SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Mathematische Annalen
Verlag: Springer
Jahr: 1983
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN235181684_0263
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN235181684_0263

Übergeordnetes Werk

Werk Id: PPN235181684
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN235181684
OPAC: http://opac.sub.uni-goettingen.de/DB=1/PPN?PPN=235181684

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

Mathematische Annalen

Band 263 1983

Begründet 1868 durch *Alfred Clebsch* und *Carl Neumann, früher herausgegeben von Alfred Clebsch* (1869 – 1872), *Carl Neumann* (1869 – 1876), *Felix Klein* (1876 – 1924), *Adolph Mayer* (1876 – 1901), *Walter v. Dyck* (1888 – 1921), *David Hilbert* (1902 – 1939), *Otto Blumenthal* (1906 – 1938), *Albert Einstein* (1920 – 1928), *Constantin Carathéodory* (1925 – 1928), *Erich Hecke* (1929 – 1947), *Bartel, L. van der Waerden* (1934 – 1968), *Franz Rellich* (1947 – 1955), *Kurt Reidemeister* (1947 – 1963), *Richard Courant* (1947 – 1968), *Heinz Hopf* (1947 – 1968), *Gottfried Köthe* (1957 – 1971), *Heinrich Behnke* (1938 – 1972), *Max Koecher* (1968 – 1976), *Lars Garding, Lund* (1970 – 1978), *Konrad Jörgens* (1972 – 1974), *Fritz John* (1968 – 1979), *Peter Dombrowski* (1970 – 1983)
Band 1 – 80 Leipzig, B. G. Teubner, ab Band 81 (1920) Berlin, Springer

Herausgegeben von
Herbert Amann, Zürich · **Heinz Bauer,** Erlangen
Louis Boutet de Monvel, Paris · **Wulf-Dieter Geyer,** Erlangen
Hans Grauert, Göttingen · **Günter Harder,** Bonn
Friedrich Hirzebruch, Bonn · **Reinhold Remmert,** Münster
Elmar Thoma, München

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

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 0025-5831, 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.

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 Verlags in irgendeiner Form – durch Fotokopie, Mikrofilm oder andere Verfahren – reproduziert oder in eine von Maschinen, insbesondere 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 gewerblich Unternehmens hergestellte oder benützte Kopie dient gewerblichen Zwecken gem. § 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.

Springer-Verlag Berlin Heidelberg New York Tokyo

Inhalt/Contents

Arendt, W., Cannière, J. De: Order Isomorphisms of Fourier-Stieltjes Algebras 145
Backes, E., Reckziegel, H.: On Symmetric Submanifolds of Spaces of Constant Curvature 419
Barrett, D. E.: Boundary Analyticity of Proper Holomorphic Maps of Domains with Non-Analytic Boundaries 479
Berndtsson, B.: A Formula for Interpolation and Division in $\mathbb{C}^n$ 399
Bland, J. S.: Local Boundary Regularity of the Canonical Einstein-Kähler Metric on Pseudoconvex Domains . 289
Borell, C.: Capacitary Inequalities of the Brunn-Minkowski Type 179
Brenner, J. L., Lyndon, R. C.: Maximal Nonparabolic Subgroups of the Modular Group 1
Cannière, J. De, see Arendt, W. 145
Cassou-Noguès, Ph., Taylor, M. J.: Local Root Numbers and Hermitian—Galois Module Structure of Rings of Integers 251
Deo, S.: On Tautness in Sheaf Cohomology 221
Esnault, H., Viehweg, E.: Sur une minoration du degré d'hypersurfaces s'annulant en certains points 75
Feingold, A. J., Frenkel, I. B.: A Hyperbolic Kac-Moody Algebra and the Theory of Siegel Modular Forms of Genus 2 87
Fitzpatrick, P. M., Massabò, I., Pejsachowicz, J.: Global Several-Parameter Bifurcation and Continuation Theorems: a Unified Approach Via Complementing Maps 61
Frenkel, I. B., see Feingold, A. J. 87
Friedberg, S.: On Maass Wave Forms and the Imaginary Quadratic Doi—Naganuma Lifting 483
Friedman, R., Miranda, R.: Smoothing Cusp Singularities of Small Length 185
Godini, G.: On a Problem of H. Berens 279
Griess, R. L., Jr.: Quotients of Infinite Reflection Groups 267
Haine, L.: Geodesic Flow on SO(4) and Abelian Surfaces 435
Hanges, N., Treves, F.: Propagation of Holomorphic Extendability of CR Functions 157
Jayne, J. E., Rogers, C. A.: Invariance of Borel Classes in Metric Spaces 323
Jennane, B.: Remarques sur les ouverts localement de Stein 371
Johnson, F. E. A.: Automorphisms of Direct Products of Groups and Their Geometric Realisations 343
Lempert, L.: Solving the Degenerate Complex Monge-Ampère Equation with One Concentrated Singularity 515
Lyndon, R. C., see Brenner, J. L. 1
Maeda, S.: Real Hypersurfaces of Complex Projective Spaces 473
Massabò, I., see Fitzpatrick, P. M., et al 61
Miranda, R., see Friedman, R. 185
Monsky, P.: The Hilbert-Kunz Function 43
Monsky, P.: p-Ranks of Class Groups in Z_p^d-Extensions 509
Nishiguchi, K.: Kodaira Dimension is not Necessarily Lower Semi-Continuous Under Degenerations of Surfaces 377
Nori, M. V.: The Fundamental Group-Scheme of an Abelian Variety 263
Patrizio, G., Wong, P.-M.: Stability of the Monge-Ampère Foliation 13
Pejsachowicz, J., see Fitzpatrick, P. M., et al 61
Rankin, R. A.: Sums of Powers of Cusp Form Coefficients 227
Reckziegel, H., see Backes, E. 419
Rogers, C. A., see Jayne, J. E. 323
Santharoubane, L. J.: Kac-Moody Lie Algebras and the Universal Element for the Category of Nilpotent Lie Algebras 365
Schottenloher, M.: Spectrum and Envelope of Holomorphy for Infinite Dimensional Riemann Domains ... 213
Strømme, S. A.: Deforming Vector Bundles on the Projective Plane 385
Stuart, C. A.: A Variational Approach to Bifurcation in L^p on an Unbounded Symmetrical Domain 51
Swartz, C.: The Schur Lemma for Bounded Multiplier Convergent Series 283
Taylor, M. J., see Cassou-Noguès, Ph. 251
Terao, H.: The Bifurcation Set and Logarithmic Vector Fields 313
Treves, F., see Hanges, N. 157
Tromba, A. J.: A Sufficient Condition for a Critical Point of a Functional to be a Minimum and Its Application to Plateau's Problem 303
Vanhecke, L., Willmore, T. J.: Interaction of Tubes and Spheres 31
Viehweg, E., see Esnault, H. 75
Willmore, T. J., see Vanhecke, L. 31
Wittmann, R.: Shilov Points and Shilov Boundaries 237
Wong, P.-M., see Patrizio, G. 13

Covered by Zentralblatt für Mathematik and Current Mathematical Publications
Printers: Brühlsche Universitätsdruckerei, Giessen

Printed in Germany © Springer-Verlag GmbH & Co KG Berlin Heidelberg 1983

Z. 5039

Mathematische Annalen

Band 263 Heft 1 1983

1-4 TD 94 308..

Begründet 1868 durch Alfred Clebsch · Carl Neumann
Fortgeführt durch Felix Klein · David Hilbert
Otto Blumenthal · Erich Hecke · Heinrich Behnke

Herausgegeben von
Herbert Amann, Zürich · **Heinz Bauer,** Erlangen
Louis Boutet de Monvel, Paris · **Wulf-Dieter Geyer,** Erlangen
Hans Grauert, Göttingen · **Günter Harder,** Bonn
Friedrich Hirzebruch, Bonn · **Reinhold Remmert,** Münster
Elmar Thoma, München



Springer-Verlag
Berlin Heidelberg New York Tokyo

8 Z Nat. 593

Niedersächsische Staats- u
Universitätsbibliothek
Göttingen
4. Mai 1983

208 Math. Ann. ISSN 0025-5831 MAANA3 263 (1) 1–144 (1983) April 1983

V 82

Mathematische Annalen

Grundsätzlich dürfen nur Arbeiten eingesandt werden, die nicht gleichzeitig an anderer Stelle zur Veröffentlichung eingereicht oder bereits veröffentlicht worden sind. Der Autor verpflichtet sich, seinen Beitrag auch nachträglich nicht an anderer Stelle zu publizieren. Mit der Annahme des Manuskripts und seiner Veröffentlichung durch den Verlag geht das Verlagsrecht für alle Sprachen und Länder einschließlich des Rechts der photomechanischen Wiedergabe oder einer sonstigen Vervielfältigung an den Verlag über.
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 benutzte 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.
Die Autoren erhalten von ihrer Arbeit zusammen 100 Sonderdrucke unentgeltlich. Zusätzliche Exemplare können bei Rücksendung der Korrekturabzüge bestellt werden.
Die Manuskripte müssen nach den in jedem Band erscheinenden „Hinweisen für die Autoren" vorbereitet sein. Die Reihenfolge der Veröffentlichungen behält sich die Redaktion vor.
Arbeiten werden nicht vom geschäftsführenden Herausgeber entgegengenommen: sie sollen an den Herausgeber gesandt werden, dessen Gebiet dem der Arbeit am nächsten steht.

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, trademarks, etc., in this publication, even if the former are not specifically identified, is not to be interpreted as a sign 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 0025-5831, 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.

100 offprints of each paper are provided free of charge; additional copies may be ordered when the proofs are returned. Manuscripts must be prepared according to the "Instructions to Authors" printed in every volume. The editors reserve the right to publish manuscripts in any order they think fit.

Papers must not be sent to the managing editor; they should only be submitted to the editor whose field is nearest to that of the paper.

Herausgeber/Editors

H. Amann — Mathematisches Institut, Universität Zürich, Rämistrasse 74, CH-8001 Zürich, Schweiz
H. Bauer — Mathematisches Institut der Universität, Bismarckstrasse 1½, D-8520 Erlangen, Federal Republic of Germany
L. Boutet de Monvel — École Normale Supérieure, Université de Paris, 45, Rue d'Ulm, F-75230 Paris Cedex 05, France
W.-D. Geyer — Mathematisches Institut der Universität, Bismarckstrasse 1½, D-8520 Erlangen, Federal Republic of Germany
G. Harder — Mathematisches Institut der Universität, Wegelerstrasse 10, D-5300 Bonn, Federal Republic of Germany
F. Hirzebruch — Max-Planck-Institut für Mathematik, Gottfried-Claren-Str. 26, D-5300 Bonn 3, Federal Republic of Germany
R. Remmert — Berliner Strasse 7, D-4540 Lengerich, Federal Republic of Germany
E. Thoma — Mathematisches Institut der Technischen Universität, Arcisstrasse 21, D-8000 München 2, Federal Republic of Germany

Geschäftsführender Herausgeber/Managing Editor

H. Grauert — Mathematisches Institut der Universität, Bunsenstrasse 3—5, D-3400 Göttingen, Federal Republic of Germany

NIEDERSÄCHS. STAATS- U. UNIV.- BIBLIOTHEK GÖTTINGEN

Subscription Information. Volumes 263—266 (4 issues each) will appear in 1983. The price for each volume is DM 424.00 or approx. US $ 177.00. Prices for back-volumes are available on request. Correspondence concerning subscriptions should be addressed to the publisher.
North America. Annual subscription rate: Approx. US $ 727.00 including carriage charges. Subscriptions are entered with prepayment only. Orders should be addressed to: Springer-Verlag New York Inc., Journal Sales Department, 44 Hartz Way, Secaucus, NJ 07094, USA, Tel. 201/348-4033, Telex 0023 125994.
All Other Countries. Annual subscription rate: DM 1696.00 plus carriage charges. Airmail delivery on request only. For Japan, carriage charges (surface airmail lifted) are DM 191.20; for India DM 85.60. Single issue price: DM 127.20 plus carriage charges. Orders can either be placed with your bookdealer or sent directly to: Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33, Tel. 030/8207-01, Telex 01 83319.
Changes of address: Allow six weeks for all changes to become effective. All communications should include both old and new addresses (with postal codes) and should be accompanied by a mailing label from a recent issue.
Microform. Microform editions are available from: University Microfilms International, 300 N. Zeeb Road, Ann Arbor, MI 48106, USA

Production
Journal Production Department I, Springer-Verlag
Postfach 10 52 80, D-6900 Heidelberg 1, Federal Republic of Germany
Tel. 06221/487-433, Telex 04 61723

Math. Ann. 263, 1–11 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Maximal Nonparabolic Subgroups of the Modular Group

J. L. Brenner[1] and R. C. Lyndon[2]

1. 10 Phillips Road, Palo Alto, CA 94303, USA
2. Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

0. Introduction

The *modular group* $M=\mathrm{PSL}(2,\mathbb{Z})$ may be defined as the group of all transformations of the set $\mathbb{C}\cup\{\infty\}$ of the form $z\mapsto\dfrac{az+c}{bz+d}$ for $a, b, c, d\in\mathbb{Z}$ and $ad-bc=1$. It is well known that M has a presentation

$$M=\langle A,B: A^2=B^3=1\rangle, \tag{0.1}$$

where A and B are the transformations

$$A: z\mapsto -1/z, \qquad B: z\mapsto\frac{z-1}{z}.$$

The *elliptic* elements of M, each with two conjugate complex fixed points, are precisely the conjugates of nontrivial powers of A and B. The *parabolic* elements, each with a single real fixed point, are precisely the conjugates of nontrivial powers of $C=AB: z\mapsto z+1$. The remaining nontrivial elements of M are *hyperbolic*, each with two real fixed points.

A subgroup S of M is torsionfree (and therefore by the Kurosh Subgroup Theorem a free group) if and only if it contains no elliptic elements. A subgroup S of M is called *nonparabolic* if it contains no parabolic elements. Neumann [9] showed that, if T is the infinite cyclic group generated by a conjugate of C, and S is a complement to T in M in the sense that $S\cap T=1$ and $ST=M$, then S is a maximal nonparabolic subgroup of M. We call such subgroups *Neumann subgroups*. Magnus [8] raised the question whether all maximal nonparabolic subgroups of the modular group are Neumann subgroups. We show that this is not the case.

We show that the kernel N of the obvious map from M onto the planar crystallographic group

$$Q=\langle A_1, B_1: A_1^2=B_1^3=(A_1B_1A_1B_1^{-1})^3=1\rangle$$

is a nonparabolic subgroup, indeed maximal in the class of normal nonparabolic subgroups. We show further that none of the maximal nonparabolic subgroups S containing N is a Neumann subgroup. Among the infinitely many such maximal nonparabolic subgroups S containing N, the simplest is perhaps the group S generated by the involutions $A_n = C^n BAB^{-1}C^{-n}$; this group is the free product $S = *\langle A_n \rangle$ of the countably many groups $\langle A_n \rangle$ of order 2.

Our method is to study subgroups S of M by reference to their coset graphs. This method has been employed by many workers, notably in the recent study by Conder [4, 5] of quotient groups of the triangle groups $(2, 3, k)$. It has been applied by Stothers [11–17] and by the present authors [1–3] to the study of nonparabolic subgroups of the modular group. In particular, the observations in Sect. 1 below are formulated by Stothers [13] and, especially, [14].

1. Coset Graphs

If S is any subgroup of M, the *coset graph* $\Gamma(S)$ of S, relative to the set of generators A and B, is defined as follows. The set $V(\Gamma(S))$ of vertices of $\Gamma(S)$ is the set of cosets Sg, $g \in M$. There are exactly three (directed) edges out of each vertex $V = Sg$: an *A-edge* from v to $vA = SgA$, a *B-edge* from v to vB, and a *B^{-1}-edge* from v to vB^{-1}. The inverses of these three edges are the A-edge from vA to v, the B^{-1}-edge from vB to v, and the B-edge from vB^{-1} to v. All graphs considered here will be embedded in the plane or in some other manifold; the edges will be represented by directed arcs, the inverse of an edge being represented by the same arc with opposite orientation.

We define a *(2, 3)-graph* to be connected graph Γ whose set of directed edges is divided into three disjoint sets, of A-edges, B-edges, and B^{-1}-edges, subject to the following conditions.

(i) At each vertex there is exactly one A-edge, one B-edge, and one B^{-1}-edge.

(ii) The inverse of an A-edge is an A-edge, the inverse of a $B^{\pm 1}$-edge is a $B^{\mp 1}$-edge.

(iii) At each vertex v, the B-edge at v is either a loop ending at v, or is one in a cycle of three B-edges.

Evidently every coset graph $\Gamma(S)$ of a subgroup S of M is a (2, 3)-graph. For the converse, let Γ be a (2, 3)-graph. Define a permutation A_1 of $V = V(\Gamma)$ to carry each vertex v to the vertex at the other end of the A-edge beginning at v, and define B_1 analogously. Then $A_1^2 = B_1^3 = 1$. In view of the presentation (0.1) of M, the map $A \mapsto A_1$, $B \mapsto B_1$ defines an action of M on V. Let v_0 be any vertex $v_0 \in V$, and let S be the stabilizer of v_0 under the action of M on V. If $g, h \in M$, then $v_0 g = v_0 h$ is equivalent to $gh^{-1} \in S$, hence to $Sg = Sh$. Thus we can define a bijection $\phi: v_0 g \mapsto Sg$ from V to $V(\Gamma(S))$. Since $(vgA)\phi = SgA = (vg)\phi A$ and $(vgB)\phi = (vg)\phi B$, the map ϕ is an isomorphism from the (2, 3)-graph Γ to the (2, 3)-graph $\Gamma(S)$. If the base point v_0 is changed, then S is replaced by a subgroup conjugate to it in M. This establishes the following.

(1.1) **Proposition.** *There is a bijective correspondence ϕ between the isomorphism classes of (2, 3)-graphs and the conjugacy classes of subgroups of M.*

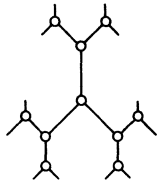


Fig. 1. The graph $\Gamma(1)$

Let $S_0 \leqq S \leqq M$. Then the inclusion map $S_0 g \mapsto Sg$ evidently defines a homomorphism from the $(2,3)$-graph $\Gamma(S_0)$ onto the $(2,3)$-graph $\Gamma(S)$. This proves the following.

(1.2) **Proposition.** *Let S_0 be a subgroup of M. Then ϕ gives a bijective correspondence between the homomorphic images of $\Gamma(S_0)$ and the conjugacy classes of subgroups of M that contain S_0.*

(1.3) *Remark.* It follows from (1.2) that the $(2,3)$-graphs are precisely the homomorphic images of the coset graph $\Gamma(1)$ of the trivial subgroup 1 of M. This graph $\Gamma(1)$, which is a Cayley graph for M, may be described as the truncation of the (infinite) cubic tree; it is shown in Fig. 1.

An *A-loop* [or *B-loop*] in Γ is an A-edge [or B-edge] beginning and ending at the same point v. A *C-orbit* of Γ is an orbit of C under the action of M on the vertex set V of Γ.

(1.4) **Proposition.** *Let S be a subgroup of M. Then the S-conjugacy classes of subgroups of S of order 2 correspond bijectively to the A-loops in $\Gamma(S)$, and the S-conjugacy classes of subgroups of order 3 to the B-loops in $\Gamma(S)$. The S-conjugacy classes of parabolic subgroups of S correspond bijectively to the finite C-orbits of $\Gamma(S)$.*

Proof. A subgroup of order 2 in S is generated by an element of the form gAg^{-1}, $g \in M$. This determines in $\Gamma(S)$ a closed path π at the vertex $v_0 = S \cdot 1$ of the form $\pi = \gamma_1 \gamma \gamma_1^{-1}$, where γ_1 is a path from v_0 to $v_0 g$ and γ is an A-loop at $v_0 g$. The same path corresponds to another subgroup of order 2 in S if and only if the two subgroups are conjugate in S.

A subgroup of order 3, with generater gBg^{-1}, corresponds in the same way to a closed path π in $\Gamma(S)$, where γ is now a B-loop. A parabolic subgroup, with a unique generator of the form $gC^k g^{-1}$, $k \geqq 1$, corresponds to a closed path π in $\Gamma(S)$ in which γ is now a closed path with label $C^k = (AB)^k$, corresponding to a C-orbit of length k. □

(1.5) **Corollary.** *A subgroup S of M is a free group if and only if $\Gamma(S)$ contains no A-loops and no B-loops.*

(1.6) **Corollary.** *A subgroup S of M is nonparabolic if and only if all C-orbits of $\Gamma(S)$ are infinite.*

(1.7) **Corollary.** *A subgroup S of M is a Neumann subgroup if and only if $\Gamma(S)$ contains only a single C-orbit, which is infinite, that is, if and only if $\Gamma(S)$ is infinite and C acts transitively on the vertices of $\Gamma(S)$.*

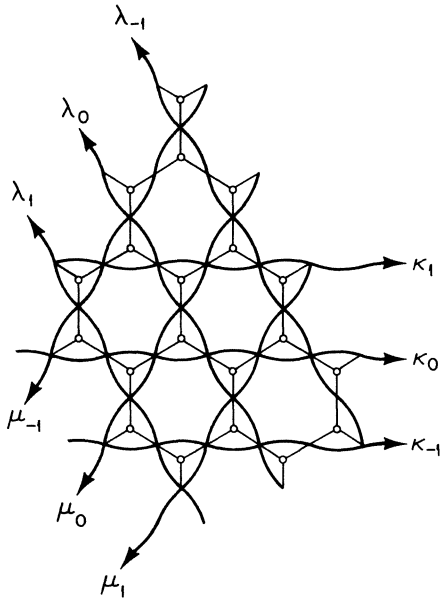


Fig. 2. The coset graph $\Gamma = \Gamma(N)$

2. The Hexagonal Graph

Let H be the 1-skeleton of the regular tessellation of the Euclidean plane by hexagons. The symmetry group of H contains, with index 2, a crystallographic group Q with a presentation

$$Q = \langle A_1, B_1 : A_1^2 = B_1^3 = (A_1 B_1 A_1 B_1^{-1})^3 = 1 \rangle .$$

(See Coxeter-Moser [7, pp. 48–49], especially formula (4.5131); see also Coxeter [6] and Sinkov [10].)

Let N be the kernel of the map from M onto Q that carries A to A_1 and B to B_1. The coset graph $\Gamma = \Gamma(N)$, relative to the generators A and B, is also a Cayley graph for Q. In fact, Γ can be obtained by "truncating" H as follows. Let ε be a positive real number less than half the common length of the edges of H. About each vertex p of H we draw a circle c_p of radius ε. We now obtain Γ by deleting the parts of H interior to the circles c_p. The straight segments remaining from the edges joining points p and q of H now become the A-edges of Γ, joining points on c_p and c_q. Since A_1 represents a reflection of H, we are led to orient the circles c_p alternately positively and negatively, so that circles joined by an A-edge are oriented oppositely. The B-edges of Γ are now the arcs on all the circles c_p, three on each circle. They are oriented in the sense of the c_p.

The graph Γ is indicated in Fig. 2. In this figure the C-orbits are indicated by wavy lines running roughly parallel to the true paths. These true paths consist of successive edges that are alternately (straight) A-edges and (curved) B-edges. The C-orbits are all infinite, and fall into three families K, Λ, M, where each family consists of infinitely many orbits that are "parallel", that is, congruent under

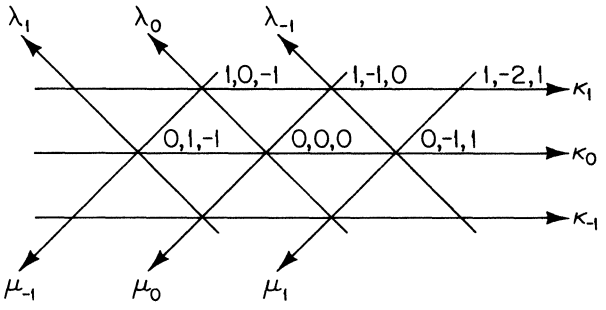


Fig. 3. The graph $\bar{\Gamma}$

translation. The figure also indicates a systematic way of indexing the elements κ_k, λ_l, μ_m of the families K, Λ, M by integers k, l, m. The indexing is arranged so that the three points on any circle c_p lie on three orbits κ_k, λ_l, μ_m, one from each family. Moreover, if c_p is any positively oriented circle, then $k+l+m=0$, and, conversely, if k, l, m are three integers such that $k+l+m=0$, then there is a unique positively oriented circle c_p such that the three points on c_p belong to κ_k, λ_l, and μ_m. We can thus describe the positively oriented circles c_p by the corresponding triple of coordinates (k, l, m). This coordinatization depends of course on the choice of direction of the families K, Λ, M and on the choice of a base point v_0.

For some purposes it is convenient to replace the graph Γ by a less complicated graph $\bar{\Gamma}$. In $\bar{\Gamma}$ each positively oriented circle c_p is replaced by a point $p=p(k,l,m)$, and the negatively oriented circles are omitted. Two points p and q in the graph $\bar{\Gamma}$ are joined by an edge in $\bar{\Gamma}$ if and only if the corresponding points p and q in the hexagonal graph H are at distance 2. The graph $\bar{\Gamma}$ is in fact a regular tessellation of the plane by triangles, and the infinite straight lines in $\bar{\Gamma}$ represent the C-orbits. The graph $\bar{\Gamma}$ is shown in Fig. 3.

We now consider a homomorphism ϕ of Γ onto a $(2,3)$-graph Γ^*. If two vertices v_1 and v_2 in Γ have the same image $v_1\phi=v_2\phi$, and if g is any element of M, then one sees that also $v_1g\phi=v_2g\phi$. It follows that the automorphism of Γ defined by $v_1g\mapsto v_2g$, for all $g\in M$, is a rigid motion of the plane. It follows further that if G is the group of all rigid motions γ such that $v\gamma\phi=v\phi$ for all v, then Γ^* is within isomorphism the quotient Γ/G of Γ by the action of G. Thus, to study homomorphic images Γ^* of Γ, and ultimately conjugacy classes of subgroups S that contain N, it suffices to study groups G of rigid motions of Γ. We shall, for brevity, call G and $\Gamma^*=\Gamma/G$ *nonparabolic* if and only if the corresponding subgroups S are nonparabolic.

We first consider rigid motions that preserve Γ and also preserve the orientation of the plane. Such a nontrivial rigid motion is either a translation or a rotation. A translation τ must carry each positively oriented circle $c_p=(k,l,m)$ to a second positively oriented circle $c_q=(k+a,l+b,m+c)$, where necessarily $a+b+c=0$. We write $\tau=\tau(a,b,c)$. Now τ carries some point of an orbit into a different point of the same orbit if and only if exactly one of a, b, c is 0. Thus a

nonparabolic group G cannot contain any nontrivial translation $\tau = \tau(a, b, c)$ one of whose coordinates is 0.

Any nontrivial rotation σ that preserves Γ, as a $(2,3)$-graph, must likewise carry every positively oriented circle c_p to another positively oriented circle c_q. Such a rotation σ must be of order 3, and can be of one of two kinds: it can carry some positively or negatively oriented circle c_p into itself, or it can have its center at the center of one of the truncated hexagons in Γ, permuting cyclically the three positively oriented circles c_p incident with that hexagon. In either case, σ permutes the three families K, Λ, M of C-orbits cyclically, hence does not carry any point of an orbit to a point on the same orbit.

We next consider a rigid motion ϱ that preserves Γ but reverses the orientation of the plane. Then ϱ must be either a reflection or a "glide reflection". Since ϱ reverses orientation, it must interchange two of the families of C-orbits while leaving the third invariant. We may suppose that ϱ leaves K invariant. Then the axis L of ϱ must be parallel to the C-orbits in the family K. Since ϱ must interchange positively oriented with negatively oriented circles, L must bisect perpendicularly the family of vertical A-edges in Fig. 2 that join each positively oriented circle c_p containing a vertex in a certain C-orbit $\kappa_k \in K$ with the negatively oriented circle c_q directly below it, which contains a vertex in the C-orbit κ_{k-1}. Thus ϱ, which reverses the linear order of the family K of orbits $(\ldots, \kappa_{-1}, \kappa_0, \kappa_1, \ldots)$, must interchange κ_k and κ_{k-1}.

Now ϱ^2, which preserves orientation, must leave invariant each orbit of the family K. If ϱ is contained in a nonparabolic group G, this can happen only if $\varrho^2 = 1$, that is, if ϱ is a reflection.

From the above we conclude that a group G of rigid motions preserving Γ is nonparabolic if and only if G contains no nontrivial translation $\tau = \tau(a, b, c)$ for which any of a, b, c is 0. In particular, although G may contain rotations and reflections, it can contain no proper glide reflection.

We are now prepared to enumerate the nonparabolic groups G of rigid motions that preserve Γ. We begin with those groups G that preserve orientation. Suppose first that $\tau_1 = \tau(a_1, b_1, c_1)$ and $\tau_2 = \tau(a_2, b_2, c_2)$ are two translations in G. By applying the Euclidean algorithm to the two vectors (a_i, b_i, c_i) in the usual way, we can replace the two generators τ_1 and τ_2 of the group $\langle \tau_1, \tau_2 \rangle$ that they generate by two generators τ_1' and τ_2' such that $a_2' = 0$. Since G is nonparabolic, this implies that $\tau_2' = 1$. We conclude that τ_1 and τ_2 are integer powers of the common element τ_1'; in particular, τ_1 and τ_2 are at least as long as τ_1'. It follows that the translation subgroup of G is either trivial or infinite cyclic, generated by one of the two translations in G of minimal length.

Suppose now that G contains a nontrivial rotation σ. If G also contains a translation τ, it contains another translation $\sigma\tau\sigma^{-1}$, in a different direction. Since τ and $\sigma\tau\sigma^{-1}$ are not powers of a common translation, this is not possible in the nonparabolic group G. Thus G contains no translation. Further, if G contained a nontrivial rotation σ_1 other than σ and σ^{-1}, then one of $\sigma_1\sigma$, $\sigma_1\sigma^{-1}$ would be a nontrivial translation. Thus G contains no nontrivial rotation other than σ and σ^{-1}.

We see that the orientation preserving groups G that preserve Γ are exactly as follows:

(i) $G_0 = 1$, the trivial group;

(ii) $G_\tau = \langle \tau \rangle$, infinite cyclic, generated by a translation τ not in the direction of any C-orbit;

(iii) $G_\sigma = \langle \sigma \rangle$, cyclic of order 3, generated by a rotation σ. Here there are two geometrically distinct cases according as: (iiia), σ preserves some circle c_p, or, (iiib), σ preserves no circle c_p.

Now suppose that G contains elements that reverse orientation. Then G contains a reflection ϱ, and G is generated by ϱ together possibly with some translations and rotations. Suppose that G contains a non trivial translation τ. Then it also contains the translation $\varrho\tau\varrho$, which must be parallel to τ. This is possible only if the direction of τ is either parallel or perpendicular to the axis L of ϱ. Since we have assumed that the axis L of ϱ is parallel to the orbits of the family K, the case that τ is parallel to L is excluded. Thus τ must be a translation in a direction perpendicular to L, that is, vertical in Fig. 2. Because τ must carry the positively oriented circle $c_p = (0,0,0)$ to another positively oriented circle c_q directly above or below c_p, inspection of Fig. 1 shows that τ must carry κ_0 to κ_n for some even $n = 2n_0$. If we choose τ in G of minimal length, then all other translations in G are powers of τ. Moreover, since G contains translations, it can contain no rotation. We conclude that G is generated by ϱ and τ, and is fully described by its action on K, on which it acts as an infinite dihedral group.

Suppose finally that G contains the reflection ϱ and also a nontrivial rotation σ. Then G contains no translation, and no nontrivial rotation other than σ and σ^{-1}. Since $\varrho\sigma\varrho$ is a rotation in a sense opposite to that of σ, it can only be σ^{-1}. The center of σ must therefore be on the axis L of ϱ, which contains no center of a circle c_p, and we conclude that the rotation subgroup of G must be of type (iiib).

From the above we see that the classes of nonparabolic groups G that preserve Γ but do not preserve orientation are exactly as follows:

(iv) $G_\varrho = \langle \varrho \rangle$, of order 2, generated by a reflection ϱ;

(v) $G_{\varrho,\tau} = \langle \varrho, \tau \rangle$, infinite dihedral, generated by a reflection ϱ together with a translation τ of even length $2n_0$ in a direction perpendicular to the axis L of ϱ;

(vi) $G_{\varrho,\sigma} = \langle \varrho, \sigma \rangle$, dihedral of order 6, generated by a reflection ϱ together with a rotation σ, of type (iiib), with center on the axis L of ϱ.

With this catalog of nonparabolic groups G in hand, we can proceed to verify that none of these groups G gives rise to a graph $\Gamma^* = \Gamma/G$ corresponding to a Neumann subgroup S of M, that is, that every case Γ^* has more than one C-orbit. This is obvious if $G = 1$. If $G = G_\tau$, a translation group, then it is clear that the images K^*, Λ^*, M^* of K, Λ, M are disjoint, and it is not difficult to check that K^*, Λ^*, M^* have the finite cardinalities $|a|$, $|b|$, $|c|$. Thus Γ^* has $n = |a| + |b| + |c|$ C-orbits, and every even integer $n \geqq 4$ occurs thus.

If $G = G_\sigma$, a rotation group, then G permutes K, Λ, M cyclically, and their common image $K^* = \Lambda^* = M^*$ in Γ^* is a single infinite family of C-orbits. If $G = G_\varrho$, we see similarly that Γ^* has two infinite families of C-orbits, K^* and $\Lambda^* = M^*$. The case $G = G_{\varrho,\tau}$, in which there are two finite families K^* and $\Lambda^* = M^*$ of cardinalities $|n|$ and $|n| - 1$, and the case $G = G_{\varrho,\sigma}$, in which there are infinitely many C-orbits, are more difficult to describe. These two cases are illustrated by Fig. 4 and 5.

We remark that the conjugacy class of groups S associated with $\Gamma^* = \Gamma/G$ will reduce to a single normal subgroup S of M if and only if G is a normal subgroup of

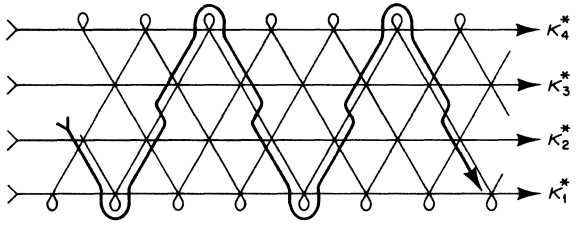


Fig. 4. The graph $\Gamma^* = \Gamma/G$ for $G = G_{\varrho,\tau}$ with $n_0 = 2$

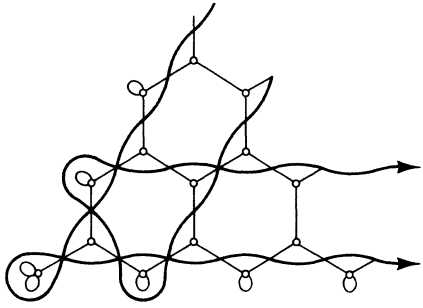


Fig. 5. The graph $\Gamma^* = \Gamma/G$ for $G = G_{\varrho,\sigma}$

the full symmetry group of Γ. Apart from the trivial group $G = 1$, none of the groups G listed above is normal in this full symmetry group. Therefore none of the nonparabolic subgroups S properly containing N is normal.

(2.1) **Theorem.** *N is maximal in the class of normal nonparabolic subgroups of M.*

We have also seen that none of the nonparabolic groups G listed above gives rise to a conjugacy class of Neumann subgroups. This proves the following.

(2.2) **Theorem.** *The nonparabolic subgroup N of M is contained in no Neumann subgroup.*

From this it follows that M contains maximal nonparabolic subgroups that are not Neumann subgroups. To find those that contain N, we determine the maximal groups among the nonparabolic groups G listed above. Clearly $G_0 = 1$ is not maximal. Let $G = G_\tau$ for $\tau = \tau(a, b, c)$ where $a + b + c = 0$ and $a, b, c \neq 0$. Evidently G_τ is not contained in a larger group of the same type if and only if a, b, and c are relatively prime, that is, if $(a, b) = 1$. However, even in this case, G may be contained in a group $G_{\varrho,\tau}$. If ϱ is chosen as before, with axis L parallel to the family K, then, in $G_{\varrho,\tau}$, the triple (a, b, c) has the form $(2n_0, -n_0, -n_0)$. In general, for G_τ not to be contained in some $G_{\varrho,\tau}$, we must further exclude the case that two of a, b, c are equal, and hence, assuming $(a, b) = 1$, have a common value $+1$ or -1.

Evidently a group $G = G_\sigma$ is maximal if and only if it is not contained in a group $G_{\varrho,\sigma}$. Thus G_σ is maximal if and only if it is of type (iiia).

Clearly no group $G = G_\varrho$ is maximal. Likewise, a group $G_{\varrho,\tau}$ is contained in a larger group $G_{\varrho,\tau'}$ if and only if the length $2n_0$ of τ is divisible by the length $2n_0'$ of

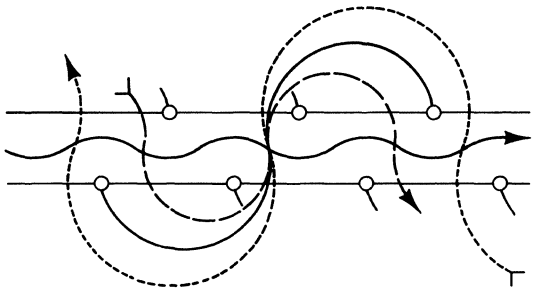


Fig. 6. The graph $\Gamma^* = \Gamma/G$ for $G = G_\tau$, with $\tau = \tau(1, -3, 2)$

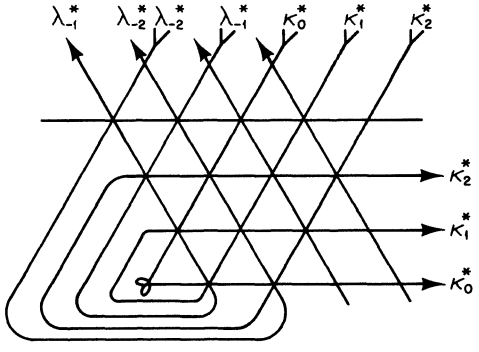


Fig. 7. The graph $\Gamma^* = \Gamma/G$ for $G = G_\sigma$

τ', and τ and τ' are parallel. Thus, among the $G_{\varrho,\tau}$, only those with τ of length $2n_0 = 2$ are maximal. Finally, the groups $G = G_{\varrho,\sigma}$ are obviously maximal. We summarize.

(2.3) **Proposition.** *The maximal nonparabolic subgroups S of M that contain N correspond to the graphs $\Gamma^* = \Gamma/G$, where G is of one of the following types:*

(i) G_τ, *where* $\tau = \tau(a, b, c)$, *a translation, such that* $a+b+c=0$, *that* $a, b, c \neq 0$, *that a and be are relatively prime, and that no two of a, b, c are $+1$ and no two are -1.*

(ii) G_σ, *where σ is a rotation of order 3 leaving invariant some B-orbit;*

(iii) $G_{\varrho,\tau}$, *where ϱ is a reflection with axis L parallel to a family of C-orbits and τ is a translation of length $2n_0 = 2$ in a direction perpendicular to L.*

(iv) $G_{\varrho,\sigma}$, *dihedral of order 6.*

We note that, in case (i), any permutation of a, b, c, or the replacement of all three by their negatives, yields an isomorphic graph Γ^*, and hence the same conjugacy class of maximal nonparabolic subgroups S of M. Nonetheless, infinitely many conjugacy classes are obtained in this case for various choices of the parameters a, b, c. Each of the remaining cases (ii), (iii), (iv) yields a single conjugacy class of maximal nonparabolic subgroups.

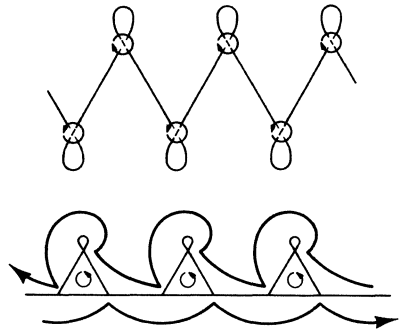


Fig. 8. Two representations of $\Gamma^* = \Gamma/G$ for $G = G_{\varrho,\tau}$ with $\tau = \tau(2, -1, -1)$, of length 2

In Fig. 6 we illustrate case (i) for the group G_τ given by $\tau = \tau(1, -3, 2)$. The resulting graph Γ^* has six C-orbits, the smallest number obtainable in case (i). Figs. 7 and 5 show the cases (ii) and (iv); these are admittedly less perspicuous.

Finally, in Fig. 8, we show Γ^* for the case (iii), first in the form in which it arises naturally, and again, redrawn in a simpler form. From the redrawn form it is clear that Γ^* has exactly two C-orbits, one running to the right and the other to the left, the second containing twice as many points of each B-orbit as the first. One can verify by direct inspection that every proper homomorphic image of Γ^* is finite, and indeed has either three vertices or a single vertex. It is also evident. upon choosing a suitable base point v_0, that the associated class of nonparabolic subgroups contains the subgroup S generated by the set of elements $A_n = C^n BAB^{-1}C^{-n}$ of order 2, for all $n \in \mathbb{Z}$, corresponding to the A-loops in Γ^*. The subgroup S is therefore the free product of the countably infinite set of subgroups of order 2 generated by the A_n. This graph Γ^* is in a reasonable sense the simplest graph yielding a nonparabolic subgroup S of M that is not a Neumann subgroup.

Acknowledgement. The second author gratefully acknowledges partial support of this research by the National Science Foundation.

References

1. Brenner, J.L., Lyndon, R.C.: Nonparabolic subgroups of the modular group. J. Algebra **77**, 311–322 (1982)
2. Brenner, J.L., Lyndon, R.C.: Permutations and cubic graphs. Pacific J. Math. **104**, 285–315 (1983)
3. Brenner, J.L., Lyndon, R.C.: The orbits of the product of two permutations. Eur. J. Com. (submitted)
4. Conder, M.D.E.: Generators for alternating and symmetric groups. J. London Math. Soc. **22**, 75–86 (1980)
5. Conder, M.D.E.: More on generators for alternating and symmetric groups. Quart. J. Math. Oxford **32**, 137–163 (1981)
6. Coxeter, H.S.M.: The groups determined by the relations $S^l = T^m = (S^{-1}T^{-1}ST)^p = 1$. Duke Math. J. **2**, 61–73 (1936)
7. Coxeter, H.S.M., Moser, W.O.J.: Generators and relations for discrete groups. In: Ergebnisse der Mathematik, No. 14. Berlin, Heidelberg, New York: 1972
8. Magnus, W.: Noneuclidean tesselations and their groups. London, New York: Academic Press 1974
9. Neumann, B.H.: Über ein gruppentheoretisch-arithmetisches Problem. Sitzungber. Preuss. Akad. Wiss. Phys. Math. Kl. No. 10 (1933)

10. Sinkov, A.: The groups determined by the relations $S^l = T^m = (S^{-1}T^{-1}ST)^p = 1$. Duke Math. J. **2**, 74–83 (1936)
11. Stothers, W.W.: Subgroups of the modular group. Proc. Camb. Phil. Soc. **75**, 139–153 (1974)
12. Stothers, W.W.: Subgroups of the (2, 3, 7) triangle group. Manus. Math. **20**, 323–334 (1977)
13. Stothers, W.W.: Subgroups of infinite index in the modular group. Glasgow Math. J. **19**, 33–43 (1978)
14. Stothers, W.W.: Diagrams associated with subgroups of Fuchsian groups. Glasgow Math. J. **20**, 103–114 (1979)
15. Stothers, W.W.: Subgroups of infinite index in the modular group. II. Glasgow Math. J. **22**, 101–118 (1981)
16. Stothers, W.W.: Subgroups of infinite index in the modular group. III. Glasgow Math. J. **22**, 119–131 (1981)
17. Stothers, W.W.: Groups of the second kind within the modular group. Ill. J. Math. **25**, 390–397 (1981)
18. Tretfoff, C.: Non-parabolic subgroups of the modular group. Glasgow Math. J. **16**, 91–102 (1975)

Received January 12, 1982; in revised form April 16, 1982

Math. Ann. 263, 13–29 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Stability of the Monge-Ampère Foliation

Giorgio Patrizio* and Pit-Mann Wong**
Department of Mathematics, University of Notre Dame, Notre Dame, IN 45556, USA

Introduction

In this paper we study complex manifold M with a C^α proper function $r: M \to (\alpha, \beta)$, $0 < \alpha < \beta \leqq \infty$ which satisfies (i) $i\partial\bar{\partial} r > 0$, (ii) $i\partial\bar{\partial} \log r \geqq 0$, and (iii) $(i\partial\bar{\partial} \log r)^{n+1} \equiv 0$, where $\dim_{\mathbb{C}} M = n+1$. More precisely, to each such pair (M, r) there is an associated foliation called the Monge-Ampère foliation, we study the stability of this foliation by considering the induced foliation on a level set $S = \{r = \text{const}\}$ which has the advantage of being compact. This is carried out in Sect. 2 by modifying a theorem of Edwards et al. [7].

The condition $(\partial\bar{\partial} \log r)^{n+1} \equiv 0$ suggested that (under suitable conditions) M^{n+1} is fibered over another manifold N^n so that the form $i\partial\bar{\partial} \log r$ is actually the pullback of a form on N. For example the form $i\partial\bar{\partial} \log |Z|^2$ on $\mathbb{C}^{n+1} - \{0\}$ is the pullback of the fundamental Kähler form on $\mathbb{C}P^n$. More generally on any negative (in the sense of Grauert) holomorphic vector bundle E over a projective variety there exists a Finsler metric r on E satisfying (i), (ii), and (iii) on $E - \{\text{zero section}\}$ where $\frac{i}{2\pi}\partial\bar{\partial} \log r$ is the Chern form of the canonical line bundle L over the projective bundle $\mathbb{P}(E)$ (a Finsler metric is identified with a hermitian metric on L).

In the above examples the associated Monge-Ampère foliations are holomorphic. However, this is not true generically. For instance if we take the unit ball B in $\mathbb{C}^{n+1}$ with $r = |Z|^2$, then the associated Monge-Ampère foliation on $B - \{0\}$ is holomorphic. However, by deforming the ball slightly so that strict convexity is still preserved, one has on the one hand domains which still possess exhaustion r satisfying (i), (ii), and (iii) outside of the origin and if the associated foliation were holomorphic, then such domains admit a one-parameter group of holomorphic automorphisms. On the other hand, results of Burns and Shnider [5] and Greene and Krantz [9] show that most (in fact an open dense set) of such domains have trivial automorphism group.

* Supported by a grant from Consiglio Nazionale delle Richerche
** Partially supported by NSF grant No. MCS 79-02571

In Sect. 2 (cf. Theorem 2.4) we give conditions so that (M, r) fibers over a compact Kähler manifold N with Kähler metric ω and $\pi^* \omega = i\partial\bar{\partial} \log r$. In Sect. 1 we begin by studying the pseudo-hermitian structure (cf. Webster [20]) on the level set S induced by the Kähler metric $i\partial\bar{\partial} r$ on M. These results are then used in Sect. 2 to relate the curvature of $i\partial\bar{\partial} r$ on M and the curvature of ω on N. By appealing to the results of Siu and Yau [15] on the Frankel conjecture we obtain conditions on r so that N is $\mathbb{C}P^n$. Of course, one can also obtain conditions on r so that N is a hyperquadric etc. by using the results of Siu [14].

In Sect. 3 we turn to the simplest situation of complete bounded proper circular domains with smooth boundaries in $\mathbb{C}^{n+1}$. These domains are defined by Finsler metrics on $\mathbb{C}^{n+1}$. We show the set of biholomorphism classes of such domains corresponds bijectively to the set $\{\Omega\}/\mathrm{Aut}\,\mathbb{C}P^n$, where $\{\Omega\}$ denotes the set of all real $(1,1)$ form ω of class C^∞ on $\mathbb{C}P^n$, which is cohomologous to the fundamental Kähler form Ω.

1. Pseudo-Hermitian Structures

Let M be a connected complex manifold of complex dimension $n+1$, $n \geqq 1$. Let $r > 0$ be a positive function of class C^∞ on M satisfying:

$$i\partial\bar{\partial} r > 0, \tag{1.1}$$

$$i\partial\bar{\partial} \log r \geqq 0, \tag{1.2}$$

$$(\partial\bar{\partial} \log r)^{n+1} \equiv 0. \tag{1.3}$$

Direct computation gives

$$\partial\bar{\partial} r / r = \partial \log r \wedge \bar{\partial} \log r + \partial\bar{\partial} \log r \tag{1.4}$$

from which we obtain via (1.1) and (1.3) that

$$0 < (i\partial\bar{\partial} r / r)^{n+1} = (n+1) i\partial \log r \wedge \bar{\partial} \log r \wedge (i\partial\bar{\partial} \log r)^n.$$

Hence we have

$$\partial \log r \neq 0 \quad \text{and} \quad (\partial\bar{\partial} \log r)^n \neq 0 \tag{1.5}$$

i.e. the function r has no critical points and the hermitian form $\partial\bar{\partial} \log r$ is of rank exactly n at every point of M.

Denote by Z the complex gradient vector field with respect to the Kähler metric $h = \sum r_{\alpha\bar{\beta}} dz^\alpha \otimes dz^{\bar{\beta}}$, where $r_{\alpha\bar{\beta}} = \partial^2 r / \partial z^\alpha \partial \bar{z}^\beta$. In other words Z is the vector field of type $(1,0)$ associated to the $(0,1)$ form $\bar{\partial} r$:

$$h(Z, W) = \bar{\partial} r(\bar{W}) \tag{1.6}$$

for all complex tangent vector W of type $(1,0)$.

The geometric meaning of the homogeneous Monge-Ampère equation (1.3) is reflected in the following properties of the complex gradient vector field Z:

$$\|Z\|_h^2 = r, \tag{1.7}$$

$$\partial \log r(Z) = 1, \tag{1.8}$$

$$\iota_Z \partial\bar{\partial} \log r = \iota_{\bar{Z}} \partial\bar{\partial} \log r = 0, \tag{1.9}$$

where ι denotes interior product;

$$L_Z \partial\bar{\partial} \log r = L_{\bar{Z}} \partial\bar{\partial} \log r = 0, \tag{1.10}$$

where L denotes Lie derivative;

$$\Delta'' r = -(n+1), \text{ where } \Delta'' = \bar{\partial}\bar{\partial}^* + \bar{\partial}^*\bar{\partial} \text{ is the } \bar{\partial}\text{-Laplacian of the Kähler metric } h; \tag{1.10'}$$

$$\Delta'' Z = 0, \tag{1.11}$$

$$\|\bar{\partial} Z\|^2 = -\varrho(Z, Z), \text{ where } \varrho \text{ is the Ricci tensor of the Kähler metric } h, \text{ in particular } Z \text{ is holomorphic iff } \varrho(Z, \bar{Z}) = 0. \tag{1.12}$$

The above identities are quite well-known, (1.7), (1.8), and (1.12) are in Stoll [16] and Burns [4]; the rest can be found in Wong [22].

Besides the metric $h = \partial\bar{\partial} r$, it is often convenient to use the hermitian metric $h' = \partial\bar{\partial} r / r$. The connection forms and curvature forms of these metrics are studied in Wong [22]. For our purpose it is necessary to extend the study in [22] a little further; in particular we will study the induced pseudo-hermitian structure as defined by Webster [20] on the level sets $\{r = \text{const}\}$.

Let $e_1, \ldots, e_{n+1}$ be a local unitary frame with respect to $h' = r^{-1} h$. By (1.7) we may choose $e_{n+1} = Z$ (in [22] Z was chosen to be e_1). Denote by $\theta^1, \ldots, \theta^{n+1}$ the unitary coframes. It follows from (1.8) that

$$\theta^{n+1} = \partial \log r. \tag{1.13}$$

Let $\omega = (\omega_\beta^\alpha)$ be the connection forms of h' with respect to the above frames, we have the structure equations:

$$d\theta^\alpha = \sum_{\beta=1}^{n+1} \theta^\beta \wedge \omega_\beta^\alpha + \eta^\alpha, \qquad 1 \leqq \alpha \leqq n+1, \tag{1.14}$$

where the torsion forms η^α are of type $(2,0)$ and $\omega = -{}^t\bar{\omega}$ is skew-hermitian; ω_β^α and ϱ^α are uniquely determined by these conditions.

On the other hand, $e^1/\sqrt{r}, \ldots, e^{n+1}/\sqrt{r}$ is unitary with respect to h and its dual $\sqrt{r}\theta^1, \ldots, \sqrt{r}\theta^{n+1}$. The structure equations are given by:

$$d(\sqrt{r}\theta^\alpha) = \sum_{\beta=1}^{n+1} \sqrt{r}\theta^\beta \wedge \varphi_\beta^\alpha, \qquad 1 \leqq \alpha \leqq n+1, \tag{1.15}$$

where $\varphi = (\varphi_\beta^\alpha)$ is skew-hermitian. The torsion forms are zero because h is Kähler.

Since $h = rh'$ the two connections D (of h) and ∇ (of h') are related by

$$D = \nabla + \partial \log r \otimes \mathrm{id}. \tag{1.16}$$

From (1.14) and (1.15) we get

$$\sum_{\beta=1}^{n+1} \theta^\beta \wedge \varphi_\beta^\alpha = \sum_{\beta=1}^{n+1} \theta^\beta \wedge \omega_\beta^\alpha + \tfrac{1}{2} d \log r \wedge \theta^\alpha + \eta^\alpha \tag{1.17}$$

and from (1.16) we obtain the following relationships:

$$\left.\begin{aligned}\varphi^{\alpha}_{\beta}&=\varphi^{\alpha}_{\beta},\\ \varphi^{\alpha}_{\alpha}&=\omega^{\alpha}_{\alpha}+\tfrac{1}{2}(\partial-\bar\partial)\log r.\end{aligned}\right\} \tag{1.18}$$

Substituting (1.18) into (1.17) we arrive at:

$$\eta^{\alpha}=\theta^{\alpha}\wedge\partial\log r=\theta^{\alpha}\wedge\theta^{n+1},\quad 1\leqq\alpha\leqq n+1, \tag{1.19}$$

so that $\eta^{n+1}=0$.

Let $\omega^{\alpha}_{\beta}=\omega^{\alpha'}_{\beta}+\omega^{\alpha''}_{\beta}$ be the decomposition into (1, 0) and (0, 1) parts, where they are related by:

$$\omega^{\alpha'}_{\beta}=-\omega^{\bar\alpha''}_{\bar\beta}, \tag{1.20}$$

where $\omega^{\bar\alpha''}_{\bar\beta}=:\overline{\omega^{\alpha''}_{\beta}}$. It was shown in Wong [22] that

$$\left.\begin{aligned}&\omega^{\alpha}_{n+1}=\theta^{\alpha}+\omega^{\alpha''}_{n+1},\quad \omega^{n+1}_{\alpha}=\omega^{n+1\,'}_{\alpha}-\theta^{\bar\alpha},\quad 1\leqq\alpha\leqq n,\\ &\omega^{n+1}_{n+1}=0\end{aligned}\right\} \tag{1.21}$$

and

$$\omega^{\alpha''}_{n+1}(\bar e_{n+1})=0,\quad 1\leqq\alpha\leqq n, \tag{1.22}$$

i.e. $\omega^{\alpha''}_{n+1}=0 \bmod \theta^{\bar 1},\ldots,\theta^{\bar n}$.

Consider now a level set of r, say $S=\{r=1\}=\{\log r=0\}$ and define a real 1-form θ:

$$\theta=\frac{i}{2}(\partial-\bar\partial)\log r, \tag{1.23}$$

so that

$$d\theta=-i\partial\bar\partial\log r=-i\sum_{\alpha=1}^{n}\theta^{\alpha}\wedge\theta^{\bar\alpha}. \tag{1.24}$$

Denote by $j:S\to M$ the inclusion map, then $j^*dr=0$ so that $j^*\theta=ij^*\theta^{n+1}$. We will drop j^* in the following discussions if no confusion arises,

$$\theta=i\theta^{n+1}\quad\text{on } S. \tag{1.25}$$

We have an induced metric on S given by:

$$ds^2=\theta\otimes\bar\theta+\operatorname{Re}\sum_{\alpha=1}^{n}\theta^{\alpha}\otimes\theta^{\bar\alpha}. \tag{1.26}$$

From the structure equations (1.14) and formulas (1.19) and (1.21) we have

$$d\theta^{\alpha}=\sum_{\beta=1}^{n}\theta^{\beta}\wedge\omega^{\alpha}_{\beta}+\theta^{n+1}\wedge\omega^{\alpha''}_{n+1},\quad 1\leqq\alpha\leqq n,$$

and by (1.25) we get

$$d\theta^{\alpha}=\sum_{\beta=1}^{n}\theta^{\beta}\wedge\omega^{\alpha}_{\beta}+\theta\wedge(-i\omega^{\alpha''}_{n+1}),\quad 1\leqq\alpha\leqq n, \tag{1.27}$$

on S. We have a basis of 1-forms $\{\theta, \theta^\alpha, \theta^{\bar{\alpha}}, \omega^\alpha_\beta, \omega^{\bar{\alpha}}_{\bar{\beta}}\}_{1\leqq\alpha,\beta\leqq n}$ of the pseudo-hermitian structure on S. The 1-forms $-i\omega^{\alpha''}_{n+1}$ [which are zero mod $\theta^{\bar{1}}, \ldots, \theta^{\bar{n}}$ by (1.22)] are the torsion forms of the pseudo-hermitian structure (denoted by τ^α in Webster [20]):

$$-i\omega^{\alpha''}_{n+1} = \tau^\alpha, \qquad 1\leqq\alpha\leqq n. \tag{1.28}$$

Let X be the real part of the complex gradient vector field $Z = e_{n+1}$, i.e. $Z = \frac{1}{2}(X - iJX)$. By the definition of θ and (1.8) we have

$$\theta(-JX) = \frac{i}{2}(\partial - \bar{\partial}) \log r(-i(Z - \bar{Z})) = 1. \tag{1.29}$$

By Proposition 2.2 in Webster [20], JX is an infinitesimal pseudo-conformal transformation iff the torsion forms $\tau^\alpha = -i\omega^{\alpha''}_{n+1}$ are zero. On the other hand, we have [keeping in mind that by (1.21) $\omega^{n+1}_{n+1} = 0$]:

$$\begin{aligned}\bar{\partial}Z &= (0,1) \text{ part of } \nabla Z = (0,1) \text{ part of } \nabla e_{n+1} \\ &= (0,1) \text{ part of } \sum_{\beta=1}^{n} \omega^\beta_{n+1} \otimes e_\beta = \sum_{\beta=1}^{n} \omega^{\beta''}_{n+1} \otimes e_\beta.\end{aligned}$$

Hence we have

Proposition 1.1. *JX is an infinitesimal pseudo-conformal transformation on S iff $\bar{\partial}Z|_S = 0$.*

Denote by $\Omega = (\Omega^\alpha_\beta)_{1\leqq\alpha,\beta\leqq n+1}$ the curvature matrix of the hermitian metric $h' = \partial\bar{\partial}r/r$, then

$$\Omega^\alpha_\beta = d\omega^\alpha_\beta - \sum_{\gamma=1}^{n+1} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma \tag{1.30}$$

and by (1.21) we obtain for $1\leqq\alpha,\beta\leqq n$

$$\begin{aligned}\Omega^\alpha_\beta &= d\omega^\alpha_\beta - \sum_{\gamma=1}^{n} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma + (\theta^{\bar{\beta}} + \omega^{\bar{\beta}''}_{n+1}) \wedge (\theta^\alpha + \omega^{\alpha''}_{n+1}) \\ &= d\omega^\alpha_\beta - \sum_{\gamma=1}^{n} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma - \theta^\alpha \wedge \theta^{\bar{\beta}} + \omega^{\bar{\beta}''}_{n+1} \wedge \theta^\alpha \\ &\quad + \theta^{\bar{\beta}} \wedge \omega^{\alpha''}_{n+1} + \omega^{\bar{\beta}''}_{n+1} \wedge \omega^{\alpha''}_{n+1}\end{aligned} \tag{1.31}$$

and if either $\alpha = n+1$ or $\beta = n+1$, then

$$\Omega^\alpha_\beta = d\omega^\alpha_\beta - \sum_{\gamma=1}^{n} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma. \tag{1.32}$$

On S we then have,

$$\Omega^\alpha_\beta = d\omega^\alpha_\beta - \sum_{\gamma=1}^{n} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma - \theta^\alpha \wedge \theta^{\bar{\beta}} - i\tau^{\bar{\beta}} \wedge \theta^\alpha + i\theta^{\bar{\beta}} \wedge \tau^\alpha + \tau^{\bar{\beta}} \wedge \tau^\alpha \tag{1.33}$$

for $1\leqq\alpha,\beta\leqq n$.

The curvature $\tilde{\Omega}=(\tilde{\Omega}^\alpha_\beta)_{1\leqq\alpha,\beta\leqq n}$ of the induced pseudo-hermitian structure as defined by Webster is given by:

$$\tilde{\Omega}^\alpha_\beta = d\omega^\alpha_\beta - \sum_{\gamma=1}^{n} \omega^\gamma_\beta \wedge \omega^\alpha_\gamma - i\theta^{\bar\beta} \wedge \tau^\alpha + i\tau^{\bar\beta} \wedge \theta^\alpha . \tag{1.34}$$

Thus the two structures on S are related by:

$$\Omega^\alpha_\beta = \tilde{\Omega}^\alpha_\beta - \theta^\alpha \wedge \theta^{\bar\beta} - \tau^\alpha \wedge \tau^{\bar\beta} + 2i\theta^\alpha \wedge \tau^{\bar\beta} - 2i\tau^\alpha \wedge \theta^{\bar\beta}, \tag{1.35}$$

where $1\leqq\alpha,\beta\leqq n$.

If JX is pseudo-conformal (e.g. Z is holomorphic), then (1.35) reduces to

$$\Omega^\alpha_\beta = \tilde{\Omega}^\alpha_\beta - \theta^\alpha \wedge \theta^{\bar\beta}, \qquad 1\leqq\alpha,\beta\leqq n, \tag{1.36}$$

on S. Furthermore, we obtain from (1.32), (1.21), (1.14), (1.19) that if JX is pseudo-conformal on S then

$$\left.\begin{aligned} &\Omega^\alpha_{n+1} = \Omega^{n+1}_\alpha = 0, \qquad 1\leqq\alpha\leqq n, \\ &\Omega^{n+1}_{n+1} = \sum_{\gamma=1}^{n} \theta^\gamma \wedge \theta^{\bar\gamma} = \partial\bar\partial \log r. \end{aligned}\right\} \tag{1.37}$$

We will also denote by $\Theta=(\Theta^\alpha_\beta)_{1\leqq\alpha,\beta\leqq n+1}$ the curvature matrix of $\partial\bar\partial r$. Then from (1.16) we see that

$$\Theta = \Omega + \partial\bar\partial \log r \otimes \mathrm{id}$$

so that we have

$$\left.\begin{aligned} &\Theta^\alpha_\beta = \Omega^\alpha_\beta, \qquad \alpha \neq \beta, \\ &\Theta^\alpha_\alpha = \Omega^\alpha_\alpha + \partial\bar\partial \log r = \Omega^\alpha_\alpha - \sum_{\alpha=1}^{n} \theta^\alpha \wedge \theta^{\bar\alpha}. \end{aligned}\right\} \tag{1.38}$$

From (1.38) and (1.37) we obtain (for JX pseudo-conformal) on S that

$$\Theta^\alpha_{n+1} = 0, \qquad 1\leqq\alpha\leqq n+1. \tag{1.39}$$

Following Webster we define the Ricci form of the pseudo-hermitian structure on S by $i\sum_{\alpha=1}^{n} \tilde{\Omega}^\alpha_\alpha$. We also say that the structure is Einstein if $\sum_{\alpha=1}^{n} \tilde{\Omega}^\alpha_\alpha = \lambda \sum_{\alpha=1}^{n} \theta^\alpha \wedge \theta^{\bar\alpha}$ for some $\lambda\in\mathbb{R}$.

From (1.35) we obtain

$$\sum_{\alpha=1}^{n} \tilde{\Omega}^\alpha_\alpha = \sum_{\alpha=1}^{n} \Omega^\alpha_\alpha + \sum_{\alpha=1}^{n} \theta^\alpha \wedge \theta^{\bar\alpha} + \sum_{\alpha=1}^{n} (\tau^\alpha \wedge \tau^{\bar\alpha} + 2i\tau^\alpha \wedge \theta^{\bar\alpha} - 2i\theta^\alpha \wedge \tau^{\bar\alpha})$$

so that for JX pseudo-conformal

$$\sum_{\alpha=1}^{n} \tilde{\Omega}^\alpha_\alpha = \sum_{\alpha=1}^{n} \Omega^\alpha_\alpha + \partial\bar\partial \log r \tag{1.40}$$

and from (1.38)

$$\sum_{\alpha=1}^{n} \tilde{\Omega}_{\alpha}^{\alpha} = \sum_{\alpha=1}^{n} \Theta_{\alpha}^{\alpha} + (n+1)\partial\bar{\partial}\log r$$
$$= \sum_{\alpha=1}^{n+1} \Theta_{\alpha}^{\alpha} + (n+1)\partial\bar{\partial}\log r. \tag{1.41}$$

Proposition 1.2. *If JX is pseudo-conformal then the pseudo-hermitian structure is Einstein iff the Ricci form ϱ of $\partial\bar{\partial}r$ satisfies $\varrho = i\lambda\partial\bar{\partial}\log r$ for some $\lambda \in \mathbb{R}$.*

As usual the holomorphic sectional curvatures are defined by:

$$\Omega(\xi) = \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \Omega_{\alpha\mu\bar{\nu}}^{\beta} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\mu} \xi^{\bar{\nu}} / \|\xi\|_h^2, \quad \xi \in T^{1,0}M,$$
$$\Theta(\xi) = \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \Theta_{\alpha\mu\bar{\nu}}^{\beta} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\mu} \xi^{\bar{\nu}} / \|\xi\|_h^2, \quad \xi \in T^{1,0}M,$$
$$\tilde{\Omega}(\xi) = \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n} \tilde{\Omega}_{\alpha\mu\bar{\nu}}^{\beta} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\mu} \xi^{\bar{\nu}} / \|\xi\|^2, \quad \xi \in HS,$$

where HS is the (maximal) complex tangent bundle of S and

$$\Omega_{\alpha}^{\beta} = \sum_{1 \leqq \mu, \nu \leqq n+1} \Omega_{\alpha\mu\bar{\nu}}^{\beta} \theta^{\mu} \wedge \theta^{\bar{\nu}}, \quad 1 \leqq \alpha, \ \beta \leqq n+1,$$
$$\Theta_{\alpha}^{\beta} = \sum_{1 \leqq \mu, \nu \leqq n+1} \Theta_{\alpha\mu\bar{\nu}}^{\beta} \theta^{\mu} \wedge \theta^{\bar{\nu}}, \quad 1 \leqq \alpha, \beta \leqq n+1,$$
$$\tilde{\Omega}_{\alpha}^{\beta} = \sum_{1 \leqq \mu, \nu \leqq n} \tilde{\Omega}_{\alpha\mu\bar{\nu}}^{\beta} \theta^{\mu} \wedge \theta^{\bar{\nu}} + \lambda_{\alpha}^{\beta} \wedge \theta, \quad 1 \leqq \alpha, \quad \beta \leqq n,$$

the last expression above is defined by Webster (cf. [20, p. 30, (1.35)]).

Under the assumption that JX is pseudo-conformal we obtain from (1.36)

$$\left.\begin{aligned} &\Omega_{\alpha\mu\bar{\nu}}^{\beta} = \tilde{\Omega}_{\alpha\mu\bar{\nu}}^{\beta}, && 1 \leqq \alpha, \beta \leqq n, \quad 1 \leqq \mu \neq \beta, \nu \neq \alpha \leqq n, \\ &\Omega_{\alpha\beta\bar{\alpha}}^{\beta} = \tilde{\Omega}_{\alpha\beta\bar{\alpha}}^{\beta} - 1, && 1 \leqq \alpha, \beta \leqq n. \end{aligned}\right\} \tag{1.42}$$

Thus for $\xi = \sum_{\alpha=1}^{n} \xi^{\alpha} e_{\alpha}$ in the holomorphic tangent bundle HS of S,

$$\Omega(\xi) = \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n} \Omega_{\alpha\mu\bar{\nu}}^{\beta} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\mu} \xi^{\bar{\nu}} / \|\xi\|^4$$
$$= \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n} \tilde{\Omega}_{\alpha\mu\bar{\nu}}^{\beta} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\mu} \xi^{\bar{\nu}} / \|\xi\|^4 - \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta \leqq n} \xi^{\alpha} \xi^{\bar{\beta}} \xi^{\beta} \xi^{\bar{\alpha}} / \|\xi\|^4,$$

so that we have

$$\Omega(\xi) = \tilde{\Omega}(\xi) - \tfrac{1}{2}, \quad \xi \in HS \tag{1.43}$$

and JX is pseudo-conformal.

Now from (1.38) we get

$$\left.\begin{aligned} &\Theta_{\alpha\mu\bar{\mu}}^{\alpha} = \Omega_{\alpha\mu\bar{\mu}}^{\alpha} - 1, && 1 \leqq \alpha \leqq n+1, \quad 1 \leqq \mu \leqq n, \\ &\Theta_{\alpha\mu\bar{\nu}}^{\beta} = \Omega_{\alpha\mu\bar{\nu}}^{\beta} && \text{otherwise} \end{aligned}\right\} \tag{1.44}$$

and so for $\xi \in T^{1,0}M$

$$\Theta(\xi) = r^{-2}(\Omega(\xi) - \tfrac{1}{2} \sum_{\substack{1 \leqq \alpha \leqq n+1 \\ 1 \leqq \mu \leqq n}} \xi^\alpha \xi^{\bar{\alpha}} \xi^\mu \xi^{\bar{\mu}} / \|\xi\|_{h'}^4) \geqq r^{-2}(\Omega(\xi) - \tfrac{1}{2}) \tag{1.45}$$

since $h = rh'$. From (1.45) we see immediately that if $\xi = \sum_{\alpha=1}^{n} \xi^\alpha e_\alpha$ is in HS, then since $r = 1$ we get

$$\Theta(\xi) = \Omega(\xi) - \tfrac{1}{2} \tag{1.46}$$

and if JX is pseudo-conformal we also have

$$\Theta(\xi) = \tilde{\Omega}(\xi) - 1, \qquad \xi \in HS. \tag{1.47}$$

Let $p = \mathrm{Re}\,\xi = \mathrm{Re} \sum \xi^\alpha e_\alpha$ and $q = \mathrm{Re}\,\eta = \mathrm{Re} \sum \eta^\alpha e_\alpha$, the (real) sectional curvature is defined by

$$\Omega(p \wedge q) = \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \Omega^\beta_{\alpha\mu\nu} (\xi^\alpha \eta^{\bar{\beta}} - \eta^\alpha \xi^{\bar{\beta}}) \overline{(\xi^\nu \eta^{\bar{\mu}} - \eta^\nu \xi^{\bar{\mu}})} / \|p \wedge q\|^2,$$

where

$$\|p \wedge q\|^2 = \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \{(\xi^\alpha \eta^{\bar{\beta}} - \eta^\alpha \xi^{\bar{\beta}}) \overline{(\xi^\nu \eta^{\bar{\mu}} - \eta^{\bar{\nu}} \xi^\mu)} + (\xi^\alpha \eta^\mu - \eta^\alpha \xi^\mu) \overline{(\xi^{\bar{\beta}} \eta^\nu - \eta^{\bar{\beta}} \xi^\nu)}\}.$$

Note that if $q = Jp$, i.e. $\eta^\alpha = i\xi^\alpha$, the above formula reduces to that of the holomorphic sectional curvature. The formulas for $\Theta(p \wedge q)$ and $\tilde{\Omega}(p \wedge q)$ are similar. A straight forward calculation gives:

$$\Omega(p \wedge q) = \tilde{\Omega}(p \wedge q) - \tfrac{1}{2} \sum_{1 \leqq \alpha, \beta \leqq n} |\xi^\alpha \eta^{\bar{\beta}} - \eta^\alpha \xi^{\bar{\beta}}|^2 / \|p \wedge q\|^2 \leqq \tilde{\Omega}(p \wedge q) - \tfrac{1}{2} \tag{1.48}$$

for $p = \mathrm{Re}\,\xi$, $q = \mathrm{Re}\,\eta$ where $\xi, \eta \in HS$; we also have

$$\Theta(p \wedge q) = \Omega(p \wedge q) - \tfrac{1}{2} \Big| \sum_{1 \leqq \alpha \leqq n+1} (\xi^\alpha \eta^{\bar{\alpha}} - \eta^\alpha \xi^{\bar{\alpha}}) \Big|^2 \Big/ \|p \wedge q\|^2 \tag{1.49}$$

for $p = \mathrm{Re}\,\xi$, $q = \mathrm{Re}\,\eta$ where $\xi, \eta \in T^{1,0}M$.

The holomorphic bisectional curvature is defined by:

$$\Omega(\xi \wedge \eta) = \tfrac{1}{2}\{\Omega(p \wedge q) + \Omega(p \wedge Jq)\},$$

where $p = \mathrm{Re}\,\xi$, $Jp = \mathrm{Re}\,i\xi$, $q = \mathrm{Re}\,\eta$, and $Jq = \mathrm{Re}\,i\eta$; $\xi, \eta \in T^{1,0}M$.

In terms of coordinates the above formula becomes:

$$\Omega(\xi \wedge \eta) = \tfrac{1}{4} \Big\{ \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \Omega^\beta_{\alpha\mu\bar{\nu}} (\xi^\alpha \eta^{\bar{\beta}} - \eta^\alpha \xi^{\bar{\beta}}) \overline{(\xi^\nu \xi^{\bar{\mu}} - \eta^\nu \xi^{\bar{\mu}})} / \|p \wedge q\|^2 + \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \Omega^\beta_{\alpha\mu\bar{\nu}} (\xi^\alpha \eta^{\bar{\beta}} + \eta^\alpha \xi^{\bar{\beta}}) \overline{(\xi^\nu \eta^{\bar{\mu}} + \eta^\nu \xi^{\bar{\mu}})} / \|p \wedge Jq\|^2 \Big\},$$

where

$$\|p \wedge Jq\|^2 = \sum_{1 \leqq \alpha, \beta, \mu, \nu \leqq n+1} \{(\xi^\alpha \eta^{\bar{\beta}} + \eta^\alpha \xi^{\bar{\beta}}) \overline{(\xi^\nu \eta^{\bar{\mu}} + \eta^\nu \xi^{\bar{\mu}})} + (\xi^\alpha \eta^\mu - \eta^\alpha \xi^\mu) \overline{(\xi^{\bar{\beta}} \eta^\nu - \eta^{\bar{\beta}} \xi^\nu)}\}.$$

The formulas for $\Theta(\xi\wedge\eta)$ and $\tilde{\Omega}(\xi\wedge\eta)$ are similar. Of course when $\xi=\eta$ these formulas reduces to those of the holomorphic sectional curvatures. We have the following relationships between the bisectional curvatures:

$$\Omega(\xi\wedge\eta)=\tilde{\Omega}(\xi\wedge\eta)-\frac{1}{4}\Big\{\sum_{1\leqq\alpha,\beta\leqq n}|\xi^\alpha\eta^{\bar\beta}-\eta^\alpha\xi^{\bar\beta}|^2/\|p\wedge q\|^2 + \sum_{1\leqq\alpha,\beta\leqq n}|\xi^\alpha\eta^{\bar\beta}+\eta^\alpha\xi^{\bar\beta}|^2/\|p\wedge Jq\|^2\Big\} \tag{1.50}$$

for $\xi,\eta\in HS$, $p=\mathrm{Re}\,\xi$, $q=\mathrm{Re}\eta$, and $Jq=\mathrm{Re}\eta$; and

$$\Theta(\xi\wedge\eta)=\Omega(\xi\wedge\eta)-\frac{1}{4}\Big\{\Big|\sum_{1\leqq\alpha\leqq n+1}(\xi^\alpha\eta^{\bar\alpha}-\eta^\alpha\xi^{\bar\alpha})\Big|^2\Big/\|p\wedge q\|^2 + \Big|\sum_{1\leqq\alpha\leqq n+1}(\xi^\alpha\eta^{\bar\alpha}+\eta^\alpha\xi^{\bar\alpha})\Big|^2\Big/\|p\wedge Jq\|^2\Big\} \tag{1.51}$$

for ξ and η in $T^{1,0}M$.

2. Stability of the Monge-Ampère Foliation

As in Sect. 1 we consider a complex manifold M of complex dimension $n+1$ with a C^∞ positive function $r>0$ satisfying (1.1), (1.2), and (1.3).

Consider the distribution

$$F=\mathrm{ann}\,\partial\bar\partial\log r=\{W\in T^{1,0}M|\partial\bar\partial\log r(W,\bar V)=0,\ \forall V\in T^{1,0}M\} \tag{2.1}$$

associated to r. By (1.3) and (1.4) the hermitian form $\partial\bar\partial\log r$ is of rank n (over $\mathbb{C}$) at every point of M. This implies that F is of rank 1 over $\mathbb{C}$, i.e. a complex line bundle. Furthermore we observe that F is integrable because $\partial\bar\partial\log r$ is a closed form. Thus it defines a foliation $\mathscr{F}$ of M whose leaves are Riemann surfaces. The foliation is holomorphic if F is a holomorphic sub-bundle (in which case so is the normal bundle). We will refer to $\mathscr{F}$ as the Monge-Ampère foliation associated to r. This foliation was first introduced by Bedford and Kalka [1] to study the complex Monge-Ampère equation. Further studies of this foliation can be found in Burns [4], Stoll [16, 17], and Wong [22]. The following results are well-known:

> F is a trivial bundle as it is spanned by the (non-vanishing) complex gradient vector field Z of r [cf. (1.7) and (1.9)]. It follows that the leaves of the foliation are the flows of Z and that F is holomorphic iff Z is holomorphic. (2.2)

> The foliation $\mathscr{F}$ is totally geodesic, i.e. each leaf is a totally geodesic submanifold, with respect to the Kähler metric $\partial\bar\partial r$. (2.3)

Define a sub-bundle of $T^{1,0}M$ by

$$H=\{W\in T^{1,0}M|\partial\log r(W)=0\}. \tag{2.4}$$

By (1.9), (1.8), and (2.2) we have

$$T^{1,0}M=F\oplus H, \tag{2.5}$$

where the hermitian metric $\partial\bar\partial/r=\partial\log r\wedge\bar\partial\log r+\partial\bar\partial\log r$ splits into a hermitian metric $\partial\log r\otimes\bar\partial\log r$ on F and a hermitian metric defined by $\partial\bar\partial\log r$ on H.

We now focus our attention on the level set $S=\{r=1\}$ *which we assume to be compact.* The bundle H restricts to S is the holomorphic tangent bundle of S. As in Sect. 1, a pseudo-hermitian structure on S is defined by

$$\theta=\frac{i}{2}(\partial-\bar{\partial})\log r \tag{2.6}$$

and the induced metric

$$ds^2=\theta\otimes\theta+\tfrac{1}{2}\operatorname{Re}\sum_{\alpha=1}^{n}\theta^{\alpha}\otimes\theta^{\bar{\alpha}}, \tag{2.7}$$

where $\partial\bar{\partial}\log r=\sum_{\alpha=1}^{n}\theta^{\alpha}\wedge\theta^{\bar{\alpha}}$.

Let $\tilde{\mathscr{F}}$ be the foliation on S induced by $\mathscr{F}$ on M. More precisely, denote by $\tilde{F}$ the subbundle of $T_{\mathbb{R}}S$ defined by:

$$\tilde{F}=\{Y\in T_{\mathbb{R}}S \mid \iota_Y\partial\bar{\partial}\log r=0\}, \tag{2.8}$$

where ι_Y denotes interior product with Y. Then $\tilde{F}$ is clearly of rank 1 (over $\mathbb{R}$), i.e. a real line bundle over S. In fact if we write $Z=\frac{1}{2}(X-iJX)$, then we have by (1.9):

$$\iota_{JX}\partial\bar{\partial}\log r=i\iota_{Z-\bar{Z}}\partial\bar{\partial}\log r=0 \tag{2.9}$$

and by (1.8):

$$JX\log r=i(Z-\bar{Z})\log r=i\{\partial\log r(Z)-\bar{\partial}\log r(\bar{Z})\}=0 \tag{2.10}$$

and also from (1.29) $\theta(-JX)=1$. These imply that $\tilde{F}$ is spanned by the non-vanishing vector field $-JX$. The leaves of $\tilde{\mathscr{F}}$ are the integral curves of $-JX$. We remark that one can show easily [cf. (2.3)] that the leaves are geodesic on S with the induced metric (2.7) which splits into fiber metric $\theta\otimes\theta$ for $\tilde{F}$ and $\operatorname{Re}\sum_{\alpha=1}^{n}\theta^{\alpha}\otimes\theta^{\bar{\alpha}}$ for $H|_S$.

We now assume furthermore that the leaves of the foliation $\tilde{\mathscr{F}}$ on S are closed. Define the volume function

$$\operatorname{vol}: S\to\mathbb{R}_+ \tag{2.11}$$

by $x\to\int_{L_x}\theta=\frac{i}{2}\int_{L_x}(\partial-\bar{\partial})\log r$, where L_x is the leaf of $\tilde{F}$ through a point $x\in M$. By our choice of orientation, the volume function is strictly positive on S.

First we recall a theorem of Edwards, Millet, and Sullivan:

Theorem [7]. *Let S be a compact differentiable manifold with a foliation whose leaves are also compact. If there exists a closed form θ on S of degree d (= dimension of each leaf) such that the volume function $\int_L\theta$ is of the same sign for all leaves L (i.e. always strictly positive or strictly negative), then the volume function is in fact bounded.*

In our situation the form $\theta = \frac{i}{2}(\partial - \bar{\partial})\log r$ is not closed so that the above theorem is not directly applicable. However we will see that the proof that leads to the above theorem works also in our case. The proof is based on the Moving Leaf Proposition in [7]. Denote by S' the set of all points $x \in S$ such that the volume function is unbounded on every neighborhood U of x. The set S' is closed in S (even when S is non-compact; the proposition does not require S to be compact) and one has:

Moving Leaf Proposition [7]. *Let S be a differentiable manifold with a compact foliation. If the set S' is compact (which is always the case if S is compact) and non-empty then there exists a homotopy of leaves $h: L \times [0, \infty] \to S - S'$ such that $\mathrm{vol} L_t \to \infty$; where $L_t = h(L, t)$ is a leave.*

In our case if S' is non-empty, we take a homotopy of leaves as in the proposition. Then for $0 < t < \infty$ we have

$$\int_{L_0} \theta - \int_{L_t} \theta = \int_{L \times [0,t]} h^* d\theta = \frac{i}{2} \int_{L \times [0,t]} h^*(\bar{\partial}\partial \log r) = 0$$

because $\iota_Y \partial\bar{\partial} \log r = 0$ for any Y tangent to the leaves [cf. (2.8)]. This shows that the volume of L_t is constant for all t, contradicting the conclusion of the Moving Leaf Proposition. Thus S' must be empty, i.e. the volume function is bounded.

A compact foliation is said to be stable if the volume function is bounded. This is equivalent to the finiteness of the holonomy group of the foliation or equivalently the Hausdorffness of the leaf space. Furthermore if the volume function is continuous then the holonomy group of the foliation is trivial (cf. [7]). In which case the canonical projection on to the leaf space $\tau: S \to N$ is a fiber bundle.

We summarize the above results in the following theorem:

Theorem 2.1. *Let r be a positive function of class C^∞ on M satisfying* (1.1), (1.2), *and* (1.3) *and let $S = \{r = 1\}$ be compact. If the Monge-Ampère foliation on S is compact, (i.e. each leaf is compact) then the volume function $x \to \int_{L_x} \theta$ is bounded, where $\theta = i(\partial - \bar{\partial})\log r$ defines the pseudo-hermitian structure on S.*

Remark. The above theorem is actually true for any strictly pseudo-convex CR hypersurface with pseudo-hermitian structure θ and the foliation defined by the vector field dual to θ.

We now consider the Monge-Ampère foliation $\mathscr{F}$ on M associated to $r: M \to (a, b)$ which we assumed to be proper. *We further assume that the leaves of the foliation are closed.* Let $Z = X - iJX$ then $Xr = \frac{1}{2}(Z + \bar{Z})r = r$ and $L_X \partial\bar{\partial} \log r = 0$ so that each leaf intersects each level set S transversally. Furthermore from (2.9), (2.10), and (1.29) we see that each component of the intersection is an integral curve of JX on S. Since X is non-vanishing and $Xr = r$, any level set S_α can be homotopically deformed onto any other level set S_β along the integral curves of X (cf. Milnor [13, Theorem 3.1]) sending the intersection of a leaf K with S_α onto the intersection of K with S_β so that the number of components of the intersections are the same. However since K is connected, we concluded that the intersection has

exactly one component. Thus the leaf space of $\mathcal{F}$ is the same as the leaf space $\tilde{\mathcal{F}}$ on any one of the level set. In particular the leaf space is Hausdorff by Theorem 2.1.

Corollary 2.2. *The Monge-Ampère foliation $\mathcal{F}$ is stable ($\Leftrightarrow$ the leaf space is Hausdorff) if the defining function r is proper and the leaves of $\mathcal{F}$ are closed in M.*

In the argument above we remark that the level sets are diffeomorphic. We will show that they are actually CR equivalent if the complex gradient vector field Z is holomorphic. We follow the proof of Theorem 3.1 in [13]. Given two level sets S_α and S_β, let $\varrho: M \to \mathbb{R}$ be a smooth function with compact support which is equal to one on an open neighborhood U of $r^{-1}([\alpha, \beta])$. Then $\tilde{X} = \varrho X$ is equal to $X = Z + \bar{Z}$ on U and is complete. Let φ_t be the 1-parameter group of diffeomorphism generated by $\tilde{X}$. Each $\varphi_t|_V : V \to \varphi_t(V)$ is biholomorphic provided that V and $\varphi_t(V)$ are contained in U. For fixed $q \in U$ we have

$$\frac{d}{dt} \log r(\varphi_r(q)) = \tilde{X} \log r = X \log r = (Xr)/r = 1$$

provided that $\varphi_t(q)$ is in U. Writing $u = \log r$, we then have

$$u(\varphi_t(q)) = t + u(q)$$

so that

$$u(\varphi_{\log\alpha - \log\beta}(q)) = \log\alpha - \log\beta + u(q) = \log\alpha$$

for all $q \in S_\beta = \{r = \beta\}$. Let $\lambda = \log\alpha - \log\beta$ then φ_λ is a biholomorphism from an open neighborhood of S_α onto an open neighborhood of S_β sending S_α onto S_β. Furthermore since φ_λ^* is biholomorphic we have $(\varphi_\lambda)^*(\partial - \bar{\partial}) \log r = (\partial - \bar{\partial}) \log r \circ \varphi_\lambda$. Thus φ_λ actually preserved the pseudo-hermitian structures defined by $i(\partial - \bar{\partial}) \log r$ on S_α and S_β.

Corollary 2.3. *Same assumption as in Corollary 2.2 and assume that the Monge-Ampère foliation $\mathcal{F}$ is holomorphic then for any two level sets $S_\alpha = \{r = \alpha\}$ and $S_\beta = \{r = \beta\} \exists CR$ isomorphism $\varphi: S_\alpha \to S_\beta$ which also preserves the pseudo-hermitian structures defined by $\frac{i}{2}(\partial - \bar{\partial}) \log r$ on S_α and S_β.*

As mentioned earlier if the volume function on a level set S is continuous then the projection $\pi: S \to N$ onto the leaf space is a fiber bundle. If the foliation F is holomorphic ($\Leftrightarrow$ Z is holomorphic) then N is a complex manifold and by (1.10) $L_Z \partial\bar{\partial} \log r = L_{\bar{Z}} \partial\bar{\partial} \log r = 0$ and since Z is holomorphic, this implies that the hermitian metric along the fiber of H is invariant along the leaves and so projects down to a Kähler metric ω on N.

Theorem 2.4. *Let $r: M \to (\alpha, \beta)$ be a proper positive function of class C^∞ satisfying* (1.1), (1.2), *and* (1.3). *Assume that the associated Monge-Ampère foliation is holomorphic and the leaves are closed. If the volume function* (2.11) *is continuous on any of the level sets of r, then the leaf space N is a compact Kähler manifold with Kähler form ω satisfying $\pi^*\omega = i\partial\bar{\partial} \log r$ where $\pi: M \to N$ is the canonical projection.*

Note that $\pi_*|_{HS}: HS \to TN$ is an isometry so that we obtain from Proposition 1.2 [(1.41) in particular] that:

Corollary 2.5. *Same assumption as in Theorem 2.4, then the leaf space (N, ω) is* (i) *Kähler Einstein if the Ricci form ϱ of the Kähler metric $\partial\bar{\partial}r$ of M is of the form $i\lambda\partial\bar{\partial}\log r$ for some $\lambda \in \mathbb{R}$,* (ii) *Ricci of $\omega = (n+1)\omega$ if $\varrho \equiv 0$,* (iii) *Ricci of ω is positive definite if $\varrho + i(n+1)\partial\bar{\partial}\log r > 0$.*

We can also use (1.47) or (1.50), (1.51) and Siu-Yau's solution to the Frenkel conjecture (cf. [15]) to obtain:

Corollary 2.6. *Same assumption as in Theorem 2.4. Then the leaf space N is biholomorphic to $\mathbb{C}P^n$ if either* (i) *the holomorphic sectional curvature of $(M, \partial\bar{\partial}r)$ is a negative constant $c < 0$ or* (ii) *the holomorphic bisectional curvature of $(M, \partial\bar{\partial}r)$ is negative.*

Of course there are other possibilities, for instance we can give conditions on $(M, \partial\bar{\partial}r)$ so that N is a hyperquadric etc. by using Siu [14].

3. Complete Proper Circular Domains

A Finsler metric on $\mathbb{C}^{n+1}$ is a non-negative function with the following properties

$$r(z) \geqq 0 \quad \text{and} \quad r(z) = 0 \quad \text{iff} \quad z = 0. \tag{3.1}$$

$$r \text{ is of class } C^\infty \text{ on } \mathbb{C}^{n+1} - \{0\} \text{ and continuous on } \mathbb{C}^{n+1} - \{0\}. \tag{3.2}$$

$$r(\lambda z) = |\lambda|^2 r(z) \quad \text{for} \quad \lambda \in \mathbb{C}. \tag{3.3}$$

From (3.3) we see that the real $(1,1)$ form $\partial\bar{\partial}\log r$ is invariant by $\mathbb{C}^*$, i.e. $\lambda^*(i\partial\bar{\partial}\log r) = i\partial\bar{\partial}\log r$, so that it projects down to a real $(1,1)$ form on $\mathbb{C}P^n$. In particular we have the Monge-Ampère condition:

$$(\partial\bar{\partial}\log r)^{n+1} \equiv 0 \quad \text{on} \quad \mathbb{C}^{n+1} - \{0\}. \tag{3.4}$$

Consider a bounded domain D in $\mathbb{C}^{n+1}$ with the following property:

if $z \in D$ and $\lambda \in C$ with $|\lambda| \leqq 1$ then

$$\lambda z = (\lambda z_1, \ldots, \lambda z_{n+1})$$

is also in D and if $z \in \bar{D}$ and $|\lambda| < 1$ then $\lambda z \in D$. (3.5)

Domains with such property are called complete proper circular domains. We will only consider those with smooth boundaries. To such a domain D let

$$m_D(z) = \inf\left\{\left|\frac{1}{t}\right| t \in \mathbb{R}^+ \text{ and } tz \notin D\right\}, \quad z \in \mathbb{C} \tag{3.6}$$

be the Minkowski functional of D and let

$$r_D = m_D^2. \tag{3.7}$$

Then it is immediate that r_D satisfies (3.1), (3.2), and (3.3) [hence also (3.4)] so that it is a Finsler metric. It is also clear that if in addition the domain D is strictly pseudo-convex then

$$i\partial\bar{\partial} r_D > 0 \quad \text{and} \quad i\partial\bar{\partial}\log r_D \geqq 0. \tag{3.8}$$

It is also clear that

$$r_D = 1 \text{ on } \partial D,\ r_D < 1 \text{ on } D \text{ and } r_D > 1 \text{ on the complement of } \bar{D}. \tag{3.9}$$

Proposition 3.1. *The complete proper bounded circular domains D_1 and D_2 with smooth boundaries are biholomorphic iff their corresponding Finsler metrics are related by $r_{D_1} = r_{D_2} \circ A$ for some $A \in \mathrm{GL}(n+1, \mathbb{C})$.*

Proof. If $r_{D_1} = r_{D_2} \circ A$ where $A \in \mathrm{GL}(n+1, \mathbb{C})$, then for $z \in D_1$ $r_{D_2} \circ A(z) = r_{D_1}(z) < 1$ so that we have $A(D_1) \subsetneqq D_2$. Similarly we have $A(D_2) \subsetneqq D_1$ thus $A|_{D_1} : D_1 \to D_2$ is a biholomorphism.

Conversely, by a theorem of Braun et al. [3], two bounded circular domains are biholomorphic iff they are linearly biholomorphic. Let A be such a linear map. It is easily verified that $r_{D_2} \circ A$ satisfies (3.1), (3.2), (3.3), and (3.4) on D_1 so that it must be the same as r_{D_1}.

Corollary 3.2. *Let D be a strictly pseudo-convex complete proper circular domain in $\mathbb{C}^{n+1}$. The following statements are equivalent:*

(i) *D is biholomorphic to the unit ball B in $\mathbb{C}^{n+1}$,*

(ii) *the associated Finsler metric r_D is of the form*

$$r_D(z) = \sum_{j=1}^{n+1} \left| \sum_{k=1}^{n+1} a_{jk} z_k \right|^2,$$

where

$$(a_{jk})_{1 \leqq j,k \leqq n+1} \in \mathrm{GL}(n+1, \mathbb{C}),$$

(iii) *r_D is of class C^∞ at the origin.*

Proof. (i)⇔(ii) follows from Proposition 3.1; (ii)⇒(iii) is immediate and (iii)⇒(i) because r_D satisfies (3.9), (3.4) and is smooth at the origin so that Stoll's characterization of the ball applies (cf. Stoll [16], Burns [4], and Wong [22]).

Example. Let $D = \{(z_1, z_2) \in \mathbb{C}^2 \mid |z_1|^2 + |z_2|^2 + |z_2|^4 < 1\}$, it is a strictly pseudo-convex complete proper circular domain. The associated Finsler metric is given by

$$r_D(z_1, z_2) = \frac{|z_1|^2 + |z_2|^2}{d^2(z_1, z_2)}$$

$$= \begin{cases} |z_1|^2, & \text{if } z_2 = 0 \\ \dfrac{2|z_2|^4}{\sqrt{(|z_1|^2 + |z_2|^2)^2 + 4|z_2|^4} - (|z_1|^2 + |z_2|^2)}, & \text{if } z_2 \neq 0, \end{cases}$$

where $d(z_1, z_2)$ = radius of the circle obtained by intersecting the complex line through the origin and $(z_1, z_2) \in \mathbb{C}^{n+1} - \{0\}$ with the domain D. Using (ii) or (iii) of Corollary 3.2 we see immediately that D is not biholomorphic to the ball.

Let L be the tautological line bundle over $\mathbb{C}P^n$. The bundle space of L-{zero section} is naturally identified with $\mathbb{C}^{n+1}-\{0\}$. A Finsler metric on $\mathbb{C}^{n+1}-\{0\}$ is then identified with a hermitian metric along the fiber of L. Furthermore the Chern form of a hermitian metric r is given by $c_1(L,r)=-\frac{i}{2\pi}\omega$ where $\pi^*\omega=\partial\bar{\partial}\log r$, $\pi:\mathbb{C}^{n+1}-\{0\}\to\mathbb{C}P^n$ the usual projection. In particular those Finsler metric which are strictly pseudo-convex are just those hermitian metric on L with negative definite Chern forms (cf. Kobayashi [11]).

On $\mathbb{C}P^n$ we have the following exact sequence:

$$\begin{array}{ccccccccc} \longrightarrow & H^1(\mathbb{C}P^n,\mathcal{O}) & \longrightarrow & H^1(\mathbb{C}P^n,\mathcal{O}^*) & \xrightarrow[\cong]{[c]} & H^2(\mathbb{C}P^n,\mathbb{Z}) & \longrightarrow & H^2(\mathbb{C}P^n,\mathcal{O}) & \longrightarrow \\ & \| & & & & & & \| & \\ & 0 & & & & & & 0 & \end{array}$$

where $[c]$ is the Chern map. The generator of $H^1(\mathbb{C}P^n,\mathcal{O}^*)$ is $[L^*]$ where L^* is the dual of the tautological bundle and $[\Omega]$ is the generator of $H^2(\mathbb{C}P^n,\mathbb{Z})$ where Ω is the fundamental Kähler form on $\mathbb{C}P^n$. It follows that $-c_1(L,r)$ is cohomologous to Ω.

Conversely, given any C^∞ real $(1,1)$ form ω on $\mathbb{C}P^n$ cohomologous to Ω, we have

$$\omega=\Omega+d\chi,$$

where χ is a real 1-form. Let $\chi=\eta+\bar{\eta}$ where η is of type $(1,0)$. Then since both $\omega=\Omega$ are of type $(1,1)$ we have

$$\omega=\Omega+\bar{\partial}\eta+\partial\bar{\eta}$$
$$\partial\eta=0 \quad \text{and} \quad \bar{\partial}\bar{\eta}=0.$$

By Dolbeault's lemma, $0=H^1(\mathbb{C}P^n,\mathcal{O})\simeq\bar{\partial}$-closed $(0,1)$ forms/$\bar{\partial}$-exact $(0,1)$ forms, we conclude that

$$\bar{\eta}=\bar{\partial}\varrho \quad \text{and} \quad \eta=\partial\bar{\varrho}$$

for some (complex-valued) C^∞ function ϱ on $\mathbb{C}P^n$. Thus we get

$$\omega=\Omega+\bar{\partial}\partial\bar{\varrho}+\partial\bar{\partial}\varrho=\Omega+\frac{i}{2\pi}\partial\bar{\partial}\tilde{g},$$

where $\tilde{g}=4\pi\,\mathrm{Im}\,\varrho$ is real valued on $\mathbb{C}P^n$.

Pulling back the previous identity to $\mathbb{C}^{n+1}-\{0\}$ we get

$$\pi^*\omega=\frac{i}{2\pi}\partial\bar{\partial}\log|z|^2+\frac{i}{2\pi}\partial\bar{\partial}g=\frac{i}{2\pi}\partial\bar{\partial}\log(|z|^2e^g),$$

where g is the pull back of $\tilde{g}$ to $\mathbb{C}^{n+1}-\{0\}$ and we have used the fact that the pull back of the Fubini-study metric is $\frac{i}{2\pi}\partial\bar{\partial}\log|z|^2$. Let $r(z)=|z|^2e^{g(z)}$ then it is C^∞ and non-vanishing on $\mathbb{C}^{n+1}-\{0\}$ and extends to 0 at the origin because g is bounded.

Since e^g is constant on each line through the origin the homogeneous condition is clear, i.e. r is a Finsler metric on $\mathbb{C}^{n+1}$.

Suppose that we have two Finsler metric r_1 and r_2 such that $\partial\bar{\partial}\log r_1 = \partial\bar{\partial}\log r_2$ on $\mathbb{C}^{n+1} - \{0\}$. Then $\log\frac{r_1}{r_2}$ is pluri-harmonic on $\mathbb{C}^{n+1} - \{0\}$ and since it is invariant by $\lambda \in \mathbb{C}^*$, it is actually pluri-harmonic on $\mathbb{C}P^n$, hence a constant. In other words we have $r_1 = cr_2$ for some positive constant $c > 0$.

We have thus shown that to a Finsler metric r, the associated $(1,1)$ form $\frac{i}{2\pi}\partial\bar{\partial}\log r$ projects down to a C^∞ real $(1,1)$ form ω on $\mathbb{C}P^n$ which is cohomologous to the fundamental Kähler form Ω. The form ω uniquely determined r up to a positive multiplicative constant. On the other hand by Proposition 3.1 two complete proper circular domains are biholomorphic iff their corresponding Finsler metric differ by a linear map $A \in \mathrm{GL}(n+1, \mathbb{C})$. Combining these two facts we have

Theorem 3.3. *Let $\mathscr{D}$ be the set of biholomorphic classes of bounded complete proper circular domains with $\mathscr{C}^\infty$ boundaries in $\mathbb{C}^{n+1}$ and $\mathscr{D}^+ \subset \mathscr{D}$ be those that are strictly pseudo-convex. Denote by $\{\Omega\}$ the set of C^∞ real forms of type $(1,1)$ on $\mathbb{C}P^n$ which are cohomologous to the fundamental Kähler form and $\{\Omega\}^+$ those that are also positive definite, then $\mathscr{D} \simeq \{\Omega\}/\mathrm{Aut}\,\mathbb{C}P^n$ and $\mathscr{D}^+ \simeq \{\Omega\}^+/\mathrm{Aut}\,\mathbb{C}P^n$.*

References

1. Bedford, E., Kalka, M.: Foliations and complex Monge-Ampère equations. Comm. Pure Appl. Math. **30**, 543–571 (1977)
2. Bedford, E., Taylor, B.A.: Variational properties of the complex Monge-Ampère equation. II. intrinsic norms. Am. J. Math. **101**, 1131–1166 (1977)
3. Braun, R., Kaup, W., Upmeier, H.: On the automorphisms of circular and Reinhardt domains in complex Banach spaces. Manuscripta Math. **25**, 97–133 (1978)
4. Burns, D.: Curvatures of Monge-Ampère foliations and parabolic manifolds (preprint)
5. Burns, D., Shnider, S.: Real hypersurfaces in complex manifolds. Proc. Symp. Pure Math. **30**, 141–168 (1977)
6. Chern, S.S., Moser, J.: Real hypersurfaces in complex manifolds. Acta Math. **133**, 219–271 (1975)
7. Edwards, R., Millett, K., Sullivan, D.: Foliations with all leaves compact. Topology **16**, 13–32 (1977)
8. Goldberg, S.I., Kobayashi, S.: On holomorphic bisectional curvature. J. Differential Geometry **1**, 225–233 (1967)
9. Greene, R., Krantz, S.: Deformation of complex structures, estimates for the $\bar{\partial}$ equation, and stability of the Bergman kernel (preprint)
10. Greene, R., Wu, H.: Function theory on manifolds which possess a pole. In: Lecture Notes in Mathematics, Vol. 699. Berlin, Heidelberg, New York: Springer 1979
11. Kobayashi, S.: Negative vector bundles and complex Finsler structure. Nagoya Math. J. **57**, 153–166 (1975)
12. Lempert, L.: Le metrique de Kobayeshi et la representation des domains sur la boule (preprint)
13. Milnor, J.: Lectures on Morse theory. Ann. Math. Studies, No. 51. Princeton. Princeton University Press 1963
14. Siu, Y.T.: Curvature characterization of hyperquadrics. Duke Math. J. **47**, 641–654 (1980)
15. Siu, Y.T., Yau, S.T.: Compact Kähler manifolds of positive bisectional curvature. Invent. Math. **59**, 189–204 (1980)

16. Stoll, W.: The characterization of strictly parabolic manifolds. Ann. Scuola Norm. Sup. Pisa, Ser. IV. **7**, 87–154 (1980)
17. Stoll, W.: The characterization of strictly parabolic spaces. Compositio Math. (1981)
18. Sunada, T.: Holomorphic equivalence problem for bounded Reinhart domains. Math. Ann. **235**, 111–128 (1978)
19. Webster, S.: On the mapping problem for algebraic real hypersurfaces. Invent. Math. **43**, 53–68 (1977)
20. Webster, S.: Pseudo-hermitian structures on a real hypersurface. J. Differential Geometry **13**, 235–250 (1978)
21. Webster, S.: On the pseudo-conformal geometry of a Kähler manifold. Math. Z. **157**, 265–270 (1977)
22. Wong, P.M.: Geometry of the homogeneous complex Monge-Ampère equation. Invent. Math. **67**, 261–274 (1982)

Received April 20, 1982

Math. Ann. 263, 31–42 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Interaction of Tubes and Spheres

L. Vanhecke[1] and T. J. Willmore[2]

1 Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200 B, B-3030 Leuven, Belgium
2 Department of Mathematics, University of Durham, South Road, Durham DH1 3LE, England

1. Introduction

Let (M, g) be a connected n-dimensional C^∞-Riemannian manifold and let σ be a curve in M. We show that the volume $V_\sigma(r)$ of a tube of radius r about σ is expressible as the sum of a term which involves only first derivatives of σ and another term which involves no derivatives of σ higher than second order. Indeed, the second order terms appear explicitly as the curvature of σ.

We define a *D'Atri space* and show that for such spaces $V_\sigma(r)$ does not involve the curvature of σ explicitly so that it depends only on first derivatives of σ. We define the concept of geodesic symmetry with respect to geodesics and obtain a number of results which give new characterizations of locally symmetric spaces and of spaces of constant curvature.

2. The Volume Density Function of $\exp_m$

Let (M, g) be a connected, n-dimensional C^∞-Riemannian manifold, and let m be a point of M. Let r_m denote the injectivity radius of $\exp_m : T_mM \to M$. Then $\exp_m$ induces a C^∞-diffeomorphism of the open ball $B_m := B_m(r_m)$ of the tangent space T_mM of radius r_m about the origin o_m onto the open geodesic ball $\mathcal{U}_m$ in M of the same radius with centre m (see, e.g., [10]). Since $\mathcal{U}_m$ is an n-cell, we can choose an orientation and the corresponding volume form ω of (M, g) on $\mathcal{U}_m$. Let ω_m denote the volume form of the coherently oriented euclidean vector space (T_mM, g_m) considered as a Riemannian manifold. Then we have

$$(\exp_m)^*(\omega) = \mathfrak{d}_m \omega_m \tag{2.1}$$

for some C^∞-function $\mathfrak{d}_m$ on $B_m \subset T_mM$. Note that if we reverse the two coherently chosen orientations in $\mathcal{U}_m$ and T_mM, we change the sign of ω and ω_m simultaneously and it follows from (2.1) that $\mathfrak{d}_m$ is independent of the choice of orientation.

It is convenient to call this (strictly positive) function $\mathfrak{d}_m$ the *volume density function of* $\exp_m$ because for every $v \in B_m$ we have, from (2.1),

$$\mathfrak{d}_m(v) = \lim_{\operatorname{diam}(K)\to 0} \frac{\operatorname{vol}_g(\exp_m(K))}{\operatorname{vol}_{g_m}(K)}, \tag{2.2}$$

where K denotes any compact ball of T_mM with centre v, contained in B_m. Formula (2.2) is reminiscent of the classical formula of Gauss for the curvature density function of surfaces.

The function $\mathfrak{d}_m$ gives rise to a uniquely determined C^∞-function $\theta_m : \mathscr{U}_m \to \mathbb{R}_+$ defined on the open geodesic ball $\mathscr{U}_m$ in M of radius r_m and centre m such that

$$\theta_m \circ \exp_m = \mathfrak{d}_m. \tag{2.3}$$

Let $r \in [0, r_m[$, and let $B_m(r)$ denote the ball in T_mM of radius r and centre o_m. Let $S_m^{n-1}(r)$ denote the sphere in T_mM of radius r and centre o_m. We now obtain a formula for the n-dimensional volume of the open geodesic ball $\exp_m(B_m(r))$ in M of radius r and centre m. Consider the function $f : [0, r] \times S_m^{n-1}(1) \to B_m(r)$ given by

$$(s, \xi) \mapsto f(s, \xi) = s\xi. \tag{2.4}$$

This is a C^∞-map with the property

$$f^* \omega_m = s^{n-1}(ds \wedge d\xi),$$

where $d\xi$ denotes the volume form of the submanifold $S_m^{n-1}(1)$ of (T_mM, g_m), and where for simplicity we have omitted the canonical pull-backs from the notation.

From (2.1) and Fubini's theorem we find for the volume of the open geodesic ball $\exp_m(B_m(r))$ in M of radius r and centre m:

$$V_m(r) = \int_{B_m(r)} \mathfrak{d}_m(v) d \operatorname{vol}_{g_m}(v) = \int_0^r s^{n-1} \int_{S_m^{n-1}(1)} \mathfrak{d}_m(s\xi) d\xi ds. \tag{2.5}$$

We wish to obtain a more explicit description of the values of $\mathfrak{d}_m$ and recall some well-known facts about the differential of $\exp_m$. The domain of definition of $\exp_m$ is an open subset E_m of T_mM, star-shaped with respect to o_m. Fix some vector $v \in E_m$ and express this as $v = s\xi$ with $\xi \in S_m^{n-1}(1)$, and $s \in [0, \infty[$. Then for every tangent vector $a \in T_v(T_mM)$ there exists a unique vector $\mathbf{a} \in T_mM$ such that

$$a = \text{ velocity vector of the curve } t \mapsto v + t\mathbf{a} \text{ at } t = 0. \tag{2.6}$$

The map $a \mapsto \mathbf{a}$ describes the standard identification of $T_v(T_mM)$ with the vector space T_mM. Then we have

$$(\exp_m)_{*|s\xi}(a) = \begin{cases} \mathbf{a}, & \text{if } s = 0, \quad (2.7) \\ \dfrac{1}{s} Y(s), & \text{if } s > 0, \quad (2.8) \end{cases}$$

where Y is a uniquely determined Jacobi field along the unit speed geodesic curve

$$\gamma : [0, s] \to M, \ \lambda \mapsto \gamma(\lambda) = \exp_m(\lambda\xi) \tag{2.9}$$

satisfying the initial conditions

$$Y(0) = o_m, \ Y'(0) = \mathbf{a}. \tag{2.10}$$

Alternatively, Y is the unique solution of the Jacobi equation $Y''=-R(\gamma', Y)\gamma'$ satisfying (2.10).

From (2.1), (2.7) we obtain immediately

$$\mathfrak{d}_m(o_m)=1, \tag{2.11}$$

and for all $(s,\xi)\in\mathbb{R}_+\times S_m^{n-1}(1)$ with $s\xi\in E_m$ we deduce:

Proposition 2.1. *If $(e_1,\ldots,e_n)$ is any positively oriented orthonormal frame of (T_mM,g) we find from* (2.1), (2.7), (2.10):

$$\mathfrak{d}_m(s\xi)=s^{-n}\omega_{\exp_m(s\xi)}(Y_1(s),\ldots,Y_n(s)), \tag{2.12}$$

where Y_i is the Jacobi vector field along γ (see (2.9)) *satisfying*

$$Y_i(0)=o_m,\ Y_i'(0)=e_i \quad \textit{for } i=1,\ldots,n. \tag{2.13}$$

We can obtain a further simplification by specializing the choice of frame $(e_1,\ldots,e_n)$ such that $e_n=\gamma'(0)=\xi$. Then evidently

$$Y_n(\lambda)=\lambda\cdot\gamma'(\lambda), \quad \text{for all } \lambda\in[0,s], \tag{2.14}$$

and hence we get from (2.12) and (2.14)

$$\mathfrak{d}_m(s\xi)=s^{1-n}\omega_{\exp_m(s\xi)}(Y_1(s),\ldots,Y_{n-1}(s),\gamma'(s)). \tag{2.15}$$

Let E_i denote the vector field parallel along γ with $E_i(0)=e_i$ for $i=1,\ldots,n$. Since $E_n=\gamma'$ [from (2.9) and (2.14)], it follows that $E_1,\ldots,E_{n-1}$ span the normal bundle along γ. But, from (2.13), $Y_1,\ldots,Y_{n-1}$ also remain orthogonal to γ' along γ, and therefore there exists a unique C^∞-endomorphism A of the normal bundle of γ with

$$Y_i(\lambda)=A(E_i)(\lambda) \quad \text{for } i=1,\ldots,n-1 \quad \text{and all} \quad \lambda\in[0,s]. \tag{2.16}$$

Consequently we get from (2.15), (2.16)

$$\mathfrak{d}_m(s\xi)=s^{1-n}(\det A)(s). \tag{2.17}$$

Moreover, since E_i is parallel and Y_i is a Jacobi field along γ satisfying (2.13) for $i=1,\ldots,n$, A must satisfy

$$A''=-(R(\gamma',\cdot)\gamma')\circ A \quad \text{with } A(0)=0,\quad A'(0)=\mathrm{Id}_{\xi^\perp}, \tag{2.18}$$

where $A':=\nabla_{d/d\lambda}A$ and $A''=\nabla_{d/d\lambda}A'$. Conversely, if A is an endomorphism of the normal bundle of γ satisfying (2.18), then it is easily verified that (2.17) is true.

We note that

$$S(s)=A'A^{-1}(s) \tag{2.19}$$

is the shape operator of the geodesic sphere $\exp_m(S_m^{n-1}(s))$ at the point $\gamma(s)=\exp_m(s\xi)$.

3. The Volume Density Function of $\exp\sigma$

Let $\sigma\colon[a,b]\to(M,g)$ be a C^∞-imbedded curve. For every $t\in[a,b]$ and $r\in\mathbb{R}_+$ let

$$B^\perp_{\sigma(t)}(r)=\{v\in T_{\sigma(t)}M \mid g(v,\dot\sigma(t))=0 \quad \text{and} \quad \|v\|<r\} \tag{3.1}$$

denote the $(n-1)$-dimensional ball of radius r in the normal space $\dot{\sigma}(t)^{\perp}(\subset T_{\sigma(t)}M)$ of σ at t about its origin.

Then we call

$$N_{\sigma}(r)=\bigcup_{t\in[a,b]} B^{\perp}_{\sigma(t)}(r) \quad (\subset \text{normal bundle } \dot{\sigma}^{\perp} \text{ of } \sigma) \tag{3.2}$$

the open solid tube of radius r about the zero section of the normal bundle of σ. The normal bundle and hence $N_{\sigma}(r)$ has a natural flat Riemannian product metric g_{σ}, namely the metric which turns the diffeomorphism

$$\left.\begin{aligned} \tau\colon [a,b]\times\dot{\sigma}(a)^{\perp}&\to\dot{\sigma}^{\perp}=\bigcup_{\lambda\in[a,b]}\dot{\sigma}(\lambda)^{\perp}\\ (t,\eta)&\mapsto\tau_t(\eta)\end{aligned}\right\} \tag{3.3}$$

into an isometry. Here $\tau_t\colon \dot{\sigma}(a)^{\perp}\to\dot{\sigma}(t)^{\perp}$ denotes the parallel translation in the normal bundle along $\sigma|[a,b]$ with respect to the induced covariant derivative $^{\perp}\nabla$, and $[a,b]\times\dot{\sigma}(a)^{\perp}$ is considered as the submanifold of the flat Riemannian product manifold $\mathbb{R}\times(T_{\sigma(a)}M, g_{\sigma(a)})$. Since parallel translation with respect to the normal connection preserves the length of the normal vectors, τ induces a C^{∞}-diffeomorphism

$$\begin{aligned}[a,b]\times B^{\perp}_{\sigma(a)}(r)&\to N_{\sigma}(r)\\ (t,\eta)&\to\tau_t(\eta).\end{aligned} \tag{3.4}$$

We now choose $r\in\mathbb{R}_+$ so small that $\exp_{\sigma}:=\exp|N_{\sigma}(r)$ is a C^{∞}-diffeomorphism of $N_{\sigma}(r)$ onto the open solid tube $\mathcal{U}_{\sigma}(r)$ of radius r about σ. That this is always possible is guaranteed since $[a,b]$ is compact and $\sigma\colon[a,b]\to M$ is an imbedding. Since $\mathcal{U}_{\sigma}(r)$ is an n-cell, we can choose an orientation and the corresponding volume form ω of (M,g) on $\mathcal{U}_{\sigma}(r)$. Let ω_{σ} denote the volume form of the coherently oriented total space of the normal bundle $\dot{\sigma}^{\perp}$ of σ with its flat metric described in (3.3). Then we have

$$\exp_{\sigma}^{*}\omega=\mathfrak{d}_{\sigma}\omega_{\sigma} \tag{3.5}$$

for some C^{∞}-function $\mathfrak{d}_{\sigma}$ defined on $N_{\sigma}(r)(\subset\dot{\sigma}^{\perp})$.

We call this (strictly positive) function $\mathfrak{d}_{\sigma}$ the *volume density function of* $\exp_{\sigma}$, analogous to our treatment in Sect. 2. Because of our choice of r, there is a unique C^{∞}-function $\theta_{\sigma}\colon\mathcal{U}_{\sigma}(r)\to\mathbb{R}_+$ defined on the open solid tube $\mathcal{U}_{\sigma}(r)$ of radius r about σ, such that

$$\theta_{\sigma}\circ\exp_{\sigma}=\mathfrak{d}_{\sigma}. \tag{3.6}$$

Then the volume $V_{\sigma}(r)$ of the tube is given by

$$\begin{aligned} V_{\sigma}(r)&=\int_{N_{\sigma}(r)}\mathfrak{d}_{\sigma}(\eta)d(\mathrm{vol}_{g_{\sigma}}(\eta))\\ &=\int_a^b\Big(\int_{B_{\sigma(t)}}\mathfrak{d}_{\sigma}(\eta)d\,\mathrm{vol}_{g_{\sigma(t)}}(\eta)\Big)dt.\end{aligned} \tag{3.7}$$

Moreover, if $\eta\in B^{\perp}_{\sigma(t)}(r)$, choose $\xi\in S^{n-2}_{\sigma(a)}(1)$ [the unit sphere in $\dot{\sigma}(a)^{\perp}$] and $s=\|\eta\|\in[0,r[$ such that

$$\eta=s\tau_t(\xi), \tag{3.8}$$

where $\tau_t(\xi)$ denotes the parallel transport of ξ with respect to the normal connection along σ from $\sigma(a)$ to $\sigma(t)$. Denote by f the C^∞-map given by

$$\left.\begin{aligned} f\colon [a,b]\times S^{n-2}_{\sigma(a)}(1)\times[0,r[&\to \mathscr{U}_\sigma(r) \\ (t,\xi,s) &\to \exp_{\sigma(t)}(s\tau_t(\xi)), \end{aligned}\right\} \tag{3.9}$$

with

$$f^*\omega_\sigma = s^{n-2}\,ds\wedge d\xi\wedge dt. \tag{3.10}$$

Using Fubini's theorem as in Sect. 2 gives, with $\eta = s\tau_t(\xi)$,

$$V_\sigma(r) = \int_a^b \int_0^r s^{n-2} \int_{S^{n-2}_{\sigma(a)}(1)} \mathfrak{d}_\sigma(s\tau_t(\xi))\,d\xi\,ds\,dt. \tag{3.11}$$

We now obtain a formula better suited for the computation of $\mathfrak{d}_\sigma(\eta)$ and hence for the computation of $V_\sigma(r)$. For fixed $\eta\in N_\sigma(r)$ choose (t,ξ,s) as in (3.8) and choose an orthonormal frame $(e_1,\dots,e_n)$ of $T_{\sigma(a)}M$ such that

$$e_1 = \dot\sigma(a),\ e_n = \xi. \tag{3.12}$$

Choose E_1 as the unit tangent field $\dot\sigma$ and $E_2,\dots,E_n$ normal vector fields along σ, parallel with respect to the normal connection $^\perp\nabla$ of the normal bundle $\dot\sigma^\perp$ such that $E_i(a)=e_i$ for $i=2,\dots,n$. Then $(E_2,\dots,E_n)$ form a $^\perp\nabla$-parallel orthonormal frame field of the normal bundle $\dot\sigma^\perp$ of σ, and evidently [see (3.8) and (3.12)]

$$\eta = sE_n(t). \tag{3.13}$$

As in Sect. 2 we compute the differential of $\exp_\sigma$ applied to tangent vectors at η. According to the definition of the Riemannian manifold $(N_\sigma(r), g_\sigma)$ in (3.3) the tangent vectors to the n curves

$$h\mapsto sE_n(t+h), \tag{3.14}$$

$$h\mapsto \eta + hE_i(t) \quad \text{for} \quad i\in\{2,\dots,n\} \tag{3.15}$$

at $h=0$ form an orthonormal frame $(a_1,\dots,a_n)$ of tangent vectors to $(N_{\sigma(r)}, g_\sigma)$ at the point η and therefore due to (3.5)

$$\mathfrak{d}_\sigma(\eta) = \omega(\exp_*|_\eta(a_1),\dots,\exp_*|_\eta(a_n)). \tag{3.16}$$

In case $\eta = o_{\sigma(t)}$ we get from (2.7), (3.14), (3.15), (3.16) immediately

$$\mathfrak{d}_\sigma(o_{\sigma(t)}) = \omega(\dot\sigma(t), E_2(t),\dots,E_n(t)) = 1. \tag{3.17}$$

Suppose now $s>0$. In order to compute $\exp_*|_\eta(a_i)$ we proceed as in Sect. 2 [see (2.8), (2.9)]. Let γ denote the unit speed geodesic curve

$$\gamma(\lambda) := \exp_{\sigma(t)}(\lambda E_n(t)) \quad \text{for} \quad \lambda\in[0,s]. \tag{3.18}$$

Then, according to (2.8) we get from (3.15)

$$\exp_*|_\eta(a_i) = \frac{1}{s}\,Y_i(s) \quad \text{for} \quad i\in\{2,\dots,n\}, \tag{3.19}$$

where Y_i is the unique Jacobi field along γ satisfying

$$Y_i(0) = o_{\sigma(t)} \quad \text{and} \quad Y_i'(0) = E_i(t) \quad \text{for} \quad i\in\{2,\dots,n\}. \tag{3.20}$$

In particular, because of (3.13), (3.15), (3.18) we have again (2.14) and therefore

$$\exp_*|_\eta(a_n)=\gamma'(s). \tag{3.21}$$

Finally, the tangent vector of the image of the curve (3.14) under exp at $h=0$ equals

$$\exp_*|_\eta(a_1)=Y_1(s), \tag{3.22}$$

where Y_1 is the unique Jacobi field along γ satisfying the initial conditions

$$Y_1(0)=\dot\sigma(t)=E_1(t) \quad \text{and} \quad Y_1'(0)=(\nabla_{d/dt}E_n)(t) \tag{3.23}$$

as follows immediately from the well-known variational description of Jacobi fields (see, for example, [5]). However, E_n was chosen to be parallel along σ with reference to the normal connection of σ, and therefore

$$\nabla_{d/dt}E_n-\langle\nabla_{d/dt}E_n,\dot\sigma\rangle\dot\sigma=0.$$

Consequently, since $\langle E_n,\dot\sigma\rangle=0$, we find

$$(\nabla_{d/dt}E_n)(t)=-\langle E_n(t),\ddot\sigma(t)\rangle\dot\sigma(t), \tag{3.24}$$

where

$$\ddot\sigma(t)=(\nabla_{d/dt}\dot\sigma)(t) \tag{3.25}$$

is the (mean) curvature normal of $\dot\sigma$ at t.

We denote by $E_a(\lambda)$, $a\in\{1,2,\ldots,n-1\}$ the orthonormal basis for vectors normal to γ obtained by parallel transport of $E_a(t)$ with respect to the Riemannian connection along γ. We write

$$Y_a(\lambda)=(BE_a)(\lambda). \tag{3.26}$$

This defines an endomorphism-valued function $\lambda\mapsto B(\lambda)$, which satisfies the Jacobi equation

$$B''+R\circ B=0. \tag{3.27}$$

The initial conditions satisfied by B corresponding to the initial conditions (3.20), (3.23) are

$$B(0)=\begin{pmatrix}1 & 0\ldots 0\\ 0 & \\ \vdots & 0\\ 0 & \end{pmatrix},\quad B'(0)=\begin{pmatrix}-\langle E_n(t),\ddot\sigma(t)\rangle & 0\ldots 0\\ 0 & \\ \vdots & I\\ 0 & \end{pmatrix}, \tag{3.28}$$

where 0 and I denote respectively the zero and unit $(n-2)\times(n-2)$ matrix.

Let $\tilde A(\lambda)$ denote the matrix solution of (3.27) which satisfies the boundary conditions

$$\tilde A(s)=0,\ \tilde A'(s)=-\mathrm{I}. \tag{3.29}$$

By a well-known property of Jacobi vector fields (the so-called Sturm trick: see for example, [1, p. 157] it follows that

$$B'(0)^*\tilde A(0)-B(0)^*\tilde A'(0)=B'(s)^*\tilde A(s)-B(s)^*\tilde A'(s), \tag{3.30}$$

where * denotes transpose of the matrix giving

$$B(s)^*=B'(0)^*\tilde A(0)-B(0)^*\tilde A'(0). \tag{3.31}$$

Using the initial conditions $B(0)$ and $B'(0)$ of (3.28) we get

$$B(s)=\begin{bmatrix} B_{11}(s) & \tilde{A}_{21}(0) & \dots \tilde{A}_{n-11}(0) \\ B_{21}(s) & \tilde{A}_{22}(0) & \dots \tilde{A}_{n-12}(0) \\ \vdots & \vdots & \vdots\vdots \\ B_{n-11}(s) & \tilde{A}_{2n-1}(0) & \dots \tilde{A}_{n-1n-1}(0) \end{bmatrix}, \tag{3.32}$$

where

$$\left.\begin{aligned} B_{11}(s)&=-\kappa_\eta \tilde{A}_{11}(0)-\tilde{A}'_{11}(0) \\ B_{\alpha 1}(s)&=-\kappa_\eta \tilde{A}_{1\alpha}(0)-\tilde{A}'_{1\alpha}(0) \end{aligned}\right\} \tag{3.33}$$

for $\alpha=2,\dots,n$, and we have written

$$\kappa_\eta=\langle \ddot{\sigma}, E_n\rangle . \tag{3.34}$$

Hence we obtain

$$det\, B(s)=-\det\begin{bmatrix} \tilde{A}'_{11}(0) & \tilde{A}_{21}(0) & \dots \tilde{A}_{n-11}(0) \\ \vdots & \vdots & \vdots\vdots \\ \tilde{A}'_{1n-1}(0) & \tilde{A}_{2n-1}(0) & \dots \tilde{A}_{n-1n-1}(0) \end{bmatrix} - \kappa_\eta \det A(0).$$

Using (2.19) we get

$$\det B(s)=-S_{11}(0)\det \tilde{A}(0)-\kappa_\eta \det \tilde{A}(0). \tag{3.35}$$

Finally we use (2.17), the corresponding relation

$$\mathfrak{d}_\sigma(s\xi)=s^{2-n}\det B(s) \tag{3.36}$$

and $\det \tilde{A}(0)=\det A(s)$ [1] to obtain

Theorem 3.1. *Let (M,g) be a Riemannian manifold and σ a curve in M. Then we have*

$$\mathfrak{d}_\sigma(\eta)=-\{\langle S_p\dot{\sigma},\dot{\sigma}\rangle+\kappa_\eta\} s\mathfrak{d}_m(\eta), \tag{3.37}$$

where S_p denotes the second fundamental form of the geodesic sphere with centre $p=\exp_m\eta$ and radius s.

We note that, in general, $\mathfrak{d}_\sigma(\eta)$ involves at most derivatives of σ of the second order. In the next section we consider spaces for which $V_\sigma(r)$ involves at most derivatives of σ of the first order.

4. D'Atri Spaces

We shall define a *D'Atri space* as a Riemannian manifold (M,g) with the property that, for every point m of M, the local geodesic symmetry $(\exp_m(\eta)\to\exp_m(-\eta))$ of (M,g) is volume-preserving. Our terminology is justified by papers by D'Atri and Nickerson [2, 3] and D'Atri [4] who previously studied Riemannian manifolds which are characterized by this property. In terms of our notation, a D'Atri space is characterized by the formula

$$\mathfrak{d}_m(\eta)=\mathfrak{d}_m(-\eta) \tag{4.1}$$

which holds for every $m \in M$ and all $\eta \in B_m(r_m)$. Equivalently we may say that $\mathfrak{d}_m$ is an even function on $B_m(r_m)$. In a D'Atri space, it follows from (3.37) that the function $\kappa_\eta s \mathfrak{d}_m(\eta)$ is odd, since κ_η changes sign with η. It follows that the contribution of this part of $\mathfrak{d}_\sigma$ to $V_\sigma(r)$ in (3.11) is zero, since for each t the integration over $S^{n-2}_{\sigma(a)}(1)$ gives zero. We have already seen that the remaining part of $\mathfrak{d}_\sigma$ depends only on the direction of σ at each point and does not involve the curvature. This completes the proof of the statement made in the introduction.

For a D'Atri space it follows from (3.11) that

$$V_\sigma(r) = -\int_a^b \int_0^r s^{n-2} \int_{S^{n-2}_{\sigma(a)}(1)} \langle S_p \dot{\sigma}(t), \dot{\sigma}(t) \rangle \mathfrak{d}_{\sigma(t)} d\xi ds dt \,. \tag{4.2}$$

We note that (4.2) gives $V_\sigma(r)$ in an arbitrary Riemannian manifold provided that the curve σ is a geodesic. This follows at once since for a geodesic $\kappa = 0$. We *believe* that if (4.2) holds for every curve σ (and not just for geodesics), then the space must be a D'Atri space, but so far we have no proof of this conjecture.

5. Geodesic Symmetry with Respect to Geodesic Curves

Let (M, g) be a Riemannian manifold and let $\sigma: = t \to \sigma(t)$ be a C^∞-imbedded curve. We consider the function $f: [0, r] \times S^{n-2}_{\sigma(a)}(1) \to B^\perp_{\sigma(a)}(r)$ given by $(s, \xi) \mapsto f(s, \xi): = s\xi$, as given in (2.4). We denote by $\tau_t: \dot{\sigma}(a)^\perp \to \dot{\sigma}(t)^\perp$ the map obtained by parallel transport along σ with respect to the normal connection. Consider now the map $\phi_\sigma: N_\sigma(r) \mapsto N_\sigma(r)$

$$\tau_t(s\xi) \mapsto \tau_t(-s\xi), \quad s \in [0, r], \quad t \in [a, b]. \tag{5.1}$$

Definition 5.1. The map

$$\exp_{\sigma(t)}(\tau_t s\xi) \mapsto \exp_{\sigma(t)}(-\tau_t s\xi), \quad s \in [0, r], \quad t \in [a, b] \tag{5.2}$$

is a geodesic symmetry with respect to the curve σ.

We shall be particularly interested in the special case when σ is a geodesic, and when the property holds for *all* geodesics.

Definition 5.2. A Riemannian manifold is said to have the D'Atri property with respect to geodesics if and only if the geodesic symmetry with respect to all geodesics is volume-preserving.

We now prove the following lemma.

Lemma 5.1. *In a Riemannian manifold the following conditions are equivalent*:

(1) $\nabla R = 0$,
(2) $\nabla_X R_{XYXY} = 0$,
(3) $\nabla_V R_{WXYZ} + \nabla_X R_{WZYV} + \nabla_Z R_{WVYZ} = 0$.

Clearly (1)⇒(2): To prove (2)⇒(3), replace X in (2) by $aU + bV + cW$ and expand. This gives a polynomial in a, b, c, which vanishes identically. In particular the coefficient of abc must vanish giving

$$\mathfrak{S}_{U,V,W} \nabla_U R_{VYWY} = 0,$$

where $\mathfrak{S}$ denotes the symmetric sum.

Polarize this equation with respect to Y to get

$$\begin{aligned}
0 &= \mathop{\mathfrak{S}}_{UVW} \{\nabla_U R_{VYWZ} + \nabla_U R_{VZWY}\} \\
&= \mathop{\mathfrak{S}}_{UVW} \{\nabla_U R_{VYWZ} - \nabla_U R_{VZYW}\} \\
&= \mathop{\mathfrak{S}}_{UVW} \{2\nabla_U R_{VYWZ} + \nabla_U R_{VWZY}\} \quad \text{(by the first Bianchi identity)} \\
&= 2 \mathop{\mathfrak{S}}_{UVW} \nabla_U R_{VYWZ} \quad \text{(by the second Bianchi identity).}
\end{aligned}$$

Thus we get (3).

We now prove that (3)⇒(1). The proof below, due to Gray [8], is an improvement on our previous proof which was restricted to definite metrics.

We write out (3) in the form

$$\nabla_V R_{WXYZ} + \nabla_Z R_{WVYX} + \nabla_X R_{WZYV} = 0.$$

Interchange X and Y and subtract the resulting equation to get

$$0 = \nabla_V (R_{WXYZ} - R_{WYXZ}) + (\nabla_Z R_{WVYX} - \nabla_Z R_{WVXY}) + (\nabla_X R_{WZYV} - \nabla_Y R_{WZXV}).$$

We use the first Bianchi identity on the first term and the second Bianchi identity on the third term, to get

$$0 = -\nabla_V R_{WZXY} - 2\nabla_Z R_{WVXY} - \nabla_V R_{WZXY},$$

i.e.

$$\nabla_V R_{WZXY} + \nabla_Z R_{WVXY} = 0.$$

Now interchange V and W in the above equation and subtract. We get

$$\nabla_V R_{WZXY} - \nabla_W R_{VZXY} + 2\nabla_Z R_{WVXY} = 0,$$

which we may rewrite as

$$\begin{aligned}
3\nabla_Z R_{WVXY} &= \nabla_Z R_{WVXY} + \nabla_W R_{VZXY} + \nabla_V R_{ZWXY} \\
&= 0
\end{aligned}$$

by the second Bianchi identity and the lemma is proved.

Lemma 5.2 [6, 7, 9]. *With the same notations as in Sect. 3 we have*

(i) $$\mathfrak{d}_\sigma(\eta) = 1 - \frac{s^2}{6}(\varrho_{\xi\xi} + 2R_{1\xi 1\xi})(m) - \frac{s^3}{12}(\nabla_\xi \varrho_{\xi\xi} + \nabla_\xi R_{1\xi 1\xi})(m) + O(s^4); \tag{5.3}$$

(ii) $$\begin{cases}
S_{\sigma 11}(\eta) = sR_{1\xi 1\xi}(m) + \dfrac{s^2}{2}\nabla_\xi R_{1\xi 1\xi}(m) + O(s^3), \\
S_{\sigma 1\alpha}(\eta) = -\dfrac{s}{2}R_{1\xi\alpha\xi}(m) + \dfrac{s^2}{3}\nabla_\xi R_{1\xi\alpha\xi}(m) + O(s^3), \\
S_{\sigma\alpha\beta}(\eta) = \dfrac{1}{s}\delta_{\alpha\beta} - \dfrac{s}{3}R_{\alpha\xi\beta\xi}(m) - \dfrac{s^2}{4}\nabla_\xi R_{\alpha\xi\beta\xi}(m) + O(s^3),
\end{cases} \tag{5.4}$$

where $\alpha, \beta = 2, \ldots, n-1$;

(iii) *the square of the length of the shape operator is given by*

$$\operatorname{tr} S_\sigma^2(\eta) = \frac{n-2}{s^2} - \frac{2}{3}(\varrho_{\xi\xi} - R_{1\xi 1\xi})(m) - \frac{s}{2}(\nabla_\xi \varrho_{\xi\xi} - \nabla_\xi R_{1\xi 1\xi})(m) + o(s^2); \tag{5.5}$$

(iv) *the scalar curvature* $\tilde{\tau}$ *of the tube* P_σ *is given by*

$$\tilde{\tau}(\eta) = \frac{(n-3)(n-2)}{s^2} + \left(\tau - \frac{2n}{3}\varrho_{\xi\xi} - \frac{4n-6}{3}R_{1\xi 1\xi}\right)(m) + s\left(\nabla_\xi \tau - \frac{n+1}{2}\nabla_\xi \varrho_{\xi\xi} - \frac{n-1}{2}\nabla_\xi R_{1\xi 1\xi}\right)(m) + o(s^2). \tag{5.6}$$

Now we are ready to give a series of theorems.

Theorem 5.3. *A Riemannian manifold* M *is locally symmetric if and only if* M *is a D'Atri space with respect to all geodesics* σ *in* M.

Proof. First let M be a locally symmetric space. Then the Jacobi equation can be solved in terms of the eigenvalues of the endomorphism R along the geodesic γ and by using a basis of eigenvectors. These eigenvalues are constant along γ and the eigenvectors are parallel. The theorem can easily be checked by computing $\mathfrak{d}_\sigma$ explicitly.

Conversely let M be a D'Atri space with respect to all geodesics. Since $\mathfrak{d}_\sigma$ has antipodal symmetry we obtain from (5.3)

$$\nabla_\xi \varrho_{\xi\xi} + \nabla_\xi R_{1\xi 1\xi} = 0.$$

Since e_1 is arbitrary we have

$$\nabla_\xi \varrho_{\xi\xi} + \nabla_\xi R_{a\xi a\xi} = 0$$

for an arbitrary basis $\{e_a,\ a = 1, 2, \ldots, (n-1)\}$. Hence contraction implies

$$\nabla_\xi \varrho_{\xi\xi} = 0$$

and consequently

$$\nabla_\xi R_{\xi a \xi a} = 0.$$

From Lemma 5.1 it follows that M is locally symmetric.

In the course of proving the above, we have incidentally shown:

Theorem 5.4. *A necessary and sufficient condition that a Riemannian manifold be locally symmetric is that it possesses the following property. Let* γ *be an arbitrary geodesic, let* $m \in \gamma$, *and let* p, p' *be any two points of* γ *equidistant from and on opposite sides of* m. *The second fundamental form at* m *of the geodesic sphere with centre* p *passing through* m *is equal to the second fundamental form at* m *of the geodesic sphere with centre* p' *passing through* m.

It is also possible to give a characterization theorem using $\operatorname{tr} S_\sigma^2$. The proof is similar. One uses here (5.5) or (5.4) when $\dim M = 2$. So we obtain

Theorem 5.5. *A Riemannian manifold M is locally symmetric if and only if the length of the second fundamental form of a tube about a geodesic σ has antipodal symmetry for all geodesics σ in M.*

For the scalar curvature of a tube the best result we have is the following.

Theorem 5.6. *Let M be a Riemannian manifold with scalar curvature τ and with $\dim M > 2$. Then M is locally symmetric if and only if the scalar curvature τ is constant and the scalar curvature $\tilde{\tau}$ of a tube about a geodesic has antipodal symmetry for all geodesics.*

Proof. First we note that $\dim M > 2$ is needed since for $\dim M = 2$ we always have antipodal symmetry because $\tilde{\tau} = 0$.

When M is locally symmetric the theorem can be checked again directly.

Next suppose that $\tilde{\tau}$ has antipodal symmetry. Then (5.6) implies

$$\nabla_\xi \tau - \frac{n+1}{2} \nabla_\xi \varrho_{\xi\xi} - \frac{n-1}{2} \nabla_\xi R_{\xi 1 \xi 1} = 0.$$

Since e_1 is arbitrary we use again contraction. This gives

$$\nabla_\xi \tau - \frac{n+2}{2} \nabla_\xi \varrho_{\xi\xi} = 0$$

and since τ is constant we obtain

$$\nabla_\xi \varrho_{\xi\xi} = 0$$

and consequently

$$\nabla_\xi R_{\xi a \xi a} = 0.$$

This means that M is locally symmetric.

As is well-known locally symmetric spaces are also characterized as those spaces such that the geodesic symmetry with respect to a point is an isometry. For geodesic symmetries with respect to geodesics we prove

Theorem 5.7. *Let M be a Riemannian manifold. Then M is a space of constant curvature if and only if the geodesic symmetries with respect to all geodesics are isometries.*

When M is assumed to be of constant curvature, the Jacobi field equations can be solved explicitly and the value of $\mathfrak{d}_\sigma(\tau_t(s\xi))$ calculated. It is readily verified that this is invariant under the geodesic symmetry map with respect to geodesics.

To prove the converse, it will be sufficient to prove that the sectional curvature at an arbitrary point m is independent of the particular 2-plane chosen. Alternatively it will be sufficient to show that there is an isometry of some neighbourhood of m whose differential maps a 2-plane A at m to a given 2-plane B at m. But, due to the special properties of the exponential map and its differential when restricted to a suitable neighbourhood U of m, our hypothesis implies that at m the differential of the geodesic symmetry with respect to a geodesic σ through m is the geodesic symmetry of $(T_m M, g_m)$ with respect to the tangent line of the geodesic σ at m. An elementary geometrical argument shows that if A and B are two different 2-planes of M_m, then it is possible to map A to B by a composition of

geodesic symmetries. If $A \cup B$ spans a 3-dimensional subspace E^3, then the reflection in the line l which, in this E^3, bisects the dihedral angle between A and B, will map A to B. If $A \cap B = o_m$, we choose lines a, b lying in A, B, respectively, passing through o_m. Then a, b span a 2-plane C, and by the previous argument we have $A \to C \to B$ by the composition of two reflections. So, in any case, there is a local isometry whose differential maps A to B, and the sectional curvature at m is independent of the 2-plane chosen. It follows from Schur's theorem that the space is of constant curvature.

Acknowledgements. We thank Luc Gheysens and John Bolton for useful discussions; in particular we thank the referee whose helpful criticisms led to an improved version of the original paper.

References

1. Besse, A.L.: Manifolds all of whose geodesics are closed. Ergebnisse der Mathematik, Bd. 93. Berlin, Heidelberg, New York: Springer 1978
2. D'Atri, J.E., Nickerson, H.K.: Divergence-preserving geodesic symmetries. J. Differential Geometry **3**, 467–476 (1969)
3. D'Atri, J.E., Nickerson, H.K.: Geodesic symmetries in spaces with special curvature tensors. J. Differential Geometry **9**, 251–262 (1974)
4. D'Atri, J.E.: Geodesic spheres and symmetries in naturally reductive homogeneous spaces. Michigan Math. J. **22**, 71–76 (1975)
5. Dombrowski, P.: Jacobi fields, totally geodesic foliations, and geodesic differential forms. Resultate der Mathematik **1**, 156–194 (1977)
6. Gheysens, L.: Riemannse differentiaalmeetkunde van buisvormige omgevingen. Doctoral Dissertation K.U. Leuven, 1981
7. Gheysens, L., Vanhecke, L.: Total scalar curvature of tubes about curves. Math. Nachr. **103**, 177–197 (1981)
8. Gray, A.: Classification des variétés approximativement kählériennes de courbure sectionnelle holomorphe constante. C. R. Acad. Sci. Paris **279**, 797–800 (1974)
9. Gray, A., Vanhecke, L.: The volumes of tubes about curves in a Riemannian manifold. Proc. London Math. Soc. **44**, 215–243 (1982)
10. Gromoll, D., Klingenberg, W., Meyer, W.: Riemannsche Geometrie im Großen. Lectures Notes in Mathematics, Vol. 55. Berlin, Heidelberg, New York: Springer 1968

Received October 20, 1981; in revised form October 22, 1982

Math. Ann. 263, 43–49 (1983)

Mathematische Annalen
© Springer-Verlag 1983

The Hilbert-Kunz Function

P. Monsky

Department of Mathematics, Brandeis University, Waltham, MA 02254, USA

Introduction

Let $(\mathcal{O}, m)$ be a complete local Noetherian Z/p-algebra and I be an m-primary ideal of $\mathcal{O}$. For each $n \geqq 0$ let I_n be the "Frobenius p^n-th power" of I, that is to say the ideal generated by all x^{p^n}, $x \in I$. Suppose that M is a finitely generated $\mathcal{O}$-module. Then the M/I_nM are $\mathcal{O}$-modules of finite length. The function $n \to l(M/I_nM)$ will be called the Hilbert-Kunz function of M. It was introduced in [1, 2] (with $I=m$).

If $\mathcal{O}$ has altitude 0 then for large n $I_nM=(0)$ and the Hilbert-Kunz function is constant. We shall assume for the rest of this note that altitude $\mathcal{O}=a \geqq 1$. In Sect. 1 we shall show that $l(M/I_nM)=c(M)p^{an}+\delta_n$ where $c(M)$ is a real constant and $\delta_n=O(p^{(a-1)n})$. We suspect, but have no idea how to prove, that $c(M)$ is always rational. (The above result contradicts Kunz's assertion on page 1011 of [2] that $\lim_{\nu \to \infty} \lambda_\nu(R)$ need not exist. But Kunz's claim does not follow from his formulae – in the examples he is considering $\lim_{\nu \to \infty} \lambda_\nu(R)$ is in fact equal to m.)

When $a=1$ we get a far more precise result. We shall show that $c(M)$ is an integer. Also, using results about "linear recurrence functions" in fields of characteristic p, developed in Sect. 2, we shall show that δ_n is a periodic function of n for large n.

In a sequel to this paper we shall use these results to get information about the growth of the p-rank of the ideal class group in a Z_p^d-tower of number fields.

1

Lemma 1.1. *Let* $c_n=p^{-an}l(\mathcal{O}/I_n)$. *Then* $0<\liminf c_n \leqq \limsup c_n<\infty$.

Proof. Suppose that I is generated by s elements. Then any monomial of degree sp^n in these elements lies in I_n. Thus $I^{sp^n} \subset I_n \subset I^{p^n}$. But for large k, $l(\mathcal{O}/I^k)$ is a polynomial in k of precise degree a. The lemma follows.

Lemma 1.2. *Let* P_i *be those height* 0 *primes of* $\mathcal{O}$ *for which* $\mathcal{O}/P_i$ *has altitude* a. *Suppose that* M *is annihilated by some* h *not in any* P_i. *Then* $l(M/I_nM)=O(p^{(a-1)n})$.

Proof. M is a module over $\mathcal{O}/h$. We argue as in Lemma 1.1, noting that altitude $\mathcal{O}/h \leqq a-1$.

Lemma 1.3. *Let M and N be finitely generated $\mathcal{O}$-modules that have isomorphic localizations, as $\mathcal{O}_{P_i}$ modules, at each of the P_i of Lemma 1.2. Then $|l(M/I_nM) - l(N/I_nN)| = O(p^{(a-1)n})$.*

Proof. Let S be the complement of the union of the P_i. $S^{-1}\mathcal{O}$ is an Artinian ring which is a product of the $\mathcal{O}_{P_i}$, and $S^{-1}M \approx S^{-1}N$ over $S^{-1}\mathcal{O}$. Since $S^{-1}\operatorname{Hom}_{\mathcal{O}}(M,N) = \operatorname{Hom}_{S^{-1}\mathcal{O}}(S^{-1}M, S^{-1}N)$, there exists an $\mathcal{O}$-linear map $\phi: M \to N$ such that $S^{-1}\phi$ is bijective. If C is the cokernel of ϕ then $S^{-1}C = (0)$, and C is annihilated by some $h \in S$. Using the exact sequence $M/I_nM \to N/I_nN \to C/I_nC \to 0$ and Lemma 1.2 we find that $l(N/I_nN) \leqq l(M/I_nM) + O(p^{(a-1)n})$. Reversing the roles of M and N gives the result.

Until further notice we shall assume that the residue class field, $\mathcal{O}/m$, of $\mathcal{O}$ is perfect. This will allow us to exploit the Frobenius morphism.

Definition 1.4. $\lambda_s: \mathcal{O} \to \mathcal{O}$ is the map $g \to g^{p^s}$. If M is an $\mathcal{O}$-module then $M_{(s)}$ is the $\mathcal{O}$-module defined as follows: $M_{(s)} = M$ as additive group, but the product of $g \in \mathcal{O}$ with $m \in M_{(s)}$ is $g^{p^s}m$.

Lemma 1.5. (a) *λ_s is a finite morphism. So if M is finite over $\mathcal{O}$, so is $M_{(s)}$.*

(b) *$M \to M_{(s)}$ is an exact functor.*

(c) *If M is m-primary so is $M_{(s)}$, and $l(M_{(s)}) = l(M)$.*

(d) *For all n, $l(M_{(s)})/I_nM_{(s)}) = l(M/I_{n+s}M)$.*

Proof. $\mathcal{O}$ is a homomorphic image of $\Lambda_d = \mathcal{O}/m[[X_1, \ldots, X_d]]$ for some d. Since the p^s-th power map $\Lambda_d \to \Lambda_d$ is finite, λ_s is finite. (b) is obvious. Since (b) is true it suffices to prove (c) when $M = \mathcal{O}/m$. Now $\mathcal{O}/m_{(s)}$ is annihilated by m and so is a vector space over $\mathcal{O}/m$. Since $(\mathcal{O}/m)^{p^s} = \mathcal{O}/m$, the element 1 spans this vector space, giving (c). Finally $M_{(s)}/I_nM_{(s)}$ evidently identifies with $(M/I_{n+s}M)_{(s)}$ and we deduce (d) from (c).

Theorem 1.6. *Let $0 \to M' \to M \to M'' \to 0$ be an exact sequence of finitely generated $\mathcal{O}$-modules. Then $l(M/I_nM) = l(M'/I_nM') + l(M''/I_nM'') + O(p^{(a-1)n})$.*

Proof. We adopt the notation of Lemma 1.2. Suppose first that $\mathcal{O}$ is reduced so that each $\mathcal{O}_{P_i}$ is a field. Then M and $M' \oplus M''$ have isomorphic localizations at each P_i, and the result follows from Lemma 1.3.

In the general case let N be the nilradical of $\mathcal{O}$ and choose an s such that $N^{p^s} = (0)$. Then if M is any $\mathcal{O}$-module, $M_{(s)}$ is annihilated by N. Consequently, we have an exact sequence of finitely generated $\mathcal{O}/N$ – modules:

$$0 \to M'_{(s)} \to M_{(s)} \to M''_{(s)} \to 0.$$

The result of the last paragraph combined with (d) of Lemma 1.5 gives the theorem.

Lemma 1.7. *Suppose that $\mathcal{O}$ is a domain. Then there is a positive real constant c such that $l(\mathcal{O}/I_n) = cp^{an} + O(p^{(a-1)n})$.*

Proof. Let r be the rank of $\mathcal{O}_{(1)}$ as $\mathcal{O}$-module. Lemma 1.3 applied to $\mathcal{O}_{(1)}$ and $\mathcal{O}^r$ shows that

$$|l(\mathcal{O}_{(1)}/I_n\mathcal{O}_{(1)}) - r\,l(\mathcal{O}/I_n)| \leqq \text{constant} \cdot (p^{(a-1)n}).$$

By Lemma 1.5 (d), $|l(\mathcal{O}/I_{n+1}) - rl(\mathcal{O}/I_n)| \leqq \text{constant} \cdot (p^{(a-1)n})$. If we set $l(\mathcal{O}/I_n) = c_n p^{an}$ as in Lemma 1.1 we find that $\left| c_{n+1} - \frac{r}{p^a} c_n \right| \leqq \frac{\text{constant}}{p^n}$. Suppose that $\frac{r}{p^a} > 1$. If we take n large with c_n close to $\limsup c_m$ we get c_{n+1} close to $\frac{r}{p^a} \limsup c_m$ which is impossible. Similarly if $\frac{r}{p^a} < 1$ we get a contradiction by taking n large with c_n close to $\liminf c_m$. So $r = p^a$, and the c_n form a Cauchy sequence converging to some c. By Lemma 1.1, $c \neq 0$. Finally $c_n = c + O\left(\frac{1}{p^n}\right)$, giving the result.

Theorem 1.8. *There is a real constant $c(M)$ such that $l(M/I_nM) = c(M)p^{an} + O(p^{(a-1)n})$. If $0 \to M' \to M \to M'' \to 0$ is an exact sequence of finitely generated $\mathcal{O}$-modules, then $c(M) = c(M') + c(M'')$.*

Proof. Suppose first that $M = \mathcal{O}/P$ with P prime. If P is one of the P_i of Lemma 1.2 we apply Lemma 1.7 to the domain $\mathcal{O}/P_i$. If $P \neq$ any P_i then M is annihilated by some h not in any P_i, and we use Lemma 1.2. Suppose now that M is arbitrary. Then M can be "built up" out of cyclic modules $\mathcal{O}/P$ with P prime, and the result follows from repeated applications of Theorem 1.6. The final assertion is immediate from Theorem 1.6.

We now drop the assumption that $\mathcal{O}/m$ is perfect and sketch a proof that Theorem 1.8 still holds. Take a co-efficient field $K \subset \mathcal{O}$ projecting onto $\mathcal{O}/m$. Let $g_1, \ldots, g_d$ generate I and set $\Lambda_d = K[[X_1, \ldots, X_d]]$. There is a finite K-algebra map $\phi : \Lambda_d \to \mathcal{O}$ with $\phi(X_j) = g_j$. Now let F be an algebraic closure of K. Let $\Lambda_d^{\#} = F[[X_1, \ldots, X_d]]$ and $\mathcal{O}^{\#}$ be the tensor product of $\Lambda_d^{\#}$ with $\mathcal{O}$ over Λ_d. It may be shown that $\mathcal{O}^{\#}$ is flat over $\mathcal{O}$, and that $\mathcal{O}^{\#}$ is a complete local ring of altitude a with maximal ideal $m\mathcal{O}^{\#}$ and residue class field F. If M is a finitely generated $\mathcal{O}$-module let $M^{\#}$ be the tensor product of M with $\mathcal{O}^{\#}$ over $\mathcal{O}$ (this is the same as the tensor product of M with $\Lambda_d^{\#}$ over Λ_d). Then $M^{\#}/I_nM^{\#}$ is just the tensor product of M/I_nM with F over K. So the $\mathcal{O}$-length of M/I_nM is equal to the $\mathcal{O}^{\#}$-length of $M^{\#}/I_nM^{\#}$, giving the first part of Theorem 1.8 and showing that $c(M) = c(M^{\#})$. Tensoring the exact sequence $0 \to M' \to M \to M'' \to 0$ with $\mathcal{O}^{\#}$ over $\mathcal{O}$, and using the flatness of $\mathcal{O}^{\#}$ over $\mathcal{O}$ we get the second part of Theorem 1.8.

Corollary. *Let K be a field of characteristic p and M be a finitely generated module over $\Lambda_d = K[[X_1, \ldots, X_d]]$. Let I_n be the ideal in Λ_d generated by the $X_j^{p^n}$ and let a be the altitude of the local ring $\mathcal{O} = \Lambda_d / \operatorname{Ann} M$. Then there is a positive real constant $c(M)$ such that $l(M/I_nM) = c(M)p^{an} + O(p^{(a-1)n})$.*

Proof. In view of Theorem 1.8 we only need show that $c(M) > 0$. Let m be the maximal ideal of $\mathcal{O}$. Then $l(M/I_nM) \geqq l(M/m^{p^n}M)$. But this is a polynomial of exact degree a in p^n for n large.

Theorem 1.9. *Suppose that $\mathcal{O}$ is a domain of altitude 1. Then $c(\mathcal{O})$ is an integer which can be described explicitly as follows. Let $\mathcal{O}'$ be the integral closure of $\mathcal{O}$ in its field of quotients, ord be the valuation function in the discrete valuation ring $(\mathcal{O}', m')$, and f be the degree of $\mathcal{O}'/m'$ over $\mathcal{O}/m$. Then $c(\mathcal{O}) = f \cdot \left(\min_{g \in I} \operatorname{ord} g \right)$.*

Proof. By Lemma 1.3, $|l(\mathcal{O}/I_n) - l(\mathcal{O}'/I_n\mathcal{O}')| = O(1)$. Now

$$l(\mathcal{O}'/I_n\mathcal{O}') = f \cdot \left(\min_{g \in I_n} \operatorname{ord} g \right) = f \cdot \left(\min_{g \in I} \operatorname{ord} g \right) \cdot p^n,$$

giving the result.

Theorem 1.10. *If altitude* $\mathcal{O}=1$, $c(M)$ *is an integer.*

Proof. M is "built up" out of cyclic modules $\mathcal{O}/P$ with P prime. So by Theorem 1.8 we may assume that $M=\mathcal{O}/P$. If $P=m$, $c(M)=0$, while if height $P=0$ we may apply Theorem 1.9 to the domain $\mathcal{O}/P$.

Remarks and Examples

When $a=1$, $l(M/I_nM)=c(M)p^n+\delta_n$ where the δ_n are bounded. δ_n need *not* be constant for large n. Suppose for example that p is a prime $\equiv \pm 2(5)$, that $\mathcal{O}=M=Z/p[[X,Y]]/(X^5-Y^5)$ and that $I=(X,Y)$. Then $l(M/I_nM)=5p^n+\delta_n$ where $\delta_n=-4$ or -6 according as n is even or odd. We shall see in Sect. 3 that when $a=1$ the δ_n are always periodic for large n.

When $a\geqq 2$, $c(M)$ need *not* be an integer. Suppose for example that $\mathcal{O}=M=Z/p[[X,Y,Z]]/(Z^m-XY)$, and that $I=(X,Y,Z)$. It may be shown that $c(M)=2-\frac{1}{m}$. We suspect that $c(M)$ is always rational. For further examples, see the last sections of [1, 2].

2

Throughout this section F is a field of characteristic p containing the algebraic closure, F_0, of the prime field.

Definition 2.1. A function ϕ from the non-negative integers to F is a "linear recurrence function" if there exist α_i and β_i in F such that $\phi(n)=\sum\alpha_i\beta_i^{p^n}$ for all n.

When $F=F_0$, a linear recurrence function is periodic. When $F\neq F_0$ this need not hold of course, but we shall prove the following. Let $A(n)$ be a matrix whose entries are linear recurrence functions. Then for n large, the rank of A is a periodic function of n.

Lemma 2.2. *Suppose that* α_i *and* $\beta_i\in F$ *with the* β_i *not all* 0. *If* $\sum\alpha_i\beta_i^{p^n}$ *vanishes for infinitely many* n *then the* α_i *are linearly dependent over* F_0.

Proof. We may replace F by any extension field of F_0 containing the α_i and the β_i. In particular, since the power series field $F_0\langle\langle T\rangle\rangle$ has infinite transcendence degree over F_0, we may replace F by a finite extension of $F_0\langle\langle T\rangle\rangle$. Then we can find a discrete valuation ring (A,m) in F with $A/m=F_0$. Multiplying each α_i by one element of F^* and each β_i by another we may assume that α_i and β_i are in A and that some β_i is a unit in A. Write $\beta_i=\gamma_i-\delta_i$ with $\gamma_i\in F_0$ and $\delta_i\in m$. Choose an r such that $\gamma_i^{p^r}=\gamma_i$ for each i, and an n_0 such that $\sum\alpha_i\beta_i^{p^n}=0$ for infinitely many $n\equiv n_0(r)$. Then for each of these n, $\sum\alpha_i\gamma_i^{p^{n_0}}=\sum\alpha_i\gamma_i^{p^n}=\sum\alpha_i(\beta_i+\delta_i)^{p^n}=\sum\alpha_i\delta_i^{p^n}$. Thus $\operatorname{ord}(\sum\alpha_i\gamma_i^{p^{n_0}})\geqq p^n\min(\operatorname{ord}\delta_i)$ for infinitely many n. So $\sum\alpha_i\gamma_i^{p^{n_0}}=0$, and the α_i are linearly dependent over F_0.

Lemma 2.3. *Let* ϕ *be a linear recurrence function. Then there exists an* r *such that* $\{n\mid\phi(n)=0\}$ *is the union of finitely many residue classes* mod r *with a finite set.*

Proof. Write $\phi(n)=\sum\alpha_i\beta_i^{p^n}$ and let w_j be a basis of the F_0-space spanned by the α_i. Then $\phi(n)=\sum\sum(\lambda_{ij}\beta_i^{p^n})w_j$ with $\lambda_{ij}\in F_0$. Choose an r so that $\lambda_{ij}^{p^r}=\lambda_{ij}$ for all i and j.

Suppose that $n \equiv a$ fixed $-n_0 \bmod r$. Then $\phi(n) = \sum\sum(\lambda_{ij}^{p^{n+n_0}}\beta_i^{p^n})w_j = \sum(w_j \cdot(\sum\lambda_{ij}^{p^{n_0}}\beta_i)^{p^n})$, which may be written as $\sum w_j\gamma_j^{p^n}$. If each γ_j vanishes then $\phi(n)=0$ for all $n \equiv -n_0 \bmod r$. If some $\gamma_j \neq 0$, Lemma 2.2 shows that $\phi(n)=0$ for only finitely many $n \equiv -n_0 \bmod r$.

Theorem 2.4. *Let ϕ_{ij} be linear recurrence functions. Then the dimension of the solution space of the system of equations $\sum_j \phi_{ij}(n)x_j = 0$ is a periodic function of n for large n.*

Proof. The determinant of an $r \times r$ matrix of linear recurrence functions is evidently a linear recurrence function. Thus the rank of the matrix $\phi_{ij}(n)$ can be calculated from a knowledge of which of a fixed finite set of linear recurrence functions vanish at n, and we use Lemma 2.3.

3

Throughout this section $(\mathcal{O}, m)$ is a complete local Noetherian Z/p-algebra of altitude 1. Until further notice we assume that $\mathcal{O}/m$ is algebraically closed. Let $K \subset \mathcal{O}$ be a co-efficient field projecting onto $\mathcal{O}/m$. Note that the length of an m-primary $\mathcal{O}$-module is just its K-dimension. Let P_i be the height 0 primes of $\mathcal{O}$; each $\mathcal{O}/P_i$ has altitude 1.

Definition 3.1. $Q_i = \{x \in \mathcal{O}$ annihilated by some $s \notin P_i\}$ $\qquad T = \cap Q_i$.

Since the annihilator of T is not contained in any P_i it must be m-primary. Thus T has finite length, and there exists a k such that $T \cap m^k = (0)$.

Lemma 3.2. *Suppose that G and H are in m^k. If G divides H in $\mathcal{O}/T$, then G divides H in $\mathcal{O}$.*

Proof. $H \equiv AG \bmod T$; since $H - AG \in T \cap m^k$, $H = AG$.

Definition 3.3. $\mathcal{O}'$ is the product of the $\mathcal{O}/Q_i$.

Lemma 3.4. *The cokernel of the inclusion $\mathcal{O}/T \hookrightarrow \mathcal{O}'$ is an m-primary $\mathcal{O}$-module.*

Proof. Since $T\mathcal{O}_{P_i} = Q_i\mathcal{O}_{P_i} = (0)$ the localization of the above map at any P_i is a bijection. Thus the cokernel localizes to (0) at each P_i, and is m-primary.

Lemma 3.5. *Suppose that $\mathcal{O}$ has precisely one height 0 prime, P, and that G and H are in $\mathcal{O}$ with* ord$G \leqq$ ord$H$, *where* ord *is the valuation function in the integral closure, $\mathcal{O}_1$, of $\mathcal{O}/P$. Then, for n large, G^{p^n} divides H^{p^n}.*

Proof. Since $\mathcal{O}/m$ is algebraically closed the residue class field of $\mathcal{O}_1$ is $\mathcal{O}/m$. Thus we can modify H by a K-multiple of G and assume that ord$G <$ ord$H$. Then $H = AG$ with $A \in \mathcal{O}_1$ and ord$A > 0$. For n large, $A^{p^n} \in \mathcal{O}/P$. So G^{p^n} divides H^{p^n} in $\mathcal{O}/P$. Since P is the nilradical of $\mathcal{O}$, $P^{p^s} = 0$ for some s. Then $G^{p^{n+s}}$ divides $H^{p^{n+s}}$.

Now let $g_1, \ldots, g_s$ be generators of I. Let $\mathcal{O}_i$ be the integral closure of $\mathcal{O}/P_i$ in its field of quotients, and ord$_i$ be the valuation function in $\mathcal{O}_i$. Since K is infinite we can find a K-linear combination, g, of the g_j such that $\text{ord}_i g = \min_j \text{ord}_i(g_j)$ for each i.

Lemma 3.6. *For n large, g^{p^n} divides each $g_j^{p^n}$ in $\mathcal{O}'$.*

Proof. It suffices to show that g^{p^n} divides $g_j^{p^n}$ in $\mathcal{O}/Q_i$. Since $\mathrm{ord}_i(g) \leqq \mathrm{ord}_i(g_j)$ this follows from Lemma 3.5 applied to $\mathcal{O}/Q_i$.

Lemma 3.7. *There is a finite subset $\{u_i\}$ of $\mathcal{O}$ with the following property. For each large n the vector space $I_n/(g^{p^n})$ is spanned over K by the elements $u_i g_j^{p^n}$.*

Proof. Take an $h \notin$ any P_i such that $h \cdot \mathcal{O}' \subset \mathcal{O}/T$; this is possible by Lemma 3.4. Suppose n is large. Since g^{p^n} divides $g_j^{p^n}$ in $\mathcal{O}'$, g^{p^n} divides $hg_j^{p^n}$ in $\mathcal{O}/T$. By Lemma 3.2, g^{p^n} divides $hg_j^{p^n}$ in $\mathcal{O}$. Let $\{u_i\}$ be any K-basis of $\mathcal{O}/(h)$. Then I_n is spanned over K by the $u_i g_j^{p^n}$ and $\mathcal{O}$-multiples of the $hg_j^{p^n}$. But the $hg_j^{p^n}$ are in (g^{p^n}).

Lemma 3.7 shows that the vector spaces $I_n/(g^{p^n})$ have bounded dimension. We now use the ideas of Sect. 2 to prove a periodicity result for these dimensions.

Lemma 3.8. *Suppose that u and δ are in $\mathcal{O}'$ and that $\phi : \mathcal{O}' \to K$ is a linear functional vanishing on $\mathcal{O}/T$. Then (for large values of n) the function $n \to \phi(u\delta^{p^n})$ is a linear recurrence function in the sense of Definition 2.1.*

Proof. Let δ_i be the projection of δ on $\mathcal{O}/Q_i$. Write $\delta_i = \beta_i + \gamma_i$ with $\beta_i \in K$ and $\gamma_i \in m/Q_i$. The β_i and the γ_i define elements β and γ of $\mathcal{O}'$ and $\delta = \beta + \gamma$. Suppose that n is large. Using Lemma 3.4 we find that $u\gamma^{p^n} \in \mathcal{O}/T$. Thus $\phi(u\delta^{p^n}) = \phi(u\beta^{p^n})$. Let u_i be the projection of u on $\mathcal{O}/Q_i$. Then $\phi(u\beta^{p^n}) = \phi(u_1, 0, \ldots) \cdot \beta_1^{p^n} + \phi(0, u_2, 0 \ldots) \cdot \beta_2^{p^n} + \ldots$, proving the lemma.

Lemma 3.9. *For large n the dimension of the K-vector space $I_n/(g^{p^n})$ is a periodic function of n.*

Proof. By Lemma 3.6 we may assume that $g_j = \delta_j g$ with $\delta_j \in \mathcal{O}'$. Choose u_i as in Lemma 3.7. Then the $u_i g_j^{p^n}$ span $I_n/(g^{p^n})$ over K. When does a K-linear combination, $\sum c_{ij} u_i g_j^{p^n}$, of these elements represent 0? This happens when g^{p^n} divides $\sum c_{ij} u_i g_j^{p^n}$ in $\mathcal{O}$. By Lemma 3.2 this is the same as g^{p^n} dividing $\sum c_{ij} u_i g_j^{p^n}$ in $\mathcal{O}/T$ and this happens precisely when $\sum c_{ij} u_i \delta_j^{p^n} \in \mathcal{O}/T$. Let ϕ_k be a basis for the space of linear functionals $\mathcal{O}' \to K$ vanishing on $\mathcal{O}/T$. Then $\sum c_{ij} u_i g_j^{p^n} \in (g^{p^n})$ if and only if $\sum c_{ij} \phi_k(u_i \delta_j^{p^n}) = 0$ for each ϕ_k. By Lemma 3.8 this amounts to finitely many linear conditions on the c_{ij}, with the co-efficients being linear recurrence functions of n. Now apply Theorem 2.4.

Theorem 3.10. *For n large, $l(\mathcal{O}/I_n) = c(\mathcal{O})p^n + a$ periodic function of n.*

Proof. (g) is an m-primary ideal in $\mathcal{O}$. So $l(\mathcal{O}/g^k)$ is a degree 1 polynomial in k for large k. Take $k = p^n$ and apply Lemma 3.9.

Observe now that Theorem 3.10 holds without the restriction that $\mathcal{O}/m$ be algebraically closed. For as in Sect. 1, we may construct an $\mathcal{O}$-algebra, $\mathcal{O}^*$, which is complete local Noetherian of altitude 1, with residue class field the algebraic closure F of K, such that the $\mathcal{O}$-length of $\mathcal{O}/I_n$ is equal to the $\mathcal{O}^*$ length of $\mathcal{O}^*/I_n\mathcal{O}^*$ for all n. We then apply Theorem 3.10 to $\mathcal{O}^*$.

Theorem 3.11. *$l(M/I_nM) = c(M)p^n + \delta_n$ where $c(M)$ is an integer and δ_n is a periodic function of n for large n.*

Proof. We shall deduce this from Theorem 3.10 by Nagata's "idealization" trick. We make $\mathcal{O}^* = \mathcal{O} \oplus M$ into a ring in such a way that M is an ideal of square 0. Then $\mathcal{O}^*$ is complete local Noetherian of altitude 1, with maximal ideal $m^* = m \oplus M$, and $I^* = I \oplus IM$ is an m^*-primary ideal. We see easily that $(I^*)_n = I_n + I_n M$. So Theorem 3.10 applied to $\mathcal{O}^*$ shows that $l(\mathcal{O}/I_n) + l(M/I_n M) = c^* p^n +$ (a periodic function of n) for large n. We conclude that $l(M/I_n M) = (c^* - c)p^n +$ (a periodic function of n) for large n.

References

1. Kunz, E.: Characterizations of regular local rings of characteristic p. Am. J. Math. **41**, 772–784 (1969)
2. Kunz, E.: On Noetherian rings of characteristic p. Am. J. Math. **98**, 999–1013 (1976)

Received October 26, 1982

Math. Ann. 263, 51–59 (1983)

Mathematische Annalen
© Springer-Verlag 1983

A Variational Approach to Bifurcation in L^p on an Unbounded Symmetrical Domain

C. A. Stuart*

Department of Mathematics, Heriot-Watt University, Edinburgh, Scotland

1. Introduction

Consider the nonlinear eigenvalue problem,

$$-\Delta u(x)-q(x)|u(x)|^{\sigma}u(x)=\lambda u(x) \quad \text{for} \quad x\in\Omega\subset\mathbb{R}^N, \tag{1.1}$$

where Ω is an unbounded domain. The constant $\sigma>0$ and the function $q:\Omega\to(0,\infty)$ are given. We are concerned with the existence of solutions $(\lambda,u)\in\mathbb{R}\times H_0^1(\Omega)$ of (1.1) and the behaviour of $|u|_p$ (the L^p norm of u on Ω) as λ approaches λ^*, the infimum of the spectrum of the linearisation of (1.1). In the cases $\Omega=\mathbb{R}^N$ and $\Omega=\mathbb{R}\times(-\pi/2,\pi/2)\subset\mathbb{R}^2$, these questions have been studied in great detail by Toland [1, 2] and Amick and Toland [3]. The method has been to study (1.1) on a sequence of bounded sub-domains $\{\Omega_k\}$ which expand to fill Ω as $k\to\infty$ and then to take limits. Although yielding very precise information, this method is rather lengthy and involves quite intricate analysis.

In this paper we present a variational approach to the problem. It yields somewhat less complete results (we treat only $p\geqq 2$ and no information about the connectedness of the set of solutions is obtained), but avoids the approximation by bounded domains and is considerably shorter than the previous approach. It also furnishes directly estimates of the type, $|u|_p\leqq K(\lambda^*-\lambda)^{\gamma(p,\sigma,N)}$ for solutions (λ,u) with $\lambda<\lambda^*$. The exponent $\gamma(p,\sigma,N)$ yields optimal results about the bifurcation of solutions from $(\lambda^*,0)$ in $L^p(\Omega)$.

The variational method works as follows. For a fixed value of λ, the solutions u of (1.1) are critical points of

$$f(u)\equiv\frac{1}{2}|\nabla u|_2^2-\frac{\lambda}{2}|u|_2^2-\frac{1}{(\sigma+2)}\int_{\Omega}q|u|^{\sigma+2}\,dx \quad \text{on} \quad H_0^1(\Omega).$$

In general, f is neither bounded above nor below on $H_0^1(\Omega)$. However, there is a smooth manifold M of codimension one in $H_0^1(\Omega)$ which contains all non-trivial

* Permanent address: Dépt. de mathématiques, EPFL., CH-1015 Ecublens, Switzerland

($u \not\equiv 0$) solutions of (1.1) and f is bounded below on M. The existence of solutions of (1.1) is established by showing that f attains its infimum on M. This involves finding a relatively compact minimising sequence and can be arranged by exploiting symmetries of the problem. Furthermore, $\inf_M f$ can easily be estimated by using suitable test functions and yields estimates of $|u|_2$ and $|u|_{\sigma+2}$ for the minimiser. Other values of $|u|_p$ are estimated by interpolation. The idea of using the natural constraint M for indefinite problems goes back to Nehari [4]. It is set out in a general form which is convenient for our purpose in Sect. 2. Sections 3 and 4 deal with (1.1) in the cases where $\Omega = \mathbb{R}^N$ and q is Schwarz symmetric and $\Omega = \mathbb{R} \times (-\pi/2, \pi/2)$ and q is doubly Steiner symmetric, respectively. Let us emphasise that the method could just as easily be applied to problems with other symmetries such as:

(a) $\Omega = \mathbb{R}^N$ and q is Steiner symmetric with respect to the plane $x_N = 0$ and $q(\cdot, x_N)$ is Schwarz symmetric on $\mathbb{R}^{N-1}$ for all $x_N \geqq 0$, or

(b) $\Omega = \mathbb{R}^2$ and q is doubly Steiner symmetric on $\mathbb{R}^2$.

Results in the L^2 theory can be obtained without using special symmetries of the problem [7, 8].

2. Existence by Minimisation for Indefinite Problems

Let H be a real Hilbert space with norm $\|\cdot\|$ and scalar product $\langle \cdot, \cdot \rangle$. We consider a functional ϕ on H which has the following properties.

$$\text{(A)} \quad \begin{cases} \phi \in C^1(H, \mathbb{R}), \\ \exists \sigma > 0 \text{ such that } \phi(tu) = t^{\sigma+2}\phi(u) \quad \forall u \in H \quad \forall t \geqq 0, \\ \exists K > 0 \text{ such that } 0 < \phi(u) \leqq K\|u\|^{\sigma+2} \quad \forall u \in H \setminus \{0\}. \end{cases}$$

The problem is to find $u \in H \setminus \{0\}$ such that

$$\langle u, v \rangle = \phi'(u)v \quad \forall v \in H. \tag{2.1}$$

Setting $f(u) = \frac{1}{2}\|u\|^2 - \phi(u)$, we see that $u \in H \setminus \{0\}$ satisfies (2.1) if and only if $f'(u) = 0$. In general, f is neither bounded above nor below on H, so we cannot expect to find solutions of (2.1) by finding global extrema of f on H. Let

$$M = \{u \in H \setminus \{0\} : \|u\|^2 = (\sigma+2)\phi(u)\}.$$

The importance of M and its properties are set out in the following remarks.

(M1) Since $\phi'(u)u = (\sigma+2)\phi(u)$ $\forall u \in H$, we see that all non-trivial solutions of (2.1) belong to M.

(M2) For $u \in M$, $f(u) = \frac{1}{2}\left(\frac{\sigma}{\sigma+2}\right)\|u\|^2 = \frac{1}{2}\sigma\phi(u) > 0$. In particular, f is bounded only if $t = t(u)$.

(M3) For $u \in M$, $\|u\|^2 \leqq (\sigma+2)K\|u\|^{\sigma+2}$ and so $\|u\| \geqq \{(\sigma+2)K\}^{-1/\sigma}$.

(M4) For $u \in H \setminus \{0\}$, set $t(u) = \left(\frac{\|u\|^2}{(\sigma+2)\phi(u)}\right)^{1/\sigma}$. Then $tu \in M$ for $t \geqq 0$ if and only if $t = t(u)$.

(M5) Setting $g(u) = (\sigma+2)\phi(u) - \|u\|^2$ we have that

$$M = \{u \in H \setminus \{0\} : g(u) = 0\} \quad \text{and} \quad g'(u)v = (\sigma+2)\phi'(u)v - 2\langle u, v \rangle.$$

For $u\in M$ and $v=u$, we have $g'(u)u=\sigma\|u\|^2>0$. Hence M is a C^1-manifold of codimension one in H.

(M6) If $u\in M$ and $f(u)=\min_M f$, then u satisfies (2.1). In fact, by (M5) and the principle of Euler-Lagrange, $\exists\xi\in\mathbb{R}$ such that $f'(u)v=\xi g'(u)v\ \forall v\in H$. Since $u\in M$, $f'(u)u=0$ and $g'(u)u=\sigma\|u\|^2$. Hence $\xi=0$ and $f'(u)=0$.

According to (M6), we can find a solution of (2.1) by showing that f attains its infimum on M. For this we need a compactness property of a minimising sequence.

$$\text{(P)}\quad\begin{cases}\exists a \text{ sequence } \{u_n\}\subset M \text{ such that } f(u_n)\to\inf_M f,\\ u_n\rightharpoonup u_\infty \text{ weakly in } H \text{ and}\\ \phi(u_n)\to\phi(u_\infty).\end{cases}$$

Remarks. Let $\{w_n\}\subset M$ be any sequence such that $f(w_n)\to\inf_M f$. By (M2), $\{w_n\}$ is bounded in H and so we can select a subsequence which converges weakly in H. It follows that (P) is satisfied whenever ϕ is weakly sequentially continuous. However we are mostly interested in cases where (P) holds but ϕ is not weakly sequentially continuous on H. The idea is to show that some special (e.g. possessing some kind of symmetry) minimising sequences satisfy (P).

Proposition 2.1. *Let ϕ satisfy the conditions* (A) *and* (P). *Then $\exists u\in M$ such that $f(u)=\min_M f$.*

Thus u satisfies (2.1) *and* $\frac{1}{2}\left(\frac{\sigma}{\sigma+2}\right)\|u\|^2=\frac{1}{2}\sigma\phi(u)=f(u)=\min_M f$.

Proof. Consider the sequence given by (P). Then,

$$\phi(u_\infty)=\lim\phi(u_n)=\frac{1}{(\sigma+2)}\lim\|u_n\|^2\geqq\frac{1}{(\sigma+2)}\{(\sigma+2)K\}^{-2/\sigma}>0,$$

by (M3). Hence we have that $u_\infty\neq 0$.

Also, $\|u_\infty\|^2\leqq\liminf\|u_n\|^2=\liminf(\sigma+2)\,\phi(u_n)=(\sigma+2)\,\phi(u_\infty)$. Hence $t_\infty\equiv t(u_\infty)\leqq 1$ and $t_\infty u_\infty\in M$, by (M4). By (M2),

$$\begin{aligned}2f(t_\infty u_\infty)&=\sigma\phi(t_\infty u_\infty)=\sigma t_\infty^{\sigma+2}\phi(u_\infty)\\&=\sigma t_\infty^{\sigma+2}\lim\phi(u_n)=2t_\infty^{\sigma+2}\lim f(u_n)\\&=2t_\infty^{\sigma+2}\inf_M f.\end{aligned}$$

Thus we see that $t_\infty=1$ and so $u_\infty\in M$. q.e.d.

3. Schwarz Symmetry on $\mathbb{R}^N$

A continuous function $u:\mathbb{R}^N\to[0,\infty)$ is said to be Schwarz symmetric on $\mathbb{R}^N$ if it is radially symmetric and radially nonincreasing (see [10–16]).

In this section we consider the nonlinear eigenvalue problem,

$$-\Delta u(x)-q(x)\,|u(x)|^\sigma u(x)=\lambda u(x)\quad\text{for}\quad x\in\mathbb{R}^N,\tag{3.1}$$

where

$$\text{(B)}\quad \begin{cases} 0<\sigma<\infty \text{ if } N=1,2 \text{ and } 0<\sigma<\dfrac{4}{N-2} \text{ if } N\geqq 3, \\ q:\mathbb{R}^N\to(0,\infty) \text{ is continuous, Schwarz symmetric on } \mathbb{R}^N \\ \text{and } q(x)\geqq A>0 \quad \forall x\in\mathbb{R}^N. \end{cases}$$

A non-trivial weak solution of (3.1) is a pair (λ, u) such that $\lambda\in\mathbb{R}$, $u\in H^1(\mathbb{R}^N)\backslash\{0\}$ and

$$\int_{\mathbb{R}^N}\{\nabla u\cdot\nabla v - q|u|^\sigma uv-\lambda uv\}\,dx=0 \quad \forall v\in H^1(\mathbb{R}^N). \tag{3.2}$$

Remarks. 1. In view of the Sobolev embedding and the restriction on σ in (B), all the integrals in (3.2) converge.

2. Using the interior L^p regularity theory of elliptic equations [3, 5] we can show that if (λ, u) is a weak solution of (3.1), then u can be identified with a function which is twice continuously differentiable on $\mathbb{R}^N$ and which satisfies (3.1) in the classical sense.

3. As is shown in [2], a scaling argument of Gidas and Spruck [6] can be used to establish the following L^∞ a-priori bound. Given $\xi>0$ $\exists K(\xi)>0$ such that

$$|u|_\infty\leqq K(\xi), \tag{3.3}$$

whenever (λ, u) is a weak solution of (3.1) such that $\lambda\in(-\xi,0)$ and u is positive and Schwarz symmetric on $\mathbb{R}^N$.

4. In general, (3.1) has no solutions for $\lambda>0$ [2, 9].

Concerning the existence of solutions for $\lambda<0$ and their behaviour as $\lambda\to 0$, we have the following result.

Theorem 3.1. *Under the hypotheses* (B), $\forall\lambda<0$ $\exists$ *a non-trivial weak solution* $(\lambda, u(\lambda))$ *of* (3.1) *such that* $u(\lambda)$ *is positive and Schwarz symmetric on* $\mathbb{R}^N$. *Furthermore,* $\exists K>0$ *such that for* $\lambda\in[-1,0)$,

$$|u(\lambda)|_p\leqq K|\lambda|^{\frac{1}{\sigma}-\frac{N}{2p}} \quad \text{if} \quad 2\leqq p\leqq\sigma+2, \tag{3.4}$$

$$\leqq K|\lambda|^{\frac{2}{p}\left\{\frac{1}{\sigma}-\frac{N-2}{4}\right\}} \quad \text{if} \quad p\geqq\sigma+2, \tag{3.5}$$

$$|\nabla u(\lambda)|_2\leqq K|\lambda|^{\frac{1}{\sigma}-\frac{N-2}{4}}. \tag{3.6}$$

Proof. Fix $\lambda<0$ and set $\|u\|=\{|\nabla u|_2^2-\lambda|u|_2^2\}^{1/2}$ for $u\in H^1(\mathbb{R}^N)$. Then $\|\cdot\|$ is equivalent to the usual norm on $H^1(\mathbb{R}^N)$. We shall apply Proposition 2.1 with $H=H^1(\mathbb{R}^N)$ and $\phi(u)=\dfrac{1}{(\sigma+2)}\displaystyle\int_{\mathbb{R}^N} q|u|^{\sigma+2}\,dx$.

Since

$$q\in L^\infty(\mathbb{R}^N) \quad \text{and} \quad 0<\sigma<4/N-2 \quad (\text{for } N\geqq 3),$$

$$\phi\in C^1(H,\mathbb{R}) \quad \text{and,} \quad \text{for} \quad u\in H\backslash\{0\},$$

$$0<\phi(u)\leqq\frac{q(0)}{(\sigma+2)}|u|_{\sigma+2}^{\sigma+2}\leqq C|\nabla u|_2^\alpha|u|_2^\beta$$

$$\leqq C|\lambda|^{-\beta/2}\|u\|^{\sigma+2},$$

where we have used the multiplicative form of the Sobolev inequality with $\alpha = \frac{N\sigma}{2}$ and $\alpha+\beta=\sigma+2$ (see [7]). This shows that the conditions (A) of Sect. 2 are satisfied. To verify the property (P), consider a sequence $\{w_n\} \subset M$ such that $f(w_n) \to \inf_M f$, where M and f are as defined in Sect. 2. Let w_n^* be the Schwarz symmetrisation of $|w_n|$, [10, 11, 13–15]. Then $w_n^* \in H$ and

$$|\nabla w_n^*|_2 \leqq |\nabla |w_n||_2 \leqq |\nabla w_n|_2,$$
$$|w_n^*|_2 = |w_n|_2, \qquad \|w_n^*\| \leqq \|w_n\|$$

and

$$\int_{\mathbb{R}^N} q(w_n^*)^{\sigma+2}\,dx = \int_{\mathbb{R}^N} q^*(w_n^*)^{\sigma+2}\,dx \geqq \int_{\mathbb{R}^N} q|w_n|^{\sigma+2}\,dx.$$

Setting $v_n = t(w_n^*)w_n^*$, it follows from (M4) that $v_n = v_n^* \in M$ and that $t(w_n^*) \leqq \left\{\frac{\|w_n\|^2}{(\sigma+2)\phi(w_n)}\right\}^{1/\sigma} \leqq 1$.

Thus, by (M2),

$$2f(v_n) = \left(\frac{\sigma}{\sigma+2}\right)\|v_n\|^2 \leqq \frac{\sigma}{(\sigma+2)}\|w_n\|^2 = 2f(w_n)$$

and so $f(v_n) \to \inf_M f$.

Since $v_n = v_n^*$, we have that $0 \leqq v_n(x) \leqq |x|^{-N/2} N^{1/2} |v_n|_2$ a.e. on $\mathbb{R}^N$. But $\{v_n\}$ is bounded in $H = H^1(\mathbb{R}^N)$, by (M2). Using (3.7) and the compactness of the Sobolev embedding on bounded domains, it is now easy to show (as in Lemma 6.3 of [7]) that $\{v_n\}$ has a subsequence $\{v_{n_k}\}$ such that $v_{n_k} \rightharpoonup v_\infty$ weakly in H and $\phi(v_{n_k}) \to \phi(v_\infty)$.

This establishes the property (P) and so according to Proposition 2.1, $\exists u(\lambda) \in M$ such that $f(u(\lambda)) = \min_M f$. From the symmetrisation arguments used above, we can suppose that $u(\lambda) \geqq 0$ and that $u(\lambda) = u(\lambda)^*$.

To estimate $f(u(\lambda))$, let $z_\alpha(x) = e^{-\alpha|x|^2}$ for $x \in \mathbb{R}^N$ and $\alpha > 0$. By calculations similar to those in Theorem 5.4 of [7], we find that

$$\|z_\alpha\|^2 \leqq C\{|\lambda|+\alpha\}\alpha^{-N/2} \quad \text{and} \quad \phi(z_\alpha) \geqq D\alpha^{-N/2},$$

where C and D are positive constants depending only on N and σ.

By (M4), $t(z_\alpha)z_\alpha \in M$ and, by (M2),

$$2f(t(z_\alpha)z_\alpha) = \sigma t(z_\alpha)^{\sigma+2}\phi(z_\alpha) = \frac{\sigma\|z_\alpha\|^{2(1+2/\sigma)}}{(\sigma+2)\{(\sigma+2)\phi(z_\alpha)\}^{2/\sigma}}$$
$$\leqq C\{|\lambda|+\alpha\}^{1+2/\sigma}\alpha^{-N/2},$$

where C is a positive constant depending only on N and σ. Putting $\alpha = |\lambda|$, we obtain,

$$f(u(\lambda)) = \min_M f \leqq K|\lambda|^{1+2/\sigma-N/2} \tag{3.8}$$

and using (M2), we have that

$$|\nabla u(\lambda)|_2^2 \leqq 2\left(\frac{\sigma+2}{\sigma}\right) f(u(\lambda)), \tag{3.9}$$

$$|u(\lambda)|_2^2 \leqq 2\left(\frac{\sigma+2}{\sigma}\right) |\lambda|^{-1} f(u(\lambda)), \tag{3.10}$$

and

$$|u(\lambda)|_{\sigma+2}^{\sigma+2} \leqq \left(\frac{\sigma+2}{A}\right) \phi(u(\lambda)) \leqq 2\left(\frac{\sigma+2}{\sigma A}\right) f(u(\lambda)). \tag{3.11}$$

The inequalities (3.4)–(3.6) now follow from (3.3) and (3.8)–(3.11) by L^p interpolation using Hölder's inequality.

4. Steiner Symmetry on a Strip

Let $\Omega = \mathbb{R} \times (-\pi/2, \pi/2)$. A continuous function $u : \Omega \to [0, \infty)$ is said to be doubly Steiner symmetric on Ω, if, for all $x_1 \in \mathbb{R}$ and $x_2 \in (-\pi/2, \pi/2)$,

$$u(x_1, x_2) = u(-x_1, x_2) = u(x_1, -x_2),$$

$$u(x_1, \cdot) \quad \text{is non-increasing on} \quad [0, \pi/2),$$

and

$$u(\cdot, x_2) \quad \text{is non-increasing on} \quad [0, \infty).$$

In this section we consider the nonlinear eigenvalue problem,

$$\begin{aligned} -\Delta u(x) - q(x)|u(x)|^\sigma u(x) &= \lambda u(x) \quad \text{for} \quad x \in \Omega \\ u(x) &= 0 \qquad \text{for} \quad x \in \partial\Omega, \end{aligned} \tag{4.1}$$

where

$$\text{(C)} \quad \begin{cases} \sigma > 0, \\ q : \Omega \to (0, \infty) \text{ is continuous, doubly Steiner symmetric} \\ \text{on } \Omega \text{ and } q(x) \geqq A > 0 \quad \forall x \in \Omega. \end{cases}$$

A non-trivial weak solution of (4.1) is a pair (λ, u) such that

$$\lambda \in \mathbb{R}, \qquad u \in H_0^1(\Omega) \setminus \{0\}$$

and

$$\int_\Omega \{\nabla u \cdot \nabla v - q|u|^\sigma uv - \lambda uv\}\, dx = 0 \quad \forall v \in H_0^1(\Omega) \tag{4.2}$$

Remarks. 1. Recall that $H_0^1(\Omega) \subset L^p(\Omega)$ for $p \in [2, \infty)$ and that

$$|u|_2 \leqq |\nabla u|_2 \quad \forall u \in H_0^1(\Omega). \tag{4.3}$$

2. As is shown in [3], the L^p regularity theory of elliptic equations implies that if (λ, u) is a weak solution of (4.1) then u can be identified with a function which is twice continuously differentiable on Ω and continuous on $\bar{\Omega}$. Furthermore, the following L^∞ a-priori bound is established in [3]. Given $\xi > 0$ $\exists K(\xi) > 0$ such that

$$|u_\infty| \leqq K(\xi) \tag{4.4}$$

whenever (λ, u) is a weak solution of (4.1) such that $\lambda \in (1-\xi, 1)$ and u is positive and doubly Steiner symmetric on Ω.

3. In general, (4.1) has no solutions for $\lambda > 1$ [3].

Concerning the existence of solutions for $\lambda < 1$ and their behaviour as $\lambda \to 1$, we have the following result.

Theorem 4.1. *Under the hypotheses* (C), $\forall \lambda < 1 \exists$ *a non-trivial weak solution* $(\lambda, u(\lambda))$ *of* (4.1) *such that* $u(\lambda)$ *is positive and doubly Steiner symmetric on* Ω. *Furthermore,* $\exists K > 0$ *such that, for* $\lambda \in [0, 1)$,

$$|u(\lambda)|_p \leqq K(1-\lambda)^{\frac{1}{\sigma} - \frac{1}{2p}} \quad \text{if} \quad 2 \leqq p \leqq \sigma + 2, \tag{4.5}$$

$$\leqq K(1-\lambda)^{\frac{2}{p}\left\{\frac{1}{\sigma} + \frac{1}{4}\right\}} \quad \text{if} \quad p \geqq \sigma + 2, \tag{4.6}$$

$$|\nabla u(\lambda)|_2 \leqq K(1-\lambda)^{\frac{1}{\sigma} - \frac{1}{4}}. \tag{4.7}$$

Proof. Since the proof is very similar to that of Theorem 3.1, we only indicate the main points.

Fix $\lambda < 1$ and set $\|u\| = \{|\nabla u|_2^2 - \lambda |u|_2^2\}^{1/2}$ for $u \in H_0^1(\Omega)$. By (4.3),

$$(1-\lambda)|u|_2^2 \leqq (1-\lambda)|\nabla u|_2^2 \leqq \|u\|^2 \tag{4.8}$$

and $\|\cdot\|$ is equivalent to the usual norm on $H_0^1(\Omega)$. We shall apply Proposition 2.1 with $H = H_0^1(\Omega)$ and $\phi(u) = \dfrac{1}{(\sigma+2)} \int_\Omega q|u|^{\sigma+2} dx$. Using the multiplicative Sobolev inequality, we see that the condition (A) is satisfied. Two verify the property (P), consider a sequence $\{w_n\} \subset M$ such that $f(w_n) \to \inf_M f$. By double Steiner symmetrisation [10, 13, 14, 16] and rescaling of $|w_n|$, we can suppose that w_n is positive and doubly Steiner symmetric on Ω. By the trace theorem for $H_0^1(\Omega)$,

$$\int_{-\infty}^{\infty} w_n(t, y)^2 dt \leqq C\|w_n\|^2 \quad \text{for a.e.} \quad y \in (-\pi/2, \pi/2).$$

But

$$\int_{-\infty}^{\infty} w_n(t, y)^2 dt \geqq 2\int_0^z w_n(t, y)^2 dt \geqq 2z w_n(z, y)^2 \quad \text{for} \quad z \geqq 0,$$

and so $0 \leqq w_n(z, y) \leqq K z^{-1/2} \|w_n\|$.

As in Sect. 3, it is now easy to show that $\{w_n\}$ has a subsequence satisfying the property (P). According to Proposition 2.1, $\exists u(\lambda) \in M$ such that $f(u(\lambda)) = \min_M f$ where $u(\lambda)$ is positive and doubly Steiner symmetric on Ω.

To estimate $f(u(\lambda))$, let $z_\alpha(x_1, x_2) = e^{-\alpha x_1^2} \cos x_2$ for $(x_1, x_2) \in \Omega$ and $\alpha > 0$. By (M4), $t(z_\alpha) z_\alpha \in M$ and

$$2f(t(z_\alpha)z_\alpha) = \frac{\sigma \|z_\alpha\|^{2(1+2/\sigma)}}{(\sigma+2)\{(\sigma+2)\phi(z_\alpha)\}^{2/\sigma}} \leqq C\{1-\lambda+\alpha\}^{1+\frac{2}{\sigma}} \alpha^{-\frac{1}{2}}.$$

Putting $\alpha = 1 - \lambda$, we obtain

$$f(u(\lambda)) = \min_M f \leqq K(1-\lambda)^{\frac{1}{2}+\frac{2}{\sigma}}.$$

Using (M2), this yields estimates for $|u(\lambda)|_2$, $|u(\lambda)|_{\sigma+2}$ and $|\nabla u(\lambda)|_2$. The proof is completed by interpolation.

References

1. Toland, J. F.: Global bifurcation for Neumann problems without eigenvalues. J. Differential Equations **44**, 82–110 (1982)
2. Toland, J. F.: Positive solutions of nonlinear elliptic equations, existence and non-existence of solutions with radial symmetry in $L_p(\mathbb{R}^N)$ (preprint)
3. Amick, C. J., Toland, J. F.: Nonlinear elliptic eigenvalue problems on an infinite strip, global theory of bifurcation and asymptotic bifurcation (preprint)
4. Nehari, Z.: On a class of nonlinear integral equations. Math. Z. **72**, 175–183 (1959)
5. Agmon, S.: The L approach to the Dirichlet problem. Ann. Scuola Norm. Sup. Pisa **13**, 405–448 (1959)
6. Gidas, B., Spruck, J.: A priori bounds for positive solutions of nonlinear elliptic equations. Comm. Pure Appl. Math. **34**, 525–598 (1981)
7. Stuart, C. A.: Bifurcation for Dirichlet problems without eigenvalues. Proc. London Math. Soc. **45**, 169–192 (1982)
8. Stuart, C. A.: Bifurcation from the essential spectrum. Proc. of Equadiff 82. In: Lecture Notes in Mathematics. Berlin, Heidelberg, New York: Springer (to appear)
9. Berestycki, H., Lions, P. L.: Nonlinear scalar field equations I (Existence of a ground state) and II (Existence of infinitely many solutions). Arch. Rational Mech. Anal. (to appear)
10. Pólya, G., Szegö, G.: Isoperimetric inequalities in mathematical physics. Princeton: Princeton University Press 1951
11. Bandle, C.: Isoperimetric inequalities and applications. London: Pitman 1980
12. Berger, M.: Nonlinearity and functional analysis. New York: Academic Press 1977
13. Lions, P. L.: Minimisation problems in $L^1(\mathbb{R}^3)$. J. Functional Analysis **41**, 236–275 (1981)
14. Brascamp, H. J., Lieb, E. H., Luttinger, J. M.: A general rearrangement inequality for multiple integrals. J. Functional Analysis **17**, 227–237 (1974)
15. Lieb, E. H.: Existence and uniqueness of minimising solutions of Choquard's nonlinear equation. Stud. Appl. Math. **57**, 93–105 (1977)
16. Berger, M. S., Fraenkel, L. E.: A global theory of steady vortex rings in an ideal fluid. Acta Math. **132**, 13–51 (1974)
17. Coffman, C.V.: A minimum-maximum principle for a class of nonlinear integral equations. J. Analyse Math. **22**, 391–419 (1969)
18. Hempel, J.A.: Superlinear variational boundary value problems and non-uniqueness. Ph.D. thesis, University of New England, 1970
19. Brown, K.J.: Some operator equations with an infinite number of solutions. Quart. J. Math. Oxford **25**, 195–212 (1974)
20. Lions, P.L.: Principe de concentration-compacité en calcul des variations. C.R. Acad. Sci. Paris **294**, 261–264 (1982)
21. Lions, P.L.: Symetrie et compacité dans les espaces de Sobolev. J. Functional Analysis (to appear)
22. Bona, J.L., Bose, D.K., Turner, R.E.L.: Finite amplitude steady waves in stratified fluids. M.R.C. report No. 2401, July 1982

Received October 26, 1982

Note added in proof. Since completing the first version of this paper a number of additional references have come to my attention and deserve comment.

First of all, since Nehari's work, the trick of working on the natural constraint M has been used by several authors, notably Coffman [17] and Hempel [18], in dealing with elliptic problems on

bounded domains. Brown [19] used the same method for similar problems on unbounded domains, but the potential q in (1.1) is required to satisfy the condition $q(x)\to 0$ as $|x|\to\infty$, thus rendering the non-linear term compact. In the problems treated in Sects. 3 and 4, we are interested in cases where $q(x)\geqq A>0$ for all x and so this compactness is not available. The existence of a precompact minimising sequence is established using symmetrisation. In this connection two recent publications of P. L. Lions must be mentioned. The paper [20] deals with the relationship between lack of compactness and symmetry. In [21], a number of decay estimates are obtained for functions enjoying various kinds of symmetry. These results imply compactness in the corresponding symmetry class and so the method of symmetrisation employed in Sects. 3 and 4 can be used in these cases.

Let us stress that the nonlinearity is taken to be homogeneous only to simplify the exposition. More general cases, such as a sum of homogeneous terms, can be treated as in [4, 17–19]. In Sect. 4, we have assumed double Steiner symmetry only to treat the same problem as Amick and Toland [3]. The method employed only really requires Steiner symmetry with respect to the line $x=0$. A problem of this type has been discussed recently by Bona et al. [22].

Math. Ann. 263, 61–73 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Global Several-Parameter Bifurcation and Continuation Theorems: a Unified Approach Via Complementing Maps

P. M. Fitzpatrick[1], I. Massabò[2], and J. Pejsachowicz[2]

1 Department of Mathematics, University of Maryland, College Park, MD 20742, USA
2 Dipartimento di Matematica, Università della Calabria, Arcavacata di Rende, Calabria, Italy

1. Introduction

In this paper we will prove a theorem whose sole assumption is that the Leray-Schauder degree of a certain compact vector field is nonzero, and whose conclusions imply several-parameter, multidimensional global refinements of the implicit function theorem, the Leray-Schauder continuation principle [3, 11], and the global bifurcation theorem of Rabinowitz [13].

Let X be a Banach space, m be a positive integer and $\mathcal{O}$ be an open subset of $\mathbb{R}^m \times X$. We will consider local and global properties of the set of solution of equations of the form

$$f(\lambda, x) \equiv x - F(\lambda, x) = 0, \quad (\lambda, x) \in \mathcal{O}, \tag{1.1}$$

where $F : \mathcal{O} \to X$ is a compact mapping. If one identifies X with $\{0\} \times X \subseteq \mathbb{R}^m \times X$, one can view f as a map of $\mathcal{O} \subseteq \mathbb{R}^m \times X$ into $\mathbb{R}^m \times X$, and it is a compact perturbation of the identity on $\mathbb{R}^m \times X$. However, the range of f is contained in a subspace of $\mathbb{R}^m \times X$ of codimension m, so that its Leray-Schauder degree, when it is defined, is zero. In a certain sense, (1.1) is underdetermined by m dimensions. As such, it is not unreasonable to expect that under stable conditions under which one has existence results for (1.1) one will also have "m-multiplicity" results for the set of solutions. Our principle aim is to give a precise realization to this expectation.

A map $g : \mathcal{O} \to \mathbb{R}^m$, which maps bounded subsets of $\mathcal{O}$ into bounded sets, will be called a *complement* for $f : \mathcal{O} \to X$ provided that the Leray-Schauder degree, $\deg((g,f), \mathcal{O}, 0)$, is defined and non-zero: the map $(g,f) : \mathcal{O} \to \mathbb{R}^m \times X$ is defined by $(g, f)((\lambda, x)) = (g(\lambda, x), f(\lambda, x))$. Since (g, f) is a compact perturbation of the identity on $\mathbb{R}^m \times X$, the above degree is defined precisely when $\{(\lambda,x) \in \mathcal{O} | (g,f)((\lambda, x)) = 0\}$ is compact. We emphasize that, at this point, we are making no assertion about the compact extendability of f to $\bar{\mathcal{O}}$, and in the case such an extension exists, no assertion about the zeros of (g,f) on $\partial\mathcal{O}$. In [6] we introduced the notion of a complementing mapping under the additional assumptions that g and f were both defined on $\bar{\mathcal{O}}$ and $(g,f)^{-1}(0) \cap \partial\mathcal{O} = \emptyset$. While, as we demonstrated in [6], these additional assumptions are crucial in obtaining sharp continuation results for

(1.1), in order to obtain a general result which is applicable to both continuation and bifurcation problems it is necessary to consider the more general notion of complementing mapping which we have defined above.

When the mapping $f: \mathcal{O} \to X$ has a complement, $g: \mathcal{O} \to \mathbb{R}^m$, one can make rather precise global m-dimensional assertions concerning the manner in which $f^{-1}(0)$ emanates from $(g,f)^{-1}(0)$: this is the point of Theorem 2.1. Moreover, and, what is, of course, of equal importance, there are a number of interesting situations when one can show that a mapping has a complement.

If one chooses $\lambda_0 \in \mathbb{R}^m$, and, letting $\mathcal{O}_{\lambda_0} = \{x \in X | (\lambda_0, x) \in \mathcal{O}\}$, defines $f_{\lambda_0}: \mathcal{O}_{\lambda_0} \to X$ by $f_{\lambda_0}(x) = f(\lambda_0, x)$, then $f: \mathcal{O} \to X$ is complemented by $g: \mathcal{O} \to \mathbb{R}^m$ defined by $g(\lambda, x) = \lambda - \lambda_0$ if and only if $\deg(f_{\lambda_0}, \mathcal{O}_{\lambda_0}, 0) \neq 0$ [6]. From this observation we are able to use Theorem 2.1 to deduce multidimensional global versions of the implicit function theorem and the Leray-Schauder continuation principle: these are Theorems 2.2 and 2.3, respectively.

In order to obtain our global bifurcation result, Theorem 2.4, from Theorem 2.1, we have to use a somewhat different type of complementing map (see Proposition 4.1).

In [6] we have given, under the assumptions that $X = \mathbb{R}^k$, that $\mathcal{O}$ is open and bounded with $\partial\mathcal{O}$ smooth, and that $f: \bar{\mathcal{O}} \to \mathbb{R}^k$ is smooth with 0 a regular value both of $f: \mathcal{O} \to \mathbb{R}^k$ and of $f: \partial\mathcal{O} \to \mathbb{R}^k$, a necessary and sufficient condition for f to have a complement: the condition is that $f^{-1}(0) \cap \mathcal{O}$ and $f^{-1}(0) \cap \partial\mathcal{O}$ be nonempty.

In [2], Alexander and Yorke proved a global version of the implicit function theorem, while in [1] Alexander and Antman proved a multidimensional version of the global bifurcation result of Rabinowitz [13]. The conclusions of [1] and [2] are stated in the form of existence results for essential mappings, and are not directly comparable to the two consequences of Theorem 2.1 to which we have just referred. On the other hand, there is a substantial difference between the topological methods uses in [1] and [2], and, in our opinion, our present techniques are simpler and more elementary. In [1] and [2], $\mathcal{O}$ is required to be homeomorphic of $\mathbb{R}^m \times X$, while no such requirement is needed here. Also, the global bifurcation conclusions which we give are more precise (see Remark 2.1).

The global continuation result we obtain, Theorem 2.3, is a refinement of the result obtained by the authors in [6], the results of [6] being an improvement of the continuation results of Massabò and Pejsachowicz [10].

After this introduction, the paper has three further sections. In Sect. 2 we state, and discuss, our results. Section 3 is devoted to a proof of our main result, Theorem 2.1. In the last section we prove some of the consequences of this theorem.

An announcement of our present results appeared in [5].

2.

Let A be a topological space and k be a non-negative integer. Then A is said to have (Lebesgue covering) dimension equal to k provided that k is the smallest integer with the property that whenever $\mathcal{F}$ is an open cover of A there exists a refinement, $\mathcal{F}'$, of $\mathcal{F}$, which also covers A, and no more than any $k+1$ members of $\mathcal{F}'$ have non-empty intersection. If A fails to have the above refinement property

for each integer k, then A is said to have infinite dimension. If $a \in A$, A will be said to have dimension at least k at a if each neighborhood, in A, of a has dimension at least k.

We will only use the notion of dimension when A is a separable metric space, and consequently, here, the concept of dimension defined above coincides with the notion of inductive dimension (see Hurewicz and Wallman [8]).

In the absence of any manifold structure on $f^{-1}(0)$, the notion of topological dimension is the natural way in which to describe its size.

When (C, D) is a pair of normal topological spaces with $D \subset C$, by $\check{H}^m(C, D)$ we will denote the m-th Čech cohomology group with integral coefficients of the pair. Moreover, a map $h:(C, D) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ will be said to be cohomologically nontrivial provided that the induced homomorphism

$$g^*: \check{H}^m(\mathbb{R}^m, \mathbb{R}^m - 0) \to \check{H}^m(C, D)$$

is non-trivial.

Recall that D is said to be cobounded in C if $\overline{C/D}$ is compact.

We refer the reader to Dold [4] and Spanier [14] for some of the results in homology and cohomology theory which we will use, and to Lloyd [9] for results concerning the Leray-Schauder degree which we will need.

Our basic result is the following.

Theorem 2.1. *Let m, X, $\mathcal{O}$, and f be as in the introduction, and assume that there exists a complement $g: \mathcal{O} \to \mathbb{R}^m$ for $f: \mathcal{O} \to X$. Then there exists a closed connected subset, C, of $f^{-1}(0)$ whose dimension at each point is at least m, and*

for each cobounded subset, D, of C with
$D \cap g^{-1}(0) = \emptyset$, $g:(C, D) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ (*)
is cohomologically nontrivial.

In particular,

$$C \cap g^{-1}(0) \neq \emptyset, \tag{2.1}$$

and

$$\text{either } C \text{ is unbounded or } \bar{C} \cap \partial\mathcal{O} \neq \emptyset. \tag{2.2}$$

Remark 2.1. Observe that (2.1) and (2.2) follow immediately from (*) by taking $D = C$ and $D = \emptyset$, respectively.

Remark 2.2. Another easy consequence of (*) is the following: if $K \subset C$ is compact, with $g^{-1}(0) \cap C \subseteq K$, then, for $1 \leqq i \leqq m$, g_i must change sign on $(C \setminus K) \cap \{(\lambda, x) | g_j(\lambda, x) = 0, j \neq i, 1 \leqq j \leqq m\}$. This will be crucial in proving our bifurcation results.

When f is continuously Fréchet differentiable, $(\lambda_0, x_0) \in \mathcal{O} \cap f^{-1}(0)$ and $\dfrac{\partial f}{\partial x}(\lambda_0, x_0)$ is a bijection, the classical implicit function theorem implies that in a neighborhood of (λ_0, x_0), $f^{-1}(0)$ is an m-manifold. On the other hand, in this situation, if one lets V be a small neighborhood of (λ_0, x_0) then $f: V \to X$ is complemented by $g: V \to \mathbb{R}^m$ defined by $g(\lambda, x) = \lambda - \lambda_0$: this follows from the well-known linearization property of the Leray-Schauder degree. Hence the following theorem, which is a straight forward consequence of Theorem 2.1, contains, as a special case, a global version of the implicit function theorem.

Theorem 2.2. *Let $m, X, \mathcal{O}$, and f be as in the introduction. Assume that $g: \mathcal{O} \to \mathbb{R}^m$ is continuous, and maps bounded sets into bounded sets. Suppose $V \subseteq \mathcal{O}$ is open and $g: V \to \mathbb{R}^m$ complements $f: V \to X$. Let $K = ((g, f)|_V)^{-1}(0)$. Then there exists a closed connected subset, C, of $f^{-1}(0)$ whose dimension at each point is at least m, $C \cap K \neq \emptyset$, and at least one of the following properties holds:*

$$\begin{aligned} &\text{(a) } C \text{ is unbounded.} \\ &\text{(b) } \bar{C} \cap \partial\mathcal{O} \neq \emptyset. \\ &\text{(c) } \bar{C} \cap \{(g,f)^{-1}(0) \backslash K\} \neq \emptyset. \end{aligned} \tag{2.3}$$

The above theorem is an immediate consequence of Theorem 2.1. Indeed, let $\mathcal{U} = \mathcal{O} \backslash \{(g,f)^{-1}(0) \backslash K\}$ and observe that $g: \mathcal{U} \to \mathbb{R}^m$ is a complement for $f: \mathcal{U} \to X$. So we apply Theorem 2.1, with $\mathcal{U}$ playing the role of $\mathcal{O}$.

When one applies Theorem 2.2 in the situation where $\dfrac{\partial f}{\partial x}(\lambda_0, x_0)$ is invertible and V is a small neighborhood of (λ_0, x_0) the conclusion is that the m-manifold, $f^{-1}(0) \cap V$, may be continued globally into a connected set, C, with $C \subseteq f^{-1}(0)$ and C has dimension at least m at each point. Moreover, either C is unbounded, $\bar{C} \cap \partial\mathcal{O} \neq \emptyset$ or C curls back to $\mathcal{O}_{\lambda_0}$ at some point distinct from x_0.

In the situation where f and g are defined on $\bar{\mathcal{O}}$ with $(g,f)^{-1}(0) \cap \partial\mathcal{O} = \emptyset$ one can considerably sharpen the conclusion of Theorem 2.1.

Recall that if A is a locally compact space and $h: A \to \mathbb{R}^m$ is a proper continuous map, then h is called *essential among proper maps* provided that $\hat{h}: \hat{A} \to S^m$ is homotopically nontrivial, where $\hat{h}$ is the extension of h to $\hat{A}$, $\hat{A}$ being the one-point compactification of A. Clearly, if h is essential among proper maps then $h(A) = \mathbb{R}^m$.

Our second consequence of Theorem 2.1 is

Theorem 2.3. *Let f and g satisfy the assumptions of Theorem* 2.1, *in addition to which we assume f and g are defined on $\bar{\mathcal{O}}$ with $(g,f)^{-1}(0) \cap \partial\mathcal{O} = \emptyset$. Then there exists a closed connected subset, C, of $f^{-1}(0) \cap \mathcal{O}$ such that $C \cap g^{-1}(0) \neq \emptyset$ and C has dimension at least m at each point. Moreover, C satisfies at least one of the following properties:*

$$\begin{aligned} &\text{(a) } C \text{ is unbounded} \\ &\text{(b) } \dim(\bar{C} \cap \partial\mathcal{O}) \geqq m-1, \text{ when } m > 1, \text{ while } \bar{C} \cap \partial\mathcal{O} \\ &\quad \text{contains at least two points when } m = 1. \end{aligned} \tag{2.4}$$

If, in addition, $g: \bar{C} \to \mathbb{R}^m$ is proper, then C satisfies at least one of the following two properties:

$$\begin{aligned} &\text{(a) } \dim(\bar{C} \cap \partial\mathcal{O}) \geqq m-1 \\ &\text{(b) } g: C \to \mathbb{R}^m \text{ is essential among proper maps, and in particu-} \\ &\quad \text{lar, } g(C) = \mathbb{R}^m. \end{aligned} \tag{2.5}$$

As a particular case of the above we obtain

Corollary 2.1. *Let $m, X, \mathcal{O}$, and f be as in the introduction, in addition to which we assume F is defined on $\bar{\mathcal{O}}$. Suppose $\lambda_0 \in \mathcal{O}$ is such that $\deg(f_{\lambda_0}, \mathcal{O}_{\lambda_0}, 0) \neq 0$ and $0 \notin f_{\lambda_0}(\partial\mathcal{O}_{\lambda_0})$. Let the following a-priori bound hold: if $\{(\lambda_n, x_n)\} \subseteq \bar{\mathcal{O}} \cap f^{-1}(0)$ and $\{\lambda_n\}$*

is bounded, then $\{x_n\}$ *is bounded. Then there exists a closed connected subset,* C, *of* $f^{-1}(0)$, *whose dimension at each point is at least* m, $C \cap \mathcal{O}_{\lambda_0} \neq \emptyset$, *and at least one of the following two properties holds:*

$$\begin{aligned} &\text{(a) } \dim(\bar{C} \cap \partial\mathcal{O}) \geqq m-1, \text{ when } m>1, \text{ while } \bar{C} \cap \partial\mathcal{O} \text{ contains at} \\ &\quad \text{least two points when } m=1. \\ &\text{(b) for each } \lambda \in \mathbb{R}^m \text{ there exists some } x \in X \text{ with } (\lambda, x) \in C. \end{aligned} \tag{2.6}$$

In the case when $m=1$ the above corollary is a slight refinement of the classical Leray-Schauder continuation principle (see [3, 11]). When $m>1$, if one slices the parameter space into one-dimensional subspaces passing through λ_0 one could apply the one dimensional result to f, with its parameter restricted to this subspace. The above corollary gives a description of how these slices of solution mesh together.

An m-dimensional continuation principle similar to the above corollary was obtained by Massabò and Pejsachowicz in [10], and by the authors in [6].

The final consequence of Theorem 2.1 which we will discuss is a bifurcation result. Here we assume, in addition to our standing assumptions on f, that

$$f(\lambda, 0)=0 \quad \text{whenever} \quad (\lambda, 0) \in \mathcal{O},$$

and we refer to $\{\mathbb{R}^m \times \{0\}\} \cap \mathcal{O}$ as the trivial solutions. We are interested in discussing the manner in which the nontrivial solutions of (1.1) branch away from the trivial solutions.

Recall that if $\lambda \in \mathbb{R}^m$ is such that 0 is an isolated point of $f_\lambda^{-1}(0)$ then the Leray-Schauder index, $\operatorname{ind}(f_\lambda, 0)$, is well-defined. For a subset, Γ, of $\mathbb{R}^m$, we will identify Γ with $\Gamma \times \{0\} \subseteqq \mathbb{R}^m \times X$, and for a subset, D, of $\mathbb{R}^m \times X$, denote by D_Γ the set $\{(\lambda, x) | (\lambda, x) \in D, \lambda \in \Gamma\}$. Moreover, for our mapping $f : \mathcal{O} \to X$, the map f_Γ will denote the restriction of f to $\mathcal{O}_\Gamma$. A point $\lambda^* \in \Gamma$ will be called a *Γ bifurcation point of f if* there exists a sequence $\{(\lambda_k, x_k)\}$ converging to $(\lambda^*, 0)$, with $\{\lambda_k\} \subseteqq \Gamma$ and $\{x_k\} \subseteqq X \backslash \{0\}$, and $\{(\lambda_k, x_k)\} \subseteqq f^{-1}(0)$.

Theorem 2.4. *Suppose* m, X, $\mathcal{O}$, *and* f *are as in the introduction. Let* $\Gamma \subseteqq \mathcal{O} \cap \{\mathbb{R}^m \times \{0\}\}$ *be described by* $\Gamma = h^{-1}(0)$, *where* $h : \mathbb{R}^m \to \mathbb{R}^{m-1}$ *is continuously differentiable and has* 0 *as a regular value. Suppose* $\underline{\lambda}$ *and* $\bar{\lambda}$ *lie in the same component of* Γ, *neither* $\underline{\lambda}$ *nor* $\bar{\lambda}$ *are* Γ *bifurcation points of* f *and* $\operatorname{ind}(f_{\underline{\lambda}}, 0) \neq \operatorname{ind}(f_{\bar{\lambda}}, 0)$. *Then there exists a closed connected subset,* C, *of* $\{(\lambda, x) | (\lambda, x) \in \mathcal{O}, x \neq 0, f(\lambda, x)=0\}$, *whose dimension at each point is at least* m, *and* $\bar{C} \cap [\underline{\lambda}, \bar{\lambda}] \neq \emptyset$, *where* $[\underline{\lambda}, \bar{\lambda}]$ *denotes any segment of* Γ *between* $\underline{\lambda}$ *and* $\bar{\lambda}$. *One of the following three properties is also satisfied by* C:

$$\begin{aligned} &\text{(a) } C \text{ is unbounded,} \\ &\text{(b) } \bar{C} \cap \partial\mathcal{O} \neq \emptyset, \\ &\text{(c) } \bar{C} \cap \{\Gamma \backslash [\underline{\lambda}, \bar{\lambda}]\} \neq \emptyset. \end{aligned} \tag{2.7}$$

Remark 2.1. The conclusion of Theorem 2.4 does *not* hold when $\Gamma = \{\gamma(t) | (t \in \mathbb{R}\}$, where $\gamma : \mathbb{R} \to \mathbb{R}^m$ is an arbitrary continuously differentiable homeomorphism. For example, it does not hold if $m=1$, $\mathcal{O}=\mathbb{R} \times X$ and Γ is a bounded interval. Rather then state the most general, necessarily technical, hypotheses on Γ under which one can prove the bifurcation result, we have asserted that Γ has a trivial normal

bundle. There are Γ, not necessarily included under this hypothesis, for which the conclusion also holds. One such situation is when $\Gamma = \{\gamma(t) | (t \in \mathbb{R}\}$, where $\gamma : \mathbb{R} \to \mathbb{R}^m$ is a continuously differentiable homeomorphism such that $|\gamma(t)| \to +\infty$ as $|t| \to +\infty$. That the theorem holds in this situation follows from the proof of Theorem 2.4 which we will give, together with the observation that for each $r > 0$, the components of $\Gamma \cap B(0, r)$ have trivial normal bundles.

Remark 2.2. When one assumes simply that $\Gamma = \{\gamma(t) | t \in \mathbb{R}\}$, γ being a C^1 homeomorphism, it is possible to obtain a weaker conclusion: namely, if one replaces (2.7) (c) by

$$\text{(c')} \quad \bar{C} \cap \{\mathbb{R}^m \backslash [\underline{\lambda}, \bar{\lambda}]\} \neq \emptyset, \tag{2.7}$$

the theorem is true. However, this latter conclusion is somewhat unsatisfactory, since, for instance, C could be a small m-disc transversal to $(\underline{\lambda}, \bar{\lambda})$, and so the conclusion, in some sense, is not global. Again, an examination of our proof of Theorem 2.4 shows how to prove this latter result.

Remark 2.3. We observe that the section of C over Γ, C_Γ, need not be connected. However, it is not difficult to see from our proof of Theorem 2.1 that there is a connected component of C_Γ which satisfies one of the alternatives listed in (2.7).

Remark 2.4. In the case when $m = 1$, Theorem 2.4 is a slight refinement of the global bifurcation theorem of Rabinowitz [13]. An m-dimensional generalization of the Rabinowitz theorem has been recently obtained by Alexander and Antman [1]. Theorem 2.4 improves the corresponding result of Alexander and Antman [1]: our assumptions are somewhat weaker, and we discuss the intersection of the component, C, with $\Gamma \backslash [\underline{\lambda}, \bar{\lambda}]$.

Remark 2.5. The conclusions of our results are not topological, in the sense that C being unbounded is not a property preserved by homeomorphic change of variables. For this reason a certain amount of care has to be taken when making simplifying assumptions.

The notion of a complementing map, obviously, can be defined whenever (g, f) falls within a class of mappings for which there is defined a topological degree. Moreover, if this degree is sufficiently well developed, the results of this paper will carry over. Our results hold for substantially larger classes of mappings then we have considered here. We will return to this in a future paper in which we will also present applications of our present results in the study of nonlinear partial differential equations. Examples of nonlinear ordinary and partial differential equations which may be formulated naturally as an equation of type (1.1) may be found in [6, 7].

3.

We will prove Theorem 2.1. To do so, and also because it will be useful in other contexts, we will first prove a preliminary proposition.

We let $S = \{(\lambda, x) \in \mathcal{O} | f(\lambda, x) = 0\}$, and observe that S is locally compact since F is a compact mapping.

Proposition 3.1. *Under the assumptions of Theorem 2.1, S has the following property:*

if D is a cobounded subset of S such that $D \cap g^{-1}(0) = \emptyset$, *then* $g:(S, D) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ *is cohomologically nontrivial.*

Proof. We first observe that $\check{H}^m(S, D) = \varinjlim \{\check{H}^m(V, W) | (V, W) \in T\}$, where T consists of pairs (V, W) with V a neighborhood, in $\mathcal{O}$, of S, and W a neighborhood, in $\mathcal{O}$, of D, such that $W \subset V$. But those pairs with the additional property that $W \cap g^{-1}(0) = \emptyset$ are a cofinal subfamily of T.

Consequently, to prove the result it will suffice to show that if (V, W) is a pair of open subsets of $\mathcal{O}$, $S \subset V$, $D \subset W$, $W \subset V$ and $g^{-1}(0) \cap W = \emptyset$, then $g:(V, W) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ is cohomologically nontrivial. Let (V, W) be such a pair.

Since $S \backslash W$ is bounded we may choose an $r > 0$ such that $S \backslash W \subseteqq B \equiv \{(\lambda, x) \in \mathbb{R}^m \times X \mid \|(\lambda, x)\| < r\}$. Then we see that

$$(g, f)^{-1}(0) \subseteqq S \backslash W \subseteqq V \cap B,$$

and so, since $\mathbb{R}^m \times X$ is normal, we may choose an open subset, $\mathcal{U}$, of $V \cap B$, such that

$$(g, f)^{-1}(0) \subseteqq \mathcal{U} \subseteqq \bar{\mathcal{U}} \subseteqq V \cap B.$$

But then $(g, f)^{-1}(0) \cap \partial \mathcal{U} = \emptyset$. Since $\partial \mathcal{U}$ is bounded and closed, and since F is compact, we may select $\varepsilon > 0$ such that $\|(g, f)(\lambda, x)\| \geqq \varepsilon > 0$, whenever $(\lambda, x) \in \partial \mathcal{U}$.

Using the basic approximation property for compact mappings we may choose a finite-dimensional subspace, H, of X and a continuous mapping $F_\varepsilon : \bar{\mathcal{U}} \to H$ such that

$$\|F(\lambda, x) - F_\varepsilon(\lambda, x)\| \leqq \varepsilon/2, \quad \text{whenever} \quad (\lambda, x) \in \bar{\mathcal{U}}.$$

It follows from the excision and homotopy invariance properties of degree that

$$\deg((g, f), \mathcal{O}, 0) = \deg((g, f), \mathcal{U}, 0) = \deg((g, f_\varepsilon), \mathcal{U}, 0),$$

where $f_\varepsilon(\lambda, x) = x - F_\varepsilon(\lambda, x)$, for $(\lambda, x) \in \bar{\mathcal{U}}$.

Consequently, if we let $\mathcal{U}' = \mathcal{U} \cap \{\mathbb{R}^m \times H\}$, and use the assumption that g complements f on $\mathcal{O}$, together with the reduction property for Leray-Schauder degree, it follows that

$$\deg((g, f_\varepsilon)|_{\bar{\mathcal{U}}'}, \mathcal{U}', 0) \neq 0,$$

where now "deg" means Brouwer degree.

We now identify H with $\mathbb{R}^n$, for some positive integer n. Let $W' = W \cap \mathcal{U}'$ and let $Z = \{(\lambda, x) \in \mathcal{U}' | f_\varepsilon(\lambda, x) \neq 0\}$. Then we have the following maps of pairs;

$$f_\varepsilon : (\mathcal{U}', Z) \to (\mathbb{R}^n, \mathbb{R}^n - 0),$$

and

$$g : (\mathcal{U}', W') \to (\mathbb{R}^m, \mathbb{R}^m - 0).$$

Since $\mathcal{U}'$, W', and Z are open subsets of euclidean spaces, we can identify the Čech cohomology groups $\check{H}^i(\mathcal{U}', Z)$ and $\check{H}^i(\mathcal{U}', W')$, with the singular cohomology groups $H^i(\mathcal{U}', Z)$ and $H^i(\mathcal{U}', W')$ respectively, by means of a natural isomorphism. Let e_k be the canonical generator of the local cohomology group, $H^k(\mathbb{R}^k, \mathbb{R}^k - 0)$.

We shall compute the cup-product

$$\eta \equiv g^*(e_m) \cup f_\varepsilon^*(e_n) \in H^{m+n}(\mathscr{U}', W' \cup Z)$$

in terms of the homomorphism induced in cohomology by the following map of pairs:

$$h \equiv (g, f_\varepsilon) \equiv (g \times f_\varepsilon) \circ \Delta : (\mathscr{U}', W' \cup Z) \to (\mathbb{R}^{m+n}, \mathbb{R}^{m+n} - 0),$$

where Δ is the diagonal map.

Indeed, from the properties of the cup-product and cross-product, we obtain

$$\eta = \Delta^*(g^*(e_m) \times f^*(e_n)) = h^*(e_m \times e_n) = h^*(e_{m+n}).$$

On the other hand, from our assumptions, the set $K \equiv \mathscr{U}' \backslash \{W' \cup Z\}$ is a compact set containing $h^{-1}(0)$, so that if we let $O_K \in H_{m+n}(\mathscr{U}', \mathscr{U}' - K)$ be the fundamental class around K and O_{m+n} denotes the fundamental class around 0 in $\mathbb{R}^{m+n}$, it follows that

$$h_*(O_K) = \deg(h, \mathscr{U}', 0) \cdot O_{m+n} \quad \text{in} \quad H_{m+n}(\mathbb{R}^{m+n}, \mathbb{R}^{m+n} - 0),$$

(see Proposition 5.5 of [4]). H_{m+n} denoting the singular homology group, and h_* being the homeomorphism induced in homology.

Now we take the Kronecker product of $\eta \in H^{m+n}(\mathscr{U}', W' \cup Z) = H^{m+n}(\mathscr{U}', \mathscr{U}' \backslash K)$ with O_K. We obtain

$$\begin{aligned} \langle \eta, O_K \rangle &= \langle h^*(e_{m+n}), O_K \rangle \\ &= \langle e_{m+n}, h_*(O_K) \rangle = \deg(h, \mathscr{U}', 0), \end{aligned}$$

since $\langle e_{m+n}, O_{m+n} \rangle = 1$.

Thus $\eta \neq 0$, and so $g^*(e_m) \in H^m(\mathscr{U}', W') \cong \check{H}^m(\mathscr{U}', W')$ is nonzero.

Letting $i : (\mathscr{U}', W') \to (V, W)$ be the inclusion map, we have the following commutative diagram

$$\begin{array}{ccc} \check{H}^m(V, W) & \xrightarrow{g^*} & \check{H}^m(\mathbb{R}^m, \mathbb{R}^m - 0) \\ \downarrow{\scriptstyle i^*} & \nearrow{\scriptstyle g^*} & \\ \check{H}^m(\mathscr{U}', W') & & \end{array}$$

It follows that $g : (V, W) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ is cohomologically nontrivial, and so the proposition is proven. □

Proof of Theorem 2.1. Recall that the m-th Čech cohomology group of S, with compact supports, $\check{H}_c^m(S)$, is defined by

$$\check{H}_c^m(S) = \varinjlim \check{H}^m(S, D),$$

where D ranges over all cobounded subsets of S. Let $D_0 = S \backslash g^{-1}(0)$, so that from our basic assumption on $(g, f)^{-1}(0)$, D_0 is a cobounded subset of S.

If we let $j_{D_0} : \check{H}^m(S, D_0) \to \check{H}_c^m(S)$ be the natural map into the direct limit, and let $\xi \equiv j_{D_0}(g^*(e_m))$, Proposition 3.1 asserts that $\xi \neq 0$.

