SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Mathematische Annalen
Verlag: Springer
Jahr: 1981
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN235181684_0256
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN235181684_0256

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 256 1981

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), *Walther v. Dyck* (1888 – 1921), *David Hilbert* (1902 bis 1939), *Otto Blumenthal* (1906 – 1938), *Albert Einstein* (1920 – 1928), *Constantin Carathéodory* (1925 – 1928), *Erich Hecke* (1929 – 1947), *Bartel, L. van der Waerden* (1934 bis 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 Gårding, Lund* (1970 – 1978), *Konrad Jörgens* (1972 – 1974), *F. John* (1968 – 1979).
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 · **Peter Dombrowski,** Köln
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.
The Article Fee Code on the first page of an article in this journal indicates the copyright owner's consent that in the USA copies may be made for personal or internal use, provided the stated fee for copyright beyond that permitted Section 107 or 108 of the United States Copyright Law is paid through the Copyright Clearance Center Inc., 21 Congress Street, Salem, MA 01970, USA. If a code does not appear copies of the article may be made without charge, provided permission is obtained from the publisher. 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 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 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

Inhalt/Contents

Abeasis, S., Fra, A. Del, Kraft, H.: The Geometry of Representations of A_m 401
Alexander, J. C.: On the De Rham Invariant of a Fibered $(4r + 1)$-Dimensional Orientable Manifold 429
Anderson, D. F., Ohm, J.: Valuations and Semi-Valuations of Graded Domains 145
Arabello, E., Cornalba, M.: Footnotes to a Paper of Beniamino Segre. The Number of g^1_d's on a General d-Gonal Curve, and the Unirationality of the Hurwitz Spaces of 4-Gonal and 5-Gonal Curves 341
Bartels, H.-J.: Zur Arithmetik von Diedergruppenerweiterungen 465
Borell, C.: A Gaussian Correlation Inequality for Certain Bodies in $\mathbb{R}^n$ 569
Bürgisser, B.: Finite p-Periodic Quotients of General Linear Groups 121
Campana, F.: Réduction algébrique d'un morphisme faiblement Kählérien propre et applications 157
Cornalba, M., see Arbarello, E. 341
Duhoux, M., see Ng, K.-F. 281
Eisenbud, D., Ven A., Van de: On the Normal Bundles of Smooth Rational Spaces Curves 453
Fra, A. Del, see Abeasis, S., et al. 401
Galovich, S., Rosen, M.: Distributions on Rational Function Fields 549
Groh, U., Neubrander, F.: Stabilität starkstetiger, positiver Operatorhalbgruppen auf C^*-Algebren 509
Hansen, F.: Selfadjoint Means and Operator Monotone Functions 29
Hauenschild, W.: Subhypergroups and Normal Subgroups 1
Hermann, C. F.: Symmetrische Hilbertsche Modulformen und Modulfunktionen zu $\mathbb{Q}(\sqrt{17})$ 191
Hofmann, K. H., Mukherjea, A.: Concentration Functions and a Class of Non-Compact Groups 535
Jacquet, H., Piatetski-Shapiro, I. I., Shalika, J.: Conducteur des représentations du groupe linéaire 199
Jänich, K.: Universal Families of C^∞-Functions on D^2 67
Jakobsen, H. P.: The Last Possible Place of Unitarity for Certain Highest Weight Modules 439
Jennane, B.: Problème de Levi et morphisme localement de Stein 37
Joshi, V. G., Saxena, R. K.: Abelian Theorems for Distributional H-Transform 311
Kanenobu, T.: A Note on 2-fold Branched Covering Spaces of S^3 449
Kantorovitz, Sh.: Spectrality Criteria for Unbounded Operators with Real Spectrum 19
König, H., Tzafriri, L.: Some Estimates for Type and Cotype Constants 85
Kosarew, S.: Das formale Prinzip und Modifikationen komplexer Räume 249
Kraft, H., see Abeasis, S., et al. 401
Kudla, S. S.: Holomorphic Siegel Modular Forms Associated to $SO(n, 1)$ 517
Lang, W. E.: Classical Godeaux Surface in Characteristic P 419
Melvin, P.: On 4-Manifolds with Singular Torus Actions 255
Meuser, D.: On the Rationality of Certain Generating Functions 303
Meyer, W., Neutsch, W.: Fermatquadrupel 51
Münzner, H. F.: Isoparametrische Hyperflächen in Sphären. II. Über die Zerlegung der Sphäre in Ballbündel 215
Mukherjea, A., see Hofmann, K. H. 535
Neubrander, F., see Groh, U. 509
Neutsch, W., see Meyer, W. 51
Ng, K.-F., Duhoux, M.: Directedness in Ordered Normed Spaces and Operators 281
Nielsen, O. A.: The Failure of the Topological Frobenius Property for Nilpotent Lie Groups. II . 561
Oda, T.: On the Poles of Andranov L-Functions 323
Ohm, J., see Anderson, D. F. 145
Piatetski-Shapiro, I. I., see Jacquet, H., et al. 199
Piene, R.: Cuspidal Projections of Space Curves 95
Pommerenke, Ch.: On Ergodic Properties of Inner Functions 43
Rosen, M., see Galovich, S. 549
Saxena, R. K., see Joshi, V. G. 311

Schmüdgen, K.: Perturbations of Selfadjoint Operators with Point Spectra by Restrictions and Selfadjoint Extensions 233
Schneider, R.: Zur optimalen Approximation konvexer Hyperflächen durch Polyeder 289
Shah, J.: Projective Degenerations of Enriques' Surfaces 475
Shalika, J., see Jacquet, H., et al. 199
Słodkowski, Z.: Analytic Set-Valued Functions and Spectra 363
Smith, M. A.: A Banach space that is MLUR but not HR 277
Spindler, H.: Die Modulräume stabiler 3-Bündel auf $\mathbb{P}_3$ mit den Chernklassen $c_1 = 0$, $c_3 = c_2^2 - c_2$. 133
Stacey, P. J.: Local and Global Splittings in the State Space of a JB-Algebra 497
Talagrand, M.: Non existence de relèvement pour certaines mesures finiement additives et retractés de $\beta\mathbb{N}$ 63
Teicher, M.: Factorization of a Birational Morphism Between 4-Folds 391
Tsuji, H.: A Generalization of Schwarz Lemma 387
Tzafriri, L., see König, H. 85
Ven, A., Van de, see Eisenbud, D. 453

Covered by Zentralblatt für Mathematik and Current Mathematical Publications
Springer-Verlag Berlin Heidelberg New York
Printer: Brühlsche Universitätsdruckerei, Giessen
Printed in Germany – © Springer-Verlag GmbH & Co KG Heidelberg 1981

Z-1982.5039

Mathematische Annalen

Band 256 Heft 1 1981

1-4 T.J.

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 · **Peter Dombrowski,** Köln
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

8 Z Nat 593

Niedersächsische Staats- u. Universitätsbibliothek Göttingen

ISSN 0025 - 5831 Math. Ann. — MAANA3 256 (1) 1—144 (1981) Juni 1981

X, 83

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. The Article Fee Code on the first page of an article in this journal indicates the copyright owner's consent that in the USA copies may be made for personal or internal use, provided the stated fee for copyright beyond that permitted Section 107 or 108 of the United States copyright Law is paid throug the Copyright Clearance Center, Inc., 21 Congress Street, Salem, MA 01970, USA. If a code does not appear, copies of an article may by made without charge, provided permission is obtained from the publisher. 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.

Herausgeber/Editors

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.

H. Amann — Mathematisches Institut der Universität, Freiestrasse 36, CH-8032 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

P. Dombrowski — Mathematisches Institut der Universität, Lindenthal 86—90, D-5000 Köln, Federal Republic of Germany

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 — Mathematisches Institut der Universität, Wegelerstrasse 10, D-5300 Bonn, 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

Subscription Information. Vols. 255—258 (3—4 issues each) will appear in 1981. The price of each volume is DM 386.00 or US $ 227.00. The publisher reserves the right to issue additional volumes during the calendar year. Prices for back-volumes are available on request. In addition, we also offer microform editions in 16mm and 35mm microform and microfiche. Correspondence concerning subscriptions should be addressed to the publisher.

Changes of Address. Six weeks should be allowed for all changes to become effective. All communications should include both old an new addresses (with postal codes) and should be accompanied by an address label from a recent issue.

North America. Annual subscription rate: US $ 937.00, including postage and handling. Subscriptions are entered with prepayment only. Orders should be addressed to: Springer-Verlag New York Inc., Service Center Secaucus, 44, Hartz Way, Secaucus, NJ 07094, USA.

All Other Countries. Annual subscription rate: DM 1544.00 plus postage and handling. Airmail delivery on request only. For Japan, postage (surface airmail lifted) and handling is DM 148.00. Orders can either be placed with your bookdealer or sent directly to: Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33.

Springer-Verlag, Journal Department 1, Postf. 105280, D-6900 Heidelberg 1, Tel. (06221) 487-1, Telex 04-6190
Springer-Verlag, Heidelberger Platz 3, D-1000 Berlin 33, Tel. (030) 8207-1, Telex 01-83319
Springer-Verlag, New York Inc., 175 Fifth Avenue, New York, NY 10010, Tel. (212) 477-8200, Telex 232235

Math. Ann. 256, 1–18 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Subhypergroups and Normal Subgroups[1]

Wilfried Hauenschild

Fachbereich Mathematik-Informatik, Universität-Gesamthochschule Paderborn, Warburger Str. 100, D-4790 Paderborn, Federal Republic of Germany

Introduction

Helgason has established a one-to-one correspondence between the set of closed normal subgroups of a compact group G and the set of subhypergroups of $\hat{G}$, the set of equivalence classes of irreducible unitary representations of G [14]. On the other side Takesaki and Tatsuuma have obtained in [31, 32] an antiisomorphism between the lattice $\mathfrak{H}$ of the closed subgroups of an arbitrary locally compact group G and the lattice $\mathfrak{B}$ of the right-invariant, weakly-$*$-closed subalgebras of $L^\infty(G)$ in the following way:

For $B\in\mathfrak{B}$ set $H_B=\{x\in G;\ {}_x\varphi=\varphi \text{ for every } \varphi\in B\}$, where ${}_x\varphi$ denotes the left translation of φ by x; likewise for $H\in\mathfrak{H}$ set $B_H=\{\varphi\in L^\infty(G);\ {}_x\varphi=\varphi \text{ for every } x\in H\}$.

H_B turnes out to be normal for a left-invariant $B\in\mathfrak{B}$ and the other way round. This setting generalizes the well-known correspondence between closed subgroups of an abelian group and those of it's dual. We want to get for Moore-groups a generalization closer to the abelian case than that of [31, 32].

For a closed normal subgroup N of the locally compact group G let now $K(N, G)\subseteq L^\infty(N)$ be the set of positive definit, continuous, G-invariant functions of norm not greater than 1, which is a simplex in case G has small invariant neighbourhoods (SIN-group) [33, Satz 1 by means of 15, 1.2. Lemma]; let further $E(N, G)$ be the set of non-trivial extremal points of $K(N, G)$.

For a SIN-group G and a closed normal subgroup N we get

$$B_N\cap E(G)=E(G/N),$$

if we write $E(G)$ instead of $E(G, G)$ and identify $E(G/N)$ with $\{\alpha\in E(G);\ \alpha(x)=1 \text{ for every } x\in N\}$ in the canonical way [34]. So we can recover N from the much smaller set $E(G/N)$ instead of B_N [26, 3.4. Corollary].

For a Moore-group G taking normalized traces of irreducible (hence finite dimensional) representations establishes a bijection between $\hat{G}$ and $E(G)$ [12]. For

1 This is the condensed first part of the author's Habilitationsschrift (Paderborn 1979)

0025-5831/81/0256/0001/$03.60

a compact group the sets of the form $E(G/N)$ (resp. $G\hat{/}N$) are just the subhypergroups as Helgason has shown [14]. Our aim in this paper is to characterize these sets in the case of a Moore-group in terms of topological and algebraic properties.

To manage this we first generalize the concept of topological hypergroups introduced by Dunkl, Jewett, and Spector [6, 18, 30] to spaces which are not necessarily Hausdorff (we cling to Jewett's definition, whenever possible). The sets $E(N, G)$ resp. $\hat{G}$ are hypergroups in this sense, and $M(E(N, G))$, the set of bounded, complex-valued (not necessarily regular) Borel measures, becomes an involutive Banach algebra [compare 18].

If we define $N_R = \{x \in G;\ \pi(x) = \mathrm{id} \text{ for every } \pi \in R\}$ for a subset R of $\hat{G}$ and $S_M = \{\pi \in \hat{G};\ \pi(x) = \mathrm{id} \text{ for every } x \in M\}$ for a subset M of G, then we can state our main theorem:

For a Moore-group G we have the following:

(i) S_M is a subhypergroup of $\hat{G}$ for every subset M of G, and N_{S_M} is the smallest closed normal subgroup of G containing M.

(ii) N_R is a closed normal subgroup of G for every subset R of $\hat{G}$, and S_{N_R} is the smallest subhypergroup of $\hat{G}$ containing R.

In particular we see, that the sets $G\hat{/}N = S_N$ are just the subhypergroups of $\hat{G}$. As in [31, 32] we also obtain an antiisomorphism between the lattice of closed normal subgroups of G and that of subhypergroups of $\hat{G}$, such that compact objects on one side correspond to open objects on the other, just as it happens for abelian groups.

1. Preliminaries and Notations

1.1. For topological spaces X and Y, subsets A, B of X and mappings $f: X \to Y$ resp. $g: A \to \mathbb{C}$ we write $f|A$ for the restriction and $\tilde{g}$ for the trivial extension. We denote by $C(X)$ the set of continuous functions on X, $C_0(X)$ resp. $C_{00}(X)$ the sets of those, which vanish at infinity resp. have compact support. We write $cl(A)$ for the closure, A^c for the complement and $A \backslash B$ instead of $A \cap B^c$.

Further we call a map proper, if the inverse image of every closed, compact subset is again compact, and we speak of $f: X \to Y$ as a quotient map, if f is a continuous surjection, and if Y carries the quotient topology. For a normal subgroup N of a group G, a function f on N, $m \in N$ and $x \in G$ we define f_m resp. f^x by: $f_m(n) = f(nm)$ resp. $f^x(n) = f(x^{-1}nx)$ for every $n \in N$. A function f, such that $f = f^x$ for every $x \in G$, is called G-invariant.

1.2. We deal with locally compact groups satisfying certain compactness conditions (see [10]):

Moore-groups: those, all of whose irreducible representations are finite dimensional.

SIN-groups: those, which possess neighbourhood bases of compact sets, invariant under inner automorphisms.

FC-groups: those, whose conjugacy classes are relatively compact.

FD-groups: those, whose commutator subgroups are relatively compact.

Z-groups: those, whose factor groups modulo their centers are compact.

The abelian as well as the compact groups belong to every class mentioned here. Every group in these classes is unimodular, and G_F, the union of relatively compact conjugacy classes, is an open normal subgroup [23]. Every Moore-group is a SIN-group [25], the index of G_F in G, $[G:G_F]$, is finite, thus G is amenable, and G_F is a FD-group [23]. For further results we refer to Palmer's compendium [27].

