldots,r$ the function q_j is in fact a function $q_j(z_n, \ldots, z_{r+1})$.

If we now again write $N_0 = \exp \mathfrak{n}_0$, where $\mathfrak{n}_0$ denotes the span of $X_n, \ldots, X_{r+1}$, and endowe N_0 with the group structure from N/Z, then the same argument as above also shows that

$$|\omega_\lambda|_{L^p(N_0)} \leqq |\hat{h}|^r_{L^\infty(\mathbb{R})} |h|^{n-r}_{L^p(\mathbb{R})} \tag{65}$$

for every $\lambda \in \mathfrak{z}^*$, if Haar measure on N_0 is chosen appropriatly. But, $\hat{h}(\xi) = \dfrac{1}{1+i\xi}$ for every $\xi \in \mathbb{R}$, and thus $|\hat{h}|_{L^\infty(\mathbb{R})} \leqq 1$, and it is also easily seen that $|h|_{L^p(\mathbb{R}^n)} \leqq 1$ for every $p \in [1, \infty]$.

So, we have proved the following

Lemma 5.1. *There exists a left-invariant differential operator D on N, which has a fundamental solution ω, such that ω and ω_λ are in $L^p(N)$ and $L^p(N_0)$ resp., for all p with $1 \leqq p \leqq \infty$, and such that*

$$|\omega|_{L^p(N)} \leqq 1 \quad \text{and} \quad |\omega_\lambda|_{L^p(N_0)} \leqq 1$$

for all $\lambda \in \mathfrak{z}^*$ and $1 \leqq p \leqq \infty$.

In the following, for convenience of the notation we will assume that N is again the group which has been considered in the first sections.

For any $\lambda \in \mathfrak{z}^*$ and any compactly supported distribution $v \in \mathscr{E}'(\mathfrak{n}_0)$ we will denote by T_v^λ the twisted convolution operator

$$(T_v^\lambda \varphi)(z) = (\varphi *_\lambda v)(z) = \int_{\mathfrak{n}_0} \varphi(z w^{-1}) \omega_\lambda(z, -w)\, v(w)\, dw, \tag{66}$$

where the integral has again to be interpreted as application of a distribution.

Further, let G_λ denote the group which is $N_0 \times T$ as a manifold, where T denotes the circle group, and with multiplication

$$(z, \xi)(w, \eta) = (z w, \omega_\lambda(z w)\, \xi \eta). \tag{67}$$

We take the Haar measure on G_λ to be $dz\, d\xi$, where $d\xi$ on T is normed to have total mass one.

Let $A_p(G_\lambda)$ denote the Figa-Talamanca-Herz algebra of G_λ (see e.g. [2]), and let $\rho \in A_p(G_\lambda)$ be any function such that $\rho(z, \xi) = \xi$ for all z in a neighborhood of the support of v and all $\xi \in T$.

If $\|T_v^\lambda\|_p$ denotes the norm of T_v^λ as an operator on $L^p(N_0)$, then in [7] H. Leptin has shown that $\|T_v^\lambda\|_p$ can be estimated by the norm of the untwisted convolution operator T_v^0 as

$$\|T_v^\lambda\|_p \leqq |\rho|_{A_p(G_\lambda)} \|T_v^0\|_p, \tag{68}$$

for $1 < p < \infty$.

Proposition 5.2. *If v is a distribution with compact support on $N_0 \simeq \mathfrak{n}_0$, if $\chi \in C_0^\infty(N_0)$ is such that $\chi(z) = 1$ for all z in a neighborhood of* supp v, *and if D is a left-invariant differential operator on N as described in Lemma* 3, *then*

$$\|T_v^\lambda\|_p \leqq |(D\delta)_\lambda *_\lambda \chi|_{L^q(N_0)} \|T_v^0\|_p, \tag{69}$$

where δ denotes the Dirac distribution on N, and where $\dfrac{1}{q} = 1 - \dfrac{1}{p}$ for $1 < p < \infty$.

Proof. Let ^{d} denote a fundamental solution of D as described in Lemma 5.1. Then $D\,{}^d\delta = \delta$, or ${}^d\delta * (D\delta) = \delta$. So, after (19), since for any $\lambda \in \mathfrak{z}^*$ $\delta_\lambda = \delta$ on $N_0 \simeq \mathfrak{n}_0$,

$$\omega_\lambda *_\lambda (D\delta)_\lambda = \delta. \tag{70}$$

This yields

$$\chi = \delta *_\lambda \chi = \omega_\lambda *_\lambda ((D\delta)_\lambda *_\lambda \chi). \tag{71}$$

Now, let $f = \omega_\lambda$, $g = (D\delta)_\lambda *_\lambda \chi$, and let $\tilde{f}(z, \xi) = f(z)\xi$ and $\tilde{g}(z, \xi) = g(z)\xi$ on G_λ. A short calculation shows that $(\tilde{f} * \tilde{g})(z, \xi) = (f *_\lambda g)(z)\xi = \chi(z)\xi$, after (71).

So, we may choose $\rho = \tilde{f} * \tilde{g}$ in (68). But, then

$$|\rho|_{A_p(G_\lambda)} \leqq |\tilde{f}|_{L^p(G_\lambda)} |\check{\tilde{g}}|_{L^q(G_\lambda)} = |f|_{L^p(N_0)} |g|_{L^q(N_0)},$$

where, after Lemma 5.1, $|f|_{L^p(N_0)} = |\omega_\lambda|_{L^p(N_0)} \leqq 1$.

So, $|\rho|_{A_p}(G_\lambda) \leqq |(D\delta)_\lambda *_\lambda \chi|_{L^q(N_0)}$, which, together with (68), proves (69).

Remark. One could use Proposition 5.2 to give another proof of Satz 1 in [8].

6. Singular Kernels on Groups Having Square Integrable Representations

In the following we will only consider nilpotent Lie groups N such that all irreducible unitary representations which are not trivial on the center are square integrable. Such groups will be called *completely square integrable groups*. From the results of Moore and Wolf it follows that these are just the groups such that

$$P(\lambda) \neq 0 \quad \text{for all } \lambda \in \mathfrak{z}^* - \{0\}. \tag{72}$$

It is known that there are nilpotent Lie groups of any nilpotent degree having square integrable representation. For example, in [9] J. Ludwig gave concrete constructions of n-step nilpotent Lie groups with exclusively flat coadjoint orbits and one-dimensional center for any positive integer n, which are completely square integrable. Moreover, it is easily seen that these groups admit dilations. Also, all generalized Heisenberg groups (see [5]) are homogeneous completely square integrable groups, which may have centers of any positive dimension.

Let ω, $\mathfrak{z}_0, \ldots$, and k be as described in Sect. 1 (so especially $\omega + \mathfrak{z} = \mathfrak{n}$).

Then we have the following main result:

Theorem 6.1. *Assume that N is completely square integrable, and let k be a Calderón-Zygmund kernel on ω having property* (H2). *If $K = k \otimes \delta$ denotes the extension of k to $\mathfrak{n}$ as described in* (22), *and if T denotes the operator T: $C_0^\infty(N) \to C^\infty(N)$ given by $Tf = f * K$, then T extends to a bounded operator from $L^2(N)$ to $L^2(N)$, and $\|T\| \leqq C A_2(k)$.*

Proof. A look at the proof of Theorem 3.1 shows that it is enough to prove that there exists a constant $C > 0$, such that all twisted convolution operators $T^\lambda_{k_{\lambda_0}}$ on $\mathfrak{n}_0$, given by

$$T^\lambda_{k_{\lambda_0}}(\varphi) = \varphi *_\lambda k_{\lambda_0},$$

are bounded for all $\lambda = (\lambda_0, \lambda_1) \in \mathfrak{z}^*$, and $\|T^\lambda_{k_{\lambda_0}}\|_2 \leqq C A_2(k)$.

Here, we make the convention that $k_{\lambda_0} = k$, if $\mathfrak{z}_0 = \omega \cap \mathfrak{z} = 0$.

After Proposition 2.3, k_{λ_0} is a Calderón-Zygmund kernel having property (H2), and $A_2(k_{\lambda_0}) \leqq C A_2(k)$ for every $\lambda_0 \in \mathfrak{z}_0^*$.

So, it will be even enough to prove the following

Claim. There exists a constant $C'>0$, such that $\|T_h^\lambda\|_2 \leqq C' A_2(k)$ for any Calderón-Zygmund kernel h on $\mathfrak{n}_0$ with $A_2(h) \leqq C A_2(k)$.

Let $\varphi \in C_0^\infty(\mathfrak{n}_0)$, and h as above. Then, by (9), for any $\lambda \in \mathfrak{z}^*$ and $r>0$,

$$\begin{aligned}
\varphi *_{D_r\lambda} h(z) &= \int_{\mathfrak{n}_0} \varphi(zw^{-1})\, e^{i\langle D_r\lambda,\, P(z,\,-w)\rangle}\, h(w)\, dw \\
&= \int_{\mathfrak{n}_0} \varphi(zw^{-1})\, e^{i\langle \lambda,\, P(D_r z,\, -(D_r w))\rangle}\, h(w)\, dw \\
&= r^{-q} \int_{\mathfrak{n}_0} (\varphi \circ D_{r^{-1}})(D_r z \cdot w^{-1})\, e^{i\langle \lambda,\, P(D_r z,\, -w)\rangle}\, h \circ D_{r^{-1}}(w)\, dw \\
&= r^{-q}[(\varphi \circ D_{r^{-1}}) *_\lambda (h \circ D_{r^{-1}})](D_r z),
\end{aligned}$$

where q denotes the homogeneous dimension of $\mathfrak{n}_0$. This calculation shows that

$$\|T_h^{D_r\lambda}\|_2 = r^{-q}\, \|T^\lambda_{h \circ D_{1/r}}\|_2 \tag{73}$$

for any $\lambda \in \mathfrak{z}^*$ and $r>0$.

Let $\lambda \neq 0$, and let $\chi \in C_0^\infty(\mathfrak{n}_0)$ be such that $0 \leqq \chi \leqq 1$, $\chi(z)=1$ if $|z| \leqq 1$, and $\chi(z)=0$ if $|z| \geqq 2$. Define

$$h' = h(\chi \circ D_{|\lambda|}), \quad \text{and} \quad h'' = h - h'. \tag{74}$$

After Lemma 2.2, h' is a Calderón-Zygmund kernel with $A_2(h') \leqq C A_2(h) \leqq C' A_2(k)$. So, after (73),

$$\|T_h^\lambda\|_2 = |\lambda|^{-q}\, \|T^{\lambda'}_{h' \circ D_{1/|\lambda|}}\|_2, \tag{75}$$

where $\lambda' = D_{1/|\lambda|}\lambda$.