We can now invoke Proposition 4.3 of [6] in order to choose a closed connected subset, C, of S, which has dimension at each point at least m, and is such that the class $\xi|_C \in \check{H}_c^m(C)$ is nonzero.

Now let D be any cobounded subset of C such that $D \cap g^{-1}(0) = \emptyset$. From the following commutative diagram,

$$\begin{array}{ccccc} & \overset{g^*}{\nearrow} & \check{H}^m(S, D_0) & \xrightarrow{j_{D_0}} & \check{H}^m_c(S) \\ \check{H}^m(\mathbb{R}^m, \mathbb{R}^m - 0) & & \downarrow \text{inclusion}^* & & \downarrow \text{inclusion}^* \\ & \underset{g^*}{\searrow} & \check{H}^m(C, D) & \xrightarrow[j_D]{} & \check{H}^m_c(C), \end{array}$$

it follows that $g : (C, D) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$ is cohomologically nontrivial. □

4.

The Proof of Theorem 2.3. Theorem 2.1 allows us to choose a closed connected subset, C, od $f = \{(\lambda, x) | (\lambda, x) \in \mathcal{O}, f(\lambda, x) = 0\}$, whose dimension at each point is at least m, and which enjoys property (*).

Let $\dot{C} = \bar{C} \cap \partial\mathcal{O}$. We will show that $g^* : \check{H}^m(\mathbb{R}^m, \mathbb{R}^m - 0) \to \check{H}^m_c(\bar{C}, \dot{C})$ is nontrivial. To do so, since $\check{H}^m_c(\bar{C}, \dot{C}) = \varinjlim \check{H}^m(\bar{C}, W)$, where W ranges over neighborhoods, in $\bar{C}$, of $\dot{C}$, which are cobounded in $\bar{C}$ and such that $g^{-1}(0) \cap W = \emptyset$, it will suffice to show that for each such W

$$g : (\bar{C}, W) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$$

is cohomologically nontrivial. But $C \cap W$ is a cobounded subset of C, on which g does not vanish, and so

$$g : (C, C \cap W) \to (\mathbb{R}^m, \mathbb{R}^m - 0)$$

is cohomologically nontrivial. Consequently, from the commutativity of the following diagram,

$$\begin{array}{ccc} (C, C \cap W) & \xrightarrow{i} & (\bar{C}, W) \\ & {\scriptstyle g} \searrow \quad \swarrow {\scriptstyle g} & \\ & (\mathbb{R}^m, \mathbb{R}^m - 0) & \end{array}$$

it follows that $g^* : H^m(\mathbb{R}^m, \mathbb{R}^m - 0) \to \check{H}^m_c(\bar{C}, \dot{C})$ is nontrivial.

To verify the first assertion of the theorem we assume that C is bounded, and will prove the claimed properties concerning $\dot{C}$. Note that $\check{H}^m_c(\bar{C}, \dot{C}) = \check{H}^m(\bar{C}, \dot{C})$.

First consider the case when $m = 1$. We claim that $g : \dot{C} \to \mathbb{R} - 0$ is essential; i.e. it assumes both positive and negative values. Indeed, if this were not so, then $g : (\bar{C}, \dot{C}) \to (\mathbb{R}, \mathbb{R} - 0)$ could be deformed, as a map of pairs, via a linear homotopy, to a constant map, in contradiction to its cohomological nontriviality. Thus $g : \dot{C} \to \mathbb{R} - 0$ is essential, and, in particular, $\dot{C}$ has at least two points.

Now consider the case when $m > 1$. We have the following commutative diagram, with the bottom map being an isomorphism:

$$\begin{array}{ccc} \check{H}^{m-1}(\dot{C}) & \longrightarrow & \check{H}^m(\bar{C}, \dot{C}) \\ \uparrow {\scriptstyle g^*} & & \uparrow {\scriptstyle g^*} \\ \check{H}^{m-1}(\mathbb{R}^m - 0) & \xrightarrow{\sim} & \check{H}^m(\mathbb{R}^m, \mathbb{R}^m - 0). \end{array}$$

Consequently, $\check{H}^{m-1}(\dot{C})$ is nontrivial, from which it follows that $\dim(\dot{C}) \geqq m - 1$. (Note also that $g : \dot{C} \to \mathbb{R}^m - 0$ is essential.)

To complete the proof we will assume $g:\bar{C}\to\mathbb{R}^m$ is proper, and that $\dim(\dot{C})<m-1$, and conclude that $g:\bar{C}\to\mathbb{R}^m$ is essential among proper maps.

Since $g:\bar{C}\to\mathbb{R}^m$ is proper it induces a homomorphism $g^*:\check{H}_c^m(\mathbb{R}^m)\to\check{H}_c^m(\bar{C})$. Thus we have the following commutative diagram, where the rows are exact:

$$\begin{array}{ccccccc}
\check{H}_c^{m-1}(\dot{C}) & \longrightarrow & \check{H}_c^m(\bar{C},\dot{C}) & \longrightarrow & \check{H}_c^m(\bar{C}) & \longrightarrow & \check{H}_c^m(\dot{C}) \\
 & & \uparrow{\scriptstyle g^*} & & \uparrow{\scriptstyle g^*} & & \\
0 & \longrightarrow & \check{H}^m(\mathbb{R}^m,\mathbb{R}^m-0) & \xrightarrow{\sim} & \check{H}_c^m(\mathbb{R}^m) & \longrightarrow & 0.
\end{array}$$

On the other hand, since $\dim(\dot{C})<m-1$ we must have $\check{H}_c^{m-1}(\dot{C})=\check{H}_c^m(\dot{C})=0$ (see [8, p. 152]), so that the top center map in the above diagram is an isomorphism.

It follows immediately that $g^*:\check{H}_c^m(\mathbb{R}^m)\to\check{H}_c^m(\bar{C})$ is nontrivial. As a consequence, $g:\bar{C}\to\mathbb{R}^m$ is essential among proper maps, and, in particular, $g(\bar{C})=\mathbb{R}^m$. □

The Proof of Theorem 2.4. The heart of the matter lies in the assertion that there exists a continuous mapping $g_1:\mathbb{R}^m\times X\to\mathbb{R}$ such that the following hold:

$$\begin{aligned}&\text{if } g:\mathbb{R}^m\times X\to\mathbb{R}^m \text{ is defined by } g(\lambda,x)=(g_1(\lambda,x),\ h_1(\lambda),\ \ldots,\\ &h_{m-1}(\lambda)),\text{ then } g \text{ complements } f \text{ on } \mathscr{U}\equiv\mathscr{O}\backslash\{\mathbb{R}^m\times\{0\}\}:\end{aligned}\tag{4.1}$$

there exist η,β in $\mathbb{R}$, with $0<\eta<\beta$, and a Γ-neighborhood, Λ, of $[\underline{\lambda},\bar{\lambda}]$, for which

$$g_1(\lambda,x)<0, \text{ when both } \lambda\in[\underline{\lambda},\bar{\lambda}] \text{ and } 0<\|x\|\leqq\eta,\tag{4.2}$$

$$\begin{aligned}&g_1(\lambda,x)>0, \text{ when both } \lambda\in\Gamma\backslash\Lambda \text{ and } x\neq 0 \text{ or when both } \lambda\in\Lambda\\ &\text{and } \|x\|\geqq\beta,\end{aligned}\tag{4.3}$$

and

$$f(\lambda,x)\neq 0, \text{ when both } \lambda\in\Lambda\backslash[\underline{\lambda},\bar{\lambda}] \text{ and } 0<\|x\|\leqq\beta.\tag{4.4}$$

Indeed, suppose, for the moment, that a mapping g_1 having the above three properties exists. Then g complements f on $\mathscr{U}$. Hence we may apply Theorem 2.1, with $\mathscr{U}$ playing the role of $\mathscr{O}$, to choose a closed connected subset, C, of $f^{-1}(0)\cap\mathscr{U}$, whose dimension at each point is at least m, and such that property (*) holds. In view of Remark 3.1 and the fact that $\Gamma=\{(\lambda,0)|g_2(\lambda)=\ldots=g_m(\lambda)=0\}$ it follows that

$$\begin{aligned}&\text{whenever } E \text{ is a compact subset of } C, \text{ with } g^{-1}(0)\cap C\subseteqq E,\\ &\text{then } g_1 \text{ changes sign on } (C\backslash E)_\Gamma.\end{aligned}\tag{4.5}$$

Observe that (4.2)–(4.4) imply that $C\cap g^{-1}(0)\subseteqq\{(\lambda,x)|\lambda\in[\underline{\lambda},\bar{\lambda}],\ \eta<\|x\|\leqq\beta\}$.

We first prove that $\bar{C}\cap[\underline{\lambda},\bar{\lambda}]\neq\emptyset$. To do so we suppose the contrary. Let $E=\{(\lambda,x)|(\lambda,x)\in C_\Gamma,\ \lambda\in[\underline{\lambda},\bar{\lambda}],\ 0<\|x\|\leqq\beta\}$. Since C is closed in $\mathscr{U}$ and $\bar{C}\cap[\underline{\lambda},\bar{\lambda}]=\emptyset$ it follows that E is compact. As noted above, $C\cap g^{-1}(0)\subseteqq E$. Thus (4.5) implies that g_1 changes sign on $(C\backslash E)_\Gamma$. Since from (4.3) and (4.4) it follows that g_1 is positive on $(C\backslash E)_\Gamma$, we obtain a contradiction. Thus $\bar{C}\cap[\underline{\lambda},\bar{\lambda}]\neq\emptyset$.

It remains to prove the global properties of C. In order to do so, we assume that none of the alternatives listed in (2.7) hold. Now let $E=\{(\lambda,x)|(\lambda,x)\in C_\Gamma,\ \|x\|\geqq\eta \text{ if } \lambda\in[\underline{\lambda},\bar{\lambda}]\}$. Since (2.7) fails to hold, and Γ is closed, E is compact. Also, $C\cap g^{-1}(0)\subseteqq E$. Thus g_1 changes sign on $(C\backslash E)_\Gamma$, in contradiction to (4.2).

Thus, at least one of the properties in (2.7) holds and the theorem is proven, once we have verified the existence of the above g_1.

We first prove the existence of the above g_1 when Γ has a simple form. Specifically, we prove the following.

Proposition 4.1. *Suppose, in addition to the assumptions of Theorem 2.4, that* $\Gamma=\{\lambda|\lambda=(\lambda_1,0,\ldots,0),\ (\lambda,0)\in\mathcal{O}\}$, *that* $\underline{\lambda}=(-1,0,\ldots,0)$ *and* $\bar{\lambda}=(1,0,\ldots,0)$. *For* $r>0$, *let* $\varphi:\mathbb{R}\to[0,r^2]$ *be continuous, equal* r^2 *on* $(-\sqrt{1+r},\sqrt{1+r})$, *and satisfy* $\varphi^{-1}(0)=\mathbb{R}\backslash[-1-r,1+r]$. *Then, if* r *is sufficiently small, the map,* g, *defined by*

$$g_1(\lambda,x)=\|x\|^2-\varphi(\lambda_1);\qquad g_i(\lambda,x)=\lambda_i,\qquad 1\leqq i\leqq m,$$

is a complement for f *on* $\mathcal{O}\backslash\{\mathbb{R}^m\times\{0\}\}$.

Proof. Since $\underline{\lambda}$ and $\bar{\lambda}$ are not Γ bifurcation points of f we may choose r such that

$$\begin{aligned}&f(\lambda,x)\neq 0 \text{ when } 0<\|x\|\leqq 2r \text{ and } \lambda=(\lambda_1,0,\ldots,0) \text{ with}\\ &|\lambda_1-1|\leqq r \text{ or } |\lambda_1+1|\leqq r,\end{aligned}\tag{4.6}$$

and we may assume

$$\left\{(\lambda,x)|\lambda=(\lambda_1,0,\ldots,0),\frac{|\lambda_1|^2}{1+r}+\frac{\|x\|^2}{r(1+r)}\leqq 1\right\}\subseteqq\mathcal{O}.\tag{4.7}$$

For such an r we let φ and g be as defined in the statement of the proposition. Let $\mathcal{U}\equiv\mathcal{O}\backslash\{\mathbb{R}^m\times\{0\}\}$.

If we let $K\equiv(g,f)^{-1}(0)\cap\mathcal{U}$, we observe that

$$K=\{(\lambda,x)|f(\lambda,x)=0,\|x\|=r,\lambda=(\lambda_1,0,\ldots,0),|\lambda_1|\leqq 1-r\}.$$

Thus K is a compact subset of $\mathcal{U}$, making $\deg((g,f),\mathcal{U},0)$ well-defined.

Now, (g,f) is a perturbation of the identity on $\mathbb{R}^m\times X$ by a compact mapping whose range lies in $\{(\lambda,x)|\lambda=(\lambda_1,0,\ldots,0)\}$. Consequently, if we identify the latter subspace with $\mathbb{R}\times X$, let $\mathcal{U}'=\mathcal{U}\cap\{\mathbb{R}\times X\}$, and define $h:\mathcal{O}\cap\{\mathbb{R}\times X\}\to\mathbb{R}\times X$ by

$$h(t,x)=(\|x\|^2-\varphi(t),f(t,x)),$$

it follows from the reduction property for Leray-Schauder degree that

$$\deg((g,f),\mathcal{U},0)=\deg(h,\mathcal{U}',0).$$

It follows from (4.7) that

$$W\equiv\left\{(t,x)\in\mathbb{R}\times X|\frac{|t^2|}{1+r}+\frac{\|x\|^2}{r(1+r)}<1\right\}\subseteqq\mathcal{O}\cap\{\mathbb{R}\times X\},$$

and one easily checks that

$$h^{-1}(0)\cap\mathcal{U}'=h^{-1}(0)\cap W.$$

Hence, from the excision property for degree,

$$\deg(h,\mathcal{U}',0)=\deg(h,W,0).$$

Now observe that $\|x\|^2-\varphi(t)=\|x\|^2-r^2$ on $\bar{W}$. It follows that $\|x\|^2-\varphi(t)=r(1-t^2)$ on ∂W, and since the degree of a mapping is only dependent on its

boundary values,

$$\deg(h, W, 0) = \deg(\psi, W, 0),$$

where $\psi(t, x) = (r(1-t^2), f(t, x))$, for $(t, x) \in \bar{W}$.

But the only zeros of ψ in W are $(+1, 0)$ and $(-1, 0)$. Hence if we let D' and D'' denote two small cubes about $(+1, 0)$ and $(-1, 0)$, from the excision and additivity properties of degree it follows that

$$\deg(\psi, W, 0) = \deg(\psi, D', 0) + \deg(\psi, D'', 0).$$

One easily sees that ψ is homotopic to $(t, x) \mapsto (r(1-t), f(1, x))$, on D', and to $(t, x) \mapsto (r(1+t), f(1, x))$, on D'', via homotopics which do not vanish on the boundaries.

Finally, from the product formulae for degree, together with the equalities established above, it follows that

$$\deg((g, f), \mathscr{U}, 0) = \operatorname{ind}(f_{\underline{\lambda}}, 0) - \operatorname{ind}(f_{\bar{\lambda}}, 0) \neq 0. \quad \square$$

To complete the proof it remains to translate the function $\|x\|^2 - \varphi(t)$, in the above proposition, onto the curve Γ.

We may choose an $\mathbb{R}^m$ neighborhood, W, of $[\underline{\lambda}, \bar{\lambda}]$, and an $\mathbb{R}^m$ neighborhood, V, of $\{t \in \mathbb{R}^m \mid t = (t_1, 0, \ldots, 0), |t_1| \leqq 1\}$, together with a mapping $\eta : \mathbb{R}^m \to \mathbb{R}$ such that the map $\psi : \mathbb{R}^m \to \mathbb{R}^m$ defined by

$$\psi(\lambda) = (\eta(\lambda), h_1(\lambda), \ldots, h_{m-1}(\lambda))$$

gives a homeomorphism of W onto V. We assume that

$$\psi(\underline{\lambda}) = (-1, 0, \ldots, 0), \qquad \psi(\bar{\lambda}) = (1, 0, \ldots, 0)$$

and

$$W \subseteqq \mathscr{O} \cap \{\mathbb{R}^m \times \{0\}\}.$$

Choose a $\sigma > 0$ such that $\hat{W} \equiv \{(\lambda, x) \mid \lambda \in W, \|x\| < \sigma\} \subseteqq \mathscr{O}$, and let $\hat{V} \equiv \{(t, x) \mid t \in V, \|x\| < \sigma\}$. Now define $\tilde{f} : \hat{V} \to X$ by

$$\tilde{f}(t, x) = f(\psi^{-1}(t), x).$$

It is clear that the assumptions of Proposition 4.1 hold, where $\tilde{f}$ and $\hat{V}$ play the roles of f and $\mathscr{O}$, respectively. Thus we may choose an $r > 0$ such that if $\varphi : \mathbb{R} \to [0, r]$ is continuous, $\varphi^{-1}(0) = \mathbb{R} \setminus [-1-r, 1+r]$, and $\varphi(t) = r^2$ when $|t|^2 \leqq 1 + r$, then

$$\tilde{g}(t, x) = (\|x\|^2 - \varphi(t_1), t_2, \ldots, t_m)$$

defines a mapping, $\tilde{g}$, which is a complement for $\tilde{f}$ on $\hat{V} \setminus \{\mathbb{R}^m \times \{0\}\}$. We also assume that (4.6) holds.

Now define $\tilde{\psi} : \hat{W} \to \hat{V}$ by $\tilde{\psi}(\lambda, x) = (\psi(\lambda), x)$. Observe that $\tilde{\psi}$ is a homeomorphism of $\hat{W} \setminus \{\mathbb{R}^m \times \{0\}\}$ onto $\hat{V} \setminus \{\mathbb{R}^m \times \{0\}\}$, and that $\tilde{\psi}$ is a compact perturbation of the identity on $\mathbb{R}^m \times X$. Consequently, from the composition property of degree, it follows that

$$\deg((\tilde{g}, \tilde{f}) \circ \tilde{\psi}, \hat{W} \setminus \{\mathbb{R}^m \times \{0\}\}, 0) \neq 0. \tag{4.8}$$

But $[(\tilde{g},\tilde{f})\circ\tilde{\varphi}](\lambda,x)=(g(\lambda,x),f(\lambda,x))$, where $g(\lambda,x)=(\|x\|^2-\varphi(\eta(\lambda)),h_1(\lambda),\ldots,\ldots,h_{m-1}(\lambda))$, for $(\lambda,x)\in\hat{W}\setminus\{\mathbb{R}^m\times\{0\}\}$.

Now let $q:\mathbb{R}^m\to\mathbb{R}$ be a continuous function which agrees with $\varphi\circ\eta$ in a W-neighborhood of $[\underline{\lambda},\bar{\lambda}]$, also agrees with $\varphi\circ\eta$ on $W\cap\Gamma$, and which vanishes on $\Gamma\setminus W$. Finally, let $g_1(\lambda,x)=\|x\|^2-q(\lambda)$, for $(\lambda,x)\in\mathbb{R}^m\times X$. We claim that g_1 satisfies (4.1)–(4.4).

But observe that $\{(\lambda,x)\in\mathcal{O}\setminus\{\mathbb{R}^m\times\{0\}\}\,|\,(g,f)(\lambda,x)=0\}=\{(\lambda,x)|\lambda\in\{\underline{\lambda},\bar{\lambda}],\ \|x\|=r,\ f(\lambda,x)=0\}$. Hence, from (4.8), together with the excision property for degree, we see that (4.1) holds. Moreover, it is clear that (4.2)–(4.4) hold with $\eta=r/2$, $\beta=2r$, and $\Lambda=\Gamma\cap\eta^{-1}((-1-r,1+r))$.

Thus the proof of the theorem is complete. □

Remark 4.1. In his proof of the one-parameter Rabinowitz bifurcation theorem, J. Ize has employed a map similar to the g of Proposition 4.1 (see [12]).

Acknowledgements. The authors were partially supported by the C.N.R. The first-named author was also partially supported by the N.S.F.

References

1. Alexander, J.C., Antman, S.S.: Global and local behavior of bifurcating multidimensional continua of solutions for multiparameter nonlinear eigenvalue problems. Arch. Rat. Mech. Anal. **76**, 339–355 (1981)
2. Alexander, J.C., Yorke, J.A.: The implicit function theorem and the global methods of cohomology. J. Functional Analysis **21**, 330–339 (1976)
3. Browder, F.E.: On continuity of fixed points under deformations of continuous mappings. Summa Brasil. Math. **4**, 183–190 (1960)
4. Dold, A.: Lectures on algebraic topology. Berlin, Heidelberg, New York: Springer 1972
5. Fitzpatrick, P.M., Massabò, I., Pejsachowicz, J.: Complementing maps, continuation and bifurcation. Bull. AMS (to appear)
6. Fitzpatrick, P.M., Massabò, I., Pejsachowicz, J.: On the covering dimension of the set of solutions of some nonlinear equations. Trans. AMS (to appear)
7. Fitzpatrick, P.M., Massabò, I., Pejsachowicz, J.: A global description of the periodic solutions of some ordinary differential equations. J. London Math. Soc. (to appear)
8. Hurewicz, W., Wallman, H.: Dimension theory. Princeton: Princeton University Press 1948
9. Lloyd, N.G.: Degree theory. Cambridge University Tracts in Mathematics, No. 78. Cambridge: Cambridge University Press 1978
10. Massabò, I., Pejsachowicz, J.: On the connectivity properties of the solution set of parameterized families of compact vector fields. J. Functional Anal. (to appear)
11. Leray, J., Schauder, J.: Topologie et equationes fonctionelles. Ann. Sci. École Norm. Sup. Pisa **51**, 45–78 (1934)
12. Nirenberg, L.: Topics in nonlinear functional analysis. Courant Inst. Lecture Notes, 1975
13. Rabinowitz, P.H.: Some global results for nonlinear eigenvalue problems. J. Functional Analysis **7**, 487–513 (1971)
14. Spanier, E.: Algebraic topology. New York: McGraw-Hill 1966

Received November 5, 1982

Math. Ann. 263, 75–86 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Sur une minoration du degré d'hypersurfaces s'annulant en certains points

Hélène Esnault[1] et Eckart Viehweg[2]

1 Université Paris VII, U.E.R. de Mathématiques, Tour 45–55, 5-ième étage, 2, Place Jussieu, F-75251 Paris Cedex 05, France
2 Institut für Mathematik und Informatik, A 5, Seminargebäude, D-6800 Mannheim 1, République Fédéral Allemande

Donnons-nous un ensemble fini S de points p_i dans une sous-variété quasi-projective lisse X^0 de $\mathbb{P}^n$ sur le corps des nombres complexes $\mathbb{C}$. Appelons Z la complétée projective de X^0 dans $\mathbb{P}^n$. On introduit les entiers

$$\omega_t(S,Z)=\inf\{d;\ \text{il existe } f\in\mathbb{C}[X_0,\ldots,X_n]-\mathscr{I}(Z),\ f \text{ homogène de degré } d, \text{ tel que } f\in\mathscr{M}_{p_i,Z}^t \text{ pour tout } p_i\in S\},$$

où $\mathscr{M}_{p_i,Z}$ est l'idéal maximal de p_i dans Z et $\mathscr{I}(Z)$ l'idéal homogène de Z.

Comme corollaire de résutats dûs à Enrico Bombieri sur la théorie des estimations L^2 et des théorèmes d'existence pour l'opérateur $\bar{\partial}$, une inégalité est démontrée entre $\omega_t(S,\mathbb{P}^n)$ et $\omega_{t'}(S,\mathbb{P}^n)$, pour deux entiers positifs t et t', que Skoda améliore sous la forme

$$\frac{\omega_{t'}(S,\mathbb{P}^n)}{t'+n-1}\leqq\frac{\omega_t(S,\mathbb{P}^n)}{t}.$$

Pour toutes ces notions et les références précises correspondantes, nous renvoyons au livre de Waldschmidt [6, Sect. 7].

Chudnovsky conjecture l'inégalité

$$\frac{\omega_1(S,\mathbb{P}^n)+n-1}{n}\leqq\frac{\omega_t(S,\mathbb{P}^n)}{t}\qquad(*)$$

qu'il prouve dans le cas $n=2$. Demailly redémontre (*) par des méthodes analytiques dans le cas $n=2$ ou bien dans celui où S contient un polytope complet à

$$\binom{\omega_1(S,\mathbb{P}^n)+n-1}{n}$$

sommets. Indépendemment, Wüstholz et Masser, l'un par des méthodes d'algèbre commutative et de théorie de l'intersection, l'autre d'algèbre linéaire, démontrent

une inégalité un peu moins précise que celle de Bombieri-Skoda-Waldschmidt, qu'ils peuvent améliorer dans le cas $n=2$ par des méthodes locales pour redonner (*). Par ailleurs, Wüstholz prouve aussi l'inégalité

$$\frac{H(\omega_1(S,Z)-1;Z)}{\omega_1(S,Z)^{m-1}\cdot\deg(Z)} \leqq \frac{\omega_t(S,Z)}{t},$$

où $H(x;Z)$ est la fonction de Hilbert de Z, $\deg(Z)$ son degré et m sa dimension. Nous renvoyons ici par exemple à [7] et [8] et aux références correspondantes.

Dans cet article nous prouvons par des méthodes de géométrie projective complexe les inégalités

$$\frac{\omega_{t'}(S,\mathbb{P}^n)+1}{t'+n-1} \leqq \frac{\omega_t(S,\mathbb{P}^n)}{t}, \quad \text{pour} \quad n\geqq 2 \tag{A}$$

$$\frac{\omega_{t'}(S,Z)+1-\deg(Z)}{t'+m-1} \leqq \frac{\omega_t(S,Z)}{t}, \quad m=\dim(Z)\geqq 1 \tag{B}$$

(voir Par. 1, Théorème II, Théorème I et Remarque (1.1), pour les énoncés précis).

Autant que nous sachions – et en tant que non spécialistes nous espérons n'avoir oublié ni de résultats ni de noms importants – l'inégalité (A) englobe et étend toutes les autres concernant $\omega_t(S,\mathbb{P}^n)$. En tenant compte des «mauvaises» singularités d'une hypersurface attachée à $\omega_t(S,\mathbb{P}^n)$ et de leur codimension, nous sommes en mesure de l'améliorer sensiblement [Remarque (1.2)]. Il nous semble que le point de vue adopté ici met en relief le lien entre d'une part le comportement de $\omega_{t'}(S,Z)$ en fonction de celui de $\omega_t(S,Z)$ et d'autre part le rôle joué par le lieu singulier de l'hypersurface associée à $\omega_t(S,Z)$. L'inégalité (B) est un cas spécial d'un résultat de type Bombieri-Skoda-Waldschmidt où l'on autorise S à être un ensemble de sous-variétés quelconques de X^0.

1. Enoncé des résultats

Soit X^0 une sous-variété quasi-projective lisse de $\mathbb{P}^n$. On appelle *degré* de X le degré de sa complétée projective Z dans $\mathbb{P}^n$.

Soit X_j^0 une sous-variété réduite et irréductible de X^0, de complétée projective Z_j dans Z. On dit qu'une section s de $H^0(\mathbb{P}^n,\mathcal{O}_{\mathbb{P}^n}(d))$ *s'annule à l'ordre* t_j *le long de* X_j^0 *et ne s'annule pas le long de* X^0 si, dans l'anneau local $\mathcal{O}_{Z_j,Z}$, s est dans $\mathcal{M}^{t_j}$ et est non nulle, où $\mathcal{M}=\mathcal{M}_{Z_j,Z}$ est l'idéal maximal. Bien sûr, s a le droit d'être dans $\mathcal{M}^{t_j+r}$ pour $r\geqq 0$.

Comme de coutume, on définit l'entier $[b]$, pour tout réel b, par $b-1<[b]\leqq b$.

Théorème I. *Soit* $S=\{X_j^0; j=1,\ldots,v\}$ *un ensemble fini de sous-variétés réduites et irréductibles* X_j^0, *de dimensions* n_j, *d'une variété quasi-projective lisse* X^0 *de dimension* $m\geqq 1$ *de* $\mathbb{P}^n$. *Supposons qu'il existe une section* s *de* $H^0(\mathbb{P}^n,\mathcal{O}_{\mathbb{P}^n}(d))$ *qui s'annule à un ordre donné* $t_j\geqq 1$ *le long de* X_j^0, *pour* $j=1,\ldots,v$, *sans s'annuler le long de* X^0. *Fixons des entiers* $t_j'\geqq 1$, *et*

$$i\geqq[(t_j'+m-n_j-1)\cdot d\cdot t_j^{-1}]+1$$

pour $j=1,\ldots,v$.

Alors il existe une section de $H^0(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(\deg(X^0)+i-2))$ *qui s'annule le long de* X_j^0 *à l'ordre* t'_j, *pour tout* $j=1,\ldots,v$, *sans s'annuler le long de* X^0.

Remarque (1.1). Si $n_1=n_2=\ldots=n_v=0$, c'est-à-dire si S est un ensemble de points, en posant $t_1=\ldots=t_v=t$ et $t'_1=\ldots=t'_v=t'$ on trouve

$$\omega_{t'}(S,Z)\leqq(t'+m-1)\cdot d\cdot t^{-1}-1+\deg(Z)$$

et donc l'inégalité (B).

Théorème II. *Soit* S *un ensemble fini de points* p_j *dans* $\mathbb{P}^n$, $n\geqq 2$. *Alors, s'il existe une section non nulle* s *de* $H^0(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(d))$ *qui s'annule à l'ordre* $t\geqq 1$ *en tout* p_j *de* S, *il existe pour* $i\geqq 2$ *une section non nulle de* $H^0(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(i-2))$ *qui s'annule en tout* p_j *de* S *à l'ordre* $t'\geqq 1$ *dès que*

$$i\geqq[(t'+n-1)\cdot d\cdot t^{-1}]+1\,.$$

Remarque (1.2). Dans la démonstration (Par. 4) il apparaîtra que les «*mauvaises singularités*» du diviseur $V(s)$ associé à s jouent un «mauvais tour». Celles que nous appelons mauvaises ne sont que les singularités en lesquelles le diviseur réduit associé à $V(s)$ n'est pas à croisements normaux. Pour $\beta\leqq n$, les mauvaises singularités étant en codimension supérieure ou égale à β par rapport à $\mathbb{P}^n$, nous démontrons l'inégalité

$$(\omega_{t'}(S,\mathbb{P}^n)+\beta-1)\cdot(t'+n-1)^{-1}\leqq d\cdot t^{-1}\,.$$

Puisque pour $n\geqq 2$, on peut toujours choisir au pire $\beta=2$, on obtient ainsi le théorème II. On atteindrait alors la conjecture de Chudnovsky (*) si l'on savait que l'hypersurface associée à $\omega_t(S,\mathbb{P}^n)$ n'a de mauvaise singularités que des singularités isolées, i.e. $\beta=n$.

Remarque (1.3). On pourrait aussi, dans le cas où le support n'est plus $\mathbb{P}^n$ mais un Z quelconque, introduire les mauvaises singularités du diviseur $V(s)\cap Z$ de Z associé à une section s de $H^0(\mathbb{P}^n, \mathcal{O}_{\mathbb{P}^n}(d))$, comme sous-variété de Z où Z et $V(s)$ ne se coupent pas transversalement. Malheureusement, même si Z est lisse, ces singularités peuvent être en codimension 1, de sorte que la formulation générale du théorème I ne s'en trouverait pas améliorée. C'est pourquoi nous ne parlerons pas de cet «essai manqué».

Notations. Rappelons que nous donnons dans les paragraphes suivants une démonstration des théorèmes I et II dans le cadre de la géométrie algébrique complexe. Pour ce, nous utilisons les notations du livre de Hartshorne [4], avec une exception: si Y est une variété algébrique et D un diviseur de Cartier, on note $\mathcal{O}_Y(D)$ le faisceau inversible associé (contrairement à la notation $\mathcal{L}(D)$ utilisée dans [4, p. 144]). Par ailleurs, on ne confondra pas le $\omega_t(S,Z)$ des arithméticiens avec le faisceau dualisant ω_Y.

2. Préliminaires

(2.1)

Soit B un diviseur effectif sur une variété projective lisse complexe Y. On pose $B=\sum v_j\cdot E_j$, où les E_j sont les composantes irréductibles de B. Si l'on suppose que

ce diviseur est associé à une puissance positive $\mathscr{L}^d$ d'un faisceau inversible $\mathscr{L}$. i.e. $\mathscr{L}^d = \mathscr{O}_Y(\sum v_j \cdot E_j)$, on pose, pour tout $i \geqq 0$,

$$\mathscr{L}^{(i)} = \mathscr{L}^i \otimes \mathscr{O}_Y(-\sum [v_j \cdot i \cdot d^{-1}] \cdot E_j).$$

Si $0 \leqq i < d$, la définition de $\mathscr{L}^{(i)}$ ne fait intervenir que les E_j pour lesquels $v_j \geqq 2$. Lorsque l'on voudra souligner le rôle de la partie réduite D de B, on écrira

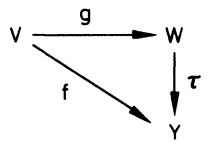
$$B = D + \sum_{v_j \geqq 2} v_j \cdot E_j, \quad \text{où} \quad D = \sum_{v_j = 1} v_j \cdot E_j.$$

(2.2)

On suppose que le diviseur B de (2.1) est à *croisement normaux*, c'est-à-dire que les E_j sont lisses et se coupent transversalement. La section s de $\mathscr{L}^d$, de support B, définit sur le faisceau de $\mathscr{O}_Y$-modules $\mathscr{A} = \bigoplus_{i=0}^{d-1} \mathscr{L}^{-i}$ une structure d'algèbre. En effet, la multiplication de A est définie par la multiplication

$$\mathscr{L}^{-i} \oplus \mathscr{L}^{-j} \to \mathscr{L}^{-i} \otimes \mathscr{L}^{-j} = \mathscr{L}^{-i-j},$$

une fois identifié $\mathscr{L}^{-d}$ à un sous-faisceau de $\mathscr{L}^0 = \mathscr{O}_Y$ par la section duale de s. On note W la normalisation de $\mathrm{Spec}_{\mathscr{O}_Y}(\mathscr{A})$, V une désingularisation de W et comme suit les morphismes associés



La variété W sera dite *extraction d-ième du diviseur B.*

Lemme ([1, Par. 1] et [2, Par. 1]). i) *W n'a que des singularités rationnelles (i.e. $R^i g_* \mathscr{O}_V = 0$ pour $i \neq 0$). En particulier W est de Cohen-Macaulay, et τ est plat.*

ii) *On a* $\tau_* \mathscr{O}_W = \bigoplus_{i=0}^{d-1} \mathscr{L}^{(i)^{-1}}$ *et donc* $R^i f_* \mathscr{O}_V = \begin{cases} \bigoplus_{i=0}^{d-1} \mathscr{L}^{(i)^{-1}}, & i = 0 \\ 0, & i \neq 0 \end{cases}$.

Théorème (2.3.1). *Soient Y, B et $\mathscr{L}$ comme dans* (2.2) *et ω_Y le faisceau canonique. Supposons que la dimension de Kodaira de $\mathscr{L}$ vérifie $\kappa(\mathscr{L}) = \dim(Y) = n$, et que $\mathscr{L}$ est engendré par ses sections globales. Alors on a*

$$H^q(Y, \omega_Y \otimes \mathscr{L}^{(i)} \otimes \mathscr{L}^k) = 0$$

pour tous $k > 0$, $q > 0$ et $i \geqq 0$.

Démonstration. D'après (2.2) et la dualité de Serre [4, III, Par. 7], on a

$$H^q(V, \omega_V \otimes f^* \mathscr{L}^k) = H^{n-q}(V, f^* \mathscr{L}^{-k}) = \bigoplus_{i=0}^{d-1} H^{n-q}(Y, \mathscr{L}^{(i)^{-1}} \otimes \mathscr{L}^{-k})$$

$$= \bigoplus_{i=0}^{d-1} H^q(Y, \omega_Y \otimes \mathscr{L}^{(i)} \otimes \mathscr{L}^k).$$

Le faisceau $f^*\mathscr{L}^k$ étant de dimension de Kodaira maximum et engendré par ses sections globales, on peut appliquer le théorème d'annulation de Grauert et Riemenschneider [3] pour trouver le résultat pour $i<d$. Si $i\geqq d$, on écrit $\mathscr{L}^{v\cdot d}=\mathscr{O}_Y(v\cdot B)$ pour $i<v\cdot d$.

En fait, le théorème (2.3.1) est un cas particulier du

Théorème (2.3.2) ([2, Par. 2] ou [5]). *Soient Y, B et $\mathscr{L}$ comme dans* (2.2). *Supposons de plus que la dimension de Kodaira de $\mathscr{O}_Y(D)$ vérifie $\kappa(\mathscr{O}_Y(D))=\dim(Y)$. Alors on a*

$$H^q(Y,\omega_Y\otimes\mathscr{L}^{(i)})=0$$

pour tous $q>0$ et $d>i>0$.

La démonstration repose sur les trois faits suivants:

i) Le calcul des $\mathscr{L}^{(i)}$ exposé en (2.2).

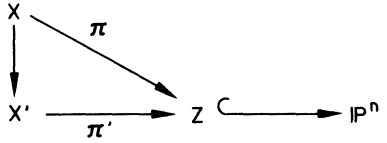
ii) La symétrie des nombres de Hodge de la variété V.

iii) La fermeture des formes différentielles globales à pôles logarithmiques le long d'un diviseur à croisements normaux.

Seul i) est purement algébrique. Jusqu'à présent, il n'y a pas de démonstration algébrique de ii) et iii).

(2.4)

Soit X^0 une sous-variété quasi-projective lisse de dimension $m\geqq 1$ de $\mathbb{P}^n$. Soient Z sa complétée projective dans $\mathbb{P}^n$ et $\pi:X\to Z$ un morphisme birationnel tel que X soit lisse et projective. Au cours de la démonstration du théorème I, nous construisons des sections de certains faisceaux inversibles sur X que nous voulons identifier à la restriction sur X^0 de fonctions polynômiales sur $\mathbb{P}^n$. Nous le faisons à l'aide de la proposition de «chasse aux sections» qui suit. Soient X' la normalisée de Z et



les morphismes correspondants. Appelons U l'ouvert des points lisses de X'.

Pour toute variété U' et tout morphisme $\varphi:U'\to\mathbb{P}^n$, on pose $\mathscr{O}_{U'}(1)=\varphi^*\mathscr{O}_{\mathbb{P}^n}(1)$. Pour tout l, appelons θ_l la composition des applications canoniques:

$$H^0(\mathbb{P}^n,\mathscr{O}_{\mathbb{P}^n}(l))\to H^0(Z,\mathscr{O}_Z(l))\xrightarrow{\pi'^*}H^0(X',\mathscr{O}_{X'}(l))\to H^0(U,\mathscr{O}_U(l)).$$

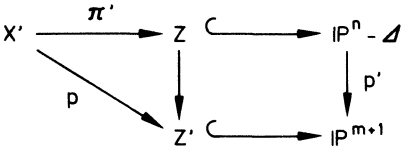
Proposition. *Il existe une injection*

$$j:\omega_U\to\mathscr{O}_U(\deg(X^0)-m-2)$$

vérifiant que, pour tout k, l'image par j de $H^0(U,\omega_U\otimes\mathscr{O}_U(k))$ dans $H^0(U,\mathscr{O}_U(\deg(X^0)-m-2+k))$ est contenue dans l'image de $\theta_{\deg(X^0)-m-2+k}$.

Démonstration. Il suffit de démontrer la proposition où l'on remplace U par tout ouvert X_1 de points lisses dont le complément $X'-X_1$ est de codimension au moins deux. Choisissons une projection générique de degré 1 de Z sur Z',

hypersurface de degré $\deg(X^0)=\deg(Z)$ dans $\mathbb{P}^{m+1}$. Notons Δ le lieu de $\mathbb{P}^n$ sur lequel cette projection n'est pas définie et comme suit les morphismes associés:



Le morphisme p est fini. Posons $X_1=X'-p^{-1}(p(\mathrm{Sing}(X')))$ et $Z_1=p(X_1)$. Alors $Z'-Z_1$ est de codimension au moins deux dans Z'.

En tant qu'hypersurface de $\mathbb{P}^{m+1}$, Z' est de Gorenstein, donc admet un faiceau dualisant inversible vérifiant [4, II, Par. 8 et III, Par. 7] $\omega_{Z'}=\mathcal{O}_{Z'}(\deg(X^0))-m-2)$.

On a donc $\omega_{Z_1}=\mathcal{O}_{Z_1}(\deg(X^0)-m-2)$.

Par ailleurs, on a [4, III, Ex. 6.10 et 7.2] $p_*\omega_{X_1}=\mathcal{H}om_{Z_1}(p_*\mathcal{O}_{X_1},\omega_{Z_1})$. Le conoyau $\mathcal{C}$ de l'inclusion $\mathcal{O}_{Z_1}\to p_*\mathcal{O}_{X_1}$ étant de torsion, on a $\mathcal{H}om_{Z_1}(\mathcal{C},\omega_{Z_1})=0$, ce qui founit une inclusion

$$p_*\omega_{X_1}\to\omega_{Z_1}=\mathcal{O}_{Z_1}(\deg(X^0)-m-2)$$

et donc un morphisme non nul

$$j':p^*p_*\omega_{X_1}\to p^*\omega_{Z_1}.$$

Le noyau $\mathcal{T}$ de la surjection canonique

$$p^*p_*\omega_{X_1}\to\omega_{X_1}$$

étant de torsion, on a $j'|_{\mathcal{T}}=0$, donc j' se factorise en

$$j:\omega_{X_1}\to p^*\omega_{Z_1}=\mathcal{O}_{X_1}(\deg(X^0)-m-2).$$

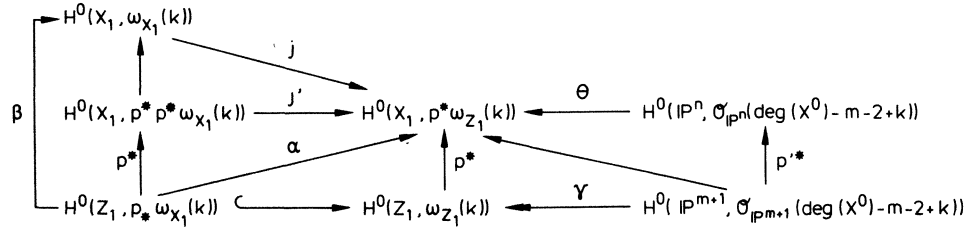
De plus, les deux faisceaux étant inversibles et j étant non nul, j est injectif. D'autre part, Δ étant de codimension au moins deux dans $\mathbb{P}^n$, on a

$$\begin{aligned}&H^0(\mathbb{P}^n-\Delta,\mathcal{O}_{\mathbb{P}^n-\Delta}(\deg(X^0)-m-2+k))\\&=H^0(\mathbb{P}^n,\mathcal{O}_{\mathbb{P}^n}(\deg(X^0)-m-2+k)).\end{aligned}$$

De même, $Z'-Z_1$ étant de codimension au moins deux dans Z', qui est de Gorenstein, a fortiori de Cohen-Macaulay, on a $H^0(Z_1,\omega_{Z_1}(k))=H^0(Z',\omega_{Z'}(k))$. La flèche de restriction

$$\gamma:H^0(\mathbb{P}^{m+1},\mathcal{O}_{\mathbb{P}^{m+1}}(\deg(X^0)-m-2+k))\to H^0(Z',\omega_{Z'}(k))$$

est surjective, son conoyau étant dans $H^1(\mathbb{P}^{m+1},\mathcal{O}_{\mathbb{P}^{m+1}}(-m-2+k))$, qui est nul car $m\geqq 1$. On obtient les diagrammes commutatifs suivants



où par définition β est l'identité et $\alpha = j' \cdot p^*$. Donc $\mathrm{Im}(j) = \mathrm{Im}(\alpha)$ est contenu dans

$$p^*(H^0(Z_1, \omega_{Z_1}(k))) = \theta \cdot p'^*(H^0(\mathbb{P}^{m+1}, \mathcal{O}_{\mathbb{P}^{m+1}}(\deg(X^0) - m - 2 + k))),$$

lui-même contenu dans $\mathrm{Im}(\theta)$.

3. Démonstration du théorème I

(3.1)

Nous considérons la situation du Par. 1. Soient X^0 quasi-projective lisse, de dimension m, Z sa complétée projective dans $\mathbb{P}^n$, $\{X_j^0\}_{j=1,\ldots,\nu}$ des sous-variétés réduites et irréductibles de X^0, de dimensions n_j, Z_j leurs complétées projectives dans Z.

Soit maintenant X' la normalisation de X. Depuis Hironaka (voir [4, V, 3.8.1], pour les résultats et références correspondantes) on sait qu'il existe une variété projective lisse X, et un morphisme birationnel $X \to X'$ obtenu par une suite finie d'éclatements à centres lisses situés au dessus du lieu singulier de X'. En particulier, on peut concevoir l'ensemble des points lisses de X' comme ouvert U de X. Dorénavant, on entendra par désingularisation de Z un tel morphisme birationnel $\pi : X \to Z$ que l'on fixe une fois pour toutes. Appelons X_j la transformée stricte de Z_j, qui contient bien sûr X_j^0.

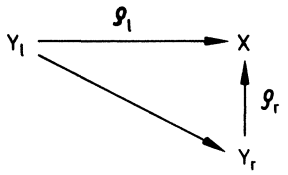
(3.2)

Nous construisons une désingularisation du diviseur $V(s)$ associé à la section s du théorème I. Afin de simplifier la suite, nous la donnons «explicitement».

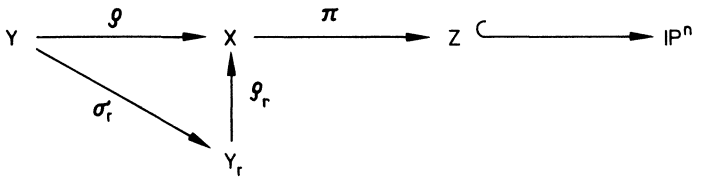
Ordonnons les dimensions n_j des X_j en $m_1 < \ldots < m_l$. Considérons la sous-variété de X de dimension m_1 : $\bigcup_{n_j = m_1} X_j$. Par une suite d'éclatements à centres lisses situés au dessus du lieu singulier de $\bigcup_{n_j = m_1} X_j$, on désingularise ce schéma réduit. En fait, ces éclatements sont la restriction d'éclatements de même centre, mais considérés dans X [4, II, 7.15], de sorte que la variété de plongement reste lisse. On a donc un éclaté de X dans lequel les transformés stricts des X_j sont lisses et deux à deux disjoints, pour $n_j = m_1$. Eclatons-les simultanément, ce qui revient à éclater leur réunion. Appelons $\varrho_1 : Y_1 \to X$ le morphisme birationnel ainsi contruit et F_j^1 le diviseur image inverse de X_j pour $n_j = m_1$. Comme ϱ_1 est composé d'éclatements à centres de dimension inférieure ou égale à m_1, l'image inverse stricte de X_j, que l'on note encore X_j, pour $n_j \geqq m_2$, est birationnelle à X_j.

Prenons $\bigcup_{n_j = m_2} X_j$ dans Y_1 et faisons la même construction. On otient ainsi un morphisme birationnel de Y_2 vers Y_1, et donc un morphisme birationnel $\varrho_2 : Y_2 \to X$.

On note F_j^2 le diviseur image inverse stricte de X_j dans Y_2 pour $n_j = m_2$. On continue jusqu'à la dimension m_l et on note F_j^r les diviseurs lisses au-dessus de la transformée stricte de X_j dans Y_r pour $n_j = m_r$ et comme suit les morphismes birationnels obtenus:



Cela posé, on considère enfin $\varrho_l^{-1}(V(s))$, dont on fait une désingularisation plongée. Pour cela, on construit une variété projective et lisse Y par une suite d'éclatements de Y_l de sorte que le transformé total de $\varrho_l^{-1}(V(s))$ soit un diviseur à croisements normaux. On note F_j les transformés stricts des F_j^r et comme suit les morphismes associés:



Posons, comme dans (2.4), $\mathcal{O}_Z(1)=\mathcal{O}_{\mathbb{P}^n}(1)|_Z$, $\mathcal{O}_X(1)=\pi^*\mathcal{O}_Z(1)$, $\mathscr{L}_r=\varrho_r^*\mathcal{O}_X(1)$ et $\mathscr{L}=\varrho^*\mathcal{O}_X(1)$. La section s étant dans $H^0(\mathbb{P}^n,\mathcal{O}_{\mathbb{P}^n}(d))$, la section $\varrho^*\pi^*s$ est dans $H^0(Y,\mathscr{L}^d)$ et son diviseur associé est un diviseur B à croisements normaux, transformé total de $V(s)$. Comme en (2.1), nous écrivons $\mathscr{L}^d=\mathcal{O}_Y(\sum v_j\cdot E_j)$. Notons que, par hypothèse du théorème I, $\varrho_r^*\pi^*s$ est dans $H^0(Y_r,\mathscr{L}_r^d\otimes\mathcal{O}_{Y_r}(-t_j\cdot F_j^r))$ pour tous r et j tels que $n_j=m_r$. En particulier, $\sum v_j\cdot E_j$ contient $\sum t_j\cdot F_j$.

Lemme (3.3). *Pour tous $1\leqq r\leqq l$ et $k\geqq 0$, on a les inclusions*

a) $\sigma_{r*}(\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^k)\to\omega_{Y_r}\otimes\mathscr{L}_r^i\otimes\mathcal{O}_{Y_r}(-\sum[t_j\cdot i\cdot d^{-1}]\cdot F_j^r)\otimes\mathscr{L}_r^k$,

b) $\varrho_*(\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^k)\to\varrho_{r*}(\omega_{Y_r}\otimes\mathscr{L}_r^i\otimes\mathcal{O}_{Y_r}(-\sum[t_j\cdot i\cdot d^{-1}]\cdot F_j^r)\otimes\mathscr{L}_r^k)$
$\to\omega_X\otimes\mathcal{O}_X(i+k)$.

Preuve. Puisque $\sum v_j\cdot E_j$ contient $\sum t_j\cdot F_j$, on a l'inclusion

$$\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^k\to\sigma_r^*(\omega_{Y_r}\otimes\mathscr{L}_r^i\otimes\mathcal{O}_{Y_r}(-\sum[t_j\cdot i\cdot d^{-1}]\cdot F_j^r)\otimes\mathscr{L}_r^k)\otimes\mathcal{O}_Y(E)$$

où E est un diviseur effectif contenu dans le lieu exceptionnel du morphisme birationnel σ_r. Donc $\sigma_{r*}\mathcal{O}_Y(E)=\mathcal{O}_{Y_r}$ et on applique la formule de projection pour obtenir a). Les insclusions b) sont alors triviales.

Rappelons-nous ici que (2.4) permet d'identifier des sections de $\omega_X\otimes\mathcal{O}_X(i+k)$, par l'intermédiaire de leurs restrictions à l'ouvert U, à des fonctions polynômiales sur $\mathbb{P}^n$, restreintes à X^0. On ramène dans ce qui suit le théorème I à l'existence de sections de $\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^k$, pour un k convenable.

Lemme (3.4). *Supposons que $c_1(\mathscr{L})^{m-1}\cdot c_1(\mathscr{L}^{(i)})>0$ (ou bien que $c_1(\mathscr{L}^{(i)})>0$ si $m=1$). Alors on a $H^0(Y,\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^m)\neq 0$.*

Preuve. Si $m=\dim(Y)\geqq 2$, choisissons une section hyperplane H de $\mathbb{P}^n$ en position générale. Appelons Y' l'image inverse dans Y de H. On a $\omega_{Y'}=\omega_Y\otimes\mathscr{L}\otimes\mathcal{O}_{Y'}$.

De la suite exacte

$$0\to\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-1}\to\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m}\to\omega_{Y'}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-1}\to 0$$

et de (2.3.1), on est ramené à prouver l'existence d'une section non nulle de $\omega_{Y'}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-1}$.

Posons $Z'=Z\cap H\subset H\cong\mathbb{P}^{n-1}$, et appelons X' son image inverse dans X. Alors, X' et Y' sont lisses, Y' rencontre transversalement $\sum v_j\cdot E_j$, le diviseur $\sum v_j\cdot E_j|_{Y'}$ est à croisements normaux, de sorte que ce diviseur, $\mathscr{L}^{d}|_{Y'}$ et $\mathscr{L}^{(i)}|_{Y'}$ sont des données de type (2.2).

Ainsi, en procédant par récurrence sur la dimension de Y, il faut prouver l'existence d'une section du faisceau $\omega_C\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}$, où C est la normalisation d'une intersection générique de Z avec $\mathbb{P}^{n-m+1}$.

Dans le cas $m=1$, C est simplement Y. On a

$$\deg_C(\mathscr{L}^{(i)})=\sum(v_j\cdot i\cdot d^{-1}-[v_j\cdot i\cdot d^{-1}])\cdot\deg_C(E_j)\geqq 0.$$

L'hypothèse signifie que $\deg_C(\mathscr{L}^{(i)})>0$. Donc on a $\deg_C(\mathscr{L}^{(i)}\otimes\mathscr{L})\geqq 2$. Cela garantit l'existence d'une section non nulle de $\omega_C\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}$, indépendamment du genre de C.

Remarque (3.5). On a montré en fait que, sans hypothèses, on a toujours $H^0(Y,\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m+1})\neq 0$.

(3.6) *Fin de la démonstration du théorème* I

Sous l'hypothèse de (3.4), on a construit une section non nulle de $\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m}$, donc (3.3, b), une section de $\omega_X\otimes\mathscr{O}_X(i+m)$. Par restriction à l'ouvert U, on obtient une section de $\omega_U\otimes\mathscr{O}_U(i+m)$. L'inclusion j de (2.4) décrit cette section comme provenant d'une section $s^{(i)}$ de $\mathscr{O}_{\mathbb{P}^n}(\deg(X^0)-2+i)$. Par ailleurs, au-dessus du point générique de chacun des X_j, avec $n_j=m_r$, on a $\omega_{Y_r}=\varrho_r^*\omega_X\otimes\mathscr{O}((m-n_j-1)\cdot F_j^r)$, car le morphisme $\varrho_r: Y_r\to X$ est le premier éclatement du point générique de X_j (cf. [4, II, Ex. 8.5]. La section $s^{(i)}$ construite s'annule donc (3.3, a) à l'ordre $[t_j\cdot i\cdot d^{-1}]-(m-n_j-1)$ le long de X_j^0. Pour qu'elle s'annule à l'ordre $t_j'\geqq 1$ le long de X_j^0, $j=1,\ldots,v$, il suffit que

$$[t_j\cdot i\cdot d^{-1}]-(m-n_j-1)\geqq t_j'$$

soit

$$i\geqq\{(t_j'+m-n_j-1)\cdot d\cdot t_j^{-1}\}$$

où pour tout réel b on pose $\{b\}=-[-b]$. Ceci prouve le théorème I sous l'hypothèse de (3.4).

En fait, on peut toujours supposer que celle-ci est vérifiée [de même d'ailleurs que celle de (2.3.2)]: Remplaçons $s\in H^0(\mathbb{P}^n,\mathscr{O}_{\mathbb{P}^n}(d))$ par $s^\eta\cdot s'$ tel que $s'\in H^0(\mathbb{P}^n,\mathscr{O}_{\mathbb{P}^n}(1))$ soit une section en position générale. Alors d est transformé en $\eta\cdot d+1$, t_j en $\eta\cdot t_j$.

Pour η suffisamment grand, on a

$$[(t_j'+m-n_j-1)\cdot d\cdot t_j^{-1}]+1=[(t_j'+m-n_j-1)\cdot(\eta\cdot d+1)\cdot(\eta\cdot t_j)^{-1}]+1.$$

$\mathscr{L}^{\eta\cdot d+1}$ admet une décomposition $\mathscr{L}^{\eta\cdot d+1}=\mathscr{O}_Y(D+\sum v_j\cdot E_j)$, où la partie réduite D est le diviseur associé à ϱ^*s', ce qui donne (3.4) et donc le théorème I. [De plus, on a $\kappa(\mathscr{O}_Y(D))=m$.]

4. Amélioration du théorème I dans le cas où $X^0=\mathbb{P}^n$ et S est un ensemble de points

Remarque (4.1). Si $X^0=\mathbb{P}^n$, alors la proposition (2.4) est inutile et la démonstration du théorème I est sensiblement simplifiée. En particulier, si de plus S est un ensemble de points, le théorème I redonne le théorème de Bombieri-Skoda-Waldschmidt. Pour démontrer le théorème II, il nous faut donc gagner une unité par rapport au théorème I.

Soit β comme dans (1.2), c'est-à-dire que β est inférieur ou égal à la codimension des mauvaises singularités de $V(s)$ et vérifie $2\leqq\beta\leqq n$. On va essayer de construire, en coupant comme dans (3.4) par des sections hyperplanes, des sections non nulles de $\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{n-\beta+1}$. Afin de pouvoir appliquer (2.3.2), on utilise le subterfuge déjà mentionné en (3.6).

(4.2)

Posons $n'=t'+n-1$ et $i=[n'\cdot d\cdot t^{-1}]+1$. On a $i>n'\cdot d\cdot t^{-1}$, et donc $i\cdot t-n'\cdot d>0$. Il existe un entier η suffisamment grand, et un entier $r>0$ tels que $i=\dfrac{n'\cdot(\eta\cdot d+r)}{\eta\cdot t}<\eta\cdot d+r$. [Prendre par exemple $\eta=k\cdot n'$ pour un entier $k\geqq 2$ et $r=(i\cdot t-n'\cdot d)\cdot k$.]

Posons $\delta=\eta\cdot d+r$ et $\tau=\eta\cdot t$. Choisissons r sections $s_1,\ldots,s_r$ de $\mathscr{O}_{\mathbb{P}^n}(1)$ en position générale. En particulier, les diviseurs associés aux s_j ne rencontrent pas les points de S. Posons $\sigma=s^\eta\cdot s_1\cdot\ldots\cdot s_r$. Alors σ est une section de $\mathscr{O}_{\mathbb{P}^n}(\delta)$ qui passe par les points de S avec la multiplicité au moins τ et dont la partie réduite n'est pas triviale.

On fait la construction (3.2) de Y pour cette section σ, ou pour l'ancienne section s, ce qui est la même chose. Ici, on n'a qu'un seul Y_1 pour $m_1=0$, et Y est simplement la désingularisation de $\varrho_1^*V(\sigma)$. La partie réduite de $V(\varrho^*\sigma)$ est le diviseur associé à $\varrho^*(s_1\cdot\ldots\cdot s_r)$ et donc de dimension de Kodaira n. Posons $V(\sigma)=\sum m_j\cdot D_j$, où les D_j sont irréductibles et réduits, de degrés d_j. En particulier $\delta=\sum m_j\cdot d_j$. De plus $V(\sigma)$ n'a de mauvaises singularités qu'en codimension β.

Lemme (4.3). *Si $q=n'\cdot\delta\cdot\tau^{-1}-\sum[m_j\cdot n'\cdot\tau^{-1}]\cdot d_j\geqq\beta$, alors $\mathscr{O}_{\mathbb{P}^n}(n'\cdot\delta\cdot\tau^{-1}-\beta)$ a une section non nulle qui s'annule en tout p_j de S à l'ordre t'.*

(Rappelons que par construction $i=n'\cdot\delta\cdot\tau^{-1}$, et donc que q est entier.)

Preuve. Coupons $V(\sigma)$ avec un $\mathbb{P}^{\beta-1}$ générique. On obtient, d'après l'hypothèse faite sur β, un diviseur $D'=\sum m_j\cdot D'_j$, de $\mathbb{P}^{\beta-1}$, localement à croisements normaux, de degré δ et de composantes D'_j de degrés d_j. Appelons $Y^{\beta-1}$ l'image inverse par $\varrho: Y\to\mathbb{P}^n$ de ce $\mathbb{P}^{\beta-1}$. On a [5, 2.3]

$$\begin{aligned}\varrho_*(\omega_{Y^{\beta-1}}\otimes\mathscr{L}^{(i)})&=\omega_{\mathbb{P}^{\beta-1}}\otimes\mathscr{O}_{\mathbb{P}^{\beta-1}}(i)\otimes\mathscr{O}_{\mathbb{P}^{\beta-1}}(-\sum[m_j\cdot i\cdot\delta^{-1}]\cdot D'_j)\\&=\mathscr{O}_{\mathbb{P}^{\beta-1}}(-\beta+n'\cdot\delta\cdot\tau^{-1}-\sum[m_j\cdot(n'\cdot\delta\cdot\tau^{-1})\cdot\delta^{-1}]\cdot d_j).\end{aligned}$$

Ce faisceau a donc une section non nulle. De la suite exacte

$$0\to\omega_{Y^\beta}\otimes\mathscr{L}^{(i)}\to\omega_{Y^\beta}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}\to\omega_{Y^{\beta-1}}\otimes\mathscr{L}^{(i)}\to 0$$

où Y^β est l'intersection de Y avec l'image inverse par ϱ d'un $\mathbb{P}^\beta$ générique, et du théorème d'annulation (2.3.2), on tire l'existence d'une section non nulle de $\omega_{Y^\beta}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}$. Puis en appliquant (2.3.1) à la suite exacte

$$0\to\omega_{Y^m}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-\beta}\to\omega_{Y^m}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-\beta+1}\to\omega_{Y^{m-1}}\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{m-\beta}\to 0$$

où Y^m est l'image inverse d'un $\mathbb{P}^n$ en position générique, $m>\beta$, on trouve par récurrence une section non nulle dans

$$H^0(Y,\omega_Y\otimes\mathscr{L}^{(i)}\otimes\mathscr{L}^{n-\beta+1})\hookrightarrow H^0(\mathbb{P}^n,\omega_{\mathbb{P}^n}\otimes\mathscr{O}_{\mathbb{P}^n}(i+n-\beta+1)),$$

dont nous savons comme en (3.6) qu'elle s'annule à l'ordre $-(n-1)+[\tau\cdot i\cdot\delta^{-1}]=-n+1+n'=t'$ en tout p_j de S.

Afin de finir la démonstration du théorème II, il nous reste à traiter le cas où l'on ne peut plus appliquer les théorèmes d'annulation. On raisonne alors directement et combinatoirement sur le diviseur $V(\sigma)$.

Lemme (4.4). *Si* $q=n'\cdot\delta\cdot\tau^{-1}-\sum[m_j\cdot n'\cdot\tau^{-1}]\cdot d_j\leqq\beta-1$, *alors* $\mathscr{O}_{\mathbb{P}^n}(n'\cdot\delta\cdot\tau^{-1}-\beta)$ *a une section non nulle qui s'annule en tout* p_j *de* S *à l'ordre* t'.

Preuve. Puisque il existe au moins un j pour lequel $m_j=1$, on a pour ce $j:[m_j\cdot n'\cdot\tau^{-1}]=[i\cdot\delta^{-1}]=0$, et donc $1\leqq q$. Appleons r_j^k la multiplicité du point p_k sur la composante D_j. Alors, pour tout p_k de S, on a $\tau\leqq\sum_j r_j^k\cdot m_j$.

Donc

$$n'\leqq\sum_j r_j^k\cdot m_j\cdot n'\cdot\tau^{-1}$$

et

$$n'-q\leqq\sum_j(r_j^k\cdot m_j\cdot n'\cdot\tau^{-1}-n'\cdot m_j\cdot d_j\cdot\tau^{-1}+[m_j\cdot n'\cdot\tau^{-1}]\cdot d_j).$$

Posons $\alpha_j=\tau\cdot(n'\cdot m_j\cdot\tau^{-1}-[n'\cdot m_j\cdot\tau^{-1}])$, $0\leqq\alpha_j<\tau$.

On obtient

$$n'-q\leqq\sum_j r_j^k\cdot[n'\cdot m_j\cdot\tau^{-1}]-\sum_j(d_j-r_j^k)\cdot\alpha_j\cdot\tau^{-1}.$$

Par définition on a $r_j^k\leqq d_j$, et par construction (4.2), il existe un j_0 pour lequel $m_{j_0}=1$ et D_{j_0} ne contient aucun point de S. Donc on a $r_{j_0}^k=0$ et $\alpha_{j_0}>0$, et donc $(d_{j_0}-r_{j_0}^k)\cdot\alpha_{j_0}\cdot\tau^{-1}>0$.

Ainsi on a l'inégalité stricte $n'-q<\sum_j r_j^k\cdot[n'\cdot m_j\cdot\tau^{-1}]$. Reformulons cette inégalité

$$t'+\beta-q\leqq t'+n-q=n'-q+1\leqq\sum_j r_j^k\cdot[n'\cdot m_j\cdot\tau^{-1}]\leqq\sum_j[n'\cdot m_j\cdot\tau^{-1}]\cdot d_j. \quad (1)$$

Posons $s'=\prod_j s_j^{[n'\cdot m_j\cdot\tau^{-1}]}$ où s_j définit le diviseur D_j. L'inégalité (1) étant vraie pour tout k, s' s'annule à l'ordre au moins $t'+\beta-q$ en chaque point p_k de S.

Différentions $\beta - q$ fois s' dans des directions génériques de telle sorte que les dérivées ne deviennent pas identiquement nulles. On obtient ainsi une section s_1 qui s'annule aux points p_j à l'ordre t' et qui est de degré

$$\sum_j [n' \cdot m_j \cdot \tau^{-1}] \cdot d_j - (\beta - q) = i - \beta .$$

Remarque (4.5). Le cas (4.4) est en fait le moins intéressant, puisque le degré du diviseur $V(\sigma)$ est très grand par rapport à ce qui serait nécessaire en exigeant la multiplicité t en les points de S. L'inégalité (1) dit en fait qu'on aurait pu différentier $n - q$ fois au lieu de $\beta - q$ fois. On aurait dans ce cas obtenu l'estimation de Chudnovsky (*) de l'introduction.

Remerciement. Michel Demazure nous a demandé si les méthodes développées dans [2] pouvaient être appliquées à la résolution de ce genre de question. Qu'il en soit ici chaleureusement remercié.

Bibliographie

1. Esnault, H.: Fibre de Milnor d'un cône sur une courbe plane. Invent. Math. **68**, 477–496 (1982)
2. Esnault, H., Viehweg, E.: Revêtements cycliques. Proc. Conf. Algebraic Threefolds, Varenna 1981, pp. 241–251. In: Lecture Notes in Mathematics, Vol. 947. Berlin, Heidelberg, New York: Springer 1982
3. Grauert, H., Riemenschneider, O.: Verschwindungssätze für analytische Kohomologiegruppen auf komplexen Räumen. Invent. Math. **11**, 263–292 (1970)
4. Hartshorne, R.: Algebraic geometry. Berlin, Heidelberg, New York: Springer 1977
5. Viehweg, E.: Vanishing theorems. J. reine angew. Math. **335**, 1–8 (1982)
6. Waldschmidt, M.: Nombres transcendants et groupes algébriques. Astérisque 69/70
7. Wüstholz, G.: On the degree of algebraic hypersurfaces with given singularities (manuscript)
8. Wüstholz, G.: Nullstellenabschätzungen auf Varietäten. Séminaire Delange-Poitou-Pisot, Octobre 1981

Reçu le 24 septembre 1982: revisé le 17 décembre 1982

Math. Ann. 263, 87–144 (1983)

Mathematische Annalen
© Springer-Verlag 1983

A Hyperbolic Kac-Moody Algebra and the Theory of Siegel Modular Forms of Genus 2

Alex J. Feingold[1,*] and Igor B. Frenkel[2,**]

1 The State University of New York at Binghamton, Binghamton, NY 13901, USA
2 Yale University, New Haven, CT 06520, USA

Dedicated to Nathan Jacobson

Table of Contents

1. Introduction 87
2. The Kac-Moody Algebra $\mathfrak{F}$ 92
3. Standard Representations and Characters of $\mathfrak{F}$ and Affine Subalgebra $\mathfrak{F}_0$ 99
4. The Construction of $\mathfrak{F}$ 110
5. Theta Series and the Space of $SL_2(\mathbb{Z})$-Invariant $\mathfrak{F}_0^e$-Characters 118
6. Siegel Modular Forms and Hecke Operators 125
7. The Correspondence of $\mathfrak{F}$-Characters with Siegel Modular Forms of Genus 2 129
8. Future Prospects 141

1. Introduction

In the last few years the theory of Kac-Moody Lie algebras has drawn considerable attention because of its numerous connections with other topics in mathematics and physics. Most of these results concern only the simplest class of infinite dimensional Kac-Moody algebras known as the affine Lie algebras. Except for some general results and the papers [4, 24, 35], little is known about the hyperbolic Lie algebras, the next class of Kac-Moody algebras after the affines. In this paper we restrict ourselves to a detailed investigation of one particular hyperbolic Lie algebra, $\mathfrak{F}$, although our techniques certainly generalize. Our main results are a construction of $\mathfrak{F}$ and the relationship between the representation theory of $\mathfrak{F}$ and the theory of Siegel modular forms of genus 2. One of the purposes of this paper is to show the richness of the algebraic structures associated with the hyperbolic Kac-Moody algebras and their newly discovered connections with classical mathematics. These lead us to expect even more important discoveries in this direction.

The definition of a Kac-Moody algebra by generators and relations (see Sect. 2) is based on an integral square matrix A, called the Cartan matrix. We will

* Research partially supported by NSF research grant MCS80-02198
** Research partially supported by NSF research grant MCS81-02534

study the case where the Cartan matrix is

$$A=(A_{ij})=\begin{pmatrix} 2 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}, \tag{1.1}$$

and the corresponding Dynkin diagram is



(1.2)

The Cartan matrix A determines a symmetric bilinear form $(\cdot,\cdot)$ on $\mathbb{R}^3=\mathbb{R}\alpha_1+\mathbb{R}\alpha_2+\mathbb{R}\alpha_3$ defined by $(\alpha_i,\alpha_j)=A_{ij}$ for $i,j\in\{1,2,3\}$. It also determines a discrete group W, called the Weyl group, which is generated by reflections with respect to α_i for $i=1,2,3$. The connections with classical mathematics start from the fact that $W\approx \mathrm{PGL}_2(\mathbb{Z})$, one of the most extensively investigated discrete groups. The root lattice

$$Q=\mathbb{Z}\alpha_1+\mathbb{Z}\alpha_2+\mathbb{Z}\alpha_3 \tag{1.3}$$

and the weight lattice

$$P=\{\lambda\in\mathbb{R}^3 \mid (\lambda,\alpha_i)\in\mathbb{Z} \text{ for } i=1,2,3\} \tag{1.4}$$

also have nice realizations,

$$Q=\left\{\begin{pmatrix} a & b \\ b & c \end{pmatrix} \middle| \, a,b,c\in\mathbb{Z}\right\} \tag{1.5}$$

and

$$P=\left\{\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix} \middle| \, a,b,c\in\mathbb{Z}\right\}, \tag{1.6}$$

so that W acts in the usual way

$$w\cdot\lambda=w\lambda w^t \tag{1.7}$$

for $w\in W$ and $\lambda\in P$. The classification of W-orbits in P can then be seen to be equivalent to the classification of equivalence classes of integral binary quadratic forms

$$F(x,y)=ax^2+bxy+cy^2. \tag{1.8}$$

In the case of positive semidefinite forms the classification is given in terms of the dominant integral weights P^{++} in Sect. 2.

One of the most important results in the theory of affine Lie algebras is their explicit realization which gives a root multiplicity formula and clarifies their structure. In Sect. 4 we give a construction of the hyperbolic algebra $\mathfrak{F}$ which can be easily generalized: Let us denote by $\mathfrak{F}_0$ the affine Lie subalgebra of $\mathfrak{F}$ corresponding to the Cartan matrix $\begin{pmatrix} 2 & -2 \\ -2 & 2 \end{pmatrix}$. Let V be a basic representation of $\mathfrak{F}_0$ (see [10, 17]) and let V^* be the dual (contragredient) representation. Let $\mathfrak{F}_0^e$ be a one-dimensional extension of $\mathfrak{F}_0$ obtained by adjoining a certain derivation (see Sect. 3) and let $\{x_i \mid i\in I\}$ be an orthonormal basis of $\mathfrak{F}_0^e$ with respect to the

invariant symmetric bilinear form $(\cdot,\cdot)$ normalized so that $(\alpha_i,\alpha_i)=2$ for $i=1,2$. Then we define a $\mathbb{Z}$-graded Lie algebra

$$\mathfrak{G}=\ldots+[V^*,V^*]+V^*+\mathfrak{F}_0^e+V+[V,V]+\ldots, \tag{1.9}$$

where for $x\in\mathfrak{F}_0^e$, $v\in V$, $v^*\in V^*$ we have

$$[x,v]=x\cdot v, \qquad [x,v^*]=x\cdot v^* \tag{1.10}$$

and

$$[v^*,v]=-\sum_{i\in I}(v^*,x_i\cdot v)x_i. \tag{1.11}$$

This algebra has a unique maximal graded ideal

$$\mathfrak{I}=\sum_{n\in\mathbb{Z}}\mathfrak{I}_n \tag{1.12}$$

and the hyperbolic algebra $\mathfrak{F}$ is isomorphic to $\mathfrak{G}/\mathfrak{I}$.

A description of the ideal $\mathfrak{I}$ allows us to reduce the problem of determining root multiplicities of $\mathfrak{F}$ to the study of weight multiplicities of some representations of $\mathfrak{F}_0^e$. The weight multiplicities of $\mathfrak{F}_0^e$ are all one, and those of V are known [6, 16] to be values of the classical partition function, and in fact we have

$$\mathrm{Mult}\begin{pmatrix} n & 0\\ 0 & 1\end{pmatrix}=p(n+1), \tag{1.13}$$

where,

$$\sum_{n\geq 0}p(n)q^n=\prod_{n\geq 1}(1-q^n)^{-1}. \tag{1.14}$$

It can be seen that $\mathfrak{I}_{-1}$, $\mathfrak{I}_0$, $\mathfrak{I}_1$ are trivial, so we have complete knowledge of the root multiplicities from those graded pieces. We are able to compute the root multiplicities from $[V,V]/\mathfrak{I}_2$ as follows. We can show that the decomposition of the tensor product

$$V\otimes V=S(V)+A(V) \tag{1.15}$$

into symmetric and antisymmetric tensors corresponds to the decomposition of $V\otimes V$ into two "strings" of equivalent $\mathfrak{F}_0$-modules [see (3.42)]. The proof of this result uses the techniques of vertex operators studied in [8, 10]. We also show that $\mathfrak{I}_2\subset A(V)\approx[V,V]$ is exactly one irreducible standard $\mathfrak{F}_0^e$-module, and using the outer multiplicity formula for $V\otimes V$ from [5] we obtain the following result. We have

$$\mathrm{Mult}\begin{pmatrix} n & 0\\ 0 & 2\end{pmatrix}=p'(2n+1) \tag{1.16}$$

and

$$\mathrm{Mult}\begin{pmatrix} n & 1\\ 1 & 2\end{pmatrix}=p'(2n), \tag{1.17}$$

where $p'(n)$ is a modified partition function

$$\sum_{n\geq 0}p'(n)t^n=\left[\prod_{n\geq 1}(1-t^n)^{-1}\right](1-t^{20}+t^{22}-t^{24}+\ldots) \tag{1.18}$$

[see (4.36)]. It is most remarkable that the first nonzero term beyond 1 in the second factor on the right side of (1.18) ist t^{20}, so that $p'(n)=p(n)$ for $0\leqq n\leqq 19$ but $p'(20)\neq p(20)$. Theoretically our construction of $\mathfrak{F}$ should allow us to find all root multiplicities by studying the weight multiplicities of certain $\mathfrak{F}_0^e$-submodules in tensor powers of V. We do not, however, have a closed formula for the multiplicity of an arbitrary root of $\mathfrak{F}$ other than the general formula given in [36].

A complete knowledge of the root multiplicities of $\mathfrak{F}$ would yield a quite remarkable identity which comes from the Weyl-Kac denominator formula for $\mathfrak{F}$. We recall first that this formula for $\mathfrak{F}_0^e$ is the classical Jacobi triple product identity which may be written as

$$\begin{aligned} J_0(z,\tau) &= \sum_{j\in\mathbb{Z}} (-1)^j e^{4\pi i(j+1/2)z} q^{(j+1/2)^2} \\ &= 2i\sin(2\pi z)\prod_{n\geqq 1}(1-q^{2n})(1-q^{2n}e^{4\pi iz})(1-q^{2n}e^{-4\pi iz}), \end{aligned} \tag{1.19}$$

where $q=e^{\pi i\tau}$ for $\mathrm{Im}(\tau)>0$ and $z\in\mathbb{C}$. The denominator formula for $\mathfrak{F}$ is

$$\begin{aligned} &\sum_{g\in \mathbf{PGL}_2(\mathbb{Z})} \det(g) e^{2\pi i \,\mathrm{Tr}((gPg^t-P)\mathfrak{Z})} \\ &= \prod_{0\leqq N\in S_2(\mathbb{Z})} (1-e^{2\pi i\,\mathrm{Tr}(N\mathfrak{Z})})^{\mathrm{Mult}(N)} \prod_{\substack{\det(N)=-1 \\ N\in v(R^-)}} (1-e^{2\pi i\,\mathrm{Tr}(N\mathfrak{Z})}), \end{aligned} \tag{1.20}$$

where $P=\begin{pmatrix} 3 & 1/2 \\ 1/2 & 2 \end{pmatrix}$, $\mathfrak{Z}=\begin{pmatrix} z_3 & z_1 \\ z_1 & z_2 \end{pmatrix}$, $S_2(\mathbb{Z})$ is the set of symmetric integral 2×2 matrices and $v(R^-)$ is the set of matrices in $S_2(\mathbb{Z})$ corresponding to the negative roots of $\mathfrak{F}$ (see Theorem 4.10). Undoubtedly this identity should have a direct analytic proof analogous to that of the famous Jacobi identity. We believe that such a proof would shed new light on the root multiplicities, $\mathrm{Mult}(N)$, perhaps showing that they have a number-theoretical meaning connected with ideal classes of imaginary quadratic fields.

The denominator $J_0(z,\tau)$ for $\mathfrak{F}_0^e$, trivially related to the classical Jacobi theta function

$$\Theta_1(z,\tau) = -i\sum_{n\in\mathbb{Z}} (-1)^n e^{2\pi i(n+1/2)z} q^{(n+1/2)^2} \tag{1.21}$$

by the formula

$$J_0(z,\tau) = i\Theta_1(2z,\tau), \tag{1.22}$$

has another remarkable property besides the product expansion (1.19). If we define

$$J(z,t,\tau) = J_0(z,\tau)e^{4\pi it} \tag{1.23}$$

then we have the following simple transformation formula for J under the involution $\tau\to\dfrac{-1}{\tau}$,

$$J(z,t,\tau) = c\tau^{-1/2} J\left(\frac{-z}{\tau}, t-\frac{z^2}{\tau}, \frac{-1}{\tau}\right), \tag{1.24}$$

where c is a certain root of unity (see Corollary 5.2). Together with a transformation formula for J under the translation $\tau\to\tau+1$, this gives a formula

for any transformation $\tau \to \frac{a\tau+b}{c\tau+d}$ where $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{PSL}_2(\mathbb{Z})$. It is an open interesting problem to determine the transformation properties of the denominator (1.20) of $\mathfrak{F}$ under the symplectic involution $\mathfrak{Z} \to -\mathfrak{Z}^{-1}$. We note that together with the group $W \ltimes Q \approx \mathrm{PGL}_2(\mathbb{Z}) \ltimes \mathbb{Z}^3$ this involution generates the symplectic group $\mathrm{Sp}_2(\mathbb{Z})$, and the invariance of (1.20) under $W \ltimes Q$ is apparent. The role of these involutions is not well understood from the Lie-theoretical point of view, although some progress in this direction has been made. Kac and Peterson in [18] indicate that the characters of standard irreducible $\mathfrak{F}_0^e$-modules are invariant with respect to certain congruence subgroups of $\mathrm{PSL}_2(\mathbb{Z})$. In Sect. 5 we study the space $\mathscr{M}_k$ of weight k $\mathrm{PSL}_2(\mathbb{Z})$-invariant $\mathfrak{F}_0^e$-characters, and its natural subspace $\mathscr{M}_k'$. We say that an $\mathfrak{F}_0^e$-module (resp., character) is $\mathfrak{F}$-dominant if its decomposition into standard irreducible modules (resp., characters) consists only of those with highest weights which are $\mathfrak{F}$-dominant. The space $\mathscr{M}_k'$ consists of the $\mathfrak{F}$-dominant characters from $\mathscr{M}_k$. We show in Sect. 7 an isomorphism between $\mathscr{M}_k'$ and the space $\mathfrak{M}_k^2$ of genus 2 Siegel modular forms of weight k, whose dimension was found by Igusa [13] to be

$$\dim(\mathfrak{M}_k^2) = \text{Cardinality}\{(a,b,c,d) \in \mathbb{Z}_+^4 \mid k = 4a+6b+10c+12d\}. \tag{1.25}$$

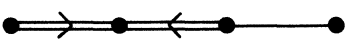
This fact implies that parallel to the correspondence between $\mathfrak{F}$-dominant standard $\mathfrak{F}_0^e$-modules and standard $\mathfrak{F}$-modules one has a correspondence between $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-characters which are $\mathrm{PSL}_2(\mathbb{Z})$-invariant and $\mathfrak{F}$-characters which are $\mathrm{Sp}_2(\mathbb{Z})$-invariant. The first correspondence is canonical, two modules being in correspondence if they have the same dominant integral highest weight, so one would expect the second correspondence to also be canonical. Such a canonical correspondence was known for functions of level 0 in the theory of Eisenstein series, and for functions of level 1 was recently studied by Maass [30–32] and others. In Sect. 7 we recall the Maass correspondence and give its generalization to higher levels.

The Maass space, which is a subspace of $\mathfrak{M}_k^2$, recently attracted considerable attention in the theory of Siegel modular forms because of its connection with the Saito-Kurokawa conjecture [2, 20, 22, 38, 43]. Our Lie-theoretical approach allows us to generalize the Maass correspondence to level 1 characters of any affine Lie algebra $\mathfrak{g}\hat{}$ and a specifically constructed Kac-Moody algebra $\mathfrak{g}\hat{}^{\hat{}}$ containing $\mathfrak{g}\hat{}$. We wish to mention one example of this generalization, where the affine algebra $\mathfrak{g}\hat{}$ is of type $C_2^{(1)}$ in the notation of Kac [16] having Dynkin diagram



(1.26)

and where the hyperbolic algebra $\mathfrak{g}\hat{}^{\hat{}}$ has Dynkin diagram



(1.27)

The Weyl group of $\mathfrak{g}\hat{}^{\hat{}}$ is isomorphic to the Klein-Fricke group Ψ_1^* which contains as a subgroup of index 4 the Picard group $\mathrm{PSL}_2(\mathbb{Z}[i])$. The semidirect product of the Weyl group with the root lattice, extended by the symplectic involution, generates $\mathrm{Sp}_2(\mathbb{Z}[i])$. The lifting of $\mathfrak{g}\hat{}^e$-characters on level 1 which are $\mathrm{PSL}_2(\mathbb{Z}[i])$-invariant by using Hecke operators in analogy with the work of Maass provides a

construction of $Sp_2(\mathbb{Z}[i])$-invariant Hermitian modular forms of genus 2. This has already been noticed by Kojima in [21]. Generalizations of these results will appear in subsequent publications.

We have recently received from Yoshida a preprint [42] in which he has worked out the straightforward classification of the finite number of rank 3 hyperbolic symmetrizable Kac-Moody algebras, and noticed that the Weyl groups of these algebras are all hyperbolic triangle groups. We had known this for some time and mentioned this observation in our talks at the 789th regional meeting of the American Mathematical Society. Yoshida has also seen that the semidirect product of the Weyl group and the root lattice will be isomorphic to a discrete subgroup of a parabolic subgroup of $Sp_2(\mathbb{R})$.

2. The Kac-Moody Algebra $\mathfrak{F}$

We will study the rank 3 hyperbolic Kac-Moody Lie algebra $\mathfrak{F}$ whose Cartan matrix is

$$A=(A_{ij})=\begin{pmatrix} 2 & -2 & 0 \\ -2 & 2 & -1 \\ 0 & -1 & 2 \end{pmatrix}. \tag{2.1}$$

We recall (see [14] or [12]) that $\mathfrak{F}$ is a Lie algebra with 9 generators e_i, f_i, h_i, $i=1,2,3$ and the following relations:

$$[h_i, h_j]=0, [e_i, f_j]=\delta_{ij}h_i, [h_i, e_j]=A_{ij}e_j, [h_i, f_j]=-A_{ij}f_j$$

for all $i,j\in\{1,2,3\}$, and (2.2)

$$(\mathrm{ad}\, e_i)^{-A_{ij}+1}e_j=0=(\mathrm{ad}\, f_i)^{-A_{ij}+1}f_j \quad \text{for} \quad i\neq j.$$

The subalgebra $\mathfrak{h}=\mathbb{C}h_1\oplus\mathbb{C}h_2\oplus\mathbb{C}h_3$ is called a Cartan subalgebra of $\mathfrak{F}$. We denote by α_i, $i=1,2,3$, the elements in the dual space $\mathfrak{h}^*$ defined by

$$\alpha_i(h_j)=A_{ij} \quad \text{for} \quad i,j\in\{1,2,3\}. \tag{2.3}$$

The algebra $\mathfrak{F}$ decomposes into eigenspaces under the simultaneous adjoint action of $\mathfrak{h}$. These eigenspaces of $\mathfrak{F}$ are called root spaces and correspond to certain elements of $\mathfrak{h}^*$ as follows. For $\alpha\in\mathfrak{h}^*$ define

$$\mathfrak{F}^\alpha=\{x\in\mathfrak{F}\mid [h,x]=\alpha(h)x, \text{ for all } h\in\mathfrak{h}\} \tag{2.4}$$

to be the α-root space of $\mathfrak{F}$. Clearly $\mathfrak{F}^0=\mathfrak{h}$. Define the root system of $\mathfrak{F}$ to be

$$R=\{0\neq\alpha\in\mathfrak{h}^*\mid \mathfrak{F}^\alpha\neq 0\}. \tag{2.5}$$

Clearly $\mathfrak{F}^{\alpha_i}=\mathbb{C}e_i$, $\mathfrak{F}^{-\alpha_i}=\mathbb{C}f_i$ and $\mathfrak{F}=\mathfrak{h}+\sum_{\alpha\in R}\mathfrak{F}^\alpha$. It is known that any α in R can be written as an integral linear combination of α_1, α_2, α_3 with all coefficients either nonnegative or nonpositive. In the first case we say that α is a positive root and we denote these by R^+. Then $R^-=-R^+$ and $R=R^+\cup R^-$.

It is known that $\mathfrak{F}$ is simple [11] and has an invariant bilinear form $(\cdot,\cdot)$, i.e. for any $x,y,z\in\mathfrak{F}$, $([x,y],z)=(x,[y,z])$. This form is uniquely determined up to a scalar

multiple. We choose this multiple so that $(h_i, h_i)=2$ for $i=1,2,3$. The set $\{\alpha_1, \alpha_2, \alpha_3\}$ of simple roots is a basis for $\mathfrak{h}^*$. Another canonical basis of $\mathfrak{h}^*$ consists of the fundamental weights ω_1, ω_2, ω_3 defined by

$$\omega_i(h_j)=\delta_{ij} \quad \text{for} \quad i,j\in\{1,2,3\}. \tag{2.6}$$

We define the root lattice of $\mathfrak{F}$ to be

$$Q=\mathbb{Z}\alpha_1\oplus\mathbb{Z}\alpha_2\oplus\mathbb{Z}\alpha_3 \tag{2.7}$$

and the weight lattice of $\mathfrak{F}$ to be

$$P=\mathbb{Z}\omega_1\oplus\mathbb{Z}\omega_2\oplus\mathbb{Z}\omega_3. \tag{2.8}$$

We also define the sets

$$Q^+=\mathbb{Z}^+\alpha_1\oplus\mathbb{Z}^+\alpha_2\oplus\mathbb{Z}^+\alpha_3, \tag{2.9}$$

$$Q^-=-Q^+, \tag{2.10}$$

and

$$P^{++}=\mathbb{Z}^+\omega_1\oplus\mathbb{Z}^+\omega_2\oplus\mathbb{Z}^+\omega_3. \tag{2.11}$$

P^{++} is called the set of dominant weights of $\mathfrak{F}$.

We identify $\mathfrak{h}$ with $\mathfrak{h}^*$ by the isomorphism η determined by

$$\eta(h_i)=\alpha_i \quad \text{for} \quad i=1,2,3. \tag{2.12}$$

Thus, $\mathfrak{h}^*$ has a bilinear form, which we continue to denote by $(\cdot,\cdot)$.

We prefer to use the following basis for $\mathfrak{h}^*$;

$$\gamma_1^*=\alpha_1/2, \quad \gamma_2^*=-\alpha_1-\alpha_2-\alpha_3, \quad \gamma_3^*=-\alpha_1-\alpha_2. \tag{2.13}$$

For this basis the matrix of the form is

$$((\gamma_i^*,\gamma_j^*))=\begin{pmatrix} 1/2 & 0 & 0 \\ 0 & 0 & -1 \\ 0 & -1 & 0 \end{pmatrix}. \tag{2.14}$$

We have

$$\omega_1=\gamma_1^*+\gamma_2^*+\gamma_3^*, \quad \omega_2=\gamma_2^*+\gamma_3^*, \quad \omega_3=\gamma_3^*, \tag{2.15}$$

and

$$\alpha_1=2\gamma_1^*, \quad \alpha_2=-2\gamma_1^*-\gamma_3^*, \quad \alpha_3=\gamma_3^*-\gamma_2^*. \tag{2.16}$$

We also see that

$$Q=\{2n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^* \mid n_1,n_2,n_3\in\mathbb{Z}\}, \tag{2.17}$$

$$Q^+=\{2n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*\in Q \mid n_1\geqq n_2+n_3, n_2+n_3\leqq 0, n_2\leqq 0\}, \tag{2.18}$$

$$P=\{n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^* \mid n_1,n_2,n_3\in\mathbb{Z}\}, \tag{2.19}$$

and

$$P^{++}=\{n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*\in P \mid n_3\geqq n_2\geqq n_1\geqq 0\}. \tag{2.20}$$

The Weyl group W acting on $\mathfrak{h}^*$ is generated by r_1, r_2, r_3 whose action on $\mathfrak{h}^*$ is given by

$$r_i(\alpha_j)=\alpha_j-A_{ij}\alpha_i. \tag{2.21}$$

It is known that W is a Coxeter group with relations

$$r_1^2=r_2^2=r_3^2=(r_1r_3)^2=(r_2r_3)^3=1. \tag{2.22}$$

Therefore, W is isomorphic to the extended modular group (see [34, p. 111]), which we think of as $\mathrm{PGL}_2(\mathbb{Z})$. The even subgroup $W^+\subset W$ generated by r_2r_1 and r_1r_3 is then isomorphic to the modular group $\Gamma=\mathrm{PSL}_2(\mathbb{Z})$. An explicit description of the isomorphism will be given below.

The set of real roots is defined to be

$$R_W=W\cdot\{\alpha_1,\alpha_2,\alpha_3\} \tag{2.23}$$

and the set of imaginary roots is then

$$R_I=\{\alpha\in R\mid\alpha\notin R_W\}. \tag{2.24}$$

We also have positive (and negative) real and imaginary roots,

$$R_W^\pm=R_W\cap R^\pm,\qquad R_I^\pm=R_I\cap R^\pm, \tag{2.25}$$

and

$$R^\pm=R\cap Q^\pm,\qquad R_W^\pm=R_W\cap Q^\pm,\qquad R_I^\pm=R_I\cap Q^\pm. \tag{2.26}$$

Moody has shown [35] that for hyperbolic algebras

$$R_I=\{\alpha\in Q\mid(\alpha,\alpha)\leqq 0\}. \tag{2.27}$$

With respect to the basis (2.13) the generators of W act as follows:

$$\begin{aligned}
r_1(z_1\gamma_1^*+z_2\gamma_2^*+z_3\gamma_3^*)&=-z_1\gamma_1^*+z_2\gamma_2^*+z_3\gamma_3^*,\\
r_2(z_1\gamma_1^*+z_2\gamma_2^*+z_3\gamma_3^*)&=(-z_1+2z_2)\gamma_1^*+z_2\gamma_2^*+(-z_1+z_2+z_3)\gamma_3^*,\\
r_3(z_1\gamma_1^*+z_2\gamma_2^*+z_3\gamma_3^*)&=z_1\gamma_1^*+z_3\gamma_2^*+z_2\gamma_3^*.
\end{aligned} \tag{2.28}$$

The elements in P, Q, and W can be realized as 2×2 matrices. Let $S_2(\mathbb{C})$ denote the complex symmetric 2×2 matrices and let $S_2(\mathbb{Z})$ denote the integral symmetric 2×2 matrices. We will also use

$$S_2'(\mathbb{Z})=\left\{\begin{pmatrix} n_2 & n_1/2\\ n_1/2 & n_3\end{pmatrix}\,\middle|\, n_1,n_2,n_3\in\mathbb{Z}\right\}\supset S_2(\mathbb{Z}). \tag{2.29}$$

Each of these sets is invariant under the following action of $\mathrm{PGL}_2(\mathbb{Z})$;

$$g\cdot x=gxg^t\quad\text{for}\quad g\in\mathrm{PGL}_2(\mathbb{Z}). \tag{2.30}$$

Define the map $v:\mathfrak{h}^*\to S_2(\mathbb{C})$ by

$$v(z_1\gamma_1^*+z_2\gamma_2^*+z_3\gamma_3^*)=\begin{pmatrix} z_3 & z_1/2\\ z_1/2 & z_2\end{pmatrix}. \tag{2.31}$$

We also define a group homomorphism $\bar{v}: W \to \mathrm{PGL}_2(\mathbb{Z})$ which is uniquely determined by $\bar{v}(r_i) = \varrho_i$ for $i = 1, 2, 3$ where

$$\varrho_1 = \pm \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, \quad \varrho_2 = \pm \begin{pmatrix} -1 & 1 \\ 0 & 1 \end{pmatrix}, \quad \varrho_3 = \pm \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{2.32}$$

are generators of $\mathrm{PGL}_2(\mathbb{Z})$. Also, $\bar{v}: W^+ \to \mathrm{PSL}_2(\mathbb{Z})$.

Proposition 2.1. (a) *The map v defines a vector space isomorphism $\mathfrak{h}^* \approx S_2(\mathbb{C})$ and lattice isomorphisms $P \approx S_2'(\mathbb{Z})$ and $Q \approx S_2(\mathbb{Z})$ so that $S_2(\mathbb{C})$ has the bilinear form $(\cdot, \cdot)$,*

(b) *the map $\bar{v}$ gives group isomorphisms*

$$PGL_2(\mathbb{Z}) \approx W \quad \text{and} \quad \mathrm{PSL}_2(\mathbb{Z}) \approx W^+,$$

(c) *for any $\lambda \in \mathfrak{h}^*$, $w \in W$ we have*

$$v(w \cdot \lambda) = \bar{v}(w) \cdot v(\lambda), \tag{2.33}$$

(d) *the form $(\cdot, \cdot)$ on $S_2(\mathbb{C})$ is invariant under $\mathrm{PGL}_2(\mathbb{Z})$,*

(e)

$$v(R_I) = \{x \in S_2(\mathbb{Z}) \mid \det x \geqq 0\},$$

$$v(R_I^-) = \left\{ \begin{pmatrix} n_3 & n_1 \\ n_1 & n_2 \end{pmatrix} \in S_2(\mathbb{Z}) \mid n_2 n_3 \geqq n_1^2, n_2 \geqq 0, n_3 \geqq 0 \right\},$$

$$v(R_W) = \{x \in S_2(\mathbb{Z}) \mid \det x = -1\}.$$

Proof. Everything is straightforward except for the fact that $\{x \in S_2(\mathbb{Z}) \mid \det x = -1\} \subseteq v(R_W)$. Note first that $r_2 r_3(\alpha_2) = \alpha_3$, so actually we have from (2.23)

$$v(R_W) = \mathrm{PGL}_2(\mathbb{Z}) \cdot \{v(\alpha_1), v(\alpha_3)\}. \tag{2.34}$$

Any element $g \in \mathrm{PGL}_2(\mathbb{Z})$ may be uniquely written as

$$g = \begin{pmatrix} a & b \\ b & d \end{pmatrix} = \begin{pmatrix} \dfrac{\alpha+\beta}{2} & \dfrac{\beta-\alpha}{2} \\ \dfrac{\gamma+\delta}{2} & \dfrac{\delta-\gamma}{2} \end{pmatrix}, \tag{2.35}$$

where $a, b, c, d \in \mathbb{Z}$, $ad - bc = \pm 1$, $\alpha, \beta, \gamma, \delta \in \mathbb{Z}$, $\alpha\delta - \beta\gamma = \pm 2$, $\alpha \equiv \beta \pmod 2$ and $\gamma \equiv \delta \pmod 2$. Also note that

$$\begin{aligned} g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g^t &= \begin{pmatrix} 2ab & ad+bc \\ ad+bc & 2cd \end{pmatrix}, \\ g \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g^t &= \begin{pmatrix} a^2-b^2 & ac-bd \\ ac-bd & c^2-d^2 \end{pmatrix} = \begin{pmatrix} \alpha\beta & (\alpha\delta+\beta\gamma)/2 \\ (\alpha\delta+\beta\gamma)/2 & \gamma\delta \end{pmatrix}. \end{aligned} \tag{2.36}$$

From (2.34) it is clear that we have

$$v(R_W) = \left\{ g \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g^t \mid g \in \mathrm{PGL}_2(\mathbb{Z}) \right\} \cup \left\{ g \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g^t \mid g \in \mathrm{PGL}_2(\mathbb{Z}) \right\}. \tag{2.37}$$

Let $N=\begin{pmatrix} n_1 & n \\ n & n_2 \end{pmatrix} \in S_2(\mathbb{Z})$ have $\det(N)=-1$. One can easily find $g_1, g_2 \in \mathrm{PGL}_2(\mathbb{Z})$ such that if $n_1, n_2 \in 2\mathbb{Z}$, $N=g_1\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g_1^t$, and if $n_1 \notin 2\mathbb{Z}$ or $n_2 \notin 2\mathbb{Z}$, $N=g_2\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g_2^t$.

Definition. Let $R_W^i=\{gv(\alpha_i)g^t \mid g \in \mathrm{PGL}_2(\mathbb{Z})\}$ for $i=1,2,3$. Then we have

$$R_W^1=\left\{g\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} g^t \mid g \in \mathrm{PGL}_2(\mathbb{Z})\right\} \quad \text{and} \quad R_W^2=R_W^3=\left\{g\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g^t \mid g \in \mathrm{PGL}_2(\mathbb{Z})\right\}.$$

Proposition 2.2. *Let $N \in R_W^i$, $i=1$ or 3, then N has exactly one presentation in the form $N=gv(\alpha_i)g^t$ for $\det(g)=+1$ and exactly one presentation in the form $N=gv(\alpha_i)g^t$ for $\det(g)=-1$.*

Proof. From (2.36) one may see that if $gv(\alpha_i)g^t=v(\alpha_i)$ then either $g=1$ or $g=v(\alpha_i)$. So from $N=g_1v(\alpha_i)g_1^t=g_2v(\alpha_i)g_2^t$ one has $g_2^{-1}g_1=1$ or $g_2^{-1}g_1=v(\alpha_i)$. Either $g_1=g_2$ or else $g_1=g_2v(\alpha_i)$.

Definition. Let

$$R_I^0=\left\{g\begin{pmatrix} m & 0 \\ 0 & 0 \end{pmatrix} g^t \mid 0 \neq m \in \mathbb{Z}, g \in \mathrm{PGL}_2(\mathbb{Z})\right\}.$$

Proposition 2.3. *There is a one-to-one correspondence between the set $(\mathbb{Q} \cup \infty) \times (\mathbb{Z}-\{0\})$ and R_I^0.*

Proof. We have that

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}\begin{pmatrix} m & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} a & c \\ b & d \end{pmatrix}=\begin{pmatrix} a^2m & acm \\ acm & c^2m \end{pmatrix},$$

where $(a,c)=1$, so that

$$R_I^0=\left\{\begin{pmatrix} a^2m & acm \\ acm & c^2m \end{pmatrix} \middle| (a,c)=1, 0 \neq m \in \mathbb{Z}\right\}.$$

But $\mathbb{Q} \cup \infty$ is in one-to-one correspondence with $\{(a;c) \in \mathbb{Z} \times \mathbb{Z} \mid (a,c)=1\}$ modulo the relation $(a;c)=(-a;-c)$. The proposition is now clear.

Definition. Let $\tilde{\mathbb{Q}}=\mathbb{Q} \cup \infty$. For each $q \in \tilde{\mathbb{Q}}$ let

$$R_I^0(q)=\left\{\begin{pmatrix} a^2m & acm \\ acm & c^2m \end{pmatrix} \middle| a, c \in \mathbb{Z}, (a,c)=1, q=a/c, 0 \neq m \in \mathbb{Z}\right\}$$

and let

$$R_W(q)=\left\{N=\begin{pmatrix} n_1 & n \\ n & n_2 \end{pmatrix} \in R_W \mid n_1c^2+n_2a^2=2nac \text{ for } q=a/c, (a,c)=1, a, c \in \mathbb{Z}\right\}.$$

Since the equation of the plane tangent to the cone $z_1z_2-z^2=0$ through the point (z_1^0, z_2^0, z^0) is just $z_1z_2^0+z_2z_1^0-2zz^0=0$, the definition of $R_W(q)$ has a clear geometrical motivation.

Definition. Let $R(q)=R_W(q)\cup R_I^0(q)$ for any $q\in\hat{\mathbb{Q}}$. If $g=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in \mathrm{PGL}_2(\mathbb{Z})$ and $q=A/C\in\hat{\mathbb{Q}}$ then define the action $g\cdot q=\dfrac{aA+bC}{cA+dC}$.

It is easy to see that

$$R_W(g\cdot q)=g\cdot R_W(q)=gR_W(q)g^t. \tag{2.38}$$

Proposition 2.4. *For any $q\in\hat{\mathbb{Q}}$ the set $R(q)$ corresponds by v to a set of points in $\mathfrak{h}^*$ which form an affine root system of type $A_1^{(1)}$.*

Proof. See Sect. 3 and [23, 25] for details about the affine algebra of type $A_1^{(1)}$ and its root system. For our present purposes it suffices to know that such a root system is of the form

$$\{mv_0\pm v_1\,|\,m\in\mathbb{Z}\}\cup\{mv_0\,|\,0\neq m\in\mathbb{Z}\}, \tag{2.39}$$

where v_0 and v_1 are fixed vectors in $\mathfrak{h}^*$. From (2.36) and Proposition 2.1, and since for $q=\infty$ we have $a=1$ and $c=0$, we see that

$$R(\infty)=\left\{\begin{pmatrix}m & \pm 1\\ \pm 1 & 0\end{pmatrix}\,\middle|\,m\in\mathbb{Z}\right\}\cup\left\{\begin{pmatrix}m & 0\\ 0 & 0\end{pmatrix}\,\middle|\,0\neq m\in\mathbb{Z}\right\}. \tag{2.40}$$

We will see later that this corresponds to the root system of the subalgebra $\mathfrak{F}_0$ of type $A_1^{(1)}$ generated by α_1 and α_2. The general result follows from the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $\hat{\mathbb{Q}}$.

Proposition 2.5. *Each element $N\in v(R_W)$ is contained in exactly two sets $R_W(q_1)$ and $R_W(q_2)$ for $q_1,q_2\in\hat{\mathbb{Q}}$.*

Proof. This follows from Proposition 2.2.

The most subtle aspect of root systems of Kac-Moody algebras is the determination of the root multiplicities. The construction of algebra $\mathfrak{F}$ given in Sect. 4 provides a significant amount of such information.

In contrast to the cases of finite dimensional and affine algebras where the union of the weights of standard representations fill up the entire weight lattice P, for hyperbolic algebras they only fill up some positive cone which one denotes P^+.

Definition. Let $P^+\subset P$ be defined by

$$v(P^+)=\{N\in S_2'(\mathbb{Z})\,|\,N\geqq 0\}. \tag{2.41}$$

Within the set P^+ we will now study the orbits under the Weyl group W. This problem was studied by Lagrange in the language of binary quadratic forms.

Lagrange considered classes of integral binary quadratic forms

$$F(x,y)=ax^2+bxy+cy^2 \quad \text{for} \quad a,b,c\in\mathbb{Z} \tag{2.42}$$

under the action of $\mathrm{PSL}_2(\mathbb{Z})$. There are finitely many classes of forms with given discriminant D. Each $N\in v(P^+)$ corresponds to a positive semi-definite form with discriminant $D\leqq 0$. If $\det(N)=0=D$ it is easy to find the $\mathrm{PSL}_2(\mathbb{Z})$-equivalent form $F_1(x,y)=nx^2$ for unique $n\in\mathbb{Z}^+$, so N is $\mathrm{PSL}_2(\mathbb{Z})$-conjugate to a unique element of $v(P^{++})$. To study the Weyl group orbits in P^+ it is clearly sufficient to study those

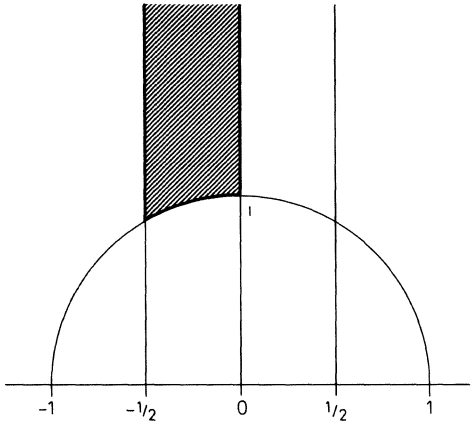


Fig. 1. Fundamental domain for $\mathrm{PGL}_2(\mathbb{Z})$

$0<N\in v(P^+)$ which are primitive. One has the well-known bijection between the set of all similarity classes of full modules in imaginary quadratic field $\mathbb{Q}(\sqrt{d})$ having fixed coefficient ring $\mathfrak{O}_f$ and the set of all equivalence classes of positive definite integral primitive binary quadratic forms of discriminant $D=df^2<0$. (See [3] for details.) A representative module associated with $N=\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ can be taken with basis

$$\left\{1, \gamma=\frac{-b+\sqrt{b^2-4ac}}{2a}\right\}, \tag{2.43}$$

where $\mathrm{Im}(\gamma)>0$. Using the action of $g=\begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix}\in\mathrm{PGL}_2(\mathbb{Z})$ on the upper half-plane

$$g\cdot z=\begin{cases} \dfrac{\alpha z+\beta}{\gamma z+\delta} & \text{if} \quad \det(g)=+1 \\ \dfrac{\alpha\bar{z}+\beta}{\gamma\bar{z}+\delta} & \text{if} \quad \det(g)=-1 \end{cases} \tag{2.44}$$

one gets as fundamental domain

$$\{\gamma\in\mathbb{C}\,|\,\mathrm{Im}(\gamma)>0,\ -1/2\leqq\mathrm{Re}(\gamma)\leqq 0,\ |\gamma|\geqq 1\} \tag{2.45}$$

(see Fig. 1). These conditions applied to (2.43) give the conditions $c\geqq a\geqq b\geqq 0$ on N. Applying reflection r_3 we see that any $N\in v(P^+)$ is $\mathrm{PGL}_2(\mathbb{Z})$-conjugate to a unique element of $v(P^{++})$.

Proposition 2.6. *We have $P^+=W\cdot P^{++}$ and every W-orbit in P^+ contains exactly one element from P^{++}.*

Remark. The map T from $v(P^+)$ to the extended complex upper half-plane given by

$$T\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}=\frac{-b+\sqrt{b^2-4ac}}{2a} \tag{2.46}$$

is such that

$$T(xNx^t)=x\cdot T(N) \quad \text{for} \quad x\in \mathrm{PGL}_2(\mathbb{Z}). \tag{2.47}$$

This gives a computational proof of Proposition 2.6, and has been noticed by Yoshida [42] and used to investigate compactifications of certain quotient spaces.

3. Standard Representations and Characters of $\mathfrak{F}$ and Affine Subalgebra $\mathfrak{F}_0$

Let $\mathscr{E}$ be the space of all complex valued functions on P^+. Then $\mathscr{E}$ is a commutative algebra whose product is the convolution operation

$$(f_1 f_2)(\lambda)=\sum_{\mu\in P^+} f_1(\lambda-\mu)f_2(\mu). \tag{3.1}$$

Only finitely many terms in this sum are nonzero because P^+ is contained in half of a cone.

Let $e^\lambda\in\mathscr{E}$ be the function defined for any $\lambda\in P^+$ by

$$e^\lambda(\mu)=\begin{cases}1 & \text{if} \quad \mu=\lambda\\ 0 & \text{if} \quad \mu\neq\lambda.\end{cases} \tag{3.2}$$

Then $e^\lambda e^\mu=e^{\lambda+\mu}$ and $e^0=1$ is the unit element in $\mathscr{E}$.

Let $\mathscr{C}$ be the category of $\mathfrak{F}$-modules V such that

$$V=\oplus\sum_{\mu\in P^+} V_\mu \quad \text{for} \quad \dim V_\mu<\infty. \tag{3.3}$$

Here we are denoting by

$$V_\mu=\{v\in V \mid h\cdot v=\mu(h)v, \text{ for all } h\in\mathfrak{h}\} \tag{3.4}$$

the μ-weight space of V. It is easy to see that $\mathfrak{F}^\alpha\cdot V_\mu\subseteq V_{\mu+\alpha}$ [recall the definition (2.4) of $\mathfrak{F}^\alpha$]. It is known that the action of the generators e_i and f_i [see (2.2)], $i=1,2,3$ on V is locally nilpotent. Kac has shown in [16] that any $\mathfrak{F}$-module V in $\mathscr{C}$ is completely reducible. Any irreducible module in $\mathscr{C}$ has a dominant integral highest weight $\lambda\in P^{++}$. There is, in fact, a one-to-one correspondence between P^{++} and the collection of irreducible modules in $\mathscr{C}$ [15]. Thus, we label these so called standard modules, V^λ.

The character of any $\mathfrak{F}$-module V in $\mathscr{C}$ is defined to be

$$X(V)=\sum_{\mu\in P^+}\dim(V_\mu)e^\mu. \tag{3.5}$$

We usually denote $X(V^\lambda)$ by X^λ. Clearly $X(V)\in\mathscr{E}$.

Kac has established in [15] the character formula for V^λ analogous to the Weyl character formula for finite dimensional modules of finite dimensional semisimple Lie algebras.

Theorem 3.1. (Weyl-Kac character formula). *For $\lambda\in P^{++}$ we have*

$$X^\lambda=\frac{\sum_{w\in W}\det(w)e^{w(\lambda+\varrho)}}{\sum_{w\in W}\det(w)e^{w\varrho}}, \tag{3.6}$$

where $\varrho=\omega_1+\omega_2+\omega_3\in\mathfrak{h}^*$ *and where*

$$\det(w)=\begin{cases}+1 & \text{if} \quad w\in W^+\\ -1 & \text{if} \quad w\notin W^+.\end{cases} \tag{3.7}$$

Let $\mathfrak{F}_0$ denote the type $A_1^{(1)}$ Kac-Moody subalgebra of $\mathfrak{F}$ with generators e_i, f_i, h_i, $i=1, 2$ and relations (2.2) whose Cartan matrix is therefore

$$\begin{pmatrix} 2 & -2\\ -2 & 2\end{pmatrix}. \tag{3.8}$$

It is known that this algebra is isomorphic to a one dimensional central extension, $\mathfrak{g}_0=\widetilde{\mathfrak{sl}_2(\mathbb{C})}\oplus\mathbb{C}c$, of

$$\widetilde{\mathfrak{sl}_2(\mathbb{C})}=\mathfrak{sl}_2(\mathbb{C})\otimes\mathbb{C}[t, t^{-1}]. \tag{3.9}$$

In order to overcome a problem caused by the degeneracy of the Cartan matrix (3.8) one can consider a bigger algebra, the semidirect product of $\mathfrak{g}_0$ with the one dimensional space spanned by the derivation $d=-t\frac{d}{dt}$. If we let

$$\mathfrak{g}_0^e=\widetilde{\mathfrak{sl}_2(\mathbb{C})}\oplus\mathbb{C}c\oplus\mathbb{C}d \tag{3.10}$$

be that extended algebra then the bracket products are as follows:

$$\begin{aligned}[x\otimes t^n, y\otimes t^m]&=[x, y]\otimes t^{n+m}+n\langle x, y\rangle\delta_{n,-m}c\\ [d, x\otimes t^n]&=-n(x\otimes t^n),\end{aligned} \tag{3.11}$$

and

where $x, y\in\mathfrak{sl}_2(\mathbb{C})$, $n, m\in\mathbb{Z}$, $\langle x, y\rangle=\frac{1}{4}\mathrm{Tr}(\mathrm{ad}_x\mathrm{ad}_y)$ is the Killing form on $\mathfrak{sl}_2(\mathbb{C})$ normalized so that $\langle h, h\rangle=2$, and c acts centrally.

We will identify $\mathfrak{g}_0^e$ as a subalgebra of $\mathfrak{F}$, $\mathfrak{F}_0^e\supset\mathfrak{F}_0$ as follows:

$$h\otimes 1=h_1, \quad c=h_1+h_2, \quad d=h_1+h_2+h_3, \tag{3.12}$$

$$e\otimes 1=e_1, \quad f\otimes 1=f_1, \quad f\otimes t=e_2, \quad e\otimes t^{-1}=f_2, \tag{3.13}$$

where $\{e, f, h\}$ is the usual basis of $\mathfrak{sl}_2(\mathbb{C})$. It is straightforward to verify the required relations. Note that (3.12) is the basis of $\mathfrak{h}_0^e$ identified with the basis of $\mathfrak{h}^*$

$$\gamma_1=\alpha_1, \quad \gamma_2=\alpha_1+\alpha_2, \quad \gamma_3=\alpha_1+\alpha_2+\alpha_3 \tag{3.14}$$

by the isomorphism η in (2.12). So the extended Cartan subalgebra $\mathfrak{h}_0^e\subset\mathfrak{g}_0^e$ has dual space $(\mathfrak{h}_0^e)^*\approx\mathfrak{h}^*$.

Let us use the following notation for elements of $\mathfrak{g}_0^e$:

$$e\otimes t^k=e(k), \quad f\otimes t^k=f(k), \quad h\otimes t^k=h(k), \tag{3.15}$$

for all $k\in\mathbb{Z}$. The Killing form $\langle\cdot,\cdot\rangle$ on $\mathfrak{sl}_2(\mathbb{C})$ extends uniquely to an invariant nondegenerate symmetric bilinear form $\langle\cdot,\cdot\rangle$ on $\mathfrak{g}_0^e$ such that $\langle h_i, h_i\rangle=2$ for $1\leqq i\leqq 3$. One may easily compute that

$$\langle e(k), f(-k)\rangle=1, \quad \langle h(k), h(-k)\rangle=2, \quad \langle c, d\rangle=-1 \tag{3.16}$$

and all other values of the form on pairs of vectors from basis

$$\{e(k), f(k), h(k), c, d | k\in\mathbb{Z}\}$$

are zero. This follows from invariance and the fact that $\langle x, y\rangle = 0$ whenever x is an α root vector, y is a β root vector and $\alpha+\beta \neq 0$. We also know that $e(k)$ is a $(k+1)\alpha_1 + k\alpha_2$ root vector, $f(k)$ is a $(k-1)\alpha_1 + k\alpha_2$ root vector and $h(k)$ is a $k\alpha_1 + k\alpha_2$ root vector for all $k \in \mathbb{Z}$.

From (3.16) it is easy to see that the union of the following three sets of vectors is an orthonormal basis for $\mathfrak{g}_0^e$ with respect to form $\langle \cdot, \cdot \rangle$;

$$\left\{\frac{1}{\sqrt{2}}(e(k)+f(-k)), \frac{1}{\sqrt{-2}}(e(k)-f(-k)) \mid k \in \mathbb{Z}\right\}, \tag{3.17}$$

$$\left\{\frac{1}{2}(h(k)+h(-k)), \frac{1}{2\sqrt{-1}}(h(k)-h(-k)) \mid 0 < k \in \mathbb{Z}\right\}, \tag{3.18}$$

$$\left\{\frac{1}{\sqrt{2}}h(0), \frac{1}{\sqrt{-2}}(c+d), \frac{1}{\sqrt{2}}(c-d)\right\}. \tag{3.19}$$

Also, (3.19) is an orthonormal basis for $\mathfrak{h}_0^e$.

The results of Kac mentioned previously, e.g. classification of standard modules and the character formula, are quite general and apply as well to $\mathfrak{F}_0^e$. The theory of representations for the algebra $\mathfrak{F}_0$ requires only the ad-hoc adjunction of derivation d, but there appears to be some value in the consideration of $\mathfrak{F}_0$ in the context of $\mathfrak{F}$. However, the condition that an element $\lambda = n_1\gamma_1^* + n_2\gamma_2^* + t\gamma_3^* \in \mathfrak{h}^*$ be integral (respectively, dominant integral) for the algebra $\mathfrak{F}_0^e$ means only that $n_1, n_2 \in \mathbb{Z}$ (respectively, $n_1, n_2 \in \mathbb{Z}^+$) while t may vary over $\mathbb{C}$.

Fundamental weights of $\mathfrak{F}_0^e$ are defined by $\lambda_i(h_j) = \delta_{ij}$ for $i = 1, 2$, which imposes no condition on the γ_3^* coefficient of λ_i, so there is a one parameter family of such weights. Define

$$\omega_1^t = \gamma_1^* + \gamma_2^* + t\gamma_3^*, \tag{3.20}$$

$$\omega_2^t = \gamma_2^* + t\gamma_3^*, \tag{3.21}$$

and

$$\omega_3^t = t\gamma_3^*. \tag{3.22}$$

The weights ω_3^t are the dominant weights of $\mathfrak{F}_0^e$ whose standard modules are the one dimensional trivial modules. Note that

$$\omega_1^1 = \omega_1, \quad \omega_2^1 = \omega_2 \quad \text{and} \quad \omega_3^1 = \omega_3. \tag{3.23}$$

The root lattice of $\mathfrak{F}_0^e$ is

$$Q_0 = \{2n_1\gamma_1^* + n_3\gamma_3^* \mid n_1, n_3 \in \mathbb{Z}\} \tag{3.24}$$

and the weight "lattice" of $\mathfrak{F}_0^e$ is

$$P_0 = \{n_1\gamma_1^* + n_2\gamma_2^* + t\gamma_3^* \mid n_1, n_2 \in \mathbb{Z}, t \in \mathbb{C}\}. \tag{3.25}$$

The Weyl group W_0 of $\mathfrak{F}_0^e$ is an infinite dihedral group with generators r_1, r_2 whose action on $\mathfrak{h}^*$ is defined by

$$r_i(\lambda) = \lambda - \lambda(h_i)\alpha_i \quad \text{for} \quad i = 1, 2. \tag{3.26}$$

So we have

$$r_1(n_1\gamma_1^* + n_2\gamma_2^* + t\gamma_3^*) = -n_1\gamma_1^* + n_2\gamma_2^* + t\gamma_3^*, \tag{3.27}$$

and

$$r_2(n_1\gamma_1^*+n_2\gamma_2^*+t\gamma_3^*)=(2n_2-n_1)\gamma_1^*+n_2\gamma_2^*+(n_2+t-n_1)\gamma_3^*. \tag{3.28}$$

Note that W_0 fixes γ_3^* so for any $w_0\in W_0$, $\lambda\in\mathfrak{h}^*$, $w_0(\lambda+t\gamma_3^*)=w_0(\lambda)+t\gamma_3^*$. It therefore suffices to compute the action of W_0 only on the span of γ_1^* and γ_2^*.

The translations in W_0 are the powers of $T=r_1r_2$. One has

$$T^i(n_1\gamma_1^*+n_2\gamma_2^*)=(2in_2+n_1)\gamma_1^*+n_2\gamma_2^*+(i^2n_2+in_1)\gamma_3^*. \tag{3.29}$$

Define $\mathscr{C}_0$ to be the category of $\mathfrak{F}_0^e$-modules which decompose into a direct sum of $\mathfrak{F}_0^e$ weight spaces each of which is finite dimensional. Define $\mathscr{E}_0$ to be the algebra of complex valued functions on

$$P_0^+=\{\lambda=n_1\gamma_1^*+n_2\gamma_2^*+t\gamma_3^*\in P_0\,|\,(\lambda,\lambda)\leqq 0, n_2\geqq 0, t\geqq 0\} \tag{3.30}$$

with convolution product. Then $\mathscr{E}\subseteqq\mathscr{E}_0$, the difference being that the γ_3^* coefficient is allowed to vary continuously over $\mathbb{R}^+$ for elements of $\mathscr{E}_0$, but must be in $\mathbb{Z}^+$ for those in $\mathscr{E}$.

Kac [15] generalized the classical Casimir element, which can be written [8] as

$$\sum_{i=1}^{\ell} h_i^2+2\sum_{\alpha>0} x_{-\alpha}x_\alpha+2h_\varrho. \tag{3.31}$$

From the condition that $\langle h_\varrho, h_i\rangle=1$ for $1\leqq i\leqq 3$ we find that for $\mathfrak{F}_0^e$

$$h_\varrho=-\tfrac{9}{2}h_1-5h_2-2h_3=\tfrac{1}{2}h(0)-3c-2d. \tag{3.32}$$

From (3.16), (3.19), (3.31), and (3.32) we see that the Casimir element of $\mathfrak{F}_0^e$ can be written as

$$C=\tfrac{1}{2}h(0)^2-2cd+2\sum_{k\geqq 0} f(-k)e(k)+2\sum_{k>0} e(-k)f(k)$$
$$+\sum_{k>0} h(-k)h(k)+h(0)-6c-4d. \tag{3.33}$$

If V^λ is a standard irreducible $\mathfrak{F}_0^e$-module with highest weight $\lambda=n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*$ then the Casimir element C acts on V^λ as a scalar which we will now compute. From (3.12) we see that basis $(\gamma_1^*,\gamma_2^*,\gamma_3^*\}$ of $\mathfrak{h}^*=(\mathfrak{h}_0^e)^*$ is dual to basis $\{h(0), c, d\}$ of $\mathfrak{h}=\mathfrak{h}_0^e$, so $\lambda(h(0))=n_1$, $\lambda(c)=n_2$ and $\lambda(d)=n_3$. Since the action of C commutes with the action of $\mathfrak{F}_0^e$ on V^λ, it suffices to compute $C\cdot v_\lambda^+$, where v_λ^+ is a highest weight vector of V^λ. Any element of $\mathfrak{F}_0^e$ from a positive root space kills v_λ^+, so

$$\begin{aligned} C\cdot v_\lambda^+ &=(\tfrac{1}{2}h(0)^2+h(0)-6c-4d-2cd)\cdot v_\lambda^+\\ &=(\tfrac{1}{2}n_1^2+n_1-6n_2-4n_3-2n_2n_3)v_\lambda^+\\ &=\langle\lambda+2\varrho,\lambda\rangle v_\lambda^+, \end{aligned} \tag{3.34}$$

where

$$\varrho=\omega_1+\omega_2+\omega_3=\gamma_1^*+2\gamma_2^*+3\gamma_3^*. \tag{3.35}$$

The standard $\mathfrak{F}_0^e$-modules have been stratified according to their "level" by several authors [6, 16, 23]. The level of V^λ for the λ given above is n_2. For standard

highest weight modules this is always a nonnegative integer. The only standard modules on level 0 are trivial one dimensional modules. Standard $\mathfrak{F}_0^e$-modules whose highest weights differ only by a multiple of $\gamma_3^* = \omega_3$ are isomorphic as $\mathfrak{F}_0$-modules, that is, they are only distinguished by the action of derivation d. Their characters differ only by a factor of $e^{t\omega_3}$ for some t.

For the affine Kac-Moody algebras a certain standard module on level 1 called the basic module plays a special role in the theory [7–9] and admits a simple construction in many cases [25, 10, 17]. For $\mathfrak{F}_0^e$ the basic module has highest weight $-\alpha_3 = \omega_2^0 - \omega_3$, whose ω_3 coefficient has been chosen because of its relevance to the structure of $\mathfrak{F}$. We will shortly discuss the construction of this basic module whose character was first computed in [6], and whose weight multiplicities were found to be the values of the classical partition function. We shall denote the character of a standard $\mathfrak{F}_0^e$-module V^λ by χ^λ. From [6] we have the following character formulas: (cf. [16])

$$\chi^{\omega_1} = e^{\omega_1}\left(\sum_{n \geqq 0} p(n) e^{n\omega_3}\right)\left(\sum_{m \in \mathbb{Z}} e^{m\alpha_1 + m(m+1)\omega_3}\right) \tag{3.36}$$

and

$$\chi^{\omega_2} = e^{\omega_2}\left(\sum_{n \geqq 0} p(n) e^{n\omega_3}\right)\left(\sum_{m \in \mathbb{Z}} e^{m\alpha_1 + m^2\omega_3}\right), \tag{3.37}$$

where p is the classical partition function of elementary number theory.

We may rewrite these as

$$\chi^{\omega_1} = e^{\gamma_2^* + \frac{3}{4}\gamma_3^*}\left(\sum_{n \geqq 0} p(n) e^{n\gamma_3^*}\right)\left(\sum_{m \in \mathbb{Z}} e^{2\left(m+\frac{1}{2}\right)\gamma_1^* + \left(m+\frac{1}{2}\right)^2\gamma_3^*}\right) \tag{3.38}$$

and

$$\chi^{\omega_2} = e^{\gamma_2^* \gamma_3^*}\left(\sum_{n \geqq 0} p(n) e^{n\gamma_3^*}\right)\left(\sum_{m \in \mathbb{Z}} e^{2m\gamma_1^* + m^2\gamma_3^*}\right). \tag{3.39}$$

Using the Weyl-Kac character formula and the technique of "principal specialization" one recovers from these formulas [6] the result of Gauss

$$\frac{\phi(q^2)^2}{\phi(q)} = \sum_{j \in \mathbb{Z}} q^{j(2j+1)}, \tag{3.40}$$

where

$$\phi(q) = \prod_{n \geqq 1} (1 - q^n). \tag{3.41}$$

There are two approaches to the study of standard modules on levels greater than 1. One approach is to restrict the basic representation to the subalgebra $\mathfrak{sl}_2(\mathbb{C}) \otimes \mathbb{C}[t^n, t^{-n}] \otimes \mathbb{C}c$ which is isomorphic to $\mathfrak{g}_0$. The basic module is then reducible and contains standard $\mathfrak{g}_0$ modules of level n [9]. Another approach to the higher levels is by studying tensor products of modules from level 1 [5, 16]. We will later require the following decomposition of the basic $\mathfrak{F}_0^e$-module tensored with itself, which can be found in [5] or [16]:

$$V^{\omega_2^0 - \omega_3} \otimes V^{\omega_2^0 - \omega_3} = \sum_{m \geqq 0} (a_m e^{(m-2)\omega_3} V^{2\omega_2^0} + b_m e^{(m-1)\omega_3} V^{2\omega_1^0}), \tag{3.42}$$

where

$$\sum_{m\geqq 0}(a_m x^{2m}+b_m x^{2m+1})=\prod_{j\geqq 1}(1+x^{2j-1}). \tag{3.43}$$

The elementary techniques used in [6] are able to give the characters of level 2 and 3 standard $\mathfrak{F}_0^e$-modules. We will later need the following formulas:

$$\chi^{2\omega_1^0}=e^{2\gamma_2^*}t^{-1}\left[\left(\sum_{m\geqq 0}E(2m)t^{2m}\right)\left(\sum_{k\in\mathbb{Z}}e^{(2k+1)\alpha_1}t^{(2k+1)^2}\right)\right.$$
$$\left.+\left(\sum_{m\geqq 0}E(2m+1)t^{2m+1}\right)\left(\sum_{k\in\mathbb{Z}}e^{2k\alpha_1}t^{4k^2}\right)\right], \tag{3.44}$$

where $t^2=q$ and the weight multiplicities are given by

$$\sum_{m\geqq 0}E(m)t^m=\prod_{j\geqq 1}(1-t^{4j})^{-1}(1-t^{4j-1})^{-1}(1-t^{4j-3})^{-1}, \tag{3.45}$$

and

$$\chi^{2\omega_2^0}=e^{2\gamma_2^*}\left[\left(\sum_{m\geqq 0}E(2m)t^{2m}\right)\left(\sum_{k\in\mathbb{Z}}e^{2k\alpha_1}t^{4k^2}\right)\right.$$
$$\left.+\left(\sum_{m\geqq 0}E(2m+1)t^{2m+1}\right)\left(\sum_{k\in\mathbb{Z}}e^{(2k+1)\alpha_1}t^{(2k+1)^2}\right)\right]. \tag{3.46}$$

For brevity let us denote the basic module $V^{\omega_2^0-\omega_3}$ by V. The tensor product (3.42) has an obvious decomposition into two "strings" of $\mathfrak{F}_0$-isomorphic modules

$$\sum_{m\geqq 0}(a_m q^{(m-2)})V^{2\omega_2^0}\quad\text{and}\quad\sum_{m\geqq 0}(b_m q^{(m-1)})V^{2\omega_1^0}, \tag{3.47}$$

where $q=e^{\omega_3}$. There is also a natural decomposition into the symmetric tensors

$$S(V)=\{v_1\otimes v_2+v_2\otimes v_1\mid v_1,v_2\in V\} \tag{3.48}$$

and the antisymmetric tensors

$$A(V)=\{v_1\otimes v_2-v_2\otimes v_1\mid v_1,v_2\in V\}, \tag{3.49}$$

both of which are reducible $\mathfrak{g}_0^e$-modules. We wish to establish

Theorem 3.2.

$$S(V)=\sum_{m\geqq 0}a_m e^{(m-2)\omega_3}V^{2\omega_2^0} \tag{3.50}$$

and

$$A(V)=\sum_{m\geqq 0}b_m e^{(m-1)\omega_3}V^{2\omega_1^0}. \tag{3.51}$$

We recall first the construction of the basic $\mathfrak{g}_0^e$-module V given by Lepowsky and Wilson [25]. Define the elements

$$h'(k)=e\otimes t^{(k-1)/2}+f\otimes t^{(k+1)/2}\quad\text{for}\quad k\in 2\mathbb{Z}+1, \tag{3.52}$$

and let

$$x(n)=\begin{cases}h\otimes t^{n/2}-\frac{1}{2}\delta_{0,n}c & \text{if}\quad n\in 2\mathbb{Z}\\ -e\otimes t^{(n-1)/2}+f\otimes t^{(n+1)/2} & \text{if}\quad n\in 2\mathbb{Z}+1.\end{cases} \tag{3.53}$$

Then from (3.11) one easily finds that

$$[h'(j), h'(k)] = j\delta_{j,-k}c \quad \text{for} \quad j, k \in 2\mathbb{Z}+1, \tag{3.54}$$

$$[h'(k), x(n)] = 2x(n+k) \quad \text{for} \quad k \in 2\mathbb{Z}+1 \quad \text{and} \quad n \in \mathbb{Z}, \tag{3.55}$$

and

$$[x(k), x(n)] = \begin{cases} k\delta_{k,-n}c & \text{if} \quad k, n \in 2\mathbb{Z} \\ -k\delta_{k,-n}c & \text{if} \quad k, n \in 2\mathbb{Z}+1 \\ 2h'(k+n) & \text{if} \quad k \in 2\mathbb{Z}+1, n \in 2\mathbb{Z}. \end{cases} \tag{3.56}$$

The brackets of derivation d with these elements are easy to compute. A basis for $\mathfrak{g}_0^e$ consists of

$$\{c, d, h'(k), x(n) \mid k \in 2\mathbb{Z}+1, n \in \mathbb{Z}\}. \tag{3.57}$$

The subalgebra

$$\mathfrak{h}' = \sum_{k \in 2\mathbb{Z}+1} \mathbb{C}h'(k) + \mathbb{C}c \tag{3.58}$$

is called the principal Heisenberg subalgebra of $\mathfrak{g}_0^e$, and the basic $\mathfrak{g}_0^e$-module V is an irreducible $\mathfrak{h}'$-module. One can identify V with the symmetric algebra

$$\mathfrak{S} = \mathfrak{S}(h'(-1), h'(-3), h'(-5), \ldots) \tag{3.59}$$

of polynomials in $\{h'(-j) \mid j \in 2\mathbb{Z}^+ + 1\}$, where the highest weight vector $v_0 \in V$ corresponds to $1 \in \mathfrak{S}$. The action of $h'(-k)$ for $k \in 2\mathbb{Z}^+ + 1$ is by multiplication on the left, that of $h'(k)$ for $k \in 2\mathbb{Z}^+ + 1$ is by the derivation $k\dfrac{\partial}{\partial h'(-k)}$, and c acts as the identity. The action of $x(n)$ for $n \in \mathbb{Z}$ is given by a vertex operator

$$X(2h', z) = -\frac{1}{2}\exp\left(\sum_{k \in 2\mathbb{Z}^+ + 1} \frac{z^k}{k} 2h'(-k)\right)\exp\left(-\sum_{k \in 2\mathbb{Z}^+ + 1} \frac{z^{-k}}{k} 2h'(k)\right) \tag{3.60}$$

in the following way. For $n \in \mathbb{Z}$ let $X_n(2h')$ denote the n^{th} homogeneous component of $X(2h', z)$, so that

$$X(2h', z) = \sum_{n \in \mathbb{Z}} X_n(2h')z^{-n}. \tag{3.61}$$

Then the operators $X_n(2h')$ are well-defined on $\mathfrak{S}$, satisfy the commutation relations (3.55) and (3.56), and act on $\mathfrak{S}$ as the operators $x(n)$ act on V.

Now consider the tensor product $V \otimes V$. We will use subscripts 1 and 2 to indicate the first and second factor, respectively. So we will let

$$h'_1(k) = h'(k) \otimes 1 \quad \text{and} \quad h'_2(k) = 1 \otimes h'(k) \quad \text{for} \quad k \in 2\mathbb{Z}+1 \tag{3.62}$$

whose action on $v_1 \otimes v_2$ is given by

$$h'_1(k) \cdot (v_1 \otimes v_2) = (h'(k) \cdot v_1) \otimes v_2 \tag{3.63}$$

and

$$h'_2(k) \cdot (v_1 \otimes v_2) = v_1 \otimes (h'(k) \cdot v_2). \tag{3.64}$$

Then the action of $\mathfrak{g}_0^e$ on $V\otimes V$ is given by the operators $h_1'(k)+h_2'(k)$ for $k\in 2\mathbb{Z}+1$, $X_n(2h_1')+X_n(2h_2')$ for $n\in\mathbb{Z}$, and c acts as the scalar 2.

We introduce the following auxiliary vertex operator on $V\otimes V$,

$$X(h_1'-h_2',z)=\exp\left(\sum_{k\in 2\mathbb{Z}^++1}\frac{z^k}{k}(h_1'(-k)-h_2'(-k))\right)\cdot\exp\left(-\sum_{k\in 2\mathbb{Z}^++1}\frac{z^{-k}}{k}(h_1'(k)-h_2'(k))\right) \tag{3.65}$$

with n^{th} homogeneous component $X_n(h_1'-h_2')$, so that

$$X(h_1'-h_2',z)=\sum_{n\in\mathbb{Z}}X_n(h_1'-h_2')z^{-n}. \tag{3.66}$$

Let us denote anti-commutators by $\{A,B\}=AB+BA$. Then direct calculations with vertex operators show that for $m\in\mathbb{Z}$, $n\in 2\mathbb{Z}+1$ we have

$$\{X_n(h_1'-h_2'),X_m(h_1'-h_2')\}=-2\delta_{n,-m}, \tag{3.67}$$

$$\{X_n(h_2'-h_1'),X_m(2h_1')+X_m(2h_2')\}=0, \tag{3.68}$$

and

$$X_m(h_1'-h_2')\cdot(1\otimes 1)=0 \quad \text{for} \quad m>0. \tag{3.69}$$

For example, in the proof of (3.68) one uses

$$\{X_n(h_2'-h_1'),X_m(2h_1')\}=\frac{1}{2\pi i}\int_{C_0}\left(\frac{1}{2\pi i}\int_{C_R}X(h_2'-h_1',z)X(2h_1',w)z^{n-1}dz+\frac{1}{2\pi i}\int_{C_r}X(2h_1',w)X(h_2'-h_1',z)z^{n-1}dz\right)w^{m-1}dw, \tag{3.70}$$

where $r<r_0<R$ and C_0 is a fixed circle of radius r_0. Elementary operations then give

$$\frac{1}{2\pi i}\int_{C_0}2X(h_2'+h_1',w)w^{n+m-1}dw=2X_{n+m}(h_2'+h_1'), \tag{3.71}$$

and by symmetry we have

$$\{X_n(h_1'-h_2'),X_m(2h_2')\}=2X_{n+m}(h_1'+h_2'). \tag{3.72}$$

These give the result. Details concerning vertex operators can be found in [8, 10].

Let $\Omega\subset V\otimes V$ be the subspace of highest weight vectors for $\mathfrak{g}_0^e$. So Ω is killed by the space of positive root vectors from $\mathfrak{g}_0^e$ which is spanned by

$$\{x(n),h'(k)|0<n\in\mathbb{Z},0<k\in 2\mathbb{Z}+1\}. \tag{3.73}$$

In $\mathfrak{S}\otimes\mathfrak{S}$ this corresponds to the space of tensors killed by

$$\{X_n(2h_1')+X_n(2h_2'),h_1'(k)+h_2'(k)|0<n\in\mathbb{Z},0<k\in 2\mathbb{Z}+1\}. \tag{3.74}$$

From (3.42) we have the character

$$\chi(\Omega)=\sum_{m\geqq 0}(a_m e^{(m-2)\omega_3+2\omega_2^0}+b_m e^{(m-1)\omega_3+2\omega_1^0}), \tag{3.75}$$

where the coefficients are given by (3.43). Let $\mathrm{ch}(\cdot)$ denote the principally specialized character of any $\mathfrak{g}_0^e$-module, that is, the result of setting $e^{-\alpha_1}=u=e^{-\alpha_2}$ in the character. Then we have

$$\mathrm{ch}(\Omega)=e^{-2\alpha_3}\prod_{j\geq 1}(1+u^{2j-1}). \tag{3.76}$$

Consider the set of vectors

$$X_{-2n_1-1}(h_1'-h_2')X_{-2n_2-1}(h_1'-h_2')\ldots X_{-2n_k-1}(h_1'-h_2')\cdot(1\otimes 1), \tag{3.77}$$

where $n_1>n_2>\ldots>n_k\geqq 0$. Using (3.67), (3.68), and (3.69) it is easy to prove that these vectors are linearly independent and are killed by $X_n(2h_1')+X_n(2h_2')$ for $0<n\in\mathbb{Z}$. If we show that they are also killed by $h_1'(k)+h_2'(k)$ for $0<k\in 2\mathbb{Z}+1$ then (3.76) shows that they are a basis for Ω. It is clear from (3.65) that $h_1'(k)+h_2'(k)$ and $X_n(h_1'-h_2')$ commute. Thus, $h_1'(k)+h_2'(k)$ applied to (3.77) will give 0 when $0<k\in 2\mathbb{Z}+1$.

Note that the vector in (3.77) is symmetric if it is unchanged when $h_1'-h_2'$ is replaced by $h_2'-h_1'$, and antisymmetric if the sign reverses. Because the subscripts are odd we see that the vector is symmetric when k is even and antisymmetric when k is odd.

Consider the element $x(0)=h\otimes 1-\frac{1}{2}c$ of the Cartan subalgebra of $\mathfrak{g}_0^e$, whose action on $V\otimes V$ is given by the operator $X_0(2h_1')+X_0(2h_2')$. If $v\in V$ has weight $\omega_1^0-m\omega_3$ for some m then

$$x(0)\cdot v=\tfrac{1}{2}v, \tag{3.78}$$

but if $v\in V$ has weight $\omega_2^0-m\omega_3$ then

$$x(0)\cdot v=-\tfrac{1}{2}v. \tag{3.79}$$

Thus, for any vector $v\in\Omega$ of the form (3.77), either v is of weight $2\omega_1^0-m\omega_3$ for some $m\in\mathbb{Z}$ and

$$(X_0(2h_1')+X_0(2h_2'))\cdot v=v, \tag{3.80}$$

or else v is of weight $2\omega_2^0-m\omega_3$ and

$$(X_0(2h_1')+X_0(2h_2'))\cdot v=-v. \tag{3.81}$$

But (3.68) with $m=0$ implies that in the first case k is odd and in the second case k is even because $1\otimes 1$ has weight $2\omega_2^0-\omega_3$. This completes the proof of Theorem 3.2.

Remark. The above result is closely related to Theorem 1.6 of [9] for the case when $n=m=2$.

Note that Theorem 3.2 along with (3.43)–(3.46) allow one to easily compute the characters of the symmetric and antisymmetric parts of the tensor product $V\otimes V$. Using (3.40) and (3.41) it is easy to obtain the principally specialized characters $\mathrm{ch}(S(V))$ and $\mathrm{ch}(A(V))$.

Corollary 3.3. *We have the principally specialized characters*

$$\mathrm{ch}(S(V))=e^{2\gamma_2}\frac{u^{-4}}{2}\frac{\phi(u^4)}{\phi(u)}\left[\prod_{j\geq 1}(1+u^{2j-1})+\prod_{j\geq 1}(1-u^{2j-1})\right] \tag{3.82}$$

and

$$\mathrm{ch}(A(V))=e^{2\gamma_3^*}\frac{u^{-4}}{2}\frac{\phi(u^4)}{\phi(u)}\left[\prod_{j\geq 1}(1+u^{2j-1})-\prod_{j\geq 1}(1-u^{2j-1})\right]. \tag{3.83}$$

From (3.34) we have the scalar by which the Casimir element C acts on an irreducible standard module V^λ. Let us consider the two "strings" of modules in (3.42). If $\lambda=2\gamma_2^*+n_3\gamma_3^*=2\omega_2^0+n_3\omega_3$ then (3.34) says

$$C\cdot v_\lambda=(-12-8n_3)v_\lambda \tag{3.84}$$

for any $v_\lambda\in V^\lambda$, while if $\lambda=2\gamma_1^*+2\gamma_2^*+n_3\gamma_3^*=2\omega_1^0+n_3\omega_3$ then

$$C\cdot v_\lambda=-8(n_3+1)v_\lambda. \tag{3.85}$$

In the first case, since $n_3\in\mathbb{Z}$, the scalar is never zero, but in the second case it is zero if and only if $n_3=-1$, that is, when $\lambda=-2\alpha_3-\alpha_2$. The coefficient in (3.42) which corresponds to this λ is $b_0=1$, so there is a unique irreducible component of $V\otimes V$ on which C acts as zero.

Corollary 3.4. *In the tensor product $V\otimes V$ of the basic module of $\mathfrak{F}_0^e$ with itself, the subspace of vectors killed by the Casimir element C forms an irreducible standard module with highest weight $2\omega_1^0-\omega_3$.*

For reasons which will not be apparent until later we would like to compute the principally specialized character of the antisymmetric "string" $A(V)$ with the first standard module (the one referred to in Corollary 3.4) removed.

Corollary 3.5. *The principally specialized character of*

$$\left(\sum_{m\geq 1}b_m q^{m-1}\right)V^{2\omega_1^0}, \tag{3.86}$$

where the coefficients b_m are given by (3.43), *is*

$$e^{2\gamma_3^*}\frac{u^{-4}}{2}\frac{\phi(u^4)}{\phi(u)}\left[\prod_{j\geq 1}(1+u^{2j-1})-\prod_{j\geq 1}(1-u^{2j-1})-2u\right]. \tag{3.87}$$

We now describe how characters in $\mathscr{E}$ can be viewed as functions of three complex variables and their domain of definition.

With respect to the bilinear form $(\cdot,\cdot)$ on $\mathfrak{h}^*$ the basis of $\mathfrak{h}^*$ dual to basis (2.13) is

$$\gamma_1=\alpha_1,\quad \gamma_2=\alpha_1+\alpha_2,\quad \gamma_3=\alpha_1+\alpha_2+\alpha_3. \tag{3.88}$$

If we write

$$z=z_1\gamma_1+z_2\gamma_2+z_3\gamma_3 \tag{3.89}$$

then for

$$\lambda=n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*\in P \tag{3.90}$$

we have

$$(\lambda,z)=n_1z_1+n_2z_2+n_3z_3. \tag{3.91}$$

For each $\lambda \in P^+$ we make $e^\lambda \in \mathscr{E}$ a function of z_1, z_2, z_3 by defining

$$e^\lambda(z_1, z_2, z_3) = e^{2\pi i(\lambda, z)} = e^{2\pi i(n_1 z_1 + n_2 z_2 + n_3 z_3)}. \tag{3.92}$$

For any character $X = \sum_{\lambda \in P^+} a(\lambda) e^\lambda \in \mathscr{E}$ we define

$$X(z_1, z_2, z_3) = \sum_{\lambda \in P^+} a(\lambda) e^\lambda(z_1, z_2, z_3) \tag{3.93}$$

for $z_j = x_j + iy_j \in \mathbb{C}$ satisfying the conditions $y_2 y_3 > y_1^2$, $y_2, y_3 > 0$. The function $X(z) = X(z_1, z_2, z_3)$ is defined on

$$\mathfrak{D} = \{z = z_1\gamma_1 + z_2\gamma_2 + z_3\gamma_3 \in \mathfrak{h}^* | z_j = x_j + iy_j, y_2 y_3 > y_1^2, y_2 > 0\}. \tag{3.94}$$

Then

$$v(\mathfrak{D}) = \left\{ \mathfrak{Z} = \begin{pmatrix} -z_2 & z_1 \\ z_1 & -z_3 \end{pmatrix} \in S_2(\mathbb{C}) | \operatorname{Im}(z_j) = y_j, y_2 y_3 > y_1^2, y_2 > 0 \right\}. \tag{3.95}$$

From our point of view this is natural because v induces a form on $S_2(\mathbb{C})$ such that

$$(v(\lambda), v(z)) = n_1 z_1 + n_2 z_2 + n_3 z_3. \tag{3.96}$$

It is more traditional, however, in the theory of Siegel modular forms (see [1]) to consider the set $v(P^+)$ as we have done, but to use instead of $v(\mathfrak{D})$ the set

$$v(\mathfrak{D})^\iota = \left\{ \mathfrak{Z} = \begin{pmatrix} z_3 & z_1 \\ z_1 & z_2 \end{pmatrix} \in S_2(\mathbb{C}) | \operatorname{Im}(\mathfrak{Z}) > 0 \right\}, \tag{3.97}$$

where ι denotes the "cannonical involution"

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\iota = \begin{pmatrix} -d & c \\ b & -a \end{pmatrix} \tag{3.98}$$

which satisfies $x^\iota y^\iota = -(xy)^\iota$ and which may be realized in $\mathfrak{h}^*$ as $-Ir_1 r_3$. The set given in (3.97) is called the Siegel domain (or Siegel upper half-plane) H_2 of genus 2. It is standard in the theory of Siegel modular forms to consider functions on H_2 which can be written as

$$\sum_{N \geqq 0} A(N) e^{2\pi i \operatorname{Tr}(N\mathfrak{Z})} \tag{3.99}$$

for $N \in v(P^+)$, $\mathfrak{Z} \in v(\mathfrak{D})^\iota = H_2$ and where

$$\operatorname{Tr}(N\mathfrak{Z}) = n_1 z_1 + n_2 z_2 + n_3 z_3. \tag{3.100}$$

This traditional use of the trace appears to obscure the indefinite bilinear form on $S_2(\mathbb{C})$ which from our point of view naturally gives rise to the form $n_1 z_1 + n_2 z_2 + n_3 z_3$ and gives a connection between the theory of Kac-Moody algebras and that of Siegel modular forms.

In order to work in the traditional domain H_2 only minor adjustments are needed. Because

$$(gxg^\iota)^\iota = g^\iota x^\iota (g^\iota)^\iota, \tag{3.101}$$

for $g \in \mathrm{PGL}_2(\mathbb{Z})$ the action of $\mathrm{PGL}_2(\mathbb{Z})$ on $x^\iota \in H_2$ is given by

$$g \cdot x^\iota = g^\iota x^\iota (g^\iota)^\iota. \tag{3.102}$$

4. The Construction of $\mathfrak{F}$

We shall now show how the algebra $\mathfrak{F}$ can be constructed from the subalgebra $\mathfrak{F}_0^e$, its basic module $V=V^{-\alpha_3}$ and the dual contragredient module V^*. This construction, inspired by the work of Kantor [19], gives $\mathfrak{F}$ as a $\mathbb{Z}$-graded algebra

$$\ldots+\mathfrak{F}_{-2}+\mathfrak{F}_{-1}+\mathfrak{F}_0^e+\mathfrak{F}_1+\mathfrak{F}_2+\ldots, \tag{4.1}$$

where $\mathfrak{F}_{-1}\approx V$ and $\mathfrak{F}_1\approx V^*$, whose weight multiplicities were shown in [6] to be values of the classical partition function p.

Let $V=V^{\omega_2^0-\omega_3}=V^{-\alpha_3}$ be the basic $\mathfrak{g}_0^e$-module with highest weight vector v_0 and let V^* be the dual space of linear functionals on V. We will write the value of the functional $v^*\in V^*$ on $v\in V$ as $\langle v^*|v\rangle$. The contragredient action of $x\in\mathfrak{g}_0^e$ on v^* is determined by

$$\langle x\cdot v^*|v\rangle=-\langle v^*|x\cdot v\rangle \tag{4.2}$$

which implies that

$$[x,y]\cdot v^*=x\cdot(y\cdot v^*)-y\cdot(x\cdot v^*) \tag{4.3}$$

for $x,y\in\mathfrak{g}_0^e$, so V^* is a $\mathfrak{g}_0^e$-module. We will define $v^*\cdot x=-x\cdot v^*$ so that $\langle v^*\cdot x|v\rangle=\langle v^*|x\cdot v\rangle$ may be simply written $\langle v^*xv\rangle$. Then V^* is an irreducible lowest weight $\mathfrak{g}_0^e$-module with lowest weight vector v_0^* of weight α_3. One has

$$\begin{aligned} &e_1\cdot v_0=0, e_2\cdot v_0=0,\\ &h_1\cdot v_0=0, h_2\cdot v_0=v_0, h_3\cdot v_0=-2v_0,\\ &f_1\cdot v_0=0, f_2\cdot(f_2\cdot v_0)=0, \end{aligned} \tag{4.4}$$

and

$$\begin{aligned} &f_1\cdot v_0^*=0, f_2\cdot v_0^*=0,\\ &h_1\cdot v_0^*=0, h_2\cdot v_0^*=-v_0^*, h_3\cdot v_0^*=2v_0^*,\\ &e_1\cdot v_0^*=0, e_2\cdot(e_2\cdot v_0^*)=0. \end{aligned} \tag{4.5}$$

Define a map $\phi: V^*\times V\to\mathfrak{g}_0^e$ by

$$\phi(v^*,v)=-\sum_{i\in I}\langle v^*x_iv\rangle x_i \tag{4.6}$$

where $\{x_i|i\in I\}$ is an orthonormal basis for $\mathfrak{g}_0^e$ with respect to the form $\langle\cdot,\cdot\rangle$. It is clear that only finitely many terms in this sum are nonzero for any particular pair (v^*,v), and that the definition of ϕ is independent of choice of orthonormal basis.

Proposition 4.1. *For any $x\in\mathfrak{g}_0^e$, $v\in V$, $v^*\in V^*$ we have*

$$[x,\phi(v^*,v)]=\phi(x\cdot v^*,v)+\phi(v^*,x\cdot v).$$

Proof. Let $[x_i,x_j]=\sum_{k\in I}C_{ij}^kx_k$ where $\{x_i|i\in I\}$ is an orthonormal basis for $\mathfrak{g}_0^e$. The invariance of the form, $\langle[x_i,x_j],x_k\rangle=\langle x_i,[x_j,x_k]\rangle$, says $C_{ij}^k=C_{jk}^i$. So we have

$$\begin{aligned} [x_j,\phi(v^*,v)]&=-\sum_{i\in I}\langle v^*x_iv\rangle[x_j,x_i]\\ &=-\sum_{i\in I}\sum_{k\in I}\langle v^*x_iv\rangle C_{ji}^kx_k=-\sum_{i\in I}\sum_{k\in I}\langle v^*x_kv\rangle C_{jk}^ix_i. \end{aligned}$$

On the other side we have

$$\phi(x_j\cdot v^*, v)+\phi(v^*, x_j\cdot v)=-\sum_{i\in I}\langle x_j\cdot v^*|x_i\cdot v\rangle x_i-\sum_{i\in I}\langle v^*\cdot x_i|x_j\cdot v\rangle x_i$$
$$=-\sum_{i\in I}\langle v^*[x_i, x_j]v\rangle x_i=-\sum_{i\in I}\sum_{k\in I}\langle v^*x_k v\rangle C_{ij}^k x_i.$$

We now define a $\mathbb{Z}$-graded Lie algebra

$$\mathfrak{G}=\mathfrak{G}(\mathfrak{g}_0^e, V)=\sum_{n\in\mathbb{Z}}\mathfrak{G}_n, \tag{4.7}$$

where $\mathfrak{G}_0=\mathfrak{g}_0^e$, $\mathfrak{G}_{-1}=V$, $\mathfrak{G}_1=V^*$ and where $\mathfrak{G}^+=\sum_{n\geqq 1}\mathfrak{G}_n$ and $\mathfrak{G}^-=\sum_{n\leqq -1}\mathfrak{G}_n$ are the free Lie algebras generated by $\mathfrak{G}_1$ and $\mathfrak{G}_{-1}$, respectively. So for $n\geqq 1$, $\mathfrak{G}_n$ (respectively, $\mathfrak{G}_{-n}$) is the space spanned by all brackets of n vectors from $\mathfrak{G}_1$ (respectively, $\mathfrak{G}_{-1}$). The Lie brackets between $\mathfrak{G}_k$ and $\mathfrak{G}_{-\ell}$ for $k\geqq 1$, $\ell>1$ are defined inductively, that is,

$$[A,[B,C]]=[[A,B],C]+[B,[A,C]] \tag{4.8}$$

for $A\in\mathfrak{G}_k$, $[B,C]\in\mathfrak{G}_{-\ell}$, B, $C\in\mathfrak{G}^-$, and similarly for $k>1$, $\ell\geqq 1$. The brackets between $\mathfrak{G}_1$ and $\mathfrak{G}_{-1}$ are given by

$$[v^*, v]=\phi(v^*, v), \tag{4.9}$$

and the brackets between $\mathfrak{G}_0$ and $\mathfrak{G}_1(\mathfrak{G}_{-1})$ are given by the $\mathfrak{g}_0^e$-module action,

$$[x, v]=x\cdot v \quad\text{and}\quad [x, v^*]=x\cdot v^*. \tag{4.10}$$

Proposition 4.1 shows that the Jacobi identity

$$[x,[v^*, v]]=[[x, v^*], v]+[v^*,[x, v]] \tag{4.11}$$

holds for all $x\in\mathfrak{G}_0$, $v^*\in\mathfrak{G}_1$, $v\in\mathfrak{G}_{-1}$. Note that $\mathfrak{G}_2$ (respectively, $\mathfrak{G}_{-2}$) and the exterior (wedge) product $V^*\wedge V^*$ (respectively, $V\wedge V$) are isomorphic as $\mathfrak{g}_0^e$-modules. The correspondence is just

$$[v_1^*, v_2^*]\leftrightarrow v_1^*\wedge v_2^* \text{ (respectively, } [v_1, v_2]\leftrightarrow v_1\wedge v_2). \tag{4.12}$$

Define the subspaces (cf. [19])

$$\mathfrak{I}_k^+=\{x\in\mathfrak{G}_k|[y_1,[y_2,\ldots,[y_{k-1}, x]\ldots]]=0 \quad\text{for all}\quad y_i\in\mathfrak{G}_{-1}\} \tag{4.13}$$

and

$$\mathfrak{I}_k^-=\{x\in\mathfrak{G}_{-k}|[y_1,[y_2,\ldots,[y_{k-1}, x]\ldots]]=0 \quad\text{for all}\quad y_i\in\mathfrak{G}_1\} \tag{4.14}$$

for $k>1$. Let

$$\mathfrak{I}^+=\sum_{k>1}\mathfrak{I}_k^+, \mathfrak{I}^-=\sum_{k>1}\mathfrak{I}_k^- \quad\text{and}\quad \mathfrak{I}=\mathfrak{I}^++\mathfrak{I}^-. \tag{4.15}$$

Proposition 4.2. *$\mathfrak{I}^+$, $\mathfrak{I}^-$, and $\mathfrak{I}$ are ideals of $\mathfrak{G}$.*

Proof. Let $x\in\mathfrak{I}_k^+$ for $k>1$. If $y\in\mathfrak{G}_{-1}$ then it is obvious that $[x, y]\in\mathfrak{I}_{k-1}^+$. If $y\in\mathfrak{G}_1$ we wish to show that $[x, y]\in\mathfrak{I}_{k+1}^+$, that is, $[z_1,[z_2,\ldots[z_k,[x, y]]\ldots]]=0$ for all $z_i\in\mathfrak{G}_{-1}$. Written out using the Jacobi identity this expression is a summation of

terms of the form $[[z_1, [z_2, \ldots [z_i, x]\ldots]], [z_{i+1}, [z_{i+2}, \ldots [z_k, y]\ldots]]]$. But since $[z_k, y] \in \mathfrak{G}_0$, $[z_{i+1}, [z_{i+2}, \ldots [z_k, y]\ldots]] \in \mathfrak{G}_{k-i-1}$, which may be distributed back into $[z_1, [z_2, \ldots [z_i, x]\ldots]]$ to yield a summation $\sum [w_1, [w_2, \ldots [w_{k-1}, x]\ldots]] = 0$ since $w_i \in \mathfrak{G}_{-1}$. The same argument shows that if $y \in \mathfrak{G}_0$ then $[x, y] \in \mathfrak{I}_k^+$. Since $\mathfrak{G}_1$, $\mathfrak{G}_0$, and $\mathfrak{G}_{-1}$ generate $\mathfrak{G}$, $\mathfrak{I}^+$ is an ideal of $\mathfrak{G}$. The argument for $\mathfrak{I}^-$ is symmetrical and so $\mathfrak{I} = \mathfrak{I}^+ + \mathfrak{I}^-$ is also an ideal of $\mathfrak{G}$.

Define the $\mathbb{Z}$-graded Lie algebra

$$\bar{\mathfrak{G}} = \bar{\mathfrak{G}}(\mathfrak{g}_0^e, V) = \mathfrak{G}/\mathfrak{I} = \mathfrak{G}^-/\mathfrak{I}^- + \mathfrak{G}_0 + \mathfrak{G}^+/\mathfrak{I}^+ \tag{4.16}$$

to be the canonical $\mathbb{Z}$-graded Lie algebra associated with $\mathfrak{g}_0^e$ and V. While our primary purpose is to show that $\bar{\mathfrak{G}}(\mathfrak{g}_0^e, V) \approx \mathfrak{F}$ it should be noted that there is no difficulty in generalizing the above construction to give $\bar{\mathfrak{G}}(\mathfrak{g}\hat{}^e, V_0)$ for V_0 the basic module of any affine Kac-Moody algebra $\mathfrak{g}\hat{}$. The affine algebra $\mathfrak{g}\hat{}$ is constructed from a finite dimensional simple Lie algebra $\mathfrak{g}$ by the addition of a new simple root $\alpha_0 = -\tilde{\alpha}$ to $\{\alpha_1, \ldots, \alpha_\ell\}$, where $\tilde{\alpha}$ is the highest root of $\mathfrak{g}$. Then we may construct from $\mathfrak{g}\hat{}$ a Kac-Moody algebra $\mathfrak{g}\hat{}^*$ by the addition of a new simple root α_{-1} to $\{\alpha_0, \alpha_1, \ldots, \alpha_\ell\}$ such that α_{-1} is connected only to α_0 by just one line in the Dynkin diagram. The Cartan subalgebra $\mathfrak{h}\hat{}^* \subset \mathfrak{g}\hat{}^*$ will then be identified with the extended Cartan subalgebra $\mathfrak{h}\hat{}^e \subset \mathfrak{g}\hat{}^e$ and $\mathfrak{g}\hat{}^* \approx \bar{\mathfrak{G}}(\mathfrak{g}\hat{}^e, V_0)$.

The simplest example of this construction is the algebra $\mathfrak{F}$ of type $A_1\hat{}^*$, which we studied first because of its Weyl group. The next cases one might consider are $A_2\hat{}^*$ and $B_2\hat{}^*$, whose Weyl groups are also quite interesting. The Weyl group of $B_2\hat{}^*$, for example, is isomorphic to the Klein-Fricke group Ψ_1^* (in Magnus' notation [34]) which contains the Picard group $\Psi_1 = \mathrm{PSL}_2(\mathbb{Z}[i])$ as a subgroup of index 4.

The algebra $\mathfrak{F}$ decomposes into $\mathfrak{F}_0^e$-modules

$$\ldots + \mathfrak{F}_{-2} + \mathfrak{F}_{-1} + \mathfrak{F}_0^e + \mathfrak{F}_1 + \mathfrak{F}_2 + \ldots, \tag{4.17}$$

where for $n \geqq 0$, $\mathfrak{F}_n$ (respectively, $\mathfrak{F}_{-n}$) is the space spanned by all brackets involving exactly n e_3's (respectively, f_3's).

Proposition 4.3. *We may identify* $\mathfrak{F}_0^e = \mathfrak{G}_0$, $\mathfrak{F}_{-1} = \mathfrak{G}_{-1}$, *and* $\mathfrak{F}_1 = \mathfrak{G}_1$ *in* $\bar{\mathfrak{G}}(\mathfrak{g}_0^e, V)$ *so that* $v_0 \in V = \mathfrak{G}_{-1}$ *corresponds to* f_3, $v_0^* \in V^* = \mathfrak{G}_1$ *corresponds to* e_3, *for* $v \in V$ *and* $v^* \in V^*$ *the bracket* $[v^*, v]$ *corresponds to* $\phi(v^*, v)$, *and* $\mathfrak{g}_0^e$*-module isomorphism* $\mathfrak{F}_{-1} + \mathfrak{F}_0^e + \mathfrak{F}_1 \approx V + \mathfrak{g}_0^e + V^*$ *respects those Lie brackets which stay within these graded pieces, provided that* $\langle v_0^* | v_0 \rangle = 1$.

Proof. From (2.2) we recall that

$$\begin{aligned} &[e_1, f_3] = 0, \quad [e_2, f_3] = 0, \\ &[h_1, f_3] = 0, \quad [h_2, f_3] = f_3, \quad [h_3, f_3] = -2f_3, \\ &[f_1, f_3] = 0, \quad [f_2[f_2, f_3]] = 0, \end{aligned} \tag{4.18}$$

and

$$\begin{aligned} &[f_1, e_3] = 0, \quad [f_2, e_3] = 0, \\ &[h_1, e_3] = 0, \quad [h_2, e_3] = -e_3, \quad [h_3, e_3] = 2e_3, \\ &[e_1, e_3] = 0, \quad [e_2[e_2, e_3]] = 0. \end{aligned} \tag{4.19}$$

It is easy to see that $\mathfrak{F}_{-1}$ (respectively, $\mathfrak{F}_1$) is generated as $\mathfrak{F}_0^e$-module by highest (respectively, lowest) weight vector f_3 (respectively, e_3) of weight $-\alpha_3$ (respectively, α_3), so $\mathfrak{F}_{-1}=V(\mathfrak{F}_1=V^*)$.

We have that $h_3=[e_3,f_3]$, which should correspond to

$$\phi(v_0^*,v_0)=-\sum_{i\in I}\langle v_0^* x_i v_0\rangle x_i$$

where $\{x_i|i\in I\}$ is any orthonormal basis of $\mathfrak{g}_0^e$. If we use the basis (3.17)–(3.19) then the only nonzero terms correspond to the basis vectors

$$\frac{1}{\sqrt{-2}}(c+d) \quad \text{and} \quad \frac{1}{\sqrt{2}}(c-d).$$

We have from (4.4) and (3.12) that

$$\begin{aligned}\phi(v_0^*,v_0)&=-\left\langle v_0^*\left(\frac{c+d}{\sqrt{-2}}\right)v_0\right\rangle\left(\frac{c+d}{\sqrt{-2}}\right)-\left\langle v_0^*\left(\frac{c-d}{\sqrt{2}}\right)v_0\right\rangle\left(\frac{c-d}{\sqrt{2}}\right)\\&=\langle v_0^* c v_0\rangle d+\langle v_0^* d v_0\rangle c\\&=\langle v_0^*|v_0\rangle(d-c)\\&=\langle v_0^*|v_0\rangle h_3,\end{aligned}$$

so the correspondence holds provided that $\langle v_0^*|v_0\rangle=1$. For arbitrary $v^*\in V^*$ and $v\in V$ we may assume that $v^*=v_0^*\cdot x_j$ and $v=x_k\cdot v_0$ for x_j,x_k basis vectors from an orthonormal basis $\{x_i|i\in I\}$ of $\mathfrak{g}_0^e$. Then $[[e_3,x_j],[x_k,f_3]]$ should correspond to $\phi(v_0^*\cdot x_j,x_k\cdot v_0)=\phi(v_0^*,x_j\cdot(x_k\cdot v_0))-[x_j,\phi(v_0^*,x_k\cdot v_0)]$ by Proposition 4.1. But from the Jacobi identity we have

$$[[e_3,x_j],[x_k,f_3]]=[e_3,[x_j,[x_k,f_3]]]-[x_j,[e_3,[x_k,f_3]]],$$

which shows that the correspondence holds.

Proposition 4.4. *The restriction of the invariant form* $(\cdot,\cdot)$ *of* $\mathfrak{F}$ *to* $\mathfrak{F}_{-1}+\mathfrak{F}_0^e+\mathfrak{F}_1$ *coincides with the invariant form* $\langle\cdot,\cdot\rangle$ *on* $\mathfrak{g}_0^e$ *and the pairing* $\langle\cdot|\cdot\rangle$ *of* V^* *and* V, *provided that* $\langle v_0^*|v_0\rangle=1$.

Proof. We see that $(e_3,f_3)=\frac{1}{2}([h_3,e_3],f_3)=\frac{1}{2}(h_3,[e_3,f_3])=\frac{1}{2}(h_3,h_3)=1$ agrees with $\langle v_0^*|v_0\rangle=1$. If $x_i,x_j\in\mathfrak{g}_0^e$ are from an orthonormal basis of $\mathfrak{g}_0^e$ we must show that $([x_i,e_3],[x_j,f_3])=\langle x_i\cdot v_0^*|x_j\cdot v_0\rangle$. From definition (4.6) we have

$$-\langle\phi(v^*,v),x_k\rangle=\langle v^* x_k v\rangle \tag{4.20}$$

for any $v^*\in V^*$, $v\in V$. From (3.12) and (3.16) it can be seen that the restriction of form $(\cdot,\cdot)$ to $\mathfrak{F}_0^e$ coincides with the form $\langle\cdot,\cdot\rangle$ on $\mathfrak{g}_0^e$. By invariance we have $([x_i,e_3],[x_j,f_3])=(x_i,[e_3,[x_j,f_3]])=\langle x_i,\phi(v_0^* x_j v_0)\rangle$ since $[e_3[x_j,f_3]]\in\mathfrak{g}_0^e$. But from (4.20) we have $\langle x_i,\phi(v_0^* x_j v_0)\rangle=-\langle v_0^* x_i|x_j v_0\rangle=\langle x_i\cdot v_0^*|x_j\cdot v_0\rangle$.

In order to see that $\mathfrak{F}\approx\mathfrak{G}(\mathfrak{g}_0^e,V)$ we must study more closely the ideal $\mathfrak{I}_2^-$. By definition

$$\begin{aligned}\mathfrak{I}_2^-&=\left\{\sum_k[v_1^k,v_2^k]\in\mathfrak{G}_{-2}\middle|\sum_k[v^*,[v_1^k,v_2^k]]=0 \quad \text{for all} \quad v^*\in V^*=\mathfrak{G}_1\right\}\\&=\left\{\sum_k[v_1^k,v_2^k]\in\mathfrak{G}_{-2}\middle|\sum_k\langle v_1^*,[v_2^*,[v_1^k,v_2^k]]\rangle=0 \quad \text{for all} \quad v_1^*,v_2^*\in V^*=\mathfrak{G}_1\right\}\end{aligned} \tag{4.21}$$

since V and V^* are dual. One has $\sum_k [v_1^k, v_2^k] \in \mathfrak{I}_2^-$ if and only if for all $v_1^*, v_2^* \in V^*$

$$0 = \sum_k (\langle v_1^*, [[v_2^*, v_1^k], v_2^k] \rangle + \langle v_1^*, [v_1^k, [v_2^*, v_2^k]] \rangle)$$

$$= \sum_k \left(\left\langle v_1^*, -\sum_{i \in I} \langle v_2^* x_i v_1^k \rangle x_i \cdot v_2^k \right\rangle + \left\langle v_1^*, \sum_{i \in I} \langle v_2^* x_i v_2^k \rangle x_i \cdot v_1^k \right\rangle \right)$$

$$= \sum_k \sum_{i \in I} (\langle v_1^* x_i v_1^k \rangle \langle v_2^* x_i v_2^k \rangle - \langle v_1^* x_i v_2^k \rangle \langle v_2^* x_i v_1^k \rangle), \tag{4.22}$$

where $\{x_i | i \in I\}$ is any orthonormal basis of $\mathfrak{g}_0^e$.

Define a bilinear pairing of $\mathfrak{G}_{-2}$ with $\mathfrak{G}_2$ as follows. For any $v_1, v_2 \in V = \mathfrak{G}_{-1}$ and any $v_1^*, v_2^* \in V^* = \mathfrak{G}_1$ let

$$\langle [v_1^*, v_2^*], [v_1, v_2] \rangle = \langle v_1^* | v_1 \rangle \langle v_2^* | v_2 \rangle - \langle v_1^* | v_2 \rangle \langle v_2^* | v_1 \rangle . \tag{4.23}$$

Then each element of $\mathfrak{G}_2$ defines a linear functional on $\mathfrak{G}_{-2}$ and each element of $\mathfrak{G}_{-2}$ defines a linear functional on $\mathfrak{G}_2$. Note that

$$\langle [v_1^*, v_2^*], [x_i \cdot v_1, x_i \cdot v_2] \rangle = \langle v_1^* x_i v_1 \rangle \langle v_2^* x_i v_2 \rangle - \langle v_1^* x_i v_2 \rangle \langle v_2^* x_i v_1 \rangle \tag{4.24}$$

so if $\sum_k \sum_{i \in I} [x_i \cdot v_1^k, x_i \cdot v_2^k] = 0$ then $\sum [v_1^k, v_2^k] \in \mathfrak{I}_2^-$. To show the converse, supposing that

$$\sum_k \sum_{i \in I} [x_i \cdot v_1^k, x_i \cdot v_2^k] \neq 0 \tag{4.25}$$

one may use dual bases of V and V^* to find elements $v_{1m}^*, v_{2m}^* \in V^*$ such that

$$\left\langle \sum_m [v_{1m}^*, v_{2m}^*], \sum_k \sum_{i \in I} [x_i \cdot v_1^k, x_i \cdot v_2^k] \right\rangle \neq 0 . \tag{4.26}$$

We have established

Lemma 4.5. *If $v_1^k, v_2^k \in V = \mathfrak{G}_{-1}$ and $\{x_i | i \in I\}$ is any orthonormal basis of $\mathfrak{G}_0 = \mathfrak{g}_0^e$, then $\sum_k [v_1^k, v_2^k] \in \mathfrak{I}_2^-$ if and only if $0 = \sum_k \sum_{i \in I} [x_i \cdot v_1^k, x_i \cdot v_2^k]$.*

Corollary 4.6. *We have $[v_0, [v_0, f_2]] \in \mathfrak{I}_2^-$.*

Proof. We must show $0 = \sum_{i \in I} [x_i \cdot v_0, x_i \cdot [v_0, f_2]]$. Using the orthonormal basis (3.17)–(3.19) and the formulas (4.4) we obtain after regrouping

$$\sum_{0 < k \in \mathbb{Z}} ([e(-k) \cdot v_0, f(k) \cdot [v^0, f_2]] + [f(-k) \cdot v_0, e(k) \cdot [v_0, f_2]]$$
$$+ [h(-k) \cdot v_0, h(k) \cdot [v_0, f_2]]) - [c \cdot v_0, d \cdot [v_0, f_2]] - [d \cdot v_0, c \cdot [v_0, f_2]] .$$

Since

$$c \cdot [v_0, f_2] = [c \cdot v_0, f_2] + [v_0, [c \cdot f_2]] = [v_0, f_2]$$

and

$$d \cdot [v_0, f_2] = [d \cdot v_0, f_2] + [v_0, [d, f_2]] = 0$$

the last two terms equal $[v_0,[v_0,f_2]]$. We have

$$\begin{aligned} f(k)\cdot[v_0,f_2] &= [f(k)\cdot v_0,f_2]+[v_0,[f(k),f_2]] \\ &= 0+[v_0,-h(k-1)+k\delta_{k,1}c] \end{aligned}$$

and

$$\begin{aligned} e(k)\cdot[v_0,f_2] &= [e(k)\cdot v_0,f_2]+[v_0,[e(k),f_2]] \\ &= 0+[v_0,[e(k),e(-1)]]=0 \end{aligned}$$

and

$$\begin{aligned} h(k)\cdot[v_0,f_2] &= [h(k)\cdot v_0,f_2]+[v_0,[h(k),f_2]] \\ &= 0+[v_0,2e(k-1)]=0 \quad \text{for} \quad k>0. \end{aligned}$$

We finally obtain

$$[e(-1)\cdot v_0,[v_0,-h(0)+c]]+[v_0,[v_0,f_2]]=[f_2\cdot v_0,-v_0]+[v_0,[v_0,f_2]]=0.$$

Theorem 4.7. $\mathfrak{F}\approx\bar{\mathfrak{G}}(\mathfrak{g}_0^e,V)$.

Proof. From Corollary 4.6 we have that $[v_0,[v_0,f_2]]\in\mathfrak{I}_2^-$ and from the symmetry between $\mathfrak{I}_2^-$ and $\mathfrak{I}_2^+$ we also have that $[v_0^*,[v_0^*,e_2]]\in\mathfrak{I}_2^+$. This shows that all of relations (2.2) hold in $\bar{\mathfrak{G}}(\mathfrak{g}_0^e,V)$, so by the simplicity of $\mathfrak{F}$ we have the isomorphism.

We would like to use Theorem 4.7 to say more about the root multiplicities of $\mathfrak{F}$. It is sufficient to understand the negative roots $\alpha=-n_1\alpha_1-n_2\alpha_2-n_3\alpha_3$, whose root space $\mathfrak{F}^\alpha$ is contained in $\mathfrak{F}_{-n_3}$. Recalling the notation of Proposition 2.1, we associate root α with $v(\alpha)=\begin{pmatrix} m_2 & m_1 \\ m_1 & m_3 \end{pmatrix}\in S_2(\mathbb{Z})$ and say that α is on level $m_3=-n_3$. From Proposition 2.1 we know that α is real if and only if $\det(v(\alpha))=-1$, in which case its multiplicity is 1. Any root β W-conjugate to an α on level 0 also has multiplicity 1. Aside from the real roots, any such root β is on the cone $\det(v(\beta))=0$. Because the ideal $\mathfrak{I}$ has trivial intersection with $\mathfrak{G}_1$ and $\mathfrak{G}_{-1}$ we have $\mathfrak{F}_1\approx V^*$ with lowest weight α_3 and $\mathfrak{F}_{-1}\approx V$ with highest weight $-\alpha_3$. Thus, we have the following

Corollary 4.8. *Any root β W-conjugate to an α on level ±1 has multiplicity $p(\det(v(\beta))+1)$ where p denotes the classical partition function.*

We would like to obtain a closed formula for the multiplicities of level ±2 roots. First we will show that $\mathfrak{I}_2^-$ is the irreducible standard $\mathfrak{g}_0^e$-module with highest weight vector $[v_0,[v_0,f_2]]$ of weight $2\omega_1^0-\omega_3=-2\alpha_3-\alpha_2$. Suppose that

$$v=\sum_{m=1}^{r}[v_1^m,v_2^m]\in\mathfrak{G}_{-2} \tag{4.27}$$

is a highest weight vector for $\mathfrak{g}_0^e$ of weight $\lambda=n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*$, so $n_2=2$ and $n_1=0$ or $n_1=2$. Using the orthonormal basis (3.17)–(3.19) and the result of Lemma 4.5 one may write out an expression $E(v)$ such that $v\in\mathfrak{I}_2^-$ if and only if $E(v)=0$. If

one uses (3.33) to write out $C\cdot v$ one obtains

$$C\cdot v=\sum_{m=1}^{r}[C\cdot v_1^m, v_2^m]+\sum_{m=1}^{r}[v_1^m, C\cdot v_2^m]+2E(v). \tag{4.28}$$

Thus we have $v\in\mathfrak{I}_2^-$ if and only if

$$C\cdot v=\sum_{m=1}^{r}([C\cdot v_1^m, v_2^m]+[v_1^m, C\cdot v_2^m]). \tag{4.29}$$

Since $v_1^m, v_2^m\in V^{-\alpha_3}$ and $-\alpha_3=\gamma_2^*-\gamma_3^*$ the scalar by which C acts on v_1^m and v_2^m is given by (3.34) to be zero. So $v\in\mathfrak{I}_2^-$ if and only if $C\cdot v=0$. From (3.84) and (3.85) this happens if and only if $\lambda=2\gamma_1^*+2\gamma_2^*-\gamma_3^*=2\omega_1^0-\omega_3=-2\alpha_3-\alpha_2$, so this is the only highest weight of a $\mathfrak{g}_0^e$-module in $\mathfrak{I}_2^-$. Corollary 4.6 shows that $[v_0,[v_0,f_2]]\in\mathfrak{I}_2^-$ is a vector of that weight.

We identify $\mathfrak{G}_{-2}$ with the wedge product $V\wedge V$ and with the antisymmetric tensors $A(V)$ in $V\otimes V$. This is quite natural since $v_1\wedge v_2=-v_2\wedge v_1$ and the action of $x\in\mathfrak{F}_0^e$ on $v_1\wedge v_2$ is given by

$$x\cdot(v_1\wedge v_2)=(x\cdot v_1)\wedge v_2+v_1\wedge(x\cdot v_2). \tag{4.30}$$

If $v_1, v_2\in\mathfrak{G}_{-1}=V$ then $[v_1, v_2]$ corresponds to $v_1\wedge v_2$ and the Jacobi identity

$$[x,[v_1,v_2]]=[[x,v_1],v_2]+[v_1,[x,v_2]] \tag{4.31}$$

corresponds to (4.30). The correspondence with $A(V)$ is given by

$$v_1\wedge v_2\leftrightarrow v_1\otimes v_2-v_2\otimes v_1 \tag{4.32}$$

which is clearly a $\mathfrak{g}_0^e$-module isomorphism.

Therefore, $\mathfrak{F}_{-2}\approx\bar{\mathfrak{G}}_{-2}=\mathfrak{G}_{-2}/\mathfrak{I}_2^-$ is just $A(V)$ with the first standard module removed (see Corollary 3.5). If we denote the root multiplicity of α by $M(v(\alpha))=M\begin{pmatrix}m_2 & m_1\\ m_1 & m_3\end{pmatrix}$ then the character of $\mathfrak{F}_{-2}$ is

$$e^{2\gamma_2^*}\left[\left(\sum_{m\geq 0}M\begin{pmatrix}m & 0\\ 0 & 2\end{pmatrix}t^{2m}\right)\left(\sum_{k\in\mathbb{Z}}e^{2k\alpha_1}t^{4k^2}\right)\right.$$
$$\left.+\left(\sum_{m\geq 0}M\begin{pmatrix}m & 1\\ 1 & 2\end{pmatrix}t^{2m-1}\right)\left(\sum_{k\in\mathbb{Z}}e^{(2k+1)\alpha_1}t^{(2k+1)^2}\right)\right], \tag{4.33}$$

where $t^2=q$. The principal specialization of this character is

$$\mathrm{ch}(\mathfrak{F}_{-2})=e^{2\gamma_2^*}\frac{\phi(u^4)^2}{\phi(u^2)}\sum_{n\geq 0}M(n-1)u^{n-1}, \tag{4.34}$$

where we have written more briefly

$$M(2m)=M\begin{pmatrix}m & 0\\ 0 & 2\end{pmatrix}\quad\text{and}\quad M(2m-1)=M\begin{pmatrix}m & 1\\ 1 & 2\end{pmatrix}. \tag{4.35}$$

Together with Corollary 3.5 this gives us the root multiplicities on the second level.

Corollary 4.9. *We have*

$$\sum_{n\geqq 0} M(n-1)u^n = \frac{\phi(u^2)}{\phi(u)\phi(u^4)}\frac{u^{-3}}{2}\left[\prod_{j\geqq 1}(1+u^{2j-1}) - \prod_{j\geqq 1}(1-u^{2j-1}) - 2u\right]$$

$$= \left[\sum_{n\geqq 0} p(n)u^n\right]\left[\prod_{j\geqq 1}(1-u^{4j-2})\right]$$

$$\cdot\frac{u^{-3}}{2}\left[\prod_{j\geqq 1}(1+u^{2j-1}) - \prod_{j\geqq 1}(1-u^{2j-1}) - 2u\right]$$

$$= \left[\sum_{n\geqq 0} p(n)u^n\right][1-u^{20}+u^{22}-u^{24}+u^{26}-2u^{28}+2u^{30}-2u^{32}+\ldots]. \tag{4.36}$$

It is quite remarkable that the first nonzero term beyond 1 in the above expansion is u^{20}. For that reason we have $M(n-1)=p(n)$ for $0\leqq n\leqq 19$, but $M(19)=p(20)-p(0)$.

We were led to our work on the root multiplicities of $\mathfrak{F}$ by computer generated data given to us by Moody. His program, based on the multiplicity formula in [36], gave data consistent with the formula

$$\dim(\mathfrak{F}^\alpha) = p(\det(N)+1), \tag{4.37}$$

where $v(\alpha)=N\in S_2(\mathbb{Z})$ from Sect. 2. This formula fails, however, for most roots not Weyl conjugate to roots from $\mathfrak{F}_{-1}+\mathfrak{F}_0^e+\mathfrak{F}_1$. A program which we developed computed multiplicities of some roots from $\mathfrak{F}_2$, $\mathfrak{F}_3$, and $\mathfrak{F}_4$. The smallest counterexample to (4.37) is when $\alpha=11\alpha_1+12\alpha_2+2\alpha_3$, corresponding to $M(19)$, for which $\dim(\mathfrak{F}^\alpha)=626$ instead of $p(20)=627$. We suspect that the root multiplicities of $\mathfrak{F}$ have some interesting number-theoretical meaning connected with ideal classes of imaginary quadratic fields. It is interesting to note that the roots α in $\mathfrak{F}_{-1}+\mathfrak{F}_0^e+\mathfrak{F}_1$ for which (4.37) holds correspond under v to binary forms associated with principal ideals.

Root multiplicities are also significant because they occur in the Weyl-Kac denominator formula for $\mathfrak{F}$.

Theorem 4.10. *The Weyl-Kac denominator formula for* $\mathfrak{F}$ *is*

$$\sum_{g\in \mathrm{PGL}_2(\mathbb{Z})} \det(g)e^{2\pi i \operatorname{Tr}(gPg^t\mathfrak{Z})}$$

$$= e^{2\pi i \operatorname{Tr}(P\mathfrak{Z})}\prod_{0\leqq N\in S_2(\mathbb{Z})}(1-e^{2\pi i\operatorname{Tr}(N\mathfrak{Z})})^{\mathrm{Mult}(N)}\prod_{N\in v(R_W^-)}(1-e^{2\pi i\operatorname{Tr}(N\mathfrak{Z})}),$$

where $P=v(\varrho)=\begin{pmatrix}3 & 1/2\\ 1/2 & 2\end{pmatrix}$, $\mathfrak{Z}=\begin{pmatrix}z_3 & z_1\\ z_1 & z_2\end{pmatrix}$

$$v(R_W^-)=\left\{N=\begin{pmatrix}n_3 & n_1\\ n_1 & n_2\end{pmatrix}\in S_2(\mathbb{Z})\,|\,n_2n_3-n_1^2=-1, n_1\leqq n_2+n_3,\ 0\leqq n_2+n_3, 0\leqq n_2\right\}.$$

Proof. The Weyl-Kac denominator formula generally is

$$\sum_{w\in W}\det(w)e^{w\varrho-\varrho} = \prod_{\alpha\in R^-}(1-e^\alpha)^{m_\alpha}$$

where m_α is the dimension of the α root space. We have written this out for algebra $\mathfrak{F}$ as a function on the Siegel domain H_2 using the results of Sects. 2 and 3.

The exponents Mult(N) are the only difficult part of this identity since we have not given them all explicitly. They are, in fact, determined by the form of the identity, and Moody and Berman [36] have shown in general how to obtain a formula for root multiplicities from the denominator formula. Since our formula involves a summation over $\mathrm{PGL}_2(\mathbb{Z})$ it may be of some interest to analysts. We challenge an analyst to prove this identity without the theory of Kac-Moody Lie algebras, or to shed more light on the meaning of the exponents Mult(N).

5. Theta Series and the Space of $\mathrm{SL}_2(\mathbb{Z})$-Invariant $\mathfrak{F}_0^e$-Characters

The Weyl-Kac character formula for $\mathfrak{F}_0^e$ may be written as

$$\chi^\lambda = \frac{\sum_{w\in W_0} \det(w) e^{w(\lambda+\varrho_0)}}{\sum_{w\in W_0} \det(w) e^{w(\varrho_0)}}, \tag{5.1}$$

where $\varrho_0 = \omega_1^0 + \omega_2^0$. We will use the notation χ^λ for the standard $\mathfrak{F}_0^e$-character with dominant highest weight λ to distinguish it from the standard $\mathfrak{F}$-character X^λ. For such λ, $\chi^\lambda \in \mathscr{E}_0$. The dominant weights of $\mathfrak{F}_0^e$ are denoted by

$$P_0^{++} = \{n_1\gamma_1^* + n_2\gamma_2^* + t\gamma_3^* | t\in\mathbb{R}^+, n_1, n_2\in\mathbb{Z}, n_2 \geqq n_1 \geqq 0\}. \tag{5.2}$$

The Weyl group W_0 decomposes into two cosets,

$$W_0^+ = \{(r_1r_2)^i | i\in\mathbb{Z}\} \tag{5.3}$$

and

$$W_0^- = \{(r_1r_2)^i r_1 | i\in\mathbb{Z}\}, \tag{5.4}$$

the even and odd elements, respectively. So we have

$$\begin{aligned}\sum_{w\in W_0} \det(w) e^{w(n_1\gamma_1^* + n_2\gamma_2^*)} &= \sum_{i\in\mathbb{Z}} (e^{T^i(n_1\gamma_1^* + n_2\gamma_2^*)} - e^{T^i(-n_1\gamma_1^* + n_2\gamma_2^*)}) \\ &= \sum_{i\in\mathbb{Z}} e^{(2in_2+n_1)\gamma_1^* + n_2\gamma_2^* + i(in_2+n_1)\gamma_3^*} \\ &\quad - \sum_{i\in\mathbb{Z}} e^{(2in_2-n_1)\gamma_1^* + n_2\gamma_2^* + i(in_2-n_1)\gamma_3^*}.\end{aligned} \tag{5.5}$$

Define the following formal theta series,

$$\Theta_{n,m} = e^{m\gamma_2^*} \sum_{j\in\mathbb{Z}+(n/2m)} e^{2jm\gamma_1^* + j^2m\gamma_3^*} \tag{5.6}$$

for $m>0$ and $n\in\mathbb{R}$. If $0<m\in\mathbb{Z}$ and $n\in\mathbb{Z}$ then $\Theta_{n,m}\in\mathscr{E}$. Then one sees that (5.5) equals

$$e^{\left(\frac{-n_1^2}{4n_2}\right)\gamma_3^*}(\Theta_{n_1,n_2} - \Theta_{-n_1,n_2}). \tag{5.7}$$

Since

$$\lambda = m_1\omega_1^0 + m_2\omega_2^0 = m_1\gamma_1^* + (m_1+m_2)\gamma_2^* \tag{5.8}$$

and

$$\lambda+\varrho_0=(m_1+1)\gamma_1^*+(m_1+m_2+2)\gamma_2^* \tag{5.9}$$

we have

$$\chi^\lambda=e^{-\left(\frac{(m_1+1)^2}{4(m_1+m_2+2)}-\frac{1}{8}\right)\gamma_3^*}\frac{(\Theta_{m_1+1,m_1+m_2+2}-\Theta_{-m_1-1,m_1+m_2+2})}{(\Theta_{1,2}-\Theta_{-1,2})}. \tag{5.10}$$

As was shown in Sect. 3, we may consider the formal theta series to be functions of three complex variables. Then we obtain from (3.92) and (5.6)

$$\Theta_{n,m}(z_1,z_2,z_3)=e^{2\pi imz_2}\sum_{j\in\mathbb{Z}+(n/2m)}e^{4\pi ijmz_1+2\pi ij^2mz_3}. \tag{5.11}$$

Lemma 5.1. *We have*

$$\Theta_{n,m}(z_1,z_2,z_3)=\frac{e^{2\pi im\left(z_2-\frac{z_1^2}{z_3}\right)}}{\sqrt{-2imz_3}}\sum_{j\in\frac{1}{2m}\mathbb{Z}}e^{2\pi i(-2jm(-z_1/z_3)+j^2m(-1/z_3))}e^{2\pi ijn}$$

$$=\frac{1}{\sqrt{-2imz_3}}\sum_{\ell=-(m-1)}^{m}\Theta_{\ell,m}\left(\frac{-z_1}{z_3},z_2-\frac{z_1^2}{z_3},\frac{-1}{z_3}\right)e^{-\pi in\ell/m}.$$

Proof. We use the Poisson summation formula which says that if $f(x)$ is a continous function of a real variable x and if

$$\hat{f}(y)=\int_{\mathbb{R}}e^{-2\pi ixy}f(x)dx \tag{5.12}$$

then

$$\sum_{n\in\mathbb{Z}}f(n)=\sum_{n\in\mathbb{Z}}\hat{f}(n). \tag{5.13}$$

If we define

$$f(x)=e^{4\pi imxz_1+2\pi i(mx^2+nx)z_3} \tag{5.14}$$

for $x\in\mathbb{R}$ then we find from (5.11) that

$$\Theta_{n,m}(z_1,0,z_3)=e^{2\pi inz_1+2\pi i(n^2/4m)z_3}\sum_{j\in\mathbb{Z}}f(j). \tag{5.15}$$

If we let $x_0=(-y+2mz_1+nz_3)/(2mz_3)$ then

$$\hat{f}(y)=e^{-2\pi imz_3x_0^2}\int_{\mathbb{R}}e^{2\pi imz_3(x+x_0)^2}dx \tag{5.16}$$

$$=\frac{1}{\sqrt{-2imz_3}}e^{-2\pi imz_3x_0^2} \tag{5.17}$$

since

$$\int_{\mathbb{R}}e^{-pu^2}du=\sqrt{(\pi/p)}.$$

After substitution of the expression for x_0 into (5.17) and some algebraic manipulation, the first part of the lemma can be obtained.

For the second part, note that

$$\Theta_{-\ell,m}\left(\frac{-z_1}{z_3}, z_2-\frac{z_1^2}{z_3}, \frac{-1}{z_3}\right)$$
$$= e^{2\pi i m(z_2-z_1^2/z_3)} \sum_{k\in\mathbb{Z}-(\ell/2m)} e^{4\pi i k m(-z_1/z_3)+2\pi i k^2 m(-1/z_3)} \tag{5.18}$$

and that

$$\frac{1}{2m}\mathbb{Z} = \bigcup_{\ell=-(m-1)}^{m}\left\{\mathbb{Z}+\frac{\ell}{2m}\right\}. \tag{5.19}$$

So using the first part we can write

$$\Theta_{n,m}(z_1,z_2,z_3)$$
$$= \frac{e^{2\pi i m(z_2-z_1^2/z_3)}}{\sqrt{-2imz_3}} \sum_{\ell=-m}^{m-1} \sum_{k\in\mathbb{Z}+(\ell/2m)} e^{-4\pi i m k(-z_1/z_3)+2\pi i m k^2(-1/z_3)} e^{2\pi i n k}.$$

By changing k to $-k$ and by using (5.18) we obtain the result.

Definition. Let $J(z_1,z_2,z_3)=\Theta_{1,2}(z_1,z_2,z_3)-\Theta_{-1,2}(z_1,z_2,z_3)$.

Corollary 5.2. $J(z_1,z_2,z_3)=\dfrac{-i}{\sqrt{-iz_3}} J\left(\dfrac{-z_1}{z_3}, z_2-\dfrac{z_1^2}{z_3}, \dfrac{-1}{z_3}\right)$.

Remark. Looijenga [26] has introduced the formal theta series for arbitrary affine Kac-Moody algebras and proved the analog of Corollary 5.2 in that generality.

Definition. If $\lambda=m_1\omega_1^0+m_2\omega_2^0+m_3\gamma_3^*=m_1\gamma_1^*+(m_1+m_2)\gamma_2^*+m_3\gamma_3^*\in\mathfrak{h}^*$ then we call λ a weight of level m_1+m_2.

If $V\in\mathscr{C}$ is an $\mathfrak{F}$-module then it is also an $\mathfrak{F}_0^e$-module ($\mathscr{C}\subseteq\mathscr{C}_0$) and decomposes into the direct sum of standard $\mathfrak{F}_0^e$-modules, each of which has character $\chi^\lambda=e^{m_3\gamma_3^*}\chi^{m_1\omega_1^0+m_2\omega_2^0}\in\mathscr{E}_0$ of level $M=m_1+m_2\geqq 0$.
Since each such λ is in P^+ we can write

$$X(V)=\sum_{M\geqq 0}\sum_{\substack{m_1+m_2=M\\ m_1,m_2\in\mathbb{Z}^+}} \chi^{m_1\omega_1^0+m_2\omega_2^0} \sum_{m_3\in\mathbb{Z}^+} A(m_1,m_2,m_3)e^{m_3\gamma_3^*}, \tag{5.20}$$

where $A(m_1,m_2,m_3)$ is the multiplicity of χ^λ in $X(V)$.

Definition. Let the level of $\mathfrak{F}$-module $V\in\mathscr{C}$ be the least positive integer M such that $A(m_1,m_2,m_3)\neq 0$ for some m_1,m_2,m_3 such that $m_1+m_2=M$.

The level of V is just the smallest level any weight λ such that χ^λ occurs in the decomposition of $X(V)$. If V^λ is the standard $\mathfrak{F}$-module with dominant highest weight

$$\lambda=n_1\omega_1+n_2\omega_2+n_3\omega_3=n_1\gamma_1^*+(n_1+n_2)\gamma_2^*+(n_1+n_2+n_3)\gamma_3^*$$

then all weights of V^λ are of the form

$$\lambda-p_1\alpha_1-p_2\alpha_2-p_3\alpha_3=\lambda-2(p_1-p_2)\gamma_1^*+p_3\gamma_2^*+(p_2-p_3)\gamma_3^*$$

for

$$p_1,p_2,p_3\in\mathbb{Z}^+.$$

So the level of any such weight is $m_1+m_2+p_3 \geqq m_1+m_2$ and the level of any $\mathfrak{F}_0^e$-module occuring in the decomposition of V^λ must be at least m_1+m_2. The vectors of weight λ generate an $\mathfrak{F}_0^e$-module of level m_1+m_2 in V^λ, so the level of V^λ is exactly m_1+m_2, the level of λ.

If we decompose the character of $\mathfrak{F}$-module $V \in \mathscr{C}$ of level M_0 into standard $\mathfrak{F}_0^e$-characters we get

$$X(V)=\sum_{M \geqq M_0} \sum_{j=0}^{M} \chi^{j\omega_1^0+(M-j)\omega_2^0} \sum_{m \in \mathbb{Z}^+} A(j, M-j, m) e^{m\gamma_3^*}. \tag{5.21}$$

Definition. An $\mathfrak{F}_0^e$-module in $\mathscr{C}_0$ is said to be of level M of each of its weights is of level M.

Note that the action of $\mathfrak{F}_0^e$ on any weight can only move it by a linear combination of α_1 and α_2 (γ_1^* and γ_3^*) so the γ_2^* coefficient, which equals the level, is unchanged. Thus, the weights of any standard $\mathfrak{F}_0^e$-module lie entirely on one level, and an $\mathfrak{F}_0^e$-module in $\mathscr{C}_0$ is of level M if and only if all of its irreducible standard $\mathfrak{F}_0^e$-components have highest weights of level M. There are $M+1$ isomorphism classes of such standard $\mathfrak{F}_0^e$-modules, representative characters being

$$\{\chi^{j\omega_1^0+(M-j)\omega_2^0} | 0 \leqq j \leqq M,\ j \in \mathbb{Z}\}. \tag{5.22}$$

Given a standard $\mathfrak{F}$-module V^λ, $\lambda \in P^{++}$, of level M, we have a unique standard $\mathfrak{F}_0^e$-module v^λ of level M whose highest weight is λ. In fact, v^λ is exactly the lowest level $\mathfrak{F}_0^e$-component of V^λ.

Conversely, given a standard $\mathfrak{F}_0^e$-module v^λ of level M with $\lambda \in P^{++}$ (note the restriction on the γ_3^* coefficient) there is a unique standard $\mathfrak{F}$-module, V^λ, whose lowest level $\mathfrak{F}_0^e$-component is v^λ.

It is easy to perform this lifting and descent for sums of standard characters. We will see later the significance of this correspondence for the theory of Siegel modular forms.

Definition. We shall call an irreducible standard $\mathfrak{F}_0^e$-module and its character $\mathfrak{F}$-dominant whenever the highest weight of the module is in P^{++}. We shall call an arbitrary $\mathfrak{F}_0^e$-module and its character $\mathfrak{F}$-dominant whenever each irreducible standard component is $\mathfrak{F}$-dominant.

We shall denote by $\mathscr{C}'_0$ the subcategory of $\mathscr{C}_0$ consisting of those $\mathfrak{F}_0^e$-modules which are $\mathfrak{F}$-dominant. The above discussion makes clear the following.

Proposition 5.3. *There is a natural one-to-one correspondence between the category of $\mathfrak{F}$-modules $\mathscr{C}$ and the category of $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-modules $\mathscr{C}'_0$, which induces a one-to-one correspondence between $\mathfrak{F}$-characters in $\mathscr{C}$ and $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-characters in $\mathscr{C}_0$.*

Remark. Given a level M $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-character, Proposition 5.3 tells us that there exist $\mathfrak{F}_0^e$-characters on the levels above M naturally associated with it. In theory we have operators which produce these higher level characters from the one on level M. If the level M character is standard irreducible then these operators construct the corresponding standard irreducible $\mathfrak{F}$-character. Specific formulas for such operators would be very interesting. For the case of level $M=1$ they might be related to the Hecke operators which will be discussed later in Sects. 6 and 7.

Consider an arbitrary level M $\mathfrak{F}_0^e$-character of the form

$$\chi_M = \sum_{j=0}^{M} \chi^{j\omega_1^0 + (M-j)\omega_2^0} \sum_{m \in \mathbb{Z}^+} A_j(m) e^{m\gamma_3^*}. \tag{5.23}$$

We understand from (3.92) that χ_M may be viewed as a function of z_1, z_2, z_3. Using (5.10) we can write

$$\begin{aligned}\chi_M(z_1, z_2, z_3) &= \sum_{j=0}^{M} \chi^{j\omega_1^0 + (M-j)\omega_2^0}(z_1, z_2, z_3) \sum_{m \geqq 0} A_j(m) e^{2\pi i m z_3} \\ &= \sum_{j=0}^{M} \bar{\chi}^{j\omega_1^0 + (M-j)\omega_2^0}(z_1, z_2, z_3) c_j(z_3),\end{aligned} \tag{5.24}$$

where

$$\bar{\chi}^{j\omega_1 + (M-j)\omega_2^0}(z_1, z_2, z_3) = \frac{\Theta_{j+1, M+2}(z_1, z_2, z_3) - \Theta_{-j-1, M+2}(z_1, z_2, z_3)}{\Theta_{1,2}(z_1, z_2, z_3) - \Theta_{-1,2}(z_1, z_2, z_3)} \tag{5.25}$$

and

$$c_j(z_3) = e^{-\left(\frac{(j+1)^2}{4(M+2)} - \frac{1}{8}\right)2\pi i z_3} \sum_{m \geqq 0} A_j(m) e^{2\pi i m z_3}. \tag{5.26}$$

Proposition 5.4. *Let $\chi_M(z_1, z_2, z_3)$ be an $\mathfrak{F}_0^e$-character of level M written in the form of* (5.24). *Let $k \in \mathbb{C}$. Then*

$$\chi_M(z_1, z_2, z_3) = (-z_3)^{-k} \chi_M\left(\frac{-z_1}{z_3}, z_2 - \frac{z_1^2}{z_3}, \frac{-1}{z_3}\right)$$

if and only if

$$c_n\left(\frac{-1}{z_3}\right)(-z_3)^{-k} = \sqrt{\frac{2}{M+2}} \sum_{j=0}^{M} c_j(z_3) \sin\left(\frac{\pi(j+1)(n+1)}{M+2}\right) \quad \text{for} \quad 0 \leqq n \leqq M.$$

Proof. For $n \in \frac{1}{2(M+2)}\mathbb{Z}$ define

$$c_n'(z_3) = \begin{cases} c_{m-1}(z_3) & \text{if } 2(M+2)n \equiv m (\mathrm{mod}\, 2(M+2)), 1 \leqq m \leqq M+1 \\ -c_{m-1}(z_3) & \text{if } 2(M+2)n \equiv -m (\mathrm{mod}\, 2(M+2)), 1 \leqq m \leqq M+1 \\ 0 & \text{if } 2(M+2)n \equiv 0, M+2 (\mathrm{mod}\, 2(M+2)). \end{cases}$$

We have

$$\chi_M(z_1, z_2, z_3) = \frac{\sum_{j=0}^{M} e^{2\pi i (M+2) z_2} (S_j^+ - S_j^-) c_j(z_3)}{J(z_1, z_2, z_3)}, \tag{5.27}$$

where

$$S_j^{\pm} = \sum_{r \in \mathbb{Z} \pm \frac{j+1}{2(M+2)}} e^{4\pi i r (M+2) z_1 + 2\pi i r^2 (M+2) z_3}.$$

This can be rewritten as

$$\frac{e^{2\pi i(M+2)z_2}\sum\limits_{r\in\frac{\mathbb{Z}}{2(M+2)}} c'_r(z_3)e^{4\pi ir(M+2)z_1+2\pi ir^2(M+2)z_3}}{J(z_1,z_2,z_3)} \tag{5.28}$$

because $\mathbb{Z}\pm\frac{j+1}{2(M+2)}\subseteq\frac{\mathbb{Z}}{2(M+2)}=\left\{p+\frac{\ell}{2(M+2)}\middle| p\in\mathbb{Z}, -(M+1)\leqq\ell\leqq M+2\right\}$ so that if $r=p\pm\frac{j+1}{2(M+2)}$ then the corresponding term in the summation is multiplied by $c_j(z_3)$, $-c_j(z_3)$ or 0 according to the following rules. We have

$$2(M+2)r=2(M+2)p\pm(j+1)$$

with $1\leqq j+1\leqq M+1$ so $\ell=j+1$. If $2(M+2)r\equiv\ell(\mathrm{mod}\,2(M+2))$ for $1\leqq\ell\leqq M+1$ then we want $c_j(z_3)=c_{\ell-1}(z_3)=c'_r(z_3)$. If $2(M+2)r\equiv-\ell(\mathrm{mod}\,2(M+2))$ for $1\leqq\ell\leqq M+1$ then we want $-c_j(z_3)=-c_{\ell-1}(z_3)=c'_r(z_3)$. The only other possibilities for $2(M+2)r(\mathrm{mod}\,2(M+2))$ are 0 and $M+2$, for which we want $0=c'_r(z_3)$.

By using Lemma 5.1 on the theta functions in the numerator of (5.25) and Corollary 5.2 on the denominator, we can obtain the following expression for $\chi_M(z_1,z_2,z_3)$:

$$\frac{\dfrac{e^{2\pi i(M+2)(z_2-z_1^2/z_3)}}{\sqrt{-2(M+2)iz_3}}\sum\limits_r\left[\sum\limits_{j=0}^{M}(e^{2\pi ir(j+1)}-e^{-2\pi ir(j+1)})c_j(z_3)\right]e^{-4\pi ir(M+2)\frac{-z_1}{z_3}+2\pi ir^2(M+2)\frac{-1}{z_3}}}{\dfrac{-i}{\sqrt{-iz_3}}J\left(\dfrac{-z_1}{z_3},z_2-\dfrac{z_1^2}{z_3},\dfrac{-1}{z_3}\right)}, \tag{5.29}$$

where the summations over r are taken over $\frac{\mathbb{Z}}{2(M+2)}$.

It is easy to check that

$$\sum_{j=0}^{M}(e^{2\pi ir(j+1)}-e^{-2\pi ir(j+1)})c_j(z_3)=\sum_{\ell=0}^{2M+3}c'_{\left(\frac{\ell}{2(M+2)}\right)}(z_3)e^{2\pi i\ell r}. \tag{5.30}$$

Let (5.31) be the expression obtained from (5.29) by substitution of (5.30), and let (5.32) be the expression for $\chi_M\left(\frac{-z_1}{z_3},z_2-\frac{z_1^2}{z_3},\frac{-1}{z_3}\right)$ obtained by substitution of variables in (5.28). After changing r to $-r$ in (5.32) and using $c'_{-r}=-c'_r$ one sees that $\chi_M(z_1,z_2,z_3)=(-z_3)^{-k}\chi_M\left(\frac{-z_1}{z_3},z_2-\frac{z_1^2}{z_3},\frac{-1}{z_3}\right)$ holds if and only if

$$\frac{1}{i\sqrt{2(M+2)}}\sum_{\ell=-(M+1)}^{M+1}c'_{\left(\frac{\ell}{2(M+2)}\right)}(z_3)e^{2\pi i\frac{\ell n}{2(M+2)}}=(-z_3)^{-k}c'_{\left(\frac{n}{2(M+2)}\right)}\left(\frac{-1}{z_3}\right)\quad\text{for}\quad 1\leqq n\leqq 2(M+2). \tag{5.33}$$

Both sides (5.33) can be seen to be 0 when $n=0$ or $M+2$ and if (5.33) holds for some n, $1\leqq n\leqq M+1$, then it also holds for $-n$, so the condition (5.33) can be

considered just for $1 \leqq n \leqq M+1$. Writing out (5.33) in terms of $c_j(z_3)$ and $c_n\left(\frac{-1}{z_3}\right)$ we finally get

$$\sum_{j=0}^{M} c_j(z_3)\left(\frac{e^{2\pi i\frac{(j+1)(n+1)}{2(M+2)}} - e^{-2\pi i\frac{(j+1)(n+1)}{2(M+2)}}}{i\sqrt{2(M+2)}}\right) = (-z_3)^{-k} c_n\left(\frac{-1}{z_3}\right)$$

$$= \sqrt{\frac{2}{M+2}} \sum_{j=0}^{M} c_j(z_3) \sin\left(\frac{\pi(j+1)(n+1)}{M+2}\right) \quad \text{for} \quad 0 \leqq n \leqq M.$$

Definition. Let $\mathcal{M}_k$ be the complex vector space spanned by those $\mathfrak{F}_0^e$-characters of the form

$$\chi(z_1, z_2, z_3) = \sum_{M \geqq 0} \chi_M(z_1, z_2, z_3) \tag{5.34}$$

with $\chi_M(z_1, z_2, z_3)$ satisfying the conditions of Proposition 5.4. Then $\mathcal{M}_k$ is naturally graded according to the level of its corresponding modules:

$$\mathcal{M}_k = \sum_{M \geqq 0} \mathcal{M}_k(M), \tag{5.35}$$

where $\mathcal{M}_k(M)$ is the space spanned by the characters of level M modules satisfying Proposition 5.4.

By definition any $\chi(z_1, z_2, z_3) \in \mathcal{M}_k$ can be written in the form

$$\sum A(m_1, m_2, m_3) e^{2\pi i(m_1 z_1 + m_2 z_2 + m_3 z_3)}, \tag{5.36}$$

where the sum is taken over integers m_1, m_2, m_3 such that $4m_2 m_3 \geqq m_1^2$, and $m_2, m_3 \geqq 0$. So any such χ remains invariant under the translations $z_j \to z_j + 1$, $j = 1, 2, 3$. From the formulas $\chi(z_1, z_2, z_3 + 1) = \chi(z_1, z_2, z_3)$ and

$$\chi(z_1, z_2, z_3) = (-z_3)^{-k} \chi\left(\frac{-z_1}{z_3}, z_2 - \frac{z_1^2}{z_3}, \frac{-1}{z_3}\right)$$

one may compute that for any $\begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ we have

$$\chi(z_1, z_2, z_3) = (cz_3 + d)^{-k} \chi\left(\frac{z_1}{cz_3 + d}, z_2 - \frac{cz_1^2}{cz_3 + d}, \frac{az_3 + b}{cz_3 + d}\right) \tag{5.37}$$

because $\begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}$ and $\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ generate $\mathrm{SL}_2(\mathbb{Z})$.

Remark. Naturally we are interested in the dimension of the space $\mathcal{M}_k(M)$. The answer to this problem, surprisingly, is connected with the theory of genus 2 Siegel modular forms which will be discussed in Sect. 7.

Now let us find the form of characters in $\mathcal{M}_k(0)$ and $\mathcal{M}_k(1)$. From (5.24), since $\bar{\chi}^{0\omega_1^0 + 0\omega_2^0} = 1$, we have for an arbitrary $\chi_0(z_1, z_2, z_3)$ of level 0,

$$\chi_0(z_1, z_2, z_3) = c_0(z_3) = \sum_{m \geqq 0} A_0(m) e^{2\pi i m t} \in \mathcal{M}_k(0) \tag{5.38}$$

if and only if it satisfies

$$(cz_3+d)^{-k}c_0\left(\frac{az_3+b}{cz_3+d}\right)=c_0(z_3) \quad \text{for any} \quad \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z}). \tag{5.39}$$

If we denote for $h=0,1$ the classical Jacobi theta functions by

$$\Theta_h(z_1,z_3)=\sum_{j\in\mathbb{Z}} e^{2\pi i\left((2j+h)z_1+\left(j+\frac{h}{2}\right)^2 z_3\right)} \tag{5.40}$$

then

$$\Theta_{h,1}(z_1,z_2,z_3)=e^{2\pi i z_2}\Theta_h(z_1,z_3), \tag{5.41}$$

$$\bar{\chi}^{\omega_1^0}(z_1,z_2,z_3)=e^{2\pi i z_2}\eta^{-1}(e^{2\pi i z_3})\Theta_1(z_1,z_3), \tag{5.42}$$

and

$$\bar{\chi}^{\omega_2^0}(z_1,z_2,z_3)=e^{2\pi i z_2}\eta^{-1}(e^{2\pi i z_3})\Theta_0(z_1,z_3), \tag{5.43}$$

where $\eta(X)=X^{1/24}\phi(X)$ denotes the Dedekind η-function. The conditions for

$$\chi(z_1,z_2,z_3)=\bar{\chi}^{\omega_2^0}(z_1,z_2,z_3)c_0(z_3)+\bar{\chi}^{\omega_1}(z_1,z_2,z_3)c_1(z_3) \tag{5.44}$$

to be in $\mathscr{M}_k(1)$ are two equations which may be stated in vector notation:

$$(-z_3)^{-k}\begin{pmatrix} c_0\left(\frac{-1}{z_3}\right) \\ c_1\left(\frac{-1}{z_3}\right) \end{pmatrix}=\frac{1}{\sqrt{2}}\begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}\begin{pmatrix} c_0(z_3) \\ c_1(z_3) \end{pmatrix}. \tag{5.45}$$

This particular vectorial modular form has occurred in the investigations of Maass [30–32] into Siegel modular forms of genus 2 in connection with the "lifting" of elliptic modular forms. In order to make the connection between the theory of algebra $\mathfrak{F}$ and the theory of Siegel modular forms we must now review some classical definitions and results.

6. Siegel Modular Forms and Hecke Operators

We recall some classical facts about Siegel modular forms of genus $n\geqq 1$ which can be found in [1, 13, 40].

Definition. The Siegel domain (upper half-plane) of genus $n\geqq 1$ is the set of all symmetric $n\times n$ complex matrices having positive definite imaginary part:

$$H_n=\{\mathfrak{Z}=X+iY\in S_n(\mathbb{C})\,|\,Y>0\}. \tag{6.1}$$

Definition. The real symplectic group of genus n is

$$\mathrm{Sp}_n(\mathbb{R})=\{N\in \mathrm{GL}_{2n}(\mathbb{R})\,|\,NJN^t=J\},$$

where J is the $2n\times 2n$ matrix whose $n\times n$ blocks are

$$\begin{pmatrix} O_n & I_n \\ -I_n & O_n \end{pmatrix}. \tag{6.2}$$

The group $\mathrm{Sp}_n(\mathbb{R})$ acts on H_n by the analytic automorphism

$$N\cdot\mathfrak{Z}=(A\mathfrak{Z}+B)(C\mathfrak{Z}+D)^{-1} \tag{6.3}$$

for $N=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\in\mathrm{Sp}_n(\mathbb{R})$ and $\mathfrak{Z}\in H_n$.

An important discrete subgroup of $\mathrm{Sp}_n(\mathbb{R})$ is the subgroup of integral symplectic matrices;

$$\mathrm{Sp}_n(\mathbb{Z})=\mathrm{Sp}_n(\mathbb{R})\cap M_{2n}(\mathbb{Z}) \tag{6.4}$$

which is called the *Siegel modular group of genus n.* One sometimes denotes it by Γ_n.

A *Siegel modular form* of genus n and weight k (natural number) is any function $f(\mathfrak{Z})$ which is holomorphic on H_n and satisfies the two conditions,

(I) For every $N=\begin{pmatrix}A & B\\ C & D\end{pmatrix}\in\mathrm{Sp}_n(\mathbb{Z})$ and every $\mathfrak{Z}\in H_n$, $\det(C\mathfrak{Z}+D)^{-k}f(N\cdot\mathfrak{Z})=f(\mathfrak{Z})$, in which case we say f is invariant under N,

(II) Function $f(\mathfrak{Z})$ is bounded in every domain of the form

$$\{\mathfrak{Z}=X+iY\in H_n \mid Y\geqq cI_n, c>0\}.$$

For $n>1$ Condition (II) is known to follow from (I) and the holomorphism of f. The set of all Siegel modular forms of genus n and weight k forms a finite dimensional complex vector space, denoted $\mathfrak{M}_k^n$.

Every $f\in\mathfrak{M}_k^n$ has a Fourier expansion

$$f(\mathfrak{Z})=\sum_{N\geqq 0}a(N)e^{2\pi i\,\mathrm{Tr}(N\mathfrak{Z})}, \tag{6.5}$$

where N runs over the positive semidefinite matrices in the set

$$S_n'(\mathbb{Z})=\{N=(n_{ij})\in M_n(\mathbb{Q}) \mid N=N^t, t_{ii}, 2t_{ij}\in\mathbb{Z} \quad \text{for} \quad i\neq j\}. \tag{6.6}$$

If we apply Condition (I) with an element $\begin{pmatrix}A & 0\\ 0 & D\end{pmatrix}\in\Gamma_n$ we get

$$\det(A)^k f(A\mathfrak{Z}A^t)=f(\mathfrak{Z}) \quad \text{for any} \quad A\in\mathrm{GL}_n(\mathbb{Z})$$

which implies the following relation for the Fourier coefficients of f;

$$a(ANA^t)=\det(A)^k a(N). \tag{6.7}$$

If all coefficients $a(N)$ such that $\det(N)=0$ vanish, then form $f(\mathfrak{Z})$ is called a *parabolic form.* We denote the space of all parabolic forms of genus n and weight k by $\mathfrak{N}_k^n$.

A very important class of examples of modular forms of genus n and weight k consists of the Eisenstein series, defined as

$$\Psi_k(\mathfrak{Z})=\sum_{C,D}\det(C\mathfrak{Z}+D)^{-k}, \tag{6.8}$$

where the summation extends over all inequivalent bottom blocks of elements of $\mathrm{Sp}_n(\mathbb{Z})$ with respect to left multiplications by matrices in $\mathrm{SL}_n(\mathbb{Z})$.

A classical theorem of H. Braun shows that the series $\Psi_k(\mathfrak{Z})$ is absolutely convergent for $k>n+1$ and is a modular form of weight k. The notation of (6.8) will be reserved for the Eisenstein series of genus 2 and we will write ϕ_k for the case of genus 1.

If we let $g_2=60\phi_4$, $g_3=140\phi_6$, $\Delta=g_2^3-27g_3^2$ then for even k one has

$$\begin{aligned}\mathfrak{M}_k^1 &= \Delta\mathfrak{M}_{k-12}^1 \oplus \mathbb{C}\phi_k \\ &= \mathfrak{N}_k^1 \oplus \mathbb{C}\phi_k .\end{aligned} \tag{6.9}$$

It is well known that ϕ_4 and ϕ_6 are algebraically independent over $\mathbb{C}$,

$$\sum_{k=0}^{\infty} \mathfrak{M}_k^1 = \mathbb{C}[\phi_4, \phi_6], \tag{6.10}$$

and that,

$$\dim \mathfrak{M}_k^1 = \text{cardinality of the set } \{(a,b)\in(\mathbb{Z}^+)^2 | 4a+6b=k\}. \tag{6.11}$$

An analogous result for genus 2 has been obtained by Igusa [13].

Theorem 6.1 (Igusa). *The genus 2 Eisenstein series* Ψ_4, Ψ_6, Ψ_{10}, Ψ_{12} *are algebraically independent over* $\mathbb{C}$ *and*

$$\sum_{k=0}^{\infty} \mathfrak{M}_k^2 = \mathbb{C}[\Psi_4, \Psi_6, \Psi_{10}, \Psi_{12}].$$

Remark (see [13]). Since $\phi_4\phi_6=\phi_{10}$ and $3^2\cdot 7^2\phi_4^3+2\cdot 5^3\phi_6^2=691\phi_{12}$ we can define normalized parabolic forms

$$\chi_{10} = -43876\cdot 2^{-12}\cdot 3^{-5}\cdot 5^{-2}\cdot 7^{-1}\cdot 53^{-1}(\Psi_4\Psi_6-\Psi_{10})$$

and

$$\chi_{12} = 131\cdot 593\cdot 2^{-13}\cdot 3^{-7}\cdot 5^{-3}\cdot 7^{-2}\cdot 337^{-1}(3^2\cdot 7^2\Psi_4^3+2\cdot 5^3\Psi_6^2-691\Psi_{12}). \tag{6.12}$$

Then we can write

$$\sum_{k=0}^{\infty} \mathfrak{M}_k^2 = \mathbb{C}[\Psi_4, \Psi_6, \chi_{10}, \chi_{12}]. \tag{6.13}$$

Corollary 6.2 $\dim \mathfrak{M}_k^2 =$ *Cardinality of the set*

$$\{(a,b,c,d)\in(\mathbb{Z}^+)^4 | k=4a+6b+10c+12d\}.$$

The Siegel operator is a map $\Phi:\mathfrak{M}_k^n\to\mathfrak{M}_k^{n-1}$ defined for $n>1$ as follows: Let $\mathfrak{Z}'\in H_{n-1}$ and $0<\lambda\in\mathbb{R}$, then $\begin{pmatrix}\mathfrak{Z}' & 0\\ 0 & i\lambda\end{pmatrix}\in H_n$, the limit

$$(\Phi f)(\mathfrak{Z}') = \lim_{\lambda\to\infty} f\left(\begin{pmatrix}\mathfrak{Z}' & 0\\ 0 & i\lambda\end{pmatrix}\right) \tag{6.14}$$

exists for $f\in\mathfrak{M}_k^n$, and $\Phi f\in\mathfrak{M}_k^{n-1}$. One can see that the parabolic forms $\mathfrak{N}_k^n$ are just the kernel of Φ. In the case of genus $n=2$, if f is written as the Fourier series (6.5) then the only terms of f which survive the limit process of (6.14) are those

corresponding to $N=\begin{pmatrix} n_3 & 0 \\ 0 & 0 \end{pmatrix}$, so if we let $a'(n_3)=a(N)$ and $z'=z_3$ then

$$(\Phi f)(z_3)=\sum_{n_3\geqq 0} a'(n_3)e^{2\pi i n_3 z_3}.$$

A very important class of operators on the space of modular forms is named for Hecke, who considered them first for genus 1. We will discuss and generalize that case here.

Let

$$\Delta_1(m)=\left\{\begin{pmatrix} a & b \\ c & d \end{pmatrix}\in M_2(\mathbb{Z})|ad-bc=m\right\} \tag{6.15}$$

for $1\leqq m\in\mathbb{Z}$. Then $\Gamma_1=\mathrm{SL}_2(\mathbb{Z})$ acts on $\Delta_1(m)$ by left multiplication and $\Delta_1(m)$ breaks up into finitely many orbits under that action. A complete set of representatives, one from each orbit, may be taken to be

$$\bar{\Delta}_1(m)=\left\{\begin{pmatrix} a & b \\ 0 & d \end{pmatrix}\in M_2(\mathbb{Z})|ad=m, a\geqq 1, 0\leqq b<d\right\}. \tag{6.16}$$

If $f(z)\in\mathfrak{M}_k^1$, define the operator $T_k(m)$ by

$$(T_k(m)f)(z)=m^{k-1}\sum_{\sigma\in\Gamma_1\backslash\bar{\Delta}_1(m)} f(\sigma\cdot z)(cz+d)^{-k} \quad \text{for} \quad \sigma=\overline{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}\in\Gamma_1\backslash\bar{\Delta}_1(m). \tag{6.17}$$

Remark. From (5.38) and (5.39) we see that $\mathfrak{M}_k^1$ coincides with $\mathscr{M}_k(0)$.

We will now extend the action of $T_k(m)$ to all of $\mathscr{M}_k$.

Definition. If $\chi(z_1,z_2,z_3)\in\mathscr{M}_k$ define

$$\begin{aligned}&(T_k(m)\chi)(z_1,z_2,z_3)\\ &=m^{k-1}\sum_{\sigma\in\Gamma_1\backslash\bar{\Delta}_1(m)}\chi\left(\frac{mz_1}{cz_3+d}, m\left(z_2-\frac{cz_1^2}{cz_3+d}\right), \frac{az_3+b}{cz_3+d}\right)(cz_3+d)^{-k}\\ &\text{for} \quad \sigma=\overline{\begin{pmatrix} a & b \\ c & d \end{pmatrix}}\in\Gamma_1\backslash\bar{\Delta}_1(m), \qquad 1\leqq m\in\mathbb{Z}, \qquad 0<k\in\mathbb{Z}.\end{aligned} \tag{6.18}$$

Clearly $T_k(1)=$ Identity operator.

We will abbreviate the expression under the summation by the slash operator:

$$\begin{aligned}(\chi|_k g)(z_1,z_2,z_3)&=\chi\left(\frac{mz_1}{cz_3+d}, m\left(z_2-\frac{cz_1^2}{cz_3+d}\right), \frac{az_3+b}{cz_3+d}\right)(cz_3+d)^{-k}\\ &\text{for} \quad g=\begin{pmatrix} a & b \\ c & d \end{pmatrix}\in\Delta_1(m).\end{aligned} \tag{6.19}$$

It is straightforward to verify that for any $g_1\in\Delta_1(m)$, $g_2\in\Delta_1(n)$ we have

$$(\chi|_k g_1)|_k g_2=(\chi|_k(g_1g_2)). \tag{6.20}$$

The condition that $\mathfrak{F}_0^e$-character $\chi\in\mathscr{M}_k$ is exactly that

$$\chi|_k g=\chi \quad \text{for all} \quad g\in\Delta_1(1)=\mathrm{SL}_2(\mathbb{Z})=\Gamma_1. \tag{6.21}$$

So

$$(T_k(m)\chi)(z_1,z_2,z_3)=m^{k-1}\sum_{\sigma\in\Gamma_1\backslash\tilde{\Delta}_1(m)}(\chi|_k\sigma)(z_1,z_2,z_3) \tag{6.22}$$

is clearly well-defined on $\mathscr{M}_k$. It is also clear that $T_k(m)\chi\in\mathscr{M}_k$ because if $\{\sigma_1,\ldots,\sigma_r\}$ is a complete set of representatives of $\Gamma_1\backslash\tilde{\Delta}_1(m)$ then so is $\{\sigma_1 g,\ldots,\sigma_r g\}$ for any $g\in\Gamma_1$.

In fact, $T_k(m):\mathscr{M}_k(M)\to\mathscr{M}_k(mM)$ can be seen as follows. The level of an $\mathfrak{F}_0^e$-character $\chi(z_1,z_2,z_3)$ is determined by the exponential factor involving z_2 since that determines the γ_2^* coefficient of each weight in the corresponding $\mathfrak{F}_0^e$-module. If $\chi\in\mathscr{M}_k(M)$ we can write

$$\chi(z_1,z_2,z_3)=e^{2\pi iMz_2}\sum a(n_1,M,n_3)e^{2\pi i(n_1z_1+n_3z_3)} \tag{6.23}$$

and using the representatives (6.16)

$$(T_k(m)\chi)(z_1,z_2,z_3)=m^{k-1}\sum_{\substack{a\geqq 1\\ ad=m}}\sum_{0\leqq b<d}d^{-k}\chi\left(az_1,mz_2,\frac{az_3+b}{d}\right) \tag{6.24}$$

in which the exponential involving z_2 is $e^{2\pi imMz_2}$. So $T_k(m)\chi\in\mathscr{M}_k(mM)$.

7. The Correspondence of $\mathfrak{F}$-Characters with Siegel Modular Forms of Genus 2

We have seen in Sects. 3 and 5 how $\mathfrak{F}$-characters $X(z_1,z_2,z_3)\in\mathscr{E}$ and $\mathfrak{F}_0^e$-characters $\chi(z_1,z_2,z_3)\in\mathscr{E}_0$ can be viewed as functions on the genus 2 Siegel domain H_2. We wish to investigate the transformation invariance of these functions and determine when they are genus 2 Siegel modular forms.

Write an $\mathfrak{F}$-character $X\in\mathscr{E}$ as

$$X(\mathfrak{Z})=X(z_1,z_2,z_3)=\sum_{\substack{n_2n_3\geqq n_1^2\\ n_2,n_3\geqq 0}}a(n_1,n_2,n_3)e^{2\pi i(n_1z_1+n_2z_2+n_3z_3)}$$

$$\text{for}\quad \mathfrak{Z}=\begin{pmatrix}z_3 & z_1\\ z_1 & z_2\end{pmatrix}\in H_2. \tag{7.1}$$

The action of the Weyl group $W\approx\mathrm{PGL}_2(\mathbb{Z})$ on H_2, given by (3.102), determines an action of $g^t\in\mathrm{PGL}_2(\mathbb{Z})$ on functions of $\mathfrak{Z}\in H_2$;

$$g^t\cdot f(\mathfrak{Z})=f(g\mathfrak{Z}g^t). \tag{7.2}$$

But $g^t\cdot X(\mathfrak{Z})=X(g\mathfrak{Z}g^t)=X(\mathfrak{Z})$ because of (3.91), the invariance of the form $(\cdot,\cdot)$ on $S_2(\mathbb{C})$ under $\mathrm{PGL}_2(\mathbb{Z})$, and the fact that weights of $\mathfrak{F}$-modules which are W-conjugate have equal multiplicities. It is also clear from (7.1) that for any $M\in S_2(\mathbb{Z})$, $X(\mathfrak{Z}+M)=X(\mathfrak{Z})$. Since the root lattice Q is isomorphic to $S_2(\mathbb{Z})$ we may consider the action of $q\in Q$ on $X(\mathfrak{Z})$ to be $q\cdot X(\mathfrak{Z})=X(\mathfrak{Z}+M)$ for $M=v(q)\in S_2(\mathbb{Z})$.