1.3. Let now N be a closed normal subgroup of the SIN-group G. $K(N, G)$, the set of continuous, positive definit, G-invariant functions on N of norm not greater than 1 equipped with the weak-$*$-topology [as a subspace of $L^\infty(N)$], is a simplex (by means of [15, 1.2 Lemma] the proof is just that of [33, Satz 1]). Thus for every $\varphi \in K(N, G)$ there is a unique maximal decomposition measure μ_φ supported by the closure of extremal points. For a non-trivial one, the set of those is called $E(N, G)$, $\|\alpha\|_\infty = \alpha(e) = 1$ is valid. That is why the compact-open topology (co-topology) on $E(N, G)$ coincides with the relative weak-$*$-topology [5, 13.5.2]. We write $K(G)$ resp. $E(G)$ instead of $K(G, G)$ resp. $E(G, G)$, and identify $E(G/N)$ in the canonical way with $\{\alpha \in E(G);\ \alpha|N = 1\}$ [34, Satz 2, Korollar]. For abelian groups $E(G)$ is nothing but the dual group $\hat{G}$.

1.4. For a Moore-group G there is a canonical bijection t between the set $\hat{G}$ of equivalence classes of irreducible representations and $E(G)$: for $\pi \in \hat{G}$ let $t(\pi)$ be the normalized trace. We write d_π for the dimension of π and $m_\varrho(\pi)$ for the multiplicity of the occurence of π in any representation ϱ of G. Then we get the following formula for $\alpha, \beta \in E(G)$:

$$\mu_{\alpha\beta} = \sum_{\gamma \in E(G)} d_{\varrho_\gamma} d_{\varrho_\alpha}^{-1} d_{\varrho_\beta}^{-1} m_{\varrho_\alpha \otimes \varrho_\beta}(\varrho_\gamma) \delta_\gamma ,$$

where we write ϱ_α instead of $t^{-1}(\alpha)$, and denote by δ_γ the Dirac measure (compare [12]). The cyclic representation π_α obtained from $\alpha \in E(G)$ is a multiple of ϱ_α.

2. Non-Hausdorff Hypergroups

2.1. In this paragraph we want to give a brief sketch of a possible non-Hausdorff hypergroup theory. In order to get good results we can't depart too far from the locally compact Hausdorff case; our spaces are expected to have the following regularity properties:

Definition. A topological space is locally compact, if every point has a neighbourhood base of compact Borel sets (compare [8, 22]).

Definition. A locally compact space X is said to admit a Hausdorffization (Y, p), if there exists a continuous, proper surjection p from X onto a locally compact Hausdorff space Y, such that for every continuous map f from X into any locally compact Hausdorff space Z, there is a continuous map f_p from Y into Z satisfying $f = f_p \circ p$.

2.2. We first prove the following

Lemma. *Let X be a locally compact space and S a closed subset of X. If X admits a Hausdorffization, then so does S.*

Proof. Clearly S is locally compact, and we define an equivalence relation $\sim$ on S by: $x \sim y$ if and only if $f(x)=f(y)$ for every $f \in C(S)$. Let Y_S be the quotient and p_S the canonical projection. At first Y_S is Hausdorff, since points in different classes can be separated by a continuous function. Thus in order to prove local compactness, we need only show the existence of a single compact neighbourhood. Since p is proper, for every $x \in X$ there is $f \in C_{00}(X)$ such that $f(x)=1$. The restriction of f to S yields a compact neighbourhood K of $p_S(x)$, namely $K=p_S(\{y \in S;\ f(y) \geqq \frac{1}{2}\})$.

For a continuous map $f: S \to Z$, where Z denotes a locally compact Hausdorff space, there exists by definition of Y_S a map $f_{p_S}: Y_S \to Z$ satisfying $f_{p_S} \circ p_S = f$, which is continuous, since Y_S carries the quotient topology. Particularly there is a continuous map $r: Y_S \to Y$, such that $p|S = r \circ p_S$. For a compact subset K of Y_S we have:

$$p_S^{-1}(K) \subseteq p^{-1}(r(K)) \cap S;$$

hence $p_S^{-1}(K)$ is compact, since it is closed; thus p_S is proper.

Remark. Since p is surjective, a Hausdorffization (Y, p) is uniquely determined within homeomorphisms, in case it exists, and Y carries the quotient topology with respect to p.

2.3. Let's agree upon the following notations for a topological space X.

$\mathfrak{B}(X)$: the σ-algebra of Borel sets (generated by all open sets).
$B(X)$: the space of bounded, complex-valued Borel functions provided with the sup-norm $\| \ \|_\infty$.
$M(X)$: the space of bounded, complex Borel measures regarded as a (closed) subspace of $B(X)^*$, the dual of $B(X)$.
$M^+(X)$ resp. $B^+(X)$: the positive elements of $M(X)$ resp. $B(X)$.

Remark. It's easy to see, that positive, finite, linear combinations of Dirac measures are $\sigma(M(X), B(X))$-dense in $M^+(X)$. We denote by $U(\mu, f_i, \varepsilon)$ a canonical $\sigma(M(X), B(X))$-neighbourhood. For general measure theoretic facts we refer to [2].

2.4. *Definition.* A measure $\mu \in M^+(X)$ is called regular, if the measure of any open set can be approximated from below by that of compact Borel sets, and if the measure of any Borel set can be approximated from above by that of open sets.

An arbitrary measure is called regular, if in the canonical decomposition $\mu=\mu_1-\mu_2+i(\mu_3-\mu_4)$ with $\mu_i \in M^+(X)$ every μ_i is regular. We denote by $M_r(X)$ the set of regular measures.

Remark. If we define as usual for $\mu \in M(X)$ the absolut value $|\mu| \in M^+(X)$ by

$$|\mu|(A)=\sup\left\{\sum_{i \in I} |\mu(E_i)|;\ (E_i)_{i \in I} \text{ a finite, disjoint partition of } A\right\};$$

then we get $|\mu|(X)=\|\mu\|$, where the norm is that of $B(X)^*$.

2.5. *Definition.* The support of a measure $\mu \in M(X)$ is defined by $\cap\{A \subseteq X;\ A \text{ closed}, \mu(f)=0 \text{ for every } f \in B(X), \text{ such that } f|A=0\} = (\cup\{U \subseteq X;\ U \text{ open and } |\mu|(U)=0\})^c$, and denoted by $\operatorname{supp}\mu$.

Remark. For a regular measure μ $\operatorname{supp}\mu$ is the complement of the largest open μ-null set.

For a continuous map $f: X \to Y$ and $\mu \in M(X)$ let $\mu_f \in M(Y)$ denote the image of μ under f. In case $\mu \in M_r^+(X)$ we get $\operatorname{cl}(f(\operatorname{supp}\mu)) \supseteq \operatorname{supp}\mu_f$, and equality holds, if μ_f is regular too. This occurs for instance, if Y is Hausdorff, and if f is a proper surjection. At last for an element x of a T_1-space X we get:

$\{x\} \in \mathfrak{B}(X)$, $\{x\} = \operatorname{supp}\delta_x$ and $\delta_x \in M_r(X)$, where δ_x denotes a Dirac measure.

2.6. In the following we will call a Hausdorff-hypergroup, what Jewett calls a "convo". Since we can't expect that an abstract convolution of regular measures still yields a regular measure, we want $M(X)$ instead of $M_r(X)$ to become a $*$-Banach-algebra, which is one of the main purposes of introducing hypergroups (compare [18]). Thus we put the following

Definition. A topological space X is called a hypergroup, if the following axioms apply:

(I) X is a locally compact T_1-space with Hausdorffization (Y, p).

(II) There is a binary operation $*$ on $M(X)$, which makes $M(X)$ an associative algebra; this operation is separately continuous and $(x, y) \to \delta_x * \delta_y(f) \in B(X \times X)$ for every $f \in B(X)$.

(III) $\mu * \nu$ is positive, if μ and ν are; for $\mu, \nu \in M_r^+(X)$, $f_i \in C_{00}^+(X) \cup \{1\}$ $(1 \leqq i \leqq m)$ and $\varepsilon > 0$, there exist $g_j \in C_{00}^+(X) \cup \{1\}$ $(1 \leqq j \leqq n)$ and $\varepsilon' > 0$, such that the following holds: $\mu' * \nu' \in U(\mu * \nu, f_i, \varepsilon)$ for every regular $\mu' \in U(\mu, g_j, \varepsilon')$ and $\nu' \in U(\nu, g_j, \varepsilon')$.

(IV) For $x, y, z \in X$ $\delta_x * \delta_y$ is a probability measure with compact support; $\delta_x * \delta_y$, as well as $(\delta_x * \delta_y) * \delta_z$, are regular.

(V) For $x, y \in X$ and for open subsets of the form $U_i = p^{-1}(U_i')$ $(i = 1, 2)$, such that $\operatorname{supp}(\delta_x * \delta_y) \subseteq U_1$ and $\operatorname{supp}(\delta_x * \delta_y) \cap U_2 \neq \phi$, there exist open neighbourhoods $V_i = p^{-1}(V_i')$ $(i = 1, 2)$ of x resp. y, such that for every $x' \in V_1$ and $y' \in V_2$ we still have: $\operatorname{supp}(\delta_{x'} * \delta_{y'}) \subseteq U_1$ and $\operatorname{supp}(\delta_{x'} * \delta_{y'}) \cap U_2 \neq \phi$.

(VI) There is an $e \in X$, such that $\delta_x * \delta_e = \delta_e * \delta_x = \delta_x$ for every $x \in X$.

(VII) There is an involutive homeomorphism $x \to \bar{x}$ of X, such that the following holds for every $x, y \in X$: $e \in \operatorname{supp}(\delta_x * \delta_y)$ if and only if $x = \bar{y}$, and for every $\mu, \nu \in M(X)$: $\mu * \nu = \bar{\nu} * \bar{\mu}$, where "$-$" denotes the canonical extension to $M(X)$ as well.

A hypergroup is called commutative if $\delta_x * \delta_y = \delta_y * \delta_x$ holds for every $x, y \in X$.

Remark. (i) For a motivation of this system of axioms see [18] and 2.7 Theorem below.

(ii) In case X is Hausdorff and $\mu * \nu$ is regular (III) means nothing but the positiv-continuity of $(\mu, \nu) \to \mu * \nu$ from $M_r^+(X) \times M_r^+(X) \to M_r^+(X)$ in the sense of [18].

Let's mention that the axioms lead to an explicit convolution formula for measures $\mu, \nu \in M^+(X)$ and functions $f \in B^+(X)$:

$$\mu * \nu(f) = \int_X \int_X \delta_x * \delta_y(f) \, d\mu(x) \, d\nu(y);$$

this equality is valid for measures with finite support, of course, and carries over by means of Fubini's theorem and axiom (II) to the general case.

2.7. **Theorem.** (i) *Every Hausdorff-hypergroup is a hypergroup.*

(ii) *Every hypergroup, that is a Hausdorff space at the same time, is a Hausdorff-hypergroup.*

Proof. (i) Because of [18, Lemma 3.1 C] by

$$\mu*\nu(A)=\int_X\int_X \delta_x*\delta_y(A)\,d\mu(x)\,d\nu(y)$$

for $\mu, \nu\in M^+(X)$ and $A\in\mathfrak{B}(X)$ a convolution on $M^+(X)$ is defined for any Hausdorff-hypergroup X, whose linear extension to $M(X)$ makes $M(X)$ an associative algebra, since the following formula is valid:

$$((\mu*\nu)*\tau)(f)=\int_X\int_X\int_X((\delta_x*\delta_y)*\delta_z)(f)\,d\mu(x)\,d\nu(y)\,d\tau(z).$$

Most of the new axioms are direct conclusions of the old ones; Fubini's theorem implies the separate continuity of the convolution.

(ii) First we introduce a "regular" convolution on $M_r^+(X)$. $f\to\mu*\nu(f)$ is a positive functional on $C_0(X)$ for $\mu, \nu\in M_r^+(X)$, and we denote by $\mu*_r\nu$ the corresponding regular measure. Then the following holds:

Lemma. $(X, *_r)$ *is a Hausdorff-hypergroup.*

Proof. Because of axiom (IV) we get $\delta_x*\delta_y=\delta_x*_r\delta_y$ and $(\delta_x*\delta_y)*\delta_z=(\delta_x*_r\delta_y)*_r\delta_z$ likewise for every $x, y, z\in X$. The validity of Jewett's axiom corresponding to our's (I), (IV), (V), (VI), and (VII) is clear at a glance. Thus the analogues of axioms (II) and (III) are left to prove [18, Sect. 3 (II) and (III)]. It's obvious that $\mu'*\nu'\in U(\mu*\nu, f_i, \varepsilon)$ implies $\mu'*_r\nu'\in U(\mu*_r\nu, f_i, \varepsilon)$ for $f_i\in C_{00}^+(X)$. We need a proof for $f=1$, i.e. $\mu*\nu(X)=\mu*_r\nu(X)$ for $\mu, \nu\in M_r^+(X)$. Since μ and ν are regular, we can choose a compact $K\subseteq X$, such that $\mu(X\setminus K)<\varepsilon$ and $\nu(X\setminus K)<\varepsilon$. So we get

$$\begin{aligned}\mu*_r\nu(X)&=\sup\{\textstyle\int\int\delta_x*\delta_y(g)\,d\mu(x)\,d\nu(y);\, g\in C_0(X), \|g\|_\infty=1\}\\&\geqq(\mu(X)-\varepsilon)\cdot(\nu(X)-\varepsilon)\geqq\mu*\nu(X)-(\mu(X)+\nu(X))\varepsilon,\end{aligned}$$

since $\bigcup_{x,y\in K}\operatorname{supp}(\delta_x*\delta_y)$ is contained in a compact subset, which is proved using [18, 2.5 D and 2.5 F] and axiom (IV) just as in [18, Lemma 3.2 B]. Finally the mapping $(x, y)\to\delta_x*_r\delta_y$ fullfills the assumption of [18, Lemma 2.4 B], hence $(\mu, \nu)\to\mu*_r\nu$ is the unique extension of $(\delta_x, \delta_y)\to\delta_x*_r\delta_y$, which is therefore associative. Thus the Lemma is proved.

Now the proof of (ii) is clear, since [18, Lemma 3.1 E] applies: for $\mu, \nu\in M_r^+(X)$ and $f\in B^+(X)$ we have

$$\mu*_r\nu(f)=\int\int\delta_x*\delta_y(f)\,d\mu(x)\,d\nu(y)=\mu*\nu(f),$$

and by linearity reasons: $\mu*_r\nu=\mu*\nu$.

2.8. Just as in [18, Theorem 6.1 G] one can prove:

Theorem. *For every hypergroup X we have: $M(X)$ is an involutive Banach algebra with unit. This algebra is commutative, in case X is.*

2.9. *Examples.* Besides locally compact groups [18, Sect. 4, Proposition 2], and double-coset spaces [18, Theorem 8.2 B] we will mention the following examples:

(i) Let G be a SIN-group and N a closed normal subgroup contained in G_F. The set of closures of G-orbits form a Hausdorff-hypergroup [18, Theorem 8.3 A].

(ii) Let G and N be as above. $E(N, G)$, which is the hypergroup dual of example (i), is a Hausdorff-hypergroup [11, Theorem 1].

(iii) We will prove in paragraph 4, that the dual $\hat{G}$ [more generally $E(N, G)$] is a hypergroup for any Moore-group G.

2.10. *Definition.* A closed, non-empty subset S of a hypergroup X [with Hausdorffization (Y_S, p_S)] is called a subhypergroup, if the following conditions hold:

(i) $S=\bar{S}$.