But, $h' \circ D_{1/|\lambda|} = (h \circ D_{1/|\lambda|})\chi$, hence $h' \circ D_{1/|\lambda|}(z)=0$ if $|z| \geqq 2$. So, if we write $\chi' = \chi \circ D_{1/3}$, then $\chi'=1$ on a neighborhood of the support of $h' \circ D_{1/|\lambda|}$, hence an application of Proposition 5.2 yields

$$\|T^{\lambda'}_{h' \circ D_{1/|\lambda|}}\|_2 \leqq |(D\delta)_{\lambda'} *_{\lambda'} \chi'|_{L^2(\mathfrak{n}_0)}\, \|T^0_{h' \circ D_{1/|\lambda|}}\|_2, \tag{76}$$

where D is a left-invariant differential operator on N as described in Lemma 5.1. We may write $(D\delta)(z,t)$ as

$$(D\delta)(z,t) = \sum_{\alpha,\beta} a_{(\alpha,\beta)}\, \delta^{(\alpha)}(z)\, \delta^{(\beta)}(t),$$

corresponding to the decomposition of $\mathfrak{n}$ as $\mathfrak{n}_0 \oplus \mathfrak{z}$, where the sum is a finite sum over multi-indices $\alpha \in \mathbb{N}_0^{\dim \mathfrak{n}_0}$ and $\beta \in \mathbb{N}_0^{\dim \mathfrak{z}}$, and where the $a_{(\alpha,\beta)}$ are constants. Thus,

$$(D\delta)_{\lambda'}(z) = \sum_{\alpha,\beta} a_{(\alpha,\beta)}(-i\lambda')^\beta\, \delta^{(\alpha)}(z).$$

Since $|\lambda'|=1$, it is clear that there exists a constant $a>0$, such that $|a_{(\alpha,\beta)}(-i\lambda')^{\beta}|\leqq a$ for all α and β, independently of $\lambda\in\mathfrak{z}^*$.

Moreover, if $z\in\operatorname{supp}\chi'$, then $|z|\leqq 6$. Since $P(z,w)$ is a polynomial function, these observations together make clear that

$$|(D\delta)_{\lambda'}*_{\lambda'}\chi'|_{L^2(\mathfrak{n}_0)}\leqq C \tag{77}$$

for some constant $C>0$, independently of $\lambda\in\mathfrak{z}^*-\{0\}$. From (75), (76) and (77) we get

$$\|T_{h'}^{\lambda}\|_2\leqq|\lambda|^{-q}\,C\,\|T^0_{h'\circ D_{1/|\lambda|}}\|_2=C\,\|T_{h'}^{\lambda}\|_2. \tag{78}$$

But, either by induction over the degree of $\mathfrak{n}$, or by applying Theorem 1 of [13], we may assume that $\|T_{h'}^0\|_2\leqq C A_2(h')\leqq C' A_2(k)$, since T_h^0 is just the operator of convolution with h' from the right on N_0.

So, finally we obtain

$$\|T_{h'}^{\lambda}\|_2\leqq C A_2(k). \tag{79}$$

To estimate $\|T_{h''}^{\lambda}\|_2$, we will apply Corollary 4.1. We have

$$\begin{aligned}|h''|^2_{L^2(\mathfrak{n}_0)}&=\int_{\mathfrak{n}_0}|1-\chi(D_{|\lambda|}z)|^2\,|h(z)|^2\,dz\\&\leqq\int_{|z|>|\lambda|^{-1}}|h(z)|^2\,dz\leqq A_2(h)^2\,|\lambda|^q.\end{aligned} \tag{80}$$

If we let $\varkappa=\inf\{|P((2\pi)^{-1}\lambda)|\colon|\lambda|=1\}$, then, by (72), $\varkappa>0$, since P is continuous. If we write $\lambda'=D_{1/|\lambda|}\lambda$, then (80) together with (58) and (59) yields

$$\begin{aligned}\|T_{h''}^{\lambda}\|_2&\leqq|P((2\pi)^{-1}\lambda)|^{-1/2}\,|h''|_{L^2(\mathfrak{n}_0)}\\&\leqq(|\lambda|^q\,|P((2\pi)^{-1}\lambda')|)^{-1/2}\,A_2(h)\,|\lambda|^{q/2}\\&\leqq C'\varkappa^{-1/2}\,A_2(k).\end{aligned} \tag{81}$$

Let us remark that we could have obtained (82) also without using the homogeneity of P, but instead by applying (76) also to h''.

(80) and (81) together yield

$$\|T_h^{\lambda}\|_2\leqq(C+C'\varkappa^{-1/2})\,A_2(k),\qquad\text{q.e.d.}$$

References

1. Corwin, L.: Criteria for solvability of left invariant operators on nilpotent Lie groups, preprint
2. Eymard, P.: Algèbres A_p et convoluteurs de L^p. Seminaire Bourbaki no. 367 (1969/70)
3. Geller, D., Stein, E.M.: Singular convolution operators on the Heisenberg group. Bull. of the A.M.S. **6**, (no. 1) 99-103 (1982)
4. Goodman, R.: Singular integral operators on nilpotent Lie groups. Ark. för Mat. **18**, 1-11 (1980)
5. Kaplan, A.: Fundamental solutions for a class of hypoelliptic PDE generated by composition of quadratic forms. Trans. of the A.M.S. **258**, 147-153 (1980)

6. Kirillov, A.A.: Unitary representations of nilpotent Lie groups. Uspehi Mat. Nauk **17**, (no. 4, 106) 57–110 (1962)
7. Leptin, H.: Another remark on twisted convolution. Rendiconti Mat. Circ. Palermo, to appear
8. Leptin, H., Müller, D.: Schranken für verschränkte Faltungsoperatoren. Math. Ann. **260**, 511–519 (1982)
9. Ludwig, J.: Der Satz von Helson-Reiter für spezielle nilpotente Lie-Gruppen. Math. Ann. **239**, 207–218 (1979)
10. Moore, C.C., Wolf, J.A.: Square integrable representations of nilpotent groups. Trans. of the A.M.S. **185**, 445–462 (1973)
11. Nagel, A., Stein, E.M.: Lectures on pseudo-differential operators: Regularity theorems and applications to nonelliptic problems, Math. notes Princeton University Press, 1979
12. Pukanszky, L.: Leçons sur les représentations des groupes, Monographies de la Société Mathématique de France, no. 2 Dunod. Paris, 1967
13. Ricci, F.: Calderón-Zygmund kernels on nilpotent Lie groups, Lecture notes in mathematics Vol. 908, pp. 217–227. Berlin-Heidelberg-New York: Springer 1982
14. Stein, E.M.: Singular Integrals and differentiability properties of functions, Princeton Univ. Press, 1970
15. Strichartz, R.: Singular integrals on nilpotent Lie groups, Proc. Amer. Math. Soc. **53**, 367–374 (1975)

Oblatum 10-III-1983

Invent. math. 73, 491–510 (1983)

Inventiones mathematicae

© Springer-Verlag 1983

The Monodromy Groups of Critical Points of Functions II

S.V. Chmutov

Chair of Differential Geometry, Department of Mathematics,
Moscow State University, Moscow, 117234, USSR

The monodromy group is one of the invariants of an isolated critical point of an analytic function. For the critical points of functions in an even number of variables this group is an infinite group of transformations of the lattice $\mathbb{Z}^\mu$ with a skew-symmetric form. The main result of the first part of this article [1] reduces the description of the monodromy group to the description of some finite group of automorphisms of some finite $\mathbb{Z}_{2\lambda}$-module, where this module and the natural number λ are determined by the skew-symmetric form. This article is devoted to the description of this finite group. We shall prove that there are invariant subsets of this finite group in the $\mathbb{Z}_{2\lambda}$-module, and any automorphism of the $\mathbb{Z}_{2\lambda}$-module, for which these sets are invariant belongs to the finite group (Theorem 1). For a critical point in two variables: $\lambda=1$; the $\mathbb{Z}_2$-module is the homology group of the nonsingular local level manifold with coefficients in $\mathbb{Z}_2$; and these invariant subsets are the following: 1) subset of all vanishing vectors; 2) subset of all nonvanishing vectors not belonging to kernel of skew-symmetric form and 3) each vector from the kernel of skew-symmetric form (see the Theorem 3). In the case of a critical point of a function in an arbitrary even number of variables these invariant subsets are defined by the skew-symmetric form and set of vanishing vectors in the homology group with coefficients in $\mathbb{Z}_2$.

In spite of an uniform formulation of the main result of this article (Theorem 1) for all critical points in an even number of variables the proof of this result consists of two, very nonsimilar cases. The first case consists of the critical points A_μ and D_μ. The second case is the main case. It contains all the other critical points. The critical point E_6 plays a surprising role in the proof of Theorem 1 in the main case. All critical points of the main case adjoin to E_6. In the purely arithmetical proof the critical point E_6 is used in the following way. Firstly, in the six-dimensional lattice corresponding to E_6 there is a symplectic basis which consists of vanishing vectors (Lemma 2). Therefore in the lattice corresponding to each critical point of the main case, there is a six-dimensional sublattice with this property. Secondly, the standard Dynkin diagram of E_6 has three "tails", two of which are sufficiently long. This

property is useful for the description of the vanishing vectors. B. Wajnryb perhaps used the existence of the three "tails" in an unobvious form (look for it in the proof of Proposition 2 in [2]). A.N. Varchenko has proved the Theorem 1 for the critical points A_μ and D_μ (1979, unpublished). This proof is given in addition to this article.

Sections 1–5 of this article are devoted to the formulations and discussions of results. In Sect. 1 we shall formulate the main theorem (Theorem 1) of this article. Section 2 is devoted to remarks to the main theorem. In Sect. 3 we shall describe the vanishing vectors. The case of critical points of functions in two variables is discussed in Sect. 4. Finally, Sect. 5 is devoted to the definition of invariant subsets and to the discussion of this definition. The proofs of results are given in Sects. 6–9.

This article and it's first part [1] are independent.

To regret, in text of the first part of this article [1] there are serious misprints. Namely: p. 124, second line from bottom instead "λ_{j-1} is a divisor of λ_j" read "λ_j is a divisor of λ_{j-1}"; p. 125, thirteenth line from top instead "$\mathrm{Aut}(\hat{H}_\Gamma/2\lambda_1 \hat{H}_\Gamma)$" read "$\mathrm{Aut}(\hat{H}_\Gamma/2\lambda_1 H_\Gamma)$".

1. Formulation of Results

1.1. Designations

Let f: $(\mathbb{C}^{2m},0)\to(\mathbb{C},0)$ be a germ of an analytic function with an isolated critical point at $0\in\mathbb{C}^{2m}$. Let H be the homology group $H_{2m-1}(X,\mathbb{Z})$ of a nonsingular local level manifold X the function f. According to Milnor's results [4] H is a lattice $\mathbb{Z}^\mu$, where μ is the Milnor number of the critical point. Let Φ be the intersection form on H. Φ is a skew-symmetric form because the number of variables is even. For $x\in H$ let λ_x be a multiplicity of the covector $\Phi(x,\cdot)$ (i.e. λ_x is a maximum of the natural numbers such that $(1/\lambda_x)\Phi(x,y)\in\mathbb{Z}$ for all $y\in H$). Let λ be the maximum of all numbers λ_x, such that the image of x in $H/\mathrm{Ker}\,\Phi$ is primitive (nondivisible). Let $\hat{H}$ be the sublattice in H consisting of all vectors x such that λ divides λ_x. It is obvious that $\hat{H}$ is a sublattice in H of the same dimension μ. Let $\hat{H}_{2\lambda}$ be the $\mathbb{Z}_{2\lambda}$-module $\hat{H}/2\lambda H$.