The Weyl group W acts on Q, so we may form the semidirect product

$$W\ltimes Q=\{w,q)\in W\times Q|(w_1,q_1)\cdot(w_2,q_2)=(w_1w_2,w_2^{-1}\cdot q_1+q_2)\}. \tag{7.3}$$

Proposition 7.1. *For $q\in Q$, let $v(q)=M\in S_2(\mathbb{Z})$ and for $w\in W$, let $\bar{v}(w)=x$ determine $x\in \mathrm{PGL}_2(\mathbb{Z})$. Then the map $\mathfrak{f}: W\ltimes Q\to \mathrm{Sp}_2(\mathbb{Z})/\pm I$ defined by $\mathfrak{f}(w,q)=\begin{pmatrix} x & xM \\ 0 & (x^t)^{-1}\end{pmatrix}$ is an isomorphism of $W\ltimes Q$ with the "upper triangular" subgroup in the projective symplectic group.*

Definition. Let the action of $W\ltimes Q$ on a function $f(\mathfrak{Z})$ be given by

$$(w,q)\cdot f(\mathfrak{Z})=\mathfrak{f}(w,q)\cdot f(\mathfrak{Z}). \tag{7.4}$$

The action of $\mathrm{Sp}_2(\mathbb{Z})$ on H_2 given in (6.3) is well-defined on the projective symplectic group $\mathrm{Sp}_2(\mathbb{Z})/\pm I$ whose elements we still denote by $\begin{pmatrix} A & B \\ C & D\end{pmatrix}$. If $N=\begin{pmatrix} A & 0 \\ 0 & (A^t)^{-1}\end{pmatrix}$ then $N\cdot\mathfrak{Z}=A\mathfrak{Z}A^t$ and $\det((A^t)^{-1})^{-k}X(A\mathfrak{Z}A^t)=\det(A)^kX(\mathfrak{Z})$. Since $A\in\mathrm{PGL}_2(\mathbb{Z})$, Condition (I) of the definition of Siegel modular form will be satisfied for all such A if and only if k is even. If $N=\begin{pmatrix} I & M \\ 0 & I\end{pmatrix}$ for $M\in S_2(\mathbb{Z})$ then $N\cdot\mathfrak{Z}=\mathfrak{Z}+M$ and $\det(I)^{-k}X(\mathfrak{Z}+M)=X(\mathfrak{Z})$. So we have

Proposition 7.2. *Let $X(\mathfrak{Z})$ be any $\mathfrak{F}$-character. Then for any $N=\begin{pmatrix} A & B \\ 0 & D\end{pmatrix}\in\mathrm{Sp}_2(\mathbb{Z})$, and any $\mathfrak{Z}\in H_2$, $X(\mathfrak{Z})=\det(D)^{-k}X(N\cdot\mathfrak{Z})=X(N\cdot\mathfrak{Z})$ if k is even.*

Proposition 7.3. *Let $f(\mathfrak{Z})\in\mathfrak{M}_k^2$. Then there exists a unique collection of standard $\mathfrak{F}$-characters $\{X^\lambda(\mathfrak{Z})\in\mathscr{E}\mid\lambda\in P^{++}\}$ and a unique collection of complex numbers $\{c_\lambda\in\mathbb{C}\mid\lambda\in P^{++}\}$ such that $f(\mathfrak{Z})=\sum_{\lambda\in P^{++}} c_\lambda X^\lambda(\mathfrak{Z})$.*

Proof. By Igusa's result (Theorem 6.1) we may assume $k\geqq 4$ is even. So if $N=\begin{pmatrix} A & B \\ 0 & D\end{pmatrix}\in\mathrm{Sp}_2(\mathbb{Z})$ we have $f(N\cdot\mathfrak{Z})=f(\mathfrak{Z})$ and $X(N\cdot\mathfrak{Z})=X(\mathfrak{Z})$ for any $X(\mathfrak{Z})\in\mathscr{E}$.

Recall the Weyl-Kac character formula (Theorem 3.1). The numerator and denominator may be considered as functions on H_2 since $\lambda+\varrho\in P^{++}$. Let us use the notation $\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z})$ for the numerator, so the denominator is just $\mathscr{A}^{\varrho}(\mathfrak{Z})$. It is easy to see that if $A\in\mathrm{PGL}_2(\mathbb{Z})$ and $M\in S_2(\mathbb{Z})$ then

$$\mathscr{A}^{\lambda+\varrho}(A\mathfrak{Z}A^t)=\det(A)\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z}) \quad\text{and}\quad \mathscr{A}^{\lambda+\varrho}(\mathfrak{Z}+M)=\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z}). \tag{7.5}$$

In fact, we can write out the Fourier expansion

$$\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z})=\sum_{A\in\mathrm{PGL}_2(\mathbb{Z})}\det(A)e^{2\pi i\,\mathrm{Tr}(ANA^t\mathfrak{Z})}, \tag{7.6}$$

where $0<N\in S_2'(\mathbb{Z})$ corresponds to $\lambda+\varrho\in P^{++}$. In the theory of Lie algebras one defines the set of strictly dominant weights to be

$$P^{+++}=\{n_1\omega_1+n_2\omega_2+n_3\omega_3\mid 0<n_i\in\mathbb{Z}\quad\text{for}\quad 1\leqq i\leqq 3\}. \tag{7.7}$$

Then

$$P^{+++}=\{n_1\gamma_1^*+n_2\gamma_2^*+n_3\gamma_3^*\in P^+\mid n_3>n_2>n_1>0\}, \tag{7.8}$$

$$v(W\cdot P^{+++})=\{N\in S_2'(\mathbb{Z})\mid N>0\}, \tag{7.9}$$

each $\mathrm{PGL}_2(\mathbb{Z})$ orbit in $v(W\cdot P^{+++})$ has a unique representative in $v(P^{+++})$, and $P^{+++}=\varrho+P^{++}$. If $\lambda\in P^{++}$ then $\lambda+\varrho\in P^{+++}$, and conversely.

Consider the function

$$f'(\mathfrak{Z})=f(\mathfrak{Z})\mathscr{A}^{\varrho}(\mathfrak{Z}). \tag{7.10}$$

Certainly $f'(\mathfrak{Z}+M)=f'(\mathfrak{Z})$ for any $M\in S_2(\mathbb{Z})$ and $f'(A\mathfrak{Z}A^t)=\det(A)f'(\mathfrak{Z})$ for any $A\in\mathrm{PGL}_2(\mathbb{Z})$. Because each term of $\mathscr{A}^{\varrho}(\mathfrak{Z})$ corresponds to some $0<N\in S_2'(\mathbb{Z})$ each nonzero term of $f'(\mathfrak{Z})$ corresponds to some $0<N_1\in S_2'(\mathbb{Z})$. If we decompose $f'(\mathfrak{Z})$ into $\mathrm{PGL}_2(\mathbb{Z})$ orbits we may write

$$f'(\mathfrak{Z})=\sum_{N\in v(P^{+++})}a(N)\sum_{A\in\mathrm{PGL}_2(\mathbb{Z})}\det(A)e^{2\pi i\,\mathrm{Tr}(ANA^t\mathfrak{Z})}, \tag{7.11}$$

where $a(N)$ is the Fourier coefficient of $f'(\mathfrak{Z})$. Since $N\in v(P^{+++})$ there is a unique $\lambda\in P^{++}$ such that $v(\lambda+\varrho)=N$ and we recognize the inner sum of (7.11) as $\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z})$. Dividing by $\mathscr{A}^{\varrho}(\mathfrak{Z})$ and using the character formula $\mathscr{A}^{\lambda+\varrho}(\mathfrak{Z})/\mathscr{A}^{\varrho}(\mathfrak{Z})=X^{\lambda}(\mathfrak{Z})$ we have $f(\mathfrak{Z})=\sum_{\lambda\in P^{++}}c_\lambda X^{\lambda}(\mathfrak{Z})$ where $c_\lambda=a(N)$ for $N=v(\lambda+\varrho)$.

Remark. In the proof of Proposition 7.3 we have used only the convergence of $f(\mathfrak{Z})$ and its invariance under the parabolic subgroup.

Following Maass [31] we consider two subgroups of $\mathrm{Sp}_2(\mathbb{Z})$.

Definition. Let

$$E_1=\left\{\left(\begin{array}{cc|cc}a&0&b&0\\0&1&0&0\\\hline c&0&d&0\\0&0&0&1\end{array}\right)\middle|\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z})\right\} \tag{7.12}$$

and

$$E_2=\left\{\left(\begin{array}{cc|cc}1&0&0&0\\0&a&0&b\\\hline 0&0&1&0\\0&c&0&d\end{array}\right)\middle|\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z})\right\}. \tag{7.13}$$

It is clear that E_1 and E_2 are subgroups of $\mathrm{Sp}_2(\mathbb{Z})$ isomorphic to $\mathrm{SL}_2(\mathbb{Z})$.

Lemma 7.4. *The subgroups E_1, E_2 and $\left\{\begin{pmatrix}I&B\\0&I\end{pmatrix}\in\mathrm{Sp}_2(\mathbb{Z})\middle|B=B^t\right\}$ suffice to generate $\mathrm{Sp}_2(\mathbb{Z})$.*

Proof. This well-known fact is an easy exercise.

Proposition 7.5. *Necessary and sufficient conditions for $f(\mathfrak{Z})\in\mathfrak{M}_k^2$ are:*

(a) $f\left(\dfrac{z_1}{cz_3+d},\ z_2-\dfrac{cz_1^2}{cz_3+d},\ \dfrac{az_3+b}{cz_3+d}\right)(cz_3+d)^{-k}=f(z_1,z_2,z_3)$,

(b) $f(z_1+m_1,z_2+m_2,z_3+m_3)=f(z_1,z_2,z_3)$, *and*

(c) $f(z_1,z_2,z_3)=f(z_1,z_3,z_2)$

for all $\begin{pmatrix}a&b\\c&d\end{pmatrix}\in\mathrm{SL}_2(\mathbb{Z})$ and all $m_1,m_2,m_3\in\mathbb{Z}$.

Proof. Condition (a) is equivalent to invariance of $f(\mathfrak{Z})$ under E_1, (b) is trivial, and (c) implies a symmetry between E_1 and E_2 so that Lemma 7.4 gives the result.

Lemma 7.6. *Write the Fourier series of a function* $f(\mathfrak{Z})=f(z_1,z_2,z_3)$ *on* H_2 *satisfying* (b) *as*

$$f(z_1,z_2,z_3)=\sum_{n_2\in\mathbb{Z}^+} f_{n_2}(z_1,z_2,z_3), \tag{7.14}$$

where

$$f_{n_2}(z_1,z_2,z_3)=e^{2\pi i n_2 z_2}\sum_{\substack{n_3\geqq 0\\ 4n_2n_3\geqq n_1^2}} a(n_1,n_2,n_3)e^{2\pi i(n_1z_1+n_3z_3)}. \tag{7.15}$$

Then $f(z_1,z_2,z_3)$ *satisfies* (a) *if and only if* $f_{n_2}(z_1,z_2,z_3)$ *satisfies* (a) *for each* $n_2\geqq 0$.

Proof. If each f_{n_2} satisfies (a) then certainly their sum f does also. The condition (a) for $f(z_1,z_2,z_3)$ is just that for each

$$\sigma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in \mathrm{SL}_2(\mathbb{Z}), f(z_1,z_2,z_3)=(f|_k\sigma)(z_1,z_2,z_3)=\sum_{n_2\geqq 0}(f_{n_2}|_k\sigma)(z_1,z_2,z_3).$$

So $\sum_{n_2\geqq 0} f_{n_2}(z_1,z_2,z_3)=\sum_{n_2\geqq 0}(f_{n_2}|_k\sigma)(z_1,z_2,z_3)$ and comparing the expansions of both sides with respect to the variable z_2, we get $f_{n_2}(z_1,z_2,z_3)=(f_{n_2}|_k\sigma)(z_1,z_2,z_3)$ for each $n_2\geqq 0$.

Theorem 7.7. *Let* $f(\mathfrak{Z})\in\mathfrak{M}_k^2$. *Then there is a unique expression*

$$f(\mathfrak{Z})=\sum_{M\geqq 0}\chi_M(z_1,z_2,z_3), \tag{7.16}$$

where $\chi_M\in\mathcal{M}_k(M)$ *is a level* M $\mathfrak{F}_0^e$*-character satisfying* (a) *and* (b). *Thus,* $\mathfrak{M}_k^2\subseteqq\mathcal{M}_k$.

Proof. By Proposition 7.3 we can uniquely write $f(\mathfrak{Z})=\sum_{\lambda\in P^{++}} c_\lambda X^\lambda(\mathfrak{Z})$. Each standard $\mathfrak{F}$-character $X^\lambda(\mathfrak{Z})$ can be uniquely decomposed into $\mathfrak{F}_0^e$-characters according to level as in (5.20). The resulting decomposition of $f(\mathfrak{Z})$ is that given in (7.14), so for each $M\geqq 0$ $f_M(z_1,z_2,z_3)=\chi_M(z_1,z_2,z_3)$ is an $\mathfrak{F}_0^e$-character of level M. Since $f(\mathfrak{Z})\in\mathfrak{M}_k^2$, $f(\mathfrak{Z})$ satisfies (a) and (b), so each $\chi_M(z_1,z_2,z_3)$ satisfies (a) and (b), by Lemma 7.6. That is, $\chi_M\in\mathcal{M}_k(M)$ and $f(\mathfrak{Z})\in\mathcal{M}_k$, so $\mathfrak{M}_k^2\subset\mathcal{M}_k$.

Corollary 7.8. $\mathfrak{M}_k^2=\{f(\mathfrak{Z})\in\mathcal{M}_k | f(z_1,z_2,z_3)=f(z_1,z_3,z_2)\}$.

Proof. Clear from Proposition 7.5, Lemma 7.6, and Theorem 7.7.

Definition. Let the level of a function $f(\mathfrak{Z})$ satisfying (b) be the least positive integer n_2 such that $f_{n_2}\neq 0$ in (7.14).

Definition. Let $\mathfrak{T}_k^2(m)$ be the subspace in $\mathfrak{M}_k^2$ of functions whose level is at least m.

Then we have the filtration of $\mathfrak{M}_k^2$

$$\mathfrak{M}_k^2=\mathfrak{T}_k^2(0)\supset\mathfrak{T}_k^2(1)\supset\mathfrak{T}_k^2(2)\supset\ldots\supset\mathfrak{T}_k^2(m)\supset\ldots \tag{7.17}$$

and it is clear that if $f(\mathfrak{Z})\in\mathfrak{T}_{k_1}^2(m_1)$ and $g(\mathfrak{Z})\in\mathfrak{T}_{k_2}^2(m_2)$ then

$$f(\mathfrak{Z})g(\mathfrak{Z})\in\mathfrak{T}_{k_1+k_2}^2(m_1+m_2).$$

Also note that because of condition (c) $\mathfrak{T}_k^2(1)$ is exactly the subspace $\mathfrak{N}_k^2$ of parabolic forms.

Recall from the remarks following Theorem 6.1 that

$$\sum_{k\geqq 0}\mathfrak{M}_k^2=\mathbb{C}[\Psi_4,\Psi_6,\chi_{10},\chi_{12}]$$

is the Igusa ring generated by the Eisenstein series Ψ_4,Ψ_6 and the parabolic forms χ_{10},χ_{12}. According to the above definition, Ψ_4 and Ψ_6 are of level 0, χ_{10}, and χ_{12} are of level 1. So the monomial $\Psi_4^a\Psi_6^b\chi_{10}^c\chi_{12}^d$ for $a,b,c,d\in\mathbb{Z}^+$ is in $\mathfrak{T}_k^2(m)$ but not in $\mathfrak{T}_k^2(m+1)$, where $m=c+d$ and $k=4a+6b+10c+12d$. We may write

$$\mathbb{C}[\Psi_4,\Psi_6,\chi_{10},\chi_{12}]=\sum_{k\geqq 0}\sum_{m\geqq 0}\mathfrak{M}_k^2(m), \tag{7.18}$$

where

$$\mathfrak{M}_k^2(m)=\sum_{\substack{a,b,c,d\in\mathbb{Z}^+\\ 4a+6b+10c+12d=k\\ c+d=m}}\mathbb{C}\Psi_4^a\Psi_6^b\chi_{10}^c\chi_{12}^d. \tag{7.19}$$

Then

$$\mathfrak{M}_k^2=\sum_{m\geqq 0}\mathfrak{M}_k^2(m) \tag{7.20}$$

and we have for each $m\geqq 0$ a vector space isomorphism

$$\mathfrak{M}_k^2(m)\approx\mathfrak{T}_k^2(m)/\mathfrak{T}_k^2(m+1). \tag{7.21}$$

From (7.19) and the linear independence of the monomials which span $\mathfrak{M}_k^2(m)$ it is clear that the dimension of $\mathfrak{M}_k^2(m)$ over $\mathbb{C}$ equals the cardinality of the set $S(k,m)$ defined to be

$$\{(a,b,c,d)\in(\mathbb{Z}^+)^4 | k=4a+6b+10c+12d, c+d=m\}. \tag{7.22}$$

We wish to study the relationship between the spaces $\mathfrak{M}_k^2(m)$ and $\mathscr{M}_k(m)$ for each $m\geqq 0$. It is obvious that $\mathfrak{M}_k^2(0)\approx\mathfrak{M}_k^1\approx\mathscr{M}_k(0)$ [see (5.38), (5.39)], where the first isomorphism is given by the Siegel operator Φ.

Definition. For each $m\geqq 0$ let the linear transformation $L_m:\mathfrak{M}_k^2\to\mathscr{M}_k$ be defined by $L_m(f(\mathfrak{Z}))=f_m(\mathfrak{Z})$. So L_0 is just the Siegel operator Φ.

Definition. Let $\mathscr{M}_k'$ [resp., $\mathscr{M}_k'(m)$] be the subspace of $\mathscr{M}_k$ [resp., $\mathscr{M}_k(m)$] consisting of those $\mathrm{PSL}_2(\mathbb{Z})$-invariant $\mathfrak{H}_0^e$-characters which are $\mathfrak{F}$-dominant. Recall that an irreducible standard $\mathfrak{H}_0^e$-module and its character are called $\mathfrak{F}$-dominant if the highest weight is in P^{++} (2.20). and that an arbitrary $\mathfrak{H}_0^e$-module and its character are called $\mathfrak{F}$-dominant if each of its irreducible standard components is $\mathfrak{F}$-dominant. One sees that

$$\mathscr{M}_k'(m)=\{f_m(\mathfrak{Z})\in\mathscr{M}_k(m) | A(n_1,m,n_3)=0 \text{ for all } n_3<m\}, \tag{7.23}$$

where

$$f_m(\mathfrak{Z})=e^{2\pi imz_2}\sum_{\substack{4mn_3\geqq n_1^2\\ n_3\geqq 0}}A(n_1,m,n_3)e^{2\pi i(n_1z_1+n_3z_3)}. \tag{7.24}$$

Because of condition (c) we see that if $f(\mathfrak{Z})\in\mathfrak{T}_k^2(m)$ then $L_m(f(\mathfrak{Z}))\in\mathscr{M}_k'(m)$. Since $\mathfrak{M}_k^2(m)\subset\mathfrak{T}_k^2(m)$, we have

$$L_m':\mathfrak{M}_k^2(m)\to\mathscr{M}_k'(m), \tag{7.25}$$

where we have denoted by L_m' the restriction of L_m to the subspace $\mathfrak{M}_k^2(m)$. We can now find the dimensions of the spaces $\mathscr{M}_k'(m)$.

Theorem 7.9. *The linear transformation $L_m':\mathfrak{M}_k^2(m)\to\mathscr{M}_k'(m)$ is an isomorphism for each $m\geqq 0$.*

Proof. We have the basis of $\mathfrak{M}_k^2(m)$

$$\{\Psi_4^a\Psi_6^b\chi_{10}^c\chi_{12}^d \mid (a,b,c,d)\in S(k,m)\}. \tag{7.26}$$

Let $(\Psi_4)_0=L_0(\Psi_4)$, $(\Psi_6)_0=L_0(\Psi_6)$, $(\chi_{10})_1=L_1(\chi_{10})$, and $(\chi_{12})_1=L_1(\chi_{12})$. Then it is clear that

$$\begin{aligned}L_m(\Psi_4^a\Psi_6^b\chi_{10}^c\chi_{12}^d)&=(\Psi_4)_0^a(\Psi_6)_0^b(\chi_{10})_1^c(\chi_{12})_1^d\\&=\varphi_4^a\varphi_6^b(\chi_{10})_1^c(\chi_{12})_1^d,\end{aligned} \tag{7.27}$$

where φ_4 and φ_6 are the genus 1 Eisenstein series involving only z_3. Also, $(\chi_{10})_1\in\mathscr{M}_{10}'(1)$ and $(\chi_{12})_1\in\mathscr{M}_{12}'(1)$ are $\mathfrak{F}_0^e$-characters which may be written [see (5.23)] as:

$$(\chi_{10})_1(\mathfrak{Z})=D_{10}^1(z_3)\chi^{\omega_1^0}(\mathfrak{Z})+D_{10}^2(z_3)\chi^{\omega_2^0}(\mathfrak{Z}), \tag{7.28}$$

$$(\chi_{12})_1(\mathfrak{Z})=D_{12}^1(z_3)\chi^{\omega_1^0}(\mathfrak{Z})+D_{12}^2(z_3)\chi^{\omega_2^0}(\mathfrak{Z}), \tag{7.29}$$

where the coefficients $D_k^j(z_3)$, $k=10,12$, $j=1,2$ are power series in $q=e^{2\pi iz_3}$ with constant term 0.

We claim that this system of equations can be inverted, expressing $\chi^{\omega_1^0}$ and $\chi^{\omega_2^0}$ as a combination of $(\chi_{10})_1$ and $(\chi_{12})_1$ with coefficients being Laurent series in q.

Let us write, for $k=10,12$, $N=\begin{pmatrix}n_3 & n_1/2\\ n_1/2 & n_2\end{pmatrix}$, $\mathfrak{Z}=\begin{pmatrix}z_3 & z_1\\ z_1 & z_2\end{pmatrix}$,

$$\chi_k(\mathfrak{Z})=\sum_{N>0}c_k(N)e^{2\pi i\,\mathrm{Tr}(N\mathfrak{Z})} \tag{7.30}$$

and

$$(\chi_k)_1(\mathfrak{Z})=e^{2\pi iz_2}\sum_{n_3\geqq 1}e^{2\pi in_3z_3}\left[c_k\begin{pmatrix}n_3 & 0\\ 0 & 1\end{pmatrix}\Theta_0(z_1,z_3)+e^{-2\pi iz_3/4}c_k\begin{pmatrix}n_3 & 1/2\\ 1/2 & 1\end{pmatrix}\Theta_1(z_1,z_3)\right], \tag{7.31}$$

where $\Theta_h(z_1,z_3)$, $h=0,1$, is given in (5.40).

Let

$$\mathscr{C}_k^0(z_3)=\sum_{n_3\geqq 1}c_k\begin{pmatrix}n_3 & 0\\ 0 & 1\end{pmatrix}e^{2\pi in_3z_3} \tag{7.32}$$

and

$$\mathscr{C}_k^1(z_3)=\sum_{n_3\geqq 1}c_k\begin{pmatrix}n_3 & 1/2\\ 1/2 & 1\end{pmatrix}e^{2\pi i\left(n_3-\frac{1}{4}\right)z_3} \tag{7.33}$$

so that we have

$$\begin{pmatrix}(\chi_{10})_1(\mathfrak{Z})\\(\chi_{12})_1(\mathfrak{Z})\end{pmatrix}=e^{2\pi i z_2}\begin{pmatrix}\mathscr{C}^0_{10}(z_3)\mathscr{C}^1_{10}(z_3)\\\mathscr{C}^0_{12}(z_3)\mathscr{C}^1_{12}(z_3)\end{pmatrix}\begin{pmatrix}\Theta_0(z_1,z_3)\\\Theta_1(z_1,z_3)\end{pmatrix}. \tag{7.34}$$

To invert this system and obtain

$$\begin{pmatrix}\Theta_0(z_1,z_3)\\\Theta_1(z_1,z_3)\end{pmatrix}=e^{-2\pi i z_2}\begin{pmatrix}\mathscr{D}^0_{10}(z_3)\mathscr{D}^0_{12}(z_3)\\\mathscr{D}^1_{10}(z_3)\mathscr{D}^1_{12}(z_3)\end{pmatrix}\begin{pmatrix}(\chi_{10})_1(\mathfrak{Z})\\(\chi_{12})_1(\mathfrak{Z})\end{pmatrix} \tag{7.35}$$

we need only that $\mathscr{C}^0_{10}(z_3)\mathscr{C}^1_{12}(z_3)-\mathscr{C}^1_{10}(z_3)\mathscr{C}^0_{12}(z_3)$ be nonzero.

From the tables of Resnikoff and Saldaña [39] we have the following expansions:

$$\left.\begin{aligned}\mathscr{C}^0_{10}(z_3)&=\tfrac{1}{4}(2q-36q^2+272q^3-\dots)\\\mathscr{C}^1_{10}(z_3)&=\tfrac{1}{4}q^{-1/4}(-1q+16q^2-99q^3+\dots)\\\mathscr{C}^0_{12}(z_3)&=\tfrac{1}{12}(10q-132q^2+736q^3+\dots)\\\mathscr{C}^1_{12}(z_3)&=\tfrac{1}{12}q^{-1/4}(1q-88q^2+1275q^3+\dots).\end{aligned}\right\} \tag{7.36}$$

From these we find that

$$\mathscr{C}^0_{10}(z_3)\mathscr{C}^1_{12}(z_3)-\mathscr{C}^1_{10}(z_3)\mathscr{C}^0_{12}(z_3)=\tfrac{1}{4}q^{-1/4}(q^2-42q^3+819q^4+\dots), \tag{7.37}$$

which shows that Θ_0 and Θ_1 can be expressed as in (7.35). From (5.42), (5.43), (5.25), and (5.26) we also have

$$\chi^{\omega_1^0}(\mathfrak{Z})=e^{2\pi i z_2}q^{-1/4}\phi^{-1}(q)\Theta_1(z_1,z_3) \tag{7.38}$$

and

$$\chi^{\omega_2^0}(\mathfrak{Z})=e^{2\pi i z_2}\phi^{-1}(q)\Theta_0(z_1,z_3). \tag{7.39}$$

Combined with (7.35) this gives

$$\chi^{\omega_1^0}(\mathfrak{Z})=4q^{-2}(1+42q+945q^2+\dots)\phi^{-1}(q)[-\mathscr{C}^0_{12}(z_3)(\chi_{10})_1(\mathfrak{Z})+\mathscr{C}^0_{10}(z_3)(\chi_{12})_1(\mathfrak{Z})]$$

and (7.40)

$$\chi^{\omega_2^0}(\mathfrak{Z})=4q^{-7/4}(1+42q+945q^2+\dots)\phi^{-1}(q)[\mathscr{C}^1_{12}(z_3)(\chi_{10})_1(\mathfrak{Z})-\mathscr{C}^1_{10}(z_3)(\chi_{12})_1(\mathfrak{Z})].$$

Of course, $\phi^{-1}(q)=\sum_{n\geq 0}p(n)q^n$, where p is the classical partition function, is a power series in q with lowest term 1. From this and (7.36) we see that we may write

$$\chi^{\omega_1^0}(\mathfrak{Z})=E^1_{10}(z_3)(\chi_{10})_1(\mathfrak{Z})+E^1_{12}(z_3)(\chi_{12})_1(\mathfrak{Z})$$

and (7.41)

$$\chi^{\omega_2^0}(\mathfrak{Z})=E^2_{10}(z_3)(\chi_{10})_1(\mathfrak{Z})+E^2_{12}(z_3)(\chi_{12})_1(\mathfrak{Z}),$$

where

$$E^j_k(z_3)=q^{-1}\sum_{n\geq 0}e^j_k(n)q^n \tag{7.42}$$

for $k=10,12$, $j=1,2$, with $e^j_k(n)\in\mathbb{Q}$ and $e^j_k(0)\neq 0$. This establishes the claim.

Furthermore, we may express

$$(\chi^{\omega_1^0})^j(\chi^{\omega_2^0})^{m-j}=q^{-m}\sum_{i=0}^{m}F_i(q)(\chi_{10})_1^{m-i}(\chi_{12})_1^i \tag{7.43}$$

for $0\leqq j\leqq m$, where $F_i(q)$ is a power series in q.

We wish to emphasize the significance of the fact that this series begins with q^{-m}. This is responsible for the isomorphism of $\mathfrak{M}_k^2(m)$ with $\mathscr{M}_k'(m)$.

We know that $(\chi^{\omega_1^0})^{n_1}$ and $(\chi^{\omega_2^0})^{n_2}$ are the characters of the $n_1{}^{\text{th}}$ and $n_2{}^{\text{th}}$ tensor powers of the standard $\mathfrak{F}_0^e$-modules with highest weights ω_1^0 and ω_2^0, respectively, and that $(\chi^{\omega_1^0})^{n_1}(\chi^{\omega_2^0})^{n_2}$ is the character of the tensor product of the two tensor powers just mentioned. This tensor product decomposes into the direct sum of standard modules [16] on level n_1+n_2. In the decomposition of this tensor product the $\mathfrak{F}_0^e$-module with highest weight $n_1\omega_1^0+n_2\omega_2^0$ is known to occur with multiplicity 1, and the only other $\mathfrak{F}_0^e$-modules which occur have highest weights of the form

$$n_1\omega_1^0+n_2\omega_2^0-i_1\alpha_1-i_2\alpha_2 \quad \text{for} \quad i_1,i_2\in\mathbb{Z}^+.$$

Thus, all weights which occur in the $\mathfrak{F}_0^e$-module with character $(\chi^{\omega_1^0})^{n_1}(\chi^{\omega_2^0})^{n_2}$ are of that form. This means that on level m, $\lambda=j\omega_1^0+(m-j)\omega_2^0$ occurs with multiplicity 1 in the module with character $(\chi^{\omega_1^0})^j(\chi^{\omega_2^0})^{m-j}$ but does not occur in any module with character $(\chi^{\omega_1^0})^k(\chi^{\omega_2^0})^{m-k}$ for $k<j$. Since

$$j\omega_1^0+(m-j)\omega_2^0=j\gamma_1^*+m\gamma_2^*$$

we may say that $e^{2\pi i(jz_1+mz_2)}$ occurs with coefficient 1 in the Fourier decomposition of $(\chi^{\omega_1^0})^j(\chi^{\omega_2^0})^{m-j}$ but does not occur in that of $(\chi^{\omega_1^0})^k(\chi^{\omega_2^0})^{m-k}$ for any $k<j$.

Therefore, any level m $\mathfrak{F}_0^e$-character $\chi(\mathfrak{Z})$ can be uniquely expressed as a homogeneous degree m polynomial in $\chi^{\omega_1^0}$ and $\chi^{\omega_2^0}$ whose coefficients are power series in $q=e^{2\pi iz_3}$, namely,

$$\chi(\mathfrak{Z})=\sum_{j=0}^{m}T_j(z_3)(\chi^{\omega_1^0})^j(\chi^{\omega_2^0})^{m-j} \tag{7.44}$$

for

$$T_j(z_3)=\sum_{n_3\geqq 0}t_j(n_3)e^{2\pi in_3z_3}. \tag{7.45}$$

The uniqueness of the decomposition (7.44) can also be shown using the fact that the fundamental characters are proportional to Jacobi theta functions [see (7.38), (7.39)].

Combining (7.44) with (7.43) we may write any level m $\mathfrak{F}_0^e$-character $\chi(\mathfrak{Z})$ as

$$\chi(\mathfrak{Z})=\sum_{j=0}^{m}H_j(q)(\chi_{10})_1^{m-j}(\chi_{12})_1^j, \tag{7.46}$$

where $H_j(q)=q^{-m}G_j(q)$ and $G_j(q)$ is a power series in q which coefficients in $\mathbb{C}$. We can also see easily that if $\chi(\mathfrak{Z})\in\mathscr{M}_k'(m)$ then the power series $G_j(q)$ for each j has lowest nonzero term of degree at least m, so that $H_j(q)$ is a power series in q.

Consider the $(m+1)$-dimensional vector space V_m spanned by the functions $(\chi^{\omega_1^0})^j(\chi^{\omega_2^0})^{m-j}$, $0\leqq j\leqq m$, over the field $\mathscr{F}$ of Laurent series in q. The discussion

leading up to (7.44) shows that these $m+1$ monomials are a basis for V_m, (7.46) shows that $\mathscr{M}_k(m) \subseteq V_m$, and (7.43) shows that the $(m+1)$ functions $(\chi_{10})_1^j (\chi_{12})_1^{m-j}$, $0 \leqq j \leqq m$, also span V_m over $\mathscr{F}$, so that they are also a basis for V_m over $\mathscr{F}$. Thus the expression (7.46) is uniquely determined for each $\chi(\mathfrak{Z}) \in \mathscr{M}_k(m)$.

For any $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{SL}_2(\mathbb{Z})$ we have $\bar{g} = \left(\begin{array}{cc|cc} a & 0 & b & 0 \\ 0 & 1 & 0 & 0 \\ \hline c & 0 & d & 0 \\ 0 & 0 & 0 & 1 \end{array}\right) \in \mathrm{Sp}_2(\mathbb{Z})$.

Replacing $\mathfrak{Z}$ by $\bar{g} \cdot \mathfrak{Z}$ and z_3 by $g \cdot z_3$ in (7.46) we get

$$\chi(\bar{g} \cdot \mathfrak{Z}) = \sum_{j=0}^{m} H_j(g \cdot z_3)(\chi_{10})_1^{m-j}(\bar{g} \cdot \mathfrak{Z})(\chi_{12})_1^j(\bar{g} \cdot \mathfrak{Z}). \tag{7.47}$$

Thus, we get

$$\chi(\mathfrak{Z}) = \sum_{j=0}^{m} H_j(g \cdot z_3)(cz_3 + d)^{10(m-j)+12j-k}(\chi_{10})_1^{m-j}(\mathfrak{Z})(\chi_{12})_1^j(\mathfrak{Z}), \tag{7.48}$$

where we have written $H_j(q)$ as $H_j(z_3)$. Comparing this with (7.46) we see that for each j, $0 \leqq j \leqq m$,

$$H_j(z_3) = H_j(g \cdot z_3)(cz_3 + d)^{10(m-j)+12j-k}. \tag{7.49}$$

That is, the Laurent series $H_j(z_3) = e^{-2\pi i m z_3} G_j(z_3)$ occurring as coefficients of the monomials $(\chi_{10})_1^{m-j}(\chi_{12})_1^j$ in the expansion of $\chi(\mathfrak{Z}) \in \mathscr{M}_k(m)$ transform as modular functions of genus 1 with weight $k-(10(m-j)+12j)$. In the case when $\chi(\mathfrak{Z}) \in \mathscr{M}'_k(m)$, $H_j(z_3)$ is a power series in $e^{2\pi i z_3}$, so is holomorphic at ∞ and is actually a genus 1 modular form of weight $k-(10(m-j)+12j)$. From the classical result (6.10), we may thus express each $H_j(z_3)$ as a polynomial in $\phi_4^a \phi_6^b$ with

$$4a + 6b = k - (10(m-j) + 12j),$$

and then we have that

$$\chi(\mathfrak{Z}) = \sum_{c+d=m} \sum_{4a+6b+10c+12d=k} \alpha_{abcd} \phi_4^a \phi_6^b (\chi_{10})_1^c (\chi_{12})_1^d \tag{7.50}$$

for some $\alpha_{abcd} \in \mathbb{C}$. Thus,

$$L'_m\left(\sum_{c+d=m} \sum_{4a+6b+10c+12d=k} \alpha_{abcd} \Psi_4^a \Psi_6^b \chi_{10}^c \chi_{12}^d \right) = \chi(\mathfrak{Z}) \tag{7.51}$$

establishes the surjectivity of L'_m onto $\mathscr{M}'_k(m)$. Injectivity was established when the monomials $(\chi_{10})_1^j (\chi_{12})_1^{m-j}$ were seen to be a basis of V_m over the field of Laurent series $\mathscr{F}$.

Corollary 7.10. *For each* $m \geqq 0$

$$\dim(\mathscr{M}'_k(m)) = \dim(\mathfrak{M}_k^2(m)) = \text{Cardinality of the set } S(k,m)$$

given in (7.22).

Corollary 7.11. *The space* $\mathscr{M}' = \sum_{k \geqq 0} \sum_{m \geqq 0} \mathscr{M}'_k(m)$ *has the structure of a graded ring and is isomorphic to the Igusa ring* $C[\Psi_4, \Psi_6, \chi_{10}, \chi_{12}]$.

Theorem 7.9 shows what a special role is played by the space $\mathscr{M}'_k$. In Proposition 5.3 we saw a natural correspondence between the characters of $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-modules in category $\mathscr{C}'_0$ and $\mathfrak{F}$-modules in $\mathscr{C}$. Parallel to that we have from Theorem 7.9 a correspondence between the space $\mathscr{M}'_k$ of weight k $\mathrm{PSL}_2(\mathbb{Z})$-invariant $\mathfrak{F}$-dominant $\mathfrak{F}_0^e$-characters and the space of $\mathfrak{F}$-characters which are weight k Siegel modular forms, namely, $\mathfrak{M}_k^2$, whose dimension is given in (1.25).

The "lifting" of elliptic modular forms to Siegel modular forms has attracted the attention of several authors [2, 20, 22, 38, 43]. Our work has been particularly inspired by the work of Maass.

In a series of papers [30–32] Maass studied the subspace $\mathfrak{Q}_k^2 \subseteq \mathfrak{M}_k^2$ of forms

$$f(\mathfrak{Z}) = \sum_{0 \leqq N \in S_2^+(\mathbb{Z})} A(N) e^{2\pi i \operatorname{Tr}(N\mathfrak{Z})} \tag{7.52}$$

satisfying the special coefficient relation

$$A\begin{pmatrix} n_3 & n_1/2 \\ n_1/2 & n_2 \end{pmatrix} = \sum_{\substack{d \mid (n_1, n_2, n_3) \\ d > 0}} d^{k-1} A\begin{pmatrix} \frac{n_2 n_3}{d^2} & \frac{n_1}{2d} \\ \frac{n_1}{2d} & 1 \end{pmatrix}. \tag{7.53}$$

Such a relation had been observed in the extensive coefficient tables of Resnikoff and Saldaña [39] for some Eisenstein series Ψ_k and was proven in [28] by Maass for all Eisenstein series. Maass then considered the subspace $\mathfrak{S}_k^2$ of parabolic forms in $\mathfrak{Q}_k^2 = \mathbb{C}\Psi_k + \mathfrak{S}_k^2$. He computed the dimension of these spaces and showed that the relation (7.53) is equivalent to a simple construction of $f(\mathfrak{Z})$ out of Hecke operators applied to the level 1 of $f(\mathfrak{Z})$, namely $f_1(\mathfrak{Z}) \in \mathscr{M}_k(1)$. Maass found the dimension of $\mathfrak{S}_k^2$ to be equal to the cardinality of the set $S(k, 1)$, and the dimension of $\mathfrak{Q}_k^2$ to be one greater than that. The construction using Hecke operators of forms in $\mathfrak{Q}_k^2$ shows that any given $f_1(\mathfrak{Z}) \in \mathscr{M}_k(1)$ determines a unique $f(\mathfrak{Z}) \in \mathfrak{Q}_k^2$ whose level 1 slice is $f_1(\mathfrak{Z})$. Let $f_1(\mathfrak{Z})$ be written as

$$e^{2\pi i z_2} \sum_{\substack{4n_3 \geqq n_1^2 \\ n_3 \geqq 0}} A(n_1, 1, n_3) e^{2\pi i (n_1 z_1 + n_3 z_3)}.$$

Then the corresponding $f(\mathfrak{Z})$ is in $\mathfrak{S}_k^2$ if and only if $A(0, 1, 0) = 0$. So on level 1 the subspace of $\mathscr{M}_k(1)$ consisting of those $f_1(\mathfrak{Z})$ such that $A(0, 1, 0) = 0$ has the same dimension as the space $\mathfrak{M}_k^2(1) \approx \mathscr{M}'_k(1)$.

The work of Maass is related to our Theorem 7.9 in the case when $m = 1$. In the case when $m > 1$ it is not known how to use Hecke operators to construct from a given $f_m(\mathfrak{Z}) \in \mathscr{M}'_k(m)$ a form $f(\mathfrak{Z}) \in \mathfrak{M}_k^2$ such that $L_m(f(\mathfrak{Z})) = f_m(\mathfrak{Z})$. Our theory indicates that such a construction may exist.

In order to relate our work to that of Maass we would like to show his level 1 construction using the Hecke operators $T_k(m)$ defined in (6.18). Recall that the genus 1 Eisenstein series of weight k is

$$\varphi_k(z_3) = 1 + \frac{(2\pi i)^k}{(k-1)!\,\zeta(k)} \sum_{n_3 \geqq 1} \sigma_{k-1}(n_3) e^{2\pi i n_3 z_3}, \tag{7.54}$$

where ζ is the Riemann zeta function and $\sigma_{k-1}(n_3) = \sum_{d \mid n_3} d^{k-1}$ is the divisor power function.

Theorem 7.12 (Maass). *Let*

$$f_1(z_1,z_2,z_3)=e^{2\pi i z_2}\sum_{\substack{4n_3\geqq n_1^2\\ n_3\geqq 0}} A(n_1,1,n_3)e^{2\pi i(n_1z_1+n_3z_3)}$$

be a function on level 1 *satisfying condition (b) of Proposition* 7.5, *and let*

$$f(z_1,z_2,z_3)=A(0,1,0)\frac{(k-1)!\,\zeta(k)}{(2\pi i)^k}\varphi_k(z_3)+\sum_{m\geqq 1}(T_k(m)f_1)(z_1,z_2,z_3).$$

Then $f_1(z_1,z_2,z_3)\in\mathscr{M}_k(1)$ *if and only if* $f(z_1,z_2,z_3)\in\Omega_k^2$. *Also,* $f_1(z_1,z_2,z_3)\in\mathscr{M}_k'(1)$ *if and only if* $f(z_1,z_2,z_3)\in\mathfrak{S}_k^2$.

Proof. As shown at the end of Sect. 6, $T_k(m):\mathscr{M}_k(1)\to\mathscr{M}_k(m)$, so if $f_1\in\mathscr{M}_k(1)$ we have that for $m\geqq 1$, $L_m(f)=T_k(m)f_1\in\mathscr{M}_k(m)$ satisfies Proposition 7.5(a) and (b), as does

$$L_0(f)=A(0,1,0)\frac{(k-1)!\,\zeta(k)}{(2\pi i)^k}\varphi_k(z_3)\in\mathscr{M}_k(0)=\mathfrak{M}_k^1.$$

So $f(z_1,z_2,z_3)$ satisfies (a) and (b). Let us write

$$f_1(z_1,z_2,z_3)=e^{2\pi i z_2}\sum_{n_3\geqq 0}\sum_{n_1\in\mathbb{Z}}A(n_1,1,n_3)e^{2\pi i(n_1z_1+n_3z_3)}. \tag{7.55}$$

Then we have

$$\begin{aligned}&(T_k(m)f_1)(z_1,z_2,z_3)\\ &=e^{2\pi imz_2}\sum_{n_3\geqq 0}\sum_{n_1\in\mathbb{Z}}\sum_{\substack{d\mid m\\ d\geqq 1}}\sum_{0\leqq b<d}m^{k-1}d^{-k}A(n_1,1,n_3)e^{2\pi i\left(n_1\frac{m}{d}z_1+n_3\frac{m}{d^2}z_3+\frac{n_3b}{d}\right)}\\ &=e^{2\pi imz_2}\sum_{n_3\geqq 0}\sum_{n_1\in\mathbb{Z}}\sum_{\substack{d\mid m,n_3\\ d\geqq 1}}\left(\frac{m}{d}\right)^{k-1}A(n_1,1,n_3)e^{2\pi i\left(n_1\frac{m}{d}z_1+n_3\frac{m}{d^2}z_3\right)}.\end{aligned} \tag{7.56}$$

We have used the fact that

$$\sum_{0\leqq b<d}e^{2\pi i\frac{n_3b}{d}}=\begin{cases}d & \text{if } d\mid n_3\\ 0 & \text{if } d\nmid n_3.\end{cases} \tag{7.57}$$

Putting $ad=m$ and replacing n_3 by dn_3 in (7.56) we have

$$(T_k(m)f_1)(\mathfrak{Z})=e^{2\pi iz_2}\sum_{n_3\geqq 0}\sum_{n_1\in\mathbb{Z}}\sum_{\substack{a\mid m\\ a\geqq 1}}a^{k-1}A\left(n_1,1,\frac{mn_3}{a}\right)e^{2\pi i(n_1az_1+n_3az_3)}. \tag{7.58}$$

Then

$$\begin{aligned}f(\mathfrak{Z})&=\sum_{m\geqq 1}(T_k(m)f_1)(\mathfrak{Z})\\ &=\sum_{m\geqq 1}e^{2\pi imz_2}\sum_{n_3\geqq 0}\sum_{n_1\in\mathbb{Z}}\sum_{\substack{a\mid m\\ a\geqq 1}}a^{k-1}A\left(n_1,1,\frac{mn_3}{a}\right)e^{2\pi i(n_1az_1+n_3az_3)}.\end{aligned} \tag{7.59}$$

If we set this equal to the Fourier series $\sum_{0\leqq N\in S_2'(\mathbb{Z})}A(N)e^{2\pi i(m_1z_1+m_2z_2+m_3z_3)}$ for $N=\begin{pmatrix}m_3 & m_1/2\\ m_1/2 & m_2\end{pmatrix}$ and compare coefficients then we get the Maass coefficient

relation (7.53). This relation shows that for $n_2>0$, $A(n_1,n_2,n_3)=A(n_1,n_3,n_2)$, and that $A(0,0,n_3)=A(0,1,0)\sigma_{k-1}(n_3)=A(0,n_3,0)$, so $f(\mathfrak{Z})$ satisfies condition (c), symmetry in z_2 and z_3, and by Proposition 7.5 we have $f(\mathfrak{Z})\in\mathfrak{M}_k^2$. Since $f(\mathfrak{Z})$ satisfies the Maass condition (7.53), $f(\mathfrak{Z})\in\mathfrak{L}_k^2$. The converse is clear, as is the fact that $f(\mathfrak{Z})$ is parabolic if and only if $A(0,1,0)=0$, that is, $f_1(\mathfrak{Z})\in\mathscr{M}_k'(1)$.

Theorem 7.13 (Maass). *We have for $k\geqq 4$ even,* $\dim(\mathscr{M}_k(1))=\dim(\mathfrak{L}_k^2)=\left[\frac{k+2}{6}\right]$ *and* $\dim(\mathscr{M}_k'(1))=\dim(\mathfrak{S}_k^2)=\left[\frac{k-4}{6}\right]$ *which equals the cardinality of the set $S(k,1)$ given in* (7.22).

Remark. Maass established these results in [30, 31] based on the computation of the dimension of the vectorial modular forms $\begin{pmatrix}c_0(z_3)\\ c_1(z_3)\end{pmatrix}$ satisfying (5.45) which determine an element of $\mathscr{M}_k(1)$. The condition (5.45) actually shows that $c_0(z_3)$ determines $c_1(z_3)$. Maass, however, used a different notation than ours. He used the classical Jacobi theta functions given in (5.40) to express a level 1 function rather than the fundamental $\mathfrak{F}_0^e$-characters $\bar{\chi}^{\omega_1^0}$ and $\bar{\chi}^{\omega_2^0}$ as we have done. One result of this difference is that Maass put the factor $\eta^{-1}(e^{2\pi i z_3})$ into the coefficients c_0 and c_1 which made his vectorial modular form $\begin{pmatrix}c_0\\ c_1\end{pmatrix}$ have weight $k-\frac{1}{2}$. For us that form has weight k.

The computation of $\dim(\mathfrak{S}_k^2)$ in [31] is based on results of Petersson in [37]. Corollary 7.10 gives a dimension formula for $\mathscr{M}_k'(m)$ for all $m\geqq 1$ which generalizes Maass' result and allows each $f(\mathfrak{Z})\in\mathfrak{M}_k^2$ to be a "lifing" of a vectorial modular form $\begin{pmatrix}c_0(z_3)\\ c_1(z_3)\\ \vdots\\ c_M(z_3)\end{pmatrix}$ of weight k satisfying the conditions of Proposition 5.4. We may show that each $c_j(z_3)$ is an automorphic form of weight k with multiplier system with respect to a subgroup of the congruence subgroup $\Gamma_0(4(M+2))$. Recalling the notation of (5.24)–(5.26) and Proposition 5.4 we see that if we write a level M function as

$$f_M(\mathfrak{Z})=\sum_{j=0}^{M}\bar{\chi}^{j\omega_1^0+(M-j)\omega_2^0}(\mathfrak{Z})c_j(z_3) \tag{7.60}$$

then $f_M(\mathfrak{Z})\in\mathscr{M}_k(M)$ if and only if the vectorial function $\mathfrak{C}(z_3)=\begin{pmatrix}c_0(z_3)\\ c_1(z_3)\\ \vdots\\ c_M(z_3)\end{pmatrix}$ satisfies the conditions

$$\mathfrak{C}\left(\frac{-1}{z_3}\right)(-z_3)^{-k}=U\mathfrak{C}(z_3) \tag{7.61}$$

and

$$\mathfrak{C}(z_3+1)=V\mathfrak{C}(z_3), \tag{7.62}$$

where $U=(U_{ij})$ and $V=(V_{ij})$ are $(M+1)\times(M+1)$ matrices such that for $1\leqq i \leqq M+1$ and $1\leqq j\leqq M+1$

$$U_{ij}=\sqrt{\frac{2}{M+2}}\sin\left(\frac{\pi ij}{M+2}\right) \tag{7.63}$$

and

$$V_{jj}=e^{-\left(\frac{j^2}{4(M+2)}-\frac{1}{8}\right)2\pi i}, \quad V_{ij}=0 \quad \text{for} \quad i\neq j. \tag{7.64}$$

This is equivalent to saying that there is a group homomorphism τ from $\mathrm{PSL}_2(\mathbb{Z})$ to $\mathrm{GL}_{M+1}(\mathbb{C})$ such that

$$\mathfrak{C}\left(\frac{az_3+b}{cz_3+d}\right)(cz_3+d)^{-k}=\tau\begin{pmatrix}a & b\\ c & d\end{pmatrix}\mathfrak{C}(z_3) \tag{7.65}$$

for all $\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\mathrm{PSL}_2(\mathbb{Z})$ such that

$$\tau\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}=U \quad \text{and} \quad \tau\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}=V. \tag{7.66}$$

The subgroup of $\mathrm{PGL}_2(\mathbb{Z})$ consisting of all elements g such that $\tau(g)$ is diagonal is the subgroup under which each function $c_j(z_3)$ will be automorphic of weight k with a certain multiplier system.

We can certainly say that $\begin{pmatrix}1 & 1\\ 0 & 1\end{pmatrix}$ is in that subgroup. Also, from (7.64) we see that $V_{jj}^{4(M+2)}=(-1)^{M+2}$. So $\tau\begin{pmatrix}1 & 4(M+2)\\ 0 & 1\end{pmatrix}=\pm I$ and therefore,

$$\tau\begin{pmatrix}1 & 0\\ 4(M+2) & 1\end{pmatrix}=\tau\left(\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}\begin{pmatrix}1 & -4(M+2)\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\right)=U(\pm I)U^{-1}=\pm I$$

which is diagonal. A deeper investigation of this subgroup is carried out by Kac and Peterson in [18].

8. Future Prospects

We would like to indicate briefly in this last section some directions for further research which we hope to pursue in subsequent publications.

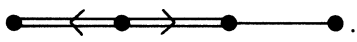
Within algebra $\mathfrak{F}$ there are other subalgebras besides $\mathfrak{F}_0$ with respect to which one may decompose $\mathfrak{F}$ and $\mathfrak{F}$-characters. We have in mind some rank 2 hyperbolic Kac-Moody algebras found by Lepowsky and Moody in [24] to be connected to Hilbert modular surfaces. We expect that such decompositions give a Lie-theoretical explanation of the connection between Siegel modular forms of genus 2 and Hilbert modular forms associated with quadratic number fields (see [39]). Geometrically one sees that slicing the Siegel domain (cone) by $\mathfrak{F}_0$ gives parabolas while slicing it by hyperbolic rank 2 subalgebras should give hyperbolas.

We may generalize the construction given in Sect. 4 of a hyperbolic algebra $\mathfrak{G}=\mathfrak{G}(\hat{\mathfrak{g}}^e, V)$ from an affine algebra $\hat{\mathfrak{g}}$ and its basic module V. In the case when $\hat{\mathfrak{g}}$

is of ADE type the basic module V has a nice construction (see [10, 17]) and action of $\mathfrak{g}\hat{}$ on V is given by vertex operators. One may expect that a result such as Theorem 3.2 can be established in that generality which would shed light on root multiplicities of $\mathfrak{G}$.

The interesting Kac-Moody algebras will be those whose Weyl groups are related to classically studied arithmetic groups. The rank 3 hyperbolic algebras have already been seen by ourselves and Yoshida [42] to have hyperbolic triangle groups for Weyl groups. The representation theory of these algebras should shed light on the theory of modular forms with respect to certain arithmetic subgroups of $\mathrm{Sp}_2(\mathbb{R})$.

Our construction of $\mathfrak{G}$ when applied to $\mathfrak{g}\hat{}$ of type $C_2^{(1)}$ gives an algebra with Dynkin diagram shown in (1.27), and when applied to $\mathfrak{g}\hat{}$ of type $D_3^{(2)}$ gives an algebra with diagram



(8.1)

Both algebras have Weyl group W with generators A, B, C, D having relations

$$\begin{aligned} &1 = A^2 = B^2 = C^2 = D^2, \\ &1 = (AB)^4 = (AC)^2 = (AD)^2 = (BC)^4 = (BD)^2 = (CD)^3. \end{aligned} \tag{8.2}$$

These are the relations among the four generators of the Klein-Fricke group Ψ_1^* which contains the Picard group $\Psi_1 = \mathrm{PSL}_2(\mathbb{Z}[i])$ as a subgroup of index 4. (See Magnus [34, p. 152], but there are errors in his list of relations.) Ψ_1^* is generated by $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, $\mathscr{D}$ where

$$\begin{aligned} &\mathscr{A}: z \to \bar{z} \\ &\mathscr{B}: z \to -i\bar{z} \\ &\mathscr{C}: z \to -\bar{z} - 1 \\ &\mathscr{D}: z \to 1/\bar{z}. \end{aligned} \tag{8.3}$$

One or both of these algebras should be closely related to the theory of Hermitian modular forms of genus 2. In particular, the lifting of $\mathfrak{g}\hat{}^e$-characters which are $\mathrm{PSL}_2(\mathbb{Z}[i])$-invariant to $\mathfrak{G}$-characters which are $\mathrm{Sp}_2(\mathbb{Z}[i])$-invariant should be similar to the lifting discussed in this paper. Already Kojima [21] has investigated this lifting for level 1 using Hecke operators as suggested in the work of Maass [30–32]. We believe that the existence of lifting and descent operations between modular forms of various types should be explained by the presence of a Kac-Moody algebra and a subalgebra.

Finally, we should mention that there are many questions concerning just the algebra $\mathfrak{F}$ which remain unanswered. One would like to have complete information about the root multiplicities and a deeper understanding of the structure of $\mathfrak{F}$.

Acknowledgements. We wish to express our gratitude to the National Science Foundation for its support of the Kac-Moody Lie algebras conference held at Rutgers University and the Jacobson algebra conference held at Yale University, during which this joint work began. We are also grateful to J. Lepowsky, A. Meurman and R. Moody for their interest in our work and their friendly help.

References

1. Andrianov, A.N.: Euler products corresponding to Siegel modular forms of genus 2. Usp. Mat. Nauk **29**, 43–110 (1974) [English transl.: Russian Math. Surveys **29**, 45–116 (1974)]
2. Andrianov, A.N.: Modular descent and the Saito-Kurokawa conjecture. Invent. Math. **53**, 267–280 (1979)
3. Borevich, Z.I., Shafarevich, I.R.: Number theory, pure and applied mathematics, Vol. 20. New York: Academic Press 1966
4. Feingold, A.J.: A hyperbolic GCM Lie algebra and the Fibonacci numbers. Proc. Am. Math. Soc. **80**, 379–385 (1980)
5. Feingold, A.J.: Tensor products of certain modules for the generalized Cartan matrix Lie algebra $A_1^{(1)}$. Comm. Algebra **9**, 1323–1341 (1981)
6. Feingold, A.J., Lepowsky, J.: The Weyl-Kac character formula and power series identities. Adv. Math. **29**, 271–309 (1978)
7. Frenkel, I.: Spinor representations of affine Lie algebras. Proc. Nat. Acad. Sci. **77**, 6303–6306 (1980)
8. Frenkel, I.: Two constructions of affine Lie algebra representations and boson-fermion correspondence in quantum field theory. J. Functional Analysis **44**, 259–327 (1981)
9. Frenkel, I.: Representations of affine Lie algebras, Hecke modular forms, and Korteweg-de Vries type equations. Proceedings of the 1981 Rutgers Conference on Lie Algebras and related topics. Lecture Notes in Mathematics, Vol. 933, pp. 71–110. Berlin, Heidelberg, New York: Springer 1982
10. Frenkel, I., Kac, V.G.: Basic representations of affine Lie algebras and dual resonance models. Invent. Math. **62**, 23–66 (1980)
11. Gabber, O., Kac, V.G.: On defining relations of certain infinite-dimensional Lie algebras, Bull. Am. Math. Soc. **5**, 185–189 (1981)
12. Garland, H., Lepowsky, J.: Lie algebra homology and the Macdonald-Kac formulas. Invent. Math. **34**, 37–76 (1976)
13. Igusa, J.: On Siegel modular forms of genus two. Am. J. Math. **84**, 175–200 (1962)
14. Kac, V.G.: Simple irreducible graded Lie algebras of finite growth. Izv. Akad. Nauk. SSSR **32**, 1323–1376 (1968) (in Russian) [English transl.: Math. USSR-Izv. **2**, 1271–1311 (1968)]
15. Kac, V.G.: Infinite-dimensional Lie algebras and Dedekind's η-function. Funkt. Anal. Priloz. **8**, 77–78 (1974) (in Russian). [English transl.: Functional Anal. Appl. **8**, 68–70 (1974)]
16. Kac, V.G.: Infinite-dimensional algebras, Dedekind's η-function, classical Mobius function, and the very strange formula. Adv. Math. **30**, 85–136 (1978)
17. Kac, V.G., Kazhdan, D.A., Lepowsky, J., Wilson, R.L.: Realization of the basic representations of the Euclidean Lie algebras. Adv. Math. **42**, 83–112 (1981)
18. Kac, V.G., Peterson, D.H.: Affine Lie algebras and Hecke modular forms. Bull. Math. Soc. **3**, 1057–1061 (1980)
19. Kantor, I.L.: Infinite dimensional simple graded Lie algebras. Dokl. Akad. Nauk. SSSR **179**, 534–537 (1968) (in Russian). [English transl. Sov. Math. Dokl. **9**, 409–412 (1968)]
20. Kohnen, W.: Modular forms of half-integral weight on $\Gamma_0(4)$, Math. Ann. **248**, 249–266 (1980)
21. Kojima, H.: An arithmetic of Hermitian modular forms of degree two. Invent. Math. **69**, 217–227 (1982)
22. Kurokawa, N.: Examples of eigenvalues of Hecke operators on Siegel cusp forms of degree two, Invent. Math. **49**, 149–165 (1978)
23. Lepowsky, J., Milne, S.: Lie algebraic approaches to classical partition identities. Adv. Math. **29**, 15–59 (1978)
24. Lepowsky, J., Moody, R.V.: Hyperbolic Lie algebras and quasi-regular cusps on Hilbert modular surfaces. Math. Ann. **245**, 63–88 (1979)
25. Lepowsky, J., Wilson, R.L.: Construction of the affine Lie algebra $A_1^{(1)}$. Commun. Math. Phys. **62**, 43–53 (1978)
26. Looijenga, E.: Root systems and elliptic curves. Invent. Math. **38**, 17–32 (1976)
27. Maaß, H.: Die Fourierkoeffizienten der Eisensteinreihen zweiten Grades. Mat. Fys. Medd. Dan. Vid. Selsk. **34**, 1–25 (1964)
28. Maaß, H. Über die Fourierkoeffizienten der Eisensteinreihen zweiten Grades. Mat. Fys. Medd. Dan. Vid. Selsk. **38**, 1–13 (1972)
29. Maaß, H.: Lineare Relationen für die Fourierkoeffizienten einiger Modulformen zweiten Grades. Math. Ann. **232**, 163–175 (1978)

30. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. Invent. Math. **52**, 95–104 (1979)
31. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. II. Invent. Math. **53**, 249–253 (1979)
32. Maaß, H.: Über eine Spezialschar von Modulformen zweiten Grades. III. Invent. Math. **53**, 255–265 (1979)
33. Macdonald, I.G.: Affine root systems and Dedekind's η-function. Invent. Math. **15**, 91–143 (1972)
34. Magnus, W.: Non-Euclidean Tesselations and their Groups. Pure Appl. Math. Vol. 61. New York, London: Academic Press 1974
35. Moody, R.V.: Root systems of hyperbolic type. Adv. Math. **33**, 144–160 (1979)
36. Moody, R.V., Berman, S.: Lie algebra multiplicities. Proc. Am. Math. Soc. **76**, 223–228 (1979)
37. Peterson, H.: Zur analytischen Theorie der Grenzkreisgruppen. II. Math. Ann. **115**, 175–204 (1938)
38. Piatetski-Shapiro, I.: On the Saito-Kurokawa lifting (preprint) (1981)
39. Resnikoff, H.L., Saldaña, R.L.: Some properties of Fourier coefficients of Eisenstein series of degree two. J. Reine Angew. Math. **265**, 90–109 (1974)
40. Shimura, G.: Introduction to the arithmetic theory of automorphic functions. Publ. Math. Soc. Japan **11** (1971)
41. Whittaker, E.T., Watson, G.N.: A course of modern analysis. New York: MacMillan 1948
42. Yoshida, M.: Discrete reflection groups in a parabolic subgroup of $Sp(2, \mathbb{R})$ and symmetrizable hyperbolic generalized Cartan matrices of rank 3 (preprint)
43. Zagier, D.: Sur la conjecture de Saito-Kurokawa (d'après H. Maß). Seminaire Delange-Pisot-Poitou (preprint) (1980)

Received August 3, 1982; in revised form January 4, 1983

Note added in proof. We have recently learned through the referee of the paper "On the theory of Jacobi forms" by M. Eichler and D. Zagier. The authors, inspired by the work of Maass [29–32], undertook the construction of a new theory. Jacobi forms (of even weight) in our language are certain characters of affine Lie algebra $\mathfrak{F}_0$. This fact naturally entailed some overlap with the second part of our paper (Sects. 5–7). These two different points of view can be very fruitful for further results and generalizations. We are grateful to the referee for informing us about the work of Eichler and Zagier, and to Zagier for sending us their preprint before publication.

We have also received a preprint "Kac-Moody symmetry of gravitation and supergravity theories" by B. Julia (l'Ecole Normale Supérieure) in which he conjectures that the hyperbolic algebra $\mathfrak{F}$ may be an algebra of internal symmetries of Einstein gravitational equations. It has been known for about ten years that the subalgebra $\mathfrak{F}_0$ is responsible for the internal symmetries of gravitational plane waves. Thus, the highly abstract Lie algebra $\mathfrak{F}$ promises to describe fundamental symmetries of nature.

Math. Ann. 263, 145–156 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Order Isomorphisms of Fourier-Stieltjes Algebras

Wolfgang Arendt[1] and Jean De Cannière[2]*

[1] Mathematisches Institut der Universität, Auf der Morgenstelle 10, D-7400 Tübingen, Federal Republic of Germany
[2] Department of Mathematics, University of California, Berkeley, CA 94720, USA

1. Introduction

The Fourier-Stieltjes algebra $B(G)$ of a locally compact group G is the space of all linear combinations of continuous positive definite functions on G. Whereas $B(G)$ has usually been investigated as a Banach algebra (with pointwise multiplication), we consider another structure. We view $B(G)$ as an ordered vector space in two senses; these orderings are determined by: the cone $P(G)$ of all continuous positive definite functions on G and the cone $B(G)_+$ of all pointwise positive functions on G.

The purpose of this article is to show that the "biordered space" $(B(G), P(G), B(G)_+)$ is a complete invariant for the locally compact group G. More precisely, if G_1 and G_2 are two locally compact groups, then G_1 is isomorphic to G_2 as a topological group if and only if there exists a bijective linear mapping T: $B(G_1) \to B(G_2)$ such that $TB(G_1)_+ = B(G_2)_+$ and $TP(G_1) = P(G_2)$.

In this way we have found a complete invariant for locally compact groups which is very simple to define. Only the structure of a topological group is necessary for its definition. In particular, the Haar measure is not needed.

The proofs, however, make use of Haar measure. In fact, they depend on Walter's theorem [9, Theorem 2], which says that two locally compact groups G_1 and G_2 are isomorphic if and only if there exists an *isometric* algebra isomorphism T from $B(G_1)$ onto $B(G_2)$. Here the norm of $B(G)$, which is defined via $L^1(G)$, is essential to obtain complete invariance.

So our first step towards the main result is to show that in Walter's theorem the condition that T be isometric can be replaced by the assumption that T is an order isomorphism for the positive definite ordering (Sect. 3). This seems to be of independent interest, since it shows that the "ordered algebra" $(B(G), \cdot, P(G))$ is a complete invariant (which is "Haar measure free").

The next step is to investigate order isomorphisms for the pointwise ordering. In Sect. 5 we obtain the following characterization: A bijective linear mapping T

---

* Bevoegdverklaard navorser N.F.W.O., Belgium
On leave from Katholieke Universiteit Leuven

from $B(G_1)$ onto $B(G_2)$ satisfies $TB(G_1)_+ = B(G_2)_+$ if and only if there exists an invertible element h in $B(G_2)_+$ and a $*$-algebra isomorphism $V: B(G_1) \to B(G_2)$ such that $Tf = h \cdot Vf$ ($f \in B(G_1)$). Both steps put together show that, up to a multiplicative positive constant, biorderisomorphisms have to be isometric $*$-isomorphisms, so that Walter's result applies.

In a previous paper [1], we made an analogous investigation of the Fourier algebra $A(G)$. We would like to comment on how those results and their proofs are related to the present paper. In [1] we showed that the "biordered space" $(A(G), A(G) \cap P(G), A(G)_+)$ is a complete invariant of G as well. This space, however, is more complicated to define than $B(G)$ (it involves the norm of $B(G)$, since $A(G)$ is the closure of all functions in $B(G)$ which have compact support; see [3] for this and equivalent definitions of $A(G)$). The proofs are related in the following way: The first step described above is a generalization of the corresponding result for $A(G)$. To deal with order isomorphisms for the pointwise ordering, however, is a completely different matter in the two cases. Actually, for $A(G)$, we used the fact that the maximal ideal space of $A(G)$ can be identified with G. This is not true for $B(G)$ in general. Hence our proof for $A(G)$ does not carry over to the $B(G)$ case. On the other hand, the method in the present paper involves the multipliers of $B(G)$, which are given by the elements of $B(G)$ itself. The same method applies to the Fourier algebra of amenable groups, mainly because in that case the multiplier algebra of $A(G)$ coincides with $B(G)$. This fails in general, however (see also the remark following the proof of Theorem 5.2 below).

Finally, we point out that our main result can be applied to the group algebra. So we prove in Sect. 6 that $L^1(G)$, too, carries the natural structure of a biordered space which, as above, is a complete invariant of G.

2. Preliminaries

As a basic reference we use Eymard's fundamental article [3]. Throughout this paper G, G_1, G_2 are locally compact groups. We denote the set of all continuous positive definite functions on G by $P(G)$. It is a proper cone in the vector space $C^b(G)$ (of all continuous bounded complex valued functions on G), i.e.:

$$\mathbb{R}_+ P(G) \subset P(G)$$
$$P(G) + P(G) \subset P(G)$$
$$P(G) \cap (-P(G)) = \{0\}.$$

Moreover, $P(G) \cdot P(G) \subset P(G)$ and $1 \in P(G)$. Thus, $B(G) = \operatorname{span} P(G)$ is a unital subalgebra of $C^b(G)$. It is called the Fourier-Stieltjes algebra of G. $B(G)$ can be canonically identified with the dual space of $C^*(G)$, the enveloping C^*-algebra of $L^1(G)$, and is a Banach algebra with respect to the dual norm $\| \ \|$. This norm satisfies

$$\|u\|_\infty \leqq \|u\| \quad (u \in B(G))$$
$$\|p\| = \|p\|_\infty = p(e) \quad (p \in P(G))$$

where $\|u\|_\infty = \sup_{t\in G}|u(t)|$ and $e\in G$ denotes the unit element of G. $B(G)$ is an involutive Banach algebra for the involution $^-$ given by

$$\bar{u}(t)=\overline{u(t)}$$

(where $\bar{z}$ denotes the complex conjugate of $z\in\mathbb{C}$).

The cone $P(G)$ defines an order relation on $B(G)$. We refer to it as the *positive definite ordering.*

There is another order relation on $B(G)$ which is defined by the cone $B(G)_+$ of all pointwise positive functions in $B(G)$. That is, for $u\in B(G)$, $u\in B(G)_+$ if and only if $u(t)\geqq 0$ for all $t\in G$.

We refer to this order relation as to the *pointwise ordering* and write

$$u\leqq v \quad \text{if and only if} \quad v-u\in B(G)_+ .$$

Of course, $B(G)_+$ is a proper cone. It is also *generating*, that is, $\operatorname{span} B(G)_+ = B(G)$. In fact, if $u\in B(G)$, then $\operatorname{Re}u=\frac{1}{2}(u+\bar{u})\in B(G)$ and $\operatorname{Im}u=\frac{1}{2i}(u-\bar{u})\in B(G)$. Since $u=\operatorname{Re}u+i\operatorname{Im}u$, it is enough to show that $v\in\operatorname{span}B(G)_+$, whenever v is a real valued function in $B(G)$. Since $-\|v\|_\infty 1\leqq v\leqq\|v\|_\infty 1$, it follows that $v_1 = v+\|v\|_\infty 1\in B(G)_+$ and $v_2=\|v\|_\infty 1-v\in B(G)_+$. So $v=\frac{1}{2}(v_1-v_2)\in\operatorname{span}B(G)_+$.

Corresponding to these two orderings are the notions of positive linear mappings: A linear mapping $T: B(G_1)\to B(G_2)$ is *positive for the positive definite ordering* (resp., *the pointwise ordering*) if $TP(G_1))\subset P(G_2)$ (respectively, $TB(G_1)_+\subset B(G_2)_+$). T is called an *order isomorphism for the positive definite ordering* (respectively, *the pointwise ordering*) if T is bijective and $TP(G_1)=P(G_2)$ (respectively, $TB(G_1)_+ = B(G_2)_+$). Finally, we say that T is a *biorder isomorphism* if T is an order isomorphism for the positive definite and the pointwise ordering.

3. Kawada's Theorem for the Fourier-Stieltjes Algebra

Kawada showed that the ordered algebra $L^1(G)$ is a complete invariant for G. More precisely: G_1 and G_2 are isomorphic if and only if there exists an algebra isomorphism T from $L^1(G_1)$ onto $L^1(G_2)$ such that $f\geqq 0$ (a.e.) if and only if $Tf\geqq 0$ (a.e.) for all $f\in L^1(G_1)$ [6, Theorem 1]. The purpose of this section is to establish an analogue of Kawada's theorem for the Fourier-Stieltjes algebra. This amounts to replacing, in Walter's theorem, the assumption that the algebra isomorphism T is isometric by the assumption that T is an order isomorphism for the positive definite ordering. Actually, we find a device (Lemma 3.1) which also allows the deduction of a "Kawada-type-theorem" from a "Wendel-type-theorem" in other situations (here we make allusion to Wendel's classical theorem [10, Theorem 1] characterizing isometric isomorphisms of group algebras).

Let $\mathfrak{A}$ be a von Neumann algebra and R its predual. By $\mathfrak{A}_+$ we denote the usual positive cone (of all hermitian elements with positive spectrum) in $\mathfrak{A}$. We denote by $R_+\subset R$ the cone of all positive, normal functionals on $\mathfrak{A}$. Then $\mathfrak{A}_+$ is the

dual cone of R_+, i.e.

$$\mathfrak{A}_+ = \{x \in \mathfrak{A} | \langle f, x \rangle \geqq 0 \quad \text{for all} \quad f \in R_+\}.$$

Let 1 be the algebra unit in $\mathfrak{A}$. Then

$$\|f\| = \langle f, 1 \rangle \quad \text{for all} \quad f \in R_+. \tag{3.1}$$

Assume now in addition, that R is a Banach algebra. We will consider the condition that 1 is a multiplicative linear functional on R, that is

$$\langle f \cdot g, 1 \rangle = \langle f, 1 \rangle \langle g, 1 \rangle \quad (f, g \in R). \tag{3.2}$$

Since $\operatorname{span} R_+ = R$, (3.2) is equivalent to

$$\|f \cdot g\| = \|f\| \cdot \|g\| \quad \text{for all} \quad f, g \in R_+. \tag{3.3}$$

Lemma 3.1. *For $i=1, 2$, let R_i be a Banach algebra and the predual of a von Neumann algebra $\mathfrak{A}_i$ such that the unit element, 1, of $\mathfrak{A}_i$ is a multiplicative linear functional on R_i. Let $T: R_1 \to R_2$ be an algebra isomorphism such that*

$$T(R_1)_+ = (R_2)_+.$$

Denote the adjoint of T by T'. Then $T'1 = 1$ and T is an isometry.

Proof. First of all, T is continuous as a consequence of the positivity condition (see, e.g., the proof of [1, 3.1]). For $i=1, 2$ let M_i denote the set of all multiplicative linear functionals on R_i. If $x \in M_i$, then x is bounded and $\|x\| \leqq 1$ [5, VIII, 2.8 Lemma]. Hence $x \leqq 1$ for every $x \in M_i \cap (\mathfrak{A}_i)_+$. By our assumption, it follows that $1 = \max\{x | x \in (\mathfrak{A}_i)_+ \cap M_i\}$. Since T is an algebra isomorphism, we have $T'M_2 = M_1$; and since $T'(\mathfrak{A}_2)_+ = (\mathfrak{A}_1)_+$, it follows that $T'((\mathfrak{A}_2)_+ \cap M_2) = (\mathfrak{A}_1)_+ \cap M_1$
. Consequently,

$$\begin{aligned} T'1 &= T'(\max\{x | x \in (\mathfrak{A}_2)_+ \cap M_2\}) = \max\{T'x | x \in (\mathfrak{A}_2)_+ \cap M_2\} \\ &= \max\{x | x \in (\mathfrak{A}_1)_+ \cap M_1\} = 1. \end{aligned}$$

It follows from [7, Corollary 1] that T' is a contraction. Similarly, $(T')^{-1}$ is contractive. Thus T' is an isometry, and consequently T is an isometry, as well.

Theorem 3.2. *Let $T: B(G_1) \to B(G_2)$ be an algebra isomorphism and an order isomorphism for the positive definite ordering. Then there exists a topological group isomorphism or antiisomorphism $\alpha: G_2 \to G_1$ such that*

$$Tf = f \circ \alpha \quad \textit{for all} \quad f \in B(G_1).$$

Proof. $R = B(G)$ is a Banach algebra which satisfies the conditions considered in Lemma 3.1. In fact, $B(G)$ is the dual of $C^*(G)$. Hence its dual is a von Neumann algebra because it is the bidual of a C^*-algebra. Moreover, $R_+ = P(G)$ and for $u \in P(G)$, $\langle u, 1 \rangle = \|u\| = u(e)$. Hence 1 is a multiplicative functional on $B(G)$. So we can apply 3.1 to conclude that T is an isometry. It follows from Walter's theorem [9, Theorem 2] that there exists a topological group isomorphism or anti-

isomorphism $\alpha: G_2 \to G_1$ and $s \in G_1$ such that $Tf(t) = f(s\alpha(t))$. In particular, $f(e) = \langle f, 1\rangle = \langle f, T'1\rangle = f(s)$ for all $f \in B(G_1)$. This implies that $s = e$, so that T has the desired form.

In the same way, one can obtain Kawada's theorem [6, Theorem 1] from Wendel's theorem [10, Theorem 1] and also our earlier result [1, Theorem 3.2] (a Kawada type theorem for $A(G)$) from [9, Theorem 3] (a Wendel type theorem for $A(G)$) with the help of Lemma 3.1.

The situation for $M(G)$ is similar. We describe it explicitly: $M(G)$ denotes the Banach space of all bounded regular complex Borel measures on G. With convolution as multiplication, $M(G)$ is a Banach algebra. For $\mu \in M(G)$, we write $\mu \geqq 0$ if μ is a positive measure in the usual sense. $M(G)$ is the predual of a von Neumann algebra and (3.3) is satisfied. So Lemma 3.1 allows to deduce the following result from Johnson's theorem [4, Theorem 3].

Theorem 3.3. *Let $T: M(G_1) \to M(G_2)$ be an algebra isomorphism such that*

$$\mu \geqq 0 \quad \text{if and only if} \quad T\mu \geqq 0$$

($\mu \in M(G_1)$). Then there exists a topological group isomorphism $\alpha: G_2 \to G_1$ such that T has the form

$$T\mu(K) = \mu(\alpha(K))$$

for all $\mu \in M(G_1)$ and all compact subsets K of G_2.

Remark. It is possible to prove 3.3 directly via Kawada's theorem: A subspace J of $M(G)$ is called an *order ideal* if $|\nu| \leqq |\mu|$, where $\mu \in J$, implies that $\nu \in J$. Now $L^1(G)$ is the smallest element (for inclusion) in the set of all closed, non-zero order and algebra ideals of $M(G)$. [Indeed, let J be a closed, non-zero order and algebra ideal in $M(G)$. Suppose, there exists $t_0 \in G$ such that $f(t_0) = 0$ for all $f \in J_0 := J \cap L^1(G) \cap C^b(G)$. Then $f(e) = (\delta_{t_0} * f)(t_0) = 0$ for all $f \in J_0$. Hence if $\mu \in J$, then $\langle f, \mu\rangle = \check{f} * \mu(e) = 0$ for all continuous functions f on G with compact support (where $\check{f}(t) = f(t^{-1})$ (for $t \in G$)), i.e. $\mu = 0$, a contradiction. Thus for every $t \in G$ there exists $f \in J_0$ such that $f(t) \neq 0$. Since J is a closed order ideal, it follows that $L^1(G) \subset J$.] Thus, if $T: M(G_1) \to M(G_2)$ is an order and algebra isomorphism, then $TL^1(G_1) = L^1(G_2)$ and we can apply Kawada's theorem to conclude that G_1 and G_2 are isomorphic. Moreover, T is uniquely determined by its restriction to $L^1(G_1)$. So the precise form of T given in 3.3 can be obtained once it is established for the restriction. This is done through Wendel's modification [10, Theorem 2] of Kawada's result.

4. Multipliers

By $M(B(G))$ we denote the space of all multipliers of $B(G)$, i.e. (since $B(G)$ has a unit)

$$M(B(G)) = \{M_h | h \in B(G)\}$$

where $M_h: B(G) \to B(G)$ is defined by $M_h f = h \cdot f$ ($f \in B(G)$).

Lemma 4.1. *Let $M: B(G) \to B(G)$ be a linear mapping. Then M is a multiplier if and only if for every $f \in B(G)$ and $t \in G$, $f(t) = 0$ implies $Mf(t) = 0$.*

Proof. The condition is clearly necessary. To prove sufficiency assume that $f(t)=0$ implies $Mf(t)=0$ ($f\in B(G), t\in G$). Let $h=M1$ and $g\in B(G)$. We show that $Mg=h\cdot g$. Let $t\in G$. Then $(g-g(t)1)(t)=0$. Hence by our assumption,

$$Mg(t)-g(t)h(t)=M(g-g(t)1)(t)=0.$$

That is,

$$Mg(t)=h(t)g(t) \quad \text{for every} \quad t\in G.$$

We give now an order theoretical characterization of multipliers which are positive for the pointwise ordering.

Proposition 4.2. *Let $M: B(G))\to B(G)$ be a linear mapping satisfying $MB(G)_+ \subset B(G)_+$. Then M is a multiplier if and only if there exists $c\geqq 0$ such that*

$$Mf\leqq cf \quad \textit{for all} \quad f\in B(G)_+. \tag{4.1}$$

Proof. If $M=M_h$ for some $h\in B(G)$, then $h=M1\in B(G)_+$. Since $h\leqq \|h\|_\infty 1$, it follows that $Mf=hf\leqq \|h\|_\infty f$ for all $f\in B(G)_+$. This establishes one implication. To prove the other one we assume that (4.1) is valid. Let $g\in B(G), t\in G$ such that $g(t)=0$. In order to apply 4.1 we have to prove that $Mg(t)=0$. Consider the scalar product

$$[k,h]=M(k\cdot\bar{h})(t) \quad (k,h\in B(G)).$$

Applying the Cauchy-Schwarz inequality to $k=g$ and $h=1$ we obtain:

$$|Mg(t)|^2=|[g,1]|^2\leqq [g,g]\,[1,1]=M(|g|^2)(t)\,[1,1]. \tag{4.2}$$

Since $g(t)=0$, it follows from (4.1) that $M(|g|^2)(t)\leqq c|g|^2(t)=c|g(t)|^2=0$, and so by (4.2) we conclude that $Mg(t)=0$.

Next, we characterize isometric multipliers and multipliers which are order isomorphisms for the positive definite ordering.

Lemma 4.3. *Let $h\in B(G)$. If $\|h\|=|h(t)|\neq 0$ for all $t\in G$, then $h/h(e)$ is an character (i.e. a homomorphism from G into $\mathbb{C}$).*

Proof. By [3, (2.14)] there exist a unitary representation $\pi: G\to\mathscr{L}(H)$ in some Hilbert space H and $\xi,\eta\in H$ such that $h(t)=(\pi(t)\xi,\eta)$ and $\|h\|=\|\xi\|\cdot\|\eta\|$. Hence

$$\|h\|=|h(t)|=|(\pi(t)\xi|\eta)|\leqq \|\xi\|\cdot\|\eta\|=\|h\| \qquad (t\in G).$$

Thus the equality in the Cauchy-Schwarz inequality holds, and so there exists $\lambda(t)\in\mathbb{C}$ such that $\pi(t)\xi=\lambda(t)\eta$. In particular, $\xi=\lambda(e)\eta$ so that $\pi(t)\xi=(\lambda(t)/\lambda(e))\xi$. Hence $\chi=\dfrac{\lambda}{\lambda(e)}$ is a character and $h=c\cdot\chi$, where $c=(\xi|\eta)=h(e)$.

Proposition 4.4. *Let $h\in B(G)$. M_h is isometric if and only if $h=c\cdot\chi$, where χ is a continuous character on G and $c\in\mathbb{C}$, $|c|=1$.*

Proof. If χ is a continuous character, then $\chi\in P(G)$. Hence $\|\chi\|=\chi(e)=1$. Similarly, $\|\chi^{-1}\|=1$. Thus $\|M_\chi\|=\|(M_\chi)^{-1}\|=1$; i.e., M_χ is an isometry. Conversely, suppose that $M=M_h$ is an isometry, where $h\in B(G)$. Then $\|h\|=\|M1\|=\|1\|=1$. For $t\in G$, let $\delta_t\in B(G)'$ be defined by $\langle u,\delta_t\rangle=u(t)$ ($u\in B(G)$). Then $\|\delta_t\|=1$. Hence

$|h(t)| = \|h(t)\delta_t\| = \|M'\delta_t\| = \|\delta_t\| = 1$ $(t \in G)$ (here M' denotes the adjoint of M). It follows from 4.3 that $h/h(e)$ is a character. Hence h has the desired form.

Proposition 4.5. *Let $h \in B(G)$. M_h is an order isomorphism for the positive definite ordering if and only if*

$$h = c \cdot \chi, \text{ where } \chi \text{ is a continuous character and } c > 0. \tag{4.3}$$

Proof. Suppose that M_h is an order isomorphism for the positive definite ordering. Then $h = M_h 1 \in P(G)$ and $\frac{1}{h} = M_h^{-1} 1 \in P(G)$. Let $\chi = h/h(e)$. Since $\chi \in P(G)$, we have $|\chi(t)| \leqq \chi(e) = \|\chi\| = 1$. Moreover, $1/\chi \in P(G)$ so that $|1/\chi(t)| \leqq 1/\chi(e) = 1$ for all $t \in G$. Consequently, $|\chi(t)| = 1 = \|\chi\|$ for all $t \in G$. It follows from 4.3 that χ is a character. Hence $h = h(e)\chi$ has the form (4.3). The converse implication is obvious.

5. Order Isomorphisms for Fourier-Stieltjes Algebras

Proposition 5.1. *Let $V: B(G_1) \to B(G_2)$ be a $*$-algebra isomorphism (that is, V is a bijective linear mapping such that $\overline{Vf} = V\bar{f}$ and $V(f \cdot g) = Vf \cdot Vg$ for all $f, g \in B(G)$). Then V is an order isomorphism for the pointwise ordering.*

Proof. Denote by $A(G_i)$ the Fourier algebra of G_i $(i = 1, 2)$ (see [3] for the definition). We show first that $Vf \geqq 0$ whenever $f \in A(G_1)$ and $f \geqq 0$. In fact, for a given $t \in G_2$, the map $f \to Vf(t)$ is a multiplicative linear functional on $A(G_1)$. Let Y be the set of all $t \in G_2$ such that this functional does not vanish identically on $A(G_1)$. Then by [3, (3.34)], for all $t \in Y$ there exists $\alpha(t) \in G_1$ such that $Vf(t) = f(\alpha(t))$. Since $Vf(t) = 0$ for $t \notin Y$, it follows that $Vf \geqq 0$ whenever $f \in A(G_1)$ and $f \geqq 0$. Replacing V with V^{-1} we also have

$$V^{-1} g \geqq 0 \quad \text{for} \quad g \in A(G_2), \quad g \geqq 0. \tag{5.1}$$

We show now that V is positive for the pointwise ordering. If this is not true, then there exists $f \in B(G_1)_+$ such that $Vf(t) < 0$ for some $t \in G_2$ (observe that $Vf\,(= V\bar{f} = \overline{Vf})$ is real). Choose $h \in A(G_2)$, $h \geqq 0$, such that $h(t) \neq 0$ and h vanishes outside $\{s \in G_2 | Vf(s) < 0\}$ (use [3, (3.2)]). Then $0 \neq Vf \cdot h \leqq 0$. Since $A(G_2)$ is an ideal in $B(G_2)$, $Vf \cdot h \in A(G_2)$. So it follows from (5.1) that $0 \neq f \cdot V^{-1} h = V^{-1}(Vf \cdot h) \leqq 0$. Since $V^{-1} h \geqq 0$ (again by (5.1)), this contradicts $f \geqq 0$. Replacing V with V^{-1} we obtain that V^{-1}, too, is positive for the pointwise ordering; so V is an order isomorphism for the pointwise ordering.

We can now characterize order isomorphisms for the pointwise ordering.

Theorem 5.2. *Let $T: B(G_1) \to B(G_2)$ be a linear mapping. T is an order isomorphism for the pointwise ordering if and only if T has the form*

$$Tf = h \cdot Vf \qquad (f \in B(G_1)), \tag{5.2}$$

where $V: B(G_1) \to B(G_2)$ is a $*$-algebra isomorphism, $h \in B(G_2)_+$ and h is invertible in $B(G_2)$ (that is, $h(t) > 0$ for all $t \in G_2$ and $\frac{1}{h} \in B(G_2)$).

Remark. Order isomorphisms for the pointwise ordering are continuous. In fact, if $S: B(G_1) \to B(G_2)$ is linear and positive, then S is continuous for the supremum norm (since $-1 \leqq u \leqq 1$ implies $-S1 \leqq Su \leqq S1$). So it follows from the closed graph theorem that S is continuous.

Proof. It is clear from 5.1 that every operator of the form (5.2) is an order isomorphism for the pointwise ordering. To prove the converse assume that $T: B(G_1) \to B(G_2)$ is an order isomorphism for the pointwise ordering. Let $f \in B(G_1)_+$. $M = TM_f T^{-1}$ is a linear mapping from $B(G_2)$ into $B(G_2)$. We show that M is a multiplier. In fact, M is positive for the pointwise ordering. Moreover, since $0 \leqq f \leqq \|f\|_\infty 1$, it follows that $f \cdot T^{-1}g \leqq \|f\|_\infty T^{-1}g$, and consequently $Mg = T(f \cdot T^{-1}g) \leqq \|f\|_\infty g$ for all $g \in B(G_2)_+$. It follows from 4.2 that $M \in M(B(G_2))$. Since span $B(G_1)_+ = B(G_1)$ we conclude that

$$TM_f T^{-1} \in M(B(G_2)) \quad \text{for every} \quad f \in B(G_1). \tag{5.3}$$

So there exists a mapping $V: B(G_1) \to B(G_2)$ such that

$$TM_f T^{-1} = M_{Vf} \quad \text{for all} \quad f \in B(G_1), \tag{5.4}$$

with the property that $Vf \geqq 0$ if $f \geqq 0$.

From (5.4) follows that V is a $*$-algebra isomorphism. In fact, for $g, h \in B(G_1)$,

$$M_{V(f \cdot g)} = TM_{f \cdot g} T^{-1} = TM_f M_g T^{-1} = TM_f T^{-1} TM_g T^{-1} = M_{Vf} \cdot M_{Vg} = M_{(Vf)(Vg)}.$$

This implies that $Vf \cdot Vg = V(f \cdot g)$. Similarly, $V(\alpha f + \beta g) = \alpha Vf + \beta Vg$, $\overline{Vf} = V\bar{f}$ $(\alpha, \beta \in \mathbb{C})$; and it is easy to see by exchanging T and T^{-1} that V is bijective.

Finally, let $h = T1$. We show that $Tf = h \cdot Vf$ for all $f \in B(G_1)$. Indeed, $Tf = TM_f 1 = TM_f T^{-1} T1 = M_{Vf} h = h \cdot Vf$ (by (5.4)). Of course, $h = T1 \geqq 0$. Let $g = T^{-1}1$. Then $h \cdot Vg = Tg = 1$. Hence h is invertible in $B(G_2)$ and $\frac{1}{h} = Vg \in B(G_2)$.

Remark. The above theorem is the analogue of [1, Proposition 4.2], where it is proved that if $T: A(G_1) \to A(G_2)$ is a continuous order isomorphism for the pointwise ordering, then there exists a bijection $\alpha: G_2 \to G_1$ and a function $h: G_2 \to (0, \infty)$ such that $Tf = h \cdot f \circ \alpha$ for all $f \in A(G_1)$ (it is easy to show that α is actually a homeomorphism and h is continuous). To see the analogy, one has to observe that $f \circ \alpha \in M(A(G_2))$ (the algebra of multipliers of $A(G_2)$, considered as functions on G_2) whenever $f \in A(G_1)$. Indeed, if $g \in A(G_2)$, there exists $g_1 \in A(G_1)$ such that $g = Tg_1$, and $(f \circ \alpha) \cdot g = T(f \cdot g_1) \in A(G_2)$. Consequently the conclusion of [1, Proposition 4.2] can be restated so as to assert the existence of an injective $*$-homomorphism $V: A(G_1) \to M(A(G_2))$, $f \to f \circ \alpha$, and a function h as above such that $Tf = h \cdot Vf$ for all $f \in A(G_1)$. In this form, Proposition 4.2 of [1] can be proved along the same lines as Theorem 5.2 above.

If moreover G_1 and G_2 are amenable, V actually maps $A(G_1)$ onto $A(G_2)$, and $h \in B(G_2)$.

The following is the main theorem.

Theorem 5.3. *Let $T: B(G_1) \to B(G_2)$ be a biorder isomorphism. Then there exists a topological group isomorphism or antiisomorphism $\alpha: G_2 \to G_1$ and a constant $c > 0$*

such that

$$Tf = c \cdot f \circ \alpha$$

for all $f \in B(G_1)$.

Proof. By 5.2, $T = M_h \cdot V$ where V is a $*$-algebra isomorphism and $h \in B(G_2)_+$ is invertible in $B(G_2)$. Moreover, $M_{Vg} = TM_g T^{-1}$ for all $g \in B(G_1)$ (by (5.4)). Let $g \in P(G_1)$. Then M_g and consequently $M_{Vg} = TM_g T^{-1}$ are positive linear mappings for the positive definite ordering. This implies that $Vg = M_{Vg} 1 \in P(G_2)$. Hence V is positive for the positive definite ordering. The same argument applied to V^{-1} shows that V is an order isomorphism for the positive definite ordering. It follows from Theorem 3.2 that there exists a topological group isomorphism or anti-isomorphism $\alpha: G_2 \to G_1$ such that $Vf = f \circ \alpha$ for all $f \in B(G_1)$. Moreover, $M_h = TV^{-1}$ is an order isomorphism for the positive definite ordering. So $h/h(e)$ is a character by 4.5. Since $h \geqq 0$, it follows that $h(t)/h(e) = 1$ for all $t \in G_2$. Thus, with $c = h(e)$, $Tf = c \cdot f \circ \alpha$ for all $f \in B(G_1)$.

Theorem 5.4. *Let* $T: B(G_1) \to B(G_2)$ *be an isometric order isomorphism for the pointwise ordering. Then there exists a topological group isomorphism or anti-isomorphism* $\alpha: G_2 \to G_1$, *and* $s \in G_1$, *such that*

$$Tf(t) = f(s \cdot \alpha(t)) \ (t \in G_2) \quad \textit{for all} \quad f \in B(G_1).$$

Proof. By 5.2 there exists a $*$-algebra isomorphism $V: B(G_1) \to B(G_2)$ and $h \in B(G_2)$ such that $T = M_h V$. Moreover, by (5.4), $TM_f T^{-1} = M_{Vf}$ for all $f \in B(G_1)$. Hence, $\|Vf\| = \|M_{Vf}\| = \|TM_f T^{-1}\| = \|M_f\| = \|f\|$ for all $f \in B(G_1)$; i.e. V is an isometry. By Walter's theorem [9, Theorem 2], there exists a topological group isomorphism or antiisomorphism $\alpha: G_2 \to G_1$ and $s \in G$, such that $Vf(t) = f(s \cdot \alpha(t))$ $(t \in G_2)$ for all $f \in B(G_1)$. Since V and T^{-1} are isometries, $M_h = TV^{-1}$ is also isometric. It follows from 4.4 that $h = c \cdot \chi$, where χ is a character and $c \in \mathbb{C}$, $|c| = 1$. Moreover, because of 5.1 V^{-1} is an order isomorphism for the pointwise order; and consequently, so is M_h. This implies that $h = 1$ so that $T = V$.

Taken together with the Theorems 3.2 and 5.3, the preceding result allows us to formulate the following conclusion, which summarizes the remarkable interplay between the algebraic, metric, and order structure of the Fourier-Stieltjes algebra: In Walter's theorem one can replace at will the multiplicative structure of $B(G)$ by the pointwise ordering, and/or its norm by the positive definite ordering.

6. Order Isomorphisms of Group Algebras

The space $L^1(G)$ is defined as usual with respect to a fixed left Haar measure. $L^1(G)$ is an involutive Banach algebra with convolution as multiplication and the involution * defined by

$$f^*(t) = \Delta(t)^{-1} \overline{f(t^{-1})} \qquad (f \in L^1(G)),$$

where Δ denotes the modular function of G. Again, there exist two orderings on $L^1(G)$. The *positive definite ordering* is defined by its involutive Banach algebra structure; i.e. by the closed convex cone $L^1(G)_p = \overline{co}\,\{g^* * g | g \in L^1(G)\}$.

By [2, 13.4.5] the dual cone of $L^1(G)_p$ in $L^\infty(G)$ is $P(G)$; that is

$$P(G)=\{p\in L^\infty(G)|\langle f,p\rangle\geqq 0 \quad \text{for all} \quad f\in L^1(G)_p\}. \tag{6.1}$$

Remark 6.1. If G is abelian, then $L^1(G)_p=\{f\in L^1(G)|\hat{f}(\gamma)\geqq 0 \text{ for all } \gamma\in\hat{G}\}$, where $\hat{G}$ denotes the dual group of G.

The *pointwise ordering* on $L^1(G)$ is defined by the positive cone $L^1(G)_+=\{f\in L^1(G)|f(t)\geqq 0 \text{ for almost every } t\in G\}$.

Corresponding to these two orderings we have again the notion of order isomorphisms: Let $T: L^1(G_1)\to L^1(G_2)$ be a bijective linear mapping. T is an *order isomorphism for the positive definite ordering* (respectively *for the pointwise ordering*) if $TL^1(G_1)_p=L^1(G_2)_p$ (respectively $TL^1(G_1)_+=L^1(G_2)_+$). T is called a *biorder isomorphism* if T is an order isomorphism for both orderings.

Let $\alpha: G_2\to G_1$ be a topological group isomorphism. Because of the uniqueness of the Haar measure, for every measurable function f on G_1, $f\in L^1(G_1)$ if and only if $f\circ\alpha\in L^1(G_2)$ and there exists a unique constant $c(\alpha)>0$ such that

$$c(\alpha)\int_{G_2} f(\alpha(t))dt=\int_{G_1} f(s)ds \quad \text{for all} \quad f\in L^1(G_1). \tag{6.2}$$

From this it is easy to see that the mapping

$$V_\alpha: L^1(G_1)\to L^1(G_2),\ V_\alpha f=c(\alpha)f\circ\alpha \tag{6.3}$$

is a $*$-algebra isomorphism. Hence V_α is an order isomorphism for the positive definite ordering (this is immediate from the definition of this ordering). Moreover, V_α is obviously an order isomorphism for the pointwise ordering. Thus V_α is a biorder isomorphism.

The mapping $W: L^1(G_2)\to L^1(G_2)$ defined by $Wf=\Delta_2^{-1}\check{f}$ $(f\in L^1(G_2))$, where $\check{f}(t)=f(t^{-1})$ and Δ_2 denotes the modular function of G_2, is a biorder isomorphism, too. Indeed, it is obviously an order isomorphism for the pointwise ordering. Moreover, the adjoint $W': L^\infty(G_2)\to L^\infty(G_2)$ of W is given by $W'f=\check{f}$. Hence $W'P(G_2)=P(G_2)$. It follows from (6.1) that W is an order isomorphism for the positive definite ordering. Thus,

$$T=WV_\alpha \text{ is a biorder isomorphism}. \tag{6.4}$$

The following theorem shows that up to a positive constant every biorder isomorphism has the form (6.3) or (6.4).

Theorem 6.2. *Let $T: L^1(G_1)\to L^1(G_2)$ be a biorder isomorphism. Then there exists a topological group isomorphism $\alpha: G_2\to G_1$ and a constant $c>0$ such that*

$$Tf(t)=c\cdot f(\alpha(t)) \quad (t\in G_2, f\in L^1(G_1)) \tag{6.5}$$

or

$$Tf(t)=c\cdot\Delta_2(t)^{-1}f(\alpha(t^{-1})) \quad (t\in G_2, f\in L^1(G_1)). \tag{6.6}$$

Note. Concerning the formulation of (6.6), one should observe that $\alpha: G_2\to G_1$ is a topological group isomorphism if and only if $\beta: G_2\to G_1$ defined by $\beta(t)=\alpha(t^{-1})$ is a topological group antiisomorphism.

Proof. By [8, II, 5.3], T is continuous. Denote by $T': L^\infty(G_2) \to L^\infty(G_1)$ the adjoint of T. Since T is an order isomorphism for the positive definite ordering, it follows from (6.1) that $T'P(G_2) = P(G_1)$. Hence $T'B(G_2) = B(G_1)$ and so the restriction S of T' to $B(G_2)$ is an order isomorphism for the positive definite ordering from $B(G_2)$ onto $B(G_1)$. Since T is an order isomorphism for the pointwise ordering, S is so, too. By Theorem 5.2, there exists a topological group isomorphism $\beta: G_1 \to G_2$ and a constant $c_1 > 0$ such that S is given by

$$Su(s) = c_1 u(\beta(s)) \quad (s \in G_1, u \in B(G_2)) \tag{6.7}$$

or

$$Su(s) = c_1 u(\beta(s^{-1})) \quad (s \in G_1, u \in B(G_2)). \tag{6.8}$$

If (6.7) is valid, then by (6.2),

$$\langle Tf, u \rangle = \langle f, Su \rangle = c_1 \int_{G_1} f(s) u(\beta(s)) ds$$

$$= c_1 \cdot \frac{1}{c(\beta)} \int_{G_2} f(\beta^{-1}(t)) u(t) dt$$

for all $f \in L^1(G_1)$, $u \in B(G_2)$. Since $B(G_2)$ separates $L^1(G_2)$, it follows that $Tf = (c_1/c(\beta)) \cdot f \circ \beta^{-1}$ for all $f \in L^1(G_1)$. Hence (6.5) is valid for $c = c_1/c(\beta)$ and $\alpha = \beta^{-1}$.

If (6.8) holds, then by (6.2) again

$$\langle Tf, u \rangle = c_1 \int_{G_1} f(s) u(\beta(s^{-1})) ds = c_1 \int_{G_1} f(s) \check{u}(\beta(s)) ds$$

$$= c_1 \cdot \frac{1}{c(\beta)} \int_{G_2} f(\beta^{-1}(t)) \check{u}(t) dt$$

$$= \frac{c_1}{c(\beta)} \int_{G_2} \Delta_2(t)^{-1} f(\beta^{-1}(t^{-1})) u(t) dt$$

for all $f \in L^1(G_1)$, $u \in B(G_2)$. Hence $Tf(t) = (c_1/c(\beta)) \Delta_2(t)^{-1} f(\beta^{-1}(t^{-1}))$ for all $f \in L^1(G_1)$, and almost all $t \in G_2$. Thus T has the form (6.6) with $c = c_1/c(\beta)$ and $\alpha = \beta^{-1}$.

Remark 6.3. Theorem 6.2 implies that the "biordered space" $(L^1(G), L^1(G)_+, L^1(G)_p)$ is a complete invariant for the locally compact group G.

Remark 6.4. For abelian groups, Theorem 6.2 has been proved in [1] (in fact, by Remark 6.1, Theorem 6.2 is another formulation of [1, 4.3]).

References

1. Arendt, W., De Cannière, J.: Order isomorphisms of Fourier algebras. J. Funct. Anal. **50**, 1–7 (1983)
2. Dixmier, J.: Les C^*-algèbres et leurs représentations. Paris: Gauthier-Villars 1969
3. Eymard, P.: L'algèbre de Fourier d'un groupe localement compact. Bull. Soc. Math. France **92**, 181–236 (1964)

4. Johnson, B.E.: Isometric isomorphisms of measure algebras. Proc. Am. Math. Soc. **15**, 186–188 (1964)
5. Katznelson, Y.: An introduction to harmonic analysis. New York: John Wiley & Sons 1968
6. Kawada, Y.: On the group ring of a topological group. Math. Japon **1**, 1–5 (1948)
7. Russo, B., Dye, H.A.: A note on unitary operators in C^*-algebras. Duke Math. J. **33**, 413–416 (1966)
8. Schaefer, H.H.: Banach lattices and positive operators. Berlin, Heidelberg, New York: Springer 1974
9. Walter, M.E.: W^*-algebras and nonabelian harmonic analysis. J. Funct. Anal. **11**, 17–38 (1972)
10. Wendel, J.G.: On isometric isomorphisms of group algebras. Pac. J. Math. **1**, 305–311 (1951)

Received August 27, 1982

Math. Ann. 263, 157–177 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Propagation of Holomorphic Extendability of CR Functions

Nicholas Hanges[1,*] and François Treves[2,**]

[1] Department of Mathematics, University of Pittsburgh, Pittsburgh, PA 15260, USA
[2] Department of Mathematics, Rutgers University, New Brunswick, NJ 08903, USA

Contents

Introduction 157
1. Hypo-Analytic Manifolds 158
2. Elliptic and Complex Submanifolds 162
3. Propagation of Hypo-Analyticity Along Elliptic Submanifolds 166
4. Applications, Examples, Singular Solutions 174
References 177

Introduction

Let Ω be a CR submanifold of $\mathbb{C}^N$ and p_0 a point of Ω. Suppose that a CR function u on Ω extends holomorphically to some neighborhood of p_0 in $\mathbb{C}^N$. Can we ever conclude from this fact that u must extend holomorphically to neighborhoods of other points p of Ω, points that may lie far from p_0? This typifies the kind of question discussed in the present article. Our answer (Theorem 4.1) is *yes* whenever p lies on a connected holomorphic (i.e., complex) submanifold $\mathscr{M}$ of $\mathbb{C}^N$, entirely contained in Ω and passing through p_0.

Our discussion is not limited to CR submanifolds of $\mathbb{C}^N$. It encompasses the hypo-analytic manifolds introduced in [1], of which embedded CR manifolds are but an example. Such a generalization enables us to deal with all first-order systems of linear *PDE* for which the question makes sense, and not just for the tangential Cauchy-Riemann equations. In the more general set-up the role of propagators of hypo-analyticity (roughly speaking, holomorphic extendability) is played by the *elliptic* submanifolds (defined in Sect. 2). In the CR case the latter are the complex submanifolds; in the general case they will contain "real" directions in addition to the "complex" ones.

These results are related to those on propagation of analytic singularities of [3] for solutions of analytic pseudodifferential equations. Here the manifold Ω and the coefficients of the *PDE* whose solutions we study are C^∞, and most of the time they

* Supported by NSF Grant MCS-8100505 and by an Alfred P. Sloan fellowship
** Supported by NSF Grant MCS-8102435

could have been taken to be C^2 – if it were not that we usually deal with distribution solutions (these are called RC distributions following the terminology of [7]).

There are two other aspects of our work we wish to emphasize: i) the manifolds we study are not necessarily of the hypersurface type. In the embedded CR case their codimension is arbitrary; ii) the elliptic submanifolds we identify as propagators do not make up any kind of foliation; they might be isolated; there might just be two (or three or four, ...) elliptic submanifolds through the central point (Example 2.1).

The general theorem (Theorem 3.1) is stated and proved in Sect. 3. Its proof, which is wholly constructive, is based on use of the Fourier-Bros-Iagolnitzer transform, whose significance in the present context was first established in [1]. Much of the argument is borrowed directly from Chap. 1 of [1].

It is conceivable that, under special circumstances, submanifolds other than the elliptic ones may serve as propagators of hypo-analyticity. In order to circumscribe such phenomena one would like to construct solutions whose set of nonextendability (their *hypo-analytic singular support*, in our terminology) is as small as possible. We show in Sect. 4 that this is not feasible if one clings to any kind of generality (even for embedded CR manifolds). But there is one special situation when the construction – on a CR manifold – is possible: when through the central point ("the origin") there passes one complex manifold of the highest possible dimension, i.e. whose complex dimension is exactly equal to the rank of the tangential Cauchy-Riemann equations. We show (Proposition 2.1) that such a submanifold is then unique and (Theorem 4.3) that one can construct a CR distribution (and, if one wishes, a CR C^∞ function), in some open neighborhood of the origin, U, which is holomorphically extendable at every point of $U\backslash M$ and at none of $U\cap M$. Such and related properties are discussed in Sect. 4.

1. Hypo-Analytic Manifolds

We shall recall, in the present section, the main concepts relating to hypo-analyticity introduced in [1], Chap. I.

Throughout the article Ω will denote a C^∞ manifold of dimension $m+n$. A *hypo-analytic structure* on Ω is the data of an open covering (U_α) of Ω and for each index α, of m C^∞ functions $Z_\alpha^1, \ldots, Z_\alpha^m$, satisfying the following two requirements:

i) $dZ_\alpha^1, \ldots, dZ_\alpha^m$ are linearly independent at each point of U;

ii) if $U_\alpha \cap U_\beta \neq \emptyset$ there is an open neighborhood $\mathcal{O}_\alpha$ of $Z_\alpha(U_\alpha \cap U_\beta)$, one, $\mathcal{O}_\beta$, of $Z_\beta(U_\alpha \cap U_\beta)$, and a holomorphism of $\mathcal{O}_\alpha$ onto $\mathcal{O}_\beta$, F_β^α, such that

$$Z_\beta = F_\beta^\alpha \circ Z_\alpha \quad \text{in} \quad U_\alpha \cap U_\beta .$$

We have used the notation $Z_\alpha = (Z_\alpha^1, \ldots, Z_\alpha^m) : U_\alpha \to \mathbb{C}^m$ (and shall use an analogous notation throughout the sequel – when there is no subscript α). Henceforth we assume that Ω is equipped with a hypo-analytic structure and refer to it as a hypo-analytic manifold. The integer $n = \dim \Omega - m$ will be called the *codimension* of the structure, and the pairs (U_α, Z_α) of the above definition will be referred to (momentarily) as the "basic hypo-analytic local charts".

Definition 1.1. We say that a distribution f defined in an open neighborhood of a point p_0 of Ω is hypo-analytic at p_0 if there is a basic hypo-analytic local chart (U_α, Z_α) whose domain contains p_0 and a holomorphic function $\tilde{f}_\alpha$ in an open neighborhood of $Z_\alpha(p_0)$ in $\mathbb{C}^m$ such that $f = \tilde{f}_\alpha \circ Z_\alpha$ in a neighborhood of p_0.

It follows from the definition of a hypo-analytic structure that in Definition 1.1 the local chart (U_α, Z_α) may be replaced by any other basic hypo-analytic local chart, (U_β, Z_β), such that $p_0 \in U_\beta$. The concept of a hypo-analytic function in an open set is self-evident. Note that such a function is automatically C^∞ in the open set. Hypo-analytic functions in an open subset of Ω make up an algebra over the complex numbers; if f is an element of that algebra and does not vanish at any point, its reciprocal $1/f$ also belongs to the algebra.

If Ω' is another hypo-analytic manifold, a mapping $F : \Omega \to \Omega'$ is said to be hypo-analytic if, given any open subset U' of Ω' and any hypo-analytic function $f : U' \to \mathbb{C}$, the pull-back $f \circ F$ is a hypo-analytic function in $F^{-1}(U')$.

We come now to the notion of *hypo-analytic local chart*, which we shall often use in the sequel: by this we mean an $(m+1)$-tuple $(U, Z^1, \ldots, Z^m)$ [most of the time abbreviated to (U, Z)] consisting of an open subset U of Ω and of m hypo-analytic functions $Z^1, \ldots, Z^m$ whose differentials are linearly independent at every point of U. We emphasize that, generally speaking, the map $Z = (Z^1, \ldots, Z^m) : U \to \mathbb{C}^m$ is not a diffeomorphism, not even a local embedding.

Definition 1.2. Let u be a distribution in an open subset Ω' of Ω. The complement in Ω' of the open set of points at which u is hypo-analytic will be called the hypo-analytic singular support of u and denoted by $\operatorname{sing\,supp}_{ha} u$.

We introduce now the RC-structure associated to the hypo-analytic structure of Ω [7]. Let p be an arbitrary point of Ω. The differentials of the germs of hypo-analytic functions at p make up a complex vector subspace of the complex cotangent space $\mathbb{C}T_p^*\Omega$, subspace which we denote by T'_p. We have $\dim_{\mathbb{C}} T'_p = m$. By availing ourselves of hypo-analytic local charts we see that the subspaces T'_p make up a smooth vector subbundle T' of the complex cotangent bundle $\mathbb{C}T^*\Omega$. We shall refer to T' as the RC-structure bundle of the hypo-analytic manifold Ω. But note that T' might correspond to different hypo-analytic structures on Ω: take for instance the case where $n = 0$, i.e., $\dim \Omega = m$; then $T' = \mathbb{C}T^*\Omega$, yet the choice of the hypo-analytic structure, through the choice of the hypo-analytic local charts (U, Z), may vary greatly.

We shall follow the notation and terminology of [7]: $T'^{\perp}$ will denote the orthogonal of T' for the duality between tangent and cotangent vectors; $T'^{\perp}$ is a complex vector subbundle (of class C^∞) of the complex tangent bundle $\mathbb{C}T\Omega$; its fibre dimension is equal to n, the codimension of the hypo-analytic structure of Ω. If (U, Z) is a hypo-analytic local chart then any smooth section of $T'^{\perp}|_U$ is a smooth complex vector field L in U such that $LZ^k = 0$ for every $k = 1, \ldots, m$. This implies at once that $[T'^{\perp}, T'^{\perp}] \subset T'^{\perp}$, i.e., the commutation bracket of any two smooth sections of $T'^{\perp}$ over one and the same open subset Ω' of Ω is a smooth section of $T'^{\perp}$.

A distribution u in an open subset Ω' of Ω is said to be an *RC distribution* if $Lu = 0$ whatever the C^∞ section L of $T'^{\perp}$ over Ω'. This is equivalent to saying that du is a (distribution) section of T' over Ω'.

We denote by T^0 the intersection of T' with the *real* cotangent bundle of Ω, $T^*\Omega$. In general T^0 is *not* a vector bundle, yet it is an object of fundamental importance: if we identify vector fields to the functions they define on the cotangent bundle (i.e., if we identify vector fields to their symbols), then T^0 is exactly the set of common zeros of all the sections of $T'^{\perp}$, that is to say, in standard *PDE* terminology, the *characteristic set* of that family of vector fields.

We must now briefly recall what are the main classes of *RC* and of hypo-analytic structures. The *RC*-structure defined by T' on Ω is said to be *essentially real* if $T'=T^0\oplus_{\mathbb{R}}\mathbb{C}$ (T' is the complexification of T^0); *elliptic* if $T^0=0$;

it is said to define a CR *structure* on Ω if $\mathbb{C}T^*\Omega=T'+\bar{T}'$ ($+$: fibrewise vector sum);

it is said to define a *complex structure* if it is elliptic and, at the same time, defines a CR structure on Ω, in other words, if

$$\mathbb{C}T^*\Omega=T'\oplus\bar{T}' \quad (\oplus\text{: direct sum}).$$

When T' is essentially real, $T'^{\perp}$ can be locally generated by *real* vector fields; since $[T'^{\perp},T'^{\perp}]\subset T'^{\perp}\Omega$ is foliated by the integral manifolds of $T'^{\perp}$ (Frobenius theorem). Locally the generators of $T'^{\perp}$ can be taken to be the partial derivatives $\partial/\partial x^j$ ($j=1,\ldots,n$) with respect to part of the local coordinates (the $\partial/\partial x^j$, $j=1,\ldots,n$, span the tangent space to the leaves).

When T' defines a complex structure on Ω we must have $n=m$, and $\dim\Omega=2m$. To say that $(U,Z^1,\ldots,Z^m)$ is a hypo-analytic local chart is to say that $Z^1,\ldots,Z^m$ can be taken as complex coordinates in U – possibly after contracting U about one of its points. Over $UT'^{\perp}$ is generated by the *Cauchy-Riemann operators* $\partial/\partial\bar{Z}^k$ defined by those coordinates.

The elliptic case looks like a mixture of the preceding two, at least locally: we can find in any sufficiently small open set $Um+n$ coordinates $x^1,\ldots,x^m$, $y^1,\ldots,y^n$ such that $Z^j=x^j+\sqrt{-1}y^j$, $j=1,\ldots,m$, are hypo-analytic (note that, when T' is elliptic, we necessarily have $m\leqq n$). Then $T'^{\perp}$ is generated over U by the operators $\dfrac{\partial}{\partial\bar{Z}^j}$, $\dfrac{\partial}{\partial y^k}$ ($j=1,\ldots,m$, $k=m+1,\ldots,n$). The essentially real case obtains when $m=0$, the complex case when $n=m$.

If T' defines a CR structure on Ω we have $n\leqq m$. Moreover

$$F\dim_{\mathbb{C}}T'\cap\bar{T}'=F\dim_{\mathbb{C}}T'+F\dim_{\mathbb{C}}\bar{T}'-F\dim_{\mathbb{C}}(T'+\bar{T}')=2m-(m+n),$$

($F\dim_{\mathbb{C}}$; fibre dimension over the complex numbers), and thus $T'\cap\bar{T}'$ is a complex vector bundle of fibre dimension $m-n$, T^0 is a real vector bundle of fibre dimension (over $\mathbb{R}$!) $m-n$. Note that the condition defining CR structures can be rewritten, by duality,

$$T'^{\perp}\cap\bar{T}'^{\perp}=0.$$

(The condition defining ellipticity, on the other hand, can be rewritten

$$\mathbb{C}T\Omega=T'^{\perp}+\bar{T}'^{\perp}.$$

That defining complex structures can be rewritten

$$\mathbb{C}T\Omega=T'^{\perp}\oplus\bar{T}'^{\perp}.)$$

So far we have been reasoning within the C^∞ category, but the same concepts could have been introduced in the *analytic* (meaning real analytic) category. In that case Ω must be an analytic manifold and the structure bundle T' must be an analytic vector subbundle of $\mathbb{C}T^*\Omega$. An analytic hypo-analytic structure on Ω will be a hypo-analytic structure in which all hypo-analytic functions are real-analytic (though possibly complex-valued). In connection with this we point out that *to every analytic RC-structure on Ω there is associated a* unique *hypo-analytic structure.* This contrasts with the C^∞ case.

However the same phenomenon occurs for complex structures, for a hypo-analytic function in an open subset of a complex manifold is simply a holomorphic function in that subset.

In general CR manifolds are not hypo-analytic (they are not even locally integrable – meaning that locally the structure bundle T' may not be generated by exact differentials, see [6, 4]). But embedded CR manifolds are naturally equipped with a hypo-analytic structure. A CR manifold Ω is said to be *embedded* if Ω is a submanifold of a complex manifold $\tilde{\Omega}$ such that the pull-back to Ω of the holomorphic bundle (i.e., the structure bundle corresponding to the complex structure) of $\tilde{\Omega}$ over Ω is identical to the structure bundle of Ω (corresponding to its CR structure). A function f in an open subset of Ω containing a point p is hypo-analytic at p if near p it is the restriction of a function $\tilde{f}$ defined and holomorphic in an open neighborhood of p in $\tilde{\Omega}$.

Let (U, Z) be a hypo-analytic local chart in Ω. We select one point of U as the "center" of the chart and refer henceforth to it as the "origin"; we denote it accordingly by 0. The following operations are permitted: i) contract U about the origin; ii) perform a complex affine substitution of the Z^j's; iii) change local (real) coordinates. By means of such operations it is possible to obtain

$$\begin{aligned} Z^j &= x^j + \sqrt{-1}\,y^j = z^j, \qquad j = 1, \ldots, r; \\ Z^k &= x^k + \sqrt{-1}\,\Phi^k(x, y), \qquad k = r+1, \ldots, m, \end{aligned} \tag{1.1}$$

with

$$\Phi^k \textit{ real-valued}, \qquad \Phi^k(0,0) = 0, \qquad d\Phi^k(0,0) = 0, \qquad r+1 \leqq k \leqq m, \tag{1.2}$$

and the tacit understanding that the coordinates x^j, y^k, and therefore the functions Z^l all vanish at the origin. (See [7, Chap. I, Sect. 2].)

The integer r in (1.1) is entirely determined. Indeed it is readily seen that

$$\dim T_0^\circ = m - r.$$

Note that

$$r \leqq \inf(m, n).$$

Let us point out the following facts:
When T' is essentially real, $r = 0$;
when T' is elliptic, $r = m$;
when T' is a CR structure, $r = n$;
when T' is a complex structure, $r = m = n$.

In the local chart (U, Z) there is a "natural" basis of T': it will consist of the vector fields L_j $(j=1,\ldots,n)$ defined by the requirements:

$$\begin{aligned} &L_j Z^k = 0, \quad k=1,\ldots,m; \\ &L_j \bar{z}^i = \delta_j^i, \quad i=1,\ldots,r; \quad L_j y^{j'} = \delta_j^{j'}, \quad j'=r+1,\ldots,n. \end{aligned} \tag{1.3}$$

The expression of L_j is

$$\begin{aligned} L_j &= \frac{\partial}{\partial \bar{z}^j} + \sum_{k=r+1}^{m} \lambda_j^k(x,y) \frac{\partial}{\partial x^k}, \quad j=1,\ldots,r; \\ L_j &= \frac{\partial}{\partial y^j} + \sum_{k=r+1}^{m} \lambda_j^k(x,y) \frac{\partial}{\partial x^k}, \quad j=r+1,\ldots,m. \end{aligned} \tag{1.4}$$

The computation of the coefficients λ_j^k is easy [7, Chap. I, Sect. 2].

We shall need a lemma about two different hypo-analytic structures, $\mathscr{A}$ and $\mathscr{A}'$, on one and the same manifold Ω. We say that $\mathscr{A}'$ is *finer* than $\mathscr{A}$ if any hypo-analytic function at a point p for $\mathscr{A}$ is hypo-analytic, at the same point, for $\mathscr{A}'$ (thus $\mathscr{A}'$ admits *more* hypo-analytic functions than $\mathscr{A}$). Call T' (resp., T'') the structure bundle defined by $\mathscr{A}$ (resp., $\mathscr{A}'$). If $\mathscr{A}'$ is finer than $\mathscr{A}$ T' is a complex vector subbundle of T''.

Lemma 1.1. *Let $\mathscr{A}$ and $\mathscr{A}'$ be two-hypo-analytic structures on Ω, with $\mathscr{A}'$ finer than $\mathscr{A}$ as just described. Any RC-distribution u in the sense of T' is an RC-distribution in the sense of T''. If such a distribution is hypo-analytic at a point p for $\mathscr{A}'$ it is also hypo-analytic at p for $\mathscr{A}$.*

Proof. The first part of the conclusion is obvious. In order to prove the last part we reason in a hypo-analytic local chart (U, Z) for $\mathscr{A}$. Possibly after contracting U we can find s hypo-analytic functions $Z^{m+1},\ldots,Z^{m+s}$ for $\mathscr{A}'$ such that $(U, Z^1,\ldots,Z^m, Z^{m+1},\ldots,Z^{m+s})$ is a hypo-analytic local chart for $\mathscr{A}'$ (as usual we call m the fibre dimension of T' over $\mathbb{C}$; that of T'' is denoted by $m+s$). Set

$$\hat{Z} = (Z^1,\ldots,Z^{m+s}) : U \to \mathbb{C}^{m+s}.$$

Let u be a hypo-analytic function in U for $\mathscr{A}'$. There is a holomorphic function $\tilde{u}$ in an open neighborhood of $\hat{Z}(U)$ in $\mathbb{C}^{m+s}$ such that $u = \tilde{u} \circ \hat{Z}$. If u is an RC-distribution for T' we have $du = \sum_{j=1}^{m} \frac{d\tilde{u}}{\partial z^j} \circ Z \, dZ^j$. This means that the partial derivatives $\partial\tilde{u}/\partial z^k$ $(k=m+1,\ldots,m+s)$ vanish identically on $\hat{Z}(U)$. But $\hat{Z}(U)$ contains a totally real submanifold of $\mathbb{C}^{m+s}$ of real dimensions $m+s$. Therefore those partial derivatives [which are holomorphic functions in an open neighborhood of $\hat{Z}(U)$] must vanish identically in a possibly smaller neighborhood of $\hat{Z}(U)$, in other words, $\tilde{u}$ must be independent of $Z^{m+1},\ldots,Z^{m+s}$ there, which is exactly what we wanted. □

2. Elliptic and Complex Submanifolds

Let, as before, Ω be a hypo-analytic manifold of dimension $m+n$. By a *submanifold* of Ω we mean a subset of Ω (not necessarily closed) endowed with a C^∞ structure

such that the natural injection into Ω is a C^∞ map with injective differential.

Let $\mathscr{M}$ be a submanifold of Ω; we shall denote by $\pi_{\mathscr{M}}$ the natural map $T^*\Omega|_{\mathscr{M}} \to T^*\mathscr{M}$ and by $\pi_{\mathscr{M}}^{\mathbb{C}}$ the analogous map of the complex cotangent bundles. Both maps are surjections. We shall write

$$T'_{\mathscr{M}} = \pi_{\mathscr{M}}^{\mathbb{C}}(T').$$

In general $T'_{\mathscr{M}}$ is *not* a vector bundle.

Definition 2.1. We say that a submanifold $\mathscr{M}$ of Ω is a hypo-analytic submanifold if $\mathscr{M}$ is equipped with a hypo-analytic structure whose structure bundle is identical to $T'_{\mathscr{M}}$ and which has the following property:

Given any hypo-analytic function f in an open set $\Omega' \subset \Omega$ which intersects $\mathscr{M}$, the restriction of f to $\mathscr{M} \cap \Omega'$ is hypo-analytic.

If $\mathscr{M}$ is a hypo-analytic submanifold of Ω we call it elliptic, CR, complex, essentially real if those adjectives apply to the hypo-analytic structure given on $\mathscr{M}$.

We shall now make the hypothesis that the submanifold $\mathscr{M}$, whose dimension we denote by v, is *elliptic*. Let m' be the fibre dimension (over $\mathbb{C}$) of the structure bundle $T'_{\mathscr{M}}$. We recall that $2m' \leqq v$.

Consider a local chart (U, Z) such that (1.1) and (1.2) hold, centered at a point ("the origin") of the submanifold $\mathscr{M}$. There must be exactly m' functions Z^j whose differentials, after pull-back to $U \cap \mathscr{M}$, span $T'_{\mathscr{M}}$ at every point. But if we call these $Z^{j_1}, \ldots, Z^{j_{m'}}$ and recall that $T'_{\mathscr{M}}$ is elliptic, we see that the pull-backs to $\mathscr{M} \cap U$ of

$$dZ^{j_1}, \ldots, dZ^{j_{m'}}, d\bar{Z}^{j_1}, \ldots, d\bar{Z}^{j_{m'}}$$

must be linearly independent. We must therefore have $j_\alpha \leqq r$ for all $\alpha = 1, \ldots, m'$. After re-ordering the Z^k we may assume that $j_1 = 1, \ldots, j_{m'} = m'$ (in particular, we must have $m' \leqq r$).

But this is not all: the restrictions to $\mathscr{M} \cap U$ of the Z^k for $k > m'$ must equal holomorphic functions of $Z' = (Z^1, \ldots, Z^{m'})$. In other words, there are such holomorphic functions, h^k, defined in an open neighborhood of the origin, such that

$$Z^k - h^k(Z') \equiv 0 \quad \text{on} \quad \mathscr{M} \cap U \quad (k = m'+1, \ldots, m).$$

Let us then set, for those same k,

$$Z^k_\# = Z^k - h^k(Z'),$$
$$x^k_\# = \operatorname{Re} Z^k_\#.$$

For $j = m'+1, \ldots, r$, we set $y^j_\# = \operatorname{Im} Z^j_\#$, and $x^i_\# = x^i$ for $i = 1, \ldots, m'$, $y^l_\# = y^l$ for $l = 1, \ldots, m'$ and for $l = r+1, \ldots, n$. It is then clear that $(x, y) \mapsto (x_\#, y_\#)$ is a *bona fide* change of variables. After deleting the subscripts $\#$ we may now assume, not only that (1.1) and (1.2) hold, but also that, on $\mathscr{M}$, we have

$$x^k = 0, \quad k = m'+1, \ldots, m, \quad y^l = 0, \quad l = m'+1, \ldots, r, \tag{2.1}$$

as well as

$$\Phi^k(x, y) = 0, \quad k = r+1, \ldots, m. \tag{2.2}$$

Note that (2.1) implies

$$m+r-2m' \leqq \operatorname{codim} \mathcal{M} (= m+n-v). \tag{2.3}$$

We now apply the implicit function theorem: there is an open neighborhood of the origin in $\mathcal{M} \cap U$ which is a *locally closed* submanifold, S, of Ω. This means the following: possibly after contraction of U about the origin we may assume that there are $m+n-v$ C^∞, real-valued functions f_j such that S is exactly defined in U by the equations

$$f_j(x,y)=0, \quad j=1,\ldots,m+n-v,$$

and that the differentials df_j are linearly independent. It is clear that we may take $m+r-2m'$ of these equations to be the Eq. (2.1), and assume that the remaining ones have the form

$$y^j = y^j(x,y'), \quad j=r+1,\ldots,n-v+2m',$$

and after a further change of coordinates, we may take S to be defined in U by the equations

$$x^k=0, \quad k=m'+1,\ldots,m, \quad y^l=0, \quad l=m'+1,\ldots,n-v+2m'. \tag{2.4}$$

Since $\Phi^k \equiv 0$ on S for $k=r+1,\ldots,m$, we also have

$$\Phi^k(x,y) = \sum_{i=m'+1}^{m} a_i^k(x,y)x^i + \sum_{j=m'+1}^{n-v+2m'} b_j^k(x,y)y^j. \tag{2.5}$$

It is convenient to use the following notation

$$x'=(x^1,\ldots,x^{m'}), \quad y'=(y^1,\ldots,y^{m'}), \quad z'=x'+\sqrt{-1}y',$$
$$x''=(x^{m'+1},\ldots,x^m), \quad y''=(y^{m'+1},\ldots,y^{n-v+2m'}),$$
$$y^*=(y^{n-v+2m'+1},\ldots,y^n),$$

and to take the neighborhood U in the form of a product

$$U=\Delta' \times V'' \times W'' \times W^*, \tag{2.6}$$

where $\Delta'=\{z\in\mathbb{C}^{m'}; |z^j|<\varrho, j=1,\ldots,m'\}$, and where V'', W'', W^* are open balls centered at the origin in the spaces of the variables x'', y'', y^* respectively. In this notation the equations which define S in U read

$$x''=0, \quad y''=0, \tag{2.7}$$

and (2.5) reads

$$\Phi^\# = Ax''+By'', \tag{2.8}$$

where $\Phi^\#=(\Phi^{r+1},\ldots,\Phi^m)$, and A is the matrix with generic entry a_i^k, B the matrix with generic entry b_j^k (note that in general these matrices are not square). By virtue of (1.2) we have $A(0,0)=0$, $B(0,0)=0$.

Suppose now that $\mathcal{M}$ is a *complex* submanifold. In this case we must have

$$v=2m'.$$

Then S is defined by the Eqs. (2.7) but y'' means here $(y^{m'+1}, \ldots, y^n)$: there are no coordinates y^*. We have

$$S = \Delta' \times \{0\} \times \{0\}. \tag{2.9}$$

Suppose furthermore that T' defines a CR structure on Ω, which means here that $n = r$. In this case the Eqs. (2.7) completely reduce to the Eqs. (2.1). Since the set defined by (2.1) contains $\mathcal{M} \cap U$ we must have $S = \mathcal{M} \cap U$:

$$\mathcal{M} \cap U = \Delta' \times \{0\} \times \{0\}. \tag{2.10}$$

The reader should not entertain the notion that through each point of Ω passes a *unique* elliptic submanifold which is *maximal* in a reasonable sense (say connected and not contained in any strictly larger connected elliptic submanifold). Counter-examples are obtainable within the class of embedded (or embeddable) CR manifolds:

Example 2.1. Define a hypo-analytic structure on $\mathbb{R}^5$ by means of the following three functions

$$Z^j = x^j + \sqrt{-1}y^j \quad (j = 1, 2), \qquad Z^3 = x^3 + iy^1y^2.$$

Let $\mathcal{M}$ be an elliptic, which necessarily means complex since we are dealing with a CR structure, submanifold passing through the origin. Suppose first that $\dim_{\mathbb{C}} \mathcal{M} = 2$. Then the pull-backs to $\mathcal{M}$ of dZ^1, dZ^2 must span the holomorphic bundle $T'_{\mathcal{M}}$ of $\mathcal{M}$ at every point of $\mathcal{M}$; in other words we may use $z^j = x^j + \sqrt{-1}y^j$ as complex coordinates on $\mathcal{M}$ (indeed, dZ^3 and $d\bar{Z}^3$ are equal at the origin). On $\mathcal{M}$ we must have $Z^3 = h(z^1, z^2)$ with h holomorphic. But then $\operatorname{Im} h(z^1, z^2) = y^1y^2$ which demands $h(z^1, z^2) = z^1y^2 + g(z^2)$: h cannot be a holomorphic function of (z^1, z^2). Thus $\dim_{\mathbb{C}} \mathcal{M} \leqq 1$. Suppose $\dim_{\mathbb{C}} \mathcal{M} = 1$. Then either dZ^1 or dZ^2 must span $T'_{\mathcal{M}}$ in some open neighborhood of the origin; suppose dZ^1 does, and use z^1 as the complex coordinate there. We must have, in that same neighborhood of 0 in $\mathcal{M}$, $Z^2 = h(z^1)$, $Z^3 = k(z^1)$, with h, k holomorphic. But this demands

$$y^1 \operatorname{Im} h(z^1) = \operatorname{Im} k(z^1),$$

which is possible only if $\operatorname{Im} h = \operatorname{Im} k = 0$ everywhere. In other words we must have $h = \text{const.}$, $k = \text{const.}$, which is equivalent to $h = k = 0$ everywhere since Z^2 and Z^3 vanish at 0. By exchanging Z^1 and Z^2 we reach the conclusion that there are two, and only two, complex submanifolds of complex dimension one passing through the origin, namely

$$\mathcal{M}^1 = \{(x, y) \in \mathbb{R}^5 ; x^2 = x^3 = y^2 = 0\},$$
$$\mathcal{M}^2 = \{(x, y) \in \mathbb{R}^5 ; x^1 = x^3 = y^1 = 0\}. \quad \square$$

However the following can be said:

Proposition 2.1. *Let Ω be a hypo-analytic* CR *manifold of dimension $m+n$, T' its structure bundle (whose fibre dimension over $\mathbb{C}$ is equal to m).*

Then, through any point of Ω, there passes at most one maximal complex submanifold of complex dimension n.

Proof. Let us reason within a hypo-analytic local chart (U, Z) centered at a point of Ω which we call the origin. Assume furthermore that (1.1) and (1.2) hold. Since Ω is a CR manifold we have $r=n$. Suppose that there passes through 0 a complex submanifold $\mathscr{M}$ of complex dimension n. Then, because $\mathbb{C}T^*\mathscr{M}=T'_{\mathscr{M}}\otimes\bar{T}'_{\mathscr{M}}$, the pull-backs to $\mathscr{M}\cap U$ of $dZ^1,\ldots,dZ^n$, $d\bar{Z}^1,\ldots,d\bar{Z}^n$ form a basis of $\mathbb{C}T^*\mathscr{M}$ over $\mathscr{M}\cap U$. We may use $z^1,\ldots,z^n$ as complex coordinates in $\mathscr{M}\cap U$. Definition 2.1 demands that the restriction to $\mathscr{M}\cap U$ of Z^k $(k=n+1,\ldots,m)$ be a holomorphic function of $z'=(z^1,\ldots,z^n)$, $H^k(z')$. This provides us with equations defining $U\cap\mathscr{M}$ in U:

$$Z^k=H^k(z'),\qquad k=n+1,\ldots,m. \tag{2.11}$$

Write $u=\mathrm{Re}(H^{n+1},\ldots,H^m)$, $\Phi''=(\Phi^{n+1},\ldots,\Phi^m)$. The Eqs. (2.11) can be rewritten

$$x''=u,\qquad \Phi''=\mathrm{Im}(H^{n+1},\ldots,H^m),$$

which is to say

$$\bar{\partial}_{z'}[u(z')+i\Phi''(x',u(z'),y)]=0. \tag{2.12}$$

If we equate (2.12) to $2(m-n)$ real equations, in the customary manner of the Cauchy-Riemann equations, we obtain a system of first-order nonlinear *PDE* in the unknown (vector-valued) function u:

$$du=F(d\Phi''(x',u,y)), \tag{2.13}$$

with F nonlinear (but C^∞ and real). In order that (2.13) have a solution it is necessary and sufficient that the exterior derivative of the right-hand side in (2.13) vanish, which requires that Φ'' satisfy a certain elliptic system of nonlinear *PDE*. *[When it does there will pass through the origin a maximal complex submanifold of complex dimension n.]* The solution of (2.13) is elementary: use integration along rays in z'-space issuing from the origin and the Picard iteration method. The latter also shows that the solution is unique. □

3. Propagation of Hypo-Analyticity Along Elliptic Submanifolds

Let Ω be a hypo-analytic manifold, as before. We recall that an *RC*-distribution is a distribution u whose differential du is a section of the structure bundle T' over the domain of definition of u.

Theorem 3.1. *If a connected elliptic submanifold $\mathscr{M}$ of Ω intersects the hypo-analytic singular support of an RC-distribution u then $\mathscr{M}$ is entirely contained in it.*

Proof. We shall prove that if $\mathscr{M}$ intersects $\Omega\backslash\mathrm{sing\,supp}_{ha}u$, $\mathscr{M}$ is entirely contained in that complement. We are going to reason within a hypo-analytic local chart (U, Z) centered at a point of $\mathscr{M}$ henceforth called the origin. We choose the local chart (U, Z) exactly as explained in Sect. 2, and we are going to prove that the "component" S of $\mathscr{M}\cap U$ $(0\in S)$ does not intersect $\mathrm{sing\,supp}_{ha}u$ if the origin does not belong to the latter. The argument will yield what we want because the "size" of

$$S=\Delta'\times\{0\}\times\{0\}\times W^* \quad \text{[cf. (2.6), (2.7)]}$$

will not depend on the distance from the origin to $\mathrm{sing\,supp}_{ha}u$.

We shall deal, in succession, with the following cases (of increasing generality):

Case I. (Ω, T') is a CR manifold and $\mathcal{M}$ is a complex submanifold of (maximum) complex dimension n;

Case II. (Ω, T') is a CR manifold, and $\mathcal{M}$ is a complex submanifold of dimension $v<n$ (but $\geqq 1$);

Case III. The general case: Ω is simply a hypo-analytic manifold and $\mathcal{M}$ is an elliptic submanifold.

Case I

In this case $m'=\dim_{\mathbb{C}}\mathcal{M}=n$, and therefore there are no coordinates $y'': y=y'$, and

$$U=\Delta'\times V'', \qquad \mathcal{M}\cap U=\Delta'\times\{0\} \quad \text{[cf. (2.6), (2.10)]}.$$

Also note that (2.8) can be rewritten

$$\Phi''(x,y)=A(x,y)x'',$$

where $\Phi''=(\Phi^{n+1},\dots,\Phi^m)$. Keep in mind that $A(0,0)=0$.

We shall now impose certain conditions on the size of U but these will never involve the RC-distribution u. Our first requirement will be that all the derivatives of Φ^k $(k=n+1,\dots,m)$ be bounded.

We shall further contract U to make sure that

$$|A(x',x_1'',y)x_1''-A(x',x_2'',y)x_2''|\leqq|x_1''-x_2''|/100$$
$$\text{for all} \quad x'+iy\in\Delta', \qquad x_j''\in V'' \quad (j=1,2). \tag{3.1}$$

We also require

$$\|A(x,y)\|\leqq 1/100 \quad \text{for all} \quad (x,y)\in U, \tag{3.2}$$

where $\|\ \|$ is the matrix norm.

To say that u is an RC-distribution in U (which is the same as saying that u is a CR distribution) in U, is to say that

$$\frac{\partial u}{\partial\bar z^j}=-\sum_{k=n+1}^{m}\lambda_j^k(x,y)\frac{\partial u}{\partial x^k}, \qquad j=1,\dots,n \quad \text{[cf. (1.4)]}, \tag{3.3}$$

which implies that u is a C^∞ function of (x',y) in Δ' valued in the space of distributions in V'', $\mathcal{D}'(V'')$ [7, Chap. I, Proposition 5.2]. To say that u is hypo-analytic at the origin means that

> *the push-forward of u under the mapping Z, $\tilde u$, extends as a holomorphic function in an open neighborhood of the origin in $\mathbb{C}^m$.* (3.4)

We write, as before, $z'=x'+iy$. We also write $z''=(z^{n+1},\dots,z^m)\in\mathbb{C}^{m-n}$, and $z=(z',z'')$, $\theta=(\theta_1,\dots,\theta_{m-n})\in\mathbb{R}_{m-n}$, $\theta\cdot z''=\theta_1 z^{n+1}+\dots+\theta_{m-n}z^m$, $[z'']^2=(z^{n+1})^2+\dots+(z^m)^2$.

We shall prove the following:

Let $\Delta_0'=\{z'\in\mathbb{C}^n; |z^j|<\varrho_0, j=1,\dots,n\}$ be any open polydisk contained in Δ' such that $\tilde u$ extends as a holomorphic function in some open neighborhood of $\Delta_0'\times\{0\}$

in $\mathbb{C}^m$. If $\varrho_0<\varrho$ then there is an open polydisk Δ'_1 contained in Δ' and containing the closure of Δ'_0 that has the same property. This of course means that the property holds for Δ'.

Let ϱ_* be a positive number such that

$$(\varrho_0/\varrho)^{1/2}<\varrho_*/\varrho_0<1. \tag{3.5}$$

There exists a number $\delta>0$ such that $\tilde{u}$ extends as a holomorphic function in the open set

$$\{z\in\mathbb{C}^m; z'=x'+iy\in\Delta'_* \text{ and } \forall x''\in V'',\ |x''|<\delta,\ |z''-Z''(x',x'',y)|<\delta\}. \tag{3.6}$$

Here and throughout the sequel Δ'_* stands for the polydisk $|z^j|<\varrho_*$, $j=1,\ldots,n$.

We introduce then a positive number κ such that

$$\kappa>\mathrm{Max}(\delta^{-1}, 1/\mathrm{diam}\, V''). \tag{3.7}$$

It is convenient to write $\mathrm{diam}\, V''=4d$. We shall make use of a cut-off function $g\in C_c^\infty(V'')$ equal to one in the ball $|x''|\leqq d$, and deal with the integral [1]

$$F^\kappa(z,\theta)=\int e^{-i\theta\cdot Z''(x,y)-\kappa|\theta|[z''-Z''(x,y)]^2}g(x'')u(x,y)dZ''(x,y),$$

where the integration is performed with respect to x'' (over V''); $z'\in\Delta'$, $z''\in\mathbb{C}^{m-n}$, $\theta\in\mathbb{R}_{m-n}$ are parameters; $F^\kappa(z,\theta)$ is a C^∞ function of (z,θ), entire holomorphic with respect to z'', in the product set

$$\Delta'\times\mathbb{C}^{m-n}\times(\mathbb{R}_{m-n}\setminus\{0\}).$$

For $z'\in\Delta'$ call $\Sigma(z')$ the image of V'' under the map $x''\mapsto Z''(x',x'',y)$. Then

$$F^\kappa(z,\theta)=\int\limits_{\Sigma(z')} e^{-i\theta\cdot w-\kappa|\theta|[w-z'']^2}g(\mathrm{Re}\, w)\tilde{u}(z',w)dw. \tag{3.8}$$

We seek upper bounds on the absolute values of F^κ and of $\bar{\partial}_{z''}F^\kappa$. In order to obtain them it is essential to estimate from below the quantity

$$Q=\mathrm{Re}\{i\dot{\theta}\cdot w+\kappa[w-z'']^2\}\quad(\dot{\theta}=\theta/|\theta|)$$

when z'' is close to 0 and w either lies on $\Sigma(z')$ or is close to it. As a rule we begin by establishing a lower bound for Q when $z''=0$, in which case we call it Q_0, and then derive a weaker lower bound for Q when z'' is allowed to vary (but stays close to 0), by a continuity argument.

1. Estimate of F^κ when $z'\in\Delta'$. Let (x,y) vary arbitrarily in U and put $w=Z''(x,y)$. We have

$$Q_0\geqq-|A(x,y)x''|+\kappa(|x''|^2-|A(x,y)x''|^2),$$

and therefore, by (3.2),

$$Q_0\geqq|x''|\left(\frac{\kappa}{2}|x''|-\frac{1}{100}\right). \tag{3.9}$$

The minimum of the right-hand side in (3.9) is equal to $-1/2.10^4\kappa$. Therefore, if the variation of z'' is restricted to a sufficiently small open neighborhood $\mathcal{O}''$ of 0 in

$\mathbb{C}^{m-n}$, we shall have

$$Q \geqq -3/4.10^4\kappa .$$

Recalling that u is a distribution we may integrate by parts as many times as we need to, as shown in [1, Chap. I, Sect. 4]. We obtain easily

$$\begin{aligned} &|F^\kappa(z,\theta)| \leqq Ce^{|\theta|/R\kappa}, \\ &\forall z' \in \Delta', \quad z'' \in \mathcal{O}'', \quad \theta \in \mathbb{R}_{m-n}, \end{aligned} \tag{3.10}$$

where C is a positive constant independent of θ and $R = 10^4$.

2. Estimate of F^κ when $z' \in \Delta'_*$. Here we shall avail ourselves of the property that $\tilde{u}$ extends holomorphically to the set (3.6). We keep z' fixed in Δ'_* and select $g_0 \in C_c^\infty(V'')$, $0 \leqq g_0 \leqq 1$ everywhere, such furthermore that

$$g_0(x'') = 1 \quad \text{if} \quad |x''| \leqq \delta/4, \qquad g_0(x'') = 0 \quad \text{if} \quad |x''| \geqq \delta/2 .$$

In the integral at the right, in (3.8), we deform the chain of integration from $\Sigma(z')$ to the image of V'' under the map

$$x'' \mapsto w = Z''(x', x'', y) - \frac{i}{4\kappa} g_0(x'')\dot{\theta} . \tag{3.11}$$

This is permitted since $(4\kappa)^{-1} < \delta/4$ by (3.7). On the new chain of integration we have

$$\begin{aligned} Q_0 &\geqq -|A(x,y)x''| + g_0(x'')/4\kappa + \kappa\left(|x''|^2 - |A(x,y)x'' - \frac{1}{4\kappa} g_0(x'')\dot{\theta}|^2\right) \\ &\geqq -|A(x,y)x''| + g_0(x'')/4\kappa + \kappa(|x''|^2 - 2|A(x,y)x''|^2 - g_0(x'')^2/8\kappa^2) \\ &\geqq g_0(x'')/8\kappa + \frac{\kappa}{2}|x''|^2 - |x''|/100 . \end{aligned}$$

If $|x''| \leqq (4\kappa)^{-1} \leqq \delta/4$ then $Q_0 \geqq \kappa^{-1}(1/8 - 1/100)$; if $|x''| \geqq 1/4\kappa$, then $Q_0 \geqq \kappa^{-1} \cdot (1/32 - 1/100)$. In any case, $Q_0 \geqq 1/68\kappa$ and the same reasoning that led to (3.10) shows that, possibly after contracting $\mathcal{O}''$ about the origin, we have, for some $C > 0$,

$$\begin{aligned} &|F^\kappa(z,\theta)| \leqq Ce^{-|\theta|/100\kappa}, \\ &\forall z' \in \Delta'_*, \quad z'' \in \mathcal{O}'', \quad \theta \in \mathbb{R}_{m-n}. \end{aligned} \tag{3.12}$$

3. Estimate of $\bar{\partial}_{z'} F^\kappa$ when $z' \in \Delta'$. We now look at

$$\frac{\partial F^\kappa}{\partial \bar{z}^j}(z,\theta) = \int \frac{\partial}{\partial \bar{z}^j}\left\{e^{-i\theta \cdot Z''(x,y) - \kappa|\theta|[z'' - Z''(x,y)]^2} u(x,y) \det \frac{\partial Z''}{\partial x''}(x,y)\right\} g(x'')dx''$$

for $j = 1, \ldots, n$. We observe that, because of (1.1),

$$\det \frac{\partial Z''}{\partial x''} = \det \frac{\partial Z}{\partial x},$$

and recall that, if h is a solution of $L_j h=0$, $j=1,\dots,n$, then

$$ {}^tL_j\left(h\det\frac{\partial Z}{\partial x}\right)=0, \quad j=1,\dots,n, $$

where ${}^tL_j$ is the transpose of L_j [7, Chap. II, Lemma 4.1]. This means that

$$ \frac{\partial}{\partial\bar z^j}\left(h\det\frac{\partial Z}{\partial x}\right)=-\sum_{k=n+1}^{m}\frac{\partial}{\partial x^k}\left(\lambda_k^j h\det\frac{\partial Z}{\partial x}\right), \quad j=1,\dots,n. $$

We put this into the preceding integral. After integration by parts we obtain

$$ \frac{\partial F^\kappa}{\partial\bar z^j}(z,\theta)=\int e^{-i\theta\cdot Z''(x,y)-\kappa|\theta|[z''-Z''(x,y)]^2} \cdot\left(\sum_{k=n+1}^{m}\lambda_k^j(x,y)\frac{\partial g}{\partial x^k}(x'')\right)u(x,y)dZ''(x,y). $$

We return to (3.9) but now we restrict the variation of x'' to the support of dg; thus $|x''|\geqq d$, whence

$$ Q_0\geqq d(\kappa d-1/50)/2\geqq d/10\geqq 1/40\kappa $$

since $4\kappa d>1$ by (3.7). In the same manner as in the proofs of (3.10) and (3.12) we reach the conclusion that, if $\mathcal{O}''$ is small enough, there is a constant $C>0$ such that

$$ |\bar\partial_{z'}F^\kappa(z,\theta)|\leqq Ce^{-|\theta|/100\kappa}, \quad \forall z'\in\Delta', \quad z''\in\mathcal{O}'', \quad \theta\in\mathbb{R}_{m-n}. \tag{3.13} $$

A remark that is used below is that estimates analogous to (3.10), (3.12), and (3.13) are valid if we replace F^κ by any one of its partial derivatives (of any order) with respect to x', y, z''. Indeed, if we differentiate F^κ under the integral sign we obtain integrals of the same kind, except that powers of θ and of $|\theta|$ might appear as factors of the exponential. But such powers have no effect on the estimates in question.

We derive now the consequences of those estimates. First of all, we note that it is elementary (see [5, 2]) to solve the inhomogeneous Cauchy-Riemann equations in the polydisk Δ' with a good estimate on the solution. Thus there is a C^∞ function $v^\kappa(z,\theta)$ in $\Delta'\times\mathcal{O}''\times(\mathbb{R}_{m-n}\backslash\{0\})$, holomorphic with respect to z'', satisfying

$$ \frac{\partial}{\partial\bar z^j}(F^\kappa-v^\kappa)=0, \quad j=1,\dots,n, \tag{3.14} $$

$$ |v^\kappa(z,\theta)|\leqq C'e^{-|\theta|/100\kappa}. \tag{3.15} $$

Set then $H^\kappa=F^\kappa-v^\kappa$; H^κ is C^∞ with respect to (z,θ) in $\Delta'\times\mathcal{O}''\times(\mathbb{R}_{m-n}\backslash\{0\})$, and is holomorphic with respect to z.

Thanks to (3.15) the analogues of (3.10) and (3.12) hold for H^κ:

$$ |H^\kappa(z,\theta)|\leqq Ce^{|\theta|/R\kappa}, \quad z'\in\Delta'; \tag{3.16} $$

$$ |H^\kappa(z,\theta)|\leqq Ce^{-|\theta|/100\kappa}, \quad z'\in\Delta'_*, \tag{3.17} $$

whatever $z''\in\mathcal{O}''$, $\theta\in\mathbb{R}_{m-n}$ (of course the constant C may have to be increased. We recall that $R=10^4$. Let us then set

$$q=(1-1/100)/(1+1/100), \tag{3.18}$$

and note that

$$\frac{1-q}{100}-\frac{q}{R}=\frac{1}{R}.$$

We define

$$\varrho_1=\varrho^q\varrho_*^{1-q}. \tag{3.19}$$

We apply Hadamard's three circles theorem, and reach the conclusion that, for some $C_1>0$,

$$|H^k(z,\theta)|\leqq C_1e^{-|\theta|/R\kappa},\quad z'\in\Delta_1',\quad z''\in\mathcal{O}'',\quad \theta\in\mathbb{R}_{m-n}, \tag{3.20}$$

where $z'\in\Delta_1'$ means that $|z^j|<\varrho_1$ for every $j=1,\ldots,n$.

We observe that, according to (3.5) and (3.19),

$$\varrho_1/\varrho_0=(\varrho/\varrho_*)^q\varrho_*/\varrho_0>(\varrho/\varrho_0)^{q-1/2}.$$

Since $q>1/2$ we obtain $\varrho_1>\varrho_0$ as announced. Let us summarize what we have found so far:

> *There is an open polydisk Δ_1' contained in Δ' and containing the closure of Δ'^0, an open neighborhood of the origin in $\mathbb{C}^{m-n}$, $\mathcal{O}''$, and a constant $C>0$ such that*
>
> $$|F^\kappa(z,\theta)|\leqq Ce^{-|\theta|/R\kappa},\quad \forall z'\in\Delta_1',\quad z''\in\mathcal{O}'',\quad \theta\in\mathbb{R}_{m-n}.$$
>
> *Similar estimates hold for all derivatives of $F^\kappa(z,\theta)$ with respect to x',y,z''.* (3.21)

By availing ourselves of (3.21) we are now going to construct a holomorphic extension of $\tilde{u}$ in an open neighborhood of $\Delta_1'\times\{0\}$ in $\mathbb{C}^{m-n}$.

It is convenient to extend $\Phi''(x,y)$ to all values of x'' in $\mathbb{R}^{m-n}$. Actually one may assume that this has been done at the start, that the matrix $A(x,y)$ has been multiplied by a cut-off function $g_1\in C_c^\infty(V'')$, $0\leqq g_1\leqq 1$ everywhere, $g_1\equiv 1$ in a neighborhood of $\operatorname{supp}g$, and that we have set $A\equiv 0$ in $\mathbb{R}^{m-n}\setminus V''$ – in such a way that (3.1) and (3.2) hold for all x''_j in $\mathbb{R}^{m-n}$ $(j=1,2)$. Calling $\mathbf{1}$ the $(m-n)\times(m-n)$ identity matrix we have

$$Z''(x,y)=(\mathbf{1}+\sqrt{-1}A(x,y))x''. \tag{3.22}$$

It follows that we can find $d_1>0$ such that $|x''|<d_1$ implies $Z''(x,y)\in\mathcal{O}''$ whatever $z'\in\Delta'$. Define then $I^\varepsilon(z)$ to be the integral

$$\begin{aligned}(\kappa/4\pi^3)^{\frac{1}{2}(m-n)}\int & e^{i\theta\cdot[z''-Z''(x,y)]-\kappa|\theta|[Z''(x',t,y)-Z''(x,y)]^2-\varepsilon|\theta|^2}\\ &\cdot g(x'')u(x,y)|\theta|^{\frac{1}{2}(m-n)}dZ''(x,y)dZ''(x',t,y)d\theta\\ =(\kappa/4\pi^3)^{\frac{1}{2}(m-n)}&\int e^{i\theta\cdot z''-\varepsilon|\theta|^2}F^\kappa(z',Z''(x',t,y),\theta)\\ &\cdot|\theta|^{\frac{1}{2}(m-n)}dZ''(x',t,y)d\theta,\end{aligned}$$

in which the integration with respect to t is performed over the ball $|t| \leqq d_1$. Define $J^\varepsilon(z)$ to be the same integral but where the t-integration is performed over the set $|t| > d_1$. In both integrals the integration with respect to (x'', θ) is performed over $V'' \times \mathbb{R}_{m-n}$.

In the sum $I^\varepsilon + J^\varepsilon$ we may carry out the integration with respect to t:

$$I^\varepsilon(z) + J^\varepsilon(z) = (2\pi)^{n-m} \int e^{i\theta \cdot [z'' - Z''(x', s, y)] - \varepsilon|\theta|^2} u(x', s, y) g(s) dZ''(x', s, y) d\theta .$$

We have preferred to write s for the variable of integration denoted by x'' above, to avoid confusion in the next formula, which is a direct consequence of the preceding one [1, Chap. 1, Corollaries 4.3, 4.4, and 4.5]:

$$g(x'')u(x, y) = \lim_{\varepsilon \to +0} (I^\varepsilon + J^\varepsilon)(Z(x, y)), \tag{3.23}$$

where the limit is valid in the distribution sense.

From (3.21) and from the expression of I^ε involving F^κ it follows, possibly after contracting $\mathcal{O}''$ about the origin, that

When $\varepsilon \to +0$, $I^\varepsilon(z)$ converges in $C^\infty(\Delta_i \times \mathcal{O}'')$ to a function which is holomorphic with respect to z''. (3.24)

We contend that the same property holds for $J^\varepsilon(z)$.

To prove this we deform the contour of integration with respect to θ from $\mathbb{R}_{m-n}$ to the image of $\mathbb{R}_{m-n}$ under the map

$$\theta \mapsto \zeta = (\mathbf{1} + i^t A(x, y))^{-1} \theta - i\tilde{\delta}|\theta| Z''(x, y), \tag{3.25}$$

where the left upper index t indicates transposition, and $\tilde{\delta}$ is a number >0, small enough that ζ remains in the region

$$|\mathrm{Im}\,\zeta| \leqq |\mathrm{Re}\,\zeta|/10 . \tag{3.26}$$

This is possible thanks to (3.2). We consider then the quantity

$$\tilde{Q} = \mathrm{Re}\{i\zeta \cdot Z''(x, y) + \kappa \langle \zeta \rangle [Z''(x', t, y) - Z''(x, y)]^2\}/|\theta|$$

where $\langle \zeta \rangle = ([\zeta]^2)^{1/2}$ (main square-root branch). We have

$$\begin{aligned} \tilde{Q} &= \mathrm{Re}\left\{ i\dot{\theta} \cdot x'' + \tilde{\delta}[Z''(x, y)]^2 + \kappa \frac{\langle \zeta \rangle}{|\theta|} [Z''(x', t, y) - Z''(x, y)]^2 \right\} . \\ &= \tilde{\delta}|x''|^2 - \tilde{\delta}|A(x, y)x''|^2 \\ &\quad + \kappa(\mathrm{Re}\langle \zeta \rangle/|\theta|)(|x'' - t|^2 - |A(x', t, y)t - A(x, y)x''|^2) \\ &\quad - 2\kappa(\mathrm{Im}\langle \zeta \rangle/|\theta|)(t - x'')(A(x', t, y)t - A(x, y)x'') . \end{aligned}$$

A simple argument based on (3.1) and (3.26) shows that there is a constant $c_0 > 0$ such that

$$\tilde{Q} \geqq c_0(|x''|^2 + |x'' - t|^2) .$$

But since $|t| \geqq d_1$ the right-hand side is bounded from below, away from zero. It follows that, if $\mathcal{O}''$ and $c_1 > 0$ are small enough, then

$$\mathrm{Re}\{-i\zeta \cdot [z'' - Z''(x, y)] + \kappa \langle \zeta \rangle [Z''(x', t, y) - Z''(x, y)]^2\} \geqq c_1|\theta|$$

for every $z' \in \Delta'$, $z'' \in \mathcal{O}''$, $\theta \in \mathbb{R}_{m-n}$. Our contention is an immediate consequence of this estimate.

If we combine (3.23), (3.24) and the analogous property of J^ε, we reach the following conclusion:

In the region

$$x' + iy \in \Delta_1', \quad |x''| < d, \tag{3.27}$$

we have

$$u(x, y) = \tilde{u}(z', Z''(x, y)), \tag{3.28}$$

where $\tilde{u}$ is a C^∞ function in the region

$$x' + iy \in \Delta_1', \quad z'' \in \mathcal{O}'', \tag{3.29}$$

holomorphic with respect to z''.

The proof in Case I is almost complete: By applying L_j to both sides in (3.28) we obtain, in the set (3.27), for each $j = 1, \ldots, n$,

$$\frac{\partial \tilde{u}}{\partial \bar{z}^j}(z', Z''(x, y)) = L_j u(x, y) - \sum_{k=n+1}^{m} \frac{\partial \tilde{u}}{\partial z^k}((z', Z''(x, y)) L_j Z^k(x, y) = 0.$$

In this equation fix arbitrarily z' in Δ_1'; $\partial \tilde{u}/\partial \bar{z}^j$ is a holomorphic function of z'' which, as we see, vanishes identically on the image of the ball $|x''| < d$ under the mapping $x'' \mapsto Z''(x', x'', y)$.

Such an image is a totally real submanifold of dimension equal to $m-n$, hence $\partial \tilde{u}/\partial \bar{z}^j \equiv 0$ in $\mathcal{O}''$ provided the latter set is connected. But then by Hartog's theorem $\tilde{u}$ is a holomorphic function of z in the set (3.29), which implies what we wanted.

Case II

We continue to suppose that $r = n$ but now $m' = \dim_{\mathbb{C}} \mathcal{M} < n$ $(m' \geqq 1)$. Thus there are coordinates $y'' = (y^{m'+1}, \ldots, y^n)$. We shall reduce this case to Case I by changing the hypo-analytic structure in U: we introduce new functions Z^k [in addition to those in (1.1)] by setting

$$Z^{m-m'+k} = y^k, \quad k = m'+1, \ldots, n. \tag{3.30}$$

We call $\mathscr{A}$ the hypo-analytic structure on U defined by $Z^1, \ldots, Z^m$, $\mathscr{A}'$ that defined by $Z^1, \ldots, Z^{m+n-m'}$, and by T'' the CR structure defined by $\mathscr{A}'$. It is obvious that $\mathscr{A}'$ is finer than $\mathscr{A}$ (Sect. 1). We may therefore apply Lemma 1.1 with U in the place of Ω.

It suffices then to observe that $\mathscr{A}'$ may be defined on U by the functions $Z_\#^1, \ldots, Z_\#^{m+n-m'}$, where

$$Z_\#^j = x^j + iy^j, \quad j = 1, \ldots, m', \quad Z_\#^k = x^k, \quad k = m'+1, \ldots, n,$$
$$Z_\#^k = x^k + i\Phi^k(x, y), \quad k = n+1, \ldots, m, \quad Z_\#^l = y^l, \quad l = m+1, \ldots, m+n-m'.$$

If we take (2.8) into account we see that $\mathcal{M} \cap U$ is a complex submanifold of U (in the sense of $\mathscr{A}'$). But the number n of the CR structure T' has been replaced, in the case of T'', by m', and therefore we find ourselves in Case I: if u is hypo-analytic for

$\mathscr{A}$ at the origin, it is also so for $\mathscr{A}'$ (trivially), hence, by the result in Case I, it is hypo-analytic for $\mathscr{A}'$ at every point of $\mathscr{M} \cap U$ and finally, by virtue of Lemma 1.1, this is also true for $\mathscr{A}$.

Case III

We do not assume any more that Ω is a hypo-analytic CR manifold. Therefore the number r attached to the hypo-analytic local chart (U, Z) is not necessarily equal to n (but always $r \leqq n$).

Call $\mathscr{N}$ the submanifold $y^* = 0$ of U, $\mathscr{N} = \Delta' \times V'' \times W'' \times \{0\}$ (Sect. 2). According to Proposition 5.2, Chap. I, [7], the trace $u_{\mathscr{N}}$ of u on $\mathscr{N}$ is well defined. The hypo-analytic structure of Ω induces on $\mathscr{N}$ a CR structure. The latter could now mean the standard C^∞ structure, which corresponds to the case $m' = 0$. When $m' \geqq 1$ the result in Case II applies: there is a holomorphic function $\tilde{u}$ in an open neighborhood $\mathscr{O}$ of $Z(\mathscr{N})$ in $\mathbb{C}^m$ such that

$$u_{\mathscr{N}}(x, y', y'') = \tilde{u}(Z(x, y', y'', 0)).$$

Since y^* enters in $\Phi^*(x, y)$ only it follows at once from (2.8) that, as soon as V'' and W'' are sufficiently small, $Z(U) \subset \mathscr{O}$.

We may therefore consider the function – obviously hypo-analytic – in U, $\tilde{u} \circ Z$. Since it equals u on $\mathscr{N}$ it equals u in the whole of U [7, Chap. II, Theorem 2.5], and thus u is hypo-analytic in a full open neighborhood of the submanifold S of U (Sect. 2).

The same argument, but without going through Cases I and II, applies when $m' = 0$. The latter means that $x = x''$, $y = (y'', y^*)$. It suffices to write that, in a neighborhood of the origin, $u = \tilde{u} \circ Z$ with $\tilde{u}$ holomorphic in an open neighborhood of the origin in $\mathbb{C}^m$, $\mathscr{O}$. We reason as when $m' \geqq 1$, with now $\mathscr{N} = V'' \times W'' \times \{0\}$.

The proof of Theorem 3.1 is complete. □

4. Applications, Examples, Singular Solutions

Our first application of Theorem 3.1 will be to *embedded* CR manifolds:

Theorem 4.1. *Let Ω be a* CR *submanifold of $\mathbb{C}^N$, $\mathscr{M}$ a connected holomorphic submanifold of $\mathbb{C}^N$ entirely contained in Ω, u a* CR *distribution in Ω.*

If u extends as a holomorphic function to an open neighborhood in $\mathbb{C}^N$ of a point of $\mathscr{M}$ the same is true at every point of $\mathscr{M}$.

Proof. $\mathscr{M}$ is an elliptic submanifold of Ω. □

The analogous statement is valid for any hypo-analytic CR manifold, even if it is not (globally) embeddable. This is Case II in the proof of Theorem 3.1.

Returning to the case of a general – not necessarily CR – hypo-analytic manifold Ω, another case of interest is that of a hypo-analytic submanifold $\mathscr{M}$ which is both *elliptic* and *essentially real* (see Sect. 1). This can only occur when the pull-back to $\mathscr{M}$ of $T'|_{\mathscr{M}}$ vanishes identically, in which case we say, following [7], that T' *vanishes on* $\mathscr{M}$. It is equivalent to saying that the restriction to $\mathscr{M}$ of any (germ of) hypo-analytic function in Ω is locally constant.

Example 4.1. Consider the hypo-analytic structure on $\mathbb{R}^2$ defined by the single function $Z=xe^{iy}$. Then T' vanishes on the submanifold $x=0$. Note that $T'^{\perp}$ is generated by the vector field

$$L=\frac{\partial}{\partial y}-ix\frac{\partial}{\partial x},$$

which reduces to $\partial/\partial y$ along the y-axis.

Theorem 4.2. *Let Ω be a hypo-analytic manifold, T' its structure bundle, $\mathcal{M}$ a C^∞ connected submanifold of Ω on which T' vanishes.*

If $\mathcal{M}$ intersects the hypo-analytic singular support of any RC-distribution $\mathcal{M}$ is entirely contained in it.

Let $\mathcal{M}$ be as in Theorem 4.2. At each one of its points the tangent space to $\mathcal{M}$ is contained in the fibre of $T'^{\perp}$ at that point. One may say that $T\mathcal{M}$ consists of "hyperbolic directions" of $T'^{\perp}$ (cf. Example 4.1).

Let us return to Theorem 3.1. One might wonder whether it represents the best possible result in its own direction. A possible test would come from the existence of *RC*-distributions whose hypo-analytic singular support is exactly equal to the elliptic submanifold $\mathcal{M}$. There is little hope of carrying out the construction of such *RC*-distributions globally. Note in passing that in general we do not even know whether there are *RC*-distributions in the whole of Ω: But even from the local viewpoint it can only be done under restrictive hypotheses about $\mathcal{M}$, as shown in the following

Example 4.2. Consider the hypo-analytic structure on $\mathbb{R}^5$ defined by the three functions

$$Z^j=x^j+\sqrt{-1}y^j \quad (j=1,2), \qquad Z^3=x^3+\sqrt{-1}\Phi^3(x^2,x^3,y^2).$$

We assume, as usual, that Φ^3 is real-valued, and that it, and its differential, vanish at the origin. It is clear that

$$\mathcal{M}=\{(x,y)\in\mathbb{R}^5;\, x^2=x^3=y^2=0\}$$

is an elliptic, in fact a complex submanifold of $\mathbb{R}^5$ for the hypo-analytic structure under consideration (it is a CR structure on $\mathbb{R}^5$).

If

$$\Phi^3(x^2,x^3,y^2)=(x^2)^2+(y^2)^2$$

then

$$(Z^3)^{1/2}=[x^3+i|Z^2|^2]^{1/2} \quad \text{(main square-root)}$$

is a continuous CR distribution in the whole space, whose hypo-analytic singular support is exactly equal to $\mathcal{M}$.

Suppose instead that

$$\Phi^3(x^2,x^3,y^2)=(y^2)^3.$$

Then it is proved in [2] (Corollary 1.4, also Theorem 2.1 and Example 2.2) that every CR distribution is a hypo-analytic function. Consequently it is not possible to find a CR distribution whose hypo-analytic singular support is contained in $\mathcal{M}$. □

The passing of a complex submanifold of complex dimension n through the point under study enables us to construct "good" local singular solutions. We recall that such a submanifold (if "maximal") is unique (Proposition 2.1).

Theorem 4.3. *Let Ω be a hypo-analytic* CR *manifold of dimension $m+n$, T' its structure bundle (whose fibre dimension is equal to m), $\mathcal{M}$ a complex submanifold of Ω whose complex dimension is equal to n.*

Then every point of $\mathcal{M}$ has an open neighborhood in which there is

i) *a* CR *distribution whose (C^∞) singular support is exactly equal to the piece of the submanifold $\mathcal{M}$ contained in the neighborhood in question;*

ii) *a C^∞* CR *function whose hypo-analytic singular support is exactly equal to that piece of $\mathcal{M}$;*

iii) *a C^∞* CR *function which vanishes to infinite order on the same piece of $\mathcal{M}$ and is different from zero at any point which does not lie on $\mathcal{M}$.*

Proof. We reason in a hypo-analytic local chart (U, Z) such that (1.1) and (1.2) hold. Since T' defines a CR structure on Ω we have $r=n$, $x=(x', x'')$, $y=y'$ (and thus there are no y'' or y^*), and $\mathcal{M}\cap U$ is defined by

$$x''=0. \tag{4.1}$$

This implies that (2.8) holds; it reads here:

$$\Phi''(x, y)=A(x, y)x'', \tag{4.2}$$

where $\Phi''=(\Phi^{n+1}, \ldots, \Phi^m)$, and A is a real $(m-n)\times(m-n)$ matrix such that $A(0,0)=0$. Possibly after contracting U about the origin we may assume that the matrix norm of A has an upper bound in U which is small compared to *one*. We may then make use of the function

$$[Z'']^2=\sum_{k=n+1}^{m}(Z^k)^2=(|x''|^2-|A(x, y)x''|^2)+2ix''\cdot A(x, y)x''.$$

Everything we want now follows easily:

i) the (main branch of the) square-root $([Z'']^2)^{1/2}$ is a continuous function which is not C^1 at the origin, yet is hypo-analytic (and a fortiori C^∞) in $U\backslash\mathcal{M}\cap U$;

ii) and iii) the function

$$f(x, y)=\exp\{-1/[Z''(x, y)]^2\}$$

is C^∞ in U, vanishes to infinite order on $\mathcal{M}\cap U$, is nowhere zero in $U\backslash\mathcal{M}\cap U$, is hypo-analytic in the latter set, but not at any point of $\mathcal{M}\cap U$. □

If one regards Theorem 4.3 as a counterpart to Theorem 4.1, there is an analogous counterpart to Theorem 4.2. In fact the proof of Theorem 4.3 suggests a generalization to an arbitrary hypo-analytic manifold, and thus a counterpart to Theorem 3.1. We leave its statement and proof as an exercise to the reader.

Instead we conclude with the example of an embeddable CR manifold in which propagation (of holomorphic extendability) from a point occurs along two distinct complex submanifolds, and the union of these two complex submanifolds constitutes in fact a "minimal propagator". The example is closely related to Examples 2.1 and 4.2:

Example 4.3. Equip $\mathbb{R}^5$ with the hypo-analytic CR structure defined by the three functions

$$Z^j = x^j + \sqrt{-1}\,y^j \quad (j=1,2); \qquad Z^3 = x^3 + \sqrt{-1}(|Z^1||Z^2|)^2 .$$

There are two, and only two, complex submanifolds passing through the origin, which are maximal (see Sect. 2): they are the complex curves $\mathscr{M}^1$ and $\mathscr{M}^2$ of Example 2.1. The continuous function $[Z^3]^{1/2}$ is obviously CR (we use, as always, the main branch of the square-root). It is hypo-analytic in $\mathbb{R}^5 \setminus (\mathscr{M}^1 \cup \mathscr{M}^2)$; it is not hypo-analytic at any point of $\mathscr{M}^1 \cup \mathscr{M}^2$.

References

1. Baouendi, M.S., Chang, C.H., Treves, F.: Microlocal hypo-analyticity and extension of CR functions. J. Differential Geometry (1983) (to appear)
2. Baouendi, M.S., Treves, F.: A microlocal version of Bochner's tube theorem. Ind. J. Math. **31**, 887–895 (1982)
3. Hanges, N., Sjostrand, J.: Propagation of analyticity for a class of non-micro-characteristic operators. Ann. Math. **116**, 559–577 (1982)
4. Jacobowitz, H., Treves, F.: Nonrealizable CR structures. Invent. Math. **66**, 231–249 (1982)
5. Newlander, A., Nirenberg, L.: Complex analytic coordinates in almost complex manifolds. Ann. Math. **65**, 391–404 (1957)
6. Nirenberg, L.: On a question of Hans Lewy. Russ. Math. Surv. **29**, 251–262 (1974)
7. Treves, F.: Approximation and representation of functions and distributions annihilated by a system of complex vector fields. Ec. Polytechnique Centre Math. publications Mai 1981

Received June 30, 1982; in revised form August 3, 1982

Math. Ann. 263, 179–184 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Capacitary Inequalities of the Brunn-Minkowski Type

Christer Borell
Department of Mathematics, Chalmers University, S-412 96 Göteborg, Sweden

1. Introduction

The main purpose of this paper is to show how so-called $:\alpha:$-concave capacities may be built up from Newton capacity c_n in $\mathbb{R}^n (n \geqq 3)$ in much the same way as $:\alpha:$-concave measures from Lebesgue measure m_n in $\mathbb{R}^n (n \geqq 1)$.

Throughout, $\Omega, \Omega_0, \ldots$ stand for non-empty, open, and convex subsets of $\mathbb{R}^n$ and $\mathscr{K}(\Omega)$ denotes the class of all convex bodies contained in Ω.

Recall that a positive Radon measure μ in Ω is said to be $:\alpha:$-concave if $-\infty \leqq \alpha < +\infty$ and

$$\begin{cases} \mu(\lambda A + (1-\lambda)B) \geqq (\lambda \mu^\alpha(A) + (1-\lambda)\mu^\alpha(B))^{1/\alpha} \\ \text{all} \quad A, B \in \mathscr{K}(\Omega), \quad 0 < \lambda < 1, \end{cases} \tag{1.1}$$

[2, 3, 6]. In a similar way a capacity μ in Ω will be called $:\alpha:$-concave if (1.1) is true.

Below, for any positive function f, we introduce the short-hand notation $f_{\alpha,n} = -f^{-1/n}$, $\alpha = -\infty$; $= (\operatorname{sgn}\alpha) f^{\alpha/(1-\alpha n)}$, $\alpha \neq 0$, $-\infty < \alpha < 1/n$; $= \ln f$, $\alpha = 0$; and $=$ an arbitrary positive constant for $\alpha = 1/n$, respectively. Using this convention, a positive Radon measure μ in Ω with $\operatorname{supp}\mu = \Omega$ is $:\alpha:$-concave precisely when $\alpha \leqq 1/n$, $\mu \ll m_n$, and a suitable version of the function $(d\mu/dm_n)_{\alpha,n}$ is concave. In a similar way, given a Borel function $f: \Omega \to]0, +\infty[$, the capacity fc_n is $:\alpha:$-concave as soon as the function $f_{\alpha, n-2}$ is concave (Theorem 3.1). However, as is readily seen, this construction does not exhaust the class of all $:\alpha:$-concave capacities in Ω.

To establish the crucical estimate

$$c_n^{1/(n-2)}(A+B) \geqq c_n^{1/(n-2)}(A) + c_n^{1/(n-2)}(B), \quad A, B \in \mathscr{K}(\mathbb{R}^n) \tag{1.2}$$

it is natural to compare the equilibrium potentials in $\mathbb{R}^n$ of the underlying convex bodies. In what follows, suppose $\bar{\Omega}, \bar{\Omega}_0, \ldots \in \mathscr{K}(\mathbb{R}^n)$ if $n \leqq 2$ and let $u_A^\Omega = P$. [Brownian motion hits A before $\Omega^c = \mathbb{R}^n \backslash \Omega$] be the equilibrium potential of $A \in \mathscr{K}(\Omega)$ relative to Ω. Then, if $A_i \in \mathscr{K}(\Omega_i)$, $i = 0, 1$, we use an idéa of Gabriel [7, 8]

to prove that

$$u^{(1-\lambda)\Omega_0+\lambda\Omega_1}_{(1-\lambda)A_0+\lambda A_1} \geqq u^{\Omega_0}_{A_0} \wedge u^{\Omega_1}_{A_1}, \qquad 0<\lambda<1, \tag{1.3}$$

(Theorem 2.1), which immediately implies (1.2).

Interestingly enough, the inequality (1.3) also enlarges the previously known collection of $:\alpha:$-concave measures, a fact we express as a simple convexity property of the Poisson kernel in a convex polyhedron (Corollary 2.1).

2. A Brunn-Minkowski Version of the Gabriel-Lewis Inequality

Theorem 2.1. *Let* $x_i \in \mathbb{R}^n$, $A_i \in \mathscr{K}(\Omega_i)$, $i=0,1$, *and set* $\xi_\lambda = (1-\lambda)\xi_0 + \lambda\xi_1$, $\xi = x, A, \Omega$, $0<\lambda<1$. *Then*

$$u^{\Omega_\lambda}_{A_\lambda}(x_\lambda) \geqq u^{\Omega_0}_{A_0}(x_0) \wedge u^{\Omega_1}_{A_1}(x_1).$$

The special case $A_0 = A_1, \Omega_0 = \Omega_1$ is due to Gabriel [7, 8] and Lewis [9]. A parabolic extension of the Gabriel-Lewis inequality has recently been settled by the author [5]. Note that Theorem 2.1 may be interpreted as a Brunn-Minkowski inequality of Brownian motion (for other estimates of this type see [6, 4], and a forthcoming paper by Antoine Ehrhard, Strasbourg).

The proof of Theorem 2.1 given below does not contain any new ingredients. Nevertheless, for the sake of completeness we will provide all details in what follows.

Lemma 2.1. *For any* $A \in \mathscr{K}(\Omega)$, $\nabla u^\Omega_A \neq 0$ *in* $\Omega \backslash A$.

Proof (assuming $0 \in A$). If $0<\alpha<1$, then by the maximum principle $u^\Omega_A(\alpha(\cdot)) \geqq u^\Omega_A$ and it follows that the non-zero harmonic map $x \cdot \nabla u^\Omega_A(x)$, $x \in \Omega \backslash A$, must be negative.

Proof of Theorem 2.1 [assuming $\bar{\Omega}_0, \bar{\Omega}_1 \in \mathscr{K}(\mathbb{R}^n)$]. Let, for brevity, $u_\lambda = u^{\Omega_\lambda}_{A_\lambda} | \bar{\Omega}_\lambda$, $0 \leqq \lambda \leqq 1$.

In what follows, suppose $0<\lambda<1$ is fixed and set

$$u^*_\lambda(x) = \sup\{u_0(x_0) \wedge u_1(x_1); x = (1-\lambda)x_0 + \lambda x_1, x_i \in \bar{\Omega}_i, i=0,1\}, \qquad x \in \bar{\Omega}_\lambda.$$

We shall prove the inequality $u^*_\lambda \leqq u_\lambda$. To this end, we show that the statement $\neg(u^*_\lambda \leqq u_\lambda)$ leads to a contradiction.

In fact, suppose there exists an $(\varepsilon, \hat{x}) \in]0,1[\times \bar{\Omega}_\lambda$ such that $\sup(u^*_\lambda - u^\varepsilon_\lambda) = u^*_\lambda(\hat{x}) - u^\varepsilon_\lambda(\hat{x})$ is strictly positive. Of course, $\hat{x} \in \Omega_\lambda \backslash A_\lambda$ because $u^*_{\lambda|\partial\Omega_\lambda} = 0$ and $u_{\lambda|A_\lambda} = 1$. Let $\hat{x} = (1-\lambda)\hat{x}_0 + \lambda\hat{x}_1$, where the $\hat{x}_i \in \bar{\Omega}_i$ and $u^*_\lambda(\hat{x}) = u_0(\hat{x}_0) \wedge u_1(\hat{x}_1)$. Certainly, $(\hat{x}_0, \hat{x}_1) \in (\Omega_0 \times \Omega_1) \backslash (A_0 \times A_1)$. Also, it is easy to see that the relation $\hat{x}_0 \notin A_0$, $\hat{x}_1 \in A_1$ is contradictory. Indeed, $\hat{x}_0 \in \partial u_0^{-1}(]u_0(\hat{x}_0), +\infty[)$ due to harmonicity and if $x_0 \in u_0^{-1}(]u_0(\hat{x}_0), +\infty[)$ is sufficently close to $\hat{x}_0$, then $u_1((\hat{x}-(1-\lambda)x_0)/\lambda) > u^*_\lambda(\hat{x})$ forcing $u^*_\lambda(\hat{x}) > u^*_\lambda(\hat{x})$, which cannot hold. Thus, by symmetry, $(\hat{x}_0, \hat{x}_1) \in (\Omega_0 \backslash A_0) \times (\Omega_1 \backslash A_1)$.

Now suppose $h \in \mathbb{R}^n$ and $h \cdot \nabla u_i(\hat{x}_i) > 0$ ($i=0$ or 1). Then, if $s>0$ is sufficiently small, $u_i(\hat{x}_i + sh) > u_i(\hat{x}_i)$ and, hence, $u^*_\lambda(\hat{x} + s\lambda_i h) \geqq u^*_\lambda(\hat{x})$, where $\lambda_i = (2i-1)\lambda + 1 - i$, so that $u^\varepsilon_\lambda(\hat{x} + s\lambda_i h) \geqq u^\varepsilon_\lambda(\hat{x})$. Accordingly, $h \cdot \nabla u_\lambda(\hat{x}) \geqq 0$ and it follows that the non-

zero vectors $\nabla u_\lambda(\hat{x})$ and $\nabla u_i(\hat{x}_i)$ are parallel. In the following, let $a=|\nabla u^\varepsilon_\lambda(\hat{x})|$, $a_i=|\nabla u_i(\hat{x}_i)|$, and $v=\nabla u^\varepsilon_\lambda(\hat{x})/a$, respectively.

From now on we assume that $u^*_\lambda(\hat{x})=u_0(\hat{x}_0)$. The case $u^*_\lambda(\hat{x})=u_1(\hat{x}_1)$ may be treated in a quite similar way.

Let $h\in\mathbb{R}^n$ be such that the scalar $\kappa=h\cdot v\neq 0$. For each real s close to 0 there exists a unique $r=r(s)$, with $|r|$ minimal, satisfying the equation

$$u_0(\hat{x}_0+sh/a_0)-u_0(\hat{x}_0)=u_1(\hat{x}_1+rh/a_1)-u_1(\hat{x}_1).$$

Writing $\hat{x}(s)=(1-\lambda)(\hat{x}_0+sh/a_0)+\lambda(\hat{x}_1+r(s)h/a_1)=\hat{x}+((1-\lambda)s/a_0+\lambda r(s)/a_1)h$ we now have

$$u_0(\hat{x}_0+sh/a_0)-u^\varepsilon_\lambda(\hat{x}(s))\leqq u^*_\lambda(\hat{x}(s))-u^\varepsilon_\lambda(\hat{x}(s))\leqq u_0(\hat{x}_0)-u^\varepsilon_\lambda(\hat{x}).$$

In particular,

$$D^2_s(u_0(\hat{x}_0+sh/a_0)-u^\varepsilon_\lambda(\hat{x}(s)))_{|s=0}\leqq 0. \tag{2.1}$$

To analyse (2.1) we introduce the following Taylor expansions

$$u_i(\hat{x}_i+sh/a_i)=u_i(\hat{x}_i)+\kappa s+b_i s^2+o(s^2)\quad\text{as}\quad s\to 0\qquad(i=0,1)$$

and get

$$r(s)=s+\kappa^{-1}(b_0-b_1)s^2+o(s^2)\quad\text{as}\quad s\to 0.$$

Moreover, setting $p=(1-\lambda)/a_0+\lambda/a_1$ and

$$u^\varepsilon_\lambda(\hat{x}+sh)=u^\varepsilon_\lambda(\hat{x})+\kappa as+Bs^2+o(s^2)\quad\text{as}\quad s\to 0,$$

respectively, a straightforward calculation yields

$$u^\varepsilon_\lambda(\hat{x}(s))=u^\varepsilon_\lambda(\hat{x})+\kappa aps+[p^2B+\lambda(a/a_1)(b_0-b_1)]s^2+o(s^2)\quad\text{as}\quad s\to 0.$$

Thus from (2.1), $\xi\cdot(b_0,b_1)-p^2B\leqq 0$ for a suitable vector $\xi\in\mathbb{R}^2$, which does not depend on h. But then, for any $h\in\mathbb{R}^n$,

$$\sum_{1\leqq j,k\leqq n}\left[\xi\cdot\left(\frac{1}{a_0^2}D^2_{x_jx_k}u_0(\hat{x}_0),\frac{1}{a_1^2}D^2_{x_jx_k}u_1(\hat{x}_1)\right)-p^2D^2_{x_jx_k}u^\varepsilon_\lambda(\hat{x})\right]h_jh_k\leqq 0$$

and we have

$$\xi\cdot\left(\frac{1}{a_0^2}\Delta u_0(\hat{x}_0),\frac{1}{a_1^2}\Delta u_1(\hat{x}_1)\right)-p^2\Delta u^\varepsilon_\lambda(\hat{x})\leqq 0,$$

or, stated otherwise, $|\nabla u_\lambda(\hat{x})|=0$. From this contradiction we conclude that $u^*_\lambda\leqq u_\lambda$ and the proof of Theorem 2.1 is completed.

Corollary 2.1. *Assume $\bar{\Omega}$ is a convex polyhedron in $\mathbb{R}^n$ and denote by F any $(n-1)$-dimensional facet of $\bar{\Omega}$. Let κ^Ω_x be the harmonic measure at $x\in\Omega$ relative to Ω. Then an appropriate version of the restricted Poisson kernel*

$$\Omega\times\mathrm{ri}(\bar{\Omega}\cap F)\to]0,+\infty[,(x,z)\rightsquigarrow[d\kappa^\Omega_x/d\sigma_{\mathrm{ri}(\bar{\Omega}\cap F)}](z)$$

raised to the power $-1/(n-1)$ is convex.

Here ri denotes the relative interior operator.
The special case $\Omega=\{x_n>0\}$ $(n\geqq 2)$ with

$$d\kappa_x^{\Omega}(z)=\frac{\Gamma(n/2)}{\pi^{n/2}}\frac{x_n}{\|x-z\|^n}dz, \qquad x_n>0, \qquad z_n=0,$$

shows that, generally, there does not exist any greater real number than $-1/(n-1)$ possessing the property stated in Corollary 2.1.

Proof. Let $A, B\subseteqq \mathrm{ri}(\bar{\Omega}\cap F)$ be arbitrary compact and convex sets. Then, by applying Theorem 2.1, we conclude that

$$\begin{cases}\kappa^{\Omega}_{\lambda x+(1-\lambda)y}(\lambda A+(1-\lambda)B)\geqq\kappa_x^{\Omega}(A)\wedge\kappa_y^{\Omega}(B)\\ \text{all} \quad x, y\in\Omega, \qquad 0<\lambda<1\end{cases}$$

which is just another way to state what we want to prove [3, Theorems 3.2, and 3.1]. □

3. :α:-Concave Capacities Induced by Newton Capacity

Recall that if μ is a capacity in Ω [11, p. 7.30] and $f:\Omega\to[0,+\infty]$ is a Borel function, then, by definition,

$$(f\mu)(A)=\int_A f d\mu=\int_0^{\infty}\mu(A\cap\{f\geqq s\})ds, \qquad A\in\mathscr{B}(\Omega).$$

Theorem 3.1. *Let $n\geqq 3$ and $-\infty\leqq\alpha\leqq 1/(n-2)$. Moreover, assume $f:\Omega\to]0,+\infty[$ is such that $f_{\alpha,n-2}$ is concave. Then the capacity fc_n is :α:-concave.*

Proof. We first consider the special case $\alpha=1/(n-2)$.

Following the convention that $c_n(B(0;r))=r^{n-2}$,

$$c_n(A)=\lim_{|x|\to+\infty}\lim_{r\to+\infty}|x|^{n-2}u_A^{B(0;r)}(x), \qquad A\in\mathscr{K}(\mathbb{R}^n),$$

and, hence, for any fixed $A, B\in\mathscr{K}(\mathbb{R}^n)$,

$$c_n(\lambda A+(1-\lambda)B)\geqq c_n(A)\wedge c_n(B), \qquad 0<\lambda<1,$$

in view of Theorem 2.1. In particular,

$$c_n(\lambda A/c_n^{1/(n-2)}(A)+(1-\lambda)B/c_n^{1/(n-2)}(B))\geqq 1, \qquad 0<\lambda<1,$$

and (1.2) follows by choosing λ as the solution of the equation

$$\lambda/c_n^{1/(n-2)}(A)=(1-\lambda)/c_n^{1/(n-2)}(B).$$

Next suppose $0<\alpha<1/(n-2)$ and set $1/\beta=\alpha/(1-\alpha(n-2))$. Given $A, B\in\mathscr{K}(\mathbb{R}^n)$ and $0<\lambda<1$, we shall prove that

$$\int_0^{\infty}c_n((\lambda A+(1-\lambda)B)\cap\{f\geqq s^{\beta}\})s^{\beta-1}ds$$

$$\geqq\left[\lambda\left(\int_0^{\infty}c_n(A\cap\{f\geqq s^{\beta}\})s^{\beta-1}ds\right)^{\alpha}+(1-\lambda)\left(\int_0^{\infty}c_n(B\cap\{f\geqq s^{\beta}\})s^{\beta-1}ds\right)^{\alpha}\right]^{1/\alpha},$$

which, however, is an immediate consequence of (1.2) and a standard inequality from the theory of $:\alpha:$-concave measures [3, Theorem 3.1 or 6, Theorem 5.1].

Finally, the case $-\infty \leqq \alpha \leqq 0$ may be treated in a similar way and we omit the details here.

This completes the proof of Theorem 3.1. □

4. Some Simple Examples

Example 4.1. The Newtonian capacity of a rectangular box in $\mathbb{R}^3$ is a concave function of the edge lengths. □

Example 4.2. If a body $A \in \mathcal{K}(\mathbb{R}^3)$ carries the total charge $p>0$, then the electrostatic energy $E(A,p)$ equals $p^2/c_3(A)$. Hence, for all $A, B \in \mathcal{K}(\mathbb{R}^3)$ and $p, q>0$,

$$E(A+B, p+q) \leqq E(A,p)+E(B,q)$$

by the Hölder inequality. □

Example 4.3. Consider an $A \in \mathcal{K}(\mathbb{R}^2)$ and suppose $g=(0,0,1)$ and $\varepsilon>0$. Following the convention that $\mathbb{R}^2 = \mathbb{R}^2 \times \{0\} \subseteqq \mathbb{R}^3$, the sets $A \pm \frac{\varepsilon}{2} g$ define a plate condenser $A(\varepsilon) = \left(A - \frac{\varepsilon}{2} g\right) \cup \left(A + \frac{\varepsilon}{2} g\right)$ possessing the capacity $c(A,\varepsilon)$ equal to the total mass of the equilibrium measure μ of $A + \frac{\varepsilon}{2} g$ relative to the half space $\{x_3>0\}$. Thus, using probabilistic normalizations (Port and Stone [10]), it follows that

$$u^H_{A+\frac{\varepsilon}{2}g}(x) = \frac{1}{2\pi} \int \left(\frac{1}{|x-y|} - \frac{1}{|x-y+2y_3 g|} \right) d\mu(y),$$

where the integrand equals

$$\frac{1}{2} \frac{\varepsilon^2}{|x-y|^3} + o\left(\frac{1}{|x-y|^3} \right)$$

uniformly in $y \in A + \frac{\varepsilon}{2} g$ as $|x| \to +\infty$ subject to the restriction $x_3 = \varepsilon/2$. By applying Theorem 2.1 we now have for all $A, B \in \mathcal{K}(\mathbb{R}^3)$ and $\varepsilon, \delta > 0, 0 < \lambda < 1$,

$$\begin{aligned} &(\lambda\varepsilon + (1-\lambda)\delta)^2 c(\lambda A + (1-\lambda)B, \lambda\varepsilon + (1-\lambda)\delta) \\ &\quad \geqq [\varepsilon^2 c(A,\varepsilon)] \wedge [\delta^2 c(B,\delta)]. \end{aligned}$$

Finally, noting that $c(\cdot,\cdot)$ is positive homogenous of degree one, we get

$$(\varepsilon+\delta)^{2/3} c^{1/3}(A+B, \varepsilon+\delta) \geqq \varepsilon^{2/3} c^{1/3}(A,\varepsilon) + \delta^{2/3} c^{1/3}(B,\delta). \quad \square$$

Example 4.4. Let $\Omega \in \mathcal{K}(\mathbb{R}^n)\,(n \geqq 3)$ be bounded and suppose $f: \Omega \to]0, +\infty[$ is concave. Then

$$\int_\Omega f^p dc_n = p \int_0^{\|f\|_\infty} s^{p-1} c_n(f \geqq s) ds,$$

where the function $c_n^{1/(n-2)}(f \geqq s)$, $0 < s < \|f\|_\infty$, is decreasing and concave. We may now apply known elementary integral inequalities [1, Lemma 3.1] to deduce that the function

$$\ln\left[\binom{n-2+p}{n-2}\int_\Omega f^p dc_n\right], \quad p>0,$$

is concave. Furthermore, the function

$$\left[\binom{n-2+p}{n-2}\int_\Omega f^p dc_n / c_n(\Omega)\right]^{1/p}, \quad p>0,$$

decreases. □

Example 4.5. Suppose μ is an $:\alpha:$-concave capacity in $\mathbb{R}^n$ and, in addition, suppose μ is symmetric, that is, $\mu(A)=\mu(-A)$, $A\in\mathscr{B}(\mathbb{R}^n)$. Then, for all barrels $A, B\subseteqq\mathbb{R}^n$ and every $x\in\mathbb{R}^n$ and $\theta>1$,

$$\mu(A\cap(B+x))\geqq\mu(A\cap(B+\theta x)).$$

Indeed, μ is increasing and the claim above follows at once from the set relation

$$A\cap(B+x)\supseteqq\frac{\theta+1}{2\theta}(A\cap(B+\theta x))+\frac{\theta-1}{2\theta}(A\cap(B-\theta x)). \quad \square$$

References

1. Borell, C.: Complements of Lyapunov's inequality. Math. Ann. **205**, 323–331 (1973)
2. Borell, C.: Convex measures on locally convex spaces. Ark. Mat. **12**, 239–352 (1974)
3. Borell, C.: Convex set functions in d-space. Period. Math. Hungar. **6**, 111–136 (1975)
4. Borell, C.: The Brunn-Minkowski inequality in Gauss space. Invent. Math. **30**, 207–216 (1975)
5. Borell, C.: Brownian motion in a convex ring and quasi-concavity. Commun. Math. Phys. **86**, 143–147 (1982)
6. Borell, C.: Convexity of measures in certain convex cones in vector space σ-algebras. Math. Scand. (to appear)
7. Gabriel, R.M.: An extended principle of the maximum for harmonic functions in 3-dimensions. J. London Math. Soc. **30**, 388–401 (1955)
8. Gabriel, R.M.: A result concerning convex level surfaces of 3-dimensional harmonic functions. J. London Math. Soc. **32**, 286–294 (1957)
9. Lewis, J.L.: Capacitary functions in convex rings. Arch. Rational Mech. Anal. **66**, 201–224 (1977)
10. Port, S.C., Stone, C.J.: Brownian motion and classical potential theory. New York, San Francisco, London: Academic Press 1978
11. Rao, M.: Brownian motion and classical potential theory. Lecture Notes Series No. 47, Aarhus Universitet 1977

Received October 8, 1982

Math. Ann. 263, 185–212 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Smoothing Cusp Singularities of Small Length

Robert Friedman[1] and Rick Miranda[2]

[1] Department of Mathematics, Columbia University, New York City, NY 10027, USA
[2] Department of Mathematics, Colorado State University, Fort Collins, CO 80523, USA

Dedicated to T.S.M., with deepest respect

0. Introduction

After the rational double points, perhaps the most tractable class of normal surface singularities is that of the minimally elliptic singularities of Laufer [7]. Such a singularity $(\bar{V}, p)$ is characterized by the properties that it is Gorenstein, and $R^1\pi_*\mathcal{O}_V \cong \mathbb{C}$, where $\pi: V \to \bar{V}$ is a resolution of the analytic germ of the singularity $(\bar{V}, p)$. Minimally elliptic singularities with reduced fundamental cycle in the minimal resolution fall into three broad classes: the simple elliptic singularities, the Dolgachev singularities $D_{p,q,r}$, and the cusp singularities. The versal deformation space of a simple elliptic singularity is well understood, by work of Pinkham and Looijenga; in particular, it is easy to describe which of them are smoothable. A considerable amount of recent work has centered on $D_{p,q,r}$ singularities with $\mathbb{C}^*$-action; smoothability for these has also been worked out by Wahl, Looijenga, and Pinkham.

In this paper, we consider the existence of smoothings of cusp singularities. It is shown that the existence of a smoothing is equivalent to a purely combinatorial statement concerning the existence of certain configurations of rational surfaces, given in Sect. 2. Our motivation has been the recent work in the degenerations of K3 surfaces, notably the classification of Kulikov and Persson-Pinkham. After some discussion of the combinatorics involved, we verify a conjective of Looijenga on the existence of smoothings of cusp singularities in special cases. As will no doubt be clear to the reader, the combinatorial problem involved in checking Looijenga's conjecture in the general case using our methods is rather daunting.

Using different techniques, Henry Pinkham and the first author have verified Looijenga's conjecture for almost all cusps of length less than or equal to three. The remaining cases are shown to be smoothable in Sect. 5, so that the two techniques combined verify the conjecture for all cusps of length $\leqq 3$.

1 Supported in part by NSF grant MCS 81-14179
2 Supported in part by NSF grant MCS 81-08814

1. Preliminaries

Let $(\bar{V}, p)$ be the germ of a cusp singularity. Thus, if $\pi: V \to \bar{V}$ is the minimal resolution, the exceptional divisor

$$\pi^{-1}(p) = D = \sum_{i=1}^{r} D_i$$

is a cycle of rational curves D_i meeting transversally. By convention, D_i meets $D_{i\pm 1}$ transversally (where we use subscripts mod r), and a cycle of length one is an irreducible nodal rational curve. We will always denote by r [or $r(D)$ if necessary] the length of the cycle. Further, let

$$m = m(D) = \begin{cases} 2 & \text{if } D^2 = -1 \\ -D^2 & \text{otherwise} \end{cases}$$

denote the multiplicity of (V, p), and set

$$d_i = -D_i^2 \quad \text{for each} \quad i = 1, \ldots, r.$$

The negative definiteness of the intersection matrix for the components of D is then equivalent to the conditions

$$\left.\begin{array}{ll} \text{(i)} & d_i \geqq 2 \quad \text{for every } i \\ \text{(ii)} & d_j \geqq 3 \quad \text{for some } j \end{array}\right\} \quad \text{if} \quad r \geqq 2$$

or just: $d_1 \geqq 1$ if $r = 1$.

The self-intersection D^2 can then be written

$$-D^2 = \begin{cases} d_1 & \text{if } r = 1 \\ \sum (d_i - 2) & \text{if } r \geqq 2, \end{cases}$$

so that

$$m(D) = \begin{cases} 2 & \text{if } D^2 = -1 \\ d_1 & \text{if } r = 1 \text{ and } d_1 \geqq 2 \\ \sum (d_i - 2) & \text{if } r \geqq 2 \text{ and } D^2 \leqq -2. \end{cases}$$

The cycle of integers $(d_1, \ldots, d_r)$ determines the analytic type of the germ of the cusp singularity $(\bar{V}, p)$. For this reason, we will abuse notation and use the letter D to denote

a) the cycle of integers $(d_1, \ldots, d_r)$,

b) the divisor D, the cycle of rational curves $\sum D_i$, with $-D_i^2 = d_i$, and

c) the germ of the cusp singularity $(\bar{V}, p)$ whose resolution is the cycle of curves D,

and we will use the cycle of integers $(d_1, \ldots, d_r)$ when we need to be specific.

It will be convenient for some purposes to represent a cusp D not by the cycle of integers but by the following:

Notation (1.1). The $2 \times k$ matrix $\begin{pmatrix} a_1 \ldots a_k \\ b_1 \ldots b_k \end{pmatrix}$ denotes a cusp D with components D_i having the following self-intersections:

(a) There are k components $D_{j_1}, \ldots, D_{j_k}$ with $D_{j_i}^2 = -a_i$, and $a_i \geqq 3$.

(b) Separating D_{j_i} and $D_{j_{i+1}}$ there are b_i components with self-intersection -2; $b_i \geqq 0$.

Here we set $D_{j_{k+1}} = D_{j_1}$.

For example, $\begin{pmatrix} 3 & 4 & 4 \\ 1 & 0 & 2 \end{pmatrix}$ is the cusp whose cycle of integers is $(3,2,4,4,2,2)$. If $r(D)=1$, we will alternately use the notation $\begin{pmatrix} a_1 \\ 0 \end{pmatrix}$ or (a_1) when $D = D_1$ and $D_1^2 = -a_1$; here $a_1 \geqq 1$.

The following is now immediate.

Lemma (1.2). *Let D be the cusp $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$. Then*

$$r(D) = \sum (b_i + 1), \tag{1.2.1}$$

$$-D^2 = \begin{cases} a_1 & \text{if } r(D)=1 \\ \sum (a_i - 2) & \text{if } r(D) \geqq 2. \end{cases} \tag{1.2.2}$$

Hence $D^2 = -1$ if and only if D is either $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$ or $\begin{pmatrix} 3 \\ b \end{pmatrix}$ for some $b \geqq 1$

$$m(D) = \begin{cases} 2 & \text{if } D^2 = -1 \\ a_1 & \text{if } r(D)=1 \text{ and } a_1 \geqq 2 \\ \sum (a_i - 2) & \text{if } r(D) \geqq 2 \text{ and } D^2 \leqq -2. \end{cases} \tag{1.2.3}$$

Definition (1.3). Let D be the cusp $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$. The *dual cusp* D' to D is $\begin{pmatrix} a_i' \\ b_i' \end{pmatrix}$, where

$$a_i' = b_i + 3 \quad \text{and} \quad b_i' = a_{i+1} - 3$$

unless $r(D)=1$ or $D^2 = -1$.

If $D = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$, then $D' = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$.

If $D = \begin{pmatrix} a \\ 0 \end{pmatrix}$ with $a \geqq 2$, then $D' = \begin{pmatrix} 3 \\ a-1 \end{pmatrix}$.

If $D = \begin{pmatrix} 3 \\ b \end{pmatrix}$ with $b \geqq 1$, *then* $D' = \begin{pmatrix} b+1 \\ 0 \end{pmatrix}$.

The following lemma is an immediate consequence of the above definition.

Lemma (1.4). *Let D be a cusp, and D' its dual. Then*

(1.4.1). *The dual of D' is D*

$$r(D') = -D^2 = \begin{cases} 1 & \text{if } D = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \text{ or } \begin{pmatrix} 3 \\ b \end{pmatrix}, \quad b \geqq 1 \\ m(D) & \text{otherwise}, \end{cases} \tag{1.4.2}$$

$$-D'^2 = \begin{cases} 1 & \text{if } D = \begin{pmatrix} a \\ 0 \end{pmatrix}, \quad a \geqq 1 \\ r(D) & \text{otherwise}. \end{cases} \tag{1.4.3}$$

Hence $D'^2 = -1$ *if and only if* $D = \begin{pmatrix} a \\ 0 \end{pmatrix}$ *for some* $a \geqq 1$

$$m(D') = \begin{cases} 2 & \text{if } D = \begin{pmatrix} a \\ 0 \end{pmatrix}, \quad a \leqq 1 \\ r(D) & \text{otherwise.} \end{cases} \tag{1.4.4}$$

We next describe a certain class of surfaces of Type VII_0 in Kodaira's list, the *Inoue-Hirzebruch* surfaces. These are surfaces V which have the following properties:

(1.5.1) $\mathcal{O}_V(-K_V) = \mathcal{O}_V(D+D')$, where D is a negative definite cycle of rational curves, D' is the dual cycle as in (1.3), and D and D' are disjoint.

(1.5.2) The only curves on V are the components of D and D'.

(1.5.3) Setting $h^{p,q} = \dim H^q(V, \Omega_V^p)$, then

$$h^{p,q} = \begin{cases} 1 & \text{if } (p,q) = (0,0), (2,2), (0,1), \text{ or } (2,1) \\ r(D)+r(D') & \text{if } (p,q) = (1,1) \\ 0 & \text{otherwise.} \end{cases}$$

(1.5.4) Every cusp D occurs on some Inoue-Hirzebruch surface.

For the construction, see [4, 3]. Briefly, V is the resolution of the completion $\bar{\bar{V}}$ of a non-compact surface which is the quotient of $\mathbb{H} \times \mathbb{C}$ by a discrete group associated to a real quadratic irrationality ω ($\mathbb{H}$ is the upper half-plane). $\bar{\bar{V}}$ has two dual cusp singularities whose resolutions yield the divisors D and D'.

Finally, we recall some facts about Type III degenerations which will motivate the constructions of the next section. Let Δ be the unit disk in $\mathbb{C}$ and $\pi : X \to \Delta$ a degeneration of K3 surfaces over Δ. By the semi-stable reduction theorem, we may assume that X is smooth and $X_0 = \pi^{-1}(0)$ is a reduced divisor with (local) normal crossings. By the theorem of Kulikov and Persson-Pinkham, if all components of X_0 are algebraic we may assume in addition that the global canonical divisor K_X of X is trivial. In this situation, the central fibers X_0 have been classified, and there are three types [6]. In the Type III case, X_0 has the following form:

(1.6.1) $X_0 = \sum V_i$, and if $\tilde{V}_i$ is the normalization of V_i, then each $\tilde{V}_i$ is a rational surface. If D_i is the inverse image on $\tilde{V}_i$ of the double curves of X_0 on V_i, then D_i is a cycle of rational curves. Moreover, the dual graph Γ of X_0 is a triangulation of S^2.

(1.6.2) $D_i \in |-K_{\tilde{V}_i}|$ for each i.

(1.6.3) If D_{ij} is an irreducible double curve joining V_i to V_j (i may equal j if $V_i = V_j$ meets itself), then

$$(D_{ij}^2)_{\tilde{V}_i} + (D_{ij}^2)_{\tilde{V}_j} = \begin{cases} 0 & \text{if } D_{ij} \text{ is a nodal curve on } \tilde{V}_i \text{ or } \tilde{V}_j \\ -2 & \text{otherwise.} \end{cases}$$

This is usually referred to as the *triple point formula.*

The main result of [1] is that (1.6.1)–(1.6.3) are the *only* combinatorial restrictions on central fibers of Type III degenerations of K3 surfaces, i.e., we can always smooth a central fiber X_0 with the desired combinatorial description subject only to (1.6.1)–(1.6.3). This result is the motivation for the next section.

2. Deformation Theory

After a brief description of Looijenga's results, we state the combinatorial problem alluded to in the introduction, and prove that its solution is equivalent to the existence of a smoothing of a cusp singularity.

In the notation of Sect. 1, let $\bar{\bar{V}}$ be the singular Inoue-Hirzebruch surface with two dual cusps D and D'; let the corresponding singular points of $\bar{\bar{V}}$ be p and p'. Looijenga proves the

Theorem (2.1) [8]. *$\bar{\bar{V}}$ has a universal deformation which is semi-universal for the (disconnected) germ of the surface singularities $(\bar{\bar{V}}, p, p')$.*

In particular, suppose that $(\bar{\bar{V}}, p)$ is smoothable. By choosing a deformation $\bar{\bar{\pi}}: \bar{\bar{X}} \to \Delta$ with $\bar{\bar{X}}_0 = \bar{\bar{V}}$ which globalizes a given local smoothing of $(\bar{\bar{V}}, p)$, we obtain a family of surfaces $\bar{\bar{X}}_t$ with a unique singular point p'_t, $t \neq 0$; moreover, $(\bar{\bar{X}}_t, p'_t)$ is analytically isomorphic to $(\bar{\bar{V}}, p')$ and we may simultaneously resolve the surfaces $\bar{\bar{X}}_t$, $t \neq 0$, and $\bar{\bar{V}} = \bar{\bar{X}}_0$ at p'. This produces a degeneration $\bar{\pi}: \bar{X} \to \Delta$ where $\bar{X}_t$ is the resolution of $\bar{\bar{X}}_t$ for $t \neq 0$, and $\bar{X}_0$ is the (partial) resolution $\bar{V}$ of $\bar{\bar{V}}$, with one singular point p and the dual cycle of rational curves D'. The surfaces X_t, $t \neq 0$, contain D' as an anticanonical cycle. It is easy to show that $\bar{X}_t$ is, in fact, a smooth rational surface [8, (2.8)]. Hence:

Corollary (2.2). *If the cusp D is smoothable, then the dual cycle D' sits as an anticanonical divisor on a smooth rational surface.*

Looijenga's Conjecture (2.3) [8, (2.11)]. *Conversely, if D' is an anticanonical divisor on a smooth rational surface, then the cusp D is smoothable.*

We now fix some notation which will be used throughout the rest of this section. Assume

$$X_0 = \bigcup_{i \geqq 0} V_i$$

is a surface with (local) normal crossings satisfying:

(2.4.1) The dual graph Γ of X_0 is a triangulation of S^2.

(2.4.2) V_0 is an Inoue-Hirzebruch surface as in 1.5.

(2.4.3) The normalization $\tilde{V}_i$ of V_i, $i > 0$, is a smooth rational surface.

(2.4.4) If D_0 is the double curve of X_0 on V_0, then $D_0 = D$ is one component of the anticanonical divisor on V_0.

(2.4.5) For $i > 0$, let D_i be the inverse image of the double curves of X_0 on $\tilde{V}_i$; then (V_i, D_i) is an *anticanonical pair*, i.e., D_i is a reduced cycle of rational curves on $\tilde{V}_i$, and $D_i \in |-K_{V_i}|$.

(2.4.6) The irreducible double curves D_{ij} of X_0 satisfy the triple point formula (1.6.3).

The main result of this section is then:

Theorem (2.5). *A surface X_0 as in* (2.4) *exists if and only if the cusp singularity D is smoothable.*

The main point of the proof will be to show that the existence of X_0 as in (2.4) implies that the variety with (local) normal crossings X_0 is smoothable, in such a way that the total space of the smoothing X is itself smooth. This part of the proof will consist almost entirely of references to [1], where the corresponding statement is proved when *all* components V_i of X_0 are rational. Since our variety X_0 has one component V_0 which is an Inoue-Hirzebruch surface, some vary minor modifications in the argument must be made. We will only write down the objects involved in the proof and state the special lemmas needed in our case, indicating where necessary the differences in the analysis.

Let $T^0_{X_0}$ and $T^1_{X_0}$ be the cotangent sheaves of Lichtenbaum-Schlessinger and $\mathbb{T}^0_{X_0}$, $\mathbb{T}^1_{X_0}$ their global counterparts. In our case,

$$T^i_{X_0} = \underline{\mathrm{Ext}}^i(\Omega^1_{X_0}, \mathcal{O}_{X_0}) \quad \text{and} \quad \mathbb{T}^i_{X_0} = \mathrm{Ext}^i(\Omega^1_{X_0}, \mathcal{O}_{X_0}),$$

where $\Omega^1_{X_0}$ is the sheaf of Kähler differentials on X_0.

A variety with local normal crossings X_0 with singular locus $Q \subsetneq X_0$ is said to be *d-semi-stable* if $T^1_{X_0} = \mathcal{O}_Q$. For a surface, this condition implies the triple point formula, which is topological in nature, but has more subtle analytic consequences as well.

Lemma (2.6). *If an X_0 exists as in* (2.4), *then an X'_0 exists, with the same Hirzebruch-Inoue component and double curve D_0, which is d-semi-stable.*

Proof. Identical to that of (5.14) of [1].

If X_0 is a d-semi-stable surface, and $n : \tilde{X}_0 \to X_0$ is the normalization map, then by [1], (3.2) and (3.5), there is an intrinsically defined subsheaf

$$\Lambda^1_{X_0} \subseteq n_* \Omega^1_{\tilde{X}_0}(\log \tilde{Q})$$

(where $\tilde{Q}$ is the normalization of Q) and a resolution

$$0 \to \Omega^1_{X_0}/\tau_{X_0} \to \Lambda^1_{X_0} \to n_* \mathcal{O}_{\tilde{Q}} \to n_* \mathcal{O}_{\tilde{T}} \to 0$$

(where $\tilde{T} = T$ is the set of the triple points of X_0). Here τ_{X_0} is the torsion part of $\Omega^1_{X_0}$, and the natural map

$$\check{\Omega}^1_{X_0} \to \Omega^1_{X_0}/\tau_{X_0}$$

is an isomorphism. The role of $\Lambda^1_{X_0}$ in deformation theory is explained by the following exact sequences. Choose a generating section $\xi \in H^0(T^1_{X_0})$ and, via Lie bracket, consider the map $[\cdot, \xi] : T^0_{X_0} \to T^1_{X_0}$. Then

$$0 \to S_{X_0} \to T^0_{X_0} \xrightarrow{[\cdot, \xi]} T^1_{X_0} \to 0$$

is exact, where $\check{S}_{X_0} = \Lambda^1_{X_0}$.

Lemma (2.7). *With X_0 as in* (2.4), *and d-semi-stable, $H^0(X_0, \Lambda^1_{X_0}) = 0$.*

Proof. Identical to that of (5.9) of [1], starting with V_0 and the negative definite cycle D_0 of double curves on it in place of the non-hexagonal component used in (5.9) of [1].

Lemma (2.8).

(2.8.1) $H^2(T^0_{X_0})=0$.

(2.8.2) The natural map $\mathbb{T}^1_{X_0}\to H^0(T^1_{X_0})$ is surjective.

(2.8.3) The natural map $H^1(T^0_{X_0})\otimes H^0(T^1_{X_0})\to H^1(T^1_{X_0})$ is surjective.

Proof. First, using the resolution [1, (1.5)]

$$0\to\Omega^1_{X_0}/\tau_{X_0}\to n_*\Omega^1_{\tilde{X}_0}\to n_*\Omega^1_{\tilde{Q}}\to 0$$

and $H^0(V_0,\Omega^1_{V_0})=0$ (1.5.3), we obtain $H^0(\Omega^1_{X_0}/\tau_{X_0})=0$. By Serre duality, $H^2(T^0_{X_0})$ is dual to $H^0((\Omega^1_{X_0}/\tau_{X_0})\otimes\omega_{X_0})$, where ω_{X_0} is the dualizing sheaf. But, by contruction,

$$n^*\omega_{X_0}|_{V_0}=\mathcal{O}_{V_0}(-D')$$

and

$$n^*\omega_{X_0}|_{V_i}=\mathcal{O}_{V_i}\quad\text{for}\quad i>0,$$

and hence

$$H^0(F\otimes\omega_{X_0})\subseteq H^0(F)$$

for any torsion-free sheaf F. Thus $H^2(T^0_{X_0})=0$, implying (2.8.1). The Ext spectral sequence immediately gives (2.8.2). As for (2.8.3), as in (5.9) of [1], we must show that $H^2(S_{X_0})=0$, or equivalently, $H^0(\Lambda^1_{X_0}\otimes\omega_{X_0})=0$. Again, this is immediate from the vanishing of $H^0(\Lambda^1_{X_0})$. Q.E.D.

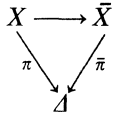
Lemma (2.9). *There exists a smooth threefold X and a proper flat map $\pi:X\to\Delta$ such that $\pi^{-1}(0)=X_0$ (as schemes).*

Proof. The proof of (5.10) of [1] applies, essentially unchanged.

Remark (2.10). The general fiber X_t of π is a rational surface.

The existence part of (2.5) now follows froma lemma due to Sheperd-Barron [9]:

Lemma (2.11). *The divisor $\sum_{i\geq 1} V_i$ of X is contractible, and, if $\bar{X}$ denotes the contraction, there is a commutative diagram*



Moreover, the map $\bar{\pi}:\bar{X}\to\Delta$ is flat, and exhibits a smoothing of $\bar{V}_0$ (which is V_0 contracted along $D_0=D$).

Remark (2.12). This lemma follows from the consideration of the map $f:X_0\to\bar{V}_0$ defined by: $f|_{V_0}$ is the contraction of D, and $f|_{V_i}$, $i>0$, is the map sending V_i to the singular point p of $\bar{V}_0$. One checks that $R^if_*\mathcal{O}_{X_0}=0$ for $i>0$, and uses standard results from deformation theory.

For the converse part of (2.5), assume that the cusp D is smoothable. Let $\bar{V}_0$ be the Inoue-Hirzebruch surface containing D, with D contracted to the cusp singularity p. By (2.1), there is a flat family $\bar{\pi}:\bar{X}\to\Delta$ with $\bar{X}_t$ smooth and

$\bar{V}_0 = \bar{\pi}^{-1}(0)$. By blowing up the singular point p of $\bar{V}_0$ in $\bar{X}$ and by making the necessary base changes, arrange $\pi : X \to \Delta$ semi-stable, with one component of X_0 being the resolution V_0 of $\bar{V}_0$. One can run through the argument of Kulikov and Persson-Pinkham, applied to the components of X_0 which occur with strictly maximal multiplicity in K_X; the point here is that the proper transform V_0 of $\bar{V}_0$ is a resolution of a minimally elliptic singularity with reduced fundamental cycle in the minimal resolution, and hence can never occur with strictly maximal multiplicity. Thus, by Kulikov's analysis, we may assume X_0 satisfies (2.4.2)–(2.4.6).

Now the dual graph Γ triangulates a closed surface. Let $e(\Gamma)$ be the Euler number. Recalling that X_t is rational and $q(V_0) = 1$, an easy argument shows that

$$1 = \chi(\mathcal{O}_{X_t}) = \chi(\mathcal{O}_{X_0}) = e(\Gamma) - 1,$$

so that $e(\Gamma) = 2$ and Γ triangulates S^2, so that X_0 satisfies (2.4.1) also. Q.E.D.

Remark (2.13). If D sits on a rational surface as an anticanonical divisor (which will not always be the case), we can play the same game as above with the appropriate rational surface S replacing the Inoue-Hirzebruch surface V_0. In this case, the smoothings will be K3 surfaces.

The following lemma shows that, in (2.4), it suffices to construct the disjoint collection of surfaces $\{\tilde{V}_i\}$, subject to the appropriate combinatorial restrictions.

Lemma (2.14). *Let $\{\tilde{V}_i\}$ be a collection of smooth surfaces, $\{D_{ij}\}$ a collection of smooth curves on $\tilde{V}_i$ and $\{\varphi_{ij} : D_{ij} \to D_{ji}\}$ a collection of isomorphisms satisfying:*

(2.14.0) $D_i = \bigcup_j D_{ij}$ *is a divisor with normal crossings,*

(2.14.1) $\varphi_{ij}^{-1} = \varphi_{ij}$,

(2.14.2) $p_{ijk} \in D_{ij} \cap D_{ik} \Leftrightarrow \varphi_{ij}(p_{ijk}) \in D_{ji} \cap D_{jk}$,

(2.14.3) $\varphi_{ik} \circ \varphi_{ij}(p_{ijk}) = \varphi_{ik}(p_{ijk})$.

Then there is a unique structure of a variety with normal crossings on the topological space

$$X_0 = \coprod \tilde{V}_i / x \in D_{ij} \sim \varphi_{ij}(x)$$

which is compatible with the inclusions $\tilde{V}_i \subseteq X_0$. A similar statement is true if the φ_{ij} are arranged to give, topologically, only local normal crossings.

Proof. We shall only sketch the proof of this elementary statement. As the question is local, assume $X_0 = \tilde{V}_i \cup \tilde{V}_j \cup \tilde{V}_k$ and let $n : \coprod \tilde{V}_i = \tilde{X}_0 \to X_0$ be the obvious map. With D_{ij}, p_{ijk} as in (2.14), we define a subsheaf

$$\mathcal{O}_{X_0} \subseteq n_* \mathcal{O}_{\tilde{X}_0}$$

by the recipe: if sections of $n_* \mathcal{O}_{\tilde{X}_0}$ at p are of the form $f = (f_1, f_2, f_3)$, $f \in \mathcal{O}_{X_0, p} \Leftrightarrow f_s \circ \varphi_{rs}(x) = f_r(x)$, $x \in D_{rs}$ $(r, s \in \{i, j, k\})$.

Alternatively, the sections of $\mathcal{O}_{X_0}$ are continuous functions on X_0 which are holomorphic on $\tilde{V}_i$. We leave to the reader the local calculation which identifies the ringed space $(X_0, \mathcal{O}_{X_0})$ locally with the germ of $\{(z_1, z_2, z_3) \in \mathbb{C}^3 ; z_1 z_2 z_3 = 0\}$. Q.E.D.

Combining (2.5) with (2.14), we obtain the following corollary to (2.5):

Proposition (2.15). *Let $D=D_1+\ldots+D_r$ be a smoothable cusp, with $-(D_1^2)\geqq 3$ and $-(D_j)^2\geqq 3$ for some $j\neq 1$. Then $C=C_1+\ldots+C_r$ is smoothable, where $(C_1^2)=(D_1^2)+1$ and $(C_i^2)=(D_i^2)$, $i\geqq 2$.*

Proof. By the argument for the converse half of (2.5), there is a semistable degeneration $\pi: X\to\Delta$ with $X_0=\pi^{-1}(0)$ as in (2.4) for the cusp D. Let V_0 be the Inoue surface and V_1 the component of X_0 meeting V_0 along D_1. Replace V_0 by $V_0'=$ the Inoue surface for C, and V_1' by V_1 blown up at a smooth point of $D_1\subsetneqq V_1$ (with obvious modifications if $V_1=V_0$). If $V_i'=V_i$ for $i\neq 0,1$, then the collection $\{V_i'\}$ yields the correct combinatorial configuration for the cusp C in (2.5). Q.E.D.

Note that (2.15) is not an adjacency relation between cusps. It would be amusing to verify directly that the dual of C sits as an anticanonical divisor on a rational surface.

3. Charge

Let V be a smooth surface, and let D be a reduced effective cycle of rational curves on V. (D need not be negative definite here.) Let $r(D)$ be the number of irreducible components of D; $r(D)\geqq 1$ and $r(D)=1$ when D is a single irreducible nodal curve.

Definition (3.1). Let V and D be as above. The *charge* of (V,D), denoted by $Q(V,D)$, is the quantity

$$Q(V,D)=12-D^2-r(D).$$

If we write

$$D=\sum_{j=1}^{r(D)} D_j,$$

then

$$Q(V,D)=\begin{cases} 11-D_1^2 & \text{if } r(D)=1 \\ 12-\sum_{j=1}^{r(D)}[(D_j^2)+3] & \text{if } r(D)\geqq 2. \end{cases}$$

We will mainly be concerned with the case where V is rational and D is an anticanonical divisor of V, i.e., (V,D) is an anticanonical pair. In this case

$$Q(V,D)=12-K_V^2-r(D).$$

Notice that $Q(V,D)$ depends only upon $r(D)$ and the D_j^2's, so that if we represent the divisor D by the cycle of integers $(d_1,\ldots,d_r)$, where $r=r(D)$ and $d_j=-D_j^2$, then $Q(V,D)$ depends only on the cycle of integers. We will sometimes use $Q(D)$ for $Q(V,D)$ when no confusion can result.

It will be useful to understand the anticanonical pairs (V,D) where V is a minimal model $\mathbb{P}^2$ or $\mathbb{F}_N$; the proof of the following lemma is left to the reader.

Lemma (3.2). *Let (V, D) be an anticanonical pair, with V isomorphic to $\mathbb{P}^2$ or $\mathbb{F}_N$. Then (V, D) is one of the following:*

V	$(-d_1, \ldots, -d_r)$	$Q(V, D)$
$\mathbb{P}^2$	$(1, 1, 1)$	0
$\mathbb{P}^2$	$(1, 4)$	1
$\mathbb{P}^2$	(9)	2
$\mathbb{F}_N$	$(-N, 0, N, 0)$	0
$\mathbb{F}_N$	$(-N, 0, N+2)$	1
$\mathbb{F}_N$	$(-N, N+4)$	2
$\mathbb{F}_0, \mathbb{F}_2$	$(2, 2)$	2
$\mathbb{F}_1$	$(1, 3)$	2
$\mathbb{F}_1$	$(1, 0, 1)$	1
$\mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2$	(8)	3

The behavior of Q under blowups and blowdowns is given by the following two elementary lemmas.

Lemma (3.3). *Let (V, D) be an anticanonical pair, let p be a point on D, and let $\pi : \tilde{V} \to V$ be the blowup at p, with exceptional divisor E.*

(3.3.1) *If p is a smooth point of D, and $\tilde{D}$ is the proper transform of D on $\tilde{V}$, then $(\tilde{V}, \tilde{D})$ is an anticanonical pair, and $Q(\tilde{V}, \tilde{D}) = Q(V, D) + 1$.*

(3.3.2) *If p is a double point of D, and $\tilde{D}$ is the proper transform of D on $\tilde{V}$, then $(\tilde{V}, \tilde{D} + E)$ is an anticanonical pair, and $Q(\tilde{V}, \tilde{D} + E) = Q(V, D)$.*

Lemma (3.4). *Let (V, D) be an anticanonical pair. Let E be an exceptional curve on V, let $\pi : V \to \bar{V}$ be the blowdown of E, and let $\bar{D} = \pi_*(D)$. Then either*

(3.4.1) *E is not a component of D, there is a unique component D_j of D which meets E once transversally, and $Q(\bar{V}, \bar{D}) = Q(V, D) - 1$ or*

(3.4.2) *E is a component of D and $Q(\bar{V}, \bar{D}) = Q(V, D)$.*

Moreover, in both cases $(\bar{V}, \bar{D})$ is an anticanonical pair.

The proofs of the above are trivial consequences of the definition of Q, the blowup formula for the canonical bundle, and the equation

$$E \cdot K_V = -1$$

where E is an exceptional curve on V; we leave them to the reader.

Lemma (3.5). *Let (V, D) be an anticanonical pair. Then*

$$Q(V, D) \geqq 0.$$

Proof. By Lemma (3.3), the charge Q cannot decrease upon blowing up; so it suffices to prove the result when V is a minimal model $\mathbb{P}^2$ or $\mathbb{F}_N$. For this the classification of Lemma (3.2) suffices. Q.E.D.

A glance at the list of Lemma (3.2) shows that if V is a minimal model, and (V, D) is an anticanonical pair, then $0 \leqq Q(V, D) \leqq 3$; hence, by Lemma (3.3), if (V, D) is an arbitrary anticanonical pair, then $Q(V, D)$ is approximately equal to the number of blowups made at smooth points of the anticanonical cycles between V and a minimal model. This can be made more precise; we will only need the following relatively crude estimate:

Lemma (3.6). *Let (V, D) be an anticanonical pair, and let $E_1, \ldots, E_k$ be a pairwise disjoint set of k exceptional curves on V which are not components of D. Then $Q(V, D) \geqq k$.*

Proof. Since the E_i's are all disjoint, they can all be blown down, to obtain an anticanonical pair $(\bar{V}, \bar{D})$. By Lemma (3.4), $Q(\bar{V}, \bar{D}) = Q(V, D) - k$, and by Lemma (3.5), $Q(\bar{V}, \bar{D}) \geqq 0$. This gives the result. Q.E.D.

The following is the principle of "conservation of charge" for Type III degenerations of K3 surfaces. Let $X \to \Delta$ be such a degeneration, and write the special fiber X_0 as

$$X_0 = \sum_i V_i.$$

Let $\tilde{V}_i$ be the normalization of V_i, let D_i be the double curve of X_0 on V_i, and let $\tilde{D}_i$ be the inverse image of D_i on $\tilde{V}_i$. [The pairs (V_i, D_i) and $(\tilde{V}_i, \tilde{D}_i)$ are equal unless V_i has self-intersection in X.] Then, by (1.6.1) and (1.6.2), each $(\tilde{V}_i, \tilde{D}_i)$ is an anticanonical pair.

Proposition (3.7). *Let $X \to \Delta$ be a Type* III *degeneration of* K3 *surfaces as above. Then*

$$\sum_i Q(\tilde{V}_i, \tilde{D}_i) = 24.$$

Proof. Let v be the number of components V_i of X_0, e be the number of irreducible double curves D_{ij}, and f be the number of triple points of X_0. Since v is the number vertices, e the number of edges, and f the number of faces of the dual graph Γ of X_0, we have $v - e + f = 2$ by Euler's formula. Let r_i be the number of components of $\tilde{D}_i$. Then $\sum_i r_i = 2e$ since each double curve lies on two surfaces. Moreover, since Γ is a triangulation, $2e = 3f$. Now compute:

$$\begin{aligned}
\sum_{i=1}^{v} Q(\tilde{V}_i, \tilde{D}_i) &= \sum_{i=1}^{v} (12 - K_{\tilde{V}_i}^2 - r_i) \\
&= \sum_{\substack{i \text{ such that} \\ r_i = 1}} (11 - D_{ij_i}^2) + \sum_{\substack{i \text{ such that} \\ r_i \geqq 2}} \left(12 - \sum_{k=1}^{r_i} [\tilde{D}_{ij_k}^2 + 3]\right) \\
&= 12v - \sum_{i,j} \tilde{D}_{ij}^2 - 3\sum_i r_i + 2m,
\end{aligned}$$

where m is the number of components V_i with $r_i = 1$.