(ii) $\operatorname{supp}(\delta_x * \delta_y) \subseteq S$ for every $x, y \in S$.

(iii) The axioms (III) and (V) of 2.6 are valid for S and p_S.

Sets of the form $x*S=\cup\{\operatorname{supp}(\delta_x*\delta_y);\ y\in S\}$ are called hypercosets.

Remarks. (i) Because of (i), (ii), and (VII) we have $e\in S$.

(ii) As a result of (i) and (ii) all axioms except (III) and (V) are valid for S, since we have $M_r(X)\cap M(S)\subseteq M_r(S)$; therefore a subhypergroup is a hypergroup.

(iii) We speak of a quasi-subhypergroup, if (i) and (ii) hold. If X is Hausdorff, every quasi-subhypergroup is allready a subhypergroup, since we need no Hausdorffization.

3. Topologies on $E(N, G)$

3.1. Before we can show that $E(N, G)$ is a hypergroup for a closed normal subgroup N of a Moore-group G, we have to examine in this paragraph two topologies on $E(N, G)$, the corresponding Borel structures, as well as properties of the restriction – resp. invariance – map.

For $\gamma\in K(N, G)$ the functional $f\to \int_N f(x)\gamma(x)dx$ on $L^1(N)$ is positive, thus can be extended to $C^*(N)$. Denote by $k(\gamma)$ resp. $I_\gamma=k(\gamma)\cap L^1(N)$ the corresponding G-invariant (closed) ideals in $C^*(N)$ resp. $L^1(N)$. As shown in [13, Lemma 2 (ii)] this yields canonical bijections between $E(N, G)$ and the sets G-Max $L^1(N)$ resp. G-Max $C^*(N)$ of maximal, G-invariant ideals in $L^1(N)$ resp. $C^*(N)$. Thus we can equip $E(N, G)$ with the respective hull-kernel-topologies [13, p. 5].

The invariance-map $s: E(M, N)\to E(M, G)$ defined as in [13, Lemma 2(i)] for closed normal subgroups M and N of G (M contained in N) by

$$s(\sigma)=[G:G_\sigma]^{-1}\sum_{x\in G/G_\sigma}\sigma^x,$$

where G_σ is the stabilizer of $\sigma\in E(M, N)$, is not only continuous for both topologies (the proof of [13, Lemma 2(ii)] applies to G-Max $C^*(N)$ as well), but is a quotient map, as is easily seen. Since [4, Satz 2] applies to Moore-groups, i.e. Prim $C^*(G)=$ Max $C^*(G)$ and $\operatorname{Prim}_*\ L^1(G)=\operatorname{Max}\ L^1(G)$ are homeomorphic, the hull-kernel-topologies on $E(N, G)$ coincide (for definitions see [4, 27]). Of course this topology (hk-topology) is coarser than the co-topology, since we have $I_\gamma\supseteq\cap\{I_{\gamma'};\ \gamma'\in A\}$ for $\gamma\in\text{co-cl}(A)$. These topologies coincide, if N is contained in G_F [26, 4.6

Corollary]. If not otherwise stated, we equip $E(N, G)$ with the hk-topology. We write $k(E)=\cap\{k(\gamma);\ \gamma\in E\}$ for any subset E of $E(N, G)$.

3.2. Let's first look at the restriction-map. Let again be M and N closed normal subgroups of G, M contained in N. $\alpha\to\alpha|M$ from $K(N, G)$ into $K(M, G)$ is of course continuous with respect to the compact-open topologies.

$r: E(N, G)\to E(M, G)$, the restriction of the above map, is surjective [15, Satz 2] and even continuous in case M is open in N, since we have $I_{r(\gamma)}=I_\gamma\cap L^1(M)$, $L^1(M)$ regarded as a subspace of $L^1(N)$.

Let's now turn to the case $r: E(G)\to E(N, G)$. In [19, Theorem 1] it is shown for any amenable SIN-group, that for an open N contained in G_F r is open and proper. We will prove the following

Lemma. *Let G and N be as in* 3.1. *$r: E(G)\to E(N, G)$ is a quotient map.*

Proof. We have to show that A is closed if and only if $r^{-1}(A)$ is so. Because of $\ker\pi_\alpha|N=\ker\pi_{\alpha|N}$ [16, formula (2.4)], $k(\alpha)=\ker\pi_\alpha$ (1.4) and the well-known fact, that relations between kernels are preserved under restrictions resp. inductions of representations, we get the following:

"only if": Let A be closed and $\alpha\in\mathrm{cl}(r^{-1}(A))$, i.e. $k(\alpha)\supseteq k(r^{-1}(A))$. Then $k(\alpha|N)\supseteq k(A)$, hence $\alpha|N\in A$, since A is closed.

"if": Let now $r^{-1}(A)$ be closed and $\gamma=\alpha|N\in\mathrm{cl}(A)$. Since G is amenable, we have $k(\alpha)\supseteq k(U^{\pi_\gamma})$, where U^{π_γ} denotes the representation induced by π_γ from N to G. From $k(\gamma)\supseteq k(A)$ it follow that $\ker U^{\pi_\gamma}\supseteq\bigcap_{\gamma'\in A}\ker U^{\pi_{\gamma'}}$ holds. Thus we get $k(\alpha)\supseteq k(r^{-1}(A))$, if we prove $\ker U^{\pi_\gamma}\supseteq k(r^{-1}(\gamma))$ for $\gamma\in E(N, G)$. But since $\ker U^{\pi_\gamma}$ is the intersection of the $k(\alpha)$'s containing it, we need only show that $k(\alpha)\supseteq\ker U^{\pi_\gamma}$ implies $\alpha|N=\gamma$. Now the following holds:

$$\ker\pi_{\alpha|N}=\ker\pi_\alpha|N\supseteq\ker U^{\pi_\gamma}|N=\ker\pi_\gamma,$$

since γ is G-invariant. Because of $\ker\pi_\gamma\in G\text{-Max } C^*(N)$ this yields $\alpha|N=\gamma$.

3.3. Let again M and N be closed normal subgroups of the Moore-group G, M contained in N. $s: E(M, N)\to E(M, G)$ is a quotient map, which is even open and proper provided that M is contained in G_F [13, Lemma 2; 26, 5.8 Theorem]. Let's take a closer look at the properties of s and r.

Theorem. (i) *$s: E(M, N)\to E(M, G)$ is an open, proper quotient map.*

(ii) *$r: E(N, G)\to E(M, G)$ is a quotient map; r is proper provided M is open in N, and even open, if in addition M in contained in N_F.*

Proof. (i) In the course of the proof we will call an open, proper quotient map good. s_2: $E(M\cap G_F, M)\to E(M\cap G_F, G)$ is good of course, since $s_3: E(M\cap G_F)\to E(M\cap G_F, M)$ and $s_4: E(M\cap G_F)\to E(M\cap G_F, G)$ are, and since $s_4=s_2\circ s_3$ [13, Proof of Lemma 2]. Now look at the following commutative diagram:

$$\begin{array}{ccc} E(M) & \xrightarrow{s_1} & E(M, G) \\ \downarrow{\scriptstyle z_1} & & \downarrow{\scriptstyle z_2} \\ E(M\cap G_F, M) & \xrightarrow{s_2} & E(M\cap G_F, G), \end{array}$$

where all maps are continuous surjections; r_1 and s_2 are even good. Let $K \subseteq E(M,G)$ be closed and compact; since $s_1^{-1}(K)$ is closed and contained in $r_1^{-1}s_2^{-1}r_2(K)$, hence compact, s_1 is proper. Let on the other side $U \subseteq E(M)$ be open. Since $s_1^{-1}s_1(U) = \cup\{U^x; x \in G\}$ is open ($\sigma \to \sigma^x$ is a homeomorphism), and since s_1 is a quotient map, $s_1(U)$ is open too. By a similar argument we see, that r_2 is also good. Finally $s_0 : E(M) \to E(M,N)$ and s_1 are good, $s_1 = s \circ s_0$, thus s is good.

(ii) Let M be open and put $r_1 : E(N) \to E(M,N)$ and $r_2 : E(M,N) \to E(M \cap N_F, N)$. Then $r_2 \circ r_1$ is good; hence as in the proof for s_1 in (i) r_1 is proper, and naturally open, if M is contained in N_F. Now have a look at the diagram:

$$\begin{array}{ccc} E(N) & \xrightarrow{r_1} & E(M,N) \\ \downarrow{\scriptstyle s_1} & & \downarrow{\scriptstyle s_2} \\ E(N,G) & \xrightarrow{r} & E(M,G); \end{array}$$

here r is proper, since s_2 and r_1 are; if M is contained in N_F, r_1 is even open, and so is r.

Remark. If domain and image of a map are locally compact, the image being even Hausdorff, then continuity and properness imply closedness.

3.4. Next we want to show, that the Borel structures received from the hk-resp. co-topologies on $E(N, G)$, are the same. For $E(G)$ this was shown in [12, Sect. 2]. We ought to mention, that, for second countable groups, this follows on one side from the well-known fact [7], that the Borel structure of $\hat{G}$ equals that obtained from a regularized topology [8]. On the other side for Moore-groups it is not too difficult to show, that this regularized topology regarded as one on $E(G)$ is not croaser than the co-topology.

We require this result on $E(G)$ to prove the corresponding fact about $E(N, G)$. At first we need some notations to examine the invariance-map $s : E(N) \to E(N, G)$ further. Put

$$B_{\geqq m} = \{\sigma \in E(N); [G : G_\sigma] \geqq m\},$$

G_σ as in 3.1 the stabilizer of σ. Define $B_{\leqq m}$ analogously and $B_m = B_{\geqq m} \cap B_{\leqq m}$. Then we get the following

Lemma. *$B_{\geqq m}$ is open in the co-topology, and $s|B_m : B_m \to E(N,G)$ is continuous with respect to the relative co-topologies.*

Proof. For $\sigma \in B_m$ choose $x_i \in G$ $(1 \leqq i \leqq m)$, such that the G-orbit $G(\sigma)$ equals $\{\sigma^{x_i}; 1 \leqq i \leqq m\}$. Then for a suitable co-neighbourhood U of σ the x_i-conjugates are pairwise different in each case, i.e. $U \subseteq B_{\geqq m}$, which is therefore open. Thus we can choose U, such that $G(\tau) = \{\tau^{x_i}; 1 \leqq i \leqq m\}$ for every $\tau \in U \cap B_m$. It is clear, that this implies continuity of $s|B_m$.

3.5 **Theorem.** *The Borel structures on $E(N, G)$ received from the hk-resp. co-topologies coincide.*

Proof. Since the co-topology is not croaser than the hk-topology, it's enough to show, that every co-open set U of $E(N, G)$ is Borel with respect to the hk-topology. One has proved this, if one knew, whether the image of a hk-Borel subset of $E(N)$

under s is hk-Borel in $E(N, G)$ again, but that's not clear at all. Thus we must take another way.

We define $E_{\geqq n}=\{\sigma\in E(N);\ d_{t^{-1}(\sigma)}\geqq n\}$ für $n\in\mathbb{N}$, which is open in both, the hk- as well as the co-topology; analogously $E_{\leqq n}$ and $E_n=E_{\geqq n}\cap E_{\leqq n}$. Mention that the relative hk- resp. co-topologies coincide on E_n [12, Sect. 2]. For our co-open $U\subseteqq E(N, G)$ there are by 3.4 Lemma co-open subsets U_m of $B_{\geqq m}(m\in\mathbb{N})$, such that $s^{-1}(U)\cap B_m=U_m\cap B_m$. Thus we get for $m, n\in\mathbb{N}$:

$$s^{-1}(U)\cap B_m\cap E_n=U_m\cap B_{\leqq m}\cap E_n.$$

Since U_m is co-open and $B_{\leqq m}$ is co-closed, there exists hk-open subsets $V_{m,n}$ and hk-closed ones $A_{m,n}$ of $E(N)$ such that:

$$V_{m,n}\cap E_n=U_m\cap E_n \quad\text{and}\quad A_{m,n}\cap E_n=B_{\leqq m}\cap E_n.$$

Notice that $s^{-1}(U)\cap B_m\cap E_n=V_{m,n}\cap B_{\leqq m}\cap E_n$ as well as $B_{\leqq m}\cap E_n$ are full inverse images under s, since conjugates have the same properties concerning G-orbits and dimensions. Hence we get on the one hand

$$s^{-1}(U)\cap B_m\cap E_n=s^{-1}s(V_{m,n})\cap B_{\leqq m}\cap E_n,$$

where $s(V_{m,n})$ is hk-open, since s is open (3.3 Theorem). On the other hand, if we define

$$A'_{m,n}=\{\sigma\in E(N);s^{-1}s(\sigma)\subseteqq A_{m,n}\},$$

we see that $A'_{m,n}$ contains $B_{\leqq m}\cap E_n$, and that $s(A'_{m,n})$ is hk-closed, since s is hk-open. Summing up we get:

$$\begin{aligned}U&=s\Big(\bigcup_{m,n\in\mathbb{N}}s^{-1}(U)\cap B_m\cap E_n\Big)\\&=\bigcup_{m,n\in\mathbb{N}}s(V_{m,n}\cap A'_{m,n}\cap E_{\geqq m}\cap E_{\leqq m})\\&=\bigcup_{m,n\in\mathbb{N}}(s(V_{m,n})\cap s(A'_{m,n})\cap s(E_{\geqq m})\cap s(E_{\leqq m})).\end{aligned}$$

Here $s(E_{\geqq m})$ resp. $s(E_{\leqq m})$ are hk-open resp. hk-closed. Thus U is Borel with respect to the hk-topology.

4. $E(N, G)$ Respectively $\hat{G}$ is a Hypergroup

4.1. Homeomorphic to G-Max $C^*(N)$ evidently $E(N, G)$ is a T_1-space.

Lemma. *Let G be a Moore-group and N a closed normal subgroup. $E(N, G)$ equipped with the hk-topology is locally compact.*

Proof. Suppose V to be a neighbourhood of $\gamma\in E(N, G)$. Then $U=s^{-1}(V)$ is one of some $\sigma\in s^{-1}(\gamma)$. For $\pi=t^{-1}(\sigma)\in\hat{N}$ there exists an $a\in C^*(N)$, such that $\pi(a)\neq 0$ and $\varrho(a)=0$ for every $\varrho\in\hat{N}\backslash t^{-1}(U)$. The compact neighbourhood of π mentioned in [8, Theorem 2.1], namely

$$C=\{\varrho\in\hat{N};\|\varrho(a)\|\geqq\tfrac{1}{2}\|\pi(a)\|\},$$

is the countable intersection of the open sets

$$W_n = \left\{ \varrho \in \hat{N} ; \|\varrho(a)\| > \left(\frac{1}{2} - \frac{1}{n}\right) \|\pi(a)\| \right\}.$$

For $K = s \circ t(C)$ and $U_n = s \circ t(W_n)$ we have: K is a compact neighbourhood of γ contained in V, and is Borel, since it is the intersection of the U_n.

4.2. Before we can prove, that $E(N, G)$ admits a Hausdorffization, we need a little topological lemma concerning this matter.