1.2. Formulation of the Main Result of the Article [1]

The monodromy group of f preserves the sublattice $\hat{H}$ because it preserves the form Φ. Consider the group of transformations of the sublattice $\hat{H}$ consisting of the restrictions of all transformations from the monodromy group on $\hat{H}$. The image of this group under the natural projection π: $\mathrm{Aut}(\hat{H})\to\mathrm{Aut}(\hat{H}_{2\lambda})$ will be called a $\mathbb{Z}_{2\lambda}$-monodromy group and will denoted by $M_{2\lambda}$.

Theorem (see [1]). *A linear transformation h: $H\to H$ belongs to the monodromy group iff h preserves the form Φ, h is the identity on the kernel Φ and $\pi(h|_{\hat{H}})$ belongs to the $\mathbb{Z}_{2\lambda}$-monodromy group $M_{2\lambda}$.*

This paper is devoted to the description of the group $M_{2\lambda}$

1.3. Formulation of the Main Result

Theorem 1. *An automorphism* $h\colon \hat{H}_{2\lambda} \to \hat{H}_{2\lambda}$ *belongs to the* $\mathbb{Z}_{2\lambda}$*-monodromy group* $M_{2\lambda}$ *iff each* $x \in \hat{H}_{2\lambda}$ *is equivalent (in the sense of Sect. 5) to* $h(x)$.

2. Remarks

2.1. In [2] B. Wajnryb has proved Theorem 1 for all critical points of functions in two variables with nondegenerate form Φ, excluding A_μ and D_μ. (More exactly see [2].) A.N. Varchenko has proved Theorem 1 for the critical points A_μ and D_μ. I have guessed the formulation of the Theorem 1 under the influence of the Varchenko's result and the article [2] of B. Wajnryb.

2.2. According to Ebeling's result (see [3]) for the critical point of functions in an odd number of variables the monodromy group is defined by the lattice with the according (symmetric) form. For the function in an even number of variables this is false. For example, the skew-symmetric form of critical points A_6 and E_6 are isomorphic, but the monodromy groups (as the groups of automorphism of a six-dimensional lattice) are not isomorphic. The classes of equivalent vectors, introduced in Sect. 5, are defined by the skew-symmetric form and by the set of vanishing vectors (mod 2). Therefore Theorem 1 involves that for the critical points of functions in an even number of variables the monodromy group is defined by the skew-symmetric form and by the set of vanishing vectors (mod 2).

2.3. If the automorphism $h\colon \hat{H}_{2\lambda} \to \hat{H}_{2\lambda}$ preserves the classes of equivalent vectors, then this automorphism preserves the induced intersection form. More exactly: let $\Phi_{2\lambda}$ be the form induced by Φ, i.e. for any $x, y \in \hat{H}_{2\lambda}$

$$\Phi_{2\lambda}(x, y) \equiv \Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y)) \bmod 2\lambda^2$$

where $\hat{p}\colon \hat{H} \to \hat{H}_{2\lambda}$ is the projection and $\hat{p}^{-1}(x)$ is any vector belonging to the inverse image of x.

Statement 1. *Let* $h\colon \hat{H}_{2\lambda} \to \hat{H}_{2\lambda}$ *be an automorphism such that any vector* $x \in \hat{H}_{2\lambda}$ *is equivalent (in the sense of Sect. 5) to it's image* $h(x)$, *then*

$$\Phi_{2\lambda}(h(x), h(y)) = \Phi_{2\lambda}(x, y)$$

for any two vectors $x, y \in \hat{H}_{2\lambda}$.

A proof of the Statement 1 is given in Sect. 6.1.

2.4. It's sufficient to verify the condition "x is equivalent to $h(x)$" (in Theorem 1) for a basis of $\hat{H}_{2\lambda}$. More exactly:

Theorem 2. *Consider an isolated critical point of an analytic function in an even number of variables, excluding critical points* A_μ *and* D_μ. *We fix any basis of* $\hat{H}_{2\lambda}$. *Then an automorphism* $h\colon \hat{H}_{2\lambda} \to \hat{H}_{2\lambda}$ *belongs to* $M_{2\lambda}$ *iff* h *preserves the form* $\Phi_{2\lambda}$, h *is the identity on the kernel* $\hat{p}\,\mathrm{Ker}\,\Phi$ *and* h *preserves the classes of equivalent vectors of this fixed basis of* $\hat{H}_{2\lambda}$.

A proof of Theorem 2 is given in Sect. 6.2. For the critical points A_μ and D_μ in the Theorem 2 a standard basis is considered instead of an arbitrary basis of $\hat{H}_{2\lambda}$. An exact formulation is given in addition (Theorem 5).

3. Vanishing Vectors (mod 2)

In this section we describe the vanishing vectors (mod 2) (i.e. the vanishing vectors in the homology group with coefficients in $\mathbb{Z}_2$) for all critical points excluding A_μ and D_μ. This description will be used in proof of Theorem 1.

Let H_2 be the $\mathbb{Z}_2$-lattice $H_2 = H/2H$ and let p: $H \to H_2$ be the natural projection. Let Φ_2: $H_2 \times H_2 \to \mathbb{Z}_2$ be the form induced by the form Φ, i.e.

$$\Phi_2(x, y) \equiv \Phi(p^{-1}(x), p^{-1}(y)) \bmod 2.$$

3.1. Wajnryb's Function

For the description of vanishing vectors we use the Wajnryb's function q: $H_2 \to \mathbb{Z}_2$. We shall prove that a vector $x \in H_2$ is a vanishing vector (mod 2) iff this vector does not belong to $\operatorname{Ker} \Phi_2$ and $q(x) = 1$. We shall define the function q in two equivalent ways.

Geometrical Way. Consider the Dynkin diagram of a critical point in a feebly distinguished basis $e_1, \ldots, e_\mu$. For any vector $a \in H$ we shall define the subdiagram Γ_a of the Dynkin diagram: the i-th vertice of Dynkin diagram belongs to the graph Γ_a iff the i-th coordinate of a in this feebly distinguished basis is odd; the i-th and j-th vertices of Γ_a are joined in Γ_a iff $\Phi(e_i, e_j)$ is odd. The Wajnryb's function is defined by the formula

$$q(p(a)) \equiv \chi(\Gamma_a) \bmod 2,$$

where $\chi(\Gamma_a)$ is the Euler characteristic of the graph Γ_a.

Algebraical Way (see [2]). Consider the feebly distinguished basis $e_1, \ldots, e_\mu$. The function q is defined by the formulas:

1) $q(p(e_i)) = 1$ for $1 \leqq i \leqq \mu$.
2) $q(x+y) = q(x) + q(y) + \Phi_2(x, y)$ for all $x, y \in H_2$.

In Sect. 7.1 we shall prove that the function q is defined only by the form Φ_2 and the set of vanishing vectors (mod 2). Hence q is undependent of the choice of a feebly distingueshed basis.

Proposition 1. *For all isolated critical points of functions in an even number of variables, excluding A_μ and D_μ, a vector $x \in H_2$ is a vanishing vector* (mod 2) *iff* $x \notin \operatorname{Ker} \Phi_2$ *and* $q(x) = 1$.

The main difference between this Proposition and the Proposition 2 of article [2] consists in the existence of a kernel $\operatorname{Ker} \Phi_2$. The proof of Proposition 1 is similar to the proof of Proposition 2 of article [2], but it isn't coincident. (For example, in the cases C) and D) of the proof of Lemma 2 in

[2] B. Wajnryb used that Φ_2 is non degenerated. In the proof of Proposition 1 in this situation we'll show that the cases C) and D) are posible only for the critical points A_μ and D_μ). The complete proof of Proposition 1 is very long and therefore we'll omit this proof.

3.2. Remarks

1. For the critical points A_μ and D_μ the conditions $x \notin \operatorname{Ker} \Phi_2$ and $q(x)=1$ are necessary but not sufficient for x to be a vanishing vector. The vanishing vectors for these critical points are described in addition.

2. The condition $x \notin \operatorname{Ker} \Phi_2$ is an essential condition. For example, in the lattice H_2 corresponding to E_7 there is a vector x such that $q(x)=1$ and $x \in \operatorname{Ker} \Phi_2$.

4. Functions in Two Variables

For a critical point of a function in two variables all results about the monodromy have a more simple formulations. Some of them are given in this section. Exactly, in this section we'll describe the classes of equivalent vectors and the monodromy group for the critical point in two variables.

For the critical points in two variables $\lambda=1$ (see [1], [2]) and the $\mathbb{Z}_2$-module $\hat{H}_2$ coincides with the $\mathbb{Z}_2$-lattice H_2.

Theorem 3. *For a critical points in two variables the classes of equivalent vectors are the following subsets: subset of all vanishing vectors* (mod 2); *subset of all nonvanishing vectors not belonging to the kernel of* Φ_2 *and each vector from the kernel of* Φ_2.

This theorem simply follows from the definition of the equivalence relation of vectors (see Sect. 5). The proof of this theorem is given in Sect. 7.2.

Theorem 4. *For a critical point in two variables the transformation* h: $H \to H$ *belongs to the monodromy group iff* h *preserves the form* Φ, h *is the identity on the kernel of* Φ *and* h *preserves the subset of vanishing vectors* (mod 2).

Theorem 4 obviously follows from Theorems 1, 3 and the main theorem of article [1] (see Sect. 1.2).

Remarks. 1. From the Theorem 2 follows that for all critical points of functions in two variables excluding the critical points A_μ and D_μ it's sufficient to verify the condition "h transforms a vanishing vector into a vanishing vector" for a basis of H_2 consisting of vanishing vector (mod 2).

2. Theorems 3 and 4 are valid also for critical points with an unimodular intersection form on $H/\operatorname{Ker} \Phi$, i.e. with $\lambda=1$.

5. Definition of Invariant Sets

In this section we give the definition of equivalent vectors.

Let $\mathbf{i}: H_2 \to \hat{H}_{2\lambda}$ be defined by the formula $\mathbf{i}(p(a)) = \hat{p}(\lambda a)$ for each $a \in H$. Clearly, $\operatorname{Ker} \mathbf{i} = 0$. Let $\hat{\Phi}: \hat{H}_{2\lambda} \oplus H_2 \to \mathbb{Z}_2$ be defined by the formula

$$\hat{\Phi}(\hat{p}(a), p(b)) \equiv \frac{1}{\lambda} \Phi(a, b) \bmod 2$$

for each $a \in \hat{H}$ and each $b \in H$.

Definition. *The vectors $x, y \in \hat{H}_{2\lambda}$ will be called equivalent ($x \sim y$) iff the pair (x, y) satisfies the following three conditions:*

1. *If $x \in \hat{p} \operatorname{Ker} \Phi$ or $y \in \hat{p} \operatorname{Ker} \Phi$ then $x = y$;*
2. *$x - y$ belongs to the image $\operatorname{Im} \mathbf{i}$ of $\mathbf{i}$;*

for all critical points excluding A_μ and D_μ the third condition has the form:

3. *If $x, y \notin \hat{p} \operatorname{Ker} \Phi$ then*

$$q(\mathbf{i}^{-1}(x - y)) = \hat{\Phi}(x, \mathbf{i}^{-1}(x - y));$$

for the critical points A_μ and D_μ the third condition has the form:

3′. *If $x, y \notin \hat{p} \operatorname{Ker} \Phi$ and $x \in \operatorname{Im} \mathbf{i}$ then both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are vanishing vectors* (mod 2) *or both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are nonvanishing vectors* (mod 2).