The quantity $\sum_{i,j} \tilde{D}_{ij}^2$ counts both terms $\tilde{D}_{ij}^2$ and $\tilde{D}_{ji}^2$

and so

$$\sum_{i,j} \tilde{D}_{ij}^2 = \sum_{\substack{\text{double curves} \\ D_{ij} \subset X_0}} \tilde{D}_{ij}^2 + \tilde{D}_{ij}^2 = -2e + 2m$$

by (1.6.3). Hence

$$\begin{aligned} \sum_i Q(\tilde{V}_i, \tilde{D}_i) &= 12v + 2e - 3(2e) \\ &= 12v - 4e \\ &= 12(2 + e - f) - 4e \\ &= 24 + 8e - 12f \\ &= 24. \quad \text{Q.E.D.} \end{aligned}$$

Remarks (3.8).

(3.8.1) If one component V_0 of X_0 is a Hirzebruch-Inoue surface, and X_0 is as in (2.4), then the same calculation as above shows that in this case

$$\sum_i Q(\tilde{V}_i, \tilde{D}_i) = 24$$

also. In particular, if $Q(V_0, D_0)$ is high, then severe restrictions are placed on the other components V_i, by Lemma (3.6).

(3.8.2) A calculation similar to the above can be made for a Type III degeneration of any $\kappa = 0$ surface. The result is:

$$\sum_i Q(\tilde{V}_i, \tilde{D}_i) = 12e(\Gamma).$$

where $e(\Gamma)$ is the topological Euler characteristic of the dual graph Γ.

(3.8.3) Using only Euler's formula, one can easily prove the formula

$$\sum_i (6 - r_i) = 12$$

for a Type III degeneration as above, which can sometimes be useful.

4. Cusps with Rational Duals

In this section, we will primarily be concerned with the case where D is a negative definite cycle of rational curves on a surface V, i.e., D is the resolution of a cusp singularity.

Lemma (4.1).

(4.1.1) *Let D be the cusp $\begin{pmatrix} a_i \\ b_i \end{pmatrix}$, then*

$$Q(V, D) = \begin{cases} 11 + a_1 & \text{if} \quad r(D) = 1 \\ 12 + \sum_i (a_i - b_i - 3) & \text{if} \quad r(D) \geqq 2. \end{cases}$$

(4.1.2) *If D' is the dual to D, then*

$$Q(V', D') = \begin{cases} 13 - a & \text{if } D = \begin{pmatrix} a \\ 0 \end{pmatrix}, \quad a \geqq 1 \\ 12 + b & \text{if } D = \begin{pmatrix} 3 \\ b \end{pmatrix}, \quad b \geqq 1 \\ 12 + \Sigma(b_i - a_i + 3) & \text{if } D = \begin{pmatrix} a_i \\ b_i \end{pmatrix} \text{ otherwise.} \end{cases}$$

Hence in all cases we have $Q(D) + Q(D') = 24$.

Proof. These statements follow immediately from Lemma (1.2), Definition (1.3), Lemma (1.4), and the Definition (3.1) of $Q(D)$. Q.E.D.

Remark (4.2). The above formula $Q + Q' = 24$ and the formula of Proposition (3.7) suggest that there may be a relationship between the combinatorics of an X_0 as in (2.4) (with one compact a Hirzebruch-Inoue surface with the cusp D) and the dual cusp D'. Looijenga's conjecture (2.3) is only qualitative; one might hope for some more precise correlation, given Theorem (2.5).

Lemma (4.3). *Let (V, D) be an anticanonical pair, and assume that D is negative definite, so that D is the resolution of a cusp singularity. Then $Q(V, D) \geqq 3$.*

Proof. Since V is a rational surface, the rank of the Neron-Severi group $NS(V)$ of V is $10 - K_V^2 = 10 - D^2$. Since D is negative definite, and $NS(V)$ contains a positive class, we must have $r(D) \leqq \operatorname{rank} NS(V) - 1$, or $9 - D^2 - r(D) \geqq 0$. Hence $Q(V, D) = 12 - D^2 - r(D) \geqq 3$. Q.E.D.

Definition (4.4). Let D be a cusp. We say that D is *rational* if there exists a rational surface V on which D is an anticanonical divisor, i.e., such that (V, D) is an anticanonical pair. If D is a cusp such that the dual D' to D is rational, we say that D *has a rational dual.*

Theorem (4.5). 1) *If D has a rational dual, $Q(D) \leqq 21$.*

2) *Conversely, let D be a cusp with $r(D) \leqq 3$ and $Q(D) \leqq 21$. Then D has a rational dual, except in the following cases:* (4, 11), (7, 8), (2, 4, 12), (2, 8, 8), (3, 3, 12), (3, 4, 11), (3, 7, 8), (4, 4, 10), (4, 6, 8), (4, 7, 7), (5, 5, 8).

We break the proof of (4.5) up into several steps.

Lemma (4.6). *If D has a rational dual, then $Q(D) \leqq 21$.*

Proof. Combine (4.1.2) with Lemma (4.3). Q.E.D.

Remark. This lemma follows from a theorem of Wahl [11], where the condition $Q(D) \leqq 21$ is restated as $m(D) \leqq r(D) + 9$.

We will now classify those cusps D with $r(D) \leqq 3$ which have rational duals. We begin with the $r = 1$ case.

Proposition (4.7). *Let D be a cusp with $r(D) = 1$. Then D has a rational dual if and only if $Q(D) \leqq 21$.*

Proof. Since $r(D)=1$, $D=\begin{pmatrix}a\\0\end{pmatrix}$ for some $a\geqq 1$. In this case $Q(D)=11+a$, so $Q(D)\leqq 21$ when $a\leqq 10$. Assume $a\leqq 10$. If $a=1$, $D'=\begin{pmatrix}1\\0\end{pmatrix}$, which is clearly rational: blow up 10 smooth points on a nodal cubic in $\mathbb{P}^2$ to obtain (V', D'). If $a\geqq 2$, then $D'=\begin{pmatrix}3\\a-1\end{pmatrix}$. To construct a rational surface V' with D' as an anticanonical divisor, take a nodal cubic C in the plane and blow up the node $a-1$ times; this produces a cycle of a curves $C_0, \ldots, C_{a-1}$, where C_0 is the proper transform of C and $C_1, \ldots, C_{a-1}$ are the $a-1$ exceptional curves. We have $(C_0^2)=7-a$, $(C_1^2)=-1$, and $(C_i^2)=-2$ for $i\geqq 2$. Finally, blow up one smooth point on C_1 and $10-a$ smooth points on C_0. This gives $\begin{pmatrix}3\\a-1\end{pmatrix}$, with the proper transform of C_0 having self-intersection -3. Q.E.D.

The situation with cusps of length two is not so simple:

Proposition (4.8). *Let D be a cusp with $r(D)=2$.*

(i) *If $Q(D)\leqq 20$, then D has a rational dual.*

(ii) *If $Q(D)=21$, then D is one of the following cusps:* (2, 13), (3, 12), (4, 11), (5, 10), (6, 9), (7, 8).

(iii) *The cusps* (2, 13), (3, 12), (5, 10), *and* (6, 9) *have rational duals.*

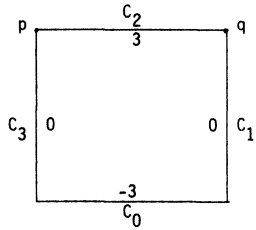
(iv) *The cusps* (4, 11) *and* (7, 8) *do not have rational duals.*

Proof. First notice that $r(D)=2$ if and only if D is either $\begin{pmatrix}a\\1\end{pmatrix}$ or $\begin{pmatrix}a_1 & a_2\\0 & 0\end{pmatrix}$. If $D=\begin{pmatrix}a\\1\end{pmatrix}$, then $Q(D)=8+a$; if $D=\begin{pmatrix}a_1 & a_2\\0 & 0\end{pmatrix}$, then $Q(D)=6+a_1+a_2$. This proves (ii).

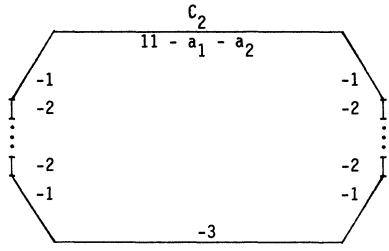
Assume $D=\begin{pmatrix}a\\1\end{pmatrix}$, with $4\leqq a\leqq 13$. Then $D'=\begin{pmatrix}4\\a-3\end{pmatrix}$. To produce an anticanonical pair (V', D'), we argue as in the proof of the previous proposition. Take C to be a nodal cubic in $\mathbb{P}^2$, and blow up the node $a-3$ times. This gives a cycle $C_0, C_1, \ldots, C_{a-3}$, with C_0 the proper transform of C, $(C_0^2)=9-a$, $(C_1^2)=-1$, and $(C_i^2)=-2$, for $i\geqq 2$. Now blow up one smooth point on C_1 and $13-a$ points on C_0, to get D'.

If $D=\begin{pmatrix}3\\1\end{pmatrix}$, $D'=\begin{pmatrix}4\\0\end{pmatrix}$, which can be obtained from $\mathbb{P}^2$ by blowing up 13 smooth points on a nodal cubic.

Assume now that $D=\begin{pmatrix}a_1 & a_2\\0 & 0\end{pmatrix}$, so that $D'=\begin{pmatrix}3 & 3\\a_1-3 & a_2-3\end{pmatrix}$, and $Q(D)\leqq 20$, so that $a_1+a_2\leqq 14$. Assume first that a_1 and a_2 are both at least 4. Start with $\mathbb{F}_3$ and the anticanonical square

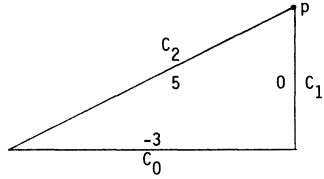


Blow up a_1-4 times at p and a_2-4 times at q to produce

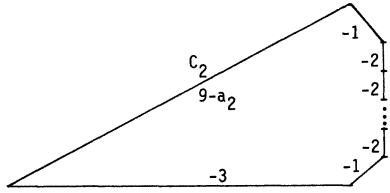


where the two fibers of the original ruling in this anticanonical cycle have lengths a_1-3 and a_2-3. (If either a_i is equal to 4, no blowups are made and the fiber is smooth.) Now blow up one point on each of the (-1)-curves in the fibers [or two points on the (0)-curve if $a_i=4$] and also blow up the proper transform of the original positive section C_2 $14-a_1-a_2$ times at smooth points. This produces (V', D').

Assume secondly that $a_1=3$ and $4\leqq a_2\leqq 12$. Start with $\mathbb{F}_3$ and the anticanonical triangle



Blow up a_2-4 times at p as above to produce



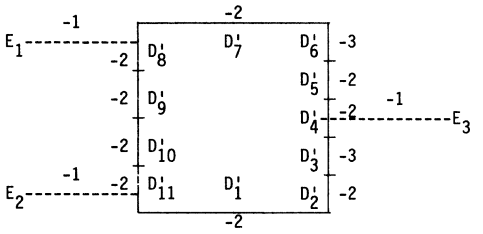
Finally, blow up one point on each of the (-1) curves in the displayed fiber [or two points on the (0)-curve if $a_2=4$] and also blow up the proper transform of C_2 $12-a_2$ times. This produces (V', D'). Notice that this construction also exhibits a rational dual for the $Q=21$ cusp $\begin{pmatrix}3 & 12\\ 0 & 0\end{pmatrix}$.

Finally, assume $a_1=a_2=3$. Then $D'=\begin{pmatrix}3 & 3\\ 0 & 0\end{pmatrix}$ and can be obtained from the anticanonical pair ($\mathbb{P}^2$, line + conic) by blowing up smooth points.

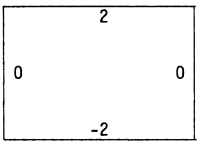
This proves (i).

To finish the proof of (iii), we must exhibit the rationality of the duals of $\begin{pmatrix}5 & 10\\ 0 & 0\end{pmatrix}$ and $\begin{pmatrix}6 & 9\\ 0 & 0\end{pmatrix}$. For these we will simply draw the surface V' on which D' sits, with the exceptional curves required to blow the surface down to a minimal model.

For $\begin{pmatrix} 5 & 10 \\ 0 & 0 \end{pmatrix}$: $D' = \begin{pmatrix} 3 & 3 \\ 2 & 7 \end{pmatrix}$

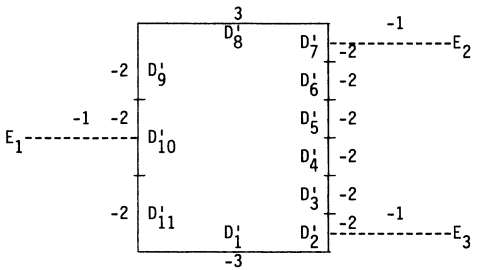


Blow down E_1, E_2, E_3, D'_3, D'_4, D'_5, D'_6, D'_8, D'_9, and D'_{10} to obtain

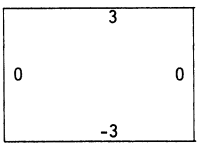


on $\mathbb{F}_2$.

For $\begin{pmatrix} 6 & 9 \\ 0 & 0 \end{pmatrix}$: $D' = \begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$

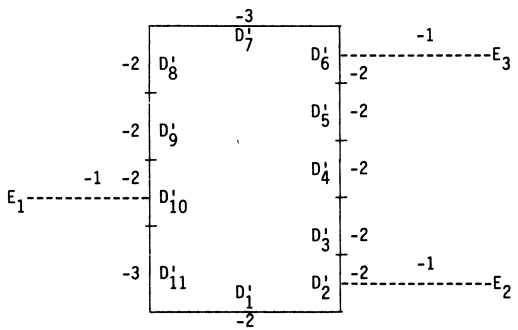


Blow down E_1, E_2, E_3, D'_3, D'_4, D'_5, D'_6, D'_7, D'_9, and D'_{10} to obtain

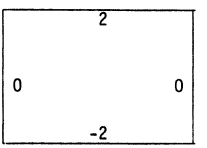


on $\mathbb{F}_3$

or, for $\begin{pmatrix} 6 & 9 \\ 0 & 0 \end{pmatrix}$:



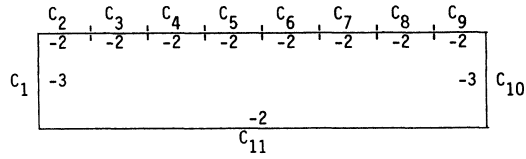
Blow down E_1, E_2, E_3, D'_3, D'_4, D'_5, D'_6, D'_8, D'_9, and D'_{10} to obtain



on $\mathbb{F}_2$.

This proves (iii).

Unfortunately, our proof of (iv) is rather ad hoc. Let us first take up the case of the $\begin{pmatrix} 4 & 11 \\ 0 & 0 \end{pmatrix}$ cusp, whose dual is the cusp $\begin{pmatrix} 3 & 3 \\ 1 & 8 \end{pmatrix}$:



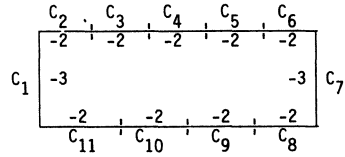
Assume this cusp is rational. Then there exists an exceptional curve E on the surface V; visibly E is not a component of the cusp, so E meets a unique C_i transversally in one point. Let us analyze the possibilities for C_i.

If E meets C_1 or C_{10}, then blowing down E produces a rational cusp with $Q=2$; this is impossible, by Lemma (4.3).

If E meets C_{11}, blowing down E and then C_{11} produces a cycle of 10 (-2)-curves. This is not minimal, so there must be a second exceptional curve F meeting one of the 10 curves. Then one can blow down F and 9 of the 10 (-2)-curves; however, this produces a nodal curve C with $(C^2)=10$. Since $C\in|-K|$, the resulting rational surface has $K^2=10$, a contradiction.

Hence E must meet C_2, C_3, ..., or C_9; by symmetry we may assume C meets C_2, C_3, C_4, or C_5. We may further assume that E meets C_j, $2\leqq j\leqq 5$, and that there is no exceptional curve on the surface meeting C_i, for $2\leqq i<j$. If so, we may blow down E, C_j, C_{j+1}, ..., C_9, producing a cycle consisting of the images of the curves C_{10}, C_{11}, C_1, C_2, ..., C_{j-1}. This is a non-minimal cycle by Lemma (3.2); however, by assumption, there are no exceptional curves meeting $C_2, \ldots, C_{j-1}$ and by the previous argument, there are none meeting C_1, C_{10}, or C_{11}. This contradiction proves that $\begin{pmatrix} 3 & 3 \\ 1 & 8 \end{pmatrix}$ is not rational.

Finally, assume that $\begin{pmatrix} 7 & 8 \\ 0 & 0 \end{pmatrix}$ has a rational dual, i.e., that $\begin{pmatrix} 3 & 3 \\ 4 & 5 \end{pmatrix}$ is rational:



Let E be an exceptional curve on the surface V which contains $\begin{pmatrix} 3 & 3 \\ 4 & 5 \end{pmatrix}$; as above, E must meet some component C_i of the cusp.

If E meets C_1 or C_7, blowing down E produces a cusp with $Q=2$, contradicting Lemma (4.3).

If E meets C_8, then $C_7+3C_8+2C_9+C_{10}+3E$ form a fiber of a ruling on V, for which C_6 and C_{11} must be sections. Hence C_1, C_2, C_3, C_4, and C_5 are components of a fiber of this ruling. They do not support a complete fiber, so there is a second exceptional curve F on V meeting one of these five curves. By the above, F meets

C_2, C_3, C_4, or C_5. If F meets C_2, then $C_1+3C_2+2C_3+C_4+3F$ is a full fiber, not containing C_5, a contradiction. If F meets C_3, $C_2+2C_3+C_4+2F$ is a full fiber, and if F meets C_4, $C_3+2C_4+C_5+2F$ is a full fiber, which are contradictions. Hence F must meet C_5. But now C_1, C_2, C_3, C_4, C_5, and F still do not support a complete fiber, so there must be a third exceptional curve G; by the above, G must also meet C_5. Now $F+G+C_5$ is a full fiber, a contradiction. Hence E does not meet C_8, and by symmetry, E does not meet C_{11} either.

If E meets C_9, then $C_8+2C_9+C_{10}+2E$ form a fiber of a ruling on V, for which C_7 and C_{11} are sections. Hence C_1, C_2, C_3, C_4, C_5, and C_6 are components of a fiber of this ruling. They do not support a full fiber, so there is an exceptional curve F, which by the previous argument cannot meet C_1. Using an analysis as above, one can easily see that if F meets C_2, C_3, C_4, or C_5, there is a full fiber supported on only a proper subset of the components; hence F must meet C_6. But, as above, C_1, C_2, C_3, C_4, C_5, C_6, and F still do not support a full fiber, so there is a third exceptional curve G, which must then meet C_6. Now $F+G+C_6$ forms a full fiber, a contradiction. Hence E cannot meet C_9 or, by symmetry, C_{10}.

Now the argument proceeds as for the previous cusp. We may assume that E meets C_j, with $2 \leqq j \leqq 4$, and that there is no exceptional curve meeting C_7, C_8, C_9, C_{10}, $C_1, \ldots, C_{j-1}$. If so, we may blow down $C_j, C_{j+1}, \ldots, C_6$ to produce a cycle consisting of C_7, C_8, C_9, C_{10}, $C_1, \ldots, C_{j-1}$; by Lemma (3.2), this cycle is not minimal, but by assumption there are no exceptional curves. This contradiction proves that $\begin{pmatrix} 3 & 3 \\ 4 & 5 \end{pmatrix}$ is not rational and completes the proof of Proposition (4.8). Q.E.D.

Finally, we have the following result for cusps of length three. Its proof is similar to that of the previous proposition, and we leave it to the reader.

Proposition (4.9). *Let D be a cusp with $r(D)=3$.*

(i) *If $Q(D) \leqq 20$, then D has a rational dual.*

(ii) *If $Q(D)=21$, then D has a rational dual if and only if D is one of the following:* (2, 2, 14), (2, 3, 13), (2, 5, 11), (2, 6, 10), (2, 7, 9), (3, 5, 10), (3, 6, 9), (4, 5, 9), (5, 6, 7), (6, 6, 6).

(iii) *The $Q=21$, $r=3$ cusps which do not have rational duals are the following:* (2, 4, 12), (2, 8, 8), (3, 3, 12), (3, 4, 11), (3, 7, 8), (4, 4, 10), (4, 6, 8), (4, 7, 7), (5, 5, 8).

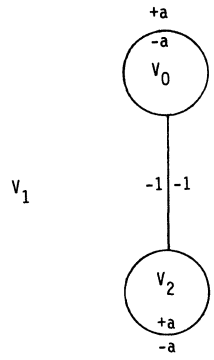
5. The Special Fibers of Smoothings of Cusps of Small Length

In this section, we will exhibit smoothings of all the cusp singularities of length one and two which have rational duals, verifying Looijenga's conjecture in these cases. We will also be able to smooth some cusps of length three. The method will be that outlined in Sect. 2; we will construct a Type III degeneration of rational surfaces whose special fiber contains one component (the Inoue surface) which has the resolution of the cusp to be smoothed on it, as the double curve.

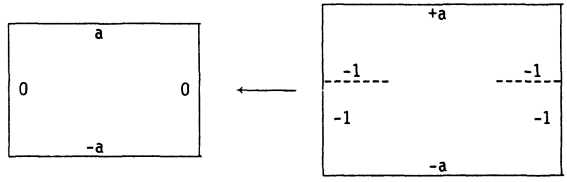
If V_i is a component of the special fiber X_0 of a given Type III degeneration, we will write $Q(V_i)$ for $Q(\tilde{V}_i, \tilde{D}_i)$ as in (3.7).

The Smoothings of Length one Cusps

By Proposition (4.7), if $r(D)=1$, then D has a rational dual if and only if $(D^2) \geqq -10$. Set $a=-(D^2)$. For $1 \leqq a \leqq 9$, use the following special fiber X_0:

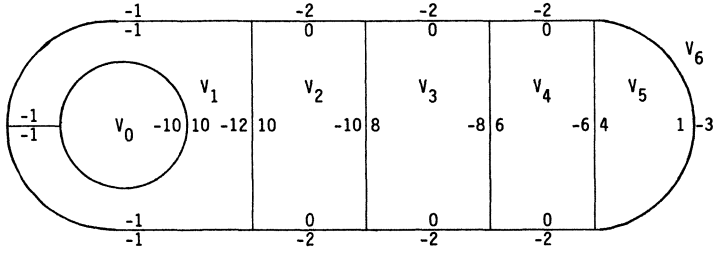


The component V_0 is the Inoue surface with double curve D. The component V_2 is the blowup of a $\mathbb{P}^2$ at $9-a$ points of a nodal cubic. The component V_1 is the blowup of $\mathbb{F}_a$ at two smooth points on different fibers,

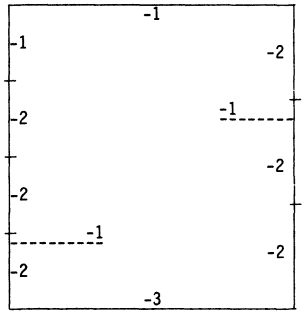


then glued to itself by joining the two (-1)-curves. $Q(V_0)=11+a$, $Q(V_1)=2$, and $Q(V_2)=11-a$ in this degeneration.

For $a=10$, we use the following X_0:



V_0 is the Inoue surface. V_1 is a two-fold blowup of $\mathbb{F}_{10}$ at two points of the negative section, then glued to itself as shown. $V_2 \cong \mathbb{F}_{10}$, $V_3 \cong \mathbb{F}_8$, and $V_4 \cong \mathbb{F}_6$. V_5 is a $\mathbb{P}^2$ with double curves a line and a conic. V_6 is a blowup of $\mathbb{F}_3$,



with the last two non-double exceptional curves shown above. Here $Q(V_0)=21$, $Q(V_5)=1$, and $Q(V_6)=2$.

This completes the analysis of the cusps of length one. Before we discuss those of length two, we need a

Lemma (5.1). *Let (V,D) be an anticanonical pair with $r(D)=2$. Write $D=D_1+D_2$. Then either*

(i) $(D_1^2)+(D_2^2)\leqq 4$, *or*

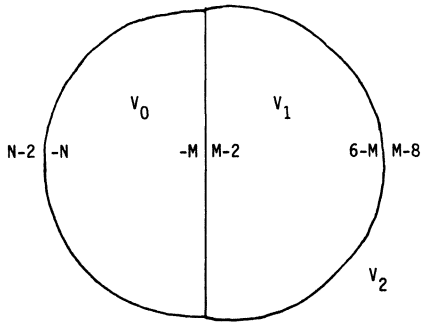
(ii) $(D_1^2)=1$, $(D_2^2)=4$, *and $V\cong\mathbb{P}^2$, with D_1 a line and D_2 a conic.*

Conversely, given two integers a_1 and a_2 such that $a_1+a_2\leqq 4$, there is an anticanonical pair (V,D) with $r(D)=2$, such that $(D_1^2)=a_1$ and $(D_2^2)=a_2$.

Proof. The first statement follows directly from the classification in Lemma (3.2) of the minimal models. The second statement follows from the existence of (V,D) where $r(D)=2$ and $(D_1^2)+(D_2^2)=4$ for any (D_1^2). [If either (D_1^2) or (D_2^2) is $\leqq 0$, use $(\mathbb{F}_N,(-N,N+4))$; if both are positive, use $(\mathbb{F}_1,(1,3))$ or $(\mathbb{F}_0,(2,2))$.] Q.E.D.

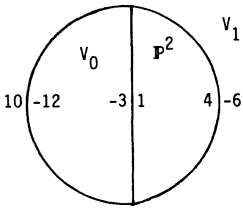
The Smoothings of the Cusps of Length Two

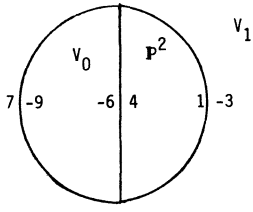
Let us begin with the cusps D with $r(D)=2$ and $Q(D)\leqq 20$, which all have rational duals, by Proposition (4.8). If $D=D_1+D_2$, with $(D_1^2)=-N$ and $(D_2^2)=-M$, then M and N are both at least two, and $Q(D)\leqq 20$ if and only if $N+M\leqq 14$. Use the following X_0:



Here $Q(V_0)=N+M+6$, $Q(V_1)=2$, and $Q(V_2)=16-N-M$. Again V_0 is the Inoue surface. V_1 is $\mathbb{F}_{6-M}$ if $M\geqq 4$ and is $\mathbb{F}_{M-2}$ if $M=2$ or 3. By Lemma (5.1), V_2 exists (as an anticanonical pair, with its two double curves) if $(N-2)+(M-8)\leqq 4$, i.e., $N+M\leqq 14$.

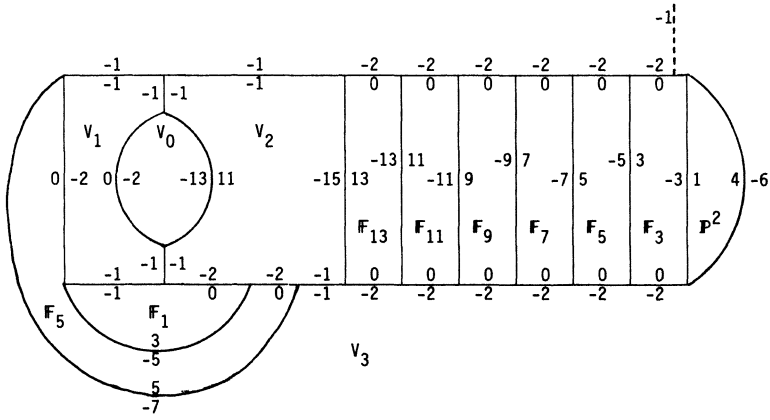
To smooth two of the four remaining $Q=21$ cusps, we may use the same type of degeneration, exploiting the existence of ($\mathbb{P}^2$, line+conic); the cusps $\begin{pmatrix}3 & 12\\ 0 & 0\end{pmatrix}$ and $\begin{pmatrix}6 & 9\\ 0 & 0\end{pmatrix}$ may be smoothed by using the following X_0's:





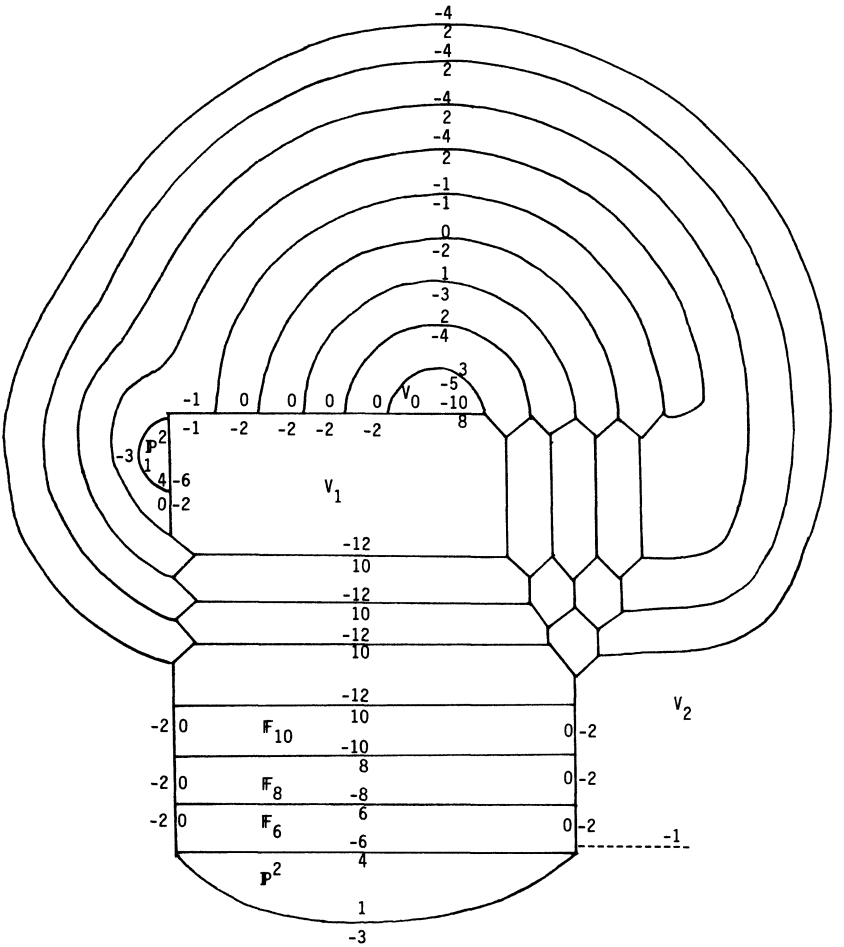
In the $\begin{pmatrix}3 & 12\\ 0 & 0\end{pmatrix}$ case, the surface V_1 is $\mathbb{F}_6$; in the $\begin{pmatrix}6 & 9\\ 0 & 0\end{pmatrix}$ case, the surface V_1 is $\mathbb{F}_3$. Both have $Q=2$.

The last two cases involve much more elaborate special fibers X_0. For the cusp $\begin{pmatrix}13\\ 1\end{pmatrix}$, use the following:

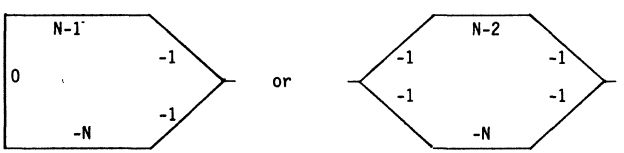


V_0 is the Inoue surface. V_1 is a two-fold blowup of $\mathbb{F}_2$, and V_2 is a four-fold blowup of $\mathbb{F}_{15}$. The other surfaces are as marked, except for V_3; V_3 is a 14-fold blow-up of $\mathbb{F}_6$, where the only exceptional curve which is not a double curve is shown above (meeting the $\mathbb{F}_3$). Here $Q(V_0)=21$, $Q(\mathbb{F}_1)=Q(\mathbb{P}^2)=Q(V_3)=1$, and $Q(\text{others})=0$.

Finally, for the cusp $\begin{pmatrix} 5 & 10 \\ 0 & 0 \end{pmatrix}$, we have



Again V_0 is the Inoue surface. V_1 is an eight-fold blowup of $\mathbb{F}_{10}$; V_2 is an eight-fold blowup of $\mathbb{F}_3$ (with one non-double exceptional curve, shown meeting the $\mathbb{F}_6$). All other surfaces are either $\mathbb{P}^2$, $\mathbb{F}_n$, or have five or six double curves and are of the form



All unmarked double curves have self-intersection -1 (on both surfaces). Here $Q(V_0)=21$, $Q(\mathbb{P}^2)=1$, and $Q(V_2)=1$.

We have therefore verified Looijenga's conjecture for cusps of length 1 or 2:

Theorem (5.2). *Let $(\bar{V}, p)$ be the germ of a cusp singularity with resolution D, and assume $r(D) \leqq 2$. If D has a rational dual, then $(\bar{V}, p)$ is smoothable.*

The Smoothings of Certain Cusps of Length Three

Combining this result with a special case of a theorem of Karras [5] and Wahl [10] we can prove the smoothability of all cusps D of length three and $Q \leqq 20$. The result for $r(D) = 3$ is the

Proposition (5.3). *Let $D = D_1 + D_2 + D_3$ be a cusp of length 3. Then D deforms to the cusp C, where $C = C_1 + C_2$ is of length 2, $(C_1^2) = (D_1^2)$, and $(C_2^2) = (D_2^2) + (D_3^2) + 2$.*

Proposition (5.4). *Let D be a cusp of length 3, with $Q(D) = 20$. Then D is smoothable.*

Proof. It suffices to show that D deforms to a smoothable cusp C of length 2. By Proposition (5.3), if $Q(D) = 20$, then $Q(C) = 21$, so that C is smoothable if and only if C is either (2, 13), (3, 12), (5, 10), or (6, 9). Therefore, $D = (d_1, d_2, d_3)$ is smoothable if and only if one of the d_i's is either 2, 3, 5, 6, 9, 10, 12, or 13; a simple check of all triples (d_1, d_2, d_3) with $d_1 + d_2 + d_3 = 17$ [which is equivalent to $Q(D) = 20$] and $d_i \geqq 2$ show that in all cases there is a d_i equal to one of the above numbers. Equivalently, not all the d_i can be chosen from among $\{4, 7, 8, 11\}$ to achieve $d_1 + d_2 + d_3 = 17$. Q.E.D.

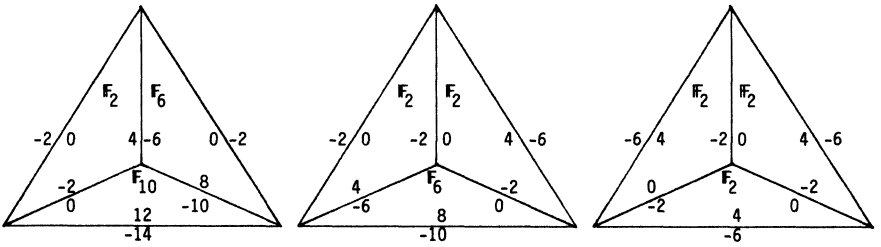
If $Q(D) < 20$, let C be as in (5.3). Then $Q(C) \leqq 20$, so, by (5.2), C is smoothable. Since D deforms to C, D is smoothable as well. Alternatively, we could deduce this from (5.4) and (2.15). Summarizing:

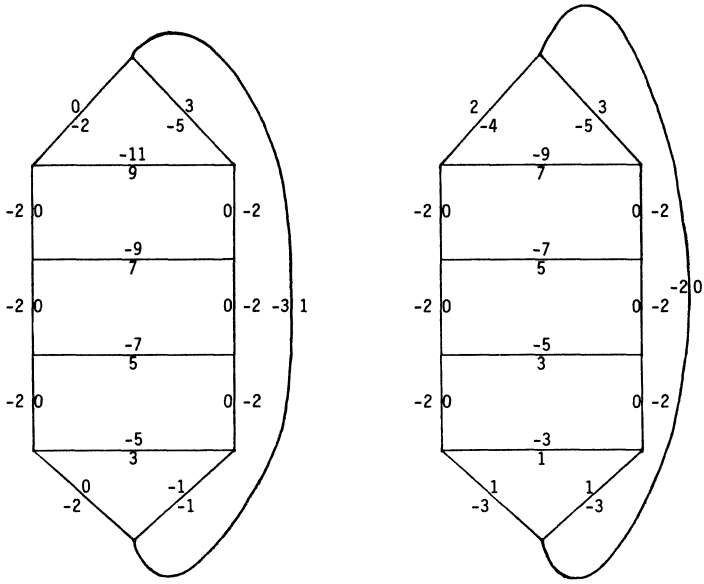
Theorem (5.5). *Let $(\bar{V}, p)$ be the germ of a cusp singularity D, with $r(D) = 3$ and $Q(D) \leqq 20$. Then $(\bar{V}, p)$ is smoothable.*

Using the methods of this paper, we have only been able to exhibit smoothings of a few $Q = 21$ cusps of length three.

Proposition (5.6). *The cusps (2, 2, 14), (2, 6, 10), and (6, 6, 6) are smoothable. The cusps (2, 5, 11) and (4, 5, 9) are smoothable.*

Proof.



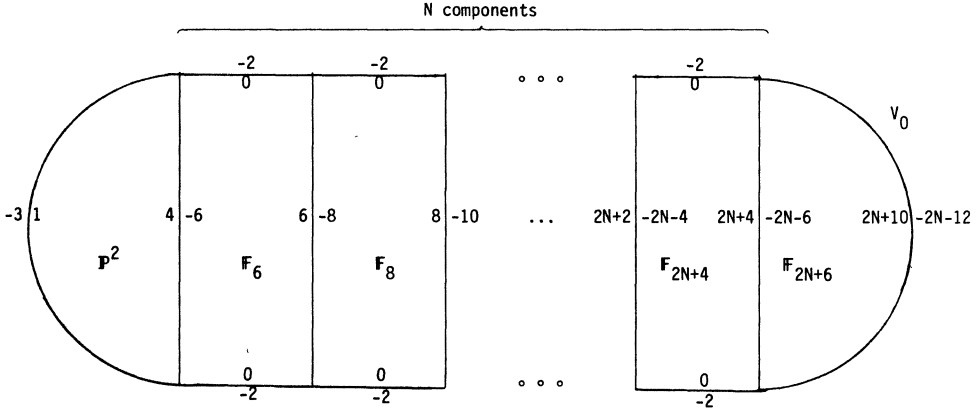


We leave the verification to the reader. Q.E.D.

Although the methods of this paper become quickly intractible if the length of the cusp is large, there are some special fibers X_0 which can be constructed with an arbitrarily large number of components, producing some interesting examples of smoothable cusps. We will just mention one such construction.

Proposition (5.7). *The* $Q=21$ *cusp* $\begin{pmatrix} 3 & 2N+12 \\ N & N \end{pmatrix}$, $N \geqq 0$, *is smoothable.*

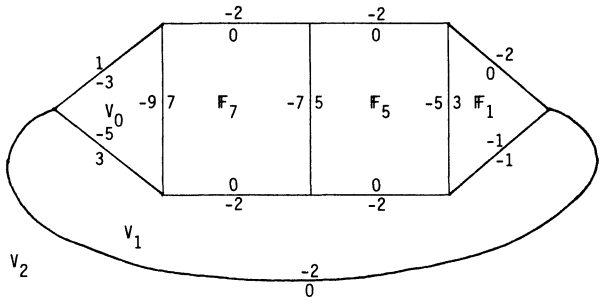
Proof. Use the special fiber



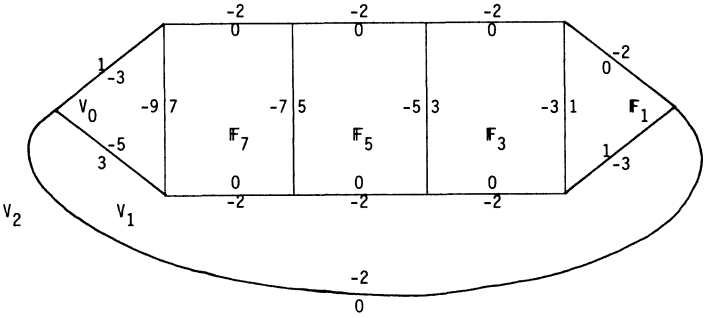
where V_0 is the Inoue surface with the given $\begin{pmatrix} 3 & 2N+12 \\ N & N \end{pmatrix}$ cusp. Here $Q(\mathbb{P}^2)=1$, $Q(\mathbb{F}_{2N+6})=2$, and $Q(V_0)=21$. Q.E.D.

6. Some Concluding Remarks

It is a theorem of Wahl [11] that the dimension of a smoothing component of the versal deformation space of a germ of a cusp singularity D is $22-Q(D)$, if one exists. Hence when $Q(D)\leqq 20$ and D is smoothable, one expects that the different directions of smoothing would imply the existence of distinct combinatorial data producing the special fiber X_0 as in (2.4). We see this empirically in many cases; one example is the cusp $(3,5,9)$ which has $Q=20$, and two different X_0's enhibiting a smoothing, which are not equivalent by base change or birational modifications:



$Q(V_1)=Q(\mathbb{F}_1)=1$
$Q(V_2)=2$
$Q(V_0)=20$



$Q(V_1)=2$
$Q(V_2)=Q(\mathbb{F}_1)=1$
$Q(V_0)=20$

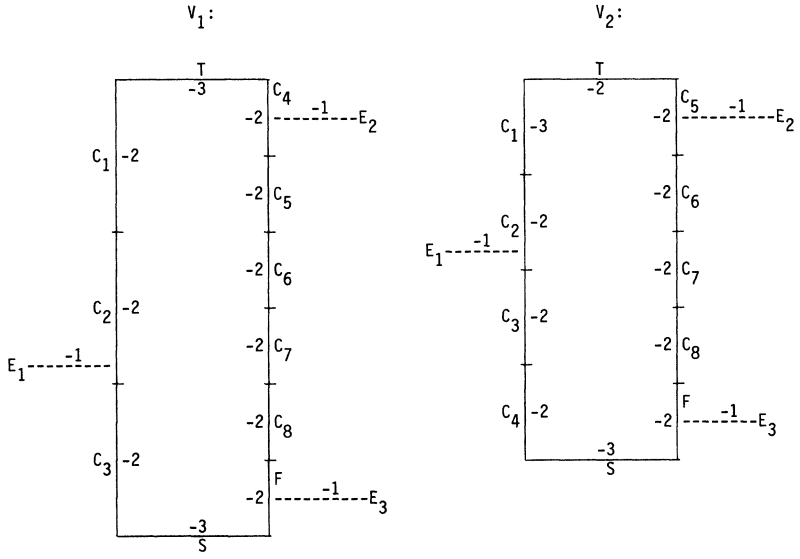
For a $Q=21$ cusp, which is smoothable, each smoothing component is one-dimensional. If there is only one smoothing component, one would expect that all special fibers X_0 exhibiting a smoothing would be base changes of a "primitive" X_0. It can be checked whether a given X_0 is the base change of another special fiber (see [2] for details) and in all cases the special fibers X_0 which we have produced for the smoothing of $Q=21$ cusps in Sect. 5 are *not* base changes of any other X_0.

If a $Q=21$ cusp had more than one smoothing component, one would expect distinct special fibers X_0 exhibiting the different smoothings, which were not related by base change or birational modifications. We have no examples of this phenomenon as yet, but we make the following

Conjecture (6.1). *Let D be a cusp, and D' its dual cusp. Assume $Q(D)=21$. Then the number of smoothing components in the versal deformation space of D is the number of non-isomorphic anticanonical pairs (V,D').*

[Two anticanonical pairs (V_1, D_1) and (V_2, D_2) are isomorphic if there is an isomorphism $f: V_1 \to V_2$ of the surfaces, such that $f^* D_2 = D_1$.]

We know of several examples of this phenomenon. Let D be the (6,9) cusp, whose dual is $D' = \begin{pmatrix} 3 & 3 \\ 3 & 6 \end{pmatrix}$. Then D' sits on two different blowups of $\mathbb{F}_3$ as an anticanonical divisor, as depicted below:



In both cases, the Neron-Severi group of V_i is generated by S, E_1, E_2, E_3, C_1 $C_2, \ldots, C_8$, and the rank $NS(V_i) = 12$. Since $r(D') = 11$, and D' is negative definite, there is exactly one primitive vector G_i (up to sign) which is orthogonal to all components of D' in both cases. A computation shows that

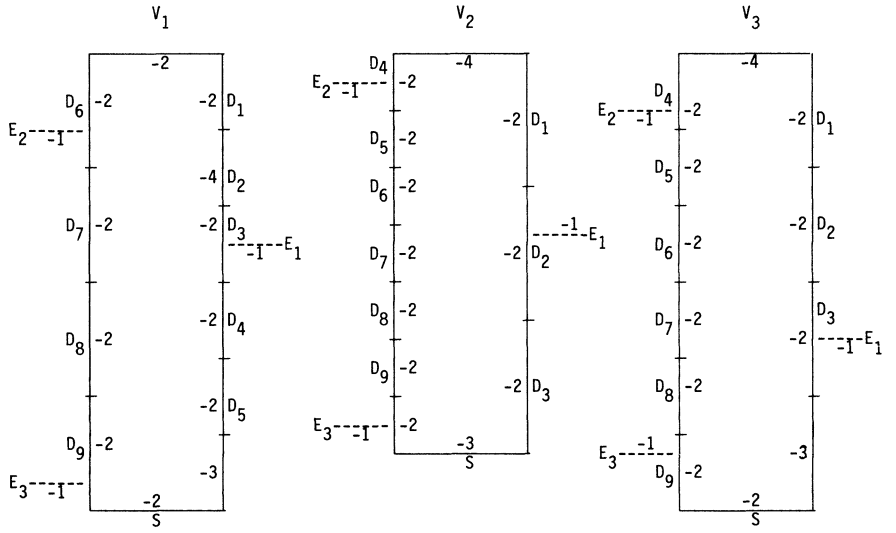
$$G_1 = 2S + 9E_1 - 6E_2 - E_3 + 5C_1 + 10C_2 + 6C_3 - 5C_4 - 4C_5 - 3C_6 - 2C_7 - C_8$$

and

$$G_2 = 12S + 80E_1 - 35E_2 - 5E_3 + 28C_1 + 84C_2 + 60C_3 + 36C_4 - 28C_5 - 21C_6 - 14C_7 - 7C_8.$$

We see that $G_1^2 = 2$ and $G_2^2 = 50$; moreover, if L is the lattice generated by the components of D', we have $|\det L| = 50$. Therefore, if the image of L in $NS(V_i)$ is the sublattice L_i, then L_2 is embedded primitively in V_2, and L_1 is embedded with index 5. (I.e., $L_2 = L_2^{\perp\perp}$ and $[L_1^{\perp\perp} : L_1] = 5$.) In particular, this proves that (V_1, D') is not isomorphic to (V_2, D'). It would be very interesting to find an alternative special fiber X_0 which exhibits a different smoothing of the (6,9) cusp than the one produced in Sect. 5, and also to be able to decide, given the special fiber X_0, which anticanonical pair is being produced as the general fiber. If it is the case that (V_1, D') is the special fiber, then there would be $\mathbb{Z}/5\mathbb{Z}$-torsion in the first homology group of the Milnor of this smoothing.

A more complicated example is afforded by the $(2,6,10)$ cusp. In matrix notation, this cusp is $D=\begin{pmatrix}6 & 10\\ 0 & 1\end{pmatrix}$, and its dual is $D'=\begin{pmatrix}3 & 4\\ 7 & 3\end{pmatrix}$, which occurs on three distinct anticanonical pairs (V_i, D'), $i=1,2,3$:



As in the previous example, the lattice L generated by the components of D' has corank one in the Neron-Severi group of V_i in each case. Let G_i be the primitive vector in $NS(V_i)$ orthogonal to L. $NS(V_i)$ is generated by $S, E_1, E_2, E_3, D_1, D_2, \ldots, D_9$ for each i (note that these letters refer to different curves, depending on i) and a computation shows that

$$G_1 = 9D_1 + 18D_2 + 63D_3 + 42D_4 + 21D_5 - 9D_6 - 20D_7 - 31D_8 - 42D_9 + 66E_1 + 2E_2 - 32E_3 - 21S,$$

$$G_2 = 30D_1 + 60D_2 + 36D_3 - 30D_4 - 25D_5 - 20D_6 - 15D_7 - 10D_8 - 5D_9 + 54E_1 - 35E_2 - 7E_3 + 12S,$$

and

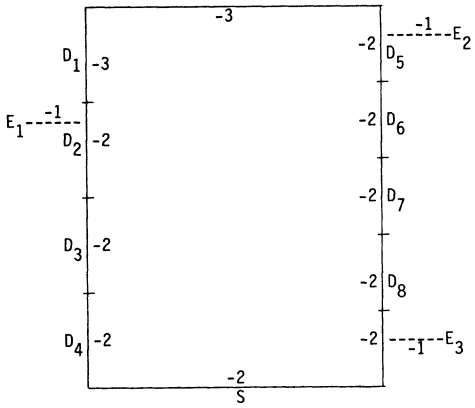
$$G_3 = D_1 + 2D_2 + 3D_3 - D_4 - 2D_5 - 3D_6 - 4D_7 - 5D_8 - 6D_9 + 4E_1 - 4E_3 - 3S.$$

One can see that $G_1^2 = G_2^2 = 100$ and $G_3^2 = 4$; since $|\det L| = 100$, L is embedded primitively in $NS(V_1)$ and in $NS(V_2)$, but has index 5 in $L^{\perp\perp}$ in $NS(V_3)$. In particular, $(V_1, D') \not\cong (V_3, D') \not\cong (V_2, D')$ as anticanonical pairs. To see that $(V_1, D') \not\cong (V_2, D')$, assume that they are isomorphic. Then the curve E_2 on V_1 must also appear as an exceptional curve E on V_2, which must meet the curve D_6 on V_2, once transversally, and be disjoint from the other components of D'. These conditions force E to be

$$E = xG_2 + 10D_1 + 20D_2 + 12D_3 - 10D_4 - 9D_5 - 8D_6 - 6D_7 - 4D_8 - 2D_9 + 18E_1 - 11E_2 - 2E_3 + 4S$$

for some integer x. However, the condition that $E^2=-1$ is equivalent to $50x^2+39x+7=0$, which has no integral solutions. Hence no such E exists on V_2 and $(V_1, D') \not\cong (V_2, D')$.

As a final example, consider the $(3,12)$ cusp D, whose dual is the cusp $D'=\begin{pmatrix}3 & 3\\ 0 & 9\end{pmatrix}$. There is only one anticanonical pair (V, D') up to isomorphism, shown below:



The classes of $S, E_1, E_2, E_3, D_1, \ldots, D_8$ generate $NS(V)$ and the primitive vector G orthogonal to the lattice L generated by the components of D' is $G=4D_1+12D_2+9D_3+6D_4-4D_5-3D_6-2D_7-D_8+11E_1-5E_2-2E_3+3S$. Here $G^2=2$, and $|\det L|=32$, so that L has index 4 in $L^{\perp\perp}$ in $NS(V)$, and there is no anticanonical pair (V, D') such that L is embedded primitively in $NS(V)$. In this case, therefore, there is in fact torsion in the first homology of the Milnor fiber, and indeed (as pointed out to us by Wahl) the fundamental group of the Milnor fiber is of order 4.

Acknowledgements. The authors would like to thank David Morrison and Nick Shepherd-Barron for many useful and lively conversations, and Jonathan Wahl for his encouragement.

References

1. Friedman, R.: Global smoothings of varieties with normal crossings (to appear)
2. Friedman, R.: Base change, automorphisms, and stable reduction of Type III K 3 surfaces. In: The birational geometry of degenerations. Progr. Math. **29** (1983) (to appear)
3. Hirzebruch, F.: Hilbert modular surfaces. Ens. Math. **19** 183–281 (1973)
4. Inoue, M.: New surfaces with no meromorphic functions. II. In: Complex analysis and algebraic geometry. Baily, W.L., Shioda, T. (eds.). Cambridge: Cambridge University Press 1977
5. Karras, U.: On pencils of elliptic curves and deformations of minimally elliptic singularities. Math. Ann. **247**, 43–65 (1980)
6. Kulikov, V.: Degenerations of K 3 surfaces and Enriques' surfaces. Math. USSR Izv. **2**, 957–989 (1977)
7. Laufer, H.: On minimally elliptic singularities. Am. J. Math. **99**, 1257–1295 (1977)
8. Looijenga, E.: Rational surfaces with an anti-canonical divisor. Ann. Math. **114**, 267–322 (1981)
9. Shepherd-Barron, N.: Extending polarizations on families of K 3 surfaces. In: The birational geometry of degenerations. Progr. Math. **29** (1983) (to appear)
10. Wahl, J.: Elliptic deformations of minimally elliptic singularities. Math. Ann. **253**, 241–262 (1980)
11. Wahl, J.: Smoothings of normal surface singularities. Topology **20**, 219–246 (1981)

Received June 18, 1982; in revised form October 10, 1982

Math. Ann. 263, 213–219 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Spectrum and Envelope of Holomorphy for Infinite Dimensional Riemann Domains

Martin Schottenloher

Mathematisches Institut der Universität, Theresienstrasse 39, 8000 München 2, Federal Republic of Germany

In this note we extend Rossi's result [15] on the construction of the envelope of holomorphy by means of the spectrum to domains Ω spread over a complex Fréchet space E with a basis: The envelope of holomorphy $\mathscr{E}(\Omega)$ of Ω can be identified with the space Σ of continuous characters on $\mathcal{O}(\Omega)$. Here, $\mathcal{O}(\Omega)$ is the algebra of holomorphic functions on Ω with the compact open topology.

This problem has previously been investigated by Alexander [1], Coeuré [3, 4], Hirschowitz [9], and the author [16, 17]. As a first step, Alexander endowed Σ with a natural structure of a manifold spread over E (in the case of a Banach space E). He showed that the Gelfand map $\delta: \Omega \to \hat{\Omega}$ is an analytic continuation of $\mathcal{O}(\Omega)$, where $\hat{\Omega} \subset \Sigma$ is that connected component of Σ containing all point evaluations $\delta(x): \mathcal{O}(\Omega) \to \mathbb{C}$, $x \in \Omega$. In general, $\hat{\Omega}$ is not the envelope of holomorphy of Ω, i.e. the analytic continuation $\delta: \Omega \to \hat{\Omega}$ is not maximal: Josefson [11] presented an example of a domain Ω in a non-separable Banach space E for which not every point of $\mathscr{E}(\Omega)$ defines a continuous character. However, Coeuré [3] and Hirschowitz [9] proved that such characters are at least bounded. Using this result the author [16] represented $\mathscr{E}(\Omega)$ as a suitably structured subset $\hat{\Omega}^b$ of the set Σ^b of all bounded characters on $\mathcal{O}(\Omega)$. Although the problem of representing $\mathscr{E}(\Omega)$ by means of the spectrum of $\mathcal{O}(\Omega)$ seems to be settled by this result, essentially all questions concerning the non-trivial relations among the spaces $\hat{\Omega}$, $\hat{\Omega}^b$, Σ, and Σ^b remained unanswered. The result of this note, which is obtained by using the approach of Gruman and Kiselman [7] to the Levi problem and its consequences [18], contributes an answer to some of these questions, since it implies $\hat{\Omega} = \hat{\Omega}^b = \Sigma \simeq \mathscr{E}(\Omega)$ whenever Ω is a domain spread over a Fréchet space E with a basis.

We begin with a description of the manifold structure of Σ and its basic properties. This seems to be indicated, although the construction of Σ can be found already in Alexander [1] (at least for the case of a Banach space E), since Alexander's thesis has not been published (the only accessible reference is Exbrayat [6]), and since the essential properties of Σ which we need for the proof of our result are not mentioned in [1]. Moreover, the construction of Σ can be applied to obtain results for domains spread over non-normed locally convex Hausdorff spaces [19, 21].

A manifold (resp. a domain) spread over a locally convex complex Hausdorff space E is a Hausdorff space (resp. a connected Hausdorff space) X together with a local homeomorphism $p:X\to E$. p defines the analytic structure on X by local sections. $\mathcal{O}(X)$ denotes the locally m-convex algebra of holomorphic functions on X endowed with the compact open topology. For a domain Ω spread over E let $\Sigma=\Sigma(\Omega)$ be the set of non-zero continuous characters $h:\mathcal{O}(\Omega)\to\mathbb{C}$. The Gelfand map $\delta:\Omega\to\Sigma$ is given by $\delta(x)=\hat{x}:\mathcal{O}(\Omega)\to\mathbb{C}$, $f\mapsto f(x)$, for $x\in\Omega$. Hence, every $f\in\mathcal{O}(\Omega)$ extends to a function $\hat{f}$ on Σ by setting $\hat{f}(h):=h(f)$ for $h\in\Sigma$.

In the sequel let E be complete (or at least quasi-complete). Every $h\in\Sigma$ defines a linear functional $E'\to\mathbb{C}$, $u\mapsto h(u\circ p)$, which is continuous with respect to the compact open topology on E'. According to the theorem of Mackey-Arens there exists a unique $qh\in E$ such that $u(qh)=h(u\circ p)$ for all $u\in E'$. Hence, we have defined a natural map $q:\Sigma\to E$.

To define the topology of Σ we need the local power series expansions of a function $f\in\mathcal{O}(\Omega)$. Given $x\in\Omega$ there exist a balanced neighbourhood $V\subset E$ of $0\in E$ and a section $s:px+V\to\Omega$ (i.e. s is continuous with $p\circ s(a)=a$ for all $a\in px+V$) such that $s(px)=x$. By the definition of the analytic structure of Ω, $f\circ s:px+V\to\mathbb{C}$ is holomorphic, hence there are continuous n-homogeneous polynomials $P^nf(x):E\to\mathbb{C}$ with

$$f\circ s(px+a)=\sum_{n=0}^{\infty}P^nf(x)\cdot a \quad \text{for all} \quad a\in V \ [5].$$

For every $n\in\mathbb{N}$ and every $a\in E$, $P_a^nf:\Omega\to\mathbb{C}$, $x\mapsto P^nf(x)\cdot a=:P_a^nf(x)$, is again a holomorphic function on Ω. Let $r>1$ with $ra\in V$. The Cauchy inequalities have the form

$$|P_a^nf(x)|\leqq r^{-n}\|f\circ s\|_{px+r\Delta a}=r^{-n}\|f\|_{s(px+r\Delta a)},$$

where $\Delta:=\{z\in\mathbb{C}\,|\,|z|<1\}$ and $\|g\|_B:=\sup\{|g(b)|\,|\,b\in B\}$.

Now, let $h\in\Sigma$. There exists a compact subset $K\subset\Omega$ such that

$$|h(f)|\leqq C\|f\|_K \quad \text{for all} \quad f\in\mathcal{O}(\Omega),$$

where C is a suitable constant. Since h is multiplicative, we can assume $C=1$. There exist an open, balanced neighbourhood $W\subset E$ of $0\in E$ and a collection $(s_x)_{x\in K}$ of sections $s_x:px+W\to\Omega$ with $s_x(px)=x$ for all $x\in K$. The Cauchy inequalities imply for $1<r$, $ra\in W$, and $x\in K$

$$|P_a^nf(x)|\leqq r^{-n}\|f\circ s_x\|_{px+r\Delta a}\leqq r^{-n}\|f\|_{K(a,r)}<\infty,$$

where $K(a,r):=\{s_x(px+\lambda a)\,|\,x\in K \text{ and } |\lambda|\leqq r\}$. Therefore,

$$|h(P_a^nf)|\leqq\|P_a^nf\|_K\leqq r^{-n}\|f\|_{K(a,r)},$$

and

$$h_a(f):=\sum_{n=0}^{\infty}h(P_a^nf), \qquad a\in\frac{1}{r}W,$$

is well-defined with $|h_a(f)| \leqq \sum r^{-n} \| f \|_{K(a,r)}$, i.e. $h_a : \mathcal{O}(\Omega) \to \mathbb{C}$ is a continuous linear form. It is easy to check that h_a is multiplicative as well. Hence, $h_a \in \Sigma$ with $|h_a| \leqq \| \ \|_{K(a,1)}$ for all $a \in W$.

Lemma 1. *($\{h_a | a \in W\} \mid h \in \Sigma$ and W as above) is the basis of a Hausdorff topology on Σ such that*

1. *$q : \Sigma \to E$ is a local homeomorphism,*
2. *$\delta : \Omega \to \Sigma$ is continuous with $p = q \circ \delta$,*
3. *$\hat{f} : \Sigma \to \mathbb{C}$ is holomorphic for all $f \in \mathcal{O}(\Omega)$.*

It is easy to show $h_{a+b} = (h_a)_b$ for $h \in \Sigma$ and appropriate $a, b \in E$. From this formula it follows that the above collection of subsets of Σ is the basis of a topology on Σ. The proofs of the other statements in the lemma are straightforward. In particular, $q(h_a) = qh + a$.

Let $\hat{\Omega} \subset \Sigma$ denote that connected component of Σ which contains $\delta(\Omega)$.

Proposition 1. *$\delta : \Omega \to \hat{\Omega}$ is an analytic continuation of $\mathcal{O}(\Omega)$ such that $\delta^* : \mathcal{O}(\hat{\Omega}) \to \mathcal{O}(\Omega)$ is open. Hence, δ^* is an isomorphism of locally convex spaces.*

Proof. δ is continuous with $p = q \circ \delta$, and δ^* is bijective with inverse $f \mapsto \hat{f}$. Hence, δ is an analytic continuation of $\mathcal{O}(\Omega)$. Clearly, δ^* is continuous. In order to prove that δ^* is also an open map it suffices to show that every compact subset $L \subset \Sigma$ is contained in some $\tilde{K}$, where $K \subset \Omega$ is compact and

$$\tilde{K} := \{h \in \Sigma \mid |h| \leqq \| \ \|_K\}.$$

By the compactness of L there are $h_1, h_2, \ldots, h_m \in L$, compact subsets $K_1, K_2, \ldots, K_m$ of Ω, balanced open neighbourhoods $W_1, W_2, \ldots, W_m$ of $0 \in E$, sections $\hat{s}_j : qh_j + W_j \to \Sigma$ with $\hat{s}_j(qh_j) = h_j$, and sections $s_{j,x} : px + W_j \to \Omega$ with $s_{j,x}(px) = x$ for $x \in K_j$, such that

$$h_j \in \tilde{K}_j \quad \text{and} \quad L \subset \bigcup_{j=1}^{m} \hat{s}_j(qh_j + \tfrac{1}{2} W_j).$$

Let L_j be the balanced hull of the compact subset $(qL - qh_j) \cap \frac{1}{2} \bar{W}_j$ of E. For each $x \in K_j$, $px + L_j \subset px + W_j$. Hence,

$$\begin{aligned} K &:= \bigcup \{s_{j,x}(px + L_j) \mid j = 1, 2, \ldots, m \text{ and } x \in K_j\} \\ &= \bigcup \{K_j(a, 1) \mid j = 1, 2, \ldots, m \text{ and } a \in L_j\} \end{aligned}$$

is a compact subset of Ω. Now, given $h \in L$ there are $j \in \{1, 2, \ldots, m\}$ and $a \in L_j$ with $h = (h_j)_a$. Thus,

$$|h(f)| \leqq \| f \|_{K_j(a,1)} \leqq \| f \|_K \quad \text{for all} \quad f \in \mathcal{O}(\Omega)$$

which implies $h \in \tilde{K}$ and hence $L \subset \tilde{K}$.

The proposition implies that $\delta : \Omega \to \hat{\Omega}$ is a maximal analytic continuation among all analytic continuations $j : \Omega \to \tilde{\Omega}$ (i.e. j continuous, $p = p \circ j$ and j^* surjective) such that $j^* : \mathcal{O}(\tilde{\Omega}) \to \mathcal{O}(\Omega)$ is open. Another consequence is:

Corollary 1. *$\Sigma(\Omega) \simeq \Sigma(\hat{\Omega})$ and $\hat{\Omega} \simeq \hat{\hat{\Omega}}$.*

In order to formulate convexity properties of Σ with respect to $\mathcal{O}(\Omega)$ let $A(\Omega) \subset \mathcal{O}(\Sigma)$ denote the algebra of Gelfand transforms $A(\Omega) := \{\hat{f} \mid f \in \mathcal{O}(\Omega)\}$ and

put, for $L \subset \Sigma$,

$$\tilde{L} = \hat{L}_{A(\Omega)} := \{h \in \Sigma \mid |\hat{f}(h)| \leqq \|\hat{f}\|_L \text{ for all } \hat{f} \in A(\Omega)\}.$$

Proposition 2. *$\delta_\Sigma(L, a) = \delta_\Sigma(\tilde{L}, a)$ for all $a \in E$ and for all compact subsets $L \subset \Sigma$, where $\delta_\Sigma : \Sigma \times E \to]0, \infty]$ is the usual directional distance function.*

Proof. For a manifold $p : X \to E$ spread over E the directional distance $\delta_X(x, a)$, $(x, a) \in X \times E$, is given by

$$\delta_X(x, a) = \sup\{t > 0 \mid \text{There is a section } s : px + t\Delta a \to X \text{ with } s(px) = x\}.$$

$\delta_X(L, a)$ is simply the infimum of all the $\delta_X(x, a)$, $x \in L$. Note that δ_X is lower semicontinuous.

$t := \delta_\Sigma(L, a)$ satisfies $t > 0$ since $L \subset \Sigma$ is compact. Consider $h \in \tilde{L}$ and $b \in t\Delta a$. For $r > 1$ with $rb \in t\Delta a$ the Cauchy inequalities imply

$$|h(P_b^n f)| \leqq \|P_b^n \hat{f}\|_L \leqq r^{-n} \|\hat{f}\|_{L(b,r)} \quad \text{for all} \quad \hat{f} \in A(\Omega).$$

$L(b, r) \subset \Sigma$ is again compact. Consequently, $h_b \in \Sigma$ is well-defined and gives a section $s : qh + t\Delta a \to \Sigma$, $qh + b \mapsto h_b =: s(qh + b)$, with $s(qh) = h$. This shows $\delta_\Sigma(h, a) \geqq t = \delta_\Sigma(L, a)$ for all $h \in \tilde{L}$.

Corollary 2. *$\hat{\Omega}$ and Σ are pseudoconvex.*

Proof. It is enough to show that for finite dimensional linear subspaces $E_0 \subset E$ the manifolds $q^{-1}(E_0)$ and $q^{-1}(E_0) \cap \hat{\Omega}$ are pseudoconvex. But this follows immediately from Proposition 2.

A stronger convexity property of Σ guarantees that Σ is already connected [21]:

Proposition 3. *Σ is connected if Σ is $A(\Omega)$-convex, i.e. if $\tilde{L}$ is compact for each compact subset $L \subset \Sigma$.*

Before proving our main result we state another natural property of Σ:

Lemma 2. *Let Ω_1, Ω_2 be domains spread over complete locally convex Hausdorff spaces E_1, E_2, and let $\pi_\iota : \Omega_1 \times \Omega_2 \to \Omega_\iota$, $\iota = 1, 2$, denote the projections. Then $\psi : \Sigma(\Omega_1 \times \Omega_2) \to \Sigma(\Omega_1) \times \Sigma(\Omega_2)$, $h \mapsto (h \circ \pi_1^*, h \circ \pi_2^*)$, is continuous and surjective.*

Theorem. *Let E be a complex Fréchet space with the Banach approximation property, i.e. there exists a sequence (T_n) of continuous linear operators $T_n : E \to E$ of finite rank such that $T_n(a) \to a$ for each $a \in E$. If Ω is a domain spread over E, then $\Sigma = \Sigma(\Omega)$ is connected and holomorphically convex. Furthermore, the Gelfand map $\delta : \Omega \to \Sigma$ is a maximal analytic continuation of $\mathcal{O}(\Omega)$, i.e. Σ coincides with the envelope of holomorphy of Ω. Finally, δ is an isomorphism whenever Ω is pseudoconvex.*

Proof. By Proposition 1, $\delta : \Omega \to \hat{\Omega}$ is an analytic continuation of $\mathcal{O}(\Omega)$, and by Corollary 2, $\hat{\Omega}$ is pseudoconvex. Since pseudoconvex domains over E are domains of existence (cf. 3.4 in [18]), δ is always maximal, and $\delta : \Omega \to \hat{\Omega}$ is an isomorphism of domains, whenever Ω is already pseudoconvex. It remains to show that Σ is connected and holomorphically convex. By Corollary 1 we can assume $\Omega = \hat{\Omega}$. In particular, we suppose Ω to be pseudoconvex (Corollary 2) in the sequel.

By a result of Pelczynski [12], E is a complemented linear subspace of a Fréchet space with a Schauder basis. According to Lemma 2 we can assume that E itself has a basis. Thus there are continuous linear projections $\pi_n : E \to E$ of finite rank such that $\pi_j \circ \pi_k = \pi_{\min\{j,k\}}$ and $\pi_n \to \mathrm{id}_E$ uniformly on each compact subset of E. Now the proof of 4.1 in [18] could be adapted to yield that Σ is $A(\Omega)$-convex, hence connected by Proposition 3. In order to avoid the whole machinery of [18] we give another proof of $\hat{\Omega} = \Sigma$, in which we require E to admit a continuous norm. This is no loss of generality by the factorization properties of pseudoconvex domains (cf. 2.3 and Sect. 1 in [18]).

Define $\eta_n(x) := \inf\{\delta_\Omega(x, \pi_j(px) - px) | j \geqq n\}$ for $x \in \Omega$ and $n \in \mathbb{N}$. It is easy to show that η_n is lower semicontinuous. Therefore $-\log \eta_n : \Omega \to [-\infty, \infty[$ is plurisubharmonic as an upper semicontinuous supremum of the plurisubharmonic functions $x \mapsto -\log \delta_\Omega(x, \pi_j(px) - px)$, $j \geqq n$. Put $x_n := \{x \in \Omega | \eta_n(x) > 1\}$. Then (X_n) is an increasing cover of Ω consisting of open subsets of Ω. For $x \in X_n$ there are $t > 1$ and a section $s : px + t\Delta(\pi_n(px) - px) \to \Omega$ with $s(px) = x$, since $\delta_\Omega(x, \pi_n(px) - px) > 1$. Setting $\tau_n(x) := s(\pi_n(px))$ we have a well-defined continuous surjection

$$\tau_n : X_n \to \Omega_n := p^{-1}(\mathrm{im}\, \pi_n)$$

with the following properties: $\pi_n \circ p | X_n = p|_{\Omega_n} \circ \tau_n$, $\tau_n|_{\Omega_n} = \mathrm{id}_{\Omega_n}$, $\tau_j \circ \tau_k = \tau_{\min\{j,k\}}$ whenever defined, and $\tau_j \to \mathrm{id}_{X_n}$ uniformly on the compact subsets of X_n. (If $\Omega \subset E$ is schlicht, $\Omega_n = \Omega \cap \mathrm{im}\, \pi_n$, $X_n \subset \Omega \cap \pi_j^{-1}(\Omega_j)$ for $j \geqq n$, and $\tau_n = \pi_n|_{X_n}$.)

Now, let $h \in \Sigma$ with $h \in \tilde{K}$ for a compact subset K of Ω. There is $n \in \mathbb{N}$ with $K \subset X_n$. Ω is holomorphically convex (cf. 4.1 in [18]). Therefore we can assume $K = \hat{K}_{\mathcal{O}(\Omega)} := \{x \in \Omega | |f(x)| \leqq \|f\|_K$ for all $f \in \mathcal{O}(\Omega)\}$. According to the Oka-Weil approximation theorem (cf. 4.5 in [18]) every $g \in \mathcal{O}(X_j)$, $j \geqq n$, can be approximated uniformly on K by holomorphic functions $f \in \mathcal{O}(\Omega)$ on Ω. As a consequence, h can be extended to a continuous character $h_j : \mathcal{O}(X_j) \to \mathbb{C}$, $j \geqq n$. Since Ω_j is a pseudoconvex manifold spread over the finite dimensional $\mathbb{C}^j \simeq \mathrm{im}\, \pi_j$, we can apply the classical result (cf. Theorem IX, D4 in [8]) to deduce that every character $h_j \circ \tau_j^* : \mathcal{O}(\Omega_j) \to \mathbb{C}$, $g \mapsto h_j(g \circ \tau_j)$, for $g \in \mathcal{O}(\Omega_j)$ is a point evaluation. Thus there are points $x_j \in \Omega_j$, $j \geqq n$, with $g(x_j) = h_j(g \circ \tau_j)$ for all $g \in \mathcal{O}(\Omega_j)$, in particular for $f|_{\Omega_j} \in \mathcal{O}(\Omega_j)$ whenever $f \in \mathcal{O}(\Omega)$. For each $f \in \mathcal{O}(\Omega)$ the sequence $(\|f \circ \tau_j - f\|_K)$ converges to zero, since $\tau_j \to \mathrm{id}_K$ uniformly on K. From

$$|f(x_j) - h(f)| = |h(f \circ \tau_j) - h(f)| \leqq \|f \circ \tau_j - f\|_K$$

it follows that

$$h(f) = \lim_{j \to \infty} f(x_j).$$

In particular, $(|f(x_j)|)_{j \geqq n}$ is bounded for every $f \in \mathcal{O}(\Omega)$. Since Ω is holomorphically convex, $(x_j)_{j \geqq n}$ has a convergent subsequence (x_{j_μ}) with limit $x \in \Omega$. Finally, for each $f \in \mathcal{O}(\Omega)$

$$h(f) = \lim_{\mu \to \infty} f(x_{j_\mu}) = f(x),$$

i.e. $h = \delta(x) \in \hat{\Omega}$. This completes the proof of $\Sigma = \hat{\Omega}$. Thus Σ is connected. Σ is holomorphically convex, since it is a pseudoconvex domain (cf. 4.1. in [18]).

A direct application of the theorem is the following:

Corollary 3. *Let $f:\Omega_1\to\Omega_2$ be a holomorphic map from a domain Ω_1 spread over a Fréchet space E_1 as in the theorem into a domain Ω_2 spread over a complete locally convex Hausdorff space E_2. Then f extends analytically to the envelopes of holomorphy $j_\iota:\Omega_\iota\to\mathscr{E}(\Omega_\iota)$, $\iota=1,2$, i.e. there exists a holomorphic map $\mathscr{E}(f):\mathscr{E}(\Omega_1)\to\mathscr{E}(\Omega_2)$ such that $\mathscr{E}(f)\circ j_1=j_2\circ f$.*

Proof. $f^{**}:\Sigma(\Omega_1)\to\Sigma(\Omega_2)$, $h\mapsto h\circ f^*$, is analytic with $f^{**}\circ\delta_1=\delta_2\circ f$. $f^{**}(\hat{\Omega}_1)\subset\Sigma(\Omega_2)$ is connected and contains $\delta_2(f(\Omega_1))$, hence $f^{**}(\hat{\Omega}_1)\subset\hat{\Omega}_2$. By Proposition 2, there exists an analytic continuation $j:\hat{\Omega}_2\to\mathscr{E}(\Omega_2)$ with $j_2=j\circ\delta_2$. Since $\delta_1:\Omega_1\to\Sigma(\Omega_1)$ can be supposed to be the envelope of holomorphy of Ω_1 by the theorem, the result follows by setting $\mathscr{E}(f):=j\circ f^{**}$.

The corollary complements the corresponding result in which the rôles of E_1 and E_2 are interchanged (cf. 3.5 in [18]).

As another application of the theorem one can show that *every ideal $\mathscr{J}\subset\mathcal{O}(\Omega)$ without a common zero in Ω is dense in $\mathcal{O}(\Omega)$*, whenever Ω is a pseudoconvex domain spread over a Fréchet space E with the Banach approximation property.

The assertion of the theorem also holds for domains spread over E'_c, the dual E' of E endowed with the compact open topology [21]. Therefore, with the aid of [2], the following classical result of Cartan can be generalized to pseudoconvex domains Ω spread over E'_c, when E is as in the theorem: *Every finitely generated, proper ideal $\mathscr{J}\subset\mathcal{O}(\Omega)$ has a common zero in Ω.* We don't know whether or not this result holds for pseudoconvex domains over a Fréchet space.

Instead of Σ or Σ^b one also can consider the space Σ^0 of finitely determined characters on $\mathcal{O}(\Omega)$, when Ω is a domain spread over a locally convex Hausdorff space E: Σ^0 is the set of continuous characters $h\in\Sigma$ with the property that there exists a compact subset $K\subset\Omega$ satisfying $|h|\leqq\|\ \|_K$ and $\dim_{\mathbb{C}}\mathrm{span}(pK)<\infty$. In the same way as Σ the set Σ^0 can be endowed with the structure of a manifold over E (even if E is not complete) such that the analogues of the Propositions 1 and 2 with their consequences are valid. In particular, the connected component $\hat{\Omega}^0$ of Σ^0 which contains $\delta(\Omega)$ is pseudoconvex and gives an analytic continuation $\delta:\Omega\to\hat{\Omega}^0$ of $\mathcal{O}(\Omega)$. In the situation of the theorem it follows that $\hat{\Omega}^0=\hat{\Omega}=\Sigma^0=\Sigma$.

For a schlicht domain $\Omega\subset E$ with the property that $\mathcal{O}(E)$ is dense in $\mathcal{O}(\Omega)$ one part of our theorem holds true if E is a complete locally convex Hausdorff space with the approximation property: Σ is connected and holomorphically convex, hence polynomially convex, and $\Sigma=\hat{\Omega}$ coincides with the smallest polynomially convex domain of E containing Ω [10, 13]. However, $\hat{\Omega}$ is in general not the envelope of holomorphy of Ω as the example of Josefson [11] shows. The only relevant result concerning the coincidence of Σ with the space Σ^b of all bounded characters on $\mathcal{O}(\Omega)$ is again known for schlicht domains $\Omega\subset E$ with $\mathcal{O}(E)$ dense in $\mathcal{O}(\Omega)$: Mujica [14] showed $\Sigma^b=\Sigma$ for such domains if E is a Fréchet space with the Banach approximation property. But even in this special situation it is unknown whether or not there exists any unbounded character on $\mathcal{O}(\Omega)$. This question is closely related to the Michael problem. From [20] one can easily deduce that the following assertions are equivalent: (1) All characters on $\mathcal{O}(\Omega)$ are bounded for every domain Ω over a locally convex Hausdorff space. (2) All characters on any m-convex commutative Fréchet algebra with unit are continuous. (3) All charac-

ters on the Fréchet algebra $\mathcal{O}(\mathcal{S}')$ are continuous, where $\mathcal{S}'$ denotes the space of tempered distributions.

References

1. Alexander, H.: Analytic functions on Banach spaces. Thesis, University of California, Berkeley, 1968
2. Arens, R.: Dense inverse limit rings. Mich. Math. J. **5**, 169–182 (1958)
3. Coeuré, G.: Fonctions plurisousharmoniques sur les espaces vectoriels topologiques et applications à l'étude des fonctions analytiques. Ann. Inst. Fourier **20**, 361–432 (1970)
4. Coeuré, G.: Analytic functions and manifolds in infinite dimensional spaces. Amsterdam: North-Holland 1974
5. Dineen, S.: Complex analysis in locally convex spaces. Amsterdam: North-Holland, 1981
6. Exbrayat, J.M.: Fonctions analytiques dans un espace de Banach (d'après Alexander). In: Séminaire P. Lelong, 1969. Lecture Notes in Mathematics, No. **116**, pp. 30–38. Berlin, Heidelberg, New York: Springer 1970
7. Gruman, L., Kiselman, C.O.: Le problème de Levi dans les espaces de Banach à base. C.R. Acad. Sci. **275** A, 1296–1299 (1972)
8. Gunning, R.C., Rossi, H.: Analytic functions of several variables. Englewood Cliffs: Prentice-Hall 1965
9. Hirschowitz: Prolongement analytique en dimension infinie. Ann. Inst. Fourier **22**, 255–296 (1972)
10. Isidro, J.M.: Characterization of the spectrum of some topological algebras of holomorphic functions. In: Advances in holomorphy. Ed. J. A. Barroso. Amsterdam · North-Holland 1979
11. Josefson, B.: A counterexample to the Levi problem. In: Proceedings on infinite dimensional holomorphy. Eds. T. L. Hayden, T. J. Suffridge. Lecture Notes in Mathematics, No. **364**, pp. 79–89. Berlin, Heidelberg, New York: Springer 1974
12. Matyszczyk, C.: Approximation of analytic and continuous mappings by polynomials in Fréchet spaces. Studia Math. **60**, 223–238 (1977)
13. Mujica, J.: The Oka-Weil theorem in locally convex spaces with the approximation property. Séminaire P. Krée, 1977/78. Exposé 3. Paris
14. Mujica, J.: Complex homomorphisms of the algebra of holomorphic functions on Fréchet spaces. Math. Ann. **241**, 73–82 (1979)
15. Rossi, H.: On envelopes of holomorphy. Comm. pure appl. Math. **16**, 9–19 (1963)
16. Schottenloher, M.: Über analytische Fortsetzung in Banachräumen. Math. Ann. **199**, 313–336 (1972)
17. Schottenloher, M.: Analytic continuation and regular classes in locally convex Hausdorff spaces. Portugal. Math. **33**, 219–250 (1974)
18. Schottenloher, M.: The Levi problem for domains spread over locally convex spaces with a finite dimensional Schauder decomposition. Ann. Inst. Fourier **26**, 207–237 (1976)
19. Schottenloher, M.: $\tau_0 = \tau_\omega$ for domains in $\mathbb{C}^{\mathbb{N}}$. In: Infinite dimensional holomorphy and applications. Ed. M. C. Matos., pp. 393–395. Amsterdam: North-Holland 1977
20. Schottenloher, M.: Michael problem and algebras of holomorphic functions. Arch. Math. **37**, 241–247 (1981)
21. Schottenloher, M.: Cartan-Thullen theorem for domains spread over DFM-spaces (to appear in J. reine angew. Math.)

Received October 25, 1982

Math. Ann. 263, 221–225 (1983)

Mathematische Annalen
© Springer-Verlag 1983

On Tautness in Sheaf Cohomology

Satya Deo
Department of Mathematics, University of Jammu, Jammu 180001, India

1. Introduction

Let us recall that if φ is a family of supports on a given topological space X and A is a subspace of X, then A is said to be φ-taut subspace of X with respect to sheaf cohomology [1, p. 52] if for each flabby sheaf $\mathscr{A}$ on X (i) the restriction homomorphism $\Gamma_{\varphi}(\mathscr{A}) \to \Gamma_{\varphi \cap A}(\mathscr{A}|A)$ is onto, and (ii) $\mathscr{A}|A$ is $(\varphi \cap A)$-acyclic. The classical notion of tautness for a contravariant functor on subsets of a space X was, first of all, formalised by Spanier [8, p. 289], and the above definition adapted to the case of sheaf cohomology functor [1, p. 52] generalises it for a given family φ of supports; if we take $\varphi = \mathrm{cld}$, to be the family of all closed sets of X, then in view of [1, Theorem 10.5, p. 52], above notion of φ-tautness reduces to that of [8]. It is well-known that the notion of tautness is fundamental in the case of Alexander-Spanier cohomology [9]. The important known results on tautness for Alexander-Spanier cohomology are the following: *a compact subset of a Hausdorff space is taut, a closed subset of a paracompact Hausdorff space is taut* [9]; *an arbitrary subset of a completely paracompact Hausdorff space is taut* [7]; *a neighbourhood retract of any space is taut* [3]. Spanier has given a very nice unified proof of all these results [9]. In sheaf cohomology the concept of φ-taut subspaces has proved to be basic not only in proving Vietoris-Begle theorems etc. but also because the final satisfactory development of relative sheaf cohomology depends mainly on this notion [1]. The following are the basic results on φ-taut subspaces in sheaf cohomology [ibd.p. 52]:

(a) if φ is the family of all closed subsets and A is compact and relatively Hausdorff in X, then A is φ-taut in X,

(b) if φ is paracompactifying, then any closed subset of X is φ-taut in X,

(c) if φ is paracompactifying and X is completely paracompact, then any subspace A of X is φ-taut in X,

(d) if φ is paracompactifying for the pair (X, A), then A is φ-taut in X.

Now, it is natural to ask whether or not, for a given family φ of supports on X, a neighbourhood retract A of a space X is φ-taut in X. Since an open set is φ-taut for any φ, one can also ask whether or not the above results on φ-tautness can be

improved for arbitrary φ. The objective of this note is to answer all these questions. One can easily see that

(b′) if φ is paracompactifying then any retract A of a Hausdorff space X is φ-taut in X,

is a special case of (b). As a corollary to our main result we accomplish that none of the results (a)–(d) and even (b′), can be improved any further. We prove a theorem which, in particular, implies that if φ is not cld or φ is not paracompactifying, then all of the corresponding results stated above fail miserably even for such a nice space as euclidean space. It also makes clear that it is the "paracompactifying" nature of the family φ which is crucial for the validity of the corresponding known results on φ-tautness.

2. Preliminaries

Unless we say otherwise, all our notations regarding sheaves and sheaf cohomology are standard, as for example given in [1]. We also assume that all our spaces are Hausdorff. The basic notion which we shall exploit in the proof of our main result is that of sheaf theoretic cohomological dimension. Let X be a space, φ be a family of supports on X and L be a ring. Then the smallest integer $n (or \infty)$ such that for every sheaf $\mathscr{A}$ of L-modules on X and for each $k>n$, $H^k_\varphi(X;\mathscr{A})=0$, is called the φ-cohomological dimension of X over L and is denoted by $\dim_{\varphi,L}(X)$. If φ is paracompactifying then it is known that $\dim_{\varphi,L}(X)\leqq n$ if and only if each sheaf $\mathscr{A}$ of L-modules on X has a φ-soft resolution of length n. For paracompactifying φ, φ-softness is a local property on the extent $E(\varphi)$ of φ. Hence it follows that for any two such families φ and ψ if $E(\varphi)\subseteqq E(\psi)$ then any ψ-soft sheaf is φ-soft, which means that $\dim_{\varphi,L}(X)\leqq\dim_{\psi,L}(X)$. This implies, in particular, that if X admits a paracompactifying family φ of supports such that $E(\varphi)=X$, then one can unambiguously define the *cohomological dimension* of X over L, denoted by $\dim_L(X)$, to be $\dim_{\varphi,L}(X)$. Locally paracompact spaces, admit of such a family of supports and hence for such spaces X, $\dim_L(X)$ is always defined.

Now the following is clear:

Lemma 1. *Suppose φ is a family of supports on a locally paracompact space X. If $\dim_L(X)=n$ and there is a sheaf $\mathscr{A}$ of L-modules on X such that $H^i_\varphi(X;\mathscr{A})\neq 0$ for some $i>n$, then φ cannot be paracompactifying.*

One can also establish the following:

Lemma 2. *Let $\{F_\alpha|\alpha\in\Lambda\}$ be a locally finite family of disjoint closed sets of a space X and for each α, let φ_α be a family of supports on F_α. If we define a family*

$$\Phi=\left\{G\subseteqq X\,|\,G=\bigcup_{\alpha\in\Lambda}A_\alpha \text{ such that } A_\alpha\in\varphi_\alpha\right\},$$

then Φ is a family of supports on X and for each sheaf $\mathscr{A}$ of L-*modules on* $\bigcup F_\alpha$, *there is a natural isomorphism*

$$H^p_\Phi(\bigcup F_\alpha;\mathscr{A})\approx\prod_{\alpha\in\Lambda}H^p_{\varphi_\alpha}(F_\alpha;\mathscr{A}|F_\alpha)$$

for each p.

3. Main Result

Suppose φ is a paracompactifying family of supports on a space X and A is a retract of X. If X is assumed to be Hausdorff, then because a retract of a Hausdorff space is always closed, A must be closed in X. Hence the following is a special case of (b) (see introduction).

(b′) *If φ is a paracompactifying family of supports, then any retract of a Hausdorff space X is φ-taut in X.*

We are mentioning this only to emphasise later on that if φ is not paracompactifying, then even this result will not be true. As a matter of fact, this work was done mainly to answer the question viz., whether or not the "paracompactifying" condition can be dropped from any or all of the known results on φ-tautness including (b′). The answer turns out to be in negative for all cases (see corollary).

Of course, if φ is assumed to be paracompactifying, then whether or not a neighbourhood retract A (which is not necessarily closed in X) of an arbitrary space X is φ-taut in X, still remains an open question.

Now, let us recall that for a given topological space X and a ring L, $\sup_{\varphi}\{\dim_{\varphi,L}(X)\}$ where φ runs over all the families of supports on X, is denoted by $\mathrm{Dim}_L(X)$, and is called the *large cohomological dimension* of X over L. We had proved in [5] that if X is a n-manifold, $n\geqq 1$, with or without boundary, then $\mathrm{Dim}_{\mathbb{Z}}(X)=n+1$. This is the basic fact which we shall use in proving our main

Theorem. *Let $X=\mathbb{R}^n$ be the euclidean n-space, $n\geqq 1$. Then for each $x\in X$ there exists a family Φ of supports on X such that $\{x\}$ is not Φ-taut in X with respect to sheaf cohomology.*

Proof. For each $x\in X$, we let $B\left(x,\frac{1}{m}\right)$ denote the open ball of radius $\frac{1}{m}$ with centre x. We also denote the corresponding closed ball by $\overline{B\left(x,\frac{1}{m}\right)}$. For each $m\geqq 1$, since $B\left(x,\frac{1}{m}\right)\Big\backslash\overline{B\left(x,\frac{1}{m+1}\right)}$ is an n-manifold, there exists a closed set (we take closed ball of sufficiently small positive radius) F_m of X such that $\dim_{\mathbb{Z}}(F_m)=n$ and $F_m\subseteqq B\left(x,\frac{1}{m}\right)\Big\backslash\overline{B\left(x,\frac{1}{m+1}\right)}$. Since F_m is a n-manifold with boundary, it follows from [5] that the large cohomological dimension of F_m over $\mathbb{Z}$ is $n+1$. Consequently, there exists a family φ_m of supports in F_m and a sheaf $\mathscr{A}_m$ of abelian groups on F_m such that $H^{n+1}_{\varphi_m}(F_m;\mathscr{A}_m)\neq 0$. By Lemma 1, φ_m is not paracompactifying for each m. Next we remark that the family $\bigcup\{\varphi_m \mid m\in\mathbb{N}\}$ of closed subsets of X is never a family of supports on X simply because it is not closed under finite unions. But, if we let φ_0 to be the family of all closed subsets of $\{x\}$ and define

$$\Phi=\left\{C\subseteqq X \mid C=\bigcup_{m\in\mathbb{N}\cup\{0\}} A_m \quad \begin{array}{l}\text{where } A_m\in\varphi_m \text{ and either } A_m \text{ is}\\ \text{empty for all } m \text{ except finitely}\\ \text{many or else } A_0=\{x\}\end{array}\right\},$$

then one can easily see that Φ is indeed a family of supports on X. We also point out that Φ is not the family of supports generated by $\bigcup\varphi_m$ but is strictly larger

than that. Finally, we observe that each member of Φ is obviously paracompact, yet it is not paracompactifying; this follows from the fact that $\Phi|F_m=\varphi_m$ and φ_m is not paracompactifying.

Now we will show that $\{x\}$ is not Φ-taut in X. To prove this let us start by observing that $F=\bigcup F_m$ is closed in $X\backslash x$ because $\{F_m\}$ is a discrete family of closed subsets of $X\backslash x$. Also it is clear that there is a unique sheaf $\mathscr{A}$ of abelian groups on F such that $\mathscr{A}|F_m=\mathscr{A}_m$. Furthermore, because F is locally closed in X, there exists a unique sheaf $\mathscr{A}^X$ on X such that $\mathscr{A}^X|F=\mathscr{A}$ and $\mathscr{A}^X|X\backslash F=0$. Let us now assume that $\{x\}$ is Φ-taut in X. To arrive at a contradiction, we compute $H^*_\Phi(X;\mathscr{A}^X)$: we consider the exact cohomology sequence for the pair $(X,\bar{V}_m)$ and the sheaf $\mathscr{A}^X$ (suppressed) where $V_m=B\left(x,\frac{1}{m}\right)$;

$$\ldots\to H^p_\Phi(X,\bar{V}_m)\to H^p_\Phi(X)\to H^p_{\Phi\cap\bar{V}_m}(\bar{V}_m)\to H^{p+1}_\Phi(X,\bar{V}_m)\to\ldots.$$

First of all we claim that each $\bar{V}_m$ is Φ-taut in X. This follows from [1, Theorem 10.5, p. 52] and the following general remark: If $A\subseteq X$ is a closed set, φ is any family of supports on X and there exists an open neighbourhood U of A such that $\varphi\cap U=\varphi\cap A$, then for any sheaf $\mathscr{B}$ on X the restriction homomorphism

$$H^*_{\varphi\cap U}(U;\mathscr{B})\to H^*_{\varphi\cap A}(A;\mathscr{B})$$

is isomorphism. Thus $\bar{V}_m$ is a cofinal system of Φ-taut neighbourhoods of x and since $\{x\}$ is assumed to be Φ-taut, we obtain [ibd.]

$$\varinjlim_m H^p_{\Phi\cap\bar{V}_m}(\bar{V}_m)\approx H^p_{\Phi\cap\{x\}}(\{x\})=0$$

for each p. Taking the direct limit of above exact sequences over m and noting the fact that a direct limit of exact sequences is exact, we find that for each p,

$$\varinjlim_m H^p_\Phi(X;\bar{V}_m)\approx H^p(X).$$

But, for each m,

$$\begin{aligned}H^p_\Phi(X,\bar{V}_m)&=H^p_\Phi(X,\bar{V}_m;\mathscr{A}^X)\\&\approx H^p_\Phi(X;(\mathscr{A}^X|(X\backslash\bar{V}_m))^X)\quad\text{by 12.2 (p. 59) of [1]}\\&\approx H^p_{\Phi|\bigcup_{k<m}F_k}\left(\bigcup_{k<m}F_k;\bigoplus_{k<m}(\mathscr{A}^X)_{F_k}\Big|\bigcup_{k<m}F_k\right)\quad\text{by 10.2 (p. 51) of [1]}\\&\approx\bigoplus_{k<m}H^p_{\varphi_k}(F_k;\mathscr{A}_k)\end{aligned}$$

since $\Phi|F_k=\varphi_k$ and $\{F_k\}_{k<m}$ is a mutually disjoint family of closed sets. Therefore,

$$H^p_\Phi(X;\mathscr{A}^X)\approx\bigoplus_{m\geqq 1}H^p_{\varphi_m}(F_m;\mathscr{A}_m).$$

We also compute $H^p_\Phi(X\backslash x;\mathscr{A}^X|(X\backslash x))$: since $\{F_m\}$ is a discrete family of closed sets of $X\backslash x$, it follows from Lemma 2 and the definition of Φ that for each p

$$H^p_{\Phi\cap(X\backslash x)}(X\backslash x;\mathscr{A}^X|(X\backslash x))\approx\prod_{m\geqq 1}H^p_{\varphi_m}(F_m;\mathscr{A}_m).$$

Now, we take the case $p=n+1$. Since for each m, $H^{n+1}_{\varphi_m}(F_m;\mathscr{A}_m)\neq 0$ no mapping from $\bigoplus_{m\geqq 1}H^{n+1}_{\varphi_m}(F_m;\mathscr{A}_m)$ to $\prod_{m\geqq 1}H^{n+1}_{\varphi_m}(F_m;\mathscr{A}_m)$ can be onto, and this atonce

implies that the restriction homomorphism

$$H_{\Phi}^{n+1}(X;\mathscr{A}^X)\to H_{\Phi\cap(X\setminus x)}^{n+1}(X\setminus x;\mathscr{A}^X|(X\setminus x))$$

cannot be onto. Therefore, from the exact cohomology sequence of the pair $(X,X\setminus x)$ for the sheaf $\mathscr{A}^X$ and the family Φ of supports, we find that

$$H_{\Phi}^{n+2}(X,X\setminus x;\mathscr{A}^X)\neq 0.$$

This, however, implies by 12.1 (p. 59) of [1] that $H_{\Phi\cap\{x\}}^{n+2}(X;\mathscr{A}^X)\neq 0$, a contradiction to the fact that $\mathrm{Dim}_{\mathbb{Z}}(X)=n+1$. Hence $\{x\}$ cannot be Φ-taut in X. Q.E.D.

Remark. We may observe that in the above theorem there is nothing special about the "point" x of $\mathbb{R}^n$; in fact it is easy to modify the same construction to prove that the theorem remains valid if the set $\{x\}$ is replaced by any compact subset K of $\mathbb{R}^n$.

In view of above theorem, the following is now immediate

Corollary. (i) *A compact subset of a Hausdorff space X need not be φ-taut in X, unless $\varphi=cld$.*

(ii) *A closed subset A of a space X* (more strongly, even a retract of a Hausdorff space X) *need not be φ-taut in X unless φ is paracompactifying.*

(iii) *An arbitrary subset of a completely paracompact space X need not be φ-taut in X, unless φ is paracompactifying.*

(iv) *A subset A of X need not be φ-taut in X even if for each $K\in\varphi$, $K\cap A$ has a fundamental system of neighbourhoods in X, unless φ is paracompactifying.*

Remark. The theorem is reminiscent of an example that we gave in [3] to show that there is compact Hausdorff space X such that no point of X is tautly embedded in X with respect to the singular cohomology. The essential difference, however, between the present result and that one is the fact that here X is a n-manifold and there the space was infinite dimensional! In fact finite dimensionality of X here is a crucial point; our technique will not work if X is an infinite dimensional manifold like Hilbert Q-manifold etc.

Acknowledgement. I wish to thank Mr. Subhash Muttepawar for pointing out an argument gap in the earlier version of this paper.

References

1. Bredon, G.E.: Sheaf theory, New York: Mc-Graw Hill 1967
2. Deo, S.: An example of non-exciveness in sheaf cohomology. Proc. Am. Math. Soc. **46**, 501–503 (1975)
3. Deo, S.: On the tautness property of Alexander-Spanier cohomology. Proc. Am. Math. Soc. **52**, 441–444 (1975)
4. Deo. S.: One dimensional manifold is of cohomological dimension two. Proc. Am. Math. Soc. **52**, 445–446 (1975)
5. Deo, S.: Cohomological dimension of a n-manifold is $n+1$. Pacific J. Math. **67**, 154–160 (1976)
6. Dugundji, J.: Topology. Boston, London, Sydney: Allyn and Bacon 1966
7. Sitnikov, K.: Combinatorial topology of non-closed sets I, the first duality law; spectral duality (Russian). Am. Math. Soc. Transl. **15**, 245–295 (1960)
8. Spanier, E.H.: Algebraic topology. New York: Mc-Graw Hill 1966
9. Spanier, E.H.: Tautness for Alexander-Spanier cohomology. Pacific J. Math. **75**, 561–563 (1978)

Received December 20, 1982

Math. Ann. 263, 227–236 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Sums of Powers of Cusp Form Coefficients

R. A. Rankin
Department of Mathematics, University of Glasgow, Glasgow G12 8QW, Scotland

1. Introduction

If a_n is the nth Fourier coefficient of a cusp form of weight k and level N, it is known that [5]

$$\sum_{n \leqq x} |a_n|^2 \sim Ax^k \tag{1}$$

for large x, where A is a positive constant depending on k. So far as I am aware, no non-trivial estimates of

$$\sum_{n \leqq x} |a_n|^{2\beta} \quad (\beta \neq 1)$$

have been obtained, although it may be noted that it follows from (1) and Hölder's inequality that positive constants A_1 and A_2 exist such that

$$\sum_{n \leqq x} |a_n|^{2\beta} \leqq A_1 x^{\beta(k-1)+1} \quad (0 < \beta < 1) \tag{2}$$

and

$$\sum_{n \leqq x} |a_n|^{2\beta} \geqq A_2 x^{\beta(k-1)+1} \quad (\beta > 1) \tag{3}$$

for large x.

Deligne's proof of the Ramanujan-Petersson conjecture implies that, for a newform of integral weight k,

$$|a_n| \leqq n^{(k-1)/2} d(n), \tag{4}$$

where $d(n)$ is the number of positive divisors of n. Since it can be shown that, for $\beta > 0$,

$$\sum_{n \leqq x} \{d(n)\}^{2\beta} = O\{x(\log x)^{2^{2\beta}-1}\}, \tag{5}$$

it follows that

$$\sum_{n \leqq x} |a_n|^{2\beta} = O\{x^{\beta(k-1)+1}(\log x)^{2^{2\beta}-1}\} \tag{6}$$

in this case.

In the present paper I show that, when $\beta > 1$, the exponent $2^{2\beta}-1$ in (6) can be replaced by

$$\gamma = 2^{2(\beta-1)}-1 \tag{7}$$

for integral weight k. It follows, for example, that

$$\sum_{n \leqq x} |\tau(n)|^4 = O(x^{23}\log^3 x), \tag{8}$$

where $\tau(n)$ is Ramanujan's function.

For $0 < \beta < 1$ a similar argument gives results in the opposite direction; see Theorem 4. In particular, it follows that, for some positive constant B,

$$\sum_{n \leqq x} |\tau(n)| \geqq Bx^{13/2}(\log x)^{-1/2} \tag{9}$$

for large x.

2. Notation

For any $\beta > 0$ define γ by (7), so that $\gamma < 0$ for $\beta < 1$ and $\gamma > 0$ for $\beta > 1$.

As in [7], we write $\{N, k, \chi, t\}_0$ for the Hilbert space of cusp forms of level N, weight k, character χ modulo N and divisor t. Here N and k are positive integers, k being even when $N = 1$ or 2; t is a positive divisor of N and $\chi(-1) = (-1)^k$.

Let T be any element of the modular group $\Gamma(1) = \mathrm{SL}(2, \mathbb{Z})$ and write

$$T = \begin{bmatrix} a & b \\ c & d \end{bmatrix}, \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

Then the principal congruence group $\Gamma(N)$ of level N is defined to be

$$\Gamma(N) = \{T \in \Gamma(1) : T \equiv I (\mathrm{mod}\, N)\}$$

and we write, as usual,

$$\Gamma_0(N) = \{T \in \Gamma(1) : c \equiv 0 (\mathrm{mod}\, N)\}.$$

We recall that a cusp form of divisor t has a Fourier series of the form

$$\sum_{\substack{n=1 \\ (n, t_1)=1}}^{\infty} a_n e^{2\pi i n z/t_1};$$

here $t_1 = N/t$. In the particular case when $t = N$ the form belongs to $\Gamma_0(N)$.

Theorem 1. *Let k and N satisfy the conditions stated above and let a_n ($n \in \mathbb{N}$) be the Fourier coefficients of a cusp form of level N, weight k and constant multiplier*

system on $\Gamma(N)$. Then, for each $\beta>1$,

$$\sum_{n\leqq x}|a_n|^{2\beta}=O\{x^{\beta(k-1)+1}\log^{\gamma}x\}.$$

A cusp form satisfying the conditions of the Theorem is a linear combination of cusp forms $f(\delta z)$, where $f(z)$ is a newform belonging to one of the spaces $\{N/d,k,\chi,t/d\}_0$. Here d and t are positive divisors of N, χ is a character modulo N/d of conductor N_χ and it is assumed that

$$d|(t,N/N_\chi),\quad \delta|d\quad(\delta>0).$$

Accordingly, since for any $M\in\mathbb{N}$ and $\alpha>0$

$$\left|\sum_{r=1}^{M}A_r\right|^{\alpha}\leqq\max(1,M^{\alpha-1})\sum_{r=1}^{M}|A_r|^{\alpha},\tag{10}$$

it suffices to prove Theorem 1 for a newform in $\{N,k,\chi,t\}_0$.

We take such a newform to be given by

$$f(z)=\sum_{n=1}^{\infty}\lambda(n)e^{2\pi inz/t_1}\quad(\mathrm{Im}\,z>0).\tag{11}$$

Here $\lambda(1)=1$ and

$$\lambda(mn)=\lambda(m)\lambda(n)\quad\text{for}\quad(m,n)=1.\tag{12}$$

Also, by Theorem 9.4.8 of [7],

$$\lambda(p^{\nu+1})=\lambda(p)\lambda(p^{\nu})-\chi(p)p^{k-1}\lambda(p^{\nu-1})\tag{13}$$

for any prime p and $\nu\in\mathbb{N}$. Moreover,

$$\lambda(n)=\chi(n)\overline{\lambda(n)}\quad\text{for}\quad(n,N)=1.\tag{14}$$

Finally, when $p|N$,

$$\lambda(p^{\nu})=\lambda^{\nu}(p)\quad\text{and}\quad|\lambda(p)|\leqq p^{(k-1)/2}.\tag{15}$$

The last inequality may be deduced, for $t=N$, from Theorem 3 of [3]. It follows in the general case from Theorem 9.4.8 (iii) of [7]. I take this opportunity to observe that, in case (iii) (d) of that theorem, the additional condition $p\nmid(N/N_\chi)$ should be added. The proof of (d) need then only be carried out for the case when $p^2\nmid N$; in this case $l=1$ and the conclusion at the bottom of p. 331 is valid, since $\lambda(l)=1\neq 0$.

It is convenient to write

$$\alpha(n)=\lambda(n)n^{-(k-1)/2}.\tag{16}$$

Then, since $|\alpha(p)|\leqq 2$ for all p, we may define for each $p\nmid N$ an angle $\theta_p\in[0,\pi]$ and a complex number ω_p of unit modulus such that

$$\omega_p^2=\chi(p),\quad\alpha(p)=2\omega_p\cos\theta_p\quad(p\nmid N).\tag{17}$$

It follows from (13) that

$$\alpha(p^\nu)=\omega_p^\nu\frac{\sin(\nu+1)\theta_p}{\sin\theta_p}\quad(\nu\in\mathbb{N},\,p\nmid N). \tag{18}$$

Note also that

$$|\alpha(p^\nu)|\leqq 1\quad(\nu\in\mathbb{N},\,p|N).$$

Theorem 1 now follows easily from

Theorem 2. *If $\beta>1$, then*

$$\sum_{n\leqq x}|\alpha(n)|^{2\beta}=O(x\log^\gamma x).$$

We also prove

Theorem 3. *If $0<\beta<1$, there exists a positive constant B_β such that*

$$\sum_{n\leqq x}|\alpha(n)|^{2\beta}\geqq B_\beta x\log^\gamma x$$

for large x.

To generalize this result we write, as in [7], M^+ for the orthogonal complement in $M=\{N,k,\chi,t\}_0$ of the subspace spanned by its oldforms; then M^+ is spanned by the newforms in M.

Theorem 4. *Let a_n $(n\in\mathbb{N})$ be the Fourier coefficients of a given non-vanishing cusp form f in M^+. Then, for each positive $\beta<1$, there exists a positive constant C_β such that*

$$\sum_{n\leqq x}|a_n|^{2\beta}\geqq C_\beta x^{\beta(k-1)+1}\log^\gamma x$$

for large x.

3. Ancillary Lemmas

Lemma 1. *For all real θ we have*

$$|2\cos\theta|^{2\beta}\leqq 2a(1+\cos 2\theta)\quad(\beta>1)$$

and

$$|2\cos\theta|^{2\beta}\geqq 2a(1+\cos 2\theta)\quad(0<\beta<1),$$

where

$$a=2^{2(\beta-1)}. \tag{19}$$

These results are easily verified.

Lemma 2. *Suppose that the power series*

$$g(z)=\sum_{\nu=0}^{\infty}c_\nu z^\nu$$

has radius of convergence 1 *and that*

$$c_0=1, \quad c_1=2c(1+\cos\varphi), \quad |c_\nu|\leqq(\nu+1)^{2\beta} \quad (\nu\geqq 2),$$

where φ *is real and* $c>0$. *Write*

$$h(z)=(1-z)^{2c}(1-2z\cos\varphi+z^2)^c g(z),$$

so that $h(z)$ *is holomorphic for* $|z|<1$. *Then there exists a positive number* A, *depending on* c *and* β *but not on* φ, *such that*

$$|h(z)-1|\leqq A|z|^2 \quad \textit{for} \quad |z|\leqq 2^{-1/2}.$$

Proof. Write

$$h(z)=\sum_{\nu=0}^{\infty} C_\nu z^\nu,$$

so that $C_0=1$ and $C_1=0$. The function $h(z)-1$ is uniformly bounded in φ for $|z|\leqq 2^{-1/2}$ and for fixed c and β. Therefore, by Schwarz's lemma, so is $\{h(z)-1\}/z$. A further application of Schwarz's lemma gives the desired result.

Before stating the next lemma we define, for any newform f,

$$A_\beta(s)=\sum_{n=1}^{\infty} |\alpha(n)|^{2\beta} n^{-s}, \tag{20}$$

so that, by (3), $A_\beta(s)$ is holomorphic for $\sigma=\mathrm{Re}\, s>1$.

Lemma 3. *For* $\sigma>1$, $A_1(s)$ *possesses the absolutely convergent Euler product*

$$A_1(s)=\prod_{p|N}(1-|\alpha(p)|^2p^{-s})^{-1}\prod_{p\nmid N}\frac{1+p^{-s}}{(1-p^{-s})(1-2p^{-s}\cos 2\theta_p+p^{-2s})}.$$

Accordingly

$$A_1(s)=\frac{\zeta^2(s)\psi(s)}{\zeta(2s)}\prod_{p|N}\frac{1-p^{-s}}{(1+p^{-s})(1-|\alpha(p)|^2p^{-s})},$$

where

$$\psi(s)=\prod_{p\nmid N}(1-2p^{-s}\cos 2\theta_p+p^{-2s})^{-1}. \tag{21}$$

The function $\zeta(s)\psi(s)$ *is holomorphic and non-zero for* $\sigma\geqq 1$, *and so consequently is* $A_1(s)/\zeta(s)$; *here* $\zeta(s)$ *is the Riemann zeta-function.*

Proof. The Euler product for $A_1(s)$ may be derived as in [4] or [6], where only functions belonging to the full modular group $\Gamma(1)$ were considered. That $\zeta(s)\psi(s)$ is holomorphic for $\sigma\geqq 1$ follows from Theorem 4 of [5]. Moreover, $\zeta(s)\psi(s)\neq 0$ for $\sigma\geqq 1$, by Theorem 1 of [6] (in which $\sigma>k$ should read $\sigma\geqq k$); the proof given in [6] extends immediately to newforms of any level and divisor.

Lemma 4. *Suppose that*

$$F(s)=\sum_{n=1}^{\infty} A_n n^{-s}$$

is a Dirichlet series with non-negative coefficients A_n $(n \geqq 1)$ *and is convergent for* $\sigma > 1$. *Suppose further that its sum function* $F(s)$ *can be continued as a holomorphic function for* $\sigma \geqq 1$ *apart from a singularity at* 1 *in the neighbourhood of which*

$$F(s) = (s-1)^{-\omega} G(s),$$

where G *is holomorphic and non-zero at* 1 *and* $\omega > 0$. *Then*

$$\sum_{n \leqq x} A_n \sim \frac{G(1)}{\Gamma(\omega)} x (\log x)^{\omega - 1}$$

as $x \to \infty$.

This extension of Ikehara's theorem is a particular case of a more general result of Delange [1].

4. Proof of Theorem 2

From (12), (15), (16), (18), and (20) we deduce that

$$A_\beta(s) = \prod_{p|N} (1 - |\alpha(p)|^{2\beta} p^{-s})^{-1} \prod_{p \nmid N} A_\beta(s, p),$$

where

$$A_\beta(s, p) = 1 + |2\cos\theta_p|^{2\beta} p^{-s} + \sum_{v=2}^{\infty} \left| \frac{\sin(v+1)\theta_p}{\sin\theta_p} \right|^{2\beta} p^{-vs}.$$

We regard this as a power series in $z = p^{-s}$ and note that, by Lemma 1, its Taylor coefficients are majorized by those of

$$g_p(z) = 1 + 2a(1 + \cos 2\theta_p) z + \sum_{v=2}^{\infty} (v+1)^{2\beta} z^v,$$

where a is given by (19). It follows from Lemma 2 that

$$g_p(p^{-s}) = (1 - p^{-s})^{-2a} (1 - 2p^{-s}\cos 2\theta_p + p^{-2s})^{-a} H_p(s),$$

where

$$|H_p(s) - 1| \leqq A p^{-2\sigma} \quad \text{for} \quad \sigma \geqq \tfrac{1}{2}.$$

Here A does not depend on θ_p nor on p.

The coefficients of the Dirichlet series $A_\beta(s)$ are therefore less in absolute value than those of

$$A_\beta^*(s) = \prod_{p|N} (1 - |\alpha(p)|^{2\beta} p^{-s})^{-1} \prod_{p \nmid N} g_p(p^{-s})$$

$$= \zeta^{2a}(s) \psi^a(s) H(s) \prod_{p|N} \frac{(1 - p^{-s})^{2a}}{1 - |\alpha(p)|^{2\beta} p^{-s}}.$$

Note that $H(s)$ is holomorphic for $\sigma > \frac{1}{2}$. It follows from Lemma 3 that $A_\beta^*(s)$ satisfies the conditions of Lemma 4 with $\omega = a$, so that, if

$$A_\beta^*(s) = \sum_{n=1}^{\infty} \alpha^*(n) n^{-s},$$

then

$$\sum_{n \leqq x} \alpha^*(n) \sim A^* x (\log x)^{a-1},$$

say. Since $|\alpha(n)|^{2\beta} \leqq \alpha^*(n)$ $(n \in \mathbb{N})$, Theorem 2 follows, and with it Theorem 1, as indicated in Sect. 1.

5. Proofs of Theorem 3 and 4

Write

$$B_\beta(s) = \sum_{n=1}^{\infty} b_n n^{-s} = \prod_p (1 + |\alpha(p)|^{2\beta} p^{-s}) \quad (\sigma > 1),$$

where $0 < \beta < 1$ and $\alpha(n)$ is defined as in Sect. 1. Then $B_\beta(s)$ represents a holomorphic function for $\sigma > 1$ and its coefficients are less than the corresponding coefficients of $A_\beta(s)$.

When $p \nmid N$ we have

$$|\alpha(p)|^{2\beta} \geqq 2a(1 + \cos 2\theta_p),$$

by Lemma 1, and so the coefficients of $A_\beta(s)$ are greater in modulus than those of

$$B_\beta^*(s) = \prod_{p \nmid N} \{1 + 2a(1 + \cos 2\theta_p) p^{-s}\}.$$

By Lemma 2, as in Sect. 4, we can write

$$\begin{aligned} B_\beta^*(s) &= \prod_{p \nmid N} (1 - p^{-s})^{-2a} \psi^a(s) K(s) \\ &= \prod_{p \mid N} (1 - p^{-s})^{2a} \zeta^{2a}(s) \psi^a(s) K(s), \end{aligned}$$

where $K(s)$ is holomorphic for $\sigma > \frac{1}{2}$. Theorem 3 now follows from this and Lemma 4.

The deduction of Theorem 4 from Theorem 3 is less simple in this case. As in Sect. 5 of [6], we take a basis $f_1, f_2, \ldots, f_\mu$ of newforms in M^+ so that, for some numbers $\lambda_1, \lambda_2, \ldots, \lambda_\mu$, not all zero,

$$f(z) = \sum_{r=1}^{\mu} \lambda_r f_r(z). \tag{22}$$

We may suppose that $\lambda_1 \neq 0$. Write $a_r(n)$ for the nth Fourier coefficient of $f_r(z)$ $(1 \leqq r \leqq \mu)$ and put

$$a_n = n^{(k-1)/2} \alpha(n), \qquad a_r(n) = n^{(k-1)/2} \alpha_r(n) \quad (1 \leqq r \leqq \mu).$$

We need to introduce the Hecke operators $T_m^{(t)}$ appropriate to the space M and write

$$f(z) | T_m^{(t)} = \sum_{n=1}^{\infty} a^{(m)}(n) e^{2\pi i n t z / N} \quad (m \in \mathbb{N}), \tag{23}$$

where

$$a^{(m)}(n)=n^{(k-1)/2}\alpha^{(m)}(n) \tag{24}$$

and

$$\alpha^{(m)}(n)=m^{(k-1)/2}\sum_{d|(m,n)}\chi(d)\alpha(mn/d^2). \tag{25}$$

Then

$$f_r(z)|T_m^{(t)}=a_r(m)f_r(z) \quad (1\leqq r\leqq\mu, m\in\mathbb{N}). \tag{26}$$

As in [6] we can find numbers $c_1, c_2, \ldots, c_\mu$ such that

$$\sum_{m=1}^{\mu} c_m a_r(m)=\begin{cases}1 & (r=1),\\ 0 & (1<r\leqq\mu),\end{cases} \tag{27}$$

since the dimension μ is less than the number of zeros of any form.

Now consider the cusp form g defined by

$$g(z)=\sum_{m=1}^{\mu} c_m f(z)|T_m^{(t)}. \tag{28}$$

By (22), (26), and (27),

$$\begin{aligned} g(z)&=\sum_{m=1}^{\mu}\sum_{r=1}^{\mu} c_m\lambda_r a_r(m)f_r(z)\\ &=\sum_{r=1}^{\mu}\lambda_r f_r(z)\sum_{m=1}^{\mu} c_m a_r(m)\\ &=\lambda_1 f_1(z). \end{aligned} \tag{29}$$

If Theorem 4 is false for f, then, given any $\varepsilon>0$, there exists X such that

$$\sum_{n\leqq x}|\alpha(n)|^{2\beta}<\varepsilon x\log^{\gamma}x \tag{30}$$

for all $x>X$. In what follows $A_1, A_2, \ldots, A_5$ denote positive constants depending only on f, μ, and β. By (29), (28), (23)–(25), and (10), we have

$$S=\sum_{n\leqq x}|\lambda_1|^{2\beta}|\alpha_1(n)|^{2\beta}\leqq A_1\sum_{m=1}^{\mu}|c_m|^{2\beta}m^{\beta(k-1)}\sum_{d|(m,n)}|\alpha(mn/d^2)|^{2\beta}.$$

Write $m=hd$, $n=ld$. Then

$$\begin{aligned} S&\leqq A_2\sum_{d\leqq\mu}\sum_{h\leqq\mu/d}\sum_{l\leqq x/d}|\alpha(hl)|\\ &\leqq A_3\sum_{d\leqq\mu}\frac{\mu}{d}\sum_{n\leqq\mu x/d^2}|\alpha(n)|\\ &\leqq\varepsilon A_4\sum_{d\leqq\mu}d^{-3}\{\log(\mu x/d^2)\}^{\gamma}\\ &\leqq\varepsilon A_5 x\log^{\gamma}x, \end{aligned}$$

by (30), provided that x is sufficiently large. This contradicts Theorem 3, if ε is sufficiently small.

6. A Result of Elliott

Let f be a newform in $\{N, k, \chi, t\}_0$ and write, as in Sect. 5,

$$a(n) = n^{(k-1)/2}\alpha(n). \tag{31}$$

Put

$$M_\beta = \lim_{x\to\infty} \frac{1}{x} \sum_{n\leqq x} |\alpha(n)|^{2\beta} \quad (0<\beta<1). \tag{32}$$

Then Elliott [2] has shown that M_β exists for each β with $0<\beta<1$; however, its value may be zero.

Moreover, he has proved that either (i) $M_\beta = 0$ for all $\beta \in]0,1[$, or (ii)

$$\sum_p \frac{1}{p} \{|\alpha(p)|^2 - 1\}^2 < \infty. \tag{33}$$

By Theorem 2 of [6] (which remains valid for a newform of any level),

$$\sum_{p\leqq x} |\alpha(p)|^2 \log p \sim x \tag{34}$$

as $x\to\infty$. Accordingly, we deduce that, if

$$\sum_{p\leqq x} \frac{|\alpha(p)|}{p} \sim \alpha \log\log x \tag{35}$$

for some $\alpha \neq 1$, then $M_\beta = 0$ for all $\beta \in]0,1[$.

If we know the probability distribution of the angles θ_p associated with f, we may be able to check whether (35) holds for some α. Elliott has shown that, if the Sato-Tate conjecture holds for $a(n) = \tau(n)$, then (35) follows for a value $\alpha \neq 1$. In fact $\alpha = 8/(3\pi)$ under this supposition; note that Elliott's evaluation of his constant c on p. 406 of [2] is in error, the correct value being $2-16/(3\pi)$. It follows that, provided the Sato-Tate conjecture holds, $M_\beta = 0$ for $\tau(n)$ for all $\beta \in]0,1[$; in particular

$$\sum_{n\leqq x} |\tau(n)| = o(x^{13/2}), \tag{36}$$

which may be compared with (9).

As I shall show in a subsequent paper, there exist cusp forms, whose Fourier coefficients are essentially *Grössencharaktere*, for which the angles θ_p are uniformly distributed, and for which (35) holds with the value $\alpha = 2/\pi \neq 1$; accordingly, in these cases, it follows that

$$\sum_{n\leqq x} |a(n)|^{2\beta} = o(x^{\beta(k-1)+1}) \quad (0<\beta<1). \tag{37}$$

For these coefficients further improvements in the results of this paper can be obtained.

References

1. Delange, H.: Généralisation du théorème de Ikehara. Ann. Sci. Ecole Norm. Sup. **71**, 213–242 (1954)
2. Elliott, P.D.T.A.: Multiplicative functions and Ramanujan's τ-function. J. Austral. Math. Soc. (Ser. A) **30**, 461–468 (1981)
3. Li, W.C.W.: Newforms and functional equations. Math. Ann. **212**, 285–315 (1975)
4. Rankin, R.A.: Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. I. The zeros of the function $\sum_{n=1}^{\infty} \frac{\tau(n)}{n^s}$ on the line $\Re s = \frac{13}{2}$. Proc. Cambridge Phil. Soc. **35**, 351–356 (1939)
5. Rankin, R.A.: Contributions etc. II. The order of the Fourier coefficients of integral modular forms. Proc. Cambridge Phil. Soc. **35**, 357–372 (1939)
6. Rankin, R.A.: An Ω-result for the coefficients of cusp forms. Math. Ann. **203**, 239–250 (1973)
7. Rankin, R.A.: Modular forms and functions. Cambridge: Cambridge University Press 1977

Received December 29, 1982

Math. Ann. 263, 237–250 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Shilov Points and Shilov Boundaries

Rainer Wittmann

Katholische Universität Eichstätt, Mathematisch-Geographische Fakultät, Ostenstrasse 26–28, D-8078 Eichstätt, Federal Republic of Germany

Introduction

Let F be a set of lower semi-continuous functions on a compact space (X, τ). A closed set $K \subset X$ is called a Shilov set iff $u \geqq 0$ on K implies $u \geqq 0$ throughout X. It is an important question whether there exists a smallest Shilov set which will then be called the Shilov boundary. Many different methods to solve this question have already been developed [2–4, 11]. In this paper we develop a new method which yields the most general existence result of the Shilov boundary containing all previous results as corollaries. The following notion is fundamental: A point $x \in X$ is called a Shilov point iff, for any closed $K \subset X$, $x \notin K$, $\alpha < 0$, and $\beta > 1$, there exists $u \in F$ such that $u(x) < 0$, $u \geqq \alpha \cdot u(x)$ on K and $u \geqq \beta \cdot u(x)$. Obviously every Shilov set contains all Shilov points. Hence, if a minimum principle holds with respect to the Shilov points, then the Shilov boundary exists. Under certain circumstances also the converse is true (Sect. 1). In Sect. 2 we prove this minimum principle for very general sets of functions F. In Sect. 3 we study the relation between the Choquet boundary and the Shilov points. As a corollary we get Bauer's minimum principle in its most general form due to Boboc and Cornea. For a special situation Shilov points were already introduced by Fuchssteiner and Hackenbroch in [9] (see also [10]) and Corollary 3.6, Proposition 3.8 are essentially due to them. Besides these two results, Theorem 3.4 is worth mentioning. It allows to give rather simple proofs and generalizations of Korovkin theorems in $C(X)$. We have omitted these applications because they are easy to verify for the expert and because they would lead too far away from our general theme.

Let now F be a convex cone and let τ' be the coarsest topology finer than τ such that all functions in F are continuous with respect to τ'. In Sect. 4 we show that the Shilov boundary with respect to this "fine" topology also exists. Moreover, in Sect. 5, we show for simplicial cones that this "fine" Shilov boundary is exactly the Choquet boundary.

1. The Relation Between Shilov Boundaries and Shilov Points

In the sequel let (X, τ) be a topological space and F be a set of numerical functions on X. For any $A \subset X$, $\bar{A}$ denotes the closure of A.

A point $x \in X$ is called a *weak Shilov point* (with respect to F and τ) iff, for any closed set $K \subset X$, $x \notin K$, there exists $u \in F$ such that $u(x) < 0$ and $u(y) \geqq 0$ for any $y \in K$. We denote by $S_w(F)$ or $S_w(F, \tau)$ the set of all weak Shilov points with respect to F and τ.

A point $x \in X$ is called a *Shilov point* (with respect to F and τ) iff, for any closed set $K \subset X$, $x \notin K$, $\alpha < 0$, and $\beta > 1$, there exists $u \in F$ such that $u(x) < 0$, $u(y) \geqq \alpha \cdot u(x)$ for any $y \in K$ and $u \geqq \beta \cdot u(x)$. We denote by $S(F)$ or $S(F, \tau)$ the set of all Shilov points with respect to F and τ.

The following lemma is obvious.

1.1. Lemma. *Assume that $u + \alpha \in F$ for any $u \in F$, $\alpha \in \mathbb{R}$. Let $u \in F$ be continuous and $x \in X$ be such that $u(x) < u(y)$ for any $y \in X$, $x \neq y$. Then x is a Shilov point.*

A set $K \subset X$ is called a *Shilov set* (with respect to F and τ) iff K is closed and $u \geqq 0$ on K implies $u \geqq 0$ for any $y \in F$. If there exists a smallest Shilov set then we call this set the Shilov boundary of F (with respect to τ) and we write $\partial_F X$ or $\partial_{F,\tau} X$.

Obviously we have for any Shilov set K.

$$S(F) \subset S_w(F) \subset K .$$

1.2. Theorem. *Assume that $\partial_F X$ exists, that the functions in F are upper semi-continuous and lower bounded, that $u + \alpha \in F$ for any $u \in F$ and $\alpha \in \mathbb{R}_+$ and let τ^* be a completely metrizable topology coarser than τ.*

Then we have $\partial_F X \subset \overline{S(F, \tau^*)}$ (the closure is taken with respect to τ).

Proof. Let U be open such that $U \cap \partial_F X \neq \emptyset$ and let $\alpha < 0$, $\beta > 1$ be given. Then there exists $u' \in F$ such that $u' \geqq 0$ on $X \setminus U$ and $\{u' < 0\} \cap \partial_F X \neq \emptyset$, otherwise $X \setminus U$ would be a Shilov set contradicting the minimality of $\partial_F X$. We denote $\delta := \inf\{u'(x) : x \in X\}$ and take $\gamma > 0$ such that $\gamma \geqq \alpha \cdot (\gamma + \delta)$, $\gamma + \delta < 0$. Then we have for $u := u' + \gamma$ and

$$U' := \left\{ u < \frac{1}{\beta} \cdot (\gamma + \delta) \right\}$$

(1) $u \in F$, $U' \neq \emptyset$, and U' is open,

(2) $u(x) < 0$ for any $x \in U'$,

(3) $u \geqq \beta \cdot u(x)$ for any $x \in U'$,

(4) $u(y) \geqq \alpha \cdot u(x)$ for any $x \in U'$ and $y \in X \setminus U$.

Since $u'' := u - \frac{1}{\beta}(\gamma + \delta) \in F$, $\{u'' < 0\} = U' \neq \emptyset$ and $\partial_F X$ is a Shilov set, we get

(5) $U' \cap \partial_F X \neq \emptyset$.

We fix now a complete metric of τ^* and, for any $A \subset X$, we denote by $\operatorname{diam}(A)$ the diameter of A with respect to this metric. Let x_0 be a point of $\partial_F X$ and U be an open neighborhood of x_0. By the above construction we get inductively a sequence (U_n) of open sets and a sequence (u_n) of functions in F such that for any $n \in \mathbb{N}$

(6) $\bar{U}_{n+1} \subset U_n \subset \bar{U}_n \subset U$,

(7) $\operatorname{diam}(U_n) \leqq \frac{1}{n}$,

(8) $U_n \cap \partial_F X \neq \emptyset$,

(9) $u_n(x) < 0$ for any $x \in U_{n+1}$,

(10) $u_n \geqq \left(1 + \frac{1}{n}\right) \cdot u_n(x)$ for any $x \in U_{n+1}$,

(11) $u_n(y) \geqq -n \cdot u_n(x)$ for any $x \in U_{n+1}$ and $y \in X \setminus U_n$.

The completeness implies that there exists a point $x' \in X$ such that $\{x'\} = \bigcap_n U_n \subset U$.

Let $K \subset X$ be τ^*-closed, $x' \notin K$, $\alpha < 0$, and $\beta > 1$. Because of (7), there exists $n \in \mathbb{N}$ such that $U_n \cap K = \emptyset$, $\alpha \geqq -n$, and $\beta \geqq 1 + \frac{1}{n}$. Then we have $u_n(x') < 0$,

$$u_n(y) \geqq -n \cdot u_n(x') \geqq \alpha \cdot u(x')$$

for any $y \in K \subset X \setminus U_n$ and

$$u_n \geqq \left(1 + \frac{1}{n}\right) \cdot u_n(x) \geqq \beta \cdot u_n(x).$$

Thus $x' \in U$ is a Shilov point.

1.3. Corollary. *Assume that X is completely metrizable that the functions in F are continuous and lower bounded and that $u + \alpha \in F$ for any $u \in F$ and $\alpha \in \mathbb{R}_+$. Then the following conditions are equivalent*

(i) *$\partial_F X$ exists.*

(ii) *$\partial_F X$ exists and $\overline{S_w(F)} = \overline{S(F)} = \partial_F X$.*

(iii) *$u \geqq 0$ on $S_w(F)$ implies $u \geqq 0$ for any $u \in F$.*

(iv) *$u \geqq 0$ on $S(F)$ implies $u \geqq 0$ for any $u \in F$.*

Proof. (i)⇒(iv). From Theorem 1.1 we get $\overline{S(F)} = \partial_F X$. Let $u \in F$ be such that $u \geqq 0$ on $S(F)$. The continuity of u yields $u \geqq 0$ on $\overline{S(F)} = \partial_F X$. Since $\partial_F X$ is a Shilov set we get $u \geqq 0$. (iv)⇒(iii) and (ii)⇒(i) are obvious. (iii)⇒(ii): Because of (ii), $\overline{S_w(F)}$ is a Shilov set. For any other Shilov set K we have $S_w(F, \tau) \subset K$. Hence $\overline{S_w(F)}$ is the Shilov boundary of F. We can now use (i)⇒(iv) to see that $\overline{S(F)}$ is also a Shilov set and $\overline{S(F)} \subset \overline{S_w(F)} = \partial_F X$. This implies condition (ii).

1.4. Remarks. (1) A further application of Theorem 1.2, closely related to Corollary 1.3, is given in Corollary 4.4.

(2) In the case when X is compact and F is a convex cone of lower semi-continuous functions containing the constants and separating the points, Bauer proved in [3] that $\partial_F X$ exists and that the Shilov boundary is the closure of the Choquet boundary. As the following example shows, the last fact may be no more true, if the convexity assumption is dropped.

1.5. Example. Let X be the closed intervall $[0, 1]$ and let U be an open subset of X. We consider the set

$$F := \{f \in C(X) : \exists a, b, c \in \mathbb{R}, y \in U, f(x) = ax^2 + bx + c, \forall x \in X, 2ay + b = 0, a \geqq 0\}.$$

We have

(1) $u+\alpha\in F$ for any $u\in F$ and $\alpha\in\mathbb{R}$.

(2) For any $u,v\in F$ there exist $\alpha\geqq 0$ such that $u+\beta v\in F$ for any $\beta\geqq\alpha$.

(3) F is point separating.

(4) X is the Choquet boundary of F.

(5) For any $u\in F$, $u\neq 0$ there exists a unique point $x\in U$ such that $u(x)=\inf\{u(y):y\in X\}$.

From (5) we get

(6) $\partial_F X$ exists and $\partial_F X=\bar{U}$.

(7) Every point in U is a Shilov point.

One may even show

(8) $\bar{U}=S(F)=S_w(F)$.

This example shows also that the Choquet boundary is in general not contained in $S(F)$ or $S_w(F)$.

2. The Minimum Principle and the Existence of Shilov Boundaries

Throughout this section X denotes a compact space and F a set of lower semi-continuous functions on X (taking values in $\mathbb{R}\cup\{+\infty\}$).

2.1. Theorem. *Assume that F satisfies the following conditions*

(C) For any $u,v\in F$, $x\in X$, and $\alpha>0$ there exists $\beta\geqq\alpha$ and $s\in F$ such that $s\geqq u+\beta\cdot v$ and $s(x)=u(x)+\beta\cdot v(x)$.

(P) There exists a strictly positive, finite function $w\in F$ such that $u+\alpha w\in F$ for any $\alpha\in\mathbb{R}_+$ and $u\in F$.

(S) For any $x,y\in X$, $x\neq y$, there exist $u,v\in F$ such that

$$u(x)\cdot v(y)\neq u(y)\cdot v(x) \quad \text{and} \quad u(x),v(x),u(y),v(y)<+\infty .$$

Then, for any $u\in F$ such that $\{u<0\}\neq\emptyset$, there exists a Shilov point $x\in X$ such that $u(x)<0$.

If also the following condition holds

(N) For any $x\in X$, there exists $u\in F$ such that $u(x)<0$.

Then, for any $u\in F$ such that $\{u\leqq 0\}\neq\emptyset$, there exists a Shilov point $x\in X$ such that $u(x)\leqq 0$.

For the proof we need the following lemma.

2.2. Lemma. *Let $D\subset X$ be closed and let f,g be two lower semi-continuous functions on X such that $g\geqq 0$ on D and $f(x)>0$ for any $x\in D\cap\{g=0\}$.*

(a) Then there exists $\alpha\in\mathbb{R}_+$ such that $(f+\beta g)(x)>0$ for any $\beta>\alpha$ and $x\in D$.

(b) If there exists $x\in D$ such that $f(x)\leqq 0$ and $g(x)<\infty$, then the above α may be chosen so that $\{f+\alpha g=0\}\neq\emptyset$.

Proof. We may assume that $D=X$. We consider the function $h:X\to\mathbb{R}$ defined by

$$h(x):=\begin{cases}0 & \text{if} \quad f(x)>0 \quad \text{or} \quad g(x)=\infty\\ -\dfrac{f(x)}{g(x)} & \text{otherwise.}\end{cases}$$

One verifies that h is upper semi-continuous. Obviously $\alpha := \sup\{h(y): y \in X\}$ possesses the required properties.

Proof of the theorem. For any $y \in X \setminus S(F)$, there exists a closed set $K_y \subset X$, $y \notin K_y$, $\alpha_y < 0$, and $\beta_y > 1$ such that there exists no function $u \in F$ having the following properties

$$u(y) < 0, \quad \inf\{u(z): z \in K_y\} > \alpha_y \cdot u(y),$$
$$\inf\{u(z): z \in X\} > \beta_y \cdot u(y).$$

We consider the system $\mathfrak{C}$ of all closed subsets $D \subset X$, $D \neq \emptyset$, such that there exist no $y \in X \setminus S(F)$ and $u \in F$ having the following properties

$$u(y) < 0, \quad \inf\{u(z): z \in K_y \cap D\} > \alpha_y \cdot u(y),$$
$$\inf\{u(z): z \in D\} > \beta_y \cdot u(y).$$

Obviously we have $X \in \mathfrak{C}$. We prove first

(1) For any $D \in \mathfrak{C}$, $v \in F$ such that $v \geqq 0$ on D and $D \cap \{v = 0\} \neq \emptyset$, $D \cap \{v = 0\}$ again belongs to $\mathfrak{C}$.

We denote $D' := D \cap \{v = 0\}$. Assume that there exists $y \in D' \setminus S(F)$ and $u \in F$ such that $u(y) < 0$,

$$\inf\{u(z): z \in K_y \cap D'\} > \alpha_y \cdot u(y)$$

and $\inf\{u(z): z \in D\} > \beta_y \cdot u(y)$. Applying Lemma 2.2 to $K_y \cap D$, $u - \alpha_y \cdot u(y)$, and v we get $\lambda_1 \in \mathbb{R}_+$ such that

(2) $(u + \lambda_1 v)(z) > \alpha_y \cdot u(y)$ for any $z \in K_y \cap D$.

Applying Lemma 2.2 to D, $u - \beta_y \cdot u(y)$ and v we get $\lambda_2 \in \mathbb{R}_+$ such that

(3) $(u + \lambda_2 v)(z) > \beta_y \cdot u(y)$ for any $z \in D$.

According to (C), there exists $\lambda \geqq \max(\lambda_1, \lambda_2)$, and $s \in F$ such that $s \geqq u + \lambda \cdot v$ and $s(y) = u(y) + \lambda v(y)$. Using (2) and (3) we get

$$\begin{aligned} s(y) = (u + \lambda v)(y) = u(y) < 0, \quad & \inf\{s(z): z \in K_y \cap D\} \\ \geqq \inf\{(u + \lambda v)(z): z \in K_y \cap D\} &> \alpha_y \cdot u(y) = \alpha_y \cdot (u + \lambda v)(y) \\ = \alpha_y \cdot s(y), \inf\{s(z): z \in D\} &\geqq \inf\{(u + \lambda v)(z): z \in D\} > \beta_y \cdot u(y) \\ = \beta_y \cdot (u + \lambda v)(y) &= \beta_y \cdot s(y). \end{aligned}$$

Since this is a contradiction to $D \in \mathfrak{C}$, (1) is proved.

Next we show

(4) $\mathfrak{C}$ is inductively ordered with respect to inclusion.

Let $\mathfrak{C}'$ be a totally ordered subset of $\mathfrak{C}$ and denote $D_0 = \bigcap \mathfrak{C}'$. Assume that there exist $y \in D_0 \setminus S(F)$ and $u \in F$ such that $u(y) < 0$,

$$\inf\{u(z): z \in K_y \cap D_0\} > \alpha_y \cdot u(y)$$

and

$$\inf\{u(z): z \in D_0\} > \beta_y \cdot u(y).$$

Since

$$\bigcap_{D\in\mathfrak{C}'} K_y\cap D = K_y\cap D_0 \subset \{u>\alpha_y\cdot u(y)\}$$

and

$$\bigcap \mathfrak{C}' = D_0 \subset \{u>\beta_y\cdot u(y)\}$$

there exists $D\in\mathfrak{C}'$ such that $K_y\cap D\subset\{u>\alpha_y\cdot u(y)\}$ and $D\subset\{u>\beta_y\cdot u(y)\}$. However, this is impossible because $D\in\mathfrak{C}'\subset\mathfrak{C}$.

Let now $u\in F$ be such that there exists $u'\in F$, $\{u\leqq 0\}\cap\{u'<0\}\neq\emptyset$. Both assertions of the theorem are proved if we have found a Shilov point x such that $u(x)\leqq 0$ [resp. $u(x)<0$ if $\{u<0\}\neq\emptyset$]. Let $w\in F$ be a function as in (*P*).

Applying Lemma 2.2 twice we get $\alpha,\alpha'\geqq 0$ such that $u+\alpha w\geqq 0$, $\{u+\alpha w=0\}\neq\emptyset$, $u'+\alpha' w\geqq 0$ on $\{u+\alpha w=0\}$ and $D:=\{u+\alpha w=0\}\cap\{u'+\alpha' w\}\neq\emptyset$. We have $D\subset\{u'<0\}$ and, by (1) and (*P*), $D\in\mathfrak{C}$. By (4) and Zorn's lemma, there exists $D'\in\mathfrak{C}$, $D'\subset D$ such that D' is minimal with respect to inclusion.

Assume that there exist two points $x_1,x_2\in D'$, $x_1\neq x_2$. According to (*S*), there exist s_1,s_2 such that

$$s_1(x_1)\cdot s_2(x_2)\neq s_1(x_2)\cdot s_2(x_1)$$

and $s_i(x_j)<\infty$ for $i,j=1,2$. Thus we have, for $i=1$ or $i=2$,

$$s_i(x_1)\cdot u'(x_2)\neq s_i(x_2)\cdot u'(x_1).$$

Hence there exist $t\in F$ and $x,y\in D'$ such that $t(x)\cdot u'(y)>t(y)\cdot u'(x)$.

Since $u'<0$ on D' we have

$$(t+\lambda u')(x)\cdot u'(y)>(t+\lambda u')(y)\cdot u'(x)$$

for any $\lambda\geqq 0$. We choose $\lambda\geqq 0$ such that $(t+\lambda u')(x)<0$. According to (*C*), there exist $\lambda'\geqq\lambda$ and $t'\in F$ such that $t'\geqq t+\lambda' u'$ and $t'(x)=(t+\lambda' u')(x)$. Now we get

$$t'(x)\cdot u'(y)=(t+\lambda' u')(x)\cdot u'(y)>(t+\lambda' u')(y)\cdot u'(x)\geqq t'(y)\cdot u'(x).$$

Hence, for $t''=t'$ or $t''=u'$, we have

$$t''(x)\cdot w(y)\neq t''(y)\cdot w(x),\qquad t''(x)<0. \tag{5}$$

From Lemma 2.2 we get $\alpha''\geqq 0$ such that $t''+\alpha'' w\geqq 0$ on D' and $D''=\{t''+\alpha'' w=0\}\cap D'\neq\emptyset$. By (1) and (*P*), $D'\in\mathfrak{C}$ and hence $D'=D''$ since D' is minimal. But $D'=D''$ is a contradiction to (5). Thus, there exists exactly one point $x_0\in X$ such that $D'=\{x_0\}$. x_0 is a Shilov point otherwise we would get

$$u'(x_0)<0,\qquad \inf\{u'(z):z\in K_{x_0}\cap D'\}=\inf\emptyset=+\infty$$
$$>\alpha_{x_0}\cdot u(x_0),\qquad \inf\{u'(z):z\in D'\}=u'(x_0)>\beta_{x_0}\cdot u'(x_0)$$

contradicting $D'\in\dot{C}$. Since $x_0\in D'\subset D\subset\{u+\alpha w=0\}$ we have $u(x_0)\leqq(u+\alpha w)(x_0)=0$. This proves the second assertion. Under the assumptions of the first assertion we have $\alpha>0$ and hence $u(x_0)<(u+\alpha w)(x_0)=0$.

2.3. Corollary. *If F satisfies the conditions (C), (P), and (S), then $\partial_F X$ exists and we have $\overline{S(F)}=\partial_F X$.*

Proof. By Theorem 2.1, $\overline{S(F)}$ is a Shilov set. On the other hand, if K is a Shilov set, then $S(F)\subset K$ and hence $\overline{S(F)}\subset K$.

2.4. Corollary. *Assume that F satisfies (S) and the following conditions*

(A) For any $u,v\in F$, we have $u+v\in F$

(P') There exists a strictly positive, finite function in F.

Then, for any $u\in F$ such that $\{u<0\}\neq\emptyset$, there exists a Shilov point x such that $u(x)<0$. Moreover $\partial_F X$ exists and $\partial_F X=\overline{S(F)}$.

The corollary is an immediate consequence of Theorem 2.1 and the following lemma:

2.5. Lemma. *Assume that F satisfies (A). Let $\tilde{F}$ be the convex cone generated by F. Then we have $S(F)=S(\tilde{F})$.*

Proof. We denote $F_1:=\{q\cdot u: q\in\mathbb{Q}, u\in F\}$. We claim that $q_1\cdot u_1+q_2\cdot u_2\in F_1$ for any $u_1,u_2\in F$ and $q_1,q_2\in\mathbb{Q}$. Indeed, by (A) there exists $n\in\mathbb{N}$ such that $n\cdot q_1\cdot u_1\in F$, $n\cdot q_2\cdot u_2\in F$. Thus, again by (A), we have

$$n\cdot(q_1\cdot u_1+q_2\cdot u_2)\in F$$

and therefore $(q_1\cdot u_1+q_2\cdot u_2)\in F_1$. Obviously we have $S(F)=S(F_1)\subset S(\tilde{F})$.

Let now x be a Shilov point with respect to $\tilde{F}$, $K\subset X$, $x\notin K$ a closed set, $\alpha<0$ and $\beta>1$. Then there exists $f=\sum_{i=1}^{n}\alpha_i u_i$ $(\alpha_i\geqq 0, u_i\in F)$ such that $f(x)<0$,

$$\inf\{f(z): z\in K\}>\alpha\cdot f(x) \quad\text{and}\quad \inf\{f(z): z\in X\}>\beta\cdot f(x).$$

We take $\gamma>1$ such that

$$\inf\{f(z): z\in K\}>\alpha\cdot\gamma\cdot f(x), \qquad \inf\{f(z): z\in X\}>\beta\cdot\frac{1}{\gamma}\cdot f(x),$$

and $\alpha_i'\in\mathbb{Q}_+$ $(1\leqq i\leqq n)$ such that $\inf\{f'(z): z\in K\}>\alpha\cdot\gamma\cdot f(x)$,

$$\inf\{f'(z): z\in X\}>\beta\cdot\gamma\cdot f(x) \qquad \gamma\cdot f(x)<f'(x)<\frac{1}{\gamma}f(x),$$

where $f':=\sum_{i=1}^{n}\alpha_i' u_i\in F_1$. Hence x is also a Shilov point with respect to F_1. Since obviously $S(F_1)\subset S(\tilde{F})$ the proof is complete.

2.6. Remark. In a similar way one may prove under the assumptions of Lemma 2.5 that $S_w(F)=S_w(\tilde{F})$ and that the Shilov sets with respect to F are exactly the Shilov sets with respect to $\tilde{F}$. Hence $\partial_F X$ exists iff $\partial_{\tilde{F}} X$ exists.

A closed subset $K\subset X$ is called a *minimum set* (with respect to F) iff, for any $u\in F$, $\inf\{u(z): z\in K\}=\inf\{u(z): z\in X\}$. If a smallest minimum set exists, we call this set the *minimum boundary* of F and we write $\partial_F^{\min} X$. In former papers $\partial_F^{\min} X$ is called the Shilov boundary.

The proof of the following lemma is left to the reader.

2.7. Lemma. *We denote $F^*:=\{u+\alpha: u\in F, \alpha\in\mathbb{R}\}$.*

Then the minimum sets with respect to F are exactly the Shilov sets with respect to F^*.

From Corollary 2.3 and Lemma 2.7 we get

2.8. Corollary. *Assume that F satisfies condition (C) and*
(S') For any $x, y \in X$, $x \neq y$ there exist $u \in F$ such that

$$u(x) \neq u(y) \quad \text{and} \quad u(x), u(y) < +\infty .$$

Then $\partial_F^{\min} X$ exists and $\overline{S(F^*)} = \partial_F^{\min} X$.

2.9. Remark. Some special cases of Corollary 2.8 have already been proved by many authors (see for example [2, 4]). The most general one is due to Bauer. In [3] he proved that $\partial_X^{\min} F$ exists if F satisfies (A) and (S'). Example 1.5 shows that Bauer's assumptions are essentially stronger than our assumptions.

2.10. Corollary. *Let H be a complex vector space of complex valued, continuous functions on a compact space X such that, for any $x, y \in X$, $x \neq y$, there exists $u \in H$ with $|u(x)| \neq |u(y)|$. Let $F := \{\alpha - |u| : u \in H\}$. Then $\partial_F X$ exists and is the smallest set of all closed sets $K \subset X$ such that*

$$(*) \qquad \sup\{|u(x)| : x \in K\} = \sup\{|u(x)| : x \in X\} \quad \text{for any} \quad u \in H .$$

Proof. Obviously, a closed set $K \subset X$ satisfies $(*)$ iff it is a Shilov set with respect to F. Hence the proof is finished if conditions (C), (P), (S) are shown for F. (P) and (S) are immediate. Let $x \in X$ and $v_1 = \alpha_1 - |u_1|$, $v_2 = \alpha_2 - |u_2|$ where $u_1, u_2 \in H$ and $\alpha_1, \alpha_2 \in \mathbb{R}$. Then there exists $\lambda_1, \lambda_2 \in \mathbb{C}$ such that $|\lambda_1| = |\lambda_2| = 1$ and $\lambda_1 u_1(x)$, $\lambda_2 u_2(x) \in \mathbb{R}_+$. Setting $u := \lambda_1 u_2 + \lambda_2 u_2$, $v := \alpha_1 + \alpha_2 - |u|$ we get

$$v \geqq \alpha_1 - |\lambda_1 \cdot u_1| + \alpha_2 - |\lambda_2 u_2| = v_1 + v_2$$

and

$$\begin{aligned} v(x) &= \alpha_1 - |\lambda_1 \cdot u_1(x)| + \alpha_2 - |\lambda_2 u_2(x)| \\ &= \alpha_1 - \lambda_1 u_1(x) + \alpha_2 - \lambda_2 u_2(x) = v_1(x) + v_2(x) . \end{aligned}$$

In particular, F satisfies (C).

2.11. Remark. Corollary 2.10 was first shown by Bear in [4] under the additional condition that the complex constants belong to H.

3. Choquet Boundary and Shilov Points

We recall the notion of Choquet boundary

3.1. Let X be a compact space and F a set of lower semi-continuous functions on X.

A Radon measure μ on X is called a representing measure of a point $x \in X$ iff, for any $u \in F$, $\int u d\mu \leqq u(x)$. For any $x \in X$, we denote by $M_x(F)$ the set of all representing measures of x. The set of all points $x \in X$ such that $M_x(F) = \{\varepsilon_x\}$ is called the Choquet boundary and we write $\mathrm{Ch}(F)$.

3.2. Proposition. *Let X be a compact space and F a set of lower semi-continuous functions on X containing a strictly positive, finite function. Then we have $S(F) \subset \mathrm{Ch}(F)$.*

Proof. Let x be a Shilov point and $\mu \in M_x(F)$. Let U be an open neighborhood of x, $\alpha < 0$, and $\beta > 1$. Since x is a Shilov point, there exists $u \in F$ such that $u(x) < 0$, $\inf\{u(z) : z \in X \setminus U\} \geqq \alpha \cdot u(x)$ and $\inf\{u(z) : z \in X\} \geqq \beta \cdot u(x)$. From

$$u(x) \geqq \int u d\mu = \int_U u d\mu + \int_{X \setminus U} u d\mu \geqq \beta \cdot u(x) \cdot \mu(U) + \alpha \cdot u(x) \cdot \mu(X \setminus U)$$

we get

$$u(x) \cdot (1 - \beta\mu(U)) \geqq u(x) \cdot \alpha \cdot \mu(X \setminus U),$$

$$\mu(X \setminus U) \leqq \frac{1 - \beta\mu(U)}{\alpha}.$$

Letting α decrease to $-\infty$ we get $\mu(X \setminus U) = 0$. Since U is an arbitrary open neighborhood of x, we get $\mu(X \setminus \{x\}) = 0$ and hence $\mu = \gamma \cdot \varepsilon_x$, $\gamma \in \mathbb{R}_+$. Let $w \in F$ be such that $0 < w(x) < \infty$. Then we have $w(x) \geqq \int w \, d\mu = \gamma \cdot w(x)$ and therefore $\gamma \leqq 1$.

Since x is a Shilov point there exists $u \in F$ such that $u(x) < 0$. Then we have $u(x) \geqq \int u \, d\mu = \gamma \cdot u(x)$ and therefore $\gamma \geqq 1$. Hence $x \in \mathrm{Ch}(F)$.

From Theorem 2.1 and Proposition 3.2 we get the most general version of Bauer's minimum principle due to Boboc and Cornea [8, p. 26].

3.3. Corollary. *Let X be a compact space and F a set of lower semi-continuous functions on X, containing a strictly positive, finite function and satisfying condition (S).*

Then, for any $u \in F$ such that $\{u < 0\} \neq \emptyset$ there exists $x \in \mathrm{Ch}(F)$ such that $u(x) < 0$.

If F satisfies also condition (N), then, for any $u \in F$ such that $\{u \leqq 0\} \neq \emptyset$, there exists $x \in \mathrm{Ch}(F)$ such that $u(x) \leqq 0$.

Proof. We consider the convex cone $\tilde{F}$ generated by F. Obviously we have $\mathrm{Ch}(\tilde{F}) = \mathrm{Ch}(F)$. An application of Theorem 2.1 and Proposition 3.2 to $\tilde{F}$ finishes the proof.

3.4. Theorem. *Let F be a set of functions on a topological space X. We assume that there exists a strictly positive, finite lower semi-continuous function $w \in F$ such that $u + \alpha w \in F$ for any $u \in F$, $\alpha \in \mathbb{R}_+$. We denote $F_c := \{\alpha \cdot u : \alpha \in \mathbb{R}_+, u \in F\}$ and, for any numerical function f on X, $f^* := \inf\{u \in F_c : f \leqq u\}$.*

Then, for any $x \in X$ the following conditions are equivalent

(i) *x is a Shilov point.*

(ii) *$f^*(x) = f(x)$ for any upper bounded, upper semi-continuous function f on X. and, if X is completely regular,*

(iii) *$f^*(x) = f(x)$ for any bounded continuous function f on X.*

Proof. (i)⇒(ii). Let f be an upper bounded, upper semi-continuous function on X. We assume first that $f(x) < 0$. Let $\varepsilon > 0$ be such that $f(x) + \varepsilon < 0$. By the upper semi-continuity of f, there exists an open neighborhood U of X and $\beta > 1$ such that $f \leqq \beta \cdot (f(x) + \varepsilon)$ on U. Since f is upper bounded, there exists $\alpha < 0$ such that $f \leqq \alpha \cdot (f(x) + \varepsilon)$ on $X \setminus U$. Since $x \in S(F)$ there exists $u \in F$ such that $u(x) < 0$,

$u \geqq \alpha \cdot u(x)$ on $X \setminus U$ and $u \geqq \beta \cdot u(x)$. By the construction, we have $f^*(x) \leqq \gamma \cdot u(x) = f(x) + \varepsilon$, where $\gamma := \dfrac{f(x)+\varepsilon}{u(x)}$.

If $f(x) \geqq 0$ and $\varepsilon > 0$ then there exists $\alpha \geqq 0$ such that $(f - \alpha w)(x) < 0$. As proved above, there exist $u \in F_c$ such that $f - \alpha w \leqq u$ and $u(x) \leqq (f - \alpha w)(x) + \varepsilon$. Now we have $u + \alpha w \in F_c$, $f \leqq u + \alpha w$ and $(u + \alpha w)(x) \leqq f(x) + \varepsilon$. Hence we have $f^*(x) \leqq (u + \alpha w)(x) \leqq f(x) + \varepsilon$.

(ii)$\Rightarrow$(i). Let $K \subset X$, $x \notin K$ be closed, $\alpha < 0$ and $\beta > 1$. Obviously the function $f : X \to \mathbb{R}$ defined by

$$f(y) := \begin{cases} 1 & \text{if} \quad y \in K \\ \dfrac{1}{\alpha} & \text{if} \quad y \notin K \end{cases}$$

is upper semi-continuous.

Hence, according to (ii), there exists $u \in F$ and $\gamma \geqq 0$ such that $f \leqq \gamma \cdot u$ and $\beta \cdot \gamma \cdot u(x) \leqq \dfrac{1}{\alpha}$. Thus we have $u(x) < 0$, $u \geqq \dfrac{1}{\gamma} \cdot f = \alpha \cdot \dfrac{1}{\gamma} \cdot \dfrac{1}{\alpha} \geqq \alpha \cdot u(x)$ on K and $u \geqq \dfrac{1}{\gamma} \cdot \dfrac{1}{\alpha} \geqq \beta \cdot u(x)$. (ii)$\Rightarrow$(iii) is obvious and (iii)$\Rightarrow$(i) may be proved in a similar way as (ii)$\Rightarrow$(i). For K, α, β as above, one has to take a continuous function f such that $\dfrac{1}{\alpha} \leqq f \leqq 1$, $f(x) = \dfrac{1}{\alpha}$ and $f = 1$ on K.

The following theorem goes back to Bauer (see [6])

3.5. Theorem. *Let F be a convex cone of lower semi-continuous functions on a compact space X satisfying condition (P').*

Then, we have for any continuous function f on X and any $x \in X$

$$[-(-f)^*(x), f^*(x)] = \{\mu(f) : \mu \in M_x(F)\}.$$

3.6. Corollary. *Let F be a set of lower semi-continuous functions on a compact space X satisfying conditions (A) and (P').*
Then we have $S(F) = \mathrm{Ch}(F)$.

Proof. If $x \in \mathrm{Ch}(F)$ then by Theorem 3.6 $f^*(x) = f(x)$, for any $f \in C(X)$ and hence $x \in S(F)$ by Theorem 3.5. $S(F) \subset \mathrm{Ch}(F)$ was already proved in Proposition 3.2.

From example (i) in [10, p. 264] we see that we have not $S_w(F) \subset \mathrm{Ch}(F)$ in general. There is however an important special case, where this is true.

3.7. Proposition. *Let X be a compact convex set of a locally convex Hausdorff space.*
We denote by X_e the set of extreme points of X, by F_1 the vector space of continuous affine functions on X and by F_2 (resp. F_3) the convex cone of continuous (resp. lower semi-continuous) concave functions on X. Then we have

$$X_e = \mathrm{Ch}(F_i) = S_w(F_i) = S(F_i) \qquad (i = 1, 2, 3).$$

Proof. $X_e = \mathrm{Ch}(F_i)$ $(i = 1, 2, 3)$ is well known (see [1] for example). Since

$$X_e = S(F_1) \subset S_w(F_1) \subset S_w(F_2) \subset S_w(F_3)$$

it remains to show that $S_w(F_3) \subset X_e$. Let $x \in S_w(F_3)$ and $y_1, y_2 \in X$ be such that $\frac{1}{2}(y_1+y_2)=x$. Then we have

$$u(x) \geqq \tfrac{1}{2}u(y_1)+\tfrac{1}{2}u(y_2) \geqq \inf\{u(z): z \in \{y_1, y_2\}\}$$

for any $u \in F_3$. Since $x \in S(F)$ we must have $x \in \{y_1, y_2\}$ and therefore $x=y_1=y_2$.

3.8. Proposition. *Let F be a set of lower semi-continuous functions on a compact space satisfying (A) and (P'). Then, for any $x \in X$, the following conditions are equivalent*

(i) *x is a weak Shilov point.*

(ii) *There exists no open neighborhood U of x and no $\mu \in M_x(F)$ such that $\mu(U)=0$.*

Proof. (i)⇒(ii). If U and μ exist as in (ii) then we have, for any $u \in F$ such that $u \geqq 0$ on $X \setminus U$, $u(x) \geqq \int u d\mu = \int\limits_{X \setminus U} u d\mu \geqq 0$. Hence $x \notin S_w(F)$.

(ii)⇒(i). According to Remark 2.6 we may assume that F is a convex cone. If x is not a weak Shilov point, then there exists a closed set $K \subset X$ such that for any $u \in F$, $u \geqq 0$ on K implies $u \geqq 0$. Thus we can define by $g \to \inf\{u(x): u \in F,\ g \leqq u \text{ on } K\}$ a sublinear functional p on $C(X)$.

Let μ be a linear functional on $C(X)$ dominated by p. For any $f \in C(X)$, $f \leqq 0$ on K, we have $\mu(f) \leqq p(f) \leqq 0$. Hence μ is a measure on X such that $\mu(X \setminus K)=0$. Since

$$\int u d\mu = \sup\{\mu(g): g \in C(K),\ g \leqq u\} \leqq \sup\{p(g): g \in C(K),\ g \leqq u\} \leqq u(x)$$

we have $\mu \in M_x(F)$ and therefore (ii) is not fulfilled.

We close this section with a measurability result of $S(F)$ and $S_w(F)$.

3.9. Proposition. *Let F be a set of continuous functions on a metric space X. Then $S(F)$ and $S_w(F)$ are G_δ-sets.*

Proof. For any $A \subset X$ we denote by $\operatorname{diam}(A)$ the diameter of A with respect to the given metric. We consider for any $\varepsilon > 0$, $\alpha < 0$, $\beta > 1$.

$$S_w^\varepsilon := \{x \in X : \exists K \subset X \text{ closed}, u \in F, \operatorname{diam}(X \setminus K) \leqq \varepsilon, u \geqq 0 \text{ on } K, u(x) < 0\},$$

$$S_{\alpha,\beta}^\varepsilon := \{x \in X : \exists K \subset X \text{ closed}, u \in F, \alpha' < \alpha, \beta > \beta' > 1, \operatorname{diam}(X \setminus K) \leqq \varepsilon,$$
$$u(x) < 0,\ u \geqq \alpha' \cdot u(x) \text{ on } K,\ u \geqq \beta' \cdot u(x)\}.$$

Because all functions $u \in F$ are continuous, S_w^ε and $S_{\alpha,\beta}^\varepsilon$ are open. The assertion follows now from $S_w(F) = \bigcap\limits_{\varepsilon > 0} S_w^\varepsilon$, $S(F) = \bigcap\limits_{\varepsilon > 0,\ \alpha < 0,\ \beta > 1} S_{\alpha,\beta}^\varepsilon$.

4. Fine Shilov Boundaries

In the sequel let F be a set of numerical functions taking values in $\mathbb{R} \cup \{+\infty\}$ on a topological space (X, τ) and $F' \subset F$ a subset of lower semi-continuous functions such that $u + \alpha v \in F$ for any $u \in F$, $v \in F''$, and $\alpha \geqq 0$.

We denote by τ' the coarsest topology on X finer than τ such that any function in F' is continuous. τ' is called the fine topology generated by F'. Obviously the

convex cone generated by F' has the same properties as F' and generates the same fine topology. Hence we may assume without loss of generality that F' is a convex cone.

The following lemma is well known in potential theory (see [7, Theorem I.3] or [8, Proposition 5.1.1]).

4.1. Lemma. *A point $x \in X$ belongs to the fine closure of a set $B \subset X$ iff the following two conditions are fulfilled*

(a) $x \in \bar{B}$ *($=\tau$-closure of B),*

(b) *there exists no function $u \in F'$ such that*

$$u(x) < \liminf_{B \ni y \to x} u(y).$$

4.2. Theorem. *Assume that the functions in F' are nonnegative. Then we have $S(F, \tau) = S(F, \tau')$.*

Proof. Obviously $S(F, \tau') \subset S(F, \tau)$.

Let $x \in S(F, \tau)$, $B \subset X$ finely closed, $x \notin B$, $\alpha < 0$, and $\beta > 1$. We have to find $u \in F$ such that $u(x) < 0$, $u \geqq \alpha \cdot u(x)$ on B and $u \geqq \beta \cdot u(x)$. If $x \notin \bar{B}$, then the existence of u follows from $x \in S(F, \tau)$. Hence we may assume that $x \in \bar{B}$. Using Lemma 4.1 we get $v \in F'$ such that $v(x) < \liminf_{B \ni y \to x} v(y)$. Then there exists $U \in \tau$, $x \in U$ such that $v(x) < \inf\{v(y) : y \in B \cap U\} =: \gamma$. We take $\beta' > 1$ such that $\beta' \cdot v(x) < \gamma$. Since $v(x) \geqq 0$ there exists $\delta \geqq 0$ such that $-1 + \delta \cdot v(x) < 0$ and $-\beta' + \delta \cdot \gamma \geqq \alpha \cdot (-1 + \delta v(x))$. Now there exist β'' and λ such that $1 < \beta'' < \beta'$, $0 \leqq \lambda < 1$ and $-\beta'' + \delta \cdot \lambda \cdot v(x) \geqq \beta \cdot (-1 + \delta v(x))$. Since v is lower semi-continuous we may assume that $v \geqq \lambda \cdot v(x)$ on U otherwise we have to take a smaller U. Since $x \in S(F, \tau)$ there exists $u' \in F$ such that $u'(x) < 0$, $u' \geqq \alpha \cdot u'(x)$ on $X \setminus U$ and $u' \geqq \beta'' \cdot u'(x)$.

The function $u := u' - u'(x) \cdot \delta \cdot v \in F$ possesses the required properties

$$u(x) = -u'(x) \cdot (-1 + \delta v(x)) < 0;$$

$$u \geqq \beta'' \cdot u'(x) - u'(x) \cdot \delta \cdot \gamma \geqq \beta' \cdot u'(x) - u'(x) \cdot \delta \cdot \gamma = -u'(x) \cdot (-\beta' + \delta \cdot \gamma)$$
$$\geqq -u'(x) \cdot \alpha \cdot (-1 + \delta \cdot v(x)) = \alpha \cdot u(x) \quad \text{on} \quad B \cap U;$$

$$u \geqq u' \geqq \alpha \cdot u'(x) \geqq \alpha \cdot u(x) \quad \text{on} \quad X \setminus U;$$

$$u \geqq \beta'' \cdot u'(x) - u'(x) \cdot \delta \cdot \lambda \cdot v(x) = -u'(x) \cdot (-\beta'' + \delta \cdot \lambda \cdot v(x))$$
$$\geqq -u'(x) \cdot \beta \cdot (-1 + \delta v(x)) = \beta \cdot u(x) \quad \text{on } U.$$

4.3. Corollary. *Assume that X is compact and that F satisfies the conditions (C), (P), (S). Then we have $S(F, \tau) = S(F, \tau')$. In particular, the fine Shilov boundary $\partial_{F, \tau} X$ exists.*

Proof. Because of (*P*) we may assume that F' contains a strictly positive finite function w. Then the fine topology τ'' generated by $F'' := \{u \in F' : u \geqq 0\}$ is exactly τ'. Indeed, for any $u \in F'$ there exists $\alpha \geqq 0$ such that $u + \alpha w \geqq 0$. Since $u + \alpha w$ and αw are τ''-continuous, $u = (u + \alpha w) - \alpha w$ is also τ''-continuous. The first assertion follows now from Theorem 4.2 and using Theorem 2.1 we get also the second assertion.

4.4. Corollary. *Assume that X is completely metrizable that $F=F'$ that the functions in F are lower bounded and that $1\in F$. Then the following conditions are equivalent.*

(i) *The fine Shilov boundary $\partial_{F,\tau'}X$ exists.*

(ii) *$\partial_{F,\tau'}X$ exists and $\partial_{F,\tau'}X$ is the τ'-closure of $S(F,\tau)$.*

(iii) *$u\geqq 0$ on $S(F,\tau)$ implies $u\geqq 0$ for any $u\in F$.*

(iv) *$u\geqq 0$ on $S(F,\tau')$ implies $u\geqq 0$ for any $u\in F$.*

Proof. Similarly as in Corollary 4.3 the assumptions imply that τ' is generated by the nonnegative functions in F'. Hence, by Theorem 4.1, $S(F,\tau)=S(F,\tau')$.

In particular, (iii) and (iv) are equivalent. The implication (iv)⇒(ii) is now also immediate. (ii)⇒(i) is trivial.

Assume now that $\partial_{F,\tau'}X$ exists. Then Theorem 1.2 implies that $\partial_{F,\tau'}X$ is contained in the τ'-closure of $S(F,\tau)$. Now, if $u\in F$ and $u\geqq 0$ on $S(F,\tau)$, then $u\geqq 0$ on τ'-closure of $S(F,\tau)$ since u is τ'-continuous. Hence $u\geqq 0$ on $\partial_{F,\tau'}X$ and therefore $u\geqq 0$.

5. The Fine Shilov Boundary of a Simplical Cone

In the sequel F denotes a convex cone of lower semi-continuous functions on a compact space X satisfying (P') and (S). τ' denotes the fine topology generated by F.

If the limit of any increasing family of functions in F belongs again to F, then some examples from potential theory [5] suggest that $\mathrm{Ch}(F)$ is exactly the fine Shilov boundary $\partial_{F,\tau'}X$. In a special case we can prove this conjecture.

Before doing this we need some facts about simplicial cones.

5.1. Let G be a further convex cone of lower semi-continuous functions on X.

A Borel measurable function q on X is called G-concave iff, for any $x\in X$ and any $\mu\in M_x(G)$ $\mu(q)\leqq q(x)$. $Q(G)$ denotes the convex cone of all G-concave functions. A function h on X is called G-affine iff h and $-h$ are G-concave.

A measure μ on X is called extremal iff, for any measure ν on X, $\nu(u)\leqq\mu(u)$ for any $u\in G$ implies $\nu=\mu$.

G is called a simplicial cone iff, for any $x\in X$ there exists at most one extremal measure $\mu\in M_x(G)$.

As in Sect. 2 we use for any function f on X the notation $f^*:=\inf\{u\in G: f\leqq u\}$.

If the functions in G are continuous and if G satisfies conditions (S) and (P') then the following results are known [6]:

(a) $\mathrm{Ch}(G)=\bigcap_{q\in C(X)\cap Q(G)}\{(-q)^*=-q\}$,

(b) G is simplicial iff $(-q)^*$ is G-affine for any G-concave function $q\in C(X)$.

5.2. Proposition. *Assume that there exists a simplicial cone G of continuous functions on X such that F is exactly the cone of lower semi-continuous G-concave functions. Then the Choquet boundary of F is exactly the fine Shilov boundary $\partial_{F,\tau'}X$.*

Proof. Obviously $\mathrm{Ch}(G)=\mathrm{Ch}(F)$. From Corollary 3.7 and Theorem 4.2 we get $\mathrm{Ch}(F)=S(F,\tau')$. Hence, since $\partial_{F,\tau'}X$ is the τ'-closure of $S(F,\tau')$ we have only to show that $\mathrm{Ch}(G)$ is τ'-closed. For any $q\in Q(G)$, $(-q)^*$ is upper semi-continuous and G-affine by 5.1(b). Hence $-(-q)^*\in F$.

In particular, $(-q)^*$ is τ'-continuous. Therefore $\{(-q)^* = -q\}$ is τ'-closed. Thus $\mathrm{Ch}(G)$ is τ'-closed by 5.1(a).

5.3. Remark. Let τ'' be the fine topology generated by all lower semi-continuous G-affine functions. Then the proof of 5.2 shows that we even have $\mathrm{Ch}(F) = \partial_{F,\tau''} X$.

Acknowledgement. I am indebted to A. Cornea for many helpful discussions.

References

1. Alfsen, E.M.: Compact convex sets and boundary integrals. Berlin, Heidelberg, New York: Springer 1971
2. Arens, R.A., Singer, I.M.: Function values as boundary integrals. Proc. Am. Math. Soc. **5**, 735–745 (1954)
3. Bauer, H.: Šilovscher Rand und Dirichletsches Problem. Ann. Inst. Fourier **11**, 89–136 (1961)
4. Bear, H.S.: The Šilov boundary for a linear space of continuous functions. Am. Math. Monthly **68**, 484–485 (1961)
5. Bliedtner, J., Hansen, W.: Simplicial cones in potential theory. Invent. Math. **29**, 83–110 (1975)
6. Boboc, N., Cornea, A.: Convex cones of lower semi-continuous functions on compact spaces. Rev. Roum. Math. Pures Appl. **12**, 471–525 (1967)
7. Brelot, M.: On topologies and boundaries in potential theory. Lecture Notes in Mathematics, No. 175. Berlin, Heidelberg, New York: Springer 1971
8. Constantinescu, C., Cornea, A.: Potential theory on harmonic spaces. Berlin, Heidelberg, New York: Springer 1972
9. Fuchssteiner, B., Hackenbroch, W.: Maximumpunkte. Arch. Math. **23**, 415–421 (1972)
10. Fuchssteiner, B., Lusky, W.: Convex cones. Amsterdam, New York, Oxford: North-Holland 1981
11. Wittmann, R.: On the existence of Shilov boundaries (to appear in Proc. Am. Math. Soc.)

Received January 20, 1983

Math. Ann. 263, 251–261 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Local Root Numbers and Hermitian–Galois Module Structure of Rings of Integers

Ph. Cassou-Noguès[1] and M. J. Taylor[2]

[1] Department of Mathematics, University of Bordeaux, F-33405 Talence Cedex, France
[2] Trinity College, Cambridge, CB2 1TQ, England

Introduction and Statement of Results

In [F 3], [F 5–7] Fröhlich considered, as a Hermitian–Galois module, the pair consisting of the ring of integers of a tame Galois extension of local fields together with the trace form. His key idea was that his generalized resolvents for symplectic characters are Pfaffians at these characters. Using this he was then able to show that the Hermitian–Galois module structure of the ring of integers determines the local root numbers of many symplectic characters of the Galois group.

The main aim of this paper is to show that, as conjectured by Fröhlich in [F 8], the local root numbers of all symplectic characters of the Galois group of any tame local extension are invariants of the Hermitian–Galois structure of the ring of integers and the trace form. Given Fröhlich's results, to do this we have to treat the singular case of 2-groups using results from [W], and then use the deep determinantal results for Galois–Gauss sums obtained in [T 1].

In order to state our main result we must introduce a certain amount of preliminary notation.

By the term local field we mean a finite extension of the p-adic field $\mathbb{Q}_p$ (here we allow the possibility $p=\infty$, in which case $\mathbb{Z}_p$, $\mathbb{Q}_p$ both mean the field of real numbers $\mathbb{R}$).

For a finite group Γ, we denote the group of virtual characters of Γ, over an algebraic closure of $\mathbb{Q}$, by R_Γ. We say that a character of Γ is symplectic if it is afforded by a symplectic representation of Γ. We denote by R_Γ^s the subgroup of R_Γ generated by the symplectic characters, and we let $T: R_\Gamma \to R_\Gamma^s$ be the homomorphism which associates to $\chi \in R_\Gamma$ the virtual character $\chi + \bar{\chi}$, where $\bar{\chi}$ is the complex conjugate of χ.

Now let N/K be a tame Galois extension of local fields (in the Archimedean case all extensions are seemed to be tame). We put $\Gamma = \mathrm{Gal}(N/K)$, then the local root number yields a homomorphism W from R_Γ with values in the group of non-zero algebraic numbers. In particular (cf. [Te] Sect. 3, Proposition 2) since all symplectic characters are real valued and have trivial determinant, we see that by restriction to R_Γ^s, $W \in \mathrm{Hom}(R_\Gamma^s, \pm 1)$. Indeed by Theorem 9(iii) in [F 1], we know W commutes with Galois action.

Next we consider the category of Hermitian $\mathbb{Z}_p\Gamma$-modules given by pairs (M,θ), where M is a free $\mathbb{Z}_p\Gamma$-module, $V=M\otimes_{\mathbb{Z}_p}\mathbb{Q}_p$ and

$$\theta: V\times V\to \mathbb{Q}_p\Gamma$$

is a non-degenerate Hermitian form, that is

(a) θ is $\mathbb{Q}_p$-linear in both variables, and $h(m,n\mu)=h(m,n)\mu$ for $m,n\in V$, $\mu\in\mathbb{Q}_p\Gamma$.

(b) For $m,n\in V$, $\overline{h(m,n)}=h(n,m)$, where $-$ is the standard involution of $\mathbb{Q}_p\Gamma$, so that $\overline{\sum_{\gamma\in\Gamma} a_\gamma\cdot\gamma}=\sum_{\gamma\in\Gamma} a_\gamma\cdot\gamma^{-1}$.

We let $K_0H(\mathbb{Z}_p\Gamma)$ be the Grothendieck group of isometry classes of Hermitian $\mathbb{Z}_p\Gamma$-modules and orthogonal sums.

In the next section we introduce a subgroup $\tilde{K}_0H(\mathbb{Z}_p\Gamma)$ of $K_0H(\mathbb{Z}_p\Gamma)$, and for each $\chi\in R_\Gamma^s$, we define natural homomorphisms

$$\eta_\chi: \tilde{K}_0H(\mathbb{Z}_p\Gamma)\to\mathbb{Q}^*,$$
$$\theta_\chi: \tilde{K}_0H(\mathbb{Z}_p\Gamma)\to\pm 1.$$

We write Tr for the Hermitian form

$$\mathrm{Tr}: N\times N\to\mathbb{Q}_p\Gamma$$

where for $m,n\in N$

$$\mathrm{Tr}(m,n)=\sum_{\gamma\in\Gamma}\mathrm{Tr}_{N/\mathbb{Q}_p}(mn^\gamma)\cdot\gamma^{-1}.$$

Here $\mathrm{Tr}_{N/\mathbb{Q}_p}$ denotes the trace form from N to $\mathbb{Q}_p$. Next we define a further Hermitian form

$$\xi: K\Gamma\times K\Gamma\to\mathbb{Q}_p\Gamma$$

where for $x,y\in K\Gamma$

$$\xi(x,y)=\mathrm{Tr}_{K/\mathbb{Q}_p}(\bar{x}\cdot y).$$

We write $\mathcal{O}_N$ for the ring of integers of a local field N (with the usual convention that in the Archimedean case $\mathcal{O}_N=N$). We define $X_{N/K}\in K_0H(\mathbb{Z}_p\Gamma)$ by

$$X_{N/K}=(\mathcal{O}_N,\mathrm{Tr})-(\mathcal{O}_K\Gamma,\xi).$$

Using Chap. VI of [F 3] and Theorems 2(a) and 2(b) in [T 1], we will show in the next section that $X_{N/K}\in\tilde{K}_0H(\mathbb{Z}_p\Gamma)$. Our main result is

Theorem 1. *For each* $\chi\in R_\Gamma^s$

$$\theta_\chi(X_{N/K})=W(\chi)$$
$$\eta_\chi(X_{N/K})=\begin{cases}N\mathfrak{f}(\chi)^{1/2} & K \text{ non-Archimedean},\\ W(\chi) & K \text{ Archimedean},\end{cases}$$

where $N\mathfrak{f}(\chi)^{1/2}$ *is the positive square root of the absolute norm of the Artin conductor of* χ.

Remark 1. It follows from the theorem that, in the non-Archimedean case, the local Galois Gauss sum $\tau(\chi)$ is characterized as a Hermitian–Galois invariant, for $\chi\in R_\Gamma^s$, by the product character $\theta_\chi\cdot\eta_\chi$.

Remark 2. The characterization of $N\mathfrak{f}(\chi)^{1/2}$ as a Hermitian–Galois invariant for $\chi \in R_\Gamma^s$ is given in Chap. VI of [F 3]. We include it here for the sake of completeness.

At the global level, one of us showed in [T 1], that, for any tame Galois extension of number fields N/K with Galois group Γ, the only obstructions to the stable freedom of $\mathcal{O}_N$ as a $\mathbb{Z}\Gamma$-module arise from the global Artin root numbers of symplectic characters of Γ. So we see that, while Taylor's theorem says that Artin root numbers of symplectic characters determine the Galois module structure of $\mathcal{O}_N$, the converse does not hold in general (cf. [F 4] for instance). Thus our result can also be seen rectifying the failure of this converse, by passing to the finer Hermitian–Galois module structure – this in fact was the motivation for Fröhlich's original work.

2. The Local Hermitian Class Group

We denote the group of units of a ring A, by A^*. Let $\bar{\mathbb{Q}}_p$ be a chosen algebraic closure of $\mathbb{Q}_p$. For $\bar{\mathbb{Q}}_p \supset M \supset \mathbb{Q}_p$, with $M/\mathbb{Q}_p$ finite, we put $\Omega_M = \mathrm{Gal}(\bar{\mathbb{Q}}_p/M)$. We write $R_{\Gamma,p}$ for the group of $\bar{\mathbb{Q}}_p$-valued virtual characters of Γ. Then Ω_M acts on $R_{\Gamma,p}$, and also on $R^s_{\Gamma,p}$ and $T(R_{\Gamma,p})$.

Let $r : \Gamma \to \mathrm{GL}_n(\bar{\mathbb{Q}}_p)$ be a representation of Γ with character $\chi \in R_{\Gamma,p}$. We extend r to an algebra homomorphism from $M\Gamma$ to $M_n(\bar{\mathbb{Q}}_p)$. Thus, for each $\alpha \in M\Gamma^*$, the element $\det(r(\alpha))$ depends only on the character of r. We now define

$$\mathrm{Det}(\alpha) : R_{\Gamma,p} \to \bar{\mathbb{Q}}_p^*$$

by $\mathbb{Z}$-linearity and by stipulating that for a character χ as above

$$\mathrm{Det}(\alpha)(\chi) = \det(r(\alpha)).$$

In the sequel we write $\mathrm{Det}(\mathcal{O}_M\Gamma^*)$ for the group of homomorphisms $\mathrm{Det}(\alpha)$, with $\alpha \in \mathcal{O}_M\Gamma^*$. It is readily verified that $\mathrm{Det}(\mathcal{O}_M\Gamma^*)$ is a subgroup of $\mathrm{Hom}_{\Omega_M}(R_{\Gamma,p}, \mathcal{O}_{\bar{\mathbb{Q}}_p}{}^*)$.

We write $\mathrm{Det}^s(\mathcal{O}_M\Gamma^*)$ for the restriction of $\mathrm{Det}(\mathcal{O}_M\Gamma^*)$ to $R^s_{\Gamma,p}$. Now we define the Hermitian class group

$$\mathrm{HCL}(\mathcal{O}_M\Gamma) = \frac{\mathrm{Hom}_{\Omega_M}(R^s_{\Gamma,p}, \bar{\mathbb{Q}}_p^*)}{\mathrm{Det}^s(\mathcal{O}_M\Gamma^*)} \tag{2.1}$$

(2.2) *Remark.* It is important not to confuse Det with the homomorphism

$$\det : R_{\Gamma,p} \to \mathrm{Hom}(\Gamma, \bar{\mathbb{Q}}_p{}^*)$$

where for a character χ, $\det(\chi) = \det \circ r|_\Gamma$ (2.3). In Sect. 4, Chap. 2 [F 3], Fröhlich used the notion of Pfaffian to define a discriminantal homomorphism from $K_0H(\mathbb{Z}_p\Gamma)$ to $\mathrm{HCL}(\mathbb{Z}_p\Gamma)$. (and, of course, more generally for larger coefficient rings too.)

3. The Adelic Hermitian Class Group

Let J denote the idele group of a chosen algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$. We write U(resp. U_+) for the group of unit ideles (resp. the group of unit ideles which are both real and positive at all infinite primes).

Let F be a numberfield and let $\Omega_F = \mathrm{Gal}(\bar{\mathbb{Q}}/F)$. Let $\mathfrak{p}$ be a prime of F above p and let $j : \bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_p$ be such that $j(\mathfrak{p}^{-1}) \notin \mathcal{O}_{\bar{\mathbb{Q}}_p}$. We write M for the closure of $j(F)$ in $\bar{\mathbb{Q}}_p$, and we set $(\bar{\mathbb{Q}})_{\mathfrak{p}} = (\bar{\mathbb{Q}} \otimes_F M)$.

Let $x \in \mathcal{O}_M \Gamma^*$, and let $r : \Gamma \to \mathrm{GL}_n(\bar{\mathbb{Q}})$ be a representation of Γ with character $\chi \in R_\Gamma$. We extend r to an algebra homomorphism from $M\Gamma$ to $M_n(\bar{\mathbb{Q}})_{\mathfrak{p}})$. As previous, we define $\mathrm{Det}(x) : R_\Gamma \to (\bar{\mathbb{Q}})^*_{\mathfrak{p}}$ by

$$\mathrm{Det}(x)(\chi) = \det(r(x)).$$

At this point then the notation $\mathrm{Det}(x)$ appears to have two meanings: one local and one semi-local. To resolve this apparent ambiguity (cf. Chap. 2 [F 2] we note that j induces an isomorphism $R_\Gamma \xrightarrow{\sim} R_{\Gamma, p}$, and also an algebra homomorphism $(\bar{\mathbb{Q}})_{\mathfrak{p}} \to \bar{\mathbb{Q}}_p$. Thus we obtain a commutative diagram where the vertical maps are inclusions, and the horizontal maps are isomorphisms j_* with $(j_* f)(\chi) = f(\chi^{\alpha^{-1}})^j$

$$\begin{array}{ccc} \mathrm{Hom}_{\Omega_F}(R_\Gamma, (\bar{\mathbb{Q}})^*_{\mathfrak{p}}) & \xrightarrow{\sim} & \mathrm{Hom}_{\Omega_M}(R_{\Gamma, p}, \bar{\mathbb{Q}}^*_p) \\ \uparrow & & \uparrow \\ \mathrm{Det}(\mathcal{O}_M \Gamma^*) & \xrightarrow{\sim} & \mathrm{Det}(\mathcal{O}_M \Gamma^*) \end{array}$$

Conversely, given $g \in \mathrm{Hom}_{\Omega_M}(R_{\Gamma, p}, \bar{\mathbb{Q}}^*_p)$ and $h : \bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_p$ an embedding with $h_{|F} = j_{|F}$, then for $\chi \in R_\Gamma$

$$(j_*^{-1} g(\chi))^h = g(\chi^h). \tag{3.1}$$

More generally we write $\check{\mathcal{O}}_F \Gamma = \Pi \mathcal{O}_{F, \mathfrak{p}} \Gamma$ where the product is taken over all primes $\mathfrak{p}$ of F, and $\mathcal{O}_{F, \mathfrak{p}}$ is the completion of $\mathcal{O}_F$ (resp. F) at $\mathfrak{p}$ if $\mathfrak{p}$ is finite (resp. infinite).

We now define the Adelic Hermitian classgroup

$$\mathrm{AHCL}(\mathcal{O}_F \Gamma) = \frac{\mathrm{Hom}_{\Omega_F}(R^s_\Gamma, J)}{\mathrm{Det}^s(\check{\mathcal{O}}_F \Gamma^*)}.$$

Again, using the Pfaffian, this group has a natural interpretation in terms of classes of Hermitian modules over $\check{\mathcal{O}}_F \Gamma$. When $F = \mathbb{Q}$, the restriction of the map j_*^{-1} embeds $\mathrm{Det}(\mathbb{Z}_p \Gamma^*)$ into $\mathrm{Det}(\mathbb{Z}\Gamma^*)$ and so induces a homomorphism

$$\mathrm{HCL}(\mathbb{Z}_p \Gamma) \xrightarrow{\alpha :} \mathrm{AHCL}(\mathbb{Z}\Gamma). \tag{3.2}$$

Remark. The embedding α corresponds to a functor from $\mathbb{Z}_p \Gamma$-Hermitian modules to $\tilde{\mathbb{Z}}\Gamma$-Hermitian modules, as described in p. 32 of [F 7].

The inclusion $\mathbb{Z}\Gamma \to \mathcal{O}_F \Gamma$ induces a homomorphism

$$\beta : \mathrm{AHCL}(\mathbb{Z}\Gamma) \to \mathrm{AHCL}(\mathcal{O}_F \Gamma). \tag{3.3}$$

Henceforth we fix F (dependent on p) as follows: If $p < \infty$, F is to contain the p^{th} roots of unity and have no real embeddings, the localisations of F at all primes over p are to contain the tame extension $T/\mathbb{Q}_p$ given in Sect. 11 of [T 1], and $F/\mathbb{Q}$ is to be at most tamely ramified at 2. (For instance we can take F to be generated by an appropriate radical element over a cyclotomic field of odd conductor).

If $p = \infty$, then we set $F = \mathbb{Q}$.

Next we define various sub-groups of $\mathrm{AHCL}(\mathcal{O}_F\Gamma)$, which again depend on p.

(1) If $p<\infty$, then we write $\mathbb{Q}_+^*$, for the group of positive rationals, and we obtain the product subgroup in $\mathrm{AHCL}(\mathcal{O}_F\Gamma)$

$$G_p(\mathcal{O}_F\Gamma)=\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,\mathbb{Q}_+^*)\times\frac{\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,U)}{\mathrm{Det}^s(\tilde{\mathcal{O}}_F\Gamma^*)}$$

and so we obtain a further direct product in $\mathrm{AHCL}(\mathcal{O}_F\Gamma)$

$$\tilde{G}_p(\mathcal{O}_F\Gamma)=\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,\mathbb{Q}_+^*)\times\frac{\mathrm{Det}^s(\tilde{\mathcal{O}}_F\Gamma^*)\,\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,\pm1)}{\mathrm{Det}^s(\tilde{\mathcal{O}}_F\Gamma^*)}\,. \tag{3.3}$$

We denote the first (resp. second) factor by $G_p'(\mathcal{O}_F\Gamma)$ (resp. $G_p''(\mathcal{O}_F\Gamma)$), and we write π' (resp. π'') for projection onto that factor. In Sect. 6 we show

$$\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,\pm1)\cap\mathrm{Det}^s(\tilde{\mathcal{O}}_F\Gamma^*)=1\,, \tag{3.4}$$

and so we obtain an isomorphism

$$G_p''(\mathcal{O}_F\Gamma)\cong\mathrm{Hom}_{\Omega_F}(R_\Gamma^s,\pm1)\,. \tag{3.5}$$

(2). If $p=\infty$, then $F=\mathbb{Q}$, and we write -1_∞ for the idele with component -1 at all finite places and component 1 at all infinite places. Then we obtain the product subgroup in $\mathrm{AHCL}(\mathbb{Z}\Gamma)$

$$G_\infty(\mathbb{Z}\Gamma)=\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\mathbb{Q}^*)\times\frac{\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,U_+)}{\mathrm{Det}^s(\tilde{\mathbb{Z}}\Gamma^*)}$$

and so we obtain a further subgroup

$$\tilde{G}_\infty(\mathbb{Z}\Gamma)=\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\mathbb{Q}^*)\times\frac{\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\pm1_\infty)\,\mathrm{Det}(\tilde{\mathbb{Z}}\Gamma^*)}{\mathrm{Det}^s(\tilde{\mathbb{Z}}\Gamma^*)}\,.$$

In Sect. 6 we show

$$\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\pm1_\infty)\cap\mathrm{Det}^s(\tilde{\mathbb{Z}}\Gamma^*)=1\,, \tag{3.6}$$

and so we get a decomposition $\tilde{G}_\infty(\mathbb{Z}\Gamma)=G_\infty'(\mathbb{Z}\Gamma)\times G_\infty''(\mathbb{Z}\Gamma)$ with

$$G_\infty'(\mathbb{Z}\Gamma)=\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\mathbb{Q}^*),\quad G_\infty''(\mathbb{Z}\Gamma)=\mathrm{Hom}_{\Omega_\mathbb{Q}}(R_\Gamma^s,\pm1_\infty)\,.$$

Let $\mathscr{A}$ denote the composite of the homomorphism given in (2.3) with $\beta\circ\alpha$ (from (3.2) and (3.3)). We define

$$\tilde{K}_0H(\mathbb{Z}_p\Gamma)=\mathscr{A}^{-1}(\tilde{G}_p(\mathcal{O}_F\Gamma))\,.$$

For any $\chi\in R_\Gamma^s$, we define

$$\theta_\chi:\tilde{K}_0H(\mathbb{Z}_p\Gamma)\to\pm1\,,\qquad \eta_\chi:\tilde{K}_0H(\mathbb{Z}_p\Gamma)\to\mathbb{Q}^*$$

to be that homomorphism obtained by restricting $\mathscr{A}$ down to $\tilde{K}_0H(\mathbb{Z}_p\Gamma)$, composing with π'' (resp. π') and then evaluating on χ.

Remark. Note that the homomorphisms θ_χ, η_χ all factor through $\mathrm{HCL}(\mathbb{Z}_p\Gamma)$. Indeed, up to now, nearly all known information about $K_0H(\mathbb{Z}_p\Gamma)$ has been obtained by passing to $\mathrm{HCL}(\mathbb{Z}_p\Gamma)$.

4. The Non-Archimedean Case

We now return to the arithmetic, and as previous, we let N/K be a tame Galois extension of local fields with $\Gamma = \mathrm{Gal}(N/K)$. We choose a to be a normal integral basis of $\mathcal{O}_N$ over $\mathcal{O}_K\Gamma$, and, for $\phi \in R_{\Gamma,p}$, we write

$$\mathcal{N}_{K/\mathbb{Q}_p}(a|\phi) = \mathrm{Det}\Big(\prod_{\sigma}(\sum a^{\gamma\sigma}\gamma^{-1})\Big)(\phi)$$

where σ runs through any transversal $\Omega_K \backslash \Omega_{\mathbb{Q}p}$. In general the above value depends on the chosen transversal to within a root of unity. However, if $\det(\phi) = 1$, this value is unique. From Theorem 15, Corollary 3, in [F 3] we know that the image of $X_{N/K}$ under (2.3) is represented by the homorphism

$$\phi \mapsto \mathcal{N}_{K/\mathbb{Q}_p}(a|\phi) \tag{4.1}$$

for all $\phi \in R^s_{\Gamma,p}$.

For the remainder of this section we take N to be non-Archimedean. (We treat the Archimedean case in Sect. 5).

Let $\tau(\chi)$ be the local Galois Gauss sum associated to $\chi \in R_\Gamma$ [M] and let $h : \bar{\mathbb{Q}} \to \bar{\mathbb{Q}}_p$. Define $b_h : R_\Gamma \to \bar{\mathbb{Q}}_p^*$, by

$$b_h(\chi) = \mathcal{N}_{K/\mathbb{Q}_p}(a|\chi^h)(\tau(\chi)^{-1})^h,$$

and let c be the idele valued homomorphism on R_Γ taking the value $b_h(\chi)$ at h, for h as above, and the value 1 at all places over $\ell \neq p$. Because $\tau|_{R^s_\Gamma}$ commutes with $\Omega_{\mathbb{Q}}$ action (see Theorem 5.1 of [M]), we know that $c|_{R^s_\Gamma}$ lies in $\mathrm{Hom}_{\Omega_F}(R^s_\Gamma, J)$. However, from the proof of (11.3) in [T 1], or directly from Theorem 31 of [F 2], we know that b_h is the product of an element in $\mathrm{Det}(\mathcal{O}_T\Gamma^*)$, for some tame extension $T/\mathbb{Q}_p$, and of a root of unity valued homomorphism which is 1 on R^s_Γ. (Note we view $\mathrm{Det}(\mathcal{O}_T\Gamma^*)$ as homomorphisms on R_Γ, via the isomorphism $h : R_\Gamma \xrightarrow{\sim} R_{\Gamma,p}$). Thus as all localizations of F at p contain T by construction, using the commutative diagram prior to (3.1), we get that

$$c|_{R^s_\Gamma} \in \mathrm{Det}^s(\tilde{\mathcal{O}}_F\Gamma^*).$$

On the other hand by (3.1) and (4.1) we know $\mathcal{A}(X_{N/K})$ is represented by the idele valued homomorphism taking the value $\mathcal{N}_{K/\mathbb{Q}_p}(a|\chi^h)$ at h, and 1 at places over $\ell \neq p$. So multiplying this homomorphism by $c^{-1}|_{R^s_\Gamma}$, we see $\mathcal{A}(X_{N/K})$ is also represented by $f \in \mathrm{Hom}_{\Omega_F}(R^s_\Gamma, J)$ where for $k : \bar{\mathbb{Q}} \to \bar{\mathbb{Q}}_\ell$, $\chi \in R^s_\Gamma$,

$$f(\chi)^k = \begin{cases} \tau(\chi)^k & \text{if } \ell = p, \\ 1 & \text{if } \ell \neq p. \end{cases}$$

On the other hand from Theorem 3 of [T 1] we know that for all $\ell \neq p$

$$(\tau|_{R^s_\Gamma})_\ell \in \prod_{\mathcal{L}|\ell} \mathrm{Det}^s(\mathcal{O}_{F,\mathcal{L}}\Gamma^*).$$

Thus we now see that $\mathcal{A}(X_{N/K})$ is represented in $\mathrm{AHCL}(\mathcal{O}_F\Gamma)$ by the homomorphism $\chi \hookrightarrow \tau(\chi)$. However, for real valued χ, of trivial determinant,

$$\tau(\chi) = N\mathfrak{f}(\chi)^{1/2} W(\chi)$$

with $W(\chi)=\pm 1$, and $N\mathfrak{f}(\chi)^{1/2}$ the positive square root of the absolute norm of the Artin conductor of χ. From Theorem 9 (ii), (iii) of [F 1],

$$\chi\mapsto N\mathfrak{f}(\chi)^{1/2}, \quad \chi\mapsto W(\chi)$$

are $\Omega_{\mathbb{Q}}$-homomorphisms, for $\chi\in R_\Gamma^s$. So we have now shown that $\mathscr{A}(X_{N/K})\in\tilde{G}_p(\mathcal{O}_F\Gamma)$ and $\pi'\circ\mathscr{A}(X_{N/K})$ (resp. $\pi''\circ\mathscr{A}(X_{N/K})$) is $\chi\mapsto N\mathfrak{f}(\chi)^{1/2}$ (resp. $\chi\mapsto W(\chi)$). Thus Theorem 1, in the non-Archimedean case, now follows.

5. The Archimedean Case

This result was obtained in Theorem 10 of [F 1]. However, for completeness, we also prove the result here.

The case $N=K$ is trivial and all invariants are equal to 1. So now we suppose that $N=\mathbb{C}$, $K=\mathbb{R}$ and $\Gamma=\mathrm{Gal}(N/K)$ is generated by complex conjugation, which we denote by σ. We write ε (resp. ϕ) for the identity (resp. the non-trivial abelian) character of Γ. Clearly $R_\Gamma^s=2R_\Gamma$.

As in Sect. 2 we have groups $K_0H(\mathbb{R}\Gamma)$, $\mathrm{HCL}(\mathbb{R}\Gamma)$. For $\chi\in R_\Gamma$, we define

$$n(\chi)=\tfrac{1}{2}(\chi(1)-\chi(\sigma)), \tag{5.1}$$

that is to say $n(\chi)$ is the number of times ϕ occurs in χ.

(5.2) **Proposition.** *The class of* $X_{\mathbb{C}/\mathbb{R}}$ *in* $\mathrm{HCL}(\mathbb{R}\Gamma)$ *is represented by the homomorphism:*

$$\chi\mapsto i^{n(\chi)} \quad \text{for} \quad \chi\in R_\Gamma^s.$$

Proof. As in (4.1) we know $X_{\mathbb{C}/\mathbb{R}}$ is represented by $\chi\mapsto(a|\chi)$ where a is a normal basis of $\mathbb{C}$ over $\mathbb{R}\Gamma$. We choose $a=\dfrac{1+i}{2}$. Then

$$(a|\varepsilon)=1, \quad (a|\phi)=i \tag{5.3}$$

and so the result follows by linearity.

We now know that $\mathscr{A}(X_{\mathbb{C}/\mathbb{R}})$ is represented in $\mathrm{AHCL}(\mathbb{Z}\Gamma)$ by the element $g\in\mathrm{Hom}_{\Omega_{\mathbb{Q}}}(R_\Gamma^s, J)$ given by

$$(g(\chi))_\ell=\begin{cases} i^{n(\chi)} & \ell=\infty, \\ 1 & \ell<\infty. \end{cases} \tag{5.4}$$

From Proposition 6.1, we will have

$$\tilde{G}_\infty(\mathbb{Z}\Gamma)=\mathrm{Hom}_{\Omega_{\mathbb{Q}}}(R_\Gamma^s, \mathbb{Q}^*)\times\mathrm{Hom}_{\Omega_{\mathbb{Q}}}(R_\Gamma^s, \pm 1_\infty)$$

with -1_∞ have the idelic meaning given previously. It is immediate that $\mathscr{A}(X_{\mathbb{C}/\mathbb{R}})\in\tilde{G}_\infty(\mathbb{Z}\Gamma)$ and that $\pi'\circ\mathscr{A}(X_{\mathbb{C}/\mathbb{R}})$ and $\pi''\circ\mathscr{A}(X_{\mathbb{C}/\mathbb{R}})$ are both represented by

$$\chi\to i^{n(\chi)}, \quad \text{for} \quad \chi\in R_\Gamma^s$$

(Note that $R_\Gamma^s=2R_\Gamma$, implies $i^{n(\chi)}=(-i)^{n(\chi)}$ for $\chi\in R_\Gamma^s$). On the other hand by definition

$$W(\chi)=(-i)^{n(\chi)}$$

and so Theorem 1, in the Archimedean case, is now shown.

6. Determinantal Results

In the sequel we write Δ_Γ for the group $\mathrm{Hom}(R_\Gamma^s, \pm 1) \cap \mathrm{Det}^s(\tilde{\mathcal{C}}_n\Gamma^*)$ if $p<\infty$ (resp. $\mathrm{Hom}(R_\Gamma^s, \pm 1_\infty) \cap \mathrm{Det}^s(\hat{\mathbb{Z}}\Gamma^*)$ if $p=\infty$). The main aim of this section is to show that in all cases:

(6.1) **Proposition** $\Delta_\Gamma = \{1\}$.

The proof of this result goes in four steps.

Step 1. Reduction to cyclic and quaternion groups. Let H be a subquotient of Γ, that is to say a quotient of a subgroup Σ of Γ. We define

$$S_\Gamma^H : \mathrm{Hom}_{\Omega_F}(R_\Gamma^s, J) \to \mathrm{Hom}_{\Omega_F}(R_H^s, J)$$

by stipulating that for $f \in \mathrm{Hom}_{\Omega_F}(R_\Gamma^s, J)$, and for $\phi \in R_H^s$

$$S_\Gamma^H(f)(\phi) = f(\mathrm{Ind}_\Sigma^\Gamma(\mathrm{Inf}_H^\Sigma(\phi))). \tag{6.2}$$

Here Ind, Inf are the standard induction and inflation homomorphisms. From Appendix VIII of [F 1] we know

$$S_\Gamma^H(\mathrm{Det}^s(\tilde{\mathcal{C}}_F\Gamma^*)) \subseteq \mathrm{Det}^s(\tilde{\mathcal{C}}_F H^*).$$

Thus S_Γ^H induces a homomorphism

$$S_\Gamma^H : \Delta_\Gamma \to \Delta_H. \tag{6.3}$$

Next we consider the structure of R_Γ^s in some detail, our main reference is [B-S] (see also Sect. 3, 5 of [M])

(6.4) R_Γ^s is generated by two types of character

(1) $\mathrm{Ind}_\Sigma^\Gamma(\phi)$ where $\phi = \psi + \bar{\psi}$ with ψ an abelian character of Σ, which is therefore inflated from a cyclic quotient H of Σ.

(2) $\mathrm{Ind}_\Sigma^\Gamma(\phi)$ where ϕ is inflated from a faithful irreducible character of a quaternion quotient H of Σ.

(6.5) **Lemma.** *The product of the S_Γ^H, from Δ_Γ to $\prod_H \Delta_H$, where H runs through all cyclic and quaternion subquotients of Γ, is injective.*

Proof. This is immediate from (6.4).

Step 2. Reduction to cyclic and quaternion two-groups. For the moment we let Γ be any finite group and ℓ be a chosen prime number.

Let $\mathscr{L}$ be the maximal ideal of $\mathcal{O}_{\bar{\mathbb{Q}}_\ell}$ and let $k : \bar{\mathbb{Q}} \hookrightarrow \bar{\mathbb{Q}}_\ell$ be an embedding. Then k composed with reduction mod $\mathscr{L}$ yields a homomorphism d_ℓ from R_Γ to the Grothendieck group of $\bar{\mathbb{F}}_\ell\Gamma$-modules, with $\bar{\mathbb{F}}_\ell$ a separable closure of $\mathbb{F}_\ell$, the field of ℓ-elements. From Theorem 42, Corollary 2 of [S], we know that $\mathrm{Ker}\, d_\ell$ is independent of k, it being the ideal of virtual characters which vanish on all elements of Γ of order prime to ℓ.

Restriction from R_Γ^s to $R_\Gamma^s \cap \mathrm{Ker}\, d_\ell$, together with taking the k-component followed by reduction mod $\mathscr{L}$ yields a homomorphism

$$v_k : \mathrm{Hom}(R_\Gamma^s, U) \to \mathrm{Hom}(R_\Gamma^s \cap \mathrm{Ker}\, d_\ell, \bar{\mathbb{F}}_\ell^*).$$

Indeed, essentially because taking determinants commutes with reduction mod $\mathscr{L}$

$$v_k(\mathrm{Det}^s(\tilde{\mathscr{O}}_F\Gamma^*))=1\,.$$

cf. Appendix III, Proposition 2 in [F 1] for details. So for $\ell\neq 2$, because ± 1 are distinguished mod $\mathscr{L}$, we see that

$$f\in\Delta_\Gamma\Rightarrow f(\chi)=1 \quad \text{for} \quad \chi\in\mathrm{Ker}\,d_\ell\cap R_\Gamma^s\,. \tag{6.6}$$

Now we consider the sum $\sum_{\ell\neq 2}(\mathrm{Ker}\,d_\ell\cap R_\Gamma^s)$.

Each $\gamma\in\Gamma$ may be written uniquely in the form $\gamma=\gamma'\gamma''$, where γ' (resp. γ'') has order prime to ℓ (resp. ℓ-power order) and where γ' and γ'' commute. For $\chi\in R_\Gamma$, we define $\delta_\ell\chi$ to be that central function with the property that for all $\gamma\in\Gamma$,

$$\delta_\ell\chi(\gamma)=\chi(\gamma')\,.$$

By use of Adams operations it is shown in Sect. 3 of [CN] that $\delta_\ell\chi\in R_\Gamma$, and moreover, if ℓ is odd, then R_Γ^s is stable under δ_ℓ. Furthermore it is immediate that $\mathrm{Ker}\,d_\ell$ contains the image of $1-\delta_\ell$. Observe that if ℓ' is also a prime number, then δ_ℓ and $\delta_{\ell'}$ commute.

Let $\ell_1,\ldots,\ell_m$ denote the different odd prime divisors of the order of Γ. We say that Γ has the two-quotient property if the set of elements in Γ of odd order form a group. In this case we write Γ_2 for the quotient of Γ by this normal subgroup. We view R_{Γ_2} as a subgroup of R_Γ by inflation.

(6.7) **Lemma.** *Let Γ be either a cyclic or a quaternion group, then Γ has the two-quotient property and*

$$R_\Gamma^s=R_{\Gamma_2}^s+\sum_{i=1}^m(\mathrm{Ker}\,d_{\ell_i}\cap R_\Gamma^s)\,.$$

Proof. The fact that Γ has the two-quotient property is a triviality. Next we put $\delta=\prod_{i=1}^m\delta_{\ell_i}$ and we show $\delta: R_\Gamma^s\to R_{\Gamma_2}^s$. A priori, it will suffice to show $\delta: R_\Gamma\to R_{\Gamma_2}$. In the cyclic case this is immediate. In the quaternion case we observe that every irreducible character of Γ can be written $\mathrm{Ind}_\Sigma^\Gamma\phi$ with ϕ an abelian character of Σ and $(\Gamma:\Sigma)\leqq 2$. The result then follows because δ commutes with induction for such pairs Γ, Σ (cf. Proposition 3.4 in [CN]).

To complete the proof of the lemma, we use the following identity

$$\chi=\chi-\delta_{\ell_1}\chi+\sum_{i=2}^m(1-\delta_{\ell_i})(\delta_{\ell_1}\ldots\delta_{\ell_{i-1}}\chi)+\delta\chi\,,$$

and so the result follows since $(1-\delta_{\ell_i})R_\Gamma\subset\mathrm{Ker}\,d_{\ell_i}$.

By (6.6) and the above lemma we see that it suffices to show that for $f\in\Delta_\Gamma$, with Γ quaternion or cyclic, then $f|_{R_\Gamma\cdot\mathfrak{F}}=1$; so, by passage to the quotient, it suffices to show $\Delta_{\Gamma_2}=1$.

Step 3. Γ a cyclic 2-group. Let $f\in\Delta_\Gamma$, let $k:\bar{\mathbb{Q}}\to\bar{\mathbb{Q}}_2$ be an embedding and let $\mathscr{O}$ denote the closure of $k(\mathscr{O}_F)$ in $\bar{\mathbb{Q}}_2$. Then, for some $x\in\mathscr{O}\Gamma^*$, we know the k-component of f is $\mathrm{Det}(x)$. Since $R_\Gamma^s\supseteq T(R_\Gamma)$, and $\mathrm{Det}(\bar{x})(\phi)=\mathrm{Det}(x)(\bar{\phi})$, for

$\phi \in R_\Gamma$, we known that $\mathrm{Det}(\bar{x}x)$ takes values ± 1 on all elements of R_Γ. So, by Theorem 4.1 in [W] (which is a local Higman result)

$$\mathrm{Det}(\bar{x}x) = \mathrm{Det}(\mu c^i)$$

for some i, where $\Gamma = \langle c \rangle$ and μ is a 2-power root of unity in $\mathcal{O}$. On the other hand, by construction, F is tame; so $\mu = \pm 1$. Further, because Γ is abelian,

$$\mathrm{Det} : \mathcal{O}\Gamma^* \to \mathrm{Det}(\mathcal{O}\Gamma^*)$$

is an isomorphism (cf. [T 2] for instance). Thus $x\bar{x} = \mu c^i$. Because $x\bar{x}$ is–symmetric, we deduce that $2^{n-1} | i$, where 2^n is the order of Γ.

First we suppose for a contradiction that $i = 2^{n-1}$. We let $x = \sum_{r=1}^{2^n} a_r c^r$, with $a_r \in \mathcal{O}$. Then

$$\Sigma a_r a_s c^{r-s} = \mu c^{2^{n-1}}. \tag{6.8}$$

However, for each pair (r, s) with $c^{r-s} = c^{2^{n-1}}$, by inversion $c^{s-r} = c^{2^{n-1}}$, whence the coefficient of $c^{2^{n-1}}$ in the left hand side of (6.8) is divisible by 2–which is a contradiction.

Next we suppose $x \cdot \bar{x} = -1$, and we let ε denote the identity character of Γ. Then $\varepsilon(\bar{x}) = \varepsilon(x)$ and so $\varepsilon(x)^2 = -1$. This again is a contradiction since $\varepsilon(x) \in \mathcal{O}$ and F is tame at 2.

Step 4. Γ a quaternion group of 2 power order. Let k and $\mathcal{O}$ be as in step 3. Let L be a direct product of a finite abelian group $t(L)$ by a finitely generated free $\mathbb{Z}_2$-module, and we put $tf(L) = L/t(L)$. Then restriction from R_Γ to R_Γ^s yields a commutative diagram

$$\begin{array}{ccccccc} 1 \to t(\mathrm{Det}\ (\mathcal{O}\Gamma^*)) & \to & \mathrm{Det}\ (\mathcal{O}\Gamma^*) & \to & tf(\mathrm{Det}\ (\mathcal{O}\Gamma^*)) \to 1 \\ \downarrow a & & \downarrow b & & \downarrow c \\ 1 \to t(\mathrm{Det}^s(\mathcal{O}\Gamma^*)) & \to & \mathrm{Det}^s(\mathcal{O}\Gamma^*) & \to & tf(\mathrm{Det}^s(\mathcal{O}\Gamma^*)) \to 1\,. \end{array}$$

By definition b is surjective. Since $R_\Gamma^s \supset T(R_\Gamma) = 2R_\Gamma$, we see $(R_\Gamma : R_\Gamma^s)$ is finite, and so c is injective. Therefore, by the Snake lemma, a is surjective.

From Theorem 4.1 of [W], together with the fact that the two torsion of $\mathcal{O}^*$ is ± 1 (because F is tame at 2), we see that the two-part of $t(\mathrm{Det}(\mathcal{O}\Gamma^*))$ is made up entirely of homomorphisms of the form $\mathrm{Det}(\pm\gamma)$, for $\gamma \in \Gamma$. However, symplectic characters have even degree and trivial determinant (i.e. lie in $\mathrm{Ker}(\det)$ with det as in (2.2)), thus a $(\mathrm{Det}(\pm\gamma)) = 1$, and so the two-part of $t(\mathrm{Det}^s(\mathcal{O}\Gamma^*))$ is null. In particular then $\Delta_\Gamma = 1$.

Proposition 6.1 is now shown.

References

[B-S] Borel, A., Serre, J.P.: Sur certains sous-groupes des groupes de Lie compacts. Comment. Math. Helv. **27**, 128–239 (1953)

[CN] Cassou-Noguès, Ph.: Structure galoisienne des anneaux d'entiers. Proc. L.M.S. **38**, 545–576 (1979)

[D] Deligne, P.: Les constantes locales de l'équation fonctionelle de la fonction L d'Artin d'une représentation orthogonale. Invent. Math. **35**, 299–316 (1976)
[F1] Fröhlich, A.: Arithmetic and Galois module structure for tame extensions. J. reine angew. Math. **286–287**, 380–440 (1976)
[F2] Fröhlich, A.: Ergebnisse der Mathematik, Bd. 1, 3. Folge. Berlin, Heidelberg, New York: Springer 1983
[F3] Fröhlich, A.: Classgroups, in particular Hermitian classgroups (to appear)
[F4] Fröhlich, A.: Module invariants and root numbers for quaternion fields of degree $4\ell^r$. Camb. Philos. Soc. **76**, 393–399 (1974)
[F5] Fröhlich, A.: Root numbers for symplectic characters and hermitian Galois module structure. Astérisque **24–25**, Soc. Math. France (1975)
[F6] Fröhlich, A.: Local Hermitian group modules. Quadratic form conference. Queen's papers on pure and applied maths, No. 46. Kingston, Ont. 1977
[F7] Fröhlich, A.: Symplectic local constants and Hermitian Galois module structure. Algebraic Number Theory, Int. Symposium Kyoto 1976, ed. S. Ivanaga. Tokyo: Japan Soc. Promotion Science, 1977
[F8] Fröhlich, A.: On parity problems. Sem. théorie des nombres. Univ. Bordeaux 1979
[M] Martinet, J.: Character theory and Artin L-functions. Algebraic number fields, ed. A. Fröhlich. London: Academic Press 1977
[S] Serre, J.-P.: Représentations linéaires des groupes finis, 2ème édition. Paris: Hermann 1971
[Te] Tate, J.: Local constants. Algebraic number fields, ed. A. Fröhlich. London: Academic Press 1977
[T1] Taylor, M.J.: On Fröhlich's conjecture for rings of integers of tame extensions. Invent. Math. **63**, 41–79 (1981)
[T2] Taylor, M.J.: A logarithmic approach to classgroups of integral group rings. J. Algebra **66**, 321–353 (1980)
[W] Wall, C.T.C.: Norms of units in group rings. Proc. AMS **29**, 593–632 (1974)

Received January 1, 1983

INSTRUCTIONS TO AUTHORS

A. General

Papers submitted for publication may be written in English, French or German.

Manuscripts must be in their **final form**, typed on one side of each sheet only, with double spacing and wide margins. Mimeographed copies are not acceptable unless clearly legible. Please include a "Note for the Printer" explaining markings used. See suggestion overleaf.

To speed up publication, authors will receive **only one set of proofs** with provisional page numbers. Authors are requested to **correct typographical errors only**; they will be charged for corrections involving changes, additions or deletions to the original manuscript.

Equations should be typewritten whenever possible, with special typefaces (alphabets) marked by underlining in color (see overleaf). If there is no difference in size, special attention should be given to the placing of subscripts and superscripts so that they are recognizable as such. Formulas should preferably be written in the "linear" form (see overleaf). Equations should be numbered sequentially with arabic numerals in parentheses on the *right-hand side* of the page. In the text they are cited simply by the arabic numeral in parentheses, e.g. (7); the form Equation (7) or Equations (7) and (8) should be used only at the beginning of a sentence.

Diagrams should be submitted on separate sheets, not included in the text. They should be drawn in Indian ink in clean uniform lines, the whole about twice the size of the finished illustration. Inscriptions should allow for the figure 1, for example, to be about 2 mm high in the final version (i.e. 4 mm for reduction $\times \frac{1}{2}$). The author should mark in the margin of the manuscript where diagrams may be inserted.

Footnotes, other than those which refer to the title heading, should be numbered consecutively and placed at the foot of the page to which they refer (not at the end of the article).

Please give on the first page of the manuscript a **running head** (condensed title), which should not exceed 70 letters including spaces.

References should be listed alphabetically according to the name of the first author. They should be numbered with arabic numerals and cited in the text by the number. Multiple references should be referred to individually in the text as follows: [5, 6, 8]. Square brackets, if not available on the typewriter, may be indicated by slashes, e.g., /8/ instead of [8]. Author(s) may be cited in the text by name, but without initials; if more than 2, only the name of the first author followed by et al. should be given (e.g., Andersen et al. [5]). The following information should be provided for journal articles: names and initials of all authors, full title of paper, name of the journal, volume, first and last page numbers and year of publication. Names of journals should be abbreviated as in "Mathematical Reviews". References to books should include name(s) of author(s), full title, edition, place of publication, publisher and year of publication.

Examples

1. Anger, B., Lembcke, J.: Hahn-Banach type theorems for hypolinear functionals. Math. Ann. **209**, 127–151 (1974)
2. Berg, C., Forst, G.: Potential theory on locally compact abelian groups. Ergebnisse der Mathematik und ihrer Grenzgebiete 87. Berlin, Heidelberg, New York: Springer 1976

B. Marking

1. Text

The words "**Theorem**", "**Lemma**", "**Corollary**", "**Proposition**" etc. are normally printed in **boldface,** followed by the formulation in italics (to be underlined in the manuscript).
The words "*Proof*", "*Remark*", "*Definition*", "*Note*" etc. are printed in *italics* with the formulation in ordinary typeface.
Words or sentences to be set in italics should be marked by single underlining.

2. Formulae

Letters in formulae are normally printed in italics, figures in ordinary typeface.
It will help the printer if in doubtful cases the position of indices and exponents is marked thus: $b\uparrow$, $a\Downarrow$. Spacing of indices and exponents must be specially indicated ($A_m{}^n{}_n{}^m$) otherwise they will be set (A_{mn}^{nm}).
Underlining for special alphabets and typefaces should be done according to the following code:

single underlining:	small letter
double underlining:	capital letter
brown:	boldface headings, boldface letters in formulae
yellow:	upright (abbreviations e.g. Re, Im, log, sin, ord, id, lim, sup, etc.)
red:	Greek
blue:	Gothic
green:	Script
violet:	the small letter *l* and the capital letter *O* (to distinguish them from the numeral 1 and zero)
orange:	Special Roman

The following are frequently confused:

$\cup$, $\mathbf{u}$, $\bigcup$, U; $\circ$, o, O, 0; $\times$, x, X, κ; $\vee$, υ, ν; $\theta, \Theta, \phi, \varphi, \Phi, \emptyset$; ψ, Ψ; $\varepsilon, \in$;

a', a^1; the symbol *a* and the indefinite article a;

also the handwritten Roman letters:

c, C; e, l; I, J; k, K; o, O; p, P; s, S; u, U; v, V; w, W; x, X; z, Z;

Please take care to distinguish them in some way.

C. Examples

1. Special alphabets or typefaces

Script	$\mathscr{A}, \mathscr{B}, \mathscr{C}, \mathscr{D}, \mathscr{E}, \mathscr{F}, \mathscr{G}, \mathscr{H}, \mathscr{I}, \mathscr{J}, \mathscr{K}, \mathscr{L}, \mathscr{M}, \mathscr{N}, \mathscr{O}, \mathscr{P}, \mathscr{Q}, \mathscr{R}, \mathscr{S}, \mathscr{T}, \mathscr{U}, \mathscr{V}, \mathscr{W}, \mathscr{X}, \mathscr{Y}, \mathscr{Z}$ $a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z$
Sanserif	A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z
Gothic	$\mathfrak{A}, \mathfrak{B}, \mathfrak{C}, \mathfrak{D}, \mathfrak{E}, \mathfrak{F}, \mathfrak{G}, \mathfrak{H}, \mathfrak{I}, \mathfrak{J}, \mathfrak{K}, \mathfrak{L}, \mathfrak{M}, \mathfrak{N}, \mathfrak{O}, \mathfrak{P}, \mathfrak{Q}, \mathfrak{R}, \mathfrak{S}, \mathfrak{T}, \mathfrak{U}, \mathfrak{V}, \mathfrak{W}, \mathfrak{X}, \mathfrak{Y}, \mathfrak{Z}$ $\mathfrak{a}, \mathfrak{b}, \mathfrak{c}, \mathfrak{d}, \mathfrak{e}, \mathfrak{f}, \mathfrak{g}, \mathfrak{h}, \mathfrak{i}, \mathfrak{j}, \mathfrak{k}, \mathfrak{l}, \mathfrak{m}, \mathfrak{n}, \mathfrak{o}, \mathfrak{p}, \mathfrak{q}, \mathfrak{r}, \mathfrak{s}, \mathfrak{t}, \mathfrak{u}, \mathfrak{v}, \mathfrak{w}, \mathfrak{x}, \mathfrak{y}, \mathfrak{z}$
Boldface	**A, B, C, D, E, F, G, H, I, J, K, L, M, N, O, P, Q, R, S, T, U, V, W, X, Y, Z** **a, b, c, d, e, f, g, h, i, j, k, l, m, n, o, p, q, r, s, t, u, v, w, x, y, z**
Special Roman	$\mathbb{A}, \mathbb{B}, \mathbb{C}, \mathbb{D}, \mathbb{E}, \mathbb{F}, \mathbb{G}, \mathbb{H}, \mathbb{I}, \mathbb{J}, \mathbb{K}, \mathbb{L}, \mathbb{M}, \mathbb{N}, \mathbb{O}, \mathbb{P}, \mathbb{Q}, \mathbb{R}, \mathbb{S}, \mathbb{T}, \mathbb{U}, \mathbb{V}, \mathbb{W}, \mathbb{X}, \mathbb{Y}, \mathbb{Z}$
Greek	$\Gamma, \Delta, \Theta, \Lambda, \Xi, \Pi, \Sigma, \Phi, \Psi, \Omega$ $\alpha, \beta, \gamma, \delta, \varepsilon, \zeta, \eta, \theta, \vartheta, \iota, \kappa, \lambda, \mu, \nu, \xi, o, \pi, \rho, \sigma, \tau, \upsilon, \varphi, \phi, \chi, \psi, \omega$

2. Notations

preferred form	instead of	preferred form	instead of
$A^*, \tilde{b}, \gamma', \mathfrak{v}, \mathbf{v}$	$\bar{A}, \hat{b}, \check{\gamma}, \vec{v}$	$f: A \to B$	$A \xrightarrow{f} B$
lim sup, lim inf	$\overline{\lim}, \underline{\lim}$		
inj lim, proj lim	$\varinjlim, \varprojlim$	$\dfrac{\cos(1/x)}{(a+b/x)^{1/2}}$	$\dfrac{\cos\dfrac{1}{x}}{\sqrt{a+\dfrac{b}{x}}}$
$\exp(-(x^2+y^2)/a^2)$	$e^{-\frac{x^2+y^2}{a^2}}$		
f^{-1}	$\overset{-1}{f}$		

Math. Ann. 263, 263–266 (1983)

Mathematische Annalen
© Springer-Verlag 1983

The Fundamental Group-Scheme of an Abelian Variety

Madhav V. Nori

School of Mathematics, Tata Institute of Fundamental Research, Homi Bhabba Road, Bombay 400005, India

Let X be an abelian variety defined over a field k, and let $n_X : X \to X$ denote multiplication by n and $X_n = \ker(n_X)$ for natural numbers n. We shall prove:

Proposition. *Let $j : P \to X$ be a principal G-bundle on X with G a finite group-scheme and V a k-rational point of P such that $j(V) = 0$. Then there is a natural number n and*

A : a homomorphism $\varrho : X_n \to G$ of group-schemes, thus defining an action of X_n on P, and

B : a morphism $h : X \to P$ which commutes with the X_n-action (the action of X_n on X is by translation), with $h(0) = V$.

Furthermore given n, the pair (ϱ, h) satisfying A and B is unique.

Remark 1. For G reduced, $j : P \to X$ is etale and this is precisely the theorem of Lang and Serre (see [M], Chap. IV) and this is used crucially in the proof.

Remark 2. For G *abelian*, this is well-known and easy; we sketch the proof. For such G, principal G-bundles P on any k-scheme X satisfying

(A) $\Gamma(X, O_X) = k$ and

(B) Pic$X$ is representable

determine, and are uniquely determined by, homomorphisms $\theta : G^* \to \mathrm{Pic}X$, where G^* is the dual group-scheme (see [M], Chap. III, §14). This is just Kummer Theory with S-valued points. Given $\chi \in G^*(S) = \mathrm{Mor}(S, G^*) = \mathrm{Hom}_S(G_S, (\mathbb{G}_m)_S)$, the principal G_S-bundle P_S on X_S induces a principal $(\mathbb{G}_m)_S$-bundle on X_S, or what is the same, a morphism $\theta(\chi) : S \to \mathrm{Pic}X$. This defines θ.

In fact, if $\varphi : Y \to X$ is an isogeny of abelian varieties, then Y is a principal H-bundle on X where $H = \ker(\varphi)$, and the corresponding θ sits in an exact sequence:

$$0 \to H^* \xrightarrow{\theta} \hat{X} \xrightarrow{\hat{\varphi}} \hat{Y} \to 0$$

(see [M], Chap. III, §15).

Given a principal G-bundle P on X with G abelian, put $\theta = p \circ q$, where $p : G^* \to H^*$ is an epimorphism and

$$0 \to H^* \xrightarrow{q} \hat{X} \to \hat{Z} \to 0$$

is exact. This dualises as follows:

$$0 \to H \xrightarrow{P^*} G \quad \text{and} \quad 0 \to H \to Z \to X \to 0$$

are exact. It is checked that P is then the quotient of $Z \times G$ by H and because the isogeny $Z \to X$ is dominated by n_X for some n, the proposition follows.

Remark 3. In the language of [N], this says that the fundamental group-scheme $\pi(X,0) \cong \varprojlim_n X_n$. If it were known that $\pi(A,a) \times \pi(B,b) \cong \pi(A \times B,(a,b))$ for complete connected reduced A and B, it would follow that $\pi(X,0)$ is abelian (in the same manner one deduces that $\pi_1(G)$ is abelian for a topological group G) and the proposition would then follow from Remark 2.

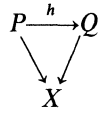
Atiyah's classification of vector bundles on elliptic curves and the results of [N] give a very direct proof of the proposition for elliptic curves.

Note that all projections are denoted simply by p_i.

We need first the following lemmas.

Lemma 1. *Let P and Q be principal bundles on any k-scheme X with affine structure groups G and H respectively. Assume that $\Gamma(P, O_P) = k$.*

A: For any commutative diagram



there is a unique homomorphism $\varrho : G \to H$ such that h intertwines the actions of G and H via ϱ.

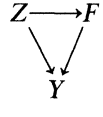
B: Given commutative diagrams as in (A) with $h_i : P \to Q$ for $i = 1, 2$, there is a unique k-rational point g of H such that $h_1(y) = h_2(y) \cdot g$.

Lemma 2. *For any k-scheme X with $\Gamma(X, O_X) = k$, any principal G-bundle on X with G affine, that becomes the trivial bundle after base-change to a field extension of k, is the product of X with a principal homogeneous G-space.*

Proof of the Proposition

Assume first that k is algebraically closed. With $j : P \to X$ and G as in the statement, let $1 \to G_0 \to G \to G_{et} \to 1$ be the usual exact sequence with G_0 local and G_{et} reduced, and let Y be any connected component of $Q = P/G_0$. Because $Q \to X$ is etale, $Y \to X$ is an isogeny of abelian varieties by the theorem of Lang-Serre.

Now $t : P \to P/G_0 = Q$ is set-theoretically injective because G_0 is local. Therefore if $F = t^{-1}(Y)_{red}$, $F \to Y$ is a finite purely inseparable surjective morphism. Let $Z \to Y$ be the isogeny with $k(Z) = k(Y)^{1/a}$, $a = p^m$ for m sufficiently large so that there is a factoring:



The composite isogeny $Z \to Y \to X$ is dominated by n_X for some natural number n, and therefore we have a commutative diagram:

$$\begin{array}{ccc} X & \longrightarrow & P \\ & {\scriptstyle n_X}\searrow \quad \swarrow{\scriptstyle j} & \\ & X & \end{array}$$

Now for k arbitrary, the above shows that $n_X^*(P)$ is the trivial G-bundle on X after a base-change to $\bar{k}$ = the algebraic closure of k, and by Lemma 2, $n_X^*(P) \cong X \times S$ with S a principal homogeneous G-space. However, $n_X^*(P)$ has a unique k-rational point W with images 0 and V in X and P respectively.

This shows that S has a k-rational point and gives moreover a commutative diagram:

$$\begin{array}{ccc} X & \overset{h}{\longrightarrow} & P \\ & {\scriptstyle n_X}\searrow \quad \swarrow{\scriptstyle j} & \\ & X & \end{array}$$

with $h(0) = V$. The Proposition now follows from Lemma 1.

Proof of Lemma 1. Denote the action-morphisms $P \times G \to P$ and $Q \times H \to Q$ by φ_1 and φ_2 respectively. Then $\Psi_1 : P \times G \to P \times_X P$ and $\Psi_2 : Q \times H \to Q \times_X Q$ given by $p_1 \circ \Psi_j = p_1$ and $p_2 \circ \Psi_j = \varphi_j$ are isomorphisms by assumption.

Let s be the composite:

$$P \times G \xrightarrow{\Psi_1} P \times_X P \xrightarrow{h \times h} Q \times_X Q \xrightarrow{\Psi_2^{-1}} Q \times H$$

and let $t = p_2 \circ s$. Clearly $p_1 \circ s = h \circ p_1$.