Lemma. *Let X and Y be locally compact spaces, Y even Hausdorff, and $p: X \to Y$ a continuous, open, proper surjection. Suppose further, that there exists a dense subset D of X, such that $p|D$ is injective; then (Y, p) is a Hausdorffization of X.*

Proof. Let Z be any locally compact Hausdorff space and $f: X \to Z$ a continuous map. Choose $x, y \in X$ such that $p(x) = p(y)$. Now there is a net $(x_\iota)_{\iota \in I}$ in D with x as a limit point. Thus because of the continuity of p the net $(p(x_\iota))_{\iota \in I}$ converges to $p(x) = p(y)$. Since p is open and $p|D$ is one-to-one, it is easily seen, that y is a limit point of $(x_\iota)_{\iota \in I}$ too. But this implies $f(x) = f(y)$. Thus we can define a map $f_p: Y \to Z$ by $f_p(p(x)) = f(x)$, which is continuous, since f is, and since p is open.

4.3. Now we can show, how the Hausdorffization of $E(N, G)$ looks like.

Lemma. *Let G and N be as in* 4.1 *and define $p: E(N, G) \to E(N \cap G_F, G)$, the restriction-map, then we have: $(E(N \cap G_F, G), p)$ is a Hausdorffization of $E(N, G)$.*

Proof. (Compare for the following [12, Lemma 1 and Corollaries 1 and 2].) The set $E = \{\gamma \in E(G_F, G) ; \tilde{\gamma} \in E(G)\}$ is dense in $E(G_F, G)$. Since the restriction-map $\alpha \to \alpha|N$ is a quotient map and $\alpha \to \alpha|G_F$ is even open (3.3 Theorem), the set $D = \{\tilde{\gamma}|N ; \gamma \in E\}$ is dense in $E(N, G)$. Clearly $p|D$ is one-to-one. Now $E(N \cap G_F, G)$ is a locally compact Hausdorff space (3.1 and 4.1) and p is a continuous, open, proper surjection (3.3); hence 4.2 Lemma yields the assertion.

4.4. **Theorem.** *Let G be a Moore-group and N a closed normal subgroup of G. Then $E(N, G)$ equipped with the hk-topology is a hypergroup; in particular this is valid for $\hat{G}$.*

Proof. (I) is allready shown.

(IV): Look at the restriction map $\alpha \to \alpha|N$ from $K(G)$ to $K(N, G)$, $\mu_{\alpha|N}$ is the image of μ_α under this map. Thus by

$$\delta_\alpha * \delta_\beta = \mu_{\alpha\beta} = \sum_{\gamma \in E(N,G)} \mu_{\alpha\beta}(\gamma) \delta_\gamma$$

there is defined a probability measure on $E(N, G)$ with finite support. In case N equals G, we can deduce from the formula in 1.4 for every $\alpha, \beta, \gamma, \tau \in E(G)$:

$$\sum_{\sigma \in E(G)} \mu_{\alpha\beta}(\sigma) \mu_{\sigma\gamma}(\tau) = \mu_{\alpha\beta\gamma}(\tau).$$

Since we have for $\sigma' \in E(N, G)$ and any $\sigma_0 \in r^{-1}(\sigma')$

$$(\mu_{\alpha\beta}(r^{-1}(\sigma'))\mu_{\sigma_0\gamma})_r = \left(\sum_{\sigma \in r^{-1}(\sigma')} \mu_{\alpha\beta}(\sigma) \mu_{\sigma\gamma} \right)_r,$$

we get for $\tau' \in E(N, G)$:

$$\sum_{\sigma' \in E(N,G)} \mu_{\alpha|N \cdot \beta|N}(\sigma') \mu_{\sigma'\gamma|N}(\tau') = \sum_{\substack{\sigma' \in E(N,G) \\ \sigma \in r^{-1}(\sigma')}} \mu_{\alpha\beta}(r^{-1}(\sigma')) \mu_{\sigma\gamma}(r^{-1}(\tau'))$$

$$= \sum_{\sigma \in E(G)} \mu_{\alpha\beta}(\sigma) \mu_{\sigma\gamma}(r^{-1}(\tau')) = \mu_{\alpha\beta\gamma}(r^{-1}(\tau'))$$

$$= \mu_{\alpha|N \cdot \beta|N \cdot \gamma|N}(\tau').$$

Thus for $\alpha, \beta, \gamma \in E(N, G)$ the following holds:

$$(\delta_\alpha * \delta_\beta) * \delta_\gamma = \mu_{\alpha\beta\gamma} = \delta_\alpha * (\delta_\beta * \delta_\gamma).$$

$(\delta_\alpha * \delta_\beta) * \delta_\gamma$ is still a probability measure with finite support, hence regular.

(II) As in the proof of [12, Lemma 2] one can show by using [1, II. 3.15] that $(\alpha, \beta) \to \mu_{\alpha\beta}(A)$ is Borel for every Borel subset A of $E(N, G)$, Borel with respect to the co-topology, hence 3.5. Theorem yields, that it is Borel with respect to the hk-topology. Thus for $\mu, \nu \in M^+(E(N, G))$ we can define a positive measure on $E(N, G)$ by

$$\mu * \nu(f) = \int\int \delta_\alpha * \delta_\beta(f) d\mu(\alpha) d\nu(\beta), f \in B^+(E(N, G));$$

the σ-additivity follows from the Monotone-Konvergence-Theorem. As in the proof of 2.7 Theorem, the linear extension to $M(E(N, G))$ is associative and separately continuous.

(III) Let $p: E(N, G) \to E(N \cap G_F, G)$ be as in 4.3 Lemma; μ_p denotes the image of a measure μ under p. For $f_k \in B(E(N \cap G_F, G))$, $\varepsilon > 0$ and $\mu \in M(E(N, G))$ we have

$$\nu \in U(\mu, f_k \circ p, \varepsilon), \quad \text{iff} \quad \nu_p \in U(\mu_p, f_k, \varepsilon).$$

On the other hand we get:

$$(\mu * \nu)_p = \mu_p * \nu_p,$$

since $(\delta_\alpha * \delta_\beta)_p = \delta_{p(\alpha)} * \delta_{p(\beta)}$ is valid, as in the proof of (IV). By 2.5 Remark we have $\mu_p * \nu_p \in M_r^+(E(N \cap G_F, G))$ for $\mu, \nu \in M_r^+(E(N, G))$. Since every $f \in C_{00}^+(E(N, G)) \cup \{1\}$ is of the form $g \circ p$ for some $g \in C_{00}^+(E(N \cap G_F, G)) \cup \{1\}$, (III) is valid, because it is for the Hausdorff-hypergroup $E(N \cap G_F, G)$ [11, Theorem 1]

(V) Because of $\operatorname{supp}(\delta_{p(\alpha)} * \delta_{p(\beta)}) = p(\operatorname{supp}(\delta_\alpha * \delta_\beta))$ [by (2.5)] this follows again from the respective property for $E(N \cap G_F, G)$.

(VI) $1 \in E(N, G)$, and we have: $\delta_1 * \delta_\alpha = \mu_{1 \cdot \alpha} = \delta_\alpha$ for every $\alpha \in E(N, G)$.

(VII) Define $\bar{\pi} \in \hat{G}$ as usual, and mention that $t(\bar{\pi}) = \overline{t(\pi)}$ is valid [17, Sect. 27]. Thus clearly $\alpha \to \bar{\alpha}$ is an involutive homeomorphism, such that $\mu * \nu(\bar{f}) = \bar{\mu} * \bar{\nu}(f)$. Now one property is left to prove. Let first G equal N. By turning to the Bohr compactification we get:

$$m_{\pi \otimes \varrho}(1) = m_{\pi \otimes 1}(\bar{\varrho}) = m_\pi(\bar{\varrho}) \quad \text{for} \quad \pi, \varrho \in \hat{G},$$

hence: $1 \in \operatorname{supp}(\delta_\alpha * \delta_\beta)$, iff $\alpha = \bar{\beta}$.

Let's turn to the general case. Let $s: E(N) \to E(N, G)$ be the invariance map and define $n_\alpha = [G : G_\alpha]$, the index of the stabilizer, for $\alpha \in E(N)$. For $\alpha, \beta, \gamma \in E(N)$ we get:

$$s(\alpha) s(\beta) = (n_\alpha n_\beta)^{-1} \sum_{\substack{x \in G/G_\alpha \\ y \in G/G_\beta}} \alpha^x \beta^y = \sum_{\delta \in E(N,G)} (n_\alpha n_\beta)^{-1} \sum_{\substack{x \in G/G_\alpha \\ y \in G/G_\beta}} \mu_{\alpha^x \beta^y}(s^{-1}(\delta)) \cdot \delta,$$

hence

$$\mu_{s(\alpha)s(\beta)}(s(\gamma)) = n_\gamma (n_\alpha n_\beta)^{-1} \sum_{\substack{x \in G/G_\alpha \\ y \in G/G_\beta}} \mu_{\alpha^x \beta^y}(\gamma).$$

Thus for $\gamma = s(\gamma) = 1$ we have: $\mu_{s(\alpha)s(\beta)}(1) > 0$, iff $\mu_{\alpha^x \beta^y}(1) > 0$ for some x, $y \in G$, iff $\alpha^x = \overline{\beta^y}$ for some x, $y \in G$, iff $s(\alpha) = s(\overline{\beta})$.

4.5. *Remark.* (i) The above formular yields an interpretation of the convolution by means of representations. Write π_α^G for the direct sum over the G-orbit of $\varrho_\alpha = t^{-1}(\alpha)$, i.e. $\pi_\alpha^G = \bigoplus_{x \in G/G_\alpha} (\varrho_\alpha)^x$; so $s(\alpha)$ is the normalized trace of π_α^G, and we get

$$\begin{aligned} \mu_{s(\alpha)s(\beta)}(s(\gamma)) &= d_{\varrho_\gamma} n_\gamma (d_{\varrho_\alpha} n_\alpha d_{\varrho_\beta} n_\beta)^{-1} \sum_{\substack{x \in G/G_\alpha \\ y \in G/G_\beta}} m_{(\varrho_\alpha)^x \otimes (\varrho_\beta)^y}(\varrho_\gamma) \\ &= d_{\pi_\gamma^G} (d_{\pi_\alpha^G} d_{\pi_\beta^G})^{-1} \cdot m_{\pi_\alpha^G \otimes \pi_\beta^G}(\pi_\gamma^G), \end{aligned}$$

since the ϱ_γ^x, $x \in G/G_\gamma$, are mutuallty inequivalent and since $\pi_\alpha^G \otimes \pi_\beta^G$ is G-invariant. (For the case $G = N$ compare 1.4).

(ii) $E(G/N) = \{\alpha \in E(G);\ \alpha | N = 1\}$ is a subhypergroup. Since (i) of the definition is clear, let's look at (ii). For $n \in N$ and α, $\beta \in E(G/N)$ we have

$$1 = \alpha \cdot \beta(n) = \sum_{\gamma \in E(G)} \mu_{\alpha\beta}(\gamma)\, \gamma(n);$$

but $|\gamma(n)| \leqq 1$ implies $\gamma(n) = 1$ for every $\gamma \in \operatorname{supp} \mu_{\alpha\beta}$, i.e. $\operatorname{supp} \mu_{\alpha\beta} \subseteq E(G/N)$. Now $E(G/N)$ is closed in $E(G)$, and G/N is a Moore-group again. Since the decomposition of $\alpha \cdot \beta$ is the same read as an element of $K(G)$ or of $K(G/N)$, (iii) follows at once.

5. The Main Theorem

5.1. We will state our results in terms of $E(G)$ rather than of $\hat{G}$.

Definition. Let G be a Moore-group and R a subset of $E(G)$, then we put

$$N_R = \{x \in G;\ \alpha(x) = 1 \quad \text{for every} \quad \alpha \in R\},$$

and for a subset M of G analogously

$$S_M = \{\alpha \in E(G),\ \alpha(x) = 1 \quad \text{for every} \quad x \in M\}.$$

Remark. For a closed normal subgroup N of G we have $S_N = E(G/N) = B_N \cap E(G)$ (compare the introduction) and because of the Gelfand-Raikov-Theorem [17, (22.12) Theorem] applied to G/N:

$$N_{S_N} = N \text{ (see also [26, 3.4. Corollary])}.$$

5.2. **Main Theorem.** *Let G be a Moore-group.*

(i) *S_M is a subhypergroup of $E(G)$ for any subset M of G, and N_{S_M} is the smallest closed normal subgroup of G containing M.*

(ii) *N_R is a closed normal subgroup of G for any subset R of $E(G)$, and S_{N_R} is the smallest subhypergroup of $E(G)$ containing R.*

Proof. Naturally N_R is a closed normal subgroup. Let now N be the smallest closed normal subgroup of G containing M. Then we have $S_N \subseteq S_M$. On the other side $N \subseteq N_{S_M}$, since $M \subseteq N_{S_M}$, i.e. $S_M \subseteq S_N$. Hence $S_M = S_N = E(G/N)$ is a subhypergroup (4.5 Remark) and we get:

$$N_{S_M} = N_{S_N} = N,$$

thus (i) is proved.

It is clear that there exists a smallest quasi-subhypergroup S containing R. If we have shown $S = S_{N_S}$ for any quasi-subhypergroup S of $E(G)$, then it follows as in the proof of (i) $S = S_{N_R}$. Thus being a subhypergroup (5.1 Remark), S is already the smallest one containing R. Regarding S as a quasi-subhypergroup of $E(G/N_S)$ we may assume S is faithful, i.e. $N_S = \{e\}$. To obtain our result we therefore need only prove the following

Theorem. *Let G be a Moore-group. Then $E(G)$ itself is the only faithful quasi-subhypergroup of $E(G)$.*

This will be done by using the structure of Moore-groups. We have to climb up the following stages: a Z-group, a compactly generated Moore-group with relatively compact conjugacy classes, then drop the "compactly generated", and finally the general case.

5.3. Previously we examine the connection between quasi-subhypergroups of $E(G)$, $E(N, G)$ and $E(N)$ respectively, where N is a closed normal subgroup of G.

Lemma. *Let S be a quasi-subhypergroup of $E(G)$. If $r(S)$ is closed, then it is a quasi-subhypergroup; $r(S)$ is faithful, in case S is.*

Proof. Since (i) of 2.10 is trivial, assume $\gamma_k \in r(S)$ $(k = 1, 2)$ such that $\gamma_k = \alpha_k | N$ for some $\alpha_k \in S$. As in the proof of 4.4 Theorem (IV) we have $\delta_{\gamma_1} * \delta_{\gamma_2} = (\delta_{\alpha_1} * \delta_{\alpha_2})_r$, hence as shown in 2.5 $\operatorname{supp}(\delta_{\gamma_1} * \delta_{\gamma_2}) = r(\operatorname{supp}(\delta_{\alpha_1} * \delta_{\alpha_2}))$, which is contained in $r(S)$. Now, if we take $x \in N$, such that $\gamma(x) = 1$ for every $\gamma \in r(S)$, then clearly $\alpha(x) = 1$ for every $\alpha \in S$, thus $x = e$, if S is faithful.

5.4. **Lemma.** *Let $T \subseteq E(N, G)$ be a quasi-hypergroup; then $s^{-1}(T)$ is a quasi-subhypergroup too, and is faithful, if T is.*

Proof. Since s is continuous, $s^{-1}(T)$ is closed, and we have $\bar{\sigma} \in s^{-1}(\overline{s(\sigma)})$ for any $\sigma \in E(N)$. Let now $\sigma \in \operatorname{supp}(\delta_{\sigma_1} * \delta_{\sigma_2})$, where $\sigma_i \in s^{-1}(T)$. Because of the decomposition formula for $s(\sigma_1) \cdot s(\sigma_2)$ in 4.4 Theorem (VII) we get $s(\sigma) \in \operatorname{supp}(\delta_{s(\sigma_1)} * \delta_{s(\sigma_2)}) \subseteq T$, thus (ii) is proved. Now look at $x \in N$, such that $\sigma(x) = 1$ for every $\sigma \in s^{-1}(T)$, then by definition $s(\sigma)(x) = 1$, since every σ^y is a member of $s^{-1}(T)$. Hence we have $x = e$, if T is faithful.