Remarks. 1. For all critical points excluding A_μ and D_μ the Condition 3′ follows (in certain sense) from the Condition 3. More exactly, let $U_1, \ldots, U_s$ be classes of equivalent vectors, i.e. two vectors $x, y \in \hat{H}_{2\lambda}$ belong to one set U_α iff the Conditions 1.–3. are satisfied for the pair (x, y). Consider the new classes of equivalent vectors $U_1', \ldots, U_t'$ which are defined by the Conditions 1, 2, 3, 3′, i.e. two vectors $x, y \in \hat{H}_{2\lambda}$ belong to one new class U_β' iff the Conditions 1, 2, 3, 3′ are satisfied for the pair (x, y). Consider the group G (resp. G') of all automorphisms of $\hat{H}_{2\lambda}$, which preserve the sets $U_1, \ldots, U_s$ (resp. $U_1', \ldots, U_t'$). Clearly, $G' \subseteq G$.

Statement 2. *For all critical points excluding A_μ and D_μ $G = G'$.*

From the Statement 2 follows that the orbits of the group G coincides with the orbits of group G', i.e. $U_\alpha \setminus U_\beta'$ is the union of the orbits of G. If λ is odd then all U_α coincides with the corresponding U_β', i.e. the Condition 3′ follows from the Condition 3.

The proof of Statement 2 is given in Sect. 7.3.

2. In Sect. 7.2. we prove that for all critical points with unimodular intersection form on $H/\operatorname{Ker} \Phi$ (i.e. with $\lambda = 1$, in particular for all critical points of a function in two variables) excluding A_μ and D_μ the Condition 3 follows from the Condition 3′.

3. If $x, y \in \operatorname{Im} \mathbf{i}$ then the Condition 3 on these vectors x, y is equivalent to condition: $q(\mathbf{i}^{-1}(x)) = q(\mathbf{i}^{-1}(y))$.

6. Proofs of Statements of Sect. 2

6.1. *Proof of Statement 1*

If $x \in \hat{p}\operatorname{Ker}\Phi$ or $y \in \hat{p}\operatorname{Ker}\Phi$ then the Statement 1 is obvious. If $x+y \in \hat{p}\operatorname{Ker}\Phi$ then $h(x+y)=x+y$. Hence

$$\Phi_{2\lambda}(h(x), h(y))=\Phi_{2\lambda}(h(x+y), h(y))=0$$

and

$$\Phi_{2\lambda}(x, y)=\Phi_{2\lambda}(x+y, y)=0.$$

Let $x, y, x+y \notin \hat{p}\operatorname{Ker}\Phi$. From the Condition 3 for the vectors $(x+y, h(x+y))$ follows that

$$q(\mathbf{i}^{-1}(x-hx)+\mathbf{i}^{-1}(y-hy))=\hat{\Phi}(x+y, \mathbf{i}^{-1}(x-hx)+\mathbf{i}^{-1}(y-hy)).$$

But

$$\begin{aligned} &q(\mathbf{i}^{-1}(x-hx)+\mathbf{i}^{-1}(y-hy)) \\ &\quad =q(\mathbf{i}^{-1}(x-hx))+q(\mathbf{i}^{-1}(y-hy))+\Phi_2(\mathbf{i}^{-1}(x-hx), \mathbf{i}^{-1}(y-hy)) \\ &\quad =\hat{\Phi}(x, \mathbf{i}^{-1}(x-hx))+\hat{\Phi}(y, \mathbf{i}^{-1}(y-hy))+\Phi_2(\mathbf{i}^{-1}(x-hx), \mathbf{i}^{-1}(y-hy)). \end{aligned}$$

Hence

$$\Phi_2(\mathbf{i}^{-1}(x-hx), \mathbf{i}^{-1}(y-hy))=\hat{\Phi}(x, \mathbf{i}^{-1}(y-hy))+\hat{\Phi}(y, \mathbf{i}^{-1}(x-hx)).$$

Therefore

$$\begin{aligned} &\Phi\left(\frac{\hat{p}^{-1}(x-hx)}{\lambda}, \frac{\hat{p}^{-1}(y-hy)}{\lambda}\right) \\ &\quad \equiv \frac{1}{\lambda}\Phi\left(\hat{p}^{-1}(x), \frac{\hat{p}^{-1}(y-hy)}{\lambda}\right)+\frac{1}{\lambda}\Phi\left(\hat{p}^{-1}(y), \frac{\hat{p}^{-1}(x-hx)}{\lambda}\right) \bmod 2. \end{aligned}$$

$$\begin{aligned} &\Phi(\hat{p}^{-1}(x-hx), \hat{p}^{-1}(y-hy)) \\ &\quad \equiv \Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y-hy))+\Phi(\hat{p}^{-1}(y), \hat{p}^{-1}(x-hx)) \bmod 2\lambda^2. \end{aligned}$$

But

$$\begin{aligned} &\Phi(\hat{p}^{-1}(x-hx), \hat{p}^{-1}(y-hy)) \\ &\quad \equiv \Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))-\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(hy)) \\ &\quad -\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y))+\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy)) \bmod 2\lambda^2. \end{aligned}$$

Therefore

$$\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))+\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy))-2\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y)) \equiv 0 \bmod 2\lambda^2.$$

Hence

$$\begin{aligned} 0 &\equiv \Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))+\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy))-2\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y)) \\ &\quad +\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy))-\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy)) \\ &\equiv \Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))-\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy)) \\ &\quad -2\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y-hy)) \bmod 2\lambda^2. \end{aligned}$$

Since λ is a divisor of the vector $\hat{p}^{-1}(y-hy)$ and $\hat{p}^{-1}(hx)\in\hat{H}$ then

$$2\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y-hy))\equiv 0 \bmod 2\lambda^2.$$

Hence $\Phi_{2\lambda}(x, y)=\Phi_{2\lambda}(hx, hy)$. For the critical points A_μ and D_μ the Statement 1 follows by a simple calculation. This proves Statement 1.

6.2. Proof of Theorem 2

We shall prove firstly the following lemma.

Lemma 1. *Consider any critical point excluding A_μ and D_μ. Let h: $\hat{H}_{2\lambda}\to\hat{H}_{2\lambda}$ be an automorphism such that $h|_{\hat{p}\operatorname{Ker}\Phi}=\mathrm{Id}$, h preserves the form $\Phi_{2\lambda}$ and $h(x)\sim x$, $h(y)\sim y$ for the some vectors x, y. Then $h(x+y)\sim x+y$.*

Proof of Lemma 1. We shall verify the Conditions 1, 2, 3 of Sect. 5 for the pair $(x+y, h(x+y))$.

1. If $x+y\in\hat{p}\operatorname{Ker}\Phi$ or $h(x+y)\in\hat{p}\operatorname{Ker}\Phi$ then $h(x+y)=x+y$, because $h|_{\hat{p}\operatorname{Ker}\Phi}=\mathrm{Id}$.

Let $x+y, h(x+y)\notin\hat{p}\operatorname{Ker}\Phi$.

2. $h(x+y)-(x+y)=(hx-x)+(hy-y)\in\operatorname{Im}\mathbf{i}$.

3. $q(\mathbf{i}^{-1}(h(x+y)-(x+y)))$

$=q(\mathbf{i}^{-1}(hx-x))+q(\mathbf{i}^{-1}(hy-y))+\Phi_2(\mathbf{i}^{-1}(hx-x), \mathbf{i}^{-1}(hy-y))$

$=\hat{\Phi}(x, \mathbf{i}^{-1}(hx-x))+\hat{\Phi}(y, \mathbf{i}^{-1}(hy-y))+\Phi_2(\mathbf{i}^{-1}(hx-x), \mathbf{i}^{-1}(hy-y))$

$\equiv(1/\lambda^2)\,\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(hx))+(1/\lambda^2)\,\Phi(\hat{p}^{-1}(y), \hat{p}^{-1}(hy))$

$+(1/\lambda^2)\,\Phi(\hat{p}^{-1}(hx-x), \hat{p}^{-1}(hy-y)) \bmod 2$

$\equiv(1/\lambda^2)\,[\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(hx))+\Phi(\hat{p}^{-1}(y), \hat{p}^{-1}(hy))$

$+\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(hy))-\Phi(\hat{p}^{-1}(hx), \hat{p}^{-1}(y))$

$-\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(hy))+\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))] \bmod 2$

$\equiv(1/\lambda^2)\,[\Phi(\hat{p}^{-1}(x+y), \hat{p}^{-1}(hx))+\Phi(\hat{p}^{-1}(x+y), \hat{p}^{-1}(hy))$

$+2\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y))-2\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(hy))] \bmod 2$

$\equiv(1/\lambda^2)\,[\Phi(\hat{p}^{-1}(x+y), \hat{p}^{-1}(h(x+y)))+2\Phi(\hat{p}^{-1}(x), \hat{p}^{-1}(y-hy))] \bmod 2$

$\equiv(1/\lambda^2)\,\Phi(\hat{p}^{-1}(x+y), \hat{p}^{-1}(h(x+y))) \bmod 2$

$=\hat{\Phi}(x+y, \mathbf{i}^{-1}(h(x+y)-(x+y)))$.

This proves Lemma 1.

Proof of Theorem 2. The necessity of conditions of Theorem 2 follows from the necessity of conditions of Theorem 1 and Statement 1.

We shall prove the sufficiency. Consider any basis $e_1, \ldots, e_\mu$ in $\hat{H}_{2\lambda}$. Then $h(e_i)\sim e_i$ by the condition of Theorem 2. Hence $h(x)\sim x$ for all $x\in\hat{H}_{2\lambda}$ by Lemma 1. Therefore $h\in M_{2\lambda}$ follows from Theorem 1. This proves Theorem 2.

7. Proofs of Statements of Sect. 3–5

7.1. Wajnryb's function is defined by the set of vanishing vectors (mod 2) *and by the form* Φ_2

We introduce the new function $q_1: H_2 \to \mathbb{Z}_2$ which is defined by the set of vanishing vectors (mod 2) and by the form Φ_2. After this we shall prove the coincidence of functions q and q_1.

Function q_1 is defined by the formulas:

1) for each vanishing vector (mod 2) $x \in H_2$ $q_1(x)=1$,

2) $q_1(x+y)=q_1(x)+q_1(y)+\Phi_2(x,y)$.

Statement 3. *Function q_1 is correctly defined and the function q_1 coincides with the function q which is defined in Sect.* 3.1 *by the feebly distinguished basis.*

Proof of Statement 3. We shall prove the coincidence of functions q and q_1. Let e be an any element of feebly distinguished basis which define the function q. We shall prove firstly that the transformation h_e (h_e is defined by the formula $h_e(a)=a-\Phi(a,e)\,e$; see [1]) preserves the function q.

Since $q(e)=1$, we have

$$\begin{aligned} q(p(h_e(a))) &= q(p(a)+p(\Phi(a,e)\,e)) \\ &= q(p(a))+q(p(\Phi(a,e)))+\Phi_2(p(a),p(\Phi(a,e)\,e)) \\ &= q(p(a)). \end{aligned}$$

Since the monodromy group is generated by transvections h_e the monodromy group preserves the function q. By definition (see [1]) each vanishing vector is an image of an element of a feebly distinguished basis under the action of monodromy group. Then for each vanishing vector x $q(x)=1$. This proves Statement 3.