Because $\Gamma(P, O_P) = k$ and H is affine, there is a factoring: $t = \varrho \circ p_1$ for a unique morphism $\varrho : G \to H$. And by the definition of s, we get the commutative diagram $(*)$ below:

$$\begin{array}{ccc} P \times G & \xrightarrow{\varphi_1} & P \\ {\scriptstyle h \times \varrho}\downarrow & & \downarrow{\scriptstyle h} \\ Q \times H & \xrightarrow{\varphi_2} & Q \end{array}$$

For any k-scheme S, $G(S)$, and $H(S)$ are groups which act on $P(S)$ and $Q(S)$ respectively, on the right. Induced by ϱ and h are set-functions $\varrho(S) : G(S) \to H(S)$ and $h(S) : P(S) \to Q(S)$. Now $(*)$ simply asserts that $h(S)(ag) = h(S)a \cdot \varrho(S)g$ for $a \in P(S)$ and $g \in G(S)$. Because $H(S)$ acts on $Q(S)$ freely, it follows that $\varrho(S)$ is a group homomorphism if $P(S) \neq \phi$.

Choose any k-scheme T with $P(T) \neq \phi$. Then $P(S \times T) \neq \phi$ and $\varrho(S \times T)$ is therefore a homomorphism. We have:

$$\begin{array}{ccc} G(S) & \xrightarrow{\alpha} & G(S \times T) \\ \downarrow & & \downarrow \\ H(S) & \xrightarrow{\beta} & H(S \times T) \end{array}$$

where α and β are the monomorphisms induced by the projection $S \times T \to S$. Thus $\varrho(S)$ is itself a homomorphism for all S; in other words ϱ is itself a homomorphism. The diagram (*) now proves part A of Lemma 1.

With h_1 and h_2 as in part B, let s be the composite:

$$P \xrightarrow{h_1 \times h_2} Q \times_X Q \xrightarrow{\Psi_2^{-1}} Q \times H \xrightarrow{p_2} H.$$

Because $\Gamma(P, O_P) = k$ and H is affine, s is a constant morphism, and this finishes the proof of Lemma 1.

Proof of Lemma 2. Let P be the principal G-bundle in question and let $S = \mathrm{Spec}\,\Gamma(P, O_P)$. Clearly G acts on S and the naturally defined $\theta: P \to X \times S$ preserves the G-action. But the hypothesis implies that θ_L is an isomorphism for some field extension L of k, and since L/k is faithfully flat, θ is itself an isomorphism.

References

[M] Mumford, D.: Abelian Varieties. Oxford: Oxford University Press
[N] Nori, M.V.: Representations of the fundamental group. Composit. Math. **33**, 29–41 (1976)
[A] Atiyah, M.F.: Vector bundles on an elliptic curve. Proc. of London Math. Soc. Third Series **VII**, 414–452 (1957)

Received June 7, 1982

Math. Ann. 263, 267–278 (1983)

Mathematische Annalen
© Springer-Verlag 1983

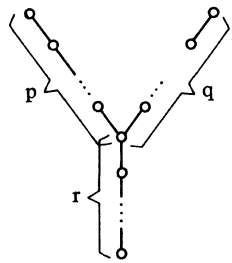
Quotients of Infinite Reflection Groups*

Robert L. Griess, Jr.
Department of Mathematics, University of Michigan, Ann Arbor, MI 48109, USA

1. Introduction

Let W be the Weyl group of a finite dimensional complex simple Lie algebra. The structure of W is quite well-known; see [2, 3] for instance. In particular, W is finite and $W/O_2(W)$ is isomorphic to a symmetric group or an orthogonal or symplectic group over the field of two elements.

It is natural to consider certain infinite analogues $W(p,q,r)$ of such Weyl groups and inquire about their structure. These are reflection groups defined by diagrams $T(p,q,r)$ of the form



Such groups come up in the theory of Kac-Moody Lie algebras [10, 13], the theory of Cremona transformations [7] and in the theory of singularities [11]. Associated to each such diagram is a $W(p,q,r)$-invariant $\mathbb{Z}$-lattice $L=L(p,q,r)$ of rank $p+q+r-2$ with an invariant bilinear form. Fixing a prime number l, we let $\bar{W}$ be the image of $W(p,q,r)$ in $GL(L/lL)$. We give the structure of all such $\bar{W}$.

The special case $2=r\leqq p,q=l=2$ was solved by Coble [5]. These are numerous interesting connections with algebraic geometry. See [4, 7].

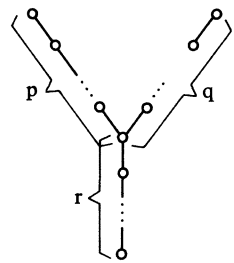
* Research supported by N.S.F. Grant MCS 80-03027, a grant from the University of Michigan and by the John Simon Guggenheim Memorial Foundation

We should make it clear that our results must not be interpreted as full description of the family of normal subgroups of finite index. There are easy examples of W with normal subgroups not of the form $\ker(W \to \bar{W})$. Unpublished work of Bernd Fischer gives strong evidence that there is an epimorphism of $W(5,5,4)$ onto F_1 (the Friendly Giant), a finite simple group of order $2^{46}3^{20}5^{9}7^{6}11^{2}13^{3}17 \cdot 19 \cdot 23 \cdot 29 \cdot 31 \cdot 41 \cdot 47 \cdot 59 \cdot 71$. It would be interesting to have an algebraic geometry interpretation of the kernel of this epimorphism.

It is pleasant that our main results are derived so easily from some existing classification results for finite linear groups. The early result of Coble mentioned above required quite a bit of work. We thank Igor Dolgachev for calling our attention to this problem and providing the above background remarks about algebraic geometry. Reflection groups defined by other diagrams beg to have similar results proven for them. It appears that suitable induction arguments may be tricky to formulate for less familiar classes of diagrams.

2. Notation and Definitions

We give notation which will be used throughout the paper. We let $T = T(p,q,r)$ be the diagram



with $n := p+q+r-2$ nodes, p, q, r all positive. Let $\Phi = \Phi(p,q,r)$ be the vertex set, $L = L(p,q,r)$ the lattice with basis Φ and symmetric bilinear form $(\alpha,\alpha) = -2$, $(\alpha,\beta) = 1$ or 0 as $\alpha, \beta \in \Phi$ are connected or not, $r_\alpha : L \to L$ the *reflection* $x \mapsto x - (2(x,\alpha)/(\alpha,\alpha))\alpha = x + (x,\alpha)\alpha$ at α and $W = W_T = W(p,q,r)$ the *associated reflection group* $\langle r_\alpha | \alpha \in \Phi \rangle$. Let $\Delta = \Delta_T := \det L$. A *deletion* of T shall be a connected subdiagram of T with $n-1$ nodes.

Unless T has type A_n, D_n or E_n, W is infinite (see [2]). We aim to determine the quotients $\bar{W}$ of W by its "congruence subgroups".

More precisely, we fix a prime number, l. Set $\bar{L} := L/lL$ and $\bar{W}$ equal to the image of W in $GL(\bar{L})$. Set $F = \mathbb{F}_l$. We use the bar convention for images in $\bar{L}$ or $\bar{W}$. Since $GL(\bar{L})$ is finite, so is $\bar{W}$. If l is odd, $\bar{W}$ is generated by the reflections $\overline{r_\alpha}$, which are -1 on $F\bar{\alpha}$ and 1 on $\bar{\alpha}^\perp$ ($\bar{L} = F\bar{\alpha} \oplus \bar{\alpha}^\perp$). When $l = 2$, a different notation is required, that of a *transvection* (by definition, a nonidentity invertible linear transformation which is trivial on a hyperplane of a vector space and on the associated quotient space).

Let $R(\bar{L})$ be the radical of the form induced on $\bar{L}$, $d := \dim_F(R(\bar{L}))$, $n' := n-d$, $K := \ker(\bar{W} \to GL(\bar{L}/R(\bar{L}))$, $G = \bar{W}/K$.

For groups A, B, we let $A \cdot B$ denote a group with normal subgroups A and quotient B. The image of B in $\mathrm{Out}(A)$ shall be understood from the context, although the extension type (i.e. split or nonsplit) of $A \cdot B$ may not be.

Group theoretic notations not explained here may be found in [9].

3. Preliminary Results

Lemma 3.1. *Every proper submodule of $\bar{L}$ is trivial. All proper submodules generate $R(\bar{L})$. As a $\bar{W}$-module, $\bar{L}/\mathrm{Rad}(\bar{L})$ is irreducible and nontrivial.*

Proof. Let $\bar{M} < \bar{L}$ be a nontrivial proper submodule. Then some generator $r = r_\alpha$ of W is nontrivial on $\bar{M}$, whence $\bar{M}(1-r) \neq 0$. But $\bar{L}(1-r) = F\bar{\alpha}$. so that $\alpha \in \bar{M}$. Then $\bar{\Phi} \subseteq M$, since two adjacent $\beta, \gamma \in \Phi$ are in a single W-orbit by virtue of the equation $\beta(r_\gamma r_\beta) = (\beta+\gamma)r_\beta = -\beta + (\gamma+\beta) = \gamma$. Since $\bar{\Phi}$ spans $\bar{L}$, $\bar{L} \leqq \bar{M}$, a contradiction which proves the first assertion.

Now let $\bar{M} < \bar{L}$ be a submodule and take $x \in \bar{M}$. If $(x, \bar{\alpha}) \neq 0$ for some $\alpha \in \Phi$, $xr_\alpha = x + (x, \bar{\alpha})\bar{\alpha}$ so that $\bar{\alpha} \in \bar{M}$. As above, this is false. So, $\bar{M} \leqq \mathrm{Rad}(\bar{L})$.

Finally, note that $\mathrm{Rad}(\bar{L}) < \bar{L}$ since $(\bar{\alpha}, \bar{\beta}) = 1$ when $\bar{\alpha}, \bar{\beta} \in \Phi$ are adjacent.

Corollary 3.2. *K is an elementary abelian l-group, isomorphic to $\bigoplus_1^{d'} (\bar{L}/R(\bar{L}))$ as $\bar{W}$-modules, where d' is an integer satisfying $0 \leqq d' \leqq d$.*

Proof. For $k \in K$, we have a linear transformation $T_k \in \mathrm{Hom}_F(\bar{L}/R(\bar{L}), R(\bar{L}))$ given by $T_k : x + R(\bar{L}) \mapsto x(1-k)$. Using Lemma 3.1, we get $T_{k_1} + T_{k_2} = T_{k_1 k_2}$ and $k \mapsto T_k$ is a $\bar{W}$-map.

Trivially, the map is injective. So, $\{T_k | k \in K\}$ is a $\bar{W}$-submodule of

$$\mathrm{Hom}_F(\bar{L}/R(\bar{L}), R(\bar{L})) \cong \mathrm{Hom}_F(\bar{L}/R(\bar{L}), F^d) \cong \bigoplus_1^d \mathrm{Hom}(\bar{L}/R(\bar{L}), F) \cong \bigoplus_1^d \bar{L}/R(\bar{L}),$$

the last isomorphism coming from the self-duality of $\bar{L}/R(\bar{L})$ as a $\bar{W}$-module. Since $\bigoplus_1^d \bar{L}/R(\bar{L})$ is a semisimple module, the corollary follows.

Dolgachev has called my attention to the following two results.

Lemma 3.3. *For any prime l, $\Delta = pqr - pq - qr - rp$.*

Proof. Induction.

Lemma 3.4. *Let l be arbitrary. Suppose that $\Delta_T \equiv 0 (\mathrm{mod}\, l)$ and $\Delta_{T'} \equiv 0 (\mathrm{mod}\, l)$ for every deletion T' of T. Then $p \equiv q \equiv r \equiv 0 (\mathrm{mod}\, l)$. In particular, $p, q, r \geqq l$.*

Proof. Use Lemma 3.3.

Corollary 3.5. (i) *$d \geqq 1$ if and only if $\Delta \equiv 0 (\mathrm{mod}\, l)$.*

(ii) *If $d \geqq 2$, $d = 2$; $d = 2$ if and only if $p, q, r \equiv 0 (\mathrm{mod}\, l)$.*

Proof. (i) is obvious.

(ii) Suppose $d > 1$; we show that $d = 2$ and $p, q, r \equiv 0 (\mathrm{mod}\, l)$. If $d > 1$, $R(\bar{L}_{T'}) \neq 0$ for every deletion T', whence $p, q, r \equiv 0 (\mathrm{mod}\, l)$, by Lemma 3.4. Again by Lemma 3.4, there is a deletion T'' of $T' = T(p-1, q, r)$ with $\Delta_{T''} \not\equiv 0 (\mathrm{mod}\, l)$. Thus, $0 = R(\bar{L}_{T''}) \geqq R(L) \cap L_{T''}$, giving $d \leqq 2$ and finally $d = 2$.

Suppose $d=1$. We argue that not all of p,q,r are in $l\mathbb{Z}$. Suppose otherwise. Remove the "middle node" β of T to get nonempty diagrams of types $A_{p-1}, A_{q-1}, A_{r-1}$. Let L_1, L_2, L_3 be the corresponding lattices. Write $\alpha_1^{(1)}, \ldots, \alpha_{p-1}^{(1)}$ for the vertices, in order, of the "p-branch" and set $z_1 = \sum_{k=1}^{p-1} k\alpha_k^{(1)}$. Then $(z_1, \alpha_1^{(1)}) = -2+2\cdot 1=0$, $(z_1, \alpha_k^{(1)}) = k-1-2k+k+1=0$ for $k\in\{2,\ldots,p-2\}$ and $(z_1, \alpha_{p-1}^{(1)}) = p-2-2(p-1) = -p \equiv 0 (\mathrm{mod}\, l)$. In other words, $\bar{z}_1$ lies in radical of $\bar{L}_1$. In similar notation, we get $z_2 = \sum_{k=1}^{q-1} k\alpha_k^{(2)}$ and $z_3 = \sum_{k=1}^{r-1} k\alpha_k^{(3)}$ for the other branches. Since $\bar{L}_0 := \bar{L}_1 \oplus \bar{L}_2 \oplus \bar{L}_3$ (orthogonal direct sum) has codimension 1 and $R(\bar{L}_0)$ contains the independent set $\{\bar{z}_1, \bar{z}_2, \bar{z}_3\}$, we get $d \geqq \dim(R(\bar{L}_0) \cap \bar{\beta}^{\perp}) \geqq 2$, a contradiction. So, l fails to divide one of p,q,r, proving (ii).

Remark 3.6. A bit more work shows that, when $d=2$, $R(\bar{L})$ is spanned by $z_1 - z_2$ and $z_2 - z_3$.

In order to state the following classification result, we need some definitions. A *reflection* means a diagonalizable element of some $GL(n,F)$, $\mathrm{char}\, F \neq 2$, with eigenvalues -1 (once), 1 ($n-1$ times). For $G \leqq GL(n,F)$, the prefix "P" shall indicate the image of G in $PGL(n,F)$. In the case of a reflection in an orthogonal group, given by the formula $r: x \mapsto x - \dfrac{2(x,y)}{(y,y)} y$ for some vector y, the class of (y,y) in $F^{\times}/\{f^2 | f \in F^{\times}\}$ indicates the conjugacy class of the reflection in the orthogonal group. Thus $R^{\delta}O(V)$ indicates the subgroup of $O(V)$ generated by reflections which correspond to $\delta \in F^{\times}$. Also $RU(V)$, $RL(V)$ denotes the subgroup of $GL(V)$, $GU(V)$ generated by reflections. The superscript "ε" in $O^{\varepsilon}(n,F)$ indicates the type of quadratic form, as in [1, p. 210]. A *wreath product* XY means a semidirect product of $X \times \ldots \times X$ with Y, where Y permutes the factors regularly (or "freely"). Finally, $\Sigma L(3,4)$ means $SL(3,4)$ extended by a field automorphism of order 2.

Proposition 3.7 (Wagner [15], Serežkin and Zalesskiĭ [14]). *Let F be a field, $l = \mathrm{char}\, F \neq 2$, V a finite dimensional vector space over F, $n = \dim V$, G an irreducible subgroup of $GL(V)$ generated by reflections. Then $\bar{G}$, the image of G in $PGL(V)$, is one of the following:*

(i) $PRL(V')$, *where F' is a subfield of F, V' an F'-subspace of V with* $\dim_{F'}(V') = n$.

(ii) $PRU(V')$, *V' as above;*

(iii) $P\Sigma L(3,4)$, $n=4$, $l=3$; *here* $\bar{G} \leqq PGU(4,3)$;

(iv) $PR^{\delta}O(V')$, *V' as above;*

(v) $A_5\mathbb{Z}_2$, $n=4$, $l \neq 5$; $\bar{G} \leqq PO^{+}(4, F_l(\sqrt{5}))$;

(vi) PW_{D_n} *or* PW_{B_n};

(vii) $2^4 \cdot \Sigma_6$ *(nonsplit);* $n=4$ *and* $\sqrt{-1} \in F$;

(viii) $2^4 \cdot \Sigma_5$ *(split);* $n=4$ *and* $\sqrt{-1} \in F$;

(ix) $P\Omega(5,3)$, $n=5$, $l \neq 3$, $\sqrt[3]{1} \in F$;

(x) $PR^{-}O^{-}(6,3)$, $n=6$, $l \neq 3$, $\sqrt[3]{1} \in F$;

(xi) $PO^{+}(4,3)$, $n=4$, $l=3$;

(xii) $PO(5,3)$;

(xiii) $\mathrm{Sp}(6,2)$, $n=7$;
(xiv) $O^+(8,2)$, $n=8$;
(xv) Σ_{n+1}, $l \nmid n+1$;
(xvi) Σ_{n+2}, $l \mid n+2$, $n>4$.

Proposition 3.8 (McLaughlin [12]). *Let $G \leqq GL(n,2)$ be an irreducible subgroup generated by transvections. Then $G \cong O^{\varepsilon}(n,2)$ (n even, $\varepsilon=+,-$, $(n,\varepsilon) \neq (4,2)$); $\mathrm{Sp}(n,2) \cong O(n-1,2)$ (n even); or Σ_m for $m=n+1$ or $n+2$, n even.*

Also, the conjugacy class of elements of G acting as transvections in some representation is unique, except for $G \cong \Sigma_6$ in which there are two such classes.

4. The Case l odd

We identify $\bar{W}$, making use of results in Sect. 3. All the "hard work" is contained in Proposition 3.7, with which many possibilities in our situation may be eliminated.

Proposition 4.1. *Let $l \neq 2$.*

(i) *If T has type A_n, D_n or E_n, $\bar{W} \cong W$. Also, $d'=1$ for A_2 and $l=3$; $d'=0$ otherwise.*

(ii) *If T has type $\tilde{E}_{n-1}$, $\bar{W} \cong \mathbb{Z}_l^{n-1} \cdot W_{T'}$, split extension, where T' is a deletion of T of type E_{n-1}, respectively. Also $d=d'$ unless T has type $\tilde{E}_6$ and $l=3$, in which case $d'=1$, $d=2$.*

(iii) *If T is not "classical" (as in (i) or (ii)), $\bar{W} \cong \mathbb{Z}_l^{n'd} \cdot G$, split extension, and $G \cong O^{\varepsilon}(n',2)$ is an orthogonal group.*

Remark 4.2. The invariant ε may be easily computed from $\Delta_{T'}$ and n', where T' is a subdiagram of T with n' nodes and $(\Delta_{T'}, l)=1$; namely, $\varepsilon=+$ if and only if $\Delta_{T'}(-1)^{n'}$ is a square in F (see [1, p. 210]).

Proof. (i) The first statement follows since the kernel of $W \to \bar{W}$ must be a normal l-subgroup of the finite group W and $O_l(W)=1$ for such T unless $l=3$ and T has type A_2. In the latter case, the result may be checked directly.

(ii) Let T_0 be the normal "translation subgroup" of W. By Corollary 3.5, $d=1$ or $n=6$, $l=3$, $d=2$. We must argue here that $\bar{T}_0 \neq 1$. However, if $\bar{T}_0=1$, $\bar{W}$ acts like $\bar{W}_{T'}$, for an appropriate deletion T'. But then, $[\bar{L}, \bar{W}]=[\bar{L}, \bar{W}_{T'}]$ has rank at most $n-1$, in contrast with Lemma 3.1. Since $\bar{\Delta}=0$, the only way for $d'<d$ to occur is when $d=2$, i.e. type $\tilde{E}_6$. There, it is clear that $d'=1$.

(iii) Since T is not classical, $n \geqq 8$. Without loss, $p \geqq q \geqq r \geqq 2$ and $p \geqq 4$.

We claim that G is an orthogonal group. We argue by induction on n.

If $n=8$, $p=4$ and $q=r=3$ since T is nonclassical. Here, $\Delta=36-12-12-9=3$. If $l \neq 3$, Proposition 3.7 applies to show that $\bar{W} \cong W_{E_8}$, Σ_9, Σ_{10} or an orthogonal group. Since $\bar{W} \geqq \bar{W}(3,3,3) \cong \mathbb{Z}_l^6 \cdot W_{E_6}$, the first three possibilities are out. If $l=3$, $d=1$ and a similar argument with $W(2,3,4) \cong W_{E_7}$ shows that $G \cong W_{E_7}$ or an orthogonal group. The shape of $T(4,3,3)$ implies that G contains a reflection group of shape $\mathbb{Z}_2 \times W_{A_6}$. So, W_{E_7} is out and the claim follows.

Suppose $n \geqq 9$. If T has a nonclassical deletion T', we may use induction, Proposition 3.7 and the fact that W_{E_8} and Σ_{n+2} do not involve a copy of the orthogonal group $G_{T'}$. Now suppose that T has no such deletion.

If $r \geqq 3$, $q=3$ and $p=4$, against $n \geqq 9$. So $r=2$.

Since $r=2$, $\Delta = 2pq - pq - 2p - 2q = pq - 2(p+q)$. We have $q \geqq 3$, or else T is classical (type D_n). Suppose that $q=3$. Then $p = n-q-r+2 \geqq 6$ and even $p \geqq 7$ or else T has type $\tilde{E}_8$. Since every deletion is classical, $p=7$ and $n=10$. Then $\Delta = 7 \cdot 3 \cdot 2 - 7 \cdot 3 - 7 \cdot 2 - 3 \cdot 2 = 1$, so that $\bar{W}$ is irreducible on $\bar{L}$ and $\dim \bar{L} > 8$. By Proposition 3.7 either $\bar{W} \cong \Sigma_{11}$ or Σ_{12} (impossible, since W_{E_8} is not embeddable in Σ_{12}) or it is an orthogonal group, as required. So, $q \geqq 4$. Since every deletion is classical but T is not, $p=5$, $q=4$ and $r=2$. Then $\Delta = 5 \cdot 4 \cdot 2 - 5 \cdot 4 - 5 \cdot 2 - 4 \cdot 2 = 2$, and $\bar{W}$ operates irreducibly in dimension $n=9$. As above, Proposition 3.7 gives the result.

Our claim that G is an orthogonal group is now verified.

We now claim that $d' = d$. Without loss, $\Delta \equiv 0 (\mathrm{mod}\, l)$. Suppose that T' is a deletion with $G_T \cong G_{T'}$. Since $[\bar{L}, W_{T'}]$ has dimension $n-1$, $\bar{W}_T > \bar{W}_{T'}$, and $d' > 0$. So, we may suppose further that $d=2$. By definition of "K", $[\bar{L}_{T'}, K_{T'}] = R(\bar{L}_{T'})$. Also, $R(\bar{L}_{T'}) \leqq R(\bar{L}_T)$ for any deletion T', since $d=2$.

If $T' \neq T''$ are deletions and $K_{T'} \neq K_{T''}$, we are done, as $d' \geqq 2$ and $d' = 2$ follow. So, we may assume that $R(\bar{L}_{T'}) = R(L_{T''})$. Then, in $\bar{L}^* = \bar{L}/\bar{L}_{T'} \cap \bar{L}_{T''}$, the three nontrivial subspaces are the images of $\bar{L}_{T'}$, $\bar{L}_{T''}$, and $R(\bar{L})$. If $X \in \{L_{T'}, L_{T''}\}$, a reflection from X in its action on L^* preserves the images of X and $R(\bar{L})$, hence of the third space as well. Therefore, $[\bar{L}^*, \bar{W}] = 0$, which conflicts with $\bar{L} = [\bar{L}, \bar{W}]$.

To complete the proof that $d' = d$, we need to consider the case that every deletion T' satisfies $G_T \ncong G_{T'}$. Then every T' is classical, whence $n \leqq 11$. If $n=8$, $p=4$, and $q=r=3=l$. Let T' and T'' be the two deletions of type E_7 and let z', z'' be the involutions in $Z(W_{T'})$, $Z(W_{T''})$, respectively. Both z', z'' are -1 on $\bar{L}/R(\bar{L})$, hence map to generators of $Z(G)$. If $d'=0$, $z' = z'' \in Z(W)$, which is certainly false since $\bar{L} = \bar{L}(1-z') \oplus \bar{L}(1+z')$ is a submodule decomposition incompatible with Lemma 3.1. So, we may assume that $n \geqq 9$. Then $r=2$ and $(\Delta, l) = 1$ (see an earlier paragraph), which is incompatible with $d' > 0$.

Finally, we show that $1 \to K \to \bar{W} \to G \to 1$ is always split. Without loss, $\Delta \equiv 0 (\mathrm{mod}\, l)$. If there is a deletion T' with $G_{T'} \cong G_T$, we use induction. If there is none, every deletion is classical and the analysis of the last paragraph shows that $r \neq 2$. Therefore, $n=8$, $p=4$, $q=r=3=l$, and $G \cong W_{E_7}$. Let z be an involution of $\bar{W}$ mapping to the element of order 2 in $Z(G)$. Then $C_{\bar{W}}(z) \cong G$ complements K. The arguments for (iii) are now complete.

5. The Case $l=2$

The analysis here, in spirit, resembles that of Sect. 4. We use Proposition 3.8 here instead of Proposition 3.7. The pleasantness of having a relatively short list of groups in the conclusion of Proposition 3.8 is balanced by the increased complexity involved in determining G and d'.

In characteristic 2, we need the notion of a quadratic form as well as that of the bilinear form induced on $\bar{L}$ by the one on L. The relevant quadratic form on L is

$Q(x)=\frac{1}{2}(x,x)\in\mathbb{Z}$. It induces a well defined map $\bar{Q}:\bar{L}\to F=\mathbb{F}_2$ with $(,)$ as its associated bilinear form. See [6] for a discussion of generalities.

Besides the earlier definition of the radical $R(\bar{L})$, we set

$$R_{\bar{Q}}(\bar{L})=\{\bar{x}\in R(\bar{L})|\bar{Q}(\bar{x})=0\}.$$

In general, $\dim R(\bar{L})/R_{\bar{Q}}(\bar{L})=0$ or 1.

The main results of this section have lengthy statements and are embodied in Lemmas 5.4, 5.5, and 5.6. Notice that $d'=d$ is "typical" for l odd, but certainly not so for $l=2$.

Lemma 5.1. *Let L be the lattice for $T=T(p,q,1)$, i.e. type A_m, $m=p+q-1$. Let $z=\sum_{k=1}^{m} k\alpha_k$, where $\alpha_1,\ldots,\alpha_m$ is the "natural" order of the roots, i.e. α_i and α_j are connected if and only if $i-j=\pm 1$. Then, if $m+1\equiv 0(\mathrm{mod}\, l)$, $\bar{z}$ spans $R(\bar{L})$. Also, $(z,z)=-(m+1)$, whatever the congruence of m modulo l.*

Proof. If $k\in\{2,\ldots,m-1\}$, $(\alpha_k,z)=k-1-2k+k+1=0$. Also

$$(\alpha_1,z)=-2+2=0 \quad \text{and} \quad (\alpha_m,z)=m-1-2m=-1-m\equiv 0(\mathrm{mod}\, l).$$

So, $\bar{z}\in R(\bar{L})$, $d\geqq 1$. Since $\Delta(p,q-1,1)=m\not\equiv 0(\mathrm{mod}\, 2)$, $d=1$.

We have

$$(z,z)=\sum_{k=1}^{m} k^2(\alpha_k,\alpha_k)+2\sum_{k=1}^{m-1} k(k+1)(\alpha_k,\alpha_{k+1})=-2\frac{m(m+1)(2m+1)}{6}$$
$$+2\left[\frac{(m-1)m(2m-1)}{6}+\frac{(m-1)m}{2}\right]=-2m^2+m^2-m=-m(m+1).$$

Lemma 5.2. (i) *If T has type A_n, $d=0$ or 1 and $d\equiv n(\mathrm{mod}\, 2)$, all n, and $d'=0$ unless $n=3$ in which case $d'=1$.*

(ii) *If T has type D_n, $n\geqq 4$, then $d'=1$, $d=1$ or 2 and $d\equiv n(\mathrm{mod}\, 2)$; also, $\bar{L}/[\bar{L},K]$ is isometric to $\bar{L}_{T'}$ where T' is a deletion of T of type A_{n-1}.*

Proof. (i) The statement about d follows from Lemma 5.1. Obviously, $d'=0$ if $O_2(W_{A_n})=1$, which is the case for $n\neq 1,3$. For $n=1$, $\bar{W}=1$. For $n=3$, $d'\neq 0$ since $\dim(\bar{L}/R(\bar{L}))=2$, $|GL(2,2)|=6$, and $|W_{A_3}|=24$.

(ii) We may take $r=q=2$. Then $\Delta=4p-2p-2p-4=-4$ and we use Corollary 3.5 to get the statements about d. We get $d'>0$ by a simple calculation, e.g. with [2, p. 256]. So, $d'=1$ if $d=1$. Suppose that $d=2$. If $d'\neq 1$, $d'=2$, and $U=O_2(W_{D_n})$ has order exceeding 2^{n-1} and lies in K. So, $n=4$ and $U\cong 2_+^{1+4}$ is extraspecial. But K is abelian, by 3.2, a contradiction.

To prove the second statement, we notice that $[\bar{L},K]\leqq R_{\bar{Q}}(\bar{L})$; for if $g\in K$ and $y\in\bar{L}$, $x=y+y^g$, then

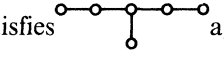
$$\bar{Q}(y)=\bar{Q}(y^g)=\bar{Q}(x+y)=\bar{Q}(x)+\bar{Q}(y)+(x,y)=\bar{Q}(x)+\bar{Q}(y)$$

and $\bar{Q}(x)=0$, as required. Thus, a quadratic form is well-defined on $\bar{L}/[\bar{L},K]$ by taking coset representatives from $\bar{L}_{T'}$, a complement to $[\bar{L},K]$ ($\bar{L}_{T'}$ is a complement: K is generated by elements which change coordinate signs at two places in

the usual description of L as in [2],: in that notation, $[\bar{L}, K] = 2L + 2e_i$, and our assertion is now clear).

By virtue of Lemma 5.2, we may and do assume $p \geqq q \geqq r \geqq 2$ and $q \geqq 3$ from now through Lemma 5.6.

Lemma 5.3. $G \ncong GL(n', 2)$, Σ_m.

Proof. Since G lies in an orthogonal group on $\bar{L}/R(\bar{L})$, $G \ncong GL(n', 2)$ (as $n' > 2$). We show $G \ncong \Sigma_m$. If so a subset of the transvections in Σ_m satisfies  against $W_{E_6} \not\hookrightarrow \Sigma_7$, contradiction.

Lemma 5.4. *If* $(\Delta, 2) = 1$, $G \cong O(\bar{Q})$.

Proof. Since $G \hookrightarrow O(\bar{Q})$, we apply McLaughlin's Theorem, 3.8, and Lemma 5.3.

Lemma 5.5. *Suppose that* $\Delta \equiv 0 (\mathrm{mod}\, 2)$ *and* $d = 1$. *Then n is odd and one or three of* p, q, r *are odd.*

If p, q, r *are odd,*

$$\bar{W} \cong \begin{cases} \mathrm{Sp}(n-1, 2) & \text{for} \quad n \equiv 1 (\mathrm{mod}\, 4), \\ 2^{n-1} \cdot O^{\varepsilon}(n-1, 2) & \text{for} \quad n \equiv 3 (\mathrm{mod}\, 4). \end{cases}$$

If $\{p, q, r\} = \{p_1, q_1, r_1\}$ *where* p_1, q_1 *are even and* r_1 *is odd,*

$$\bar{W} \cong \begin{cases} \mathrm{Sp}(n-1, 2) & \text{for} \quad p_1 + q_1 \equiv 2 (\mathrm{mod}\, 4) \\ 2^{n-1} \cdot O^{\varepsilon}(n-1, 2) & \text{for} \quad p_1 + q_1 \equiv 0 (\mathrm{mod}\, 4). \end{cases}$$

In all cases, $\bar{W}$ is a split extension.

Proof. The assertions about n, p, q, r are clear from Lemmas 3.3 and 3.4.

Since $d = 1$, there is a deletion T_1 of T with $\Delta_1 = \Delta(T_1) \not\equiv 0 (\mathrm{mod}\, 2)$. We may arrange for $T_1 = T(p_1, q_1, r_1)$ to not have type A_{n-1} (easy). Note that T_1 does not have type D_{n-1} since Δ_1 is odd. Without loss, $p_1 \geqq q_1 \geqq 3$. If L_1 and W_1 are the corresponding lattice and group, $\bar{W}_1 \cong O^{\varepsilon}(n-1, 2)$, by Lemma 5.4. Also, $\bar{L} = [\bar{L}, \bar{W}]$, $\bar{L}_1 = [\bar{L}_1, \bar{W}_1] = [\bar{L}, \bar{W}_1]$, irreducible action of W_1 on $\bar{L}_1$ and Proposition 3.8 imply that $\bar{W} \cong \mathrm{Sp}(n-1, 2)$, $2^{n-1} \cdot \mathrm{Sp}(n-1, 2)$ or $2^{n-1} \cdot O^{\varepsilon}(n-1, 2)$.

The parities of n, p, q, and r imply that there is an extension S of T by one node so that $\Delta(S) \not\equiv 0 (\mathrm{mod}\, 2)$. Thus, $W(S) \cong O^{\varepsilon}(n+1, 2)$ and the embedding $\bar{W} \hookrightarrow W(S)$ imply that $\bar{W} \cong \mathrm{Sp}(n-1, 2)$ or $2^{n-1} \cdot O^{\varepsilon}(n-1, 2)$, according to whether $\bar{W}$ stabilizes a nonsingular or singular vector of $\bar{L}(S)$. (Also, $\varepsilon = \varepsilon'$ in the latter case). Visibly, the extensions are split.

Case 1. p, q, r odd. Let Φ_1 be the nodes of T along the union of the p and q-branches and let $z_1 = \sum_{k=1}^{p+q+1} k\alpha_k$ be a sum taken in a natural way along Φ_1 (in either direction). Let Φ_2 be the $r-2$ outermost roots along the r-branch ($r \geqq 3$ here) and let $z_2 = \sum_{k=1}^{r-2} k\beta_k$ be a sum taken along Φ_2. Let L_1, L_2 be the corresponding lattices. Then $\bar{z}_i$ spans $R(\bar{L}_i)$, $i = 1, 2$, by Lemma 5.1. Let $\{\delta\} = \Phi - \Phi_1 - \Phi_2$. Then

$(z_1, \delta) \equiv (z_2, \delta) \equiv 1 (\mathrm{mod}\, 2)$, whence $R(\bar{L}) = F(\bar{z}_1 + \bar{z}_2)$. Also,

$$(z_1 + z_2, z_1 + z_2) = (z_1, z_1) + (z_2, z_2) \equiv p + q + r - 1 \equiv n + 1 (\mathrm{mod}\, 4),$$

using Lemma 5.1 and the fact that $p+q$ and $r-1$ are even. So, in $\overline{L(S)}$, $\overline{z_1 + z_2}$ represents a singular, nonsingular vector as $n+1 \equiv 0, 2 (\mathrm{mod}\, 4)$, respectively. Since $\bar{W}$ stabilizes $\overline{z_1 + z_2}$, we get $\bar{W} \cong 2^{n-1} \cdot O^{\varepsilon}(n-1, 2)$ or $\mathrm{Sp}(n-1, 2)$, accordingly.

Case 2. r_1 is odd and p_1, q_1 are even. We copy the argument for Case 1, except that we let Φ_2 be the $r_1 - 1$ outermost roots of the r_1-branch. We do not have $(L_1, L_2) = 0$, though $(\bar{z}_1, \bar{L}_2) = 0$ since $p_1 \equiv q_1 \equiv 0 (\mathrm{mod}\, 2)$. Thus, $\bar{z}_1$ spans $\mathrm{Rad}\, \bar{L}$. We have $(z_1, z_1) \equiv p_1 + q_1 (\mathrm{mod}\, 4)$, and we finish as before.

Lemma 5.6. *Suppose $d = 2$. Then n is even and so are p, q, r. We have*

$$\begin{cases} d' = 2, & \bar{W} \cong (2^{n-2})^2 \cdot O^{\varepsilon}(n-2, 2) \quad \text{if} \quad p \equiv q \equiv r (\mathrm{mod}\, 4); \\ d' = 1, & \bar{W} \cong 2^{n-2} \cdot \mathrm{Sp}(n-2, 2) \quad \text{otherwise}. \end{cases}$$

These extensions are all split.

Proof. By Lemma 3.4, all of p, q, r are even. Without loss, $p \geqq q \geqq r$. Let T_1, T_2, T_3 be the three deletions of T (at the p, q, and r branches, respectively, and let $T_{ij} = T_i \cap T_j$, $\Delta_i = \Delta(T_i), \ldots$. Then Δ_{ij} is odd. Let L, W, G, L_i, W_i, G_i, etc. have the obvious meanings.

Note that there is an enlargement S of T by two nodes (along different branches) so that $\Delta(S)$ is odd. Thus, $\bar{W}$ maps into the stabilizer H in $\overline{W(S)} \cong O^{\varepsilon}(n+2, 2)$ of a 2-dimensional subspace $R(\bar{L})$, which is singular with respect to the bilinear form on $\overline{L(S)}$. Consequently, by Lemma 5.4, $G \cong O^{\varepsilon}(n, 2)$ or $\mathrm{Sp}(n, 2)$, as $R_{\bar{Q}}(\bar{L})$ has codimension 0 or 1 in $R(\bar{L})$, respectively.

Suppose $R_Q(\bar{L}) = R(\bar{L})$. As in Case 2 of Lemma 5.5, $\bar{Q}(\bar{z}_i) = 0$ and

$$(z_i, z_i) \equiv p_i + q_i \equiv 0 (\mathrm{mod}\, 4),$$

for all i. Thus, $p \equiv q \equiv r (\mathrm{mod}\, 4)$ and we have $\bar{W}_i \cong 2^{n-2} \cdot O^{\varepsilon_i}(n-2, 2)$ and $\langle \bar{z}_i \rangle = [\bar{L}_i, O_2(\bar{W}_1)]$, for at least two i, since $p \geqq q \geqq 4$ (but $r \geqq 2$, only). Let i and j be two such indices. If $\bar{z}_i, \bar{z}_j$ are linearly independent, we get $d' = 2$, as required. This is proven by an argument like the one for $d' = d$ in Proposition 4.1 (iii).

Suppose $R_{\bar{Q}}(\bar{L}) \neq R(\bar{L})$. We claim that the even integers p, q, r are not pairwise congruent modulo 4. Suppose otherwise. Let $\alpha_1^{(i)}, \alpha_2^{(i)}, \ldots$ be the roots along the p, q, and r branches for $i = 1, 2, 3$, respectively, listed in the inward direction and excluding the node β of valency 3. Set $z_i = \sum_{k=1}^{t_i - 1} k \alpha_k^{(i)}$, $t_1 = p$, $t_2 = \gamma$, $t_j = r$. Then $(z_i, z_i) = -t_i(t_i - 1)$ and $(z_i, \alpha_k^{(j)}) \equiv 0 (\mathrm{mod}\, 2)$ for all i, j, k. Since $(z_i, \beta) \equiv 1 (\mathrm{mod}\, 2)$ for all i, $R(\bar{L})$ is spanned by the $z_i - z_j$. Since

$$(z_i - z_j, z_i - z_j) \equiv t_i + t_j \equiv 0 (\mathrm{mod}\, 4),$$

we get $\dim R_{\bar{Q}}(\bar{L}) \geqq 2$, against our assumption for this paragraph. Now arrange notation so that $\{p, q, r\} = \{p_1, q_1, r_1\}$, $p_1 \not\equiv q_1 \equiv r_1 (\mathrm{mod}\, 4)$. If $r > 2$, $p \geqq q \geqq r \geqq 4$, and we argue as follows: By Lemma 5.5,

$$\overline{W(p_1, q_1, r_1 - 1)} \cong \mathrm{Sp}(n-2, 2), \qquad W(p_1 - 1, q_1, r_1) \cong 2^{n-2} \cdot O^{\varepsilon}(n-2, 2).$$

Since $G \cong \mathrm{Sp}(n-2,2)$, this forces $d'=1$, $O_2(\bar{W}) \cong 2^{n-2}$, $[\bar{L}, O_2(W)] = \mathrm{Rad}_Q(\bar{L})$ and we have the required split extension. Now suppose that $r=2$. Then $n \geqq 8$. If $r=p_1$, then

$$\overline{W(p_1, q_1-1, r_1)} \cong \overline{W(p_1, q_1, r_1-1)} \cong \mathrm{Sp}(n-2,2) \cong G$$

(see the second paragraph of the proof), corcing $d' \neq 0$, whence $d'=1$ and we have split extension. So, we may assume that $r \neq p_1$. Without loss, $r=q_1$. Then $r_1 \equiv 2 (\mathrm{mod}\, 4)$ implies $r_1 \geqq 6$, and so $n \geqq 2+4+6-2=10$. The argument for $r>2$ now works. The proof is now complete.

If $\bar{L}$ is nonsingular with respect to the form, the *type* of the quadratic form is $\varepsilon = \pm$, as the Witt index is maximal of nonmaximal; see [6]. We say that the geometry on $\bar{L}$ has *type* $\varepsilon = \pm$ if $R(\bar{L}) = R_{\bar{Q}}(\bar{L})$ and the type of $(\bar{L}/R(\bar{L}), \bar{Q})$ is ε, and we say that the type is $\varepsilon = 0$ if $R(\bar{L}) \neq R_{\bar{Q}}(\bar{L})$. During the previous arguments for $l=2$, the need to know ε (when $\varepsilon = \pm$) was avoided. Here, we give a simple inductive procedure for determining ε (see Corollary 5.9).

Lemma 5.7. *If L is the A_4-root lattice (A_2-root lattice), $\Delta = -5 (\Delta = -3)$ and $L/2L$ has type $\varepsilon = -$.*

Proof. Easily, $\Delta = -5$ (respectively, -3). Also, $O^+(4,2) \cong \Sigma_3 \mathbb{Z}_2$ (order $2^3 \cdot 3^2$) and $O^-(4,2) \cong \Sigma_5$ [3, 9]. Since $W(A_4) \cong \Sigma_5$ acts faithfully on the lattice modulo 2, $\varepsilon = -$. Since $O^+(2,2) \cong \mathbb{Z}_2$ and $O^-(2,2) \cong \Sigma_3$ [3, 9], similar remarks get $\varepsilon = -$ in this case. An alternate argument is to compute the relevant invariant in the Clifford algebra [6].

Lemma 5.8. *Suppose $p \geqq 5$. Let $\alpha_1, \alpha_2, \ldots$ be the roots on the p-branch listed beginning with the outermost one. Let $L_1 = \sum_{i=1}^{4} \mathbb{Z}\alpha_i$, $\beta = \sum_{k=1}^{5} k\alpha_k$, $L_2 = \mathbb{Z}\beta + \sum_{\alpha \in \Phi_1} \mathbb{Z}\alpha$, where*

$$\Phi_1 = \Phi - \{\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5\}.$$

Then $(L_1, L_2) = 0$, $\bar{L} = \bar{L}_1 \oplus \bar{L}_2$, and $\bar{L}_2$ is isometric to the lattice for $T(p-4, q, r)$ taken modulo 2.

Proof. We have $(\alpha_1, \beta) = -2+2=0$, $(\alpha_i, \beta) = i-1-2i+i+1=0$ for $1<i<5$, so that $(L_1, L_2)=0$. Also $\beta \in \alpha_5 + L_1 + 2L$, whence $\bar{L} = \bar{L}_1 + \bar{L}_2$ and dimension considerations give $\bar{L} = \bar{L}_1 \oplus \bar{L}_2$. Finally, note that if $\alpha \in \Phi_1$, $(\alpha, \beta) \equiv (\alpha, \alpha_5) (\mathrm{mod}\, 2)$ and $(\beta, \beta) = -5.6$ so that $\bar{Q}(\bar{\beta}) = 1 = \bar{Q}(\bar{\alpha}_5)$; see Lemma 5.1.

Corollary 5.9. *If $p \geqq 5$, $\varepsilon = -\varepsilon'$, where ε' is the type of the geometry for $L(p-4, q, r)$ modulo 2.*

Proof. The function "ε" is multiplicative over orthogonal direct sums of nonsingular quadratic spaces over $\mathbb{F}_2$.

We now complete our description of the procedure by discussing the "small cases", i.e. p, q, r all less than 5. They are checked with easy calculations.

$r=1$: $n=p+q+1-2=p+q-1$, L is just the A_n-lattice, and $\Delta=n+1=p+q$. Since we may as well take $q=1$ here, we may assume that $n=p\leqq 4$. The situation is:

$$\begin{array}{rl} n=1 & \varepsilon=0 \\ 2 & - \\ 3 & - \\ 4 & -. \end{array}$$

$r=2$: $n=p+q+2-2=p+q$, $2\leqq q\leqq p\leqq 4$. For $q=2$, the lattice is the D_n-lattice, and the sitaution is the following, according to Lemma 5.2(ii), Corollary 5.9, and the discussion for $r=1$:

$$\begin{array}{rl} n=4 & - \\ 5 & - \\ 6 & 0 \end{array}$$

When $q=3$, we get the E_6-lattice and $\varepsilon=-$ for $p=3$ (use $E_{E_6}\cong O^-(6,2)\ncong O^+(6,2)$ [3,9]), the E_7-lattice and $\varepsilon=0$ for $p=4$ (use $W_{E_7}\cong \mathrm{Sp}(6,2)\times\mathbb{Z}_2$ [3]). When $q=4=p$, we get the "extended E_7"-lattice and $\varepsilon=0$ by noting, as in Lemma 5.2(ii), that $[\bar{L},K]\leqq R_{\bar{Q}}(\bar{L})$ and $[\bar{L},K]$ is complemented by $\bar{L}_T$, where T' has type E_7.

In the remaining case $r=3$ and 4, we get ε by observing that when $(\Delta,2)=1$, the stabilizer in the orthogonal group $O^\varepsilon(n,2)$ of a singular vector looks like $2^{n-2}\cdot O^\varepsilon(n-2,2)$.

$r=3$: $n=p+q+3-2=p+q+1$, $3\leqq q\leqq p\leqq 4$. The situation is

$$\begin{array}{rll} (q,p)=(3,3) & \varepsilon=- & \text{("extended } E_6\text{"-lattice)} \\ (3,4) & \varepsilon=- & \\ (4,4) & \varepsilon=-. & \end{array}$$

The case (3,4) may be handled as follows. Here, $\Delta=3$ and the p-deletion lends to the group $2^6\cdot W_{E_6}=2^6\cdot O^-(6,2)$ and then $\varepsilon=-$, using the observation. The case (4,4) leads to $\varepsilon=-$ via a q-deletion, the case for (3,4) and the observation.

$r=4$: $q=p=4$. A p-deletion and the observation gives $\varepsilon=-$.

This concludes the discussion for $l=2$.

Summary of the Structure of $\bar{W}$

$p\geqq q\geqq r\geqq 1$;

$d=0$, 1 *or* 2 *(see* 3.5*)*; ε *may be computed using* 4.2 *for* $l\neq 2$ *and* 5.9 *for* $l=2$.

$l\neq 2$: (see 4.1)

$l=2$: $r=1$, $d=0$ or 1, $d\equiv n+1\,(\mathrm{mod}\,2)$, $d'=\begin{cases}0 & n\neq 3\\ 1 & n=3\end{cases}$, $W\cong\bar{W}\cong W_{A_n}$,

$r=2$, $q=2$: $d\equiv n(\mathrm{mod}\,2)$, $d'=1$, $W\cong\bar{W}\cong W_{D_n}$,

$r=2$, $q>2$ *or* $r>2$: $\Delta\equiv 1(\mathrm{mod}\,2)$, $d=0$, $\bar{W}\cong O^\varepsilon(n,2)$.

$$\Delta \equiv 0(\mathrm{mod}\,2), \quad d=1, \quad p,q,r \text{ odd}, \quad \bar{W} \cong \begin{cases} \mathrm{Sp}(n-1,2), & n \equiv 1(\mathrm{mod}\,4), \\ 2^{n-1} \cdot O^{\varepsilon}(n-1,2), & n \equiv 3(\mathrm{mod}\,4); \end{cases}$$

exactly two of p,q,r, say p_1 and q_1, are even,

$$\bar{W} \cong \begin{cases} \mathrm{Sp}(n-1,2), & p_1+q_1 \equiv 2(\mathrm{mod}\,4), \\ 2^{n-1} \cdot O^{\varepsilon}(n-1,2), & p_1+q_1 \equiv 0(\mathrm{mod}\,4). \end{cases}$$

$d=2$, p, q, r even,

$$\begin{cases} p \equiv q \equiv r(\mathrm{mod}\,4), & d'=2, \quad \bar{W} \cong (2^{n-2})^2 \cdot O^{\varepsilon}(n-2,2), \\ \text{otherwise} & d'=1, \quad \bar{W} \cong 2^{n-2} \cdot \mathrm{Sp}(n-2,2). \end{cases}$$

References

1. Artin, E.: Geometric algebra. New York: Interscience 1957
2. Bourbaki, N.: Groupes et algebres de Lie. Paris: Hermann 1968, Chaps. 4–6
3. Carter, R.: Simple groups of Lie type. New York: Wiley 1972
4. Coble, A.: Algebraic geometry and theta functions. A.M.S. Colloq. Publications, X, Providence 1929 (2nd edn. 1961)
5. Coble, A.: Theta modular groups determined by point sets. Am. J. Math. **40**, 317–340 (1918)
6. Diendonne, J.: La géometrié des groupes classiques. Berlin, Heidelberg, New York: Springer 1971
7. Dolgachev, I.: Weyl groups and Cremona transformations. Proc. Summer Institute on Singularities, Arcata, 1981 (to appear)
8. Griess, R.: Automorphisms of extraspecial groups and nonvanishing degree 2 cohomology. Pac. J. Math. **48**, 403–422 (1973)
9. Gorenstein, D.: Finite groups. New York: Harper and Row 1968
10. Kac, V.G.: Simple irreducible graded Lie algebras of finite growth (Russian). Izv. Akad. Nauk. SSSR **32**, 1323–1367 (1968); (English translation) Math. USSR-Izv. **2**, 1271–1311 (1968)
11. Looijenga, E.: Rational surfaces with an anticanonical cycle. Ann. Math. **114**, 267–322 (1981)
12. McLaughlin, J.E.: Some subgroups of $\mathrm{SL}_n(\mathbb{F}_2)$. Ill. J. Math. **13**, 108–115 (1969)
13. Moody, R.: A new class of Lie algebras: J. Algebra **10**, 211–230 (1968)
14. Serežkin, V.N., Zalesskii, A.E.: Izv. Acad. Sci. USSR Ser. Mat. **44**, 1279–1307 (1980)
15. Wagner, A.: Determination of the finite primitive reflection groups over an arbitrary field of characteristic not 2. Parts I–III. Geometriae Dedicata **9**, 239–253 (1980); **10**, 191–203 (1981); **10**, 475–523 (1981)

Received June 14, 1982

Math. Ann. 263, 279–281 (1983)

Mathematische Annalen
© Springer-Verlag 1983

On a Problem of H. Berens

G. Godini
Department of Mathematics, INCREST, B-dul Pacii 220, R-79622 Bukarest, Romania

Let H be a real Hilbert space, G a nonempty subset of H and $x \in H$. Define

$$P_G(x) = \{g_0 \in G \mid \|x - g_0\| = d(x, G)\}$$

where $d(x, G)$ denotes the distance of x to G;

$$\Phi_G(x) = \bigcap_{r > d(x,G)} \overline{co}\{g \in G \mid \|g - x\| < r\}$$

where $\overline{co}$ denotes the closed convex hull. The set G is a *Chebyshev* set (respectively is *proximinal*) in H if for each $x \in H$ the set $P_G(x)$ is a singleton (respectively is nonempty). For each $x \in H$, $\Phi_G(x)$ is a nonempty, closed, convex set, and we always have $\overline{co}\, P_G(x) \subset \Phi_G(x)$ [1–3]. Following [5], a proximinal set G in H has *property* (β) if $\Phi_G(x) = \overline{co}\, P_G(x)$ for each $x \in H$. As it was observed in [1–3], if H is finite-dimensional then each closed set of H has this property.

Berens [1, 2] raised the following question: *Does each proximinal set in H have property* (β)? He noticed that an affirmative answer to this question would disprove Klee's conjecture [4] on the existence of nonconvex Chebyshev sets in a (possibly nonseparable) Hilbert space. A negative answer is established by Klee ([5], Theorem 2.7) for the Hilbert space $l^2(n)$ when n is a regular infinite cardinal such that $n^{\aleph_0} = n$ (e.g., an inaccesible cardinal), and Berens' question remains open for separable infinite-dimensional H. In this note a negative answer is established for *any* infinite-dimensional Hilbert space, and the proof uses only elements of functional analysis.

Lemma. *Let H be a Hilbert space, $u \in H$, $\|u\| = 1$ and $r > 0$. The set $G = \{y \in H \mid \langle u, y\rangle \leqq 0,\ \|y\| \geqq r\}$ is proximinal in H.*

Proof. Let $M = \{y \in H \mid \langle u, y\rangle = 0\}$. Let $x \in H \setminus G$ and $m_0 \in M$ such that $P_M(x) = \{m_0\}$.

Case 1. $\langle u, x\rangle > 0$ and $\|m_0\| \geqq r$.

Then $m_0 \in G$ and we have $m_0 \in P_G(x)$.

Case 2. $\langle u, x\rangle > 0$ and $\|m_0\| < r$.

If $m_0=0$, choose $g_0\in M$, $\|g_0\|=r$. If $m_0\neq 0$ let $g_0=rm_0/\|m_0\|$. In both cases we have $g_0\in M\cap G$. Let $g\in G$ and let $y=\lambda x+(1-\lambda)g$ for $\lambda=\langle u,g\rangle/(\langle u,g\rangle-\langle u,x\rangle)$. Then $y\in M$ and since $0\leqq\lambda<1$ it follows that $\|x-g\|\geqq\|x-y\|$. If $\|y-m_0\|\geqq\|m_0-g_0\|$ then $\|x-g\|^2\geqq\|x-y\|^2=\|x-m_0\|^2+\|m_0-y\|^2\geqq\|x-m_0\|^2+\|m_0-g_0\|^2=\|x-g_0\|^2$. If $\|y-m_0\|<\|m_0-g_0\|$, then $\|y\|<r$ and since $\|g\|\geqq r$ there is $0<\mu\leqq 1$ such that for the element $z=\mu g+(1-\mu)y$ to have $\|z\|=r$. Then

$$\|x-g\|\geqq\|x-z\|=\|x-y\|+\|y-z\| \tag{1}$$

$$\|z-m_0\|\geqq\|m_0-g_0\| \tag{2}$$

$$\|y-z\|\geqq\|z-m_0\|-\|y-m_0\|\geqq\|m_0-g_0\|-\|y-m_0\|>0. \tag{3}$$

Using (1), (3) and (2) it is easy to show that $\|x-g\|\geqq\|x-g_0\|$. Therefore $g_0\in P_G(x)$.

Case 3. $\langle u,x\rangle\leqq 0$.

If $x=0$, choose any $g_0\in G$, $\|g_0\|=r$. If $x\neq 0$, let $g_0=rx/\|x\|$. In both cases we have $g_0\in P_G(x)$, which completes the proof.

Theorem. *In any infinite-dimensional Hilbert space there is a proximinal set which lacks property* (β).

Proof. Let H be an infinite-dimensional Hilbert space. We distinguish two cases:

Case 1. H is separable.

It suffices to assume $H=l^2$. For $1\leqq n<\infty$, let e_n be the point of l^2 whose nth coordinate is 1 and all other coordinates are 0. Let $r_n>0$, $r_n\to 0$, and $x_0\in l^2$, $\|x_0\|=1$. Let us put $G_n=\{y\in l^2\mid\langle e_n,y\rangle\leqq -1,\ \|y+e_n\|\geqq r_n\}$. We show that the set

$$G=\{x_0\}\cup\bigcup_{n=1}^{\infty}(G_n\cup -G_n)$$

is proximinal in l^2 and that $\Phi_G(0)\neq\overline{co}\ P_G(0)$.

Obviously, $P_G(0)=\{x_0\}$. Let $r>d(0,G)=1$. Since $r_n\to 0$, for sufficiently large n there is an element $g\in G_n\subset G$ such that $\|g\|<r$. Since $-g\in G$ it follows that $0\in\overline{co}\,\{g\in G\mid\|g\|<r\}$ and so $0\in\Phi_G(0)\backslash\overline{co}\ P_G(0)$.

To see that G is proximinal, consider $x=(\xi_i)\in l^2$, $x\neq 0$, and let n_0 be such that

$$\xi_n^2<\min\left\{1,\frac{\|x\|^2}{2}\right\}\qquad(n>n_0) \tag{4}$$

$$r_n^2<\frac{\|x\|^2}{2}\qquad(n>n_0). \tag{5}$$

By (4) and (5) we get

$$\sum_{i\neq n}\xi_i^2=\|x\|^2-\xi_n^2>r_n^2\qquad(n>n_0) \tag{6}$$

Now we show that

$$d(x,G_n\cup -G_n)=1-|\xi_n|\qquad(n>n_0). \tag{7}$$

Indeed, we have

$$d(x, G_n \cup -G_n) = \min\{d(x, G_n), d(-x, G_n)\}.$$

Let $M_n = \{y \in l^2 \mid \langle e_n, y\rangle = -1\}$ and let $P_{M_n}(x) = \{m_n\}$, $P_{M_n}(-x) = \{m_n'\}$. Then $m_n = (\alpha_i)$ where $\alpha_i = \xi_i$ for $i \neq n$, $\alpha_n = -1$, and $m_n' = (\beta_i)$ where $\beta_i = -\xi_i$ for $i \neq n$, $\beta_n = -1$. For $n > n_0$, we have by (4) that $\langle e_n, \pm x\rangle > -1$, and by (6), $\|m_n + e_n\|^2 = \|m_n' + e_n\|^2 = \sum_{i \neq n} \xi_i^2 > r_n^2$, and so we are in the Case 1 from the proof of the lemma. Hence $m_n \in P_{G_n}(x)$ and $m_n' \in P_{G_n}(-x)$, for $n > n_0$, and $d(x, G_n \cup -G_n) = \min\{\|x - m_n\|, \|x + m_n'\|\} = \min\{1 + \xi_n, 1 - \xi_n\} = 1 - |\xi_n|$ which proves (7).

Let $n_1 > n_0$ be such that $|\xi_{n_1}| \geqq |\xi_n|$ for each $n > n_0$. By (7) we get $d(x, G_{n_1} \cup -G_{n_1}) = 1 - |\xi_{n_1}| \leqq 1 - |\xi_n| = d(x, G_n \cup -G_n)$, $n > n_0$. Hence

$$d(x, G_{n_1} \cup -G_{n_1}) = d\Big(x, \bigcup_{n > n_0} (G_n \cup -G_n)\Big).$$

Let

$$G_0 = \{x_0\} \cup (G_{n_1} \cup -G_{n_1}) \cup \bigcup_{n=1}^{n_0} (G_n \cup -G_n)$$

Then $G_0 \subset G$ and $d(x, G) = d(x, G_0)$. Hence, since G_0 is proximinal in l^2 (as the union of a finite number of proximinal sets (use Lemma)), it follows that $P_G(x) \neq \emptyset$.

Case 2. H is nonseparable.

Let H_0 be an infinite dimensional separable Hilbert subspace of H. By the above, there is a proximinal set G in H_0 which lacks property (β). It remaines to show that G is proximinal in H. Let $x \in H$ and $P_{H_0}(x) = \{h_0\}$. Then any $g_0 \in P_G(h_0)$ belongs to $P_G(x)$, which completes the proof of the theorem.

Corollary. *A Hilbert space H is finite-dimensional if and only if each proximinal set in H has property (β).*

In view of this result, all sufficient conditions known to imply property (β) for proximinal sets in a Hilbert space, are of interest [1, 2].

We conclude this note with the following question: Does there exist an infinite-dimensional Banach space E such that all its proximinal sets have property (β)? Notice that all finite-dimensional Banach spaces have this property [3].

References

1. Berens, H.: Best approximation in Hilbert space. Approximation theory. III. Cheney, E.W. (ed.), pp. 1–20. New York: Academic Press 1980
2. Berens, H.: Ein Problem über die beste Approximation in Hilberträumen. Functional Analysis and Approximation. ISNM, Vol. 60. Butzer, P.L., Görlich, E., Sz-Nagy, B. (eds.), pp. 247–254. Basel: Birkhäuser 1981
3. Franchetti, C., Papini, P.L.: Approximation properties of sets with bounded complements. Proc. R. Soc. Edinburgh **89**A, 75–86 (1981)
4. Klee, V.: Remarks on nearest points in normed linear spaces. Proceedings of the Colloquium on Convexity, Fenchel, W. (ed.), pp. 168–176. Univ. of Copenhagen, 1965
5. Klee, V.: Dispersed Chebyshev sets and coverings by balls. Math. Ann. **257**, 251–260 (1981)

Received August 23, 1982

Math. Ann. 263, 283–288 (1983)

Mathematische Annalen
© Springer-Verlag 1983

The Schur Lemma for Bounded Multiplier Convergent Series

Charles Swartz
Department of Mathematics, New Mexico State University, Las Cruces, NM 88003, USA

The Schur lemma [2, Sect. IX.2] is one of the best known and most useful results in functional analysis. For example, many of the proofs of the Orlicz-Pettis Theorem, including the original proof of Pettis, use some form of the Schur Lemma (see for example [6]). There have been several generalizations of the Schur Lemma given for series with values in B-spaces, locally convex spaces and topological groups [3, 8, 4]. In this note we present a version of the Schur Lemma for bounded multiplier convergent series in metric linear spaces. Our version of the Schur Lemma is based on an observation that the conclusion of the Schur Lemma for B-space valued series can be somewhat strengthened, and this strengthened conclusion is used in the bounded multiplier version.

We first recall the B-space version of the Schur Lemma and show how the conclusion can be somewhat strengthened. Let X be a B-space (real vector spaces will be considered for simplicity). A series $\sum x_j$ in X is said to be subseries convergent if for each subsequence $\{x_{j_k}\}$ the series $\sum x_{j_k}$ is convergent. If $\sum x_j$ is subseries convergent and σ is an infinite subset of $\mathbb{N}$, the positive integers, then $\sum_{j\in\sigma} x_j$ denotes the sum $\sum_k x_{j_k}$ where the elements of σ are arranged in the subsequence $\sigma=\{j_k : k\in\mathbb{N}\}, j_k<j_{k+1}$. For σ finite the meaning of $\sum_{j\in\sigma} x_{ij}$ is clear. Let $x_{ij}\in X$ for $i, j\in\mathbb{N}$ be such that the series $\sum_j x_{ij}$ is subseries convergent for each i. The Schur Lemma for X-valued series is given by the following statement: if $\lim_i \sum_{j\in\sigma} x_{ij}$ exists for each $\sigma\subseteq\mathbb{N}$ and if $x_j=\lim_i x_{ij}$, then the series $\sum x_j$ is subseries convergent and

$$\lim_i \sum_{j\in\sigma} x_{ij}=\sum_{j\in\sigma} x_j \quad \text{uniformly for} \quad \sigma\subseteq\mathbb{N}. \tag{1}$$

(See [3, Corollary 2].)

We now show that the conclusion (1) can be strengthened. Let $\varepsilon>0$. By (1) there exists N such that $\left\|\sum_{j\in\sigma}(x_{ij}-x_j)\right\|<\varepsilon$ for $i\geqq N$ and $\sigma\subseteq\mathbb{N}$. Thus, for $x'\in X'$ and $\|x'\|\leqq 1$, $\left|\sum_{j\in\sigma}\langle x', x_{ij}-x_j\rangle\right|<\varepsilon$ for $i\geqq N$, $\sigma\subseteq\mathbb{N}$. This implies that

$$\sum_{j=1}^{\infty} |\langle x', x_{ij} - x_j \rangle| \leqq 2\varepsilon \quad \text{for} \quad i \geqq N, \ \|x'\| \leqq 1 \tag{2}$$

[7, Sect. 11.12]. Now for $\{t_j\} \in l^\infty$, the space of bounded sequences equipped with the sup-norm, the series $\sum_j t_j x_{ij}$ is convergent for each i [9, Sect. III.6.5] and if $\|\{t_j\}\| = \sup_j |t_j| \leqq 1$, we have from (2)

$$\left\| \sum_{j=1}^{\infty} t_j (x_{ij} - x_j) \right\| = \sup \left\{ \left| \left\langle x', \sum_{j=1}^{\infty} t_j (x_{ij} - x_j) \right\rangle \right| : \|x'\| \leqq 1 \right\}$$

$$\leqq \sup \left\{ \sum_{j=1}^{\infty} |t_j| \, |\langle x', x_{ij} - x_j \rangle| : \|x'\| \leqq 1 \right\} \leqq 2\varepsilon$$

for $i \geqq N$. That is, (1) can be strengthened to

$$\lim_i \sum_{j=1}^{\infty} t_j x_{ij} = \sum_{j=1}^{\infty} t_j x_j \quad \text{uniformly for} \quad \{t_j\} \in l^\infty, \|\{t_j\}\| \leqq 1. \tag{3}$$

[(3) clearly implies (1) by taking the sequences $\{t_j\}$ with $t_j = 1$ for $j \in \sigma$ and $t_j = 0$ for $j \notin \sigma$.]

If X is a topological vector space, a (formal) series $\sum x_j$ in X is said to be bounded multiplier convergent if the series $\sum t_j x_j$ is convergent for each $\{t_j\} \in l^\infty$. A bounded multiplier convergent series is clearly subseries convergent, and in locally convex spaces the converse is valid [9, Sect. III.6.5]. In certain non-locally convex spaces, there exist series which are subseries convergent but not bounded multiplier convergent [9, Sect. III.6.9]. (Sect. III.6 of [9] contains a detailed discussion of such series in F-spaces.)

We now present a Schur-type lemma for bounded multiplier convergent series. Our version of the Schur Lemma is motivated by the conclusion given in (3) for B-space valued series.

Throughout the remainder of the paper, let X be a metric linear space, i.e., X is a topological vector space whose topology is given by an invariant metric d. (These spaces are called F^* spaces in [9].) We set $|x| = d(x, 0)$, the F-norm of X in the terminology of [9].

In order to establish our main result we require the following matrix theorem. This theorem has been established in [1], but we include it here for the sake of completeness and for the convenience of the reader since [1] is unpublished at this time.

Theorem 1. *Let $x_{ij} \in X$ for $i, j \in \mathbb{N}$ satisfy:*

$$\lim_i x_{ij} = 0 \ \textit{for each} \ j \tag{4}$$

and

for each subsequence $\{m_i\}$ there is a subsequence $\{n_i\}$ of $\{m_i\}$ such that $\lim_i \sum_{j=1}^{\infty} x_{in_j} = 0.$ (5)

Then $\lim_i x_{ii} = 0.$

Proof. We show that $\{x_{ii}\}$ has a subsequence which converges to 0; since the same argument can be applied to any subsequence of $\{x_{ii}\}$, this will establish the result.

Set $k_1=1$. Since $\lim_j x_{ij}=0$ for each i by (5), there exists $k_2>k_1$ such that $|x_{k_1k_2}|<2^{-3}$ and $|x_{k_2k_1}|<2^{-3}$. Proceeding by induction produces a subsequence $\{k_i\}$ such that $|x_{k_ik_j}|<2^{-(i+j)}$ for $i\neq j$. To avoid double subscripts later, assume that $k_i=i$. By (5) there is a subsequence $\{n_i\}$ such that $\lim_i \sum_{j=1}^{\infty} x_{in_j}=0$. Hence,

$$|x_{n_in_i}|\leqq\left|\sum_{j\neq i} x_{n_in_j}\right|+\left|\sum_{j=1}^{\infty} x_{n_in_j}\right|\leqq\sum_{j=1}^{\infty}2^{-(i+j)}+\left|\sum_{j=1}^{\infty} x_{n_in_j}\right|. \tag{6}$$

The first term on the right hand side of (6) goes to zero since it is 2^{-i} and the second term goes to zero by hypothesis. Hence, $\lim_i x_{n_in_i}=0$, and this establishes the result.

We next establish a preliminary lemma. For this lemma we require the following property of convergence in X. If the sequence $\{x_j\}$ in X converges to 0, then $\lim_j |tx_j|=0$ uniformly for $|t|\leqq 1$ [10, Sect. I.2.2].

Lemma 2. *Let $x_{ij}\in X$ for $i,j\in\mathbb{N}$ be such that $\sum_j x_{ij}$ is bounded multiplier convergent for each i. If $\lim_i \sum_{j=1}^{\infty} t_jx_{ij}=0$ for each $\{t_j\}\in l^\infty$, then $\lim_i \sum_{j=1}^{\infty} t_jx_{ij}=0$ uniformly for $\|\{t_j\}\|\leqq 1$.*

Proof. If the conclusion fails to hold, we may assume (by passing to a subsequence if necessary) that there is a $\delta>0$ such that $\sup\left\{\left|\sum_{j=1}^{\infty} t_jx_{ij}\right| : |t_j|\leqq 1\right\}>\delta$ for each i. Set $i_1=1$. Then there exists $\alpha_1=\{t_{1j}\}\in l^\infty$ *such that* $\|\alpha_1\|\leqq 1$ and $\left|\sum_{j=1}^{\infty} t_{1j}x_{i_1j}\right|>\delta$. There exists M_1 such that $\left|\sum_{j=1}^{M_1} t_{1j}x_{i_1j}\right|>\delta$. Since $\lim_i x_{ij}=0$ for each j, by the observation above there exists $i_2>i_1$ such that $i\geqq i_2$ implies $\sum_{j=1}^{M_1}|t_jx_{ij}|<\delta/2$ for $|t_j|\leqq 1$. Now there exists $\alpha_2=\{t_{2j}\}\in l^\infty$ such that $\|\alpha_2\|\leqq 1$ and $\left|\sum_{j=1}^{\infty} t_{2j}x_{i_2j}\right|>\delta$. There exists $M_2>M_1$, such that $\left|\sum_{j=1}^{M_2} t_{2j}x_{i_2j}\right|>\delta$. Note

$$\left|\sum_{j=M_1+1}^{M_2} t_{2j}x_{i_2j}\right|\geqq\left|\sum_{j=1}^{M_2} t_{2j}x_{i_2j}\right|-\sum_{j=1}^{M_1}|t_{2j}x_{i_2j}|>\delta/2.$$

Continuing this construction inductively gives subsequences $\{i_k\}$ and $\{M_k\}$ of positive integers such that

$$\left|\sum_{j=M_{k-1}+1}^{M_k} t_{kj}x_{i_kj}\right|>\delta/2 \quad \text{for all } k, \tag{7}$$

where $M_0=0$.

Now consider the matrix $Z_{jk}=\sum_{i=M_{k-1}+1}^{M_k} t_{ki}x_{i_j i}$. Since $\lim_j x_{ji}=0$ for each i, $\lim_j Z_{jk}=0$ for each k since $|t_{ki}|\leqq 1$. Let $\sigma\subseteq\mathbb{N}$. Define a sequence $\{s_i\}\in l^\infty$ by $s_i=t_{ki}$ if $M_{k-1}+1\leqq i\leqq M_k$ and $k\in\sigma$ and $s_i=0$ otherwise. Then $\sum_{k\in\sigma} Z_{jk}=\sum_{i=1}^{\infty} s_i x_{i_j i}$ converges to 0 as $j\to\infty$ by hypothesis. The matrix $\{Z_{jk}\}$ satisfies the hypothesis of Theorem 1 so that $\lim_k Z_{kk}=0$. But this contradicts (7).

We now present our Schur-type lemma for bounded multiplier convergent series. Note that conclusion (ii) in the theorem below is just condition (3) in the strengthened conclusion of the Schur Lemma for B-spaces.

Theorem 3. *Let $x_{ij}\in X$ for $i,j\in\mathbb{N}$ be such that $\sum_j x_{ij}$ is bounded multiplier convergent for each i. Assume that $\lim_i \sum_{j=1}^{\infty} t_j x_{ij}$ exists for each $\{t_j\}\in l^\infty$. If $\lim_i x_{ij}=x_j$, then*

(i) *the series $\sum x_j$ is bounded multiplier convergent and*

(ii) $\lim_i \sum_{j=1}^{\infty} t_j x_{ij}=\sum_{j=1}^{\infty} t_j x_j$ *uniformly for* $\|\{t_j\}\|\leqq 1$.

Proof. First we claim that the sequence $\left\{\sum_{j=1}^{\infty} t_j x_{ij}\right\}$ satisfies a Cauchy condition uniformly for $\|\{t_j\}\|\leqq 1$. If this is not the case, there exist a subsequence $\{i_k\}$ and a $\delta>0$ such that

$$\sup\left\{\left|\sum_{j=1}^{\infty} t_j(x_{i_{k+1}j}-x_{i_k j})\right| : \|\{t_j\}\|\leqq 1\right\}>\delta. \tag{8}$$

Consider the series $\sum_j (x_{i_{k+1}j}-x_{i_k j})$. This sequence of series satisfies the hypothesis of Lemma 2 so that $\lim_k \sum_{j=1}^{\infty} t_j(x_{i_{k+1}j}-x_{i_k j})=0$ uniformly for $\|\{t_j\}\|\leqq 1$. This contradicts (8) and establishes the claim.

We now establish (i) and also

$$\lim_i \sum_{j=1}^{\infty} t_j x_{ij}=\sum_{j=1}^{\infty} t_j x_j \quad \text{for each} \quad \{t_j\}\in l^\infty. \tag{9}$$

Let $\varepsilon>0$. By what has been established above, there exists an N such that $i,k\geqq N$ implies $\left|\sum_{j\in\sigma} t_j(x_{ij}-x_{kj})\right|<\varepsilon$ for $\sigma\subseteq\mathbb{N}$. Hence, for every M and $i\geqq N$ $\left|\sum_{j=1}^{M} t_j(x_{ij}-x_j)\right|\leqq\varepsilon$. Thus, for $i\geqq N$,

$$\left|\sum_{j=1}^{M} t_j x_j-\sum_{j=1}^{\infty} t_j x_{ij}\right|\leqq\left|\sum_{j=1}^{M} t_j(x_j-x_{ij})\right|+\left|\sum_{j=M+1}^{\infty} t_j(x_{Nj}-x_{ij})\right|$$
$$+\left|\sum_{j=M+1}^{\infty} t_j x_{Nj}\right|<2\varepsilon+\left|\sum_{j=M+1}^{\infty} t_j x_{Nj}\right|. \tag{10}$$

The last term on the right hand side of (10) goes to 0 as $M\to\infty$ for N fixed. Condition (i) and (9) follow from this estimate.

Condition (ii) follows by applying Lemma 2 to the series $\sum_j (x_{ij}-x_j)$.

Theorem 3 has as a corollary a generalization of another result of Schur on summability matrices. Let $[a_{ij}]$ be a real matrix such that the sequence $\left\{\sum_{j=1}^{\infty} a_{ij}t_j\right\}$ is convergent for each $\{t_j\}\in l^\infty$. Then Schur's result asserts that the series $\sum_j |a_{ij}|$ converge uniformly with respect to i [5, Sect. 7.1.6]. This condition (for real series) clearly implies that the series $\sum_j t_j a_{ij}$ are uniformly convergent for $\|\{t_j\}\|\leqq 1$. Using Theorem 3, we can obtain an analogous result for bounded multiplier convergent series in X.

Corollary 4. *Let x_{ij} satisfy the hypothesis of Theorem 3. Then the series $\sum_j t_j x_{ij}$ converge uniformly for $\|\{t_j\}\|\leqq 1$.*

Proof. First note that if the series $\sum_j x_j$ is bounded multiplier convergent, then the series $\sum_j t_j x_j$ converge uniformly for $\|\{t_j\}\|\leqq 1$. (This follows from Theorem 3 but is also easily checked directly.)

Let $\varepsilon>0$. By Theorem 3, there exists N such that $i\geqq N$ implies $\left|\sum_{j=1}^{\infty} t_j(x_{ij}-x_j)\right|<\varepsilon$ for $\|\{t_j\}\|\leqq 1$.

By the observation above, there exists M such that $m\geqq M$ implies $\left|\sum_{j=m}^{\infty} t_j x_{ij}\right|<\varepsilon$, $\left|\sum_{j=m}^{\infty} t_j x_j\right|<\varepsilon$ for $1\leqq i\leqq N$, $\|\{t_j\}\|\leqq 1$. Hence, $m\geqq M$ implies

$$\left|\sum_{j=m}^{\infty} t_j x_{ij}\right| \leqq \left|\sum_{j=m}^{\infty} t_j(x_{ij}-x_j)\right| + \left|\sum_{j=m}^{\infty} t_j x_j\right| < 2\varepsilon$$

for $i\geqq N$, $\|\{t_j\}\|\leqq 1$, and the result follows.

A natural question that arises at this point is whether the hypothesis in Theorem 3 can be replaced by the weaker hypothesis that $\lim_i \sum_{j\in\sigma} x_{ij}$ exists for each $\sigma\subseteqq\mathbb{N}$. This is, of course, the usual hypothesis in the Schur Lemma for subseries convergent series. Using the example of Rolewicz [9, Sect. III.6.9] it is easy to construct an example that shows that this cannot be done.

Let $\sum_j x_j$ be a series which is subseries convergent but not bounded multiplier convergent [9, Sect. III.6.9]. Set $x_{ij}=x_j$ if $1\leqq j\leqq i$ and $x_{ij}=0$ if $j>i$. Then $\sum_j x_{ij}$ is subseries convergent, and for each i the sequence $\{x_{ij}\}_{j=1}^{\infty}$ lies in a finite dimensional subspace so $\sum_j x_{ij}$ is also bounded multiplier convergent. By the subseries convergence of $\sum x_j$, we have $\lim_i \sum_{j\in\sigma} x_{ij} = \sum_{j\in\sigma} x_j$ for each $\sigma\subseteqq\mathbb{N}$. But, since $\sum x_j$ is not bounded multiplier convergent, condition (i) of Theorem 3 fails.

In addition to discussing series which are bounded multiplier convergent, Rolewicz in [9] also discusses series which he calls C-sequences [9, Sect. III.8]. A sequence $\{x_j\}$ is a C-sequence if the series $\sum_j t_j x_j$ is convergent for each sequence $\{t_j\} \in c_0$. It is natural to ask if the analogue of Theorem 3 is valid for series which are C-convergent. (A series $\sum x_j$ is C-convergent if the sequence $\{x_j\}$ is a C-sequence.) We first note that the analogue of condition (i) of Theorem 3 holds for C-convergent series.

Proposition 5. *Let $x_{ij} \in X$ be such that $\sum_j x_{ij}$ is C-convergent for each i. Assume that $\lim_i \sum_{j=1}^{\infty} t_j x_{ij}$ exists for each $\{t_j\} \in c_0$. If $x_j = \lim_i x_{ij}$, then the series $\sum x_j$ is C-convergent.*

Proof. Let $\{t_j\} \in c_0$. Note that for each $\{s_j\} \in l^\infty$, the sequence $\{s_j t_j\} \in c_0$. Therefore, Theorem 3 can be applied to the series $\sum_j t_j x_{ij}$, and condition (i) implies that the series $\sum t_j x_j$ is convergent.

Whereas the analogue of Theorem 3 (i) does hold for C-convergent series, it is easy to see that the analogue of Theorem 3 (ii) does not hold for such series. For an example, let e_k be the unit vector in c_0 which has a 1 in the k^{th} coordinate and 0 elsewhere. Set $e_{ij} = e_j$ for $1 \leqq j \leqq i$ and $e_{ij} = 0$ for $j > i$. Then each series $\sum_j e_{ij}$ is C-convergent. For $\{t_j\} \in c_0$, $\lim_i \sum_{j=1}^{\infty} t_j e_{ij} = \lim_i \sum_{j=1}^{i} t_j e_j = \sum_{j=1}^{\infty} t_j e_j$ but the convergence is not uniform for $\|\{t_j\}\| \leqq 1$.

The author would like to acknowledge discussions with Professor Piotr Antosik which were of considerable assistance in obtaining the results of this paper.

References

1. Antosik, P., Swartz, C.: Matrix Methods in Analysis (preprint)
2. Banach, S.: Theorie des operations linéaires. Warsaw 1932
3. Brooks, J.K.: Sur les suites uniformément convergentes dans un espace de Banach. C.R. Acad. Sci. (Paris) **274**, A 1037–A 1040 (1972)
4. Constantinescu, C.: Spaces of multipliable families in Hausdorff topological groups. In: Lecture Notes in Mathematics, Vol. 794. Berlin, Heidelberg, New York: Springer 1979
5. Maddox, I.: Elements of functional analysis. Cambridge: Cambridge University Press
6. Pettis, B.J.: On integration in vector spaces. Trans. Am. Math. Soc. **44**, 277–304 (1938)
7. Pietsch, A.: Nukleare lokalkonvexe Räume. Berlin 1965
8. Robertson, A.: Unconditional convergence and the Vitali-Hahn-Saks theorem. Bull. Soc. Math. France, Suppl., Mem. **31–32**, 335–341 (1972)
9. Rolewicz, S.: Metric linear spaces. Warsaw: Polish Scientific Publishers 1972
10. Yosida, K.: Functional analysis. Berlin, Heidelberg, New York: Springer 1966

Received August 26, 1982

Math. Ann. 263, 289–301 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Local Boundary Regularity of the Canonical Einstein-Kähler Metric on Pseudoconvex Domains

John S. Bland

Institute for Advanced Study, School of Mathematics, Princeton, NJ 08540, USA

Introduction

In [3], S. Y. Cheng and S. T. Yau established the existence and uniqueness of a complete Einstein-Kähler metric on bounded pseudoconvex domains in $\mathbb{C}^n$ with C^2 boundary; N. Mok and S. T. Yau have extended this result to arbitrary bounded pseudoconvex domains in $\mathbb{C}^n$ in [12]. In case the domain Ω was C^{3n+7} bounded, strictly pseudoconvex, Cheng and Yau [3] also showed that $\exp(-g)\in C^{n+3/2-\varepsilon}(\bar{\Omega})$, $\varepsilon>0$, where $\dfrac{\partial^2}{\partial z_i \partial z_{\bar{j}}} g dz_i \otimes dz_{\bar{j}}$ defines the Einstein-Kähler metric and $\exp((n+1)g)$ is the volume. More recently, J. Lee and R. Melrose [10] established the exact form of the singularity for $\exp(-g)$ at the boundary of strictly pseudoconvex bounded domains in $\mathbb{C}^n$. The main purpose of this paper is to give somewhat weaker regularity results for $-\exp(-g)$ at strictly pseudoconvex boundary points of arbitrary pseudoconvex domains in $\mathbb{C}^n$ possessing a complete Einstein-Kähler metric.

Throughout the paper, operations with respect to a specific metric will be subscripted by the name of the metric (e.g. Δ_g will refer to the Laplacian with respect to the metric $[g_{i\bar{j}}]$). The subscript E refers to the extrinsic Euclidean metric. Also, the following notation will be used: ϱ_g = distance in the metric $[g_{i\bar{j}}]$,

$$[\varphi_{i\bar{j}}] = \left[\frac{\partial^2 \varphi}{\partial z_i \partial z_{\bar{j}}}\right], \quad [\varphi^{\bar{j}k}] = [\varphi_{i\bar{j}}]^{-1}, \quad \varphi^k = \sum_j \varphi_{\bar{j}} \varphi^{\bar{j}k},$$

$$|d\varphi|^2 = \sum_{i,j} \varphi_i \varphi_{\bar{j}} \varphi^{i\bar{j}}, \quad |d\varphi|_g^2 = \sum_{i,j} \varphi_i \varphi_{\bar{j}} g^{i\bar{j}}, \quad \langle d\varphi, df \rangle_g = \sum_{i,j} \varphi_i f_{\bar{j}} g^{i\bar{j}}.$$

The results in this paper were contained in the author's doctoral thesis at the University of California, Los Angeles. The author would like to express his sincere appreciation to S. Y. Cheng and R. E. Greene for their considerable help and constant encouragement while he was engaged in this work.

1. Localized Maximum Principles and Schwarz Lemmas

This section will develop two basic tools needed later in the study of the complete Einstein-Kähler metric near the boundary of pseudoconvex domains. The first lemma will be a localized maximum principle. This lemma allows localization of many of Yau's recent results about harmonic functions on complete Riemannian manifolds [14]. The localized maximum principle will then be used to prove a local version of the Schwarz lemma – that under certain curvature conditions, holomorphic maps are metric decreasing for appropriate metrics.

The first lemma roughly states that under certain conditions on the Ricci curvature, a bounded, non-negative function u satisfying $\Delta u \geqq ku$ in a neighborhood of the boundary must vanish to a certain order near the boundary. The idea is as follows. The function $u(z)$ may be compared to a radially symmetric (using the Kähler distance) function $v_{z_i}(r)$ about z_i by using the Laplacian Comparison Theorem of Greene-Wu [7]. Upon doing so, one determines that the maximum value that u attains on spheres of Kähler radius r about z_i must grow as $v_{z_i}(r)$ (exponentially). This fact, together with an upper bound on u, yields a strong upper bound on the value of u at the center of the ball, $u(z_i)$. In particular, if z_i is a sequence of points, and $r(z_i)$ is the radius of the largest ball which is centered at z_i and contained in the region where $\Delta u \geqq ku$, then $r(z_i) \to \infty$ implies that $u(z_i) \to 0$. The function φ in the statement of the lemma measures $r(z_i)$. For completeness, the statement of Greene-Wu's Laplacian Comparison Theorem [Proposition 2.15, 7] precedes the lemma.

Laplacian Comparison Theorem. *Let M be the unit ball in $\mathbb{C}^n$ with constant negative holomorphic sectional curvature, and let N be a complete Riemannian manifold of the same dimension. Suppose that for $o \in M$ and $p \in N$, the Ricci curvature satisfies* $\mathrm{Ric}_N\left(\frac{\partial}{\partial \varrho_N}, \frac{\partial}{\partial \varrho_N}\right)(y) \geqq \mathrm{Ric}_M\left(\frac{\partial}{\partial \varrho_M}, \frac{\partial}{\partial \varrho_M}\right)(x)$ *whenever* $\varrho_N(y, p) = \varrho_M(x, o)$. *Let f be a C^2 nondecreasing function on $[0, \infty)$. Then for all such x and y with x inside the cut locus of p, $\Delta_N f(\varrho(y)) \leqq \Delta_M f(\varrho(x))$.*

Remark. It is well known that the one sided comparison (i.e. when M is the unit ball) also holds in an appropriate weak sense outside the cut locus of p [2]. This weak sense is sufficient for the following applications.

Lemma I. *Suppose p is a boundary point of the pseudoconvex domain Ω, and D is a neighborhood of p. Suppose $g_{i\bar{j}}$ is a Kähler metric on $D \cap \Omega$, with Ricci curvature $\geqq -c$. Define $\varphi(z) = \exp[-\varrho_g(z, \partial D \cap \Omega)]$.*

Suppose $u \in C^2(D \cap \Omega)$, $u \leqq M$, $\Delta u \geqq ku$ for $k > 0$. Then $u(z) \leqq 2M|\varphi(z)|^\alpha$ where

$$\alpha = \begin{cases} \sqrt{k/n} & k \geqq c^2 n \\ \frac{1}{2}\left(\sqrt{c^2(n-1)^2 + 4k} - c(n-1)\right) & k < c^2 n. \end{cases}$$

Proof. Fix $z_0 \in D \cap \Omega$. Define $r(z) = \varrho_g(z, z_0)$. Let B be the standard model of constant negative curvature, with Ricci curvature $= -c$. By the Laplacian Comparison Theorem:

$$\Delta_g \cosh \alpha r(z) \leqq \Delta_B \cosh \alpha r(z)$$

$$= \alpha^2 \cosh \alpha r + \alpha c(n-1) \frac{\cosh cr}{\sinh cr} \frac{\sinh \alpha r}{\cosh \alpha r} \cosh \alpha r$$

$$\leqq \begin{cases} \left(\alpha^2 + \alpha c(n-1)\frac{\alpha}{c}\right)\cosh \alpha r & \alpha \geqq c \\ (\alpha^2 + \alpha c(n-1))\cosh \alpha r & \alpha < c \end{cases}$$

$$= \begin{cases} n\alpha^2 \cosh \alpha r & \alpha \geqq c \\ (\alpha^2 + \alpha c(n-1))\cosh \alpha r & \alpha < c. \end{cases}$$

Set

$$\alpha = \begin{cases} \sqrt{k/n} & k \geqq c^2 n \\ \frac{1}{2}(\sqrt{c^2(n-1)^2 + 4k} - c(n-1)) & k < c^2 n \end{cases}$$

and for arbitrary $\varepsilon > 0$, consider the difference $(u(z) - u(z_0)\cosh((\alpha - \varepsilon)r(z)))$, defined for $r(z) \leqq -\ln|\varphi(z_0)|$.

$$(u(z) - u(z_0)\cosh((\alpha-\varepsilon)r(z)))|_{z=z_0} = 0$$

$$\Delta(u(z) - u(z_0)\cosh((\alpha-\varepsilon)r(z))) > k(u(z) - u(z_0)\cosh((\alpha-\varepsilon)r(z)))$$

so the difference must be non-negative at some point for which $r(z) = -\ln|\varphi(z_0)|$. Since $u(z) \leqq M$

$$M \geqq \max_{r(z) = -\ln|\varphi(z_0)|} u(z)$$

$$\geqq u(z_0)\cosh((\alpha-\varepsilon)(-\ln|\varphi(z_0)|))$$

$$> \tfrac{1}{2}u(z_0)|\varphi(z_0)|^{\varepsilon - \alpha} \quad \text{with } \varepsilon \text{ arbitrary.}$$

Hence $u(z_0) < 2M|\varphi(z_0)|^{\alpha}$. □

In practice, the interesting cases are when $\varphi(z) \to 0$ as $z \to p$, referred to as locally complete sectors.

The proof of the next lemma uses the Laplacian Comparison Theorem of [7] and comparison functions from Lemma I to derive a much stronger statement in the Generalized Maximum Principle of Cheng and Yau [3].

Lemma II. *Suppose Ω is an open domain in $\mathbb{C}^n$; let $g_{i\bar{j}}$ be a Kähler metric on Ω with Ricci curvature $\geqq -c$. Suppose D is a subdomain of Ω with a set of points $z_i \in D$ satisfying $\varrho_g(z_i, \partial D) \geqq R$ for some fixed $R > 0$. Suppose further that $u \in C^{2+\alpha}(\bar{D})$, $u \geqq 0$ on D, and $u(z_i) \to 0$. Then there is a sequence $\tilde{z}_i \in D$ such that*

(1) $u(\tilde{z}_i) \leqq u(z_i)$

(2) $|du(\tilde{z}_i)|_g^2 \leqq u^2(z_i)\left(\dfrac{c^2 \sinh^2 cR}{4\sinh^4(cR/2)}\right)$

(3) $\Delta u(\tilde{z}_i) \geqq -u(z_i)\left(\dfrac{nc^2 \cosh cR}{\cosh cR - 1}\right)$.

Proof. Fix z_i. Define $r(z) = \varrho_g(z, z_i)$. Define

$$v(z) = \frac{\cosh cR - \cosh(cr(z))}{\cosh cR - 1}.$$

Then

$$v(z_i)=1$$

$$v(z)=0 \quad \text{for } z \text{ satisfying} \quad r(z)=R$$

$$dv(z)=\frac{-c\sinh cr(z)}{\cosh cR-1}$$

$$\Delta_g v(z)\geqq \Delta_B\left(\frac{\cosh cR-\cosh cr(z)}{\cosh cR-1}\right)=\frac{nc^2\cosh cr(z)}{1-\cosh cR},$$

where B is the standard model of constant negative sectional curvature with Ricci curvature $=-c$. Consider the function $\frac{u(z)}{v(z)}$. It is non-negative, and $u/v(z)\to\infty$ as $r(z)\to R$. Therefore, it has a minimum, $\tilde{z}_i\in D$. At $\tilde{z}_i$:

(1) $u/v(\tilde{z}_i)\leqq u/v(z_i)=u(z_i)$

$$\therefore u(\tilde{z}_i)\leqq u(z_i)$$

(2) $d(u/v)(\tilde{z}_i)=0$

$$\therefore du(\tilde{z}_i)=u/v(\tilde{z}_i)dv(\tilde{z}_i)$$

$$|du(\tilde{z}_i)|_g^2\leqq u^2(z_i)\frac{c^2\sinh^2 cR}{4\sinh^4(cR/2)}$$

(3) $\Delta_g(u/v)(\tilde{z}_i)=\frac{\Delta_g u(\tilde{z}_i)}{v(\tilde{z}_i)}-\frac{u(\tilde{z}_i)}{v(\tilde{z}_i)}\frac{\Delta_g v(\tilde{z}_i)}{v(\tilde{z}_i)}\geqq 0$

$$\therefore \Delta_g u(\tilde{z}_i)\geqq \frac{u(\tilde{z}_i)}{v(\tilde{z}_i)}\Delta_g v(\tilde{z}_i)\geqq -u(z_i)\frac{nc^2\cosh cr(z)}{\cosh cR-1}$$

$$\geqq -u(z_i)\frac{nc^2\cosh cR}{\cosh cR-1}. \quad \square$$

The next two lemmas will apply these results to deduce local versions of a Schwarz lemma. They state that under certain curvature assumptions on the domain and image manifold, holomorphic maps are metric (volume) decreasing. Yau [15] proved global versions of these results.

Lemma III. *Let $(M, g_{i\bar{j}})$ be a Kähler manifold with Ricci curvature $\geqq -K_M$; $(N, h_{i\bar{j}})$ a hermitian manifold with holomorphic bisectional curvature $\leqq -K_N$. Let $p\in D\cap\partial M$ for an open set D. Define $\varphi(z)=\exp[-\varrho_g(z,\partial(D\cap M))]$, and suppose $\varphi(z)\to 0$ as $z\to p$. (Hence, g is "locally complete" at p.)*

Suppose $f:(M\cap D)\to N$ is holomorphic. Then on any $D_\varepsilon=\{z|\varphi(z)\leqq 1-\varepsilon\}$, $\varepsilon>0$,

$$(f^*[h_{i\bar{j}}])g^{i\bar{j}}\leqq\frac{nK_M}{(n+1)K_N}+O(|\varphi|^\alpha)$$

and

$$f^*[h_{i\bar{j}}]\leqq[g_{i\bar{j}}]\frac{nK_M}{(n+1)K_N}+O(|\varphi|^\alpha)$$

for some $\alpha>0$.

Proof. Let $R_{i\bar{j}}$ = Ricci curvature of $g_{i\bar{j}}$

$$S_{\alpha\bar{\beta}\bar{\gamma}\eta} = \text{holomorphic bisectional curvature of } h_{i\bar{j}}.$$

Choose co-frames θ^i, ω^i satisfying

$$g_{i\bar{j}}dz^i dz^{\bar{j}} = \theta^i\overline{\theta^i} \quad \text{and} \quad h_{i\bar{j}}dw^i dw^{\bar{j}} = \omega^i\overline{\omega^i}.$$

Define $a_{i\alpha}$, u respectively by:

$$f^*\omega^i = a_{i\alpha}\theta^\alpha, \qquad u = \sum_{i,\alpha} |a_{i\alpha}|^2.$$

By [11], $\frac{1}{2}\Delta_g u = 2\frac{|du|_g^2}{u} + \sum_{i,j,\alpha} a_{\alpha i}\overline{a_{\alpha j}}R_{j\bar{i}} - \sum_{\substack{\alpha,\beta,\gamma,\eta \\ i,k}} \overline{a_{\alpha i}}a_{\beta i}a_{\gamma k}\overline{a_{\eta k}}S_{\alpha\bar{\beta}\bar{\gamma}\eta}$

$$\begin{aligned}
\Delta_g(u+c)^{-1/2} &= -\frac{1}{2}\frac{\Delta_g u}{(u+c)^{3/2}} + \frac{3}{4}\frac{|du|_g^2}{(u+c)^{5/2}} \\
&= -\frac{a_{\alpha i}\overline{a_{\alpha j}}R_{j\bar{i}} - \overline{a_{\alpha i}}a_{\beta i}a_{\gamma k}\overline{a_{\eta k}}S_{\alpha\bar{\beta}\bar{\gamma}\eta}}{(u+c)^{3/2}} \\
&\quad + \frac{3}{4}\frac{|du|_g^2}{(u+c)^{5/2}} - 2\frac{|du|_g^2}{u(u+c)^{3/2}} \\
&\leqq \frac{K_M u - \frac{1}{n}(n+1)K_N u^2}{(u+c)^{3/2}}.
\end{aligned}$$

Suppose that $z_i \in D_\varepsilon$, $u(z_i) \to \infty$ [$z_i \in D_\varepsilon$ means that $\varrho_g(z_i, \partial(D \cap M))$ is bounded strictly positive]. Then for $c > 0$, $(u+c)^{-1/2}(z_i) \to 0$, and $\Delta(u+c)^{-1/2}(z_i) \to -\infty$. This contradicts Lemma II. Hence, u is bounded on D_ε. Lemma I now applies.

$$\frac{1}{2}\Delta_g u \geqq -K_M u + \frac{(n+1)}{n}K_N u^2$$

$$\therefore \frac{1}{2}\Delta_g u \geqq K_M\left(u - \frac{nK_M}{(n+1)K_N}\right)$$

or

$$\frac{1}{2}\Delta_g\left(u - \frac{nK_M}{(n+1)K_N}\right) \geqq K_M\left(u - \frac{nK_M}{(n+1)K_N}\right).$$

By Lemma I, $u \leqq \frac{nK_M}{(n+1)K_N} + O(|\varphi|^\alpha)$ for some $\alpha > 0$. By the definition of u

$$f^*[h_{i\bar{j}}] \leqq [g_{i\bar{j}}]\left(\frac{nK_M}{(n+1)K_N} + O(|\varphi|^\alpha)\right) \quad \text{on } D_\varepsilon. \quad \square$$

Notice that the Localized Schwarz Lemma does not actually say that holomorphic maps are metric decreasing, but only gives an upper bound on their expansion factor. This is because the trace of the pull back metric is the quantity being analyzed. However, if the map is a biholomorphic map between complete

manifolds with the appropriate curvature assumptions, the next lemma will combine with the last to give the actual metric decreasing property.

Lemma IV. *Let $(M, g_{i\bar{j}})$ be a Kähler manifold with Ricci curvature $\geqq -c$ and with scalar curvature $\geqq -K_M$; $(N, h_{i\bar{j}})$ be a Hermitian manifold with Ricci curvature $\leqq -K_N$. Let $p \in D \cap \partial M$ for an open set D, and define $\varphi(z) = \exp[-\varrho_g(z, \partial(D \cap M))]$, $D_\varepsilon = \{z | \varphi(z) \leqq 1-\varepsilon\}$, $\varepsilon > 0$. Suppose $\varphi(z) \to 0$ as $z \to p$. (Hence, g is "locally complete" at p, and D_ε is an open neighborhood of p.)*

Let $f: (M \cap D) \to N$ be holomorphic. Then

$$f^* dV_N \leqq dV_M \left(\left(\frac{K_M}{nK_N} \right)^n + O(|\varphi|^\alpha) \right) \quad \text{for some} \quad \alpha > 0$$

where dV_M, dV_N are the volume elements for $(M, g_{i\bar{j}})$, $(N, h_{i\bar{j}})$ respectively.

Proof. With the notation of the above lemma, define $v = |\det(a_{i\alpha})|^2$. By [11],

$$\frac{1}{2} \Delta_g v = 2 \frac{|dv|_g^2}{v} + v \left(R - \sum_{\alpha, k, \beta} a_{\alpha k} \bar{a}_{\beta k} S_{\bar{\alpha}\beta} \right),$$

where R is the scalar curvature of M, and $S_{\bar{\alpha}\beta}$ is the Ricci curvature of N. The result follows from a direct imitation of the above proof. □

The assumptions made in the lemmas are quite reasonable. Let M be a domain in $\mathbb{C}^n$ with a smooth, strictly pseudoconvex boundary point p. Choose coordinates so that near p, ∂M is smooth and strictly convex. Define a function d on M by $d(x) =$ the Euclidean distance from x to ∂M. Near ∂M, $(-d)$ is strictly convex. Let D be any bounded neighborhood of p for which $(-d)$ is strictly convex on $D \cap M$. Then $g_{i\bar{j}} = (-\log d)_{i\bar{j}}$ defines a metric on $D \cap M$. For this metric, there is a constant $k > 0$ such that the ratio metric length/Euclidean length $\geqq k/\sqrt{d(x)}$ for any tangent vector to M at x; if the tangent vector is parallel to the gradient of d, then the ratio metric length/Euclidean length $\geqq k/d$. Let $D_1 \subset \bar{D}_1 \subset D$ be any smaller neighborhood of p. By comparing measurements in the metric $g_{i\bar{j}}$ to measurements on the unit ball with its standard metric, it is clear that

$$\varrho_g(\{x \in D_1 \cap M | d(x) \leqq \varepsilon\}, \partial D \cap M) \to \infty$$

as $\varepsilon \to 0$. Finally, for any smooth defining function φ on M, $(-\log -\varphi)_{i\bar{j}}$ defines a metric near $\partial M \cap D$ which is uniformly equivalent to g.

2. Local Boundary Regularity

This section establishes some regularity results for the potential for the complete Einstein-Kähler metric which rely entirely upon local data. Let $(-\varphi)^{-(n+1)}$ be the volume for the Einstein-Kähler metric, in which case $\dfrac{\partial^2(-\log -\varphi)}{\partial z_i z_{\bar{j}}} dz_i \otimes dz_{\bar{j}}$ is the metric itself. In [3], Cheng and Yau showed that if the domain is strictly pseudoconvex, then φ is $C^{n+3/2-\varepsilon}$, $\varepsilon > 0$, up to the boundary. This section will show that at strictly pseudoconvex boundary points of arbitrary domains, φ is $C^{n/2+\varepsilon}$ up to the boundary for some small $\varepsilon > 0$; in the case of certain "nice" strictly

pseudoconvex boundary points (defined later), φ is $C^{n+1-\varepsilon}$ up to the boundary. The value of this local result is that it is now possible to study the metric near strictly pseudoconvex points of arbitrary bounded pseudoconvex domains in $\mathbb{C}^n$.

The regularity results will be proved by comparing the Einstein-Kähler metric to an asymptotically Ricci constant Kähler metric. In [6], C. Fefferman describes an explicit method for obtaining such a metric on a strictly pseudoconvex domain Ω. He defines the operator J on $C^{2+\alpha}$ negative functions u by

$$J(u) = -\det\begin{pmatrix} u & u_{\bar{j}} \\ u_i & u_{i\bar{j}} \end{pmatrix}.$$

Then $J(u) = (-u)^{n+1}\det((-\log -u)_{i\bar{j}})$, and if $(-\log -u)_{i\bar{j}}$ defines a metric, $(\log J(u))_{i\bar{j}}$ is the difference between the Ricci curvature and $-(n+1)(-\log -u)_{i\bar{j}}$. If $u=0$ on $\partial\Omega$, then the metric is complete. In particular, if u is a defining function for Ω (that is, $u=0$ on $\partial\Omega$, $|\nabla u| \neq 0$ on $\partial\Omega$, and $u<0$ on Ω) and $J(u)$ is asymptotically equal to 1, then the metric is asymptotically Ricci constant. If $J(u) \equiv 1$, the metric is Einstein-Kähler.

Fefferman's method is to start with a defining function, and iteratively modify it into a better approximate solution to $J(u)=1$. For a C^k domain $\Omega \subset\subset \mathbb{C}^n$, and for any $l \leqq \min\left\{n+1, \left[\frac{k-2}{2}\right]\right\}$, he produced a $C^{k-2l}(\bar{\Omega})$ function φ_l such that $J(\varphi_l) = 1 + \eta u^l$, where u is a defining function for Ω, and $\eta \in C^{k-2l-2}(\bar{\Omega})$. Throughout the remainder of this section φ_l will be referred to as the l^{th} order approximate solution (metric).

The first result will give some a priori estimates on the Einstein-Kähler metric at strictly pseudoconvex points. These will be obtained by comparing it to the extrinsic Euclidean metric on $\mathbb{C}^n$ for $\Omega \subset \mathbb{C}^n$, and using the localized Schwarz Lemma. Then, by localizing some of the arguments used by Cheng-Yau [3], regularity for the potential for the Einstein-Kähler metric will be established.

Proposition. *Suppose Ω is a domain in $\mathbb{C}^n$ with a complete Einstein-Kähler metric $g_{i\bar{j}}$ (normalized to have Ricci curvature $\equiv -1$). Let p be a strictly pseudoconvex boundary point. Then for every $M>0$, there is a neighborhood N_M of p such that $[g_{i\bar{j}}] > M[\delta_{ij}]$ on $N_M \cap \Omega$.*

Proof. Choose a coframe θ^i, and functions $a_{\alpha i}$ such that

$$\sum_i \theta^i \theta^{\bar{i}} = \sum_{i,j} g_{i\bar{j}} dz^i dz^{\bar{j}}$$

$$\sum_{i,\alpha,\beta} a_{i\alpha}\theta^\alpha \overline{a_{i\beta}}\theta^{\bar{\beta}} = \sum_{i,j} \delta_{ij} dz^i dz^{\bar{j}}.$$

Define $u = \sum_{i,\alpha} |a_{i\alpha}|^2$. By [11]

$$\frac{1}{2}\Delta_g u = 2\frac{|du|_g^2}{u} + \sum_{i,\alpha,j} a_{\alpha i}\overline{a_{\alpha j}} R_{j\bar{i}} - \sum_{\alpha,\beta,\gamma,\delta}\sum_i \overline{a_{\alpha i}} a_{\beta i} a_{\gamma i}\overline{a_{\delta i}} S_{\bar{\alpha}\beta\gamma\bar{\delta}}$$

where $R_{j\bar{i}} = (\text{Ricci curvature of } g_{i\bar{j}}) = -g_{j\bar{i}}$

$$S_{\bar{\alpha}\beta\gamma\bar{\delta}} = \text{holomorphic bisectional curvature of } \mathbb{C}^n \equiv 0$$

so

$$\frac{1}{2}\Delta_g u = 2\frac{|du|_g^2}{u} - u$$

or

$$\Delta_g(u^{-3}) = 6u^{-3}.$$

Set $f(z) = u^{-3}(z) \geqq 0$, and let $\lambda(z)$ be the smallest eigenvalue of $g_{i\bar{j}}$ at z. Then $\frac{\lambda^3}{n^3} \leqq f \leqq \lambda^3$, and the statement of the proposition is that $f(z) \to \infty$ as $z \to p$. Suppose not. Then $0 < f(z_i) + 1$ is bounded on some sequence $z_i \to p$. Define a C^3 function φ on $\bar{\Omega}$ such that $\varphi(p) = 0$, $\varphi(\bar{\Omega} \backslash p) > 0$, $\nabla\varphi \neq 0$ near p, $[\varphi_{i\bar{j}}] < 0$. [For instance, $\varphi = -$ defining function $+ \varrho_E(z, p)$.] Then $f + 1$ being bounded on $z_i \to p$ implies that $\varphi \cdot (f+1) \to 0$ as $z_i \to p$, a unique minimum. By Lemma II, there is a sequence $\tilde{z}_i \to p$ such that

$$\varphi \cdot (f+1)(\tilde{z}_i) \leqq \varphi \cdot (f+1)(z_i)$$

$$|d\varphi \cdot (f+1)|_g(\tilde{z}_i) \leqq \varphi \cdot (f+1)(z_i)$$

$$\Delta_g \varphi \cdot (f+1)(\tilde{z}_i) \geqq -2n\varphi \cdot (f+1)(z_i).$$

Thus

$$(f+1)|d\varphi|_g(z_i) \leqq \varphi |df|_g(z_i) + \varphi \cdot (f+1)(z_i)$$

$$|\langle d\varphi, df\rangle_g(\tilde{z}_i)| \leqq |d\varphi|_g |df|_g(\tilde{z}_i) \leqq |df|_g(\tilde{z}_i)\left(\frac{\varphi |df|_g(\tilde{z}_i) + \varphi \cdot (f+1)(z_i)}{(f+1)(\tilde{z}_i)}\right).$$

By [14], our known bounds on the Ricci curvature of $(\Omega, g_{i\bar{j}})$ and $\Delta_g f$ give bounds on $|df|_g$. So $|df|_g(\tilde{z}_i) \leqq k f(\tilde{z}_i)$ for some constant k. Thus

$$|\langle d\varphi, df\rangle_g(\tilde{z}_i)| \to 0.$$

Now

$$\Delta_g \varphi (f+1)(\tilde{z}_i) = (\Delta_g \varphi)(f+1)(\tilde{z}_i) + 2Re\langle d\varphi, df\rangle_g(\tilde{z}_i) + \varphi \Delta_g f(\tilde{z}_i)$$

$$\to (\Delta_g \varphi)(f+1)(\tilde{z}_i) < -\varepsilon < 0$$

since the largest eigenvalue of $g^{i\bar{j}} = (\lambda(\tilde{z}_i))^{-1}$ and

$$(\Delta_g \varphi)(f+1)(\tilde{z}_i) \sim -(\lambda(\tilde{z}_i))^{-1}(\lambda(\tilde{z}_i)^3 + 1).$$

This contradicts the maximum principle, which states that

$$\Delta_g \varphi (f+1)(\tilde{z}_i) \geqq -2n\varphi (f+1)(\tilde{z}_i).$$

Therefore, f cannot be bounded on any sequence $z_i \to p$. □

The proposition states that at strictly pseudoconvex points, the metric $g_{i\bar{j}}$ is locally complete, and grows large in all directions.

Let $[w_{i\bar{j}}]$ represent the metric defined by Fefferman's approximate solution. By Klembeck [8], $[w_{i\bar{j}}]$ has holomorphic sectional curvature arbitrarily close to -1 for points arbitrarily close to p. By the above proposition, $[g_{i\bar{j}}]$ becomes uniformly large at points near p. The arguments of the localized Schwarz Lemma apply, and $g_{i\bar{j}}$ may be compared to $w_{i\bar{j}}$.

Proposition. *If Ω is a domain in $\mathbb{C}^n$ with complete Einstein-Kähler metric $g_{i\bar{j}}$, and p is a strictly pseudoconvex point of $\partial\Omega$, then there is some neighborhood of p on which*

$$[g_{i\bar{j}}]=[w_{i\bar{j}}](1+O(|\varphi|^{\alpha})) \quad \textit{for some} \quad \alpha>0,$$

where $[w_{i\bar{j}}]$ is as defined above, and $\varphi=-e^{-w}$.

Proof. By Lemma III, for every $\varepsilon>0$,

$$\sum_{i,\bar{j}} w_{i\bar{j}}g^{i\bar{j}}\leqq n+\varepsilon+O(|\varphi|^{\alpha}) \quad \text{for some} \quad \alpha>0.$$

By Lemma IV, $\det[g_{i\bar{j}}]\leqq\det[w_{i\bar{j}}](1+\varepsilon+O(|\varphi|^{\alpha}))$ for some $\alpha>0$. Combining these statements, and noticing that ε may be allowed to decrease as $O(|\varphi|)$ (by Klembeck [8]) gives the result. □

It is now time to localize some regularity results of Cheng and Yau [3]. For a bounded, strictly pseudoconvex domain in $\mathbb{C}^n$, they started with Fefferman's approximate metric $w_{i\bar{j}}=(-\log-\varphi)_{i\bar{j}}$, set $F=\log(J(\varphi))$, and solved the partial differential equation:

$$* \begin{cases} \det(w_{i\bar{j}}+u_{i\bar{j}})=e^{(n+1)u}e^{-F}\det(w_{i\bar{j}}) \\ \dfrac{1}{c}[w_{i\bar{j}}]\leqq[w_{i\bar{j}}+u_{i\bar{j}}]\leqq c[w_{i\bar{j}}] \quad \text{for some} \quad c>0. \end{cases}$$

The complete Einstein-Kähler metric was then given by $g_{i\bar{j}}=w_{i\bar{j}}+u_{i\bar{j}}$, and regularity was established by showing that u was appropriately small.

Let $\Omega\subset\mathbb{C}^n$ be a domain with complete Einstein-Kähler metric $[g_{i\bar{j}}]$, and let D be an open set in $\mathbb{C}^n$ such that $\bar{D}\cap\partial\Omega$ is C^k strictly pseudoconvex. When φ is Fefferman's l^{th} order approximate solution on $\bar{D}\cap\Omega$, $l\leqq\min\left\{n+1,\left[\dfrac{k-4}{2}\right]\right\}$, then the above situation applies, and $\varphi\in C^{k-2l}(\bar{D}\cap\Omega)$. By previous estimates, $[u_{i\bar{j}}]=O(|\varphi|^{\alpha})[w_{i\bar{j}}]$ for some $\alpha>0$, and hence u, $\Delta_g u$, $\Delta_w u=O(|\varphi|^{\alpha})$. Since $u=g-w$, $|dg|_g\leqq 1$, $|dw|_w\leqq 1$, and the metrics $g_{i\bar{j}}$, $w_{i\bar{j}}$ are uniformly equivalent, then $|du|_g$ and $|du|_w$ are bounded. Finally, since $e^F=J(\varphi)=1+\eta\varphi^l$ for some $\eta\in C^{k-2l-2}(\bar{D}\cap\Omega)$, then $F=O(|\varphi|^l)$, $|\nabla F|_g=O(|\varphi|^l)$, and $\Delta_g F=O(|\varphi|^l)$. These estimates on u and F will be utilized in the following propositions.

Proposition. *Suppose that in $*$, $F\in C^{3+\varepsilon}(D\cap\bar{\Omega})$, $F=O(|\varphi|)$, $\Delta_w F=O(|\varphi|)$, $|dF|_g^2=O(|\varphi|^l)$. Then $|du|_w^2=O(|\varphi|^{\gamma})$ on every $D'\subset\overline{D'}\subset D$, where $\gamma=\min\left\{l,n+\dfrac{n-1}{n(n+1)}\right\}$.*

Proof. Using Eq. (6.16) of [3] for $\Delta_g(|du|_w^2)$, where $R_{i\bar{j}k\bar{l}}$, $u_{i,k}$ are the curvature and covariant derivatives with respect to the metric $w_{i\bar{j}}$,

$$\Delta_g\left(\frac{|du|_w^2}{(-\varphi)^\beta}\right)=(-\varphi)^{-\beta}(2(n+1)|du|_w^2-2Rew^{i\bar{j}}u_iF_{\bar{j}}+g^{k\bar{l}}w^{i\bar{p}}w^{\bar{j}q}R_{k\bar{l}i\bar{j}}u_{\bar{p}}u_q$$
$$+w^{i\bar{j}}g^{k\bar{l}}u_{k,i}u_{\bar{l},\bar{j}}+w^{i\bar{j}}g^{k\bar{l}}u_{k,j}u_{\bar{l},\bar{i}})+2\beta(-\varphi)^{-\beta-1}Re(g^{i\bar{j}}(|du|_w^2)_i\varphi_{\bar{j}})$$
$$+|du|_w^2(\beta(-\varphi)^{-\beta-1}\Delta_g\varphi+\beta(\beta+1)(-\varphi)^{-\beta-2}|d\varphi|_g^2)$$
$$\geqq(-\varphi)^{-\beta}\left(2(n+1)|du|_w^2-2Rew^{i\bar{j}}u_iF_{\bar{j}}+g^{k\bar{l}}w^{i\bar{p}}w^{\bar{j}q}R_{k\bar{l}i\bar{j}}u_{\bar{p}}u_q+\frac{1}{2}\frac{|d|du|_w^2|_g^2}{|du|_w^2}\right)$$
$$-2\beta(-\varphi)^{-\beta}|d|du|_w^2|_g\left|\frac{d\varphi}{\varphi}\right|_g$$
$$+|du|_w^2(-\varphi)^{-\beta}\left(\beta\Delta_g(-\log-\varphi)+\beta^2\left|\frac{d\varphi}{\varphi}\right|_g^2\right)$$

(by a double application of the Schwarz inequality)

$$\geqq(-\varphi)^{-\beta}(2(n+1)|du|_w^2-2Rew^{i\bar{j}}u_iF_{\bar{j}}+g^{k\bar{l}}w^{i\bar{p}}w^{\bar{j}q}R_{k\bar{l}i\bar{j}}u_{\bar{p}}u_q)$$
$$+\frac{1}{2}(-\varphi)^{-\beta}|du|_w^2\left(\frac{|d|du|_w^2|_g}{|du|_w^2}-2\beta\left|\frac{d\varphi}{\varphi}\right|_g\right)^2$$
$$+(-\varphi)^{-\beta}\beta|du|_w^2(n-c_1(-\varphi)^\delta)-\beta^2(-\varphi)^{-\beta}|du|_w^2\left|\frac{d\varphi}{\varphi}\right|_g^2$$
$$\geqq(-\varphi)^{-\beta}|du|_w^2\left(2(n+1)+n\beta-\beta^2\left|\frac{d\varphi}{\varphi}\right|_g^2-\beta c_1(-\varphi)^\delta\right)$$
$$+(-\varphi)^{-\beta}(-|du|_w^2-|dF|_w^2+(1+c_2(-\varphi)^\delta)w^{k\bar{l}}w^{i\bar{p}}w^{\bar{j}q}R_{k\bar{l}i\bar{j}}u_{\bar{p}}u_q)$$
$$\geqq(-\varphi)^{-\beta}(|du|_w^2(2n+1+n\beta-\beta^2-c_3(-\varphi)^\delta)-|dF|_w^2$$
$$-(1+c_2(-\varphi)^\delta)|du|_w^2(n+1)(1+c_4(-\varphi)))$$
$$\geqq(-\varphi)^{-\beta}(|du|_w^2(n+n\beta-\beta^2-c_5(-\varphi)^\delta)-|dF|_w^2)$$

$$-\Delta_g(-\varphi)^\alpha=\alpha(-\varphi)^\alpha\frac{\Delta_g\varphi}{-\varphi}-\alpha(\alpha-1)(-\varphi)^\alpha\left|\frac{d\varphi}{\varphi}\right|_g^2$$
$$=\alpha(-\varphi)^\alpha\Delta_g(-\log-\varphi)-\alpha^2(-\varphi)^\alpha\left|\frac{d\varphi}{\varphi}\right|_g^2$$
$$\geqq\alpha(-\varphi)^\alpha(n-c_1(-\varphi)^\delta)-\alpha^2(-\varphi)^\alpha(1+c_6(-\varphi)^\delta)$$
$$\geqq(n\alpha-\alpha^2)(-\varphi)^\alpha-c_7(-\varphi)^{\alpha+\delta}.$$

If $|dF|_w^2=0(|\varphi|^{\alpha+\beta})$, then for large c_8

$$\Delta_g\left(\frac{|du|_w^2}{(-\varphi)^\beta}-c_8(-\varphi)^\alpha\right)\geqq(n-1+n\beta-\beta^2)\left(\frac{|du|_w^2}{(-\varphi)^\beta}-c_8(-\varphi)^\alpha\right).$$

Setting $k=(n-1+n\beta-\beta^2)$, and noting that $k<n(n+1)$, Lemma I says that $\left(\frac{|du|_w^2}{(-\varphi)^\beta}\right)$ being bounded implies that $|du|_w^2=O(|\varphi|^{\beta+\min\{\alpha,b\}})$ where

$$b=\frac{1}{2}(\sqrt{(n^2-1)^2+4k}-(n^2-1)) \quad k<n(n+1)^2$$

$$\geqq \frac{k}{n(n+1)}.$$

If $\beta \leqq n$, then $b \geqq \frac{n-1}{n(n+1)}$. Using $\alpha=\frac{n-1}{n(n+1)}$, it follows easily that $|du|_w^2=O(|\varphi|^\gamma)$ with $\gamma=\min\left\{l, n+\frac{n-1}{n(n+1)}\right\}$. □

A slightly better bound can be obtained for γ. If $n \geqq 3$, then it is easily seen that $\gamma=\min\{l, n+\frac{1}{2}\}$; however, the above technique gives no better an estimate than $\gamma=\min\{l, n+1\}$.

Corollary. *Under the hypotheses stated above,* $u=O(|\varphi|^{\gamma/2})$.

Proof.

$$|du|_w^2=O(|\varphi|^\gamma) \quad \text{implies} \quad |du|_E=O(|\varphi|^{\gamma/2-1}).$$

Since $u=O(|\varphi|)$, then $u=O(|\varphi|^{\gamma/2})$. □

These results give some a priori knowledge about the potential for a complete Einstein-Kähler metric at strictly pseudoconvex boundary points. For certain "nice" strictly pseudoconvex boundary points, stronger information can be obtained. Let $p \in \partial\Omega$ where Ω is an arbitrary pseudoconvex domain. We shall call p "nice" if there is a neighborhood N of p and a bounded, C^k strictly pseudoconvex domain $\tilde{\Omega}$ such that $\Omega \subset \tilde{\Omega}$ and $\Omega \cap N=\tilde{\Omega} \cap N$.

Suppose $p \in \partial\Omega$ is a "nice" strictly pseudoconvex boundary point. Let N be a neighborhood of p, and Ω_1, Ω_2 be bounded strictly pseudoconvex C^k domains ($k>2n+2$) such that $\Omega_1 \subset \Omega \subset \Omega_2$ and $\Omega_1 \cap N=\Omega_2 \cap N$. Cheng and Yau showed that the potential for the Einstein-Kähler metrics on the domains Ω_1, Ω_2 each agree with Fefferman's approximate solution up to order $(n+1)$; further, since Fefferman's method was local, the approximate solution is the same for both domains. Finally, since the potential for the Einstein-Kähler metric for Ω is caught between those for Ω_1 and Ω_2, it must also agree with Fefferman's approximate solution to order $(n+1)$ near p. This is the desired stronger information.

The main result of this section is the following.

Theorem. *Suppose* $\Omega \subset \mathbb{C}^n$, *and* $q \in \partial\Omega$ *is strictly pseudoconvex, with* $\partial\Omega$ C^k *near* q. *Let* $w=-\log-\varphi$, *where* φ *is the* l^{th} *order approximate solution to* $J(\varphi)=1$ *on*

$$D \cap \Omega \left(l \leqq \min\left\{\left[\frac{k-4}{2}\right], n+1\right\}\right).$$

Let u *be a solution to* $*$. *Then* $u=O(|\varphi|^{\gamma/2})$,

$$|D^p u|_w=O(|\varphi|^{\gamma/2}) \quad p=0, \ldots, k-2l$$

where

$$\gamma = \begin{cases} 2l & \text{if } q \text{ is "nice"} \\ \min\left\{l, n+\dfrac{n-1}{n(n+1)}\right\} & \text{otherwise} \end{cases}$$

and $|D^p u|_w$ *is the length of the* p^{th} *covariant derivative of* u *with respect to* $[w_{i\bar{j}}]$.

Proof. The first statement was established above. To obtain the other estimates, use $*$ for

$$(n+1)u - F = \log|w_{i\bar{j}} + u_{i\bar{j}}| - \log|w_{i\bar{j}}| = \int_0^1 \frac{d}{dt}(\log|w_{i\bar{j}} + tu_{i\bar{j}}|)dt$$

$$= \sum\left(\int_0^1 (w+tu)^{i\bar{j}}dt\right)u_{i\bar{j}} = \sum A^{i\bar{j}}u_{i\bar{j}}$$

where

$$A^{i\bar{j}} = \int_0^1 (w+tu)^{i\bar{j}}dt, \qquad (w+tu)^{i\bar{j}} = [w_{i\bar{j}} + tu_{i\bar{j}}]^{-1}, \qquad [A_{i\bar{j}}] = [A^{i\bar{j}}]^{-1}.$$

Then

$$\frac{1}{c}[w_{i\bar{j}}] \leqq [A_{i\bar{j}}] \leqq c[w_{i\bar{j}}]$$

and

$$\Delta_A u = (n+1)u - F$$

with $F = O(|\varphi|^l)$. Standard Schauder theory implies the result. □

Corollary. *Under the hypothesis above,*

$$u \in C^{\min(\gamma/2-\varepsilon,\, k-2l-\varepsilon)}(D \cap \bar{\Omega}), \quad \text{for} \quad \varepsilon > 0.$$

Corollary. *If* $n \geqq 6$, $k \geqq 18$, *and* $g_{i\bar{j}}$ *is a complete Einstein-Kähler metric on* Ω, *then* $e^{-g} \in C^{3+\varepsilon}(D \cap \bar{\Omega})$ *for some* $\varepsilon > 0$, *where* g *is the potential for the metric.*

Proof. Choose $l = 7$. Then $u \in C^{3+\varepsilon}(D \cap \bar{\Omega})$ and $\varphi \in C^{3+\varepsilon}(D \cap \bar{\Omega})$ for some $\varepsilon > 0$ implies that $e^{-g} = -\varphi e^{-u} \in C^{3+\varepsilon}(D \cap \bar{\Omega})$.

Corollary. *If* q *is a "nice" strictly pseudoconvex boundary point,* $k \geqq 12$, *and* $g_{i\bar{j}}$ *is the restriction of a complete Einstein-Kähler metric to* $D \cap \bar{\Omega}$, *then* $e^{-g} \in C^{3+\varepsilon}(D \cap \bar{\Omega})$ *for some* $\varepsilon > 0$.

Suppose that $F: M \to N$ is a biholomorphic map between domains in $\mathbb{C}^n$ that extends continuously to the boundary in a neighborhood of the strictly pseudoconvex boundary points $p \in \partial M$ and $p' \in \partial N$, and that $F(p) = p'$. Then F restricts to a biholomorphic map between bounded domains with complete Einstein-Kähler metrics for which p and p' are "nice" boundary points. The regularity results of this section show that if ∂M and ∂N are C^{12+k} in neighborhoods of p, p' respectively, then the respective potentials for the Einstein-Kähler metrics are $C^{3+k+\varepsilon}$ in neighborhoods of p and p'. In a forthcoming paper, the author will show that under these circumstances, the geodesics for the Einstein-Kähler metric will give $C^{k+\varepsilon}$ coordinate systems in neighborhoods of p and p' respectively. Thus, F extends $C^{k+\varepsilon}$ to a neighborhood of p. It should be noted here that this proof of the local boundary regularity of biholomorphic mappings at strictly pseudoconvex boundary points requires slightly stronger boundary smoothness than the recent result by Lempert [9].

References

1. Aubin, T.: Equations du type Monge-Ampère sur les variétés Kählerienne compactes. Bull. Sci. Math. (2^e série) **102**, 63–95 (1978)
2. Calabi, E.: An extension of E. Hopf's maximum principle with an application to Riemannian geometry. Duke Math. J. **25**, 45–56 (1958)
3. Cheng, S.Y., Yau, S.T.: On the existence of a complete Kähler metric on non-compact complex manifolds and the regularity of Fefferman's equation. Commun. Pure Appl. Math. **33**, 507–544 (1980)
4. Chern, S.S., Moser, J.K.: Real hypersurfaces in complex manifolds. Acta Math. **133**, 219–271 (1974) (Appendix by S. Webster)
5. Fefferman, C.: The Bergman kernel and biholomorphic mappings of pseudoconvex domains. Invent. Math. **26**, 1–65 (1974)
6. Fefferman, C.: Monge-Ampère equations, the Bergman kernel, and geometry of pseudoconvex domains. Ann. Math. **103**, 395–417 (1976)
7. Greene, R.E., Wu, H.: Function theory on manifolds which possess a pole. Lecture Notes in Mathematics, No. 699. Berlin, Heidelberg, New York: Springer 1979
8. Klembeck, P.: Kähler metrics of negative curvature, the Bergman metric near the boundary and the Kobayashi metric on smooth bounded strictly pseudoconvex sets. Indiana Univ. Math. J. **27** (2), 275–282 (1978)
9. Lempert, L.: La métrique de Kobayashi et la représentation des domaines sur la boule. Bull. Soc. Math. France **109**, 427–474 (1981)
10. Lee, J., Melrose, R.: Boundary behavior of the complex Monge-Ampère equation. Acta Math. **148**, 159–192 (1982)
11. Lu, V.C.: Holomorphic mappings of complex manifolds. J. Differential Geometry **2**, 299–312 (1968)
12. Mok, N., Yau, S.T.: Completeness of the Kähler-Einstein metric on bounded domains and the characterization of domains of holomorphy by curvature conditions. Preprint
13. Yang, P.: Automorphism of tube domains. Preprint
14. Yau, S.T.: Harmonic functions on complete Riemannian manifolds. Commun. Pure Appl. Math. **28**, 201–228 (1975)
15. Yau, S.T.: A general Schwarz lemma for Kähler manifolds. Am. J. Math. **100**, 197–203 (1978)
16. Yau, S.T.: On the Ricci curvature of a compact Kähler manifold and the complex Monge-Ampère equation I^*. Commun. Pure Appl. Math. **31**, 339–411 (1978)

Received September 13, 1982

Math. Ann. 263, 303–312 (1983)

Mathematische Annalen
© Springer-Verlag 1983

A Sufficient Condition for a Critical Point of a Functional to be a Minimum and Its Application to Plateau's Problem

A. J. Tromba
Department of Mathematics, University of California, CA 95064, USA

0. Introduction

It is a classical consequence of Taylor's theorem that if $U \subset \mathbb{R}^n$ is an open subset, $q \in U$ a critical point for a C^2 mapping $f: U \to R$ and the second derivative of f at q or "Hessian" of f, $D^2 f(q): \mathbb{R}^n \times \mathbb{R}^n \to R$ is a positive quadratic form, then q is a strict minimum for f.

If $U \subset H$ is open in a Hilbert space H, q a critical point for $f: U \to R$, and $D^2 f(q)(v, v) \geqq c\|v\|^2$, $c > 0$, then q is also a strict minimum for f.

If we merely have $D^2 f(q)(v, v) > 0$ for all v, the implication that q is a strict minimum is, in general, false. To illustrate an example of such an f let $H = l_2$ the Hilbert space of square summable sequences. For $x = (x_1, x_2, \ldots)$ define $f(x) = -\sum_{i=1}^{\infty} \frac{1}{i^4} \cos i x_i$. $0 \in l_2$ is clearly a critical point and moreover $D^2 f(0)(v, v) = \sum_i \frac{1}{i^2} v_i^2$ which is obviously positive. $f(0) = -\sum_i 1/i^4$. However if $y_n = (0, \ldots, 2\pi/n, 0, \ldots)$, where $2\pi/n$ occurs in the n^{th} place then $f(y_n) = f(0)$. Thus we have a sequence $y_n \to 0$ with $f(y_n) = f(0)$ and 0 is *not* a strict minimum for f.

It is also possible to find an f with $D^2 f(0)(v, v) > 0$ for $v \neq 0$ with 0 not even a minimum [consider for example $f(x) = \sum_{i=1}^{\infty} \left(-\frac{1}{i^4} + \frac{x_i^4}{i^4} \right) \cos i x_i$].

Despite these examples there are many many problems in non-linear functional analysis where all we know is that the Hessian is positive but *not* positive definite. In these cases it would be useful to know when the critical point is a minimum. The situation which we have in mind is Plateau's problem where this occurs. In the section below we shall describe a theory which shows how to deal with such situations and which also lays the foundation for an infinite dimensional catastrophe theory for a class of problems heretofore intractible.

An approach to such questions was given by the author in [8] and by Magnus in [4]. However the results in [8] were phrased in much too abstract a setting for them to be reachly applicable and in [4] applications seem not to be the immediate objective.

I. The Local Theory in Banach Space

For clarity of exposition of our ideas we shall start first with a discussion of functionals defined on open subsets of certain Banach spaces, and then only later globalize to the somewhat more complicated case of functionals defined on open subsets of Banach manifolds.

Let E be a Banach space with norm $\| \ \|$ and suppose $\langle\langle , \rangle\rangle : E \times E \to R$ is an inner product on E; i.e. $\langle\langle , \rangle\rangle$ is continuous, symmetric, bilinear and satisfies $\langle\langle u, u \rangle\rangle \geqq 0$ and equals 0 iff $u = 0$. The inner product $\langle\langle , \rangle\rangle$ induces a *weak* norm on E given by $\|u\|_w = \sqrt{\langle\langle u, u \rangle\rangle}$ with $\|u\|_w \leqq \text{const} \|u\|$. We can complete E in this norm obtaining a Hilbert space E_w with a natural inclusion map $i : E \to E_w$. By construction i will have a dense range in E_w.

Definition 1.1. A smooth real valued C^r map, $r \geqq 1$, $f : U \to R$ with $U \subset E$, E a Banach space with an inner product $\langle\langle , \rangle\rangle$, is called *$C^r$-acceptable* with respect to $\langle\langle , \rangle\rangle$ if there is a C^{r-1} vector field $X : U \to E$ such that the differential of f, df, satisfies

$$df(p)h = \langle\langle X(p), h \rangle\rangle$$

for all $h \in E$ and $p \in U$. Such an X can be thought of as the gradient of f with respect to $\langle\langle , \rangle\rangle$. In Sect. 3 we shall see that this class of functionals arises naturally in variational problems.

Suppose we have an inner product $\langle\langle , \rangle\rangle$ on a Banach space E. Let $\mathscr{A}$ denote the class of continuous linear maps from E to itself which have continuous linear adjoints $A^* : E \to E$ with respect to $\langle\langle , \rangle\rangle$ with $A^* = A$. "$\mathscr{A}$" is clearly a closed subspace of the ring of all linear maps. We shall call the members of $\mathscr{A}$ *self-adjoint* linear maps. Thus $A \in \mathscr{A}$ iff $\langle\langle Au, v \rangle\rangle = \langle\langle u, Av \rangle\rangle$ for all $u, v \in E$.

Lemma 1.2. *If $f : U \to R$ is a C^2-acceptable functional then the Frechét derivative at any point $p \in U$ of the gradient X of f with respect to $\langle\langle , \rangle\rangle$ is self-adjoint.*

Proof. Since $d^2 f(p)(u, v) = \langle\langle DX(p)u, v \rangle\rangle$, the result follows immediately from the symmetry of the second derivative of f.

We are now ready to state and prove the main result of this section.

Theorem 1.3. *Let $f : U \to R$ be a C^3-acceptable mapping on a Banach space E with respect to an inner product structure $\langle\langle , \rangle\rangle$ on E. Suppose q is a critical point of f ($df(q) = 0$, $X(q) = 0$) with $D^2 f(q)(v, v) > 0$ for $v \neq 0$ and the Frechét derivative $DX(q) : E \xleftarrow{\sim}$ an isomorphism. Then q is a strict minimum for f.*

This result will follow directly from the next lemma, a generalization of the Morse lemma. The essential computational ideas are due to Palais [6], however see also [8].

Theorem 1.4 (Morse Lemma). *Let $f : U \to R$ be a C^3-acceptable mapping with q a critical point of f with $DX(q) : E \xleftarrow{\sim}$ an isomorphism. (Such a critical point will be said to be non-degenerate.) Then there exists a diffeomorphism φ of a neighborhood of q such that*

$$f \circ \varphi(\mathfrak{x}) = f(q) + \tfrac{1}{2} D^2 f(q)(\mathfrak{x}, \mathfrak{x}).$$

Proof. For simplicity let us assume that $q = 0$.

By the fundamental theorem of integral calculus we have that

$$f(x)=f(0)+\int_0^1 df(tx)x\,dt$$

$$=f(0)+\left\langle\!\!\left\langle \int_0^1 X(tx)dt, x \right\rangle\!\!\right\rangle.$$

Again applying the fundamental theorem this equals

$$f(0)+\tfrac{1}{2}\langle\!\langle A(x)x, x\rangle\!\rangle \tag{1.5}$$

where

$$A(x)=\int_0^1\int_0^1 DX(stx)ds dt\,.$$

Since $DX(p)\in\mathscr{A}$ for each $p\in U$ it follows that $A(x)\in\mathscr{A}$ for each $x\in U$. Moreover $A(0)$ is an isomorphism and $x\to A(x)$ is C^1.

Let $Q(x)=A^{-1}(x)A(0)$. Then for each x, $x\to Q(x)$ is C^1 and $Q(0)=I$, the identity. Thus for x sufficiently close to 0 $Q(x)$ has a "square root" $S(x)$ which is a power series in Q with $S(x)^2=Q(x)$, $x\mapsto S(x)C^1$. Now Q has an adjoint Q^* with respect to $\langle\!\langle\,,\rangle\!\rangle$ defined by $Q^*=A(0)A^{-1}(x)$, and

$$Q^*(x)A(x)=A(x)Q(x)=A(0)\,.$$

Moreover locally Q^* has a square root $\tilde{S}$ with $\tilde{S}=S^*$, and since $S(x)$ is given as a power series in $Q(x)$ we have that

$$S^*(x)A(x)=A(x)S(x)\,,$$

and so

$$S^*(x)A(x)S(x)=A(x)S^2(x)=A(0)\,.$$

Let $R(x)=S(x)^{-1}$. Then from formula (1.5) we have

$$f(x)=f(0)+\tfrac{1}{2}\langle\!\langle R(x)^*A(0)R(x)x, x\rangle\!\rangle\,,$$

where $R(0)=I$ and $x\mapsto R(x)$ is C^1. Let $\Psi(x)=R(x)x$. It follows immediately that $D\Psi(0)=I$ and thus Ψ has a local inverse φ. Clearly φ satisfies the construction of the lemma.

II. The Global Theory

In this section we prove the manifold analogs to Theorems 1.3 and 1.4, analogs which we apply to Plateau's problem in the next section. For the purpose of exposition we shall not present the theory in its most abstract possible formulation (as we did in [8]) but in a formulation motivated by the example we wish to present in the next section.

Let E be a Banach space with a continuous symmetric bilinear pairing $\langle\!\langle\,,\rangle\!\rangle$ with the property that $\langle\!\langle v, v\rangle\!\rangle\geqq 0$. We are *not* assuming that $\langle\!\langle v, v\rangle\!\rangle=0$ implies $v=0$. The reason for this will be clear in Sect. 3. Let $M\subset E$ be a C^3 submanifold. The pairing on E induces a pairing on the tangent space $T_pM\subset E$ for each $p\in M$.

We shall assume that this is an inner product on T_pM, i.e. $\langle\langle v, v\rangle\rangle = 0$ and $v \in T_pM$ implies $v = 0$.

Since M is a submanifold of E there is for each $q \in M$ a diffeomorphism of a neighborhood V of q in E onto a neighborhood of E which locally "flattens" M, i.e. $\varphi(M \cap V) \subset T_qM \subset E$, an open subset of the tangent space at q.

Definition 2.1. A C^r submanifold $M \subset E$ will be called a C^r *weak Riemannian submanifold* if a collection of locally flattening diffeomorphisms $\{\varphi, V\}$ as above can be found such that

(i) $\{\varphi | V \cap M, V \cap M\}$ is a coordinate cover for M

(ii) for each $p \in M$ $\langle\langle , \rangle\rangle$ is an inner product (although not necessarily complete) for $T_pM \subset E$. If (φ, V) is a coordinate chart about $q \in M$ then for each $p \in V$, by definition $D\varphi(p) : E \to E$. Let $\overline{D\varphi}(p) = D\varphi(p) | T_pM$, the restriction of $D\varphi(p)$ to T_pM. So $\overline{D\varphi}(p) : T_pM \to T_qM$.

(iii) For each $p \in V$ we require that $\overline{D\varphi}(p)$ have an adjoint with respect to $\langle\langle , \rangle\rangle$, such that $\overline{D\varphi}(p)^* : T_qM \to T_pM$, isomorphically. Consequently $\overline{D\varphi}(p)^* \in \mathscr{L}(T_qM, E)$, the Banach space of linear maps from T_qM to E. We shall require that $p \to \overline{D\varphi}(p)^*$ is C^{r-1}.

A maximal atlas of such φ's will be called a weak-Riemannian structure for M.

Definition 2.2. Let $U \subset M$ be open with $f : U \to R$ a C^r map, $r \geqq 1$, $M \subset E$ as above. f is called C^r*-acceptable* with respect to the induced inner product structure on the tangent bundle TM, if for each $p \in U$ there is a C^{r-1} vector field $X : U \to TM | U$ such the differential of f, df, satisfies

$$df(p)[h] = \langle\langle X(p), h\rangle\rangle$$

for all $h \in T_pM$ and $p \in U$. X is then the gradient of f with respect to $\langle\langle , \rangle\rangle$.

We would like to prove a version of 1.3 for such an f about a critical point $q \in U \subset M$ under the assumption that $DX(q) : T_qM \leftarrow$ is an isomorphism. At first glance it may seem that we could do this without any special coordinate structures on M. We would begin by choosing any coordinate chart (φ, V), $V \subset U$, about q to "pull things back" to $\tilde{V} = \varphi(V) \subset T_qM$, and then attempt to apply 1.3. Unfortunately we would not be in the situation of 1.3 since the pull back (via φ) of X will be a vector field on $\tilde{V}$ but it will not be the gradient of f with respect to the "constant" inner product $\langle\langle , \rangle\rangle$ on T_qM (or in the language of Differential Geometry, with respect to the flat metric on $\tilde{V}$).

It will, however, be the gradient of the pull back of the given metric on TM induced from the inner product on E. Unfortunately this pull back metric is not in general flat, or constant on $\tilde{V}$ and the methods of 1.4 do not apply. Fortunately, the assumptions that M is a smooth weak Riemannian submanifold of E and that f is C^1 acceptable garauntee that f does indeed have a gradient $\tilde{X}$ on $\tilde{V}$ with respect to the flat or constant metric $\langle\langle , \rangle\rangle$ on T_qM.

For $\tilde{p} \in \tilde{V}, \tilde{X}(\tilde{p})$ will be given by the formula

$$\tilde{X}(\tilde{p}) = [\overline{D\varphi}(p)^*]^{-1} X(\varphi^{-1}(\tilde{p})),$$

where $\overline{D\varphi}^*(p) : T_qM \to T_pM$ is the adjoint assumed in Definition 2.1 (iii).

To see that this formula for $\tilde{X}$ is indeed correct note that by definition

$$\tilde{X}(\tilde{p}) = D\varphi(p)X(\varphi^{-1}(\tilde{p}))$$

and

$$df(p)[h] = \langle\langle X(p), h\rangle\rangle.$$

Therefore for $\tilde{h} \in T_q M$

$$\begin{aligned} d(f\circ\varphi^{-1})(\tilde{p})[\tilde{h}] &= \langle\langle X(\varphi^{-1}(\tilde{p})), D\varphi^{-1}(\tilde{p})[\tilde{h}]\rangle\rangle \\ &= \langle\langle X(\varphi^{-1}(\tilde{p})), [\overline{D\varphi}(p)]^{-1}[\tilde{h}]\rangle\rangle \\ &= \langle\langle [\overline{D\varphi}(p)^*]^{-1} X(\varphi^{-1}(p)), \tilde{h}\rangle\rangle^1. \end{aligned}$$

This verifies the formula for $\tilde{X}$.

Applying these remarks and Theorem 1.4 we obtain

Theorem 2.3 (Morse Lemma). *Let $f: U \to R$ where U is an open subset of a weak Riemannian submanifold M of E. Assume that f is C^3-acceptable and that q is a critical point of f with $DX(q): T_q M \leftarrow$ an isomorphism. Then for any weak Riemannian coordinate* chart φ *about q (as defined by 2.1), there exists a C^1 diffeomorphism Ψ of a neighborhood of $\varphi(q)$ with a neighborhood of $\varphi(q)$ in $T_q M \subset E$, preserving $\varphi(q)$ and such that*

$$(f\circ\varphi^{-1})(\Psi(\mathfrak{x})) = f(q) + D^2(f\circ\varphi^{-1})(\varphi(q))(\mathfrak{x},\mathfrak{x}).$$

We immediately obtain

Theorem 2.4. *Let $f: U \to R$ be a C^3-acceptable mapping on an open subset U of a weak-Riemannian submanifold M of E. Let q be a non-degenerate critical point of f (that is $DX(q): T_q M \leftarrow$ is an isomorphism), where X is the weak gradient of f. If $D^2 f(q)(v,v) > 0$ for all $v \in T_q M$, $v \neq 0$, then q is a strict local minimum for f.*

We now proceed to apply this to Plateau's problem.

III. The Problem of Plateau

For most of this section we shall be following [7].

Let $H^{r+6}(S^1, R^n)$ be the Sobelev Hilbert space of H^{r+6} maps of the unit circle S^1 into R^n, with $r \geqq 5$. Let $\mathscr{A} = \mathrm{Emb}(S^1, R^3)$ be the open submanifold of $H^{r+6}(S^1, R^n)$ which consists of embeddings of S^1 into R^n, $n \geqq 2$. Let Γ be the image of such an embedding $\alpha \in \mathscr{A}$. Set η_α to be the component of $H^2(S^1, \Gamma)$ {the C^r Hilbert manifold of H^2 maps from S^1 to Γ} determined by the embedding α. Let $\mathscr{M}_\alpha$ be the open submanifold of η_α consisting of the diffeomorphisms. For every $u \in H^2(S^1, \Gamma) \subset H^2(S^1, R^n) = \mathscr{H}$ we can extend $u = (u_1, \ldots, u_n)$ harmonically to the disc $\mathscr{D}$. Let $\mathscr{P}_\alpha$ denote the open submanifold of $\mathscr{M}_\alpha$ consisting of those maps whose harmonic extensions are immersions. Define the C^r smooth Dirichlet energy functional $E_\alpha: \eta_\alpha \to R$ by

$$E_\alpha(u) = \frac{1}{2}\sum_{i=1}^{3} \int_{\mathscr{D}} \left[\left(\frac{\partial u_i}{\partial x}\right)^2 + \left(\frac{\partial u_i}{\partial y}\right)^2\right] dx dy.$$

1 Of course we used here the easily checked fact that if A has an adjoint and is invertible A^{-1} has an adjoint given by $(A^{-1})^* = (A^*)^{-1}$

We can write the Dirichlet functional as an integral over $S^1 = \partial \mathscr{D}$ as follows. Let $\theta \to \dfrac{\partial u}{\partial r}(\theta)$ represent the partial dertivative with respect to the polar coordinate r of the harmonic extension of u to $\mathscr{D}$ evaluated at $r=1$. This agrees with the normal derivative on S^1 of the harmonic extension. Let $\langle , \rangle$ denote the $\mathbb{R}^n$ inner product. Then applying Green's theorem to (1) we find that

$$E_\alpha(u) = \frac{1}{2} \int_{S^1} \left\langle \frac{\partial u}{\partial r}, u \right\rangle d\theta .$$

Denote by $\bar{\mathscr{M}}_\alpha$ the closure of $\mathscr{M}_\alpha$ in η_α.

Douglas showed in his pionneering work [2] that the harmonic extensions of the critical points of E_α in $\bar{\mathscr{M}}_\alpha$ are the simply connected minimal surfaces spanning Γ. For a somewhat simpler proof see [7]. Let $h, k \in \mathscr{H} = H^2(S^1, \mathbb{R}^n)$. Define the symmetric bilinear pairing $\langle\langle , \rangle\rangle : \mathscr{H} \times \mathscr{H} \to R$ by setting

$$\langle\langle h, k \rangle\rangle = \int_{S^1} \left\langle \frac{\partial h}{\partial r}, k \right\rangle d\theta .$$

This is *not* an inner product on $\mathscr{H}$ since $\langle\langle h, h \rangle\rangle = 0$ iff h is constant. However it will be an inner product when restricted to $T_u \eta_\alpha$ for any $u \in \eta_\alpha$ (cf. [7, p. 36]). Also from results in [7] it follows that η_α is a weak Riemannian submanifold of $\mathscr{H}$.

This fact follows from the standard construction of coordinate charts for η_α and some basic lemmas proved in [7]. We shall outline a proof of

Theorem 3.1. *η_α is a weak Riemannian submanifold of $\mathscr{H}$.*

Proof. In the construction of coordinate charts for η_α we follow Palais [6]. Let $u_0 \in \eta_\alpha$, $\eta_\alpha \subset H^2(S^1, \Gamma)$ where Γ is a C^{r+5} submanifold of $\mathbb{R}^n$. We can therefore find a C^{r+4} Riemannian metric for $\mathbb{R}^n$ such that Γ is totally geodesic. Then if $\mathscr{E} : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n$ is the corresponding exponential map [i.e. $t \to \mathscr{E}(x, tv)$ is the geodesic emanating from x with tangent vector v] $\mathscr{E}$ is a C^{r+3} map. Let $u_0 \in H^2(S^1, \Gamma)$, let $\varphi(\lambda(\theta)) = \mathscr{E}(u_0(\theta), \lambda(\theta))$, $0 \leqq \theta \leqq 2\pi$, representing points in S^1, and define $\gamma : H^2(S, \mathbb{R}^n) \leftarrow$ by $\gamma(\lambda)(\theta) = \varphi(\lambda(\theta))$. Then γ is C^r and $\gamma(0) = u_0$. The notational use of both γ and φ here is to emphasize the important fact that γ is a map on function space which arises as a composition.

From the property of the exponential function we know that $d\gamma(0) =$ identity, so by the inverse function theorem as a C^r local inverse φ. Since Γ is totally geodesic it follows that for λ near zero in $H^2(S^1, R^n)$, $\gamma(\lambda) \in H^2(S^1, \Gamma)$, iff $\lambda \in T_{u_0} \eta_\alpha = \{\lambda \in H^2 | \lambda(\theta) \in T_{u_0(\theta)} \Gamma\}$. Consequently φ restricted to a suitable neighborhood of u_0 is a locally flattening coordinate chart for η_α. We must now show that for $u \in \eta_\alpha$, u near u_0 the differential $\overline{d\varphi}(u) : T_u \eta_\alpha \to T_{u_0} \eta_\alpha$ [where again $\overline{d\varphi}(u)$ denotes restriction; cf. Definition 2.1], has a C^{r-1} adjoint $\overline{d\varphi}(u)^* : T_{u_0} \eta_\alpha \to T_u \eta_\alpha$.

We note that $d\varphi(u) = d\gamma(\lambda)^{-1}$, where $\lambda = \gamma^{-1}(u)$ and that for $k \in T_{u_0} \eta_\alpha$ $(d\gamma(\lambda)k)(\theta) = d\varphi(\lambda(\theta)) \cdot k(\theta)$, where for each θ, $d\varphi(\lambda(\theta)) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$, the linear maps from $\mathbb{R}^n$ to $\mathbb{R}^n$. This implies that $[d\varphi(u)h](\theta)$ is a map of the form $w(\theta) \cdot h(\theta)$, where $w(\theta) \in \mathscr{L}(\mathbb{R}^n, \mathbb{R}^n)$.

Let us now compute. Let $h \in T_u\eta_\alpha$, $k \in T_{u_0}\eta_\alpha$.

$$\ll \overline{D\varphi}(u)h, k \gg = \ll wh, k \gg$$
$$= \int_{S^1} \left\langle \frac{\partial}{\partial r}(wh), k \right\rangle_{\mathbb{R}^n} d\theta = \int_{S^1} \left\langle h, w^* \frac{\partial k}{\partial r} \right\rangle d\theta,$$

where for each θ, $w^*(\theta)$ is the $\mathbb{R}^n$ adjoint of w. Now $w^* \dfrac{\partial k}{\partial r} \in H^1(S^1, \mathbb{R}^n)$. For $p \in \Gamma$ let $\Omega(p)$ be the orthogonal with respect to the Euclidean $\mathbb{R}^n$ metric) projection of $\mathbb{R}^n$ onto $T_p\Gamma$. Since $h \in T_u\eta_\alpha$ $\Omega(u(\theta))h(\theta) = h(\theta)$ for all θ. Consequently

$$\ll \overline{D\varphi}(u)h, k \gg = \int_{S^1} \left\langle h, \Omega(u) w^* \frac{\partial k}{\partial r} \right\rangle d\theta.$$

For $u = u_0$, $\Omega(u)w^* \dfrac{\partial k}{\partial r} = \Omega(u_0) \dfrac{\partial k}{\partial r}$ ($w^* = \mathrm{id}$). It is shown in [7, p. 23] that $k \to \Omega(u_0) \dfrac{\partial k}{\partial r}$ is an isomorphism from $T_{u_0}\eta_\alpha$ to $(T_{u_0}\eta_\alpha)^1$, the later representing the H^1 completion of $T_{u_0}\eta_\alpha$. From this it follows that for u close to u_0, $k \to \Omega(u)w^* \dfrac{\partial k}{\partial r}$ is an isomorphism of $T_{u_0}\eta_\alpha$ to $T_u\eta_\alpha$. Define $v(w)[k]$ as the unique (cf. [6, p. 23]) solution to the boundary value problem

$$\Omega(u) \frac{\partial}{\partial r} [v(w)[k]] = \Omega(u) w^* \frac{\partial k}{\partial r} \quad \text{on} \quad S^1, \tag{3.2}$$

where of course $v(w)[k]$ is taken to be harmonic in the disc. Clearly for fixed w, $k \to v(w)[k]$ is a linear continuous map from $T_{u_0}\eta_\alpha$ to $T_u\eta_\alpha$. It is obvious that $v(w) = \overline{D\varphi}^*(u)$.

The smoothness of $u \to \overline{D\gamma}(u)^*$ follows from the smoothness of the right hand side of (3.2) as a function of u and the uniqueness of solutions to the Neuman system (3.2). This concludes the lemma.

We must still show that E_α is C^r acceptable. This was also shown in [7, Theorem 6, p. 23]. We state it here as

Theorem 3.3. *$E_\alpha : \eta_\alpha \to R$ is a C^r acceptable functional on η_α. Thus there exists a C^{r-1} vector field $X^\alpha : \eta_\alpha \to T\eta_\alpha$ which is the gradient of E_α with respect to $\ll , \gg$.*

Let G be the three dimensional non-compact Lie group of bijective holomorphic maps of the disc onto itself. G acts on η_α (and on $\mathcal{M}_\alpha$) in the following natural way. If $g \in G$ and $z \in \mathcal{D}$ denotes a point in the unit disc then for $u \in \eta_\alpha$ we define $g_\#(u) \in \eta_\alpha$ by $g_\#(u)(z) = u(g(z))$. The maps $g \mapsto g_\#$ defines the G-action on η_α (and on $\mathcal{M}_\alpha$). This action is not smooth and this causes technical difficulties in the development of the theory [6]. The functional E_α and the vector field X^α of Theorem 3.3 will be equivariant with respect to this action [6, p. 38]. Therefore, there is no hope that the critical points of E_α in $\bar{\mathcal{M}}_\alpha$ will be isolated or non-degenerate in the sence described earlier since the orbits of any critical point will consist of critical points. Let $u \in \bar{\mathcal{M}}_\alpha$ be a zero of X^α, and let $\mathcal{O}_G(u)$ denote the orbit of u under G. Then $\mathcal{O}_G(u)$ is an immersed three-dimensional submanifold of η_α consisting of critical points of E_α. We call $\mathcal{O}_G(u)$ a *critical submanifold* of η_α. The derivative of X^α at u induces a homomorphism of the normal bundle of $\mathcal{O}_G(u)$ into

itself. A critical point u, or more precisely $\mathcal{O}_G(u)$, is *non-degenerate* if this induced homomorphism is an isomorphism.

Remark. It follows from the work in [7] that if u is non-degenerate then u must be an immersed minimal surface, and hence $u \in \mathscr{P}_\alpha$. The proofs of Theorems 3.4 through 3.9 can also be found in [7].

Theorem 3.4. *For an open dense set of embeddings $\mathscr{A}_0 \subset \mathscr{A}$ the zeros of X^α in $\mathscr{P}_\alpha, \alpha \in \mathscr{A}_0$ are non-degenerate (and therefore isolated) three dimensional submanifolds of η_α. Moreover for such $\alpha \in \mathscr{N}$ there are only finitely many such critical submanifolds.*

Theorem 3.5. *The weak Riemannian structure $\langle\langle , \rangle\rangle$ on $T\eta$ induces a smooth C^{r-2} geodesic spray (see [7, 3] for relevant definitions). The corresponding exponential map $\exp_u$ restricted to $T_u\eta_\alpha$ gives a local diffeomorphism of a neighborhood of zero in $T_u\eta_\alpha$ onto a neighborhood of u in η_α. The principal reason for using the geodesic spray of the weak metric is to construct local C^{r-2} codimension three submanifolds of η_α transverse to the orbit of a minimal surface u. Let $u \in \bar{\mathscr{M}}_\alpha$ be a zero of X^α. The orbit $\mathcal{O}_G(u)$ of u under G is the image of an injective immersion. The tangent space $\mathbb{F}_u = T_u[\mathcal{O}_G(u)]$ is the three dimensional subspace of $T_u\eta_\alpha$ spanned by u_θ, $\cos\theta \cdot u_\theta$, and $\sin\theta \cdot u_\theta$. Let $\mathbb{F}_u^\perp$ denote the orthogonal complement of $\mathbb{F}_u$ with respect to the weak inner product on $T_u\eta_\alpha$. $\mathbb{F}_u^\perp$ is a codimension three subspace and $\exp_u$ maps a neighborhood of $\mathcal{O}_u \in \mathbb{F}_u^\perp$ diffeomorphically onto a codimension three submanifold ε of η_α transverse to $\mathcal{O}_G(u)$ at u.*

ε is called the *local exponential surface about a zero u of X^α*. Such an exponential surface provides a useful way of locally "factoring out" the action of G.

If $g \in G$ let $g_\# : \eta_\alpha \leftarrow$ again denote the induced diffeomorphism on η_α. The following result reflects the equivariance of the exponential map.

Theorem 3.6. *If ε is a local exponential surface about a zero u of X^α then for fixed $g \in G$, $g_\#(\varepsilon) = \{g_\#(w) | w \in \varepsilon\}$ is a local exponential surface about $g_\#(u)$.*

We also have the next product neighborhood theorem.

Theorem 3.7. *If N is a neighborhood of the identity of G define $N_{g_0} = \{gg_0 | g \in N\}$. Then N_{g_0} is a neighborhood of g. For a subset $A \subset \eta_\alpha$ set $\mathcal{O}_{N_{g_0}}(A) = \{g_\#(u) | g \in N_{g_0}, u \in A\}$. If $u \in \bar{\mathscr{M}}_\alpha$ is a zero of X^α, then for a sufficiently small local exponential surface ε about u, and a sufficiently small neighborhood N, $\mathcal{O}_{N_{g_0}}(\varepsilon)$ is an open neighborhood of $(g_0)_\# u$; and moreover $\mathcal{O}_{N_{g_0}}(\varepsilon)$ is diffeomorphic to the product $N_{g_0} \times \varepsilon$.*

Remark. Theorem 3.7 is actually quite surprising in view of the fact that for $\omega \in \varepsilon$, $\mathcal{O}_N(\omega)$ need not be a submanifold of η_α.

A minimal surface u is said to be *isolated* if there exists an open neighborhood $\mathscr{N}$ of $\mathcal{O}_G(u)$ in η_α such that if $\omega \in \mathscr{N}$ and $X^\alpha(\omega) = 0$ then $\omega \in \mathcal{O}_G(u)$.

Theorem 3.8. *Let u be a zero of X^α in $\bar{\mathscr{M}}_\alpha$ and suppose that for some sufficiently small local exponential surface ε through u there is no other zero of X^α in ε. Then u is isolated.*

We would now like to show that if u is a zero of X^α and ε is a sufficiently small exponential surface containing u then there is a vector field Y^α on ε with the property that the zeros of Y^α on ε are precisely the zeros of X^α on ε.

For $\omega \in \varepsilon$,

$$T_\omega \eta_\alpha = T_\omega \varepsilon \oplus (T_\omega \varepsilon)^\perp .$$

Let P_ω be the weak orthogonal projection of $T_\omega \eta_\alpha$ onto $T_\omega \varepsilon$ with $\tilde{P}_\omega = I - P_\omega$, I the identity linear map on $T_\omega \eta_\alpha$.

The vector field Y^α on ε defined by $Y^\alpha(\omega) = P_\omega X^\alpha(\omega)$ is a C^{r-3} vector field.

Theorem 3.9. *If ε is a sufficiently small exponential surface about a zero u of X^α then the zeros of Y^α are precisely the zeros of X^α on ε. Y^α can be interpreted as the weak gradient of the restriction to ε of the Dirichlet functional E_α, and moreover is a weak-Riemannian submanifold of $\mathscr{H}$, since it is a finite codimensional submanifold of η_α. The point $u \in \varepsilon \subset \eta^\alpha$ will be a non-degenerate minimal surface iff u is a non-degenerate zero of Y^α in the sense that the Frechét derivative of Y^α at u will be an isomorphism of $T_u \varepsilon$ to itself. A non-degenerate minimal surface u will be isolated in ε and by Theorem 3.8 will also be isolated in η_α.*

Let's see what all this means in our context. Suppose u is a non-degenerate minimal surface and denote by $\bar{E}_\alpha$ the restriction of E_α to a local exponential surface ε about u. Since ε is a C^{r-2} manifold, a weak Riemannian submanifold of $\mathscr{H}$ and for $r \geqq 5$, it follows that $\bar{E}_\alpha : \varepsilon \to R$ is at least C^3 smooth. Conditions of Theorem 2.3 with $M = \varepsilon$, $f = \bar{E}_\alpha$, and $X = Y^\alpha$, are obviously satisfied and therefore the Morse lemma for $\bar{E}_\alpha$ holds about a neighborhood of u in ε.

By earlier comments we may identify ε with an open neighborhood of $\mathscr{O}_u \in \mathbb{F}_u^\perp$ (the weak orthogonal complement of the tangent space to the orbit of u, $T_u[\mathscr{O}_G(u)]$). Making this identification and applying the Morse lemma we have

Theorem 3.10. *There is a local diffeomorphism Ψ of a neighborhood of u in ε into ε which fixes u and such that for $\omega \in \varepsilon$*

$$\bar{E}_\alpha \cdot \Psi(\omega) = E_\alpha(u) + \tfrac{1}{2} d^2 \bar{E}_\alpha(u)(\omega, \omega),$$

where $d^2 \bar{E}_\alpha(u) : T_u \varepsilon \times T_u \varepsilon \to R$ is the restriction of the Hessian of E_α to $T_u \varepsilon \subset T_u \eta^\alpha$.

Let N be a small enough neighborhood of the identity of G and U a sufficiently small neighborhood of u in ε. By Theorem 3.7 the map $(g, \omega) \mapsto g_\#(\omega)$ maps the product $N \times U$ diffeomorphically onto a neighborhood V of u in $\mathscr{M}^\alpha$ (in fact in P^α). Therefore every $v \in V$ can be written as $v = g_\#(\omega), g \in N, \omega \in U$. We already know that E_α is equivariant with respect to the action of G. Consequently $E_\alpha(v) = E_\alpha(g_\#(\omega)) = \bar{E}_\alpha(\omega)$. If U is chosen to be in the domain of Ψ we have an induced diffeomorphism $\Psi_\#$ on V defined by $\Psi_\#(v) = g_\#(\Psi(\omega))$, where $v = g_\#(\omega)$. Applying these remarks to Theorem 3.8 and again making the appropriate identification we obtain the last result of this paper.

Theorem 3.11. *Let $u \in \mathscr{M}_\alpha$ be a non-degenerate minimal surface, and ε a sufficiently small local exponential surface through u. Then there is a local diffeomorphism $\Psi_\#$ of a neighborhood V of u in $\mathscr{M}_\alpha$ which fixes u and such that*

$$E_\alpha(\Psi_\#(v)) = E_\alpha(u) + \tfrac{1}{2} d^2 E_\alpha(u)(\omega, \omega),$$

where

$$v = g_{\#}(\omega), \qquad \omega \in \varepsilon .$$

The next corollary follows immediately.

Corollary 3.12. *If u is a minimal surface such that the Hessian $d^2E_\alpha(u)$ is positive in the sense that $d^2E_\alpha(u)$ is positive definite on $\mathbb{F}_u^\perp$, then u is a strict local relative minimum for E_α in the sense that $E_\alpha(v) > E_\alpha(u)$ unless $v \in \mathcal{O}_G(u) \cap V$.*

Remarks. Nitsche [5] has proved a result related to Corollary 3.12. He showed that if the second normal variation of area at an immersed surface is positive than a critical point of the area functional is a minimum in the C^0 topology. That it is a minimum in the C^1 topology follows from classical differential calculus methods.

In our work the positivity of the Hessian of Energy implies that the minimal surface must be immersed.

References

1. Böhme, R., Tromba, A.J.: The index theorem for classical minimal surfaces. Ann. Math. **113**, 447–499 (1981)
2. Douglas, T.: Solution to the problem of Plateau. Trans. Am. Math. Soc. **33**, 263–321 (1931)
3. Lang, S.: Introduction to differentiable manifolds. New York: Interscience 1962
4. Magnus, R.J.: On universal unifoldings of certain real functions on a Banach space. Math. Proc. Camb. Phil. Soc. **81**, 91–95 (1977)
5. Nitsche, J.C.C.: A new uniqueness theorem for minimal surfaces. Arch. RAT. Mech. **52**, 319–329 (1973)
6. Palais, R.: Morse theory on Hilbert manifolds. Topology **2**, 299–340 (1963)
7. Tromba, A.J.: On the number of simply connected minimal surfaces spanning a curve. AMS Memo. No. 194, November 1977
8. Tromba, A.J.: Almost Riemannian structures on Banach manifolds, the Morse lemma, and the Darboux theorem. Canadian J. Math. **28**, 640–652 (1976)

Received September 1, 1982; in revised form September 9, 1982

Math. Ann. 263, 313–321 (1983)

Mathematische Annalen
© Springer-Verlag 1983

The Bifurcation Set and Logarithmic Vector Fields

Hiroaki Terao*
Department of Mathematics, University of Wisconsin, Madison, WI 53706, USA

The logarithmic vector fields (holomorphic vector fields tangent to given hypersurface) have been studied by Saito, Arnol'd, Lyashko, and Terao. Especially they studied the logarithmic vector fields i) along the arrangement of hyperplanes [6, 9, 10] and ii) along the discriminant of semiuniversal deformation of an isolated hypersurface singularity [6, 1, 4, 11]. Here we shall study the module of the logarithmic vector fields along the bifurcation set of a semiuniversal deformation of an isolated hypersurface singularity and decide its structure. We shall give a method of explicit computation for its free base and a defining equation of the bifurcation set.

1. Free Divisors

(1.1) *Definition.* Let $(D, 0)$ be (a germ of) a reduced hypersurface of $(\mathbf{C}^{n+1}, 0)$. Define

$$\mathrm{Der}_{\mathbf{C}^{n+1}}(\log D)_0 = \{\theta;\ \text{a germ at 0 in } \mathbf{C}^{n+1} \text{ of holomorphic vector field such that } \theta \cdot h \in h\mathcal{O}\},$$

where $\mathcal{O} = \mathcal{O}_{\mathbf{C}^{n+1}, 0}$ and $h \in \mathcal{O}$ is a reduced defining equation for D. The element of $\mathrm{Der}_{\mathbf{C}^{n+1}}(\log D)$ is called a *logarithmic vector field along D*. Sometimes we simply write $\mathrm{Der}_{\mathbf{C}^{n+1}}(\log D)$ or $\mathrm{Der}(\log D)$ instead of $\mathrm{Der}_{\mathbf{C}^{n+1}}(\log D)_0$.

(1.2) *Definition.* Under the same notation as above, $(D, 0)$ is said to be *free* if $\mathrm{Der}(\log D)$ is an $\mathcal{O}$-free module. Free divisor is called Saito divisor according to the terminology of Cartier [2].

2. Bifurcation Set and Theorem A and B

From now on, denote the indeterminates $(x_1, \ldots, x_m)$, $(y_2, \ldots, y_k)$, and $(y_1, \ldots, y_k)$ simply by (x), (y'), and (y), respectively. Then $(y) = (y_1, y')$.

---

* The author is a JMS research fellow

Assume that $f \in \mathbf{C}\{x\}$ satisfies $f(0)=0$ and has an isolated critical point at 0. Consider a semiuniversal deformation of f. In other words, choose $F \in \mathbf{C}\{x, y'\}$ such that

i) $F(x,0)=f(x)$,

ii) $(\partial F/\partial y_i)(x,0)$ $(i=2,\ldots,k)$ together with 1 determine $\mathbf{C}$-base for $\mathbf{C}\{x\}/(\partial f/\partial x_1,\ldots,\partial f/\partial x_m)$.

Put $Z=(\mathbf{C}^{k-1},0)$, $Y=(\mathbf{C},0)\times Z$ and $X=(\mathbf{C}^m,0)\times Z$ with their coordinates (y'), (y), and (x,y') respectively.

Define a holomorphic map

$$\varphi : X \to Y$$

by $\varphi^* y_1 = F(x,y')$ and $\varphi^*(y')=(y')$. Denote the critical set and the discriminant (the set of critical values) of φ by C and D respectively. Then the structure sheaf of C is defined as

$$\mathcal{O}_C = \mathbf{C}\{x,y'\}/(\partial F/\partial x_1,\ldots,\partial F/\partial x_m).$$

The defining ideal of D is the 0-th fitting ideal of $\mathcal{O}_C$ as an $\mathcal{O}_Y$-module.

Let

$$\pi : Y=(C,0)\times Z \to Z$$

be the natural projection.

(2.1) *Definition. The bifurcation set* B is defined by $\pi(\mathrm{Sing}D)\subset Z$.

(2.2) **Theorem A.** *The bifurcation set* B *is a free hypersurface, i.e.,* $\mathrm{Der}_Z(\log B)$ *is* $\mathcal{O}_Z$*-free.*

We shall show the explicit way to compute an $\mathcal{O}_Z$-free base for $\mathrm{Der}_Z(\log B)$ in Theorem C.

Define

$$\mathcal{K} = \{\theta \in \mathrm{Der}_Y ; \theta \cdot y_i \in \mathbf{C}\{y'\}\ (i=2,\ldots,k)\}.$$

Then $\mathcal{K}$ is an $\mathcal{O}_Z$-submodule of Der_Y and equal to the set of "lowerable" vector fields to Z by π. Thus π induces an $\mathcal{O}_Z$-linear map

$$\Pi = \pi_* : \mathcal{K} \cap \mathrm{Der}_Y(\log D) \to \mathrm{Der}_Z$$

by $\Pi\left(\sum_{i=1}^{k} f_i(\partial/\partial y_i)\right) = \sum_{i=2}^{k} f_i(\partial/\partial y_i)$. The following theorem is a by-product of the proof of Theorem A.

(2.3) **Theorem B.** $\mathrm{Im}(\Pi)=\mathrm{Der}_Z(\log B)$, *in other words, every element of* $\mathrm{Der}_Z(\log B)$ *is liftable to an element of* $\mathcal{K}\cap\mathrm{Der}_Y(\log D)$.

(2.4) *Remark.* The result of this type was first proved for the discriminants and the bifurcation sets of reflecting groups A_2, A_3, B_3 by Arnol'd [1, Theorem 4] through direct computation. He also conjectured the general theorem for any real irreducible reflection groups. Later Lyashko proved it in much more generalized form [4]. Our Theorem B is nothing but a special case of the result by Lyashko, but his proof is not available as far as the author knows.

3. Proof of Theorems A and B

To prove Theorems A and B, we should recall four properties (3.1)–(3.4) of the discriminant D:

(3.1) The types of singularities of D are $(2,3)$-cusps and ordinary double points outside some lower-dimensional analytic subset of D (over which the fiber has an A_2-singularity or A_1+A_1-singularities).

(3.2) D is a free hypersurface in Y and $\mathrm{Der}_Y=\mathrm{Der}_Y(\log D)\oplus\mathscr{G}$ (as $\mathcal{O}_Z$-modules), where $\mathscr{G}=\sum_{i=1}^{k}\mathcal{O}_Z(\partial/\partial y_i)$ [7].

(3.3) D can be defined by a reduced defining equation $\delta=0$, $\delta\in\mathbf{C}\{y\}$, with

$$\delta=y_1^k+\delta',$$

where $\delta'\in(y_2,\ldots,y_k)\mathbf{C}\{y\}$.

(3.4) There exist a real constant R and an neighborhood V of 0 in D such that

$$|\partial\delta/\partial y_i(p)|\leqq R|\partial\delta/\partial y_1(p)| \qquad (i=1,\ldots,k)$$

for any $p\in V$ (see [8, (5.5)]). It follows from this fact that

$$\begin{aligned}\mathrm{Sing}(D)&=\{\delta=\partial\delta/\partial y_1=0\}\\&=\text{the critical set of }\pi|_D,\end{aligned}$$

and

$$B=\text{the discriminant of }\pi|_D.$$

(3.5) **Lemma.** $\mathrm{Im}(\Pi)$ $(\subset\mathrm{Der}_Z)$ *is a free* $\mathcal{O}_Z$*-module of rank* $(k-1)$.

Proof. Notice $\mathscr{K}\supset\mathscr{G}$. We contend that the natural map

$$\mathscr{K}\cap\mathrm{Der}_Y(\log D)\to\mathscr{K}/\mathscr{G}$$

is an $\mathcal{O}_Z$-isomorphism. It is injective because

$$\mathscr{G}\cap\mathscr{K}\cap\mathrm{Der}_Y(\log D)=\mathscr{G}\cap\mathrm{Der}_Y(\log D)=0$$

on account of (3.2). It is surjective because

$$\begin{aligned}\mathscr{K}&=\mathscr{K}\cap(\mathscr{G}+\mathrm{Der}_Y(\log D))\\&=\mathscr{G}+(\mathscr{K}\cap\mathrm{Der}_Y(\log D))\end{aligned}$$

in the light of (3.2) and $\mathscr{K}\supset\mathscr{G}$. Since $\mathscr{K}/\mathscr{G}\tilde{\to}\mathcal{O}_Y/\mathcal{O}_Z$, we have an $\mathcal{O}_Z$-isomorphism

$$\mathscr{K}\cap\mathrm{Der}_Y(\log D)\tilde{\to}\mathcal{O}_Y/\mathcal{O}_Z$$

by corresponding $\theta\in\mathscr{K}\cap\mathrm{Der}_Y(\log D)$ to $[\theta\cdot y_1]\in\mathcal{O}_Y/\mathcal{O}_Z$. If δ and $\partial\delta/\partial\delta y_1$ have a common divisor in $\mathcal{O}_Y$, then $\dim\mathrm{Sing}(D)=\dim D$ by (3.4). This contradicts the fact that δ is reduced. Thus δ and $\partial\delta/\partial y_1$ have no common factor. Therefore

$$\mathrm{Ker}(\Pi)=\mathcal{O}_Y(\delta(\partial/\partial y_1)).$$

Consider a diagram

$$\begin{array}{ccccccc} 0 \to & \mathcal{O}_Y(\delta(\partial/\partial y_1)) & \to & \mathscr{K}\cap \mathrm{Der}_Y(\log D) & \xrightarrow{\Pi} & \mathrm{Im}(\Pi) & \to 0 \\ & \| & & \| & & & \\ 0 \to & \delta\mathcal{O}_Y & \to & \mathcal{O}_Y/\mathcal{O}_Z & \to & \mathcal{O}_D/\mathcal{O}_Z & \to 0. \end{array}$$

where $\mathcal{O}_D=\mathcal{O}_Y/\delta\mathcal{O}_Y$. So we have an $\mathcal{O}_Z$-isomorphism

$$\varepsilon : \mathrm{Im}(\Pi)\to\mathcal{O}_D/\mathcal{O}_Z \tag{3.6}$$

with $\varepsilon(\Pi(\theta))=[\theta\cdot y_1]\in\mathcal{O}_D/\mathcal{O}_Z$ for $\theta\in\mathscr{K}\cap\mathrm{Der}_Y(\log D)$. Because of (3.3), we have

$$\mathcal{O}_D\underset{\mathcal{O}_Z}{\otimes}\mathcal{O}_Z/(y_2,\ldots,y_k)\xrightarrow{\sim}\mathbf{C}\{y_1\}/(y_1^k).$$

This together with Nakayama's lemma implies that $\mathcal{O}_D$ is a finite $\mathcal{O}_Z$-module with generators $[1],[y_1],\ldots,[y_1^{k-1}]$. Clearly these generators are $\mathcal{O}_Z$-independent in $\mathcal{O}_D$. Thus $\mathcal{O}_D$ is $\mathcal{O}_Z$-free with base $[1],[y_1],\ldots,[y_1^{k-1}]$. It follows from this that $\mathcal{O}_D/\mathcal{O}_Z$ is $\mathcal{O}_Z$-free module of rank $k-1$ with base $[y_1],\ldots,[y_1^{k-1}]$. Since $\mathrm{Im}(\Pi)\cong\mathcal{O}_D/\mathcal{O}_Z$, we have the desired result. □

(3.7) **Lemma.** *The two sheaves* $\mathrm{Im}(\Pi)$ *and* $\mathrm{Der}_Z(\log B)$ *are equal outside some* (k–3)*-dimensional analytic set of* B.

(3.7) is an immediate consequence from (3.1) and (3.4) and the following

(3.8) **Sublemma.** *Assume that a hypersurface* $H\subset Y$ *has multiplicity two at the origin. If* $\pi|_H : H\to Z$ *is finite of degree two at the origin, then*

$$\pi_*(\mathscr{K}\cap\mathrm{Der}_Y(\log H))=\mathrm{Der}_Z(\log K),$$

where K *is the discriminant of* $\pi|_H$.

Proof. By the Weierstrass preparation theorem, we can assume that H is defined by

$$h=y_1^2+k \quad (k\in\mathbf{C}\{y'\})$$

after suitable coordinate changes of Z and Y. Then K is defined by $k=0$. For any

$$\theta=f_1(y)(\partial/\partial y_1)+\sum_{i=2}^{k} f_i(y')(\partial/\partial y_i)\in\mathscr{K}\cap\mathrm{Der}_Y(\log H),$$

$$\begin{aligned}(\pi_*(\theta))\cdot k &= \sum_{i=2}^{k} f_i(y')(\partial k/\partial y_i)\\ &= \sum_{i=2}^{k} f_i(y')(\partial h/\partial y_i)\\ &=(\theta\cdot h)-f_1(y)(\partial h/\partial y_1)\\ &=(\theta\cdot h)-2y_1f_1(y).\end{aligned}$$

Since $(\pi_*(\theta))\cdot k\in\mathbf{C}\{y'\}$, we obtain

$$(\pi_*(\theta))\cdot k=(\theta\cdot h)(0,y')\in h(0,y')\mathcal{O}_Z=k\mathcal{O}_Z.$$

This shows $\pi_*(\theta)\in \mathrm{Der}_Z(\log K)$ and thus

$$\pi_*(\mathscr{K}\cap \mathrm{Der}_Y(\log H))\subset \mathrm{Der}_Z(\log B).$$

Conversely if

$$\psi=\sum_{i=2}^{k} f_i(y')(\partial/\partial y_i)\in \mathrm{Der}_Z(\log K)$$

with $\psi\cdot k=p(y')k$, put

$$\eta=(y_1 p/2)(\partial/\partial y_1)+\psi\in \mathrm{Der}_Y.$$

Then we have

$$\eta\cdot h=py_1^2+pk=ph,$$

and thus $\eta\in \mathrm{Der}_Y(\log H)$. Since

$$\pi_*(\eta)=\psi,$$

this proves

$$\mathrm{Der}_Z(\log K)\subset \pi_*(\mathscr{K}\cap \mathrm{Der}_Y(\log H)). \quad \square$$

Proof of Theorems A and B. By (3.5), we can take an $\mathcal{O}_Z$-base

$$\psi_i=\sum_{j=2}^{k} f_{ji}(\partial/\partial y_j) \qquad (i=2,\ldots,k)$$

for $\mathrm{Im}(\Pi)$. Next by (3.7), $\psi_2,\ldots,\psi_k$ make a base for $\mathrm{Der}_Z(\log B)$ outside a (k–3)-dimensional analytic subset of B. By Saito's criterion [6, (1.8)], $\beta=\det(f_{ij})_{2\leqq i,j\leqq k}$ vanishes at order one generically along B. This implies that β vanishes at order one entirely along B and that $\psi_2,\ldots,\psi_k$ make a base for $\mathrm{Der}_Z(\log B)$. Thus $\mathrm{Der}_Z(\log B)=\mathrm{Im}(\Pi)$ is $\mathcal{O}_Z$-free. $\square$

4. Explicit Computation of Free Base for $\mathrm{Der}_Z(\log B)$

Under the same notation in Sects. 2 and 3, put $g_1=-1$, $g_i=\partial F/\partial y_i (i=2,\ldots,k)$ and $\mathscr{J}=(\partial F/\partial x_1,\ldots,\partial F/\partial x_m)\mathcal{O}_X$. First we shall recall a free base for $\mathrm{Der}_Y(\log D)$ [7, 11]. Define a $k\times k$-matrix $A=(a_{ij})$ with entries in $\mathcal{O}_Z$ by

$$Fg_j-\sum_{i=1}^{k} a_{ij}g_i\in\mathscr{J} \qquad (j=1,\ldots,k).$$

This definition makes sense because $g_1,\ldots,g_k$ give an $\mathcal{O}_Z$-free base for $\mathcal{O}_C=\mathcal{O}_X/\mathscr{J}$. Then an $\mathcal{O}_Y$-free base $\theta_1,\ldots,\theta_k$ for $\mathrm{Der}_Y(\log D)$ is given by

$$\theta_j=y_1(\partial/\partial y_j)-\sum_{i=1}^{k} a_{ij}(y')(\partial/\partial y_i)$$

$(j=1,\ldots,k)$. A reduced defining equation for D is given by $|\theta_1\ldots\theta_k|:=\det(y_1 I-A)$.

We shall explicitly determine a base for $\mathrm{Der}(\log B)$. Define

$$\tilde{A}=(a_{ij})_{2\leqq i,j\leqq k},$$
$$e={}^t[a_{21},\ldots,a_{k1}].$$

Then

$$A=\left[\begin{array}{c|c} a_{11} & \cdots\ a_{1k} \\ \hline e & \tilde{A} \end{array}\right].$$

Put

$${}^t[b_{2i},\ldots,b_{ki}]=\tilde{A}^{i-2}e \qquad (i=2,\ldots,k).$$

Then we have

(4.1) **Theorem C.** *The vector fields*

$$\omega_i=\sum_{j=2}^{k} b_{ij}(\partial/\partial y_j) \qquad (i=2,\ldots,k)$$

give an $\mathcal{O}_Z$-free base for $\mathrm{Der}_Z(\log B)$.

Proof. Recall an $\mathcal{O}_Z$-isomorphism

$$\varepsilon:\mathrm{Der}_Z(\log B)=\mathrm{Im}(\Pi)\tilde{\to}\mathcal{O}_D/\mathcal{O}_Z$$

in (3.6). Then we know that, for $\varphi_2,\ldots,\varphi_k\in\mathrm{Der}_Y(\log D)\cap\mathscr{K}$, $\{\Pi(\varphi_2),\ldots,\Pi(\varphi_k)\}$ makes an $\mathcal{O}_Z$-free base for $\mathrm{Der}_Z(\log B)$ iff $\{[\varphi_2\cdot y_1],\ldots,[\varphi_k\cdot y_1]\}$ makes an $\mathcal{O}_Z$-free base for $\mathcal{O}_D/\mathcal{O}_Z$. Take $\theta_1=y_1(\partial/\partial y_1)-\sum_{j=2}^{k} a_{j1}(y')\,(\partial/\partial y_j)$. Define φ_i $(i=2,\ldots,k)$ inductively by

$$\varphi_2=\theta_1,$$
$$\varphi_i=y_1\varphi_{i-1}-\sum_{j=2}^{k}(\varphi_{i-1}\cdot y_j)\theta_j \qquad (i=3,\ldots,k-2).$$

Obviously $\varphi_2,\ldots,\varphi_k\in\mathrm{Der}_Y(\log D)$. For $2\leqq t\leqq k$, we have

$$\begin{aligned}\varphi_i\cdot y_t&=y_1(\varphi_{i-1}\cdot y_t)-\sum_{j=2}^{k}(\varphi_{i-1}\cdot y_j)(\theta_j\cdot y_t)\\&=y_1(\varphi_{i-1}\cdot y_t)-(\varphi_{i-1}\cdot y_t)(y_1-a_{tt})\\&\qquad-\sum_{\substack{j=2\\ j\neq t}}^{k}(\varphi_{i-1}\cdot y_j)(-a_{tj})\\&=\sum_{j=2}^{k}a_{tj}(\varphi_{i-1}\cdot y_j).\end{aligned}$$

Thus we can inductively prove that

$$\varphi_i\in\mathscr{K}\cap\mathrm{Der}_Y(\log D)$$

and that $\Pi(\varphi_i)=\omega_i (i=2,\ldots,k)$. Notice that

$$\begin{aligned}\varphi_i\cdot y_1 &= y_1(\varphi_{i-1}\cdot y_1)-\sum_{j=2}^{k}(\varphi_{i-1}\cdot y_j)a_{1j}\\ &\equiv y_1(\varphi_{i-1}\cdot y_1) \bmod \mathcal{O}_Z .\end{aligned}$$

This implies that

$$\begin{aligned}[\varphi_i\cdot y_1] &= [y_1^{i-2}(\varphi_2\cdot y_1)]\\ &= [y_1^{i-2}(y_1-a_{11})]\in \mathcal{O}_D/\mathcal{O}_Z\end{aligned}$$

and that $\{[\varphi_2\cdot y_1],\ldots,[\varphi_k\cdot y_1]\}$ makes an $\mathcal{O}_Z$-free base for $\mathcal{O}_D/\mathcal{O}_Z$. □

(4.2) **Corollary.** *A reduced defining equation for B is given by*

$$|\omega_2\ldots\omega_k| := \det(b_{ij})=\det[e,\tilde{A}e,\ldots,\tilde{A}^{k-2}e].$$

Proof. By Theorem C and Saito's criterion [6, (1.8)], $|\omega_2\ldots\omega_k|$ is a reduced defining equation. □

Fix a weight of the coordinates (x,y). Let $e_i\in\mathbf{Z}$, $e_i>0$, be the weight of $y_i (i=1,\ldots,k)$. In what follows, "homogeneous" always implies homogeneous according to this weight. When $g\in\mathbf{C}\{x,y\}$ is homogeneous, the degree of g is counted according to this weight.

Assume that $y_1-F(x,y')$ is homogeneous. Then

$$\deg(F)=e_1,\quad \deg(g_i)=e_1-e_i\quad (i=1,\ldots,k).$$

Since $\mathcal{I}$ is homogeneous (with respect to our weight), we can assume that

$$\deg(a_{ij})=e_1+e_j-e_i\quad (1\leqq i,j\leqq k).$$

A polynomial vector field $\theta=\sum_{i=1}^{k} f_i(y)(\partial/\partial y_i)$ is said to be homogeneous of degree n if

$$n=\deg f_i-\deg y_i=\deg f_i-e_i \quad\text{or}\quad f_i=0 \quad\text{for any } i,\ 1\leqq i\leqq k,$$

and is written as $n=\deg\theta$. Then $\deg\theta_i=e_1-e_i (i=1,\ldots,k)$. The set of integers $(\deg\theta_1,\ldots,\deg\theta_k)$ does not depend upon the choice of homogeneous free base for $\mathrm{Der}_Y(\log D)$. It is easy to see that

$$\deg|\theta_1\ldots\theta_k|=\sum_{i=1}^{k}\deg\theta_i+\sum_{i=1}^{k}e_i=ke_1.$$

Under the notation in (4.1), we have

$$\begin{aligned}&\deg\omega_2=\deg\varphi_2=\deg\theta_1=0,\\ &\deg\omega_i=\deg\varphi_i=e_1+\deg\varphi_{i-1}\\ &\qquad=e_1+\deg\omega_{i-1}\quad (i=3,\ldots,k-2).\end{aligned}$$

This implies that $\deg\omega_i=(i-2)e_1$ $(i=2,\ldots,k)$.

Therefore the degree of $|\omega_2 \ldots \omega_k|$ (4.2) is given by

$$\sum_{i=2}^{k} (\deg \omega_i) + \sum_{i=2}^{k} e_i = \sum_{i=2}^{k} (i-2)e_1 + \sum_{i=2}^{k} e_i$$
$$= \sum_{i=1}^{k} e_i + (k(k-3)e_1/2).$$

(4.3) *Example* (A_3-singularity)

$$m=1, \quad k=3, \quad F=x^4+y_3x^2+y_2x,$$
$$\mathcal{J}=(\partial F/\partial x)=(4x^3+2y_3x+y_2)\mathbf{C}\{x, y_2, y_3\},$$
$$g_1=1, \quad g_2=x, \quad g_3=x^2.$$

One can compute

$$Fg_1 \equiv (3y_2/4)g_2+(y_3/2)g_3,$$
$$Fg_2 \equiv (-y_2y_3/8)g_1+(-y_3^2/4)g_2+(3y_2/4)g_3,$$
$$Fg_3 \equiv (-3y_2^2/16)g_1+(-y_2y_3/2)g_2+(-y_3^2/4)g_3$$

modulo $\mathcal{J}$. Thus we have

$$A=\begin{bmatrix} 0 & -y_2y_3/8 & -3y_2^2/16 \\ 3y_2/4 & -y_3^2/4 & -y_2y_3/2 \\ y_3/2 & 3y_2/4 & -y_3^2/4 \end{bmatrix},$$

$$\tilde{A}=\begin{bmatrix} -y_3^2/4 & -y_2y_3/2 \\ 3y_2/4 & -y_3^2/4 \end{bmatrix}, \quad e=\begin{bmatrix} 3y_2/4 \\ y_3/2 \end{bmatrix},$$

$$\tilde{A}e=\begin{bmatrix} -7y_2y_3^2/16 \\ 9y_2^2/16-y_3^3/8 \end{bmatrix},$$

$$\omega_2=(3y_2/4)(\partial/\partial y_2)+(y_3/2)(\partial/\partial y_3),$$
$$\omega_3=(-7y_2y_3^2/16)(\partial/\partial y_2)+(9y_2^2/16-y_3^3/8)(\partial/\partial y_3).$$

By Theorem C, ω_2, ω_3 make a free base for $\mathrm{Der}_Z(\log B)$. A reduced equation for B is given by

$$|\omega_2\omega_3|=\det[e, \tilde{A}e]=y_2(27y_2^2+8y_3^3)/64.$$

Its degree is

$$\sum_{i=1}^{k} e_i+(k(k-3)e_1/2)=4+3+2+3(3-3)4/2=9$$

with respect to the weight corresponding x, y_1, y_2, y_3 to 1, 4, 3, 2, respectively.

(4.4) Considering Saito's conjecture [5] that the complement of a free divisor is $K(\pi, 1)$ space, it may be natural to ask.

Problem. Is the complement of bifurcation set (of a semiuniversal deformation of an isolated hypersurface singularity) $K(\pi, 1)$?

This has been positively answered for a simple singularity by Looijenga [3].

References

1. Arnol'd, V.I.: Indices of singular points of 1-forms on a manifold with boundary, convolution of invariants of reflection groups, and singular projections of smooth surfaces. Usp. Mat. Nauk **34**, 3–38 (1979); Russian Math. Surveys **34**, 1–42 (1979)
2. Cartier, P.: Les arrangements d'hyperplans: un chapitre de géometrie combinatoire. Séminaire Bourbaki 33e année, 1980/1981, n° 561. Lecture Notes in Mathematics, No. 901. Berlin, Heidelberg, New York: Springer 1982
3. Looijenga, E.: The complement of the bifurcation variety of a simple singularity. Invent. Math. **32**, 105–116 (1974)
4. Lyashko, O.V.: The geometry of bifurcation diagrams. Usp. Mat. Nauk **34**, 205–206 (1979); Russian Math. Surveys **34**, 209–210 (1979)
5. Saito, K.: On the uniformization of complements of discriminant loci. Symp. in Pure Math., Williams College, 1975. Several Complex variables. Providence: AMS 1977
6. Saito, K.: Theory of logarithmic differential forms and logarithmic vector fields. J. Fac. Sci. Univ. Tokyo Sect. IA **27**, 265–291 (1980)
7. Saito, K.: Primitive forms for an unfolding of a function with an isolated critical point. J. Fac. Sci. Univ. Tokyo Sect. IA **28**, 775–792 (1982)
8. Teissier, B.: The hunting of invariants in the geometry of discriminants. Real and complex singularities (Proc. Ninth Nordic Summer School/NAVF Sympos. Math., Oslo, 1976), pp. 565–678. Alphen aan den Rijn: Sijthoff and Noordhoff 1977
9. Terao, H.: Arrangements of hyperplanes and their freeness. I. J. Fac. Sci. Univ. Tokyo Sect. IA **27**, 293–312 (1980)
10. Terao, H.: Generalized exponents of a free arrangement of hyperplanes and Shephard-Todd-Brieskorn formula. Invent. Math. **63**, 159–179 (1981)
11. Terao, H.: Discriminant of a holomorphic map and logarithmic vector fields (to appear in J. Fac. Sci. Univ. Tokyo Sect. IA)

Received September 30, 1982

Math. Ann. 263, 323–341 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Invariance of Borel Classes in Metric Spaces

J. E. Jayne and C. A. Rogers
Department of Mathematics, University College London, Gower Street, London WC1E6BT, UK

1. Introduction

In a remarkable paper [15] J. Saint-Raymond has shown that, if p is a continuous map of a compact metric space L onto a compact metric space M, and C is a subset of M, with $B=p^{-1}(C)$ a Borel set of additive class α or of multiplicative class α, then C is a Borel set of the same class as B. This, of course, implies that, if C is of exact additive class α or of exact multiplicative class α, then $p^{-1}(C)$ is of the same exact class as C.

Our aim in this paper is to generalize and extend Saint-Raymond's result. In another paper [11] we apply the results obtained here to obtain the invariance of absolute Borel classes under maps of various types; we state some of these results at the end of this introduction.

We need to recall some definitions before we state our results. In a metric space, the Borel sets of additive class 0 are just the open sets, and the Borel sets of multiplicative class 0 are just the closed sets. For each ordinal α, with $0<\alpha<\omega_1$, the Borel sets of additive class α are just the countable unions of Borel sets of multiplicative class less than α, and the Borel sets of multiplicative class α are just the countable intersections of Borel sets of additive class less than α. The Borel sets of ambiguous class α are those that are simultaneously additive class α and multiplicative class α. In particular the sets of additive class 1 are the $\mathscr{F}_\sigma$-sets and the sets of multiplicative class 1 are the $\mathscr{G}_\delta$-sets. As in [8, 9] a map f of a metric space L into a metric space M will be called an $\mathscr{F}_\sigma$ map, if (a) f maps $\mathscr{F}_\sigma$-sets in L to $\mathscr{F}_\sigma$-sets in M, and (b) f^{-1} maps $\mathscr{F}_\sigma$-sets in M back to $\mathscr{F}_\sigma$-sets in L. A map f of L onto M will be said to be piece-wise closed, if L and M can be expressed in the forms

$$L=\bigcup_{n=1}^{\infty} L_n \quad \text{and} \quad M=\bigcup_{n=1}^{\infty} M_n,$$

with $L_1, L_2, \ldots$ closed in L, and $M_1, M_2, \ldots$ closed in M, and with the restriction of f to L_n a closed map of L_n onto M_n for each $n \geqq 1$. A set in a metric space N is said to be a Souslin-$\mathscr{F}$ set, if it has a representation in the form

$$\bigcup_{\sigma \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F(\sigma|n),$$

where each set $F(\sigma|n)$ is closed, $\mathbb{I}=\mathbb{N}^{\mathbb{N}}$ is the space of all infinite sequences

$$\sigma=\sigma_1,\sigma_2,\sigma_3,\ldots$$

of positive integers, and

$$\sigma|n=\sigma_1,\sigma_2,\ldots,\sigma_n.$$

A metric space N that is a Souslin-$\mathscr{F}$ set in its completion will be called an absolute Souslin-$\mathscr{F}$ set. It is clear that a piece-wise closed map between metric spaces is necessarily an $\mathscr{F}_\sigma$ map. We explain in [9] that, provided Martin's axiom holds but the continuum hypothesis is false, there will be $\mathscr{F}_\sigma$ maps that are not piece-wise closed. In [9], as corrected with the help of Hansell [6], we give the following alternative conditions that are sufficient to ensure that an $\mathscr{F}_\sigma$ map p of metric space L that is an absolute Souslin-$\mathscr{F}$ set onto a metric space M that is also an absolute Souslin-$\mathscr{F}$ set should be piece-wise closed:

(i) Fleissner's axiom $SC\omega_2$ (see [3, p. 322]) holds;

(ii) each point of L has a neighbourhood that is mapped by p onto a set in M that is σ-locally of weight at most $\aleph_1$;

(iii) p is an open map (i.e. maps open sets to open sets).

We can now state a theorem on the invariance of Borel classes.

Theorem 4. *Let p be a piece-wise closed map of a metric space L onto a metric space M. Let C be a subset of M with $B=p^{-1}(C)$ a Borel set of additive class α or of multiplicative class α, with $1\leqq\alpha<\omega_1$, then C is a Borel set of the same class as B.*

In order to state a more complicated result operating at a higher level of the Borel hierarchy we need further definitions. A function f is said [14, p. 174] to be an α-th level function, if $f^{-1}(E)$ is of additive class α whenever E is of additive class α. A function f will be said to preserve the α-th level, if f is an α-th level function and f maps sets of additive class α to sets of additive class α. In the special case when $\alpha=1$, such a map preserving the first level is an $\mathscr{F}_\sigma$ map.

A family $\{E_\xi:\xi\in\Xi\}$ of sets in a space L is said to be discrete if each point of L has a neighbourhood that meets E_ξ for at most one ξ in Ξ. A family $\{E_\xi:\xi\in\Xi\}$ is said to be σ-discrete if it is possible to write

$$\Xi=\bigcup_{n=1}^{\infty}\Xi(n)$$

with each family $\{E_\xi:\xi\in\Xi(n)\}$, $n=1,2,\ldots$ discrete. The family $\{F_\theta:\theta\in\Theta\}$ is said to be a refinement of the family $\{E_\xi:\xi\in\Xi\}$ if

$$\bigcup_{\theta\in\Theta}F_\theta=\bigcup_{\xi\in\Xi}E_\xi,$$

and, for each θ in Θ, there is a ξ in Ξ with $F_\theta\subset E_\xi$.

Theorem 3. *Let L and M be metric spaces. Let $1\leqq\alpha<\omega_1$. Suppose that p maps L onto M and that p preserves the α-th level. Suppose, further, that:*

(a) *p maps any discrete family of closed sets in L into a family in M having a σ-discrete refinement; and*

(b) *p^{-1} maps any point of M back to a set in L that is closed in the completion of L.*

Let C be a subset of M with $B=p^{-1}(C)$ a Borel set of additive class β or of multiplicative class β, with $\alpha \leqq \beta < \omega_1$. Then C is a Borel set of the same class as B.

We remark (see Lemma 3 below) that the condition (a) is satisfied by any map p that maps closed sets in L into closed sets in M.

Recall that a map is said to be of Borel class α, if the inverse image of each open set is of additive class α.

If p maps the space L onto the space M, a map s from M back to L, satisfying the condition

$$p \circ s(m) = m$$

for each m in M, is called a selection function for the set-valued function p^{-1}. Saint-Raymond obtains his result by first proving the existence of such a selection function of Borel class 1, satisfying a certain side condition. Our proofs work in a similar way, but make use of side conditions that seem to be more natural. We prove three such theorems.

Theorem 1. *Let L and M be metric spaces. Let $1 \leqq \alpha < \omega_1$, suppose that p maps L onto M and that p preserves the α-th level. Suppose, further, that:*

(a) *p maps any discrete family of closed sets in L into a family in M having a σ-discrete refinement; and*

(b) *p^{-1} maps any point of M back to a set in L that is closed in the completion of L.*

Let $\{D_\xi : \xi \in \Xi\}$ be a σ-discrete family of sets in L of additive class α. Then there is a selection function s for p^{-1}, of the α-th Borel class, such that

$$s^{-1}\left(\bigcup_{\theta \in \Theta} D_\theta\right)$$

is of additive class α in M, for each subset Θ of Ξ.

Theorem 2. *Let p be a piece-wise closed map of a metric space L onto a metric space M. Let $\{F_\xi : \xi \in \Xi\}$ be a σ-discrete family of $\mathscr{F}_\sigma$-sets in L. Then there is a selection function s for p^{-1}, of the first Borel class, such that*

$$s^{-1}\left(\bigcup_{\theta \in \Theta} F_\theta\right)$$

is an $\mathscr{F}_\sigma$-set in M for each subset Θ of Ξ.

As we have already remarked, these theorems have consequences for the theory of absolute Borel classes. Here we give a brief summary of results treated in detail in [11]; but note that Theorems 2 and 3 of [11] need the corrections outlined in [6]. A metric space is said to be a Borel set of absolute additive, multiplicative or ambiguous class α if it is of this class whenever it is topologically embedded in a metric space.

Theorem A. *Let f be a piece-wise closed map of a metric space X onto a metric space Y that is an absolute Souslin-$\mathscr{F}$ set. If X is of absolute additive Borel class α, with $1 \leqq \alpha < \omega_1$, or is of absolute multiplicative Borel class α, with $2 \leqq \alpha < \omega_1$, then Y is of the same absolute Borel class.*

Theorem B. *Let f be a piece-wise closed map of a metric space X that is an absolute Souslin-$\mathscr{F}$ set onto a metric space Y. Suppose that, either*

(ai) $1 \leqq \alpha < \omega_1$ *and* Y *is of absolute additive Borel class* α, *and*
(aii) $f^{-1}(y)$ *is of absolute additive Borel class* α *for each* y *in* Y, *or*
(bi) $2 \leqq \alpha < \omega_1$ *and* Y *is of absolute multiplicative Borel class* α, *and*
(bii) $f^{-1}(y)$ *is of absolute multiplicative Borel class* α *for each* y *in* Y.
Then, in each case, X *is of the same absolute additive or multiplicative Borel class as* Y.

If f is a proper map of a metric space X onto a metric space Y, then in Theorem A the condition that Y is an absolute Souslin-$\mathscr{F}$ set can be dropped, and in Theorem B the conditions (aii) and (bii) are automatically satisfied. In this case Theorems A and B reduce to especially simple forms. In [11] we also obtain rather more complicated results when f is a Borel measurable map of class β, $0 \leqq \beta < \omega_1$, and f either maps closed sets to closed sets or maps open sets to open sets. As we explain in [11], these results add precision to older results of Vaĭnšteĭn [17, 18], Taĭmanov [16], Čoban [1], and Freiwald [2] and they add generality to a recent result of Saint-Raymond [15]. In [11] we also give a sketch of the history of this type of result.

Saint-Raymond's paper [15] was motivated by his applications to the study of the Borel classes of certain affine functions. In Sect. 4 we discuss ways in which these applications can be generalized.

2. Selection Theorems

In order to prove Theorem 1 we need a generalization of the standard reduction theorem for countable unions of sets of additive class α (see [13, p. 350]). It follows from a proof of Kaniewski and Pol [12, Lemma 4, p. 1047]; it can also be obtained as a corollary to Theorem 1 of Hansell in [5]. Recall that Hansell [4] says that a family $\{E_\xi : \xi \in \Xi\}$ is discretely σ-decomposable, if

$$E_\xi = \bigcup_{n=1}^{\infty} E_\xi^{(n)},$$

for each ξ in Ξ, with each family $\{E_\xi^{(n)} : \xi \in \Xi\}$, $n = 1, 2, \ldots$, discrete.

Lemma 1. *Let* $1 \leqq \alpha < \omega_1$. *Let* $\{F_\xi : \xi \in \Xi\}$ *be a discretely* σ*-decomposable family of sets of additive class* α *with union* F *in a metric space. Then there is a disjoint* σ*-discrete family* $\{H_\phi : \phi \in \Phi\}$ *of sets of additive class* α, *refining the family* $\{F_\xi : \xi \in \Xi\}$, *and with*

$$\bigcup_{\theta \in \Theta} H_\theta$$

a set of additive class α, *for each subset* Θ *of* Φ.

It will be convenient to have a lemma concerning the representation of Borel sets as Souslin-$\mathscr{F}$ sets. In addition to the notation introduced in the introduction we use $\mathbb{I}(\tau|m)$ to denote the "Baire interval" of sequences σ in $\mathbb{I}$ with

$$\sigma|m = \tau|m.$$

Lemma 2. *In a metric space, a Borel set of additive or multiplicative class* α *has a representation*

$$B = \bigcup_{\sigma \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F(\sigma|n)$$

as a Souslin-$\mathscr{F}$ set, with each set $F(\sigma|n)$ closed and with each "fragment"

$$\bigcup_{\sigma \in \mathbb{I}(\tau|m)} \bigcap_{n \in \mathbb{N}} F(\sigma|n),$$

with $\tau \in \mathbb{I}$, $m \in \mathbb{N}$, of the corresponding additive or multiplicative class α.

Proof. If B is closed we just take

$$F(\sigma|n) = B$$

for all σ in $\mathbb{I}$ and n in $\mathbb{N}$.

Now suppose that B is open. For each $\varepsilon > 0$ and each x in the metric space let $B(x;\varepsilon)$ denote the open ball of points y within distance ε of x. For each sequence $\sigma_1, \sigma_2, \ldots, \sigma_n$ of positive integers write

$$G(\sigma_1, \sigma_2, \ldots, \sigma_n) = \left\{ x : \text{for some } \varepsilon > \frac{1}{\sigma_1} + \frac{1}{\sigma_2} + \ldots + \frac{1}{\sigma_n} \text{ we have } B(x;\varepsilon) \subset B \right\},$$

and

$$F(\sigma_1, \sigma_2, \ldots, \sigma_n) = \operatorname{cl} G(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

Clearly each set $G(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is open and each set $F(\sigma_1, \sigma_2, \ldots, \sigma_n)$ is closed. It is now easy to verify that

$$B = \bigcup_{\sigma \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F(\sigma|n)$$

and that

$$\bigcup_{\sigma \in \mathbb{I}(\tau|m)} \bigcap_{n \in \mathbb{N}} F(\sigma|n) = G(\tau|m)$$

for all $\tau \in \mathbb{I}$ and $m \in \mathbb{N}$. Thus the lemma holds for the class of open sets.

We now proceed by transfinite induction. Suppose that $1 \leqq \alpha < \omega_1$ and that the lemma has been proved for all β with $0 \leqq \beta < \alpha$. First consider the case when B is of multiplicative class α. Then

$$B = \bigcap_{l=0}^{\infty} B^{(l)},$$

with each set $B^{(l)}$ of additive class less than α. So we have

$$B^{(l)} = \bigcup_{\sigma \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F^{(l)}(\sigma|n),$$

with each set $F^{(l)}(\sigma|n)$ closed and with each set

$$\bigcup_{\sigma \in \mathbb{I}(q|k)} \bigcap_{n \in \mathbb{N}} F^{(l)}(\sigma|n)$$

of additive class less than α.

For each positive integer n there are unique integers $h(n) \geqq 0$, $t(n) \geqq 1$, and $q(n) \geqq 1$, satisfying the diophantine equations

$$n = [2q(n) - 1]t(n), \qquad t(n) = 2^{h(n)}.$$

Conversely, given any $h \geqq 0$ and $q \geqq 1$, there is a unique integer n, given by

$$n = [2q-1]2^h,$$

with $h(n) = h$ and $q(n) = q$.

For each σ in $\mathbb{I}$, define $\tau(h) = \tau_\sigma(h)$ in $\mathbb{I}$ by

$$\tau(h) = \sigma_t, \sigma_{3t}, \tau_{5t}, \ldots,$$

with $t = 2^h$. Then for each σ in $\mathbb{I}$ we have a corresponding sequence

$$T = T_\sigma = \tau(0), \tau(1), \tau(2), \ldots$$

in $\mathbb{I}^{\mathbb{N}}$, and the map taking σ to T is a homeomorphic map of $\mathbb{I}$ onto $\mathbb{I}^{\mathbb{N}}$.

Define closed sets by writing

$$F(\sigma|n) = F^{(h(n))}(\tau(h(n))|q(n)).$$

Then

$$\begin{aligned}
\bigcup_{\sigma\in\mathbb{I}} \bigcap_{n\in\mathbb{N}} F(\sigma|n) &= \bigcup_{\sigma\in\mathbb{I}} \bigcap_{n\in\mathbb{N}} F^{(h(n))}(\tau(h(n))|q(n)) \\
&= \bigcup_{\sigma\in\mathbb{I}} \bigcap_{h=0}^{\infty} \bigcap_{q=1}^{\infty} F^{(h)}(\tau(h)|q) \\
&= \bigcap_{h=0}^{\infty} \bigcup_{\tau(h)\in\mathbb{I}} \bigcap_{q=1}^{\infty} F^{(h)}(\tau(h)|q) \\
&= \bigcap_{h=0}^{\infty} B^{(h)} = B.
\end{aligned}$$

Similarly, if ϱ in $\mathbb{I}$ is given, write

$$\kappa(h) = \kappa_\varrho(h) = \varrho_t, \varrho_{3t}, \varrho_{5t}, \ldots,$$

with $t = 2^h$. Let

$$s(h) = s_k(h) = \max\{s : [2s-1]2^h \leqq k\}.$$

Then

$$\begin{aligned}
\bigcup_{\sigma\in\mathbb{I}(\varrho|k)} \bigcap_{n\in\mathbb{N}} F(\sigma|n) &= \bigcup_{\sigma\in\mathbb{I}(\varrho|k)} \bigcap_{n\in\mathbb{N}} F^{(h(n))}(\tau(h(n))|q(n)) \\
&= \bigcup_{\sigma\in\mathbb{I}(\varrho|k)} \bigcap_{h=0}^{\infty} \bigcap_{q=1}^{\infty} F^{(h)}(\tau(h)|q) \\
&= \bigcap_{h=0}^{\infty} \bigcap_{\tau(h)\in\mathbb{I}(\kappa(h)|s(h))} \bigcap_{q=1}^{\infty} F^{(h)}(\tau(h)|q),
\end{aligned}$$

and this set, being the countable intersection of sets of additive class less than α, is of multiplicative class α.

If B is of additive class α, then

$$B = \bigcup_{l=0}^{\infty} B^{(l)},$$

with each set $B^{(l)}$ of multiplicative class less than α. As before we have Souslin-$\mathscr{F}$ representations for these sets, but this time the fragments are of multiplicative class

less than α. Now write

$$F(\sigma|n) = F^{(h(\sigma_1))}(q(\sigma_1), \sigma_2, \ldots, \sigma_n),$$

with h and q defined as before. It is easy to verify that

$$B = \bigcup_{\sigma \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F(\sigma|n)$$

gives a suitable Souslin-$\mathscr{F}$ representation for B. Hence the result follows by induction.

Before we prove Theorem 1, generalizing Saint-Raymond's selection theorem, we remark that he based his proof on the principle of deferred choice; choices are deferred until the last possible opportunity. While we follow his methods in many respects, all our selection operations will be made by application of Lemma 1, and in this respect our proof follows the method of Kaniewski and Pol in their proof of their selection theorem [12, Theorem 2, p. 1048].

Proof of Theorem 1. As each set D of additive class α can be expressed as a countable union of sets of ambiguous class α it is clear that the general case of Theorem 1 will follow provided we establish the special case when each set of the family $\{D_\xi : \xi \in \Xi\}$ is of ambiguous class α. We confine our attention to this special case.

We can choose a σ-discrete family $\{V_\phi : \phi \in \Phi\}$ of open sets forming a base for the open sets of L. Further, each set V_ϕ, $\phi \in \Phi$, can be written as the union of a countable sequence of closed sets, and the family of these closed sets will be σ-discrete in L. By adjoining this family to the σ-discrete family $\{D_\xi : \xi \in \Xi\}$ of sets of ambiguous class α, if necessary, we may suppose that each set V_ϕ, $\phi \in \Phi$, is a countable union of closed sets chosen from the family $\{D_\xi : \xi \in \Xi\}$. Write

$$\Xi = \bigcup_{n=1}^{\infty} \Xi(n),$$

with each family $\{D_\xi : \xi \in \Xi(n)\}$, $n = 1, 2, \ldots$ discrete in L.

We say that a set A respects a set H if we have

$$\text{either} \quad A \cap H = A \quad \text{or} \quad A \cap H = \emptyset.$$

We say that a set A respects a family of sets, if it respects all the sets of the family.

We first prove a preliminary result. *Let A be a Borel set of the ambiguous class α in L, and suppose that $p(A)$ is of ambiguous class α in M. Let $\{C_n : n \geqq 1\}$ be a sequence of sets of additive class α, with union A. Let $\{S_\sigma : \sigma \in \Sigma(k)\}$ be a discrete family of sets of ambiguous class α in L, for $k = 1, 2, \ldots, n$. Let $\Sigma(1), \Sigma(2), \ldots, \Sigma(k)$ be disjoint and write*

$$\Sigma = \bigcup_{k=1}^{n} \Sigma(k).$$

Then, if $\varepsilon > 0$ is given, there is a family $\{A_\gamma : \gamma \in \Gamma\}$ with the properties:

(c) *$\{A_\gamma : \gamma \in \Gamma\}$ is a disjoint family of sets of ambiguous class α, each set A_γ being a subset of A and also a subset of one of the sets $\{C_n : n \geqq 1\}$;*

(d) *diam $A_\gamma < \varepsilon$, for all $\gamma \in \Gamma$;*

(e) *A_γ respects $\{S_\sigma : \sigma \in \Sigma\}$, for all γ in Γ; and*

(f) $\{p(A_\gamma):\gamma\in\Gamma\}$ *is a* σ*-discrete partition of* $p(A)$ *into sets of ambiguous class* α *in* M.

We construct a transfinite sequence

$$X_0, X_1, \ldots, X_\omega, \ldots, X_\beta, \ldots$$

of closed sets in L. Take $X_0 = L$. Once X_β has been constructed, we take $X_{\beta+1} = \emptyset$, if $X_\beta = \emptyset$, otherwise, we choose a point x_β in X_β, take U_β to be any open set in the base $\{V_\phi : \phi\in\Phi\}$, containing x_β, having diameter less than ε, and meeting at most one set from each of the discrete families $\{S_\sigma : \sigma\in\Sigma(k)\}$, $k=1,2,\ldots,n$, and define $X_{\beta+1}$ by

$$X_{\beta+1} = X_\beta \setminus U_\beta,$$

so that

$$X_\beta \setminus X_{\beta+1} = X_\beta \cap U_\beta.$$

This ensures that $X_{\beta+1}$ is a proper subset of X_β, when $X_\beta \neq \emptyset$. When λ is a limit ordinal, and X_β has been defined for $\beta<\lambda$, we take

$$X_\lambda = \bigcap_{\beta<\lambda} X_\beta.$$

This process defines a transfinite sequence with the following properties.

(g) $X_0 = L$.

(h) $\beta\leqq\gamma$ ensures $X_\beta \supset X_\gamma$.

(i) $X_{\beta+1} \neq X_\beta$, if $X_\beta \neq \emptyset$.

(j) $\operatorname{diam}(X_\alpha \setminus X_{\alpha+1}) \leqq \varepsilon$.

(k) $X_\beta \setminus X_{\beta+1}$ meets at most one set from each of the families $\{S_\sigma : \sigma\in\Sigma(k)\}$, $k=1,2,\ldots,n$.

(l) $X_\lambda = \bigcap_{\beta<\lambda} X_\beta$, when λ is a limit ordinal.

(m) $X_\beta = \emptyset$, for all sufficiently large β.

(n) $L = \bigcup_{\beta>0} X_\beta \setminus X_{\beta+1}$.

For each $\beta\geqq 0$, we can choose a finite subset $F(\beta)$ of Σ so that S_σ meets U_β, if, and only if, $\sigma\in F(\beta)$. For each partition $\{P, N\}$ of $F(\beta)$ into two parts, the set

$$U_\beta \cap \left(\bigcap_{\sigma\in P} S_\sigma\right) \cap \left(\bigcap_{\sigma\in N} L\setminus S_\sigma\right)$$

is a set of ambiguous class α respecting the family $\{S_\sigma : \sigma\in\Sigma\}$. Let $O_{\beta,q}$, $q=1,2,\ldots,n(\beta)$ be an enumeration of the non-empty sets of this form. Then $\{O_{\beta,q} : 1\leqq q\leqq(\beta)\}$ is a partition of U_β. Note that the same set V_ϕ, $\phi\in\Phi$, of the base cannot be taken twice to be a set U_β, $\beta\geqq 0$. As the base is σ-discrete, so is the family

$$\{X_\beta \cap U_\beta : \beta\geqq 0\}.$$

Hence the family

$$\{(X_\beta \setminus X_{\beta+1}) \cap O_{\beta,q} : \beta\geqq 0 \text{ and } 1\leqq q\leqq n(\beta)\}$$

is a σ-discrete partition of L into sets of ambiguous class α that respect the family $\{S_\sigma : \sigma\in\Sigma\}$.

By the standard reduction theorem, as the countable family $\{C_n : n \geqq 1\}$ of additive class α sets has as union the ambiguous class α set A, we can choose a disjoint family $\{E_n : n \geqq 1\}$ of ambiguous class α sets with

$$E_n \subset C_n, \quad \text{for} \quad n \geqq 1,$$

and

$$\bigcup_{n=1}^{\infty} E_n = A.$$

So the family

$$\{E_k \cap (X_\beta \backslash X_{\beta+1}) \cap O_{\beta,q} : k \geqq 1,\ \beta \geqq 0 \text{ and } 1 \leqq q \leqq n(\beta)\}$$

is a σ-discrete partition of A into sets of ambiguous class α that respects the family $\{S_\sigma : \sigma \in \Sigma\}$ and this partition refines the family $\{C_n : n \geqq 1\}$. For convenience, we re-index this family as

$$\{Q_\delta : \delta \in \Delta\}.$$

For each δ in Δ, we write

$$P_\delta = p(Q_\delta).$$

As $\bigcup_{\delta \in \Delta} Q_\delta = A$, we see that $\{P_\delta : \delta \in \Delta\}$ has as union the set $p(A)$ of ambiguous class α. By our hypothesis (a), as the family $\{\mathrm{cl}\, Q_\delta : \delta \in \Delta\}$ is a σ-discrete family of closed sets, the family $\{p(\mathrm{cl}\, Q_\delta) : \delta \in \Delta\}$, and so also the family $\{P_\delta : \delta \in \Delta\}$, has a σ-discrete refinement. Let $\{R_\lambda : \lambda \in \Lambda\}$ be a σ-discrete refinement of $\{P_\delta : \delta \in \Delta\}$. For each λ in Λ, let $\delta(\lambda)$ be in Δ with $R_\lambda \subset P_{\delta(\lambda)}$. Write $\Lambda = \bigcup_{n=1}^{\infty} \Lambda(n)$ with each family

$$\{R_\lambda : \lambda \in \Lambda(n)\},$$

$n = 1, 2, \ldots$, discrete. Then

$$\{\mathrm{cl}\, R_\lambda : \lambda \in \Lambda(n)\},$$

is also discrete for $n = 1, 2, \ldots$. Hence, as p preserves the α-th level,

$$\{(\mathrm{cl}\, R_\lambda) \cap P_{\delta(\lambda)} : \lambda \in \Lambda\}$$

is a σ-discrete refinement of $\{P_\delta : \delta \in \Delta\}$ by sets of additive class α. By Lemma 1, there is a σ-discrete partition of $p(A)$ into a family $\{T_\gamma : \gamma \in \Gamma\}$ of sets of ambiguous class α, refining $\{P_\delta : \delta \in \Delta\}$. For each γ in Γ choose $\delta(\gamma)$ in Δ with $T_\gamma \subset P_{\delta(\gamma)}$. Now define the family $\{A_\gamma : \gamma \in \Gamma\}$ by taking

$$A_\gamma = Q_{\delta(\gamma)} \cap p^{-1}(T_\gamma).$$

As p is an α-th level function, it is easy to verify that $\{A_\gamma : \gamma \in \Gamma\}$ satisfies the requirements (c), (d), (e), and (f).

We now give an inductive construction of a sequence of families

$$\{A(\gamma|k) : \gamma|k \in \Pi(k)\} \qquad k = 1, 2, \ldots,$$

of subsets of L, indexed by multi-indices

$$\gamma|k = \gamma_1, \gamma_2, \ldots, \gamma_k,$$

with $\gamma_1, \gamma_2, \ldots, \gamma_k$ in certain index sets $\Gamma(1), \Gamma(2), \ldots, \Gamma(k)$, so that $\gamma|k$ lies in the cartesian product

$$\Pi(k) = \prod_{i=1}^{k} \Gamma(i),$$

$k = 1, 2, \ldots$. Here the index sets $\Gamma(1), \Gamma(2), \ldots$ will themselves be constructed inductively. At the same time we also construct a double sequence of families

$$\{\tau(h; \gamma|k)|(k-1) : \gamma|k \in \Pi(k)\}, \qquad 1 \leqq h \leqq k-1,$$

of finite sequences

$$\tau(h; \gamma|k)|(k-1) = \tau_1(h; \gamma|k), \tau_2(h; \gamma|k), \ldots, \tau_{k-1}(h; \gamma|k),$$

of positive integers. We use ϕ to denote the "empty sequence" of zero length. It will also be convenient to write $\Pi(0) = \{\phi\}$. We shall satisfy the following conditions.

(o) $A(\phi) = L$.

(p) For $k \geqq 1$ and $\gamma|k$ in $\Pi(k)$ the set $A(\gamma|k)$ has a Souslin-$\mathscr{F}$ representation

$$A(\gamma|k) = \bigcup_{\tau \in \mathbb{I}} \bigcap_{n \in \mathbb{N}} F(\gamma|k; \sigma|n),$$

with each set $F(\gamma|k; \sigma|n)$ closed in L and with each fragment

$$A(\gamma|k; \tau|m) = \bigcup_{\sigma \in \mathbb{I}(\tau|m)} \bigcap_{n \in \mathbb{N}} F(\gamma|k; \sigma|n)$$

a set of additive class α in L.

(q) For $k \geqq 1$ and

$$\gamma|(k-1) = \gamma_1, \gamma_2, \ldots, \gamma_{k-1} \Pi(k-1),$$

with $\gamma|0 = \phi$ in the case $k = 1$, the subfamily

$$\{A(\gamma|k) : \gamma_k \in \Gamma(k)\}$$

is a disjoint family of sets of ambiguous class α, each set $A(\gamma|k)$ being contained both in $A(\gamma|(k-1))$ and also in

$$\bigcap_{1 \leqq h < k} A(\gamma|h; \tau(h; \gamma|k)|(k-1)),$$

when $k \geqq 2$.

(r) For $k \geqq 1$ and $\gamma|k \in \Pi(k)$, we have

$$\operatorname{diam} A(\gamma|k) < 1/k.$$

(s) For $k \geqq 1$ and $\gamma|k \in \Pi(k)$, the set $A(\gamma|k)$ respects the family

$$\left\{D_\xi : \xi \in \bigcup_{h=1}^{k} \Xi(h)\right\}.$$

(t) For each $k \geqq 1$, the family

$$\{p(A(\gamma|k)) : \gamma|k \in \Pi(k)\}$$

is a σ-discrete partition of M into sets of ambiguous class α.

(u) For $k>l\geqq h\geqq 1$ and $\gamma|k\in\Pi(k)$, $\tau_1(h;\gamma|l)=\tau_1(h;\gamma|k),\ldots,\tau_l(h;\gamma|l)=\tau_l(h;\gamma|k)$.

The construction is initiated by writing $A(\phi)=L$. Note that the conditions (o) to (u) are satisfied trivially when $k=0$; they will need special care in the case $k=1$. Suppose that $k\geqq 0$ and that the index sets and families have been constructed for this value of k and all smaller values, satisfying the conditions (o) to (u). We define the family

$$\{A(\gamma|(k+1)):\gamma|(k+1)\in\Pi(k+1)\}$$

by defining the family

$$\{A(\gamma|(k+1)):\gamma_{k+1}\in\Gamma(k+1)\}$$

separately, for each choice of $\gamma|k$ in $\Pi(k)$.

When $k\geqq 1$ in the italicized result proved above, we take $A=A(\gamma|k)$, we take $\varepsilon=1/(k+1)$, we take $\varepsilon=1/(k+1)$, we take $\{C_n:n\geqq 1\}$ to be an enumeration of the countable family

$$\Big\{A(\gamma|k;\varrho(1),\varrho(2),\ldots,\varrho(k))\cap\bigcap_{1\leqq h<k}A(\gamma|h;\tau(h;\gamma|k)|(k-1),\sigma(h)):$$
$$\varrho(1),\varrho(2),\ldots,\varrho(k),\sigma(1),\sigma(2),\ldots,\sigma(k-1)\in\mathbb{N}\Big\},\tag{1}$$

of sets of additive class α, with union $A(\gamma|k)$, and we take $\{S_\sigma:\sigma\in\Sigma\}$ to be the family

$$\Big\{D_\xi:\xi\in\bigcup_{h=1}^{k}\Xi(h)\Big\}.$$

In the initial case $k=0$, we take $A=A(\phi)=L$, we take $\varepsilon=1$, we take $C_n=L$ for $n\geqq 1$, and we take $\{S_\sigma:\sigma\in\Sigma\}$ to be the empty family. Let

$$\{B(\gamma|k;\lambda):\lambda\in\Lambda(\gamma|k)\}$$

be the family, provided by the italicized result, and satisfying the conditions (c) to (f). We may suppose that the index sets $\Lambda(\gamma|k)$, $\gamma|k\in\Pi(k)$, are all disjoint. Write

$$\Gamma(k+1)=\bigcup_{\gamma|k\in\Pi(k)}\Lambda(\gamma|k),\qquad \Pi(k+1)=\prod_{h=1}^{k+1}\Gamma(h),$$

and take

$$A(\gamma|(k+1))=B(\gamma|k;\gamma_{k+1})$$

if $\gamma_{k+1}\in\Lambda(\gamma|k)$, and take

$$A(\gamma|(k+1))=\emptyset,$$

otherwise. Each set $A(\gamma|(k+1))$ with $\gamma|(k+1)\in\Pi(k+1)$ is now of ambiguous class α and so has a Souslin-$\mathscr{F}$ representation in the form required for condition (p) by Lemma 1. As each member of the family

$$\{A(\gamma|k,\gamma_{k+1}):\gamma_{k+1}\in\Gamma(k+1)\}$$

lies in at least one member of the family (1), we can choose the k-th component of

$\tau(h;\gamma|(k+1))|k$, for $1 \leqq h < k$, and the k components of $\tau(k;\gamma|(k+1)|k$ so that (q) holds. It is now easy to verify that all the conditions (p) to (u) are satisfied, and the inductive construction is complete.

We can now introduce the index set

$$\Gamma = \prod_{k=1}^{\infty} \Gamma(k),$$

and, for γ in Γ, we can use $\gamma|k$ to denote the sequence composed of the first k components $\gamma_1, \gamma_2, \ldots, \gamma_k$ of γ. Write

$$N(k) = \bigcup_{\gamma \in \Gamma} A(\gamma|k)$$

for each $k \geqq 0$, and

$$N = \bigcap_{k=1}^{\infty} N(k).$$

Then

$$L = N(0) \supset N(1) \supset \ldots \supset N(k) \supset \ldots \supset N.$$

By condition (t),

$$p(N(k)) = M,$$

for $k \geqq 0$, and further, for each point y of M, there is a unique sequence $\gamma(y)$ in Γ with

$$y \in p(A(\gamma(y)|k)), \qquad k = 1, 2, \ldots .$$

So we can choose $x(y,k)$ in $p^{-1}(y) \cap A(\gamma(y)|k)$ for each $k \geqq 1$. As $\operatorname{diam} A(\gamma(y)|k)$ is at most $1/k$, and the sequence

$$A(\gamma(y)|k), \qquad k = 1, 2, \ldots,$$

is decreasing, the sequence $x(y,k)$, $k = 1, 2, \ldots$ is a Cauchy sequence in $p^{-1}(y)$ which is closed in the completion of X. So there is a point $s(y)$ in $p^{-1}(y)$ with

$$s(y) = \lim_{k \to \infty} x(y,k).$$

As $s(y) \in p^{-1}(y)$ we have

$$p \circ s(y) = y$$

for all y in M.