5.5. We now turn to the first step of proving 5.2 Theorem, thus assume G is a Z-group, and S is a subhypergroup [$E(G)$ is Hausdorff]. Since any quotient of a Z-group is a Z-group again, we have to show

Lemma. 5.2. *Theorem is true for any Z-group.*

Proof. Following an idea of [29, Proposition 3.5] we consider the set

$$E = \cup \{\operatorname{supp} \mu_{\overline{\alpha\alpha}}, \alpha \in S\} \subseteq S.$$

By means of the character formula [26, 4.4 Proposition] and the fact, that for a Z-group G/Z is isomorphic to the group of inner automorphisms (Z denotes the

center of G), we get the following equality for $\alpha \in E(G)$ and $x \in G$:

$$\sum_{\beta \in E(G)} \mu_{\bar{\alpha}\alpha}(\beta)\beta(x) = \alpha(x)\alpha(x^{-1}) = \int_{G/Z} \alpha(yxy^{-1}x^{-1})d\dot{y},$$

where $d\dot{y}$ denotes the normalized Haar measure on G/Z. Now $x \in N_E$, iff the left side equals 1 for every $\alpha \in S$, thus, iff the right side equals 1 for every $\alpha \in S$. But this means, that $\alpha(yxy^{-1}x^{-1}) = 1$ for every $\alpha \in S$ and $y \in G$, since α is continuous. Hence $N_E = Z$, S taken to be faithful. So we have

$$E \subseteq S_{N_E} = E(G/Z),$$

and $S \cap E(G/Z)$ is a faithful subhypergroup of $E(G/Z)$, hence equals $E(G/Z)$ because of Helgason's result about compact groups.

Now look at the map $r: E(G) \to E(Z, G) = E(Z)$. We apply [29, 3.7 Proposition] to the hypergroup of conjugacy classes of G and the dual hypergroup $E(G)$ (see also [29, Sect. 5]) to have:

$$r^{-1}r(\alpha) = \cup\{\mathrm{supp}(\delta_\alpha * \delta_\beta); \beta \in E(G/Z)\}.$$

Thus we have $S = r^{-1}r(S)$, since $E(G/Z) \subseteq S$, and $r(S)$ is closed, since r is a lokal homeomorphism [9, 1.4 Theorem]. Hence as a faithful subhypergroup of $E(Z)\, r(S)$ equals $E(Z)$ because of the validity of 5.2 Theorem in the abelian case. Summing up we get:

$$S = r^{-1}(r(S)) = r^{-1}(E(Z)) = E(G).$$

5.6. In preparation of the second step, where we look at a compactly generated Moore-group, that is a FC-group at the same time, we prove an auxiliary result. Such a group is a FD-group, and we denote by P the normal subgroup of compact elements, which turns out to be compact in this case [10, (3.20) Theorem].

Lemma. *Let G be a FD-Moore-group and S a faithful subhypergroup of $E(G)$, such that $S \cap E(G/P) = \{1\}$; then we have $G = P$.*

Proof. By [23, 5.2, and 5.3] G is a projective limit of Z-groups, i.e. for every U from a neighbourhood base $\mathfrak{U}$ of e, there is a compact normal subgroup K_U contained in U, such that G/K_U is a Z-group. Moore has shown [25, Proposition 2.2 and Remark]:

$$E(G) = \bigcup_{U \in \mathfrak{U}} E(G/K_U).$$

$S_U = S \cap E(G/K_U)$ is a subhypergroup of $E(G/K_U)$, but it isn't clear straight away, whether S_U is faithful or not. But, since we are in the Z-group case, there are normal subgroups of G, such that $K_U \subseteq N_U$ and $S_U = E(G/N_U)$. Because of $S = \bigcup_{U \in \mathfrak{U}} S_U$ we get $\{e\} = \bigcap_{U \in \mathfrak{U}} N_U$. Now for the closed normal subgroups $N_U \cdot P$ we have:

$$S_{N_U \cdot P} \subseteq S_{N_U} \cap S_P = S_U \cap E(G/P) \subseteq S \cap E(G/P) = \{1\},$$

thus $N_U \cdot P = G$ for every $U \in \mathfrak{U}$. Let x be any element of G; there exist $n_U \in N_U$ and $x_U \in P$, such that $x = n_U x_U$. Since P is compact, we may assume that $(x_U)_{U \in \mathfrak{U}}$ converges. Hence $(n_U)_{U \in \mathfrak{U}}$ converges too, and that against e, since $\bigcap_{U \in \mathfrak{U}} N_U = \{e\}$. But that means: $x = \lim n_U \cdot x_U = \lim x_U \in P$.

5.7. **Lemma.** 5.2. *Theorem is true for any compactly generated FC-Moore-group.*

Proof. Sticking to the notation of 5.6 we have G/P abelian. Then because of [21, Lemma 4]:

$$E(G)=\bigcup_{\alpha\in\Delta}\alpha\cdot E(G/P));$$

here $\Delta\subseteq E(G)$ is choosen in such a way, that the map $r: E(G)\to E(P, G)$ restricted to Δ is a bijection, and that $\alpha\in\Delta$ is in S in case $\alpha\cdot E(G/P)\cap S\neq\phi$. Since $\lambda\in\operatorname{supp}\mu_{\alpha\beta}\subseteq S$ for any $\beta=\alpha\cdot\lambda\in\alpha\cdot E(G/P)\cap S$, we have

$$S=\bigcup_{\alpha\in\Delta\cap S}\alpha\cdot(S\cap E(G/P)\,;$$

we write S_1 for the subhypergroup $S\cap E(G/P)$ of $E(G/P)$. G/P is abelian, thus $S_1=E(G/L)$ for $L=N_{S_1}\supseteq P$. Now look at

$$r_1: E(G)\to E(L, G), r_2: E(L, G)\to E(P, G) \quad \text{and} \quad s: E(L)\to E(L, G).$$

$E(P, G)$ is discrete, since P is compact, r_2 is continuous and injective on $r_1(S)$; hence $r_1(S)$ is closed and the lemmata 5.3 and 5.4 apply: $T=s^{-1}(r_1(S))$ is a faithful subhypergroup of $E(L)$. Moreover L and G have the same compact elements and $T\cap E(L/P)=\{1\}$. Take for that $\alpha\in T\cap E(L/P)$; then there exists $\beta\in S$, such that $s(\alpha)=r_1(\beta)$, and for that reason $\beta\in S_1$. Thus $r_1(\beta)=1$ and $\alpha\in s^{-1}(r_1(\beta))=\{1\}$. Lemma 5.6 yields $L=P$. On the one hand this means: $S_1=E(G/P)$; on the other: $T=E(L)$, since $L=P$ is compact and T is faithful. Therefore we get $r(S)=s(T)=E(P, G)$, i.e. $\Delta\subseteq S$, which the assertion follows from.

5.8. **Lemma.** 5.2 *is true for any FC-Moore-group.*

Proof. Since $r: E(G)\to E(N, G)$ is closed for every open normal subgroup N of G (3.3 Theorem and Remark), it follows, that $s^{-1}(r(S))$ is a faithful subhypergroup of $E(N)$, if S is such a one of $E(G)$. If in addition N is compactly generated, we have $r(S)=E(N, G)$ by 5.7. If we take for any $\alpha\in E(G)$ and any neighbourhood U of α an open, compactly generated normal subgroup N, such that $r(\alpha)=r(\beta)$ implies $\beta\in U$, we see that S is dense in $E(G)$, hence equals it.

5.9. *Proof of the Theorem.* Look at $r: E(G)\to E(G_F, G)$ and $s: E(G_F)\to E(G_F, G)$. For any faithful quasi-subhypergroup S of $E(G)$ it turns out, that $s^{-1}(r(S))$ is a faithful subhypergroup of $E(G_F)$, since r is closed (3.3 Theorem and Remark). Thus $r(S)=E(G_F, G)$. In 4.3 Lemma we have allready mentioned a dense subset of $E(G_F, G)$, namely $E=\{\gamma\in E(G_F, G);\ \tilde{\gamma}\in E(G)\}$. For every $\gamma\in E$ we have: $r^{-1}(\gamma)=\{\tilde{\gamma}\}$ [12, Korollar p. 243]. Thus $U=\{\tilde{\gamma};\ \gamma\in E\}$ is contained in S, but it is dense, since r is open. This proves our theorem.

5.10. Now we want to describe the hypercosets as is done for compact groups in [6, Proposition 3.16] and in a special case for Z-groups in [29, 3.7 Proposition].

Theorem. *Let G be a Moore-group, $S=E(G/N)$ a subhypergroup of $E(G)$ and $\alpha\in E(G)$. Then we have*

$$\alpha * S=\{\beta\in E(G);\ \beta|N=\alpha|N\}.$$

Proof. For $\beta\in S$ we have $\alpha|N=\alpha|N\cdot\beta|N=\sum_{\gamma\in E(G)}\mu_{\alpha\beta}(\gamma)\,\gamma|N$, and therefore $\alpha|N=\gamma|N$ for every $\gamma\in\operatorname{supp}\mu_{\alpha\beta}$, since $\alpha|N\in E(N, G)$, and that means $\alpha * S\subseteq\{\beta\in E(G);\ \beta|N=\alpha|N\}$.

The other inclusion is proved by passing to the Bohr compactification. Let α and β be such that $\alpha|N=\beta|N$. α resp. β possess unique continuous extensions α' resp. β' to the Bohr compactification bG of G. Since in this case the closure of N in bG may be regarded as the Bohr compactification bN [28, (2.3) Satz and (3.5) Korollar], we have $\alpha'|bN=\beta'|bN$. Thus [4, Proposition 3.16] applies putting $H=E(bG)$ and $K=E(bG|bN)$, i.e. there exists $\gamma'\in E(bG|bN)$, such that $\mu_{\alpha'\beta'}(\gamma')>0$. But then $\mu_{\alpha\beta}(\gamma)>0$ is valid for $\gamma=\gamma'|G$ (compare [12, Proof of Lemma 5]). Thus we get $\beta\in\alpha*S$.

5.11. *Remark.* If we equip the set of hypercosets [with respect to a subhypergroup S_N of $E(G)$] with the quotient-topology, it is homeomorphic to $E(N, G)$, since we have $\alpha*S_N=r^{-1}r(\alpha)$, where $r: E(G)\to E(N, G)$ is a quotient-map. In the abelian case one obtains in this manner the dual of N, namely $\hat{N}=E(N)$ [17, (24.11) Theorem]. In the case of a Moore-group we get the dual "almost":

There is a continuous, open, proper surjection (finite-to-one) from $E(N)$ to the set of hypercosets (3.3 Theorem).

5.12. At the end we obtain results corresponding to those in [31, Theorem 9, 10; 32, Theorem 4].

Theorem. *Let G be a Moore-group. $N\to S_N$ is a lattice-antiisomorphism between the lattice of closed normal subgroup of G and that of subhypergroups of $E(G)$, whose inverse is $S\to N_S$.*

Proof. In view of the main theorem (5.2) there exists for a given family $(S_i)_{i\in I}$ of subhypergroups a smallest one containing these, namely $S_{N_{\cup S_i}}$. On the other side one can easily show, that $\cap S_i=S_{\cup N_{S_i}}$ is valid; thus $\cap S_i$ is a subhypergroup. Hence $VS_i=S_{N_{\cup S_i}}$ and $\wedge S_i=\cap S_i$ define lattice operations. Furthermore we get for a family $(N_i)_{i\in I}$ of closed normal subgroups the equality: $\cap N_i=N_{\cup S_{N_i}}$. Thus the following is valid:

$$VN_{S_i}=N_{S_{\cup N_{S_i}}}=N_{\cap S_i},\quad \text{hence}\quad \cap S_{N_i}=S_{N_{\cap S_{N_i}}}=S_{VN_{S_{N_i}}}=S_{VN_i}$$

as well as:

$$VS_{N_i}=S_{N_{\cup S_{N_i}}}=S_{\cap N_i},\quad \text{thus}\quad \cap N_{S_i}=N_{S_{\cap N_{S_i}}}=N_{VS_{N_{S_i}}}=N_{VS_i}.$$

5.13. *Remark.* As in the abelian case we get: N is open if and only if S_N is compact, and N is compact if and only if S_N is open.

Proof. N is open if and only if G/N is discrete, which is for SIN-groups equivalent to the compactness of $\hat{G/N}=S_N$ [19, Theorem 2 (1)]. For compact N we get from [22, Lemma 5.2], that S_N is open, while the other implication follows from [20, Theorem 1].

References

1. Alfsen, E.: Compact convex sets and boundary integrals. Berlin, Heidelberg, New York: Springer 1971
2. Bauer, H.: Wahrscheinlichkeitstheorie. Berlin: de Gruyter 1968
3. Blattner, R.: On induced representations. Am. J. Math. **83**, 79–98 (1961)
4. Boidol, J., Leptin, H., Schürmann, J., Vahle, D.: Räume primitiver Ideale von Gruppenalgebren. Math. Ann. **236**, 1–13 (1978)
5. Dixmier, J.: Les C^*-algèbres et leurs représentations. Paris: Gauthier-Villars 1964