7.2. Proof of Theorem 3

Since $\lambda=1$, $\mathbf{i}$: $H_2 \to H_2$ is the identity. Therefore in this case the Condition 2 of definition of equivalent vectors is satisfied. From the Remark 1 of Sect. 5 follows that for the all critical points with $\lambda=1$ excluding the A_μ and D_μ Condition 3′ follows from the Condition 3. We shall prove that for the critical points with $\lambda=1$ excluding the A_μ and D_μ the Condition 3 follows from the Condition 3′ (see Remark 2 of Sect. 5). Obviously it follows from the following fact (see Remark 3 of Sect. 5): the condition $q(\mathbf{i}^{-1}(x))=q(\mathbf{i}^{-1}(y))$ follows from the condition "both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are vanishing vectors (mod 2) or both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are non-vanishing vectors (mod 2)". We shall prove this fact. Consider two cases:

1) both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are vanishing vectors (mod 2). Then (see the Proposition 1) $q(\mathbf{i}^{-1}(x))=1=q(\mathbf{i}^{-1}(y))$.

2) both $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ are not vanishing vectors (mod 2) and $\mathbf{i}^{-1}(x)$, $\mathbf{i}^{-1}(y)$ do not belong to $\operatorname{Ker}\Phi_2$. It follows from the Proposition 1 that $q(\mathbf{i}^{-1}(x))=0=q(\mathbf{i}^{-1}(y))$. This proves Theorem 3.

7.3. Proof of Statement 2

$G' \subseteq G$ (see the Remark 1 of Sect. 5). We shall prove that $G \subseteq G'$ for all critical point excluding A_μ and D_μ. Let $h \in G$. We shall prove that the Condition 3′ in the definition of Sect. 5 is satisfied for the pair (x, hx), where $x \in \operatorname{Im} \mathbf{i}$. Consider three cases:

1) $\mathbf{i}^{-1}(x) \in \operatorname{Ker} \Phi_2$. Then 2 divides the covector $\Phi_2(\mathbf{i}^{-1}(x), \cdot) \equiv \Phi\left(\frac{\hat{p}^{-1}(x)}{\lambda}, \cdot\right) \bmod 2$. Hence 2λ divides the covector $\Phi(\hat{p}^{-1}(x), \cdot)$. Then 2 divides the vector $\hat{p}^{-1}(x)$ by the definition of λ. Let $\hat{p}^{-1}(x) = 2a$ then $x = 2\hat{p}(a)$ and $h(x) - x = 2(h(\hat{p}(a)) - \hat{p}(a))$. Since $h \in G$ then $y = h(\hat{p}(a)) - \hat{p}(a) \in \operatorname{Im} \mathbf{i}$. Hence for the some

$$b \in H\colon\ 2y = 2\mathbf{i}\, p(b) = 2\hat{p}(\lambda b) = \hat{p}(2\lambda b) = 0.$$

Therefore $h(x) = x$ and Condition 3′ of the definition of Sect. 5 is satisfied.

Let $\mathbf{i}^{-1}(x) \notin \operatorname{Ker} \Phi_2$.

2) $q(\mathbf{i}^{-1}(x)) = 1$. Then $\mathbf{i}^{-1}(x)$ is a vanishing vector. Hence

$$q(\mathbf{i}^{-1}(hx)) = 1 \quad \text{and} \quad \mathbf{i}^{-1}(hx) \notin \operatorname{Ker} \Phi_2.$$

Therefore $\mathbf{i}^{-1}(h(x))$ is a vanishing vector.

3) $q(\mathbf{i}^{-1}(x)) = 0$. Then $\mathbf{i}^{-1}(x)$ is a nonvanishing vector. Hence $q(\mathbf{i}^{-1}(h(x)) = 0$. Therefore $\mathbf{i}^{-1}(h(x))$ is a nonvanishing vector.

This proves Statement 2.

8. Proof of Theorem 1 Necessity

In this section we shall prove a necessity of the conditions of the main theorem for the all critical point excluding A_μ and D_μ. For A_μ and D_μ the necessity is obvious.

Let $h_a\colon H \to H$ be a transvection in the vector a. The monodromy group M is generated by the transvections in all vanishing vectors a. Let $\bar{h}_a$ be the image of transvection h_a under the natural projection $M \to \operatorname{Aut}(\hat{H}_{2\lambda})$. We shall verify the Conditions 1-3 of the definition of Sect. 5 for all pair $(x, \bar{h}_a x)$, where a is a vanishing vector.

1. It is obvious that $\bar{h}_a x = x$ for $x \in \hat{p} \operatorname{Ker} \Phi$.

2.
$$\begin{aligned} \bar{h}_a(x) &= \hat{p}(h_a(\hat{p}^{-1}(x))) \\ &= \hat{p}(\hat{p}^{-1}(x) + \Phi(\hat{p}^{-1}(x), a)\, a) \\ &= x + \hat{p}\left(\lambda \cdot \frac{\Phi(\hat{p}^{-1}(x), a)}{\lambda} a\right), \end{aligned}$$

i.e. $\bar{h}_a(x) - x \in \operatorname{Im} \mathbf{i}$.

3. Let $(1/\lambda)\, \Phi(\hat{p}^{-1}(x), a) \equiv \alpha \bmod 2$. If $\alpha = 0$ then $\bar{h}_a(x) = x$. If $\alpha = 1$ then $\mathbf{i}^{-1}(\bar{h}_a(x) - x) = p(a)$, i.e. $\mathbf{i}^{-1}(\bar{h}_a(x) - x)$ is a vanishing vector. Hence $q(\mathbf{i}^{-1}(\bar{h}_a(x) - x)) = 1$. But also

$$\hat{\Phi}(x, p(a)) \equiv \frac{1}{\lambda} \Phi(\hat{p}^{-1}(x), a) \bmod 2 = \alpha = 1.$$

This proves the necessity of the conditions of Theorem 1.

9. Proof of Theorem 1 Sufficiency

In this section we shall prove a sufficiency of the condition of the main theorem for all critical points excluding A_μ and D_μ.

There is a basis $e_1, f_1, \ldots, e_n, f_n, g_1, \ldots, g_l$ of H, where $2n+l=\mu$ such that:

$$\left.\begin{aligned} &\Phi(e_i, f_i) = \lambda_i \neq 0;\ \lambda_1 = \lambda; \\ &\Phi(e_i, e_j) = \Phi(e_i, f_j) = \Phi(e_i, g_k) = \Phi(f_i, f_j) = \Phi(f_i, g_k) = \Phi(g_k, g_s) = 0 \end{aligned}\right\} (*)$$

where $1 \leqq i, j \leqq n$; $i \neq j$; $1 \leqq k, s \leqq l$ and λ_i divides λ_{i-1} for $2 \leqq i \leqq n$.

This basis plays an important role in the proof of Theorem 1. (Also as in the proof of main theorem of article [1].)

It is obvious (see the proof of Lemma 3 in [1]) that

$$\hat{H} = \bigoplus_{i=1}^{n} \frac{\lambda}{\lambda_i} \langle e_i, f_i \rangle \oplus \operatorname{Ker} \Phi,$$

where $\langle e_i, f_i \rangle$ is the sublattice in H generated by vectors e_i and f_i. Hence the vectors

$$a_i = \hat{p}\left(\frac{\lambda}{\lambda_i} e_i\right), \quad b_i = \hat{p}\left(\frac{\lambda}{\lambda_i} f_i\right), \quad \hat{p}(g_k),$$

$(1 \leqq i, j \leqq n, 1 \leqq k \leqq l)$ are a basis of the $\mathbb{Z}_{2\lambda}$-module $\hat{H}_{2\lambda}$.

Let k $(1 \leqq k \leqq n)$ be a natural number such that $\lambda_1, \ldots, \lambda_{k-1}$ are even and $\lambda_k, \ldots, \lambda_n$ are odd. The Theorem 1 obviously follows from the Statements 4 and 5.

Statement 4. *Let $h(x) \sim x$ for the all $x \in \hat{H}_{2\lambda}$. Then there is $h_1 \in M_{2\lambda}$ such that $h_1 h(a_i) = a_i$ and $h_1 h(b_i) = b_i$ for $1 \leqq i < k$.*

Statement 5. *Let $h(x) \sim x$ for all $x \in \hat{H}_{2\lambda}$ and $h(a_i) = a_i$, $h(b_i) = b_i$ for $1 \leqq i < k$. Then $h \in M_{2\lambda}$.*

9.1. Auxiliary Lemmas

In the proof of Theorem 1 we used the following fact: in the lattice H corresponding to E_6 there is a symplectic basis which consists of vanishing vectors. More exactly, in the lattice H consider a distinguished basis with the standard Dynkin diagram, i.e. basis $\Delta_1, \ldots, \Delta_6$ such that:

$$\Phi(\Delta_1, \Delta_2) = \Phi(\Delta_2, \Delta_3) = \Phi(\Delta_3, \Delta_5) = \Phi(\Delta_5, \Delta_6) = \Phi(\Delta_4, \Delta_3) = 1$$

and all other intersection number are zero.

Lemma 2. *The basis* Δ_1, Δ_2, $\Delta_1+\Delta_3+\Delta_6$, Δ_4, Δ_5, Δ_6 *of the lattice* H *corresponding to* E_6 *is a symplectic basis which consists of vanishing vectors.*

Proof of Lemma 2 is obvious.

Remark. According to this lemma we may suppose that the vectors e_{n-2}, f_{n-2}, e_{n-1}, f_{n-1}, e_n, f_n of the basis (*) are vanishing vectors. Hence $\lambda_{n-2}=\lambda_{n-1}=\lambda_n=1$. Therefore $k\leqq n-2$.

In the proof of Statement 4 we shall search the automorphism h_1 in the form of composition of transvections in the special vanishing vectors. In Lemma 3 we shall prove the possibility of choosing such vectors.

Consider the vectors $c, d, y\in H$ such that:

1) the vectors c, d are primitive (nondivisible), $c, d\in\langle e_i, f_i\rangle$ for some i, $1\leqq i<k$ and $\Phi(c,d)\equiv\lambda_i \bmod 2\lambda_i$.

2) $p(y)\notin \operatorname{Ker}\Phi_2$ and $q(p(y))=0$.

Lemma 3. *For such* $c, d, y\in H$ *there is* $z\in H$ *such that:*

1. $\Phi(z,y)\equiv 0 \bmod 2$,
2. $\Phi(z,c)\equiv\lambda_i \bmod 2\lambda_i$,
3. $p(y+z)\notin\operatorname{Ker}\Phi_2$,
4. $p(z)$ *is a vanishing vector* (mod 2),
5. $\Phi(z,d)\equiv 0 \bmod 2\lambda_i$,
6. $z\in\langle e_i, f_i, e_{i+1}, f_{i+1}, \ldots, e_n, f_n\rangle$.

In the proof of this lemma we shall use the following Wajnryb's result.