Using the consistency conditions (u), for each γ in Γ there will be sequences

$$\tau(1;\gamma), \tau(2;\gamma), \ldots$$

in $\mathbb{I}$ with

$$\tau(h;\gamma|k)|(k-1) = \tau(h;\gamma)|(k-1),$$

for $k > h \geqq 1$. Using the condition (q), we have, for $l > k > h$,

$$\begin{aligned} x(y,l) &\in A(\gamma|y)|l) \\ &\subset A(\gamma|y)|h;\ \tau(h;\gamma(y)|l)|(l-1)) \\ &\subset A(\gamma(y)|h;\ \tau(h;\gamma(y))|k) \\ &\subset F(\gamma(y)|h;\ \tau(h;\gamma(y))|k). \end{aligned}$$

As $F(\gamma(y)|h;\ \tau(h;\gamma(y))|k)$ is closed, on letting l tend to infinity we obtain

$$s(y) \in F(\gamma(y)|h;\ \tau(h;\gamma(y))|k),$$

for $k > h$. Letting k tend to infinity we obtain

$$\begin{aligned} s(y) &\in \bigcap_{k=1}^{\infty} F(\gamma(y)|h;\ \tau(h;\gamma(y))|k) \\ &\subset A(\gamma(y)|h) \end{aligned}$$

for $h \geqq 1$. Hence

$$\{s(y)\} = \bigcap_{h=1}^{\infty} A(\gamma(y)|h).$$

Thus, for each y in M, there is an $s(y)$ in $N = \bigcap_{k=1}^{\infty} N(k)$ with $p \circ s(y) = y$. As the families $\{p(A(\gamma|k)) : \gamma \in \Gamma\}$, $k = 1, 2, \ldots$ are disjoint there is just the one point $s(y)$ in N that is mapped onto y by p. On the other hand, given any x in N, as

$$p(x) \in p(A(\gamma(p(x))|k)),$$

for $k \geqq 1$, and the families $\{A(\gamma|k) : \gamma \in \Gamma\}$, $k = 1, 2, \ldots$, are disjoint, we have

$$x \in A(\gamma(p(x))|k),$$

for $k \geqq 1$, and so $x = s(p(x))$. Hence s maps M bijectively onto N.

Now let Θ be any subset of Ξ and write

$$\Theta(n) = \Theta \cap \Xi(n),$$

for $n \geqq 1$. Write

$$D^{(k)} = \bigcup_{\theta \in \Theta} D_\theta,$$

for $k \geqq 1$. Let $\Delta(k)$ denote the set of those δ in Γ with

$$D^{(k)} \cap A(\delta|k) \neq \emptyset.$$

For each δ in $\Delta(k)$, the set $A(\delta|k)$ respects the disjoint family

$$\{D_\theta : \theta \in \Theta(k)\}.$$

So $A(\delta|k)$ is contained in any set of this family that it meets.

Hence

$$\begin{aligned} D^{(k)} \cap N_k &= \left\{ \bigcup_{\theta \in \Theta(k)} D_\theta \right\} \cap \left\{ \bigcup_{\delta \in \Gamma} A(\delta|k) \right\} \\ &= \bigcup_{\delta \in \Delta(k)} A(\delta|k). \end{aligned}$$

Now

$$s^{-1}(D^{(k)}) = s^{-1}(D^{(k)} \cap N_k) = \bigcup_{\delta \in \Delta(k)} A(\delta|k).$$

As the family $\{p(A(\delta|k)) : \delta \in \Gamma\}$ is a σ-discrete family of sets of ambiguous class α, for each $k \geqq 1$, the set

$$s^{-1}\left(\bigcup_{\theta \in \Theta} D_\theta\right) = \bigcup_{k=1}^{\infty} \bigcup_{\delta \in \Delta(k)} p(A(\delta|k))$$

is of additive class α.

By our adjustment to the family $\{D_\xi : \xi \in \Xi\}$ to ensure that each set V_ϕ, $\phi \in \Phi$ is a countable union of sets chosen from $\{D_\xi : \xi \in \Xi\}$, we see that each open set G in X is of the form

$$G = \bigcup_{\theta \in \Theta} D_\theta$$

for some subset Θ of Ξ, so that $s^{-1}(G)$ is a set of additive class α in M. Hence s is a map of Borel class α. This completes the proof of Theorem 1.

We next need a lemma (cf. Hansell [19, Proposition 3.10]) showing that maps that carry closed sets to closed sets satisfy the condition (a) of Theorem 1.

Lemma 3. *Let L and M be metric spaces and let p be a map of L onto M. Suppose that p maps closed sets in L into closed sets in M. Then p maps any discrete family of sets in L into a family of sets in M having a σ-discrete refinement.*

Proof. Let $\{A_\gamma : \gamma \in \Gamma\}$ be a discrete family of sets in L. Then the family $\{\mathrm{cl}\, A_\gamma : \gamma \in \Gamma\}$ is also discrete in L. Let D be the set of points d of M for which $d \in p(\mathrm{cl}\, A)$ for infinitely many γ in Γ. We prove that D is discrete in M. Suppose that $d_1, d_2, \ldots$ was sequence of distinct points of D converging to a point m of M distinct from each of $d_1, d_2, \ldots$. Then we could choose a sequence $\gamma(1), \gamma(2), \ldots$ of distinct elements of Γ, inductively, with

$$d_i \in p(\mathrm{cl}\, A_{\gamma(i)}), \quad i = 1, 2, \ldots.$$

Then, for suitable points $x_1, x_2, \ldots$ in L, we would have

$$x_i \in \mathrm{cl}\, A_{\gamma(i)}, \quad p(x_i) = d_i, \quad i = 1, 2, \ldots.$$

As the family $\{\mathrm{cl}\, A_{\gamma(i)} : i \geqq 1\}$ would be closed in L. So the set

$$\{p(x_1), p(x_2), \ldots\} = \{d_1, d_2, \ldots\}$$

would be closed in M. This is not the case, as m is not in this set, but is one of its limit points. Thus D is discrete in M. In particular, D is closed in M.

For each point d in $D' = D \cap B$, with

$$B = \bigcup_{\gamma \in \Gamma} p(A_\gamma),$$

we choose one element $\gamma(d)$ of Γ with $d \in p(A_{\gamma(d)})$. This done, define

$$B_\gamma^{(0)} = \{d : \gamma(d) = \gamma\}.$$

Then the family

$$\{B_\gamma^{(0)} : \gamma \in \Gamma\}$$

is a discrete family of subsets of D', refining the family

$$\{D \cap p(A_\gamma) : \gamma \in \Gamma\}.$$

Now the open set $E = M \setminus D$ is the set of points e of M for which $e \in p(\operatorname{cl} A_\gamma)$ for only finitely many γ in Γ. Hence

$$E = \bigcup_{\Phi \subset \Gamma} \left\{ M \setminus \bigcup_{\gamma \in \Gamma \setminus \Phi} p(\operatorname{cl} A_\gamma) \right\},$$

the union being taken over all finite subsets Φ of Γ. Now, for each finite set Φ in Γ,

$$\bigcup_{\gamma \in \Gamma \setminus \Phi} p(\operatorname{cl} A_\gamma) = p\left(\bigcup_{\gamma \in \Gamma \setminus \Phi} \operatorname{cl} A_\gamma \right),$$

and so is closed in M. Thus, for each finite set Φ in Γ, the set

$$M \setminus \bigcup_{\gamma \in \Gamma \setminus \phi} p(\operatorname{cl} A_\gamma)$$

is open. Choose a σ-discrete open refinement $\{U_\xi : \xi \in \Xi\}$ for the family

$$\left\{ M \setminus \bigcup_{\gamma \in \Gamma \setminus \Phi} p(\operatorname{cl} A_\gamma) : \Phi \text{ finite in } \Gamma \right\}$$

of open sets with union E. For each ξ in Ξ choose a finite set

$$\Phi(\xi) = \{\phi(1, \xi), \phi(2, \xi), \ldots, \phi(f(\xi), \xi)\}$$

in Γ such that

$$U_\xi \cap B \subset \bigcup_{\phi \in \Phi(\xi)} p(A_\phi).$$

Write

$$V_\xi^{(r)} = U_\xi \cap p(A_{\phi(r,\xi)}), \quad \text{if} \quad 1 \leqq r \leqq f(\xi),$$
$$V_\xi^{(r)} = \emptyset, \quad \text{if} \quad f(\xi) < r,$$

for each ξ in Ξ. Then

$$\{V_\xi^{(r)} : 1 \leqq r, \xi \in \Xi\}$$

is a σ-discrete refinement of the family

$$\{p(A_\gamma) \cap E : \gamma \in \Gamma\}.$$

Hence the union of the families

$$\{B_\gamma^{(0)} : \gamma \in \Gamma\}$$

and

$$\{V_\xi^{(r)} : 1 \leqq r, \xi \in \Xi\}$$

forms a σ-discrete refinement of the family $\{p(A_\gamma) : \gamma \in \Gamma\}$.

To enable us to obtain Theorem 2 from Theorem 1 and Lemma 3 we quote part of one of our recent results [9, Theorem 3] derived immediately from work of Vaĭnšteĭn [17, 18], as our next lemma.

Lemma 4. *Let p be a piece-wise closed map of a metric space L onto a metric space M. Then L has a subset L_∞, with $p(L_\infty)=M$ and with*

$$L_\infty = \bigcup_{n=1}^{\infty} L_n,$$

and

$$L_1 \subset L_2 \subset \ldots \subset L_n \subset \ldots,$$

and the restriction of p to L_n is a proper map of L_n onto a closed set $p(L_n)$ in M, for $n \geqq 1$.

Proof of Theorem 2. Let $L_\infty, L_1, L_2, \ldots$, be the $\mathscr{F}_\sigma$-set and the closed sets provided by Lemma 4. For each $n \geqq 1$, Lemma 3 shows that the restriction of p to L_n satisfies the conditions of Theorem 1, with $L=L_n$, $M=p(L_n)$, and $x=1$. In Theorem 1, we take $\{D_\xi : \xi \in \Xi\}$ to be the σ-discrete family $\{F_\xi \cap L_n : \xi \in \Xi\}$ of closed sets; let s_n be the selection function defined on $p(L_n)$ by this theorem. Define s on M by

$$s(y)=s_n(y), \quad \text{if} \quad y \in p(L_n) \setminus p(L_{n-1}),$$

with the convention that $L_0=\emptyset$. It is easy to verify that s has the required properties.

3. The Invariance of Exact Borel Classes under Inverse Image Maps

We give the proofs for Theorems 3 and 4 stated in the introduction.

Proof of Theorem 3. First suppose that $\alpha=\beta$. If B is of additive class α, the fact that p preserves the α-th level, ensures that $p(B)$ is also of additive class α. If B is of multiplicative class α, then $p(L \setminus B)=p(L \setminus p^{-1}(C))=M \setminus C$ is of additive class α, and so C is of multiplicative class α.

Now suppose that $\alpha<\beta$, and that $B=p^{-1}(C)$ is of additive or multiplicative class β. We can regard B as being constructed by a transfinite sequence of steps, indexed by the ordinals η with $0 \leqq \eta \leqq \beta$, as follows. The "zero-th" step consists of a listing (as an ordinary sequence of type ω) all the zero-th class Borel sets; i.e., all the open sets and all the closed sets, that are to be used at any stage of the construction of B. When $1 \leqq \eta < \beta$, the η-th stage consists of constructing, from sets listed at previous stages, those Borel sets of class η that will be used at any later stage of the construction, and then listing these newly constructed sets. The final β-th stage consists of constructing the final set of additive class β as a countable union of previously constructed sets, or of constructing the final set of multiplicative class β as a countable intersection of previously constructed sets.

Now at the α-th stage of the construction of B we have a list of certain sets of additive class α and certain sets of multiplicative class α. Let $B_1, B_2, \ldots$ be a

sequence including all the sets listed at the previous η-th stages, $0 \leqq \eta < \alpha$, together with all the additive class α sets arising at the α-th stage, and all the complements of all the multiplicative class α sets arising at the α-th stage. Then all the sets $B_1, B_2, \ldots$ are of additive class α, and so can be expressed as countable unions of sets of ambiguous class α. Let $D_1, D_2, \ldots$ be a list of all these component sets of ambiguous class α. We now apply Theorem 1 to the function p from L to M and to the σ-discrete family $\{D_\xi : \xi \in \Xi\} = \{D_n : n \in \mathbb{N}\}$.

Let s be the selection function provided by Theorem 1. Then, for each n in $\mathbb{N}$, $s^{-1}(D_n)$ is of additive class α. Hence $s^{-1}(B_n)$ is of the same additive or multiplicative class α as is B_n, for $n \in \mathbb{N}$. Now a listing of these sets of additive and multiplicative class α, can be constructed in M by using a sequence of stages indexed by the ordinals η with $0 \leqq \eta \leqq \alpha$. Provided the subsequent stages with $\alpha < \eta \leqq \beta$ of the construction of B in L from the sets listed up to and including the α-th stage are exactly copied in M, the result at the β-th stage will be precisely the set

$$s^{-1}(B) = C.$$

Hence C is of the corresponding additive or multiplicative class β, as required.

Proof of Theorem 4. This follows from Theorem 3 just as Theorem 2 follows from Theorem 1.

4. Applications to Affine Functions

In [15] Saint-Raymond gives the following theorems (his Corollaries 7 and 8) as applications of his results.

Theorem A. *Let V be a locally convex Hausdorff topological vector space and let K_1 and K_2 be two compact convex metrizable subsets of V. Let K be the least convex cover of K_1 and K_2 and let f be an affine form defined on K. If the restrictions of f to K_1 and K_2 are both of Borel class α, then f is also of Borel class α.*

Theorem B. *Let V be a locally convex Hausdorff topological vector space and let K be a compact convex metrizable subset of V. Let f be a Borel measurable affine form defined on K and satisfying the barycentric calculus on K. If the restriction of f to the closure of the set of extreme points of K is of Borel class α, then f is also of Borel class α on K.*

Our results enable us to obtain a result like Theorem A when one of the sets K_1, K_2 is compact and the other is closed.

Theorem A1. *Let V be a locally convex Hausdorff topological vector space. Let K be the least convex cover of two convex sets K_1 and K_2 in V, with K_1 compact and K_2 closed in K. Suppose that the topology on K is metrizable. Let f be an affine form defined on K. If the restrictions of f to K_1 and to K_2 are both of Borel class α with $\alpha \geqq 1$, then f is also of Borel class α.*

We have not obtained any corresponding generalization of Theorem B.

Proof of Theorem A1. Write

$$L = [0, 1) \times K_1 \times K_2.$$

Define maps:

$$p:L\to K, \quad \text{by} \quad p(t,x,y,)=tx+(1-t)y;$$

$$\phi:L\to\mathbb{R}, \quad \text{by} \quad \phi=f\circ p;$$

$$F:L\to[0,1)\times\mathbb{R}\times\mathbb{R}, \quad \text{by} \quad F(t,x,y)=(t,f(x),f(y));$$

and

$$r:[0,1)\times\mathbb{R}\times\mathbb{R}\to\mathbb{R}, \quad \text{by} \quad r(t,\xi,\eta)=t\xi+(1-t)\eta.$$

Write

$$K^{(1)}=K_1, \qquad K^{(2)}=p(L),$$

and let $f^{(1)}$, $f^{(2)}$ denote the restrictions of f to $K^{(1)}$ and to $K^{(2)}$. Of course, $f^{(1)}$ and $f^{(2)}$ agree on the intersection of their domains of definition. As $f^{(1)}$ is of class α on the compact set $K^{(1)}$ and $\alpha\geqq 1$, it will suffice to prove that $K^{(2)}$ is an $\mathscr{F}_\sigma$-set in K and that $f^{(2)}$ is of class α on $K^{(2)}$.

As f is affine, we have

$$\phi=f\circ p=r\circ F$$

on L. As F is of Borel class α on L and as r is continuous, we see that ϕ is of Borel class α on L. As $(f^{(2)})^{-1}=p\circ\phi^{-1}$ on $\mathbb{R}$, the map $f^{(2)}$ will be of Borel class α on $K^{(2)}$, provided p maps Borel sets of additive class α of the form $p^{-1}(B)$ with B in $K^{(2)}$, into Borel sets of additive class α. By Theorem 4, above, it suffices to prove that p is a piece-wise closed map of L to $K^{(2)}$ and that $K^{(2)}$ is an $\mathscr{F}_\sigma$-set in K.

Write

$$L=\bigcup_{n=1}^{\infty} L_n,$$

with

$$L_n=[0,n/(n+1)]\times K_1\times K_2,$$

for $n=1,2,\ldots$. Then each set L_n is closed in L. Let H be any closed set in L. Suppose that $n\geqq 1$, and that x is any limit point of $p(L_n\cap H)$ in K. Then there is a sequence $\{(\tau_m,\xi_m,\eta_m)\}_{m=1}^{\infty}$ of points in $L_n\cap H$, with

$$p(\tau_m,\xi_m,\eta_m)\to x \quad \text{as} \quad m\to\infty.$$

As K_1 is compact, we may pass to a subsequent and ensure that

$$\tau_m\to\tau, \quad \text{as} \quad m\to\infty,$$

and

$$\xi_m\to\xi, \quad \text{as} \quad m\to\infty,$$

with $0\leqq\tau\leqq n/(n+1)$ and $\xi\in K_1$. Now

$$\eta_m=(1-\tau_m)^{-1}[p(\tau_m,\xi_m,\eta_m)-\tau_m\xi_m]$$

for $m\geqq 1$, so that

$$\eta_m\to\eta=(1-\tau)^{-1}[x-\tau\xi]$$

as $m \to \infty$. Now $(\tau, \xi, \eta) \in L_n \cap H$ as K_2 is closed in K and so $x = p(\tau, \xi, \eta)$ is in $p(L_n \cap H)$. Thus $p(L_n \cap H)$ is closed. Hence p is piece-wise closed on L, and further, $K^{(2)}$ is an $\mathscr{F}_\sigma$-set in K. This completes the proof.

Theorems A and B can be generalized by relaxing the metrizability conditions. In [14] we use a 'bootstrapping' operation to deduce a version of one of Saint-Raymond's main results (his Theorem 5), that works with Baire classes but applies to compact Hausdorff spaces, from his original result; see [14, Theorem 5.9.13, p. 153]. Using this result together with Saint-Raymond's arguments we obtain the following generalizations.

Theorem A2. *Theorem A continues to hold when the condition that K_1 and K_2 be metrizable is deleted, provided the Borel classes are replaced by the Baire classes.*

Theorem B2. *Theorem B continues to hold when the condition that K be metrizable is deleted, provided the Borel classes are replaced by the Baire classes.*

References

1. Čoban, M.M.: Multivalued mappings and Borel sets I. Trudy Moskov. Mat. Obšč. **22**, 229–250 (1970) [=Transl. Moscow Math. Soc. **22**, 258–280 (1970)]
2. Freiwald, R.C.: Images of Borel sets and k-analytic sets. Fund. Math. **75**. 35–46 (1972)
3. Fleissner, W.G.: An axiom for nonseparable Borel theory. Trans AMS **251**, 309–328 (1979)
4. Hansell, R.W.: Borel measurable mappings for nonseparable metric spaces. Trans. AMS **161**, 145–169 (1971)
5. Hansell, R.W.: On Borel mappings and Baire functions. Trans. AMS **194**, 195–211 (1974)
6. Hansell, R.W., Jayne, J.E., Rogers, C.A.: Piece-wise closed functions and almost discretely σ-decomposable families. To appear
7. Jayne, J.E., Rogers, C.A.: Fonctions et isomorphismes boréliens du premier niveau. C.R. Acad. Sci. Paris Sér. A **291**, 351–354 (1980)
8. Jayne, J.E., Rogers, C.A.: Fonctions fermées en partie. C.R. Acad. Sci. Paris Sér. A **291**, 667–670 (1980)
9. Jayne, J.E., Rogers, C.A.: Piece-wise closed functions. Math. Ann. **255**, 499–518 (1981)
10. Jayne, J.E., Rogers, C.A.: First level Borel functions and isomorphisms. J. Math. Pure Appl. **61**, 177–205 (1982)
11. Jayne, J.E., Rogers, C.A.: The invariance of the absolute Borel classes, Convex analysis and optimization. Research Notes in Math., No. 57, pp. 118–151. London: Pitman 1982
12. Kaniewski, J., Pol, R.: Borel measurable selectors for compact-valued mappings in the non-separable case. Bull. Acad. Polon. **23**, 1043–1050 (1975)
13. Kuratowski, K.: Topology I. New York: Academic Press 1966
14. Rogers, C.A., Jayne, J.E., Dellacherie, C., Topsøe, F., Hoffmann-Jørgensen, J., Martin, D.A., Kechris, A.S., Stone, A.H.: Analytic sets. London: Academic Press 1980
15. Saint-Raymond, J.: Fonctions boréliennes sur un quotient. Bull. Sci. Math. **100**, 141–147 (1976)
16. Taĭmanov, A.D.: On closed mappings I. Mat. Sb., **36**, 349–352 (1955)
17. Vaĭnšteĭn, I.A.: On closed mappings on metric spaces. Dokl. Akad. Nauk. SSSR **57**, 319–321 (1947)
18. Vaĭnšteĭn, I.A.: On closed mappings. Učen. Zap. Moskov. Univ. **155**, 3–53 (1952)
19. Hansell, R.W.: On characterizing non-separable analytic and extended Borel sets as types of continuous images. Proc. London Math. Soc. **28**, 683–699 (1974)

Received April 13, 1982; in revised form December 10, 1982

Math. Ann. 263, 343–364 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Automorphisms of Direct Products of Groups and Their Geometric Realisations

Francis E. A. Johnson
Department of Mathematics, University College, Gower Street, London WC1E 6BT, UK

0. Introduction

An outstanding problem in the theory of Poincaré Duality (=PD) groups is to decide whether each PD group can be realised as the fundamental group of a smooth closed aspherical manifold. See [2, 15–17] for previous work on this problem.

There are two main ways of constructing new PD groups from old [18];
(i) by extension of one PD group by another, and
(ii) by torsion free extension of a PD group by a finite group.

One approach to the above problem is to apply the constructions (i) and (ii) to known realisable PD groups and try to decide whether the resulting PD groups are realisable. Though other examples are known, for instance the examples constructed by Mostow and Siu in [25], the most convenient PD groups with which to start are discrete uniform subgroups of noncompact Lie groups. If Γ is a torsion free discrete uniform subgroup of a connected noncompact Lie group G, and if K is a maximal compact subgroup of G, then $\Gamma\backslash G/K$ is a smooth closed manifold of type $K(\Gamma, 1)$.

The present paper is a continuation of both [15] and [16]. Whereas in [16] we allowed only Surface groups, that is, torsion free discrete uniform subgroups of $PGL_2(\mathbb{R})$, and in [15] allowed only discrete uniform subgroups of Lie groups with no $PSL_2(\mathbb{R})$ factors as building blocks, here we go some way towards combining the two.

The bulk of the paper is a purely algebraic study of the automorphisms of products of certain groups, notably Surface groups. This may be thought of as a non-abelian variant of Wedderburn Theory. Its chief method is to introduce a quasi-ring structure into the set $\mathrm{End}(G)$ of endomorphisms of a group G, and to regard $\mathrm{End}(G^{(n)})$ as a quasi-ring of $n \times n$ matrices over $\mathrm{End}(G)$, where $G^{(n)}$ is the n-fold product of G with itself. Our main purely algebraic result [Corollary (2.2) below] may be stated simple as follows, where $A \int \sigma_n$ denotes the wreath product $A^{(n)} \rtimes \sigma_n$.

Theorem. *Let* $G \cong G_1^{(p_1)} \times \ldots \times G_n^{(p_n)}$ *be a direct product of groups such that* (i) *each* G_i *has trivial centre and is not decomposable as a nontrivial direct product, and* (ii) *for* $i<j$, *there is no nontrivial homomorphism* $\theta: G_i \to G_j$ *with* $\theta(G_i)$ *normal in* G_j. *Then there is a natural isomorphism*

$$\mathrm{Aut}(G) \cong \prod_{i=1}^{n} \mathrm{Aut}(G_i) \wr \Sigma_{p_i}.$$

This result is used in Sects. 4 and 5 to prove our main topological result, Theorem (5.4) below, which we state here thus:

Theorem. *Let* $1 \to H \times H' \to G \to Q \to 1$ *be an exact sequence of groups where*

(i) H' *is a finite product of Surface groups,*

(ii) H *is a torsion free discrete uniform subgroup of a connected linear semisimple Lie group* L *with no compact factors, such that* $\pi(H)$ *is non-discrete for any epimorphism* π *from* L *to a Lie group of* $\mathbb{R}$-rank 1, *and*

(iii) *there exists a smooth closed manifold of type* $K(Q,1)$.

Then there exists a smooth closed manifold of type $K(G,1)$.

The link between the algebra and the realisation results is provided by the concept of weak rigidity. A PD group Γ is said to be weakly rigid if there exists a smooth closed manifold X of type $K(\Gamma,1)$ such that the natural map $\mathrm{BDIFF}(X_\Gamma) \to BG(X_\Gamma)$ admits a right inverse up to homotopy, where $\mathrm{DIFF}(X_\Gamma)$ is the diffeomorphism group of X_Γ, and $G(X_\Gamma)$ is the monoid of homotopy equivalences of X_Γ. Our discussion of weak rigidity is carried out in Sect. 3.

In the above theorem we allow the degenerate cases $H' = \{1\}$ and $H = \{1\}$. At the referee's suggestion, we first treat the case where $H = \{1\}$. This is done in Sect. 4. We use this result to realise some 6-dimensional poly-Surface groups neglected in [16]. Finally, in Sect. 5 we treat the general case where $H \neq \{1\}$.

1. Automorphisms of $G^{(n)}$

To distinguish between the notions of internal and external direct product, we will use the notation "$G = A_1 \circ \ldots \circ A_k$" to mean that the group G is the internal direct product of its normal subgroups $A_1, \ldots, A_k$, reserving "$A_1 \times \ldots \times A_k$" for the external product, and "$G = A_1 \ldots A_k$" merely to mean that G is generated by its normal subgroups $A_1, \ldots, A_k$. For any group G, $G^{(n)}$ will denote the n-fold external direct product

$$G^{(n)} = \underbrace{G \times \ldots \times G}_{n}.$$

Suppose that $G = H^{(n)}$. Then there is an obvious injective homomorphism $\natural: \mathrm{Aut}(H) \wr \Sigma_n \to \mathrm{Aut}(H^{(n)})$ defined by

$$(\natural(a_1, \ldots, a_n, \sigma))(x_1, \ldots, x_n) = (a_1(x_{\sigma^{-1}(1)}), \ldots, a_n(x_{\sigma^{-1}(n)})),$$

where, of course, $\mathrm{Aut}(H) \wr \Sigma_n$ is just wreath product, i.e. semidirect product $\mathrm{Aut}(H)^{(n)} \rtimes \Sigma_n$, where the action of the symmetric group Σ_n on $\mathrm{Aut}(H)^{(n)}$ is by

$${}^{\sigma}(a_1, \ldots, a_n) = (a_{\sigma^{-1}(1)}, \ldots, a_{\sigma^{-1}(n)}).$$

Similarly, in the case where $G = A_1 \times \ldots \times A_k$, even with no repeated factors, there is an injective group homomorphism, still denoted by $\natural$,

$$\natural : \mathrm{Aut}(A_1) \times \ldots \times \mathrm{Aut}(A_k) \to \mathrm{Aut}(G)$$

given by $(\natural(a_1, \ldots, a_k))(x_1, \ldots, x_k) = (a_1(x_1), \ldots, a_n(x_n))$. We shall show in a number of geometrically interesting cases, including products of Surface groups, that these natural maps

$$\natural : \mathrm{Aut}(H) \wr \sigma_n \to \mathrm{Aut}(H^{(n)})$$

$$\natural : \mathrm{Aut}(A_1) \times \ldots \times \mathrm{Aut}(A_k) \to \mathrm{Aut}(A_1 \times \ldots \times A_k)$$

are isomorphisms. We may think of these as being semisimplicity results. The theory has a number of points of similarity with Wedderburn theory. The case where all factors are distinct is dealt with in Sect. 2. In this section we concern ourselves only with the case where all factors are the same, and we will prove

Theorem 1.1. *Let G be a nontrivial group having the following properties;*

(i) *the centre, $\zeta_1(G)$, of G is trivial, and*

(ii) *G cannot be decomposed as a nontrivial direct product. Then, for any $n \in \mathbb{Z}_+$, the natural map*

$$\natural : \mathrm{Aut}(G) \wr \sigma_n \to \mathrm{Aut}(G^{(n)})$$

is an isomorphism.

For any group G denote by $\mathrm{End}(G)$ the set of all group homomorphisms $G \to G$. $\mathrm{End}(G)$ is, of course, a monoid under composition and $\mathrm{Aut}(G)$ is its group of invertible elements. We recall briefly how to describe elements of $\mathrm{End}(G^{(n)})$ by means of matrices. This is commonplace when G is abelian, less so when G is nonabelian.

First note that, in addition to its composition "product", $\mathrm{End}(G)$ also has a partially defined "sum" which we now describe.

Definition. Let $G = (G, *, 1)$ be a group and let $\alpha, \beta \in \mathrm{End}(G)$. Say that α centralises β iff $\mathrm{Im}\alpha$ centralises $\mathrm{Im}\beta$ i.e. iff for all $x, y \in G$, $\alpha(x) * \beta(y) = \beta(y) * \alpha(x)$.

Define $C(G) = \{(\alpha, \beta) \in \mathrm{End}(G) \times \mathrm{End}(G) : \alpha \text{ centralises } \beta\}$. Clearly $C(G)$ is a symmetric subset of $\mathrm{End}(G) \times \mathrm{End}(G)$. The map

$$\Delta : C(G) \to \mathrm{End}(G), \quad (\alpha \Delta \beta)(x) = \alpha(x) * \beta(x),$$

is our partially defined sum. In particular, denoting composition by juxtaposition of symbols, and abusing notation by writing "0" for the trivial homomorphism into the multiplicative group $(G, *, 1)$, the following laws hold whenever all terms are defined.

(0) $\alpha 1 = 1\alpha = \alpha$, (I) $(\alpha \Delta \beta)\gamma = (\alpha\gamma) \Delta (\beta\gamma)$,

(II) $\gamma(\alpha \Delta \beta) = (\gamma\alpha) \Delta (\gamma\beta)$, (III) $\alpha \Delta (\beta \Delta \gamma) = (\alpha \Delta \beta) \Delta \gamma$,

(IV) $\alpha \Delta \beta = \beta \Delta \alpha$, (V) $\alpha \Delta 0 = 0 \Delta \alpha = \alpha$.

Definition. By a quasi-ring we mean a sextuple $(A, C, \Delta, \cdot, 1, 0)$ where A is a set, $C \subset A \times A$ is a symmetric subset, $0, 1 \in A$, and $\Delta : C \to A$, $\cdot : A \times A \to A$ are mappings satisfying

(i) for each $a, b, c \in A$, $(a, b) \in C$, and $(a \Delta b, c) \in C$ if and only if $(b, c) \in C$ and $(a, b \Delta c) \in C$,

(ii) for each $a, b, c \in A$, $(a, b) \in C$ iff $(ca, cb) \in C$ iff $(ac, bc) \in C$,

(iii) for each $a \in A$, $(a, 0) \in C$, and

(iv) the laws (0)–(V) above are satisfied when all terms are defined.

Clearly $(\mathrm{End}(G), C(G), \Delta, \cdot, 1, 0)$ is a quasi-ring for any group G.

Definition. Define $\mathfrak{N}_n^G = \left\{\alpha = (\alpha_{ij})_{\substack{1 \leqq i \leqq n \\ 1 \leqq j \leqq n}} : \alpha_{ij} \in \mathrm{End}(G)\right\}$, the set of all $n \times n$ matrices in $\mathrm{End}(G)$ for a group G, and put

$$M_n^G = \{\in \mathfrak{N}_n^G : \text{for each } i, j, k \in \{1, \ldots, n\},\ \alpha_{ij} \text{ centralises } \alpha_{ik} \text{ provided } j \neq k\}.$$

Now let $\alpha, \beta \in M_n^G$. Then for each r, s and each i, j with $i \neq j$, observe that $\alpha_{ri}\beta_{is}$ centralises $\alpha_{rj}\beta_{js}$. Thus we may define a map

$$\square : M_n^G \times M_n^G \to \mathfrak{N}_n^G$$

by

$$(\alpha \square \beta)_{rs} = (\alpha_{r1}\beta_{1s}) \Delta (\alpha_{r2}\beta_{2s}) \Delta \ldots \Delta (\alpha_{rn}\beta_{ns})$$

i.e. for all $x \in G$

$$(\alpha \square \beta)_{rs}(x) = \alpha_{r1}(\beta_{1s}(x)) * \ldots * \alpha_{rn}(\beta_{ns}(x)),$$

where "$*$" is multiplication in G. Of course, the above is by straightforward analogy with ordinary matrix multiplication over a genuine ring.

Proposition 1.2. *$\alpha \square \beta \in M_n^G$ for all $\alpha, \beta \in M_n^G$.*

Proof. We must show that $(\alpha \square \beta)_{rs}$ centralises $(\alpha \square \beta)_{rt}$ for $s \neq t$. For $x, y \in G$ we have

$$(\alpha \square \beta)_{rs}(x) * (\alpha \square \beta)_{rt}(y) = g_1 * g_2 * \ldots * g_n * h_1 * h_2 * \ldots * h_n,$$

where $g_i = \alpha_{ri}(\beta_{is}(x))$ and $h_j = \alpha_{rj}(\beta_{jt}(y))$. Since α_{ri} centralises α_{rj} for $i \neq j$, then $g_i * h_j = h_j * g_i$ for $i \neq j$, and we may rearrange the above product to get

$$(\alpha \square \beta)_{rs}(x) * (\alpha \square \beta)_{rt}(y) = g_1 * h_1 * g_2 * h_2 * \ldots * g_n * h_n.$$

However,

$$g_i * h_i = \alpha_{ri}(\beta_{is}(x) * \beta_{it}(y)) = \alpha_{ri}(\beta_{it}(y) * \beta_{is}(x))$$

since β_{is} centralises β_{it} for $s \neq t$. Thus $g_i * h_i = h_i * g_i$, and we may rearrange the above product again to get

$$\begin{aligned}(\alpha \square \beta)_{rs}(x) * (\alpha \square \beta)_{rt}(y) &= h_1 * g_1 * h_2 * g_2 * \ldots * h_n * g_n \\ &= h_1 * h_2 * \ldots * h_n * g_1 * g_2 * \ldots * g_n \\ &= (\alpha \square \beta)_{rt}(y) * (\alpha \square \beta)_{rs}(x). \quad \text{Q.E.D.}\end{aligned}$$

Thus $\square$ is actually a map $\square : M_n^G \times M_n^G \to M_n^G$. Moreover, defining $I_n \in M_n^G$ by

$$(I_n)_{rs} = \begin{cases} \mathrm{Id}_G & \text{if } r = s \\ 0 & \text{if } r \neq s \end{cases}$$

then I_n is a two-sided identity for $(M_n^G, \square)$.

Proposition 1.3. *$M_n^G = (M_n^G, \square, I_n)$ is a monoid, for any group G.*

Proof. It remains only to show that $\square$ is associative, the proof of which follows the usual proof that matrix multiplication is associative. The only properties that we need of $(\mathrm{End}(G), \Delta)$ are (I)–(IV) mentioned above. Q.E.D.

For any group G, let $\pi_r : G^{(n)} \to G$ be the r^{th} projection, $\pi_r(x_1, \ldots, x_n) = x_r$, and let $i_s : G \to G^{(n)}$ be the canonical inclusion of the s^{th} factor. $i_s(x) = (1, 1, \ldots, x, \ldots, 1)$. We define a mapping s^{th} place

$$\Phi : \mathrm{End}(G^{(n)}) \to \mathfrak{N}_n^G \quad \text{by} \quad \Phi(\alpha)_{rs} = \pi_r \alpha i_s .$$

Proposition 1.4. *For any group G, $\Phi : \mathrm{End}(G^{(n)}) \to M_n^G$ is a monoid isomorphism.*

Proof. Observe that αi_s centralises αi_t for $s \neq t$, so that $\pi_r \alpha i_s$ centralises $\pi_r \alpha i_t$ for $s \neq t$. Hence $\mathrm{Im}(\Phi)$ is certainly contained in M_n^G. To see that Φ is a monoid homomorphism, first note that each mapping

$$\varphi_{rs} : \mathrm{End}(G^{(n)}) \to \mathrm{End}(G), \qquad \varphi_{rs}(\alpha) = \pi_r \alpha i_s ,$$

is Δ-homomorphism. If $1 \in \mathrm{End}(G^{(n)})$ is the identity map we have

$$1 = (i_1 \pi_1) \Delta (i_2 \pi_2) \Delta \ldots \Delta (i_n \pi_n),$$

so that

$$\alpha\beta = (\alpha i_1 \pi_1 \beta) \Delta (\alpha i_2 \pi_2 \beta) \Delta \ldots \Delta (\alpha i_n \pi_n \beta),$$

for all $\alpha, \beta \in \mathrm{End}(G^{(n)})$. And since φ_{rs} is a Δ-homomorphism we get

$$\varphi_{rs}(\alpha\beta) = \varphi_{rs}(\alpha i_1 \pi_1 \beta) \Delta \ldots \Delta \varphi_{rs}(\alpha i_n \pi_n \beta)$$

But $\varphi_{rs}(\alpha i_p \pi_p \beta) = \varphi_{rp}(\alpha) \varphi_{ps}(\beta)$. Hence

$$\begin{aligned} \Phi(\alpha\beta)_{rs} = \varphi_{rs}(\alpha\beta) &= \Phi(\alpha)_{r1} \Phi(\beta)_{1s} \Delta \ldots \Delta \Phi(\alpha)_{rn} \Phi(\beta)_{ns} \\ &= (\Phi(\alpha) \square \Phi(\beta))_{rs} . \end{aligned}$$

That is, Φ is a monoid homomorphism.

Injectivity of Φ is straightforward, and we leave it as an exercise. Surjectivity, also straightforward, is more instructive. Let $(\alpha_{rs})_{\substack{1 \leqq r \leqq n \\ 1 \leqq s \leqq n}} \in M_n^G$, so that α_{rp} centralises α_{rq} for $p \neq q$. For each r define $\alpha_r : G^{(n)} \to G$ by $\alpha_r(x_1, \ldots, x_n) = \alpha_{r1}(x_1) * \ldots * \alpha_{rn}(x_n)$. Then α_r is a homomorphism by the centralising hypothesis, and we get a homomorphism $\alpha : G^{(n)} \to G^{(n)}$ by $\alpha(\mathbf{x}) = (\alpha_1(\mathbf{x}), \ldots, \alpha_n(\mathbf{x}))$. Moreover $\Phi(\alpha)_{rs} = \alpha_{rs}$ so that Φ is surjective. Q.E.D.

In the proof of Theorem (1.1), we shall need

Lemma 1.5. *Let G be a nontrivial group having the following properties;*

(i) *G has trivial centre,*

(ii) *G is not expressible as a non-trivial direct product, and*

(iii) *$G = A_1 \dots A_n$ where each A_i is normal in G, and A_i centralises A_j for $i \neq j$. Then for some unique $r \in \{1, \dots, n\}$ we have $G = A_r$, and $A_s = \{1\}$, for $s \neq r$.*

Proof. By induction on n. For $n = 1$, there is nothing to prove. Suppose proved for $n-1$, and put $B = A_1 \dots A_{n-1}$. Then B is normal and centralises A_n, so that $B \cap A_n$ centralises $G = B \cdot A_n$. But G has trivial centre, so that $B \cap A_n = \{1\}$, and $G \cong B \times A_n$. Since G is not isomorphic to a non-trivial direct product, we get either that $B = \{1\}$ or $A_n = \{1\}$. If $A_n = \{1\}$, then $G = A_1 \dots A_{n-1}$, and we get the result by induction. If $B = \{1\}$, then $G = A_n$, and the result follows. Q.E.D.

Consider the mapping $\Psi : \mathrm{Aut}(G)^{(n)} \times \sigma_n \to M_n^G$ defined by

$$\Psi(\alpha, \sigma)_{rs} = \begin{cases} \alpha_r & r = \sigma(s), \\ 0 & r \neq \sigma(s), \end{cases} \quad \text{where} \quad \alpha = (\alpha_i)_{1 \leqq i \leqq n}.$$

It is straightforward to verify that Ψ induces an injective group homomorphism $\Psi : \mathrm{Aut}(G) \wr \sigma_n \to (M_n^G)^*$, where $(M_n^G)^*$ denotes the group of invertible elements of the monoid M_n^G.

Proof of Theorem 1.1. Let G be a nontrivial group with trivial centre, and suppose that G is not expressible as a nontrivial direct product. Let $\alpha \in \mathrm{Aut}(G^{(n)})$ and write $\Phi(\alpha) = (\alpha_{rs})_{\substack{1 \leqq r \leqq n \\ 1 \leqq s \leqq n}}$. Fix $i \in \{1, \dots, n\}$. Now $\pi_i(G^{(n)}) = G$ so that $G = \mathrm{Im}(\alpha_{i1}) \cdot \mathrm{Im}(\alpha_{i2}) \dots \mathrm{Im}(\alpha_{in})$. Since $\mathrm{Im}(\alpha_{is})$ centralises $\mathrm{Im}(\alpha_{it})$ for $s \neq t$, then by Lemma (1.5), there exists a unique $\sigma(i) \in \{1, \dots, n\}$ such that $\mathrm{Im}(\alpha_{i\sigma(i)}) = G$, with $\mathrm{Im}(\alpha_{ij}) = \{1\}$ if $i \neq j$. We claim that the mapping $\sigma : \{1, \dots, n\} \to \{1, \dots, n\}$ is bijective. For if not, then σ is not surjective, and there exists $m \in \{1, \dots, n\}$ such that $\alpha_{rm} = 0$ for all r. Thus $G \cong \mathrm{Im}(i_m) \subset \mathrm{Ker}(\pi_r \alpha)$ for all r, so that G is trivial. Contradiction. Hence σ is bijective.

Also note that each $\alpha_{p\sigma(p)}$ must be an automorphism of G, for if $\alpha_{p\sigma(p)}$ is not surjective neither is $\pi_p \alpha$, which is a contradiction, whilst if $\alpha_{p\sigma(p)}$ is not injective neither is αi_p, also a contradiction. Thus $\alpha_{p\sigma(p)} \in \mathrm{Aut}(G)$ and we have shown that

$$\Phi(\mathrm{Aut}(G^{(n)}) = \Psi(\mathrm{Aut}(G \wr \sigma_n).$$

Since Φ maps $\mathrm{Aut}(G^{(n)})$ bijectively onto $(M_n^G)^*$, we see that $\Phi^{-1}\Psi : \mathrm{Aut}(G) \wr \sigma_n \to \mathrm{Aut}(G^{(n)})$ is a group isomorphism. It is easy to see that $\Phi^{-1}\Psi$ is in fact the previously defined map $\natural$. Q.E.D. [Theorem (1.1)]

By a Surface group we mean the fundamental group of a closed connected 2-manifold of genus $\geqq 2$. The hypotheses imposed on G in (1.1) are satisfied by Surface groups and by many others. The following list is representative, without in any way attempting to be exhaustive.

Proposition 1.6. *Any group G from the classes $\mathfrak{C}_1, \mathfrak{C}_2, \mathfrak{C}_3$ below satisfies the hypotheses of* (1.1), *namely that G has trivial centre and is not decomposable as a nontrivial direct product.*

$\mathfrak{C}_1$: *Surface groups*

$\mathfrak{C}_2$: *Nontrivial free products. In particular, free groups of* rank $\geqq 2$.

$\mathfrak{C}_3$: *Discrete subgroups of finite covolume in simple noncompact Lie groups with trivial centre.*

Proof. The proof for $\mathfrak{C}_3$ is a by-now-standard argument using Borel's Density Theorem, and is almost, but not quite, proved in Chap. V of Raghunathan's book [26]. If $G=G_1\circ G_2$ is discrete with finite covolume in $L\subset \mathrm{GL}(n,\mathbb{R})\subset \mathrm{GL}(n,\mathbb{C})$, where L is a simple noncompact Lie group with $\zeta_1(L)=1$, put $L_i=$(Zariski closure of $G_i)\cap L$. Then $G_i\lhd G$ so that $L_i\lhd L$ and either $L_i=L$ or $L_i=\{1\}$. However $G_i\neq\{1\}$, so that $L_i=L$. But G_1 centralises G_2 so that L_1 centralises L_2 hence $L_1=L_2=L$ is abelian. Contradiction. The argument showing $\zeta_1(G)=\{1\}$ is similar, and occurs in [26, Corollary 5.18, p. 84].

For $\mathfrak{C}_2$, that free products are centreless is clear, or follows from the Kurosh Theorem. The observation that a free product cannot also be a direct product is made in [21, p. 177].

For Surface groups, we can, by passing to a subgroup of finite index, suppose that G is orientable. G is then a discrete uniform subgroup of $PSL_2(\mathbb{R})$, by Uniformisation Theory, and the result follows from the $\mathfrak{C}_3$ case already treated. Q.E.D.

2. Product Structures: Uniqueness and Automorphisms

Let $\mathfrak{C}$ be a class of groups. By a $\mathfrak{C}$-product structure on a group G we mean a finite sequence $\mathfrak{S}=(A_i, G_i, p_i)_{1\leqq i\leqq n}$, where

(i) A_i is a normal subgroup of G
(ii) $G_i\in\mathfrak{C}$, and
(iii) p_i is a positive integer such that
(iv) $G_i\ncong G_j$ if $i\neq j$
(v) $G=A_1\circ\ldots\circ A_n$ and
(vi) $A_i\cong G_i^{(p_i)}$.

There are a number of possible notions of uniqueness for $\mathfrak{C}$-product structures. We consider two.

Definition. If $\mathfrak{S}=(A_i, G_i, p_i)_{1\leqq i\leqq n}$ and $\mathfrak{T}=(B_i, H_i, q_i)_{1\leqq i\leqq m}$ are $\mathfrak{C}$-product structures on a group G say that $\mathfrak{S}$ and $\mathfrak{T}$ are *equivalent* if and only if $n=m$ and for some (necessarily unique) $\sigma\in\sigma_n$ we have $H_i\cong G_{\sigma(i)}$ and $q_i=p_{\sigma(i)}$. Moreover, we say that $\mathfrak{S}$ and $\mathfrak{T}$ are *strongly equivalent* when they are equivalent and in addition $B_i=A_{\sigma(i)}$, for all i, $1\leqq i\leqq n$. Note that we take "equals" here, not merely "$\cong$".

In this section we give conditions on a class of groups $\mathfrak{C}$ which ensure that when G has a $\mathfrak{C}$-product structure, it is unique in the strong sense. Thus, *without further mention*, for the remainder of this section, $\mathfrak{C}$ will denote a class of *nontrivial* groups satisfying the following three properties;

($\mathfrak{P}1$) Each $G\in\mathfrak{C}$ has trivial centre

($\mathfrak{P}2$) If $G\in\mathfrak{C}$ then G is not decomposable as a nontrivial direct product

($\mathfrak{P}3$) There is a total ordering $\leqq$ on the isomorphism types of $\mathfrak{C}$ such that, if $H<K$ then there is no nontrivial homomorphism $\theta: H\to K$ with $\theta(H)$ normal in K. Here by $H<K$ we mean that $H\leqq K$ and $H\ncong K$.

In this section we shall prove

Theorem 2.1. *If $\mathfrak{S}$, $\mathfrak{T}$ are $\mathfrak{C}$-product structures on a group G, then $\mathfrak{S}$ and $\mathfrak{T}$ are strongly equivalent.*

Before embarking on the proof of (2.1), we give some explanations and examples. Observe that in order to get any sort of uniqueness result it is necessary to exclude, as we have done, the trivial group from the class $\mathfrak{C}$. However it is useful to be able to refer to the trivial group, which we do by putting $\mathfrak{C}_+ = \mathfrak{C} \cup \{\text{trivial group}\}$ and extending $\leqq$ to a total ordering on $\mathfrak{C}_+$ by insisting that $H < \{1\}$ for all $H \in \mathfrak{C}$. Then the statement "$H < K$" will mean at least "$H \neq \{1\}$".

The simplest example of a class satisfying ($\mathfrak{P}1$), ($\mathfrak{P}2$), ($\mathfrak{P}3$) is the class of finitely generated free groups of rank $\geqq 2$. In this case we take $H \leqq K$ iff $\text{rank}(H) \leqq \text{rank}(K)$. The class which interests us most is the class of Surface groups, though here the ordering is slightly more complicated, and we refer the reader to Proposition (2.7), where it is explained in detail.

Theorem (2.1) has the following obvious corollary, which is, in fact, the main purpose of this section.

Corollary 2.2. *Let $(A_i, G_i, p_i)_{1 \leqq i \leqq n}$ be a $\mathfrak{C}$-product structure on G. Then the natural maps $\natural$ defined in Sect. 1 give isomorphisms*

$$\text{Aut}(G) \cong \prod_{i=1}^{n} \text{Aut}(A_i) \cong \prod_{i=1}^{n} \text{Aut}(G_i) \int \sigma_{p_i}.$$

On the way to proving (2.1), we first prove

Proposition 2.3. *If $A, B \in \mathfrak{C}$ and $A^{(n)} \cong B^{(m)}$ then $A \cong B$ and $n = m$.*

Proof. Let $f : A^{(n)} \to B^{(m)}$ be an isomorphism, and let $\pi_r : B^{(m)} \to B$ and $i_s : A \to A^{(n)}$ denote the canonical projections and inclusions. Put $B_{rs} = \text{Im}(\pi_r f i_s)$. Then $B = B_{r1} \cdot B_{r2} \ldots B_{rn}$ and B_{rp} centralises B_{rq} for $p \neq q$. By Lemma (1.5), there is a function $\sigma : \{1, \ldots, m\} \to \{1, \ldots, n\}$ with the property that, for all $r \in \{1, \ldots, m\}$, $B = B_{r, \sigma(r)}$ whilst $B_{rs} = 1$ for $s \neq \sigma(r)$. Note that $\pi_r f i_{\sigma(r)} : A \to B$ is onto, so that $B \leqq A$. Now by symmetry $A \leqq B$, so that $A \cong B$.

We claim that $\sigma : \{1, \ldots, m\} \to \{1, \ldots, n\}$ is surjective. For if not, then there exists $s \in \{1, \ldots, n\}$ such that $B_{rs} = 1$ for all r. Hence $\text{Im}(i_s)$ is contained in the intersection of the $\text{Ker}(\pi_r f)$ for $1 \leqq r \leqq m$. Hence $\text{Im}(i_s) \cong A$ must be trivial. Contradiction. Thus $n \leqq m$, and by symmetry, $m \leqq n$. Hence $n = m$. Q.E.D.

Lemma 2.4. *Suppose that $G = A \circ B$, where A is a centreless characteristic subgroup of G. Then B is also characteristic in G, and*

$$\natural : \text{Aut}(A) \times \text{Aut}(B) \to \text{Aut}(G)$$

is an isomorphism.

Proof. Let $\text{Aut}(\mathfrak{E})$ be the group of all automorphisms of the exact sequence

$$\mathfrak{E} = (1 \to A \xrightarrow{i} G \xrightarrow{p} B \to 1).$$

Since $\mathfrak{E}$ is a trivial extension, there is a split exact sequence

$$1 \to C(\mathfrak{E}) \to \text{Aut}(\mathfrak{E}) \underset{\natural}{\overset{\varrho}{\rightleftarrows}} \text{Aut}(A) \times \text{Aut}(B) \to 1$$

where ϱ is the obvious restriction map, the splitting map $\natural$ is the one defined in Sect. 1, and $C(\mathfrak{E})$ is the group of congruences of $\mathfrak{E}$, i.e. the group of automorphisms which induce the identity on A and B, the "ends" of $\mathfrak{E}$.

However, since A has trivial centre, $C(\mathfrak{E})$ is trivial. For if $h \in C(\mathfrak{E})$, then for each $x \in G$,

$$p(h(x)x^{-1}) = ph(x)p(x)^{-1} = p(x)p(x)^{-1} = 1$$

hence $h(x)x^{-1} \in \mathrm{Ker}(p) = A$. Consider the map $\hat{h}: G \to A$ defined by $\hat{h}(x) = h(x)x^{-1}$. If $a \in A$ and $x \in G$ then $\hat{h}(xa) = \hat{h}(x)$, since

$$\hat{h}(xa) = h(x)h(a)a^{-1}x^{-1} = \hat{h}(x),$$

h inducing the identity on A. Now if $x \in G$, then for all $y \in A$,

$$y\hat{h}(x)y^{-1} = h(y)h(x)x^{-1}y^{-1} = \hat{h}(xy),$$

and, by the above, $\hat{h}(yx) = \hat{h}(x(x^{-1}yx)) = \hat{h}(x)$, since $x^{-1}yx \in A$. That is, $y\hat{h}(x)y^{-1} = \hat{h}(x)$ for all $y \in A$, so that each $\hat{h}(x) \in \mathfrak{Z}(A)$, the centre of A. Since $\mathfrak{Z}(A) = \{1\}$, then for all $x \in G$, $h(x) = x$. Thus $C(\mathfrak{E}) = \{1\}$. Now $\mathrm{Aut}(\mathfrak{E}) = \mathrm{Aut}(G)$, since A is characteristic in G, so that $\natural$ is an isomorphism

$$\natural: \mathrm{Aut}(A) \times \mathrm{Aut}(B) \to \mathrm{Aut}(G)$$

with inverse ϱ, and B is characteristic since it is visibly invariant under the whole of $\mathrm{Aut}(G) \cong \mathrm{Aut}(A) \times \mathrm{Aut}(B)$. Q.E.D.

Lemma 2.5. *Suppose that* $G = A \circ B = A' \circ B'$ *and* $A \cong G_1^{(p)}$ $(p \geqq 1)$, $B \cong \prod_{i=2}^{n} G_i$, $A' \cong H_1^{(q)}$ $(q \geqq 1)$, $B' \cong \prod_{j=2}^{m} H_j$ *where* $G_i, H_j \in \mathfrak{C}_+$, *possibly allowing repetitions amongst the* G_i, H_j, *and such that* $G_1 < G_i$ *and* $H_1 < H_j$ *for all* $i, j \geqq 2$. *Then*

(i) $A = A'$, $G_1 \cong H_1$, *and* $p = q$, *and*

(ii) $B = B'$.

(Note that we really do mean $A = A'$ and $B = B'$, and not just $A \cong A'$, $B \cong B'$.)

Proof. First note that everything will follow if we can prove that $A \subset A'$. For then it follows by symmetry that $A = A'$. The conclusion that $G_1 \cong H_1$ and $p = q$ will follow from Proposition (2.3). Also note that A is then characteristic, for if α is an automorphism of G, putting $A'' = \alpha(A)$ and $B'' = \alpha(B)$ we see that the decomposition $G = A'' \circ B''$ satisfies our hypotheses, so that $A = A'' = \alpha(A)$. Since a product of centreless groups is centreless, A has trivial centre so that B is also characteristic by (2.4). To show that $B = B'$, note that $B \cong G/A = G/A' \cong B'$. If $f: B \to B'$ is an isomorphism consider the automorphism $h: G \to G$ defined by $h(ab) = af(b)$, $a \in A$, $b \in B$. Since B is characteristic, $B = h(B) = f(B) = B'$.

Thus it suffices to show, as we now do, that $A \subset A'$. Let $\pi_j: G \to H_j$ $(2 \leqq j \leqq m)$ and $\pi_1^t: G \to H_1$ $(1 \leqq t \leqq q)$ be the canonical projections and $i_s: G_1 \to A \subset G$ $(1 \leqq s \leqq p)$ the canonical inclusions. Then each $\pi_r i_s(G_1)$ [resp. $\pi_1^t i_s(G_1)$] is normal in H_r (resp. H_1). Suppose that $G_1 < H_1$. Then $G_1 < H_r$ for each r, so that each $i_s(G_1)$ is trivial. Contradiction, hence we must have $H_1 \leqq G_1$. However, if $H_1 < G_1$, then by the

above argument and symmetry we again get a contradiction. Thus $G_1 \cong H_1$, so that $G_1 < H_r$ for all $r \geqq 2$, and so each $\pi_r i_s(G_1) = \{1\}$ for $2 \leqq r \leqq q$. Thus for each s, $1 \leqq s \leqq p$, we see that

$$\mathrm{Im}(i_s) \subset \bigcap_{r=2}^{p} \mathrm{Ker}(\pi_r) = A'$$

and hence that $A \subset A'$. Q.E.D.

Proof of (2.1). Let $\mathfrak{S} = (A_i, G_i, p_i)_{1 \leqq i \leqq k}$ and $\mathfrak{I} = (A'_i, H_i, q_i)_{1 \leqq i \leqq m}$ be $\mathfrak{C}$-product structures on the group G. Without loss of generality, we can suppose that

$$G_k < G_{k-1} < \ldots < G_1, \qquad H_m < H_{m-1} < \ldots < H_1$$

and that $k \leqq m$. We first prove that $\mathfrak{S}$ and $\mathfrak{I}$ are equivalent in the weaker sense by proving $P(k)$ for all $k \geqq 1$, where

$P(k)$: $k = m$, and for all i, $1 \leqq i \leqq k$, $G_i \cong H_i$, and $p_i = q_i$.

Proof of $P(1)$. Suppose that $m > 1$. Put $A = A_1$, $B = \{1\}$, $A' = A'_m$, and $B' = A'_{m-1} \circ \ldots \circ A'_1$, and apply Lemma (2.5). Incidentally, the whole point of introducing the class $\mathfrak{C}_+$ is to allow the trivial factor B at this point. From (2.5) we get that $B' = B = \{1\}$, hence that $A'_r = \{1\}$ for $1 \leqq r \leqq m-1$. This is a contradiction, since $A'_r \cong H_r^{(q_r)} \neq \{1\}$, since $H_r \neq \{1\}$, and $q_r \geqq 1$. Thus $m = 1$, and $G_1^{(p_1)} \cong H_1^{(q_1)}$, whence $G_1 \cong H_1$ and $p_1 = q_1$ by (2.3). Q.E.D. $P(1)$

Proof that $P(k-1) \Rightarrow P(k)$. Put $A = A_k$, $B = A_{k-1} \circ \ldots \circ A_1$, $A' = A'_m$, and

$$B' = A'_{m-1} \circ \ldots \circ A'_1 .$$

By Lemma (2.5), we get that $B = B'$, so that, by induction, $(k-1) = (m-1)$, and for all i, $1 \leqq i \leqq k-1$, $G_i \cong H_i$, and $p_i = q_i$. However, (2.5) also gives that $A = A'$, whence, by (2.3), $G_k \cong H_k$ and $p_k = q_k$. Q.E.D. $P(k-1) \Rightarrow P(k)$.

Thus $P(k)$ is true for all $k \geqq 1$. To get the strong equivalence of $\mathfrak{S}$ and $\mathfrak{I}$, i.e. that $A_i = A'_i$ for all i, we use induction on k. For $k = 1$ there is nothing to prove. Suppose true for $(k-1)$ and still suppose that $G_k < G_{k-1} < \ldots < G_1$ and $H_k < H_{k-1} < \ldots < H_1$. Again put $A = A_k$, $B = A_{k-1} \circ \ldots \circ A_1$, $A' = A'_k$, and $B' = A'_{k-1} \circ \ldots \circ A'_1$. Lemma (2.5) gives $B = B'$, whence by induction we get that $A_i = A'_i$ for all i with $1 \leqq i \leqq k-1$. However, Lemma (2.5) also gives $A = A'$, so that $A_k = A'_k$, and $\mathfrak{S}$ and $\mathfrak{I}$ are strongly equivalent. Q.E.D. [Theorem (2.1)]

Proof of (2.2). If $G = A_1 \circ \ldots \circ A_k$ where each A_i is characteristic and centreless, then it follows by induction on (2.5) that the natural map

$$\natural : \prod_{i=1}^{k} \mathrm{Aut}(A_i) \to \mathrm{Aut}(G)$$

is an isomorphism. Thus if $(A_i, G_i, p_i)_{1 \leqq i \leqq k}$ is a $\mathfrak{C}$-product structure on G it suffices to show that each A_i is characteristic. However, if $\alpha : G \to G$ is an automorphism then $(\alpha(A_i), G_i, p_i)_{1 \leqq i \leqq k}$ is also a $\mathfrak{C}$-product structure on G. Hence $A_i = \alpha(A_i)$ by (2.1).

The isomorphism

$$\prod_{i=1}^{k} \mathrm{Aut}(G_i) \int \sigma_{p_i} \cong \mathrm{Aut}(G)$$

follows from (1.1). Q.E.D. [Corollary (2.2)]

Using results of Borel [4] and Margulis [29], one can construct large classes of discrete subgroups of Lie groups which satisfy ($\mathfrak{P}$.1), ($\mathfrak{P}$.2), and ($\mathfrak{P}$.3). Without attempting to be exhaustive, let $\mathfrak{L}$ be the following class.

$\mathfrak{L} = \{H : H$ is a discrete uniform subgroup in a noncompact connected simple adjoint Lie group L with $\mathrm{rank}_{\mathbb{R}}(L) \geqq 2.\}$

Proposition 2.6. *Any finite subclass of* $\mathfrak{L}$ *satisfies* ($\mathfrak{P}$.1), ($\mathfrak{P}$.2), *and* ($\mathfrak{P}$.3).

Proof. Let $\mathfrak{R}$ be the following relation on $\mathfrak{L}$: $H\mathfrak{R}K$ iff H is isomorphic to a subgroup of K. Trivially $\mathfrak{R}$ is reflexive and transitive. We also note that $\mathfrak{R}$ is antisymmetric in the following strong sense:

$H\mathfrak{R}K$ and $K\mathfrak{R}H \Rightarrow H \cong K$ and any injective homomorphism $i : H \to K$ is an isomorphism.

To see this, take an imbedding $i(H) \hookrightarrow H$. If $[H : i(H)]$ is infinite, let X be the symmetric space of $L \supset H$. By passing to a subgroup of finite index, $H\backslash X$ and $i(H)\backslash X$ are manifolds of type $K(H, 1)$ and dimension $n = \dim(X)$. But

$$H_n(H\backslash X, \mathbb{Z}_2) \cong \mathbb{Z}_2$$

since $H\backslash X$ is closed, and

$$H_n(i(H)\backslash X, \mathbb{Z}_2) \cong H_c^0(i(H)\backslash X, \mathbb{Z}_2) \cong 0$$

since $i(H)\backslash X$ is noncompact connected. Thus $[H : i(H)]$ is finite, and the result follows from [4, Proposition 1.7].

Now let $\mathfrak{D} = \{H_1, \ldots, H_n\}$ be a finite subclass of $\mathfrak{L}$. Without loss of generality we may suppose that $H_i \not\leqq H_j$ for $i \neq j$. Define $h : \mathfrak{D} \to \mathbb{N}$ by

$$h(H) = 0 \quad \text{iff} \quad K\mathfrak{R}H \Rightarrow H \cong K.$$

$$h(H) = n+1 \quad \text{iff} \quad K\mathfrak{R}H \quad \text{and} \quad H \not\leqq K \Rightarrow h(K) \leqq n.$$

Let $N = \max\{h(H) : H \in \mathfrak{D}\}$, and define

$$\mathfrak{D}_r = \{H \in \mathfrak{D} : h(H) = N - r\}.$$

Choose an arbitrary total ordering $\leqq_r$ on each $\mathfrak{D}_r$ and let $\leqq$ be the total ordering on $\mathfrak{D}$ constructed from the $\leqq_r$ with the additional requirement that

$$H \in \mathfrak{D}_r, \quad K \in \mathfrak{D}_s \quad \text{and} \quad r < s \Rightarrow H < K.$$

Observe that if H and K are distinct elements of the same $\mathfrak{D}_r$ then neither $H\mathfrak{R}K$ nor $K\mathfrak{R}H$. Also note that if H is isomorphic to a subgroup of K then $K \leqq H$.

Now suppose that $H, K \in \mathfrak{D}$ satisfy $H < K$, and let $\theta : H \to K$ be a homomorphism with $\theta(H) \lhd K$. Claim that $\theta(H) = \{1\}$. For any finite normal subgroup of K is central (see, for example the argument on p. 146 of [15]), and K has trivial centre by (1.6). Thus if $\theta(H) \neq \{1\}$, then $\mathrm{Ker}(\theta)$ is a normal subgroup of infinite index in H. By a theorem of Margulis [27], $\mathrm{Ker}(\theta) = \{1\}$, since $\mathrm{rank}_{\mathbb{R}}(L) \geqq 2$. Thus H is

isomorphic to a subgroup of K, hence $K \leqq H$, which contradicts $H < K$. Thus $\theta(H) = \{1\}$. This proves ($\mathfrak{P}$.3) for $\mathfrak{D}$, and we have already proved ($\mathfrak{P}$.1) and ($\mathfrak{P}$.2) in (1.6). Q.E.D.

Our primary interest is in products of Surface groups. Let SURFACE (resp. FREE) denote the class of fundamental groups of closed surfaces of genus $\geqq 2$ (resp. finitely generated free groups of rank $\geqq 2$). Note that the union of SURFACE and FREE is just the class of torsion free discrete subgroups of finite covolume in $PGL_2(\mathbb{R})$.

Proposition 2.7. SURFACE$\cup$FREE *satisfies* ($\mathfrak{P}$.1), ($\mathfrak{P}$.2), *and* ($\mathfrak{P}$.3).

Proof. Let F_n denote the free group of rank n, and put

$$S^-_{n+1} = \langle c_0, c_1, \ldots, c_n : c_0^2 c_1^2 \ldots c_n^2 = 1 \rangle$$

$$S^+_{2n} = \left\langle x_1, \ldots, x_n, y_1, \ldots, y_n : \prod_{i=1}^{n} x_i y_i x_i^{-1} y_i^{-1} = 1 \right\rangle.$$

Then SURFACE$\cup$FREE $= \{F_n, S^-_{n+1}, S^+_{2n} : n \geqq 2\}$. The ordering we take is

$$F_2 < S^-_3 < F_3 < S^-_4 < S^+_4 < F_4 < \ldots < F_{2n-1} < S^-_{2n} < S^+_{2n} < F_{2n} < S^-_{2n+1} < F_{2n+1}.$$

Note that, in each case, the suffix records the rank of the given group. Now there are no epimorphisms of the form

(i) $\varphi : S^-_{2n} \twoheadrightarrow S^+_{2n}$ $(n \geqq 2)$,

(ii) $\varphi : S^+_{2n} \twoheadrightarrow F_{2n}$ $(n \geqq 2)$

or

(iii) $\varphi : S^-_n \twoheadrightarrow F_n$ $(n \geqq 3)$.

To see (i), observe that an epimorphism $S^-_{2n} \twoheadrightarrow S^+_{2n}$ would induce an epimorphism

$$H_1(S^-_{2n}; \mathbb{Z}) \twoheadrightarrow H_1(S^+_{2n}; \mathbb{Z}).$$

However this is impossible since

$$H_1(S^-_{2n}; \mathbb{Z}) \cong \mathbb{Z}^{2n-1} \oplus (\mathbb{Z}/2\mathbb{Z}) \quad \text{and} \quad H_1(S^+_{2n}; \mathbb{Z}) \cong \mathbb{Z}^{2n}.$$

To see (ii), let $\psi : F_{2n} \twoheadrightarrow S^+_{2n}$ be the epimorphism defined by the standard presentation. Then the existence of an epimorphism $\varphi : S^+_{2n} \twoheadrightarrow F_{2n}$ would imply the existence of an epimorphism $\varphi\psi : F_{2n} \twoheadrightarrow F_{2n}$ with nontrivial kernel, contradicting the Hopfian property of F_{2n} [22, p. 109]. (iii) is similar to (ii).

Thus to prove ($\mathfrak{P}$.3), it suffices to show that

(iv) There is no nontrivial homomorphism $\theta : H \to K$ with $\theta(H) \lhd K$ and $\theta(H) \neq K$, if $H \leqq K$.

To show (iv) it is enough to prove

(v) For all $K \in$ SURFACE$\cup$FREE, if $K_1 \lhd K$ with $K_1 \neq \{1\}$ and $K_1 \neq K$, then $\operatorname{rank}(K_1) > \operatorname{rank}(K)$.

To prove (v), first suppose that K_1 has infinite index in K. Then K_1 is a free group of infinite rank, whence the result. For K free this is proved in [22, Theorem 2.10, p. 104], whilst for K a Surface group this is proved in [12]. Thus it

suffices to consider the case where K_1 has finite index in K, and here the result follows easily from the formula for the Euler characteristic of a finite covering, in this case

$$\chi(K_1)=[K:K_1]\chi(K)$$

and the relation between rank and Euler characteristic. For a free group K we have $\chi(K)=1-\text{rank}(K)$, whilst for a Surface group we get $\chi(K)=2-\text{rank}(K)$, noting of course that the type of a group, FREE or SURFACE, is preserved upon passing to subgroups of finite index.

This proves ($\mathfrak{P}$.3), and we have already established ($\mathfrak{P}$.1) and ($\mathfrak{P}$.2) in (1.6). Q.E.D.

Finally, we can combine the results of (2.6) and (2.7) in

Proposition 2.8. *For any finite subclass* $\mathfrak{D}$ *of* $\mathfrak{L}$, $\mathfrak{D}\cup$SURFACE$\cup$FREE *satisfies* ($\mathfrak{P}$.1), ($\mathfrak{P}$.2), *and* ($\mathfrak{P}$.3).

Proof. Take the individual orderings on $\mathfrak{D}$ and SURFACE$\cup$FREE given by (2.6) and (2.7) and combine them by making $\mathfrak{D}$ entirely precede SURFACE$\cup$FREE. To check condition ($\mathfrak{P}$.3), let $H\in\mathfrak{D}$ and $K\in$SURFACE$\cup$FREE, and let $\theta:H\to K$ be a homomorphism with $\theta(H)\lhd K$. θ cannot be injective otherwise H would be isomorphic to a Surface group or free group of rank $\geqq 2$, and would thus have a proper normal subgroup of infinite index, contradicting a Theorem of Margulis for $\mathfrak{D}$ [29]. Thus $\text{Ker}(\theta)\neq\{1\}$, and by the aforesaid Theorem of Margulis, must have finite index in H. But K is torsion free, so that $\text{Ker}(\theta)=H$, and $\theta(H)=\{1\}$. Q.E.D.

Finally, let $[\ ]\colon \text{Aut}(G)\to\text{Aut}(G)/\text{Inn}(G)=\text{Out}(G)$ be the identification map. The following is obvious from (2.2).

Corollary 2.9. *Let* $(A_i, G_i, p_i)_{1\leq i\leq n}$ *be a* $\mathfrak{C}$*-product structure on a group* G, *where* $\mathfrak{C}$ *is a class of groups satisfying* ($\mathfrak{P}$.1), ($\mathfrak{P}$.2), *and* ($\mathfrak{P}$.3). *Then there is an isomorphism* φ *making the following diagram commute;*

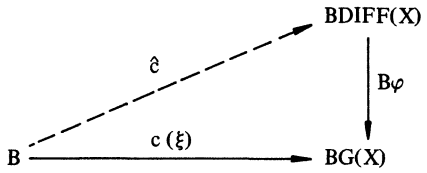
$$\begin{array}{ccc} \prod_{i=1}^{n} \text{Aut}(G_i)\wr\sigma_{p_i} & \xrightarrow{\natural} & \text{Aut}(G) \\ \downarrow{\scriptstyle \prod[\ ]_i} & & \downarrow{\scriptstyle [\ ]} \\ \prod_{i=1}^{n} \text{Out}(G_i)\wr\sigma_{p_i} & \xrightarrow{\varphi} & \text{Out}(G). \end{array}$$

3. Weak Rigidity and Realisations of Extensions

Let X be a smooth closed manifold, $G(X)$ its monoid of homotopy equivalences with the compact open topology, and DIFF(X) its group of diffeomorphisms, topologised as in [20]. The inclusion $\varphi:\text{DIFF}(X)\hookrightarrow G(X)$ induces a map between classifying spaces, $B\varphi:B\,\text{DIFF}(X)\to BG(X)$, where BM is Milgram's classifying space for a topological monoid M [28].

Definition. A smooth closed manifold X is said to be *weakly rigid* iff $B\varphi:B\,\text{DIFF}(X)\to BG(X)$ admits a homotopy right inverse.

This definition is motivated by the following considerations. Let $\xi=(E_1\to B)$ be a Hurewicz fibration over a CW complex B with fibre homotopy equivalent to X. ξ is classified up to fibre homotopy equivalence by a homotopy class $c(\xi):B\to BG(X)$. Now suppose that X is weakly rigid, with $s:BG(X)\to B\,\mathrm{DIFF}(X)$ a homotopy right inverse to $B\varphi$. If $\hat{c}=s\circ c(\xi)$, then the following diagram commutes up to homotopy.



Let η' be the locally trivial X bundle classified by $\hat{c}$. If, in addition, B is a smooth manifold, then by standard approximation arguments, we can replace η' by an equivalent smooth fibre bundle $\eta=(E_2\to B)$ with fibre diffeomorphic to X. Note that η is still fibre homotopy equivalent to ξ. Summarising, we obtain

Proposition 3.1. *Let X be a smooth closed weakly rigid manifold, B a smooth manifold, and $\xi=(E_1\to B)$ a Hurewicz fibration whose fibre is homotopy equivalent to X. Then ξ is fibre homotopy equivalent to a smooth locally trivial fibre bundle $\eta=(E_2\to B)$ with fibre diffeomorphic to X.*

Definition. We say that a Poincaré Duality (=PD) group G is *smoothly realisable* when there is a smooth closed aspherical manifold X_G with $\pi_1(X_G)=G$. X_G is called a smooth model for G. More generally, we say that an exact sequence of PD groups $\mathfrak{E}=(K\rightarrowtail\Gamma\twoheadrightarrow Q)$ is smoothly realisable when there exists a smooth locally trivial fibre bundle $p:X\to X_Q$ with fibre X_K, where X_K, X_Γ, X_Q are smooth models for K, Γ, and Q respectively, and such that the long homotopy exact sequence of $\eta=(X_\Gamma, X_Q, p, X_K)$ reduces to $\mathfrak{E}$.

A PD group Γ is said to be weakly rigid with X_Γ as model iff Γ has a smooth realisation with a weakly rigid manifold X_Γ as model.

Theorem 3.2. *Let $\mathfrak{E}=(K\rightarrowtail\Gamma\twoheadrightarrow Q)$ be a short exact sequence of* PD *groups in which*
(i) *K and Q are smoothly realisable with models X_K and X_Q respectively, and*
(ii) *X_K is weakly rigid.*
Then $\mathfrak{E}$ has a smooth realisation with X_K, X_Q as models for the fibre and base respectively, and in particular, Γ has a smooth realisation.

Proof. If H is a discrete group, let $(H\to WH\to\overline{WH})$ be the Eilenberg-Mac-lane principal simplicial H-bundle with contractible total space. Applying $\bar{W}$ to a short exact sequence $\mathfrak{E}=(K\rightarrowtail\Gamma\twoheadrightarrow Q)$ yields a minimal Kan fibration $\overline{W\mathfrak{E}}=(\overline{WK}\to\overline{W\Gamma}\to\overline{WQ})$ which in turn, upon taking geometrical realisations, yields a fibre bundle which is locally trivial in the category of compactly generated spaces [9, p. 55] thus; $|\overline{W\mathfrak{E}}|=(|\overline{WK}|\to|\overline{W\Gamma}|\to|\overline{WQ}|)$. Now suppose that there is a smooth manifold X_Q of homotopy type $K(Q,1)\simeq|\overline{WQ}|$. Replacing the base $|\overline{WQ}|$ of $|\overline{W\mathfrak{E}}|$ by X_Q, we obtain a fibre bundle $\xi=(|\overline{WK}|\to E_1\to X_Q)$, which, since X_Q is locally compact, is locally trivial in the category of (all) topological spaces, hence is

a Hurewicz fibration. Observe that the long homotopy exact sequence of ξ reduces to the original exact sequence $\mathfrak{E}$.

Finally suppose that $\mathfrak{E}=(K\rightarrowtail\Gamma\twoheadrightarrow Q)$ is a short exact sequence of PD groups such that K and Q are smoothly realisable by X_K and X_Q respectively, and such that X_K is weakly rigid. By (3.1), we may replace the fibration ξ constructed above by some smooth locally trivial fibre bundle $\eta=(X_\Gamma, X_Q, p, X_K)$. In particular, Γ is smoothly realisable with X_Γ as model. Q.E.D.

At this point it is appropriate to observe that, for any $K(\Gamma,1)$ space X_Γ, the homotopy type of $BG(X_\Gamma)$ is quite simple to describe. Namely, it is the total space of a fibration $(K(\mathfrak{Z},2)\rightarrow BG(X_\Gamma)\rightarrow B\mathrm{Out}(\Gamma))$, where $\mathfrak{Z}$ is the centre of Γ,

$$\mathrm{Out}(\Gamma)=\mathrm{Aut}(\Gamma)/\mathrm{Inn}(\Gamma)$$

is the outer automorphism group of Γ, and where the action of $\mathrm{Out}(\Gamma)$ on $\mathfrak{Z}$ is the obvious one, induced by evaluation of $\mathrm{Aut}(\Gamma)$ on Γ. In the case where Γ is centreless, $BG(X_\Gamma)$ is homotopy equivalent to $B\mathrm{Out}(\Gamma)$. Compare [10].

Many geometrically interesting groups are weakly rigid. We give examples in Sects. 4 and 5. We end this section by proving some basic results on weak rigidity and smooth realisations.

Proposition 3.3. *Suppose that H, K are centreless weakly rigid groups with models X_H and X_K, and suppose that the natural map*

$$\natural:\mathrm{Aut}(H)\times\mathrm{Aut}(K)\rightarrow\mathrm{Aut}(H\times K)$$

is an isomorphism. Then $H\times K$ is weakly rigid with $X_H\times X_K$ as model.

Proof. Clearly $\natural:\mathrm{Aut}(H)\times\mathrm{Aut}(K)\xrightarrow{\cong}\mathrm{Aut}(H\times K)$ induces an isomorphism

$$\natural:\mathrm{Out}(H)\times\mathrm{Out}(K)\xrightarrow{\cong}\mathrm{Out}(H\times K).$$

Let j denote the canonical inclusion

$$j:\mathrm{DIFF}(X_H)\times\mathrm{DIFF}(X_K)\hookrightarrow\mathrm{DIFF}(X_H\times X_K).$$

We get a homotopy commutative diagram as below

$$\begin{array}{ccccc}
B\mathrm{DIFF}(X_H)\times B\mathrm{DIFF}(X_K) & \xrightarrow{i} & B(\mathrm{DIFF}(X_H)\times\mathrm{DIFF}(X_K)) & \xrightarrow{j} & B\mathrm{DIFF}(X_H\times X_K)\\
\downarrow{\scriptstyle\lambda_1} & & \downarrow{\scriptstyle\lambda_2} & & \downarrow{\scriptstyle\lambda_3}\\
B\mathrm{Out}(H)\times B\mathrm{Out}(K) & \xrightarrow{h} & B(\mathrm{Out}(H)\times\mathrm{Out}(K)) & \xleftarrow{B(\natural^{-1})} & B\mathrm{Out}(H\times K),
\end{array}$$

where the λ_r are induced by the natural maps $\mathrm{DIFF}(Y)\rightarrow\mathrm{Out}(\pi_1(Y))$ and where i and h are homotopy equivalences.

By hypothesis, λ_1 has a homotopy right inverse. Hence so also does λ_2. Let μ be a homotopy right inverse for λ_2. Then $\nu=j\circ\mu\circ B(\natural^{-1})$ is the required homotopy right inverse. Q.E.D.

Proposition 3.4. *Let Γ be a* PD *group with trivial centre, which is not decomposable as a nontrivial direct product, and which is smoothly realisable with X_Γ as model. Suppose also that $\varphi:\mathrm{DIFF}(X_\Gamma)\rightarrow\mathrm{Out}(\Gamma)\simeq G(X_\Gamma)$ is a homotopy equivalence. Then*

for each $n \geqq 1$, $\Gamma^{(n)}$ *is weakly rigid with*

$$X_{\Gamma^{(n)}} = \underbrace{X_\Gamma \times \ldots \times X_\Gamma}_{n}$$

as model.

Proof. First note that the hypothesis that $\varphi : \mathrm{DIFF}(X_\Gamma) \to \mathrm{Out}(\Gamma)$ is a homotopy equivalence implies that $B\varphi : B\,\mathrm{DIFF}(X_\Gamma) \to B\,\mathrm{Out}(\Gamma)$ is a homotopy equivalence, and hence that X_Γ is weakly rigid. $\mathrm{DIFF}(X_{\Gamma^{(n)}})$ has an obvious subgroup $S(\Gamma, n)$ defined as follows; $f \in S(\Gamma, n)$ iff there are diffeomorphisms $(f_i)_{1 \leqq i \leqq n}$, $f_i \in \mathrm{DIFF}(X_\Gamma)$ such that, for some $\sigma \in \sigma_n$

$$f(x_1, \ldots, x_n) = (f_1(x_{\sigma(1)}, \ldots, f_n(x_{\sigma(n)}).$$

Clearly $S(\Gamma, n) \cong \mathrm{DIFF}(X_\Gamma) \wr \sigma_n$. Indeed there is a commutative diagram

$$\begin{array}{ccccc} \mathrm{DIFF}(X_\Gamma) \wr \sigma_n & \xrightarrow{\simeq} & S(\Gamma, n) & \xrightarrow{i} & \mathrm{DIFF}(X_{\Gamma^{(n)}}) \\ \downarrow{\scriptstyle \varphi \wr i} & & \downarrow{\scriptstyle \varphi_*} & \swarrow & \\ \mathrm{Out}(\Gamma) \wr \sigma_n & \xrightarrow{\natural} & \mathrm{Out}(\Gamma^{(n)}) & & \end{array}$$

in which $\natural$ is an isomorphism by (1.1) and (2.9), and in which the vertical arrows are homotopy equivalences by virtue of the fact that the natural map

$$\varphi : \mathrm{DIFF}(X_\Gamma) \to \mathrm{Out}(\Gamma)$$

is a homotopy equivalence. Let

$$\mu : B\,\mathrm{Out}(\Gamma^{(n)}) \to BS(\Gamma, n)$$

be a homotopy inverse for $B\varphi_* : BS(\Gamma, n) \to B\,\mathrm{Out}(\Gamma^{(n)})$. Then $j = (Bi)\mu$ is the desired homotopy right inverse. Q.E.D.

Recall that if $\mathfrak{A}$ is a class of groups, then by a poly-$\mathfrak{A}$ filtration of length n on a group G we mean a sequence $(G_r)_{0 \leqq r \leqq n}$ of subgroups such that

(i) $1 = G_0 \subset G_1 \subset \ldots G_{n-1} \subset G_n = G$, and

(ii) for each r, $G_r \lhd G_{r+1}$ and $G_{r+1/G_r} \in \mathfrak{A}$ for $0 \leqq r \leqq n-1$.

Such a filtration is called characteristic if, in addition, each G_r is a characteristic subgroup of G_{r+1}. Moreover, we shall call a filtration *strongly poly-*$\mathfrak{A}$ of length n iff $G_1 \lhd G$ and $(G_{r+1}/G_1)_{0 \leqq r \leqq n-1}$ is a strongly poly-$\mathfrak{A}$ filtration of length $(n-1)$ on G/G_1.

Clearly we have (characteristically poly-$\mathfrak{A}$) $\Rightarrow$ (strongly poly-$\mathfrak{A}$) $\Rightarrow$ (poly-$\mathfrak{A}$). From (3.2) it is immediate by induction on the length of a filtration that

Proposition 3.5. *A strongly poly-{weakly rigid} group is smoothly realisable.*

To conclude this section, we note that the concept of weak rigidity given here is different from, and supercedes, the now obsolete concept with the same name given in [14].

4. Weak Rigidity of Products of Surface Groups

Let Γ be a finite product of Surface groups. We may write Γ in the form $\Gamma=\Gamma_1^{(n_1)}\times\ldots\times\Gamma_m^{(n_m)}$ where $(\Gamma_i)_{1\leqq i\leqq m}$ is a sequence of Surface groups, and $(n_i)_{1\leqq i\leqq m}$ is a sequence of positive integers, and $\Gamma_i\ncong\Gamma_j$ if $i\neq j$. Let X_{Γ_i} be the smooth closed 2-manifold with $\pi_1(X_{\Gamma_i})=\Gamma_i$. Then Γ is smoothly realisable with the following standard model for $K(\Gamma,1)$;

$$X_\Gamma=X_{\Gamma_1}^{(n_1)}\times\ \ldots\ \times X_{\Gamma_m}^{(n_m)}.$$

Proposition 4.1. *Let Γ be a finite product of Surface groups. Then Γ is weakly rigid in its standard model.*

Proof. Adhering to the conventions above, write

$$\Gamma=\Gamma_1^{(n_1)}\times\ \ldots\ \times\Gamma_m^{(n_m)}\quad\text{and}\quad X_\Gamma=X_{\Gamma_1}^{(n_1)}\times\ \ldots\ \times X_{\Gamma_m}^{(n_m)}.$$

By definition of "Surface group", each X_{Γ_i} has genus $\geqq 2$. In particular, each Γ_i has trivial centre, and so Γ has trivial centre. We show that Γ is weakly rigid by induction on m.

If $m=1$, put $K=\Gamma_1$ and $n=n_1$. Then $G(X_K)\simeq\mathrm{Out}(K)$ since K has trivial centre. Also the identity component of $\mathrm{DIFF}(X_K)$ is contractible, by the theorem of Gramain [11] and Eells and Earle [6]. The topological and *PL* versions of this result are due to Hamstrom [13], Scott [27], and Akiba [1]. It follows that $\mathrm{DIFF}(X_K)$ is homotopy equivalent to $\pi_0(\mathrm{DIFF}(X_K))$ and by earlier work of Baer [3] and Mangler [23], $\pi_0(\mathrm{DIFF}(X_K)\cong\mathrm{Out}(K)$. See also the paper of Epstein [7]. Thus $\mathrm{DIFF}(X_K)$ is homotopy equivalent to $\mathrm{Out}(K)$. By (1.6) and (3.4), $K^{(n)}$ is weakly rigid with $X_K^{(n)}$ as model, and this proves the case where $m=1$.

Suppose that the theorem is proved for $H=\Gamma_1^{(n_1)}\times\ \ldots\ \times\Gamma_{m-1}^{(n_{m-1})}$ and let $\Gamma=H\times K$ where $K=\Gamma_m^{(n_m)}$, for some Surface groups Γ_i, with $\Gamma_i\ncong\Gamma_j$ if $i\neq j$, and where each $n_i>0$. By induction and the case $m=1$ already proved, the theorem is true for H and K individually. Moreover, by (2.8) and (2.2), the natural map

$$\natural:\mathrm{Aut}(H)\times\mathrm{Aut}(K)\to\mathrm{Aut}(\Gamma)$$

is an isomorphism, so that the result follows from (3.3). Q.E.D.

Corollary 4.2. *Let $1\to K\to\Gamma\to Q\to 1$ be an exact sequence of* PD *groups in which K is a finite product of Surface groups, and Q is smoothly realisable. Then Γ is smoothly realisable.*

Proof. Immediate from (3.2) and (4.1). Q.E.D.

In [16] we showed that every poly-Surface group has a subgroup of finite index which is smoothly realisable. The problem of realising all poly-Surface groups is still outstanding however, and seems at present to be rather intricate.

The problem begins in dimension six, that is, with poly-Surface length three: a poly-Surface group of length n has cohomological dimension $2n$, and the four-dimensional groups were realised in [16]. To end this section, we use the results of Sects. 1–3 to realise some 6-dimensional groups neglected in [16].

Let $\leqq$ be the ordering on SURFACE introduced in (2.7). Then the following is true.

Proposition 4.3. *Let* $\mathfrak{E}=\left(1\to G_1 \xrightarrow{i} G \xrightarrow{p} Q\to 1\right)$ *be an extension with* G_1, Q *Surface groups satisfying* $G_1 \leqq Q$. *Then either* (i) $i(G_1)$ *is characteristic in* G *or or* (ii) $\mathfrak{E}$ *is trivial,* $G_1 \cong Q$ *and* $G \cong G_1 \times G_1$.

Proof. Suppose that $i(G_1)$ is not characteristic in G and let $\alpha: G \to G$ be an automorphism which does not preserve $\mathfrak{E}$. Put $\theta = p\alpha i$. Then $\theta(G_1) \neq \{1\}$. However, $\theta(G_1) \lhd Q$ so that if $G_1 < Q$ we get a contradiction. Hence $G_1 \cong Q$. Now $\theta(G_1)$ cannot have index greater than one in Q, since as in the proof of (2.7), this would imply that

$$\operatorname{rank}(\theta(G_1)) > \operatorname{rank}(G_1) = \operatorname{rank}(Q).$$

Hence $\theta(G_1) = Q$. Put $H_1 = i(G_1)$ and $H_2 = \alpha i(G_1)$. Then $H_1, H_2 \lhd G$ and $H_1 \cdot H_2 = G$. Also we have an exact sequence

$$1 \to H_1 \cap H_2 \to H_2 \xrightarrow{p} Q \to 1.$$

If $H_1 \cap H_2 \neq \{1\}$, then Q is a proper quotient of $H_2 \cong Q$, contradicting the Hopfian property of the Surface group Q [8]. Hence $G = H_1 \circ H_2 \cong G_1 \times Q$, and the extension $\mathfrak{E}$ is trivial. An isomorphism $\gamma: G_1 \times G_1 \to G$ is given by

$$\gamma(x, y) = i(x) \cdot \alpha i(y). \quad \text{Q.E.D.}$$

Corollary 4.4. *Let* G *be a poly-Surface group constructed from exact sequences as follows;*

(i) $1 \to G_1 \to G_2 \to Q_1 \to 1$

(ii) $1 \to G_2 \to G \to Q_2 \to 1$,

where G_1, Q_1, *and* Q_2 *are Surface groups with* $G_1 \leqq Q_1$.

Then G *is smoothly realisable.*

Proof. Suppose that G_1 is characteristic in G_2. Then G is strongly poly-Surface, hence is smoothly realisable by (3.5). If G_1 is not characteristic, then by (4.3), $G_2 \cong G_1 \times G_1$, so that G_2 is weakly rigid by (3.3), and G is smoothly realisable by (3.2). Q.E.D.

This is an improvement on the results of [16], where we needed to assume that $\operatorname{rank}(G_1) < \operatorname{rank}(Q_1)$. The remaining cases with $Q_1 < G_1$ will be treated in a subsequent paper.

5. Weak Rigidity of Discrete Uniform Subgroups

In Sect. 4 we showed that finite products of Surface groups are weakly rigid. By Uniformisation Theory for complex curves, orientable Surface groups can be regarded as discrete uniform subgroups of the connected group $PSL_2(\mathbb{R})$. Non-orientable Surface groups are discrete uniform subgroups of the two component group $PGL_2(\mathbb{R})$. Consequently, a finite product of Surface groups is a discrete uniform subgroup of a finite product of copies of $PGL_2(\mathbb{R})$. To conclude this paper, we investigate the weak rigidity of subgroups of more general Lie groups. To begin with, we note the following, which is, ignoring holomorphic arguments, the essential content of [5].

Proposition 5.1 (Conner and Raymond [5]). *$\mathbb{Z}^n$ is weakly rigid with $\mathbb{Z}^n\backslash\mathbb{R}^n$ as model.*

From now on, for simplicity, we concentrate on discrete uniform subgroups in connected semisimple Lie groups. Since we shall make heavy use of Mostow's Strong Rigidity theorem [24], to clarify matters we make the following definition.

Definition. By a *Mostow pair* (G, Γ) we mean

(i) a connected semisimple adjoint Lie group G, having no compact factors

(ii) a torsion free discrete uniform subgroup Γ of G such that

(iii) for any epimorphism $\pi: G \to PSL_2(\mathbb{R})$, $\pi(\Gamma)$ is non-discrete in $PSL_2(\mathbb{R})$.

If (G, Γ) is a Mostow pair, we put $X_\Gamma = \Gamma\backslash G/K$ where K is a maximal compact subgroup of G. Our first step is to show that Γ is weakly rigid with X_Γ as model. This is such a routine consequence of Mostow's theorem that we may reasonably attribute it to Mostow. However, since the proof does not seem to be written down anywhere, we give it below. To begin with, we need the following which, though presumably very well known, likewise seems to lack a convenient reference in the literature.

Lemma 5.2. *Let G be a connected semisimple noncompact Lie group, and let K be a maximal compact subgroup. Then K is self-normalising.*

Proof. Let $\mathfrak{g} = \mathfrak{k} \oplus \mathfrak{p}$ be the corresponding Cartan decomposition of the Lie algebra $\mathfrak{g}$ of G. The Killing form $B_\mathfrak{g}$ of $\mathfrak{g}$ is positive definite on $\mathfrak{p}$ and negative definite on $\mathfrak{k}$. First suppose that $y \in \mathfrak{p}$ centralises $\mathfrak{k}$. We show that $y = 0$. For $[y, \mathfrak{p}] \subset \mathfrak{k}$ since $[\mathfrak{p}, \mathfrak{p}] \subset \mathfrak{k}$, and $[y, \mathfrak{k}] = 0$. Thus $\mathrm{ad}(y)\,\mathrm{ad}(y) = 0$ so that $B_\mathfrak{g}(y, y) = \mathrm{Tr}(\mathrm{ad}(y)^2) = 0$. Hence $y = 0$.

Now let $x \in N_\mathfrak{g}(\mathfrak{k}) = \{z \in \mathfrak{g} : [z, \mathfrak{k}] \subset \mathfrak{k}\}$. Write $x = x_0 + x_1$, $x_0 \in \mathfrak{k}$, $x_1 \in \mathfrak{p}$. Since $[\mathfrak{k}, \mathfrak{k}] \subset \mathfrak{k}$ and $[\mathfrak{k}, \mathfrak{p}] \subset \mathfrak{p}$ then x_1 centralises $\mathfrak{k}$ so that $x_1 = 0$ by above. Hence $x = x_0 \in \mathfrak{k}$. Thus $\mathfrak{k} = N_\mathfrak{g}(\mathfrak{k})$. Consequently $K = N_G(K) = \{g \in G : gKg^{-1} = K\}$. Q.E.D.

Theorem 5.3. *Let (G, Γ) be a Mostow pair. Then Γ is weakly rigid with $X_\Gamma = \Gamma\backslash G/K$ as model, where K is a maximal compact subgroup of G.*

Proof. Put $X = G/K$, $X_\Gamma = \Gamma\backslash X$, and let $x = [1] = K \in G/K$, so that

$$K = G_x = \{g \in G : g \cdot x = x\}.$$

If $\alpha \in \mathrm{Aut}(G)$ then $\alpha(K)$ is a maximal compact subgroup of G, and since any two maximal compact subgroups are conjugate, we may choose $h_\alpha \in G$ such that $\alpha(K) = h_\alpha K h_\alpha^{-1}$. We define a diffeomorphism $\alpha_* : X \to X$ by means of

$$\alpha_*(g \cdot x) = \alpha(g) \cdot h_\alpha \cdot x.$$

This definition does not depend on the particular representation of a point in X in the form $g \cdot x$. No more does it depend on the particular choice of $h_\alpha \in G$. For if $h_\alpha K h_\alpha^{-1} = k_\alpha K k_\alpha^{-1}$, then since K is self-normalising, $k_\alpha = h_\alpha \cdot k$ for some $k \in K$. Thus $\alpha(g) \cdot k_\alpha \cdot x = \alpha(g) \cdot h_\alpha \cdot x$ since $k \cdot x = x$. Moreover, it is straightforward to see that $(\alpha\beta)_* = \alpha_* \beta_*$. Hence we have a homomorphism $\mathrm{Aut}(G) \to \mathrm{Diff}(X)$, $\alpha \mapsto \alpha_*$.

Now appeal to Mostow's theorem. Under our hypotheses, an automorphism $\alpha : \Gamma \to \Gamma$ extends uniquely to an automorphism $\bar{\alpha} : G \to G$. Thus we get a homomorphism $\mathrm{Aut}(\Gamma) \to \mathrm{Diff}(X)$ by means of $\alpha \mapsto (\bar{\alpha})_*$. This in turn gives rise to a

homomorphism $\psi : \mathrm{Aut}(\Gamma) \to \mathrm{Diff}(X_\Gamma)$ as follows:

$$\psi(\alpha)[\Gamma \cdot g \cdot x] = \Gamma \cdot (\bar{\alpha})_*(g \cdot x) = \Gamma \cdot \bar{\alpha}(g) \cdot h_{\bar{\alpha}} \cdot x .$$

Finally note that if $\alpha \in \mathrm{Inn}(G)$ then $\psi(\alpha) = \mathrm{Id}$. For if $\alpha(g) = zgz^{-1}$, $z \in \Gamma$, we may take $h_{\bar{\alpha}} = z$ so that

$$\psi(\alpha)[\Gamma \cdot g \cdot x] = \Gamma \cdot z \cdot g \cdot x = \Gamma \cdot g \cdot x, \quad \text{since} \quad z \in \Gamma .$$

It is now easy to check that the homomorphism $\psi_* : \mathrm{Out}(\Gamma) \to \mathrm{DIFF}(X_\Gamma)$ induced by ψ is a right inverse to the canonical map $\varphi : \mathrm{DIFF}(X_\Gamma) \to \pi_0(G(X_\Gamma)) = \mathrm{Out}(\Gamma)$. Since Γ is centreless (see Corollary (5.18), p. 84 of [26]), $B\mathrm{Out}(\Gamma)$ is homotopy equivalent to $BG(X_\Gamma))$, so that the induced map

$$B\psi_* : BG(X_\Gamma)) \cong B\,\mathrm{Out}(\Gamma) \to B\,\mathrm{DIFF}(X_\Gamma)$$

is the required homotopy right inverse. Q.E.D.

Of the two approaches we have given to weak rigidity, one uses low dimensional techniques, generalisation of which to higher dimensions seem very difficult, and one uses Mostow's Strong Rigidity theorem, which works well except in low dimensions, precisely except when G has a $PSL_2(\mathbb{R})$ factor in which the projection of Γ is discrete, in which case it is known to be false. Let us examine how far we are from being able to treat the general case. Recall [26, p. 86] that if G is a connected adjoint semisimple Lie group and P is a discrete uniform subgroup of G then P is called irreducible iff $\pi(P)$ is non-discrete for any projection π of G onto a proper direct factor. In general, if P is a discrete uniform subgroup in such a G, then G admits a direct product decomposition $G = G_1 \circ \ldots \circ G_m$ into connected normal subgroups G_i, such that $P_i = G_i \cap P$ is irreducible in G_i, and such that $P_0 = P_1 \circ \ldots \circ P_m$ has finite index in P. G_i need not be simple. By collecting factors, we may re-express this as follows, separating factors G_i with $\mathbb{R}$-rank $\geqq 2$ from factors with $\mathbb{R}$-rank $= 1$; P has a normal subgroup P_0 of finite index in P such that P_0 is a direct product $P_0 = H \times H' \times H''$ where H is a product of irreducible discrete uniform subgroups in factors G_i with $\mathbb{R}$-rank$(G_i) \geqq 2$, H' is a product of Surface groups, and H'' is a finite product $L_1 \times \ldots \times L_p$, where each L_i is a discrete uniform subgroup in a connected adjoint Lie group of $\mathbb{R}$-rank 1 which is *not* isomorphic to $PSL_2(\mathbb{R})$. That is, L_i is a discrete uniform subgroup in some $PSO(n,1)$ $(n \geqq 3)$, $PSU(n,1)$ $(n \geqq 2)$, $P\mathrm{Sp}(n,1)$ $(n \geqq 1)$, or in the isometry group of the noncompact symmetric space $(f_{4(-20)}, \mathfrak{SO}(9))$. Then P/P_0 and H'' measure the deviation of our hypotheses from the ideal in the following sense.

Theorem 5.4. *Let $P = H \times H'$ be a product group in which the factors satisfy the following:*

(i) *(G, H) is a Mostow pair for some G such that $\pi(H)$ is non-discrete for any epimorphism $\pi : G \to L$ onto a Lie group L with $\mathbb{R}$-rank$(L) = 1$, and*

(ii) *H' is a finite product of Surface groups.*

Then P is weakly rigid. In particular, for any exact sequence

$$\mathfrak{E} = (1 \to H \times H' \to G \to Q \to 1),$$

if Q is smoothly realisable, then so is G.

Proof. The hypothesis on H in (i) above is equivalent to saying that in the decomposition of a subgroup H_0 of finite index in H, into irreducible factors, each of the irreducible factors is irreducible in a Lie group of $\mathbb{R}$-rank $\geqq 2$.

Let $q: P \to H'$ be the projection. Observe that any non-trivial subgroup A of H' has an infinite cyclic summand in $H_1(A:\mathbb{Z})$. This is obvious by projection of A into the various Surface group factors of H'. However, by a theorem of Kazdan [19] and Wang [30], $H_1(H:\mathbb{Z})$ is finite. $H_1(-:\mathbb{Z})$ preserves surjections however, whence it follows that for each automorphism α of P, $q\alpha(H \times \{1\})$ is trivial, and hence that $H \times \{1\}$ is characteristic in P. However, H has trivial centre (Corollary (5.18) of [26]), so that the natural map $\natural: \mathrm{Aut}(H) \times \mathrm{Aut}(H') \to \mathrm{Aut}(P)$ is an isomorphism, by (2.4). Hence, by (3.3), P is weakly rigid. The realisation clause now follows from (3.2). Q.E.D.

Observe that the degenerate case of the above where $H = \{1\}$ is true by (4.1), and the degenerate case where $H' = \{1\}$ follows from (5.3), which is a more general statement.

References

1. Akiba, T.: On the homotopy type of PL_2. J. Fac. Sci. Tokyo Univ. **14**, 197–204 (1967)
2. Auslander, L., Johnson, F.E.A.: On a conjecture of C.T.C. Wall. J. London Math. Soc. **14**, 331–332 (1976)
3. Baer, R.: Isotopie von Kurven auf orientierbaren, geschlossenen Flächen und ihr Zusammenhang mit der topologischen Deformation der Flächen. J. Reine angew. Math. **159**, 101–111 (1928)
4. Borel, A.: On the automorphisms of certain subgroups of semisimple Lie groups. In: Algebraic geometry. Papers presented at the Bombay Colloquium, 1968, pp. 43–73. London, Bombay: Oxford University Press 1969
5. Conner, P.E., Raymond, F.: Holomorphic Seifert fibering. In: Proceedings of the Second Conference on Compact Transformation Groups (Amherst, 1971), pp. 124–204. Lecture Notes in Mathematics, No. 299. Berlin, Heidelberg, New York: Springer 1972
6. Earle, C.J., Eells, J.: The diffeomorphism group of a compact Riemann surface. Bull. Am. Math. Soc. **73**, 557–559 (1967)
7. Epstein, D.B.A.: Curves on 2-manifolds and isotopies. Acta Math. **115**, 83–107 (1966)
8. Frederick, K.N.: The Hopfian property for a class of fundamental groups. Comm. Pure Appl. Math. **16**, 1–8 (1963)
9. Gabriel, P., Zisman, M.: Calculus of fractions and homotopy theory. Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 35. Berlin, Heidelberg, New York: Springer 1967
10. Gottlieb, D.H.: A certain subgroup of the fundamental group. Am. J. Math. **87**, 840–856 (1965)
11. Gramain, A.: Le type d'homotopie du groupe des diffeomorphismes d'une surface compacte. Ann. Sci. Ecole Norm. Sup. **6**, 53–66 (1973)
12. Greenberg, L.: Discrete groups of motions. Canad. J. Math. **12**, 415–425 (1960)
13. Hamstrom, M.E.: Homotopy groups of the space of homeomorphisms on a 2-manifold. Ill. J. Math. **10**, 563–573 (1966)
14. Johnson, F.E.A.: On the first end invariant of an exact sequence. Mathematika **22**, 60–70 (1975)
15. Johnson, F.E.A.: On Poincaré duality groups of poly-linear type and their realisations. Math. Z. **163**, 145–148 (1978)
16. Johnson, F.E.A.: On the realisability of poly-Surface groups. J. Pure Appl. Algebra **15**, 235–241 (1979)
17. Johnson, F.E.A.: Two conjectures in the theory of Poincaré Duality groups. Proc. Am. Math. Soc. **81**, 353–354 (1981)
18. Johnson, F.E.A., Wall, C.T.C.: On groups satisfying Poincaré duality. Ann. Math. **96**, 592–598 (1972)
19. Kazdan, D.A.: Connection of the dual space of a group with the structure of its closed subgroups. Functional Anal. Appl. **1**, 63–65 (1967)

20. Leslie, J.A.: On a differential structure for the group of diffeomorphisms. Topology **6**, 263–271 (1967)
21. Lyndon, R.C., Schupp, P.E.: Combinatorial group theory. Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 89, Berlin, Heidelberg, New York: Springer 1977
22. Magnus, W., Karass, A., Solitar, D.: Combinatorial group theory. New York: Dover 1976
23. Mangler, W.: Die Klassen topologischer Abbildungen einer geschlossenen Fläche auf sich. Math. Z. **44**, 541–554 (1939)
24. Mostow, G.D.: Strong rigidity of locally symmetric spaces. Anneals of Mathematics Studies, No. 78. Princeton: Princeton University Press 1973
25. Mostow, G.D., Siu, Y.-T.: A compact Kähler surface of negative curvature not covered by the ball. Ann. Math. **112**, 321–360 (1980)
26. Raghunathan, M.S.: Discrete subgroups of Lie groups. Ergebnisse der Mathematik und ihrer Grenzgebiete, No. 68. Berlin, Heidelberg, New York: Springer 1972
27. Scott, G.P.: The space of homeomorphisms of a 2-manifold. Topology **9**, 97–109 (1970)
28. Steenrod, N.E.: Milgram's classifying space of a topological group. Topology **7**, 349–368 (1968)
29. Tits, J.: The work of Gregori Alexandrovitch Margulis. Proc. International Congress of Mathematicians. Helsinki 1978
30. Wang, S.P.: The dual space of semisimple Lie groups. Am. J. Math. **91**, 921–937 (1969)

Received October 10, 1982; in revised form January 7, 1983

Math. Ann. 263, 365–370 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Kac-Moody Lie Algebras and the Universal Element for the Category of Nilpotent Lie Algebras

L. J. Santharoubane

Département de Mathématiques, Université de Poitiers, 40, Avenue de Recteur Pineau, F-86022 Poitiers, France

0. Introduction

The problem of classifying nilpotent Lie algebras is studied for the first time in 1891 by a student of Engel, Umlauf, who gives the complete list over C up to the dimension 6 and a certain complex family at the dimensions 7, 8 and 9 [15]. Later on, several authors worked out the subject: Chevalley [3], Dixmier [4], Morozov [11], Vergne [16].

The introduction in 1972 by Bratzlavsky [2] and Favre [5, 6] of the root systems for the nilpotent Lie algebras constitutes an important step in the classification (Gurevic applied already this notion to the study of metabelian Lie algebras [7]).

By using these root systems we establish a link between nilpotent Lie algebras and Kac-Moody Lie Algebras (which generalise semi-simple Lie algebras and are of infinite dimension).

In [12] (published in [13]) we started with the maximal rank case and in [14] we continued with certain nilpotent Lie algebras [see 2.11(b)]. Here we give the main theorem (2.10) in the general case.

All the structures are over an algebraically closed field K of characteristic 0.

1. Root System and Cartan Matrix Associated to a Nilpotent Lie Algebra

1.1. All through 1, $\mathfrak{g}$ is a nilpotent Lie algebra of finite dimension, $\mathrm{Der}\,\mathfrak{g}$ its derivation algebra, $\mathrm{Aut}\,\mathfrak{g}$ its automorphism group.

1.2. A torus on $\mathfrak{g}$ is a commutative subalgebra of $\mathrm{Der}\,\mathfrak{g}$ whose elements are semi-simple endomorphisms. We denote $\mathscr{T}_{\mathfrak{g}}$ the set of maximal tori on $\mathfrak{g}$ (i.e. tori not contained in any other tori).

1.3. **Mostow's Theorem** (4.1 of [10]). *If $T, T' \in \mathscr{T}_{\mathfrak{g}}$ there exists $\theta \in \mathrm{Aut}\,\mathfrak{g}$ such that $\theta T \theta^{-1} = T'$.*

Current address: Department of Mathematics Massachusetts, Institute of Technology, Cambridge, MA 02139, USA

1.4. Let $T\in\mathscr{T}_{\mathfrak{g}}$ and consider the root space decomposition of $\mathfrak{g}$ relatively to T:

$$\mathfrak{g}=\sum_{\beta\in R(T)}\mathfrak{g}^{\beta},$$

$$\mathfrak{g}^{\beta}=\{x\in\mathfrak{g};\,tx=\beta(t)x\ \ \forall t\in T\},$$

$$R(T)=\{\beta\in T^{*};\,\mathfrak{g}^{\beta}\neq(0)\}\quad\text{(root system)}.$$

We denote:

$$R^{1}(T)=\{\beta\in R(T);\,\mathfrak{g}^{\beta}\not\subset[\mathfrak{g},\mathfrak{g}]\},$$

$$l_{\beta}=\dim(\mathfrak{g}^{\beta}/[\mathfrak{g},\mathfrak{g}]\cap\mathfrak{g}^{\beta})\quad\forall\beta\in R^{1}(T).$$

$$d_{\beta}=\dim\mathfrak{g}^{\beta}\quad\forall\beta\in R^{1}(T).$$

The map: $\beta\mapsto d_{\beta}$ $R^{1}(T)\to\mathbb{N}^{*}$ gives the partition:

$$R^{1}(T)=R^{1}(T)_{p_1}\cup\ldots\cup R^{1}(T)_{p_q},$$

$$p_1<\ldots<p_q\quad R^{1}(T)_{p_i}\neq\emptyset,$$

$$R^{1}(T)_{p}=\{\beta\in R^{1}(T);\,d_{\beta}=p\}.$$

Let $s_i=\#R^{1}(T)_{p_i}$ and $s=s_1+\ldots+s_q$; we number the elements of $R^{1}(T):\beta_1\ldots\beta_s$ in such a way that:

$$R^{1}(T)_{p_1}=\{\beta_1\ldots\beta_{s_1}\},\qquad R^{1}(T)_{p_2}=\{\beta_{s_1+1}\ldots\beta_{s_2}\},\ldots.$$

Let $d_i=d_{\beta_i}$, $l_i=l_{\beta_i}$ and $l=l_1+\ldots+l_s$ (one checks that $l=\dim\mathfrak{g}/[\mathfrak{g},\mathfrak{g}]$). Let $\mathfrak{S}_s^{s_1\ldots s_q}$ be the group of permutations of $\{1\ldots s\}$ which leave $\{1\ldots s_1\}$, $\{s_1+1\ldots s_2\}\ldots$ invariant.

1.5. Let $T'\in\mathscr{T}_{\mathfrak{g}}$ be another maximal torus; then there exists $\theta\in\mathrm{Aut}\,\mathfrak{g}$ such that $\theta T\theta^{-1}=T'$ (1.3). For T', we use the notations of 1.4 with prime. The map:

$$\check{\theta}:T^{*}\to T'^{*},\qquad\beta\mapsto\check{\theta}\beta,\qquad\check{\theta}\beta(\theta t\theta^{-1})=\beta(t)\quad\forall\beta\in T^{*},\quad\forall t\in T,$$

is a vector space isomorphism and one has obviously:

$$\theta\mathfrak{g}^{\beta}=\mathfrak{g}^{\check{\theta}\beta}\quad\forall\beta\in R(T)$$

therefore $d'_{\check{\theta}\beta}=d_{\beta}$ $\forall\beta\in R^{1}(T)$ which gives:

$$q'=q,\qquad p'_i=p_i,\qquad s'_i=s_i,\qquad 1\leqq i\leqq q,\qquad s'=s.$$

Since $\theta[\mathfrak{g},\mathfrak{g}]=[\mathfrak{g},\mathfrak{g}]$, one has $l'_{\check{\theta}\beta}=l_{\beta}$ $\forall\beta\in R^{1}(T)$. The map $\check{\theta}$ induces a bijection between: $R(T)$ and $R(T')$, $R^{1}(T)$ and $R^{1}(T')$, $R^{1}(T)_{p_i}$ and $R^{1}(T')_{p_i}$ $1\leqq i\leqq q$; thus there exists $\tau\in\mathfrak{S}_s^{s_1\ldots s_q}$ such that

$$\check{\theta}\beta_a=\beta'_{\tau a}\qquad 1\leqq a\leqq s.$$

In conclusion, we have the

Lemma. *The integers* q, $p_1\ldots p_q$, $s_1\ldots s_q$, s, $d_1\ldots d_s$, $l_1\ldots l_s$, l *defined in* 1.4. *are invariants of* $\mathfrak{g}$.

If T, $T'\in\mathscr{T}_{\mathfrak{g}}$ *there exists* $\theta\in\mathrm{Aut}\,\mathfrak{g}$ *and* $\tau\in\mathfrak{S}_s^{s_1\ldots s_q}$ *such that* $\theta\mathfrak{g}^{\beta_a}=\mathfrak{g}^{\beta'_{\tau a}}$ $1\leqq a\leqq s$.

1.6. Let $f:\{1,\ldots,l\}\to\{1,\ldots,s\}$ be defined by

$$\begin{aligned} f(i)&=1 \quad \text{if} \quad 1\leqq i\leqq l_1\\ &=2 \quad \text{if} \quad l_1<i\leqq l_1+l_2\\ &\cdots\cdots\cdots\cdots\cdots\\ &=s \quad \text{if} \quad l_1+\ldots+l_{s-1}<i\leqq l. \end{aligned}$$

(I am thankful to the referee for his suggestion to introduce f instead of a double numbering in the original version.)

Since f is onto, there exists $g:\{1,\ldots,s\}\to\{1,\ldots,l\}$ such that $f\circ g=\mathrm{Id}$. Obviously

$$\mathfrak{S}_s^{s_1\cdots s_q}\to\mathfrak{S}_l,$$
$$\sigma\to\hat{\sigma}=g\circ\sigma\circ f,$$

is an one-to-one group homomorphism. Define an action of $\mathfrak{S}_s^{s_1\cdots s_q}$ on the set of $l\times l$ matrices by setting:

$$\sigma(A_{ij})_{1\leqq i,j\leqq l}=(A_{\hat{\sigma}i\hat{\sigma}j})_{1\leqq i,j\leqq l}$$

1.7. **Theorem.** *For $i,j\in\{1,\ldots,l\}$ $i\neq j$ let*

$$-A_{ij}(T)=\mathrm{Min}\{-n\in\mathbb{N};\ (\mathrm{adv})^{-n+1}w=0\ \forall v\in\mathfrak{g}^{\beta_{f(i)}}\ \forall w\in\mathfrak{g}^{\beta_{f(j)}}\}$$

with $(\mathrm{ad}0)^0=0$ and let $A_{ii}(T)=2$ for $i=1,\ldots,l$. Then

(1) *$A(T)=(A_{ij}(T))_{1\leqq i,j\leqq l}$ is a Cartan matrix* (2.1).

(2) *The $\mathfrak{S}_s^{s_1\cdots s_q}$-orbit of $A(T)$ is an invariant of $\mathfrak{g}$.*

Proof. (1) Since adv is nilpotent, $A_{ij}(T)$ is a well-defined non-positive integer for $i\neq j$; if $[v,w]=0$ then $[w,v]=0$, therefore $A_{ij}(T)=0$ implies $A_{ji}(T)=0$. Since $A_{ii}(T)=2$ by definition, $A(T)$ is a Cartan matrix.

(2) Let $T'\in\mathscr{T}_{\mathfrak{g}}$, by 1.5 there exist $\theta\in\mathrm{Aut}\mathfrak{g}$ and $\tau\in\mathfrak{S}_s^{s_1\cdots s_q}$ such that $\theta\mathfrak{g}^{\beta_a}=\mathfrak{g}^{\beta'_{\tau a}}$ $1\leqq a\leqq s$; if $v\in\mathfrak{g}^{\beta_a}$ and $w\in\mathfrak{g}^{\beta_b}$ and if $i,j\in\{1,\ldots,l\}$ are such that $f(i)=a$, $f(i)=b$, then $(\mathrm{adv})^{-A_{ij}(T)+1}w=0$; thus $(\mathrm{ad}\theta v)^{-A_{ij}(T)+1}\theta w=0$ with $\theta v\in\mathfrak{g}^{\beta'_{\tau a}}=\mathfrak{g}^{\beta'_{f(\hat{\tau}i)}}$ and $\theta w\in\mathfrak{g}^{\beta'_{\tau b}}=\mathfrak{g}^{\beta'_{f(\hat{\tau}j)}}$; therefore $-A_{\hat{\tau}i\hat{\tau}j}(T')\leqq -A_{ij}(T)$, and by symmetry $A_{\hat{\tau}i\hat{\tau}j}(T')=A_{ij}(T)$ which proves that $\tau A(T')=A(T)$.

1.8. *Definition.* We choose arbitrarily A in the $\mathfrak{S}_s^{s_1\cdots s_q}$-orbit of $A(T)$ $\left(\text{which has at most } \dfrac{s!}{s_1!\ldots s_q!} \text{ elements}\right)$ and we say by an abuse of language: "$\mathfrak{g}$ is of type A" or "A is the Cartan matrix of $\mathfrak{g}$". We denote:

$$\mathscr{T}_{\mathfrak{g}}(A)=\{T\in\mathscr{T}_{\mathfrak{g}};\ A(T)=A\},$$
$$\mathfrak{S}_s^{s_1\cdots s_q}(A)=\{\sigma\in\mathfrak{S}_s^{s_1\cdots s_q};\ \sigma A=A\}.$$

1.9. **Lemma.** *If T, $T'\in\mathscr{T}_{\mathfrak{g}}(A)$ there exist $\theta\in\mathrm{Aut}\mathfrak{g}$ and $\tau\in\mathfrak{S}_s^{s_1\cdots s_q}(A)$ such that:*

$$\theta\mathfrak{g}^{\beta_a}=\mathfrak{g}^{\beta'_{\tau a}}\quad \forall a=1\ldots s.$$

Proof. By 1.5 there exists $\theta\in\mathrm{Aut}\mathfrak{g}$ and $\tau\in\mathfrak{S}_s^{s_1\cdots s_q}$ such that $\theta\mathfrak{g}^{\beta_a}=\mathfrak{g}^{\beta'_{\tau a}}$; by the proof of 1.7 (2), $\tau A(T')=A(T)$; therefore $\tau\mathrm{A}=\tau\mathrm{A}(\mathrm{T}')=\mathrm{A}(\mathrm{T})=\mathrm{A}$.

1.10. We denote by $msg(\mathfrak{g})$ the set of minimal systems of generators of $\mathfrak{g}$; by [1, Sect. 4, ex. 4, p. 119]: $(x_1 \ldots) \in msg(\mathfrak{g})$ if and only if $(x_1 + [\mathfrak{g}, \mathfrak{g}], \ldots)$ is a basis of $\mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$ therefore each element of $msg(\mathfrak{g})$ is an l-tuple $(x_1 \ldots x_l)$ where $l = \dim \mathfrak{g}/[\mathfrak{g}, \mathfrak{g}]$.

Let $T \in \mathscr{T}_{\mathfrak{g}}(A)$ and denote:

$$msg(T) = msg(\mathfrak{g}) \cap ((\mathfrak{g}^{\beta_1})^{l_1} \times \ldots \times (\mathfrak{g}^{\beta_s})^{l_s}).$$

For all $(x_1 \ldots x_l) \in msg(T)$ one has:

$$(\mathrm{ad}\, x_i)^{-A_{ij}+1} x_j = 0, \quad 1 \leqq i \neq j \leqq l.$$

2. Kac-Moody Lie Algebras and the Universal Element for the Category of Nilpotent Lie Algebras

2.1. An $l \times l$ matrix with entries in $\mathbb{Z}$, $A = (A_{ij})$ is a Cartan matrix if:

a) $A_{ii} = 2 \quad \forall i = 1, \ldots, l$,

b) $A_{ij} \leqq 0 \quad \forall i, j = 1, \ldots, l, \quad i \neq j$,

c) $(A_{ij} = 0$ if $A_{ij} = 0) \quad \forall i, j = 1, \ldots, l, \quad i \neq j$.

2.2. Let $\|A\| = \mathrm{Sup}\{|A_{ij}|;\ i \neq j\}$; denote h the Coxeter number of A if A is of finite type and $+\infty$ if not.

For $n \in \mathbb{N}$ such that $\|A\| < n \leqq h+1$ let $\mathrm{Nil}p_n(A)$ be the category of nilpotent Lie algebras of nilpotency index n and of type A.

2.3. To A one can associate a Lie algebra (called a Kac-Moody Lie algebra) defined by generators:

$$\{f_1 \ldots f_l\ h_1 \ldots h_l\ e_1 \ldots e_l\},$$

and relations:

$$[h_i e_j] = A_{ij} e_j, \quad [h_i f_j] = -A_{ij} f_j, \quad (\forall i, j = 1 \ldots l),$$

$$[h_i h_j] = 0, \quad [e_i f_i] = h_i, \quad (\forall i, j = 1 \ldots l),$$

$$[e_i f_j] = 0, \quad (\mathrm{ad}\, e_i)^{-A_{ij}+1} e_j = 0 = (\mathrm{ad}\, f_i)^{-A_{ij}+1} f_j, \quad (\forall i \neq j).$$

[8, 9]. Actually, this definition is not the usual one (ordinarily one has to take a quotient of the above algebra).

2.4. Let $H = Kh_1 + \ldots + Kh_l$; denote $\mathscr{L}_+(A)$ (resp. $\mathscr{L}_-(A)$) the subalgebra of $\mathscr{L}(A)$ generated by $\{e_1 \ldots e_l\}$ (resp. $\{f_1 \ldots f_l\}$). One shows that:

$$\mathscr{L}(A) = \mathscr{L}_-(A) \oplus H \oplus \mathscr{L}_+(A).$$

One can define $\mathscr{L}_+(A)$ by generators: $\{e_1 \ldots e_l\}$ and relations:

$$(\mathrm{ad}\, e_i)^{-A_{ij}+1} e_j = 0 \quad \forall i, j = 1 \ldots l, \quad i \neq j.$$

2.5. Let $L \equiv L_n(A) \equiv \mathscr{L}_+(A)/C^{n+1}\mathscr{L}_+(A)$ where $(C^p \mathscr{L}_+(A))_{p \geq 1}$ is the descending central series of $\mathscr{L}_+(A)$ with $C^1 \mathscr{L}_+(A) = \mathscr{L}_+(A)$.

By abuse of notation, write e_i instead of $e_i + C^{n+1}\mathscr{L}_+(A)$. One can define $L_n(A)$ by generators: $\{e_1 \ldots e_l\}$ and relations:

$$(\mathrm{ad}\, e_i)^{-A_{ij}+1} e_j = 0 \quad \forall i, j = 1 \ldots l, \quad i \neq j;$$

$$[e_{i_1} \ldots e_{i_{n+1}}] = 0 \quad \forall i_k = 1 \ldots l \quad \forall k = 1 \ldots n+1.$$

2.6. If A is of finite type $\mathscr{L}_+(A)$ is a nilpotent Lie algebra of nilpotency index $h+1$; therefore $C^n\mathscr{L}_+(A) \supsetneq C^{n+1}\mathscr{L}_+(A)$ (since $n \leqq h+1$) thus $C^nL \neq (0)$. If A is not of finite type then $C^p\mathscr{L}_+(A) \supsetneq C^{p+1}\mathscr{L}_+(A)$ $\forall p \geqq 1$ (one says that $\mathscr{L}_+(A)$ is pro-nilpotent) thus $C^nL \neq (0)$.

In both cases, one has $C^{n+1}L=(0)$; therefore $L_n(A)$ is a nilpotent Lie algebra of nilpotency index n.

2.7. One defines derivations $\partial_1 \ldots \partial_l$ of L by setting $\forall i,j=1\ldots l$

$$\partial_i e_j = \begin{cases} e_i & i=j \\ 0 & i \neq j \end{cases}.$$

Let $(\alpha_1 \ldots \alpha_l)$ be the dual basis of $(\partial_1 \ldots \partial_l)$ and denote $D = K\partial_1 + \ldots + K\partial_l$. It is easy to show that:

$$D \in \mathscr{T}_L, \quad R^1(D) = \{\alpha_1 \ldots \alpha_l\}, \quad L^{\alpha_i} = Ke_i \quad \forall i = 1\ldots l,$$
$$d_i = l_i = 1 \quad \forall i = 1\ldots l, \quad q=1, \quad p_1=1, \quad s_1=l, \quad \mathfrak{S}_s^{s_1\ldots s_q} = \mathfrak{S}_l$$

(permutation group of $\{1\ldots l\}$). We are therefore in the case 2.11(b).

2.8. Let $i,j \in \{1\ldots l\}$ $i \neq j$ and assume that $(\mathrm{ad}\, e_i)^{-A_{ij}} e_j = 0$ in L, then $(\mathrm{ad}\, e_i)^{-A_{ij}} e_j \in C^{n+1}\mathscr{L}_+(A)$ in $\mathscr{L}_+(A)$; but $C^{n+1}\mathscr{L}_+(A)$ is spanned as a vector space by the $[e_{i_1}\ldots e_{i_r}]$'s such that $r \geqq n+1$, thus $-A_{ij}+1 \geqq n+1$ (since $(\mathrm{ad}\, e_i)^{-A_{ij}} e_j = [e_i \ldots e_i e_j]$ with $-A_{ij}+1$ terms); this contradicts $\|A\| < n$; therefore $(\mathrm{ad}\, e_i)^{-A_{ij}} e_j \neq 0$, it follows by 2.11(b) that $A(D)=A$; so L is of type A and $D \in \mathscr{T}_L(A)$.

2.9. **Lemma.** *The Lie algebra $L_n(A)$ is an object of* $\mathrm{Nil}p_n(A)$.

Proof. It follows from 2.6 and 2.8.

2.10. **Theorem.** *The Lie algebra $L_n(A)$ is an universal element for the category* $\mathrm{Nil}p_n(A)$ *i.e. any object of* $\mathrm{Nil}p_n(A)$ *is a quotient of $L_n(A)$.*

More precisely: for any object $\mathfrak{g}$ *of* $\mathrm{Nil}p_n(A)$, for any $T \in \mathscr{T}_{\mathfrak{g}}(A)$ *and any* $x=(x_1\ldots x_l) \in msg(T)$, *the map* $\{e_1\ldots e_l\} \to \mathfrak{g}$ $e_i \mapsto x_i$ *extends uniquely to an homomorphism* $\pi \equiv \pi(T,x) : L_n(A) \to \mathfrak{g}$ *which is onto; and one has* $\mathfrak{g} \cong L_n(A)/\mathrm{Ker}\,\pi$.

Proof. In $\mathfrak{g}$ one has at least the relations: $(\mathrm{ad}\, x_i)^{-A_{ij}+1} x_j = 0$ (1.10) and $[x_{i_1}\ldots x_{i_{n+1}}]=0$; whereas in L the defining relations are $(\mathrm{ad}\, e_i)^{-A_{ij}+1} e_j = 0$ and $[e_{i_1}\ldots e_{i_{n+1}}]=0$ (2.5); whence the existence of an homomorphism $\pi : L \to \mathfrak{g}$ such that $\pi e_i = x_i$; π is onto because $(x_1\ldots x_l) \in msg(\mathfrak{g})$; π is unique because $(e_1\ldots e_l) \in msg(L)$.

2.11. *Particular cases.* (a) Assume

$$l_a = 1 \quad \forall a = 1, \ldots, s,$$

then $s=l$ (since $l_1 + \ldots + l_s = l$), $f = \mathrm{Id}$, $\hat{\sigma} = \sigma$ $\forall \sigma \in \mathfrak{S}_l^{s_1\ldots s_q}$ and for $i \neq j$:

$$-A_{ij}(T) = \mathrm{Min}\{-n \in \mathbb{N}; (\mathrm{ad}\, v)^{-n+1} w = 0 \quad \forall v \in \mathfrak{g}^{\beta_i} \ \forall w \in \mathfrak{g}^{\beta_j}\}.$$

(b) Assume that

$$d_a = 1 \quad \forall a = 1, \ldots, s,$$

(since $l_a \leqq d_a$, it is a subcase of the preceeding one). Then, with the notations of 1.4, we have $q=1$, $p_1=1$, $s_1=l$, $\mathfrak{S}_s^{s_1 \ldots s_q}=\mathfrak{S}_l$. Furthermore for all $i=1 \ldots l$ there exists $x_i \in \mathfrak{g}^{\beta_i} \backslash (0)$ such that $\mathfrak{g}^{\beta_i}=Kx_i$, therefore

$$\forall i \neq j, \quad -A_{ij}(T)=\operatorname{Min}\{-n \in \mathbb{N}; (\operatorname{ad} x_i)^{-n+1} x_j=0\},$$

thus $-A_{ij}(T)$ is the unique integer such that $(\operatorname{ad} x_i)^{-Aij(T)} x_j \neq 0$ and $(\operatorname{ad} x_i)^{-A_{ij}(T)+1} x_j=0$ which is the definition given in [12–14]. This subcase is studied in detail in [14].

Acknowledgements. All my thanks to Professors M. Vergne and V. G. Kac for their kind invitation to the Massachusetts Institute of Technology (Cambridge, USA). It is during this stay that this work began with the help of Professor V. G. Kac. I also thank D. H. Peterson (MIT, Dept. of Math.) for the interest he took in my work.

References

1. Bourbaki, N.: Groupes et algèbres de Lie. Chap. 1. Paris: Hermann 1968
2. Bratzlavsky, F.: Sur les algèbres admettant un tore d'automorphismes donné. J. Algebra **30**, 305–316 (1974)
3. Chevalley, C.: Unpublished work [quoted by J. Dixmier in (4)]
4. Dixmier, J.: Sur les représentations unitaires des groupes de Lie nilpotentes III. Canad. J. Math. **10**, 321–348 (1958)
5. Favre, G.: Système de poids sur une algèbre de Lie nilpotente. Thesis 1972, Ecole Polytechnique Fédérale de Lausanne
6. Favre, G.: Système de poids sur une algèbre de Lie nilpotente. Manus. Math. **9**, 53–90 (1973)
7. Gurevic, G.R.: On metabelian Lie algebras. Trudy Sem. Vektor. Tenzor. Anal. **12**, 9–61 (1963)
8. Kac, V.G.: Simple irreducible graded Lie algebras of finite growth. Math. USSR-Izv. **2**, 1271–1311 (1968)
9. Moody, R.V.: A new class of Lie algebras. J. Algebra **10**, 211–230 (1968)
10. Mostow, G.D.: Fully reducible subgroups of algebraic groups. Am. J. Math. **78**, 200–221 (1956)
11. Morozov, V.V.: Classification of nilpotent Lie algebras of sixth order (in Russian; MR 24A 190). Ivz. Ucebn. Zave denii Mat. n° 4 (5), 161 (1958), USSR
12. Santharoubane, L.J.: Structure et cohomologie des algèbres de Lie nilpotentes. Thesis (1979), University of Paris 6
13. Santharoubane, L.J.: Kac-Moody Lie algebras and the classification of Nilpotent Lie algebras of maximal rank. Canadian J. of Math. (to appear)
14. Santharoubane, L.J.: Kac-Moody Lie algebras and the classification of certain nilpotent Lie algebras. Manuscript
15. Umlauf, K.A.: Über die Zusammensetzung der endlichen kontinuierlichen Transformationsgruppen, insbesondere der Gruppen vom Range Null. Leipzig SS 80, 1891
16. Vergne, M.: Variétés des algèbres de Lie nilpotentes. Thesis, University of Paris (1966)

Received November 30, 1982; in revised form January 11, 1983

Math. Ann. 263, 371–375 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Remarques sur les ouverts localement de Stein

B. Jennane
Département de Mathématiques, Université M^ed V, Rabat, Maroc

Dans un espace analytique complexe X, un ouvert Ω est dit localement de Stein si tout point z de la frontière de Ω posséde un voisinage U dans X tel que $U \cap \Omega$ soit un ouvert de Stein.

Le problème de savoir si lorsque X est de Stein, Ω est-il de Stein demeure ouvert. On sait toutefois que lorsque X est une variété ou que les singularités de X sont isolées alors la réponse est positive [1]. Il a été aussi démontré [2] que si $H^q(\Omega, O)=0$ pour $q>0$, alors Ω est de Stein. Nous allons voir qu'en fait $H^q(\Omega, F)=0$ pour $q \geqq 2$ et pour tout faisceau cohérent F. Nous obtenons en conséquence un résultat similaire à celui de [6] pour les réunions croissantes d'ouverts de Stein ou de [4] pour les fibrés holomorphes, à savoir qu'il y a équivalence entre les assertions suivantes

$$H^1(\Omega, F)=0,$$

$$H^1(\Omega, F) \text{ est séparé},$$

$$H^1(\Omega, F) \text{ est de dimension finie},$$

$$\Omega \text{ est de Stein}.$$

Une simple lecture des démonstrations des théorèmes 1 et 2 de [3] permet de voir qu'en fait nous avons

Théorème 1. *Soit Y un espace analytique complexe. Si Y posséde un recouvrement localement fini par des ouverts de Stein U_i tels que $\left(\bigcup_{i \leqq j} U_i\right) \cap U_{j+1}$ est de Stein pour tout j, alors $H^p(Y, F)=0$ pour tout $p \geqq 2$ et tout faisceau cohérent F sur Y.*

Si de plus $\left(\bigcup_{i \leqq j} U_i\right) \cap U_{j+1}$ est de Runge dans U_{j+1}, alors il y a équivalence entre

$$H^1(Y, F)=0,$$

$$H^1(Y, F) \textit{ est de dimension finie},$$

$$H^1(Y, F) \textit{ est séparé}$$

et ce pour tout faisceau cohérent F.

Or un ouvert localement de Stein dans espace de Stein de dimension bornée posséde bien un tel recouvrement. Plus exactement, nous avons

Lemme. *Soit Ω un ouvert localement de Stein dans un espace de Stein X de dimension bornée. Alors $\bar{\Omega}$ admet un recouvrement localement fini par des ouverts U_i de X tels que $U_i \cap \Omega$ est de Stein et*

$$\Omega_j = \left(\bigcup_{i \leqq j} U_i\right) \cap \Omega \quad \textit{est localement de Stein pour tout } j,$$

$$\Omega_j \cap U_{j+1} \quad \textit{est de Runge dans} \quad U_{j+1} \cap \Omega.$$

Preuve. Considerons un recouvrement (V_i) de $\bar{\Omega}$ par des ouverts V_i de X telsque $V_i \cap \Omega$ soit de Stein pour tout i. On lui adjoind l'ouvert $\complement\bar{\Omega}$. Soit V le recouvrement de X ainsi obtenu. D'après le lemme de recouvrement d'un espace de Stein de Stehlé [8], il existe un recouvrement plus fin, localement fini par des ouverts de Stein $(U_i)_{i \in N}$ tels que $\bigcup_{i \leqq j} U_i$ soit de Stein pour tout j. De plus il existe dans $\bigcup_{i \leqq j} U_i$ une fonction φ_{j+1} strictement *psh* continue vérifiant

$$\left(\bigcup_{i \leqq j} U_i\right) \cap U_{j+1} = \{x \in U_{j+1} \mid \varphi_{j+1}(x) < 0\}.$$

Si des ouverts U_i nous ne considérons que ceux qui rencontrent $\bar{\Omega}$, soit (U'_i), nous obtenons un recouvrement de $\bar{\Omega}$ plus fin que le recouvrement (V_i), donc

- $U'_i \cap \Omega$ est de Stein
- $\left(\bigcup_{i \leqq j} U'_i\right) \cap \Omega$ est localement de Stein comme intersection d'un ouvert de Stein

et d'un ouvert localement de Stein

- $\left(\bigcup_{i \leqq j} U'_i\right) \cap \Omega \cap U'_{j+1} = \{x \in U'_{j+1} \cap \Omega \mid \varphi_{j+1}(x) < 0\}$ est de Runge dans $U_{j+1} \cap \Omega$

d'après le théorème de Narasimhan [7]. C.Q.F.D.

Pour se passer de l'hypothèse «X de dimension bornée», nous démontrons la proposition suivante

Proposition. *Soit p un entier $\geqq 1$, et soit X un espace analytique complexe dont les composantes irréductibles X_k vérifient la propriété (P) suivante : «Pour tout Faisceau cohérent F et tout entier $p \geqq q$, $H^p(X_k, F) = 0$». Alors X vérifie lui aussi la propriété (P).*

Démonstration. a) On commence par montrer que si X est l'union de deux ensembles analytiques fermés vérifiant chacun la propriété (P), alors X vérifie lui aussi (P).

Soit Y_1 and Y_2 ces ensembles analytiques et soit $\mathscr{I}_i$ le faisceau des germes de fonctions holomorphes s'annulant sur Y_i. On pose $O_i = O/\mathscr{I}_i$ et si F est un faisceau cohérent, on définit $F_i = F \otimes O_i$. Ainsi, si e_i est l'image de $1 \in O$ dans O_i, tout $\xi \in (F_i)_x$ se met sous la forme $\xi = \alpha \otimes (e_i)_x$ où $\alpha \in F_x$.

Considérons l'homomorphisme

$$\eta : F \to F_1 \oplus F_2$$

défini par $\eta(\alpha) = \alpha \otimes e_1 + \alpha \otimes e_2$.

η est injectif et si l'on note $\mathscr{H}$ son coker, nous avons la suite exacte

$$0 \to F \xrightarrow{\eta} F_1 \oplus F_2 \xrightarrow{\varrho} \mathscr{H} \to 0. \qquad (*)$$

$\mathscr{H}$ est un faisceau cohérent puisque F et F_i le sont. $\mathscr{H}$ est à support dans $Y_1 \cap Y_2$ car si $x \notin Y_i$, alors $(O_i)_x = 0$ et η est surjectif en x et par conséquent $\mathscr{H}_x = 0$.

ϱ induit l'application

$$\varrho^* : H^p(X, F_1) \oplus H^p(X, F_2) \to H^p(X, \mathscr{H}).$$

Nous allons voir que ϱ^* est surjectif. La composée ϱ_i de l'injection $F_i \to F_1 \oplus F_2$ et de ϱ est une surjéction. De la suite exacte

$$0 \to \ker \varrho_i \to F_i \xrightarrow{\varrho_i} \mathscr{H} \to 0$$

nous déduisons les suites longues de cohomologie

$$H^p(X, F_i) \to H^p(X, \mathscr{H}) \to H^{p+1}(X, \ker \varrho_i)$$

et

$$H^p(Y_i, F_i) \to H^p(Y_i, \mathscr{H}) \to H^{p+1}(Y_i, \ker \varrho_i).$$

Or d'une part $H^{p+1}(Y_i, \ker \varrho_i) = 0$ pour $p \geqq q$ et d'autre part $H^p(X, F_i) = H^p(Y_i, F_i)$. Donc

$$H^p(X, F_i) \to H^p(X, \mathscr{H}) \to 0$$

et il s'ensuit que ϱ^* est surjectif.

Considérons maintenant la suite longue de cohomologie associée à la suite (*):

$$\to H^p(X, F_1) \oplus H^p(X, F_2) \xrightarrow{\varrho^*} H^p(X, \mathscr{H}) \to H^{p+1}(X, F) \to 0$$

où il a été tenu compte du fait que $H^{p+1}(X, F_i) = H^{p+1}(Y_i, F_i) = 0$. ϱ^* étant surjectif, il en ressort que $H^{p+1}(X, F) = 0$ pour $p \geqq q$.

b) Soit $V_k = \bigcup_{i \leqq k} X_i$. D'après a), $H^p(V_k, F) = 0$ pour $p \geqq q$ et tout faisceau cohérent F. Soit $\mathscr{U} = (U_i)$ un recouvrement de Stein de X et soit $\mathscr{U}_k = (U_i \cap V_k)$. Pour tout faisceau cohérent F sur X l'application

$$Z^p(\mathscr{U}_{k+1}, F) \to Z^p(\mathscr{U}_k, F) \qquad (**)$$

est surjective pour $p \geqq q-1$ comme on va le voir.

V_k est un sous ensemble analytique de V_{k+1}. Notons $\mathscr{J}$ le faisceau idéal associé, O' le faisceau des germes de fonctions holomorphes dans V_{k+1} O'' le faisceau quotient $O'/\mathscr{J}$ et $F' = F \otimes O''$. Nous avons la suite exacte

$$0 \to \ker \eta' \to F \xrightarrow{\eta'} F' \to 0$$

où est définie par $\eta'(\alpha) = \alpha \otimes e$, e étant l'image de $1 \in O'$ dans O''. On en déduit que pour $p \geqq q-1$ on a

$$\to H^p(V_{k+1}, F) \to H^p(V_{k+1}, F') \to H^{p+1}(V_{k+1}, \ker \eta').$$

Or $H^{p+1}(V_{k+1}, \ker \eta')$ est nul et $H^p(V_{k+1}, F') = H^p(V_k, F)$. On a donc une surjection

$$H^p(\mathscr{U}_{k+1}, F) \to H^p(\mathscr{U}_k, F) \to 0$$

qui est induite par l'application restriction à V_k.

Soit $c \in Z^p(\mathcal{U}_k, F)$. Il existe alors $\tilde{c} \in Z^p(\mathcal{U}_{k+1}, F)$ et $\alpha \in C^{p-1}(\mathcal{U}_k, F)$ tels que la restriction à $\mathcal{U}_k$ de $\tilde{c}$ soit $c + \delta\alpha$. V_k est un sous ensemble analytique de V_{k+1} et le recouvrement $\mathcal{U}_{k+1}$ est de Stein; on peut étendre α à $\mathcal{U}_{k+1}$ en une cochaine $\tilde{\alpha} \in C^{p-1}(\mathcal{U}_{k+1}, F)$, c apparait alors comme la restriction du p-cycle $\tilde{c} + \delta\tilde{\alpha}$, d'où la surjection de l'application (**).

c) Démontrons maintenant la proposition. Soit $c \in Z^p(\mathcal{U}, F)$, et soit $c^k \in Z^p(\mathcal{U}_k, F)$ sa restriction à $\mathcal{U}_k$. $H^p(\mathcal{U}_k, F)$ etant nul c^k est un cobord: $c^k = \delta\alpha^k$. $\alpha^{k+1} - \alpha^k$ est un élément de $Z^{p-1}(\mathcal{U}_k, F)$ où $p \geqq q$; d'après b) il existe β^{k+1} dans $Z^{p-1}(\mathcal{U}_{k+1}, F)$ de restriction $\alpha^{k+1} - \alpha^k$. En remplaçant α^{k+1} par $\alpha^{k+1} - \beta^{k+1}$ on construit ainsi de proche en proche une suite convergente α^k telle que $\delta\alpha^k = c^k$.

Remarque. Cette proposition généralise le théorème selon lequel un espace analytique dont les composantes irréductibles sont de Stein est lui même de Stein [1].

Nous sommes maintenant en mesure de démontrer le théorème principal de cet article.

Théorème 2. *Soit Ω un ouvert localement de Stein dans un espace de Stein. Alors pour tout faisceau cohérent F et tout entier $p \geqq 2$ nous avons $H^p(\Omega, F) = 0$. D'autre part il y a équivalence entre Ω est de Stein*

$$H^1(\Omega, O_\Omega) = 0,$$

$$H^1(\Omega, O_\Omega) \text{ est séparé},$$

$$H^1(\Omega, O_\Omega) \text{ est de dimension finie}.$$

Démonstration. La première assertion est une conséquence immédiate du Théorème 1 et de la proposition précédente.

Pour la seconde notons d'abord que lorsque Ω est de dimension bornée, le résultat est une conséquence immédiate du Théorème 1, et du 2.6.2 de [2]. Pour le cas géneral on utilise les faits suivants: i) Ω est de Stein si et seulement si ses composantes irréductibles Ω_ν sont de Stein. ii) Si $H^1(\Omega, O_\Omega) = 0$ alors pour tout ν, $H^1(\Omega_\nu, O_{\Omega_\nu}) = 0$ d'après l'exactitude de la suite

$$0 \to J_\nu \to O_\Omega \to O_{\Omega_\nu} \to 0$$

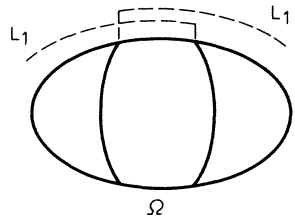
et la nullité de $H^2(\Omega, J_\nu)$. Inversement si $H^1(\Omega_\nu, O_{\Omega_\nu}) = 0$ pour tout ν, Ω_ν est alors de Stein et il est de même alors pour Ω; donc $H^1(\Omega, O_\Omega) = 0$. iii) Par le même argument que précédemment on montre l'équivalence entre $H^1(\Omega, O_\Omega)$ est de dimension finie et $\dim H^1(\Omega_\nu, O_{\Omega_\nu}) < \infty$ pour tout ν. iv) Si tous les $H^1(\Omega_\nu, O_{\Omega_\nu})$ sont séparés, les Ω_ν sont alors de Stein et il en est de même pour Ω lequel vérifie alors $H^1(\Omega, O_\Omega) = 0$. Inversement, si $H^1(\Omega, O_\Omega)$ est séparé, $\delta C^0(\mathcal{U}, O_\Omega)$ est fermé dans $Z^1(\mathcal{U}, O_\Omega)$ pour tout recouvrement $\mathcal{U}$ de Ω par des ouverts de Stein (U_i). Posons $\Omega'_\nu = \bigcup_{\mu \leqq \nu} \Omega_\nu$ et $\mathcal{U}_\nu = (U_i \cap \Omega'_\nu)$. Par une démonstration analogue à celle du théorème 2 de [3], on montre que $\delta C^0(\mathcal{U}_\nu, O_\Omega)$ est dense dans $Z^1(\mathcal{U}_\nu, O_\Omega)$. Soit $c \in Z^1(\mathcal{U}, O_\Omega)$. Pour tout ν, sa restriction c_ν à $\mathcal{U}_\nu$ peut être approchée d'aussi près que l'on veut par une suite $\delta\alpha_k$, $\alpha_k \in C^0(\mathcal{U}_\nu, O)$. Si l'on note $\tilde{\alpha}_k$ une extension de α_k à $\mathcal{U}$, $\delta\tilde{\alpha}_k$ approche c. Donc $\delta C^0(\mathcal{U}, O)$ est dense dans $Z^1(\mathcal{U}, O)$ et puisqu'il est fermé, $H^1(\mathcal{U}, O) = 0$.

Remarques. 1) Pour la démonstration de la seconde assertion dans le théorème précédent, du fait que Ω est holomorphiquement séparable et que $H^p(\Omega, O_\Omega)=0$ pour $p\leqq 2$ on peut obtenir le resultat comme conséquence immédiate des théorèmes 2.6.2 de [2] et de [5]. Mais la démonstration précédente a l'avantage de s'appliquer à d'autres cas où l'espace n'est pas holomorphiquement séparable: le cas d'un fibré holomorphe à base et à fibre de Stein par exemple (voir [3]).

2) Le lemma de recouverment ci-dessus permet de ramener le problème de savoir si un ouvert localement de Stein dans un espace de Stein est de Stein au cas où cet ouvert est l'union de deux ouverts de Stein L_1 et L_2 tels que

$$L_1 \cap L_2 \neq \emptyset$$
$$(\Omega - L_1) \cap (\Omega - L_2) = \emptyset$$

et ce grace à une démonstration par récurrence sur l'indice des ouverts du recouvrement.



Bibliographie

1. Andreotti, A., Narasimhan, R.: Oka's Heftungslemma and the Levi problem for complex spaces. Trans. Am. Math. Soc. **111**, 345–366 (1964)
2. Fornaess, J.E., Narasimhan, R.: The Levi problem on complex spaces with singularities. Math. Ann. **248**, 47–72 (1980)
3. Jennane, B.: Groupes de cohomologie d'un fibré à base de Stein. Séminaire Lelong-Skoda, 1978–1979. Lecture Notes in Mathematics, No. 822. Berlin, Heidelberg, New York: Springer 1980
4. Jennane, B.: Problème de Levi etmorphisme localement de Stein. Math. Ann. **256**, 37–42 (1981)
5. Alessandrini, L.: On the cohomology of a holomorphically separable complex anallytic space. Boll. U.M.I. **6**, 261–268 (1982)
6. Markoe, A.: Runge families on inductive limits of Stein spaces. Ann. Inst. Fourier, Grenoble **27**, 117–127 (1977)
7. Narasimhan, R.: The Levi problem for complex spaces. II. Math. Ann. **136**, 245–318 (1958)
8. Stéhle, J.-L.: Fonction plurisousharmoniques et convexité holomorphe de certains fibrés analytiques. Séminaire Lelong, 1973/74, pp. 155–179. Lecture Notes in Mathematics, No. 474. Berlin, Heidelberg, New York: Springer 1975

Received January 18, 1983

Math. Ann. 263, 377–383 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Kodaira Dimension is not Necessarily Lower Semi-Continuous Under Degenerations of Surfaces

Kenji Nishiguchi

Department of Mathematics, Faculty of Science, Kyoto University, Kyoto, Japan

Introduction

Let $\pi: X \to \Delta$ be a proper surjective holomorphic map of a three dimensional complex manifold X to a disc $\Delta = \{z \in \mathbf{C} \mid |z| < \varepsilon\}$ with connected fibers. Assume that π is smooth at each point on $\pi^{-1}(\Delta^*)$, $\Delta^* = \Delta - \{0\}$. We call such a holomorphic map $\pi: X \to \Delta$ a *degeneration*. Then we have the following

Conjecture. *Let $\kappa(X_t)$ (resp. $\kappa(V)$) be the Kodaira dimension of X_t (resp. V). Then we have $\kappa(X_t) \geqq \kappa(V)$ for every irreducible component V of X_0, where X_t is a general fiber $\pi^{-1}(t)$ $(t \neq 0)$, and X_0 is a special fiber $\pi^{-1}(0)$.*

Under the assumption that π is projective, this conjecture was partly proved by Persson [9], and completely proved by Ueno [11] using the result of Kulikov [6] and Persson-Pinkham [10]. It is easy to generalize their arguments to obtain the affirmative answer to the conjecture when X is Kähler. Moreover, if $\kappa(X_t) \geqq 1$, their arguments show that the kählerian assumption is not needed to prove the conjecture.

The main purpose of the present paper is to show that the above conjecture does not necessarily hold if $\kappa(X_t) \leqq 0$ and X is non-Kähler. In Sect. 3, we shall give such examples.

Friedman's paper [1] is a general reference throughout this paper, and we freely use some results of it.

I would like to express my thanks to Professor K. Ueno, whose insight and encouragement are invaluable.

1.

In this section, we shall briefly review the deformation theory of varieties with normal crossings, due to Friedman [1]. Let $X_0 = V_1 \cup_E V_2$ be a two dimensional compact analytic space with normal crossings, where V_1 and V_2 are smooth surfaces (not necessarily Kähler) crossing transversely along a smooth curve E. (Most of the following arguments hold for more general varieties with normal crossings. But for our present purpose, it is enough to consider only the variety

X_0.) Assume that $N_{E/V_1} \otimes N_{E/V_2} \cong \mathcal{O}_E$, namely X_0 is D semi-stable in the sense of Friedman [1].

We put

$$T^i_{X_0} = \mathcal{E}xt^i(\Omega^1_{X_0}, \mathcal{O}_{X_0}),\ \mathbb{T}^i_{X_0} = Ext^i(\Omega^1_{X_0}, \mathcal{O}_{X_0}).$$

Since X_0 is a variety with normal crossings, we have

$$T^1_{X_0} = N_{E/V_1} \otimes N_{E/V_2} (= \mathcal{O}_E),$$

and

$$T^i_{X_0} = 0 \quad \text{for} \quad i \geqq 2.$$

From the spectral sequence

$$E_2^{pq} = H^p(X_0, \mathcal{E}xt^q(\Omega^1_{X_0}, \mathcal{O}_{X_0})) \Rightarrow Ext^{p+q}(\Omega^1_{X_0}, \mathcal{O}_{X_0}),$$

we obtain the exact sequence

$$0 \to H^1(T^0_{X_0}) \to \mathbb{T}^1_{X_0} \to H^0(T^1_{X_0}) \to H^2(T^0_{X_0}) \to \mathbb{T}^2_{X_0} \to H^1(T^1_{X_0}) \to 0.$$

Theorem (Friedman [1]). *Assume that the D semi-stable variety X_0 satisfies the following two conditions*

(i) $H^2(T^0_{X_0}) = 0$,

(ii) $H^1(T^0_{X_0}) \otimes H^0(T^1_{X_0}) \to H^1(T^1_{X_0})$ *is surjective, where this "natural" pairing can be obtained, under assumption* (i), *from the Lie bracket* $[\,,\,]: \mathbb{T}^1_{X_0} \otimes \mathbb{T}^1_{X_0} \to \mathbb{T}^2_{X_0}$ *which is always induced from that on* $\bigoplus_{i=0}^{1} T^i_{X_0}$ *(see Friedman* [1, *Sects.* 4 *and* 5] *or Palamodov* [8] *for more detailed definitions of these products).*

Then X_0 has a smoothing, more precisely, there exists a degeneration $X \to \Delta$ whose singular fiber is isomorphic to X_0.

The proof is exactly same as that of the smoothability of "D semi-stable $K3$ surfaces" in Friedman [1, Theorem (5.10)]. (See also Friedman-Miranda [2], Lemma (2.9).) So we don't repeat it, but give an informal explanation in the following.

The conditions (i) and (ii), combining the above exact sequence, imply that there exists a first order deformation $e \in T^1_{X_0}$ such that e comes, in fact, from a global deformation and the image of e in $H^0(T^1_{X_0})$ is not zero. This implication was proved in [1, (5.10)] where the theorem of Palamodov et al. [8] was used. On the other hand, since X_0 is D semi-stable, i.e., $T^1_{X_0} = \mathcal{O}_E$, this e generates $T^1_{X_0}$ at all points. Hence e corresponds to a smoothing. This was proved in [1, Proposition (2.5)].

Remark. The result due to Friedman [1] has a lot of applications other than those given in [1] and in this paper. For example, using it, Friedman-Miranda [2] have studied the smoothability of certain cusp singularities, and the author has constructed new types of non-Kähler degenerations of $K3$ surfaces (see [7]).

2.

Let V_1 be a relatively minimal elliptic surface, and E_1 be a multiple fiber of type ${}_mI_0$ on V_1 (see Kodaira [3, II]). By the canonical bundle formula (Kodaira [4, I]), K_{V_1}

is written in the form, $K_{V_1}=(m-1)E_1+F$, where F is composed of fibers and does not contain E_1. We assume the following

$$H^2(V_1, \Theta_{V_1})=0, \tag{0}$$

where Θ is a sheaf of germs of holomorphic vector fields.

Let V_2 be a Hopf surface defined as follows. $V_2=\mathbf{C}^2-\{0\}/\langle g\rangle$, where g is the automorphism of $\mathbf{C}^2-\{0\}$ in the form

$$g:(z_1, z_2)\to(\alpha_1 z_1+z_2^m, \alpha_2 z_2);\ \alpha_1, \alpha_2\in\mathbf{C};\ 0<|\alpha_1|, |\alpha_2|<1;\ \alpha_2^m=\alpha_1.$$

Let E_2 be an elliptic curve defined by $z_2=0$. E is isomorphic to $\mathbf{C}^*/\langle\alpha_1\rangle$, and $K_{V_2}=(-m-1)E_2$. (See Kodaira [4, II].) We assume that E_1 is isomorphic to E_2. (In the below, we identify E_1 and E_2, and denote them by E.) We also assume that $N_{E/V_1}\otimes N_{E/V_2}\cong\mathcal{O}_E$. (Note that both N_{E/V_1} and N_{E/V_2} are of order m.) We set $X_0=V_1\cup_E V_2$. X_0 is a variety with normal crossings.

Proposition. *The variety X_0 has a smoothing. That is, there exists a degeneration $\pi: X\to\Delta$ such that $X_0=\pi^{-1}(0)$.*

Proof. As seen in Sect. 1, it suffices to prove the following Lemmas 1 and 2. We use the same notation as in Sect. 1.

Lemma 1. $H^2(T^0_{X_0})=0$.

Proof. Let $i:\tilde{X}_0=V_1\cup V_2\to X_0$ be a normalization of X_0. By Friedman [1, Sect. 2, (2.10)], we have

$$H^2(T^0_{X_0})=H^0(\Omega^1_{X_0}/\tau_{X_0}\otimes\omega_{X_0})^*.$$

Here τ_{X_0} is the torsion differentials, i.e., $\tau_{X_0}=\mathrm{Ker}(\Omega^1_{X_0}\to i_*\Omega^1_{\tilde{X}_0})$, and ω_{X_0} is the dualizing sheaf of X_0. By the definition of τ_{X_0}, we get the following exact sequence,

$$0\to\Omega^1_{X_0}/\tau_{X_0}\to i_*\Omega^1_{\tilde{X}_0}\to\Omega^1_E\to 0$$

Hence

$$0\to H^0(\Omega^1_{X_0}/\tau_{X_0}\otimes\omega_{X_0})\to H^0(i_*\Omega^1_{\tilde{X}_0}\otimes\omega_{X_0})\to H^0(\Omega^1_E\otimes\omega_{X_0})\to$$

is exact. Let ϕ be an element of $H^0(\Omega^1_{X_0}/\tau_{X_0}\otimes\omega_{X_0})$. We can also consider ϕ as an element of $H^0(i_*\Omega^1_{\tilde{X}_0}\otimes\omega_{X_0})$. Let $\phi_{|V_j}$ be a restriction of ϕ to V_j $(j=1,2)$. Then $\phi_{|V_2}=0$, since

$$\phi_{|V_2}\in H^0(\Omega^1_{V_2}\otimes\mathcal{O}_{V_2}(-mE))\subset H^0(\Omega^1_{V_2})=0.$$

On the other hand, we have

$$\phi_{|V_1}\in H^0(\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}(mE+F)).$$

As E is a singular fiber of type ${}_mI_0$ on V_1, it is clear that $\mathcal{O}_{V_1}(mE+F)\otimes\mathcal{O}_E\cong\mathcal{O}_E$. Hence there is an exact sequence

$$0\to\mathcal{N}\to\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}(mE+F)\xrightarrow{r}\Omega^1_E\to 0,$$

where $\mathcal{N}$ is the kernel of r. Moreover we have an exact sequence

$$0\to\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}((m-1)E+F)\to\mathcal{N}\to\mathcal{O}_E((m-1)E+F)\to 0.$$

For these sequences, see Kodaira and Spencer [5]. From $\phi_{|V_2}=0$, it follows that $\phi_{|V_1}\in H^0(\mathcal{N})$. Since

$$H^0(\mathcal{O}_E((m-1)E+F))=H^0((m-1)N_{E/V_1})=0,$$

we have

$$\phi_{|V_1}\in H^0(\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}((m-1)E+F)).$$

Therefore we infer that $\phi_{|V_1}=0$, since

$$\begin{aligned} H^0(\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}((m-1)E+F)) &= H^0(\Omega^1_{V_1}\otimes\Omega^2_{V_1}) \\ &= H^2(\Theta_{V_1})^* && \text{(Serre duality)} \\ &= 0. && \text{(the assumption (0))} \end{aligned}$$

Lemma 2. *The following natural pairing is surjective,*

$$H^1(T^0_{X_0})\otimes H^0(T^1_{X_0})\to H^1(T^1_{X_0}).$$

Proof. Recall that $T^1_{X_0}=\mathcal{O}_E$. Hence $H^0(T^1_{X_0})=H^1(T^1_{X_0})=\mathbf{C}$. Let ξ be a non-zero element of $H^0(T^1_{X_0})$. Then,

(1) $$T^0_{X_0}\xrightarrow{[\ ,\xi]}T^1_{X_0}$$

is surjective, where $[\ ,\ ]$ is the bracket. Let K be the kernel of this map, and $\check{K}$ be its dual. Then,

(2) $$\check{K}=\Lambda^1_{X_0},$$

where $\Lambda^1_{X_0}$ is the abstract log complex. These (1) and (2) are the main two lemmas of Friedman [1]. See [1] for the definitions and proofs. By (1), we have the following exact sequence

$$0\to K\to T^0_{X_0}\xrightarrow{[\ ,\xi]}T^1_{X_0}\to 0$$

and hence

$$H^1(T^0_{X_0})\to H^1(T^1_{X_0})\to H^2(K)$$

is exact. By Serre duality and (2), we have $H^2(K)=H^0(\Lambda^1_{X_0}\otimes\omega_{X_0})^*$. So it suffices to show that

$$H^0(\Lambda^1_{X_0}\otimes\omega_{X_0})=0.$$

Let ϕ be an element of $H^0(\Lambda^1_{X_0}\otimes\omega_{X_0})$. $\phi_{|V_2}$ vanishes everywhere on V_2, since by the definition of $\Lambda^1_{X_0}$,

$$\phi_{|V_2}\in H^0(\Omega^1_{V_2}(\log E)\otimes\mathcal{O}_{V_2}(-mE))\subset H^0(\Omega^1_{V_2})=0.$$

On the other hand, we have

$$\phi_{|V_1}\in H^0(\Omega^1_{V_1}(\log E)\otimes\mathcal{O}_{V_1}(mE+F)).$$

We consider the residue sequence

$$0\to\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}(mE+F)\to\Omega^1_{V_1}(\log E)\otimes\mathcal{O}_{V_1}(mE+F)\to\mathcal{O}_E\to 0.$$

From $\phi_{|V_2}=0$, it follows that

$$\phi_{|V_1}\in H^0(\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}(mE+F)).$$

Therefore the proof of this Lemma 2 is reduced to that of Lemma 1.

3.

We use the notation of Sect. 2.

Example 1. Let V_1 be a relatively minimal elliptic surface over $\mathbf{P}^1$ with a multiple fiber E of type ${}_mI_0$, $K_{V_1}=(m-1)E$, $p_g(V_1)=1$, $q(V_1)=0$, and $\kappa(V_1)=1$. For example, V_1 is obtained from an elliptic $K3$ surface by means of a logarithmic transformation. It is easy to see that $H^2(V_1,\Theta_{V_1})=0$. Let V_2 be a Hopf surface with $K_{V_2}=-(m+1)E$. As easily seen, we can choose V_1 in such a way that $N_{E/V_1}\otimes N_{E/V_2}\cong\mathcal{O}_E$. We set $X_0=V_1\cup_E V_2$. In the degeneration $\pi:X\to\Delta$ of the above Proposition, a general fiber X_t is a $K3$ surface. Thus we obtain a family of $K3$ surfaces which degenerates to a surface with $\kappa=1$ and a Hopf surface.

Proof. There is an exact sequence

$$0\to\mathcal{O}_{X_0}\to i_*\mathcal{O}_{\tilde{X}_0}\to\mathcal{O}_E\to 0,$$

hence the following one is exact,

$$\begin{aligned}0\to H^1(\mathcal{O}_{X_0})\to H^1(\mathcal{O}_{V_1})\oplus H^1(\mathcal{O}_{V_2})\to H^1(\mathcal{O}_E)\to\\ \to H^2(\mathcal{O}_{X_0})\to H^2(\mathcal{O}_{V_1})\oplus H^2(\mathcal{O}_{V_2})\to 0.\end{aligned}$$

Here, $H^1(\mathcal{O}_{V_1})=0$, $H^2(\mathcal{O}_{V_1})=\mathbf{C}$, $H^2(\mathcal{O}_{V_2})=0$, and $H^1(\mathcal{O}_{V_1})\to H^1(\mathcal{O}_E)$ is isomorphic, hence $H^1(\mathcal{O}_{X_0})=0$, $H^2(\mathcal{O}_{X_0})=\mathbf{C}$. By the upper semi-continuity, we have $q(X_t)=0$ and $p_g(X_t)=1$ from $\chi(\mathcal{O}_{X_t})=\chi(\mathcal{O}_{X_0})=2$. Thus we can take a non-zero element ϕ of $H^0(K_X)$ (see Friedman [1]). The divisor (ϕ) can be written in the form

$$n_1V_1+n_2V_2+D,\ \text{where } D \text{ does not contain } V_1 \text{ nor } V_2.$$

Both $D_{|V_1}=-kE+F$ and $D_{|V_2}=kE$ are linearly equivalent to effective divisors. (In particular, $k\geqq 0$). By the adjunction formula, we have $K_{V_1}=(n_2-n_1-1-k)E+F$ and $K_{V_2}=(n_1-n_2-1+k)E$. So $n_1-n_2-1+k=-(m+1)$, and hence $K_{V_1}=(m-1)E+F$. Therefore we infer that $F=0$ and $k=0$, so that K_{X_t} is trivial for non-zero t. (Here we abuse the symbol "=". By "=", we also mean a linear equivalence.) Q.E.D.

Similarly, we obtain the following two examples.

Example 2. Let V_1 be an Enriques surface. It is an elliptic surface over $\mathbf{P}^1$ with two multiple fibers E, E' of type ${}_2I_0$, and $K_{V_1}=E+E'-C$, where C is a general fiber. It is well-known that $H^2(V_1,\Theta_{V_1})=0$. Let V_2 be a Hopf surface with $K_{V_2}=-3E$. In the degeneration $\pi:X\to\Delta$, the general fiber X_t $(t\neq 0)$ is a rational surface. Thus we obtain a family of rational surfaces which degenerates to an Enriques surface and a Hopf surface.

Proof. We can prove $q(X_t)=0$ as in Example 1. Let us assume that $h^0(rK_{X_t})\geqq 1$ for some $r\geqq 1$. We proceed as in Example 1. We can then take a non-zero element ϕ of

$H^0(rK_X)$. (ϕ) can be written in the form $n_1V_1+n_2V_2+D$, and both $D_{|V_1}=-kE+F$ and $D_{|V_2}=kE$ are linearly equivalent to effective divisors. We have

$$rK_{V_1}=(n_2-n_1-r-k)E+F, \quad \text{and} \quad rK_{V_2}=(n_1-n_2-r+k)E,$$

hence $n_1-n_2-r+k=-3r$, and so $rK_{V_1}=rE+F$. Therefore we infer that $F=rE'-rC$, and so that $D_{|V_1}=-kE+rE'-rC$ is never linearly equivalent to an effective divisor. This is a contradiction. Thus we have $h^0(rK_{X_t})=0$ for any $r\geqq 1$. By Castelnuovo's criterion, we conclude that X_t is a rational surface. Q.E.D.

Example 3. Let V_1 be an elliptic surface of class VII_0 with a general fiber C and just three multiple fibers E, E', E'' of type ${}_mI_0$, ${}_{m'}I_0$, ${}_{m''}I_0$ respectively. V_1 is obtained from $\mathbf{P}^1\times C$ by means of three times logarithmic transformations. If $\frac{1}{m}+\frac{1}{m'}+\frac{1}{m''}<1$ (resp. $=1$, >1), then the Kodaira dimension of V_1 is 1 (resp. 0, $-\infty$). We can also show that $H^2(V_1,\Theta_{V_1})=0$. See Kodaira [4, II] for the proofs of these facts. Let V_2 be a Hopf surface with $K_{V_2}=-(m+1)E$. We set $X_0=V_1\cup_E V_2$. In the degeneration $\pi: X\to\Delta$, the general fiber is a surface of class VII with $\kappa(X_t)=-\infty$.

Proof. We can prove that $q(X_t)=1$ and $\kappa(X_t)=-\infty$ as in Example 1 and Example 2. So it suffices to show that $b_1(X_t)=1$. But we can prove this, using the standard topological argument. We shall review it briefly. Let W be a sufficiently small open neighborhood of E in X, and W_0 (resp. W_t) be the intersection of W and X_0 (resp. X_t). The boundary ∂W_0 of W_0 in X_0 is naturally homeomorphic to ∂W_t. Then X_t is homeomorphic to the union of X_0-W_0 and $\bar{W}_t$ with ∂W_0 and ∂W_t naturally identified where $\bar{W}_t$ is the closure of W_t in X_t. Hence we easily see that $b_1(X_t)=1$. Here we must note that one of the connected components of ∂W_t is a strong deformation retract of $\bar{W}_t$. For more details, we refer the reader to Nakamura, "On surfaces of class VII_0" (preprint).

Remark. In the above three Examples, we have not checked that each elliptic surface V_1 satisfies

(0) $$H^2(V_1,\Theta_{V_1})=0.$$

As for Example 3, Kodaira showed that if an elliptic surface S of class VII_0 contains just $r(\geqq 3)$ multiple fibers, then

$$\dim H^2(S,\Theta_S)=r-3.$$

(See [4, II, p. 692].)

We shall briefly consider the case of Example 1. Recall that V_1 in Example 1 is an elliptic surface over $\mathbf{P}^1$ with a multiple fiber E of type ${}_mI_0$, and $K_{V_1}=(m-1)E$. By Serre duality,

$$\dim H^2(V_1,\Theta_{V_1})=\dim H^0(V_1,\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}((m-1)E)).$$

We take a meromorphic 1-form

$$\phi\in H^0(V_1,\Omega^1_{V_1}\otimes\mathcal{O}_{V_1}((m-1)E)).$$

Local calculations show that ϕ is, in fact, the pull back of a meromorphic 1-form on the base curve $\mathbf{P}^1$ with only one pole of order 1 at most. But this must vanish everywhere. Thus we have $H^2(V_1, \Theta_{V_1})=0$. Note that if we consider an elliptic surface V_1' with *at least two* multiple fibers which is obtained from an elliptic $K3$ surface by means of logarithmic transformations, then $H^2(V_1', \Theta_{V_1'})\neq 0$.

Similarly in Example 2, considering the canonical bundle, we see that an Enriques surface V_1 satisfies (0), but an elliptic surface with *at least three* multiple fibers which is obtained from a rational elliptic surface, does not satisfy (0).

Thus, in each Example, the number of multiple fibers on each V_1 is maximum for it to satisfy the assumption (0).

Note. The author has constructed another counter-example to the Conjecture in the Introduction, more precisely, a degeneration of abelian surfaces whose singular fiber contains a component of a surface with Kodaira dimension 1. The main tool of its construction is not the deformation theory but rather complicated "logarithmic transformations" (see [7]).

References

1. Friedman, R.: Global smoothings of varieties with normal crossings (to appear)
2. Friedman, R., Miranda, R.: Smoothing cusp singularities of small length (to appear)
3. Kodaira, K.: On compact analytic surfaces. II. Ann. Math. **77**, 563–626 (1963)
4. Kodaira, K.: On the structure of compact analytic surfaces. I, II. Am. J. Math. **86**, 751–798 (1964); **88**, 682–721 (1966)
5. Kodaira, K., Spencer, D.C.: On a theorem of Lefschetz and the lemma of Enriques-Severi-Zariski. Proc. Natl. Acad. Sci. **39**, 1273–1278 (1953)
6. Kulikov, V.S.: Degenerations of $K3$ surfaces and Enriques surfaces. Math. USSR Izvestija **11**, 957–989 (1977)
7. Nishiguchi, K.: Degeneration of surfaces with trivial canonical bundles (in preparation)
8. Palamodov, V.P.: Deformations of complex spaces. Russ. Math. Survays **31**, 129–197 (1976)
9. Persson, U.: Degenerations of algebraic surfaces. Memoirs of A.M.S. **189** (1977)
10. Persson, U., Pinkham, H.: Degeneration of surfaces with trivial canonical bundle. Ann. Math. **113**, 45–66 (1981)
11. Ueno, K.: Appendix to T. Ashikaga's paper. Tohoku Math. J. **33**, 210–214 (1981)

Received October 12, 1982; in revised form January 31, 1983

Math. Ann. 263, 385–397 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Deforming Vector Bundles on the Projective Plane

Stein Arild Strømme
Department of Mathematics, University of Bergen, N-5014 Bergen, Norway

0. Introduction

The starting point of the present work was the following problem, as communicated to me by Bănică [1]: Characterize those rank-2 vector bundles E on $\mathbb{P}^2$ that can be deformed to the trivial bundle $2\mathcal{O}_{\mathbb{P}^2}$. A necessary condition is clearly that the Chern classes of E be trivial. If we denote by $d(E)$ the largest integer such that $E(-d(E))$ has a global section, it is known that E is always deformable if $d(E) \leqq 2$ [1, 11]. Even in this case, to my knowledge, no reasonably explicit families are known.

The topologically trivial 2-bundles E on $\mathbb{P}^2$ with given type $d(E) = d \geqq 1$ are parametrized in a natural way by an irreducible nonsingular variety $M(d)$ of dimension $(3d^2 - 1)$, see (2.7) below, or [1] or [11]. Denoting by $M(d; 0) \subseteqq M(d)$ the set corresponding to bundles that can be deformed to the trivial one, we show in this paper that $M(d; 0)$ is closed of codimension $(d-1)(d-2)$ in $M(d)$, thus covering the result of Bănică and Schafft mentioned above (3.13) and (4.7). We also give one sufficient and one necessary condition for deformability to the trivial bundle (4.6), (5.2). In turns out that the methods we use work for arbitrary Chern classes c_1 and c_2, and hence give similar results on questions like: which bundles are deformable to a stable bundle, or a decomposable bundle, or just to any bundle which is "more stable".

The material is organized as follows: Section 1 contains basic preliminaries on 2-bundles on $\mathbb{P}^2$, the invariant $d(E)$ is defined, and dimensions of some cohomology groups are computed. Section 2 contains a study of general families of bundles. In Sect. 3 some extra structure is added to a bundle in order to overcome the difficulties arising from the non-existence of moduli for unstable bundles. We describe a certain Quot scheme and determine its irreducible components, allowing us to conclude that if d is large enough, the general bundle of type d is not deformable to smaller type. In Sect. 4 we give an explicit construction of many deformable bundles. Section 5 contains some conjectures.

Notation and *Conventions*

k – an algebraically closed field of characteristic $\neq 2$.
$\mathbb{P}^2$ – projective plane over k.
$\mathcal{O}$ – structure sheaf on $\mathbb{P}^2$.
p_T: $\mathbb{P}^2_T = \mathbb{P}^2 \times T \to T$ is the second projection.
Bundle – see (1.1).

If F is a coherent sheaf on a k-scheme X, then $H^n(F) = H^n(X, F)$ and $h^n(F) = \dim_k H^n(F)$. $\chi(F) = \sum_i (-1)^i h^i(F)$ is the Euler characteristic.

P – the polynomial $P(x) = (x-1)(x-2-c_1) - c_2$.
χ = the number $4 - 4c_2 - c_1$, see (1.1) for c_1 and c_2.

1. Preliminaries

(1.1) Let c_1 and c_2 be two integers such that $c_1(c_1+1) = 0$. We shall use the term *bundle* (with no further qualification) to designate a locally free sheaf E of rank 2 on $\mathbb{P}^2$ with Chern classes $c_1(E) = c_1$ and $c_2(E) = c_2$. Recall that a bundle E is *stable* (in the sense of Mumford-Takemoto) if and only if $h^0(E) = 0$. If E is *unstable* (not stable), denote by $d = d(E)$ the largest integer such that $h^0(E(-d)) \neq 0$. Then $d(E)$ is called the *type* of E. If E is stable, we agree to put $d(E) = -1$. If $d \geqq -1$ is an integer, we shall denote by $B(d)$ the set of isomorphism classes of bundles of the given type.

(1.2) **Proposition.** *$B(d)$ is nonempty if and only if one of the following two conditions hold:*

(i) $d \geqq 0$ *and* $d^2 - dc_1 + c_2 \geqq 0$,
(ii) $d = -1$ *and* $c_2 - c_1 \geqq 2$.

Proof. For stable bundles, this is well known, see for example [9]. If E is a bundle of type $d \geqq 0$, pick a non-zero section s of $E(-d)$. Then its zero-scheme $Y = V(s)$ has codimension two, hence it has finite length $c_2(E(-d)) = d^2 - dc_1 + c_2$. Conversely, suppose (i) holds, and let Y be a closed, locally complete intersection subscheme of finite length $= (d^2 - dc_1 + c_2)$. Then a sufficiently general extension of $I_Y(c_1 - d)$ by $\mathcal{O}(d)$ gives a bundle of type d. See for example [5, Theorem 1.1] for details.

(1.3) From the above proof we obtain, for each unstable bundle E, a basic exact sequence

$$0 \to \mathcal{O}(d) \xrightarrow{s} E \to I_Y(c_1 - d) \to 0.$$

If E is not the trivial bundle $2\mathcal{O}$, the minimal section s is unique up to a non-zero scalar. In particular, the subscheme $Y = Y(E)$ is uniquely determined by E. The element $\xi(E) = \xi$ of $\operatorname{Ext}^1(I_Y(c_1 - d), \mathcal{O}(d))$ corresponding to the basic exact sequence is zero if and only if E is decomposable (and $Y = \emptyset$), otherwise it is determined up to a non-zero scalar only. It has the property that the induced global section of the local Ext-sheaf $\underline{\operatorname{Ext}}^1(I_Y(c_1 - d), \mathcal{O}(d)) \cong \omega_Y$ generates this sheaf.

(1.4) **Proposition.** *Let E be a bundle of type $d \geqq -1$. Then the following equalities hold:*

(i) $\chi(\underline{\operatorname{End}} E) = \chi = 4 - c_1 - 4c_2$.

(ii) *If* $d \geqq 0$, *then*

$$h^0(\underline{\mathrm{End}}\, E) = 2d(d-c_1)+3d-2c_1+2+\alpha+\alpha\beta$$

$$h^1(\underline{\mathrm{End}}\, E) = 4(d^2-dc_1+c_2)-1+\alpha-\beta+\alpha\beta$$

$$h^2(\underline{\mathrm{End}}\, E) = 2d(d-c_1)-3d+c_1+1-\beta$$

where $\alpha = \begin{cases} 0 & \text{if } d^2-dc_1+c_2>0 \\ 1 & \text{if } d^2-dc_1+c_2=0 \end{cases}$

and $\beta = \begin{cases} 0 & \text{if } (d,c_1)\neq(0,0) \\ 1 & \text{if } d=c_1=0. \end{cases}$

(iii) *If* $d=-1$, *then*

$$h^0(\underline{\mathrm{End}}\, E) = 1$$

$$h^1(\underline{\mathrm{End}}\, E) = c_1+4c_2-3$$

$$h^2(\underline{\mathrm{End}}\, E) = 0.$$

Proof. (i) Follows from Riemann-Roch, and (iii) is well known. (ii) is an easy computation, which is left to the reader with the remark that, if $E \neq 2\mathcal{O}$, then any endomorphism of E must leave the image of the map $s: \mathcal{O}(d) \to E$ invariant.

2. Families of Bundles

(2.1) Let T be a k-scheme, and $\mathscr{E}$ a locally free sheaf of rank 2 on $\mathbb{P}^2_T$. If t is a k-point of T, we denote by $\mathscr{E}_t$ the restriction of $\mathscr{E}$ to $\mathbb{P}^2 = \mathbb{P}^2 \times \{t\} \subseteqq \mathbb{P}^2_T$, and call it the *fiber of* $\mathscr{E}$ *over* t. If all the fibers of $\mathscr{E}$ are bundles [recall the convention of (1.1)] we shall say that $\mathscr{E}/T$, or simply $\mathscr{E}$, is a *family of bundles* (parametrized by T). The family is called *irreducible* if T is an irreducible scheme of finite type over k. If $\mathscr{E}/T$ is an irreducible family, the *type* of $\mathscr{E}$, denoted by $d(\mathscr{E})$, is defined to be the minimum of the $d(\mathscr{E}_t)$, $t \in T$. A bundle E is *deformable to type* d if it is a fiber in some irreducible family of type d.

(2.2) By Serre duality $h^0(E(-d)) = h^2(E^\vee(d-3))$, hence if $d \geqq 0$ we have $d(E) \geqq d$ if and only if $h^2(E^\vee(d-3)) \neq 0$. If $\mathscr{E}/T$ is a family of bundles, the coherent sheaf $R^2p_T{*}\mathscr{E}^\vee(d-3)$ commutes with base change on T, thus the type of $\mathscr{E}_t$ is an upper semicontinuous function on T. From [10, lecture 8] we also obtain a unique maximal locally closed subscheme $T(d) = T(d; \mathscr{E}) \subseteqq T$ with the property that the restriction to $T(d)$ of $R^2p_T{*}\mathscr{E}^\vee(d-3)$ is locally free of rank 1. Let us make the following definition:

(2.3) Let $d \geqq 0$, and let $\mathscr{E}/T$ be a family. We say that $\mathscr{E}$ has *pure type* d if $R^2p_T{*}\mathscr{E}^\vee(d-3)$ is locally free of rank 1. We agree that $\mathscr{E}$ has *pure type* -1 if all the fibres are stable bundles.

(2.4) *Remark.* If $\mathscr{E}/T$ has pure type $d \geqq 0$, then all the fibers $\mathscr{E}_t$ have type d. If T is reduced of finite type over k, the converse also holds.

(2.5) **Lemma.** *Let $\mathscr{E}/T$ be a family and $d \geqq 0$ an integer. Then $\mathscr{E}$ has pure type d if and only if $p_T*\mathscr{E}(-d)$ is locally free of rank 1 and commutes with base change on T.*

Proof. Follows from the relative duality isomorphism, functorial in N,

$$D^0 : p_T*(\mathscr{E}(-d) \otimes p_T^* N) \to \underline{\mathrm{Hom}}(R^2 p_T*\mathscr{E}^\vee (d-3), N)$$

where N is any quasicoherent sheaf on T [7, Theorem 21].

(2.6) Consider the following (contravariant) functors $\underline{B}(d)$: (schemes/k)$\to$(sets) defined by $\underline{B}(d)(T) = \{\text{families } \mathscr{E}/T \text{ of pure type } d\}/\sim$, where $\mathscr{E}_1 \sim \mathscr{E}_2$ if there exists a linebundle $\mathscr{L}$ on T and an isomorphism $\mathscr{E}_1 \cong \mathscr{E}_2 \otimes p_T^* \mathscr{L}$.

(2.7) **Proposition.** *There is an irreducible, nonsingular, quasiprojective and rational variety $M(d)$ acting as a coarse moduli space for $\underline{B}(d)$. If $d = -1$, then $\dim M(d) = 4c_2 + c_1 - 3$, and if $d \geqq 0$ and $d^2 - dc_1 + c_2 = 0$, then $M(d)$ is a point. If $d \geqq 0$ and $d^2 - dc_1 + c_2 > 0$, then $\dim M(d) = 3(d^2 - dc_1 + c_2) - 1$.*

Proof. For the case $d = -1$, see [2, 6, 3]. The case $d^2 - dc_1 + c_2 = 0$ is trivial. If $d^2 - dc_1 + c_2 > 0$, we apply the same construction as is carried out in [1] and [11] for topologically trivial bundles. Note that the correspondence $E \leftrightarrow (Y(E), \xi(E))$ of (1.4) generalizes to families of pure type, by (2.5). Thus, if H denotes the Hilbert scheme parametrizing the $Y(E)$'s, one can obtain $M(d)$ as an open subset of $\mathbb{P}_H(D^\vee)$, where D is a certain locally free sheaf on H inducing $\mathrm{Ext}^1(I_Y(c_1 - d), \mathscr{O}(d))$ in each point $\{Y\} \in H$. We omit the details.

(2.8) *Remark.* If E is a bundle, the tangent space of its infinitesimal deformation functor is canonically isomorphic to $\mathrm{Ext}^1(E, E) = H^1(\underline{\mathrm{End}}(E))$. Letting E correspond to the point $m \in M(d)$, $d \geqq 0$, the tangent space of $M(d)$ at m can be identified with the kernel of the map

$$\mathrm{Ext}^1(E, E) \to \mathrm{Ext}^1(\mathscr{O}(d), I_Y(c_1 - d))$$

induced by the basic sequence of (1.3).

3. Bundles with Sections

(3.1) Unfortunately, the functor of all families of bundles is not even prorepresentable. To remedy this we do the fairly standard trick of adding enough structure to the bundles to obtain a representable functor. One way of doing this is to choose a k-basis for the vector space $H^0(E(n))$ for some large but fixed n, leading to the Quot scheme of Grothendieck. To start with, fix a type $\bar{d}$ and an integer n large enough that $E(n)$ is generated by its global sections and $H^i(E(n)) = 0$ for $i \neq 0$ for all $E \in B(d)$ with $d \leqq \bar{d}$. Put $N = (n+1)(n+2+c_1) - c_2$, then, by Riemann-Roch, $h^0(E(n)) = N$ for all these E. Let V be a vector space of dimension N. As a matter of notation, put $V(-n) = V \bigotimes_k \mathscr{O}(-n)$ and $V_T(-n) = V \bigotimes_k \mathscr{O}_{\mathbb{P}^3_T}(-n)$ for any k-scheme T.

(3.2) Denote by Quot the Grothendieck Quot scheme parametrizing quotients $V(-n) \to F$ with Hilbert polynomial $\chi(F(\nu)) = (\nu+1)(\nu+2+c_1) - c_2$ [4] and let $Q \subseteq \mathrm{Quot}$ be the open subscheme defined by the conditions (i) F is locally free of

type $\leqq \bar{d}$, and (ii) the induced map $V \to H^0(F(n))$ is an isomorphism. Q comes equipped with a universal exact sequence

$$0 \to \mathscr{K} \to V_Q(-n) \xrightarrow{\tilde{q}} \mathscr{F} \to 0.$$

From deformation theory we have the following facts about the local structure of Q:

(3.3) **Proposition.** *Let* $q = (0 \to K \to V(-n) \to F \to 0)$ *be a closed point of* Q. *Let* $\dim(Q_q)$ *be the dimension of* Q *at* q, *and denote by* A *the sheaf* $\underline{\mathrm{Hom}}(K, F)$.

(i) *The tangent space* $T_{Q,q}$ *is canonically isomorphic to* $H^0(A)$.

(ii) $h^0(A) - h^1(A) \leqq \dim(Q_q) \leqq h^0(A)$.

For a proof, see for example [8, (4.2.4)].

(3.4) **Corollary.** *Let* $\chi = 4 - 4c_2 - c_1$. *Then* $N^2 - \chi \leqq \dim Q_q \leqq N^2 - \chi + h^2(\underline{\mathrm{End}}\, F)$. Q *is smooth at* q *if and only if the rightmost inequality is an equality.*

Proof. This is essentially a rephrasing of (3.3), applying $\underline{\mathrm{Hom}}(-, F)$ to the sequence q of (3.3), and noting that $\chi = \chi(\underline{\mathrm{End}}\, F)$.

(3.5) $\mathscr{F}$ is a family of vector bundles parametrized by Q. Recalling the definition of (2.3), denote by $Q(d) = Q(d; \mathscr{F})$ the maximal locally closed subscheme such that the restriction $\mathscr{F}^d$ of $\mathscr{F}$ to $\mathbb{P}^2_{Q(d)}$ is of pure type $d \geqq 0$. Also denote by $Q(-1)$ the open subscheme parametrizing stable quotients, and by $\mathscr{F}^{-1}$ the corresponding restriction. $\bar{Q}(d)$ will denote the closure of $Q(d)$ in Q.

(3.6) Since for each d $(-1 \leqq d \leqq \bar{d})$, $M(d)$ is a coarse moduli space, we get a uniquely defined morphism

$$\phi = \phi(d) : Q(d) \to M(d)$$

defined by the family $\mathscr{F}^d$. It is clearly surjective by our choice of n.

(3.7) **Proposition.** $\phi(d)$ *is smooth of relative dimension* $N^2 - h^0(\underline{\mathrm{End}}\, E)$, *where* E *is any bundle in* $B(d)$.

Proof. Let $q : V(-n) \to E$ be any k-point of $Q(d)$ and put $K = \ker q \subseteq V(-n)$. The smoothness follows easily from the infinitesimal lifting criterion [12], using that $H^1(E(n)) = 0$. For the dimension assertion, apply the functor $\mathrm{Hom}(-, E)$ to the sequence $0 \to K \to V(-n) \to E \to 0$ to get

$$0 \to \mathrm{End}\, E \to \mathrm{Hom}(V(-n), E) \to \mathrm{Hom}(K, E) \xrightarrow{\delta} H^1(\underline{\mathrm{End}}\, E) \to 0.$$

Clearly, $T_{Q(d),q} = \delta^{-1} T_{M(d),\{E\}}$ as a subspace of $T_{Q,q} = \mathrm{Hom}(K, E)$. The assertion follows.

(3.8) *Remark.* There is an obvious action of $\mathrm{Aut}(V) = GL(N)$ on $Q(d)$, transitive on the fibers of $\phi(d)$, and $M(d)$ is the (geometric) quotient of $Q(d)$ by this action. The isotropy subgroup of a point $q : V(-n) \to E$ is easily seen to be isomorphic to $\mathrm{Aut}(E)$.

(3.9) **Theorem.** *Assume that* $-1 \leqq d \leqq \bar{d}$ *and that* $B(d)$ *is nonempty. Let* P *be the polynomial* $P(x) = (x-1)(x-2-c_1) - c_2$ *and* $\chi = 4 - 4c_2 - c_1$.

(i) *$Q(d)$ is irreducible and smooth of dimension*

$$\dim Q(d) = \begin{cases} N^2-\chi & \text{if } d=-1 \text{ or } d=c_1=c_2=0 \\ N^2-\chi+P(d)-1 & \text{otherwise.} \end{cases}$$

(ii) *$\bar{Q}(d)$ is an irreducible component of Q_{red} if and only if* $\dim Q(d) \geqq N^2-\chi$.

(iii) *$\bar{Q}(d)$ is not an irreducible component of Q_{red} if and only if one of the following two conditions holds:*

(a) $0 \leqq c_2 \leqq 1+c_1$ *and* $1 \leqq d \leqq 2+c_1$.

(b) $c_2 \geqq 2+c_1$ *and* $d \geqq 0$ *and* $P(d) \leqq 0$.

(iv) *If $d=-1$ or $d^2-dc_1+c_2=0$, then Q is nonsingular along $Q(d)$. In the remaining cases, if $\bar{Q}(d)$ is a component of Q_{red}, the embedding codimension of Q at any point of $Q(d)$ equals $(d^2-dc_1+c_2)$, hence Q is not reduced along $Q(d)$.*

Proof. (i) follows from (3.7), (2.7), and (1.3). (ii) The necessity of the condition has already been noted (3.4). Conversely, if $\bar{Q}(d)$ is not a component, it must be contained in $\bar{Q}(e)$ for some $e<d$. Consequently $\dim Q(d) < \dim Q(e)$. If $e=-1$ or $e=c_1=c_2=0$, we are done, since $\dim Q(e)=N^2-\chi$ in this case, by (i). In the remaining cases we must have $P(d)<P(e)$. Looking at the definition of P, we must have $e=0$, and either $c_1=0$ and $1 \leqq d \leqq 2$, or $c_1=-1$ and $d=1$. In either case $P(d) = -c_2 \leqq 0$ since $Q(0)$ is nonempty. Hence $\dim Q(d) \leqq N^2-\chi-1$. (iii) is merely a rephrasing of (ii). (iv) follows from (3.4) and (i) above.

(3.10) We want to interpret Theorem (3.9) in terms of the bundles themselves, without reference to the Quot scheme. Recalling the notion of deformability (2.1), let us consider the following subset of $B(d)$, where $-1 \leqq e < d$ are integers:

$$B(d;e) = \{D \in B(d) \text{ such that } D \text{ is deformable to type } e\}.$$

(3.11) **Lemma.** *Let $q: V(-n) \to D$ be a point of $Q(d)$. Then q is in $\bar{Q}(e)$ if and only if D is in $B(d;e)$.*

Proof. If $\mathscr{D}/T$ is an irreducible family of type e containing D, the sheaf $p_T*\mathscr{D}(n)$ is locally free of rank N. Shrinking T if necessary, we may assume it is free. Then the isomorphism $V \to H^0(D(n))$ extends to an isomorphism $V \otimes \mathscr{O}_T \to p_T*\mathscr{D}(n)$, hence gives a morphism $T \to Q$ hitting the point q. Since the generic point of T lands in Q(e), we are done.

(3.12) It follows that the subset $M(d;e) \subseteq M(d)$ corresponding to $B(d;e) \subseteq B(d)$ is closed and that $Q(d) \cap \bar{Q}(e) = \phi(d)^{-1} M(d;e)$. In particular, the number of irreducible components of $M(d;e)$ and their codimensions in $M(d)$ are the same as those for $Q(d) \cap \bar{Q}(e)$ in $Q(d)$. Theorem (3.9) thus translates as follows:

(3.13) **Theorem.** *Let $-1 \leqq e < d$, and put*

$$\gamma = \begin{cases} P(d) & \text{if } e=-1 \text{ or } e=c_1=c_2=0 \\ P(d)-P(e)+1 & \text{otherwise.} \end{cases}$$

(i) *Each irreducible component of $M(d;e)$ has codimension at least γ in $M(d)$.*

(ii) *Assume $\gamma \leqq 0$. Then $e \leqq 0$. Furthermore, if $e_0 \leqq e$ is the smallest integer such that $M(e_0)$ is nonempty, then $M(d;e_0)=M(d)$.*

(iii) *Let E be a bundle of type d, and let H denote the propresentable hull of the deformation functor of E* [12] *("the base of the miniversal deformation of E").*

Then, if $d=-1$ *or* $d^2-dc_1+c_2=0$, *then* H *is nonsingular. Otherwise,* H *is non-reduced of embedding codimension* $(d^2-dc_1+c_2)$, *unless* $M(d)=M(d;e)$ *for some* $e<d$.

Proof. (i) By (3.9, i) $\gamma=\dim Q(d)-\dim Q(e)+1$, clearly a lower bound for the codimension of $Q(d)\cap\bar{Q}(e)$ in $Q(d)$. Then use (3.12). (ii) follows from (3.9, ii) in the same way. For (iii), one uses that the fiber functor of Q at a point $q: V(-n)\to E$ is smooth over H, and (3.9, iv).

4. Existence

(4.1) Our results so far are mainly negative: With a few exceptions (3.13, ii) the general bundle of a given type is not deformable to smaller type. The purpose of this section is to give an explicit family of deformable bundles, large enough to show that all the $M(d;e)$ are non-empty. For our description we shall use monads, i.e. three-term complexes resolving the bundles.

(4.2) Let E be a given bundle of type e, and $\tau\in H^0(E(d-c_1))$ a global section vanishing in a subscheme $Y\subseteq\mathbb{P}^2$ of codimension 2, where $d>e$ is some integer. Furthermore, let $F\in H^0(\mathcal{O}(2d-c_1))$ be the equation of a curve disjoint from Y. Denote by D the cohomology bundle of the following monad:

$$\mathcal{O}(c_1-d)\xrightarrow{b}\mathcal{O}(c_1-d)\oplus E\oplus\mathcal{O}(d)\xrightarrow{a}\mathcal{O}(d)$$

where the maps a and b are given by

$$a=(F,\tau\wedge,0)\quad\text{and}\quad b=(0,\tau,-F)^t.$$

Let $\xi\in\mathrm{Ext}^1(I_Y,\mathcal{O}(c_1-2d))$ correspond to the short exact sequence induced by the section τ.

(4.3) **Proposition.** (i) D *is deformable to type* e.
(ii) $d(D)=d$.
(iii) $Y(D)=Y$.
(iv) $\xi(D)=F^2\xi$.

Proof. (i) Substitute a non-zero scalar λ for the zero in the maps a and b of the monad. Then the cohomology bundle is clearly E. Thus we have a family deforming D to E, letting $\lambda\to 0$.

(ii)–(iv): Consider the display of the monad (4.2):

$$\begin{array}{ccccccccc}
 & & & & 0 & & 0 & & \\
 & & & & \downarrow & & \downarrow & & \\
0 & \to & \mathcal{O}(c_1-d) & \longrightarrow & B^\vee(c_1) & \longrightarrow & D & \to & 0 \\
 & & \| & & \downarrow & & \downarrow & & \\
0 & \to & \mathcal{O}(c_1-d) & \to & \mathcal{O}(c_1-d)\oplus E\oplus\mathcal{O}(d) & \to & B & \to & 0 \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & \mathcal{O}(d) & & =\mathcal{O}(d) & & \\
 & & & & \downarrow & & \downarrow & & \\
 & & & & 0 & & 0 & &
\end{array}$$

The projection of the middle term onto the first factor $\mathcal{O}(c_1-d)$ factors through B, hence $\mathcal{O}(c_1-d)$ is a direct summand of B. In fact, we have $B=\mathcal{O}(c_1-d)\oplus N(d)$, where $N(d)$ is the cokernel of the map $(\tau, -F)^t: \mathcal{O}(c_1-d)\to E\oplus\mathcal{O}(d)$. Hence N is locally free of rank 2 [with $c_1(N)=0$ and $c_2(N)=d^2-dc_1+c_2$]. Now the proposition is a direct consequence of the following

(4.4) **Lemma.** *There is a commutative diagram with exact rows and cartesian (pushout) left squares:*

$$\begin{array}{ccccccc}
0 & \to & \mathcal{O}(c_1-2d) & \longrightarrow & E(-d) & \longrightarrow & I_Y \to 0 \\
 & & \downarrow F & & \downarrow & & \| \\
0 & \longrightarrow & \mathcal{C} & \longrightarrow & N & \longrightarrow & I_Y \to 0 \\
 & & \downarrow F & & \downarrow & & \| \\
0 & \to & \mathcal{O}(2d-c_1) & \to & D(d-c_1) & \to & I_Y \to 0.
\end{array}$$

Proof. The first two rows follows directly from the definition of N above. The last two rows follow from the top row of the display in (4.3), noting that $B^\vee \cong \mathcal{O}(d-c_1)\oplus N(-d)$.

(4.5) **Theorem.** *Let D be a bundle of type $d>0$, and assume that there exists an exact sequence*

$$0\to\mathcal{O}(c_1-d)\to E\to I_Y(d)\to 0$$

where $Y=Y(D)$ and E is some bundle of type e. Furthermore, assume that there exists a form $F_0\in H^0(\mathcal{O}(2d-c_1))$ such that multiplication with F_0 induces a surjection $H^1(E(-d))\to H^1(E(d-c_1))$. Then D is deformable to type e.

Proof. By (4.3) it suffices to show that the map

$$H^0(\mathcal{O}(2d-c_1))\times \operatorname{Ext}^1(I_Y(d),\mathcal{O}(c_1-d))\to \operatorname{Ext}^1(I_Y(c_1-d),\mathcal{O}(d))$$

given by $(F,\xi)\to F^2\xi$ is dominating. Let ξ_0 be the extension given by the exact sequence above, and fix a form F_0 as in the theorem such that $V(F_0)$ is disjoint from Y.

Differentiating the above map at the point (F_0,ξ_0), we obtain the linear map $(F,\xi)\to F_0(F_0\xi+2F\xi_0)$. Since the map

$$F_0: \operatorname{Ext}^1(I_Y,\mathcal{O})\to \operatorname{Ext}^1(I_Y(c_1-d),\mathcal{O}(d))$$

is an isomorphism by our choice of F_0, it suffices to show that the map

$$\beta: H^0(\mathcal{O}(2d-c_1))\times \operatorname{Ext}^1(I_Y(d),\mathcal{O}(c_1-d))\to \operatorname{Ext}^1(I_Y,\mathcal{O})$$

given by $\beta(F,\xi)=F_0\xi+2F\xi_0$ is surjective. To see that this is the case, use the exact sequence above to construct the right-hand column in the following diagram:

$$\begin{array}{ccc}
H^0(\mathcal{O}(2d-c_1)) & \xrightarrow{\alpha} & H^0(\mathcal{O}(2d-c_1)) \\
\downarrow & & \downarrow \\
H^0(\mathcal{O}(2d-c_1))\times \operatorname{Ext}^1(I_Y(2d-c_1),\mathcal{O}) & \xrightarrow{\beta} & \operatorname{Ext}^1(I_Y,\mathcal{O}) \\
\downarrow & & \downarrow p \\
\operatorname{Ext}^1(I_Y(2d-c_1),\mathcal{O}) & \xrightarrow{\gamma} & \operatorname{Ext}^1(E(-d),\mathcal{O}).
\end{array}$$

The upper right vertical map is $F \to F\xi_0$, the map α is multiplication by 2, the left vertical maps are the obvious ones, and γ is multiplication by F_0 followed by p. The diagram clearly commutes and the columns are exact. Since $\mathrm{char}(k) \neq 2$, α is an isomorphism. Thus it suffices to show that γ is surjective. Note that γ must factor through the map

$$F_0 : \mathrm{Ext}^1(E(d-c_1), \mathcal{O}) \to \mathrm{Ext}^1(E(-d), \mathcal{O}).$$

Using that E is locally free, we identify this map with the map

$$F_0 : H^1(E^\vee(c_1-d)) = H^1(E(-d)) \to H^1(E(d-c_1)) = H^1(E^\vee(d)),$$

which is surjective by assumption.

(4.6) *Remark.* As a special case ($e = c_1 = c_2 = 0$) we find that D is deformable to the trivial bundle if $Y(D)$ is a complete intersection of two curves of degree d (or a flat specialization of such complete intersections). For $d = 1$ or 2, we recover once more the result of Bănică and Schafft mentioned in the introduction. It is also a special case of (3.13, ii), our proof of which is essentially equivalent to those in [1] and [11]. We shall see later that if D is deformable to the trivial bundle, then at least $h^0(I_{Y(D)}(d)) \geqq 2$, thus giving a partial converse to (4.5) in this case. See (5.2) below.

(4.7) **Theorem.** *Let $-1 \leqq e < d$ be integers such that $M(e)$ is nonempty. Assume that either*

(a) $e = -1$ *or*

(b) $e \geqq 0$ *and* $\binom{d-e-1}{2} \geqq e^2 - ec_1 + c_2$.

Then $M(d; e)$ contains an irreducible component for which the bound of (3.13, i) *is sharp.*

Proof. In case (a), if $P(d) = (d-1)(d-2-c_1) - c_2 \leqq 0$, the result $M(d; -1) = M(d)$ is covered by (3.13, ii). We will thus assume that $P(d) \geqq 0$ in this case. In both cases, let E be a general bundle of $M(e)$. I claim that $H^i(E(d-c_1-3)) = 0$ for $i \neq 0$. Indeed, in case (a), E has natural cohomology, and $\chi(E(d-c_1-3)) = P(d) \geqq 0$ by Riemann-Roch. In case (b), $H^i(E(d-c_1-3)) = H^i(I_{Y(E)}(d-e-3))$ by the basic exact sequence. The assumption in (b) is that $\chi(I_{Y(E)}(d-e-3)) \geqq 0$. Since $Y(E)$ is in general position, the claim follows also in this case.

Now let $A(d; e) \subseteq M(e)$ be the open subset of those E for which $H^i(E(d-c_1-3)) = 0$ for $i \neq 0$. Denote by Z the set of pairs (E, τ) where $E \in A(d; e)$ and $\tau \in H^0(E(d-c_1))$ is a section vanishing in codimension two, and let W be the subset of the Hilbert scheme of $\mathbb{P}^2$ consisting of the $V(\tau)$ for $(E, \tau) \in Z$. From $H^i(E(d-c_1-3)) = 0$ for $i \neq 0$ one deduces easily that $\dim \mathrm{Ext}^1(I_Y(d), \mathcal{O}(c_1-d)) = 1$ for any $Y \in W$. Hence, if (E, τ) and (E', τ') are in Z with $V(\tau) = V(\tau')$, then $E \cong E'$ and τ and τ' are equivalent under the natural action of $\mathrm{Aut}(E)$ on $H^0(E(d-c_1))$. Hence the dimension of W is $\dim W = \dim M(e) + \chi(E(d-c_1)) - h^0(\underline{\mathrm{End}}\, E)$, hence its codimension in the Hilbert scheme is

$$\begin{aligned}&(2(d^2 - dc_1 + c_2) - \chi(E(d-c_1)) - (\dim M(e) - h^0(\underline{\mathrm{End}}\, E))\\ &\quad = (P(d) - \chi) - (\dim Q(e) - N^2).\end{aligned}$$

Using (3.9), this is precisely the bound γ of (3.13). The proof is completed using (4.5). Note that the assumption on F_0 is automatically satisfied, since $H^1(E(d-c_1))=0$ for $E\in A(d;e)$.

(4.8) In the remaining case, $e\geqq 0$ and $\binom{d-e-1}{2}<e^2-ec_1+c_2$, a little more care is needed to compute the dimension of the family satisfying the hypothesis of (4.5). The differences from the above proof are the following: For $A(d;e)\subsetneqq M(e)$ take the (locally closed) set of those E for which $Y(E)$ is contained in some nonsingular curve of degree $(d-e)$, but otherwise in general position. Then proceed as above, noting that

$$\dim \operatorname{Ext}^1(I_Y(d),\mathcal{O}(c_1-d))=1+h^1(I_{Y(E)}(d-e-3)).$$

The codimension of $A(d;e)$ in $M(e)$ is positive if $\chi(I_{Y(E)}(d-e))\leqq 0$, but the dimension of $H^0(E(d-c_1))$ increases by the same amount, so that the formula $\dim Z=\dim M(e)+\chi(E(d-c_1))$ still holds.

(4.9) **Theorem.** *Let $0\leqq e<d$ be integers such that*

$$\delta=e^2-ec_1+c_2-\binom{d-e-1}{2}>0.$$

Then $M(d;e)$ contains an irreducible component of codimension at most $P(d)-P(e)+1+\delta$ in $M(d)$.

(4.10) *Example.* If $d=e+1$, a bundle $D\in M(d)$ such that $Y(D)$ is on a line is in $M(d;d-1)$. The family of such D has codimension $(d^2-dc_1+c_2-2)$, which agrees with the formula of (4.9).

Proof of (4.9). In view of the above, it suffices to show that for a general E in $A(d;e)$, there exists a form F_0 satisfying the condition of (4.5). It thus suffices to prove the following:

(4.11) **Lemma.** *Let X be a group of $(e^2-ec_1+c_2)$ simple points, contained in some curve of degree $(d-e)$, but otherwise in general position. Let F_0 be a form of degree $(2d-c_1)$ such that the curve $B=V(F_0)$ is disjoint from X. Then a sufficiently general extension E of $I_X(c_1-e)$ by $\mathcal{O}(e)$ has the property that*

$$F_0: H^1(E(-d))\to H^1(E(d-c_1))$$

is surjective.

Proof. Consider the following exact commutative diagram

$$\begin{array}{ccccccc}
 & & & & H^0(\mathcal{O}_B(d-e)) & & \\
 & & & & \downarrow{\scriptstyle\beta} & & \\
0 & \to & H^1(E(-d)) & \to & H^1(I_X(c_1-e-d)) & \xrightarrow{\gamma} & H^2(\mathcal{O}(e-d)) \\
 & & \downarrow{\scriptstyle\alpha} & & \downarrow{\scriptstyle\delta} & & \\
0 & \to & H^1(E(d-c_1)) & \to & H^1(I_X(d-e)) & \to & 0.
\end{array}$$

The vertical lower maps are multiplication by F_0, and the horizontal maps come from the extension $\xi(E)$. A trivial diagram chase shows that, since δ is obviously surjective, α is also surjective if $im(\gamma\beta)=im(\gamma)$. To show that this is the case, consider the multiplicative group $T=H^0(\mathcal{O}_X^*)$. It acts in a natural way as a maximal torus on the vector space $\mathrm{Ext}^1(I_X(c_1-e),\mathcal{O}(e))$ in such a way that the open orbit is precisely the set of locally free extensions. It also acts compatibly on $H^1(I_X(c_1-e-d))$ and the set of maps γ as in the diagram above. Since X is in general position, subject only to the condition $H^0(I_X(d-e))\neq 0$, one sees that the image of β is transversal to each T-invariant subspace of $H^1(I_X(c_1-e-d))$. Hence the lemma follows from the following transversality result:

(4.12) **Proposition.** *Let $W\subseteq V$ be vector spaces of dimensions w, v resp., and let K be a quotient of V of dimension k. Let T be a maximal torus in $\mathrm{GL}(V)$ and assume that W is transversal to each T-invariant subspace of V. Then, for a general $t\in T$, the composed map $W\to V\xrightarrow{t} V\to K$ is of maximal rank.*

Proof. Without loss of generality we may assume that $w=k$. Let $\mathrm{Grass}_k(V)\subseteq\mathbb{P}(\wedge^k V)$ be the Plücker embedding. T acts naturally on both these varieties. A suitable 1-parameter subgroup T_1 of T will even act on $\wedge^k V$ with distinct weights. The transversality assumption in the theorem implies that all the $\binom{n}{k}$ Plücker coordinates of $(V\to V/W)$ are nonzero (with respect to a basis diagonalizing T). Hence consideration of a certain Vandermonde determinant shows that its orbit is not contained in any hyperplane of $\wedge^k V$, hence in particular not in the special Schubert cycle defined by K.

5. Open Questions

(5.1) Some problems arising naturally from our results so far are the following:

(I) Is the family constructed by the method of (4.2) dense in $M(d;e)$?

(II) If D is of type $d\geqq 0$ and $H^1(I_{Y(D)}(2d-c_1-3))=0$, does it follow that D is not deformable to smaller type?

(III) In the same situation, will $H^1(I_{Y(D)}(2d-c_1-3))\neq 0$ imply that D is deformable?

(IV) If Y is a finite subscheme of P^2 with $H^1(I_Y(n))\neq 0$, does it follow that Y is the limit of a family $Y_t\subseteq\mathbb{P}^2$ such that (Y_t,n) satisfy the Cayley-Bacharach condition for $t\neq 0$? (The CB condition is that any curve of degree n containing all but one point of Y contains Y.)

Clearly, (I) $\Rightarrow$ (II) and (IV) $\Rightarrow$ (III) (using 4.5). As evidence for (I), at least in the case that D can be deformed to a decomposable bundle, let us prove the following necessary condition for deformability.

(5.2) **Proposition.** *Let $D\in M(d;e)$ and suppose that $e^2-ec_1+c_2=0$. Then there are curves C_1 and C_2 of degrees $(d-e)$, $(d+e-c_1)$ respectively, C_2 not containing C_1, such that $Y(D)_{\mathrm{red}}\subseteq C_1\cap C_2$.*

Proof. One shows easily that there is a nonsingular affine curve T with distinguished point $0\in T$ parametrizing a family $\mathcal{E}$ with $\mathcal{E}_0\cong D$ and $\mathcal{E}_t\cong\mathcal{O}(e)\oplus\mathcal{O}(c_1-e)$

for $t \in T - \{0\}$. Let λ be a uniformizing parameter at 0. For each integer m the base change map $H^0(\mathscr{E}(m)) \bigotimes_T k(0) \to H^0(\mathscr{E}_0(m))$ is injective. Let $\sigma^i \in H^0(\mathscr{E}(m_i))$ $i = 1, 2$ be chosen so that: (i) $m_1 = -e$, $m_2 = e - c_1$, (ii) the map $\sigma^1 \oplus \sigma^2$: $\mathscr{O}_{\mathbb{P}^2_T}(e) \oplus \mathscr{O}_{\mathbb{P}^2_T}(c_1 - e) \to \mathscr{E}$ is an isomorphism over $T - \{0\}$, and (iii) σ^i is not divisible by λ. Let σ^i_0 be the image of σ^i in $H^0(\mathscr{E}_0(m_i))$. If $s \in H^0(\mathscr{E}_0(-d))$ is the minimal section, there are forms G^i of degree $d + m_i$ such that $\sigma^i_0 = G^i s$. Let $C_i = V(G^i)$. Since $V(\sigma^i)$ has no component of codimension 3, and furthermore has support in $V(\sigma^i_0) = Y(D) \cup C_i$, we conclude that $Y(D)_{\text{red}} \subseteq C_i$. The assertion that C_1 is not contained in C_2 follows from the injectivety of the base-change map for m_2.

(5.3) *Remark.* We have not ruled out the possibility of C_1 and C_2 having a common component. Under the additional hypothesis that the curve of minimal degree containing $Y(D)$ be irreducible, it does follow that $Y(D)$ is a complete intersection. In particular, if $d = e + 1$ (and $e^2 - ec_1 + c_2 = 0$) then we can answer (I) affirmatively. In the general case ($e^2 - ec_1 + c_2 > 0$) this method seems only to show that at least $(d^2 - dc_1 + c_2) - (e^2 - ec_1 + c_2)$ points of $Y(D)$ are contained in a curve of degree $(d - e)$. A slightly different approach would be to consider the case $m = (d - c_1)$ and consider the exact sequence

$$H^0(\mathscr{E}(d - c_1)) \to H^0(\mathscr{E}_0(d - c_1)) \to H^1(\mathscr{E}(d - c_1)).$$

All sections of $H^0(\mathscr{E}_0(d - c_1))$ are of the form Fs, where $F \in H^0(\mathscr{O}(2d - c_1))$. The strategy would be to find an F such that (i) Fs is the image of a section $\sigma \in H^0(\mathscr{E}(d - c_1))$, and (ii) $V(F) \cap Y(\mathscr{E}_0)$ is empty. It would follow that $V(\sigma) = X$ would have codimension 2, and that $Y(\mathscr{E}_0)$ would be the (flat) limit of X_t. A first-order study of the coboundary map seems to indicate that in "most cases" such an F does in fact exist, but there are technical difficulties involved. In the case of the families of (4.2), the F that occurs there will do the job.

(5.4) The question (II) is a weaker version of (I). A possible approach is to study the base of the versal deformation by actually computing Massey products in the algebra $\operatorname{Ext}^{\cdot}(D, D)$. The first of these, the cup product, shows an intimate relation to the condition of (II). (III) is the converse of (II), and might be approached in a similar way. Question (IV) is a stronger version of (III). The answer is yes if there is an irreducible curve of minimal degree containing Y. A possible approach would be to compute dimensions of the locally closed subsets of the Hilbert scheme with a given Hilbert function or minimal resolution.

(5.5) Another question of interest would be to describe intersections $M(d; e) \cap M(d; e^1)$, or more generally to determine, for a given $D \in M(d)$, the set of integers $e < d$ for which $D \in M(d; e)$. An extreme example is given by (4.10): If $Y(D)$ is on a line, then D is in $M(d; e)$ for all $e < d$ for which $M(e)$ is nonempty. On the other hand, if $c_1 = c_2 = 0$ and $d = 2$, then D is in $M(2; 0)$, but not in $M(2; 1)$ unless $Y(D)$ is on a line.

Acknowledgements. I would like to express thanks to C. Bănică for reawakening my interest in bundles on $\mathbb{P}^2$, and to O. A. Laudal for inspiration and patient listening.

References

1. Bănică, C.: Topologisch triviale holomorphe Vektorbündel auf $\mathbb{P}_n(C)$. Reine Angew. Math. (to appear)
2. Barth, W.: Moduli of vector bundles on the projective plane. Invent. Math. **42**, 63–91 (1977)
3. Ellingsrud, G., Strømme, S.A.: On the moduli space for stable rank-2 vector bundles on $\mathbb{P}^2$. Preprint, Oslo (1979)
4. Grothendieck, A.: Techniques de construction et théorèmes d'existence en géométrie algébrique IV: Les schémas de Hilbert. Séminaire Bourbaki 221 (1960/61)
5. Hartshorne, R.: Stable vector bundles of rank 2 on $\mathbb{P}^3$. Math. Ann. **238**, 229–280 (1978)
6. Hulek, K.: Stable rank-2 vector bundles on $\mathbb{P}^2$ with c_1 odd. Math. Ann. **242**, 241–266 (1979)
7. Kleiman, S.: Relative duality for quasi-coherent sheaves. Compositio Math. **41**, 39–60 (1980)
8. Laudal, O.A.: Formal moduli of algebraic structures. Lecture Notes in Mathematics, No. 754. Berlin, Heidelberg, New York: Springer 1979
9. Maruyama, M.: On a family of algebraic vector bundles. In: Number theory, algebraic geometry and commutative algebra, in honor of Y. Akizuki, pp. 95–146. Tokyo 1973
10. Mumford, D.: Lectures on curves on an algebraic surface. Annals of Math. Studies 59. Princeton: Princeton University Press 1966
11. Schafft, U.: Nichtseparierthеit instabiler Rang-2-Vektorbündel auf $\mathbb{P}^2$. Dissertation, Göttingen (1981)

11a. Schafft, U.: Nichtseparierthеit instabiler Rang-2-Vektorbündel auf $\mathbb{P}_2$. J. Reine Angew. Math. **338**, 136–143 (1983)

12. Schlessinger, M.: Functors of Artin rings. Trans. Am. Math. Soc. **130**, 208–222 (1968)

Received October 7, 1982; in revised form March 21, 1983

Math. Ann. 263, 399–418 (1983)

Mathematische Annalen
© Springer-Verlag 1983

A Formula for Interpolation and Division in $\mathbb{C}^n$

Bo Berndtsson

Departamento de Matematicas, CIEA del IPN, Apartado Postal 14-740 Mexico 14 DF and
Department of Mathematics, University of Gothenburg and CTH, S-41296 Göteborg, Sweden

Introduction

The aim of this paper is primarily to prove an explicit formula for interpolation in $\mathbb{C}^n$. If V is a regular analytic subvariety of a domain D in $\mathbb{C}^n$ and h is a function holomorphic on V, our formula gives a holomorphic function, H, in D whose restriction to V is equal to h. H is given as

$$H(z)=\int_V K(z,\zeta)h(\zeta), \tag{1}$$

where K is a certain integral kernel.

When $n=1$ and $D=\mathbb{C}$ there is a classical such formula of Lagrange. In this case V is a discrete set, $V=\{a_k\}$. Let f be an entire function which vanishes to order one in the points a_k. Then the series

$$H(z)=\sum_k \frac{f(z)b_k}{(z-a_k)f'(a_k)}$$

represents (if it converges) an entire function which assumes the value b_k in a_k.

There is a generalization of this formula to arbitrary n, known as Jacobi's formula ([2]). Here one still assumes that $V=\{a_k\}$, is a discrete set. We also make the assumption that one can find n functions $f_1,\dots,f_n$, such that $V=\{z;\ f_j(z)=0\,\forall_j\}$. Then there are holomorphic functions g_i^j $1\leqq i,j\leqq n$ of $2n$ variables (ζ,z) such that

$$f_j(z)-f_j(\zeta)=\sum_j g_i^j(\zeta,z)(z_i-\zeta_i). \tag{2}$$

Jacobi's formula is

$$H(z)=\sum_k b_k \det(g_j^i(a_k,z))/\det(\partial f_i/\partial\zeta_j(a_k)).$$

It follows immediately from (2) that

$$g_i^j(\zeta,\zeta)=\frac{\partial f_j}{\partial\zeta_i}(\zeta)$$

and also that the vectors $(g_1^j,\dots,g_n^j)(a_k,a_m)$ must be linearily dependent for $a_k\neq a_m$.

Hence $H(a_k)=b_k$. Note also that the formulas of Lagrange and Jacobi coincide for $n=1$.

Considering the case $\dim V>0$ we first note that in particular (1) has to hold for $z\in V$. Thus K must be a representing kernel for holomorphic functions on V. Conversely, if we can find a representing kernel for holomorphic functions on V, which has a holomorphic extension to z outside of V, we will have a kernel for interpolation. This is what we will do. Our formula is a generalization of the Henkin-Ramirez representation formula for holomorphic functions which we now recall.

Let
$$s(\zeta,z)=(s_1,\ldots,s_n):\bar{D}\times\bar{D}\to\mathbb{C}^n$$
be a function that satisfies essentially
$$\langle s,\zeta-z\rangle=\sum s_j(\zeta_j-z_j)\neq 0 \quad \text{for} \quad \zeta\neq z, \quad \text{and}$$
$$s(\zeta,\cdot) \quad \text{holomorphic for} \quad \zeta\in\partial D.$$

Denote also by s the differential form
$$s=\sum s_j d\zeta_j.$$

Then the Henkin-Ramirez formula is
$$h(z)=C_n\int_{\partial D} h(\zeta)\frac{s\wedge(\bar{\partial}s)^{n-1}}{\langle s,z-\zeta\rangle^n} \quad \text{for} \quad h\in H(\bar{D})$$
(see Sect. 1).

Now, let V be a regular analytic sub-variety in D of codimension p. Assume also that V is given as
$$V=\{z\in D: f_1(z)=\ldots f_p(z)=0\},$$
where, say, $f\in H(\bar{D})$. As before we find holomorphic functions g_i^j, $1\leqq i\leqq n$, $1\leqq j\leqq p$, such that
$$f_j(z)-f_j(\zeta)=\sum g_i^j(\zeta,z)(z_i-\zeta_i). \tag{3}$$

Define
$$\mu=C_p\frac{g^1\wedge\ldots g^p\wedge\overline{\partial f_1}\ldots\overline{\partial f_p}}{\|\partial f\|^2}dV.$$

Here $g^j=\sum g_i^j d\zeta_i$, $\|\partial f\|$ is the euclidean norm of the form $\partial f_1\wedge\ldots\partial f_p$ and dV is surface measure on V. Note that μ is a (p,p) current in ζ whose coefficients are measures supported on V, depending holomorphically on $z\in D$. Our generalization of the Henkin-Ramirez formula is
$$\forall z\in V\; h(z)=C_n\int_{V\cap\partial D} h(\zeta)\frac{s\wedge(\bar{\partial}s)^{n-1-p}\wedge\mu}{\langle s,\zeta-z\rangle^{n-p}} \quad \text{for} \quad h\in H(\bar{D}\cap V). \tag{4}$$

Evidently, the same formula can be used to define an extension of h outside V.

The method of proof is based on the results of [4], by M. Andersson and the author. There we constructed formulas with "weights". The present paper corresponds to a choice of weights which are singular on V, already indicated in

example (2) of [4]. These weights turn out to give certain residue terms, and the residues are precisely the kernels on V. Actually, one can also incorporate other weights in the kernels on V and get a general version of (4). This is useful in interpolation problems with growth conditions.

In Sect. 1 we give several instances of the interpolation formula, corresponding to the case D convex, strictly pseudoconvex or $D=\mathbb{C}^n$. In the latter case we show that our formula gives a condition for interpolation which is less restrictive than the one previously obtained by Berenstein and Taylor [2, 3], Yoshioka [17] and Nishimura [12]. Their methods were all based on Hörmanders L^2-estimates for the $\bar{\partial}$-operator. Most recently, this method has been further refined by Demailly [19], who obtained a condition similar but not identical to ours.

In the last part of Sect. 1 we carry out the construction of the current μ in the case when V is not a complete intersection. This means that we do not assume that V can be represented globally as the intersection of p hypersurfaces, only locally. Then one gets different formulas for μ, valid locally in ζ. It turns out, however, that one can solve (3) in such a manner that the different $\mu : s$ agree on overlaps. We do this using the Oka-Cartan theory for coherent sheaves and hence the results are not as explicit as before. Via the theorem that any Stein manifold can be imbedded in $\mathbb{C}^n$ we thus get kernels on general Stein manifolds. Such kernels have been constructed earlier with other methods by Henkin and Leiterer [8] and Hortmann [20].

We should also mention that different constructions of kernels on subvarieties have previously been given in Stout [16], Palm [13] and Fornaess [5].

We also point out that just like the classical Henkin-Ramirez formulas and the formulas devellopped in [4] we also get solution formulas for the $\bar{\partial}$-equation on subvarieties (for $(0, q)$ forms). However, the stress in this paper lies on the representation-interpolation formulas.

In the last section we show that with the kernels on V are also associated kernels which give explicit solutions to some division problems previously treated in Hörmander [10] and Skoda [15]. Our results cover essentially the results in [10], but are somewhat weaker than those in [15]. Nevertheless, we feel they might have some interest because they are explicit and not restricted to L^2-norms.

A few words about notation: We often use the same letter to denote different constants. This occurs even if the letter has subscripts. Thus the subscripts do not serve to name the constant, but only to say what parameters it may depend on.

We frequently identify an n-tuple of functions $t=(t_1, \ldots, t_n)$ with the form $t=\sum t_j d\zeta_j$. In booth cases the notation

$$\langle t, a\rangle$$

means

$$\sum t_j a_j .$$

1. Interpolation

We begin by recalling the main result of [4]. Let D be a domain in $\mathbb{C}^n$ with C^2-boundary. Let

$$s=(s_1, \ldots, s_n) : \bar{D} \times \bar{D} \to \mathbb{C}^n$$

be a C^1 function satisfying

$$|s(\zeta,z)|\leqq C|\zeta-z| \quad \text{and} \quad |\langle s,\zeta-z\rangle|\geqq c|\zeta-z|^2 \tag{1}$$

uniformly for $\zeta\in\bar{D}$ and z in any compact subset of D. Here, for ξ and η in $\mathbb{C}^n$

$$\langle\xi,\eta\rangle=\sum_1^n \xi_j\eta_j.$$

Then let

$$Q:\bar{D}\times\bar{D}\to\mathbb{C}^n$$

be a C^1 function such that for ζ fixed in $\bar{D}$ $Q(\zeta,\cdot)$ is holomorphic in D. With s and Q we associate the $(1,0)$-forms

$$\sum s_j d\zeta_j \quad \text{and} \quad \sum Q_j d\zeta_j$$

which we also denote by s and Q respectively.

Let G be a function (of one variable) holomorphic in a simply connected domain containing the image of $\bar{D}\times\bar{D}$ under the map

$$(\zeta,z)\to\langle Q,z-\zeta\rangle+1$$

and assume $G(1)=1$.

Now we can define

$$K=\sum_{k=0}^{n-1}\frac{(n-1)!}{k!}G^{(k)}(\langle Q,z-\zeta\rangle+1)\frac{s\wedge(\bar{\partial}s)^{n-1-k}\wedge(\bar{\partial}Q)^k}{\langle s,\zeta-z\rangle^{n-k}} \tag{2}$$

and

$$P=-\frac{1}{n}G^{(n)}(\langle Q,z-\zeta\rangle+1)(\bar{\partial}Q)^n \tag{3}$$

(here $G^{(k)}$ is the $k:th$ derivative of G).

Finally, we denote by K_q the component of K which is of bidegree $(0,q)$ in z and $(n,n-q-1)$ in ζ, and P_q the component of P that is of bidegree $(0,q)$ in z. Since Q is holomorphic in z we have in fact $P_q=0$ for $q>0$, so $P=P_0$.

Theorem 0 [4]. *Let u be a $(0,q)$ form in $C^1(\bar{D})$. Then for $z\in D$*

$$u(z)=C_{q,n}\left\{\int_{\partial D}u\wedge K_q+(-1)^{q+1}\left(\int_D\bar{\partial}u\wedge K_q-\bar{\partial}_z\int_D u\wedge K_{q-1}\right)\right\} \tag{4}$$

if $q>0$ and

$$u(z)=C_n\left(\int_{\partial D}uK_0-\int_D\bar{\partial}u\wedge K_0-\int_D uP_0\right) \tag{5}$$

for $q=0$ (integrations are with respect to ζ).

This is a generalization of the classical "Koppelman's formula". The classical case corresponds to $Q\equiv 0$ (or $G\equiv 1$). Our present aim is to find an analogue of this formula where D is replaced by a subvariety of D. Therefore, let $f_1,\ldots,f_p$ be p functions that are holomorphic in D and (say) in $C^1(\bar{D})$, and denote by V the set of common zeros, i.e.

$$V=\{\zeta\in D, f_1(\zeta)=\ldots f_p(\zeta)=0\}$$

We also make the assumption that $\partial f_1 \wedge \ldots \partial f_p \neq 0$ on $V \cap \bar{D}$, so that V is a regular subvariety of codimension p in D. Next, choose functions g_j^k for $k=1, \ldots, p$, $j=1, \ldots, n$ that are holomorphic in $D \times D$ and satisfy

$$f_k(z) - f_k(\zeta) = \sum g_j^k(\zeta, z)(z_j - \zeta_j). \tag{6}$$

We assume in addition that the $g_j^k : s$ are continuous up to the boundary. We will not go into any details about how to solve the problem (6), since primarily we only want to show how the construction of the kernels is carried out in principle. Just observe that in case D is a convex domain we may take

$$g_j^k(\zeta, z) = \int_0^1 \frac{\partial f_k}{\partial \zeta_j}(\zeta + t(z-\zeta))\, dt \tag{7}$$

since

$$f_k(z) - f_k(\zeta) = \int_0^1 \frac{d}{dt} f_k(\zeta + t(z-\zeta))\, dt.$$

Set

$$g^k = \sum_1^n g_j^k d\zeta_j.$$

We can now make the definition

$$\mu = \frac{g^1 \wedge \ldots g^p \wedge \overline{\partial f_1} \wedge \ldots \overline{\partial f_p}}{\|\partial f\|^2} dV(\zeta).$$

Here dV is surface measure on V and $\|\partial f\|$ is the euclidean norm of $\partial f = \partial f_1 \wedge \ldots \partial f_p$. μ is a current (in ζ) whose coefficients depend holomorphically on z. Before stating our result we need one more concept. We say that V intersects ∂D transversally if

$$\partial f_1 \wedge \ldots \partial f_p \wedge \partial \varrho \neq 0 \quad \text{on} \quad V \cap \partial D$$

where ϱ is a defining function for ∂D, i.e. a realvalued function that equals zero on ∂D, and satisfies $\partial \varrho \neq 0$ on ∂D.

If V intersects ∂D transversally integrals like

$$\int_{\zeta \in \partial D} \alpha \wedge \mu$$

where α is a $(2(n-p)-1)$-form, are well defined. This will be clear from the proofs later on where μ appears as the limit of smooth forms. Now, let G, s and Q be as above and set

$$K = \sum_{k=0}^{n-p-1} \frac{(n-1)!}{p!\,k!} G^{(k)}(\langle Q, z-\zeta\rangle + 1) \frac{s \wedge (\bar{\partial} s)^{n-p-1-k} \wedge (\bar{\partial} Q)^k \wedge \mu}{\langle s, \zeta - z\rangle^{n-p-k}} \tag{8}$$

and

$$P = -\frac{(n-1)!}{p!(n-p)!} G^{(n-p)}(\langle Q, z-\zeta\rangle + 1)(\bar{\partial} Q)^{n-p} \wedge \mu, \tag{9}$$

and let K_q and P_q mean the same as before. We can then formulate our basic result:

Theorem 1. *Let u be a $(0,q)$ form on V which has a C^1 extension to $\bar{V}\cap\bar{D}$. Then for $z\in V$*

$$u(z)=C_{q,n}\left\{\int_{\partial D} u\wedge K_q+(-1)^{q+1}\left(\int_D \bar{\partial}u\wedge K_q-\bar{\partial}_z\int_D u\wedge K_q\right)\right\} \tag{10}$$

if $q>0$ and

$$u(z)=c_n\left(\int_{\partial D} uK_0-\int_D \bar{\partial}u\wedge K_0-\int_D uP_0\right) \tag{11}$$

for $q=0$.