6. Dunkl, C.F.: The measure algebra of a locally compact hypergroup. Trans. Amer. Math. Soc. **179**, 331–348 (1973)
7. Effros, E.G.: A decomposition theory for representations of C^*-algebras. Trans. Am. Math. Soc. **107**, 83–106 (1963)
8. Fell, J.M.G.: The structure of algebras of operator fields. Acta Math. **106**, 233–280 (1961)
9. Grosser, S., Mosak, R.D., Moskowitz, M.: Duality and harmonic analysis on central topological groups. Indag. Math. **35**, 65–91 (1973)
10. Grosser, S., Moskowitz, M.: Compactness conditions in topological groups. J. Reine Angew. Math. **246**, 1–40 (1967)
11. Hartmann, K., Henrichs, R.W., Lasser, R.: Duals of orbit spaces in groups with relatively compact inner automorphism groups are hypergroups. Mh. Math. **88**, 229–238 (1979)
12. Hauenschild, W., Kaniuth, E.: Harmonische Analyse auf Gruppen mit endlich-dimensionalen irreduziblen Darstellungen. Math. Z. **144**, 239–256 (1975)
13. Hauenschild, W., Kaniuth, E.: Ideale mit beschränkter approximativer Eins in Gruppenalgebren von Moore-Gruppen. Manuscripta Math. **27**, 167–182 (1979)
14. Helgason, S.: Lacunary Fourier series on noncommutative groups. Proc. Am. Math. Soc. **9**, 782–790 (1958)
15. Henrichs, R.W.: Über Fortsetzung positiv-definiter Funktionen. Math. Ann. **232**, 131–150 (1978)
16. Henrichs, R.W.: Weak Frobenius reciprocity and compactness conditions in topological groups. Pacific J. Math. **82**, 387–406 (1979)
17. Hewitt, E., Ross, K.A.: Abstract harmonic analysis. I, II. Berlin, Heidelberg, New York: Springer 1963, 1969
18. Jewett, R.I.: Spaces with an abstract convolution of measures. Adv. Math. **18**, 1–101 (1975)
19. Kaniuth, E.: Topology in duals of SIN-groups. Math. Z. **134**, 67–80 (1973)
20. Kaniuth, E.: A note on reduced duals of certain locally compact groups. Math. Z. **150**, 189–194 (1976)
21. Kaniuth, E., Schlichting, G.: Zur harmonischen Analyse klassenkompakter Gruppen. II. Invent. Math. **10**, 332–345 (1970)
22. Lipsman, R.: Representation theory of almost connected groups. Pacific J. Math. **42**, 453–467 (1972)
23. Liukkonen, J.R.: Dual spaces of groups with relatively compact conjugacy classes. Trans. Am. Math. Soc. **180**, 85–108 (1973)
24. Ludwig, J.: Good ideals in the group algebra of a nilpotent Lie group. Math. Z. **161**, 195–210 (1978)
25. Moore, C.C.: Groups with finite dimensional irreducible representations. Trans. Am. Math. Soc. **166**, 401–410 (1972)
26. Mosak, R.D.: The L^1- and C^*-algebras of $[\mathrm{FIA}]_B^-$-groups and their representations. Trans. Am. Math. Soc. **163**, 277–310 (1972)
27. Palmer, T.W.: Classes of nonabelian, noncompact, locally compact groups. Rocky Mountain J. Math. **8**, 683–741 (1978)
28. Poguntke,D.: Zwei Klassen lokalkompakter maximal fastperiodischer Gruppen. Mh. Math. **81**, 15–40 (1976)
29. Ross, K.A.: Centers of hypergroups. Trans. Am. Math. Soc. **243**, 251–269 (1978)
30. Spector, R.: Aperçu de la théorie des hypergroupes. In: Analyse Harmonic sur les Groupes de Lie, Séminaire Nancy-Strasbourg 1973–75. Lecture Notes in Mathematics, Vol. 497. Berlin, Heidelberg, New York: Springer 1975
31. Takesaki, M., Tatsuuma, N.: Duality and subgroups. Ann. Math. **93**, 344–364 (1971)
32. Takesaki, M., Tatsuuma, N.: Duality and subgroups. II. J. Functional Analysis **11**, 184–190 (1972)
33. Thoma, E.: Über unitäre Darstellungen abzählbarer, diskreter Gruppen. Math. Ann. **153**, 111–138 (1964)
34. Thoma, E.: Über positiv-definite Klassenfunktionen abzählbarer Gruppen. Math. Z. **84**, 389–402 (1964)

Received June 11, 1980

Math. Ann. 256, 19–28 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Spectrality Criteria for Unbounded Operators with Real Spectrum

Shmuel Kantorovitz

Forschungsinstitut für Mathematik, ETH, CH-8000 Zürich, Schweiz and Bar-Ilan University, Ramat-Gan, Israel

Theorem. *T is a scalar-type spectral operator if and only if it is closed with dense domain $D(T)$ and* $\sup_{n\geqq 0}\|R^n\|_B<\infty$.

For each $x\in X$, let $|||x|||=\sup_{n\geqq 0}\|R^n x\|_B$, $Z=\{x\in X;\,|||x|||<\infty\}$, *and*

$$D(T|Z)=\{x\in D(T);\,x,\,Tx\in Z\}.$$

If T is a group generator, it is shown that its restriction to $D(T|Z)$ possesses a spectral resolution analogous to that of scalar-type spectral operators.

Introduction

Let T be a possibly unbounded linear operator with domain $D(T)$ in a Banach space X. Suppose there exists a (strongly) countably additive spectral measure E on the Borel σ-field $\mathfrak{B}(\mathbb{R})$ of the real line $\mathbb{R}$, such that

$$D(T)=\left\{x\in X;\,\lim_{n\to\infty}\int_{-n}^{n}sE(ds)x \text{ exists}\right\},\quad \text{and } Tx \text{ equals the above limit for } x\in D(T),$$

where limits are always strong limits in X. [Necessarily, T is then closed, densely defined, and has spectrum $\sigma(T)$ lying in $\mathbb{R}$.] Such operators T, which include the self-adjoint operators in Hilbert space, are called scalar-type spectral operators (cf. [2] as a standard reference).

Continuing the study we started in [7], we are looking for criteria of scalar-type spectrality involving adequate functions of the resolvent $R(\lambda;T)$ of T. In [7], we considered the case when $\sigma(T)$ is contained in $[0,\infty)$. We shall presently lift this restriction on the spectrum, and obtain a criterion analogous to the Hille-Yosida criterion for semi-group generation. Introducting the "Bochner norm" $\|\cdot\|_B$ on vector-valued functions defined on $\mathbb{R}$ and letting $R(t)=(I-itT)^{-1}$, our characteri-

* Research partially done while the author was a visiting member at the Forschungsinstitut für Mathematik, ETH, Zürich. Research supported by the Israel Academy of Science, Bar-Ilan University Research Authority and the National Science Foundation

0025-5831/81/0256/0019/02.00

zation of scalar-type spectral operators T with real spectrum in reflexive Banach spaces is the following: T is closed, densely defined, and

$$\sup_{n\geqq 0}\|R^n\|_B<\infty \quad \text{(Theorem 1.2)}.$$

In the spirit of [5, 6], we obtain also a "local" version of this result (Theorem 2.6).

We conclude this introduction with some notational conventions. We abbreviate the expression "all parameters involved" by a.p.i. If X and Y are Banach spaces, $B(X, Y)$ denotes the Banach space of all bounded linear operators from X to Y with the operator norm. We write $B(X)$ for $B(X,X)$.

$\mathbb{B}(\mathbb{R})$ denotes the Banach algebra of all bounded complex Borel functions on $\mathbb{R}$ with pointwise operations and supremum norm $\|\cdot\|_\infty$; it contains as a dense subalgebra the linear span$\mathbb{B}_0(\mathbb{R})$ of all characteristic function χ_δ, $\delta\in\mathfrak{B}(\mathbb{R})$. If m is a (strongly) countably additive Y-valued measure on $\mathfrak{B}(\mathbb{R})$, the map $h\to\int_{\mathbb{R}} h(s)m(ds)$ is a well-defined element of $B(\mathbb{B}(\mathbb{R}), Y)$ with norm denoted by $v(m)$, that is

$$\left\|\int_{\mathbb{R}} h(s)m(ds)\right\|\leqq\|h\|_\infty v(m), \qquad h\in\mathbb{B}(\mathbb{R}).$$

A $B(X)$-valued strongly countably additive measure E on $\mathfrak{B}(\mathbb{R})$ is called a spectral measure if $E(\mathbb{R})=I$ (the identity operator) and $E(\sigma\cap\delta)=E(\sigma)E(\delta)$ for all $\sigma,\delta\in\mathfrak{B}(\mathbb{R})$. For spectral measures, the map $h\to\int_{\mathbb{R}} h(s)E(ds)$ is a continuous algebra homomorphism of $\mathbb{B}(\mathbb{R})$ into $B(X)$.

If A, B are unbounded operators in X and $U\in B(X)$, then $A\subset B$ means $D(A)\subset D(B)$ and $Ax=Bx$ for $x\in D(A)$; $A=B$ means $A\subset B\subset A$; U commutes with A if $UA\subset AU$.

1. The Spectrality Criterion

Let Ω denote the set of all vectors

$$(\mathbf{c};\mathbf{t})=(c_1,\ldots,c_m;\,t_1,\ldots,t_m)\in\mathbb{C}^m\times\mathbb{R}^m$$

with m varying ($m=1,2,\ldots$), such that

$$\left\|\sum_{k=1}^{m} c_k\exp(it_k s)\right\|_\infty\leqq 1$$

For any function F defined on $\mathbb{R}$ with values in a Banach space Y, we let

$$\|F\|_B=\sup_{\Omega}\|\textstyle\sum c_k F(t_k)\|, \tag{1.1}$$

where $\sup_{\Omega}$ stands for the supremum over all vectors $(\mathbf{c},\mathbf{t})$ in Ω.

The set of all $F:\mathbb{R}\to Y$ with $\|F\|_B<\infty$ is a vector space $\mathfrak{F}(Y)$ for the pointwise operations. The following elementary properties are easily verified:

(a) $\mathfrak{F}(Y)$ contains the constant functions $F(t)=C$ ($C\in Y$) and $\|C\|_B=\|C\|$.

(b) $\|F\|_B\geqq\|F\|_\infty=\sup_{\mathbb{R}}\|F(t)\|$.

(c) $(\mathfrak{F}(Y),\|\cdot\|_B)$ is a Banach space.

(d) The norm $\|\cdot\|_B$ on $\mathfrak{F}(Y)$ is invariant under additive translations $F(t)\to F(t+c)$ ($c\in\mathbb{R}$) and multiplicative translations $F(t)\to F(ct)$ ($0\neq c\in\mathbb{R}$).

(e) For any $U\in B(Y_1, Y_2)$, $\|UF\|_B\leqq\|U\|\,\|F\|_B$, so that $U\mathfrak{F}(Y_1)\subset\mathfrak{F}(Y_2)$.

(f) If $F:\mathbb{R}\to B(X, Y)$ for Banach spaces X and Y, then the following three statements are equivalent:

(i) $F\in\mathfrak{F}(B(X, Y))$;

(ii) $Fx\in\mathfrak{F}(Y)$ for each $x\in X$;

(iii) $y^*Fx\in\mathfrak{F}(\mathbb{C})$ for each $x\in X$, $y^*\in Y^*$.

In addition, if X_1 and Y_1^* denote the unit balls of X and Y^* resp.

$$\begin{aligned}\|F\|_B&=\sup\{\|Fx\|_B;\ x\in X_1\}\\&=\sup\{\|y^*Fx\|_B;\ x\in X_1,\ y^*\in Y_1^*\}.\end{aligned}$$

Property (f) follows from the Uniform Boundedness Theorem.

If T is a possibly unbounded linear operator in the Banach space X with spectrum $\sigma(T)$ contained in $\mathbb{R}$, the function

$$R(t)=(I-itT)^{-1} \tag{1.2}$$

is a well-defined $B(X)$-valued function on $\mathbb{R}$. We let

$$V_T=\sup_{n\geqq 0}\|R^n\|_B. \tag{1.3}$$

Note that $V_T\geqq\|I\|_B=\|I\|=1$ and $V_{cT}=V_T$ for $0\neq c\in\mathbb{R}$.

Lemma 1.1. *Let T be a possibly unbounded linear operator in a Banach space X. Then T is closed with dense domain, real spectrum, and $V_T<\infty$ if and only if iT generates a strongly continuous group of operators $\{G(t);\ t\in\mathbb{R}\}$ with $\|G\|_B<\infty$. In this case, $V_T=\|G\|_B$.*

Proof. 1. If T has real spectrum, then V_T is certainly well-defined; assume $V_T<\infty$. Then, for $0\neq\lambda\in\mathbb{R}$ and $n\geqq 1$,

$$\|[\lambda R(\lambda;iT)]^n\|=\|R^n(\lambda^{-1})\|\leqq\|R^n\|_B\leqq V_T<\infty.$$

Therefore, if T is also closed with dense domain, the Hille-Yosida theorem (cf. [4, Theorem 12.3.2]) implies that iT generates a strongly continuous group of operators $\{G(t);t\in\mathbb{R}\}$.

By [4, Theorem 11.6.6],

$$G(t)x=\lim_{n\to\infty}R^n(t/n)x\quad(t\in\mathbb{R}, x\in X).$$

Therefore, for each $(\mathbf{c}, \mathbf{t})\in\Omega$ and $x\in X$,

$$\sum c_kG(t_k)x=\lim_{n\to\infty}\sum c_kR^n(t_k/n)x.$$

The sums on the right have norms bounded by

$$\|R^n(t/n)x\|_B=\|R^nx\|_B,$$

so that

$$\|Gx\|_B\leqq\sup_{n\geqq 0}\|R^nx\|_B$$

and $\|G\|_B\leqq V_T$ by Property (f).

2. Conversely, if iT generates a strongly continuous group of operators $\{G(t); t\in\mathbb{R}\}$ with $\|G\|_B<\infty$, then since $\|G\|_\infty\leqq\|G\|_B$, we have

$$\lim_{t\to\infty} t^{-1}\log\|G^{\pm}(t)\|\leqq 0,$$

where $G^{\pm}(t)=G(\pm t)$ $(t\geqq 0)$ are the semi-groups with generators $\pm iT$ respectively. By [4, p. 622], $\sigma(\pm iT)$ is contained in the closed left half-plane. Hence $\sigma(iT)$ is pure imaginary and $\sigma(T)\subset\mathbb{R}$. As a (semi)group generator, iT is closed with dense domain.

We now use the formula

$$R(\lambda;\pm iT)^n x=\Gamma(n)^{-1}\int_0^\infty e^{-\lambda t}t^{n-1}G^{\pm}(t)x\,dt,$$

valid for all $x\in X$, $n\geqq 1$ and $\lambda>0$, cf. [4, formula (5.8.5) and Theorem 11.5.2]. Since

$$R^n(\pm 1/\lambda)=[\lambda R(\lambda;\pm iT)]^n,$$

a simple calculation leads to the formula

$$R^n(t)x=\Gamma(n)^{-1}\int_0^\infty e^{-s}s^{n-1}G(ts)x\,ds \tag{1.4}$$

valid for all $x\in X$, $n\geqq 1$, and $t\in\mathbb{R}$.

Since $\|G(ts)x\|_B=\|G(t)x\|_B\leqq\|G\|_B\|x\|<\infty$ for each fixed $s>0$ [Property (d) of $\|\cdot\|_B$], we obtain for all $x\in X$ and $(\mathbf{c},\mathbf{t})\in\Omega$,

$$\begin{aligned}\|\textstyle\sum c_k R^n(t_k)x\| &= \Gamma(n)^{-1}\left\|\int_0^\infty e^{-s}s^{n-1}\sum c_k G(t_k s)x\,dx\right\|\\ &\leqq\Gamma(n)^{-1}\int_0^\infty e^{-s}s^{n-1}\|G(ts)x\|_B ds\\ &=\|Gx\|_B.\end{aligned}$$

Hence $\sup_{n\geqq 0}\|R^n x\|_B\leqq\|Gx\|_B$ since trivially

$$\|R^0 x\|_B=\|x\|=\|G(0)x\|\leqq\|Gx\|_\infty\leqq\|Gx\|_B,$$

and therefore $V_T\leqq\|G\|_B<\infty$ [cf. Property (f)]. Parts 1 and 2 of the proof imply the equality $V_T=\|G\|_B$.

We now state the main theorem of this section.

Theorem 1.2. *Let T be a possibly unbounded linear operator with real spectrum, acting in a reflexive Banach space X. Then T is a scalar-type spectral operator if and only if it is closed, densely defined, and $V_T<\infty$.*

Proof. Necessity. Let E be the resolution of the identity for T. Then

$$R^n(t)=\int_{\mathbb{R}}(1-its)^{-n}E(ds)\quad(t\in\mathbb{R}, n\geqq 0).$$

Let

$$f_n(u)=\Gamma(n)^{-1}u^{n-1}e^{-u}\quad(u\geqq 0; n=1,2,\ldots).$$

Then

$$\int_0^\infty e^{ivu} f_n(u)du = (1-iv)^{-n} \quad (v \in \mathbb{R})$$

(see e.g. [3, p. 152]).