Lemma 4 (B. Wajnryb [2]). *In the* $\mathbb{Z}_2$*-lattice*

$$\langle p(e_k), p(f_k), \ldots, p(e_n), p(f_n)\rangle$$

there exists a symplectic basis $\bar{e}_k, \bar{f}_k, \ldots, \bar{e}_n, \bar{f}_n$ *such that either*

1) *all* $\bar{e}_i, \bar{f}_i$ $(k\leqq i\leqq n)$ *are the vanishing vectors* (mod 2) *or*

2) *all* $\bar{e}_i, \bar{f}_i$ *are the vanishing vectors* (mod 2), *except for* $\bar{e}_k, \bar{f}_k$.

Proof of Lemma 3.

Case I. $q(p(d))=1$.

Consider the triplet of vectors $\bar{e}_s$, $\bar{f}_s$, $\bar{e}_s+\bar{f}_s$. The vector $p(y)$ does not intersect at least one vector from this triplet.

Ia. There exist two triplets $(\bar{e}_s, \bar{f}_s, \bar{e}_s+\bar{f}_s)$ and $(\bar{e}_t, \bar{f}_t, \bar{e}_t+\bar{f}_t)$ such that all six vector from these triplets are vanishing vectors (mod 2) and $p(y)$ does not intersect exactly one vector from each triplet. Let

$$\Phi_2(p(y), \bar{e}_s)=\Phi_2(p(y), \bar{e}_t)=0.$$

Then $z=d+p^{-1}(\bar{f}_s+\bar{f}_t)$.

Ib. There exists only one triplet $(\bar{e}_s, \bar{f}_s, \bar{e}_s+\bar{f}_s)$ such that all three vectors from this triplet are vanishing vectors (mod 2) and $p(y)$ does not intersect exactly one vector from this triplet. Let $\Phi_2(\mathrm{p(y)}, \bar{\mathrm{e}}_s)=0$.

Consider the another triplet $(\bar{e}_t, \bar{f}_t, \bar{e}_t+\bar{f}_t)$ such that all $\bar{e}_t$, $\bar{f}_t$, $\bar{e}_t+\bar{f}_t$ are vanishing vectors. The vector $p(y)$ does not intersect all three vectors from this triplet. Then $z=d+p^{-1}(\bar{e}_s+\bar{e}_t)$.

Ic. For each triplet $(\bar{e}_s, \bar{f}_s, \bar{e}_s+\bar{f}_s)$ such that all $\bar{e}_s$, $\bar{f}_s$, $\bar{e}_s+\bar{f}_s$ are vanishing vectors, the vector $p(y)$ does not intersect the all vectors from that triplet. Since $p(y)\notin \operatorname{Ker}\Phi_2$ then $\bar{e}_k$, $\bar{f}_k$ are nonvanishing vectors and $p(y)$ does not intersect exactly one vector from $\bar{e}_k$, $\bar{f}_k$, $\bar{e}_k+\bar{f}_k$. Let $\Phi_2(p(y), \bar{e}_k)=1$. Consider the pair of vectors $(\bar{e}_k+\bar{e}_{n-1}, \bar{f}_k+\bar{e}_{n-1})$. The conditions of case Ib are satisfied for two triplets $(\bar{e}_n, \bar{f}_n, \bar{e}_n+\bar{f}_n)$ and $(\bar{e}_k+\bar{e}_{n-1}, \bar{f}_k+\bar{e}_{n-1}, \bar{e}_k+\bar{f}_k)$.

Case II. $q(p(d))=0$.

From Lemma 2 follows that there exist at least three pairs of vanishing vectors $\bar{e}_{n-2}, \bar{f}_{n-2}; \bar{e}_{n-1}, \bar{f}_{n-1}; \bar{e}_n, \bar{f}_n$.

If at least for one pair (e, f) from this three pair the vector $p(y)$ does not intersect all vectors e, f, $e+f$, then $z=d+p^{-1}(e)$.

If $p(y)$ does not intersect exactly one vector from each triplet $(\bar{e}_{n-2}, \bar{f}_{n-2}, \bar{e}_{n-2}+\bar{f}_{n-2})$, $(\bar{e}_{n-1}, \bar{f}_{n-1}, \bar{e}_{n-1}+\bar{f}_{n-1})$, $(\bar{e}_n, \bar{f}_n, \bar{e}_n+\bar{f}_n)$ then we may suppose that

$$\Phi(p(y), \bar{e}_{n-2})=\Phi(p(y), \bar{f}_{n-2})=\Phi(p(y), \bar{e}_{n-1})=1$$

and $\Phi(p(y), u)=0$, where $u\in\langle \bar{e}_n, \bar{f}_n\rangle$. Let

$$v=\bar{e}_{n-2}+\bar{e}_{n-1}+u, \qquad w=\bar{f}_{n-2}+\bar{e}_{n-1}+u.$$

Hence (v, w) is a pair of vanishing vectors and $p(y)$ does not intersect all vectors v, w, $v+w$. Then $z=d+p^{-1}(v)$. This proves Lemma 3.

9.2. *Proof of Statement 4*

We may suppose that there exists $h_2\in M_{2\lambda}$ such that $h_2h(a_j)=a_j$ and $h_2h(b_j)=b_j$ for $j<i$, where $1\leqq i, j<k$.

I. We shall search $h_3\in M_{2\lambda}$ such that:

$$h_3h_2h(a_j)=a_j, \qquad h_3h_2h(b_s)=b_s$$

for $j\leqq i$, $s<i$.

Since

$$h_2h(a_i)\sim a_i, \qquad h_2h(a_i)-a_i=\hat{p}(\lambda y)$$

for some $y\in H$. From the statement 1 follows that

$$\Phi_{2\lambda}(h_2h(a_i), a_j)=\Phi_{2\lambda}(h_2h(a_i), b_j)=0$$

for $j<i$. Hence

$$\begin{aligned}\Phi_{2\lambda}(a_i+\hat{p}(\lambda y), a_j)&=\Phi_{2\lambda}(\hat{p}(\lambda y), a_j)\\ &\equiv \Phi(\lambda y, \hat{p}^{-1}(a_j)) \bmod 2\lambda^2\\ &\equiv 0 \bmod 2\lambda^2.\end{aligned}$$

Then $\Phi(y, \hat{p}^{-1}(a_j))\equiv 0 \bmod 2\lambda$. But $\hat{p}^{-1}(a_j)=(\lambda/\lambda_j)e_j$. Then $\Phi(y, e_j)\equiv 0 \bmod 2\lambda_j$ for $j<i$. By analogy $\Phi(y, f_j)\equiv 0 \bmod 2\lambda_j$ for $j<i$. Hence $p(y)\in\langle p(e_i), p(f_i), \ldots, p(g_l)\rangle$.

Ia. If $\mathbf{i}^{-1}(h_2 h(a_i)-a_i)=p(y)$ is a vanishing vector (mod 2), then $q(p(y))=1$ and $p(y)\notin \operatorname{Ker}\Phi_2$. Since

$$1=q(p(y))=\hat{\Phi}(a_i,p(y))\equiv(1/\lambda)\,\Phi((\lambda/\lambda_i)\,e_i,y)\bmod 2$$

we have $\Phi(e_i,y)\equiv\lambda_i \bmod 2\lambda_i$.

Designation. Let $a\in H_2$ be a vanishing vector (mod 2), then $h_a\in M_{2\lambda}$ is an image $\bar{h}_{a'}$ of a transvection $h_{a'}$ under the natural projection $M\to M_{2\lambda}$, where $a'\in H$ is a vanishing vector and $p(a')=a$.

We may put $h_3=h_{p(y)}$.

Ib. If $p(y)$ is a nonvanishing vector (mod 2) then $q(p(y)=0$ or $p(y)\in\operatorname{Ker}\Phi_2$.

We may suppose that $p(y)\notin\operatorname{Ker}\Phi_2$, because of if $p(y)\in\operatorname{Ker}\Phi_2$ then consider a vector

$$\Delta=\begin{cases} f_i+e_{n-1}+e_n, & \text{if } q(p(f_i))=1\\ f_i+e_n, & \text{if } q(p(f_i))=0\end{cases}.$$

(All six vectors e_{n-2}, f_{n-2}, e_{n-1}, f_{n-1}, e_n, f_n are vanishing vectors.) Hence $p(\Delta)$ is a vanishing vector (mod 2) and $\Phi(\Delta,e_i)\equiv\lambda_i \bmod 2\lambda_i$. Therefore

$$h_{p(\Delta)}h_2 h(a_i)=a_i+\hat{p}(\lambda(y+\Delta)) \quad \text{and} \quad p(y+\Delta)\notin\operatorname{Ker}\Phi_2.$$

Let $p(y)\notin\operatorname{Ker}\Phi_2$ then $q(p(y)=0$. Hence

$$0=q(p(y))\equiv\frac{1}{\lambda_i}\Phi(e_i,y)\bmod 2.$$

Therefore $\Phi(e_i,y)\equiv 0\bmod 2\lambda_i$. We put $c=e_i$, $d=f_i$. From Lemma 3 follows that there exists z such that:

1) $\Phi(z,y)\equiv 0\bmod 2$,
2) $\Phi(z,e_i)\equiv\lambda_i \bmod 2\lambda_i$,
3) $p(z+y)\notin\operatorname{Ker}\Phi_2$,
4) $p(z)$ is a vanishing vector (mod 2),
5) $\Phi(z,f_i)\equiv 0\bmod 2\lambda_i$,
6) $z\in\langle e_i,f_i,\ldots,e_n,f_n\rangle$.

Then

$$h_{p(z)}\in M_{2\lambda};\qquad h_{p(z)}h_2 h(a_i)=a_i+\hat{p}(\lambda(y+z))$$

and

$$q(p(y+z))=\hat{\Phi}(a_i,p(y+z))\equiv\frac{1}{\lambda_i}\Phi(e_i,y+z)\bmod 2$$

$$\equiv\frac{1}{\lambda_i}\Phi(e_i,z)\bmod 2\equiv 1\bmod 2.$$

Hence $p(y+z)$ is a vanishing vector (mod 2). This is case Ia.

II. Now we may suppose that $h(a_j)=a_j$ for $j\leqq i$ and $h(b_s)=b_s$ for $s<i$. We shall search $h_2\in M_{2\lambda}$ such that $h_2 h(a_j)=a_j$ and $h_2 h(b_j)=b_j$ for $j\leqq i$.

Let $h(b_j)-b_j=\hat{p}(\lambda y)$. By analogy with Ib we may suppose that $p(y)\notin\operatorname{Ker}\Phi_2$ (here in the definition of the vanishing vector Δ the vector e_i is considered

instead of f_i). From Statement 1 follows that $\Phi(e_j, y) \equiv 0 \bmod 2\lambda_j$ for $j \leqq i$ and $\Phi(f_s, y) \equiv 0 \bmod 2\lambda_s$ for $s < i$. Therefore

$$p(y) \in \langle p(e_i),\ p(e_{i+1}),\ p(f_{i+1}), \ldots, p(e_n),\ p(f_n),\ p(g_1), \ldots, p(g_l) \rangle.$$

IIa. Let $p(y) = p(e_i) + p(y')$, where

$$p(y') \in \langle p(e_{i+1}), p(f_{i+1}), \ldots, p(g_l) \rangle.$$

Then

$$\Phi(y', f_i) \equiv \Phi(y', e_i) \equiv 0 \bmod 2\lambda_i$$

and

$$q(p(y)) = \hat{\Phi}(b_i, p(y)) \equiv \frac{1}{\lambda_i} \Phi(f_i, y) \bmod 2 \equiv 1 \bmod 2.$$

Hence $p(y)$ is a vanishing vector (mod 2). Since $\Phi(e_i, y) \equiv 0 \bmod 2\lambda_i$, then we may put $h_2 = h_{p(y)}$.