We postpone the proof of theorem 1 to the end of this section. Later on we will also discuss the case when V is not a complete intersection.

A few remarks are in order. First, to define the integrals we need to take a C^1 extension of u, and since μ, and hence K and P, is supported on V, the result does not depend on the choice of extension. This also holds for the integrals containing $\bar{\partial}u$ since μ contains all the "normal" covectors $\overline{\partial f_1}\ldots\overline{\partial f_p}$. Moreover, the operator $\bar{\partial}_z$ in (10) is the $\bar{\partial}$-operator on V.

If D is strictly pseudeconvex (with C^2 boundary) we can choose s (or Q) in such a way that, for $\zeta\in\partial D$, $K_q=0$ for $q>0$, and K_0 is holomorphic in $z\in D$. This is precisely as in the classical case (see [6], [14] or [4]). Thus theorem 1 gives a solution formula for $\bar{\partial}$ on V and also a representation formula with holomorphic kernel for holomorphic functions on V.

Note also that the kernels are well defined for all $z\in D$. The formulas (10) and (11) define extensions of forms defined on V to all of D. Thus, if $\bar{\partial}u=0$ on V, (10) gives an extension which is closed (even exact) in $D\,(q>0)$, and (11) gives an extension of a holomorphic function on V to a holomorphic function in D. We shall now study this more closely in a few concrete situations. We treat only the case $q=0$, i.e. u is a function.

First, let $p=n$. Then $K=0$ and $V=\{a_m\}$ is a finite subset of D so the assumption that u be holomorphic on V is vacuous. Say $u(a_m)=b_m$. Our extension, U, is given as

$$U(z)=C_n\int_D uP_0\,.$$

Taking $G\equiv 1$ we get

$$P=C\mu=C\sum_m \frac{\det(g_j^k(a_m,z))}{\det(\partial f_k/\partial\zeta_j(a_m))}\delta a_m\,,$$

where δa_m is dirac-measure in a_m. Hence

$$U(z)=C_n\sum_m b_m\frac{\det(g_j^k(a_m,z))}{\det(\partial f_k/\partial\zeta_j(a_m))}\,,$$

so we have obtained Jacobi's formula (see Introduction).

Next we take V to be a submanifold of positive dimension in a C^2 convex set D. Say

$$D=\{\zeta\in\mathbb{C}^n,\varrho(\zeta)<0\}\,,$$

where ϱ is a convex C^2 function such that $\nabla\varrho \neq 0$ on ∂D. Then, as in [4], take

$$Q = \frac{\partial\varrho}{\varrho - \varepsilon} \quad \text{and} \quad G(\alpha) = \alpha^{-N}, \quad N > 0.$$

As in [4], we can let $\varepsilon \to 0$ and get that

$$P = C_{n,p} \frac{\varrho^{N+n-p}}{(\langle \partial\varrho(\zeta), z-\zeta\rangle + \varrho(\zeta))^{N+n-p}} \left(\partial\bar{\partial}\log - \frac{1}{\varrho}\right)^{n-p} \wedge \mu$$

is also an admissible kernel in Theorem 1. Since the kernel K now vanishes on ∂D, for any choice of section s, (11) gives the following representation-interpolation formula:

Theorem 2. *Suppose h is holomorphic on V and in $C^1(\bar{V})$. Then the formula*

$$H(z) = C_{n,p} \int_V h(\zeta) \frac{\varrho^{N+n-p}}{(\langle \partial\varrho, z-\zeta\rangle + \varrho)^{N+n-p}} \left(\partial\bar{\partial}\log - \frac{1}{\varrho}\right)^{n-p} \wedge \mu \tag{12}$$

defines a holomorphic function in D which coincides with h on V.

Actually, one can see by exhausting D from inside with smaller domains that it is enough to assume that h does not grow more rapidly near ∂D, than that the integral converges.

A similar formula can be given for a strictly pseudoconvex domain with C^2-boundary. Let $D = \{\zeta, \varrho(\zeta) < 0\}$ where ϱ is a strictly pluri-subharmonic defining function. It is shown in [5] that there is a function $H(\zeta, z)$ defined on $\bar{D} \times \bar{D}$, of class C^2, such that

$$H(\zeta, \cdot) \text{ is holomorphic in } D$$
$$\operatorname{Re} H(\zeta, z) \leqq \varrho(z) - \varrho(\zeta) - \delta|\zeta - z|^2 \quad \text{for some} \quad \delta > 0 \text{ and}$$
$$H(\zeta, z) = \sum h_j(\zeta, z)(z_j - \zeta_j) \text{ with}$$
$$h_j \in C^1(\bar{D} \times \bar{D}),\ h_j(\zeta, \cdot) \text{ holomorphic in } D.$$

We can then let

$$Q = \frac{\sum h_j d\zeta_j}{\varrho} \quad \text{and} \quad G(\alpha) = \alpha^{-N}, \quad N > 0,$$

and get the kernel

$$P = C_{n,p} \frac{\varrho^{N+n-p}}{(H(\zeta, z) + \varrho(\zeta))^{N+n-p}} (\bar{\partial} Q)^{n-p} \wedge \mu.$$

Our next example concerns the case $D = \mathbb{C}^n$. Let $\varphi \geqq 0$ be a C^2 convex function on $\mathbb{C}^n$. For simplicity we assume φ satisfies two technical conditions

(i) for some C $\max_{\frac{R}{2} \leqq |\zeta| \leqq R} \varphi \leqq C \max_{R \leqq |\zeta| \leqq 2R} \varphi$ for all $R > 0$,

and

(ii) for some C $\|\partial\bar{\partial}\varphi\| \leqq C \exp C\varphi$.

The main example is $\varphi(\zeta)=|\zeta|^a$ $a\geqq 1$, but for instance the support function of any sufficiently regular convex set will also do. Suppose now that $f_1,\dots,f_p$ are entire and such that

(*) $$|f_j|\leqq C\exp C\varphi \quad \text{for some} \quad C$$

and that $\partial f_1\wedge\dots\partial f_p\neq 0$ on $V=\{\zeta, f_1(\zeta)=\dots f_p(\zeta)=0\}$.

Note that (i) implies that the ring of entire functions satisfying (*) is closed under differentiation. Moreover, if we define g_j^k by (7), we get

$$|g_j^k(\zeta,z)|\leqq C\exp C(\varphi(\zeta)+\varphi(z)).$$

We can now prove

Theorem 3. *Let h be a holomorphic function on V that satisfies the condition*

$$\int_V |h(\zeta)|\exp -B\varphi(\zeta)\frac{dV(\zeta)}{\|\partial f\|}<\infty \quad \textit{for some} \quad B. \tag{13}$$

Then

$$H(z)=C_{n,p}\int_V h(\zeta)(\exp A\langle\partial\varphi, z-\zeta\rangle)(A\partial\bar\partial\varphi)^{n-p}\wedge\mu \tag{14}$$

defines (for A large enough) an entire function whose restriction to V is equal to h and moreover H satisfies the estimate

$$|H(z)|\leqq C_1\exp C_2\varphi(z) \quad \textit{for some} \quad C_j. \tag{15}$$

Proof. Set

$$s=(\bar\zeta-\bar z),\ Q=A\partial\varphi\ (A>0), \quad \text{and} \quad G=\exp\alpha.$$

From the inequality for convex functions

$$\varphi(z)-\varphi(\zeta)\geqq 2\,\mathrm{Re}\langle\partial\varphi(\zeta), z-\zeta\rangle$$

we obtain

$$|\exp\langle Q, z-\zeta\rangle|\leqq\exp A\varphi(z)/2\exp -A\varphi(\zeta)/2.$$

With g_j^k chosen as above we put

$$P=\exp\langle Q, z-\zeta\rangle(A\partial\bar\partial\varphi)^{n-p}\wedge\mu,$$

and have the estimate

$$\|P\|\leqq C_1\exp A/2(\varphi(z)-\varphi(\zeta))\exp C_2(\varphi(z)+\varphi(\zeta))\frac{dV(\zeta)}{\|\partial f\|},$$

where C_2 depends only on the functions f_j.

Then we let K be defined by (8) with $s=\bar\zeta-\bar z$. Take a C^∞ function χ with compact support in $\{\zeta;|\zeta|<2\}$, such that $\chi(\zeta)=1$ for $|\zeta|\leqq 1$, and set $\chi_R(\zeta)=\chi(\zeta/R)$. Apply (11) with $u=\chi_R h$ and $D=\{\zeta, |\zeta|<2R\}$. This yields

$$h(z)=C_{n,p}\left\{\int_V h\bar\partial\chi_R\wedge K+\int_V h\chi_R P\right\} \tag{16}$$

for $z\in V$, $|z|<R$.

K can be estimated by

$$\|K\| \leqq C_1 \exp A/2(\varphi(z)-\varphi(\zeta)) \exp C_2(\varphi(z)+\varphi(\zeta)) \frac{dV}{\|\partial f\| \, |\zeta - z|^{2(n-p)-1}}$$

Hence, if $A \geqq 2(C_2 + B)$ we can let $R \to \infty$ in (16), and the first term on the right hand side in (16) then goes to zero. This shows that $H(z)$ (as defined in (14)) is equal to $h(z)$ for $z \in V$, and the theorem is proved. We note that the constant C_2 in (14) can be taken equal to B plus a constant that depends only on the variety.

In particular theorem 3 shows that if (13)

$$\int_V \exp - B\varphi \frac{dV}{\|\partial f\|} < \infty \quad \text{for some} \quad B, \tag{17}$$

then V is an interpolating variety in the sense of Berenstein-Taylor [3]. This means that every holomorphic h on V satisfying a condition

$$|h(\zeta)| \leqq C \exp C\varphi$$

can be extended to an entire function satisfying the same bound (possibly with another constant). In [3] it was shown that V is interpolating if

$$\|\partial f\| \geqq C \exp - C\varphi \quad \text{on} \quad V. \tag{18}$$

(This was proved in a more general setting with φ plurisubharmonic.) Results in the same genre have also been obtained by Nishimura [12] and Yoshioka [17]. A counterexample by Nishimura shows that condition (18) is not necessary for V to be interpolating. There is also a recent paper by Demailly [19], who proves, among other things, that

$$\int_V \exp - B\varphi \frac{dV}{\|\partial f\|^2} < \infty \quad \text{for some} \quad B \tag{19}$$

is sufficient for V to be an interpolating variety.

In case $p=1$ it is easy to see that condition (18) implies (17) since one can estimate the surface measure of V by Jensens formula. For $p>1$ it is known to be impossible to estimate dV by the growth of the $f_j : s$ only (see [18]). Nevertheless it seems quite probable that (18) implies (17) in general although I have not proved it.

We shall now use Nishimura's counterexample to show that (17) in general does not imply (18) or (19).

Example. *Let $\varphi = |\zeta|^a$, $a \geqq 1$. There is a hypersurface in $\mathbb{C}^2$ defined by a function of order zero, which satisfies* (17) *but not* (19) *(or* (18)*).*

Proof. Following Nishimura [12], we choose a sequence $a_k \in \mathbb{R}$ tending to $+\infty$ rapidly, and a sequence of ε_k tending to zero rapidly. Then the canonical product

$$p(\xi) = \Pi \left(1 - \frac{\xi}{a_k}\right)\left(1 - \frac{\xi}{a_k + \varepsilon_k}\right)$$

is an entire function in $\mathbb{C}$ of order zero, having simple zeros in the points $\xi = a_k$, $\xi = a_k + \varepsilon_k$. Put

$$f(\zeta_1, \zeta_2) = \zeta_2^2 - p(\zeta_1).$$

Then

$$\partial f = -p'(\zeta_1)d\zeta_1 + 2\zeta_2 d\zeta_2$$

which is never zero on $V = \{\zeta, f(\zeta) = 0\}$.

If $|\xi - a_m| \leqq 1$ and a_k tends to $+\infty$ sufficiently rapidly we have

$$|p(\xi)| \leqq C \exp|\xi| \, (|\xi - a_m| \, |\xi - a_m - \varepsilon_m|)$$

and

$$|p'(\xi)| \leqq C \exp|\xi| \, (|\xi - a_m| + |\xi - a_m - \varepsilon|).$$

On V

$$\|\partial f\|^2 \approx |p'(\zeta_1)|^2 + |\zeta_2|^2 = |p'(\zeta_1)|^2 + |p(\zeta_1)|.$$

Surface measure on V can be decomposed

$$dV = \tfrac{i}{2} d\zeta_1 \wedge d\bar{\zeta}_1 + \tfrac{i}{2} d\zeta_2 \wedge d\bar{\zeta}_2 = dV_1 + dV_2,$$

and it is easy to see that by choosing ε_m small enough we will make the integral

$$\sum_m \int_{|\zeta_1 - a_m| \leqq 1} \exp - B|\zeta|^a \frac{dV_1}{\|\partial f\|^2}$$

divergent for all B so that (18) does not hold, and of course neither does (17).

On the other hand

$$\int_V \exp - B|\zeta|^a \frac{dV_1}{\|\partial f\|} \leqq 2 \int_V \exp - B|\zeta_1|^a \frac{dV_1}{\sqrt{|p(\zeta_1)|}}$$

(for every value of ζ_1 there are two values of ζ_2 such that $(\zeta_1, \zeta_2) \in V$), and the later integral is easily seen to converge. It remains to check that the integral with respect to dV_2 converges. This we do by observing that $\|\partial f\| \geqq |\zeta_2|$, and then estimating the number of points on V that lie above a fixed value of ζ_2 with Jensens formula but we omit the details.

Finally we briefly consider the case of polynomial growth in $\mathbb{C}^n$. Then we assume that V is algebraic and the functions $f_1, \ldots, f_p$ are polynomials. Set

$$Q = \partial \log(1 + |\zeta|^2) \quad \text{and} \quad G(\alpha) = \alpha^N, \qquad N > 0.$$

Then if h is a function of polynomial growth on V we get the representation formula

$$h(z) = C_{n,p} \int_V h(\zeta) \left(\frac{1 + \bar{\zeta} \cdot z}{1 + |\zeta|^2} \right)^N (\partial \bar{\partial} \log(1 + |\zeta|^2))^{n-p} \wedge \mu \tag{20}$$

which will converge for N large enough. This gives the well known result that h is the restriction to V of a polynomial.

In the last two examples we have had the same kind of growth condition on the functions that define V as on the function we want to extend. This is clearly not necessary. Actually if the functions that define V grow slowly compared to the function h, formulas (15) and (20) give very precise estimates of the extension.

We now turn to the proof of Theorem 1. The first step is to see that the kernels K and P in (2) and (3) can be generalized somewhat. Namely, take two choices of Q, Q^1 and Q^2 both holomorphic in the second variable, z. Also, let G_1 and G_2 be two functions holomorphic in a neighborhood of the images of $\bar{D}\times\bar{D}$ under the maps

$$(\zeta, z)\to\langle Q^j, z-\zeta\rangle+1$$

respectively, and such that $G_j(1)=1$. Set

$$K=\sum_{\substack{(\alpha_0,\alpha_1,\alpha_2)\\ \alpha_0+\alpha_1+\alpha_2=n-1}}\frac{(n-1!}{\alpha_1!\,\alpha_2!}G_1^{(\alpha_1)}G_2^{(\alpha_2)}\frac{s\wedge(\bar\partial s)^{\alpha_0}\wedge(\bar\partial Q^1)^{\alpha_1}\wedge(\bar\partial Q^2)^{\alpha_2}}{\langle s,\zeta-z\rangle^{\alpha_0+1}} \tag{21}$$

$$P=-\sum_{\substack{\alpha=(\alpha_1,\alpha_2)\\ \alpha_1+\alpha_2=n}}\frac{(n-1)!}{\alpha_1!\,\alpha_2!}G_1^{(\alpha_1)}G_2^{(\alpha_2)}(\bar\partial Q^1)^{\alpha_1}\wedge(\bar\partial Q^2)^{\alpha_2} \tag{22}$$

(here $G_j^{(\alpha_j)}$ means $G_j^{(\alpha_j)}(\langle Q^j, z-\zeta\rangle+1)$).

Then formulas (4) and (5) will hold with this new definition of K and P. The proof of this is entirely analogous to [4], so we will only sketch it. Put

$$Q=\lambda_1 Q^1+\lambda_2 Q^2$$

Use this definition of Q in (2) and (3) with $G(\alpha)=\exp\alpha$. This gives (21) and (22) with $G_j(\alpha)=\exp\lambda_j\alpha$. Then take two functions $g_1, g_2\in C_0^\infty(\mathbb{R})$ and take averages with respect to the measure

$$e^{-(\lambda_1+\lambda_2)}g_1(\lambda_1)g_2(\lambda_2)d\lambda_1 d\lambda_1\,.$$

Now we get (21) and (22) with G_j equal to the Laplace transform of g_j, and approximating arbitrary functions with Laplace transform we get (21) and (22) in general.

With f_j and g_j^k as before we take Q^1 arbitrary and set

$$Q_\varepsilon^2=\sum_k\overline{f_k(\zeta)}g^k\Big/\sum_k|f_k(\zeta)|^2+\varepsilon \tag{23}$$

for $\varepsilon>0$ (remember $g^k=\sum g_j^k d\zeta_j$). From now on we use the notation

$$|f|^2=\sum|f_k(\zeta)|^2,\ \overline{f(\zeta)}f(z)=\sum\overline{f_k(\zeta)}f_k(z)\,.$$

Then

$$\langle Q_\varepsilon^2, z-\zeta\rangle+1=\frac{\overline{f(\zeta)}f(z)+\varepsilon}{|f|^2+\varepsilon}$$

Furthermore

$$\bar\partial Q_\varepsilon^2=\frac{\sum\overline{\partial f_k}\wedge g^k}{|f|^2+\varepsilon}-\frac{\bar\partial|f|^2\wedge\sum\overline{f_k}g^k}{(|f|^2+\varepsilon)^2}$$

and

$$(\bar\partial Q^2)^m = \frac{(\sum \overline{\partial f_k}\wedge g^k)^m}{(|f|^2+\varepsilon)^m} - m\frac{\bar\partial|f|^2\wedge\sum \overline{f_k}g^k\wedge(\sum\overline{\partial f}\wedge g^k)^{m-1}}{(|f|^2+\varepsilon)^{m+1}}.$$

In the last computation we used the fact that forms of even degree commute and that $(\partial|f|^2)^2=0$.

Let us introduce the notation

$$\partial f=\partial f_1\wedge\ldots\partial f_p \quad\text{and}\quad g=g^1\wedge\ldots g^p.$$

Recall that the standard metric on $\mathbb{C}^n$ induces a norm on (p,q) forms as follows:

If $\bar\psi=\sum\psi_{IJ}dz_I\wedge d\overline{z_J}$ then

$$\|\psi\|^2=\sum|\psi_{IJ}|^2.$$

Here I and J are increasing multiindices and $dz_I=dz_{i_1}\wedge\ldots dz_{i_p}$ if $I=(i_1,\ldots,i_p)$.

This norm satisfies the inequality

$$\|\psi\wedge\Phi\|\leqq C\|\psi\|\,\|\Phi\|$$

Let Θ_V be the current of integration on V (see [11]). Then

$$\Theta_\varepsilon=(\partial\bar\partial\log(|f|^2+\varepsilon))^p=C_p\frac{\varepsilon\partial f\wedge\overline{\partial f}}{(|f|^2+\varepsilon)^{p+1}}\to\Theta_V \tag{24}$$

as $\varepsilon\to0$ (modulo a multiplicative constant). Hence the trace measure of Θ_ε

$$\tau(\Theta_\varepsilon)=\frac{\varepsilon}{(|f|^2+\varepsilon)^{p+1}}\|\partial f\|^2$$

tends to the trace measure of Θ, which is dV, surface measure on V. This yields

$$\frac{\varepsilon}{(|f|^2+\varepsilon)^{p+1}}\to\frac{dV}{\|\partial f\|^2} \tag{25}$$

Since $\bar\partial Q_\varepsilon^2$ is of the form

$$\bar\partial Q_\varepsilon^2=\sum_1^p\overline{\partial f_k}\wedge\alpha_k$$

we have $(\bar\partial Q_\varepsilon^2)^m=0$ if $m>p$.

Moreover

$$(\bar\partial Q_\varepsilon^2)^p=C_p\left(\frac{\overline{\partial f}\wedge g}{(|f|^2+\varepsilon)^p}-\frac{|f|^2\overline{\partial f}\wedge g}{(|f|^2+\varepsilon)^{p+1}}\right)=C_p\frac{\varepsilon\overline{\partial f}\wedge g}{(|f|^2+\varepsilon)^{p+1}},$$

so (24) implies

Lemma 4. *If $\partial f\neq0$ in D, then*

$$\lim_{\varepsilon\to0}(\bar\partial Q_\varepsilon^2)^p=a_pg\wedge\frac{\overline{\partial f}}{\|\partial f\|^2}dV=\mu,$$

a current in ζ with measure coefficients supported on V, depending holomorphically on z.

Observe that

$$\left\| g \wedge \frac{\overline{\partial f}}{\|\partial f\|^2} \right\| \leqq C \frac{\|g\|}{\|\partial f\|}.$$

For $p=1$

$$\frac{\overline{\partial f}}{\|\partial f\|^2} dV = \bar{\partial} \frac{1}{f}$$

as follows from

$$\frac{1}{f} = \lim_{\varepsilon \to 0} \frac{\bar{f}}{|f|^2+\varepsilon}, \quad \bar{\partial} \frac{\bar{f}}{|f|^2+\varepsilon} = \frac{\varepsilon \bar{\partial} f}{(|f|^2+\varepsilon)^2}$$

together with (25). The importance of this current for interpolation problems has been showed by Amar [1]. Actually, in the case $p=1$, one could probably obtain results equivalent to ours using Amars method together with the kernels from [4].

Take G_1 arbitrary and $G_2(\alpha)=\alpha^p$. From (21) and (22) we get

$$K = \sum_{|\alpha|=n-1} \frac{(n-1)!}{\alpha_1!\alpha_2!} \frac{p!}{(p-\alpha_2)!} G_1^{(\alpha_1)} \left(\frac{\overline{f(\zeta)} f(z)+\varepsilon}{|f|^2+\varepsilon} \right)^{p-\alpha_2} \frac{s \wedge (\bar{\partial} s)^{\alpha_0} \wedge (\bar{\partial} Q^1)^{\alpha_1} \wedge (\bar{\partial} Q_\varepsilon^2)^{\alpha_2}}{\langle s, \zeta - z \rangle^{\alpha_0+1}} \tag{26}$$

and

$$P = \sum_{|\alpha|=n} \frac{(n-1)!}{\alpha_1!\alpha_2!} \frac{p!}{(p-\alpha_2)!} G_2^{(\alpha_1)} \left(\frac{\bar{f}(\zeta) f(z)+\varepsilon}{|f|^2+\varepsilon} \right)^{(p-\alpha_2)} (\bar{\partial} Q^1)^{\alpha_1} \wedge (\bar{\partial} Q^2)^{\alpha_2}. \tag{27}$$

We want to let ε tend to zero. From previous computations we get the estimates

$$\|(\bar{\partial} Q_\varepsilon^2)^k\| \leqq C \left(\frac{|f|^2}{(|f|^2+\varepsilon)^{k+1}} + \frac{1}{(|f|^2+\varepsilon)^k} \right)$$

$$\cdot \frac{\|s\|}{|\langle s, \zeta - z \rangle|^{n-k}} \leqq C \frac{1}{|\zeta - z|^{2n-2k+1}}$$

and clearly

$$|\overline{f(\zeta)} f(z)| \leqq C|f|(|f|+|\zeta - z|).$$

From these estimates it follows that the terms with $\alpha_2 < p$ in (26) and (27) are for fixed $z \in D$, bounded by a function integrable over D independently of ε. Similarily, for $z \in D$ fixed, they are bounded by a function integrable over ∂D if V intersects ∂D transversally.

Moreover, the term $\alpha_2 = p$, will converge in D to a form with measure coefficients in D, as we have seen in Lemma 4. It is easy to see that, under the assumption of transversality, this term will also converge to a form with measure coefficients on the boundary, but we don't compute this form explicitly since in most applications the integral over the boundary will vanish for other reasons.

Now take z on V, so that $f_k(z)$ vanishes for $k=1, \ldots, p$. When ε tends to zero all the terms in (26) and (27) with $\alpha_2 < p$ will tend to zero, so by Lebesgue's theorem on

dominated convergence, we get "Koppelman's formulas" (10) and (11) with K and P given by (8) and (9) respectively. The only thing thing that could possibly cause trouble is the term in (10) that involves the operator $\bar{\partial}_z$. But since we have convergence pointwise for $z \in V$ and moreover dominated convergence when z belongs to a compact subset of V we get the result by the continuity of differentiation in distribution theory (remember $\bar{\partial}_z$ is the $\bar{\partial}$-operator on V). This completes the proof of theorem 1.

We shall now discuss the case when V is not a complete intersection in D. For convenience we assume that V extends to a variety in a neighbourhood, $\tilde{D}$, of $\bar{D}$. The problem we have is to construct the current μ. We shall see that this construction can be made locally, and moreover in such a way that the results patch together.

Since the ideal of functions that vanish on V is generated by global sections we can, for each point in $\tilde{D}$, find a p-tuple of functions $(f_1, \ldots, f_p)$ in $H(\tilde{D})$ such that $f_j = 0$ on V and $\partial f_1 \wedge \ldots \partial f_p \neq 0$ in a neighbourhood of the given point. Let $\{U_\alpha\}$ be a finite collection of such neighbourhoods that cover $V \cap \bar{D}$ and denote the corresponding p-tuples $f^\alpha = (f_1^\alpha, \ldots, f_p^\alpha)$. Since on $U_\alpha \cap U_\beta$ booth p-tuples generate the ideal of functions vanishing on V, there is a holomorphic $p \times p$ matrix, $A_{\alpha\beta}$, on $U_\alpha \cap U_\beta$ such that

$$A_{\alpha\beta} f^\beta = f^\alpha \text{ i.e. } f_k^\alpha = \sum_j (A_{\alpha\beta})_{kj} f_j^\beta .$$

This is clear locally and since we can assume U_α is convex we even get it globally on $U_\alpha \cap U_\beta$.

For each α choose functions $(g^\alpha)_j^k$ holomorphic on $\tilde{D} \times \tilde{D}$ such that

$$f_k^\alpha(z) - f_k^\alpha(\zeta) = \sum_j (g^\alpha)_j^k(\zeta, z)(z_j - \zeta_j) .$$

Let

$$(g^\alpha)^k = \sum (g^\alpha)_j^k d\zeta_j \quad \text{and} \quad g^\alpha = (g^\alpha)^1 \wedge \ldots (g^\alpha)^p . \tag{28}$$

For each α put

$$\mu^\alpha = \frac{g^\alpha \wedge \overline{\partial f^\alpha}}{\|\partial f^\alpha\|^2} dV .$$

Then μ^α is a current defined for $\zeta \in U_\alpha$, with coefficients depending holomorphically on $z \in D$. We want to arrange things so that the $\mu^\alpha : s$ agree on overlaps. Let the matrices $A_{\alpha\beta}$ act on $(1,0)$-forms componentwise. Then we get for $\zeta \in U_\alpha \cap U_\beta \cap V$

$$A_{\alpha\beta} \begin{pmatrix} \partial f_1^\beta \\ \vdots \\ \partial f_p^\beta \end{pmatrix} = \begin{pmatrix} \partial f_1^\alpha \\ \vdots \\ \partial f_p^\alpha \end{pmatrix} \tag{29}$$

so that

$$\partial f^\alpha = \det A_{\alpha\beta} \partial f^\beta .$$

If we can arrange things so that

$$A\begin{pmatrix}(g^\beta)^1\\ \vdots\\ (g^\beta)^p\end{pmatrix}=\begin{pmatrix}(g^\alpha)^1\\ \vdots\\ (g^\alpha)^p\end{pmatrix} \quad \text{for} \quad \zeta\in U_\alpha\cap U_\beta\cap V \tag{30}$$

we get

$$\mu^\alpha=\frac{\det A_{\alpha\beta}\det\overline{A_{\alpha\beta}}}{|\det A_{\alpha\beta}|^2}\frac{g^\beta\wedge\overline{\partial f^\beta}}{\|\partial f^\beta\|^2}dV=\mu^\beta$$

so the $\mu^\alpha : s$ define a global current.

Because of (29) the matrix $A_{\alpha\beta}$ is uniquely determined on $U_\alpha\cap U_\beta\cap V$ (since booth $\{\partial f_j^\alpha\}$ and $\{\partial f_j^\beta\}$ form basisis for the conormal space of V). This also implies

$$A_{\alpha\beta}=A_{\beta\alpha}^{-1} \quad \text{and} \quad A_{\alpha\beta}A_{\beta\gamma}A_{\gamma\alpha}=1 \quad \text{on} \quad U_\alpha\cap U_\beta\cap U_\gamma\cap V$$

Hence we can use the $A_{\alpha\beta} : s$ as transition functions to define a vector bundle, E, over V. Now (30) means precisely that

$$((g^\alpha)^1,\ldots,(g^\alpha)^\beta)$$

should define a section of $E\otimes C^\infty_{(1,0)}$ over V($C_{1,0}$ is the bundle of $(1,0)$-forms on $\mathbb{C}^n$ which is trivial so $E\otimes C^\infty_{(1,0)}\simeq E^n$, the cartesian product of n copies of E). We will achieve this for $z\in V$, which is all we need. But we must be careful to ensure that the $g^\alpha : s$ have holomorphic extensions for ζ outside of V satisfying (27) since otherwise we might for instance get the trivial solution $g^\alpha=0$. Therefore we define a sheaf, $\mathscr{F}$, over $V\times V$ in the following way:

First define a vector bundle, F, over $V\times V$ by using the cover

$$\tilde{U}_\alpha=U_\alpha\times V$$

and the transition matrices $A_{\alpha\beta}(\zeta)$. Thus, F is the pullback of E to $V\times V$ under the first coordinate map. Then let $\mathscr{F}$ be the presheaf whose local sections over $W\times U$ (W, U open in V) are sections, s, to $F\times C^\infty_{(1,0)}$, satisfying the condition:

For each α there is an open subset N_α of U_α such that $N_\alpha\cap V\supseteq U_\alpha\cap U$, and if s is

$$s=\sum s_j d\zeta_j$$

there is a holomorphic extension to $W\times N_\alpha$

$$\tilde{s}=\sum\tilde{s}_j d\zeta_j$$

satisfying

$$\sum\tilde{s}_j(z_j-\zeta_j)=\langle\tilde{s},z-\zeta\rangle=0. \tag{31}$$

This clearly defines a presheaf over $V\times U$. In fact, it is also a coherent analytic sheaf, but we postpone a proof of this in order not to break the general argument.

Now, for each α choose $g^\alpha : s$ as above. Define a section, $s_{\alpha\beta}$, to $F\times C^\infty_{(1,0)}$ over $(U_\alpha\cap U_\beta\cap V)\times V$ by

$$(s_{\alpha\beta})_\gamma=A_{\gamma\alpha}\begin{pmatrix}(g^\alpha)^1\\ \vdots\\ (g^\alpha)^p\end{pmatrix}-A_{\gamma\beta}\begin{pmatrix}(g^\beta)^1\\ \vdots\\ (g^\beta)^p\end{pmatrix} \tag{32}$$

where $(s_{\alpha\beta})_\gamma$ means the representation of $s_{\alpha\beta}$ in the frame over $\tilde{V}_\gamma$. Clearly

$$A_{\tau\gamma}(s_{\alpha\beta})_\gamma = (s_{\alpha\beta})_\tau$$

to, indeed, we get a section. Moreover, we can use the same formula to define an extension of $s_{\alpha\beta}$ to $\zeta \in U_\alpha \cap U_\beta$. Then, if $z \in V$,

$$\langle (s_{\alpha\beta})_\beta, z-\zeta\rangle = A_{\beta\alpha}\langle g^\alpha, z-\zeta\rangle - \langle g^\beta, z-\zeta\rangle = -f^\beta(\zeta)+f^\beta(\zeta)=0.$$

Hence $s_{\alpha\beta}$ is also a section to $\mathscr{F}$. Finally observe that

$$s_{\alpha\beta}+s_{\beta\gamma}+s_{\gamma\alpha}=0 \quad \text{on} \quad \tilde{U}_\alpha \cap \tilde{U}_\beta \cap \tilde{U}_\gamma$$

to $s_{\alpha\beta}$ is a cocycle with values in $\mathscr{F}$. It follows from Cartan's theorem B that we can find sections s_α over $\tilde{U}_\alpha$ such that

$$s_{\alpha\beta} = s_\alpha - s_\beta \quad \text{on} \quad \tilde{U}_\alpha \cap \tilde{U}_\beta.$$

Put

$$\begin{pmatrix} (\gamma^\alpha)^1 \\ \vdots \\ (\gamma^\alpha)^p \end{pmatrix} = \begin{pmatrix} (g^\alpha)^1 \\ \vdots \\ (g^\alpha)^p \end{pmatrix} - (s_\alpha)_\alpha \quad \text{on} \quad \tilde{U}_\alpha.$$

Then

$$A_{\beta\alpha}\begin{pmatrix} (\gamma^\alpha)^1 \\ \vdots \\ (\gamma^\alpha)^p \end{pmatrix} = A_{\beta\alpha}\begin{pmatrix} (g^\alpha)^1 \\ \vdots \\ (g^\alpha)^p \end{pmatrix} - A_{\beta\alpha}((s_{\alpha\beta})_\alpha - (s_\beta)_\alpha)$$

$$= \begin{pmatrix} (g^\beta)^1 \\ \vdots \\ (g^\beta)^p \end{pmatrix} - (s_\beta)_\beta = \begin{pmatrix} (\gamma^\beta)^1 \\ \vdots \\ (\gamma^\beta)^p \end{pmatrix} \quad \text{on} \quad \tilde{U}_\alpha \cap \tilde{U}_\beta.$$

Moreover, by the definition of $\mathscr{F}$ γ^α has an extension, $\tilde{\gamma}^\alpha$, for ζ in a neighbourhood of $U_\alpha \cap V$ satisfying

$$\langle (\tilde{\gamma}^\alpha)^k, z-\zeta\rangle = -f_k^\alpha(\zeta) \quad \text{for} \quad z \in V. \tag{33}$$

This is all we need to construct the kernels. First extend $(\tilde{\gamma}^\alpha)^k$ holomorphically to $z \in \bar{D}$. Then take an arbitrary C^1 extension to $\zeta \in \bar{D}$. Use the result, which we still call $\tilde{\gamma}^\alpha$ to define a kernel on D. That is, put

$$Q = \frac{\sum \bar{f}_k^\alpha (\tilde{\gamma}^\alpha)_j^k d\zeta_j}{|f^\alpha|^2 + \varepsilon}$$

and proceed as before. If we then apply Theorem 0 to a form u with support in U, the neighbourhood of $V \cap U_\alpha$ where $\tilde{\gamma}^\alpha$ is holomorphic and satisfies (33), the same argument as before shows that we can let $\varepsilon \to 0$ and get a kernel supported on V. Since the kernels are given by (8) and (9) with μ replaced by μ^α, and since the $\mu^\alpha : s$ agree on overlaps we get in fact a global representation formula for $z \in V$ (a form u can always be decomposed as a sum of forms with support in U_α). This completes the construction of the kernels in the general case.

It is clear that in this case too, we can use the representation formulas for holomorphic functions to define holomorphic extensions outside of V. One just has to partition the integral in a sum of integrals over $U_\alpha \cap V$ and extend each term in the sum. However, since the current μ is constructed in such an abstract way, it seems that one can get concrete information only in the case when V extends over the boundary. We conclude this section with the proof that $\mathscr{F}$ is a coherent sheaf.

To prove that $\mathscr{F}$ is a sheaf we must prove that if s_a are sections over open sets U_a and $s_a = s_b$ on $U_a \cap U_b$, then there is a global section s on $\cup U_a$ whose restriction to each U_a is s_a. It is enough to prove the following

Claim. *If $s=(s^1,\dots,s^n)$ is defined on $U \cup V$, $U \subseteqq U_\alpha$, $U = \cup U_a$, and in each U_a there is an extension s_a satisfying*

$$\langle s_a, z-\zeta\rangle = 0$$

then s has an extension to satisfying the same thing.

This follows from standard cohomology arguments if we can show that the sheaf of n-tuples $t=(t^1,\dots,t^n)$ satisfying

$$\langle t, z-\zeta\rangle = 0 \quad \text{and} \quad t=0 \quad \text{on} \quad V$$

is coherent. For this, by Oka's lemma it is enough to prove that it is locally finitely generated. But a local n-tuple t vanishes on V exactly when it can be written

$$t^j = \sum_k a_k^j f_k .$$

The $a_k^j : s$ must satisfy

$$\sum a_k^j f_k (z_j - \zeta_j) = 0 ,$$

i.e. a_k^j must be a section to the sheaf of relations between $\{f_k(z_j-\zeta_j)\}$. As this sheaf is locally finitely generated by Oka's lemma, we are done.

That $\mathscr{F}$ is locally finitely generated, hence coherent follows since the sheaf of n-tuples, t, over $\mathbb{C}^n$ satisfying

$$\langle t, z-\zeta\rangle = 0$$

has this property.

2. Division

In this section we shall use our kernels to get explicit solutions to some division problems. More precisely, given functions $f_1,\dots,f_p$ and a function h, we want to find functions $g_1,\dots,g_p$ such that

$$h = \sum f_j g_j \tag{34}$$

Very precise existence theorems for this kind of problems exist (see [10] and [15]). Our results cover only a part of those theorems but we still feel they may have some interest because of the explicit nature of the solution.

Again, let $f_1,\dots,f_p$ be holomorphic in a domain D and, say, in $C(\bar{D})$. This time we do not assume $\partial f_1 \dots \partial f_p \neq 0$, the most interesting case is when this does not

hold. As before take functions g_j^k such that

$$f_k(z)-f_k(\zeta)=\sum_1^n g_j^k(\zeta,z)(z_j-\zeta_j),$$

and define Q_ε^2 as in (22)

$$Q_\varepsilon^2=\frac{\sum\overline{f_k(\zeta)}g_j^k d\zeta_j}{|f|^2+\varepsilon}.$$

Then put $q=\min(n+1,p)$ and $G_2(\alpha)=\alpha^q$. The choice of G_1 and Q^1 will be different depending on the growth condition we impose, and the domain. Let us start with the case of a bounded domain D which is strictly pseudeconvex with C^2-boundary.

Take, as in Sect. 1,

$$G_1(\alpha)=\alpha^{-N} \quad \text{and} \quad Q^1=\frac{\sum h_j d\zeta_j}{\varrho}$$

when ϱ is a strictly plurisubharmonic defining function for D and $h_j(\zeta,z)$ are holomorphic in z and satisfy

$$\operatorname{Re}H(\zeta,z)=\operatorname{Re}\sum h_j(\zeta,z)(z_j-\zeta_j)\leqq\varrho(z)-\varrho(\zeta)-\delta|\zeta-z|^2.$$

For the case of a convex domain $h_j=\partial\varrho/\partial\zeta_j$ will do.

The kernel we obtain is

$$P=\sum_{k=0}^{\min(n,q)} C_k\left(\frac{\varrho(\zeta)}{H(\zeta,z)+\varrho(\zeta)}\right)^{N+n+k}\left(\frac{\overline{f(\zeta)}f(z)+\varepsilon}{|f|^2+\varepsilon}\right)^{q-k}(\bar\partial Q^1)^{n-k}\wedge(\bar\partial Q_\varepsilon^2)^k$$

and the representation formula

$$h(z)=C_n\int h(\zeta)P(\zeta,z) \tag{35}$$

for say $h\in H(D)\cap C^1(\bar D)$. As usual ε shall tend to zero, and this will give us a holomorphic representation kernel which is divisible by the $f_k(z)$:s. This means that (35) will give an explicit solution to (34).

From the estimates in section 1 we have

$$\left|\frac{\overline{f(\zeta)}f(z)+\varepsilon}{|f|^2+\varepsilon}\right|^{q-k}\cdot\|(\bar\partial Q_\varepsilon^2)^k\|\leqq C\frac{(|f|^2+\varepsilon)^{q-k}|f|^2}{(|f|^2+\varepsilon)^{q+1}} \quad \text{for} \quad k<q \tag{36}$$

and

$$\|(\bar\partial Q_\varepsilon^2)^p\|\leqq C\frac{\varepsilon}{(|f|^2+\varepsilon)^{p+1}}.$$

Since

$$\|(\bar\partial Q^1)^{n-k}\|\leqq C(-\varrho)^{-(n-k+1)}$$

we have that

$$\|P\|\leqq C\varrho^{N-1}|f|^{-2q}$$

uniformly for z in a compact subset of D. Assume

$$\int_D \varrho^{N-1} \frac{|h|}{|f|^{2q}} d\lambda < \infty . \tag{37}$$

Then Lebesgues theorem on dominated convergence shows that we can put $\varepsilon = 0$ in (35). The term with $k = p$ in the definition of P becomes zero so we get

Proposition 1. *Under the above assumptions we have the representation formula*

$$h(z) = C_n \int_D h(\zeta) \sum_{k=0}^{q-1} C_k \left(\frac{\varrho}{H(\zeta, z) + \varrho} \right)^{N+n-k} \left(\frac{\overline{f(\zeta)} f(z)}{|f|^2} \right)^{q-k} (\bar{\partial} Q^1)^{n-k} \wedge (\bar{\partial} Q_0^2)^k .$$

Since $k < q$ for all terms in the sum we have an explicit solution to (34).

Next, take $D = \mathbb{C}^n$, and let φ be a convex function as in Sect. 1. Put

$$Q^1 = A \partial \varphi \quad \text{and} \quad G_1(\alpha) = \exp \alpha .$$

Assume $|h|$, $|f_k| \leqq C \exp C\varphi$ and that for some C

$$\int_{\mathbb{C}^n} |h| \exp - C\varphi |f|^{-2q} d\lambda < \infty . \tag{38}$$

With a similar argument as above we then get

Proposition 2. *If A is large enough*

$$h(z) = C_n \int_{\mathbb{C}^n} h(\zeta) \exp A \langle \partial \varphi, z - \zeta \rangle \sum_{k=0}^{q-1} \left(\frac{\overline{f(\zeta)} f(z)}{|f|^2} \right)^{q-k} (A \partial \bar{\partial} \varphi)^{n-k} \wedge (\bar{\partial} Q_0^2)^k$$

We thus get a solution to (34) with

$$|g_j| \leqq C \exp C\varphi .$$

Similarily one can treat the algebraic case and get a proof of Hilbert's "Nullstellensatz", but we omit the details.

The exponent q in (37) and (38) can not be taken smaller (this is clear at least if $p \leqq n$ since otherwise h need not even vanish on the set of common zeros of $f_1 \ldots f_p$). Nevertheless, the assumption in Skoda [15]

$$\exists \alpha > 1 \int_D \varrho^{N-1} \frac{|h|^2}{|f|^{2\alpha q}} d\lambda < \infty$$

is less restrictive since it is a condition on $|h|^2$ instead of $|h|$. We refer to [15] for the statement of Skoda's general result, which also includes a careful estimate of the solution.

References

1. Amar, E.: Extension de fonctions holomorphe et courants. Preprint Orsay 1980
2. Berenstein, C.A., Taylor, B.A.: Interpolation problems in $\mathbb{C}^n$ with applications to harmonic analysis. J. Analyse Math. **38** (1980)
3. Berenstein, C.A., Taylor, B.A.: On the geometry of interpolating varieties. Sem. Lelong-Skoda 1980–1981. Lecture Notes in Mathematics 919. Berlin, Heidelberg, New York: Springer 1982

4. Berndtsson, B., Andersson, M.: Henkin-Ramirez formulas with weight factors. Ann. Inst. Fourier **32**, 3 (1982)
5. Fornaess, J.E.: Embedding strictly pseudoconvex domains in convex domains. Am. J. Math. **98** (1976)
6. Henkin, G.M.: Integral representation of a function in a strictly pseudoconvex domain and applications to the $\bar{\partial}$-problem (Russian) Mat. Sb. (1982)
7. Henkin, G.M.: Continuation of bounded holomorphic functions from submanifolds in general position to strictly pseudoconvex domains (Russian). Izv. Akad. Nauk. SSSR, Ser. Mat. **36** (1972)
8. Henkin, G.M., Leiterer, J.: Global integral formulas for solving the $\bar{\partial}$-equation on Stein manifolds. Ann. Pol. Math. (to appear)
9. Hörmander, L.: An introduction to complex analysis in several variables. Amsterdam: Van Nostrand 1966
10. Hörmander, L.: Generators for some rings of analytic functions. Bull. Am. Math. Soc. **73** (1967)
11. Lelong, P.: Fonctions plurisousharmoniques et formes différentielles positives. London: Gordon and Breach 1968
12. Nishimura, Y.: Probléme d'extension dans la théorie des fonctions entiére d'ordre fini. J. Math. Kyoto Univ. **20** (1980)
13. Palm, A.: Integral representation formulas on strictly pseudoconvex domains in Stein manifolds. Duke Math. J. **43** (1976)
14. Ramirez de Arellao, E.: Ein Divisionsproblem und Randintegraldarstellungen in der komplexen Analysis. Math. Ann. **184**, 172 (1969)
15. Skoda, H.: Application des techniques L^2 a la theorie des ideaux d'une algébre de fonctions holomorphes avec poids. Ann. Sci. Ecole Norm. Sup. **5** (1972)
16. Stout, E.L.: An integral formula for holomorphic functions on strictly pseudoconvex hypersurfaces. Duke Math. J. **42** (1975)
17. Yoshioka, T.: Cohomologie a estimation L^2 et extension des fonctions holomorphes avec controle de la croissance. Proc. Jpn. Acad. **57** (1981)
18. Cornalba, M., Shiffman, B.: A counterexample to the "Transcendental Bezov-problem". Ann. Math. **96** (1972)
19. Demailly, J.P.: Scindage holomorphe d'un morphisme de fibrés vectoriels semi-positifs avec estimation L^2. Sem. Lelong-Skoda 1980–81. Lecture Notes in Mathematics 919. Berlin, Heidelberg, New York: Springer 1982
20. Hortmann, M.: Globale holomorphe Kerne zur Lösung der Cauchy-Riemannschen Differentialgleichungen. Annals Math. Studies 100. Princeton: Princeton University Press 1981

Received November 22, 1982

Math. Ann. 263, 419–433 (1983)

Mathematische Annalen
© Springer-Verlag 1983

On Symmetric Submanifolds of Spaces of Constant Curvature

Erich Backes and Helmut Reckziegel

Mathematisches Institut der Universität, Weyertal 90, D-5000 Köln, Federal Republic of Germany

The article presents an algebraic characterization of the second fundamental form of symmetric submanifolds in standard spaces of constant curvature, and as application the classification of the symmetric submanifolds and more generally of all isometric immersions with parallel second fundamental form into the standard spaces.

Introduction

In 1980 D. Ferus introduced the concept of symmetric submanifolds of euclidean space [7], he "demonstrated the prominent rôle played by them in extrinsic differential geometry", and he classified them. A submanifold of $\mathbb{R}^n$ is symmetric iff it is invariant with respect to the reflections at its normal spaces. This definition makes sense in every standard space N_κ^n of constant curvature κ (see Def. 1); and in the non-euclidean spaces the symmetric submanifolds are of the same importance as in $\mathbb{R}^n$.

As the symmetry of submanifolds is intimately related to the parallelity of the second fundamental form (see Sect. 1, Prop. of Ferus/Strübing), Theorem 2 of [12] shows that a symmetric submanifold S is completely determined by one point $p \in S$, the tangent space $\mathbf{E} = T_pS$ and the second fundamental form $\mathbf{h} = h_p^S$ of S at p. It is therefore a very natural question to ask for a characterization of all pairs $(\mathbf{E}, \mathbf{h})$ which occur as *initial data* of symmetric submanifolds. The main purpose of the article consists in solving this problem. For that we associate a trilinear operation $\mathbf{L}$ on $\mathbf{E}$ with the pair $(\mathbf{E}, \mathbf{h})$, a procedure which was suggested by Ferus' investigation [7]; see our Def. 2. Our main result (Theorem 3) can now be formulated as follows: $(\mathbf{E}, \mathbf{h})$ *are initial data of a symmetric submanifold, if and only if* $\mathbf{L}$ *is a euclidean Jordan triple system* (see Def. 3). For this test you have only to check one algebraic equation, namely (9). (It should be mentioned that in spaces of non-constant curvature it is already difficult to characterize the tangent spaces of totally geodesic submanifolds.) For the proof we decompose the Jordan triple system $\mathbf{L}$ (Lemma 3) and reduce the problem thereby to the two special cases

handled in Theorem 1 and 2. Theorem 1 corresponds to the *classification of full minimal symmetric submanifolds* $S \subset N_\kappa^n$: Either $S = N_\kappa^n$, or N_κ^n is a euclidean sphere (i.e., $\kappa > 0$) and S is a standard imbedded symmetric R-space (see [5, 15, 2] for $\kappa = 0$, [15, 2] for $\kappa < 0$, and [7] for $\kappa > 0$; symmetric R-spaces are classified in [8, 11]).

In Theorem 2 we construct symmetric submanifolds $S \subset N_\kappa^n$ with the following properties:

(i) S is isometric to a Riemannian product $N_{\kappa_1}^{m_1} \times S_{\kappa_2}^{m_2} \times \ldots \times S_{\kappa_k}^{m_k}$ $(S_{\kappa_i}^{m_i} := N_{\kappa_i > 0}^{m_i})$,

(ii) the shape of S is in accord with the product structure of S [in the sense of Def. 4, formula (13)], and

(iii) each slice of S parallel to one of the factors is a spherical submanifold of N_κ^n.

These symmetric submanifolds are called *extrinsic products of spherical submanifolds* (see Def. 4). In Theorem 2 they are described as image of isometric imbeddings, which can be composed of well-known canonical maps. In the non-euclidean case these extrinsic products are substitutes for the canonical decompositions of the euclidean space into affine subspaces. By means of them it is possible also to generalize other results of euclidean geometry to arbitrary standard spaces. In the Corollary of Theorem 2 these extrinsic products are characterized as the symmetric submanifolds with flat normal bundle, a result which completes investigations of Walden [16] and Akiba [1].

In the proof for the implication (iii)⇒(i) of Theorem 3 we give a very special construction of symmetric submanifolds, and it is shown that in this way every symmetric submanifold is obtained. Thus we have proved also the following classification of symmetric submanifolds.

Theorem 4. *A submanifold S of the standard space N_κ^n is symmetric, if and only if it can be represented as image $g(M_1 \times \ldots \times M_k)$, where*

$$g : N_{\kappa_1}^{m_1} \times S_{\kappa_2}^{m_2} \times \ldots \times S_{\kappa_k}^{m_k} \to N_\kappa^n$$

is an extrinsic product of spherical submanifolds (as described in Theorem 2) and each M_i is a full minimal symmetric submanifold of $N_{\kappa_i}^{m_i}$ $(= S_{\kappa_i}^{m_i}$ for $i \geqq 2)$; i.e., $M_i = N_{\kappa_i}^{m_i}$, or $\kappa_i > 0$ and M_i is a standard imbedded symmetric R-space of $S_{\kappa_i}^{m_i}$.

As mentioned above Ferus has already given the classification in the euclidean case [7], and the spherical case can be reduced to it; the hyperbolic case however requires new investigations. They were also carried out independently and differently by M. Takeuchi [15]. He follows Ferus' ideas carrying over a factorization lemma of J. D. Moore to the Lorentz space. In contrast to our Theorem 4 he lists the symmetric submanifolds in the various standard spaces separately. For instance, it can easily be seen from his list or Theorem 4 that every symmetric submanifold of $H_\kappa^n := N_{\kappa<0}^n$ is one of the extrinsic products of Theorem 2 in case $n \leqq 4$; in H_κ^5 the only exceptions are the 2-dimensional Veronese surfaces of geodesic hyperspheres of H_κ^5 (see [16], Satz 2).

As a further application of Theorem 3 we generalize the well-known fact that the image of a totally umbilical immersion $f : M \to \mathbb{R}^n$ (M connected, $\dim M \geqq 2$) is an open part of an affine subspace or a sphere of $\mathbb{R}^n$:

Theorem 5. *If M is a connected Riemannian manifold and $f : M \to N_\kappa^n$ an isometric immersion with parallel second fundamental form, then $f(M)$ is an open part of*

a symmetric submanifold S of N_κ^n. If in addition M is complete, then f is a covering map onto S (see [9] I, p. 176); as S is an (intrinsic) Riemannian symmetric space (see the Remark of Sect. 1), the manifold M then can be proved to be a symmetric space, too.

For the proof we choose $p \in M$ and apply Prop. 1 and Theorem 3. Therefrom we obtain a symmetric submanifold S of N_κ^n such that $f(p) \in S$, $T_{f(p)}S = f_* T_p M$ and the second fundamental forms h_p^f and $h_{f(p)}^S$ coincide, more exactly: $h_p^f(v, w) = h_{f(p)}^S(f_* v, f_* w)$. According to [12] Theorem 2, f is then an isometric immersion into S.

A last remark is in order: Theorem 5 shows that in N_κ^n the symmetric submanifolds are exactly the *reduced, maximally expanded submanifolds* with parallel second fundamental form. In [13] exact definitions are given, especially compare Theorem 5 and [13] Theorem 1. On the other hand, as any space N of constant curvature κ is locally isometric to N_κ^n it is possible to generalize Theorem 3 to N replacing the expression "symmetric submanifold" by "reduced, maximally expanded isometric immersion with parallel second fundamental form".

1. Preliminaries

In this article all manifolds, maps, ... are supposed to be of class C^∞. – First we will fix some notations and recall some definitions. Let M and N be Riemannian manifolds, $f: M \to N$ an isometric immersion and ξ a vector bundle over M. Then $T_p M$ denotes the tangent space and $T_p^1 M$ the space of unit vectors of M at $p \in M$; $\Gamma(\xi)$ the modul of global C^∞ sections of ξ; $\nu(f)$ the normal bundle, $h = h^f$ the second fundamental form, $A = A^f$ the shape operator, and $H = H^f$ the mean curvature vector field of f (see [9] Ch. VII.3). To say that h is *parallel* means

$$D_X(h(Y, Z)) = h(\nabla_X Y, Z) + h(Y, \nabla_X Z) \quad \text{for all} \quad X, Y, Z \in \Gamma(TM),$$

where ∇ denotes the Levi-Civita derivative of M and D the canonical covariant derivative of $\nu(f)$. The immersion f is said to be *spherical*, iff H is parallel in $\nu(f)$ and $h(X, Y) = \langle X, Y \rangle H$ for all $X, Y \in \Gamma(TM)$. Obviously every totally geodesic immersion is spherical, and every spherical one has parallel second fundamental form. – In case M is connected, the immersion f is said to be *full*, iff there exists no totally geodesic submanifold S of N (see below) with $\dim S < \dim N$ such that M is differentiably immersed into S by f. – Let L be a *submanifold* of M, i.e. the underlying set of L is a subset of M and the inclusion map $i_L: L \hookrightarrow M$ is an immersion. If i_L is spherical, then L is called a spherical submanifold (or an extrinsic sphere). Analogously submanifolds with parallel second fundamental form and full submanifolds are defined. Of course, $f|L = f \circ i_L$ is also an isometric immersion; in particular we use this notation if L is a foliation of M (see [3] Sect. 9.2). Finally, L is called a *regular submanifold* of M iff the topology of L is induced by the topology of M. – The n-dimensional *standard space* (= space form) of constant sectional curvature κ is denoted by N_κ^n. Besides we use the notation $N_0^n = \mathbb{R}^n$, $N_{\kappa>0}^n = S_\kappa^n$ ($= n$-dimensional sphere $\subset \mathbb{R}^{n+1}$ of radius $1/\sqrt{\kappa}$), and $N_{\kappa<0}^n = H_\kappa^n$ ($= n$-dimensional hyperbolic space = upper half space $\mathbb{R}^{n-1} \times \mathbb{R}_+$ with Riemannian metric $(\sqrt{-\kappa} \cdot x_n)^{-2} \cdot$ euclidean metric). Using this description, N_κ^n is also defined for $n = 1$ (of course, H_κ^1 and $\mathbb{R}$ are isometric spaces).

The spherical submanifolds of the standard spaces N_κ^n are well-known; see Prop. 2. The following lemmas are of interest, because they prove the parallelity of the second fundamental form of further isometric immersions.

Lemma 1. *Let $L \xrightarrow{g} M \xrightarrow{f} N$ be isometric immersions, $p \in L$ and $v, w \in T_pL$. Then*

$$\nu_p(f \circ g) = f_* \nu_p(g) \oplus \nu_{g(p)}(f) \tag{1}$$

and

$$h^{f \circ g}(v, w) = f_* h^g(v, w) + h^f(g_* v, g_* w). \tag{2}$$

If moreover h^f is parallel and

$$h^f(g_* v, \eta) = 0 \quad \text{for all} \quad p \in L, \quad v \in T_pL, \quad \eta \in \nu_p(g), \tag{3}$$

then we have the equivalence

$$h^g \text{ is parallel} \Leftrightarrow h^{f \circ g} \text{ is parallel}.$$

Obviously (3) is satisfied, if f is spherical.

Lemma 2. *Let $M = M_1 \times \ldots \times M_k$ and $N = N_1 \times \ldots \times N_k$ be products of Riemannian manifolds M_i resp. N_i, and let $g = g_1 \times \ldots \times g_k : M \to N$ be the product of isometric immersions $g_i : M_i \to N_i$. The second fundamental form h^g is then related to the second fundamental forms h_i of g_i by*

$$\tau_{i*} h^g(v, w) = h_i(\pi_{i*} v, \pi_{i*} w)$$

for $i = 1, \ldots, k$, and $v, w \in T_pM$, $p \in M$, where $\pi_i : M \to M_i$ and $\tau_i : N \to N_i$ denote the canonical projections. Furthermore we have the implication

$$(\forall i : h_i \text{ is parallel}) \Rightarrow h^g \text{ is parallel}.$$

Now we generalize Ferus' definition of symmetric submanifolds to arbitrary standard spaces:

Definition 1. Let S be a *connected, regular* submanifold of N_κ^n, and for every $p \in S$ let $s_p : N_\kappa^n \to N_\kappa^n$ denote the isometry with the initial data

$$s_p(p) = p \quad \text{and} \quad s_{p*} v = \begin{cases} -v & \text{for} \quad v \in T_pS \\ v & \text{for} \quad v \in \nu_p(S). \end{cases}$$

If we have $s_p(S) = S$ for all $p \in S$, then we call S a *symmetric submanifold* of N_κ^n.

Examples of symmetric submanifolds

a) standard imbedded symmetric R-spaces in S_κ^n, see [7, 10],

b) the spherical submanifolds of Prop. 2.

Remark. Every symmetric submanifold S of N_κ^n is an intrinsic Riemannian symmetric space. Indeed, $s_p | S$ is the geodesic symmetry of S at p. In particular S is a complete Riemannian subspace of N_κ^n.

The investigation of symmetric submanifolds is essentially facilitated by the following discovery due to Ferus [7] and Strübing [14], see also [12] p. 101, Cor. 1.

Proposition (Ferus/Strübing). *For every connected, complete, regular submanifold S of N_κ^n the following statements are equivalent:*

(i) *S is symmetric,*

(ii) *S has parallel second fundamental form.*

2. Euclidean Jordan Triple Systems

In this section we make the algebraic preparations for the construction of symmetric submanifolds to initial data and solve the problem in a special case (Theorem 1). The principal algebraic idea is the following construction of a triple system.

Definition 2. Let V be a euclidean vector space, $\mathbf{E}$ a linear subspace of V, $\mathbf{h}: \mathbf{E} \times \mathbf{E} \to \mathbf{E}^\perp$ (= the orthogonal complement of $\mathbf{E}$ in V) a symmetric bilinear map, and $\kappa \in \mathbb{R}$ an arbitrary real number. Then we define the *triple system*[1] $\mathbf{L}$ *of the initial data* $(\mathbf{E}, \mathbf{h})$ with respect to (V, κ) to be the bilinear map

$$\mathbf{L} = \mathbf{S} + \mathbf{R} : \mathbf{E} \times \mathbf{E} \to \mathrm{End}(\mathbf{E}),$$

where the symmetric and skew-symmetric parts $\mathbf{S}$ and $\mathbf{R}$ are given by

$$\langle \mathbf{S}(x, y)v, w \rangle = \kappa \cdot \langle x, y \rangle \langle v, w \rangle + \langle \mathbf{h}(x, y), \mathbf{h}(v, w) \rangle \tag{4}$$

and

$$\mathbf{R}(x, y)z = \mathbf{S}(y, z)x - \mathbf{S}(x, z)y. \tag{5}$$

Combining the formulas (4) and (5) we obtain the "Gauß equation"

$$\begin{aligned} \langle \mathbf{R}(x, y)v, w \rangle = \kappa \cdot (&\langle x, w \rangle \cdot \langle y, v \rangle - \langle x, v \rangle \cdot \langle y, w \rangle) \\ &+ \langle \mathbf{h}(x, w), \mathbf{h}(y, v) \rangle - \langle \mathbf{h}(x, v), \mathbf{h}(y, w) \rangle . \end{aligned} \tag{6}$$

Therefore we have

$$\mathbf{L}(x, y)^* = \mathbf{L}(y, x) \quad \text{for all} \quad x, y \in \mathbf{E}, \tag{7}$$

where * marks the adjoint with respect to the inner product of $\mathbf{E}$. A further trivial calculation shows

$$\mathbf{L}(x, y)z = \mathbf{L}(z, y)x \quad \text{for all} \quad x, y, z \in \mathbf{E}. \tag{8}$$

Definition 3. If $\mathbf{E}$ is a euclidean vector space, a bilinear map

$$\mathbf{L} : \mathbf{E} \times \mathbf{E} \to \mathrm{End}(\mathbf{E})$$

is called a *euclidean Jordan triple system* (in this article abbreviated by JTS) on $\mathbf{E}$, iff the formulas (7), (8) and the following formula (9) for the commutator are satisfied:

$$[\mathbf{L}(x, y), \mathbf{L}(v, w)] = \mathbf{L}(\mathbf{L}(x, y)v, w) - \mathbf{L}(v, \mathbf{L}(y, x)w). \tag{9}$$

---

1 The notation suggests that $\{x, y, z\} = \mathbf{L}(x, y)z$ may be regarded as trilinear operation on $\mathbf{E}$

In [2] the following example is presented:

Proposition 1. *Let $f: M \to N$ be an isometric immersion with parallel second fundamental form h into a space N of constant curvature κ. If $p \in M$, $V = T_{f(p)}N$, $\mathbf{E} = f_* T_p M$, and $\mathbf{h}: \mathbf{E} \times \mathbf{E} \to \mathbf{E}^\perp$ is given by $\mathbf{h}(f_* x, f_* y) = h(x, y)$, then the triple system $\mathbf{L}$ of $(\mathbf{E}, \mathbf{h})$ with respect to (V, κ) is a JTS.*[2]

Conversely, we will try to find symmetric submanifolds for prescribed initial data $(\mathbf{E}, \mathbf{h})$. A first well-known result is:

Proposition 2. *Let p be a point of the standard space $N_\kappa^n =: N$, $\mathbf{E}$ an m-dimensional linear subspace of $T_p N$ and $\boldsymbol{\eta}$ a vector of $T_p N$ orthogonal to $\mathbf{E}$. Then there exists one and only one regular, complete spherical submanifold S of N such that $p \in S$, $T_p S = \mathbf{E}$, and $\boldsymbol{\eta}$ is the mean curvature normal of S at p. The space S is isometric to $N_{\bar\kappa}^m$ with $\bar\kappa = \kappa + \langle \boldsymbol{\eta}, \boldsymbol{\eta} \rangle$. In particular, S is a symmetric submanifold of N with the second fundamental form $h(x, y) = \langle x, y \rangle \boldsymbol{\eta}$ at p.*

A strong generalization of Prop. 2 is Theorem 2 on the existence of "extrinsic products of spherical submanifolds". Another basic result is the following theorem on "minimal" initial data.

Theorem 1. *Let p be a point of a standard space $N_\kappa^n =: N$, $\mathbf{E}$ a subspace of $T_p N$ and $\mathbf{h}: \mathbf{E} \times \mathbf{E} \to \mathbf{E}^\perp$ a "minimal" symmetric bilinear map, i.e., $\sum \mathbf{h}(x_i, x_i) = 0$ for any orthonormal basis (x_i) of $\mathbf{E}$. Furthermore, we suppose that the triple system of $(\mathbf{E}, \mathbf{h})$ is a JTS, see Def. 2 and 3. Then the following assertions are true:*

(i) *If $\kappa \leqq 0$, then $\mathbf{h} = 0$; therefore there exists a (regular) complete totally geodesic submanifold M of N such that $p \in M$ and $T_p M = \mathbf{E}$.*

(ii) *If $\kappa > 0$ (i.e., $N = S_\kappa^n$) and $T_p N = \mathbf{E} \oplus \operatorname{span}\{\mathbf{h}(x, x) | x \in \mathbf{E}\}$, then either $\mathbf{h} = 0$ or there exists a standard imbedded symmetric R-space M in N (see [10]) with $p \in M$, $T_p M = \mathbf{E}$ and which has the prescribed second fundamental form $\mathbf{h}$ at p. In particular, M is a minimal symmetric submanifold, fully imbedded in N.*

Proof. As $\mathbf{h}$ is minimal, we obtain $\mathbf{h} = 0$ in case $m := \dim \mathbf{E} = 1$ immediately. Therefore we suppose $m \geqq 2$ from now.

For (i) we use Theorem 1 of [2] which states that

$$\operatorname{trace} \mathbf{S}(\mathbf{S}(x, y)x, y) \geqq 0 \quad \text{for all} \quad x, y \in \mathbf{E}, \tag{10}$$

and that equality holds if and only if $\mathbf{S}(x, y) \equiv 0$; see formula (4). As $\mathbf{h}$ is minimal, we obtain

$$\operatorname{trace} \mathbf{S}(x, y) = m \cdot \kappa \cdot \langle x, y \rangle, \tag{11}$$

hence

$$\operatorname{trace} \mathbf{S}(\mathbf{S}(x, y)x, y) = m \cdot \kappa \cdot (\kappa \cdot \langle x, y \rangle^2 + \|\mathbf{h}(x, y)\|^2). \tag{12}$$

In case $\kappa < 0$ there follows $\|\mathbf{h}(x, y)\|^2 \leqq -\kappa \langle x, y \rangle^2$ from (10), i.e., $\mathbf{h}(x, y) = 0$ if x, y are orthonormal; hence $\mathbf{h}(x, x) = \mathbf{h}(y, y)$ for any two unit vectors. As $\mathbf{h}$ is minimal, $\mathbf{h} = 0$.

In case $\kappa = 0$, formula (12) shows that equality holds in (10) and that therefore $\mathbf{S}(x, y) \equiv 0$ for all $x, y \in \mathbf{E}$. Hence we have $\mathbf{h} = 0$ again.

2 According to formula (6), the skew-symmetric part $\mathbf{R}$ of $\mathbf{L}$ is related to the curvature tensor R_p^M of M at p by $\mathbf{R}(f_* x, f_* y) f_* z = f_* R_p^M(x, y)z$

The existence of the totally geodesic submanifold M follows now from Prop. 2.

For (ii) we regard S^n_κ as subspace of $\mathbb{R}^{n+1}$ and $\mathbf{E}$ as linear subspace of $T_p\mathbb{R}^{n+1}$. Then we have to replace the "second fundamental form" $\mathbf{h}$ by

$$\alpha(x, y) := \mathbf{h}(x, y) + \langle x, y\rangle \eta,$$

where η denotes the mean curvature normal of S^n_κ at p. If we define the linear map $A: \mathbf{E}^\perp \oplus \mathbb{R}\eta \to \mathrm{End}(\mathbf{E})$, $\xi \to A_\xi$ as usually by $\langle A_\xi x, y\rangle = \langle \alpha(x, y), \xi\rangle$, we obtain from (4)

$$A_{\alpha(x,y)} = \mathbf{S}(x, y) \quad \text{and} \quad A_\eta = \kappa \cdot id_{\mathbf{E}}.$$

Now, assertion (ii) can be proved by repeating Ferus' proof of Theorem 1 in [7], using the following lexicon

Ferus	$n+p$	n	x	T	r
here	n	m	p	$\mathbf{E}$	$1/\kappa$

One remark is necessary: According to the Gauß equation (see [9] II, p. 23) the map L in [7] Lemma 2 corresponds to our triple system, which we have *supposed* to be a JTS. Furthermore, Lemma 3 of [7] is an immediate consequence of (11) in our situation. Hence the analogon of [7] Lemma 4 is the crucial point of the proof. □

In the following Lemma we describe a canonical decomposition of euclidean Jordan triple systems, which will be used in the proof of the main Theorem 3.

Lemma 3. *Let* $\mathbf{L}$ *be a JTS on a euclidean vector space* $\mathbf{E}$, $\mathbf{S}$ *its symmetric part,* $\mathbf{A}$ *the self-adjoint operator defined by*

$$\langle \mathbf{A}x, y\rangle = \mathrm{trace}\,\mathbf{S}(x, y),$$

and $\mathbf{E}_1, \ldots, \mathbf{E}_k$ *the eigenspace of* $\mathbf{A}$. *Then the following assertions are true:*

(i) $\mathbf{A} = \sum \mathbf{S}(x_i, x_i)$ *where* (x_i) *is an orthonormal basis of* $\mathbf{E}$,

(ii) $\mathbf{L}(x, y)(\mathbf{E}_i) \subseteq \mathbf{E}_i$ *for all* $x, y \in \mathbf{E}$, *therefore we obtain a JTS on* $\mathbf{E}_i$ *by restricting* $\mathbf{L}$ *to* $\mathbf{E}_i \times \mathbf{E}_i$.

(iii) $\mathbf{L}(x, y)z = \mathbf{L}(z, y)x = 0$ *if* $x \in \mathbf{E}_i$, $y \in \mathbf{E}_j$, $i \neq j$.

Proof. In the following we use the elementary equations $\mathbf{L}(x, x) = \mathbf{S}(x, x)$, $\mathrm{trace}\,\mathbf{L}(x, y) = \mathrm{trace}\,\mathbf{S}(x, y)$, and $\langle \mathbf{L}(x, y)v, w\rangle = \langle \mathbf{L}(v, w)x, y\rangle$, where the last one follows from (7) and (8). Now the proof of (i) is obvious. In order to prove (ii) it suffices to show

$$[\mathbf{A}, \mathbf{L}(x, y)] = 0 \quad \text{for all} \quad x, y = \mathbf{E}.$$

For that we use (7) and (9):

$$\begin{aligned}\langle [\mathbf{A}, \mathbf{L}(x, y)]v, w\rangle &= \langle \mathbf{A}\mathbf{L}(x, y)v, w\rangle - \langle \mathbf{A}v, \mathbf{L}(y, x)w\rangle \\ &= \mathrm{trace}(\mathbf{L}(\mathbf{L}(x, y)v, w) - \mathbf{L}(v, \mathbf{L}(y, x)w)) = \mathrm{trace}[\mathbf{L}(x, y), \mathbf{L}(v, w)] = 0.\end{aligned}$$

Finally, assertion (iii) follows from (ii) and (8). □

3. Extrinsic Products

The aim of this section is to prove the existence of an important class of symmetric submanifolds, see Theorem 2.

Definition 4. Let $M=M_1\times\ldots\times M_k$ be a product of Riemannian manifolds and $L_1,\ldots,L_k$ the totally geodesic foliations of M, canonically induced by the product structure of M. (For instance, in the case $k=2$ the leaf of L_2 through $p=(p_1,p_2)$ is $\{p_1\}\times M_2$.) An isometric immersion $g:M\to N$ into an arbitrary Riemannian manifold N is called an *extrinsic product* in N, iff

$$\forall p\in M, i\neq j\ :\ h^g|T_pL_i\times T_pL_j\equiv 0. \tag{13}$$

If in addition g is an imbedding, h^g is parallel and the "components" $g|L_1,\ldots,g|L_k$ (see Sect. 1) are spherical, then we call g and $g(M_1\times\ldots\times M_k)$ an *extrinsic product of spherical submanifolds* of N.

The case $k=1$ shall be included; thus every isometric immersion can be regarded as an extrinsic product with one factor only. The product $g_1\times\ldots\times g_k$ of isometric immersions $g_i:M_i\to N_i$ is obviously another trivial example of an extrinsic product (in $N_1\times\ldots\times N_k$). An important generalization is the following

Proposition 3. *Let $g:M_1\times\ldots\times M_k\to N$ be an extrinsic product of spherical submanifolds and $g_1,\ldots,g_k$ arbitrary extrinsic products in $M_1,\ldots,M_k$, respectively. Then the following assertions are true:*

a) *$f:=g\circ(g_1\times\ldots\times g_k)$ is also an extrinsic product in N.*

b) *If $g_1,\ldots,g_k$ have parallel second fundamental form, then h^f is also parallel.*

c) *Suppose N to be a space of constant curvature. f has then a flat normal bundle, if and only if the normal bundles $\nu(g_1),\ldots,\nu(g_k)$ are flat.*

Proof. For $p=(p_1,\ldots,p_k)$ let us identify $T_p(\prod M_i)$ and $\oplus T_{p_i}M_i$. Using the notations of Def. 4 we have the equivalence

$$v\in T_pL_j \Leftrightarrow v_i=0 \quad \text{for} \quad i\neq j$$

for every vector $v=(v_i)\in T_p(\prod M_i)$. If we denote the mean curvature normal of $g|L_i$ at p with η_i, we obtain in our special situation

$$h^g(v,w)=\sum\langle v_i,w_i\rangle\cdot\eta_i, \tag{14}$$

because the foliations L_i are totally geodesic.

Now, let g_i be defined on the product $\tilde{M}_i:=\prod M_{ij}$. For every point

$$q=(q_1,\ldots,q_k)\in\prod\tilde{M}_i$$

and vectors $\tilde{v},\tilde{w}\in T_q(\prod\tilde{M}_i)$ we derive then

$$h^f(\tilde{v},\tilde{w})=\sum\langle\tilde{v}_i,\tilde{w}_i\rangle\cdot\eta_i+g_*(h^{g_1}(\tilde{v}_1,\tilde{w}_1),\ldots,h^{g_k}(\tilde{v}_k,\tilde{w}_k)) \tag{15}$$

from (14) if $p:=(g_i(q_i))$ (see Lemma 1 and 2). As the immersions g_i satisfy the characteristic equation (13), the same follows for f; hence a) is proved. – Furthermore, as consequence of the identification of $T_p(\prod M_i)$ and $\oplus T_{p_i}M_i$ we have

$$\nu_q(g_1\times\ldots\times g_k)=\prod\nu_{q_i}(g_i). \tag{16}$$

From (14) we obtain therefore $h^g((g_1\times\ldots\times g_k)_*\tilde{v},z)=0$ for all $\tilde{v}\in T_q(\prod\tilde{M}_i)$ and $z\in\nu_q(g_1\times\ldots\times g_k)$. Thus the assertion b) follows immediately from the Lemmas 1 and 2.

For c): Because of (16) and (1) every vector $\xi \in \nu_q(f)$ has a unique representation

$$\xi = \eta + g_*(z_1, \ldots, z_k) \quad \text{with} \quad \eta \in \nu_{(g_i(q_i))}(g) \quad \text{and} \quad z_i \in \nu_{q_i}(g_i).$$

Multiplying it with (15) we obtain for all vectors $\tilde{v}, \tilde{w} \in T_q(\prod \tilde{M}_i)$

$$\langle A^f_\xi \tilde{v}, \tilde{w} \rangle = \langle h^f(\tilde{v}, \tilde{w}), \xi \rangle = \sum (\langle \eta_i, \eta \rangle \langle \tilde{v}_i, \tilde{w}_i \rangle + \langle h^{g_i}(\tilde{v}_i, \tilde{w}_i), z_i \rangle),$$

hence

$$A^f_\xi \tilde{v} = (A^{g_i}_{z_i} \tilde{v}_i + \langle \eta_i, \eta \rangle \cdot \tilde{v}_i)_{i=1,\ldots,k}. \tag{17}$$

We may now suppose $M_1, \ldots, M_k$ to be connected. If N has constant curvature, the spaces M_i are also of constant curvature because the components $g|L_i$ are spherical. Therefore the assertion c) is a consequence from (17), the equation of Ricci ([4] p. 47 formula 2.16) and the definition of flatness (see [4] p. 99). □

Theorem 2. *Let the following set of data be given: a point p of a standard space $N^n_\kappa =: N$; linear subspaces $\mathbf{E}_1, \ldots, \mathbf{E}_k$ of T_pN which are orthogonal one to each other; and distinct vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_k$ of T_pN orthogonal to $\mathbf{E} := \oplus \mathbf{E}_i$. Put $\kappa_i := \kappa + \langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_i \rangle$ and suppose $\kappa_1 \leqq \kappa_2 \leqq \ldots \leqq \kappa_k$. Furthermore, let $\mathbf{h}: \mathbf{E} \times \mathbf{E} \to \mathbf{E}^\perp$ be the symmetric bilinear map defined by*

$$\mathbf{h}(x, y) = \sum \langle x_i, y_i \rangle \cdot \boldsymbol{\eta}_i, \tag{18}$$

where x_i resp. y_i is the component of x resp. y in $\mathbf{E}_i$. Then the following assertions are equivalent:

(i) $$\langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_j \rangle = -\kappa \quad \textit{for} \quad i \neq j.$$

(ii) *The real numbers $\kappa_2, \ldots, \kappa_k$ are strictly positive, and there exists an extrinsic product*

$$g: N^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k} \to N$$

of spherical submanifolds such that

$$p \in S := g(N^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k}), \qquad T_pS = \mathbf{E}, \quad \textit{and} \quad h^S_p = \mathbf{h};$$

more exactly, if $q := g^{-1}(p)$ and if $L_1, \ldots, L_k$ are the canonical foliations of the product $\prod N^{m_i}_{\kappa_i}$ (see Def. 4), then $g_* T_q L_i = \mathbf{E}_i$, and $\boldsymbol{\eta}_i$ is the mean curvature normal of the component $g|L_i$ at q. In particular, S is a symmetric submanifold.

(iii) *The triple system $\mathbf{L}$ of $(\mathbf{E}, \mathbf{h})$ is a JTS (see Def. 2, 3).*

Remarks. a) Assertion (ii) describes all complete extrinsic products of spherical submanifolds in N^n_κ. Indeed, the initial data $(\mathbf{E}, \mathbf{h})$ of any one of these products induce a JTS $\mathbf{L}$ on $\mathbf{E}$ (see Prop. 1), they determine this submanifold completely (see [12] Theorem 2), and $\mathbf{h}$ satisfies formula (18), see (14).

b) If we define $m := k - 1 + \sum \dim \mathbf{E}_i$ and $\bar{\kappa}$ as in (23) then (ii) is also equivalent to formula (27), i.e., the codimension $n - \sum \dim \mathbf{E}_i$ and the curvatures $\kappa_1, \ldots, \kappa_k$ have to be sufficient great for the existence of g.

Proof. First we assume (i) and prove (ii). The case $k = 1$ is treated in Prop. 2. Therefore we may suppose $k \geqq 2$.

1st Step. Using $\eta_2 \neq \eta_1$ we calculate

$$0 < \langle \eta_2 - \eta_1, \eta_2 - \eta_1 \rangle = \kappa_1 + \kappa_2 \leqq 2\kappa_2,$$

hence:

$$\kappa_i > 0 \quad \text{for} \quad i \geqq 2. \tag{19}$$

Next we define:

$$D_k(x_0, \ldots, x_k) := \det(x_0 + \delta_{ij} x_i)_{i,j=1,\ldots,k} \quad \text{for} \quad x_0, \ldots, x_k \in \mathbb{R}$$

and prove by induction

$$D_k(x_0, \ldots, x_k) = \sum_{i=0}^{k} \prod_{j \neq i} x_j.$$

Then we calculate the Gram determinants:

$$G' := \det(\langle \eta_i - \eta_1, \eta_j - \eta_1 \rangle)_{i,j \geqq 2} = D_{k-1}(\kappa_1, \ldots, \kappa_k) = \sum_{i=1}^{k} \prod_{j \neq i} \kappa_j \tag{20}$$

and

$$G := \det(\langle \eta_i, \eta_j \rangle)_{i,j=1,\ldots,k} = D_k(-\kappa, \kappa_1, \ldots, \kappa_k) = (\prod \kappa_i) - \kappa \cdot G'. \tag{21}$$

As important consequence we obtain $G' > 0$. Indeed, $G' = 0$ means the linear dependence of $\eta_2 - \eta_1, \ldots, \eta_k - \eta_1$, which would imply the linear dependence of $\eta_1, \ldots, \eta_k$, hence $G = 0$; using (19) and (21) we would obtain $\kappa_1 = 0$, hence $G' = \kappa_2 \ldots \kappa_k > 0$. – From $G' > 0$ we obtain immediately

$$\dim \operatorname{span}\{\eta_i | i = 1, \ldots, k\} \geqq k - 1. \tag{22}$$

If we define

$$\bar{\kappa} := \begin{cases} 0 & \text{if} \quad \kappa_1 = 0 \\ (\sum 1/\kappa_i)^{-1} & \text{if} \quad \kappa_1 \neq 0, \end{cases} \tag{23}$$

then we derive from (20) and (21)

$$\bar{\kappa} = (\prod \kappa_i)/G' = \kappa + G/G'.$$

Therefore (because of $G/G' \geqq 0$) we have

$$\bar{\kappa} \geqq \kappa. \tag{24}$$

Moreover, we see (in addition to (22)):

$$\bar{\kappa} = \kappa \Leftrightarrow \eta_1, \ldots, \eta_k \text{ are linearly dependent}, \tag{25}$$

and

$$\kappa_1 < 0 \Rightarrow \bar{\kappa} < 0. \tag{26}$$

If we put $m_i := \dim \mathbf{E}_i$ and $m := k - 1 + \sum m_i$, a combination of (22), (24), and (25) yields

$$(\bar{\kappa} = \kappa \text{ and } m \leqq n) \quad \text{or} \quad (\bar{\kappa} > \kappa \text{ and } m + 1 \leqq n). \tag{27}$$

2nd Step. Here we prove the existence of an extrinsic product

$$g: N^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k} \to N^n_\kappa$$

of spherical submanifolds with the following properties:

$$\kappa + \langle \eta_i, \eta_j \rangle = 0 \quad \text{for} \quad i \neq j, \tag{28}$$

and

$$\kappa + \langle \eta_i, \eta_i \rangle = \kappa_i \text{ for every } i, \tag{29}$$

where η_i denotes the mean curvature vector field of the component $g|L_i$ (notations as in Theorem 2):

According to Prop. 2 and (27) there exists a spherical imbedding $N^m_{\bar\kappa} \to N^n_\kappa$ (for $\bar\kappa = \kappa$ it is totally geodesic). Hence it suffices (see Prop. 3) to prove the existence of g for the case $n = m$ and $\kappa = \bar\kappa$.

a) If $\kappa_1 = 0$, the canonical imbedding $\mathbb{R}^{m_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k} \to \mathbb{R}^m$ has the desired properties.

b) For $\kappa_1 > 0$ the canonical imbedding $S^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k} \to \mathbb{R}^{m+1}$ is an extrinsic product of spherical submanifolds, which goes into $S^m_{\bar\kappa}$. In fact, this imbedding is also an extrinsic product in $S^m_{\bar\kappa}$.

c) For $\kappa_1 < 0$, see (26): According to the last case and Prop. 3 it suffices to treat the case $k = 2$. If we choose the half space model for the hyperbolic space (see Sect. 1), then

$$g: \mathbb{R}^{m_1 - 1} \times \mathbb{R}_+ \times S^{m_2}_{\kappa_2} \to \mathbb{R}^{m_1 - 1} \times \mathbb{R}^{m_2 + 1} \times \mathbb{R}_+$$

$$(x, y, z) \mapsto (\sqrt{\bar\kappa/\kappa_1} \cdot x, \sqrt{-\bar\kappa} \cdot yz, y)$$

describes an extrinsic product $H^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \to H^m_{\bar\kappa}$ with the desired properties.[3]

As for the formulas (28) and (29), we note that the curvature tensor of our product space is given by

$$R(x, y)z = \sum \kappa_i \cdot (\langle y_i, z_i \rangle x_i - \langle x_i, z_i \rangle y_i)$$

where $x_i, \ldots$ denotes the component of $x, \ldots$ with respect to the product structure. According to the Gauß equation [9] II and formula (14), the equation (28) is valid for every extrinsic product of spherical submanifolds. The same is true for formula (29) if $m_i \geqq 2$. In case $m_i = 1$, $\kappa_i > 0$, an elementary calculation proves (29). Only the case $m_i = 1$, $\kappa_i \leqq 0$ has to be verified for the special examples described in a) and c).

3rd Step. Last we construct an isometry F of N which moves g in the right position. For that let q be an arbitrary point of the product space. As the canonical foliations L_i of the product space are totally geodesic the mean curvature normals $\eta_i(q)$ of $g|L_i$ are orthogonal to $g_* T_q(\prod N^{m_i}_{\kappa_i}) = \oplus g_* T_q L_i$. As in addition $\langle \eta_i(q), \eta_j(q) \rangle = \langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_j \rangle$ for all i, j, there exists a linear isometry $\Phi: T_{g(q)} N \to T_p N$ such that $\Phi(g_* T_q L_i) = \mathbf{E}_i$ and $\Phi(\boldsymbol{\eta}_i(q)) = \boldsymbol{\eta}_i$ for all i. If F is the isometry of N with $F(q) = p$ and $F_*|_q = \Phi$, then $F \circ g$ is the desired extrinsic product; and (ii) is proved.

---

3 g has the two constant principal curvatures $\sqrt{\kappa_1 - \bar\kappa}$ and $\sqrt{\kappa_2 - \bar\kappa}$. Therefore g parameterizes an isoparametric hypersurface of $H^m_{\bar\kappa}$, namely a hyperbolic cylinder

The implication (ii) $\Rightarrow$ (iii) is an immediate consequence from Prop. 1. In order to deduce (i) from (iii), we derive

$$[\mathbf{S}(x, y), \mathbf{R}(v, w)]z = \mathbf{S}(\mathbf{S}(x, y)v, w)z - \mathbf{S}(\mathbf{S}(x, y)w, v)z \tag{30}$$

from (9); then we calculate

$$\mathbf{S}(x, y)z = \sum_{i,j} \kappa_{ij} \langle x_i, y_i \rangle z_j \quad \text{and} \quad \mathbf{R}(x, y)z = \sum_{i,j} \kappa_{ij} (\langle y_i, z_i \rangle x_j - \langle x_i, z_i \rangle y_j)$$

where we have put $\kappa_{ij} = \kappa + \langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_j \rangle$, see (4), (5); and then we substitute these expressions in (30). As the right side of (30) vanishes, we obtain:

$$\sum_{i,j,l} \kappa_{ij} (\kappa_{jl} - \kappa_{il}) \langle x_l, y_l \rangle (\langle w_i, z_i \rangle v_j - \langle v_i, z_i \rangle w_j) = 0,$$

hence:

$$\kappa_{ij}(\kappa_{jl} - \kappa_{il}) = 0 \text{ for all } i, j, l.$$

Using $\boldsymbol{\eta}_i \neq \boldsymbol{\eta}_j$ for $i \neq j$, we find

$$0 < \|\boldsymbol{\eta}_i - \boldsymbol{\eta}_j\|^2 = (\kappa_{ii} - \kappa_{ij}) + (\kappa_{jj} - \kappa_{ij}).$$

Therefore $\kappa_{ij} = 0$ for $i \neq j$, whereby (i) is proved. □

Corollary. *A symmetric submanifold S of a standard space N_κ^n has a flat normal bundle if and only if S is an extrinsic product*

$$g(N_{\kappa_1}^{m_1} \times S_{\kappa_2}^{m_2} \times \ldots \times S_{\kappa_k}^{m_k})$$

of spherical submanifolds as described in Theorem 2(ii).

Proof. According to Prop. 3 every extrinsic product of spherical submanifolds in N_κ^n has a flat normal bundle. Conversely, let S be a symmetric submanifold of N_κ^n with flat normal bundle $\nu(S)$ and choose a point $p \in M$. As consequence of the Ricci equation (see [4], p. 47) every shape operator A_ξ, $\xi \in \nu_p(S)$, of S commutes with each other. Therefore there exist distinct normal vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_k \in \nu_p(S)$ such that $T_pS = \oplus \mathbf{E}_i$ with

$$\mathbf{E}_i := \{v \in T_pS \mid A_\xi v = \langle \boldsymbol{\eta}_i, \xi \rangle v \text{ for all } \xi \in \nu_p(S)\}.$$

The second fundamental form $\mathbf{h} = h_p^S$ is then described by formula (18), and assertion (iii) of Theorem 2 is true (see Prop. 1). Hence there exists an extrinsic product $g: \prod N_{\kappa_i}^{m_i} \to N_\kappa^n$, as described in Theorem 2(ii). Indeed $S = g(\prod N_{\kappa_i}^{m_i})$, as follows immediately from [12] Theorem 2. □

In case $\kappa \geqq 0$ the Corollary is known (see [16], Satz 2, 3); our proof, however, has the advantage to work for arbitrary κ (also $\kappa < 0$) simultaneously. The results of Akiba [1] are also included.

4. The Main Result

Theorem 3. *Let p be a point of the standard space $N_\kappa^n =: N$, $\mathbf{E}$ a linear subspace of T_pN and $\mathbf{h}: \mathbf{E} \times \mathbf{E} \to \mathbf{E}^\perp$ a symmetric bilinear map. Then the following assertions are equivalent:*

(i) *There exists a symmetric submanifold S of N (see Def. 1) such that $p \in M$, $T_pS = \mathbf{E}$ and $\mathbf{h}$ is the second fundamental form of S at p.*

(ii) *The triple system $\mathbf{L}$ of $(\mathbf{E}, \mathbf{h})$ is a JTS (see Def. 2,3). Moreover, the symmetric submanifold S in* (i) *is unique.*

Proof. The implication (i) $\Rightarrow$ (ii) follows from Prop. 1. The uniqueness of S in (i) is proved in [12] Theorem 2. Therefore we suppose now (ii) and prove (i). For that put $\eta := \frac{1}{m}\sum \mathbf{h}(x_i, x_i)$, where $m := \dim \mathbf{E}$ and x_i runs over an orthonormal basis of $\mathbf{E}$, and let A_η be the self-adjoint operator of $\mathbf{E}$, defined by

$$\langle A_\eta x, y\rangle = \langle \mathbf{h}(x, y), \eta\rangle .$$

We use now the decomposition of $\mathbf{L}$ as described in Lemma 3. From Lemma 3(i) we derive $\mathbf{A} = m(\kappa \cdot id_{\mathbf{E}} + A_\eta)$. Therefore $\mathbf{E}_1, \ldots, \mathbf{E}_k$ are also the eigenspaces of A_η. If $x \in \mathbf{E}_i$, $y \in \mathbf{E}_j$ and $i \neq j$, then Lemma 3(iii) yields $\mathbf{S}(x, y) = 0$ and therefore (see (4))

$$\mathbf{h}(x, y) = 0 \quad \text{for} \quad x \in \mathbf{E}_i, y \in \mathbf{E}_j, \quad i \neq j . \tag{31}$$

Similarly we obtain

$$\langle \mathbf{h}(x, x), \mathbf{h}(y, y)\rangle = -\kappa \cdot \|x\|^2 \cdot \|y\|^2 \quad \text{for} \quad x \in \mathbf{E}_i, \quad y \in \mathbf{E}_j, \quad i \neq j . \tag{32}$$

Let us define $n_i := \dim \mathbf{E}_i$; $\boldsymbol{\eta}_i := \frac{1}{n_i}\sum \mathbf{h}(x_j, x_j)$ where x_j only runs over an orthonormal basis of $\mathbf{E}_i$, $\kappa_i := \kappa + \langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_i\rangle$, and

$$\tilde{\mathbf{E}}_i := \mathbf{E}_i \oplus \operatorname{span}\{\mathbf{h}(x, x) - \langle x, x\rangle \boldsymbol{\eta}_i | x \in \mathbf{E}_i\} .$$

Of course, we may suppose $\kappa_1 \leqq \kappa_2 \leqq \ldots \leqq \kappa_k$. – Then we have obviously

$$m \cdot \eta = \sum n_i \boldsymbol{\eta}_i , \tag{33}$$

and by means of (32)

$$\langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_j\rangle = -\kappa \quad \text{and} \quad \langle \mathbf{h}(x, x), \boldsymbol{\eta}_j\rangle = -\kappa \cdot \|x\|^2 \quad \text{for} \quad x \in \mathbf{E}_i, \quad i \neq j . \tag{34}$$

Furthermore, if λ_i is the eigenvalue of A_η corresponding to $\mathbf{E}_i$, we get

$$\langle \mathbf{h}(x, y), \eta\rangle = \lambda_i \langle x, y\rangle \quad \text{and} \quad \langle \boldsymbol{\eta}_i, \eta\rangle = \lambda_i \quad \text{for} \quad x, y \in \mathbf{E}_i , \tag{35}$$

in particular

$$\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_k \text{ are distinct vectors.} \tag{36}$$

Combining (33), (34), and (35) we derive

$$\langle \mathbf{h}(x, y), \boldsymbol{\eta}_i\rangle = \langle x, y\rangle \langle \boldsymbol{\eta}_i, \boldsymbol{\eta}_i\rangle , \tag{37}$$

and from (34) and (37) we obtain

$$\tilde{\mathbf{E}}_i \perp \tilde{\mathbf{E}}_j \quad \text{for} \quad i \neq j, \quad \text{and} \quad \boldsymbol{\eta}_j \perp \oplus \tilde{\mathbf{E}}_i .$$

Thus all hypotheses of Theorem 2 are satisfied; therefore $\kappa_2, \ldots, \kappa_k > 0$ and there

exists an extrinsic product

$$g: N^{m_1}_{\kappa_1} \times S^{m_2}_{\kappa_2} \times \ldots \times S^{m_k}_{\kappa_k} \to N$$

and a point $q \in \prod N^{m_i}_{\kappa_i}$ such that $g(q) = p$,

$$g_* T_q L_i = \tilde{\mathbf{E}}_i \text{ and } \boldsymbol{\eta}_i \text{ is the mean curvature normal of } g|L_i \text{ at } q, \tag{38}$$

where $L_1, \ldots, L_k$ denote the canonical foliations of the product space; to recall it, the leaf $L_i(q)$ of L_i through q is canonically isomorphic to $N^{m_i}_{\kappa_i}$.

We consider now the symmetric bilinear form

$$\mathbf{h}_i(x, y) := \mathbf{h}(x, y) - \langle x, y \rangle \boldsymbol{\eta}_i, \qquad x, y \in \mathbf{E}_i;$$

its values lie in $\mathbf{E}_i^{\perp} \cap \tilde{\mathbf{E}}_i$ (the orthogonal complement of $\mathbf{E}_i$ in $\tilde{\mathbf{E}}_i$). Therefore we can define the triple system $\mathbf{L}_i$ of $(\mathbf{E}_i, \mathbf{h}_i)$ with respect to $(\tilde{\mathbf{E}}_i, \kappa_i)$, see Def. 2. By means of (37) we see that $\mathbf{L}_i$ is the restriction of $\mathbf{L}$ to $\mathbf{E}_i \times \mathbf{E}_i$. According to Lemma 3(ii) $\mathbf{L}_i$ is therefore a JTS. Furthermore, the definition of $\boldsymbol{\eta}_i$ implies trivially that $\sum \mathbf{h}_i(x_j, x_j) = 0$ for any orthonormal basis (x_j) of $\mathbf{E}_i$. Hence, we are in the situation of Theorem 1, if we lift the initial data $(\mathbf{E}_i, \mathbf{h}_i)$ in $N^{m_i}_{\kappa_i}$ by means of the linear isometry $g_*: T_q L_i(q) \to \tilde{\mathbf{E}}_i$ and the canonical isometry $L_i(q) \to N^{m_i}_{\kappa_i}$. Therefore there exists a full minimal symmetric submanifold M_i of $N^{m_i}_{\kappa_i}$ which conversely produces a full minimal symmetric submanifold $\tilde{M}_i$ of the spherical submanifold $S_i := g(L_i(q))$.

From the construction we obtain

$$p \in \tilde{M}_i, \qquad T_p \tilde{M}_i = \mathbf{E}_i \subseteq \tilde{\mathbf{E}}_i = T_p S_i, \tag{39}$$

moreover, $\mathbf{h}_i$ is the second fundamental form of $\tilde{M}_i$ at p with respect to S_i and $\boldsymbol{\eta}_i$ the mean curvature normal of S_i at p (see (38)); therefore

$$\mathbf{h}|\mathbf{E}_i \times \mathbf{E}_i \text{ is the second fundamental form of } \tilde{M}_i \text{ at } p \text{ with respect to } N. \tag{40}$$

The last step enlightens completely the relation between our algebraic construction and the geometric situation. Formula (37), for instance, says that $\tilde{M}_i$ is pseudo-umbilical in N (see [4], p. 43).

To complete the proof, according to Lemma 2 $M_1 \times \ldots \times M_k$ is a (minimal) complete submanifold with parallel second fundamental form of the product $\prod N^{m_i}_{\kappa_i}$. Applying Prop. 3 we see that $S := g(M_1 \times \ldots \times M_k)$, therefore, is a symmetric submanifold of N with $p \in S$ and $T_p S = \oplus T_p \tilde{M}_i = \mathbf{E}$, see (39). Since the submanifolds $\tilde{M}_i$ are totally geodesic in S (as well as the M_i are totally geodesic in $M_1 \times \ldots \times M_k$) the second fundamental form h^S_p of S at p is given by $h^S_p(x, y) = \mathbf{h}(x, y)$, see (31) and (40). Thus the proof is complete. □

References

1. Akiba, S.: Submanifolds with flat normal connection and parallel second fundamental tensor. Sci. Rep. Yokohama Natl. Univ. Sect. I No. **23**, 7–14 (1976)
2. Backes, E.: Geometrie applications of euclidean Jordan triple systems. Manuscripta Math. **42**, 265–272 (1983)
3. Bourbaki, N.: Variétés différentielles et analytiques, Fascicule de résultats. Sects. 8–15. Paris: Hermann 1971
4. Chen, B.-Y.: Geometry of submanifolds. New York: Marcel Dekker 1973

5. Ferus, D.: Produkt-Zerlegung von Immersionen mit paralleler zweiter Fundamentalform. Math. Ann. **211**, 1–5 (1974)
6. Ferus, D.: Immersions with parallel second fundamental form. Math. Z. **140**, 87–93 (1974)
7. Ferus, D.: Symmetric submanifolds of euclidean space. Math. Ann. **247**, 81–93 (1980)
8. Kobayashi, S., Nagano, T.: On filtered Lie algebras and geometric structures. I. J. Math. Mech. **13**, 875–907 (1964)
9. Kobayashi, S., Nomizu, K.: Foundations of differential geometry I, II. New York: Interscience Publishers 1963/69
10. Kobayashi, S., Takeuchi, M.: Minimal imbeddings of R-spaces. J. Diff. Geometry **2**, 203–215 (1968)
11. Nagano, T.: Transformation groups on compact symmetric spaces. Trans. Am. Math. Soc. **118**, 428–453 (1965)
12. Reckziegel, H.: On the problem whether the image of a given differentiable map into a Riemannian manifold is contained in a submanifold with parallel second fundamental form. J. reine angew. Math. **325**, 87–103 (1981)
13. Reckziegel, H.: A class of distinguished isometric immersions with parallel second fundamental form. Resultate der Math. (to appear)
14. Strübing, W.: Symmetric submanifolds of Riemannian manifolds. Math. Ann. **245**, 37–44 (1979)
15. Takeuchi, M.: Parallel submanifolds of space forms. Manifolds and Lie Groups. Papers in Honour of Y. Matsushima, pp. 429–447. Basel: Birkhäuser 1981
16. Walden, R.: Untermannigfaltigkeiten mit paralleler zweiter Fundamentalform in euklidischen Räumen und Sphären. Manuscripta Math. **10**, 91–102 (1973)

Received October 18, 1982; in revised form January 13, 1983

Math. Ann. 263, 435–472 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Geodesic Flow on SO(4) and Abelian Surfaces

L. Haine*

Institut de Mathématique Pure et Appliquée, Université de Louvain, Chemin du Cyclotron, 2, B-1348 Louvain-la-Neuve, Belgium

Many classical Hamiltonian systems known to be completely integrable have rational integrals. In recent years many other such systems have been discovered. The algebraic nature of the integrals is usually responsible for a much richer structure than for an arbitrary completely integrable system: namely the invariant tori obtained from Liouville's theorem are part of algebraic complex tori (abelian varieties) with algebraic addition law consistent with the linearization of the motion. This fact lies at the origin of a fruitful interaction of the subject with the algebraic geometry of abelian varieties. For example, Knörrer [8] and Moser [13] have shown that the geodesic flow on a n-dimensional ellipsoid, a problem integrated by Jacobi, is linked to the results of Reid [16] on linear subspaces of complex intersections of two quadrics.

In this paper, we discuss another interesting interaction between algebraic geometry and dynamics: the Euler-Arnold equations associated to a class of geodesic flows on SO(4) for a left invariant diagonal metric. These equations take place in the 6-dimensional Lie algebra so(4) and are Hamiltonian on a 4-dimensional phase space (a Kostant-Kirillov orbit) defined by two "trivial" quadratic invariants $Q_0=c_0$ and $Q_1=c_1$. Under certain conditions on the metric (which are natural extensions of the case where geodesic motion on SO(4) corresponds to the free rotation of a 4-dimensional rigid body) Dikii [6] proved that this system is completely integrable with two other non trivial quadratic invariants $Q_2=c_2$ and $Q_3=c_3$, one of which is the Hamiltonian. In this paper we provide a careful study of the abelian variety $\tilde{A}_c$ which completes the complex affine variety $A_c=\bigcap_{i=0}^{3}\{Q_i=c_i\}\subseteq\mathbb{C}^6$, and this, making strong use of the differential equations. To be precise $A_c=\tilde{A}_c\backslash$some curve on $\tilde{A}_c$, where $\tilde{A}_c$ can be identified as the Prym variety of a certain double cover C of an elliptic curve C_0 [1]. The curve C_0

* Aspirant Fonds National Belge de la Recherche Scientifique (F.N.R.S.). Supported in part by N.S.F. Grant 8102696 while visiting Brandeis University

1 A Prym variety is an abelian variety constructed from a double cover $C\to C_0$ with the property that $\mathrm{Jac}(C)=\mathrm{Jac}(C_0)+\mathrm{Prym}(C/C_0)$. This concept is recalled in an appendix to this paper

arises naturally in the linear system of quadrics $Q_t = \sum_{i=0}^{3} t_i Q_i$, $t = [(t_0, t_1, t_2, t_3)] \in \mathbb{P}^3(\mathbb{C})$: namely this linear system contains a whole curve $C_0 \subseteq \mathbb{P}^3(\mathbb{C})$ of rank 3 quadrics[2]. This is quite exceptional since, in general, you expect only a finite number of points $t \in \mathbb{P}^3(\mathbb{C})$ such that rank $(Q_t) = 3$. Moreover the double covering $C \to C_0$ is ramified over $C_0 \cap \left\{ \sum_{i=0}^{3} t_i c_i = 0 \right\} = \{4 \text{ points}\}$ (c_i the four constants of the motion, i.e. $Q_i = c_i$) which makes the genus of C equal to 3. This is the content of Sects. 2, 3, 4.

Now in [2] Adler and van Moerbeke linearize the Euler-Arnold equations (for the class of metrics considered in this paper) on a 2-dimensional Prym variety $\mathrm{Prym}(K/K_0)$ of another genus 3 curve K over an elliptic curve K_0. This comes from the fact that, as was first observed by Manakov [11], under the time evolution of this system, for each $h \in \mathbb{C}$, the spectrum of a certain matrix $X(h) = X + ah$ [$X \in \mathrm{so}(4)$, a some constant diagonal matrix related to the metric] remains fixed. In recent years many completely integrable systems have been related to similar isospectral deformations. This procedure provides not only the algebraic integrals of the system [as eigenvalues of $X(h)$] but also, in many examples, the flow can be realized as a straight line motion on the Jacobi variety of the curve K in (h, z) space K: $\det(X(h) - zI) \equiv P(z, h) = 0$ (or some subabelian variety of it). In this case the flow is linearized on $\mathrm{Prym}(K/K_0)$ with $K_0 = K/\iota$ the quotient of K by the involution $\iota : K \to K : (z, h) \curvearrowright (-z, -h)$. We recall this result in Sect. 1. We point out that this method merely proves that a certain quotient of the variety $A_c = \bigcap_{i=0}^{3} \{Q_i = c_i\} \subseteq \mathbb{C}^6$ by a finite group of automorphisms G is the affine part of $\mathrm{Prym}(K/K_0)$ and, as such, it does not characterize the geometrical object formed by the intersection of the four quadratic invariants of the flow. Indeed, it turns out that the two abelian surfaces $\mathrm{Prym}(C/C_0)$ and $\mathrm{Prym}(K/K_0)$ are isomorphic only up to multiplication of some *(not all)* periods by 2. However, these two abelian surfaces are intimately related in the precise sense that they are dual of each other. This will be explained in Sect. 5.

To prove our result we use a method suggested by Kowalewski [9, 10] in her study of the integrable cases among the rigid body motions in the presence of gravity. This method is explained at the beginning of Sect. 3 and has been discussed and used more systematically in two recent papers by Adler and van Moerbeke [3, 4]. It is based on the idea that if a completely integrable system can be linearized on an abelian variety then, for fixed values of the constants of the motion, the phase variables are meromorphic (abelian) functions on a compact complex torus which forces them to blow up along a codimension one divisor. Hence the solutions of the system admit Laurent series expansions in $t \in \mathbb{C}$ near this divisor; since the initial conditions are not specified these expansions depend on the constants of the motion and the point where the trajectory hits the divisor ($= 5$ free parameters in our example). In [3] and [4] the authors exploit Kowalewski's method as a criterion to derive *necessary conditions* for a given Hamiltonian system to be *algebraically completely integrable*, which means that the system can

2 This property of Q_t was discovered by Adler and van Moerbeke

be linearized on an abelian variety. *The sufficiency of this criterion has never been established in general.* In this paper we show the sufficiency of these conditions on the family of Hamiltonian systems formed by the Euler-Arnold equations associated to a class of integrable geodesic flows on SO(4). The main point is that the knowledge of the Laurent series solutions, which is a problem of linear algebra, provides a systematic method to build in an explicit way the abelian surface $\tilde{A}_c$ at infinity, the free parameters playing the role of regularizing variables when the equations blow up. This explicit information is then employed to identify $\tilde{A}_c$ as a Prym variety (Sect. 4) and to understand its relation with the Prym variety coming from the method of isospectral deformations (Sect. 5).

Mumford [5] has given an abstract algebro-geometrical proof that these four quadrics intersect in the affine part of an abelian surface using Enriques classification of algebraic surfaces.

Table of Contents

1. Euler-Arnold Equations 437
A. Euler's Rigid Body Motion 438
B. The Case $n=4$ by the Method of Isospectral Deformations 440
2. The Curve of Rank 3 Quadrics 446
3. Asymptotic Expansions 448
4. The Abelian Surface $\tilde{A}_c$ as a Prym Variety 456
5. The Relation Between Prym(C/C_0) and Prym(K/K_0) 466
Appendix 468
References 471

1. Euler-Arnold Equations

Consider the group SO(n) and its Lie algebra so(n) paired to itself by the customary inner product

$$K(X, Y) = -\tfrac{1}{2}\mathrm{Tr}(X \cdot Y) \quad \text{where} \quad X = (X_{ij}) \equiv \sum_{i<j} X_{ij} e_{ij} \in \mathrm{so}(n).$$

A left invariant metric on SO(n) is defined by a non-singular symmetric linear map $\lambda : \mathrm{so}(n) \to \mathrm{so}(n) : X \curvearrowright \lambda(X)$ and by the following inner product: given two vectors gX and gY in the tangent space to SO(n) at the point $g \in \mathrm{SO}(n)$

$$\langle gX, gY \rangle = K(X, \lambda^{-1}(Y))$$

regardless of g. Consider geodesic motion on SO(n) with regard to this metric and define the element $X = \lambda^{-1}(g^{-1}\dot{g})$ in the Lie algebra so(n). Then the Euler-Arnold equations take the form

$$\dot{X} = [X, \lambda(X)]. \tag{1.1}$$

It is known that (1.1) is a Hamiltonian system with regard to the Kostant-Kirillov symplectic structure induced on each orbit O formed by the coadjoint action of the group SO(n) on the dual Lie algebra so$(n)^* \approx$ so(n):

$$O = \{\mathrm{Ad}^*(g)(X) = g^{-1} X g \mid g \in \mathrm{SO}(n)\}. \tag{1.2}$$

For $\xi_1=[X,Z_1]$ and $\xi_2=[X,Z_2]$ (with $Z_1,Z_2\in\mathrm{so}(n)$) two tangent vectors to the orbit at the point X, the symplectic structure is defined by

$$\Omega(X)(\xi_1,\xi_2)=K(X,[Z_1,Z_2]). \tag{1.3}$$

Functions H defined on O lead to Hamiltonian vector fields

$$\dot{X}=[X,\nabla H(X)]. \tag{1.4}$$

In particular $H=\frac{1}{2}K(X,\lambda(X))$ leads to (1.1).

In this paper, we are concerned with the case $n=4$ for these equations and the problem of solving them in terms of abelian functions under certain conditions on the metric λ to be specified later. As motivation, let me first recall what this means in the case $n=3$.

A) Euler's Rigid Body Motion

For $n=3$, if we set

$$X=\begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}, \quad \lambda(X)=\begin{pmatrix} 0 & -\lambda_3 x_3 & \lambda_2 x_2 \\ \lambda_3 x_3 & 0 & -\lambda_1 x_1 \\ -\lambda_2 x_2 & \lambda_1 x_1 & 0 \end{pmatrix}$$

(1.1) reduce to the classical Euler's equations of the free rotation of a rigid body namely,

$$\begin{aligned} \dot{x}_1&=(\lambda_3-\lambda_2)x_2x_3 \\ \dot{x}_2&=(\lambda_1-\lambda_3)x_1x_3 \\ \dot{x}_3&=(\lambda_2-\lambda_1)x_1x_2 \end{aligned} \tag{1.5}$$

with (x_1,x_2,x_3) the angular momentum in body coordinates and $\lambda_i=1/I_i$ the inverse of the principal moments of inertia ($\lambda_1<\lambda_2<\lambda_3$). They have two invariants

$$\begin{aligned} Q_0(X)&\equiv x_1^2+x_2^2+x_3^2=c_0 \\ Q_1(X)&\equiv \lambda_1x_1^2+\lambda_2x_2^2+\lambda_3x_3^2=c_1. \end{aligned} \tag{1.6}$$

On each sphere $Q_0(X)=c_0$, (1.5) is a Hamiltonian system with Hamiltonian $H=\frac{1}{2}Q_1(X)$ and as such it is a completely integrable system. Indeed, the sphere $Q_0(X)=c_0$ intersects the ellipsoid $Q_1(X)=c_1$ in two "circles" (at least if $c_1/\lambda_3<c_0<c_1/\lambda_1$, otherwise this intersection is empty). But, as it is well known, there is a much richer structure involved in the problem. Namely, the complex intersection A_c of this sphere and this ellipsoid

$$A_c=\{Q_0(X)=c_0\}\cap\{Q_1(X)=c_1\}\subseteq\mathbb{C}^3$$

is the affine part of an elliptic curve $\bar{A}_c\subseteq\mathbb{P}^3(\mathbb{C})$ and if $p(t)=(x_1(t),x_2(t),x_3(t))$ is a solution of (1.5), the law connecting $p(t_1+t_2)$ to $p(t_1)$ and $p(t_2)$ is the addition law on the elliptic curve. This is readily seen as follows. From the Eqs. (1.5), the unique holomorphic differential on $\bar{A}_c$ is given by

$$\omega=\frac{dx_1}{(\lambda_3-\lambda_2)x_2x_3}=\frac{dx_2}{(\lambda_1-\lambda_3)x_1x_3}=\frac{dx_3}{(\lambda_2-\lambda_1)x_1x_2} \tag{1.7}$$

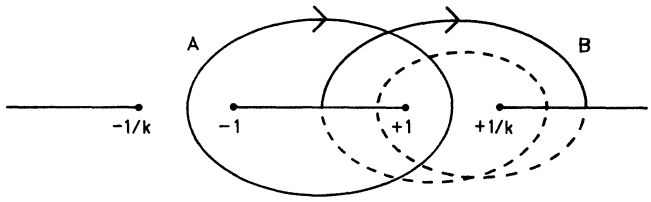


Fig. 1

leading to

$$t = \int_{p_0}^{p(t)} \omega$$

for a suitable choice of the origin $p_0 \in \bar{A}_c$; also, for fixed values of the constants of the motion, the coordinates x_i are elliptic functions on a complex torus $\mathbb{C}/L_\Omega$. In fact, the classical way of solving Euler's equations in terms of elliptic functions can be understood as a characterization of this complex torus. Since the elliptic curve $\bar{A}_c$ will play an important role in what follows, let me recall this classical result in a somewhat unusual form. Assume for example $\lambda_2 c_0 < c_1$, $\bar{A}_c$ is then characterized as follows:

Theorem 1. *The elliptic curve $\bar{A}_c \subseteq \mathbb{P}^3(\mathbb{C})$ is a 4-sheeted unramified covering of the elliptic plane curve*

$$K: w^2 = z(1-z)(1-k^2 z)$$

with $0 < k^2 = \dfrac{(\lambda_2 - \lambda_1)(\lambda_3 c_0 - c_1)}{(\lambda_3 - \lambda_2)(c_1 - \lambda_1 c_0)} < 1$. *Moreover, $\bar{A}_c$ and K have proportional period lattices, so that they are in fact isomorphic.*

Proof. Solving (1.6) for x_1 and x_3, one finds

$$x_1 = \pm \left[\frac{(\lambda_3 c_0 - c_1)(1-s^2)}{(\lambda_3 - \lambda_1)}\right]^{1/2} \quad \text{and} \quad x_3 = \pm \left[\frac{(c_1 - \lambda_1 c_0)(1-k^2 s^2)}{(\lambda_3 - \lambda_1)}\right]^{1/2} \tag{1.8}$$

with

$$s = \left(\frac{\lambda_3 - \lambda_2}{\lambda_3 c_0 - c_1}\right)^{1/2} x_2 \quad \text{and} \quad 0 < k^2 = \frac{(\lambda_2 - \lambda_1)(\lambda_3 c_0 - c_1)}{(\lambda_3 - \lambda_2)(c_1 - \lambda_1 c_0)} < 1.$$

Because of (1.8) we may think of $\bar{A}_c$ as a 4-sheeted ramified covering of the s-plane with two simple branch points over each of the four points $s = -k^{-1} < -1 < 1 < k^{-1}$ and homology cycles A and B as in Fig. 1. In the local parameter s, up to some constant factor, the holomorphic differential ω (1.7) takes the form

$$\omega = \pm \frac{ds}{[(1-s^2)(1-k^2 s^2)]^{1/2}}.$$

From Fig. 1, it is clear that

$$\int_A \omega = 4 \int_0^1 \frac{ds}{[(1-s^2)(1-k^2 s^2)]^{1/2}} \quad \text{and} \quad \int_B \omega = 4 \int_1^{1/k} \frac{ds}{[(1-s^2)(1-k^2 s^2)]^{1/2}}.$$

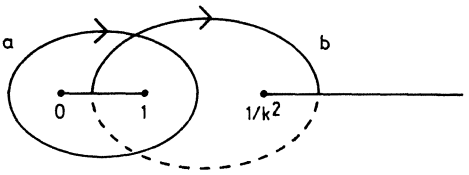


Fig. 2

On the other hand, the curve $\bar{A}_c$ can also be viewed as a 4-sheeted unramified covering of the elliptic plane curve

$$K: w^2 = z(1-z)(1-k^2 z)$$

with covering projection defined by

$$\Pi: A_c \to K: (x_1, x_2, x_3) \curvearrowright (w, z)$$
$$z \equiv \left(\frac{\lambda_3 - \lambda_2}{\lambda_3 c_0 - c_1}\right) x_2^2 \quad \text{and} \quad w \equiv \left(\frac{\lambda_3 - \lambda_1}{\lambda_3 c_0 - c_1}\right)\left(\frac{\lambda_3 - \lambda_2}{c_1 - \lambda_1 c_0}\right)^{1/2} x_1 x_2 x_3 . \tag{1.9}$$

Let θ be the unique holomorphic differential on K such that the pullback $\Pi^*(\theta) = \omega$

$$\theta = \frac{1}{2} \frac{dz}{[z(1-z)(1-k^2 z)]^{1/2}}$$

and a, b the fundamental cycles of K as in Fig. 2. Making the change of variables $z = s^2$, one has obviously

$$\int_A \omega = 2 \int_a \theta \quad \text{and} \quad \int_B \omega = 2 \int_b \theta .$$

Let us call I_ω (resp. I_θ) the uniformizing map

$$I_\omega: \bar{A}_c \to \mathbb{C}/L_\Omega : p \curvearrowright \int_{p_0}^{p} \omega$$
$$\left(\text{resp. } I_\theta: K \to \mathbb{C}/L_\Delta : \Pi(p) \curvearrowright \int_{\Pi(p_0)}^{\Pi(p)} \theta\right)$$

it is now clear that we have the commutative diagram

$$\begin{array}{ccccc} \bar{A}_c & \xrightarrow{I_\omega} & \mathbb{C}/L_\Omega & \longleftarrow & \mathbb{C} \\ \Big\downarrow \Pi & & \Big\downarrow & & \Big\| \\ K & \xrightarrow{I_\theta} & \mathbb{C}/L_\Delta & \longleftarrow & \mathbb{C} \end{array} \tag{1.10}$$

with $L_\Omega = 2L_\Delta$. Since the elliptic curves $\bar{A}_c$ and K have proportional period lattices, they are isomorphic. This establishes Theorem 1.

B) The Case $n=4$ by the Method of Isospectral Deformations

Consider now the case $n=4$ of (1.1). A skew-symmetric matrix $X \in \mathrm{so}(4)$ will be written as

$$X=(X_{ij})_{1\leqq i,\,j\leqq 4}\equiv\sum_1^6 x_i e_i\equiv\begin{pmatrix}0 & -x_3 & x_2 & -x_4\\ x_3 & 0 & -x_1 & -x_5\\ -x_2 & x_1 & 0 & -x_6\\ x_4 & x_5 & x_6 & 0\end{pmatrix}.$$

We shall restrict ourselves to the case where λ is diagonal in the basis (e_i) i.e.

$$\lambda(X)=(\lambda_{ij}X_{ij})_{1\leqq i,\,j\leqq 4}=\sum_1^6\lambda_i x_i e_i.$$

The phase space (1.2) is now defined by the two "trivial" invariants

$$\begin{aligned}Q_0(X)&\equiv(\det X)^{1/2}=x_1x_4+x_2x_5+x_3x_6\\ Q_1(X)&\equiv-\tfrac{1}{2}\operatorname{Tr}X^2=\sum_{i=1}^6 x_i^2.\end{aligned}\tag{1.11}$$

$Q_0(X)$ and $Q_1(X)$ are called trivial invariants because they give rise to identically zero Hamiltonian vector fields with regard to the symplectic structure (1.3). Since the phase space is 4-dimensional and the only non trivial invariant that we know is the Hamiltonian

$$H=\tfrac{1}{2}K(X,\lambda(X))=\frac{1}{2}\sum_{i=1}^6\lambda_i x_i^2$$

we need one extra independent invariant to make the system completely integrable. The first step towards the complete integrability of (1.1) was done by Manakov [11] who observed that if

$$\begin{aligned}&\lambda_{ij}=\frac{b_i-b_j}{a_i-a_j}\Leftrightarrow[X,b]+[a,\lambda(X)]=0,\ \forall X\\ &\text{with}\quad a=\operatorname{diag}(a_1,\ldots,a_n),\qquad b=\operatorname{diag}(b_1,\ldots,b_n)\qquad(n=4)\end{aligned}\tag{1.12}$$

(1.1) is equivalent to

$$(X+ah)^{\cdot}=[X+ah,\lambda(X)+bh]\tag{1.13}$$

with a formal indeterminate h. Equation (1.13) means that, for each $h\in\mathbb{C}$, the spectrum of the matrix $X+ah$ is time independent or, what amounts to the same information, if we form the characteristic polynomial

$$\det(X+ah-zI)=\prod_1^4(z-a_ih)+Q_1(X)z^2-Q_2(X)zh+Q_3(X)h^2+Q_0^2(X)$$

the $Q_i(X)$, $0\leqq i\leqq 3$, are constants of the motion. $Q_0(X)$ and $Q_1(X)$ are the trivial invariants defined in (1.11) while $Q_2(X)$ and $Q_3(X)$ are two independent non trivial algebraic invariants given by

$$\begin{aligned}Q_2(X)\equiv&(a_1+a_4)x_1^2+(a_2+a_4)x_2^2+(a_3+a_4)x_3^2\\ &+(a_2+a_3)x_4^2+(a_1+a_3)x_5^2+(a_1+a_2)x_6^2\\ Q_3(X)\equiv&\,a_1a_4x_1^2+a_2a_4x_2^2+a_3a_4x_3^2+a_2a_3x_4^2+a_1a_3x_5^2+a_1a_2x_6^2.\end{aligned}\tag{1.14}$$

Remark. Given a rigid body in $\mathbb{R}^n$, there exists an orthonormal basis in $\mathbb{R}^n$, completely determined by the mass-distribution of the body, in which the equations of the motion of the free rigid body about a fixed point have the form $\dot{X}=[X,\Omega]$ where $X,\Omega\in so(n)$, $X=\Omega J+J\Omega$, for J a diagonal matrix $J=\mathrm{diag}(b_1,\ldots,b_n)$ satisfying $b_i+b_j>0$ for $i\neq j$. The free rotation of a n-dimensional rigid body corresponds thus to the case $a=b^2$ of (1.12). Observe that, when $n=3$, (1.12) is not a restriction on the metric and can always be fulfilled with $b=\frac{1}{2}(I_1+I_2+I_3)\,\mathrm{id}-\mathrm{diag}(I_1,I_2,I_3)$ and $a=b^2$ (recall that $\lambda_i=1/I_i$, with I_i the principal moments of inertia). When $n\geqq 4$, (1.12) is a restriction on the metric and Manakov's trick (1.13) leads in fact to the right number ($=\frac{1}{2}\dim$ phase) space) of commuting independent integrals for any n [12, 15]. However the case $n=4$ is distinguished by the fact that it is the only case where all the algebraic invariants are quadratic.

In [2], the authors use the form (1.13) of the equations to solve the problem by quadrature as a special case of a general theory. We recall here their result in a somewhat improved form. Consider the algebraic curve

$$K=\{(z,h)\in\mathbb{C}^2 \mid \det(X+ah-zI)=0 \text{ with } Q_i(X)=c_i,\ 0\leqq i\leqq 3\}.$$

For generic choice of the c_i, K is an algebraic plane curve of degree 4. Its genus is thus 3. In the coordinates at infinity ($v=1/h$, $u=z/h$), K is written as

$$K: c_0^2v^4+P(u)v^2+\prod_{i=1}^{4}(u-a_i)=0 \tag{1.15}$$

with $P(u)=c_1u^2-c_2u+c_3$. Moreover, the map $\iota: K\to K:(v,u)\rightsquigarrow(-v,u)$ is an involution on K giving rise to a 2-dimensional Prym variety $\mathrm{Prym}_\iota(K)$[3] in the Jacobi variety of the curve K. The quotient $K_0=K/\iota$ by the involution ι, is an elliptic curve defined by

$$K_0=K/\iota: t^2=P^2(u)-4c_0^2\prod_{i=1}^{4}(u-a_i).$$

Solving (1.15) for v^2, the curve K appears as a 2-sheeted ramified covering of K_0

$$K:\begin{cases}2c_0^2v^2+P(u)-t=0\\ t^2=P^2(u)-4c_0^2\prod_{i=1}^{4}(u-a_i)\end{cases}$$

with 4 simple branch points $p_i\equiv(v=0,\ t=P(a_i),\ u=a_i)$, $1\leqq i\leqq 4$, and 4 points at infinity q_i, $1\leqq i\leqq 4$, satisfying $q_1^\iota=q_3$ and $q_2^\iota=q_4$. Clearly the divisor structure of u and v is

$$(v)=-\sum_1^4 q_i+\sum_1^4 p_i \quad\text{and}\quad (u)=-\sum_1^4 q_i+4\ \text{zeroes}.$$

Let θ, ψ_1, ψ_2 be a basis of holomorphic differentials on K such that $\iota^*\theta=\theta$ and $\iota^*\psi_j=-\psi_j$ $(j=1,2)$. Let A_c be the complex affine variety defined by the in-

3 For the definition of a Prym variety, see appendix

tersection of the constants of the motion

$$A_c = \bigcap_{i=0}^{3} \{Q_i(X) = c_i\} \subseteq \mathbb{C}^6 \tag{1.16}$$

we have then the following

Theorem 2. *The Euler-Arnold equations with λ satisfying* (1.12) *can be linearized on* $\mathrm{Prym}_t(K)$. *More precisely, there is a* $4-1$ *unramified algebraic map Π from the affine variety A_c onto a Zariski open set $\mathscr{U} \subset \mathrm{Prym}_t(K)$*

$$\Pi : A_c \xrightarrow{-} \mathscr{U} \subset \mathrm{Prym}_t(K) : X \in A_c \rightsquigarrow D = \sum_{i=1}^{3} \mu_i \in \mathscr{U}$$

such that any smooth Hamiltonian $H(Q_0, Q_1, Q_2, Q_3)$ leads to a linear flow on $\mathrm{Prym}_t(K)$ *i.e.*

$$\frac{d}{dt} \sum_{i=1}^{3} \int_{0_i}^{\mu_i(t)} (\theta, \psi_1, \psi_2) = (0, k_1, k_2).$$

Moreover,

$$\mathrm{Prym}_t(K) \backslash \mathscr{U} = \Theta \cap \mathrm{Prym}_t(K)$$

with Θ some translate of the Θ-divisor in $\mathrm{Jac}(K)$.

Finally, the functions x_i^2 $(1 \leqq i \leqq 6)$, $x_1x_2x_3$, $x_3x_4x_5$, $x_2x_4x_6$, $x_1x_5x_6$, $x_4x_5x_6$, $x_1x_2x_6$, $x_1x_3x_5$, and $x_2x_3x_4$ restricted to A_c are abelian functions on $\mathrm{Prym}_t(K)$.

Remarks. 1. Theorem 2 extends word for word for arbitrary n, provided the metric λ satisfies (1.12) (see [2], p. 337–346). The precise characterization of $\mathrm{Prym}_t(K) \backslash \Pi(A_c)$ is new.

2. When $n=3$, the curve $K : \det(X + ah - zI) = 0$ is an elliptic plane curve which, in the coordinates at infinity $(v = 1/h,\ u = z/h)$, is written as $K : (c_0u - c_1)v^2 - \prod_{i=1}^{3} (u - a_i) = 0$. Obviously $\mathrm{Prym}_t(K) = K$. Moreover, after the fractional linear transformation

$$z = \left(\frac{a_3 - a_2}{a_3c_0 - c_1}\right)\left(\frac{c_0u - c_1}{u - a_2}\right).$$

K becomes the elliptic plane curve of Theorem 1 $K : w^2 = z(1-z)(1-k^2z)$, with λ_i replaced by a_i. The algebraic map Π, of which it is question in Theorem 2, is then nothing but the projection map $\Pi : A_c \to K$ defined in (1.9). Remember also that x_i^2 and $x_1x_2x_3$ are elliptic functions on K, but not the x_i which are elliptic functions only on $\bar{A}_c$.

3. From the proof of Theorem 2, it will become clear that, given a divisor $D \in \mathscr{U}$, the matrix $X \in A_c$ can be recovered up to a finite free group action $G = \{1, \sigma_1, \sigma_2, \sigma_3\}$ on A_c, each σ_i, $1 \leqq i \leqq 3$, corresponding to some sign changes on the x_k, $1 \leqq k \leqq 6$, (see formula 1.17 below). The upshot is that Theorem 2 only proves that the affine variety A_c/G is the affine part of the abelian variety $\mathrm{Prym}_t(K)$. As a consequence, any G-invariant rational function on A_c is a rational function on A_c/G and thus an abelian function on $\mathrm{Prym}_t(K)$. In particular, the functions x_i^2 restricted to A_c/G are abelian functions on $\mathrm{Prym}_t(K)$, but not the x_i themselves.

Now, suppose for a moment we have proved that A_c itself is the affine part of an abelian surface $\tilde{A}_c$, the addition law $\oplus$ on $\tilde{A}_c$ being defined by the two commuting everywhere independent vector fields X_{Q_2} and X_{Q_3}, with Q_2 and Q_3 the two non trivial invariants defined in (1.14). Then, using (1.17), it is easy to check that X_{Q_2} and X_{Q_3} are G-invariant, which means that each σ_i, $1 \leqq i \leqq 3$, is in fact a translation $\sigma_i : \tilde{A}_c \to \tilde{A}_c : p \curvearrowright p \oplus a_i$ on $\tilde{A}_c$. Moreover, since $\sigma_i^2 = 1$, $1 \leqq i \leqq 3$, $2a_i = 0$ and $\mathrm{Prym}_t(K)$ is thus obtained from $\tilde{A}_c$ by dividing two of the four periods of $\tilde{A}_c$ by 2. In particular, this indicates that $\tilde{A}_c$ cannot be isomorphic to $\mathrm{Prym}_t(K)$. The precise relation between $\tilde{A}_c$ and $\mathrm{Prym}_t(K)$ will be discussed in Theorem 5 of Sect. 5.

Sketch of the Proof of Theorem 2. Start with an $X \in A_c$. For each $(z, h) \in K$, solve the spectral problem

$$(X + ah) f = zf \Leftrightarrow (Xv + a) f = uf$$

with $f = (f_1, f_2, f_3, f_4)^T$ normalized at $f_1 = 1$. The f_k are defined by

$$f_k = \frac{\Delta_{1k}}{\Delta_{11}}$$

where Δ_{ij} stands for the minor of $(Xv + a - uI)$ corresponding to the i,j^{th} entry. Each f_k can be regarded as a meromorphic function on K. One sees easily that

$$(\Delta_{1k}) = -3 \sum_1^4 q_i + 2 \sum_1^4 p_i - p_1 - p_k + E_k$$

with E_k a positive divisor of degree 6, so that

$$(f_k) = p_1 - p_k + E_k - E_1 .$$

Let D be the minimal positive divisor such that

$$(f_k) + D \geqq p_1 - p_k .$$

One can show[4] that D is a general divisor [i.e. $\dim \mathscr{L}(D) = 1$] of degree 3. Since

$$\Delta_{11}(p_1) = \prod_{i=2}^{3} (a_1 - a_i) \neq 0$$

$p_1 \notin D$. Moreover, because of the antisymmetry of X, $E_1 = D + D'$. Hence,

$$\left(\frac{\Delta_{11}}{v^2} \right) = - \sum_1^4 q_i - 2p_1 + D + D' .$$

By Abel's theorem, this implies

$$\int_0^D \omega + \int_{0'}^{D'} \omega = 0 \text{ (mod periods)}$$

with $\omega = (\theta, \psi_1, \psi_2)$ and $0 = p_1 + q_1 + q_2$. This shows that $\Pi(X) \equiv D \in \mathrm{Prym}_t(K)$.

4 In this example, this can be done "by hand"

Conversely, let D be a general divisor of degree 3 not containing p_1 and which moreover belongs to $\mathrm{Prym}_t(K)$. Then

$$1 \leqq \dim \mathscr{L}(D-p_1+p_k) \leqq \dim \mathscr{L}(D-p_1)+1=1$$

the first inequality is a consequence of Riemann-Roch, the second a consequence of the fact that adding a pole increases the dimension by at most one and the third (equality) a consequence of the fact that $\dim \mathscr{L}(D)=1$ and $p_1 \notin D$. Define $f_1=1$ and f_k $(k \neq 1)$ the unique function (up to multiplication by a non zero constant) which belongs to $\mathscr{L}(D-p_1+p_k)$. Because D is general and $p_1 \notin D$, f_k $(k \neq 1)$ must have a pole of exact order $\mathrm{ord}_{p_k}(D+p_k)$ at p_k. Since

$$(z-a_k h)=\left(\frac{u-a_k}{v}\right)=p_k+\mu_k+\mu_k^{\prime}-\sum_{i \neq k} p_i$$

one has

$$(zf_k-a_k h f_k) \geqq -D-\sum_{i \neq 1,k} p_i .$$

From the expression in brackets, substract some multiple of f_i $(i \neq 1, k)$ if it has a pole at p_i $(i \neq 1, k)$. Proceeding in this way and using again the fact that D is general, zf_k is seen to be expressible as a linear combination

$$zf_k=a_k h f_k+\sum_{i \neq k} X_{ki} f_i .$$

Clearly the f_k $(k \neq 1)$ can be renormalized so as to have $X_{1i}+X_{i1}=0$. Using the fact that $D \in \mathrm{Prym}_t(K)$ one shows then that the matrix $X=(X_{ij})$ must be skew-symmetric. Since the f_k $(k \neq 1)$ can still be multiplied by ± 1, this shows that the matrix X is determined up to 8 sign changes. However, only 4 of them lead to $x_1 x_4+x_2 x_5+x_3 x_6=c_0$, namely

$$\begin{array}{cccccc} x_1 & x_2 & x_3 & x_4 & x_5 & x_6 \\ + & + & + & + & + & + \\ - & - & + & - & - & + \\ - & + & - & - & + & - \\ + & - & - & + & - & - \end{array} . \tag{1.17}$$

The upshot is that we have constructed a 4-1 surjective algebraic map

$$\Pi : A_c \xrightarrow{-} \mathscr{U}=\{D \in \mathrm{Prym}_t(K) \mid D \text{ is general, } \deg D=3 \text{ and } p_1 \notin D\} .$$

Also, it is easy to check that Π is unramified for generic choice of the c_i, $0 \leqq i \leqq 3$.

Now, by Riemann's theorem (see [7], p. 338), the set of divisors $D=\sum_{i=1}^{3} \mu_i \in \mathrm{Jac}(K)$, considered modulo linear equivalence, such that D is equivalent to $p_1+v_2+v_3$ (p_1 fixed) is a translate of the theta-divisor of $\mathrm{Jac}(K)$. Let us call this set Θ. It is equivalent to say that Θ^c is the set of divisors $D=\sum_{i=1}^{3} \mu_i \in \mathrm{Jac}(K)$

such that D is general and $p_1 \notin D$. This shows that

$$\mathrm{Prym}_t(K) \backslash \mathscr{U} = \Theta \cap \mathrm{Prym}_t(K).$$

Moreover, any smooth function of the Hamiltonians $Q_i(X)$, $0 \leqq i \leqq 3$, leads to a linear flow on $\mathrm{Prym}_t(K)$ [2]. This finishes the proof of Theorem 2.

2. The Curve of Rank 3 Quadrics

In this section we explain a remarkable property of the linear system

$$\left\{ Q_t(X) = \sum_{i=0}^{3} t_i Q_i(X) | t = [(t_0, t_1, t_2, t_3)] \in \mathbb{P}^3(\mathbb{C}) \right\}$$

generated by the 4 quadrics $Q_i(X)$ $(0 \leqq i \leqq 3)$ defined in (1.11) and (1.14). Namely this linear system contains a whole curve of rank 3 quadrics. As explained in the introduction, this is an exceptional situation. The precise statement is the following. Let Q_i be the matrix of the quadratic form $Q_i(X)$. Consider the discriminant surface

$$\Delta = \left\{ [(t_0, t_1, t_2, t_3)] \in \mathbb{P}^3(\mathbb{C}) | \det\left(\sum_{i=0}^{3} t_i Q_i \right) = 0 \right\}. \tag{2.1}$$

Proposition 2.1 (Adler-van Moerbeke [4]). *Δ is a union of 3 quadratic cones which intersect in an elliptic curve C_0 of degree 4. For each point $t = [(t_0, t_1, t_2, t_3)] \in C_0$, the quadric Q_t has rank 3.*

Proof. Let us write for a moment the two quadrics $Q_2(X)$ and $Q_3(X)$ as

$$Q_2(X) = \sum_{i=1}^{6} \mu_i x_i^2 \quad \text{and} \quad Q_3(X) = \sum_{i=1}^{6} \nu_i x_i^2.$$

After elementary operations on the rows and the columns of the matrix $\sum_{i=0}^{3} t_i Q_i$, one finds

$$\det\left(\sum_{i=0}^{3} t_i Q_i \right) = \prod_{j=1}^{3} \left[(t_1 + t_2 \mu_j + t_3 \nu_j)(t_1 + t_2 \mu_{j+3} + t_3 \nu_{j+3}) - \frac{t_0^2}{4} \right].$$

The discriminant surface Δ (2.1) splits thus into three quadratic cones $\Delta = K_1 \cup K_2 \cup K_3$. Using the explicit form of the μ_i and ν_i (1.14), their equations are

$$\begin{aligned} K_j : & t_1 t_2 \sigma_1 + t_1 t_3 (a_j a_4 + a_{j'} a_{4'}) + t_2^2 (a_j + a_4)(a_{j'} + a_{4'}) \\ & + t_2 t_3 \sigma_3 + t_3^2 \sigma_4 = \frac{t_0^2}{4} - t_1^2 \end{aligned} \tag{2.2j} \quad (1 \leqq j \leqq 3)$$

where, for each $1 \leqq j \leqq 3$, the indices $j', 4'$ are such that $\{j, 4, j', 4'\} = \{1, 2, 3, 4\}$ and the σ_i, $1 \leqq i \leqq 4$, are the symmetric functions of the a_i given by $\sigma_1 = \sum a_i$, $\sigma_2 = \sum a_i a_j$, $\sigma_3 = \sum a_i a_j a_k$, and $\sigma_4 = \prod a_i$.

Now, substracting $(2.2\,j=2)$ from $(2.2\,j=1)$ or $(2.2\,j=3)$ from $(2.2\,j=2)$, we always obtain, up to some constant factor,

$$t_1 t_3 - t_2^2 = 0$$

which shows dependence of the three equations $(2.2\,j=1,2,3)$. The upshot is that the intersection $C_0 = K_1 \cap K_2 \cap K_3$ is a curve defined by only two equations

$$C_0 : \begin{cases} t_0^2 = 4(t_1^2 + t_1 t_2 \sigma_1 + t_2^2 \sigma_2 + t_2 t_3 \sigma_3 + t_3^2 \sigma_4) \\ t_1 t_3 = t_2^2 . \end{cases} \tag{2.3}$$

In the affine piece $t_3 = 1$, one has

$$C_0 : t_0^2 = 4 \prod_{i=1}^{4} (t_2 + a_i); \qquad t_1 = t_2^2$$

which shows that C_0 is an elliptic curve of degree 4. For each

$$t = [(t_0, t_1, t_2, t_3)] \in C_0 \quad (t_3 = 1),$$

one has

$$\begin{aligned} Q_t &= t_2^2 Q_1 + t_2 Q_2 + Q_3 \pm 2 \sqrt{\prod_{i=1}^{4} (t_2 + a_i) Q_0} \\ &= \sum_{j=1}^{3} (\sqrt{(t_2 + a_j)(t_2 + a_4)}\, x_j \pm \sqrt{(t_2 + a_{j'})(t_2 + a_{4'})}\, x_{j+3})^2 \end{aligned} \tag{2.4}$$

and so the quadric Q_t has rank 3. This establishes Proposition 2.1.

To the four branch points $t_2 = -a_i$ of C_0, correspond four distinguished rank 3 quadrics. If we define

$$Q_{a_i}(X) = \frac{a_i^2 Q_1(X) - a_i Q_2(X) + Q_3(X)}{\prod_{k \neq i} (a_i - a_k)} \tag{2.5}$$

they take the particularly nice form

$$\begin{aligned} Q_{a_1}(X) &= \frac{x_2^2}{a_1 - a_3} + \frac{x_3^2}{a_1 - a_2} + \frac{x_4^2}{a_1 - a_4} \\ Q_{a_2}(X) &= \frac{x_1^2}{a_2 - a_3} + \frac{x_3^2}{a_2 - a_1} + \frac{x_5^2}{a_2 - a_4} \\ Q_{a_3}(X) &= \frac{x_1^2}{a_3 - a_2} + \frac{x_2^2}{a_3 - a_1} + \frac{x_6^2}{a_3 - a_4} \\ Q_{a_4}(X) &= \frac{x_4^2}{a_4 - a_1} + \frac{x_5^2}{a_4 - a_2} + \frac{x_6^2}{a_4 - a_3} \end{aligned} \tag{2.6}$$

with $\sum_{i=1}^{4} Q_{a_i}(X) = 0$. Clearly, the four equations $Q_i(X) = c_i$, $0 \leqq i \leqq 3$, are equivalent to $Q_0(X) = c_0$ and $Q_{a_i}(X) = d_i$, $1 \leqq i \leqq 3$.

3. Asymptotic Expansions

Since Euler's rigid body motion can be solved in terms of elliptic functions x_i $(1 \leqq i \leqq 3)$, the question arises naturally whether the Euler-Arnold equations

$$\dot{X} = [X, \lambda(X)], \quad n=4, \quad \lambda \text{ satisfying (1.12)} \tag{3.1}$$

can be solved in terms of abelian functions x_i $(1 \leqq i \leqq 6)$. Recall that the main observation of Sect. 1 is that the linearization of (3.1) by the method of isospectral deformations (Theorem 2) does not answer this question. The precise question breaks up into two parts:

1) Is the variety $A_c \subseteq \mathbb{C}^6$ (1.16), defined by setting the four quadratic invariants of (3.1) equal to constants, generically the affine part of an abelian surface $\tilde{A}_c$ such that the uniformizing map $\tilde{A}_c \to \mathbb{C}^2/L_\Omega : p \curvearrowright \int_{p_0}^{p} (\omega_1, \omega_2)$ (ω_1, ω_2 a basis of holomorphic differential 1-forms on $\tilde{A}_c$) linearizes (3.1) and the other commuting flows i.e.

$$\frac{d}{dt} \int_{p_0}^{p(t)} (\omega_1, \omega_2) = (k_1, k_2). \tag{3.2}$$

If so, we shall say that (3.1) is *algebraically completely integrable with abelian functions* x_i.

2) Try to identify the abelian surface $\tilde{A}_c$ and to understand how it is related to the Prym variety $\mathrm{Prym}_t(K)$ of Theorem 2.

Before to state the main result of this section we need to explain the following *necessary criterion of algebraic complete integrability.*

Proposition 3.1. *If* (3.1) *is algebraically completely integrable with abelian functions* x_i $(1 \leqq i \leqq 6)$, *then* (3.1) *and any of its commuting flows must have Laurent series solutions in* $t \in \mathbb{C}$ *depending on* 5 *free parameters.*

Proof. Pick a typical 2-dimensional variety A_c whose compactification $\tilde{A}_c$ is an abelian surface on which (3.1) and the other commuting flows are straight line motions. Then, the coordinates x_i restricted to A_c are meromorphic (abelian) functions on a compact complex torus $\mathbb{C}^2/L_\Omega$, $x_i = f_i(t_1, t_2)$ and thus, they each blow up along one or several irreducible pieces of a codimension one divisor

$$\mathscr{D} = \{(t_1, t_2) \in \mathbb{C}^2/L_\Omega \mid f_i^{-1} = 0 \text{ for some } i\}.$$

Assume first that, for a generic set of tori, each irreducible piece of $\mathscr{D}$ is not an invariant set of the flow (3.1). Pick then an irreducible component of $\mathscr{D}$ and a typical point p on it. By the definition of algebraic complete integrability (3.2), the solution of (3.1) which goes through p is $x_i(t) = f_i(p + t(k_1, k_2))$ and can thus be expanded in a Laurent series. Now, think of $\mathscr{D}$ as a set of initial conditions for the differential equations and allow the 4 constants of the motion $c = (c_0, c_1, c_2, c_3)$ to vary. The upshot is that the differential equations (3.1) must have enough pole expansions in $t \in \mathbb{C}$ such that each x_i blows up at least once, and such that the expansions depend on $4 + 1 = 5$ free parameters.

In the case where, for an open set of tori, the divisor $\mathscr{D}$ has a component which is an invariant set of the flow (3.1), the solution $x_i(t)$ which starts from a point p on this component remains infinite and can not be expanded in a Laurent series.

However since, by hypothesis, the flow (3.1) may not vanish, such a component must be an embedded elliptic curve $E \subseteq \mathbb{C}^2/L_\Omega$. By the Poincaré complete reducibility theorem (see [14], p. 173) there exists then another elliptic curve E' such that $\mathbb{C}^2/L_\Omega = E + E'$ with $E \cap E' = \{\text{finite number of points}\}$. Also by hypothesis, the flow (3.1) is translation-invariant which implies that for each $\gamma \in E'$, the elliptic curve $E\gamma = E + \gamma$ is an invariant set of (3.1). Since this happens for an open set of tori, the upshot is that (3.1) has a fifth independent algebraic invariant and can be integrated in terms of elliptic functions. As a consequence, in such a situation, (3.1) also possesses Laurent series solutions depending on 5 free parameters. Now, the whole argument can be repeated for any flow commuting with (3.1). This establishes the proposition.

Remark. Obviously, the above necessary criterion of algebraic complete integrability has nothing to do with the special form of the equations (3.1). For a general formulation of this criterion see [4], Proposition 1. The authors informed me that their proof is incorrect. The above proof works for systems which linearize on a 2-dimensional abelian variety. In [4], the criterion is used to prove that, for $n=4$ and λ diagonal in the usual basis of so(4), the Euler-Arnold equations can be algebraically completely integrable with abelian functions x_i $(1 \leqq i \leqq 6)$, only if λ satisfies (1.12). However the sufficiency of (1.12) is proved there by reference to Theorem 2 which, as explained before, does not prove that the x_i are abelian functions.

As announced in the introduction, our aim is to show that this necessary condition of algebraic complete integrability is also sufficient for the class of Hamiltonian systems (3.1). For this we shall only need to know the expansions of the x_i $(1 \leqq i \leqq 6)$ in two independent directions. Therefore, we may pick specific linear combinations of the invariants $Q_i(X)$ $(0 \leqq i \leqq 3)$ giving rise to Hamiltonian vector fields for which these expansions are easy to compute. In this section we exhibit these expansions for the Hamiltonian flows corresponding to the four distinguished rank 3 quadrics $Q_{a_i}(X)$ $(1 \leqq i \leqq 4)$ obtained in (2.6). After we use them to find the equations of the divisor $\mathscr{D}$ along which the "presumed" abelian functions x_i $(1 \leqq i \leqq 6)$ blow up.

Consider (for example) the rank 3 quadric

$$Q_{a_4}(X) = \mu_4 x_4^2 + \mu_5 x_5^2 + \mu_6 x_6^2$$

where

$$\mu_4 = \frac{1}{a_4 - a_1}, \quad \mu_5 = \frac{1}{a_4 - a_2}, \quad \mu_6 = \frac{1}{a_4 - a_3}.$$

From (1.4), the corresponding Hamiltonian vector field is

$$\dot{X} = [X, \nabla Q_{a_4}(X)] = [X, \mu(X)]. \tag{3.3}$$

Then upon setting $B_{ij} = \mu_i - \mu_j$, (3.3) takes the simple form:

$$\begin{aligned} \dot{x}_1 &= B_{65} x_5 x_6, & \dot{x}_4 &= \mu_6 x_2 x_6 - \mu_5 x_3 x_5 \\ \dot{x}_2 &= B_{46} x_4 x_6, & \dot{x}_5 &= \mu_4 x_3 x_4 - \mu_6 x_1 x_6 \\ \dot{x}_3 &= B_{54} x_4 x_5, & \dot{x}_6 &= \mu_5 x_1 x_5 - \mu_4 x_2 x_4. \end{aligned} \tag{3.4}$$

The reader will find in [4, p. 323] a proof of the following tedious lemma.

Lemma 3.2. *A Laurent series solution to* (1.1) ($n=4$) *with a pole harder than* t^{-1} *can depend of at most* 4 *free parameters.*

From this lemma, it follows that Laurent series solutions to (3.4) which depend on 5 free parameters must have the form

$$X(t)=t^{-1}(X^0+X^1t+\ldots). \tag{3.5}$$

Putting (3.5) into (3.3), one solves inductively for the X^ℓ and one finds

$$F(X^0)\equiv X^0+[X^0,\mu(X^0)]=0, \quad X^0\not\equiv 0 \tag{3.6}$$

and

$$(\mathscr{L}-\ell I)(X^\ell)=-\sum_{k=1}^{\ell-1}[X^k,\mu(X^{\ell-k})] \quad (\ell\geqq 1) \tag{3.7}$$

where $\mathscr{L}$ is the linear operator $\mathscr{L}:\mathrm{so}(4)\to\mathrm{so}(4)$ defined by

$$\mathscr{L}(Y)=Y+[X^0,\mu(Y)]+[Y,\mu(X^0)]$$

and the matrix X^0 appearing in $\mathscr{L}$ is a solution of the non-linear equation (3.6). In the basis $(e_i)_{1\leqq i\leqq 6}$, $\mathscr{L}$ is given by the matrix

$$\mathscr{L}=\begin{pmatrix} 1 & 0 & 0 & 0 & B_{65}x_6^0 & B_{65}x_5^0 \\ 0 & 1 & 0 & B_{46}x_6^0 & 0 & B_{46}x_4^0 \\ 0 & 0 & 1 & B_{54}x_5^0 & B_{54}x_4^0 & 0 \\ 0 & \mu_6x_6^0 & -\mu_5x_5^0 & 1 & -\mu_5x_3^0 & \mu_6x_2^0 \\ -\mu_6x_6^0 & 0 & \mu_4x_4^0 & \mu_4x_3^0 & 1 & -\mu_6x_1^0 \\ \mu_5x_5^0 & -\mu_4x_4^0 & 0 & -\mu_4x_2^0 & \mu_5x_1^0 & 1 \end{pmatrix}.$$

Note the Jacobian of $F(X^0)=0$ (3.6) is $\mathscr{L}$. Thus degrees of freedom in the determination of X^0 can occur only if $\det\mathscr{L}=0$. For $\ell\geqq 1$, the X^ℓ are determined by solving linear systems of equations (3.7) and freedom can occur only if $\det(\mathscr{L}-\ell I)=0$. Observe also that by (3.6) $\mathscr{L}(X^0)+X^0=0$ and so -1 is automatically an eigenvalue of $\mathscr{L}$. We now prove that there are effectively 5 degrees of freedom in the determination of the X^ℓ, $\ell\geqq 0$.

Lemma 3.3. *The non identically zero solutions of* (3.6) *define the affine part of an elliptic curve* E *which is a* 4*-sheeted unramified covering of the elliptic curve* $y^2=\prod_{i=1}^{4}(z-a_i)$.

Proof. Equations (3.6) have the following explicit form

$$\begin{aligned} &x_1^0+B_{65}x_5^0x_6^0=0, \quad && x_4^0+\mu_6x_2^0x_6^0-\mu_5x_3^0x_5^0=0 \\ &x_2^0+B_{46}x_4^0x_6^0=0, && x_5^0+\mu_4x_3^0x_4^0-\mu_6x_1^0x_6^0=0 \\ &x_3^0+B_{54}x_4^0x_5^0=0, && x_6^0+\mu_5x_1^0x_5^0-\mu_4x_2^0x_4^0=0. \end{aligned} \tag{3.8}$$

Consider first the solutions of (3.8) such that $\Pi x_i^0 \neq 0$. Substituting the first three equations of (3.8) into the three last ones we obtain the equivalent system

$$\begin{aligned} -\mu_5 B_{54}(x_5^0)^2 + \mu_6 B_{46}(x_6^0)^2 = 1\,, \quad & x_1^0 = -B_{65} x_5^0 x_6^0 \\ \mu_4 B_{54}(x_4^0)^2 - \mu_6 B_{65}(x_6^0)^2 = 1\,, \quad & x_2^0 = -B_{46} x_4^0 x_6^0 \\ -\mu_4 B_{46}(x_4^0)^2 + \mu_5 B_{65}(x_5^0)^2 = 1\,, \quad & x_3^0 = -B_{54} x_4^0 x_5^0\,. \end{aligned} \tag{3.9}$$

It is now easy to check that the only nonidentically zero solutions of (3.8) with $\Pi x_i^0 = 0$ are given by ($x_2^0 = x_3^0 = x_4^0 = 0$ and $x_1^0 x_5^0 x_6^0 \neq 0$) or ($x_1^0 = x_3^0 = x_5^0 = 0$ and $x_2^0 x_4^0 x_6^0 \neq 0$) or ($x_1^0 = x_2^0 = x_6^0 = 0$ and $x_3^0 x_4^0 x_5^0 \neq 0$) and that the system (3.9) possesses these solutions. The system (3.8) with $X^0 \neq 0$ is thus completely equivalent to (3.9). Now, among the three first equations of (3.9), only two are independent giving us the equation of an algebraic curve

$$E: x_5^0 = \pm \left(\frac{1 + \mu_4 B_{46}(x_4^0)^2}{\mu_5 B_{65}} \right)^{1/2}; \quad x_6^0 = \pm \left(\frac{\mu_4 B_{54}(x_4^0)^2 - 1}{\mu_6 B_{65}} \right)^{1/2} \tag{3.10}$$

which, because of the three last equations of (3.9), determines in fact all the solutions of this system.

Now, the algebraic curve E has exactly the same structure as the Euler's curve (1.8) studied in Sect. 1. Namely, it can be viewed as a 4-sheeted unramified covering of the elliptic plane curve

$$C_0: \zeta^2 = (\mu_5 \mu_6)^{-1} \alpha_1 (\mu_4 \alpha_1 + B_{54})(\mu_4 \alpha_1 - B_{46}) \tag{3.11}$$

with covering projection $\Pi: E \to C_0$ defined by

$$\Pi(x_4^0, x_5^0, x_6^0) = (\zeta, \alpha_1) \equiv (B_{54} B_{46} B_{65} x_4^0 x_5^0 x_6^0, B_{54} B_{46} (x_4^0)^2)\,.$$

Since Π is unramified, E is also an elliptic curve. The fractional linear transformation

$$\alpha_1 = \frac{z - a_1}{z - a_4}, \qquad \zeta = \frac{y}{(z - a_4)^2} \tag{3.12}$$

maps the curve C_0 into

$$C_0: y^2 = \prod_{i=1}^{4} (z - a_i)\,.$$

According to (1.10), we get the following commutative diagram

$$\begin{array}{ccccc} E & \xrightarrow{I_\varphi} & \mathbb{C}/2L_\Delta & \longleftarrow & \mathbb{C} \\ \Big\downarrow \Pi & & \Big\downarrow & & \Big\| \\ C_0 & \xrightarrow{I_\theta} & \mathbb{C}/L_\Delta & \longleftarrow & \mathbb{C} \end{array} \tag{3.13}$$

with $\theta = \dfrac{dz}{y}$ and $\varphi = \Pi^*(\theta)$. This ends the proof of Lemma 3.3.

Lemma 3.4. (i) *The system* $(\mathscr{L} - I)(X^1) = 0$ *has* 1 *degree of freedom.*

(ii) *The system* $(\mathscr{L} - 2I)(X^2) = -[X^1, \mu(X^1)]$ *has* 3 *degrees of freedom.*

Proof. (i) The system $(\mathscr{L}-I)(X^1)=0$ is explicitely given by

$$\begin{aligned}
&x_6^0x_5^1+x_5^0x_6^1=0, \quad && \mu_6x_6^0x_2^1-\mu_5x_5^0x_3^1=\mu_5x_3^0x_5^1-\mu_6x_2^0x_6^1\\
&x_6^0x_4^1+x_4^0x_6^1=0, && -\mu_6x_6^0x_1^1+\mu_4x_4^0x_3^1=-\mu_4x_3^0x_4^1+\mu_6x_1^0x_6^1\\
&x_5^0x_4^1+x_4^0x_5^1=0, && \mu_5x_5^0x_1^1-\mu_4x_4^0x_2^1=\mu_4x_2^0x_4^1-\mu_5x_1^0x_5^1.
\end{aligned}$$

Suppose first that $x_4^0x_5^0x_6^0\neq 0$. Then the only solution of the three first equations is $x_4^1=x_5^1=x_6^1=0$. The remaining three last equations constitute a linear homogeneous system in x_1^1, x_2^1, and x_3^1 with solution

$$x_1^1=\mu_4x_4^0\beta, \quad x_2^1=\mu_5x_5^0\beta, \quad x_3^1=\mu_6x_6^0\beta$$

with a free parameter β. In the case where $x_4^0x_5^0x_6^0=0$, using again the fact that the only possibilities are ($x_2^0=x_3^0=x_4^0=0$ and $x_1^0x_5^0x_6^0\neq 0$) or ($x_1^0=x_3^0=x_5^0=0$ and $x_2^0x_4^0x_6^0\neq 0$) or ($x_1^0=x_2^0=x_6^0=0$ and $x_3^0x_4^0x_5^0\neq 0$), the reader can easily check that the above solution is still valid.

(ii) Since $x_4^1=x_5^1=x_6^1=0$, the system $(\mathscr{L}-2I)(X^2)=-[X^1,\mu(X^1)]$ becomes

$$\begin{aligned}
x_1^2&=B_{65}(x_6^0x_5^2+x_5^0x_6^2)\\
x_2^2&=B_{46}(x_6^0x_4^2+x_4^0x_6^2)\\
x_3^2&=B_{54}(x_5^0x_4^2+x_4^0x_5^2)\\
\mu_6x_6^0x_2^2-\mu_5x_5^0x_3^2-x_4^2-\mu_5x_3^0x_5^2+\mu_6x_2^0x_6^2&=0\\
-\mu_6x_6^0x_1^2+\mu_4x_4^0x_3^2+\mu_4x_3^0x_4^2-x_5^2-\mu_6x_1^0x_6^2&=0\\
\mu_5x_5^0x_1^2-\mu_4x_4^0x_2^2-\mu_4x_2^0x_4^2+\mu_5x_1^0x_5^2-x_6^2&=0.
\end{aligned}$$

The three first equations give x_1^2, x_2^2, and x_3^2 in terms of x_4^2, x_5^2, and x_6^2. If we substitute these expressions into the three last equations, we find that they are identically satisfied. The solution of this system is thus given by

$$\begin{gathered}
x_4^2=\gamma, \quad x_5^2=\delta, \quad x_6^2=\varepsilon\\
x_1^2=B_{65}(x_6^0\delta+x_5^0\varepsilon), \quad x_2^2=B_{46}(x_6^0\gamma+x_4^0\varepsilon), \quad x_3^2=B_{54}(x_5^0\gamma+x_4^0\delta)
\end{gathered}$$

where γ, δ, ε are free parameters. This finishes the proof of Lemma 3.4.

Remark. A remarkable feature in the computation of the asymptotic expansions of the x_i in the distinguished directions $X_{Q_{a_j}}$, $1\leqq j\leqq 4$, is that, whenever $\det(\mathscr{L}-\ell I)=0$, the right member of (3.7) is identically zero so that this linear system is automatically compatible.

Now, from Lemmas 3.3 and 3.4 it follows that $\det(\mathscr{L}-\ell I)=(\ell^2-1)\ell(\ell-2)^3$ and so the X^ℓ, $\ell\geqq 3$, are uniquely determined by solving inductively (3.7). Since formal series solutions of differential equations with quadratic right hand side are automatically convergent as a consequence of the majorant method, we have thus proved the

Proposition 3.5. *The flow* $X_{Q_{a_1}}$ (3.3) *possesses Laurent series solutions which depend on* 5 *free parameters. The first free parameter appears in the form of an elliptic curve* (3.10) *while the* 4 *other ones* β, γ, δ, ε *are affine parameters. These Laurent series*

solutions are explicitely given by

$$\begin{aligned}
x_1(t) &= t^{-1}(x_1^0 + \mu_4 x_4^0 \beta t + B_{65}(x_6^0\delta + x_5^0\varepsilon)t^2 + \ldots)\\
x_2(t) &= t^{-1}(x_2^0 + \mu_5 x_5^0 \beta t + B_{46}(x_6^0\gamma + x_4^0\varepsilon)t^2 + \ldots)\\
x_3(t) &= t^{-1}(x_3^0 + \mu_6 x_6^0 \beta t + B_{54}(x_5^0\gamma + x_4^0\delta)t^2 + \ldots)\\
x_4(t) &= t^{-1}(x_4^0 + ot + \gamma t^2 + \ldots)\\
x_5(t) &= t^{-1}(x_5^0 + ot + \delta t^2 + \ldots)\\
x_6(t) &= t^{-1}(x_6^0 + ot + \varepsilon t^2 + \ldots).
\end{aligned} \tag{3.14}$$

Similar expansions (modulo permutation of indices) can be obtained for $X_{Q_{a_1}}$, $X_{Q_{a_2}}$ *and* $X_{Q_{a_3}}$.

We are now in position to compute the divisor $\mathscr{D}$ along which the "presumed" abelian functions x_i blow up. For this, we substitute the expansions (3.14) into the first integrals of the flow, the only remaining term being the term in t^0. After substitution into $Q_{a_i}(X) = d_i$, $1 \leqq i \leqq 4$, $\sum_{i=1}^{4} d_i = 0$ (2.6), we obtain

$$\begin{aligned}
-2\gamma' + \mu_5 B_{54}(x_4^0)^2\delta' - \mu_6 B_{46}(x_4^0)^2\varepsilon' &= \frac{d_1}{\mu_4} + \frac{\mu_5\mu_6\beta^2}{B_{54}B_{46}}\\
-\mu_4 B_{54}(x_5^0)^2\gamma' - 2\delta' + \mu_6 B_{65}(x_5^0)^2\varepsilon' &= \frac{d_2}{\mu_5} + \frac{\mu_4\mu_6\beta^2}{B_{65}B_{54}}\\
\mu_4 B_{46}(x_6^0)^2\gamma' - \mu_5 B_{65}(x_6^0)^2\delta' - 2\varepsilon' &= \frac{d_3}{\mu_6} + \frac{\mu_4\mu_5\beta^2}{B_{46}B_{65}}\\
\mu_4\gamma' + \mu_5\delta' + \mu_6\varepsilon' &= d_4
\end{aligned} \tag{3.15}$$

with $\gamma' = 2x_4^0\gamma$, $\delta' = 2x_5^0\delta$, and $\varepsilon' = 2x_6^0\varepsilon$. This is a linear system in γ', δ', ε', the fourth equation being a consequence of the three first ones. The last invariant $Q_0(X) = c_0$ (1.11) gives

$$-2B_{65}x_5^0x_6^0\gamma - 2B_{46}x_4^0x_6^0\delta - 2B_{54}x_4^0x_5^0\varepsilon = c_0.$$

If $x_4^0x_5^0x_6^0 \neq 0$, this equation is equivalent to

$$-B_{65}(x_5^0)^2(x_6^0)^2\gamma' - B_{46}(x_4^0)^2(x_6^0)^2\delta' - B_{54}(x_4^0)^2(x_5^0)^2\varepsilon' = c_0x_4^0x_5^0x_6^0. \tag{3.16}$$

Solving (3.15) for γ', δ', ε' and substituting the solution into (3.16), by making repeated use of (3.9), one finally obtains an equation connecting the two remaining parameters β and the "parameter" of the elliptic curve E (3.10) which is the equation of an affine piece $\mathscr{D}^a$ [5] of the divisor $\mathscr{D}$ along which the x_i blow up:

$$\begin{aligned}
\mathscr{D}^a : \{(\beta, x_4^0, x_5^0, x_6^0) \in \mathbb{C}^4 \mid \beta^2 + R(x_4^0, x_5^0, x_6^0) = 0\\
\mu_4 B_{54}(x_4^0)^2 - \mu_6 B_{65}(x_6^0)^2 = 1\\
-\mu_4 B_{46}(x_4^0)^2 + \mu_5 B_{65}(x_5^0)^2 = 1\}
\end{aligned} \tag{3.17}$$

5 We shall use the notation $\mathscr{D}^a$ when we want to insist on the fact that we consider the affine piece of $\mathscr{D}$ defined by (3.17)

where

$$R(x_4^0, x_5^0, x_6^0) = 2c_0\zeta - \frac{d_1}{\mu_4}\alpha_2\alpha_3 - \frac{d_2}{\mu_5}\alpha_1\alpha_3 - \frac{d_3}{\mu_6}\alpha_1\alpha_2$$

$$\zeta = B_{54}B_{46}B_{65}x_4^0x_5^0x_6^0$$

$$\alpha_1 = B_{54}B_{46}(x_4^0)^2, \quad \alpha_2 = B_{65}B_{54}(x_5^0)^2, \quad \alpha_3 = B_{46}B_{65}(x_6^0)^2.$$

It is now easy to check that the expression (3.17) remains valid when $x_4^0x_5^0x_6^0 = 0$. The upshot is that $\mathscr{D}$ has only one irreducible component which can be seen as a 2-sheeted ramified covering of the elliptic curve E (3.10):

$$\tilde{p}: \mathscr{D} \to E : (\beta, x_4^0, x_5^0, x_6^0) \rightsquigarrow (x_4^0, x_5^0, x_6^0). \tag{3.18}$$

As we know from Sect. 1, the elliptic curve E embeds into $\mathbb{P}^3(\mathbb{C})$. The non singular compactification $\mathscr{D}$ of $\mathscr{D}^a$ is obtained by making the following change of coordinates at infinity:

$$\begin{gathered}(x_4^0, x_5^0, x_6^0, \beta) \rightsquigarrow (y_2, y_3, y_4, \beta') \\ y_2 \equiv \frac{x_5^0}{x_4^0}, \quad y_3 \equiv \frac{x_6^0}{x_4^0}, \quad y_4 \equiv \frac{1}{x_4^0}, \quad \beta' \equiv \frac{\beta}{(x_4^0)^2}.\end{gathered} \tag{3.19}$$

One sees easily that there are 8 distinct points at infinity, 2 over each of the 4 points at infinity of E so that $\tilde{p}$ is unramified at infinity.

Remark. At this point, it may be not so clear for the reader why the curve $\mathscr{D}$ must really live on an abelian surface. Let us say, for the moment, that the equations of the divisor $\mathscr{D}$, as a curve traced on the abelian surface $\tilde{A}_c$ (to be constructed in the next section), must be understood as relations connecting the free parameters as they appear firstly in the expansions (3.14) i.e. linearly. In the present situation, this means that (3.17) must be understood as relations connecting β and x_4^0, x_5^0, x_6^0. This will be fully justified in the next section [see Sect. 4 formula (4.7)].

Proposition 3.6. *The divisor of poles* $\mathscr{D}$ (3.17) *of the "presumed" abelian functions* x_i $(1 \leqq i \leqq 6)$ *is a non singular irreducible genus* 9 *curve.*

Proof. Since the covering $\tilde{p}: \mathscr{D} \to E$ (3.18) is unramified at infinity, the branch points of $\tilde{p}$ are defined by the zeroes of $R(x_4^0, x_5^0, x_6^0)$ on the affine part of E. Observe that, by (3.10), we have

$$\alpha_2 = \frac{\mu_4\alpha_1 + B_{54}}{\mu_5} \quad \text{and} \quad \alpha_3 = \frac{\mu_4\alpha_1 - B_{46}}{\mu_6} \quad \text{on } E$$

where α_1, α_2, and α_3 are defined in (3.17). As a consequence

$$R(x_4^0, x_5^0, x_6^0) = 2c_0\zeta - \frac{d_1}{\mu_4}\alpha_2\alpha_3 - \frac{d_2}{\mu_5}\alpha_1\alpha_3 - \frac{d_3}{\mu_6}\alpha_1\alpha_2 = 2c_0\zeta + P(\alpha_1) \text{ on } E \tag{3.20}$$

with $P(\alpha_1)$ a polynomial of degree 2 in α_1. Recall now the picture (3.13). Formula (3.20) means that the elliptic function $R(x_4^0, x_5^0, x_6^0)$ defined on $E \approx \mathbb{C}/2L_\Delta$ is also an elliptic function on the "smaller" elliptic curve $C_0 \approx \mathbb{C}/L_\Delta$ defined in (3.11). Clearly R has 4 zeroes on C_0 and thus 16 zeroes on E

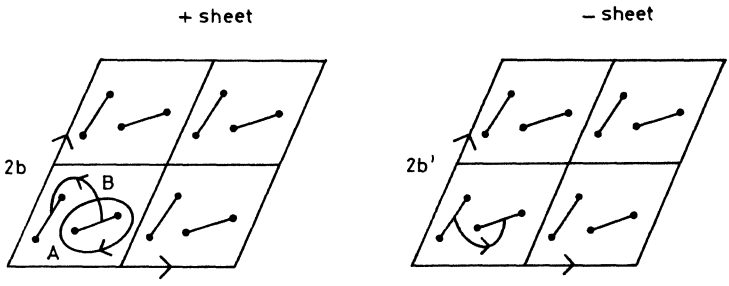


Fig. 3

represented as in Fig. 3. Using the Riemann-Hurwitz formula, we find

$$g(\mathscr{D}) = n(g(E)-1)+1+\frac{b}{2} = 9$$

where, as usual, n is the number of sheets and b the ramification index of the covering $\tilde{p}:\mathscr{D}\to E$. This establishes Proposition 3.6.

Now, the previous discussion leads us naturally to consider also the algebraic curve C defined by

$$C:\{(\beta,\zeta,\alpha_1)\in\mathbb{C}^3 | \beta^2+2c_0\zeta+P(\alpha_1)=0,$$
$$\zeta^2=(\mu_5\mu_6)^{-1}\alpha_1(\mu_4\alpha_1+B_{54})(\mu_4\alpha_1-B_{46})\}$$

where $P(\alpha_1)$ is the degree 2 polynomial in α_1 defined by (3.20). C is thus a 2-sheeted ramified covering of the elliptic curve C_0 (3.11)

$$p:C\to C_0:(\beta,\zeta,\alpha_1)\curvearrowright(\zeta,\alpha_1) \tag{3.21}$$

ramified over the 4 zeroes of $2c_0\zeta+P(\alpha_1)$ on C_0. Its genus is thus $g(C)=3$. We summarize all the relations existing between the four algebraic curves $\mathscr{D}$, E, C, and C_0 in the following commutative diagram:

$$\begin{array}{ccccccc}
\mathscr{D} & \xrightarrow[2-1r]{\tilde{p}} & E & \xrightarrow{I_\varphi} & \mathbb{C}/2L_\Delta & \longleftarrow & \mathbb{C} \\
{\scriptstyle 4-1\atop ur}\downarrow & \tilde{\Pi} & {\scriptstyle 4-1\atop ur}\downarrow & \Pi & \downarrow & & \| \\
C & \xrightarrow[2-1r]{p} & C_0 & \xrightarrow{I_\theta} & \mathbb{C}/L_\Delta & \longleftarrow & \mathbb{C}
\end{array} \tag{3.22}$$

r = ramified covering; ur = unramified covering

where all the maps have been defined previously except $\tilde{\Pi}$ which is defined by

$$\tilde{\Pi}:\mathscr{D}\to C:(\beta,x_4^0,x_5^0,x_6^0)\curvearrowright(\beta,\zeta,\alpha_1)$$
$$\beta\equiv\beta,\quad \zeta\equiv B_{54}B_{46}B_{65}x_4^0x_5^0x_6^0,\quad \alpha_1\equiv B_{54}B_{46}(x_4^0)^2. \tag{3.23}$$

The genus 3 curve C will play a very important role in the next section. It can be written in a particularly nice form using the (y,z) coordinates introduced in

(3.12). Namely, make the following change of coordinates

$$(\beta, \zeta, \alpha_1) \curvearrowright (w, y, z)$$

$$\alpha_1 \equiv \frac{z-a_1}{z-a_4}, \quad \zeta \equiv \frac{y}{(z-a_4)^2}, \quad \beta \equiv \frac{w}{z-a_4}.$$

Then using relation (2.5) with $Q_i(X)=c_i$ and $Q_{a_i}(X)=d_i$ $(1 \leqq i \leqq 3)$, in the new coordinates (w, y, z), the curve C is written as

$$C: \begin{cases} w^2 + c_1 z^2 - c_2 z + c_3 + 2c_0 y = 0 \\ y^2 = \prod_{i=1}^{4} (z-a_i) \end{cases} \tag{3.24}$$

and the covering projection (3.21) is given by $p: C \to C_0 : (w, y, z) \curvearrowright (y, z)$. In the sequel we shall need to know a basis of holomorphic differentials on C. Using Poincaré residue formula (see [7], p. 221) we find

$$\theta = \frac{dz}{y}, \quad \psi_1 = \frac{z\,dz}{wy}, \quad \psi_2 = \frac{dz}{wy}. \tag{3.25}$$

In (β, ζ, α_1) coordinates, we shall use the basis

$$\theta' = \frac{d\alpha_1}{\zeta}, \quad \psi'_1 = \frac{\alpha_1 d\alpha_1}{\beta\zeta}, \quad \psi'_2 = \frac{d\alpha_1}{\beta\zeta}. \tag{3.26}$$

Observe that $\iota^*(\theta) = \theta$ and $\iota^*(\psi_j) = -\psi_j$ $(j=1,2)$ for the involution

$$\iota: C \to C: (w, y, z) \curvearrowright (-w, y, z).$$

Remark. Suppose for a moment we have proved that A_c is the affine part of an abelian surface $\tilde{A}_c$. Then, by the Riemann-Roch theorem, $\dim \mathscr{L}(\mathscr{D}) = g(\mathscr{D}) - 1 = 8$ where, as usual, $\mathscr{L}(\mathscr{D})$ denotes the space of meromorphic functions on $\tilde{A}_c$ with at most a simple pole along $\mathscr{D}$. Trivially 1, x_i $(1 \leqq i \leqq 6)$ belong to $\mathscr{L}(\mathscr{D})$. The last abelian function in $\mathscr{L}(\mathscr{D})$ is easily found by inspection of the expansions (3.14):

$$\begin{aligned} x_7 &= (a_3 - a_1)(a_4 - a_2) x_1 x_4 + (a_3 - a_2)(a_4 - a_1) x_2 x_5 \\ &= \frac{(a_4 - a_1)(a_4 - a_2)(a_4 - a_3)\beta}{t} + \ldots \end{aligned} \tag{3.27}$$

4. The Abelian Surface $\tilde{A}_c$ as a Prym Variety

In this section we use the expansions (3.14) to compactify the affine variety A_c (1.16) in an abelian surface $\tilde{A}_c$. We identify this abelian surface as a Prym variety of a certain ramified double cover C of the elliptic curve C_0 of rank 3 quadrics found in (2.3). The precise statements are embodied in the following theorems:

Theorem 3. *The variety* $A_c = \bigcap_{i=0}^{3} \{Q_i(X) = c_i\} \subseteq \mathbb{C}^6$ *generically is the affine part of an abelian surface* $\tilde{A}_c$ *with* $\tilde{A}_c \backslash A_c =$ *the genus* 9 *curve* $\mathscr{D}$ *found in* (3.17). *Moreover, any Hamiltonian flow commuting with* (3.1) *is a straight line motion on* $\tilde{A}_c$.

Theorem 4. (i) $\tilde{A}_c$ *is the Prym variety of the genus* 3 *curve* C *defined in* (3.24) *for the involution* $\iota: C \to C: (w, y, z) \curvearrowright (-w, y, z)$.

(ii) *The map*

$$\mathbb{C}^2/L_\Omega \approx \tilde{A}_c \to \mathbb{P}^7(\mathbb{C}): (t_1, t_2) \curvearrowright [(1, x_1(t_1, t_2), \ldots, x_6(t_1, t_2), x_7(t_1, t_2))]$$

with x_7 *the last abelian function in* $\mathscr{L}(\mathscr{D})$ *found in* (3.27) *is an embedding. One needs* 6 *quadratic equations to describe* $\tilde{A}_c \approx \mathbb{C}^2/L_\Omega$ *as embedded into* $\mathbb{P}^7(\mathbb{C})$.

Connection with the Curve of Rank 3 *Quadrics.* Let $C_0 \subseteq \mathbb{P}^3(\mathbb{C})$ be the elliptic curve obtained in (2.3) such that the quadric $Q_t = \sum_{i=0}^{3} t_i Q_i$ has rank 3 for each $t \in C_0$. Putting $t_3 = 1$, $t_2 = -z$, and $t_0 = 2y$ into (2.3),

$$C_0 = \left\{ [(2y, z^2, -z, 1)] \in \mathbb{P}^3(\mathbb{C}) \mid y^2 = \prod_{i=1}^{4} (z - a_i) \right\}.$$

Let $\Pi \subseteq \mathbb{P}^3(\mathbb{C})$ be the hyperplane

$$\Pi = \left\{ t \in \mathbb{P}^3(\mathbb{C}) \mid \sum_{i=0}^{3} t_i c_i = 0 \right\}$$

where the c_i $(0 \leqq i \leqq 3)$ denote the 4 constants of the motion, i.e. $Q_i(X) = c_i$. Clearly

$$C_0 \cap \Pi = \left\{ [(2y, z^2, -z, 1)] \in \mathbb{P}^3(\mathbb{C}) \mid y^2 = \prod_{i=1}^{4} (z - a_i) \text{ and } c_1 z^2 - c_2 z + c_3 + 2c_0 y = 0 \right\}$$

which shows that the genus 3 curve C (3.24) can be obtained as the 2-sheeted covering of C_0 ramified over $C_0 \cap \Pi = \{4 \text{ points}\}$.

We first outline the *idea of the proof of Theorem 3.* We must glue the divisor $\mathscr{D}$ (3.17) to the affine piece A_c in an "algebraic fashion" i.e. in such a way that the transitions from the affine coordinates x_i $(1 \leqq i \leqq 6)$ to the coordinates at infinity are algebraic. For this, we use the expansions (3.14) to construct (locally) rational functions τ, $\hat{x}_4^0$, $\hat{x}_5^0$, $\hat{x}_6^0$, $\hat{\beta}$, $\hat{\beta}'$[6] of the x_i $(1 \leqq i \leqq 6)$ such that $\tau(t) = t + o(t^2)$, $\hat{x}_i^0 = x_i^0 + o(t)$ $(4 \leqq i \leqq 6)$, $\hat{\beta} = \beta + o(t)$, $\hat{\beta}' = \beta + o(t)$. In these new coordinates an open affine piece of A_c is defined by algebraic equations $F_k(\tau, \hat{x}_4^0, \hat{x}_5^0, \hat{x}_6^0, \hat{\beta}, \hat{\beta}') = 0$ $(0 \leqq k \leqq 3)$ which reduce to the equations of an affine piece of $\mathscr{D}$ (3.17) when $\tau = 0$. In this way we construct, by mean of an algebraic atlas, a non singular compact variety $\tilde{A}_c = A_c \cup \mathscr{D}$. Showing that, for example, the two commuting vector fields $Y_1 = X_{Q_{a_4}}$ and $Y_2 = X_{Q_{a_1}}$ are regular and remain independent all along $\mathscr{D}$ proves then, by the same argument as in the classical Liouville's theorem, that $\tilde{A}_c$ is a torus. Now, a complex torus which is given locally by algebraic equations with algebraic transition functions is an abelian variety (see Mumford [14], p. 35).

Let $\bar{A}_c$ be the projective closure of $A_c \subseteq \mathbb{C}^6$ into $\mathbb{P}^6(\mathbb{C})$. Because of (2.5) $\bar{A}_c$ can be defined by

$$\begin{aligned} &(4.1)(0)\quad Q_0(X) = c_0 x_0^2 \\ &(4.1)(i)\quad Q_{a_i}(X) = d_i x_0^2 \qquad \left(1 \leqq i \leqq 4,\ \sum_{i=1}^{4} d_i = 0\right). \end{aligned} \tag{4.1}$$

6 The consideration of $\hat{\beta}'$ is inessential as will appear from the proof of Theorem 3

Before proceeding to the proof of Theorem 3, it is necessary to analyze the locus at infinity $I=\bar{A}_c\cap\{x_0=0\}$. This will now be done in the following sequence of lemmas.

Lemma 4.1. *On I one never has $(x_1=x_4=0)$ or $(x_2=x_5=0)$ or $(x_3=x_6=0)$.*

Lemma 4.2. *The locus I can be covered by the 4 affine open sets $U_1\equiv\{x_1x_5x_6\neq 0\}$, $U_2\equiv\{x_2x_4x_6\neq 0\}$, $U_3\equiv\{x_3x_4x_5\neq 0\}$, and $U_4\equiv\{x_1x_2x_3\neq 0\}$.*

The proof of these two lemmas is delayed to the end of this section.

Lemma 4.3. *In U_1, the four Eqs.* (4.1)(0)(2)(3)(4) *are equivalent to* (4.1)(0)(2)(3) *and*

$$(4.1)(5)\quad F^2+(a_4-a_1)(a_3-a_1)Gx_0^2=0$$

with

$$F=(a_3-a_1)(a_4-a_2)x_1x_4+(a_3-a_2)(a_4-a_1)x_2x_5$$

and

$$G=(a_2-a_1)[d_4(a_4-a_2)(a_3-a_4)x_1^2+d_3(a_4-a_3)(a_3-a_2)x_5^2+d_2(a_2-a_4)(a_3-a_2)x_6^2]$$
$$+c_0(a_4-a_2)(a_3-a_2)[2x_3x_6-c_0x_0^2].$$

One has of course a similar statement in U_2, U_3, and U_4.

Proof. In U_1 Eqs. (4.1)(0)(2)(3)(4) may be written as

$$\begin{aligned}
&(0)\quad x_1x_4+x_2x_5+x_3x_6=c_0x_0^2\\
&(2)\quad \mu_6B_{65}\left(\frac{1}{B_{65}}\frac{x_1}{x_5}\right)^2-\mu_4B_{54}\left(\frac{1}{B_{54}}\frac{x_3}{x_5}\right)^2+1=-\frac{d_2}{\mu_5}\left(\frac{x_0}{x_5}\right)^2\\
&(3)\quad -\mu_5B_{65}\left(\frac{1}{B_{65}}\frac{x_1}{x_6}\right)^2+\mu_4B_{46}\left(\frac{1}{B_{46}}\frac{x_2}{x_6}\right)^2+1=-\frac{d_3}{\mu_6}\left(\frac{x_0}{x_6}\right)^2\\
&(4)\quad \mu_4\left(\frac{x_1x_4}{x_5x_6}\right)^2+\mu_5\left(\frac{x_1}{x_6}\right)^2+\mu_6\left(\frac{x_1}{x_5}\right)^2=d_4\left(\frac{x_0x_1}{x_5x_6}\right)^2.
\end{aligned}\tag{4.2}$$

Subtracting (4.2)(3) from (4.2)(2) we find

$$\frac{1}{B_{65}}\left[\mu_5\left(\frac{x_1}{x_6}\right)^2+\mu_6\left(\frac{x_1}{x_5}\right)^2\right]=\frac{\mu_4}{B_{54}}\left(\frac{x_3}{x_5}\right)^2+\frac{\mu_4}{B_{46}}\left(\frac{x_2}{x_6}\right)^2-\frac{d_2}{\mu_5}\left(\frac{x_0}{x_5}\right)^2+\frac{d_3}{\mu_6}\left(\frac{x_0}{x_6}\right)^2.$$

Substituting this expression into (4.2)(4) and using (4.2)(0) we obtain after some manipulation

$$(4.1)(5)\quad F^2+(a_4-a_1)(a_3-a_1)Gx_0^2=0$$

with F and G defined as in the lemma. It is now easy to check that in U_1 the four Eqs. (4.1)(0)(2)(3)(4) are equivalent to (4.1)(0)(2)(3)(5). This establishes Lemma 4.3.

Lemma 4.4. *$I=\bar{A}_c\cap\{x_0=0\}$ is an ordinary double curve of $\bar{A}_c$ except at 16 points which are ordinary pinch points. The reduced curve I_{red} is non singular and is exactly the elliptic curve E* (3.10) *which appears at the 0 step in the expansions* (3.14).

Proof. We analyze $I \cap U_1$. The other pieces are similar. By Lemma 4.3, $I \cap U_1$ is defined by

$$\begin{aligned}
&(0)\ x_1x_4 + x_2x_5 + x_3x_6 = 0\\
&(2)\ \mu_6 B_{65}\left(\frac{1}{B_{65}}\frac{x_1}{x_5}\right)^2 - \mu_4 B_{54}\left(\frac{1}{B_{54}}\frac{x_3}{x_5}\right)^2 + 1 = 0\\
&(3)\ -\mu_5 B_{65}\left(\frac{1}{B_{65}}\frac{x_1}{x_6}\right)^2 + \mu_4 B_{46}\left(\frac{1}{B_{46}}\frac{x_2}{x_6}\right)^2 + 1 = 0\\
&(4)\ (B_{46}x_1x_4 - B_{65}x_2x_5)^2 = 0.
\end{aligned} \tag{4.3}$$

From (4.3)(0) and (4.3)(4) we deduce

$$\frac{1}{B_{65}}\frac{x_1x_4}{x_5x_6} = \frac{1}{B_{46}}\frac{x_2}{x_6} = \frac{1}{B_{54}}\frac{x_3}{x_5}.$$

Defining

$$\hat{x}_4^0 = \frac{1}{B_{65}}\frac{x_1x_4}{x_5x_6}, \quad \hat{x}_5^0 = \frac{1}{B_{65}}\frac{x_1}{x_6}, \quad \hat{x}_6^0 = \frac{1}{B_{65}}\frac{x_1}{x_5}$$

the reduced curve $I_{\text{red}} \cap U_1$ can be parametrized as follows:

$$\frac{x_0}{x_1} = 0, \quad \frac{x_2}{x_1} = \frac{B_{46}\hat{x}_4^0}{B_{65}\hat{x}_5^0}, \quad \frac{x_3}{x_1} = \frac{B_{54}\hat{x}_4^0}{B_{65}\hat{x}_6^0}$$

$$\frac{x_4}{x_1} = \frac{\hat{x}_4^0}{B_{65}\hat{x}_5^0\hat{x}_6^0}, \quad \frac{x_5}{x_1} = \frac{1}{B_{65}\hat{x}_6^0}, \quad \frac{x_6}{x_1} = \frac{1}{B_{65}\hat{x}_5^0}$$

with $\hat{x}_4^0$, $\hat{x}_5^0$, $\hat{x}_6^0$ constrained by the equations

$$E: \begin{aligned} &\mu_6 B_{65}\hat{x}_6^{0^2} - \mu_4 B_{54}\hat{x}_4^{0^2} + 1 = 0\\ &-\mu_5 B_{65}\hat{x}_5^{0^2} + \mu_4 B_{46}\hat{x}_4^{0^2} + 1 = 0. \end{aligned}$$

These equations are exactly the equations of the elliptic curve E obtained in (3.10). Since the reduced curve $I_{\text{red}} = E$ is non singular, it follows that I is an ordinary double curve of $\bar{A}_c$ except at the points where $G=0$ [see (4.1)(5)] which are ordinary pinch points of $\bar{A}_c$ (i.e. $\bar{A}_c$ has a local analytic equation $x^2 = yz^2$). It is now easy to check that $G=0$ on I is equivalent to $R(\hat{x}_4^0, \hat{x}_5^0, \hat{x}_6^0) = 0$ on E with $R(\hat{x}_4^0, \hat{x}_5^0, \hat{x}_6^0)$ defined as in (3.17). From Proposition 3.6 we know that this expression has 16 zeroes on E, so the number of pinch points is 16. This finishes the proof of Lemma 4.4.

We now come to the

Proof of Theorem 3

Step 1. Construction of $\tilde{A}_c$. Define $U_{oj} \equiv U_j \cap \{x_0 \neq 0\}$ $(1 \leqq j \leqq 4)$. First we construct in each U_{oj} an algebraic change of coordinates. Let us begin with U_{01}. In this patch the time t and the parameters x_4^0, x_5^0, x_6^0, β involved in the expansions (3.14)

are easily approximated by

$$\tau=\frac{1}{x_1}=\frac{t}{x_1^0}+o(t^2),\qquad \hat{x}_4^0=\frac{1}{B_{65}}\frac{x_1x_4}{x_5x_6}=-x_4^0+o(t)$$
$$\hat{x}_5^0=\frac{1}{B_{65}}\frac{x_1}{x_6}=-x_5^0+o(t),\qquad \hat{x}_6^0=\frac{1}{B_{65}}\frac{x_1}{x_5}=-x_6^0+o(t)$$
$$\hat{\beta}=B_{46}x_4-B_{65}\frac{x_2x_5}{x_1}=\frac{\beta}{B_{65}x_5^0x_6^0}+o(t) \tag{4.4}$$
$$\hat{\beta}'=B_{65}\frac{x_3x_6}{x_1}-B_{54}x_4=\frac{\beta}{B_{65}x_5^0x_6^0}+o(t)$$

defining an algebraic change of coordinates

$$\phi_{01}: U_{01}=\mathbb{C}^6\cap\{x_1x_5x_6\neq 0\}\dashrightarrow U_{10}=\mathbb{C}^6\cap\{\tau\hat{x}_5^0\hat{x}_6^0\neq 0\}$$
$$(x_1,x_2,x_3,x_4,x_5,x_6)\rightsquigarrow(\tau,\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta},\hat{\beta}') \tag{4.5}$$

the inverse of which is given by

$$x_1=\frac{1}{\tau},\qquad x_2=\left(\frac{B_{46}\hat{x}_4^0}{\tau B_{65}\hat{x}_5^0\hat{x}_6^0}-\hat{\beta}\right)\hat{x}_6^0,\qquad x_3=\left(\frac{B_{54}\hat{x}_4^0}{\tau B_{65}\hat{x}_5^0\hat{x}_6^0}+\hat{\beta}'\right)\hat{x}_5^0$$
$$x_4=\frac{\hat{x}_4^0}{\tau B_{65}\hat{x}_5^0\hat{x}_6^0},\qquad x_5=\frac{1}{\tau B_{65}\hat{x}_6^0},\qquad x_6=\frac{1}{\tau B_{65}\hat{x}_5^0}. \tag{4.6}$$

According to Lemma 4.3, the affine piece $V_{01}\equiv A_c\cap U_{01}$ is defined by the 4 Eqs. (4.1)(0)(2)(3) and (4.1)(5). An easy computation shows that, in the new coordinates, these 4 equations are written as

$$\begin{aligned}
&(4.1)(0)\Leftrightarrow F_0^{(1)}\equiv\hat{\beta}'-\hat{\beta}-c_0B_{65}\tau=0\\
&(4.1)(2)\Leftrightarrow F_1^{(1)}\equiv\mu_6B_{65}\hat{x}_6^{0^2}-\mu_4B_{54}\hat{x}_4^{0^2}+1+\tau R(\tau,\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta})=0\\
&(4.1)(3)\Leftrightarrow F_2^{(1)}\equiv-\mu_5B_{65}\hat{x}_5^{0^2}+\mu_4B_{46}\hat{x}_4^{0^2}+1+\tau S(\tau,\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta})=0\\
&(4.1)(5)\Leftrightarrow F_3^{(1)}\equiv\hat{\beta}^2+\frac{B_{54}B_{46}}{B_{65}}\left[\frac{d_2}{\mu_4\mu_5}\frac{1}{\hat{x}_5^{0^2}}-\frac{d_3}{\mu_4\mu_6}\frac{1}{\hat{x}_6^{0^2}}+\frac{d_4B_{65}}{\mu_4}-2c_0\frac{\hat{x}_4^0}{\hat{x}_5^0\hat{x}_6^0}\right]\\
&\qquad\qquad-\tau c_0B_{46}[2\hat{\beta}'-c_0B_{65}\tau]=0
\end{aligned} \tag{4.7}$$

with R and S polynomial expressions in τ, $\hat{x}_4^0$, $\hat{x}_5^0$, $\hat{x}_6^0$, $\hat{\beta}$. Putting $\tau=0$ in the Eqs. (4.7), they reduce to the equations of an affine piece $\mathscr{D}_1\equiv\mathscr{D}\cap\{x_5^0x_6^0\neq 0\}$ of the genus 9 curve $\mathscr{D}$ found in (3.17). Since generically the curve $\mathscr{D}$ is non singular we have

$$rk\left.\frac{\partial(F_0^{(1)},F_1^{(1)},F_2^{(1)},F_3^{(1)})}{\partial(\tau,\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta},\hat{\beta}')}\right|_{\tau=0}=rk\left.\frac{\partial(F_0^{(1)},F_1^{(1)},F_2^{(1)},F_3^{(1)})}{\partial(\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta},\hat{\beta}')}\right|_{\tau=0}=4 \tag{4.8}$$

which shows that the affine variety V_1 defined by

$$V_1=\bigcap_{k=0}^{3}\{F_k^{(1)}(\tau,\hat{x}_4^0,\hat{x}_5^0,\hat{x}_6^0,\hat{\beta},\hat{\beta}')=0\}\subseteq\mathbb{C}^6\cap\{\hat{x}_5^0\hat{x}_6^0\neq 0\}$$

is non singular along the piece $\mathscr{D}_1 \equiv V_1 \cap \{\tau = 0\}$ added at infinity.

The whole process of compactification of A_c is achieved by performing similar algebraic changes of coordinates in U_{02}, U_{03}, and U_{04}. Nothing is special about U_{02} and U_{03}. Namely the expansions (3.14) lead to algebraic changes of coordinates

$$\phi_{oj} : U_{oj} \xrightarrow{\sim} U_{jo} : (x_i) \rightsquigarrow (\tau, \hat{x}_4^0, \hat{x}_5^0, \hat{x}_6^0, \hat{\beta}, \hat{\beta}') \qquad (j = 2, 3)$$

with τ, $\hat{x}_4^0$, $\hat{x}_5^0$, $\hat{x}_6^0$, $\hat{\beta}$, $\hat{\beta}'$ suitable rational approximations of t, x_4^0, x_5^0, x_6^0, β as in (4.4). Using again Lemma 4.3, the pieces $V_{oj} \equiv A_c \cap U_{oj}$ $(j = 2, 3)$ are described by algebraic equations $F_k^{(j)}(\tau, \hat{x}_4^0, \hat{x}_5^0, \hat{x}_6^0, \hat{\beta}, \hat{\beta}') = 0$, $\tau \neq 0$ $(0 \leqq k \leqq 3, j = 2, 3)$ defining affine varieties $V_j = \bigcap_{k=0}^{3} \{F_k^{(j)} = 0\}$ which are non singular along $\mathscr{D}_j \equiv V_j \cap \{\tau = 0\}$. Up to now we have glued the affine piece $\mathscr{D}^a = \mathscr{D}_1 \cup \mathscr{D}_2 \cup \mathscr{D}_3 = \mathscr{D} \setminus \{8 \text{ points at infinity}\}$ to A_c. By construction (4.4) $X_{Q_{a_4}}$ is regular along $\mathscr{D}^a$. Moreover $\dot{\tau} \neq 0$ along $\mathscr{D}^a$ which shows that $X_{Q_{a_4}}$ is transversal to $\mathscr{D}^a$ at every point. Because of the Riemann-Roch theorem, the last piece $\mathscr{D}_4$ to be glued to A_c will thus contain the 8 points where $X_{Q_{a_4}}$ must become tangent to $\mathscr{D}$[7]. Since the expansions (3.14) do not "see" these points, it is appropriate to use another independent commuting Hamiltonian vector field to define regularizing variables around these points. The reader will easily check that the expansions corresponding to $X_{Q_{a_1}}$ (or more simply the symmetry of the problem) lead to the following algebraic change of coordinates in U_{04}:

$$\phi_{04} : U_{04} = \mathbb{C}^6 \cap \{x_1 x_2 x_3 \neq 0\} \xrightarrow{\sim} U_{40} = \mathbb{C}^6 \cap \{\tau \hat{x}_2^0 \hat{x}_3^0 \neq 0\}$$

$$(x_1, x_2, x_3, x_4, x_5, x_6) \rightsquigarrow (\tau, \hat{x}_2^0, \hat{x}_3^0, \hat{x}_4^0, \hat{\beta}, \hat{\beta}')$$

$$\tau = \frac{1}{x_1}, \qquad \hat{x}_2^0 = \frac{1}{B}\frac{x_1}{x_3}, \qquad \hat{x}_3^0 = \frac{1}{B}\frac{x_1}{x_2}, \qquad \hat{x}_4^0 = \frac{1}{B}\frac{x_1 x_4}{x_2 x_3}$$

$$\hat{\beta} = \frac{(a_4 - a_2)}{(a_1 - a_4)(a_1 - a_2)} x_4 - B\frac{x_2 x_5}{x_1}, \qquad \hat{\beta}' = B\frac{x_3 x_6}{x_1} - \frac{(a_3 - a_4)}{(a_1 - a_3)(a_1 - a_4)} x_4$$

$$\text{with} \quad B = \frac{(a_2 - a_3)}{(a_1 - a_2)(a_1 - a_3)}.$$

By Lemma 4.3 we obtain then a $V_4 = \bigcap_{k=0}^{3} \{F_k^{(4)}(\tau, \hat{x}_2^0, \hat{x}_3^0, \hat{x}_4^0, \hat{\beta}, \hat{\beta}') = 0\}$ which is non singular along $\mathscr{D}_4 = V_4 \cap \{\tau = 0\}$. Let $\tilde{A}_c$ be the algebraic variety defined by the atlas of affine varieties $\{V_0 = A_c, V_1, V_2, V_3, V_4\}$. The transitions $\tilde{\phi}_{oj} : V_0 = A_c \supset V_{oj} \xrightarrow{\sim} V_{jo} \subset V_j$ are induced by the $\phi_{oj} : U_{oj} \xrightarrow{\sim} U_{jo}$ (4.5) $(1 \leqq j \leqq 4)$ while the $\tilde{\phi}_{jk} : V_j \supset V_{jk} \xrightarrow{\sim} V_{kj} \subset V_k$ $(1 \leqq j, k \leqq 4)$ are defined by observing that the map $\tilde{\phi}_{ok} \circ \tilde{\phi}_{oj}^{-1}$ extends to infinity by the identity if $1 \leqq j$, $k \leqq 3$ and by the algebraic change of coordinates at infinity on $\mathscr{D}$ (3.19) if j or $k = 4$. We know that $\tilde{A}_c$ is non singular along $\mathscr{D}$. We now prove that it is compact. Using formulae like (4.6), it is easy to define a map $\tilde{p} : \tilde{A}_c \to \bar{A}_c$ such that $\tilde{p}|A_c = \text{identity}$ and $\tilde{p}|\mathscr{D}$ is the 2-sheeted ramified covering $\mathscr{D} \to E$ (3.18). Since, by Lemma 4.4, the (reduced) curve at infinity on $\bar{A}_c$ is exactly the elliptic curve E, $\tilde{p}$ is surjective and so $\tilde{A}_c$ is compact.

7 This is because $\dim \mathscr{L}(\mathscr{D}) = \text{number of points on } \mathscr{D} \text{ where } X_{Q_{a_4}} \text{ is tangent to } \mathscr{D} = g(\mathscr{D}) - 1 = 8$

Step 2. $\tilde{A}_c$ is an abelian variety. To prove that $\tilde{A}_c$ is an abelian variety we must still show that, for example, the two commuting vector fields $Y_1 = X_{Q_{a_4}}$ and $Y_2 = X_{Q_{a_1}}$ extend holomorphically and remain independent at every point of $\mathscr{D}$. Let us denote by φ_1^t and φ_2^t the corresponding flows on the affine part A_c.

1) *Regularity.* As observed in Step 1, φ_1 is by construction regular and transversal to the divisor along $\mathscr{D}^a = \mathscr{D}_1 \cup \mathscr{D}_2 \cup \mathscr{D}_3$. The same is true for φ_2 along $\mathscr{D}_2 \cup \mathscr{D}_3 \cup \mathscr{D}_4$. Let $p \in \mathscr{D}_1$. For ε fixed and small enough, $\varphi_1^\varepsilon(p)$ is well defined and $\varphi_1^\varepsilon(p) \in A_c$. Then φ_2 can be holomorphically extended at p by $\varphi_2^t(p) = \varphi_1^{-\varepsilon} \circ \varphi_2^t \circ \varphi_1^\varepsilon(p)$. Indeed, $\varphi_2^t \circ \varphi_1^\varepsilon(p)$ is holomorphic around $t=0$ and, since φ_1^ε maps bi-holomorphically a neighbourhood of p onto one of $\varphi_1^\varepsilon(p)$, it follows that also $\varphi_2^t(p)$ is holomorphic around $t=0$. Repeating the same argument for φ_1, proves then that φ_1 and φ_2 are regular all along $\mathscr{D}$.

2) *Independence.* Consider first the affine piece $U \equiv \mathscr{D} \cap \{x_4^0 x_5^0 x_6^0 \beta\} \neq 0$. Around a point $p \in U$, we may take as local coordinates on $\tilde{A}_c$

$$\tau = \frac{1}{x_1} \quad \text{and} \quad \hat{\gamma} = \hat{x}_6^0 = \frac{1}{B_{65}} \frac{x_1}{x_5}.$$

Define $J(\tau, \hat{\gamma})$ to be the matrix

$$J(\tau, \hat{\gamma}) = \begin{pmatrix} Y_1(\tau) & Y_2(\tau) \\ Y_1(\hat{\gamma}) & Y_2(\hat{\gamma}) \end{pmatrix}$$

where $Y_j(\tau)$, $Y_j(\hat{\gamma})$ denote the derivatives of τ, $\hat{\gamma}$ in the directions Y_j $(j=1,2)$. We compute

$$|J(\tau, \hat{\gamma})|_{|\tau=0} = \det \begin{pmatrix} -\dfrac{1}{B_{65} x_5^0 x_6^0} & -\dfrac{\mu_4^2 (x_4^0)^2}{B_{65} x_5^0 x_6^0} \\ \dfrac{\mu_4 x_4^0 \beta}{B_{65} x_5^0} & -\dfrac{\mu_4^2 x_4^0 \beta}{B_{46} B_{65} x_5^0} \end{pmatrix} \tag{4.9}$$

$$= \mu_4^2 \mu_5 (B_{46} B_{65} x_6^0)^{-1} x_4^0 \beta \neq 0 \quad \text{along } U.$$

One could now check that Y_1 and Y_2 are also independent at every point of $\mathscr{D} \backslash U$ by choosing appropriate local coordinates around these points. However the following argument will be probably more illuminating. Suppose there is a point $q \in \mathscr{D} \backslash U$ such that Y_1 and Y_2 become dependent at q. Since $\mathscr{D}$ is non singular, it is always possible to introduce local coordinates $(\tau', \hat{\gamma}')$ in a neighbourhood V of q with $\dfrac{\partial}{\partial \tau'}$ a direction transversal to $\mathscr{D}$. Choose $p \in V \cap U$. By the compactness of $\tilde{A}_c$ the map

$$\varphi : \mathbb{C}^2 \to \tilde{A}_c : (t_1, t_2) \rightsquigarrow \varphi(t_1, t_2) = \varphi_1^{t_1} \circ \varphi_2^{t_2}(p) \tag{4.10}$$

is well defined and, since Y_1 and Y_2 are independent at p, φ maps a neighbourhood of the origin in $\mathbb{C}^2$ bi-holomorphically onto a neighbourhood $W \subset V$ of p, $q \notin W$. In particular dt_1 and dt_2 are well defined holomorphic differential 1-forms in W which can be computed as

$$\begin{pmatrix} dt_1 | W \\ dt_2 | W \end{pmatrix} = J^{-1}(\tau, \hat{\gamma}) \begin{pmatrix} d\tau \\ d\hat{\gamma} \end{pmatrix} = J^{-1}(\tau', \hat{\gamma}') \begin{pmatrix} d\tau' \\ d\hat{\gamma}' \end{pmatrix}. \tag{4.11}$$

Since $d\tau|\mathscr{D}\cap W = d\tau'|\mathscr{D}\cap W = 0$ we obtain

$$dt_1|\mathscr{D}\cap W = -(|J(\tau',\hat{\gamma}')|^{-1}Y_2(\tau'))_{|\tau'=0}d\hat{\gamma}'$$
$$dt_2|\mathscr{D}\cap W = (|J(\tau',\hat{\gamma}')|^{-1}Y_1(\tau'))_{|\tau'=0}d\hat{\gamma}'.$$

By construction we know that at least one of the two vector fields Y_1 or Y_2 is transversal to $\mathscr{D}$ at q, i.e. $Y_1(\tau')$ or $Y_2(\tau')\neq 0$ at q. Thus if Y_1 and Y_2 become dependent at q, $|J(\tau',\hat{\gamma}')|=0$ at q and at least one of the two holomorphic differentials $dt_1|\mathscr{D}\cap W$ or $dt_2|\mathscr{D}\cap W$ cannot be holomorphically extended to $\mathscr{D}\cap V$. Conversely if we prove that, around any point $p\in U$, these two differentials extend holomorphically to $\mathscr{D}$, Y_1 and Y_2 must be everywhere independent. Using (4.9), (4.11) and the relation $\mu_4 B_{54}x_4^0 dx_4^0 = \mu_6 B_{65}x_6^0 dx_6^0$ on $\mathscr{D}$, we compute

$$\begin{aligned} dt_1|\mathscr{D}\cap W &= k_1\frac{(x_4^0)^3dx_4^0}{x_4^0x_5^0x_6^0\beta} = \frac{k_1}{2}\frac{(x_4^0)^2d(x_4^0)^2}{x_4^0x_5^0x_6^0\beta} \\ dt_2|\mathscr{D}\cap W &= k_2\frac{x_4^0dx_4^0}{x_4^0x_5^0x_6^0\beta} = \frac{k_2}{2}\frac{d(x_4^0)^2}{x_4^0x_5^0x_6^0\beta} \end{aligned} \tag{4.12}$$

with k_1 and k_2 some constants depending on the a_i $(1\leqq i\leqq 4)$. This shows that $dt_1|\mathscr{D}\cap W$ and $dt_2|\mathscr{D}\cap W$ extend holomorphically to $\mathscr{D}$ by $k_1'\tilde{\pi}^*(\psi_1')$ and $k_2'\tilde{\pi}^*(\psi_2')$ (k_1', k_2' appropriate constants) where $\tilde{\pi}$ is the covering map $\tilde{\pi}:\mathscr{D}\to C$ defined in (3.23) and ψ_1', ψ_2' are the two holomorphic differentials on C computed in (3.26)

$$\psi_1' = \frac{\alpha_1 d\alpha_1}{\zeta\beta} \quad \text{and} \quad \psi_2' = \frac{d\alpha_1}{\zeta\beta}$$

which satisfy $\iota^*(\psi_j') = -\psi_j'$ $(j=1,2)$ for the involution

$$\iota: C\to C: (\beta,\zeta,\alpha_1)\curvearrowright(-\beta,\zeta,\alpha_1).$$

As a consequence Y_1 and Y_2 are everywhere independent on $\tilde{A}_c$ and the map φ (4.10) is onto. Dividing $\mathbb{C}^2$ by the lattice of points (t_1,t_2) which are mapped into the base point p and using the compactness of $\tilde{A}_c$, it follows that $\tilde{A}_c$ is analytically isomorphic to a complex torus $\mathbb{C}^2/L_\Omega$ and so $\tilde{A}_c$ is an abelian variety. Clearly Y_1 and Y_2 are straight line motions on $\mathbb{C}^2/L_\Omega$. Moreover dt_1 and dt_2 are now well defined holomorphic differentials on $\tilde{A}_c$ which can be computed locally by formulae like (4.11) so that their restrictions to the divisor $\mathscr{D}$ are given by (4.12). This ends the proof of Theorem 3.

Corollary 4.5. *The vector space of holomorphic differential* 1*-forms on* $\tilde{A}_c$ *restricted to* $\mathscr{D}$ *is generated by* $\tilde{\pi}^*(\psi_1)$ *and* $\tilde{\pi}^*(\psi_2)$ *with* $\tilde{\pi}:\mathscr{D}\to C$ *defined as in* (3.23) *and* ψ_1,ψ_2 *given by* (3.25) *(or* (3.26)*), a basis of holomorphic differentials on* C *satisfying* $\iota^*(\psi_j)=-\psi_j$ *$(j=1,2)$ for the involution* ι *interchanging sheets of the covering* $p: C\to C_0$ (3.21).

We now turn to the

Proof of Theorem 4. (i) Let dt_1, dt_2 be a basis of holomorphic differential 1-forms on $\tilde{A}_c$ such that $dt_1|\mathscr{D}=\tilde{\pi}^*(\psi_1)$ and $dt_2|\mathscr{D}=\tilde{\pi}^*(\psi_2)$ as in Corollary 4.5. Consider the uniformizing map

$$\tilde{A}_c\to\mathbb{C}^2/L_\Omega: p\curvearrowright\int_{p_0}^{p}\begin{pmatrix}dt_1\\ dt_2\end{pmatrix}.$$

The period lattice L_Ω is by definition

$$L_\Omega = \left\{ \sum_{i=1}^{4} n_i \int_{\gamma_i} \begin{pmatrix} dt_1 \\ dt_2 \end{pmatrix} \middle| n_i \in \mathbb{Z}, \gamma_i \ (1 \leqq i \leqq 4) \text{ a basis of } H_1(\tilde{A}_c, \mathbb{Z}) \right\}.$$

Now, by the Lefschetz hyperplane section theorem (see [7] p. 156) the inclusion map $i: \mathscr{D} \hookrightarrow \tilde{A}_c$ induces a surjective map i_* in 1-dimensional homology $i_*: H_1(\mathscr{D}, \mathbb{Z}) \to H_1(\tilde{A}_c, \mathbb{Z})$. This means we can find 4 cycles γ_i $(1 \leqq i \leqq 4)$ in the algebraic curve $\mathscr{D}$ such that

$$L_\Omega = \left\{ \sum_{1}^{4} n_i \int_{\gamma_i} \begin{pmatrix} dt_1|\mathscr{D} \\ dt_2|\mathscr{D} \end{pmatrix} \right\} = \left\{ \sum_{1}^{4} n_i \int_{\gamma_i} \begin{pmatrix} \tilde{\pi}^*(\psi_1) \\ \tilde{\pi}^*(\psi_2) \end{pmatrix} \right\}.$$

To pick such 4 cycles seems arduous in view of the fact that $\mathscr{D}$ is a genus 9 curve. This is overcome by exhibiting 4 cycles γ_i in $\mathscr{D}$ such that

$$\Omega = (\tilde{\pi}^*(\psi_j)(\gamma_i))_{\substack{1 \leqq j \leqq 2 \\ 1 \leqq i \leqq 4}} \equiv \left(\int_{\gamma_i} \tilde{\pi}^*(\psi_j) \right)_{\substack{1 \leqq j \leqq 2 \\ 1 \leqq i \leqq 4}}$$

is a Riemann matrix. Indeed, if so, the γ_i $(1 \leqq i \leqq 4)$ must be a basis of $H_1(\tilde{A}_c, \mathbb{Z})$. Otherwise there would exist a non trivial linear combination of the γ_i with integer coefficients such that $\sum_1^4 n_i \gamma_i = \partial\sigma$, σ a 2-chain on $\tilde{A}_c$. Since dt_1 and dt_2 are closed differentials on $\tilde{A}_c$, this implies

$$\sum_{1}^{4} n_i \int_{\gamma_i} \begin{pmatrix} dt_1|\mathscr{D} \\ dt_2|\mathscr{D} \end{pmatrix} = \int_{\partial\sigma} \begin{pmatrix} dt_1 \\ dt_2 \end{pmatrix} = 0$$

which contradicts the linear independence of the columns of Ω.

Remember now Fig. 3 (Sect. 3) and the diagram (3.22). Let a, A, a', b, B, b' be a canonical homology basis of the genus 3 curve C such that $\iota(a) = a'$, $\iota(b) = b'$, $p(A)$, and $p(B)$ are homologous to zero on C_0. Then we claim that the 4 cycles $2a$, A, $2b$, B in $\mathscr{D}$, represented in Fig. 3 (Sect. 3) generate $H_1(\tilde{A}_c, \mathbb{Z})$. As observed before it is sufficient to prove that the matrix

$$\Omega = \begin{pmatrix} 2\psi_1(a) & \psi_1(A) & 2\psi_1(b) & \psi_1(B) \\ 2\psi_2(a) & \psi_2(A) & 2\psi_2(b) & \psi_2(B) \end{pmatrix} \tag{4.13}$$

is a Riemann matrix. Now Ω is nothing but the period matrix of the Prym variety $\mathrm{Prym}_\iota(C)$ for the involution ι interchanging sheets of the covering $p: C \to C_0$ (3.21)[8]. Thus $\tilde{A}_c$ and $\mathrm{Prym}_\iota(C)$ are two abelian varieties analytically isomorphic to the same complex torus $\mathbb{C}^2/L_\Omega$. By Chow's theorem, $\tilde{A}_c$ and $\mathrm{Prym}_\iota(C)$ are then algebraically isomorphic.

(ii) Since (for fixed values of the constants of the motion) the x_i $(1 \leqq i \leqq 7)$ are abelian functions on $\mathbb{C}^2/L_\Omega$ with a simple pole along the divisor $\mathscr{D}$, there exists 8 theta-functions θ_i, $0 \leqq i \leqq 7$, associated to this divisor such that $x_i = \dfrac{\theta_i}{\theta_0}$ $(1 \leqq i \leqq 7)$.

8 See Appendix, Theorem 1

From the definition of x_7 (3.27) we deduce

$$(a_3-a_1)(a_4-a_2)\theta_1\theta_4+(a_3-a_2)(a_4-a_1)\theta_2\theta_5=\theta_0\theta_7. \tag{4.14}$$

Substituting this expression into (4.1)(5) we find

$$\theta_7^2+(a_4-a_1)(a_3-a_1)\begin{bmatrix} d_2(a_2-a_4)(a_3-a_2)(a_2-a_1)\theta_6^2 \\ +d_3(a_4-a_3)(a_3-a_2)(a_2-a_1)\theta_5^2 \\ +d_4(a_4-a_2)(a_3-a_4)(a_2-a_1)\theta_1^2 \\ +2c_0(a_4-a_2)(a_3-a_2)\theta_3\theta_6 \end{bmatrix} = \begin{matrix} c_0^2(a_3-a_1)(a_3-a_2) \\ (a_4-a_1)(a_4-a_2)\theta_0^2. \end{matrix} \tag{4.15}$$

From the proof of Theorem 3 (formula 4.7) we know that the 6 equations $Q_0(\theta)=c_0\theta_0^2$, $Q_{a_i}(\theta)=d_i\theta_0^2$ $(2\leqq i\leqq 4)$, (4.14) and (4.15) suffice to describe $\tilde{A}_c$ in the neighbourhood of $\mathscr{D}$ at least in the patch $\theta_1\theta_5\theta_6\neq 0$. Similar equations like (4.15) involving in the bracket $(\theta_2^2,\theta_4^2,\theta_6^2)$ or $(\theta_3^2,\theta_4^2,\theta_5^2)$ or $(\theta_1^2,\theta_2^2,\theta_3^2)$ instead of $(\theta_1^2,\theta_5^2,\theta_6^2)$ can be obtained, but are all equivalent modulo $Q_{a_i}(\theta)=d_i\theta_0^2$ $(2\leqq i\leqq 4)$. This observation makes clear that the 6 above quadratic equations suffice to describe $\tilde{A}_c\approx\mathbb{C}^2/L_\Omega$ as embedded into $\mathbb{P}^7(\mathbb{C})$. This establishes Theorem 4.

We end this section with the proof of Lemmas 4.1 and 4.2.

Proof of Lemma 4.1. Suppose for example that $x_1=x_4=0$. Putting $x_0=0$, Eqs. (4.1)(0)(1)(2)(3) become

$$\begin{aligned} &(0)\ x_2x_5+x_3x_6=0 && (2)\ \frac{x_3^2}{a_2-a_1}+\frac{x_5^2}{a_2-a_4}=0 \\ &(1)\ \frac{x_2^2}{a_1-a_3}+\frac{x_3^2}{a_1-a_2}=0 && (3)\ \frac{x_2^2}{a_3-a_1}+\frac{x_6^2}{a_3-a_4}=0. \end{aligned} \tag{4.16}$$

Since these are projective equations we must have $x_2x_3x_5x_6\neq 0$. Multiplying (4.16)(3) by x_3^2 and using (4.16)(0) we find

$$\frac{x_3^2}{a_3-a_1}+\frac{x_5^2}{a_3-a_4}=0.$$

This equation combined with (4.16)(2) leads to

$$\frac{x_3^2}{x_5^2}=\frac{a_3-a_1}{a_4-a_3}=\frac{a_2-a_1}{a_4-a_2}$$

which is incompatible with $\prod_{i<j}(a_i-a_j)\neq 0$. A similar argument is used for the other cases. This ends the proof of Lemma 4.1.

Proof of Lemma 4.2. More precisely, we show that the only points in $I\cap\left\{\prod_{i=1}^{6}x_i=0\right\}$ are such that ($x_2=x_3=x_4=0$ and $x_1x_5x_6\neq 0$) or ($x_1=x_3=x_5=0$ and $x_2x_4x_6\neq 0$) or ($x_1=x_2=x_6=0$ and $x_3x_4x_5\neq 0$) or ($x_4=x_5=x_6=0$ and $x_1x_2x_3\neq 0$). The reader will easily check that this amounts to 16 points. Suppose $x_1=0$. Then, by Lemma 4.1, $x_4\neq 0$ which implies, because of (4.1)(1) with $x_0=0$,

$x_2 \neq 0$ or $x_3 \neq 0$. Putting $x_0 = 0$, Eqs. (4.1)(0)(2)(3) become

$$(0)\ x_2x_5 + x_3x_6 = 0; \quad (2)\ \frac{x_3^2}{a_2 - a_1} + \frac{x_5^2}{a_2 - a_4} = 0; \quad (3)\ \frac{x_2^2}{a_3 - a_1} + \frac{x_6^2}{a_3 - a_4} = 0. \tag{4.17}$$

If $x_2 \neq 0$, (4.17)(3) implies $x_6 \neq 0$ and thus $x_2x_4x_6 \neq 0$. Multiplying (4.17)(3) by x_3^2 and using (4.17)(0) we obtain

$$\frac{x_3^2}{a_3 - a_1} + \frac{x_5^2}{a_3 - a_4} = 0.$$

If $x_3 \neq 0$ or $x_5 \neq 0$, this equation is incompatible with (4.17)(2). Thus we must have ($x_1 = x_3 = x_5 = 0$ and $x_2x_4x_6 \neq 0$). If $x_3 \neq 0$, we find in a similar way ($x_1 = x_2 = x_6 = 0$ and $x_3x_4x_5 \neq 0$). Now, the whole process can be repeated starting with $x_4 = 0$ which leads to the two last possibilities ($x_2 = x_3 = x_4 = 0$ and $x_1x_5x_6 \neq 0$) or ($x_4 = x_5 = x_6 = 0$ and $x_1x_2x_3 \neq 0$). This finishes the proof of Lemma 4.2.

5. The Relation Between Prym(C/C_0) and Prym(K/K_0)

We now prove the

Theorem 5. *Let* Prym(K/K_0) *be the abelian variety on which the Euler-Arnold equations have been linearized in Sect. 1, Theorem 2. Then* Prym(K/K_0) *is not isomorphic to* Prym(C/C_0) *but only isogenous, the precise relation between these two abelian varieties being that they are dual of each other.*

Proof. Using the definition of the coordinates at infinity (4.4), it is easy to see that the group action G defined on the affine part A_c of $\tilde{A}_c$ by (1.17) extends to the divisor $\mathscr{D}$ by

$$\begin{array}{cccc} x_4^0 & x_5^0 & x_6^0 & \beta \\ + & + & + & + \\ - & - & + & + \\ - & + & - & + \\ + & - & - & + \end{array}. \tag{5.1}$$

Let us denote by $\tilde{G}$ the action of G extended to $\tilde{A}_c$. It is easy to check that, for example, the two vector fields $X_{Q_{a_1}}$ and $X_{Q_{a_4}}$ are $\tilde{G}$-invariant. This implies that each $\tilde{\sigma} \in \tilde{G}$ is in fact a translation $\tilde{\sigma}: \tilde{A}_c \to \tilde{A}_c : p \curvearrowright p \oplus a$ with $2a = 0$, since $\tilde{\sigma}^2 = 1$. The quotient variety $\tilde{A}_c/\tilde{G}$ is thus an abelian variety and the covering map $\tilde{\Pi}: \tilde{A}_c \to \tilde{A}_c/\tilde{G}$ is an isogeny of degree 4. Since $\tilde{A}_c/\tilde{G}$ and Prym(K/K_0) are two abelian varieties which each contain a Zariski open set isomorphic to A_c/G, they must be isomorphic. Now, from (5.1) it is easy to see that $\tilde{\Pi}$ restricted to $\mathscr{D}$ is nothing but the covering $\tilde{\Pi}: \mathscr{D} \to C$ defined in (3.23). Remembering Corollary 4.5, it is then clear that we can find two holomorphic differential 1-forms ω_1 and ω_2 on $\tilde{A}_c/\tilde{G}$ such

that $\omega_1|C=\psi_1$ and $\omega_2|C=\psi_2$ with ψ_1 and ψ_2 the two holomorphic differentials on C such that $\iota^*(\psi_j)=-\psi_j$ defined in (3.25). Applying Lefschetz theorem to the hyperplane section $C\hookrightarrow\tilde{A}_c/\tilde{G}$, we find that the period matrix Ω^* of $\tilde{A}_c/\tilde{G}\approx\mathrm{Prym}(K/K_0)$ is given by

$$\Omega^*=\begin{pmatrix}\psi_1(a) & \psi_1(A) & \psi_1(b) & \psi_1(B)\\ \psi_2(a) & \psi_2(A) & \psi_2(b) & \psi_2(B)\end{pmatrix}$$

with a, A, b, B the 4 cycles of C defined as in the proof of Theorem 4. Now Ω^* is the period matrix of the dual abelian variety of $\mathrm{Prym}(C/C_0)$[9], which establishes Theorem 5.

Remarks. 1. Remembering Theorem 2 (Sect.1), we know that $\mathrm{Prym}_\iota(K)\backslash\mathscr{U}=\Theta\cap\mathrm{Prym}_\iota(K)$, where $\mathscr{U}$ is an open affine set of $\mathrm{Prym}_\iota(K)$ isomorphic to A_c/G and Θ is some translate of the theta-divisor of $\mathrm{Jac}(K)$. Combining this with the above proof, we obtain the interesting fact that the dual abelian variety of $\mathrm{Prym}_\iota(K)$ is $\mathrm{Prym}_\iota(C)$ where $C=\Theta\cap\mathrm{Prym}_\iota(K)$. Also it is easy to see that the involution ι on C is induced by the involution $\iota:\mathrm{Jac}(K)\to\mathrm{Jac}(K)$ coming from the involution ι on K.

2. When $c_0=0$, i.e. $Q_0(X)=0$, Moser linearizes the Euler-Arnold equations (3.1) on the Jacobi variety of the genus 2 hyperelliptic curve

$$Y: v^2+(c_1z^2-c_2z+c_3)\prod_{i=1}^{4}(z-a_i)=0$$

and shows that the x_i^2 are abelian functions on $\mathrm{Jac}(Y)$ (see [13], p. 182). The connection with our approach goes as follows. Consider the genus 3 curve C in (w, y, z) coordinates (3.24) and the differentials $\psi_1=\dfrac{zdz}{wy}$ and $\psi_2=\dfrac{dz}{wy}$ (3.25) on it such that $\iota^*(\psi_j)=-\psi_j$ $(j=1,2)$. When $c_0\to 0$, C can be viewed as a 4-sheeted ramified covering of the z-plane:

$$w=\pm\left[-(c_1z^2-c_2z+c_3)\pm 2c_0\sqrt{\prod_{i=1}^{4}(z-a_i)}\right]^{1/2}$$

with branch points $a_1, a_2, a_3, a_4, p_1^+, p_2^+, p_1^-, p_2^-$ and homology cycles a, A, a', b, B, b' as in Fig. 4. The period matrix Ω (4.13) becomes:

$$\Omega=\begin{pmatrix}4\int_{a_3}^{a_4}\frac{zdz}{yw} & 2\int_{p_1^+}^{p_2^+}\frac{zdz}{yw} & 4\int_{a_3}^{a_2}\frac{zdz}{yw} & 2\int_{p_1^+}^{a_2}+2\int_{p_1^-}^{a_2}\frac{zdz}{yw}\\ 4\int_{a_3}^{a_4}\frac{dz}{yw} & 2\int_{p_1^+}^{p_2^+}\frac{dz}{yw} & 4\int_{a_3}^{a_2}\frac{dz}{yw} & 2\int_{p_1^+}^{a_2}+2\int_{p_1^-}^{a_2}\frac{dz}{yw}\end{pmatrix}. \tag{5.2}$$

When $c_0=0$, one has obviously $p_1^+=p_1^-$, $p_2^+=p_2^-$, and $yw=v$, which shows that $\mathrm{Prym}_\iota(C)$ is then an 8–1 unramified covering of $\mathrm{Jac}(Y)$. In particular, since only 3 periods of $\mathrm{Jac}(Y)$ are multiplied by 2, $\tilde{A}_c=\mathrm{Prym}_\iota(C)$ is not isomorphic to $\mathrm{Jac}(Y)$.

9 See Appendix, Corollary 1

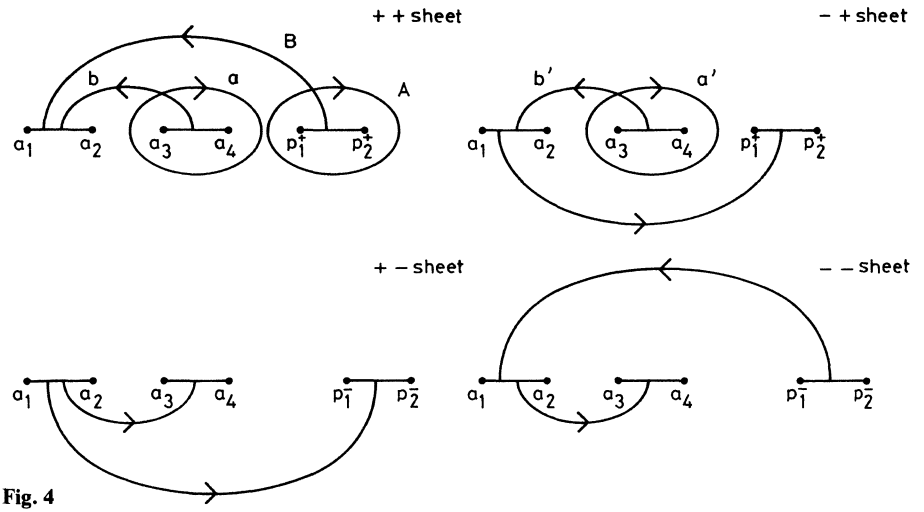


Fig. 4

Appendix

Since the concept of a Prym variety appears only in specialized papers in algebraic geometry, we explain it in this appendix.

Definition 1. Let $p: C \to C_0$ be a double covering with $2n$ branch points, where C and C_0 are non singular compact algebraic curves. Let g_0 be the genus of C_0. Then by Hurwitz formula, the genus of C is $g = 2g_0 + n - 1$. Let ι be the involution $\iota: C \to C$ which interchanges the two sheets. Fix $0 = \sum_{j=1}^{g} o_j$ some origin in the Jacobi variety $\mathrm{Jac}(C)$ of C. Then the Prym variety $\mathrm{Prym}(C/C_0)$ [or $\mathrm{Prym}_\iota(C)$] of C over C_0 (quotient of C by ι) may be defined as the set of $D \in \mathrm{Jac}(C)$, D a divisor $\sum_{j=1}^{g} p_j$, considered modulo linear equivalence, such that

$$\int_0^D \omega + \int_{0'}^{D'} \omega = \sum_{j=1}^{g} \int_{o_j}^{p_j} \omega + \sum_{j=1}^{g} \int_{o_j'}^{p_j'} \omega = 0 \pmod{\text{periods}}$$

where $\omega = (\omega_1, \ldots, \omega_g)^T$ is a basis of holomorphic differentials on C.