Hence, for all $(\mathbf{c}, \mathbf{t}) \in \Omega$ and $n \geqq 1$,

$$\sum c_k R^n(t_k) = \int_{\mathbb{R}} \int_0^\infty \sum c_k \exp(it_k su) \cdot f_n(u)du E(ds).$$

The sum in the integrand in bounded by 1 for a.p.i.; $f_n \geqq 0$ and $\int_0^\infty f_n(u)du = 1$. It follows that the expression above has norm $\leqq v(E)$, and therefore $V_T \leqq v(E) < \infty$.

The fact that spectral operators are closed and densely defined is well-known. Observe that the reflexivity assumption is not needed for the necessity part of the theorem.

Sufficiency. By Lemma 1.1, iT generates a strongly continuous group of operators $\{G(t); t \in \mathbb{R}\}$ with $\|G\|_B = V_T < \infty$. For each $x \in X$ and $x^* \in X^*$, the complex valued function $x^*G(\cdot)x$ is continuous on $\mathbb{R}$ and satisfies the "Bochner criterion" (cf. [1])

$$\left| \sum_{k=1}^m c_k x^* G(t_k) x \right| \leqq \|G\|_B \|x\| \, \|x^*\| \left\| \sum_{k=1}^m c_k \exp(it_k s) \right\|_\infty$$

for all vectors $(\mathbf{c}; \mathbf{t}) \in \mathbb{C}^m \times \mathbb{R}^m$ $(m = 1, 2, \ldots)$. There exists therefore an unique regular complex Borel measure $\mu(\cdot | x, x^*)$ on $\mathfrak{B}(\mathbb{R})$ such that

$$x^*G(t)x = \int_{\mathbb{R}} e^{its} \mu(ds | x, x^*)$$

and

$$\operatorname{var} \mu(\cdot | x, x^*) \leqq \|G\|_B \|x\| \, \|x^*\| .$$

Proceeding in the routine way (using uniqueness, reflexivity, etc. ...), we obtain a unique function $E: \mathfrak{B}(\mathbb{R}) \to B(X)$ with the following properties:

(i) E is (strongly) countably additive;

(ii) $\mu(\cdot | x, x^*) = x^* E(\cdot) x$;

(iii) $G(t) = \int_{\mathbb{R}} e^{its} E(ds)$ [usual meaning; cf. (i)].

Taking $t = 0$, we have in particular $E(\mathbb{R}) = I$. If $U \in B(X)$ commutes with T $(UT \subset TU)$, it is well known that U commutes then with $G(t)$ for all $t \in \mathbb{R}$. Therefore, by uniqueness of the representation (iii), we obtain that U commutes with $E(\delta)$ for all $\delta \in \mathfrak{B}(\mathbb{R})$. Since for a.p.i.

$$\begin{aligned} \int_{\mathbb{R}} e^{ius} \mu(ds | G(t)x, x^*) &= x^* G(u) G(t) x = x^* G(u+t) x \\ &= \int_{\mathbb{R}} e^{ius} e^{its} \mu(ds | x, x^*), \end{aligned}$$

we have (uniqueness for Fourier-Stieltjes transforms)

$$\mu(\delta | G(t)x, x^*) = \int_{\mathbb{R}} e^{its} \chi_\delta(s) \mu(ds | x, x^*) \quad \text{(a.p.i.)}.$$

But the left hand side is equal to

$$x^*E(\delta)G(t)x = x^*G(t)E(\delta)x$$
$$= \int_{\mathbb{R}} e^{its}\mu(ds|E(\delta)x, x^*) \quad \text{(a.p.i.)},$$

hence (uniqueness for Fourier-Stieltjes transforms)

$$x^*E(\sigma)E(\delta)x = \mu(\sigma|E(\delta)x, x^*) = \int_{\mathbb{R}} \chi_\sigma(s)\chi_\delta(s)\mu(ds|x, x^*)$$
$$= \mu(\sigma\cap\delta|x, x^*) = x^*E(\sigma\cap\delta)x \quad \text{(a.p.i.)},$$

that is $E(\sigma)E(\delta) = E(\sigma\cap\delta)$ for all $\sigma, \delta \in \mathfrak{B}(\mathbb{R})$. Thus E is a spectral measure on $\mathfrak{B}(\mathbb{R})$.

For each $x\in X$, $x^*\in X^*$, and $\lambda > 0$,

$$x^*R(\lambda; iT)x = x^* \int_0^\infty e^{-\lambda t}G(t)x\, dt$$
$$= \int_0^\infty e^{-\lambda t} \int_{\mathbb{R}} e^{its}x^*E(ds)x\, dt.$$

Since

$$\int_0^\infty \int_{\mathbb{R}} e^{-\lambda t}|x^*E(ds)x|dt = \lambda^{-1}\operatorname{var}(x^*Ex) < \infty,$$

Fubini's theorem implies that

$$x^*R(\lambda; iT)x = \int_{\mathbb{R}} \int_0^\infty e^{-\lambda t}e^{its}dt\, x^*E(ds)x$$
$$= x^* \int_{\mathbb{R}} (\lambda - is)^{-1}E(ds)x \quad \text{(a.p.i.)},$$

i.e.,

$$R(\lambda, iT) = \int_{\mathbb{R}} (\lambda - is)^{-1}E(ds)$$

for all $\lambda > 0$, and similarly for all $\lambda < 0$.

Let $x\in D(T) = D(iT)$. Then $x = R(\lambda; iT)y$ for $\lambda > 0$ fixed and $y = \lambda x - iTx$. Since the function $s(\lambda - is)^{-1}$ is bounded on $\mathbb{R}$,

$$\int_{-n}^{n} sE(ds)x = \int_{-n}^{n} s(\lambda - is)^{-1}E(ds)y$$
$$\xrightarrow[n\to\infty]{} \int_{-\infty}^{\infty} s(\lambda - is)^{-1}E(ds)y$$
$$= i \int_{-\infty}^{\infty} [1 - \lambda(\lambda - is)^{-1}]E(ds)y$$
$$= i[y - \lambda R(\lambda; iT)y] = i[y - \lambda x] = Tx,$$

that is

$$D(T) \subset \left\{x\in X \mid \lim_{n\to\infty} \int_{-n}^{n} sE(ds)x \text{ exists}\right\} \tag{1.5}$$

and

$$Tx = \lim_{n\to\infty} \int_{-n}^{n} sE(ds)x, \qquad x \in D(T).$$

On the other hand, if y belongs to the set on the right of (1.5), then denoting

$$z = \lim_{n\to\infty} \int_{-n}^{n} sE(ds)y,$$

we have as before (for $\lambda > 0$ fixed)

$$\begin{aligned} R(\lambda; iT)z &= \lim_{n\to\infty} R(\lambda; iT) \int_{-n}^{n} sE(ds)y \\ &= \int_{-\infty}^{\infty} s(\lambda - is)^{-1} E(ds)y \\ &= i[y - \lambda R(\lambda; iT)y]. \end{aligned}$$

Hence $y = R(\lambda; iT)[\lambda y - iz] \in D(iT) = D(T)$, proving equality in (1.5), Q.E.D.

Note that the proof shows also that $V_T = v(E)$.

2. The Semi-Simplicity Manifold

In this section, T will be a fixed possibly unbounded linear operator in a Banach space X, with spectrum lying in $\mathbb{R}$. As before,

$$R(t) = (I - itT)^{-1} \quad (t \in \mathbb{R}).$$

Definition 2.1. Let

$$|||x||| = \sup_{n \geqq 0} \|R^n x\|_B \quad (x \in X).$$

The *semi-simplicity manifold for* T is the set

$$Z = \{x \in X;\ |||x||| < \infty\}.$$

Lemma 2.2. (i) *Z is a linear manifold in X and $|||\cdot|||$ is a norm on Z which majorizes the given norm;* (ii) *$(Z, |||\cdot|||)$ is a Banach space;* (iii) *if $U \in B(X)$ commutes with T (that is $UT \subset TU$), then $UZ \subset Z$ and $|||Ux||| \leqq \|U\|\,|||x|||$ for all $x \in Z$.*

Proof. We show only the completeness of $(Z, |||\cdot|||)$. If $\{x_p\} \subset Z$ is $|||\cdot|||$-Cauchy, it is also $\|\cdot\|$-Cauchy. Let $x = \lim_{p\to\infty} x_p$ (in X). Given $\varepsilon > 0$, let p_0 be such that $|||x_p - x_q||| < \varepsilon$ for $p > q > p_0$. For each $(\mathbf{c}, \mathbf{t}) \in \Omega$ and $n \geqq 0$,

$$\|\sum c_k R^n(t_k)(x_p - x_q)\| < \varepsilon, \qquad p > q > p_0.$$

Keeping q fixed and letting $p \to \infty$, we obtain for a.p.i.

$$\|\sum c_k R^n(t_k)(x - x_q)\| \leqq \varepsilon, \qquad q > p_0,$$

that is $\|x - x_q\| \leqq \varepsilon$ for $q > p_0$.

Proposition 2.3. *The following statements are equivalent.*

(i) $Z=X$;

(ii) Z *is of the second category in* X;

(iii) $|||\cdot|||$ *is equivalent to* $\|\cdot\|$;

(iv) $V_T<\infty$;

(v) *(for* T *closed with dense domain)* T *is a scalar-type spectral operator.*

Proof. Apply the Uniform Boundedness Theorem and Theorem 1.2.

Proposition 2.4. *If* iT *generates a strongly continuous group of operators* $\{G(t);t\in\mathbb{R}\}$, *then*

$$|||x|||=\|Gx\|_B \quad (x\in X).$$

Proof. It suffices to verify the following statements:

(a) if $\|Gx\|_B<\infty$, then $|||x|||\leqq\|Gx\|_B$;

(b) if $|||x|||<\infty$, then $\|Gx\|_B\leqq|||x|||$.

This is explicitly done in the proof of Lemma 1.1.

Definition 2.5. The restriction $T|Z$ of T to its semi-simplicity manifold Z is the operator T with "maximal domain" in Z

$$D(T|Z)=\{x\in D(T);x, Tx\in Z\}.$$

The algebra of all linear operators in X with domain Z and range contained in Z shall be denoted by $\mathbb{T}(Z)$.

Theorem 2.6. *Suppose* iT *is the infinitesimal generator of a strongly continuous group of operators in a reflexive Banach space* X, *and let* Z *be the semi-simplicity manifold for* T. *Then there exists an algebra homomorphism* E *of the Boolean algebra* $\mathfrak{B}(\mathbb{R})$ *into* $\mathbb{T}(Z)$ *with the following properties* (1)–(4):

(1) $E(\mathbb{R})=I|Z$ *and for each* $x\in Z$, $E(\cdot)x$ *is a countably additive measure on* $\mathfrak{B}(\mathbb{R})$;

(2) *for each* $\delta\in\mathfrak{B}(\mathbb{R})$, $E(\delta)$ *commutes with every* $U\in B(X)$ *which commutes with* T;

(3) $D(T|Z)=\left\{x\in Z;\ \lim\limits_{n\to\infty}\int\limits_{-n}^{n}sE(ds)x \text{ exists and belongs to } Z\right\}$;

(4) $Tx=\lim\limits_{n\to\infty}\int\limits_{-n}^{n}sE(ds)x,\ x\in D(T|Z)$.

(The limits in (3) *and* (4) *are strong limits in* X.*)*

Proof. Let $\{G(t);t\in\mathbb{R}\}$ be the group generated by iT. We apply Proposition 2.4 and the Bochner criterion as in the proof of Theorem 1.2 to obtain a uniquely determined family $\{\mu(\cdot|x,x^*);x\in Z,x^*\in X^*\}$ of regular complex Borel measures on $\mathfrak{B}(\mathbb{R})$ such that

$$x^*G(t)x=\int_{\mathbb{R}}e^{its}\mu(ds|x,x^*) \tag{2.1}$$

and

$$\operatorname{var}\mu(\cdot|x,x^*)\leqq|||x|||\,\|x^*\| \tag{2.2}$$

for all $x\in Z$ and $x^*\in X^*$.

By reflexivity of X, we then obtain a uniquely determined family of X-valued countably additive functions $\{E(\cdot)x|x\in Z\}$ on $\mathfrak{B}(\mathbb{R})$ such that $\mu(\delta|x,x^*)=x^*E(\delta)x$

for all $x\in Z$ and $x^*\in X^*$. For each $\delta\in\mathfrak{B}(\mathbb{R})$, $E(\delta)$ is a linear operator with domain Z which commutes with each $U\in B(X)$ that commutes with T (see proof of Theorem 1.2). Taking $t=0$ in (2.1), we see that $E(\mathbb{R})=I|Z$.

For $h\in\mathbb{B}(\mathbb{R})$, define

$$\psi(h)x=\int_{\mathbb{R}} h(s)E(ds)x\,,\qquad x\in Z\,.$$

Since $E(\cdot)x$ is countably additive (for each $x\in Z$), $\psi(h)$ is a well-defined linear operator in X with domain Z, and $\|\psi(h)x\|\leqq\|h\|_\infty|||x|||$ for all $h\in\mathbb{B}(\mathbb{R})$ and $x\in Z$ [cf. (2.2)]. Furthermore, $\psi(h)$ commutes with every $U\in B(X)$ which commutes with T.

Lemma 2.7. $|||\psi(h)x|||\leqq\|h\|_\infty|||x|||$ $(x\in Z, h\in\mathbb{B}(\mathbb{R}))$. *In particular,* $\psi(h)Z\subset Z$.

Proof. Fix $x\in Z$ and $(\mathbf{c},\mathbf{t})\in\Omega$. For each $(\mathbf{c}',\mathbf{t}')\in\Omega$,

$$\begin{aligned}\Big\|\sum_j c'_jG(t'_j)\sum_k c_kG(t_k)x\Big\| &= \Big\|\sum_{j,k} c'_jc_kG(t'_j+t_k)x\Big\|\\ &\leqq |||x|||\,\Big\|\sum_{j,k} c'_jc_k\exp i(t'_j+t_k)s\Big\|_\infty\\ &\leqq |||x|||\,\Big\|\sum_j c'_j\exp it'_js\Big\|_\infty\Big\|\sum_k c_k\exp it_ks\Big\|_\infty\leqq|||x|||\,.\end{aligned}$$

Hence, by Proposition 2.4,

$$|||\textstyle\sum c_kG(t_k)x|||\leqq|||x|||\,,\qquad x\in Z\,,\qquad(\mathbf{c},\mathbf{t})\in\Omega\,.$$

Fix $x\in Z$ and $h\in\mathbb{B}(\mathbb{R})$. For each $(\mathbf{c},\mathbf{t})\in\Omega$,

$$\begin{aligned}\|\textstyle\sum c_kG(t_k)\psi(h)x\| &= \|\psi(h)\sum c_kG(t_k)x\|\\ &\leqq\|h\|_\infty|||\textstyle\sum c_kG(t_k)x|||\leqq\|h\|_\infty|||x|||\end{aligned}$$

[since $\sum c_kG(t_k)$ commutes with T]. Taking the supremum over all $(\mathbf{c},\mathbf{t})\in\Omega$, the lemma follows. □

Taking in particular $h=\chi_\delta$, we see that $|||E(\delta)x|||\leqq|||x|||$ for all $\delta\in\mathfrak{B}(\mathbb{R})$ and $x\in Z$. Thus E maps $\mathfrak{B}(\mathbb{R})$ into $\mathbb{T}(Z)$. The group property of G and the uniqueness theorem for Fourier-Stieltjes transforms imply (as in the proof of Theorem 1.2) that $E(\sigma\cap\delta)=E(\sigma)E(\delta)$ for all $\sigma,\delta\in\mathfrak{B}(\mathbb{R})$. Equivalently

$$E(ds)E(\delta)x=\chi_\delta(s)E(ds)x$$

for all $\delta\in\mathfrak{B}(\mathbb{R})$ and $x\in Z$.