IIb. Let

$$p(y) \in \langle p(e_{i+1}), p(f_{i+1}), \ldots, p(g_l) \rangle.$$

Then $\Phi(y, e_i) \equiv \Phi(y, f_i) \equiv 0 \bmod 2\lambda_i$. Therefore $q(p(y)) = 0$. We put $c = f_i$, $d = e_i$. From Lemma 3 (we may suppose that $p(y) \notin \operatorname{Ker} \Phi_2$) follows that there exists z such that:

1) $\Phi(z, y) \equiv 0 \bmod 2$,
2) $\Phi(z, f_i) \equiv \lambda_i \bmod 2\lambda_i$,
3) $p(y+z) \notin \operatorname{Ker} \Phi_2$,
4) $p(z)$ is a vanishing vector (mod 2),
5) $\Phi(z, e_i) \equiv 0 \bmod 2\lambda_i$,
6) $z \in \langle e_i, f_i, \ldots, e_n, f_n \rangle$.

Then $h_{p(z)} h(b_i) = b_i + \hat{p}(\lambda(y+z))$ and

$$h_{p(z)} h(a_j) = a_j \quad \text{for } j \leqq i,$$
$$h_{p(z)} h(b_s) = b_s \quad \text{for } s < i.$$

Since $p(y+z) \notin \operatorname{Ker} \Phi_2$ and

$$q(y+z) = q(y) + q(z) + \Phi_2(p(z), p(y)) = 1,$$

then $p(y+z)$ is a vanishing vector (mod 2) and we may put $h_2 = h_{p(y+z)} h_{p(z)}$. This proves Statement 4.

9.3. *Proof of Statement 5*

Consider the $\mathbb{Z}_2$-lattice

$$V = \langle p(e_k), p(f_k), \ldots, p(e_n), p(f_n), p(g_1), \ldots, p(g_l) \rangle.$$

Consider a subgroup of the monodromy group which is generated by transvections in vanishing vectors from a lattice

$$L=\langle e_k,f_k,\ldots,e_n,f_n,g_1,\ldots,g_l\rangle.$$

Let M' be an image of this subgroup under the natural projection $\mathrm{Aut}(L)\to\mathrm{Aut}(V)$. Let $h\colon \hat{H}_{2\lambda}\to\hat{H}_{2\lambda}$ be an automorphism such that the conditions of Statement 5 are satisfied. This automorphism h defines an automorphism h' of the $\mathbb{Z}_2$-lattice V. We shall prove that if $h'\in M'$ then $h\in M_{2\lambda}$. In Lemma 5 we shall prove that $h'\in M'$.

Definition. Suppose conditions of Statement 5 are satisfied for the automorphism $h\colon \hat{H}_{2\lambda}\to\hat{H}_{2\lambda}$, i.e. $h(x)\sim x$ for all $x\in\hat{H}_{2\lambda}$ and $h(a_j)=a_j$, $h(b_j)=b_j$ for $j<k$. The automorphism h' is defined by the formulas:

$$h'(p(e_j))=p(e_j)+\lambda_j\mathbf{i}^{-1}(h(a_j)-a_j),$$
$$h'(p(f_j))=p(f_j)+\lambda_j\mathbf{i}^{-1}(h(b_j)-b_j)$$

for $k\leqq j\leqq n$; $h'(p(g_s))=p(g_s)$ for $1\leqq s\leqq l$.

Lemma 5. $h'\in M'$.

Proof of Lemma 5. First we shall prove that h' preserves the function $q|_V$.

$$\begin{aligned}q(h'(p(e_j)))&=q(p(e_j))+q(\lambda_j\mathbf{i}^{-1}(h(a_j)-a_j))\\&\quad+\Phi_2(p(e_j),\lambda_j\mathbf{i}^{-1}(h(a_j)-a_j))\\&=q(p(e_j))+q(\mathbf{i}^{-1}(h(a_j)-a_j))\\&\quad+\hat{\Phi}(a_j,\mathbf{i}^{-1}(h(a_j)-a_j))\\&=q(p(e_j)).\end{aligned}$$

By analogy $q(h'(p(f_j)))=q(p(f_j))$ for all $k\leqq j\leqq n$.

Now our situation is similar to Wajnryb's situation (in proof of the corollary of Proposition 2 of [2]). Therefore the end of proof of Lemma 5 is analogous to Wajnryb's proof.

Consider a basis

$$\bar{e}_k,\bar{f}_k,\ldots,\bar{e}_n,\bar{f}_n,\ p(g_1),\ldots,p(g_l)$$

of V as in Lemma 4. For each α $(k\leqq\alpha\leqq n)$ let

$$V_\alpha=\langle\bar{e}_k,\bar{f}_k,\ldots,\bar{e}_\alpha,\bar{f}_\alpha,p(g_1),\ldots,p(g_l)\rangle.$$

Let M'_α be a subgroup of M' which preserves V_α.

Remark. If $\bar{e}_k,\bar{f}_k$ are nonvanishing vectors then the group $M'_\alpha(\alpha>k)$ acts transitively on the vanishing vectors from V_α. If $\bar{e}_k$, $\bar{f}_k$ are the vanishing vectors then group M'_α $(\alpha>k+1)$ acts transitively on the vanishing vectors from V_α.

Induction Hypothesis. Let

$$h'(\bar{e}_{\alpha+1})=\bar{e}_{\alpha+1},\quad h'(\bar{f}_{\alpha+1})=\bar{f}_{\alpha+1},\ldots,$$
$$h'(\bar{e}_n)=\bar{e}_n,\quad h'(f_n)=\bar{f}_n\quad\text{and}\quad\alpha>k+1.$$

Induction Step. Since M'_α acts transitively on the vanishing vectors, we may suppose that $h'(\bar{e}_\alpha)=\bar{e}_\alpha$. Let $h'(\bar{f}_\alpha)=f$. Then $\Phi_2(\bar{e}_\alpha,f)=1$ because h' preserves the form Φ_2 (since h' preserves q). If $\Phi_2(\bar{f}_\alpha,f)=1$, then we put $u=\bar{f}_\alpha+f$. If $\Phi_2(\bar{f}_\alpha,f)=0$, then we put $u=\bar{e}_\alpha+\bar{f}_\alpha+f$. Hence $h'_u h'(\bar{e}_\alpha)=\bar{e}_\alpha$ and $h'_u h'(\bar{f}_\alpha)=\bar{f}_\alpha$. This reduces the proof of Lemma 5 to following two cases:

1) $\alpha=k+1$ and $\bar{e}_k,\bar{f}_k$ are vanishing vectors;
2) $\alpha=k$ and $\bar{e}_k,\bar{f}_k$ are nonvanishing vectors.

Case 1. (Cf. [2].) This case is nontrivial only in following situation

$$h'(\bar{e}_{k+1})=\bar{e}_k;\quad h'(\bar{f}_{k+1})=\bar{f}_k;\quad h'(\bar{e}_k)=\bar{e}_{k+1};\quad h'(\bar{f}_k)=\bar{f}_{k+1}.$$

Then h' is the composition of transvection in the following vanishing vectors:

$$\bar{e}_{k+2}+\bar{e}_{k+1}+\bar{f}_k;\quad \bar{e}_{k+2}+\bar{f}_{k+1}+\bar{e}_k+\bar{f}_k;\quad \bar{f}_{k+1};$$
$$\bar{e}_{k+2}+\bar{e}_{k+1}+\bar{e}_k;\quad \bar{e}_{k+2};\quad \bar{e}_{k+1};\quad \bar{e}_k+\bar{f}_k;\quad \bar{e}_k.$$

Case 2. The vector $u=\bar{e}_k+\bar{f}_k$ is a vanishing vector.

If $h'(u)=u$ and $h'(\bar{e}_k)=e$ then $\Phi_2(e,u)=1$ because h' preserves the form Φ_2, and $q(e)=0$. Hence $e=v+x$, where $x\in\operatorname{Ker}\Phi_2\cap V$ and $v=\bar{e}_k$ or $v=\bar{f}_k$. Therefore $q(x)=0$. Then $u+x$ is a vanishing vector and

$$h'_{u+x}h'(u)=u;\quad h'_{u+x}h'(\bar{e}_k)=v+u.$$

But $u+v=\bar{e}_k$ or $u+v=\bar{f}_k$. If $u+v=\bar{f}_k$, then

$$h'_u h'_{u+x}h'(u)=u;\quad h'_u h'_{u+x}h'(\bar{e}_k)=\bar{e}_k.$$

Now let $h'(u)\neq u$. Then $h'(u)=v+x$, where $x\in\operatorname{Ker}\Phi_2\cap V$ and $v\in\langle\bar{e}_k,\bar{f}_k\rangle$. If $v=u$, then

$$h'_{\bar{e}_{k+1}+\bar{e}_k}h'_{\bar{e}_{k+1}+\bar{e}_k+x}h'(u)=u.$$

If $v\neq u$ then $v=\bar{e}_k$ or $v=\bar{f}_k$. Hence $q(x)=1$. If $v=\bar{e}_k$ then $h'_{\bar{f}_k+x}h'(u)=u$. If $v=\bar{f}_k$ then $h'_{\bar{e}_k+x}h'(u)=u$. This proves Lemma 5.

Proof of Statement 5. Let $h'\in M'$. Hence there exists $h_1: L\to L$ such that $h_1\in M$ and the image h'_1 of transformation h_1 in a group M' is the inverse transformation to h', i.e. $h'_1 h=\mathrm{Id}$. Hence

$$p(e_j)=h'_1h'(p(e_j))=h'_1(p(e_j)+\lambda_j\mathbf{i}^{-1}(ha_j-a_j))$$
$$=p(h_1(e_j))+\lambda_j p\left(h_1\left(\frac{\hat{p}^{-1}(ha_j-a_j)}{\lambda}\right)\right)$$
$$=p(h_1(e_j))+p\left(\frac{\lambda_j}{\lambda}h_1(\hat{p}^{-1}(ha_j-a_j))\right).$$

Let $\hat{h}_1$ be the image of the transformation $h_1\in M$ in the group $M_{2\lambda}$, i.e. $\hat{h}_1(x)=\hat{p}(h_1(\hat{p}^{-1}(x)))$ for all $x\in\hat{H}_{2\lambda}$.