Observe that every holomorphic differential on C_0 can be lifted to a holomorphic differential on C which is invariant under ι. Therefore it is possible to find a basis

$$\omega = \begin{pmatrix} \theta \\ \psi \end{pmatrix} = (\theta_1, \ldots, \theta_{g_0}, \psi_1, \ldots, \psi_{g-g_0})^T$$

such that $\iota^*(\theta_j) = \theta_j$ and $\iota^*(\psi_j) = -\psi_j$. Consider now a canonical homology basis of C

$$a_1, \ldots, a_{g_0}, A_1, \ldots, A_{n-1}, a_1', \ldots, a_{g_0}', b_1, \ldots, b_{g_0}, B_1, \ldots, B_{n-1}, b_1', \ldots, b_{g_0}'$$

such that $\iota(a_j)=a'_j$, $\iota(b_j)=b'_j$, $\iota(A_j)=-A_j$, $\iota(B_j)=-B_j$, $p(A_j)$, and $p(B_j)$ are homologous to zero on C_0.

We shall use the notation $\psi(A)$ etc... to design the $(g-g_0, n-1)$ matrix

$$\psi(A)=(\psi_j(A_k))_{\substack{1\leqq j\leqq g-g_0\\ 1\leqq k\leqq n-1}}\equiv\left(\int_{A_k}\psi_j\right)_{j,k} \text{ etc. ...}$$

We have then the

Theorem 1. (i) $\mathrm{Prym}_\iota(C)$ *is a subabelian variety of* $\mathrm{Jac}(C)$. *Moreover* $\mathrm{Jac}(C)$ *decomposes into a sum*

$$\mathrm{Jac}(C)=\mathrm{Prym}_\iota(C)+\mathrm{Jac}(C_0)$$

with $\mathrm{Prym}_\iota(C)\cap\mathrm{Jac}(C_0)=$*finite number of points.*

(ii) *The Abel-Jacobi map*

$$I_\omega:\mathrm{Jac}(C)\to\mathbb{C}^g/L_\Gamma:D\rightsquigarrow\int_0^D\omega$$

maps $\mathrm{Prym}_\iota(C)$ *bi-analytically onto a complex subtorus* $\mathbb{C}^{g-g_0}/L_\Omega\hookrightarrow\mathbb{C}^g/L_\Gamma$ *with period matrix* Ω *given by*

$$\Omega=(2\psi(a)\quad\psi(A)\quad 2\psi(b)\quad\psi(B)).$$

Proof. (i) and (ii) will be proved together. First we compute the period matrix Γ of $\mathrm{Jac}(C)$:

$$\begin{aligned}\Gamma&=(\omega(a)\quad\omega(A)\quad\omega(a')\quad\omega(b)\quad\omega(B)\quad\omega(b'))\\&=\begin{pmatrix}\theta(a)&0&\Omega(a)&\theta(b)&0&\theta(b)\\ \psi(a)&\psi(A)&-\psi(a)&\psi(b)&\psi(B)&-\psi(b)\end{pmatrix}.\end{aligned}$$

Consider now the following matrices

$$\begin{aligned}\Gamma_1&=(\omega(a+a')\quad\omega(b+b')\quad\omega(a)\quad\omega(A)\quad\omega(b)\quad\omega(B))\\&=\begin{pmatrix}2\theta(a)&2\theta(b)&\theta(a)&0&\theta(b)&0\\ 0&0&\psi(a)&\psi(A)&\psi(b)&\psi(B)\end{pmatrix}\end{aligned}$$

and

$$\begin{aligned}\Gamma_2&=(\omega(a)\quad\omega(b)\quad\omega(a-a')\quad\omega(A)\quad\omega(b-b')\quad\omega(B))\\&=\begin{pmatrix}\theta(a)&\theta(b)&0&0&0&0\\ \psi(a)&\psi(b)&2\psi(a)&\psi(A)&2\psi(b)&\psi(B)\end{pmatrix}.\end{aligned}$$

Let us call

$$\Delta=(\theta(a)\quad\theta(b))\quad\text{and}\quad\Omega=(2\psi(a)\quad\psi(A)\quad2\psi(b)\quad\psi(B)).$$

Observe that Δ is the period matrix of $\mathrm{Jac}(C_0)$. Since $L_\Gamma=L_{\Gamma_1}=L_{\Gamma_2}$, the structure of the two matrices Γ_1 and Γ_2 makes then clear that the two tori $\mathbb{C}^{g_0}/2L_\Delta$ and

$\mathbb{C}^{g-g_0}/L_\Omega$ can be embedded into $\mathbb{C}^g/L_\Gamma$ as follows

$$\mathbb{C}^{g_0}/2L_\Delta \hookrightarrow \mathbb{C}^g/L_\Gamma : \begin{pmatrix} t_1 \\ \vdots \\ t_{g_0} \end{pmatrix} \bmod 2L_\Delta \curvearrowright \begin{pmatrix} t_1 \\ \vdots \\ t_{g_0} \\ 0 \\ \vdots \\ 0 \end{pmatrix} \bmod L_\Gamma$$

and

$$\mathbb{C}^{g-g_0}/L_\Omega \hookrightarrow \mathbb{C}^g/L_\Gamma : \begin{pmatrix} t_1 \\ \vdots \\ t_{g-g_0} \end{pmatrix} \bmod L_\Omega \curvearrowright \begin{pmatrix} 0 \\ \vdots \\ 0 \\ t_1 \\ \vdots \\ t_{g-g_0} \end{pmatrix} \bmod L_\Gamma .$$

Now, if we replace in Γ_1 the two "columns" $\omega(a)$ and $\omega(b)$ by $\omega(a) \curvearrowright 2\omega(a) - \omega(a+a')$ and $\omega(b) \curvearrowright 2\omega(b) - \omega(b+b')$, we obtain a matrix Γ_3

$$\Gamma_3 = \begin{pmatrix} 2\Delta & 0 \\ 0 & \Omega \end{pmatrix}$$

which shows that the covering map

$$\mathbb{C}^g/L_{\Gamma_3} = \mathbb{C}^{g_0}/2L_\Delta \oplus \mathbb{C}^{g-g_0}/L_\Omega \to \mathbb{C}^g/L_\Gamma : t \bmod L_{\Gamma_3} \curvearrowright t \bmod L_\Gamma$$

is of degree 2^{2g_0}.

On the other hand, from the structure of Γ_1 and Γ_2, one deduces easily that

$$D \in \mathrm{Prym}_\iota(C) \Leftrightarrow \begin{pmatrix} 2\int_0^D \theta \\ 0 \end{pmatrix} \in L_\Gamma \Leftrightarrow \int_0^D \theta \in L_\Delta \Leftrightarrow \int_0^D \begin{pmatrix} \theta \\ \psi \end{pmatrix} = \begin{pmatrix} 0 \\ t \end{pmatrix} \bmod L_\Gamma .$$

This shows that the Abel-Jacobi map I_ω maps $\mathrm{Prym}_\iota(C)$ bi-analytically onto $\mathbb{C}^{g-g_0}/L_\Omega$. By Chow's theorem, $\mathrm{Prym}_\iota(C)$ is thus a subabelian variety of $\mathrm{Jac}(C)$. Also, it is now clear that $\mathrm{Jac}(C_0)$ is embedded in $\mathrm{Jac}(C)$ as the set of divisors $D = \sum_{j=1}^{g} p_j \in \mathrm{Jac}(C)$, considered modulo linear equivalence, such that

$$\int_0^D \begin{pmatrix} \theta \\ \psi \end{pmatrix} = \begin{pmatrix} t \\ 0 \end{pmatrix} \bmod L_\Gamma$$

and $\mathrm{Prym}_\iota(C) \cap \mathrm{Jac}(C_0) = 2^{2g_0}$ points. This establishes the theorem.

A direct computation shows that the Riemann period relations for Γ induce period relations for Ω, namely

$$0 = \Gamma J^{-1} \Gamma^T = \begin{pmatrix} 2\Delta J^{-1} \Delta^T & 0 \\ 0 & \Omega Q^{-1} \Omega^T \end{pmatrix}$$

and

$$0<-\sqrt{-1}\Gamma J^{-1}\bar{\Gamma}^T=-\sqrt{-1}\begin{pmatrix}2\Delta J^{-1}\bar{\Delta}^T & 0\\ 0 & \Omega Q^{-1}\bar{\Omega}^T\end{pmatrix}$$

where J and Q are integral skew-symmetric matrices given by

$$J=\begin{pmatrix}0 & I\\ -I & 0\end{pmatrix} \quad \text{and} \quad Q=\begin{pmatrix}0 & A\\ -A & 0\end{pmatrix}$$

$$I=\text{identity}; \qquad A=\text{diag}(\delta_1,\ldots,\delta_{g-g_0})=\text{diag}(\overbrace{2,\ldots,2}^{g_0},\overbrace{1,\ldots,1}^{g-2g_0}).$$

Write $\Omega=(E,F)$ with E, F $(g-g_0, g-g_0)$ matrices and denote by λ_α $(1\leqq\alpha\leqq g-g_0)$ the $g-g_0$ columns of E. Then, in the new basis $e_\alpha=\lambda_\alpha/\delta_\alpha$ of $\mathbb{C}^{g-g_0}$, the period matrix Ω takes the canonical form

$$\Omega=(A,Z) \quad \text{with} \quad Z=AE^{-1}F \quad \text{symmetric}, \quad \text{Im}Z>0.$$

Definition 2 (see [7], p. 314). Let M be an abelian variety. Then $\text{Pic}^0(M)=\{\text{holomorphic line bundles on } M \text{ with first Chern class zero}\}$ is also an abelian variety called the *dual abelian variety* of M. The dual of M is denoted by $\hat{M}$.

Also in [7], p. 315, one finds the following information. If the period matrix Ω of M is in the canonical form $\Omega=(A,Z)$ with $A=\text{diag}(\delta_1,\ldots,\delta_\ell)(\delta_\ell|\delta_{\ell-1}|\ldots|\delta_1)$ and Z symmetric, $\text{Im}Z>0$, then the period matrix Ω^* of $\hat{M}$ is given by

$$\Omega^*=(\delta_1A^{-1},\delta_1A^{-1}ZA^{-1}). \tag{1}$$

Corollary 1. *With the notations of the previous theorem, the period matrix Ω^* of the dual abelian variety of* $\text{Prym}_t(C)$ *is* $\Omega^*=(E^*,F^*)=(\psi(a)\,\psi(A)\,\psi(b)\,\psi(B))$.

Proof. Because of (1) it suffices to show that

$$\delta_1A^{-1}AE^{-1}FA^{-1}=\delta_1A^{-1}(E^*)^{-1}F^*$$

which follows immediately from $E=E^*A$ and $F=F^*A$. This establishes the corollary.

Acknowledgements. I wish to thank Mark Adler and Pierre van Moerbeke for their constant encouragement during this work and the many discussions which have helped to clarify the overall picture. My thanks also go to J. Moser for stimulating questions.

References

1. Adler, M., van Moerbeke, P.: Completely integrable systems, Euclidean Lie algebras and curves. Adv. Math. **38**, 267–317 (1980)
2. Adler, M., van Moerbeke, P.: Linearization of Hamiltonian systems, Jacobi varieties and representation theory. Adv. Math. **38**, 318–379 (1980)
3. Adler, M., van Moerbeke, P.: Kowalewski's asymptotic method, Kac-Moody Lie algebras and regularization. Commun. Math. Phys. **83**, 83–106 (1982)
4. Adler, M., van Moerbeke, P.: The algebraic integrability of geodesic flow on SO(4). Invent. Math. **67**, 297–331 (1982)
5. Mumford, D.: Appendix to [4]

6. Dikii, L.A.: Hamiltonian systems connected with the rotation group. Funct. Anal. Appl. **6**, 83–84 (1972)
7. Griffiths, P., Harris, J.: Principles of algebraic geometry. New York: Wiley-Interscience 1978
8. Knörrer, H.: Geodesics on the ellipsoid. Invent. Math. **59**, 119–144 (1980)
9. Kowalewski, S.: Sur le problème de la rotation d'un corps solide autour d'un point fixe. Acta Math. **12**, 177–232 (1889)
10. Kowalewski, S.: Sur une propriété du système d'équations différentielles qui définit la rotation d'un corps solide autour d'un point fixe. Acta Math. **14**, 81–83 (1889)
11. Manakov, S.V.: Remarks on the integrals of the Euler equations of the n-dimensional heavy top. Funct. Anal. Appl. **10**, 93–94 (1976)
12. Mischenko, A.S., Fomenko, A.T.: Euler's equation on finite dimensional Lie groups. Math. USSR Izv. **12**, 371–389 (1978)
13. Moser, J.: Geometry of quadrics and spectral theory. In: The Chern Symposium 1979, pp. 147–188. Berlin, Heidelberg, New York: Springer 1980
14. Mumford, D.: Abelian varieties. Bombay, Oxford: Oxford University Press 1974
15. Ratiu, T.: The motion of the free n-dimensional rigid body, Indiana Univ. Math. J. **29**, 609–629 (1980)
16. Reid, M.: The complete intersection of two or more quadrics. Thesis, Cambridge Univ. 1972

Received January 26, 1983

Math. Ann. 263, 473–478 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Real Hypersurfaces of Complex Projective Spaces

Sadahiro Maeda

Department of Mathematics, Tokyo Metropolitan University, Tokyo 158, Japan

0. Introduction

Let $P^m(\mathbb{C})$ denote the complex projective space with Fubini-Study metric of constant holomorphic sectional curvature 4 and suppose $m \geqq 2$. It is well-known that there does not exist a totally umbilical real hypersurface M of $P^m(\mathbb{C})$ (see Tashiro and Tachibana [7]). Moreover, as a consequence of Codazzi equation, there does not exist a real hypersurface M of $P^m(\mathbb{C})$ with parallel second fundamental form. To get this, Y. Maeda [4] estimated the norm of the derivative of the second fundamental form. He showed

Theorem 1. *Let M be a compact real hypersurface of $P^m(\mathbb{C})$. Then $\|\nabla A\|^2 \geqq 4(m-1)$ and the equality holds if and only if M is congruent to $M^{\mathbb{C}}_{p,q}$ for some non-negative integers p, q.*

The model space $M^{\mathbb{C}}_{p,q}$ in the above theorem is established as follows: Lawson [3] considered a family of minimal hypersurfaces

$$M_{n,n'} := S^n\left(\sqrt{\frac{n}{2m}}\right) \times S^{n'}\left(\sqrt{\frac{n'}{2m}}\right)$$

in $S^{2m+1}(1)$,where $n+n'=2m$. But, in our paper, we do not necessarily need the condition that $M_{n,n'}$ are minimal in $S^{2m+1}(1)$ so that the radii of S^n and $S^{n'}$ are not unique. Choosing n and n' to be odd, namely $n=2p+1$, $n'=2q+1$, we get a fibration $S^1 \to M_{2p+1,2q+1} \to M^{\mathbb{C}}_{p,q}$ which is compatible with the Hopf fibration $S^1 \to S^{2m+1} \to P^m(\mathbb{C})$. $M^{\mathbb{C}}_{p,q}$ thus obtained has a characteristic property, which can be used to prove M to be congruent to $M^{\mathbb{C}}_{p,q}$. In general, a real hypersurface M of $P^m(\mathbb{C})$ has two structures, namely the contact structure induced from the complex structure J of $P^m(\mathbb{C})$ and the submanifold structure represented by the second fundamental form which is induced from the Fubini-Study metric of $P^m(\mathbb{C})$. Concerning the relation between these two structures on M, Okumura [5] proved the following theorem which is a characterization of $M^{\mathbb{C}}_{p,q}$.

Theorem 2. *Let M be a real hypersurface of $P^m(\mathbb{C})$ and $\bar{M}$ its inverse image with respect to the Hopf fibration $S^1 \to S^{2m+1} \to P^m(\mathbb{C})$. Then the second fundamental form of $\bar{M}$ is parallel if and only if the contact structure of M induced from $P^m(\mathbb{C})$ commutes with the second fundamental form of M.*

The purpose of this paper is to give another characterization of $M^{\mathbb{C}}_{p,q}$. We get

Theorem A. *Let M be a compact real hypersurface of $P^m(\mathbb{C})$. Then every geodesic γ of M, considered as a curve in $P^m(\mathbb{C})$, has constant first curvature if and only if M is congruent to $M^{\mathbb{C}}_{p,q}$ for some non-negative integers p, q.*

In Sect. 3, we consider the condition $RA=0$ which is weaker than $\nabla A=0$, where R is the curvature tensor of M.

The author thanks Professor K. Ogiue and Mr. K. Tsukada for their valuable suggestions and encouragement during the preparation of this paper. And also the author wishes to express his appreciation to Professor P. Dombrowski and the referee for useful comments.

1. Preliminaries

Let M be a real hypersurface of $P^m(\mathbb{C})$. In a neighborhood of each point, we choose a unit normal vector field N in $P^m(\mathbb{C})$. The Riemannian connections $\tilde{\nabla}$ in $P^m(\mathbb{C})$ and ∇ in M are related by the following formulas for arbitrary vector fields X and Y on M:

$$\tilde{\nabla}_X Y = \nabla_X Y + g(AX, Y)N, \tag{1.1}$$

$$\tilde{\nabla}_X N = -AX, \tag{1.2}$$

where g denotes the Riemannian metric on M induced from the Fubini-Study metric G of $P^m(\mathbb{C})$ and A is the (local) second fundamental form of M in $P^m(\mathbb{C})$. An eigenvector X of the second fundamental form A is called a *principal curvature vector*. Also an eigenvalue μ of A is called a *principal curvature*. In what follows, we denote by V_μ the eigenspace of A with eigenvalue μ.

It is known that M has an almost contact metric structure induced from the complex structure J of $P^m(\mathbb{C})$ (cf. [5]), i.e., we have a tensor field ϕ of type (1.1) on M, given by

$$g(\phi X, Y) = G(JX, Y) \quad \text{i.e.} \quad \phi(X) := JX - G(JX, N)N \tag{1.3}$$

for all tangent vectors X, Y of M, and – depending on the local choice of N – one has the unit tangent vector field ξ and the 1-form η of M defined by

$$\xi := -JN \quad \text{resp.} \quad \eta(X) := g(\xi, X) = G(JX, N). \tag{1.4}$$

This implies by use of (1.1):

$$(\nabla_X \phi)(Y) = \eta(Y)AX - g(AX, Y)\xi. \tag{1.5}$$

Let $\tilde{R}$ and R be the curvature tensor of $P^m(\mathbb{C})$ and M, respectively. Since the curvature tensor $\tilde{R}$ has a nice form, we have the following Gauss and Codazzi equations:

$$g(R(X, Y)Z, W) = g(Y, Z)g(X, W) - g(X, Z)g(Y, W) + g(\phi Y, Z)g(\phi X, W) - g(\phi X, Z)g(\phi Y, W) - 2g(\phi X, Y)g(\phi Z, W) + g(AY, Z)g(AX, W) - g(AX, Z)g(AY, W) \tag{1.6}$$

and

$$(\nabla_X A)Y - (\nabla_Y A)X = \eta(X)\phi Y - \eta(Y)\phi X - 2g(\phi X, Y)\xi . \tag{1.7}$$

In the following, we use the same terminology and notations as above unless otherwise stated.

For a submersion $\bar{M} \to M$ compatible with the Hopf fibration $S^1 \to S^{2m+1} \to P^m(\mathbb{C})$, it is well-known (cf. Ishihara and Konishi [2]) that if $\bar{M}$ has the parallel second fundamental form, M satisfies

$$(\nabla_X A)Y = -\eta(Y)\phi X - g(\phi X, Y)\xi \tag{1.8}$$

for all tangent vectors X, Y of M.

Y. Maeda [4] considered the converse problem, namely he determined the hypersurfaces M satisfying (1.8). He showed

Theorem 3. *The manifolds $M^{\mathbb{C}}_{p,q}$ (with non-negative integers) are the only compact hypersurfaces of $P^m(\mathbb{C})$ satisfying* (1.8).

If T denotes the tensor field given by

$$T(X, Y, Z) := g((\nabla_Z A)X, Y) + \eta(X)g(\phi Z, Y) + \eta(Y)g(\phi Z, X),$$

he calculated $\|T\|^2 = \|\nabla A\|^2 - 4(m-1)$, thereby deducing Theorem 1 from Theorem 3.

2. Proof of Theorem A

Now we prove Theorem A again by using Theorem 3. First of all we consider the condition of Theorem A:

$$(C)\begin{cases}\text{Every geodesic } \gamma : I \to M \text{ of } M \text{ has, considered as a curve}\\ \text{in } P^m(\mathbb{C}), \text{ constant first curvature.}\end{cases}$$

(C) means by definition, that $\tilde{\nabla}_{\dot\gamma}\dot\gamma$ has constant length in $P^m(\mathbb{C})$, i.e. $G(\tilde{\nabla}_{\dot\gamma}\dot\gamma, \tilde{\nabla}_{\dot\gamma}\dot\gamma)$ is constant on the interval I. But from (1.1) follows $G(\tilde{\nabla}_{\dot\gamma}\dot\gamma, \tilde{\nabla}_{\dot\gamma}\dot\gamma) = g(A\dot\gamma, \dot\gamma)^2$, hence (C) is equivalent to the constancy of $g(A\dot\gamma, \dot\gamma)$ on I, i.e. by differentiation (and using $\nabla_{\dot\gamma}\dot\gamma = 0$): $g((\nabla_{\dot\gamma}A)\dot\gamma, \dot\gamma) = 0$ on I. Therefore (C) is equivalent to

$$g((\nabla_X A)X, X) = 0 \text{ for every tangent vector } X \text{ of } M . \tag{2.1}$$

Let M be a compact real hypersurface satisfying (2.1). We may easily check that (2.1) is equivalent to

$$g((\nabla_X A)Y, Z) + g((\nabla_Y A)Z, X) + g((\nabla_Z A)X, Y) = 0 \tag{2.2}$$

for any X, Y and Z tangent to M. On the other hand, by virtue of Codazzi equation (1.7), we have

$$g((\nabla_Z A)X, Y) - g((\nabla_X A)Z, Y) = \eta(Z)g(\phi X, Y) - \eta(X)g(\phi Z, Y) - 2\eta(Y)g(\phi Z, X). \tag{2.3}$$

Exchanging X and Y, we get

$$g((\nabla_Z A)Y, X) - g((\nabla_Y A)Z, X) = \eta(Z)g(\phi Y, X) - \eta(Y)g(\phi Z, X) - 2\eta(X)g(\phi Z, Y). \tag{2.4}$$

Summing up (2.2), (2.3), and (2.4), we immediately obtain (1.8). Therefore, by using Theorem 3, M is congruent to $M^{\mathbb{C}}_{p,q}$ for some p, q. Since (2.1) is induced directly from (1.8), the converse is trivial. Q.E.D.

3. The Curvature Operator of M^{2m-1} in $P^m(\mathbb{C})$

We consider the action of the derivation $R(X, Y)$ on the algebra of tensor fields of M^{2m-1}. We recall that if T is a tensor field of type (r, s) and X and Y are vector fields, $R(X, Y) \cdot T = \nabla_X \nabla_Y T - \nabla_Y \nabla_X T - \nabla_{[X,Y]} T$. For brevity of notation, we denote by RT the tensor of type $(r, s+2)$ defined by

$$(RT)(X_1, X_2, \ldots, X_s, X, Y) = (R(X, Y) \cdot T)(X_1, X_2, \ldots, X_s).$$

In this section, in particular, we shall study real hypersurfaces of a complex projective space which satisfy $RA \equiv 0$. Here, for later use, we recall Cecil and Ryan's recent work [1]. The statement is as follows.

Theorem 4. *Let M be a connected real hypersurface in $P^m(\mathbb{C})$, $m \geqq 3$, with at most two distinct principal curvatures at each point. Then M is an open subset of a geodesic hypersphere.*

We are now in a position to prove the following.

Theorem B. *If $m \geqq 3$, there does not exist a real hypersurface in $P^m(\mathbb{C})$ satisfying $RA \equiv 0$.*

Proof. Assume that there exists a real hypersurface M satisfying $RA \equiv 0$ in $P^m(\mathbb{C})$. Let x be an arbitrary point of M and let $\{e_i\}_{i=1}^{2m-1}$ be an orthonormal basis for $T_x(M)$ such that $Ae_i = \lambda_i e_i$. For each λ, let $V_\lambda = \{X; AX = \lambda X\}$. Since A is symmetric, $V_\lambda \subseteq V_\mu^\perp$ whenever $\lambda \neq \mu$. Since $RA = 0$, we see that $R(X, Y)$ and A commute for all X and Y. In particular,

$$R(e_i, e_j)(Ae_j) = AR(e_i, e_j)e_j,$$

and hence

$$\lambda_j R(e_i, e_j)e_j = AR(e_i, e_j)e_j.$$

Thus, $R(e_i, e_j)e_j \in V_{\lambda_j}$ and hence $g(R(e_i, e_j)e_j, e_i) = 0$, whenever $\lambda_i \neq \lambda_j$. On the other hand, the Gauss equation (1.6) shows that

$$g(R(e_i, e_j)e_j, e_i) = 1 + 3\phi_{ij}\phi_{ij} + \lambda_i\lambda_j,$$

where $\phi_{ij}=g(\phi e_j, e_i)$. Therefore, if $\lambda_i \neq \lambda_j$, then necessarily $\lambda_i\lambda_j<0$, i.e. each two different principal curvatures must have opposite signs. This proves that M has at most two different principal curvatures at each of its points. Hence Theorem 4 shows that every connected component of M is an open subset of a geodesic hypersphere of $P^m(\mathbb{C})$. Nevertheless, an easy calculation yields that RA is not zero at an arbitrary point of any geodesic hypersphere (cf. [6]). This is a contradiction. So there does not exist a real hypersurface satisfying $RA \equiv 0$ in $P^m(\mathbb{C})$, $m \geqq 3$. Q.E.D.

We now remark that Theorem B does not say whether there exist real hypersurfaces which have a point x where $RA=0$. We shall give a negative partial answer to this problem. For this purpose we review Y. Maeda's lemma (for details, see p. 532 in [4]).

Lemma. *Let M be an orientable real hypersurface of $P^m(\mathbb{C})$ and assume that ξ is a principal curvature vector field on M with a C^∞ principal curvature function α. If X belongs to V_r and is orthogonal to ξ, then $(2r-\alpha)A\phi X=(\alpha r+2)\phi X$.*

Finally we prove the following.

Theorem C. *Let M be an orientable real hypersurface of $P^m(\mathbb{C})$ and assume that ξ is a principal curvature vector field on M for a C^∞ principal curvature function α on M. Then M has no point where $RA=0$.*

Proof. We assume that there exists a point x of M where $RA=0$. Due to the proof of Theorem B, we see that M must have at most two different principal curvatures at x. Moreover two different principal curvatures must have opposite signs. Now, in the following, we shall consider our problem at a fixed point x. Let $X \neq 0$ be a principal curvature vector with principal curvature r, which is orthogonal to ξ and therefore (since $\ker(\Phi)=\mathbb{R}\xi$) $\Phi X \neq 0$: Then, in view of Lemma, ϕX belongs to $V_{(\alpha r+2)/(2r-\alpha)}$ (we note that $2r-\alpha \neq 0$. In fact, if $2r-\alpha=0$, then $\alpha \neq r$ but $\alpha r>0$. This is a contradiction.).

So we must consider the following four cases:

Case 1; $\alpha=r=(\alpha r+2)/(2r-\alpha)$,

Case 2; $\alpha=r\neq(\alpha r+2)/(2r-\alpha)$,

Case 3; $(\alpha r+2)/(2r-\alpha)=\alpha\neq r$,

Case 4; $(\alpha r+2)/(2r-\alpha)=r\neq\alpha$.

The rest of proof is to show that the above four cases never happen.

Case 1; We immediately obtain a contradiction.

Case 2; We have $r\cdot(\alpha r+2)/(2r-\alpha)=r\cdot(r^2+2)/r=r^2+2>0$, which is a contradiction.

Case 3; The equation $(\alpha r+2)/(2r-\alpha)=\alpha$ yields $\alpha r=\alpha^2+2>0$, which is a contradiction.

Case 4; The equation $(\alpha r+2)/(2r-\alpha)=r$ yields $r\alpha=r^2-1$. On the other hand, we get (see, the proof of Theorem B)

$$\begin{aligned} r\alpha &= -1-3(g(\phi\xi,X))^2 \\ &= -1\,. \end{aligned}$$

Hence we have $r=0$. This is a contradiction. Q.E.D.

References

1. Cecil, E., Ryan, J.: Focal sets and real hypersurfaces in complex projective space. Trans. Am. Math. Soc. **269**, 481–499 (1982)
2. Ishihara, S., Konishi, M.: Differential geometry of fibred space. Publ. S.G.G. 8, 1973
3. Lawson, H.B.: Rigidity theorems in rank-1 symmetric spaces. J. Differential Geom. **4**, 349–357 (1970)
4. Maeda, Y.: On real hypersurfaces of a complex projective spaces, J. Math. Soc. Jpn. **28**, 529–540 (1976)
5. Okumura, M.: On some real hypersurfaces of a complex projective space. Trans. Am. Math. Soc. **212**, 355–364 (1975)
6. Takagi, R.: Real hypersurfaces in a complex projective space with constant principal curvatures I, II. J. Math. Soc. Jpn. **27**, 43–53, 507–516 (1975)
7. Tashiro, Y., Tachibana, S.: On Fubinian and C-Fubinian manifolds. Kōdai Math. Sem. Rep. **15**, 176–183 (1963)

Received November 2, 1981; in revised form February 2, 1983

Math. Ann. 263, 479–482 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Boundary Analyticity of Proper Holomorphic Maps of Domains with Non-Analytic Boundaries

David E. Barrett

Department of Mathematics, Princeton University, Princeton, NJ 08544, USA

Ligocka showed in [7] that boundary analyticity properties of the Bergman kernel function can be exploited to produce results on boundary analyticity of biholomorphic maps of bounded domains in $\mathbb{C}^n$. In the case of proper maps Bell showed [3] that a related boundary analyticity condition on the Bergman projection operator can be used to obtain corresponding results if the target domain has real analytic boundary. If $n \geqq 2$, though, there are many examples of domains for which the Bergman kernel function has good analytic boundary behavior even though the domain fails to have a real analytic boundary. The purpose of this paper is to show that boundary analyticity results can still be obtained for proper maps between such domains.

Let D be a bounded domain in $\mathbb{C}^n$. Let P_D denote the Bergman projection operator from $L^2(D)$ onto the subspace consisting of square-integrable holomorphic functions. The Bergman kernel function $K_D(z, w)$ for D satisfies

$$P_D f(z) = \int_D K_D(z, w) f(w) dv(w)$$

for $f \in L^2(D)$ and $z \in D$. D is said to satisfy condition R if P_D maps $C^\infty(\bar{D})$ continuously onto $A^\infty(\bar{D})$, the subspace of $C^\infty(\bar{D})$ consisting of functions holomorphic in D. The main result of this paper is the

Theorem. *Let $F: D_1 \to D_2$ be a proper holomorphic map of bounded domains in $\mathbb{C}^n$ with C^∞ boundaries which satisfy condition R. Suppose that the functions $K_{D_i}(\cdot, z)$ extend holomorphically to a neighborhood of $\bar{D}_i$ for all $z \in D_i$, $i = 1, 2$. Then F extends holomorphically to a neighborhood of $\bar{D}_1$.*

In [1] it is shown that the hypotheses of the theorem are satisfied by all smooth bounded domains which have so-called "transverse symmetries extending beyond the boundary." In particular the hypotheses are satisfied by all smooth bounded Reinhardt domains and by many Cartan domains.

The proof of the theorem makes use of several lemmas which were developed in order to study boundary smoothness of proper holomorphic maps.

Lemma 1. *Let D be a smooth bounded domain in $\mathbb{C}^n$ and let E be a nowhere dense subset of D. If D satisfies condition R then the complex linear span of the functions*

$$\{K(\cdot, z): z \in D - E\}$$

is a dense subspace of $A^\infty(\bar{D})$.

Lemma 2. *If $F: D_1 \to D_2$ is a proper holomorphic map of bounded domains in $\mathbb{C}^n$ then the Bergman kernel functions for D_1 and D_2 satisfy the transformation law*

$$\sum_{w \in F^{-1}(\zeta)} \frac{K_{D_1}(z, w)}{\det F'(w)} = \det F'(z) K_{D_2}(F(z), \zeta)$$

for $z \in D_1$ and $\zeta \in D_2 - V$, where $V \subset D_2$ is the variety $\{F(w): \det F'(w) = 0,\ w \in D_1\}$.

Lemma 3. *If $F: D_1 \to D_2$ is a proper holomorphic map of smooth bounded domains in $\mathbb{C}^n$ both of which satisfy condition R then $\det F' \cdot (g \circ F) \in A^\infty(\bar{D})$ for all $g \in A^\infty(\bar{D}_2)$.*

Lemma 1 is proved in [5]. (The proof there assumes that $E = \emptyset$ but only a trivial modification is required to obtain the version stated here. Actually, it is sufficient that $D - E$ be a set of uniqueness [4].) Lemma 2 is proved in [2]. Lemma 3 is a consequence of the first two lemmas and is proved in [4].

One more lemma is required in the proof of the theorem.

Lemma 4. *If D is a domain in $\mathbb{C}^n$ with smooth boundary and if $f \in A^\infty(\bar{D})$ is meromorphic in a neighborhood of $\bar{D}$ then f is holomorphic in a neighborhood of $\bar{D}$.*

Proof. Let $p \in bD$. By a suitable change of coordinates we may assume that p is the origin and that the curve $z' = 0$ is transverse to bD at p, where $z' = (z_1, \ldots, z_{n-1})$. By the Weierstrass Preparation Theorem we may write f in the form

$$q(z) + \frac{\sum_{j=0}^{m-1} b_j(z') z_n^{\,j}}{z_n^{\,m} + \sum_{j=0}^{m-1} a_j(z') z_n^{\,j}},$$

where q is holomorphic and the a_j's and b_j's are holomorphic functions of z' which vanish at $0' \in \mathbb{C}^{n-1}$.

It will suffice to show that each function b_j vanishes to infinite order at $0'$. Suppose this to be false. Pick α and k so that $\frac{\partial^\alpha b_k}{\partial z^\alpha}(0') \neq 0$ and $m|\alpha| + k$ is minimal. Then

$$\frac{\partial^\alpha f}{\partial z^\alpha}(0', z_n) = \frac{\partial^\alpha q}{\partial z^\alpha}(0', z_n) + \sum_{j=k}^{m-1} \frac{\partial^\alpha b_j}{\partial z^\alpha}(0') z_n^{\,j-m}$$

tends to infinity as $z_n \to 0$. This contradicts the smoothness of f, proving the lemma.

We turn now to the proof of the theorem. Let V be the variety defined in Lemma 2. By Lemma 1 there are functions $g_0, g_1, \ldots, g_n$ in the complex linear span of $\{K_{D_2}(\cdot, z): z \in D_2 - V\}$ which are C^1 – close to the functions $1, z_1, \ldots, z_n$ respectively. Then $G = \left(\frac{g_1}{g_0}, \ldots, \frac{g_n}{g_0}\right)$ is biholomorphic in a neighborhood of $\bar{D}_2$. Let $D_3 = G(D_2)$ and $H = G \circ F$. If H extends holomorphically to a neighborhood of $\bar{D}_1$, then so does $F = G^{-1} \circ H$.

Applying the transformation law from Lemma 2 to the functions

$$\frac{g_j \circ F}{g_0 \circ F} = \frac{\det F' \cdot g_j \circ F}{\det F' \cdot g_0 \circ F}$$

we conclude that each coordinate function h of H is a quotient of functions in the complex linear span of $\{K_{D_1}(\cdot, z): z \in D_1\}$; hence H is meromorphic in a neighborhood of $\bar{D}_1$. By Lemma 3 we see that $\det H' \cdot h^s$ is in $A^\infty(\bar{D}_1)$ for $s = 0, 1, \ldots$. Thus by Lemma 4 $\det H' \cdot h^s$ is holomorphic in a neighborhood of $\bar{D}_1$.

Now the ring $\mathcal{O}(\bar{D}_1)$ of functions holomorphic in a neighborhood of $\bar{D}_1$ is a unique factorization domain, so $\det H'$ is a product of irreducible factors. Let γ be any of these factors and suppose that γ^k divides $\det H'$. Then $\gamma^{k(s-1)}$ divides $(\det H')^{s-1}$. But $(\det H')^{s-1}$ divides $(\det H' \cdot h)^s$ so γ must appear at least $\frac{k(s-1)}{s}$ times in the factorization of $\det H' \cdot h$. Taking s large we find that γ^k must divide $\det H' \cdot h$. Repeating the argument for the other irreducible factors of $\det H'$ we conclude that $\det H'$ divides $\det H' \cdot h$; that is, h is holomorphic in a neighborhood of $\bar{D}_1$. (This factorization argument appears in Bell [3].) The theorem is proved.

Remarks. 1) It seems reasonable to ask if the theorem above still holds with reduced assumptions on the smoothness of the boundaries of the two domains. Indeed, examination of the proof shows that the smoothness of the boundary of D_1 is used only in the invocation of Lemma 4; a simple cone condition on the boundary of D_1 would suffice.

With regard to the smoothness of bD_2, however, the situation appears to be more troublesome. The difficulty comes in the construction of "Bell's operators" (denoted Φ^s in [5] and [1]) which play a crucial role in the proof of Lemma 1. These operators take a function $f \in C^\infty(\bar{D}_2)$ and produce a function $g \in C^\infty(\bar{D}_2)$ vanishing to high order on bD_2 such that $P_{D_2} f = P_{D_2} g$.

I know of no way to construct an adequate collection of these operators in general without assuming bD_2 to be of class C^∞. Such a construction can be carried out, however, in the case that bD_2 is C^1 and D_2 has "transverse symmetries extending beyond the boundary." As defined in [1], this means that D_2 has a Lie algebra $\mathfrak{g}$ of complete vector fields which are holomorphic in a neighborhood of $\bar{D}_2$ such that for each $p \in bD_2$ there is $X \in \mathfrak{g}$ such that $(JX)(r)(p) \neq 0$, where r is a C^1 defining function for D_2. (Compare with Tanaka's notion of "regularity" of a CR hypersurface in [8].) Following Proposition 4 of [8], we may assume after a holomorphic change of coordinates in a neighborhood of a point $p \in bD_2$ that r is of the form $y_n - R(z_1, \ldots, z_{n-1})$, where $y_n = \operatorname{Im} z_n$. If f is supported near p then we may let $\Phi^s f = ((-2\sqrt{-1}r)^s/s!)\left(\frac{\partial}{\partial z_n}\right)^{*s} f$, where $\left(\frac{\partial}{\partial z_n}\right)^*$ is the first order differential operator formally adjoint to $\frac{\partial}{\partial z_n}$ with respect to the original volume element dv on D_2. Integration by parts and induction reveal that $\int_{D_2} \Phi^s f \bar{h} dv = \int_{D_2} f \bar{h} dv$ for all holomorphic $h \in L^2(D_2)$, hence $P_{D_2} \Phi^s f = P_{D_2} f$. A partition of unity argument as in [5] or [1] completes the construction.

Combining this construction with the arguments in this paper and the results of [1], we have proved the

Theorem. *Let $F: D_1 \to D_2$ be a proper holomorphic map of bounded domains in $\mathbb{C}^n$. If D_1 and D_2 both have C^1 boundaries and have transverse symmetries extending beyond the respective boundaries then F extends holomorphically to a neighborhood of $\bar{D}_1$.*

2) Han [6] has discovered an extension principle for *CR* maps which is also based upon "transverse symmetry" considerations. It requires real analyticity of the relevant *CR* hypersurfaces as well as a pseudoconvexity assumption but has the advantage of being local in nature.

Most of the material in this paper is contained in my doctoral dissertation at the University of Chicago.

References

1. Barrett, D.: Regularity of the Bergman projection on domains with transverse symmetries. Math. Ann. **258**, 441–446 (1982)
2. Bell, S.: Proper holomorphic mappings and the Bergman projection. Duke Math. J. **48**, 167–175 (1981)
3. Bell, S.: Analytic hypoellipticity of the $\bar{\partial}$-Neumann problem and extendability of holomorphic mappings. Acta Math. **147**, 109–116 (1981)
4. Bell, S.: Boundary behavior of proper holomorphic mappings between non-pseudoconvex domains. Am. J. Math. (to appear)
5. Bell, S., Ligocka, E.: A simplification and extension of Fefferman's theorem on biholomorphic mappings. Invent. Math. **57**, 283–289 (1980)
6. Han, C.K.: Analyticity of *CR* equivalences between some real hypersurfaces in $\mathbb{C}^n$ with degenerate Levi forms. (To appear)
7. Ligocka, E.: Some remarks on extension of biholomorphic mappings. Lecture Notes in Mathematics, Vol. 798, pp. 350–363. Berlin, Heidelberg, New York: Springer 1980
8. Tanaka, N.: On the pseudo-conformal geometry of hypersurfaces of the space of n complex variables. J. Math. Soc. Jpn. **14**, 397–429 (1962)

Received February 10, 1983

Math. Ann. 263, 483–508 (1983)

Mathematische Annalen
© Springer-Verlag 1983

On Maass Wave Forms and the Imaginary Quadratic Doi-Naganuma Lifting

Solomon Friedberg

Department of Mathematics, Harvard University, Science Center 325, 1 Oxford Street, Cambridge, MA 02138, USA

Introduction

Let K be an imaginary quadratic field. There is a map, the Doi-Naganuma lifting, taking classical (holomorphic) modular forms with respect to a congruence subgroup of $SL(2,\mathbb{Z})$ to modular forms for K (cf. Asai [1] and Friedberg [3]). The aim of this paper is two-fold: first, to extend this lifting to include also non-holomorphic modular forms (Maass wave forms), and second, to study the notion of a modular form for K in some detail.

We begin in Sect. 1 by developing the properties of Maass wave forms on the complex upper half plane

$$\mathfrak{H} = \{x+iy \in \mathbb{C} | y > 0\};$$

in particular, we discuss certain (not necessarily holomorphic) forms 'coming from K' which play a key role in the lifting. In Sect. 2 we study modular forms associated to an imaginary quadratic field. These are vector valued functions of the *quaternionic* upper half space

$$\mathfrak{H}^1 = \{x+y\mathbf{k} | x \in \mathbb{C},\ 0 < y \in \mathbb{R}\}$$

which satisfy a transformation law and pair of differential equations, based upon the action of $SL(2,\mathbb{C})$ on $\mathfrak{H}^1$ by linear fractional transformation. We give the differential equations explicitly, and separate variables to produce the vector Fourier expansion.

The third section gives the definition of the Doi-Naganuma lifting, and some of its properties. We employ the theta function technique to produce an integral lifting kernel from an indefinite quadratic form of type (3,1) associated to K. By comparing the actions of the $\mathfrak{H}$ and $\mathfrak{H}^1$ Casimir operators on this kernel, we show that a Maass form on $\mathfrak{H}$ of weight $k \geqq 1$ and eigenvalue λ lifts to one of weight $2k-2$ and eigenvalue $(k^2-2k)/4+\lambda$ on $\mathfrak{H}^1$.

Finally, in Sect. 4 we compute the lifted Fourier expansion and Dirichlet series. From the explicit formula it is immediate that the lifting carries arithmetic forms to forms with arithmetic coefficients (compare Kudla [7]). Further, the lifting is

cuspidal if and only if the original form is orthogonal to a certain list of forms (those of Sect. 1 'coming from K') with respect to the Petersson inner product. In particular, the lift of a Maass wave form on $\mathfrak{H}$ of weight k and eigenvalue λ is *always* cuspidal *unless* λ belongs to the (finite) set

$$\{(\alpha^2-(k-1)^2)/4 | \alpha \in \mathbb{Z},\ 0 \leqq \alpha \leqq k-1\}.$$

Since the space of Maass wave forms on a fixed congruence subgroup of weight k and any fixed eigenvalue λ is finite dimensional, the lift is 'almost always' cuspidal. We conclude the paper with a brief discussion of extending the lifting to non-cusp forms, indicating how to compute the lifts of (non-holomorphic) Eisenstein series.

When $K/\mathbb{Q}$ is a cyclic Galois extension, a theorem of Langlands states that every automorphic representation π of $GL(2, \mathbf{A}_{\mathbb{Q}})$ ($\mathbf{A}$ = adeles) can be lifted to an automorphic representation $\tilde{\pi}$ of $GL(2, \mathbf{A}_K)$. Our results on the Doi-Naganuma lifting here describe a special case of this theorem, using a different method. In addition, we hope that our use of a more classical language, and our derivation of explicit formulas, will be of interest.

I would like to warmly thank Harold Stark for his advice and helpful suggestions, and the University of Chicago, the University of California, San Diego, the Massachusetts Institute of Technology, and Sigma-Xi, the Scientific Research Society, for financial support. Many of the results here have also been reported on in the author's University of Chicago doctoral dissertation ("Theta functions, liftings, and generalized Hilbert modular forms", June 1982).

1. Wave Forms on $\mathfrak{H}$

For a function $f(z)$ on $\mathfrak{H}$ and a matrix $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in $GL^+(2, \mathbb{R})$, we define the slash operator of weight k, an integer, by

$$(f|[\sigma]_k)(z) = (\det \sigma)^{k/2}(cz+d)^{-k} f(\sigma z)$$

where as usual $\sigma z = (az+b)(cz+d)^{-1}$. The differential operator corresponding to minus the Casimir operator on $SL(2, \mathbb{R})$, shifted by $(k^2-2k)/4+\lambda$, is

$$\Omega = \Omega^{k,\lambda} = -y^2\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right) + iky\left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right) + \lambda$$

(cf. Sect. 2). As is well known (see for example Roelcke [10]), Ω is invariant with respect to $|[\ ]_k$, i.e.,

$$(\Omega^{k,\lambda} f)|[\sigma]_k = \Omega^{k,\lambda}(f|[\sigma]_k)$$

for all σ in $GL^+(2, \mathbb{R})$ and twice differentiable $f: \mathfrak{H} \to \mathbb{C}$. The wave forms on $\mathfrak{H}$, whose properties we shall give in this section, are defined by

Definition 1.1. *The space of Maass wave forms of weight k on $\Gamma(N)$ and eigenvalue λ, $G\{N, k, \lambda\}$, is the space of all complex valued real analytic functions f on $\mathfrak{H}$ satisfying*

(1) $\Omega^{k,\lambda} f = 0$

(2) $f|[\sigma]_k = f$ *for all* σ *in* $\Gamma(N)$

(3) *For all* A *in* $SL(2,\mathbb{Z})$, $z = x+iy$ *in* $\mathfrak{H}$, $f|[A]_k(z) = O(y^\delta)$ *as* $y \to \infty$, *uniformly in* x, *for some fixed real* δ.

The subspace of Maass cusp wave forms, $S\{N,k,\lambda\}$, *is the subspace of* $G\{N,k,\lambda\}$ *of forms* f *satisfying*

(4) *For all* A *in* $SL(2,\mathbb{Z})$, $f|[A]_k(z) = O(e^{-\delta y})$ *as* $y \to \infty$, *uniformly in* x, *for some fixed* $\delta > 0$.

For a character $\chi \bmod N$, we also set

$$G\{N,k,\lambda,\chi\}_0 = \{f \in G\{N,k,\lambda\} \mid f|[\sigma]_k = \chi(d) f \quad \text{for all} \quad \sigma \in \Gamma_0(N)\},$$
$$G\{N,k\} = \bigcup_\lambda G\{N,k,\lambda\}, \quad \text{and} \quad G\{N,k,\chi\}_0 = \bigcup_\lambda G\{N,k,\lambda,\chi\}_0,$$

and define analogously $S\{N,k,\lambda,\chi\}_0$, $S\{N,k\}$, and $S\{N,k,\chi\}_0$. Notice that holomorphic forms on $\Gamma(N)$ are contained in $G\{N,k,0\}$.

Let $W_{m,n}(y)$ be Whittaker's function, a solution to the differential equation

$$\left\{4y^2 \frac{d^2}{dy^2} + 1 - 4n^2 + 4my - y^2\right\} W_{m,n}(y) = 0$$

which goes to zero exponentially as $y \to \infty$. Set, for $\varepsilon = \pm 1$, $y > 0$, k, λ as above,

$$\omega = [(k-1)^2 + 4\lambda]^{1/2}/2$$
$$W(\varepsilon y, k, \lambda) = y^{-k/2} W_{\varepsilon k/2, \omega}(2y),$$

and

$$u(y,k,\lambda) = \begin{cases} y^{[(1-k)/2-\omega]} & \omega \neq 0 \\ y^{(1-k)/2} \log y & \omega = 0. \end{cases}$$

Also, put $e(w) = \exp(\pi i w)$. The next two results, Propositions 1.2 and 1.3, are essentially due to Maass [8] and Roelcke [10], though with different normalizations of the slash operator than ours.

Proposition 1.2. *Let* f *be in* $G\{N,k,\lambda\}$. *Then* f *has a Fourier expansion at the cusp corresponding to* A *in* $SL(2,\mathbb{Z})$ *given by*

$$f|[A^{-1}]_k(z) = a_A(0) y^{[(1-k)/2+\omega]} + a'_A(0) u(y,k,\lambda) + \sum_{n \neq 0} a_A(n) W(2\pi n y/N, k, \lambda) e(2nx/N). \tag{1.1}$$

The Whittaker functions arise through condition (1) above, when one substitutes $\sum a_A(n,y) e(2nx/N)$ for f. For convenience, we drop the subscript A below when $A = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$.

Two differential operators play a key role in the study of Maass forms, allowing one to relate forms whose weights differ by an even integer. They are

$$\mathcal{K}_k = 2i \frac{\partial}{\partial z} + k y^{-1}, \qquad \Lambda = 2iy^2 \frac{\partial}{\partial \bar{z}}.$$

Proposition 1.3. (1) *For any continuously differentiable* $f:\mathfrak{H}\to\mathbb{C}$, *and* σ *in* $GL^+(2,\mathbb{R})$, $(\mathcal{K}_k f)|[\sigma]_{k+2}=\mathcal{K}_k(f|[\sigma]_k)$ *and* $(\Lambda f)|[\sigma]_{k-2}=\Lambda(f|[\sigma]_k)$.

(2) $\mathcal{K}_k\Omega^{k,\lambda}=\Omega^{k+2,\lambda-k}\mathcal{K}_k$, *and* $\Lambda\Omega^{k,\lambda}=\Omega^{k-2,\lambda+k-2}\Lambda$.

(3) $\mathcal{K}_k$ *gives a linear map of* $G\{N,k,\lambda\}$ *into* $G\{N,k+2,\lambda-k\}$ *which is bijective for* $\lambda\neq k$, *and takes cusp forms to cusp forms.*

Λ *gives a linear map of* $G\{N,k,\lambda\}$ *into* $G\{N,k-2,\lambda+k-2\}$ *which is bijective for* $\lambda\neq 0$, *and takes cusp forms to cusp forms.*

(4) $\Omega^{k,\lambda}=\Lambda\mathcal{K}_k+\lambda-k=\mathcal{K}_{k-2}\Lambda+\lambda$.

(5) *If* $S\{N,k,\lambda\}\neq 0$, *then* λ *is real, and* $\lambda\leqq\min(k,0)$.

(6) *Let* $f:\mathfrak{H}\to\mathbb{C}$ *be twice continuously differentiable. Then* $\mathcal{K}_k f=0$ *if and only if* $y^k f$ *is antiholomorphic, and in this case* $\Omega^{k,k}f=0$.

The proof of this Proposition may be found in [10]. For the spectral analysis of Ω and applications to the trace formula, see Hejhal [6] and Terras [13].

Given two functions f and g satisfying (2) and (3) of Definition 1.1, with at least one satisfying (4), we denote their Petersson inner product by

$$\langle f,g\rangle=\int_{\Gamma(N)\backslash\mathfrak{H}} f(z)\overline{g(z)}y^{k-2}dxdy.$$

Another property of the differential operators is given by

Proposition 1.4. *Suppose* f *and* g *are continuously differentiable functions on* $\mathfrak{H}$ *satisfying* $f|[\sigma]_{k+2}=f$ *and* $g|[\sigma]_k=g$ *for all* σ *in* $\Gamma(N)$. *Then* $\langle\Lambda f,g\rangle=\langle f,\mathcal{K}_k g\rangle$ *(when the inner product is defined).*

Proof. We apply Green's theorem to the fundamental domain for $\Gamma(N)\backslash\mathfrak{H}$, cancelling contributions from boundary pieces which are $\Gamma(N)$ equivalent (actually, there is some additional justification needed because of cusps, but it is well known). This gives

$$i\int_{\Gamma(N)\backslash\mathfrak{H}}(f_x\bar{g}+f\bar{g}_x)y^k dxdy=\int_{\Gamma(N)\backslash\mathfrak{H}}[f_y\bar{g}y^k+f(\bar{g}_y y^k+ky^{k-1}\bar{g})]dxdy$$

and the Proposition follows.

This Proposition is essentially a special case of a result of Shimura ([11], Sect. 2), though our proof is different. As an immediate corollary, we see that $\Omega^{k,0}$ is self-adjoint. In addition,

Corollary 1.5. *Let* f *be a holomorphic modular form of weight* k *on* $\Gamma(N)$, g *be in* $G\{N,k,\lambda\}$, $\lambda\neq 0$, *and suppose that one of* f *and* g *is a cusp form. Then* $\langle f,g\rangle=0$.

Proof. $\langle f,g\rangle=\langle f,-\mathcal{K}_{k-2}\Lambda g/\lambda\rangle=(-1/\bar{\lambda})\langle\Lambda f,\Lambda g\rangle=0$.

A similar argument, using the self-adjointness of $\Omega^{k,0}$, shows that in fact $S\{N,k,\lambda\}\perp G\{N,k,\mu\}$ for any $\lambda\neq\mu$. In particular, a twice continuously differentiable function f of weight k on $\Gamma(N)$ such that $y^k f$ is antiholomorphic is orthogonal to $S\{N,k,\lambda\}$ for any $\lambda\neq k$.

Next, we compute the action of $\mathcal{K}_k$ on certain functions, for later use. For a positive integer n, denote by $\mathcal{K}_k^n$ the n-th iterate of $\mathcal{K}_k$: $\mathcal{K}_k^n=\mathcal{K}_{k+2(n-1)}\cdots\mathcal{K}_{k+2}\mathcal{K}_k$. Let L and H denote the Laguerre and Hermite

polynomials, respectively, i.e.

$$L_n^{\alpha}(y) = e^y(n!y^{\alpha})^{-1}\frac{d^n}{dy^n}(e^{-y}y^{n+\alpha}),$$

$$H_n(y) = (-1)^n e^{y^2/2}\frac{d^n}{dy^n}(e^{-y^2/2}).$$

Lemma 1.6. (1) *Set*

$$t_{\gamma}(z,m,p) = y^{1/2}(8\pi y)^{-\gamma/2}H_{\gamma}((2\pi y)^{1/2}(m-p))e(-2mpx+iy(m^2+p^2)).$$

Then $\mathscr{K}_{\gamma}^n t_{\gamma} = (-4\pi)^n t_{\gamma+2n}$ *for* $0 \leqq \gamma$ *in* $\mathbb{Z}, m, p$ *in* $\mathbb{R}$.

(2) $\mathscr{K}_k^n e(2az) = n!\,y^{-n}L_n^{(k-1)}(4\pi ay)e(2az)$ *for* a *in* $\mathbb{C}$.

(3) $\mathscr{K}_0^n e(iv^2y^{-1}) = (\pi y^{-2}v^2)^n e(iv^2y^{-1})$ *for* v *in* $\mathbb{C}$.

(4) $\mathscr{K}_1^n(y^{-1}e(iv^2y^{-1})) = y^{-1}(\pi y^{-2}v^2)^n e(iv^2y^{-1})$ *for* v *in* $\mathbb{C}$.

Proof. Parts (1) and (2) may be found in Asai ([1]), Lemma 9; note δ_k^n there is $(-4\pi)^{-n}\mathscr{K}_k^n$. Equation (3) is proved by induction on n, and (4) follows from (3) since

$$\mathscr{K}_1^n(y^{-1}e(iv^2y^{-1})) = y^{-1}\mathscr{K}_0^n e(iv^2y^{-1}).$$

Fix now an imaginary quadratic field K of discriminant $-D$, with ring of integers $\mathcal{O}_K$, and an ideal I of $\mathcal{O}_K$. We shall describe below certain Maass wave forms "coming from K" which play a key role in the lift. Let $\mathfrak{E}_1 = \mathcal{O}_K, \ldots, \mathfrak{E}_{h(K)}$ be a set of wide ideal class representatives, and take $\mathfrak{e}_1 = 1, \ldots, \mathfrak{e}_{h(K)}$ ideal numbers for these ideals in the sense of Hecke [5] (14, Sect. 2). We set $J_i = I\mathfrak{E}_i\mathfrak{e}_i^{-1}$, and let J be one of the J_i. Write tr for the trace from $\mathbb{C}$ to $\mathbb{R}$, and $\mathbf{N} = \mathrm{Norm}(\mathbb{C}/\mathbb{R})$. For $\alpha, \beta \geqq 0$, z in $\mathfrak{H}$, and u, v in $\mathbb{C}$, we define

$$\phi_{\alpha,\beta}(z,u,v) = \sum_{r\in J} y^{-\beta}L_{\beta}^{(\alpha)}(4\pi y\mathbf{N}(r-v))(r-v)^{\alpha}e(2\mathbf{N}(r-v)z + 2\,\mathrm{tr}(r\bar{u}) - \mathrm{tr}(u\bar{v})).$$

Also, put $\chi_{-D}(n) = \left(\frac{-D}{n}\right)$, where $(-)$ is the Kronecker symbol.

Proposition 1.7. *For nonnegative integers* α *and* β

(1) $\mathscr{K}_{\alpha+2\beta+1}\phi_{\alpha,\beta}(z,u,v) = (\beta+1)\phi_{\alpha,\beta+1}(z,u,v)$.

(2) $\Omega^{\alpha+2\beta+1,\,-\beta(\alpha+\beta)}\phi_{\alpha,\beta}(z,u,v) = 0$.

(3) $\Lambda\phi_{\alpha,\beta}(z,u,v) = (\alpha+\beta)\phi_{\alpha,\beta-1}(z,u,v)$.

(4) *For* $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *in* $\Gamma_0(D\mathbf{N}(I))$,

$$\phi_{\alpha,\beta}(\sigma z, au+bv, cu+dv) = \chi_{-D}(d)(cz+d)^{\alpha+2\beta+1}\phi_{\alpha,\beta}(z,u,v).$$

Proof. From Lemma 1.6, (2), we see that $\beta!\,\phi_{\alpha,\beta} = \mathscr{K}_{\alpha+1}^{\beta}\phi_{\alpha,0}$, and part (1) follows. Then (1) implies (2), since $\phi_{\alpha,0}$ is holomorphic and hence has eigenvalue zero. Combining (1) and (2) with Proposition 1.3, (4), gives (3). Finally, (4) is well known when $\beta = 0$, and then the general case is a consequence of (1) and Proposition 1.3(1).

Notice that $\phi_{\alpha,\beta}(z,0,v)$ is in $G\{N,\alpha+2\beta+1,-\beta(\alpha+\beta)\}$ for v in $\bar{J}^{-1}\mathscr{D}_K^{-1}$, where $\mathscr{D}_K^{-1}$ is the inverse different. As an easy application of Corollary 1.5 (which will be useful in the sequel), we then have

Corollary 1.8. *Let v be in $\bar{J}^{-1}\mathscr{D}_K^{-1}$, $D\mathbf{N}(I)|N$, and g be in $S\{N,\alpha+2\beta+1,\lambda\}$, $\lambda \neq -\beta(\alpha+\beta)$. Then $\langle\phi_{\alpha,\beta}(z,0,v),g\rangle=0$.*

Proof. Write Λ^β for the β-th iterate of Λ. We have, by Propositions 1.4 and 1.7, $\langle\phi_{\alpha,\beta}(z,0,v),g\rangle=\beta!^{-1}\langle\phi_{\alpha,0}(z,0,v),\Lambda^\beta g\rangle$. Thus the Corollary follows from Corollary 1.5 provided that $\Lambda^\beta g$ is not in $S\{N,\alpha+1,0\}$. This happens precisely when $\lambda \neq -\beta(\alpha+\beta)$.

Finally, we close this section with a Proposition which allows us to 'exchange' the positive and negative Fourier coefficients $a(n)$ of a Maass wave form, given as in (1.1). Combining this with the results of Sect. 4, we shall be able to obtain a functional equation not only for

$$\sum_{n=1}^{\infty} a(n)n^{-s}\cdot\sum_{n=1}^{\infty} a(n)\chi_{-D}(n)n^{-s}$$

(a result analogous to those of Doi and Naganuma [2, 9]), but also for

$$\sum_{n=1}^{\infty} a(-n)n^{-s}\cdot\sum_{n=1}^{\infty} a(-n)\chi_{-D}(n)n^{-s}. \tag{1.2}$$

Proposition 1.9. *Say $f(z)$ in $G\{N,k,\lambda\}$ has the Fourier expansion* (1.1). *Then the function*

$$g(z)=b(0)y^{[(1-k)/2+\omega]}+b'(0)u(y,k,\lambda)+\sum_{n\neq 0} b(n)W(2\pi ny/N,k,\lambda)e(2nx/N),$$

where

$$b(n)=\begin{cases} a(-n) & n>0\\ a(-n)\prod_{c=0}^{k-1}(\lambda+c(k-c-1)) & n<0\\ (-1)^k\Big\{a(0)\prod_{c=0}^{k-1}([1-k]/2+\omega+c) & \\ \quad +a'(0)\delta_{\omega,0}\sum_{d=0}^{k-1}\prod_{\substack{c=0\\c\neq d}}^{k-1}([1-k]/2+c)\Big\} & n=0\end{cases}$$

and $b'(0)=a'(0)(-1)^k\prod_{c=0}^{k-1}([1-k]/2-\omega+c)$, is also in $G\{N,k,\lambda\}$. In fact,

$$g(z)=(-1)^k\mathscr{K}_{-k}^k(y^k f(-\bar{z})) \tag{1.3}$$

and if k is even, one also has

$$g(z)=(-1)^{k/2}\mathscr{K}_0^{(k/2)}[(\Lambda^{(k/2)}f)(-\bar{z})]. \tag{1.4}$$

Proof. Checking the conditions of Definition 1.1, one sees immediately that $y^k f(-\bar{z})$ is in $G\{N,-k,\lambda-k\}$. A computation similar to Maass [8] shows that

$$\mathscr{K}_k[W(2\pi ny/N,k,\lambda)e(2nx/N)] = \begin{cases}(-2\pi n/N)W(2\pi ny/N,k+2,\lambda-k)e(2nx/N) & n>0\\ (2\pi n/N)(\lambda-k)W(2\pi ny/N,k+2,\lambda-k)e(2nx/N) & n<0\end{cases}$$

and also

$$\Lambda[W(2\pi ny/N,k,\lambda)e(2nx/N)] = \begin{cases}(\lambda N/2\pi n)W(2\pi ny/N,k-2,\lambda+k-2)e(2nx/N) & n>0\\ (-N/2\pi n)W(2\pi ny/N,k-2,\lambda+k-2)e(2nx/N) & n<0.\end{cases}$$

These allow one to calculate the right hand sides of (1.3) and (1.4) directly, proving the Proposition.

2. Wave Forms on $\mathfrak{H}^1$

In this section we shall define and study waveforms on the quaternionic upper half space, $\mathfrak{H}^1$. Recall $\sigma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in SL(2,\mathbb{C})=\mathbf{G}$ acts on z in $\mathfrak{H}^1$ by the usual action $\sigma z=(az+b)(cz+d)^{-1}$; the stabilizer of $\mathbf{k}$ is just $SU(2,\mathbb{C})=\mathbf{K}$. The notion of wave form on $\mathfrak{H}^1$ is more complicated than it is on $\mathfrak{H}$ since $\mathbf{K}$ is not abelian – it is necessary to take representations of $\mathbf{K}$ which are not one-dimensional into account.

Namely, let $(a\ b)_n={}^t(a^n, a^{n-1}b, \ldots, ab^{n-1}, b^n)$, where ^{t} denotes the transpose, and define the n-fold symmetric representation $\varrho_n(g)$, for g in $\mathbf{G}$, by $((a\ b)^t g)_n=\varrho_n(g)(a\ b)_n$. The ϱ_n give irreducible $(n+1)$-dimensional representations of $\mathbf{K}$. We also write E_n for the $n\times n$ identity.

As is well known, the Lie algebra of $\mathbf{G}$ is $sl(2,\mathbb{C})$, the set of all 2×2 matrices of trace zero. Corresponding to $S\in sl(2,\mathbb{C})$ are differential operators L_S, L'_S; for $f\in C^\infty(\mathbf{G})$,

$$L_S f(g)=\left[\frac{\partial}{\partial t} f(g\exp(tS))\right]_{t=0}, \qquad L'_S f(g)=\left[\frac{\partial}{\partial \bar{t}} f(g\exp(tS))\right]_{t=0}.$$

The matrices $A=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}$, $B=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}$, and $U=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}$ are a $\mathbb{C}$-basis for $sl(2,\mathbb{C})$, and using them one can form the Casimir operators

$$\Psi=\tfrac{1}{2}(L_A L_B+L_B L_A+\tfrac{1}{2}L_U L_U), \qquad \tilde{\Psi}=\tfrac{1}{2}(L'_A L'_B+L'_B L'_A+\tfrac{1}{2}L'_U L'_U),$$

which generate the center of the universal enveloping algebra. Translating Ψ and $\tilde{\Psi}$ to the symmetric space level will give the analogue of Ω. We give it first on the

group **G** itself. Take Iwasawa coordinates for **G**, writing $g = y^{-1/2}\begin{pmatrix} y & x \\ 0 & 1 \end{pmatrix}\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}$ with a, b, x in $\mathbb{C}$, y in $\mathbb{R}^+$, $|a|^2+|b|^2=1$. It suffices to use local coordinates x, $\bar{x}$, y, a, $\bar{a}$, and b (resp. $\bar{b}$), giving Ψ (resp. $\tilde{\Psi}$) on the open subset of **G** where these are valid, by the homogeneity of **G**. For convenience write ∂_x for $\frac{\partial}{\partial x}$, etc. A simple computation gives

Proposition 2.1.

$$\Psi = y^2(\partial^2_{x\bar{x}} + \partial^2_{yy}/4) + yb\partial^2_{x\bar{a}} + y\partial_y(a\partial_a - \bar{a}\partial_{\bar{a}} + b\partial_b - 1)/4$$
$$+ a(a\partial_a - 2\bar{a}\partial_{\bar{a}} + 2b\partial_b - 3)\partial_a/16 + \bar{a}(\bar{a}\partial_{\bar{a}} - 2b\partial_b + 5)\partial_{\bar{a}}/16 + b(b\partial_b - 3)\partial_b/16,$$

and $\tilde{\Psi} = \bar{\Psi}$.

To make the correspondence of Casimir operators between group and symmetric space precise, it is first necessary to relate functions on **G** and $\mathfrak{H}^1$. Let ϱ be the additive and multiplicative representation of the quaternions given by $\varrho(x+y\mathbf{k}) = \begin{pmatrix} x & -y \\ \bar{y} & \bar{x} \end{pmatrix}$ for x, y in $\mathbb{C}$; also, for $g = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ in **G**, set $J(g, z) = \varrho(cz+d)$. Since we shall no longer need a notation for the differential operator Λ of Sect. 1, let Λ now denote a subgroup of **G** [usually $SL(2, \mathcal{O}_K)$ or one of its subgroups]. Take an integer $f \geqq 0$, and let $\varrho^1_f(g) = {}^t\varrho_f({}^t g)$ for g in **G**; each ϱ^1_f also gives an irreducible $(f+1)$-dimensional representation of **K**.

Proposition 2.2. *There is a one-to-one correspondence between*

$$\{F : \mathbf{G} \to \mathbb{C}^{f+1} | F(\gamma g \kappa) = {}^t\varrho_f(\kappa) F(g) \text{ for all } \gamma \in \Lambda, \kappa \in \mathbf{K}\}$$

and

$$\{F^1 : \mathfrak{H}^1 \to \mathbb{C}^{f+1} | F^1(\gamma z) = \varrho^1_f(J(\gamma, z)) F^1(z) \text{ for all } \gamma \in \Lambda\}$$

given by $F^1(z) = \varrho^1_f(J(g, \mathbf{k}))F(g)$ *where* $z = g\mathbf{k}$, *and* $F(g) = \varrho^1_f(J(g, \mathbf{k}))^{-1} F^1(g\mathbf{k})$.

The proof is a straightforward use of the cocycle equation

$$\varrho^1_f(J(g_1 g_2, z)) = \varrho^1_f(J(g_1, g_2 z))\varrho^1_f(J(g_2, z))$$

combined with the observation that $J(\kappa, \mathbf{k})^t\kappa = E_2$ for all κ in **K**.

Let now $F^1 : \mathfrak{H}^1 \to \mathbb{C}^{f+1}$ be as in Proposition 2.2, with corresponding F in $C^\infty(\mathbf{G})$. Then the Casimir operators of weight f on $\mathfrak{H}^1$ are given by

$$(\Psi^f F^1)(z) = \varrho^1_f(J(g, \mathbf{k}))(\Psi F)(g), \quad (\tilde{\Psi}^f F^1)(z) = \varrho^1_f(J(g, \mathbf{k}))(\tilde{\Psi} F)(g) \quad \text{where} \quad g\mathbf{k} = z.$$

Define a slash operator by $F^1 \| [g]_f(z) = \varrho^1_f(J(g, z))^{-1} F^1(gz)$ for g in **G**. As above, write $z = x + y\mathbf{k}$.

Proposition 2.3.

(1) $(\Psi^f F^1) \| [g]_f = \Psi^f (F^1 \| [g]_f)$; $\quad (\tilde{\Psi}^f F^1) \| [g]_f = \tilde{\Psi}^f (F^1 \| [g]_f)$.

$$\Psi^f = y^2(\partial^2_{x\bar{x}} + \partial^2_{yy}/4)E_{f+1} + y\partial_x \begin{pmatrix} 0 & 1 & & \\ & \ddots & \ddots & \\ & & 2 & \\ & & & f \\ & & & 0 \end{pmatrix}$$

$$+ y\partial_y \cdot \tfrac{1}{4} \begin{pmatrix} 2f-1 & & & \\ & 2f-3 & & \\ & & \ddots & \\ & & & -1 \end{pmatrix}$$

$$+ \tfrac{1}{4} \begin{pmatrix} f(f-2) & & & \\ & (f-1)(f-3) & & \\ & & \ddots & \\ & & & 0 \end{pmatrix}. \tag{2.1}$$

$$\tilde{\Psi}^f = y^2(\partial^2_{x\bar{x}} + \partial^2_{yy}/4)E_{f+1} - y\partial_{\bar{x}} \begin{pmatrix} 0 & & & \\ f & & & \\ & f-1 & \ddots & \\ & & 1 & 0 \end{pmatrix}$$

$$+ y\partial_y \cdot \tfrac{1}{4} \begin{pmatrix} -1 & \ddots & & \\ & 2f-3 & & \\ & & & 2f-1 \end{pmatrix}$$

$$+ \tfrac{1}{4} \begin{pmatrix} 0 & \ddots & & \\ & (f-1)(f-3) & & \\ & & & f(f-2) \end{pmatrix}.$$

Proof. The proof of (1) follows by combining the left invariance of Ψ and $\tilde{\Psi}$ with the definitions of Ψ^f, $\tilde{\Psi}^f$ and the cocycle equation. To get (2) and (3), it is necessary to combine Proposition 2.1 with a computation, using the definition of Ψ^f. We do the $y\partial_x$ term of (2) to indicate the method. Let

$$M = M(a, b, \bar{a}) = \varrho^1_f\left(J\left(\begin{pmatrix} a & b \\ -\bar{b} & \bar{a} \end{pmatrix}, \mathbf{k}\right)\right)^{-1}.$$

We need to find $M^{-1} b\partial_{\bar{a}} M$. But $(au + bv \;\; -\bar{b}u + \bar{a}v)_f = {}^t M(u\; v)_f$. Differentiating with respect to $\bar{a}$ and multiplying by b, it is easy to get

$$\begin{aligned} &{}^t(0 \;\; (au+bv)^f \, 2(au+bv)^{f-1}(-\bar{b}u + \bar{a}v) \ldots f(au+bv)(-\bar{b}u+\bar{a}v)^{f-1}) \\ &= {}^t(M^{-1} b\partial_{\bar{a}} M)(au + bv \;\; -\bar{b}u + \bar{a}v)_f, \end{aligned}$$

and the result follows (compare Weil ([14])).

We extend the slash operator to A in $GL(2, \mathbb{C})$, by putting

$$F\|[A]_f = F\|[(\det A)^{-1/2} A]_f$$

(the ambiguity of sign caused by the choice of square root will not cause any harm in the applications). Let Λ be $SL(2, \mathcal{O}_K)$ or one of its congruence subgroups. The notion of a modular form on Λ is made precise by the following definition.

Definition 2.4. *A wave form, or generalized Hilbert modular form, for K of weight f on Λ and eigenvalues (λ, λ') is a real analytic function $F^1: \mathfrak{H}^1 \to \mathbb{C}^{f+1}$ satisfying*

(1) $\Psi^f F^1 = \lambda F^1$. (1′) $\tilde{\Psi}^f F^1 = \lambda' F^1$.

(2) $F^1 \| [\sigma]_f = F^1$ *for all* σ *in* Λ.

(3) *For all* A *in* $GL(2, \mathcal{O}_K)$, $z = x + y\mathbf{k} \in \mathfrak{H}^1$, $F^1 \| [A]_f(z) = O(y^\delta)$ *as* $y \to \infty$, *uniformly in* x, *for some fixed real* δ.

If for all A *in* $GL(2, \mathcal{O}_K)$, $F^1 \| [A]_f(z) = O(e^{-\delta y})$ *as* $y \to \infty$, *uniformly in* x, *for some fixed* $\delta > 0$, F^1 *is said to be cuspidal or a cusp form. If* $\lambda = \lambda'$, *we shall say simply that* F^1 *has eigenvalue* λ.

In a similar way, we define a modular form on a subgroup Λ of

$$SL(2, \mathcal{O}_K)_i = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \mathbf{G} \mid \alpha, \delta \in \mathcal{O}_K, \beta \in \bar{\mathfrak{e}}_i \mathcal{O}_K, \gamma \in \mathfrak{e}_i \mathcal{O}_K \right\}$$

by replacing $GL(2, \mathcal{O}_K)$ in the definition above by

$$GL(2, \mathcal{O}_K)_i = \left\{ \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in GL(2, \mathbb{C}) \mid \alpha, \delta \in \mathcal{O}_K, \beta \in \bar{\mathfrak{e}}_i \mathcal{O}_K, \gamma \in \mathfrak{e}_i \mathcal{O}_K \right\}.$$

In all cases, one can of course deal with the corresponding function on $\mathbf{G}$, by Proposition 2.2. We shall do so when convenient without further comment in Sects. 3 and 4 below.

Example. Let $\mu = (1 + 4\lambda)^{1/2}$, $\mu' = (1 + 4\lambda')^{1/2}$. Choose λ, λ' such that $(f + \mu + \mu')/2$ is an integer between 0 and f, and set $F_j^1(z) = y^{1-j+\mu}$ if $j = (f + \mu + \mu')/2$, and 0 otherwise. Then one sees easily that

$$\sum_{\Lambda_\infty \backslash \Lambda \ni \gamma} F_j^1 \| [\gamma]_f(z)$$

is a modular form of weight f and eigenvalues (λ, λ') (whenever the sum converges).

In order to exploit the symmetry between Ψ^f and $\tilde{\Psi}^f$, we define an involution ι_f on the space of modular forms of weight f by

$$\iota_f F^1(z) = \varrho_f \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} F^1(-\bar{z}).$$

We have

Proposition 2.5.

(1) *If* $\Psi^f F^1 = \lambda F^1$, *then* $\tilde{\Psi}^f(\iota_f F^1) = \lambda \iota_f F^1$.
If $\tilde{\Psi}^f F^1 = \lambda' F^1$, *then* $\Psi^f(\iota_f F^1) = \lambda' \iota_f F^1$.

(2) *If* $F^1 \| [\sigma]_f = F^1$, *then* $(\iota_f F^1) \| \left[\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \bar{\sigma} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} \right]_f = \iota_f F^1$.

(3) ι_f *takes modular forms for* K *of weight* f *and eigenvalues* (λ, λ') *bijectively to modular forms for* K *of weight* f *and eigenvalues* (λ', λ).

Proof. Equation (1) is clear from the explicit formulas of Proposition 2.3. To see (2), it suffices to verify that

$$\varrho_f \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \varrho_f^1 \left(J\left(\begin{pmatrix} \bar{a} & -\bar{b} \\ -\bar{c} & \bar{d} \end{pmatrix}, -\bar{z} \right) \right) \varrho_f \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \varrho_f^1 \left(J\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}, z \right) \right),$$

and this follows from a direct computation. Then (3) is a consequence of (1) and (2) and Definition 2.4.

In the remaining part of this section, we discuss the Fourier expansion of a generalized Hilbert modular form F of weight f. Without loss we shall confine ourselves to the case $\Lambda = SL(2, \mathcal{O}_K)$. Let ξ^j be the Grössencharacter $\xi^j(r) = (r/|r|)^j$, and $K_\nu^{(\mu)}$ denote the μ^{th}-derivative of the ν^{th} K-Bessel function.

Proposition 2.6. *The Fourier expansion of a function F^1 satisfying* (1), (2), *and* (3) *of Definition 2.4 is given by*

$$F^1(z) = \sum_r Q(r, y) e(2 \operatorname{tr}(rx)) \tag{2.2}$$

where the sum is over r in the inverse different $\mathcal{D}_K^{-1}$, and

$$(Q(r, y)) = {}^t(Q_f(r, y), \ldots, Q_0(r, y))$$

is the following column vector: for $r \neq 0$,

$$Q_j(r, y) = \xi^j(r) \sum_{n=0}^{j} C_{n,j}(r) (4\pi|r|y)^{1-n} K^{(j-n)}_{(1+4\lambda)^{1/2}}(4\pi|r|y)$$

where $C_{j,j}(r)$ is an arbitrary complex number for each $0 \leqq j \leqq f$, and

$$\begin{pmatrix} C_{0,j}(r) \\ \vdots \\ C_{j-1,j}(r) \end{pmatrix} = -2i(f+1-j) \begin{pmatrix} 2(j) & & & \\ (j)(j-1) & 2(j-1) & & \\ & (j-1)(j-2) & \ddots & \\ & & (2)(1) & 2(1) \end{pmatrix}^{-1}$$

$$\cdot \begin{pmatrix} C_{0,j-1}(r) \\ \vdots \\ C_{j-1,j-1}(r) \end{pmatrix}. \tag{2.3}$$

Also, for $r = 0$,

$$Q_j(0, y) = C_{1,j}(0) y^{1-j+(4\lambda+1)^{1/2}} + C_{2,j}(0) q(y, j, \lambda)$$

where $C_{i,j}(0)$ are arbitrary complex numbers, $i = 1, 2$, and

$$q(y, j, \lambda) = \begin{cases} y^{1-j-(4\lambda+1)^{1/2}} & \lambda \neq -1/4 \\ y^{1-j} \log y & \lambda = -1/4. \end{cases}$$

The functions $y^{1-n} K^{(j-n)}_{(1+4\lambda)^{1/2}}(y)$ are $\mathbb{C}$-linearly independent for fixed j and $0 \leqq n \leqq j$.

Remark. Alternate appearances of this Fourier expansion can be obtained by use of identities which involve the derivatives of K-Bessel functions (cf. [4]). One such alternate is given in the theorems of Sect. 4.

Proof. First, say $r \neq 0$. For convenience put $Q_j(r, y) = \xi^j(r) q_j(r, y_1)$ where $y_1 = 4\pi|r|y$. Substituting (2.2) into (2.1) and equating coefficients, we see that $v = q_j(r, y_1)$ satisfies

$$y_1^2 v'' + (2j-1) y_1 v' + (j(j-2) - 4\lambda - y_1^2) v + 2i(f+1-j) y_1 q_{j-1}(r, y_1) = 0 \tag{2.4}$$

(with $q_{-1}=0$). It is straightforward to verify that the q_j corresponding to the Q_j given in the Proposition does satisfy this equation, by a direct computation, as $u=y^{1-n}K^{(j-n)}_{(1+4\lambda)^{1/2}}(y)$ is a solution to

$$y^2u''+(2j-1)yu'+(j(j-2)-4\lambda-y^2)u-2(j-n)y^{2-n}K^{(j-n-1)}_{(1+4\lambda)^{1/2}}$$
$$-(j-n)(j-n-1)y^{1-n}K^{(j-n-2)}_{(1+4\lambda)^{1/2}}=0. \tag{2.5}$$

Since the other solutions to (2.4) are combinations of I-Bessel functions, which grow exponentially at infinity, the only possible solutions which may appear in the Fourier expansion of F are the Q_j's.

Similarly, in the case $r=0$, $v=Q_j(0,y)$ must satisfy

$$y^2v''+y(2j-1)v'+(j(j-2)-4\lambda)v=0.$$

The solutions are those given in the Proposition, as is well known.

Finally, to get the linear independence of the $y^{1-n}K^{(j-n)}_{(1+4\lambda)^{1/2}}(y)$, we induct on j. Let $\mathbf{D}$ denote the differential operator

$$\mathbf{D}=y^2\frac{d^2}{dy^2}+(2j-1)y\frac{d}{dy}+j(j-2)-4\lambda-y^2.$$

Suppose that $\sum_{n=0}^{j}\alpha(n)y^{1-n}K^{(j-n)}_{(1+4\lambda)^{1/2}}(y)=0$ and not all $\alpha(n)$ are zero. Then at least two are non-zero. Applying $\mathbf{D}$ and using (2.5), we conclude that

$$\sum_{n=0}^{j-1}\beta(n)y^{1-n}K^{((j-1)-n)}_{(1+4\lambda)^{1/2}}(y)=0$$

where

$$\beta(n)=2\alpha(n)(j-n)+\alpha(n-1)(j-n+1)(j-n)$$

with $\alpha(-1)=0$. Since the square matrix in (2.3) is nonsingular, the $\beta(n)$ are not all zero, and this contradicts the inductive hypothesis.

It is of course not difficult to compute the inverse matrix appearing in (2.3); call it $(b_{mn}(j))$, $1\leqq m,\ n\leqq j$. It is:

$$b_{mn}(j)=\begin{cases}0 & n>m\\ (2(j-m+1))^{-1} & n=m\\ -1/4 & m=n+1\\ (-1)^{k-1}(j-n)\dots(j-n-k+1)/2^{k+2} & m-n=k+1,\ k>0.\end{cases}$$

Observe that by using Proposition 2.5 along with this result, we also obtain a Fourier expansion for any function F^1 satisfying (1′) of Definition 2.4. For a modular form of eigenvalues (λ,λ'), this enables us to determine all the $C_{j,j}(r)$ from $C_{0,0}(r)$. For example, since $\iota_f F^1$ is also an eigenfunction of Ψ^f, we get an additional equation relating $C_{0,0}(r)$ and $C_{1,1}(r)$ from (2.4) when $j=f$, and we find that

$$C_{1,1}(r)=-iC_{0,0}(r)(f^2+4(\lambda-\lambda'))/2.$$

Then (2.4) with $j=f-1$ allows us to determine $C_{2,2}(r)$, and so on. Comparing also the constant terms, we have

Theorem 2.7. *Let F^1 be a modular form of weight f and eigenvalues (λ, λ').*

(1) *In the Fourier expansion* (2.2) *of F^1,*

$$Q_j(r,y)=\xi^j(r)C(r)Q_j(4\pi|r|y) \quad f\geqq j\geqq 0 \tag{2.6}$$

where the functions Q_j are completely determined from the results above. In particular, only one Dirichlet series

$$\sum C(r)\mathbf{N}(r)^{-s}$$

and its twists by the Grössencharacters $\xi^j(r)$ arises from F^1 via Mellin transform.

(2) *F^1 is a cusp form unless*

$$f-2j=\pm(1+4\lambda)^{1/2}\pm(1+4\lambda')^{1/2} \tag{2.7}$$

for some integer j, $0\leqq j\leqq f$, and (independent) choices of signs. In particular, there are only finitely many λ for which there exist non-cusp forms of eigenvalue λ.

Remark. The lifting studied in Sects. 3 and 4 below *can* fail to be cuspidal whenever a j satisfying (2.7) does exist.

3. The Doi-Naganuma Lifting

In order to define the lifting, we shall first introduce a theta function which is based upon an indefinite quadratic form of type (3, 1). Namely, let $\mathscr{V}$ denote the real 4-dimensional vector space

$$\mathscr{V}=\{X\in M(2,\mathbb{C})|{}^tX=\bar{X}\},$$

with $\mathbf{G}$-action given by

$$X^g={}^t\bar{g}Xg,$$

and Q be the quadratic form defined by $Q[X]=-2\det X$. The majorants of Q are given by the (positive definite) quadratic forms $R^g[X]=\mathrm{tr}((X^g)^2)$ for g in $\mathbf{G}$; notice $R^g=R^{g\kappa}$ for all κ in $\mathbf{K}$. The associated spherical harmonic polynomials have been studied by Asai [1]. Let $X=\begin{pmatrix} m & r \\ \bar{r} & p \end{pmatrix}$ be in $\mathscr{V}$. Then a $\mathbb{C}$-basis for the polynomials of degree $k-1$ is the set of

$$\eta_{k,\alpha}(X)=(k-1)!\sum_{\beta,\gamma}(\alpha+\beta)!^{-1}\gamma!^{-1}2^{-\beta}r^{\alpha}L_{\beta}^{(\alpha)}(2r\bar{r})H_{\gamma}(m-p), \tag{3.1}$$

where α runs over integers with $|\alpha|\leqq k-1$, L and H are Laguerre and Hermite polynomials, respectively, and the sum is over nonnegative integers β and γ such that $2\beta+\gamma=k-\alpha-1$ and $\alpha+\beta\geqq 0$.

Given nonzero integers M and P, and an ideal I, we also introduce the lattices $\mathscr{L}_i$, $1\leqq i\leqq h(K)$:

$$\mathscr{L}_i=\left\{\begin{pmatrix} m & r \\ \bar{r} & p \end{pmatrix}\in\mathscr{V}\,|\,m\in M\mathbb{Z},\ p\in P\mathbb{Z},\ r\in J_i\right\}$$

(where J_i is as in section 1), and their Q-duals:

$$\mathscr{L}_i^* = \left\{ \begin{pmatrix} m & r \\ \bar{r} & p \end{pmatrix} \in \mathscr{V} \mid Pm \in \mathbb{Z},\ Mp \in \mathbb{Z},\ r \in \bar{J}_i^{-1} \mathscr{D}_K^{-1} \right\}.$$

Finally, fix one such lattice, $\mathscr{L}$, and choose V in $\mathscr{L}^*$. Then the theta functions we shall use are the functions of z in $\mathfrak{H}$ and g in $\mathbf{G}$

$$\theta_{k,\alpha}(z, V, g, \mathscr{L}) = y^{1/2} \sum_{X \in \mathscr{L}} \eta_{k,\alpha}((X-V)^g) e((xQ + iyR^g)[X-V]).$$

We sometimes drop the $\mathscr{L}$ to shorten the notation. Set $N_1 = \text{l.c.m.}\ (MP, \mathbf{N}(I)D)$. We have

Proposition 3.1.

For all $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ *in* $\Gamma_0(N_1)$

$$\theta_{k,\alpha}(z, dV, g) | [\sigma]_k = \chi_{-D}(d) \theta_{k,\alpha}(z, V, g).$$

In particular, for σ *in* $\Gamma(N_1)$

$$\theta_{k,\alpha}(z, V, g) | [\sigma]_k = \theta_{k,\alpha}(z, V, g).$$

We omit the proof, as well as a more complete description of the properties of $\theta_{k,\alpha}$; see Friedberg [3] for details.

For the rest of this paper fix an f in $S\{N, k, \lambda\}$, $k \geqq 1$.

Definition 3.2. *Given a lattice* $\mathscr{L}$ *such that* $N_1 | N$, *and* V *in* $\mathscr{L}^*$, *the lift of* f *associated to* $\mathscr{L}$ *and* V *is given for* $|\alpha| \leqq k-1$ *by*

$$F_\alpha(g, V, \mathscr{L}) = \int_{\Gamma(N) \backslash \mathfrak{H}} \theta_{k,\alpha}(z, V, g, \mathscr{L}) \overline{f(z)} y^{k-2} dx dy, \tag{3.2}$$

with associated vector form

$$F(g, V, \mathscr{L}) = {}^t(F_{1-k}, \ldots, F_{k-1}).$$

The integral makes sense by the Proposition above.

Let $G(\mathscr{L}) = \{\gamma \in \mathbf{G} \mid \mathscr{L}^\gamma = \mathscr{L}\}$, and denote by ι the main involution of $M(2, \mathbb{C})$: $\begin{pmatrix} a & b \\ c & d \end{pmatrix}^\iota = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}$. Each $G(\mathscr{L}_i)$ is a congruence subgroup of $SL(2, \mathcal{O}_K)_i$ – cf. [3]. Basic properties of the lifted form F are collected in the following Theorem; in particular, by part (1) and Proposition 2.2, corresponding to F is a function F^1 on $\mathfrak{H}^1$ of weight $2k-2$.

Theorem 3.3.

(1) *For* κ *in* $\mathbf{K}$, γ *in* $G(\mathscr{L})$, $F(\gamma g \kappa, V) = {}^t\varrho_{2k-2}(\kappa) F(g, V^\gamma)$.

(2) $F(g, V, \mathscr{L}) \doteq (-1)^{k-1} {}^t\varrho_{2k-2} \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} F(\bar{g}, \bar{V}, \bar{\mathscr{L}})$.

(3) $F({}^t g^{-1}, V, \mathscr{L}) = (-1)^{k-1} F(\bar{g}, V^\iota, \mathscr{L}^\iota)$.

(4) $F\left(\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} g \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}, V\right) = (-1)^{k-1} {}^t\varrho_{2k-2} \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix} F\left(g, V^{\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)}\right)$.

(5) *For* z *in* $\mathfrak{H}^1$, $\iota_{2k-2} F^1(z, V, \mathscr{L}) = F^1\left(z, \bar{V}^{\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)}, \bar{\mathscr{L}}\right)$.

Proof. Because these properties follow formally from properties of $\theta_{k,\alpha}$, the proofs of (1) to (4) are exactly the same as those given in [3] for the case of holomorphic forms f; we omit the details here. Then (5) follows from (2) and (4).

To show that the lifted form F is an eigenfunction of the Casimir operator, and thus a generalized Hilbert modular form in the sense of section 2, we compare the actions of the $\mathfrak{H}$ and $SL(2,\mathbb{C})$ Casimir operators on the theta kernel $\theta_{k,\alpha}(z,g)$ $(=\theta_{k,\alpha}(z,V,g,\mathscr{L}_i))$ (recall that the $SL(2,\mathbb{R})$ Casimir operator gives $-\Omega^{k,(2k-k^2)/4}$ on $\mathfrak{H}$).

Theorem 3.4. $\Psi\theta_{k,\alpha}(z,g)=\tilde{\Psi}\theta_{k,\alpha}(z,g)=-\Omega^{k,(2k-k^2)/4}\theta_{k,\alpha}(z,g)$.

We defer the proof briefly to note two corollaries.

Corollary 3.5. $(\Psi F_\alpha)(g)=(\tilde{\Psi}F_\alpha)(g)=((k^2-2k)/4+\lambda)F_\alpha(g)$.

Proof. By the theorem above and Propositions 1.3(4) and 1.4,

$$\begin{aligned}(\Psi F_\alpha)(g)=(\tilde{\Psi}F_\alpha)(g)&=\langle\Psi\theta_{k,\alpha}(z,g),f(z)\rangle\\&=\langle-\Omega^{k,(2k-k^2)/4}\theta_{k,\alpha}(z,g),f(z)\rangle\\&=-\langle\theta_{k,\alpha}(z,g),\Omega^{k,(2k-k^2)/4}f(z)\rangle\\&=-\langle\theta_{k,\alpha}(z,g),((2k-k^2)/4-\lambda)f(z)\rangle\\&=((k^2-2k)/4+\lambda)F_\alpha(g).\end{aligned}$$

Summing over a $G(\mathscr{L})$-invariant collection of V's, we have:

Corollary 3.6. *The lifted forms on $\mathfrak{H}^1$ corresponding to f in $S\{N,k,\lambda\}$ are generalized Hilbert modular forms for K of weight $2k-2$ and eigenvalue $(k^2-2k)/4+\lambda$.*

Proof. This follows by combining the corollary above with Theorem 3.3 and some estimates. We shall study the growth of F (and in particular determine when it is a cusp form) in section 4 below, so we do not reproduce them here.

The proof of Theorem 3.4 requires the following two lemmas. We let $X=\begin{pmatrix}m & r\\ \bar{r} & p\end{pmatrix}\in\mathscr{V}$, and $\eta_{k,\alpha}=\eta_{k,\alpha}(X)$.

Lemma 3.7.

$$\frac{-\partial}{\partial\bar{r}}\eta_{k,\alpha-1}=\frac{\partial}{\partial r}\eta_{k,\alpha+1}=\frac{\partial}{\partial m}\eta_{k,\alpha}=-\frac{\partial}{\partial p}\eta_{k,\alpha}=(k-1)\eta_{k-1,\alpha}.$$

Proof. Recall the following formulas for Laguerre and Hermite polynomials:

(i) $H_\gamma'(x)=\gamma H_{\gamma-1}(x)$
(ii) $xL_\beta^{(\alpha)\prime}(x)=\beta L_\beta^{(\alpha)}(x)-(\beta+\alpha)L_{\beta-1}^{(\alpha)}(x)$
(iii) $L_\beta^{(\alpha+1)}(x)-L_{\beta-1}^{(\alpha+1)}(x)=L_\beta^{(\alpha)}(x)$
(iv) $L_\beta^{(\alpha-1)\prime}(x)=-L_{\beta-1}^{(\alpha)}(x)$

(where $\beta,\gamma\geqq 0$, and we take $L_{-1}^{(\alpha)}=H_{-1}=0$). To prove the lemma, one combines these with (3.1). For example, let $\sum$ (respectively $\sum'$) denote the sum over β and γ nonnegative integers such that $2\beta+\gamma=k-\alpha-1$ (respectively $2\beta+\gamma=k-\alpha-2$)

and $\alpha+\beta\geqq 0$. Using (i), we see

$$\begin{aligned}\frac{\partial}{\partial m}\eta_{k,\alpha}=-\frac{\partial}{\partial p}\eta_{k,\alpha}&=(k-1)!\sum(\alpha+\beta)!^{-1}\gamma!^{-1}2^{-\beta}r^{\alpha}L_{\beta}^{(\alpha)}(2r\bar{r})H_{\gamma}'(m-p)\\&=(k-1)!\sum{}'(\alpha+\beta)!^{-1}\gamma!^{-1}2^{-\beta}r^{\alpha}L_{\beta}^{(\alpha)}(2r\bar{r})H_{\gamma}(m-p)\\&=(k-1)\eta_{k-1,\alpha}.\end{aligned}$$

The other equalities are similar.

Lemma 3.8. *Let* $\eta_{k,\alpha}^{*}(X)=\eta_{k,\alpha}(X)e(iyR[X])$, *and* $\Delta=\dfrac{\partial^2}{\partial r\partial\bar{r}}-\dfrac{\partial^2}{\partial m\partial p}$. *Then*

$$\Delta\eta_{k,\alpha}^{*}(X)=2\pi y(\pi yQ[X]-k)\eta_{k,\alpha}^{*}(X).$$

Proof. We induct on k; a direct computation gives the result for $k=1$, $\alpha=0$. Given the formula up to $k-1$ and all corresponding α, we use the following recursion formula for the $\eta_{k,\alpha}$, due to Asai [1], to do the k-th case: For $X=\begin{pmatrix} m & r\\ \bar{r} & p\end{pmatrix}$ in $\mathscr{V}$, $k\geqq 1$

$$\eta_{k,\alpha}(X)=-\bar{r}\eta_{k-1,\alpha+1}(X)+(m-p)\eta_{k-1,\alpha}(X)+r\eta_{k-1,\alpha-1}(X)$$

(where $\eta_{k,n}=0$ if $|n|\geqq k$). We get (with $Q=Q[X]$, etc.)

$$\begin{aligned}\Delta\eta_{k,\alpha}^{*}&=\Delta(-\bar{r}\eta_{k-1,\alpha+1}^{*}+(m-p)\eta_{k-1,\alpha}^{*}+r\eta_{k-1,\alpha-1}^{*})\\&=\{2\pi y(\pi yQ-(k-1))\eta_{k,\alpha}^{*}\}+2\pi y\{\bar{r}\eta_{k-1,\alpha+1}^{*}+(p-m)\eta_{k-1,\alpha}^{*}-r\eta_{k-1,\alpha-1}^{*}\}\\&\quad+e(iyR)\left\{-\frac{\partial}{\partial r}\eta_{k-1,\alpha+1}+\left(\frac{\partial}{\partial m}-\frac{\partial}{\partial p}\right)\eta_{k-1,\alpha}+\frac{\partial}{\partial\bar{r}}\eta_{k-1,\alpha-1}\right\}\\&=2\pi y(\pi yQ-k)\eta_{k,\alpha}^{*}.\end{aligned}$$

Note that we have used Lemma 3.7 in the computation.

Finally, to prove Theorem 3.4, observe that $C^{\infty}(\mathscr{V})$ embeds in $C^{\infty}(\mathbf{G})$ by associating to h in $C^{\infty}(\mathscr{V})$ the function $h_X(g)=h(X^g)$ in $C^{\infty}(\mathbf{G})$. The Casimir operator Ψ acts on $C^{\infty}(\mathscr{V})$ via an operator Ψ_0, such that $(\Psi_0 h)_X(g)=\Psi(h_X(g))$. Asai ([1], page 153) has computed Ψ_0 and the similarly defined operator $\tilde{\Psi}_0$, and found that they are equal, and given by

$$\Psi_0=\tilde{\Psi}_0=\tfrac{1}{2}(\tfrac{1}{2}L^2+L-Q[X]\Delta),$$

where $L=m\dfrac{\partial}{\partial m}+n\dfrac{\partial}{\partial n}+r\dfrac{\partial}{\partial r}+\bar{r}\dfrac{\partial}{\partial\bar{r}}$. Since $\theta_{k,\alpha}(z,g)$ is in the image of $C^{\infty}(\mathscr{V})$, it is enough to show that

$$\Psi_0(\eta_{k,\alpha}^{*})y^{1/2}e(xQ)=-\Omega^{k,(2k-k^2)/4}(y^{1/2}\eta_{k,\alpha}^{*}e(xQ))$$

since Q is $\mathbf{G}$-invariant. Now $\eta_{k,\alpha}$ is homogeneous of degree $k-1$, so

$$L\eta_{k,\alpha}^{*}=(k-1-2\pi yR)\eta_{k,\alpha}^{*}$$

and thus

$$\Psi_0\eta_{k,\alpha}^{*}=\{\pi^2y^2(R^2-Q^2)+k\pi y(Q-R)-\pi yR+(k-1)^2/4+(k-1)/2\}\eta_{k,\alpha}^{*}. \quad (3.3)$$

Since

$$-2i\frac{\partial}{\partial x}(y^{1/2}\eta^*_{k,\alpha}e(xQ))=2\pi Q(y^{1/2}\eta^*_{k,\alpha}e(xQ))$$

$$\left(-2\frac{\partial}{\partial y}+y^{-1}\right)(y^{1/2}\eta^*_{k,\alpha}e(xQ))=2\pi R(y^{1/2}\eta^*_{k,\alpha}e(xQ))$$

a direct computation of $\Omega(y^{1/2}\eta^*_{k,\alpha}e(xQ))$ is simple, and comparing this with (3.3) completes the proof.

It is interesting to compare our proof above, which uses Asai's formulas for the $\eta_{k,\alpha}$ to carry out a direct computation, with the Weil representation approach used by Shintani in [12] to get a related result.

As a closing remark, the extra variable V in $\theta_{k,\alpha}$ allows us to mix in characters, just as in the holomorphic lift case. We give one example; for others, see [3]. Suppose f is in $S\{N,k,\chi\}_0$. Fix lattice $\mathscr{L}$. If we embed the set

$$S=\{v\in\mathbb{Z}/MP\mathbb{Z}|\text{g.c.d.}(v,MP)=1\}$$

in $\mathscr{L}^*$ via $\mathscr{S}(v)=\begin{pmatrix} v/P & 0\\ 0 & 0\end{pmatrix}$, then a short computation shows that $S \bmod \mathscr{L}$ is fixed by $G(\mathscr{L})$. Set

$$\theta_{k,\alpha}(z,g,\chi)=\sum_{v\in S}(\chi\chi_{-D})(v)\theta_{k,\alpha}(z,\mathscr{S}(v),g).$$

Then

$$\theta_{k,\alpha}(z,g,\chi)|[\sigma]_k=\chi(d)\theta_{k,\alpha}(z,g,\chi)$$

for all σ in $\Gamma_0(N_1)$, and further

$$\theta_{k,\alpha}(z,g'g,\chi)=(\chi\chi_{-D})(\mathbf{N}(\delta))\theta_{k,\alpha}(z,g,\chi)$$

for $g'=\begin{pmatrix}\alpha' & \beta\\ \gamma & \delta\end{pmatrix}$ in $G(\mathscr{L})$ such that β is in PJ_i. Thus the associated lift given by Petersson inner product with θ (a sum of the lifts of Definition 3.2) satisfies

$$F(g'g)=(\chi\chi_{-D})(\mathbf{N}(\delta))F(g)$$

for all g' as indicated above. We shall call this lift *the Doi-Naganuma lift with character.*

4. Computing the Lift

Throughout this section let $\mathscr{L}$ be one of the $\mathscr{L}_i$'s above, and J be the corresponding J_i. Write V in $\mathscr{L}^*$ as $V=\begin{pmatrix}v_1 & v_2\\ \bar{v}_2 & v_4\end{pmatrix}$, and for $X=\begin{pmatrix}m & r\\ \bar{r} & p\end{pmatrix}$ in $\mathscr{V}$, put $X-V=\begin{pmatrix}m_1 & r_2\\ \bar{r}_2 & p_4\end{pmatrix}$. Take a set of coset representatives $\mathscr{R}$ for $\begin{pmatrix}1 & \mathbb{Z}\\ 0 & 1\end{pmatrix}\backslash SL(2,\mathbb{Z})/\Gamma(N)$ such that each $\varrho=\begin{pmatrix}\varrho_1 & \varrho_2\\ \varrho_3 & \varrho_4\end{pmatrix}$ in $\mathscr{R}$ has $\varrho_3\neq 0$. Let

$$C_\alpha=(k-1)!\,N(\mathbf{N}(I)^2D)^{-1/2}\pi^{-k}2^{2-3k/2}i^{k-\alpha},$$

and for r, v_2 in $\bar{J}^{-1}\mathscr{D}_K^{-1}$, ϱ in $\mathscr{R}$, put

$$\tau(\varrho, r, v_2) = \varrho_3^{-1} e(-2\varrho_3^{-1}(\varrho_1 \mathbf{N}(r) + \mathrm{tr}(v_2 \bar{r}) + \varrho_4 \mathbf{N}(v_2)))$$
$$\times \sum_{j \in J/\varrho_3 J} e(-2\varrho_3^{-1}(\mathrm{tr}(\bar{j}(r + \varrho_4 v_2)) + \varrho_4 \mathbf{N}(j))).$$

We consider f in $S\{N, k, \lambda\}$, $k \geqq 1$, whose Fourier expansion is given by (1.1), and if f is in $S\{N, k, \lambda, \chi\}_0$, we write $f|\begin{bmatrix} 0 & -1 \\ N & 0 \end{bmatrix}_k$ in $S\{N, k, \lambda, \bar{\chi}\}_0$ as

$$f_N(z) = f|\begin{bmatrix} 0 & -1 \\ N & 0 \end{bmatrix}(z) = \sum b(n) W(2\pi n y, k, \lambda) e(2nx).$$

For μ in $\mathbb{C}$, $0 < v$ in $\mathbb{R}$, we set $g(\mu, v) = v^{-1/2}\begin{pmatrix} v & \mu \\ 0 & 1 \end{pmatrix}$, and for nonnegative integers β and γ, we define the functions

$$q_{\alpha,\beta,\gamma}(y, \lambda) = \sum_{m=0}^{\beta} \sum_{e=0}^{\beta+\gamma-m} (-1)^{\gamma+e} \binom{\beta+\gamma-m}{e} [(\beta-m)!(\alpha+m)! m! \gamma!]^{-1} 2^{\gamma}$$
$$\cdot \binom{\frac{k-1}{2} - \omega}{\beta+\gamma-m-e} y^{2m+e+\alpha+1} K_{2\omega-e}(2y)$$

and

$$q_\alpha(y, \lambda) = \sum_{\beta,\gamma} q_{\alpha,\beta,\gamma}(y, \lambda),$$

where K is again a K-Bessel function, $(a)_n = a(a-1)\ldots(a-n+1)$ (with $(a)_0 = 1$), and the sum is over nonnegative integers β and γ as in (3.1). In this section we shall show

Theorem 4.1. *The lift of f is given for $\alpha \geqq 0$ by*

$$F_\alpha(g(\mu, v)) = C_\alpha \sum_{\varrho \in \mathscr{R}} \sum_{\substack{t=1 \\ t\varrho_3 \equiv -Pv_1 \bmod MP}}^{\infty} t^{-1} \sum_{r \in \bar{J}^{-1}\mathscr{D}_K^{-1} - 0} |r|^{-k} \xi^{t\alpha}(r)$$
$$\cdot \overline{a_\varrho(N\mathbf{N}(r))} \tau(\varrho, r, v_2) q_\alpha(2\pi P^{-1} t v |r|, \lambda) e(2P^{-1} t(\mathrm{tr}(\bar{r}\mu) - \varrho_4 v_4))$$
$$+ (k-1)! \left[\binom{\frac{k-1+\alpha}{2}}{} ! (4\pi)^{(k-1-\alpha)/2} P\right]^{-1}$$
$$\cdot \delta_{v_1 \bmod M, 0} v \langle \phi_{\alpha, (k-1-\alpha)/2}(z, 0, v_2), f \rangle.$$

If $\alpha \leqq 0$, $F_\alpha(g, V, \mathscr{L}) = (-1)^\alpha F_{-\alpha}(\bar{g}, \bar{V}, \bar{\mathscr{L}})$.

Here $\delta_{v_1 \bmod M, 0}$ is one if $v_1 \in M\mathbb{Z}$, and zero otherwise, and the last term is zero if $\binom{\frac{k-1-\alpha}{2}}{} \notin \mathbb{Z}$. Notice that in fact precisely those cosets such that $(\varrho_3, MP) | Pv_1$ occur nontrivially in the sum. Though this expression for F is not in the same form as that of Proposition 2.6, as we have remarked above, it can be put in that form by use of the differential equation for the K-Bessel function together with identities involving the $K_{2\omega-e}$ such as (4.3) below; see also Lemma 4.8.

In many cases we can simplify this Theorem. For example, we will prove

Theorem 4.2. *The Doi-Naganuma lift with character of f in $S\{N,k,\lambda,\chi\}_0$, with $P=1$ and $M=N$, is given by*

$$F_\alpha(g(\mu,v)) = C'_\alpha \sum_{t=1}^{\infty} t^{-1}\chi\chi_{-D}(t) \sum_{r\in J-0} |r|^{-k}\xi^\alpha(r)\overline{b(N\cdot\mathbf{N}(r)/D\mathbf{N}(I)^2)} \cdot q_{|\alpha|}(2\pi vt|r|/[D\mathbf{N}(I)^2]^{1/2},\lambda)e(2\mathrm{tr}(r\partial^{-1}\mathbf{N}(I)^{-1}t\bar{\mu})). \tag{4.1}$$

Here $\partial\in\mathfrak{H}$ is the purely imaginary generator of $\mathscr{D}_K$, $\chi\chi_{-D}$ is a (not necessarily primitive) character mod N, *and*

$$C'_\alpha = C_{|\alpha|}[D\mathbf{N}(I)^2/N]^{k/2}Ni^{2-\alpha}\#\{\varrho\in\mathscr{R}|(\varrho_3,N)=1\}.$$

[It is of course simple to compute $\#\{\varrho\in\mathscr{R}|(\varrho_3,N)=1\}$, but its value is irrelevant for us.] Observe that Theorem 4.1 involves the Fourier expansion for f at *all the cusps*, while in Theorem 4.2 *only a single cusp is involved.*

In the case of a holomorphic form, these expressions can be simplified. Also, these theorems can be used to compute the Fourier expansion of F at any cusp, and to deal with lattices containing some $\mathscr{L}_i$ as a sublattice of finite index. Consult [3] for details.

Notice that using the well-known asymptotic properties of K-Bessel functions, one sees that F does satisfy the required growth estimate of Definition 2.4(3). Also, the criterion that F have no constant term at a certain cusp is always an orthogonality condition [with respect to the Petersson inner product on $\Gamma(N')$ for some N' such that $N|N'$ (note f is in $S\{N',k\}$)], involving $\phi_{\alpha,\beta}(z,0,v_2;J_i)$ for some J_i and v_2. Using Corollary 1.8, we have

Corollary 4.3. *All lifts* (3.2) *of a Maass wave form f of weight k and eigenvalue λ are cuspidal if and only if f is orthogonal to the space spanned by*

$$\{\phi_{\alpha,(k-1-\alpha)/2}(z,0,v;J_i)|I \text{ ideal of } K,\ J_i=I\mathfrak{E}_i\mathfrak{e}_i^{-1},\ v \text{ in } \bar{J}_i^{-1}\mathscr{D}_K^{-1},\ 1\leqq i\leqq h(K)\}.$$

In particular, this holds when $\lambda\notin\{(\alpha^2-(k-1)^2)/4|\alpha\in\mathbb{Z},\ 0\leqq\alpha\leqq k-1\}$.

The last sentence here may also be proven by combining Theorem 2.7(2) and Corollary 3.6.

Corollary 4.4. *Almost all lifts of $S\{N,k\}$ are cuspidal.*

As an additional corollary of these formulas, the lift is 'arithmetic' in the following sense (the notion in this context is somewhat tentative, due to lack of Hermitian structure; compare Kudla [7]). Let $\mathbb{Q}_{ab}$ and K_{ab} denote the maximal abelian extensions of $\mathbb{Q}$ and K.

Corollary 4.5. *There exists a constant C such that for all f in $S\{N,k,\chi\}_0$ with Fourier coefficients in $\mathbb{Q}_{ab}$, the lift* (4.1) *of f has Fourier coefficients in $C\cdot K_{ab}$.*

Serre has conjectured to us that there exists such f only for special eigenvalues (such as $\lambda=0$).

To prove the Theorems, we first recall two results from Friedberg [3], whose proofs we omit.

Proposition 4.6. *For* $\alpha \geqq 0$,

$$\theta_{k,\alpha}(z, V, g(\mu, v)) = (k-1)! \sum_{\beta,\gamma} (\alpha+\beta)!^{-1}\gamma!^{-1} i^{\gamma} v^{\gamma+1} (4\pi)^{-\beta} P^{-1}$$

$$\cdot \Big\{ \delta_{v_1 \bmod M, 0} \delta_{\gamma,0} \phi_{\alpha,\beta}(z, 0, v_2)$$

$$+ \sum_{\varrho \in \mathscr{R}} \sum_{\substack{t=1 \\ t\varrho_3 \equiv -Pv_1 \bmod MP}}^{\infty} (P^{-1}t)^{\gamma} \sum_{\sigma \in \Gamma_\infty \backslash \Gamma(N)} (\phi_{\alpha,\beta}(z, P^{-1}t\mu\varrho_4, v_2 - P^{-1}t\mu\varrho_3)$$

$$|[\varrho^{-1}]_{\alpha+2\beta+1} y^{-\gamma} e(iP^{-2}t^2v^2y^{-1} - P^{-1}t(\mathrm{tr}(v_2\bar{\mu}\varrho_4) + 2\varrho_4 v_4)))|[\sigma\varrho]_k \Big\}. \quad (4.2)$$

This result will allow us to use Rankin's method to compute the Doi-Naganuma lift, when combined with the computation of the Fourier expansion of the $\phi_{\alpha,\beta}$ at an arbitrary cusp. This is given in

Proposition 4.7. *Let* $\sigma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in SL(2, \mathbb{Z})$, $c \neq 0$, *and for* v' *in* $\bar{J}^{-1}\mathscr{D}_K^{-1}$, *set*

$$\tau'(\sigma, u, v, v') = (\mathbf{N}(I)^2 D)^{-1/2} i \sum_{v'' \in \bar{J}/cJ} e(2c^{-1}(a\mathbf{N}(v'') + d\mathbf{N}(v'))$$
$$- \mathrm{tr}(\bar{v}'(bv - du - 2v''c^{-1}))).$$

Then

$$\phi_{\alpha,\beta}(z, u, v)|[\sigma]_{\alpha+2\beta+1} = (-1)^{\alpha+1} c^{\alpha+2\beta-1} \sum_{v' \in \bar{J}^{-1}\mathscr{D}_K^{-1}/cJ} \tau'(\sigma, u, v, v')$$
$$\cdot \phi_{\alpha,\beta}(c^2 z, c(bv - du), u - c^{-1}(av - v'))$$

where the sum is over a fixed set of coset representatives v'.

Simplification of this Proposition is sometimes possible; cf. [3].

In order to get the formula of Theorem 4.1, we proceed as follows. The computation of the 'constant' term of F_α is immediate from Proposition 4.6. As for the rest, one first writes

$$\langle \theta_{k,\alpha}, f \rangle = \langle \theta_{k,\alpha}|[\varrho^{-1}], f|[\varrho^{-1}] \rangle.$$

Using Rankin's method and (4.2) the integral (3.2) defining F_α reduces to an integral over a strip of width N, and thus, by Fubini's theorem, to a double integral. After applying Proposition 4.7, everything in the integral is completely explicit, and so we evaluate. The x integral is simple. The y integral is computed by the following lemma.

Lemma 4.8. *Let* a, b *be in* $\mathbb{R}^+$, *and* α, β, γ *be as above, with* $\alpha \geqq 0$. *Then*

$$\int_0^\infty L_\beta^{(\alpha)}(ay) e^{-ay/2 - b/y} W_{k/2,\omega}(ay) y^{\alpha+\beta-k/2-1} dy$$

$$= 2 \sum_{m=0}^{\beta} (-1)^{\beta+\gamma} \binom{\beta+\alpha}{\beta-m} a^{m+1/2} m!^{-1} \left(\frac{d}{db}\right)^{\beta+\gamma-m} \{b^{(k-1)/2} K_{2\omega}(2(ab)^{1/2})\}$$

$$= 2 \sum_{m=0}^{\beta} \sum_{e=0}^{\beta+\gamma-m} (-1)^{\beta+\gamma+e} \binom{\beta+\alpha}{\beta-m} \binom{\beta+\gamma-m}{e} ((k-1)/2-\omega)_{\gamma+\beta-m-e} (m!)^{-1}$$
$$\cdot a^{m+(e+1)/2} b^{m+(\alpha-\gamma+e)/2} K_{2\omega-e}(2(ab)^{1/2}).$$

Proof. This follows by expanding out the Laguerre polynomial, taking derivatives with respect to b of the equality ([4], pg. 868)

$$\int_0^\infty e^{-ay/2-b/y} W_{k/2,\omega}(ay) y^{k/2-2} dy = 2a^{1/2} b^{(k-1)/2} K_{2\omega}(2(ab)^{1/2})$$

and then, for the second formula, writing the resulting derivatives of K-Bessel functions in terms of K-Bessel functions, using

$$yK_w'(y) + wK_w(y) = -yK_{w-1}(y) \tag{4.3}$$

(one can obtain other expressions for the integral by using different formulas at this stage).

This result allows one to easily complete the evaluation of F_α for $\alpha \geqq 0$. Since the relationship between $\alpha \geqq 0$ and $\alpha \leqq 0$ is simply a restatement of Theorem 3.3(2), this finishes the proof of Theorem 4.1.

Next we state and prove a more general version of Theorem 4.2. Let $P=1$, $M=N$. We compute the lift associated to V in $\mathscr{L}^*$ such that $(v_1, N)=1$ in another way, avoiding the sum of Proposition 4.7. Namely, for ϱ in $\mathscr{R}$, we may assume $(\varrho_3, N)=1$, or else ϱ does not occur in the sum (4.2). Then there are integers w and w' such that $w\varrho_3 + Nw'\varrho_4 = 1$, so

$$\varrho^{-1} = \tau_1 \tau_2,$$

$$\tau_1 = \begin{pmatrix} 0 & -1 \\ N & 0 \end{pmatrix} \begin{pmatrix} -\varrho_3 & w' \\ -N\varrho_4 & -w \end{pmatrix}, \qquad \tau_2 = \begin{pmatrix} N^{-1} & -N^{-1}\varrho_3^{-1}\varrho_1 + w'\varrho_3^{-1} \\ 0 & 1 \end{pmatrix}.$$

One sees that for f in $S\{N, k, \chi\}_0$,

$$\left\langle \sum_{\sigma \in \Gamma_\infty \backslash \Gamma(N)} \left\{ \phi_{\alpha,\beta}(z, t\mu\varrho_4, v_2 - t\mu\varrho_3) | [\varrho^{-1}] y^{-\gamma} e(it^2 v^2 y^{-1}) \right\} | [\sigma\varrho], f \right\rangle$$

$$= \bar{\chi}(-\varrho_3) N^{-\gamma/2} \int_0^\infty \int_0^1 \phi_{\alpha,\beta}(z, t\mu\varrho_4, v_2 - t\mu\varrho_3) | [\tau_1] y^{-\gamma}$$

$$\cdot e(it^2 v^2 (Ny)^{-1}) \overline{f_N(z)} y^{k-2} dx dy.$$

Also, in Friedberg [3] we show

$$\phi_{\alpha,\beta}(z, t\mu\varrho_4, v_2 - t\mu\varrho_3) | [\tau_1]_{\alpha+2\beta+1}$$

$$= C_1 \chi_{-D}(-\varrho_3) \phi_{\alpha,\beta}(Nz/D\mathbf{N}(I)^2, \bar{\partial}^{-1} \mathbf{N}(I)^{-1}(wv_2 - t\mu), -\partial \mathbf{N}(I) \varrho_4 v_2)$$

where

$$C_1 = i^{\alpha-1} (N/D\mathbf{N}(I)^2)^{(\alpha+2\beta+1)/2}.$$

Combining these last two formulas with (4.2) and a computation similar to that sketched above, one gets the following result from which Theorem 4.2 immediately follows.

Theorem 4.9. *Let $P=1$, $M=N$, $(v_1, N)=1$, and f be in $S\{N, k, \lambda, \chi\}_0$. For each $m\in\mathbb{Z}$, $(m, N)=1$, let $m\tilde{m}\equiv 1 \bmod N$ for some $\tilde{m}\in\mathbb{Z}$. Then the lifting* (3.2) *is given by*

$$F_\alpha(g(\mu, v), V)=C''_\alpha \sum_{\varrho\in\mathscr{R}} \overline{\chi\chi_{-D}(-\varrho_3)} \sum_{\substack{t=1 \\ t\varrho_3\equiv -v_1 \bmod N}}^{\infty} t^{-1} \sum_{r\in J} \overline{b\left(\frac{N\mathbf{N}(r)}{D\mathbf{N}(I)^2}\right)} \xi^\alpha(r)|r|^{-k}$$
$$\cdot q_{|\alpha|}(2\pi tv|r|/(D\mathbf{N}(I)^2)^{1/2}, \lambda)$$
$$\cdot e(2\,\mathrm{tr}(r\partial^{-1}\mathbf{N}(I)^{-1}\overline{(t\mu-\tilde{\varrho}_3 v_2)})-2\tilde{\varrho}_3\varrho_4\mathbf{N}(v_2)-2t\varrho_4 v_4),$$

where $C''_\alpha=C'_\alpha(\#\{\varrho\in\mathscr{R}|(\varrho_3, N)=1\})^{-1}$.

Remark. By comparing the combinations of Bessel functions in the Theorems of this section with the Q_j of Theorem 2.7, one gets an alternate proof that the lifted form F *is* an eigenfunction of the Casimir operators, with eigenvalue $\lambda+(k^2-2k)/4$.

Next, we discuss Dirichlet series. For convenience we write

$$F_\alpha(g(\mu, v), V, \mathscr{L})=C_\alpha(0, V, \mathscr{L})v+c_\alpha\sum_r C(r, V, \mathscr{L})\xi^\alpha(r)|r|^{-k}q_{|\alpha|}(2\pi v|r|, \lambda)e(2\,\mathrm{tr}(r\bar{\mu})),$$

where c_α depends only on α, and the sum is over r in $P^{-1}\bar{J}^{-1}\mathscr{D}_K^{-1}$. $C(r, V, \mathscr{L})$ can of course be read off directly from the theorems above; note $C(\bar{r}, \bar{V}, \bar{\mathscr{L}})=C(r, V, \mathscr{L})$. This is not accidental, for by Theorem 3.3(2), $F_{-\alpha}(g, V, \mathscr{L})=(-1)^\alpha F_\alpha(\bar{g}, \bar{V}, \bar{\mathscr{L}})$, and yet the Fourier coefficients must be scalars (cf. (2.6)). Also, set

$$\varphi_\alpha(s)=\int_0^\infty v^{2s-k-1}q_{|\alpha|}(v, \lambda)dv.$$

A standard argument using the Mellin transform

$$\int_0^\infty (F_\alpha(g(0, v), V, \mathscr{L})-C_\alpha(0, V, \mathscr{L})v)v^{2s-k-1}dv$$

along with Theorem 3.3 (3), gives

Proposition 4.10. *The function*

$$\Phi_\alpha(s, V, \mathscr{L})=(2\pi)^{-2s}\varphi_\alpha(s)\sum_r C(r, V, \mathscr{L})\xi^\alpha(r)\mathbf{N}(r)^{-s}$$

has a meromorphic continuation to the entire s plane, which is holomorphic except for simple poles of residue $-C_\alpha(0, V, \mathscr{L})/2^{1+k}\pi^k c_\alpha$ *and* $(-1)^{k-1}C_\alpha(0, V^\iota, \mathscr{L}^\iota)/2^{1+k}\pi^k c_\alpha$ *at* $s=(k-1)/2$ *and* $(k+1)/2$ *respectively when these quantities are nonzero.* Φ_α *satisfies the functional equation*

$$\Phi_\alpha(s, V, \mathscr{L})=(-1)^{k-1}\Phi_\alpha(k-s, V^\iota, \mathscr{L}^\iota).$$

Now, the Mellin transform of a K-Bessel function is two to a power times the product of two gamma functions. Combining this with the van der Monde identity, one sees that

$$\varphi_\alpha(s)=P_{|\alpha|}(s)\Gamma(s+\omega+(|\alpha|+1-k)/2)\Gamma(s-\omega+(|\alpha|+1-k)/2),$$

where the polynomial $P_\alpha(s)$ is given for $\alpha \geqq 0$ by

$$P_\alpha(s) = \sum_{\beta,\gamma} 2^\gamma (-1)^{\beta+1} (\beta!\gamma!)^{-1} P_{\alpha,\beta,\gamma}(s)$$

with

$$P_{\alpha,\beta,\gamma}(s) = \sum_{m=0}^{\beta} (-1)^m \binom{\beta}{m} (\alpha+m)!^{-1} (s+m+\omega+(\alpha-1-k)/2)_m \cdot (s+m-\omega+(\alpha-1-k)/2)_m (s-\alpha/2-\beta-1)_{\beta+\gamma-m}.$$

Further, $P_\alpha(s)$ is invariant (up to sign) under the substitution $s \to k-s$:

$$P_\alpha(s) = (-1)^{k-\alpha-1} P_\alpha(k-s) \quad (\text{for } \alpha \geqq 0).$$

Though this can be proven by use of a generalized hypergeometric function, we prefer a number theoretic proof. In fact, we have:

Proposition 4.11. $P_{\alpha,\beta,\gamma}(s) = (-1)^{k-\alpha-1} P_{\alpha,\beta,\gamma}(k-s)$.

Proof. Fix any $\alpha \geqq 0$. We write

$$F_\alpha(g, V) = \sum_{\beta,\gamma} F_{\alpha,\beta,\gamma}(g, V)$$

where each $F_{\alpha,\beta,\gamma}(g, V)$ is defined by replacing q_α by $q_{\alpha,\beta,\gamma}$ in Theorem 4.1. Although $\{F_{\alpha,\beta,\gamma}(g, V)\}$ does *not* give a modular form on $\mathfrak{H}^1$, one can form its Mellin transform, exactly as above, obtaining $P_{\alpha,\beta,\gamma}(s)$ times the same gamma factors, $(2\pi)^{-2s}$, and Dirichlet series as in Proposition 4.10. Surprisingly, this Mellin transform still has an analytic continuation and functional equation. To see this, it is enough to prove that

$$F_{\alpha,\beta,\gamma}(g(0, v)^{-1}, V, \mathscr{L}) = (-1)^{k-1} F_{\alpha,\beta,\gamma}(g(0, v), V^\iota, \mathscr{L}^\iota). \tag{4.4}$$

Let $\eta_{\alpha,\beta,\gamma}$ be the corresponding (β,γ) term in the expression (3.1) for $\eta_{k,\alpha}$. A computation shows that

$$\eta_{\alpha,\beta,\gamma}\left(\bar{X}^{g(0,v)\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}}\right) = (-1)^{k-1} \eta_{\alpha,\beta,\gamma}(X^{g(0,v)}),$$

and then (4.4) follows, since ${}^t g^{-1} = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} g \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$ for g in $\mathbf{G}$, and $V^{\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}} = \bar{V}^\iota$ for V in $\mathscr{V}$. However, observe that if for one choice of (β,γ), the polynomial $P_{\alpha,\beta,\gamma}(s)$ satisfies Proposition 4.11, then these functional equations imply the result for all choices, since besides $P_{\alpha,\beta,\gamma}(s)$ identically equal terms arise in all Mellin transforms of the $F_{\alpha,\beta,\gamma}$. The $\beta = 0$ case always occurs, and there the Proposition is easy to verify. Thus it holds in general.

Combining this with Proposition 4.10, we get

Corollary 4.12. *Let*

$$\Phi_\alpha^*(s, V, \mathscr{L}) = (2\pi)^{-2s} \Gamma(s+\omega+(|\alpha|+1-k)/2) \Gamma(s-\omega+(|\alpha|+1-k)/2) \cdot \sum_r C(r, V, \mathscr{L}) \xi^\alpha(r) \mathbf{N}(r)^{-s}.$$

Then $\Phi_\alpha^*(s)$ has a meromorphic continuation to the entire s plane, and satisfies the functional equation

$$\Phi_\alpha^*(s, V, \mathscr{L}) = (-1)^\alpha \Phi_\alpha^*(k-s, V^\iota, \mathscr{L}^\iota).$$

Consider Proposition 4.10 in the case given by Theorem 4.2. There

$$C(r) = \sum_{t|r} t^{k-1} \chi\chi_{-D}(t)\overline{b(N\mathbf{N}(rt^{-1}))},$$

where the sum is over positive rational integers t such that $(t)|(r)$ in $\bar{J}^{-1}\mathscr{D}_K^{-1}$. When f is a normalized new (eigen)form, so that $b(n) = (\text{constant})\overline{a(n)}$, and K has class number one, the relevant Dirichlet series is

$$D_\alpha(s) = U^{-1} \sum_{r \in \mathscr{O}_K} C_1(r)\xi^\alpha(r)\mathbf{N}(r)^{-s}$$

where

$$C_1(r) = \sum_{t|r \text{ in } \mathscr{O}_K} t^{k-1}\chi\chi_{-D}(t)a(\mathbf{N}(rt^{-1})),$$

and U is the number of units of K (this changes Φ_α by constants and $(D\mathbf{N}(I))^{-s}$). Note C_1 is multiplicative, and for $\mathfrak{p}|p$, $p \nmid N$, we have

$$C_1(\mathfrak{p}^n) = \begin{cases} a(p^n) & \chi_{-D}(p) = 1 \\ \sum_{f=0}^{n} (-1)^f \chi(p)^f p^{f(k-1)} a(p^{2n-2f}) & \chi_{-D}(p) = -1 \end{cases}$$

Thus we can compute the Euler product for D_α. For example, when p remains prime, we have

$$\sum_{n \geq 0} \xi^\alpha(p^n) C_1(p^n) p^{-2ns} = \sum_{f \geq 0} a(p^{2f}) \sum_{n \geq f} p^{(n-f)(k-1)-2ns}(-\chi(p))^{n-f}$$

and combining this with the formula

$$\sum_{f \geq 0} a(p^{2f}) p^{-2fs} = (1 + \chi(p)p^{k-1-2s})(1 - (a(p^2) - \chi(p)p^{k-1})p^{-2s} + \chi(p^2)p^{2(k-1-2s)})^{-1}$$

one obtains the Euler p-factor. The other cases are similar, and so, for $\mathrm{Re}(s)$ sufficiently large, we find that

$$D_\alpha(s) = \prod_{\mathfrak{p}} (1 - \xi^\alpha(\mathfrak{p})C_1(\mathfrak{p})\mathbf{N}(\mathfrak{p})^{-s} + \xi^{2\alpha}(\mathfrak{p})\chi \circ \mathbf{N}(\mathfrak{p})\mathbf{N}(\mathfrak{p})^{k-1-2s})^{-1},$$

so that F is an eigenfunction of all the Hecke operators $T(\mathfrak{p})$, and also

$$D_0(s) = \sum_{n=1}^{\infty} a(n)n^{-s} \cdot \sum_{n=1}^{\infty} a(n)\chi_{-D}(n)n^{-s}.$$

Thus we have arrived at a functional equation for $D_0(s)$, analogous to the one discovered in the real quadratic case for holomorphic forms by Doi and Naganuma [2, 9].

Observe that by using Proposition 1.9, we also have a functional equation for the Dirichlet series (1.2) formed from the *negative* Fourier coefficients, since if $f(z)$ is in $S\{N, k, \lambda\}$, then so is the $g(z)$ of that Proposition. Also, as in the holomorphic

case, when $h(K)>1$ we can add the Dirichlet series corresponding to the $\mathscr{L}_i$ to obtain

$$\sum_{n=1}^{\infty} a(n)\psi(n)n^{-s}\cdot\sum_{n=1}^{\infty} a(n)\bar{\psi}(n)\chi_{-D}(n)n^{-s}, \tag{4.5}$$

where ψ_0 is any character of the ideal class group, and for each rational prime p we fix a prime $\mathfrak{p}_0|p$ in K and then set

$$\psi(p)=\psi_0(\mathfrak{p}_0),$$

and extend ψ multiplicatively (because (4.5) has an Euler product, it is independent of the choice of $\mathfrak{p}_0$).

Though we have so far treated only cusp forms of weight at least 1, several additional generalizations should be noted. Cusp forms of weight $k\leqq -1$ can be lifted, since then, as noted in Proposition 1.9 above, $y^k f(-\bar{z})$ is in $S\{N,-k\}$. Further, we can use the differential operator $\mathscr{K}$ of section one to take almost all forms of low weight bijectively onto forms of higher weight (Proposition 1.3), which we can then lift. Finally, we can extend the lifting to $G\{N,k\}$, $k>1$, by using (3.2) as usual, since $\theta_{k,\alpha}$ satisfies (4) of Definition 1.1; however, for $k=1$ this is not true, and the integral (3.2) need not converge.

As an illustration, we indicate how to compute the lifting of the (non-holomorphic) Eisenstein series of (integer) weight $k>1$,

$$E_{w,k,\sigma}=\sum_{\gamma\in\Gamma_\infty\backslash\Gamma(N)}\operatorname{Im}(\gamma\sigma z)^w j(\gamma\sigma,z)^{-k}\quad(2\operatorname{Re}(w)+k>2)$$

where w is in $\mathbb{C}$, σ is in $\Gamma(N)\backslash\Gamma(1)$, $j\left(\begin{pmatrix}a & b\\ c & d\end{pmatrix},z\right)=cz+d$, and $\operatorname{Re}(w)+k>3/2$. By Rankin's method, the integral (3.2) reduces to

$$\int_{\Gamma_\infty\backslash\mathfrak{H}}\theta_{k,\alpha}(z,V,g,\mathscr{L})|[\sigma^{-1}]_k y^{\bar{w}+k-2}dx\,dy.$$

Without loss of generality, $c\neq 0$.

Write $\sigma^{-1}=\begin{pmatrix}a & b\\ c & d\end{pmatrix}=\begin{pmatrix}1 & a/c\\ 0 & 1\end{pmatrix}\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}\begin{pmatrix}c & d\\ 0 & 1/c\end{pmatrix}$. Then we can compute the action of σ^{-1} in the corresponding three parts (as in the proof of Proposition 4.7 given in [3]), in order to get the Fourier expansion of $\theta_{k,\alpha}$ at the cusp $\sigma^{-1}(\infty)$. Finally, doing the integral, we get (up to some numerical terms involving σ, k, w and the Q-volume of $\mathscr{L}$, but not g or V)

$$F_\alpha(g,V,\mathscr{L})=(\text{constant})\sum_{\substack{X\in\mathscr{L}^*-0\\ Q[X]=0}}\eta_{k,\alpha}(X^g)R^g[X]^{1/2-\bar{w}-k}\tau_1(V,\sigma,X)$$

where

$$\tau_1(V,\sigma,X)=e(Q(aV-X,d(aV-X)/c-2bV)+abQ[V])\cdot\sum_{v'\in\mathscr{L}/c\mathscr{L}}e(Q(v',av'/c+2(aV-X)/c)).$$

References

1. Asai, T.: On the Doi-Naganuma lifting associated with imaginary quadratic fields. Nagoya Math. J. **71**, 149–167 (1978)
2. Doi, K., Naganuma, H.: On the functional equation of certain Dirichlet series. Invent. Math. **9**, 1–14 (1969)
3. Friedberg, S.: On the imaginary quadratic Doi-Naganuma lifting of modular forms of arbitrary level. Nagoya Math. J. **92** (1983)
4. Gradshteyn, I., Ryzhik, I.: Tables of integrals, series, and products. Corrected and enlarged edition. New York: Academic Press 1980
5. Hecke, E.: Mathematische Werke. Göttingen: Vandenhoeck and Ruprecht 1970
6. Hejhal, D.: The Selberg trace formula for $PSL(2, \mathbb{R})$, Vol. I. Lecture Notes in Math. 548. Berlin, Heidelberg, New York: Springer 1976; Vol. II (to appear)
7. Kudla, S.: On certain arithmetic automorphic forms for $SU(1, q)$. Invent. Math. **52**, 1–25 (1979)
8. Maass, H.: Lectures on modular functions of one complex variable. Bombay: Tata Institute of Fundamental Research 1964
9. Naganuma, H.: On the coincidence of two Dirichlet series associated with cusp forms of Hecke's "Neben"-type and Hilbert modular forms over a real quadratic field. J. Math. Soc. Jpn. **25**, 547–555 (1973)
10. Roelcke, W.: Das Eigenwertproblem der automorphen Formen in der hyperbolischen Ebene. I. Math. Ann. **167**, 292–337 (1966). II. Math. Ann. **168**, 261–324 (1967)
11. Shimura, G.: On certain zeta functions attached to two Hilbert modular forms. I. The case of Hecke characters. Ann. Math. **114**, 127–164 (1981)
12. Shintani, T.: On construction of holomorphic cusp forms of half integral weight. Nagoya Math. J. **58**, 83–126 (1975)
13. Terras, A.: Fourier analysis on symmetric spaces and applications to number theory. (To appear)
14. Weil, A.: Dirichlet series and automorphic forms. Lecture Notes in Mathematics 189. Berlin, Heidelberg, New York: Springer 1971

Received October 25, 1982; in revised form February 23, 1983

Math. Ann. 263, 509–514 (1983)

Mathematische Annalen
© Springer-Verlag 1983

p-Ranks of Class Groups in Z_p^d-Extensions

Paul Monsky

Department of Mathematics Brandeis University Waltham, MA 02254, USA

Introduction

Let L be a Galois extension of a number field k, with $E=G(L/k)$ isomorphic to a product of d copies of the additive group of p-adic integers. Let k_n be the fixed field of E^{p^n} and G_n be the ideal-class group of k_n. Let $\bar{e}_n(L/k)$ be the dimension of the Z/p vector space G_n/pG_n. We wish to study how $\bar{e}_n(L/k)$ grows with n.

To state our results we need to introduce a certain module $\bar{X}$ over the d-variable power series ring with co-efficients in Z/p. Explicitly we let $\bar{M}_L$ be the maximal abelian unramified extension of L whose Galois group is annihilated by p, and we set $\bar{X}=G(\bar{M}_L/L)$. We take $\bar{\Lambda}_d$ to be the completion, with respect to powers of the augmentation ideal, of the Z/p group ring of E; it is isomorphic to $Z/p[[\mathbf{X}_1,\ldots,\mathbf{X}_d]]$. $\bar{X}$ has the structure of finitely generated (but not necessarily torsion) $\bar{\Lambda}_d$-module; in fact $\bar{X}$ is just the reduction mod p of the usual Iwasawa-Greenberg module as defined in [1].

When $\bar{X}$ is a finite group we shall see that the $\bar{e}_n(L/k)$ are constant for large n. When $\bar{X}$ is infinite the situation is more interesting. Let a be the altitude of the local ring $\bar{\Lambda}_d/\mathrm{Ann}\bar{X}$, so that $1\leqq a\leqq d$. We shall prove the following results:

There is a real constant $c>0$ such that $\bar{e}_n(L/k)=cp^{an}+\delta_n$, where $\delta_n=O(p^{(a-1)n})$. When $a=d$, c is just the rank of $\bar{X}$ over $\bar{\Lambda}_d$. When $a=1$, c is an integer and δ_n is a periodic function of n for large n.

The proof uses [1] to reduce to a module-theoretic problem which may be handled by the methods of our paper, [2], on the Hilbert-Kunz function. Various interesting and difficult questions remain untouched. What can be said about δ_n when $a>1$? When $1<a<d$, is the constant c rational? And finally, do analogous results hold for the groups G_n/p^jG_n for fixed $j>1$ and large n?

1.

We show how to calculate the $\bar{e}_n(L/k)$ from a knowledge of $\bar{X}$ and some supplementary information about the inertia groups in $\bar{M}_L/k$. Our arguments closely parallel those made in Sects. 5 and 6 of [1]. We adopt the notation of the

introduction. Let $Y=G(\bar{M}_L/k)$. Y maps onto E with kernel $\bar{X}$. Let $Y^{(n)}\subset Y$ be the inverse image of E^{p^n}; then $Y^{(n)}=G(\bar{M}_L/k_n)$. For each prime of k which ramifies in L choose a prime of $\bar{M}_L$ lying over it. Let $\mathscr{J}_1,\ldots,\mathscr{J}_r\subset Y$ be the corresponding inertia groups and $\mathscr{J}_j^{(n)}=\mathscr{J}_j\cap Y^{(n)}$. The $\mathscr{J}_j^{(n)}$ are inertia groups in $\bar{M}_L/k_n$, and every inertia group in this extension is a conjugate of some $\mathscr{J}_j^{(n)}$ by some element of Y. $\mathscr{J}_j$ maps bijectively on a Z_p-submodule $\bar{\mathscr{J}}_j$ of E.

Now take $\sigma_1,\ldots,\sigma_d\in Y$ mapping on a basis of E. Let $\mathscr{J}_{r+j}$ be the closed subgroup of Y generated by σ_j^p, and let $\mathscr{J}_{r+j}^{(n)}=\mathscr{J}_{r+j}^{p^n}$.

Definition 1.1. *C_n is the closed subgroup of $Y^{(n)}$ generated by $[Y^{(n)},Y^{(n)}]$, all p^{th} powers of elements of $Y^{(n)}$, and all the inertia groups in $G(\bar{M}_L/k_n)$. B_n is the intersection of C_n with $\bar{X}$.*

Using class-field theory, we find as in [1]:

Lemma 1.2. *The Z/p-dimension of $Y^{(n)}/C_n$ is $\bar{e}_n(L/k)$. B_n is a $\bar{\Lambda}_d$-submodule of $\bar{X}$, and for n large, $\bar{e}_n(L/k)-l(\bar{X}/B_n)$ is constant.*

Observe now that C_n is the closed subgroup of $Y^{(n)}$ generated by $[Y^{(n)},Y^{(n)}]$ and all the conjugates under Y of the $\mathscr{J}_j^{(n)}$, $1\leqq j\leqq r+d$. So the situation we are in is virtually the same as that of Sect. 5 of [1] and we can define submodules A_n, A_n' and A_n'' of B_n as in that section. In particular we find:

Lemma 1.3. *Let $A_n=[\bar{X},Y^{(n)}]$. Then A_n is a $\bar{\Lambda}_d$-submodule of B_n. In fact, $A_n=I_n\bar{X}$ where I_n is the ideal generated by the $(1+\mathbf{X}_j)^{p^n}-1$ in $\bar{\Lambda}_d$.*

Since $p=0$ in $\bar{\Lambda}_d$, I_n is actually the ideal generated by the $\mathbf{X}_j^{p^n}$, or alternatively the ideal generated by all g^{p^n} with g in the maximal ideal, m, of $\bar{\Lambda}_d$. This enables us to use the results of [2] to estimate $l(\bar{X}/A_n)$. We next get information about B_n/A_n.

Definition 1.4. $$A_n'=A_n+\sum_1^{r+d}\bar{\Lambda}_d[Y,\mathscr{J}_j^{(n)}].$$

Lemma 1.5. *$A_n''=A_n'+[Y^{(n)},Y^{(n)}]$ is a $\bar{\Lambda}_d$-submodule of B_n. For each n, $(\sigma,\tau)\to[\sigma,\tau]$ induces a Z_p-bilinear map $E^{p^n}\times E^{p^n}\to A_n''/A_n'$ whose image generates A_n''/A_n'. Thus the groups A_n''/A_n' have bounded order.*

Lemma 1.6. *Let $J^{(n)}\subset\mathscr{J}_1^{(n)}\times\ldots\times\mathscr{J}_{r+d}^{(n)}$ consist of all $(\varrho_1,\ldots,\varrho_{r+d})$ with $\Pi\bar{\varrho}_i=1$. Then $(\varrho_1,\ldots,\varrho_{r+d})\to\varrho_1\cdot\varrho_2\ldots\varrho_{r+d}$ induces a Z_p-linear map of $J^{(n)}$ onto B_n/A_n''. Thus the groups B_n/A_n'' have bounded order.*

Lemma 1.7. *Let $(\mathcal{O},m)$ be a local Noetherian Z/p-algebra of altitude $a\geqq 1$ and M be a finitely generated $\mathcal{O}$-module. Let I_n be the ideal of $\mathcal{O}$ generated by all the x^{p^n}, $x\in m$. Suppose that $g\in m$ and that b is a fixed integer. Then $l((g^{p^n-b},I_n)M/I_nM)=O(p^{(a-1)n})$.*

Proof. We may assume that $M=\mathcal{O}$. Let $N\subset\mathcal{O}=\{x\mid x$ is annihilated by some power of $g\}$ and set $\mathcal{O}'=\mathcal{O}/(N,g^b)$. Since g^b is not a zero-divisor on $\mathcal{O}/N$, the altitude of $\mathcal{O}'$ is $\leqq a-1$. Now if m is generated by s elements, $I_n\supset m^{sp^n}$. We conclude that $l(\mathcal{O}'/I_n\mathcal{O}')\leqq l(\mathcal{O}'/m^{sp^n})=O(p^{(a-1)n})$. Note finally that, for n large, multiplication by g^{p^n-b} annihilates N and maps $\mathcal{O}'/I_n\mathcal{O}'=\mathcal{O}/(I_n,N,g^b)$ onto $(g^{p^n-b},I_n)/I_n$, completing the proof.

Theorem 1.8. *Let $\mathcal{O}=\bar{\Lambda}_d/\mathrm{Ann}\bar{X}$, and $a=$ altitude $\mathcal{O}$. Suppose that $a\geqq 1$. Then $\bar{e}_n(L/k)=l(\bar{X}/A_n)+O(p^{(a-1)n})$.*

Proof. $\bar{e}_n(L/k) = l(\bar{X}/B_n) +$ constant for large n. In view of lemmas 1.5 and 1.6 it suffices to show that $l(A'_n/A_n) = O(p^{(a-1)n})$. Now there exists a v such that for $n \geqq v$, $\mathscr{J}_j^{(n)} = (\mathscr{J}_j^{(v)})^{p^{n-v}}$. Using definition 1.4, and arguing as in the proof of Theorem 5.11 of [1] we find that A'_n/A_n is generated as $\bar{\Lambda}_d$-module by finitely many elements $((\bar{\sigma}^{p^n}-1)/(\bar{\sigma}^{p^v}-1))\cdot u$ where $\bar{\sigma}$ and u are independent of n. Set $g = \bar{\sigma} - 1$ and $b = p^v$. Then $(\bar{\sigma}^{p^n}-1)/(\bar{\sigma}^{p^v}-1) = g^{p^n-b}$. Lemma 1.7 then shows that A'_n/A_n is a homomorphic image of a product of a fixed finite number of $\mathcal{O}$-modules, each of which has length $O(p^{(a-1)n})$. The theorem follows.

Theorem 1.9. *Notation as in Theorem* 1.8. *Suppose that* $a \geqq 1$. *Then there is a real constant* $c > 0$ *such that* $\bar{e}_n(L/k) = cp^{an} + \delta_n$ *where* $\delta_n = O(p^{(a-1)n})$. *When* $a = 1$, c *is an integer. When* $a = d$, c *is just the rank of* $\bar{X}$ *over* $\bar{\Lambda}_d$.

Proof. The corollary to Theorem 1.8 of [2] shows that $l(\bar{X}/A_n) = l(\bar{X}/I_n\bar{X}) = c(\bar{X})p^{an} + O(p^{(a-\alpha)n})$ with $c(\bar{X}) > 0$, and we apply Theorem 1.8 of this paper. When $a = 1$, Theorem 1.10 of [2] shows that $c(\bar{X})$ is an integer. When $a = d$, $\operatorname{Ann}\bar{X} = (0)$ and $\mathcal{O} = \bar{\Lambda}_d$. Let r be the rank of $\bar{X}$ as $\mathcal{O}$-module. Then Lemma 1.3 of [2] shows that $c(\bar{X}) = c(\mathcal{O}^r)$. Since $l(\bar{\Lambda}_d/I_n) = p^{dn}$, $c(\mathcal{O}^r) = r$, completing the proof.

When $a = 1$, Theorem 3.11 of [2] shows that $l(\bar{X}/A_n) = c(\bar{X})p^n +$ (a periodic function of n) for large n. To show that a similar result holds for $l(\bar{X}/B_n)$, and consequently for the $\bar{e}_n(L/k)$, is our final task.

We shall assume until the last theorem of this section that the inertia groups in L/k are direct summands of E. Then $\mathscr{J}_j^{(n)} = \mathscr{J}_j^{p^n}$ for $1 \leqq j \leqq r+d$ and all n. We now make some constructions similar to those of Sect. 6 of [1]. We define functions $f_{\sigma,\tau}$, $g_{\sigma,\tau}$ and h_ϱ from the set of positive integers, N, to $\bar{X}$ as follows:

Definition 1.10. *If* σ *and* τ *are in* Y, $f_{\sigma,\tau}(l) = [\sigma^l, \tau]$, $g_{\sigma,\tau}(l) = [\sigma^l, \tau^l]$.
If $\varrho = (\varrho_1, \ldots, \varrho_{r+d})$ *is in* $J^{(0)}$, *(see Lemma* 1.6*), then* $h_\varrho(l) = \varrho_1^l \cdots \varrho_{r+d}^l$.

As in [1] we prove:

Lemma 1.11. A'_n *is generated as* $\bar{\Lambda}_d$*-module by* $I_n\bar{X}$ *and finitely many* $f_{\sigma,\tau}(p^n)$ *where the* σ *and* τ *do not depend on* n. A''_n *is generated as* $\bar{\Lambda}_d$*-module by* A'_n *and finitely many* $g_{\sigma,\tau}(p^n)$ *where the* σ *and* τ *do not depend on* n. B_n *is generated as* $\bar{\Lambda}_d$*-module by* A''_n *and finitely many* $h_\varrho(p^n)$ *where the* $\varrho \in J^{(0)}$ *do not depend on* n.

We now abstract a property shared by the functions $f_{\sigma,\tau}$, $g_{\sigma,\tau}$ and h_ϱ.

Definition 1.12. *Let* (A, m) *be a local* Z/p*-algebra. If* $g \in A$ *we write* $g_{(n)}$ *and* $g_{\langle n \rangle}$ *to denote* g^{p^n} *and* g^{p^n-1} *respectively. If* g *and* h *are in* A *we set* $(g,h)_n = \Sigma g^i h^j$, *the sum extending over all* i *and* j *with* $i+j = p^n - 2$. *Note that* $(g-h)\cdot(g,h)_n = g_{\langle n \rangle} - h_{\langle n \rangle}$.

Definition 1.13. *A function from* N *to an* A*-module* M *is "special" if it is of the form* $n \to \Sigma(g_i, h_i)_n u_i$ *where the* g_i *and* h_i *are in* m *and the* u_i *are in* M.

Lemma 1.14. *View* $\bar{X}$ *as a module over* $\bar{\Lambda}_d$, *or alternatively over* $\mathcal{O} = \bar{\Lambda}_d/\operatorname{Ann}\bar{X}$. *Then the functions* $n \to f_{\sigma,\tau}(p^n)$ *of Lemma* 1.11 *are special functions* $N \to \bar{X}$.

Proof. Suppose that σ and τ are in Y. Let g be the element $\bar{\sigma} - 1$ of the maximal ideal m of $\bar{\Lambda}_d$. By Lemma 6.5 of [1], $[\sigma^l, \tau] = \sum_{j=1}^{l} \binom{l}{j} g^{j-1}[\sigma, \tau]$. Taking $l = p^n$ we find that $f_{\sigma,\tau}(p^n) = g_{\langle n \rangle} \cdot [\sigma, \tau] = (g,g)_n \cdot (-g[\sigma,\tau])$, proving the result.

Lemma 1.15. *The functions* $n \to g_{\sigma,\tau}(p^n)$ *and* $n \to h_\varrho(p^n)$ *of Lemma* 1.11 *are special functions* $N \to \bar{X}$,

Proof. We give the proof for the h_ϱ; the argument for the $g_{\sigma,\tau}$ is the same. We adopt the notation of Lemma 6.6 of [1]. Then $h_\varrho(l) = lu_1 + \sum_{k=2}^{l} \binom{l}{k} u_k$. So if $n \geqq 1$, $h_\varrho(p^n) = u_{p^n}$. But Lemma 6.6 of [1] shows that $u_{p^n} = \sum_{1 \leqq i < j \leqq r+d} (z_j, z_{j+1})_n \bar{\tau}_i \bar{\tau}_j [\varrho_i, \varrho_j]$ with the $z_j \in m$. The lemma follows.

Combining the above results we find:

Theorem 1.16. *There is a finite set of special functions* $\psi_i : N \to \bar{X}$ *such that, for each* n, B_n *is the submodule of* $\bar{X}$ *generated by* $I_n \cdot \bar{X}$ *and the* $\psi_i(n)$.

Remark. Suppose that $\bar{X}$ is a finite group so that $\mathcal{O} = \bar{\Lambda}_d / \mathrm{Ann}\bar{X}$ is Artinian. Then, for n large, $I_n = (0)$ and $(g,h)_n = 0$ in $\mathcal{O}$. So $B_n = (0)$ and $\bar{e}_n(L/k)$ is constant for large n. (There are simpler proofs of this, of course.)

Section 2 will be devoted to a proof of the following variation on Theorem 3.11 of [2]:

Theorem 1.17. *Let* $(\mathcal{O}, m)$ *be a complete local Noetherian* Z/p*-algebra of altitude* 1, *reduced, with finite residue class field and let* I_n *be the ideal generated by the* g^{p^n}, $g \in m$. *Let* M *be a finitely generated* $\mathcal{O}$*-module and* ψ_i *be finitely many special functions* $N \to M$. *Let* B_n *be the submodule of* M *generated by* $I_n M$ *and the* $\psi_i(n)$. *Then* $l(M/B_n) = c(M) \cdot p^n + \delta_n$, *where* δ_n *is a periodic function of* n *for large* n.

The above results allow a precise description of the growth of the $\bar{e}_n(L/k)$ when $a = 1$:

Theorem 1.18. *Suppose we are in the situation of Theorem* 1.9 *with* $a = 1$. *Then* δ_n *is a periodic function of* n *for large* n.

Proof. Replacing the Z_p^d-extension L/k by L/k_s for some s does not change $\bar{X}$ as a group. But its structure as a module over $Z/p[[\mathbf{X}_1, \ldots, \mathbf{X}_d]]$ is changed, and the new $\mathrm{Ann}\bar{X}$ is precisely $\{g \mid g^{p^s} \in \text{the old } \mathrm{Ann}\bar{X}\}$. So if we replace k by k_s with s large we may assume that $\mathrm{Ann}\bar{X}$ is a radical ideal and that $\mathcal{O}$ is reduced. We may further assume that the inertia groups in $G(L/k)$ are direct summands of E, so that the conclusion of Theorem 1.16 holds. Then Theorem 1.17 and Lemma 1.2 give the result.

2.

This section is devoted to the proof of Theorem 1.17. Throughout $\mathcal{O}$, I_n, M and B_n are as in the statement of that theorem.

Definition 2.1. s *is the number of height* 0 *primes of* $\mathcal{O}$; $P_1, \ldots, P_s$ *are those primes.* $\mathcal{O}_i$ *is the integral closure of* $\mathcal{O}/P_i$ *in its field of quotients.* ord_i *is the valuation function in the complete discrete valuation ring* $\mathcal{O}_i$. $\mathcal{O}' = \prod_1^s \mathcal{O}_i$.

Remark. Suppose that M is a finitely generated $\mathcal{O}$-module. To say that M is a finite group is the same as saying that it has finite length over $\mathcal{O}$; that is to say that it is m-primary. This also amounts to the localization of M at each P_i being (0).

Now let ϕ_i be the map $\mathcal{O}\to\mathcal{O}_i$ and $\phi:\mathcal{O}\to\mathcal{O}'$ be the product of the ϕ_i. Since $\mathcal{O}$ is reduced, ϕ is injective.

Lemma 2.2. *Suppose that $u_i\in\mathcal{O}_i$ with each $\operatorname{ord}_i(u_i)$ large. Then $u=(u_1,\ldots,u_s)$ is in the image of ϕ.*

Proof. We show first that the cokernel of ϕ is m-primary. Since $\mathcal{O}_i$ is finite over $\mathcal{O}/P_i$, $\mathcal{O}'$ is a finite $\mathcal{O}$-module. So it suffices to show that the localization of ϕ at each P_i is onto. But this is easy – the localizations of $\mathcal{O}$ and $\mathcal{O}'$ at P_i are each just the field of quotients of $\mathcal{O}/P_i$. It follows that cokernel ϕ is annihilated by some h not in any P_i. Suppose now that $\operatorname{ord}_i(u_i)\geqq\operatorname{ord}_i(h)$. (By $\operatorname{ord}_i(h)$ we mean $\operatorname{ord}_i(\varphi_i(h))$; we shall consistently abuse language in this way). Then $u\in h\mathcal{O}'\subset(\text{image }\phi)$, and we're done.

Lemma 2.3. *Suppose that (A,m) is a complete Noetherian local Z/p-algebra with finite residue class field F, of degree r over Z/p. Suppose g and h are in A. Then as $n\to\infty$, $n\equiv$ a fixed $n_0(r)$, the sequences $g_{(n)}$, $g_{\langle n\rangle}$ and $(g,h)_n$ converge in A.*

Proof. A is a homomorphic image of $F[[\mathbf{X}_1,\ldots,\mathbf{X}_d]]$ for some d, and we can evidently replace A by that ring. If $g\in m$, $g_{(n)}$ and $g_{\langle n\rangle}\to 0$. If $g\notin m$, $g=\alpha\beta$ with $\alpha\in F$ and $\beta\equiv 1(m)$. Then if $n\equiv n_0\ (r)$, $g_{(n)}=\alpha_{(n)}\beta_{(n)}=\alpha_{(n_0)}\beta_{(n)}$ and this $\to\alpha_{(n_0)}$ as $n\to\infty$; if follows that $g_{\langle n\rangle}\to\alpha_{(n_0)}\cdot g^{-1}$. Suppose now that $g\neq h$. Since $(g-h)\cdot(g,h)_n=g_{\langle n\rangle}-h_{\langle n\rangle}$, we see from the above that $(g-h)\cdot(g,h)_n\to$ some u in A. Since the ideal $(g-h)A$ is closed, $u=(g-h)$ (some v) and we see easily that the $(g,h)\to v$. Finally $(g,g)_n=-g^{p^n-2}$ and the argument given for $g_{\langle n\rangle}$ shows that the $(g,g)_n$ converge.

Definition 2.4. *r_i is the Z/p-dimension of the residue class field of $\mathcal{O}_i$. r is the least common multiple of the r_i.*

We shall show that if r is taken as in Definition 2.4, then $l(M/B_n)=c(M)p^n+\delta_n$ where $\delta_{n+r}=\delta_n$ for large n. Until further notice we shall assume that the number of elements in the finite field $\mathcal{O}/m$ is $\geqq$ the number of height 0 primes s, of $\mathcal{O}$.

Lemma 2.5. *Under the above assumption there exists an $f\in m$ such that for each i, $\operatorname{ord}_i(f)=\min_{g\in m}(\operatorname{ord}_i(g))$.*

Proof. Let $K\subset\mathcal{O}$ be a co-efficient field projecting onto $\mathcal{O}/m$. Take an $f\in m$ such that $\operatorname{ord}_i(f)=\min_{g\in m}(\operatorname{ord}_i g)$ for a maximal subset J of the i's.

Suppose there is some index j such that $j\notin J$. Choose an h such that $\operatorname{ord}_j(h)=\min_{g\in m}(\operatorname{ord}_j(g))$. For $\lambda\in K$ set $h_\lambda=h+\lambda f$. Then $\operatorname{ord}_j(h_\lambda)=\operatorname{ord}_j(h)$.

Furthermore for each $i\in J$ there is at most one λ such that $\operatorname{ord}_i(h_\lambda)>\operatorname{ord}_i(f)$. Since $|K|>|J|$ we can find a λ such that $\operatorname{ord}_i(h_\lambda)=\operatorname{ord}_i(f)$ for each $i\in J$; this contradicts the definition of f.

Definition 2.6. *With f as in Lemma 2.5 and r as in Definition 2.4, we set $F=f_{\langle r\rangle}$. (Note that $F\cdot f=f_{(r)}$ and that $F_{(n)}\cdot f_{(n)}=f_{(n+r)}$.)*

Lemma 2.7. *Suppose that g and h are in m and that n is large. Then:*

(a) $F_{(n)} \cdot g_{(n)} \equiv g_{(n+r)} \bmod f_{(n+r)}$

(b) $F_{(n)} \cdot (g,h)_n \equiv (g,h)_{n+r} \bmod f_{(n+r)}$.

Proof. By our assumptions on f we may write $g = \alpha^{(i)} f$, $h = \beta^{(i)} f$ with $\alpha^{(i)}$ and $\beta^{(i)}$ in $\mathcal{O}_i$. (To simplify notation we write g for $\phi_i(g)$, h for $\phi_i(h)$). Then, in $\mathcal{O}_i$, $F_{(n)} \cdot g_{(n)} - g_{(n+r)} = f_{(n+r)}(\alpha^{(i)}_{(n)} - \alpha^{(i)}_{(n+r)})$. Let $u_i = \alpha^{(i)}_{(n)} - \alpha^{(i)}_{(n+r)}$ and $u = (u_1, \ldots, u_s)$. Since each r_i divides r, Lemma 2.3 applied to $\mathcal{O}_i$ shows that $\mathrm{ord}_i(u_i)$ is large. By Lemma 2.2, $u = \phi(x)$ for some $x \in \mathcal{O}$. Then $F_{(n)} g_{(n)} - g_{(n+r)} = f_{(n+r)} \cdot x$. Similarly, $(F_{(n)} \cdot (g,h)_n) - (g,h)_{n+r} = \frac{1}{f^2} \cdot f_{(n+r)}((\alpha^{(i)}, \beta^{(i)})_n - (\alpha^{(i)}, \beta^{(i)})_{n+r})$ in $\mathcal{O}_i$. Setting $u_i = \frac{1}{(\phi_i(f))^2}((\alpha^{(i)}, \beta^{(i)})_n - (\alpha^{(i)}, \beta^{(i)})_{n+r})$ and arguing as above, using Lemmas 2.3 and 2.2, gives the result. An immediate consequence is:

Corollary. (a) *Suppose $g \in m$, $u \in M$. For n large, $F_{(n)} \cdot (g_{(n)} u) \equiv g_{(n+r)} u$ mod the submodule $f_{(n+r)} M$ of M.*

(b) *Let $\psi : N \to M$ be a special function. For n large, $F_{(n)} \cdot (\psi(n)) \equiv \psi(n+r) \bmod f_{(n+r)} M$.*

Lemma 2.8. *For n large, $l(B_n / f_{(n)} M) = l(B_{n+r} / f_{(n+r)} M)$.*

Proof. (f) is an m-primary ideal; it follows that the above modules have finite length. So it suffices to construct an $\mathcal{O}$-linear map of $B_n / f_{(n)} M$ onto $B_{n+r} / f_{(n+r)} M$ for all large n. Now multiplication by $F_{(n)}$ induces a map $M / f_{(n)} M \to M / f_{(n+r)} M$. Let $g^{(i)}$ be generators of m and g_j be generators of M. The image of $B_n / f_{(n)} M$ under the above map is the submodule of $M / f_{(n+r)} M$ generated by the $F_{(n)} \cdot (g^{(i)}_{(n)} u_j)$ and the $F_{(n)} \cdot \psi_i(n)$. But the above corollary shows this submodule to be precisely $B_{n+r} / f_{(n+r)} M$ for large n.

We shall now prove Theorem 1.17; in fact we shall show that $\delta_{n+r} = \delta_n$ for large n if r is chosen as in Definition 2.4. Suppose first that $|\mathcal{O}/m| \geqq s$, so that f may be chosen as in Lemma 2.5. Since (f) is m-primary, $l(M / f^k M)$ is a polynomial of degree $\leqq 1$ in k for large k. Taking $k = p^n$ and applying Lemma 2.8 gives the result.

The general case may be handled by an "extension of ground field" argument. Let $K \subset \mathcal{O}$ be a co-efficient field projecting onto $\mathcal{O}/m$. Let l be a large prime, $K^{\#}$ the degree l extension of K, and $\mathcal{O}^{\#}$ the tensor product of $\mathcal{O}$ with $K^{\#}$ over K. Then $\mathcal{O}^{\#}$ is complete local Noetherian of altitude 1, reduced, with residue class field $K^{\#}$. We may assume that l does not divide r. It is then easily seen that the height 0 primes of $\mathcal{O}^{\#}$ are just the $P_i \mathcal{O}^{\#}$, and that the "r_i" for this prime ideal is lr_i. Now let $M^{\#}$ be the tensor product of M with $\mathcal{O}^{\#}$ over $\mathcal{O}$ and $B_n^{\#}$ be the $\mathcal{O}^{\#}$-submodule of $M^{\#}$ generated by $I_n M^{\#}$ and the $\psi_i(n)$. Then the $\mathcal{O}$-length of M / B_n is just the $\mathcal{O}^{\#}$-length of $M^{\#} / B_n^{\#}$. So by our above result, applied to $\mathcal{O}^{\#}$, $l(M/B_n) = c(M^{\#}) p^n + \delta_n$ where $\delta_{n+lr} = \delta_n$ for large n. Replacing l by another large prime l' we get the desired result.

References

1. Cuoco, A., Monsky, P.: Class numbers in Z_p^d-extensions. Math. Ann. **255**, 235–258 (1981)
2. Monsky, P.: The Hilbert-Kunz function

Received March 18, 1983

Math. Ann. 263, 515–532 (1983)

Mathematische Annalen
© Springer-Verlag 1983

Solving the Degenerate Complex Monge-Ampère Equation with One Concentrated Singularity

László Lempert

Department of Analysis I, Eötvös University, Muzeum krt 6–8, H-1088 Budapest, Hungary

1. Introduction

In this paper we shall be concerned with the Dirichlet problem for the degenerate Monge-Ampère equation in a strictly convex smooth domain $D \subset \mathbb{C}^n$ $(n>1)$. Let $\varphi : \partial D \to \mathbb{R}$ be a smooth function. Then the problem is to find a continuous $u : \bar{D} \to \mathbb{R}$ such that

$$u \text{ is plurisubharmonic in } D \tag{1.1}$$

$$\det \left| \frac{\partial^2 u}{\partial z_j \partial \bar{z}_k} \right| = 0 \quad \text{in } D \tag{1.2}$$

$$u|\widehat{\partial D} = \varphi \tag{1.3}$$

We shall briefly describe what is known about the solution of the above system when D is a ball. Then there is a unique generalized solution u (in the sense explained in [BT]). This u is in $C^{1,1}(D)$ (that is, grad$u$ is locally Lipschitz), satisfies (1.1), (1.3), and almost everywhere (1.2) (see Bedford and Taylor [BT], where more general results can also be found).

On the other hand u need not be C^2 however smooth φ is. A particularly simple counterexample is due to T. Gamelin and N. Sibony:

Let D be the unit ball of $\mathbb{C}^2$,

$$u(z_1, z_2) = \max\{(|z_1|^2 - 1/2)_+^2, (|z_2|^2 - 1/2)_+^2\}.$$

Here a_+ stands for $\max(a, 0)$. Then u is a solution of the Dirichlet problem corresponding to

$$\varphi(z) = (|z_1|^2 - 1/2)^2 = (|z_2|^2 - 1/2)^2 \qquad (z \in \partial D).$$

φ is analytic but u is not C^2 at the hypersurfaces $|z_1|^2 = 1/2$ and $|z_2|^2 = 1/2$.

This example shows that the singular support of u may be fairly large. It is not known how large it can be. In fact, at present the singularities of a solution of (1.1)...(1.3) seem to be beyond control.

In this paper we shall show that by introducing a sufficiently large singularity at one point $Z\in D$, the system (1.1)...(1.3) will admit a solution u that is smooth in $\bar{D}\backslash\{Z\}$. More precisely we shall prove the following

Theorem. *Let D be a strictly convex, analytically bounded domain in $\mathbb{C}^n$, let $\varphi:\partial D\to\mathbb{R}$ be a real analytic function, and let $Z\in D$. Then the problem*

$$\begin{cases} u \text{ is plurisubharmonic in } D \\ \det|\partial^2 u(z)/\partial z_j\partial\bar{z}_k|=0 & \text{if } z\in D\backslash\{Z\} \\ u(z)=c\log|z-Z|+0(1) & \text{as } z\to Z \\ u(z)=\varphi(z) & \text{if } z\in\partial D, \end{cases} \tag{1.4}$$

where c is a positive number, admits a unique real analytic solution $u:\bar{D}\backslash\{Z\}\to\mathbb{R}$, provided $c>c_0$. Here c_0 depends on D, Z, and φ.

The condition on D means that there is a real analytic $r:\mathbb{C}^n\to\mathbb{R}$ such that $r(z)\to\infty$ as $|z|\to\infty$, the real hessian of r is positive definite, and $D=\{z:r(z)<0\}$.

The idea of the proof is quite simple. When $\varphi\equiv 0$, the assertion has been proved in [L1] (Theorems 3 and 4); the general statement will be obtained from this special case by a perturbation. – Perturbation is a common device for dealing with nonlinear equations. In particular, it has recently been applied to the Monge-Ampère equation by Bedford [B] and Moriyón [M]. In our proof there will be common points with Bedford's proof, although the setup is rather different.

Neverthless we feel that the above theorem is worth establishing. We interprete it in the following way. As we have seen above, a given smooth boundary value φ cannot always be continued to a smooth solution of the Monge-Ampère equation. Such a φ has some irregularity imperceptible for the naked eye that forces every continuation to be singular somewhere. Nevertheless, every φ can be continued to a solution that is singular only in one point. The singularity at that point is then necessarily large.

We shall now complement the Theorem by describing properties of the solution u. All these will show that for $c>c_0$ (1.4) is as well behaved as any nondegenerate equation can be. In what follows, u will denote the solution of (1.4), where $c>c_0$. Furthermore $U=\{\zeta\in\mathbb{C}:|\zeta|<1\}$.

Complement 1. *The rank of the matrix $u_{z_j\bar{z}_k}(z)$ is $n-1$ if $z\in\bar{D}\backslash\{Z\}$.*

Complement 2. *For every $z\in\bar{D}\backslash\{Z\}$ there is a holomorphic imbedding $f_z=f:\bar{U}\to\bar{D}$ such that $z, Z\in f(\bar{U})$ and u is harmonic restricted to $f(U)\backslash\{Z\}$. This f is unique up to automorphisms of U. Consequently the corresponding analytic discs $f_z(\bar{U})$ yield a foliation of $\bar{D}$, singular at Z.*

Complement 3. *For every $z\in\bar{D}$*

$$u(z)=\sup_F c\log|F(z)|,$$

where the supremum is taken over holomorphic functions $F:\bar{D}\to\mathbb{C}$ such that $F(Z)=0$ and $c\log|F(w)|\leqq\varphi(w)$ for $w\in\partial D$.

We shall now discuss the behaviour of u at the singular point Z. Since the only singularity occurs at Z, one might expect that singularity to be particularly nasty. Surprisingly this is not the case and the singularity can essentially be resolved by blowing $\bar{D}$ up.

Complement 4. *Let $\mu:\bar{D}^*\to\bar{D}$ denote the blowing up of $\bar{D}$ at Z. Put $u^*=u\circ\mu$. Then* $\exp(2u^*/c)$ *is real analytic on $\bar{D}^*$.*

Finally we shall consider dependence of u (or rather of $\exp(2u^*/c)$) on φ. Instead of working in complicated topological vector spaces we shall introduce some Banach spaces of functions. Thus, let V be an open neighbourhood of ∂D in the complexification of $\mathbb{C}^n\cong\mathbb{R}^{2n}$. Every φ analytic on ∂D extends to a holomorphic function on some V. Let $\Phi(V)$ denote the Banach space of those functions $\varphi:\partial D\to\mathbb{R}$ that extend to a bounded holomorphic function of V. The norm in $\Phi(V)$ is the sup-norm over V of the smallest extension.

In a similar way, consider first the blowing up $(\mathbb{C}^n)^*$ of $\mathbb{C}^n$ at Z, and, regarded as a real manifold, complexify it. Let W be a neighbourhood of $\bar{D}^*$ in this complexification. Let $\Psi(W)$ denote the Banach space of those functions $\bar{D}^*\to\mathbb{R}$ that extend to a bounded holomorphic function on W, the norm being the sup-norm of the extension.

In the following complement V, W, $\Phi(V)$, and $\Psi(W)$ will have the meaning explained above.

Complement 5. *For any V and $\varphi_0\in\Phi(V)$ there are a c_0 and a W such that:*

1. *If $c>c_0$ and $\varphi\in\Phi(V)$ is near φ_0, then* (1.4) *has a (unique) solution u, such that* $\exp(2u^*/c)\in\Psi(W)$.

2. *The mapping*

$$\Phi(V)\ni\varphi\mapsto\exp(2u^*/c)\in\Psi(W)$$

is real analytic near φ_0

We conclude this introduction by remarking that all the above results have analogues in the C^∞ category. The proofs are similar but more complicated at some points.

2. About the Proofs

Replacing φ by $c\varphi$ and u by cu in the Theorem, one finds that it is sufficient to prove that (1.4) admits a unique real analytic solution $u:\bar{D}\setminus\{Z\}\to\mathbb{R}$ when $c=1$ and φ is small enough (i.e. sufficiently near to 0 in some space $\Phi(V)$). Also the complements need only be proved when $c=1$ and φ is small. This, along with $Z=0$ *will be assumed throughout the rest of the paper.*

Suppose now that u is a solution and consider a holomorphic imbedding $f:\bar{U}\to\bar{D}$ such that $f(0)=Z=0$. (It is understood that $f(\partial U)\subset\partial D$, in accordance with the concept of imbedding one manifold-with-boundary into another.) Then $u\circ f:\bar{U}\to\mathbb{R}$ is subharmonic, has a logarithmic pole at 0 and its boundary value is $\varphi\circ f$. Denoting the Poisson kernel of U by $P(\zeta,e^{it})$ ($\zeta\in U$, $t\in\mathbb{R}$), this implies

$$u(f(\zeta))\leqq\log|\zeta|+\int_0^{2\pi}\varphi(f(e^{it}))P(\zeta,e^{it})dt\,.$$

Hence, given a point $z\in D$,

$$u(z)\leqq \inf_f\left\{\log|\zeta|+\int_0^{2\pi}\varphi(f(e^{it}))P(\zeta,e^{it})dt\right\}. \tag{2.1}$$

where the infimum is taken over all holomorphic imbeddings $f:\bar{U}\to\bar{D}$ such that $f(0)=0$, $f(\zeta)=z$.

Now, it turns out that for small φ equality holds in (2.1) and there is an essentially unique f for which the minimal value is attained. These extremal imbeddings will yield the foliation figuring in Complement 2.

When $\varphi\equiv 0$, the extremal mappings were characterized in [L1] and more explicitly in [L2]: f is extremal if and only if there exists a holomorphic mapping $\tilde{f}:\bar{U}\to\mathbb{C}^n$ and a positive analytic function $p:\partial U\to\mathbb{R}^+=\{x\in\mathbb{R}:x>0\}$ such that

$$\tilde{f}(\zeta)=\zeta p(\zeta)r_z(f(\zeta))\qquad(\zeta\in\partial U).$$

Here r is a defining function of D so that $r_{\bar{z}}(f(\zeta))$ is a normal vector to ∂D in $f(\zeta)$.

In order to obtain a similar characterization for a nonzero φ, it will be convenient to assume that φ is defined (and real analytic) in a neighbourhood of $\partial D\subset\mathbb{C}^n$ and that it is small even in that neighbourhood. Then it will turn out that an imbedding f is extremal if and only if there is a holomorphic mapping $\tilde{f}:\bar{U}\to\mathbb{C}^n$ and a positive (real) analytic function $p:\partial U\to\mathbb{R}^+$ such that

$$\tilde{f}(\zeta)=\zeta[\varphi_z(f(\zeta))+p(\zeta)r_z(f(\zeta))]\qquad(\zeta\in\partial U) \tag{2.2}$$

and

$$\operatorname{Re} f'(0)\tilde{f}(0)=1/2. \tag{2.3}$$

Here we write ab for $\sum a_jb_j$ if $a=(a_j)\in\mathbb{C}^n$, $b=(b_j)\in\mathbb{C}^n$. As to (2.2) cf. also [BK].

Definition. A holomorphic imbedding $f:\bar{U}\to\bar{D}$ such that $f(0)=0$ will be called a φ-mapping if there is a holomorphic $\tilde{f}:\bar{U}\to\mathbb{C}^n$ and an analytic $p:\partial U\to\mathbb{R}^+$ such that (2.2) and (2.3) hold.

The first step in the proof of the Theorem will be to show that, when φ is small, to any given $z\in\bar{D}\setminus\{0\}$ there is a φ-mapping $f=f_z$ such that $f(\bar{U})\ni z$. This will depend on the corresponding result for $\varphi\equiv 0$ and the implicit function theorem (Chap. 4). To prepare this latter, in Chap. 3. we shall investigate a functional equation.

Once the φ-mappings f_z are constructed, u will be defined by

$$u(z)=\log|\zeta|+\int_0^{2\pi}\varphi(f_z(e^{it}))P(\zeta,e^{it})dt,$$

where $\zeta=f_z^{-1}(z)$.

From this definition it will be evident that u satisfies all conditions of (1.4) but the first; namely the plurisubharmonicity is not trivial. In Chap. 5. we shall essentially prove Complement 3, whence it will follow that u is indeed p.s.h. This will then complete the proof of the Theorem. Also the proof of the other complements will be easy using the information concerning φ-mappings obtained in Chapters 4. and 5.

3. A Functional Equation

The following problem will be discussed here. Given an $m \times m$ matrix valued function γ on ∂U and a function $\psi : \partial U \to \mathbb{C}^m$, a holomorphic $h : \bar{U} \to \mathbb{C}^m$ is wanted such that the function

$$\bar{h} + \gamma h - \psi,$$

defined on ∂U, extends to a holomorphic function in $\bar{U}$. Here h is considered as a column vector. The more general case will also be treated when $\bar{h}$ too has a matrix coefficient. The result we shall need in the next chapter is contained in Lemma 3.

This problem is rather implicitly solved in [L1] (in the proof of Proposition 10) and more explicitly in [L2] (in the proof of Proposition 13) but in neither case in the function spaces we shall need here and which we are now going to introduce.

Given a real number $\sigma > 1$, let

$$U_\sigma = \{\zeta \in \mathbb{C} : \sigma > |\zeta| > \sigma^{-1}\}.$$

Let furthermore $A = A_m = A_{m,\sigma}$ denote the Banach space of those mappings $\psi : \partial U \to \mathbb{C}^m$ that extend to a continuous mapping $\psi : \bar{U}_\sigma \to \mathbb{C}^m$ so that $\check{\psi}$ is holomorphic in U_σ and the generalized derivatives

$$\frac{\partial}{\partial t}\psi(\sigma e^{it}), \qquad \frac{\partial}{\partial t}\psi(\sigma^{-1} e^{it})$$

are square integrable on $0 \leqq t \leqq 2\pi$. Equivalently speaking $\check{\psi}'$ is assumed to belong to the Hardy space $H^2(U_\sigma)$. A function $\psi(e^{it}) = \sum_{-\infty}^{\infty} a_k e^{ikt}$, where $a_k \in \mathbb{C}^m$. is in the space A if and only if

$$\sum_{-\infty}^{\infty} (1 + k^2)|a_k|^2 \sigma^{2|k|} < \infty \tag{3.1}$$

The norm $[\psi] = [\psi]_\varepsilon$ of ψ will be defined by

$$[\psi] = \left(\int_0^{2\pi} |\check{\psi}(\sigma e^{it})|^2 dt\right)^{1/2} + \left(\int_0^{2\pi} |\check{\psi}(\sigma^{-1} e^{it})|^2 dt\right)^{1/2}$$
$$+ \varepsilon\left(\int_0^{2\pi} \left|\frac{\partial}{\partial t}\check{\psi}(\sigma e^{it})\right|^2 dt\right)^{1/2} + \varepsilon\left(\int_0^{2\pi} \left|\frac{\partial}{\partial t}\check{\psi}(\sigma^{-1} e^{it})\right|^2 dt\right)^{1/2},$$

where ε is a positive number to be fixed later.

Let furthermore $B = B_m = B_{m,\sigma}$ denote the space of those $h \in A$ that extend to a function holomorphic in U; $\bar{B} = \{\bar{h} : h \in B\}$. We stress that elements of B, $\bar{B}$, and A are considered only as defined on ∂U (and not on $\bar{U}_\sigma \cup U$ or $\bar{U}_\sigma$).

One more piece of notation: if M is a numerical matrix of size $m \times m$, $\|M\|$ stands for its (operator) norm.

Lemma 1. *Let $\gamma = (\gamma_{ij})$ be an $m \times m$ matrix valued function on ∂U, whose complex valued elements γ_{ij} are analytic on ∂U, and*

$$\max\{\|\gamma(\zeta)\| : \zeta \in \partial U\} < 1.$$

Let furthermore $\psi \in A_{m,\sigma}$. Consider the equation

$$\bar{h} + \gamma h - \psi \quad \text{extends holomorphically from } \partial U \text{ to } \bar{U}, \tag{3.2}$$

where $h \in B_{m,\sigma}$ is the unknown function.

Then there is a number $\sigma_0 > 1$ depending only on γ such that for $1 < \sigma < \sigma_0$

*a. The solutions of the homogeneous (3.2) (*i.e. *when $\psi \equiv 0$) constitute a $2m$-dimensional vector space over $\mathbb{R}$;*

b. The inhomogeneous (3.2) admits a solution $h \in B_{m,\sigma}$ (and then the solutions depend on $2m$ real parameters).

Proof. We shall prove that (3.2) has a unique solution h that assumes a given value $a \in \mathbb{C}^m$ at the origin. This will then prove both *a.* and *b.*

Define a (non surjective) projection $\pi : A \to \bar{B}$ by

$$\pi\left(\sum_{-\infty}^{\infty} a_k e^{ikt}\right) = \sum_{-\infty}^{-1} a_k e^{ikt}.$$

Let furthermore $P = \bar{\pi}$, i.e.

$$P\left(\sum_{-\infty}^{\infty} a_k e^{ikt}\right) = \sum_{-\infty}^{-1} \bar{a}_k e^{-ikt}.$$

A $g \in A$ extends to a holomorphic function on $\bar{U}$ if and only if $\pi(g) = 0$. Therefore (3.2) can be written

$$\pi(\bar{h} + \gamma h - \psi) = 0.$$

In view of the condition $h(0) = a$, this means

$$h = P(\psi - \gamma h) + a. \tag{3.3}$$

Thus the function we are looking for is a fixed point of the mapping $K : B \to B$ given by

$$K(h) = P(\psi - \gamma h) + a.$$

We shall now show that if $\sigma > 1$ is near to 1 and ε is small then K is a contraction. To this end

$$[P(\gamma g)] < q[g] \tag{3.4}$$

should be ensured for all $g \in B$, with $q < 1$.

Extend γ to a holomorphic function in a neighbourhood of ∂U and choose σ so near to 1 that γ be holomorphic on $\bar{U}_\sigma$ and $\mu = \max\{\|\gamma(\zeta)\| : \zeta \in \bar{U}_\sigma\} < 1$.

Then

$$\int_0^{2\pi} |\gamma(\sigma e^{it}) g(\sigma e^{it})|^2 dt \leqq \mu^2 \int_0^{2\pi} |g(\sigma e^{it})|^2 dt,$$

$$\left(\int_0^{2\pi} \left|\frac{\partial}{\partial t}\{\gamma(\sigma e^{it}) g(\sigma e^{it})\}\right|^2 dt\right)^{1/2}$$

$$\leqq \left(\int_0^{2\pi} \left|\frac{\partial}{\partial t} \gamma(\sigma e^{it}) g(\sigma e^{it})\right|^2 dt\right)^{1/2} + \left(\int_0^{2\pi} \left|\gamma(\sigma e^{it}) \frac{\partial}{\partial t} g(\sigma e^{it})\right|^2 dt\right)^{1/2}$$

$$\leqq M\left(\int_0^{2\pi} |g(\sigma e^{it})|^2 dt\right)^{1/2} + \mu\left(\int_0^{2\pi} \left|\frac{\partial}{\partial t} g(\sigma e^{it})\right|^2 dt\right)^{1/2},$$

where $M=\sigma\max\{\|\gamma'(\zeta)\|:\zeta\in\bar{U}_\sigma\}$. Similar estimates hold when σ is replaced by σ^{-1}. Hence

$$[\gamma g]\leqq\mu[g]+\varepsilon M[g].$$

Choose now ε so small that $\mu+\varepsilon M<q<1$, then $[\gamma g]<q[g]$. The norm of the projection P being 1, (3.4) follows. Thus K is a contraction, admitting exactly one fixed point. Therefore (3.3) has exactly one solution, q.e.d.

Lemma 2. *If $1<\sigma<\sigma_0$ and $\zeta_0\in\bar{U}$ is arbitrary then (3.2) has a unique solution $h\in B_{m,\sigma}$ that assumes a given value b at ζ_0.*

Proof. By virtue of the previous lemma it is sufficient to treat the homogeneous case. Also it will be enough to prove unicity, namely that $h(\zeta_0)=0$ implies $h\equiv 0$.

Now this follows from Lemma 1 (or, better to say, from its proof) if $\zeta_0=0$. The case $\zeta_0\in U$ can be reduced to $\zeta_0=0$ by applying an automorphism of U. What is left is the case $\zeta_0\in\partial U$.

Let e.g. $\zeta_0=1$, i.e. suppose $h(1)=0$. Let furthermore g denote the holomorphic continuation of $\bar{h}+\gamma h$ to U; then g is holomorphic on $\bar{U}$, and, of course, $g(1)=0$. Therefore the functions

$$\overline{\left(\frac{h(\zeta)}{\zeta-1}\right)},\qquad \gamma(\zeta)h(\zeta)\frac{\zeta}{1-\zeta},\qquad g(\zeta)\frac{\zeta}{1-\zeta} \tag{3.5}$$

are analytic on ∂U, and the first and the third extends antiholomorphically resp. holomorphically to $\bar{U}$. Let the expansions of these three functions be

$$\sum_{-\infty}^{0}a_k\zeta^k,\qquad \sum_{-\infty}^{\infty}b_k\zeta^k,\qquad \sum_{1}^{\infty}c_k\zeta^k$$

$(\zeta\in\partial U)$.

Now $\overline{\zeta-1}=(1-\zeta)/\zeta$ if $\zeta\in\partial U$, so that the sum of the first two functions in (3.5) is the third. Therefore $a_k=-b_k$ $(k\leqq 0)$ so that by Parseval's formula

$$\int_0^{2\pi}\left|\frac{h(e^{it})}{e^{it}-1}\right|^2dt=2\pi\sum_{-\infty}^{0}|a_k|^2\leqq 2\pi\sum_{-\infty}^{\infty}|b_k|^2$$
$$=\int_0^{2\pi}\left|\gamma(e^{it})\frac{h(e^{it})}{e^{it}-1}e^{it}\right|^2dt<\int_0^{2\pi}\left|\frac{h(e^{it})}{e^{it}-1}\right|^2dt,$$

at least if $h\not\equiv 0$. This is impossible, thus $h\equiv 0$, q.e.d.

Lemma 3. *Let α,β be $m\times m$ matrix valued analytic functions on ∂U, such that for every $\zeta\in\partial U$ $\alpha(\zeta)$ is self adjoint and positive definite, and $\beta(\zeta)$ is symmetric. Suppose that for an arbitrary $v\in\mathbb{C}^m\setminus\{0\}$ (conceived as a column vector) and $\zeta\in\partial U$*

$$v^*\overline{\alpha(\zeta)}v>|v^t\beta(\zeta)v|, \tag{3.6}$$

where * and ^{t} denote adjoint resp. transpose.

Given a $\psi\in A_{m,\sigma}$, consider the equation for the unknown $h\in B_{m,\sigma}$

$$\alpha\bar{h}+\beta h-\psi \text{ extends holomorphically from } \partial U \text{ to } \bar{U}. \tag{3.7}$$

Then there is a $\sigma_0>1$ depending only on α,β such that if $1<\sigma<\sigma_0$, (3.7) admits a unique solution h that assumes a given value $b\in\mathbb{C}^m$ in a given point $\zeta_0\in\bar{U}$.

The proof of this lemma is a reduction to the previous lemma with the aid of

Lemma 4. *Let α be as in Lemma 3. Then there is a holomorphic function $\delta : \bar{U} \to \mathrm{GL}(m, \mathbb{C})$ such that $\delta\delta^* = \alpha$ on ∂U.*

This is Theorème *B* of [L 1].

Proof of Lemma 3. Let δ be as guaranteed by Lemma 4. Upon multiplying by δ^{-1} from the left in (3.7) we obtain

$$\delta^* \bar{h} + \delta^{-1}\beta h - \delta^{-1}\psi \text{ extends holomorphically to } \bar{U}.$$

Putting $h_1 = \delta^t h$, $\gamma = \delta^{-1}\beta(\delta^t)^{-1}$, and $\psi_1 = \delta^{-1}\psi$, we infer

$$\overline{h_1} + \gamma h_1 - \psi_1 \text{ extends holomorphically to } \bar{U}. \tag{3.8}$$

Now from (3.6) one deduces $|v^t \gamma(\zeta) v| < 1$, if $|v| \leqq 1$, $\zeta \in \partial U$. Since γ is symmetric, this implies $\|\gamma(\zeta)\| < 1$, so that Lemma 2. can be applied to (3.8), and the claim of Lemma 3. follows.

4. Construction of φ-Mappings

Return now to the Monge-Ampère equation (1.4) (where $Z = 0$, $c = 1$, and φ is small). In Chap. 2 φ-mappings were defined and their relevance to the solution of (1.4) was pointed out. Here we are going to show that given a $z \in \bar{D} \setminus \{0\}$, there is a φ-mapping $f : \bar{U} \to \bar{D}$ such that $z = f(\xi)$ with some $\xi \in (0, 1]$. Moreover f depends analytically on z.

The case $\varphi \equiv 0$ was settled in [L 1]. Unfortunately, φ-mapping with $\varphi \equiv 0$ (i.e. 0-mappings) were called there "stationnaire". (To increase the confusion, in [L 2] the same mappings were termed *E*-mappings.) In these two papers it was shown that 0-mappings are extremal (in the sense discussed in Chapter 2) and that for any $z \in \bar{D} \setminus \{0\}$ there is a unique 0-mapping $f : \bar{U} \to \bar{D}$ such that $f(\xi) = z$ with some real $\xi \in (0, 1]$. (See Theorème 2 of [L 1] or Theorem 3 of [L 2].)

The existence of φ-mappings with $\varphi \not\equiv 0$ will be proved by perturbing the above result. In the proof we shall use the Banach spaces A, B defined in Chap. 3, the space $\Phi(V)$ defined in the Introduction, the space

$$Q = Q_\sigma = \{q \in A_{1,\sigma} : q \text{ is real on } \partial U\},$$

and the space

$$B^0 = B^0_{m,\sigma} = \{f \in B_{m,\sigma} : \text{the holomorphic extension of } f \text{ vanishes at } 0\}.$$

In the following, whenever f denotes a φ-mapping, p and $\tilde{f}$ will denote the mappings occuring in the definition of φ-mappings. At present we do not know that f determines p and $\tilde{f}$ uniquely (although this will turn out later on), so for the moment let us say that p and $\tilde{f}$ denotes any pair of mappings satisfying the conditions of the definition.

Lemma 5. *Let $f_0 : \bar{U} \to \bar{D}$ be a 0-mapping, $f_0(\xi_0) = z_0$ with $0 < \xi_0 \leqq 1$. Then for any neighbourhood V of ∂D in the complexification of $\mathbb{C}^n \cong \mathbb{R}^{2n}$ there is a $\sigma > 1$, a neighbourhood N of $(0, z_0) \in \Phi(V) \times \mathbb{C}^n$, and a real analytic mapping*

$$\mathscr{G} : N \to B_{n,\sigma} \times \mathbb{R}$$

such that

1. $\mathscr{G}(0, z_0)=(f_0, \xi_0)$, *and*
2. *if* $(\varphi, z)\in N$ *and* $\mathscr{G}(\varphi, z)=(f, \xi)$ *then* f *is a* φ*-mapping and* $f(\xi)=z$.

Proof. Define a mapping

by
$$\mathscr{F}: \Phi(V)\times \mathbb{C}^n \times B^0_{n,\sigma}\times B_{n,\sigma}\times Q_\sigma \times \mathbb{R}\to Q_\sigma \times \mathbb{C}^n \times \mathbb{R}\times A_{n,\sigma}$$

$$\mathscr{F}(\varphi, z, f, g, q, \xi)$$
$$=(r\circ f, f(\xi)-z, \operatorname{Re} f'(0)g(0)-1/2, \zeta(\varphi_z\circ f+qr_z\circ f)-g).$$

A little explanation may be needed here. f and g are considered as defined only on ∂U (in accordance with the definition of $B_{n,\sigma}$ and $A_{n,\sigma}$); $f(\xi)$ is understood as the value of the holomorphic extension of f in $\xi\in U\cup \bar{U}_\sigma$. The meaning of $f'(0)$ and $g(0)$ is analoguous. In the last term above ζ denotes the identity function on ∂U, which multiplies the expression in the brackets.

Let $\tilde{f}_0$ and p_0 denote the functions associated with the 0-mapping f_0. It is an easy exercise to check that for $\sigma>1$ sufficiently near to 1, $\mathscr{F}$ is indeed defined in a neighbourhood of

$$(0, z_0, f_0, \tilde{f}_0, p_0, \xi_0)\in \Phi(V)\times \mathbb{C}^n \times B^0_{n,\sigma}\times B_{n,\sigma}\times Q_\sigma \times \mathbb{R}$$

and $\mathscr{F}$ is analytic there. On the other hand

$$\mathscr{F}(\varphi, z, f, g, q, \xi)=0 \tag{4.1}$$

means that f is a φ-mapping, $f(\xi)=z$, (and $g=\tilde{f}, q=p$), at least if $(\varphi, z, f, g, q, \xi)$ is near to $(0, z_0, f_0, \tilde{f}_0, p_0, \xi_0)$.

Thus all we have to do is to solve (4.1) with respect to f, ξ (and g, q), when φ and z are given. We shall prove that the partial derivative of $\mathscr{F}$ with respect to (f, g, q, ξ) at the point $(0, z_0, f_0, \tilde{f}_0, p_0, \xi)$ is invertible. Therefore by the implicit function theorem it will follow that (4.1) can indeed be solved if (φ, z) is near $(0, z_0)$; moreover, the solution depends analytically on φ, z. This will then prove the lemma.

So we will have to compute the partial derivate

$$L=\mathscr{F}_{(f,g,q,\xi)}(0, z_0, f_0, \tilde{f}_0, p_0, \xi_0)$$
$$: B^0_{n,\sigma}\times B_{n,\sigma}\times Q_\sigma \times \mathbb{R}\to Q_\sigma \times \mathbb{C}^n \times \mathbb{R}\times A_{n,\sigma}$$

and prove that for any $(R, Z, X, F)\in Q_\sigma \times \mathbb{C}^n \times \mathbb{R}\times A_{n,\sigma}$ the equation

$$L(\dot{f}, \dot{g}, \dot{q}, \dot{\xi})=(R, Z, X, F) \tag{4.2}$$

admits exactly one solution $(\dot{f}, \dot{g}, \dot{q}, \dot{\xi})$. The computation of L is straightforward and one finds that (4.2) means more explicitly

$$2\operatorname{Re}\{(r_z\circ f_0)\dot{f}\}=R, \tag{4.3}$$

$$\dot{f}(\xi_0)+f_0'(\xi_0)\dot{\xi}=Z, \tag{4.4}$$

$$\operatorname{Re}\{\dot{f}'(0)\tilde{f}_0(0)+f_0'(0)\dot{g}(0)=X, \tag{4.5}$$

$$\xi\{\dot{q}(r_z\circ f_0)+p_0(r_{zz}\circ f_0)\dot{f}+p_0(r_{z\bar{z}}\circ f_0)\bar{\dot{f}}\}-\dot{g}=F. \tag{4.6}$$

A word of explanation may be needed again. The products in (4.3) and (4.5) are scalar products of vectors in $\mathbb{C}^n$, as agreed upon in Chap. 2. On the other hand, in (4.6) $r_{zz}=\partial^2 r/\partial z_j \partial z_k$ and $r_{z\bar{z}}=\partial^2 r/\partial z_j \partial \bar{z}_k$ are considered as matrices (j being the row index and k the column index), r_z, $\dot{f}$, and $\bar{\dot{f}}$ are column vectors and $r_{zz}\dot{f}$ is usual matrix multiplication.

To solve the system (4.3)...(4.6) we shall need the following fact, proved in [L 2] (Proposition 11): It is possible to find a real analytic defining function of D – which we shall go on calling r –, and to perform a biholomorphic change of coordinates in a neighbourhood of D in such a way that in the new coordinates – which again we shall denote by the same letters $z=(z_1, z_2, \ldots, z_n)$ –

1. $f_0(\zeta)=(\zeta, 0, \ldots, 0)^t$;
2. $r_z(f_0(\zeta))=(\bar{\zeta}, 0, \ldots, 0)$ if $\zeta \in \partial U$;
3. D is strictly convex even in the new system z, at least near $f_0(\partial U)$.

In fact only strict *linear* convexity is claimed in Proposition 11. of [L 2] but this is all that we shall need later on; namely the cicumstance that for any $v=(v_2, \ldots, v_n)^t \in \mathbb{C}^{n-1}\setminus\{0\}$ and $\zeta \in \partial U$

$$\left| \sum_{j,k=2}^{n} r_{z_j z_k}(f_0(\zeta)) v_j v_k \right| < \sum_{j,k=2}^{n} r_{\bar{z}_j z_k}(f_0(\zeta)) \bar{v}_j v_k \tag{4.7}$$

holds. This is also proved in [L 1] in the course of the proof of Proposition 9. – Note that from 2. follows $p_0=1/2$ and $\tilde{f}(\zeta)=(1/2, 0, \ldots, 0)^t$.

Now, in this adapted system of coordinates (4.3)...(4.6) are much simplified. Let $\dot{f}_1$ denote the first and $\dot{f}_+$ the last $n-1$ coordinates of $\dot{f}$; similarly let $\dot{g}^t=(\dot{g}_1, \dot{g}^t_+)$. Then (4.3) becomes $\mathrm{Re}\,\bar{\zeta}\dot{f}_1=R/2$ or

$$\mathrm{Re}(\dot{f}_1/\zeta)=R/2. \tag{4.8}$$

Since $\dot{f}_1/\zeta$ is holomorphic even at 0, (4.8) determines it up to an additive imaginary constant as an element of $B_{1,\sigma}$. This imaginary constant can be computed from (4.4), which then also yields the values of $\dot{\xi}$ and $\dot{f}_+(\xi_0)$.

So far $\dot{\xi}$, $\dot{f}_1$, and $\dot{f}_+(\xi_0)$ are determined and (4.3) and (4.4) are satisfied. Consider (4.6). In reality, this is a system of n equations. Fortunately, one of the unknowns, $\dot{q}$, occurs only in the first equation. Substituting the already known value of $\dot{f}_1$ into the last $n-1$

$$\zeta(r_{z_j z_k} \circ f_0)_{2 \leqq j,k \leqq n} \cdot \dot{f}_+ + \zeta(r_{z_j \bar{z}_k} \circ f_0)_{2 \leqq j,k \leqq n} \cdot \bar{\dot{f}}_+ - \dot{g}_+ = \psi \tag{4.9}$$

is obtained, where $\psi \in A_{n-1,\sigma}$ can be expressed in terms of F and $\dot{f}_1$. Put

$$\dot{f}_+(\zeta)/\zeta = h(\zeta) \in B_{n-1,\sigma},$$
$$(r_{z_j \bar{z}_k}(f_0(\zeta)))_{2 \leqq j,k \leqq n} = \alpha(\zeta),$$
$$\zeta^2 (r_{z_j z_k}(f_0(\zeta)))_{2 \leqq j,k \leqq n} = \beta(\zeta).$$

Then (4.9) expresses the fact that

$$\alpha \bar{h} + \beta h - \psi \text{ extends holomorphically from } \partial U \text{ to } \bar{U}. \tag{4.10}$$

Moreover, $h(\xi_0)$ is known ($=\dot{f}_+(\xi_0)/\xi_0$). Because of (4.7), condition (3.6) of Lemma 3. is satisfied, so that by this lemma (4.10) has a unique solution $h \in B_{n-1,\sigma}$ assuming the right value in ξ_0 – provided $\sigma > 1$ is small enough.

With h also $\dot{f}_+ \in B^0_{n-1,\sigma}$ and $\dot{g}_+ \in B_{n-1,\sigma}$ are determined. Finally it is easily checked that there is a unique way to choose $\dot{q}$ and $\dot{g}_1$ so that (4.5) and the first of the n equations in (4.6) be satisfied.

Thus the invertibility of L is proved, q.e.d.

In conclusion, we see that every $z_0 \in \bar{D} \setminus \{0\}$ has a neighbourhood G_{z_0} such that if $\varphi \in \Phi(V)$ is sufficiently small and $z \in G_{z_0}$ then there is a φ-mapping f such that $f(U \cup U_\sigma) \ni z$. We shall now show that the same holds for any z_0 in the blown up $\overline{D^*}$ of $\bar{D}$.

To this end, first note that every φ-mapping $f: \bar{U} \to \bar{D}$ can be lifted to a holomorphic mapping $f^\mu : \bar{U} \to \overline{D^*}$. Indeed, f takes only $0 \in U$ to $0 \in D$ and $f'(0) \neq 0$. Thus $f^\mu(0)$ can be defined as $(f_1'(0) : f_2'(0) : \ldots : f_n'(0)) \in \mathbb{P}_{n-1}$. From Theorème 2 of [L1] or from Theorem 3 of [L2] it follows that given a $z_0 \in \overline{D^*}$ there is a 0-mapping $f: \bar{U} \to \bar{D}$ such that $f^\mu(\bar{U}) \ni z_0$. Moreover, f is unique up to composition by a rotation of U.

Lemma 6. *Let $z_0 \in \overline{D^*}$, f_0 be a 0-mapping such that $f_0^\mu(\xi_0) = z_0$, where $\xi_0 \in \bar{U}$, Then for any neighbourhood V of ∂D in the complexification of $\mathbb{C}^n \cong \mathbb{R}^{2n}$ there is a $\sigma > 1$, a neighbourhood N of $(0, z_0) \in \Phi(V) \times (\mathbb{C}^n)^*((\mathbb{C}^n)^*$ denoting the blow up of $\mathbb{C}^n$ at 0), and a real analytic mapping $\mathscr{H} : N \to B_{n,\sigma} \times \mathbb{C}$ such that*

1. *$\mathscr{H}(0, z_0) = (f_0, \xi_0)$, and*
2. *If $(\varphi, z) \in N$ and $\mathscr{H}(\varphi, z) = (f, \xi)$ then f is a φ-mapping and $f^\mu(\xi) = z$.*

Observe the minor change in the use of the letter ξ: Previously ξ always denoted a positive number, here it can be complex.

Proof. If $z_0 \in \bar{D}^*$ does not lie over 0 (i.e. $\mu(z_0) \neq 0$) then the claim is the same as that of Lemma 5. Assume therefore that $\mu(z_0) = 0$. In this case $\xi_0 = 0$ and $f_0^\mu(0) = z_0$. Let, for instance, z_0 be represented by the tangent vector $v = (1, 0, \ldots, 0) \in T_0(D) \cong \mathbb{C}^n$.

As explained in the course of the proof of Lemma 5, it can be assumed that, firstly, $f_0(\zeta) = (\zeta, 0, \ldots, 0)^t$, and secondly, $r_z(f_0(\zeta)) = (\bar{\zeta}, 0, \ldots, 0)$ for $\zeta \in \partial U$.

Now, analytically and in a neighbourhood of z_0, blowing up means the introduction of new coordinates in D:

$$y_1 = z_1, \; y_2 = z_2/z_1 \ldots, \; y_n = z_n/z_1 .$$

The coordinates of z_0 are $y_1 = y_2 = \ldots = y_n = 0$. Thus a $z \in D^*$ is near z_0 if its y coordinates are near zero.

Define again a mapping

$$\mathscr{J} : \Phi(V) \times \mathbb{C}^n \times B^0_{n,\sigma} \times B_{n,\sigma} \times Q_\sigma \times \mathbb{C} \to Q_\sigma \times \mathbb{R} \times \mathbb{C} \times \mathbb{C}^{n-1} \times \mathbb{R} \times A_{n,\sigma}$$

by

$$\mathscr{J}(\varphi, y, f, g, q, \xi) = (r \circ f, \operatorname{Im} f_1'(0), f_1(\xi) - y_1,$$

$$\frac{f_+(\xi)}{f_1(\xi)} - y_+, \operatorname{Re} f'(0) g(0) - 1/2, \zeta\{\varphi_z \circ f + q(r_z \circ f)\} - g).$$

Here f_1 resp. y_1 stand for the first and f_+ resp. y_+ stand for the last $n-1$ coordinates of f resp. y.

If

$$\mathscr{J}(\varphi, y, f, g, q, \xi) = 0 \tag{4.11}$$

then f is a φ-mapping, the coordinates of $z=f^{\mu}(\xi)\in D^*$ are $y_1,\dots,y_n$ (and $\tilde{f}=g, p=q$). Thus to prove the lemma it will be sufficient to show that from (4.11) f and ξ can be expressed as analytic functions of φ, y, at least when $(\varphi, y)\in\Phi(V)\times\mathbb{C}^n$ is near $(0,0)$.

To this end one computes the partial derivative

$$\mathscr{J}_{(f,g,q,\xi)}(0,0,f_0,\tilde{f}_0,p_0,0)$$

and finds that it is invertible, exactly as in the case of $\mathscr{F}$ of the previous lemma. Hence the claim of Lemma 6 follows again from the implicit function theorem.

Thus for every $z_0\in\bar{D}^*$ there is a $\sigma>1$ and a neighbourhood G_{z_0} of z_0 such that if $\varphi\in\Phi(V)$ is sufficiently small, and $z\in G_{z_0}$ then there is a φ-mapping f such that $z\in f^{\mu}(U\cup\bar{U}_{\sigma})$. By compactness follows therefore

Lemma 7. *For every V there is a neighbourhood X of $\bar{D}^*$ in $(\mathbb{C}^n)^*$ and a $\sigma>1$ such that if $\varphi\in\Phi(V)$ is sufficiently small then for every $z\in X$ there is a φ-mapping f such that $z=f^{\mu}(\xi)$ with some $\xi\in U\cup\bar{U}_{\sigma}$. Moreover, locally it is possible to choose f and ξ to depend analytically on z (and φ).*

If $f(\zeta)$ is a φ-mapping then $f(e^{i\theta}\zeta)$ is also one so that z does not determine the corresponding φ-mapping. For a large φ it is also possible that there exist two essentially different φ-mappings f and g such that $f(U)$ and $g(U)$ have common points (other than 0). This may be true even for small φ – we do not know. Nevertheless, in the next chapter we shall show that the φ-mappings constructed in this chapter do have the property that if f, g are two such mappings and $f^{\mu}(U\cup\bar{U}_{\sigma})\cap g^{\mu}(U\cup\bar{U}_{\sigma})\neq\emptyset$ then $f(\zeta)=g(e^{i\theta}\zeta)$. Thus the images of $\bar{U}$ under these φ-mappings yield a foliation of $\bar{D}^*$.

In the following the φ-mappings constructed in this chapter will be called special φ-mappings. Note that if $f(\zeta)$ is a special φ-mapping then so is $f(e^{i\theta}\zeta)$. Note also that if $z\in\bar{D}^*$ is any point and f_0 is a 0-mapping, f a special φ-mapping such that $f_0^{\mu}(\xi)=f^{\mu}(\xi')=z$, then with a suitable θ the φ-mapping $g(\zeta)=f(e^{i\theta}\zeta)$ will be near to $f_0(\zeta)$, $\tilde{g}$ will be near to $\tilde{f}_0$ (both in the space $B_{n,\sigma}$), and $e^{-i\theta}\xi'$ will be near to ξ – provided φ is small. This follows from the implicit function theorem and the uniqueness of 0-mappings. It is this property that distinguishes special φ-mappings among plain φ-mappings.

5. Uniqueness of Special φ-Mappings

It turns out that the natural domain of definition of special φ-mappings is not the unit disc but a perturbation thereof. Let $f:\bar{U}\to\bar{D}$ be a special φ-mapping. Let j be a nowhere vanishing holomorphic function on $\bar{U}$ such that $j(0)$ is real and positive and

$$\varphi(f(\zeta))=\log|j(\zeta)| \qquad (\zeta\in\partial U). \tag{5.1}$$

Then $k(\zeta)=\zeta j(\zeta)$ will be univalent in a neighbourhood of $\bar{U}$ (if φ is small). Put $\Omega=k(U)$ and

$$g(\omega)=f(k^{-1}(\omega)).$$

Thus g is a holomorphic imbedding $\bar{\Omega}\to\bar{D}$. We shall call the function g thus obtained a natural φ-mapping. Note that the range of a natural φ-mapping is the same as that of the initial special φ-mapping. Furthermore $\varphi(g(\omega))=\log|\omega|$ if $\omega\in\partial\Omega$, which explains the term "natural".

In the above situation put also $\tilde{g}(\omega)=j(\zeta)\tilde{f}(\zeta)$ where $\omega=k(\zeta)\,(=\zeta j(\zeta))$. Then

$$\tilde{g}(\omega)=\zeta j(\zeta)\{\varphi_z(f(\zeta))+p(\zeta)r_z(f(\zeta))\}=\omega\{\varphi_z(g(\omega))+p(\zeta)r_z(g(\omega))\}\,. \tag{5.2}$$

Lemma 8. $\tilde{g}(\omega)g'(\omega)=1/2$ $(\omega\in\Omega)$.

Proof. In view of (5.1)

$$\varphi(f(\zeta))=\log|k(\zeta)|+a(\zeta)r(f(\zeta))$$

for $\zeta\in\bar{U}$, with a real valued smooth function $a(\zeta)$. Taking $\partial/\partial z$ of both sides

$$f'(\zeta)\varphi_z(f(\zeta))=\frac{1}{2}\frac{k'(\zeta)}{k(\zeta)}+a(\zeta)f'(\zeta)r_z(f(\zeta))\quad(\zeta\in\partial U)\,. \tag{5.3}$$

Now for a fixed $\zeta\in\partial U$ the curve $\{f(e^{it}\zeta):t\in\mathbb{R}\}$ is in ∂D, so that its tangent in $f(\zeta)$ is orthogonal to $r_{\bar{z}}(f(\zeta))$:

$$\mathrm{Re}\,i\zeta f'(\zeta)r_z(f(\zeta))=0\,.$$

Therefore

$$\begin{aligned}\mathrm{Im}\,f'(\zeta)\tilde{f}(\zeta)&=\mathrm{Im}\,f'(\zeta)\zeta\{\varphi_z(f(\zeta))+p(\zeta)r_z(f(\zeta))\}\\&=\mathrm{Im}\,\zeta f'(\zeta)\varphi_z(f(\zeta))\\&=\mathrm{Im}\,\zeta k'(\zeta)/(2k(\zeta))\,.\end{aligned}$$

Hence $f'(\zeta)\tilde{f}(\zeta)=\zeta k'(\zeta)/(2k(\zeta))+$ a real constant, which must be zero since the real part of both $f'(0)f(0)$ and $\zeta k'(\zeta)/(2k(\zeta))|_{\zeta=0}$ is $1/2$. Thus, putting $\omega=k(\zeta)$

$$\tilde{g}(\omega)g'(\omega)=j(\zeta)\tilde{f}(\zeta)f'(\zeta)\frac{d\zeta}{d\omega}=\frac{j(\zeta)\zeta k'(\zeta)}{2k(\zeta)k'(\zeta)}=1/2\,,$$

q.e.d.

Lemma 9. *Let f be a special φ-mapping, where φ is small. Then there is a holomorphic function $F:\bar{D}\to\mathbb{C}$ vanishing at 0 such that*

1. *$F(f(\zeta))$ has just one zero in $\bar{U}$, namely in $\zeta=0$, and this zero is simple;*
2. *$\log|F(z)|=\varphi(z)$ if $z\in f(\partial U)$;*
3. *$\log|F(z)|<\varphi(z)$ if $z\in\partial D\backslash f(\partial U)$.*

Proof. When $\varphi=0$, this is proved in [L2] (Proposition 1). The proof there consists of showing that for a given $z\in\bar{D}$ the equation

$$(z-f(\zeta))\tilde{f}(\zeta)=0$$

for the unknown $\zeta\in\bar{U}$ admits a unique solution $\zeta=F(z)$. This easily follows from the argument principle. F is clearly holomorphic in D and one checks that it can be continued analytically past ∂D.

Let now φ be just small. Then there is a 0-mapping f_0 near f such that also $\tilde{f}_0$ and $\tilde{f}$ are near to one another. Construct the natural φ-mapping $g:\bar{\Omega}\to\bar{D}$ corresponding to f. Then f_0 (resp. $\tilde{f}_0$) and g (resp. $\tilde{g}$) will also be near to one another. Therefore the functions

$$\mathscr{F}(z,\omega)=(z-f_0(\omega))\tilde{f}_0(\omega)$$

and

$$\mathscr{G}(z,\omega)=(z-g(\omega))\tilde{g}(\omega)$$

are near when z is in a small neighbourhood of $\bar{D}$ and ω is in a small neighbourhood of $\bar{U}$, (which can be assumed to contain $\bar{\Omega}$ and be such that g is holomorphic on it).

Since the equation $\mathscr{F}(z,\omega)=0$ can be solved for ω when z is near $\bar{D}$, the same must be true for $\mathscr{G}(z,\omega)=0$ (for z in a possibly smaller neighbourhood of $\bar{D}$). Let $\omega=F(z)$ be the solution. Then $F(g(\omega))=\omega$. We claim that F satisfies conditions 1, 2, and 3.

1. is trivial, and so is 2., since upon putting $\omega=k(\zeta)$ with k as above, we have

$$\varphi(f(\zeta))=\varphi(g(\omega))=\log|\omega|=\log|F(g(\omega))|=\log|F(f(\zeta))| \tag{5.4}$$

if $\zeta\in\partial U$.

As for 3., a little computation is needed. Take a point $z_0\in f(\partial U)$, and consider the intersection of ∂D with a 2-dimensional real plane spanned by the normal vector in z_0 and a complex tangent vector in z_0. (v is complex tangent if both v and iv are tangent to ∂D.) The intersection near z_0 will be a curve $z(t)$, which will be assumed to be parametrized by arc length and z_0 to correspond to $t=0$.

We shall compute the Taylor expansion of order 2 of both $\varphi(z(t))$ and $\log|F(z(t))|$ (about $t=0$) and compare the two. For the sake of brevity we shall put $\omega=\omega(t)=F(z(t))$, $\dot{z}=dz/dt$, $\dot{\omega}=d\omega/dt$ etc.

Step 1. In view of (5.4) the constant terms of the two expansions agree.

Step 2. From the definition of ω:

$$0=(z-g(\omega))\tilde{g}(\omega),$$

by differentiation follows, using also $g'(\omega)\tilde{g}(\omega)=1/2$

$$\begin{aligned}0&=(\dot{z}-g'(\omega)\dot{\omega})\tilde{g}(\omega)+(z-g(\omega))\tilde{g}'(\omega)\dot{\omega}\\&=\dot{z}\tilde{g}(\omega)-\dot{\omega}\{\tfrac{1}{2}-(z-g(\omega))\tilde{g}'(\omega)\}.\end{aligned} \tag{5.5}$$

When $t=0$, i.e. $z=z_0=g(\omega)$, we have

$$\dot{\omega}=2\dot{z}\tilde{g}(\omega) \tag{5.6}$$

Thus (again for $t=0$)

$$\frac{d\log|\omega|}{dt}=\operatorname{Re}\frac{\dot{\omega}}{\omega}=2\operatorname{Re}\frac{\dot{z}\tilde{g}(\omega)}{\omega}=2\operatorname{Re}\dot{z}\varphi_z(z_0)=\frac{d\varphi(z)}{dt}.$$

Here we used (5.2) and the fact that $r_{\bar{z}}(z_0)$ is complex-orthogonal to $\dot{z}(0)$. We conclude that also the first order terms of the two expansions agree.

Step 3. Differentiating (5.5) once more and putting $t=0$ so that certain terms vanish we obtain

$$\ddot{\omega}=2\ddot{z}\tilde{g}(\omega)+2\dot{\omega}\{2\dot{z}\tilde{g}'(\omega)-g'(\omega)\tilde{g}'(\omega)\}. \tag{5.7}$$

Suppose now that φ goes to zero in some space $\Phi(V)$. Then g goes to a 0-mapping f_0 and $\tilde{g}$ goes to $\tilde{f}_0$. Since $\tilde{f}_0(\zeta)$ is a normal vector to ∂D in $f_0(\zeta)$, if $\zeta\in\partial U$, this means in view of (5.6) that $\dot{\omega}(0)$ goes to zero. Thus from (5.6) and (5.7) follows that

$$\dot{\omega}(0)=o(1), \qquad \ddot{\omega}(0)=2\ddot{z}\tilde{g}(\omega(0))+o(1),$$

where $o(1)$ means a term that goes to 0 as $\varphi\to 0$. Hence for $t=0$

$$\frac{d^2\log|\omega|}{dt^2}=\mathrm{Re}\left(\frac{\ddot{\omega}}{\omega}-\frac{\dot{\omega}^2}{\omega^2}\right)=2\,\mathrm{Re}\,\ddot{z}\tilde{g}(\omega)/\omega+o(1)$$

$$=2\,\mathrm{Re}\,\ddot{z}p_0(\zeta)r_z(z_0)+o(1),$$

where p_0 is the limit of p as $\varphi\to 0$.

Now the principal term in this last expression is a negative constant times the normal curvature of ∂D in the direction of the vector $\dot{z}(0)$. D being strictly convex we have therefore

$$\left.\frac{d^2\log|F(z(t))|}{dt^2}\right|_{t=0}<-\text{const}<0$$

when φ is small.

On the other hand, $d^2\varphi(z(t))/dt^2$ goes to zero as $\varphi\to 0$. Thus for small φ the second order term in the expansion of $\log|F(z(t))|$ is smaller than the corresponding term in the expansion of $\varphi(z(t))$, whence

$$\log|F(z(t))|<\varphi(z(t)) \tag{5.8}$$

if t is small.

Observe at this point that $f(\partial U)$ is transverse to complex tangential directions (since f' is nonzero e.g. because of Lemma 8., and an analytic disc cannot be tangent to a strictly convex surface). Therefore the curves $z(t)$ constructed above issuing from all possible points $z_0\in f(\partial U)$ in all possible complex tangential directions cover an entire neighbourhood of $f(\partial U)$ on ∂D. Hence it follows from (5.8) that

$$\log|F(z)|<\varphi(z) \tag{5.9}$$

when $z\in\partial D\backslash f(\partial U)$ is near to $f(\partial U)$. Here the meaning of "near" is independent of the small φ.

To settle the case of points bounded away from $f(\partial U)$, introduce the function $\omega=F_0(z)$ corresponding to the 0-mapping f_0. That is, $\mathscr{F}(z,F_0(z))=0$. When z is bounded away from $f_0(\partial U)$, $\log|F_0(z)|<0$ is bounded away from zero. Since F is near F_0, this implies (5.9) when z is outside a fixed neighbourhood of $f(\partial U)$, provided φ is small. Thus (5.9) is proved for all $z\in\partial D\backslash f(\partial U)$ if φ is small, q.e.d.

Observe that we have proved more than (5.9), namely

$$\log|F(z)| \leqq \varphi(z) - \eta \operatorname{dist}^2(z, f(\partial U)) \tag{5.10}$$

if $z \in \partial D$, where $\eta > 0$ is independent of the small φ.

Now we can prove the uniqueness of special φ-mappings:

Lemma 10. *Suppose φ is small. Let f and f_1 be two special φ-mappings. If $f^\mu(\bar{U}) \cap f_1^\mu(\bar{U}) \neq \emptyset$ then $f_1(\zeta) = f(e^{i\theta}\zeta)$ with some real θ.*

Proof. Suppose that f and f_1 are essentially different and still $f^\mu(\xi) = f_1^\mu(\xi_1) = z \in \bar{D}^*$, where $\xi, \xi_1 \in \bar{U}$. Assume first that $0 < |\xi|, |\xi_1| < 1$.

Let F denote the function constructed in Lemma 9, starting with the special φ-mapping f, and let F_1 be the corresponding function when starting with f_1. Then if $\zeta \in \partial U$

$$\log|F_1(f(\zeta))| \leqq \varphi(f(\zeta)) = \log|F(f(\zeta))|.$$

Since $f(\partial U)$ and $f_1(\partial U)$ have only a finite number of common points, we have here strict inequality with the eventual exception of finitely many $\zeta \in \partial U$. In particular

$$\log|F_1(f(\zeta))| - \log|F(f(\zeta))| < 0 \quad \text{a.e. on } \partial U.$$

The left hand side here is a subharmonic function of $\zeta \in \bar{U}$, since $\log|F(f(\zeta))|$ is harmonic but for the origin, where it is $\log|\zeta| + 0(1)$, and $\log|F_1(f(\zeta))| \leqq \log|\zeta| + 0(1)$. Hence by the maximum principle

$$0 > \log|F_1(f(\xi))| - \log|F(f(\xi))| = \log|F_1(z)| - \log|F(z)|$$

The rôle of F and F_1 being symmetric, $0 > \log|F(z)| - \log|F_1(z)|$ also holds, which is a contradiction. This settles the case when $0 < |\xi|, |\xi_1| < 1$.

If ξ and ξ_1 are 0 or on ∂U, take a $z' \in f_1^\mu(U \setminus \{0\})$ near z. Then by Lemma 7 it is possible to find a special φ-mapping f_2 near f, such that $f_2(U) \ni z'$. Then f_1 and f_2 are essentially different since f_2 is near to f but f_1 is at a fixed distance from f; nevertheless $f_2(\xi_2) = f_1(\xi') = z'$, where $0 < |\xi_2|, |\xi'| < 1$. This contradicts to what has already been proved.

Lemma 11. *There is a $\sigma > 1$ such that if φ is small and f, f_1 are two special φ-mappings such that $f^\mu(U \cup \bar{U}_\sigma) \cap f_1^\mu(U \cup \bar{U}_\sigma) \neq \emptyset$ then $f_1(\zeta) = f(e^{i\theta}\zeta)$ with some real θ.*

Proof. Let B_n denote the unit ball in $\mathbb{C}^n$, and construct a mapping λ from a neighbourhood of $\bar{B}_n \setminus \{0\}$ into a neighbourhood of $\bar{D} \setminus \{0\}$ in the following way. If w is near $\bar{B}_n$, let g be the special φ-mapping $\bar{U} \to \bar{D}$ such that $g'(0)$ is a positive constant times w. The existence and uniqueness of this g follows from Lemmas 7 and 10. Put $\lambda(w) = g(|w|)$. Then λ is real analytic in a neighbourhood of $\bar{B}_n \setminus \{0\}$ (by Lemma 7) and $\lambda(\bar{B}_n \setminus \{0\}) = \bar{D} \setminus \{0\}$.

Moreover λ restricted to $\bar{B}_n \setminus \{0\}$ is a diffeomorphism. This again is a consequence of Lemmas 7 and 10. Therefore λ is diffeomorphic on a neighbourhood of $\bar{B}_n \setminus \{0\}$, whence the statement of the lemma follows.

6. Proof of the Theorem

As explained in Chapter 2, it is sufficient to prove our theorem when $c=1$ and φ is small (as an element of the Banach space $\Phi(V)$).

In that case there is a $\sigma>1$ and a neighbourhood Y of $\bar{D}^*$ in $(\mathbb{C}^n)$ with the property that for every $z\in Y$ there is an essentially unique special φ-mapping f (i.e. unique up to a rotation of ζ) such that $f^\mu(\xi)=z$ with some $\xi\in U\cup\bar{U}_\sigma$. This follows from Lemmas 7 and 11. Let $g:\bar{\Omega}\to\bar{D}$ be the corresponding natural φ-mapping, and let $g(\omega)=z$. (When z lies outside $\bar{D}$ but not far from it, ω still can be defined since g extends analytically past $\partial\Omega$.) Put $u(z)=\log|\omega|$. In other words

$$u(z)=\log|F(z)| \tag{6.1}$$

where F is the function constructed in Lemma 9. This definition makes sense because if $f(\zeta)$ is replaced by $f(e^{i\theta}\zeta)$ then $g(\omega)$ will be replaced by $g(e^{i\theta}\omega)$.

Now $\exp(2u(z))=|\omega|^2$ is real analytic on $\bar{D}^*$ since the special φ-mapping f and ξ can be chosen to depend analytically on z, so that then g and ω depend analytically on z. Accepting that this u is the solution of (1.4), Complement 4 is proved.

We claim that u is indeed the solution of (1.4). First note that u is harmonic along special φ-discs $f(\bar{U})$ except at 0, where $u(z)=\log|z|+0(1)$. (This proves Complement 2.) Hence $\det|u_{z_j\bar{z}_k}|=0$ on $\bar{D}\setminus\{0\}$ follows if we can show that u is *psh*.

Now we prove that u is plurisubharmonic. Let $G:\bar{D}\to\mathbb{C}$ be a holomorphic function vanishing in 0 such that $\log|G(z)|\leqq\varphi(z)$ if $z\in\partial D$. Then

$$\log|G(z)|\leqq u(z) \qquad (z\in\bar{D}) \tag{6.2}$$

Indeed, let f be a special φ-mapping such that $f(\bar{U})\ni z$. Then $\log|G(w)|-u(w)$ is subharmonic along $f(\bar{U})$, is nonpositive on $f(\partial U)$, so that it is $\leqq 0$ even for $w=z$.

It follows now from (6.1) and (6.2) that

$$u(z)=\sup_G \log|G(z)| \qquad (z\in\bar{D}),$$

where the supremum is taken over all holomorphic functions $G:\bar{D}\to\mathbb{C}$ that vanish in 0 and are such that $\log|G(z)|\leqq\varphi(z)$ if $z\in\partial D$. This is the claim of Complement 3. It also implies that u is plurisubharmonic.

Finally it is easily checked that u satisfies the last two conditions of (1.4). The details are left to the reader.

It remains to prove the uniqueness of the solution of (1.4). To this end we may assume that φ is negative. Let v be another solution of (1.4). Then for any $\varepsilon>0$

$$(1+\varepsilon)v(z)\leqq u(z)\leqq(1-\varepsilon)v(z) \tag{6.3}$$

if $z\in\partial D$ or z is on a sufficiently small sphere S centred at 0. By the minimum principle of [BT] it follows that (6.3) also holds for z in the domain bounded by S and ∂D. Letting $\varepsilon\to 0$ we conclude $u\equiv v$, q.e.d.

7. Proof of Complement 1

By now all complements have been proved but the first and the last. However, this latter is a consequence of Lemma 7. Here we shall prove Complement 1.

Let $z_0 \in \bar{D} \setminus \{0\}$, f a special φ-mapping such that $f(\bar{U}) \ni z_0$. It will be sufficient to show that for any vector x of type $(1,0)$, not tangent to $f(\bar{U})$ in z_0,

$$\langle \partial\bar{\partial} u(z_0), x \otimes \bar{x} \rangle > 0 \tag{7.1}$$

holds. To prove this we shall compare u with $\log|F|$ in a neighbourhood of z_0. (Here F is the function constructed in Lemma 9.)

Let $z(t)$ be a curve in $\bar{D}$, not tangent to $f(\bar{U})$, such that $z_0 = z(0)$, and let f_t be special φ-mappings such that $f_t(\xi_t) = z(t)$, $f_0 = f$. It can be assumed that f_t and ξ_t depend smoothly on t. Since $\operatorname{dist}(z(t), f(\bar{U})) > \text{const} \cdot t$, it follows that

$$\operatorname{dist}(f_t(\partial U), f(\partial U)) > \text{const} \cdot t$$

so that by (5.10)

$$\log|F(f_t(\zeta))| \leqq \varphi(f_t(\zeta)) - \text{const} \cdot t^2$$

if $\zeta \in \partial D$. Thus

$$\log|F(f_t(\zeta))| + \text{const} \cdot t^2 - u(f_t(\zeta))$$

is a subharmonic function of $\zeta \in \bar{U}$, which is $\leqq 0$ for $\zeta \in \partial U$. Hence $\log|F(f_t(\xi_t))| + \text{const} \cdot t^2 \leqq u(f_t(\xi_t))$, i.e.

$$\log|F(z(t))| + \text{const} \cdot t^2 \leqq u(z(t)).$$

For $t = 0$ in this formula equality holds, whence

$$\left.\frac{d^2}{dt^2}\right|_{t=0} \log|F(z(t))| < \left.\frac{d^2}{dt^2}\right|_{t=0} u(z(t)).$$

This is true for all curves $z(t)$ not tangent to $f(\bar{U})$, so that for any vector x of type $(1,0)$ not tangent to $f(\bar{U})$

$$\langle \partial\bar{\partial} u(z_0), x \otimes \bar{x} \rangle > \langle \partial\bar{\partial} \log|F(z_0)|, x \otimes \bar{x} \rangle = 0,$$

q.e.d.

References

[B] Bedford, E.: Stability of envelopes of holomorphy and the degenerate Monge-Ampère equation. Math. Ann. **259**, 1–28 (1982)

[BK] Bedford, E., Kalka, M.: Foliations and complex Monge-Ampère equations. Comm. Pure Appl. Math. **30**, 543–571 (1977)

[BT] Bedford, E., Taylor, B.A.: The Dirichlet problem for a complex Monge-Ampère equation. Invent. Math. **37**, 1–44 (1976)

[L1] Lempert, L.: La métrique de Kobayashi et la représentation des domaines sur la boule. Bull. Soc. Math. France **109**, 427–474 (1981)

[L2] Lempert, L.: Intrinsic distances and holomorphic retracts, Proc. Conf. Varna, 1981, to appear

[M] Moriyón, R.: Regularity of the Dirichlet problem for the complex Monge-Ampère equation $\det u_{j\bar{k}} = 0$, Proc. Nat. Acad. Sci. USA, **76**, 1022–1023 (1979)

Received August 20, 1982; in revised form March 23, 1983