Then

$$p(e_j)=p\left(h_1(e_j)+\frac{\lambda_j}{\lambda}\hat{p}^{-1}(\hat{h}_1(h a_j-a_j))\right)$$

$$=p\left(h_1(e_j)+\frac{\lambda_j}{\lambda}[\hat{p}^{-1}(\hat{h}_1 h a_j)-\hat{p}^{-1}\hat{h}_1 a_j]\right)$$

$$=p\left(h_1(e_j)+\frac{\lambda_j}{\lambda}\left[\hat{p}^{-1}(\hat{h}_1 h a_j)-h_1\left(\frac{\lambda}{\lambda_j}e_j\right)\right]\right)$$

$$=p\left(\frac{\lambda_j}{\lambda}\hat{p}^{-1}(\hat{h}_1 h(a_j))\right).$$

Hence

$$p\left(\frac{\lambda_j}{\lambda}\hat{p}^{-1}(\hat{h}_1 h(a_j))-e_j\right)=0.$$

But $e_j=\frac{\lambda_j}{\lambda}\hat{p}^{-1}(a_j)$. Therefore $p\left(\frac{\lambda_j}{\lambda}\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]\right)=0$. Then 2 divides the vector $\frac{\lambda_j}{\lambda}\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]$. Since $\hat{h}_1 h a_j\sim a_j$, λ divides the vector $\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]$. Then λ_j divides the vector $\frac{\lambda_j}{\lambda}\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]$. Since λ_j is odd, $2\lambda_j$ divides the vector $\frac{\lambda_j}{\lambda}\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]$. Then 2λ divides the vector $\hat{p}^{-1}[\hat{h}_1 h a_j-a_j]$. Hence $\hat{h}_1 h a_j=a_j$. By analogy $\hat{h}_1 h b_j=b_j$. This proves Statement 5.

Addition. On the Monodromy of A_μ and D_μ

In this addition we shall give Varchenko's results about the monodromy of critical points of A_μ and D_μ.

For all critical points A_μ and D_μ $\lambda=1$ and the $\mathbb{Z}_2$-module $\hat{H}_2$ coincides with $\mathbb{Z}_2$-lattice H_2 (see Sect. 4).

Theorem 5. *For the critical points A_μ and D_μ the transformation h: $H_2\to H_2$ belongs to the monodromy group M_2 iff h preserves the form Φ_2; $h|_{p(\mathrm{Ker}\Phi)}=\mathrm{Id}$ and the image $h(x)$ is a vanishing vector for each element x of a standard basis (see Sect. 1 of this addition).*

This theorem is analogous to Theorem 2. Theorem 1 for the critical points A_μ, D_μ follows from Statement 1 and Theorem 5.

The necessity of conditions of Theorem 5 is obvious. We shall prove the sufficiency.

1. Standard Basis

A_μ. The basis $e_1,\ldots,e_\mu$ of the lattice H will be called a standard basis of the critical point A_μ if $e_1,\ldots,e_\mu$ are vanishing vectors, $\Phi(e_i,e_{i-1})=1$ for $2\leqq i\leqq\mu$ and all other intersection numbers are zero.

D_μ. The basis $e_1, \ldots, e_\mu$ of the lattice H will be called a standard basis of the critical point D_μ if $e_1, \ldots, e_\mu$ are vanishing vectors, $\Phi(e_i, e_{i-1})=1$ for $2 \leqq i \leqq \mu-1$, $\Phi(e_\mu, e_{\mu-2})=1$ and all other intersection numbers are zero.

2. Vanishing Vectors

A_μ. **Proposition 2.** *The vector* $x=\sum_{i=1}^{\mu} \varepsilon_i p(e_i) \in H_2$ $(\varepsilon_i=0$ *or* $1)$ *is a vanishing vector* (mod 2) *iff*

$$\sum_{i=0}^{\mu} |\varepsilon_i - \varepsilon_{i+1}| = 2, \quad \text{where } \varepsilon_0 = \varepsilon_{\mu+1} = 0.$$

D_μ. **Proposition 3.** *The vector* $x=\sum_{i=1}^{\mu} \varepsilon_i p(e_i) \in H_2$ $(\varepsilon_i=0$ *or* $1)$ *is a vanishing vector* (mod 2) *iff the coordinates* $\varepsilon_1, \ldots, \varepsilon_\mu$ *satisfy one of the following three conditions:*

1) $\varepsilon_1 + \sum_{i=1}^{\mu-4} |\varepsilon_i - \varepsilon_{i+1}| + \varepsilon_{\mu-3} = 2$; $\varepsilon_{\mu-2} = \varepsilon_\mu - \varepsilon_{\mu-1} = 0$.
2) $\varepsilon_1 + \sum_{i=1}^{\mu-4} |\varepsilon_i - \varepsilon_{i+1}| + \varepsilon_{\mu-3} = 2$; $\varepsilon_{\mu-3} = 1$; $\varepsilon_{\mu-2} = 1$.
3) $\varepsilon_i = 0$ for $i \leqq \mu-3$; $\varepsilon_\mu + |\varepsilon_\mu - \varepsilon_{\mu-2}| + |\varepsilon_{\mu-2} - \varepsilon_{\mu-1}| + \varepsilon_\mu = 2$.

Proofs of Propositions 2 and 3 are simple and therefore omitted.

3. Proof of Theorem 5 for A_μ

The theorem is proved by induction on μ. If $\mu=1,2$ the theorem is proved by direct computations. Let $\mu>2$ and $\bar{e}_1=p(e_1), \ldots, \bar{e}_\mu=p(e_\mu)$. Since $h(\bar{e}_\mu)$ is a vanishing vector, there exists $h_1 \in M_2$ such that $h_1 h(\bar{e}_\mu)=\bar{e}_\mu$. Since $h_1 h$ preserves the form Φ_2, the $(\mu-1)$-th coordinate of the vector $h_1 h(\bar{e}_{\mu-1})$ is equal to 1 and the $(\mu-1)$-th coordinates of vectors $h_1 h(\bar{e}_1), \ldots, h_1 h(\bar{e}_{\mu-2})$ are equal to zero. Since all vectors $h_1 h(e_i)$ are vanishing vectors, it follows from Proposition 2 that the μ-th coordinates of vectors $h_1 h(\bar{e}_1), \ldots, h_1 h(\bar{e}_{\mu-2})$ are equal to zero. The left upper $(\mu-2)\times(\mu-2)$ corner of the matrix of $h_1 h$ satisfies the hypothesis of Theorem 5 for $A_{\mu-2}$. Hence there exists $h_2 \in M_2$ such that $h_2 h(e_i) = e_i$ for $i \leqq \mu-2$ and $i=\mu$. From the preservation of the form Φ_2 follows that $h_2 h(\bar{e}_{\mu-1}) = \bar{e}_{\mu-1} + \varepsilon \bar{e}_\mu$, where $\varepsilon=0$ or 1. If $\varepsilon=1$ then $h=h_2^{-1} h_\mu^{-1} \in M_2$. If $\varepsilon=0$ then $h=h_2^{-1} \in M_2$. This proves Theorem 5 for A_μ.

4. Proof of Theorem 5 for D_μ

Lemma 6. *Let* h *be a matrix (in a standard basis) of the transformation* h: $H_2 \to H_2$ *which satisfies the hypothesis of Theorem* 5 *for the critical point* D_μ. *Let* $v=(v_1, \ldots, v_\mu)$ *be a* μ*-th row of* $h(v_i \in \mathbb{Z}_2)$. *Then* v *can be transformed into* $(0, 0, \ldots, 0, 1)$ *with the help of the multiplication of* h *by* $h_{\bar{e}_i}$ $(i=1, \ldots, \mu)$ *from the left.*

Proof of Lemma 6. The multiplication of h by $h_{\bar{e}_1}$ from the left is the addition of v_1 to v_2. The multiplication of h by $h_{\bar{e}_\mu}$ (resp. $h_{\bar{e}_{\mu-1}}$) from the left is the addition of v_μ (resp. $v_{\mu-1}$) to $v_{\mu-2}$. The multiplication of h by $h_{\bar{e}_{\mu-2}}$ from the left is the addition of $v_{\mu-2}$ to $v_{\mu-1}$, v_μ, $v_{\mu-3}$. The multiplication of h by $h_{\bar{e}_i}$ $(2 \leqq i \leqq \mu-3)$ from the left is the addition of v_i to v_{i-1} and v_{i+1}. Since $h(\bar{e}_{\mu-1}+\bar{e}_\mu)=\bar{e}_{\mu-1}+\bar{e}_\mu$, $v_\mu+v_{\mu-1}=1$. Hence multiplying of h by $h_{\bar{e}_\mu}$ or $h_{\bar{e}_{\mu-1}}$ we can attach to $v_{\mu-2}$ an arbitrary value. So multiplying of h by $h_{\bar{e}_\mu}$, $h_{\bar{e}_{\mu-1}}$, $h_{\bar{e}_{\mu-2}}$ we can attach to $v_{\mu-3}$ an arbitrary value and so on. So we can transform coordinates $v_1, \ldots, v_{\mu-2}$ into arbitrary coordinates. Now consider two cases.

1. μ is even. Then

$$h(\bar{e}_1+\bar{e}_3+\ldots+\bar{e}_{\mu-1})=\bar{e}_1+\bar{e}_3+\ldots+\bar{e}_{\mu-1}.$$

Hence $v_1+v_3+\ldots+v_{\mu-1}=0$. Let $v_1=\ldots=v_{\mu-2}=0$. Then $v_{\mu-1}=0$. But $v_\mu+v_{\mu-1}=1$, so $v_\mu=1$.

2. μ is odd. We can attach v to one of the following forms: $(0,0,\ldots,0,1,0)$ or $(0,\ldots,0,1)$. If $v=(0,\ldots,0,1,0)$ then there exists $h_1 \in M_2$ such that multiplying of h by h_1 from the left we can attach v to $(1,\ldots,1,a,b)$, where $a,b \in \mathbb{Z}_2$. So multiplying of hh_1 by the matrixes of transvections in vectors $\bar{e}_{\mu-2}$, $\bar{e}_{\mu-4}$, $\bar{e}_{\mu-6}, \ldots, \bar{e}_3, \bar{e}_1$ we can attach v to $(1,1,\ldots,1,b,a)$. Then multiplying by h_1^{-1} from the left we can attach v to $(0,0,\ldots,0,1)$. This proves Lemma 6.

Now we may suppose that the μ-th row of h has the form $(0,\ldots,0,1)$. From the Propositions 2 and 3 follows that the left upper $(\mu-1)\times(\mu-1)$ corner of h satisfies the hypothesis of Theorem 5 for $A_{\mu-1}$. Therefore multiplying of h by the matrices of transvections $h_{\bar{e}_1}, \ldots, h_{\bar{e}_{\mu-1}}$ we can transform this left upper $(\mu-1)\times(\mu-1)$ corner into the unit matrix. Then $h=\mathrm{Id}+\delta$ where all elements of matrix δ, excluding the elements of the last column, are equal to zero. From the preservation of form Φ_2 and of the kernel of Φ_2 it follows that $\delta=0$. This proves Theorem 5.

Acknowledgements. The author thanks A.N. Varchenko and V.I. Arnold for valuable discussions and B. Martin, G.A. Nikitina for the help in preparing the manuscript.

References

1. Chmutov, S.V.: Monodromy groups of critical points of function. Invent. Math. **67**, 123-131 (1982)
2. Wajnryb, B.: On the monodromy groups of plane curve singularities. Math. Ann. **246** (N°2), 141-154 (1980)
3. Ebeling, W.: On the monodromy groups of singularities. Preprint (1981)
4. Milnor, J.: Singular points of complex hypersurfaces. Princeton: University Press, 1968

Oblatum 30-V-1983