Hence $\psi(h\chi_\delta)=\psi(h)\psi(\chi_\delta)$ for all $h\in\mathbb{B}(\mathbb{R})$ and $\delta\in\mathfrak{B}(\mathbb{R})$, and so $\psi(hg)=\psi(h)\psi(g)$ for all $h\in\mathbb{B}(\mathbb{R})$ and $g\in\mathbb{B}_0(\mathbb{R})$ [the subalgebra of simple functions in $\mathbb{B}(\mathbb{R})$].

For $h,g\in\mathbb{B}(\mathbb{R})$, choose $g_n\in\mathbb{B}_0(\mathbb{R})$ such that $\|g_n-g\|_\infty\to0$. Then by Lemma 2.7, for $x\in Z$,

$$\begin{aligned}\|\psi(hg)x-\psi(h)\psi(g)x\| &\leqq\|\psi[h(g-g_n)]x\|+\|\psi(h)\psi(g_n-g)x\|\\ &\leqq\|h(g-g_n)\|_\infty|||x|||+\|h\|_\infty|||\psi(g_n-g)x|||\\ &\leqq2\|h\|_\infty\|g-g_n\|_\infty|||x|||\xrightarrow[n\to\infty]{}0\,.\end{aligned}$$

We proved (cf. also Lemma 2.7):

Lemma 2.8. $\psi:\mathbb{B}(\mathbb{R})\to\mathbb{T}(Z)$ *is an algebra homomorphism, continuous with respect to the* $B(Z,|||\cdot|||)$*-topology on* $T(Z)$.

Lemma 2.9. *For any* $0 \neq t \in \mathbb{R}$, $D(T|Z) = R(t)Z$.

Proof. If $x \in D(T|Z)$, then $x \in Z \cap D(T)$ and $Tx \in Z$. Hence $z = (I - itT)x \in Z$, and therefore $x = R(t)z \in R(t)Z$ for any $t \in \mathbb{R}$. Conversely, if $x \in R(t)Z$ for some $0 \neq t \in \mathbb{R}$, say $x = R(t)z$ with $z = x - itTx \in Z$, then $x \in Z \cap D(T)$ (cf. Lemma 2.2) and $Tx = it^{-1}(z - x) \in Z$, that is $x \in D(T|Z)$. □ We may complete now the proof of Theorem 2.6. Let $x \in D(T|Z)$, so that $x = R(t)z$ with $z \in Z$ and $t \neq 0$ fixed. By Lemma 2.8,

$$\int_{-n}^{n} sE(ds)x = \int_{-n}^{n} s(1 - ist)^{-1} E(ds)z \xrightarrow[n \to \infty]{} \int_{\mathbb{R}} s(1 - ist)^{-1} E(ds)z. \tag{2.3}$$

By (2.1), (1.4) with $n = 1$, and Fubini's theorem,

$$R(t)z = \int_{\mathbb{R}} (1 - ist)^{-1} E(ds)z \tag{2.4}$$

for all $t \in \mathbb{R}$ and $z \in Z$.

Writing $s(1 - ist)^{-1} = it^{-1}[1 - (1 - ist)^{-1}]$ for $t \neq 0$, the integral in (2.3) is seen to equal

$$it^{-1}[z - R(t)z] = it^{-1}(z - x) = Tx \in Z.$$

Thus the inclusion $\subset$ in (3) and Relation (4) are proved (the numbers refer to the statement of Theorem 2.6).

On the other hand, if x belongs to the set on the right of (3), then denoting the limit in (3) by $z \in Z$, we have by Property (2), (2.4), and Lemma (2.8) (for $t \neq 0$):

$$R(t)z = \lim_{n \to \infty} \int_{-n}^{n} sE(ds)R(t)x = \lim_{n \to \infty} \int_{-n}^{n} s(1 - its)^{-1} E(ds)x$$
$$= \int_{\mathbb{R}} s(1 - its)^{-1} E(ds)x = it^{-1}[x - R(t)x]$$

(cf. preceding calculation). Therefore

$$x = R(t)(x - itz) \in R(t)Z = D(T|Z)$$

by Lemma 2.9, and the proof of Theorem 2.6 is complete.

References

1. Bochner, S.: A theorem on Fourier-Stieltjes integrals. Bull. Am. Math. Soc. **40**, 271–276 (1934)
2. Dunford, N., Schwartz, J.T.: Linear operators. Part III: Spectral operators. New York: Wiley 1971
3. Fisz, M.: Probability theory and mathematical statistics. New York: Wiley 1963
4. Hille, E., Phillips, R.S.: Functional analysis and semi-groups. Am. Math. Soc. Coll. Publ. **31** (1957)
5. Kantorovitz, S.: The semi-simplicity manifold of arbitrary operators. Trans. Am. Math. Soc. **123**, 241–252 (1966)
6. Kantorovitz, S.: Characterization of C^n-operators. Indiana Univ. Math. J. **25**, 119–133 (1976)
7. Kantorovitz, S.: Characterization of unbounded spectral operators with spectrum in a half-line. Comm. Math. Helv. (in press)

Received June 30, 1980

Math. Ann. 256, 29–35 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Selfadjoint Means and Operator Monotone Functions

Frank Hansen*

Research Institute for Mathematical Sciences, Kyoto University, Kyoto 606, Japan

1. Introduction and Main Result

The axiomatic theory for connections and means of pairs of positive operators has been developed by Nishio and Ando [12] and Kubo and Ando [10]. Let $A, B, C, \ldots$ denote bounded, linear and positive operators on a Hilbert space $\mathscr{H}$. A binary operation σ defined on and with values in the class of positive operators on $\mathscr{H}$ is called a *connection* if

(I) $A \leqq C$ and $B \leqq D$ imply $A \sigma B \leqq C \sigma D$,

(II) $C(A \sigma B)C \leqq (CAC) \sigma (CBC)$,

(III) $A_n \downarrow A$ and $B_n \downarrow B$ imply $(A_n \sigma B_n) \downarrow A \sigma B$.

A *mean* is a connection with normalization condition

(IV) $1 \sigma 1 = 1$.

If $\mathscr{H}$ is infinite dimensional Kubo and Ando [10] showed the existence of an affine order isomorphism between the class of connections and the class of positive operator monotone functions defined on $\mathbb{R}_+$. This isomorphism $\sigma \leftrightarrow f$ is characterized by the relation

$$A \sigma B = A^{1/2} f(A^{-1/2} B A^{-1/2}) A^{1/2}$$

for invertible A and B. A connection σ is a mean if and only if the *representing function* f under the above isomorphism satisfies $f(1) = 1$. A mean σ is said to be *selfadjoint* if it satisfies

(V) $(A \sigma B)^{-1} = A^{-1} \sigma B^{-1}$

for invertible A and B. It follows that a mean σ is selfadjoint if and only if the representing function f satisfies

$$f(x)^{-1} = f(x^{-1}) \qquad \forall x \in \mathbb{R}_+ .$$

The aim of this article is to determine the class of selfadjoint means of pairs of positive operators on an infinite dimensional Hilbert space. We achieve it by the following characterization of their representing functions.

---

* On leave from Matematisk Institut, Universitetsparken 5, DK-2100 København ϕ, Denmark

0025-5831/81/0256/0029/$01.40

Theorem 1.1. *Let $f:\mathbb{R}_+\to\mathbb{R}_+$ be a function satisfying*
(i) *f is operator monotone,*
(ii) $f(x)^{-1}=f(x^{-1})\quad \forall x\in\mathbb{R}_+$.
Then f admits a canonical representation

$$f(x)=\exp\int_{-1}^{0}\left(\frac{1}{\lambda-x}+\frac{x}{1-\lambda x}\right)h(\lambda)\,d\lambda \qquad x\in\mathbb{R}_+,$$

where $h:[-1,0]\to[0,1]$ is a measurable function and $d\lambda$ denotes Lebesgue measure. The equivalence class containing h is uniquely determined by f. Any function of the indicated form maps $\mathbb{R}_+$ into $\mathbb{R}_+$ and satisfies (i) *and* (ii).

Choosing h to be a constant $\alpha\in[0,1]$ we obtain the function $f(x)=x^\alpha$, $x\in\mathbb{R}_+$. In [10, Sect. 4] the question was raised whether the selfadjoint means with these representing functions exhaust the class of selfadjoint means. We can now see that this is not the case. Taking $h(\lambda)=\lambda+1$, $\lambda\in[-1,0]$ we obtain a new solution

$$f(x)=\left(\frac{x}{1+x}\right)^{1+x}(1+x)^{1+\frac{1}{x}}:\mathbb{R}_+\to\mathbb{R}_+$$

satisfying conditions (i) and (ii).

2. Exponential Ordering Monotone Functions

We introduce an order relation $\preccurlyeq$ called *exponential ordering* between linear selfadjoint operators on a Hilbert space by

Definition 2.1. Let A, B denote linear selfadjoint operators on a Hilbert space $\mathscr{H}$. We write $A\preccurlyeq B$ if $\exp A\leqq\exp B$.

It is easily verified that exponential ordering satisfies
(i) $A\preccurlyeq A$,
(ii) $A\preccurlyeq B$ and $B\preccurlyeq C$ imply $A\preccurlyeq C$,
(iii) $A\preccurlyeq B$ and $B\preccurlyeq A$ imply $A=B$.

Furthermore, since the functions $f(x)=x^\lambda$, $x\in\mathbb{R}_+$ are operator monotone for $\lambda\in[0,1]$ we obtain
(iv) $A\preccurlyeq B$ implies $\lambda A\preccurlyeq\lambda B$ for $\lambda\in[0,1]$.

Finally, using that A and 1 commute we get
(v) $A\preccurlyeq B$ implies $A+\lambda\preccurlyeq B+\lambda$ for $\lambda\in\mathbb{R}$.

Conditions (i), (ii), and (iii) show that $\preccurlyeq$ is an order relation. Since the logarithm is operator monotone on $\mathbb{R}_+$, it follows that $\preccurlyeq$ is stronger than $\leqq$. Indeed, $A\preccurlyeq B$ implies $A\leqq B$. We denote by $\mathscr{E}$ the class of functions $F:\mathbb{R}\to\mathbb{R}$ such that $A\preccurlyeq B$ implies $F(A)\preccurlyeq F(B)$ for all bounded, linear and selfadjoint operators A, B on an infinite dimensional Hilbert space $\mathscr{H}$. $\mathscr{E}$ is thus the class of *exponential ordering monotone functions* defined on $\mathbb{R}$.

Lemma 2.2. *Let $F\in\mathscr{E}$. The function*

$$\Phi(F)(x)=\exp F(\log x),\qquad x\in\mathbb{R}_+$$

maps $\mathbb{R}_+$ into $\mathbb{R}_+$. The mapping Φ is a bijection of $\mathscr{E}$ onto the class $\mathscr{P}$ of operator monotone functions defined on and with values in $\mathbb{R}_+$.

Proof. Let $F \in \mathscr{E}$. It is clear that $\Phi(F)$ maps $\mathbb{R}_+$ into $\mathbb{R}_+$. Any pair of bounded, linear and selfadjoint operators on $\mathscr{H}$ with spectrum contained in $\mathbb{R}_+$ can be written in the form $\exp A$, $\exp B$ where A and B are selfadjoint. Suppose that $\exp A \leqq \exp B$, that is $A \leqslant B$. Since $F \in \mathscr{E}$ we have $F(A) \leqslant F(B)$ or $\exp F(A) \leqq \exp F(B)$. However, $\exp F(A) = \Phi(F)(\exp A)$ hence $\Phi(F)(\exp A) \leqq \Phi(F)(\exp B)$ which shows that $\Phi(F)$ is operator monotone. Let now $f \in \mathscr{P}$. The function $F(x) = \log f(\exp x)$ maps $\mathbb{R}$ into $\mathbb{R}$ and $\Phi(F) = f$. It remains to be proven that $F \in \mathscr{E}$. If $A \leqslant B$, that is $\exp A \leqq \exp B$, then $f(\exp A) \leqq f(\exp B)$. But $f(\exp A) = \Phi(F)(\exp A) = \exp F(A)$, thus $\exp F(A) \leqq \exp F(B)$ or $F(A) \leqslant F(B)$. Q.E.D.

As a corollary of Löwner's theory we obtain

Theorem 2.3. *A non-constant function $F : \mathbb{R} \to \mathbb{R}$ belongs to the class $\mathscr{E}$ if and only if F admits an analytic continuation into the strip $\{z \in \mathbb{C} \mid 0 < \mathrm{Im} z < \pi\}$ which leaves the strip invariant.*

Proof. The non-constant functions in the class $\mathscr{P}$ are characterized by Löwner [11] as those functions $f : \mathbb{R}_+ \to \mathbb{R}_+$ which admit an analytic continuation into the upper half plane $\{z \in \mathbb{C} \mid \mathrm{Im} z > 0\}$ which leave the half plane invariant. The statement of the theorem is now a consequence of Lemma 2.2, because the logarithm admits an analytic continuation into the upper half plane which maps the upper half plane onto the strip $\{z \in \mathbb{C} \mid 0 < \mathrm{Im} z < \pi\}$. Q.E.D.

Similarly we can prove that a non-constant function $F :]a, b[\to \mathbb{R}$ is exponential ordering monotone on the interval $]a, b[$ (that is, $A \leqslant B$ implies $F(A) \leqslant F(B)$ for all bounded, linear and selfadjoint operators A, B with spectrum contained in $]a, b[$) if and only if F admits an analytic continuation into the strip $\{z \in \mathbb{C} \mid 0 < \mathrm{Im} z < \pi\}$ which leaves the strip invariant. However, we shall not need this result. It does not follow in any easy way from the definition of $\mathscr{E}$ that $\mathscr{E}$ is convex, but this is clear from the characterization given in Theorem 2.3. The mapping $\Phi : \mathscr{E} \to \mathscr{P}$ is thus a bijection between two convex sets. We want to derive an integral representation for the functions in $\mathscr{E}$ similar to the well-known expression for the operator monotone functions [4]. Since Φ is not affine, this can not be done merely by transport of structure.

Theorem 2.4. *Let $F \in \mathscr{E}$. Then F admits a canonical representation*

$$F(x) = \beta + \int_{-\infty}^{0} \left(\frac{1}{\lambda - \exp x} - \frac{\lambda}{\lambda^2 + 1} \right) h(\lambda)\, d\lambda \qquad \forall x \in \mathbb{R},$$

where $\beta \in \mathbb{R}$, $h :]-\infty, 0] \to [0, 1]$ is a measurable function and $d\lambda$ denotes Lebesgue measure. The equivalence class containing h is uniquely determined by F. Any function F given on the indicated form belongs to the class $\mathscr{E}$.

Proof. Let $F \in \mathscr{E}$. The function

$$\phi(x) = F(\log x) : \mathbb{R}_+ \to \mathbb{R}$$

admits an analytic continuation into the upper half plane which maps the upper half plane into the subset $\{z \in \mathbb{C} \mid 0 \leqq \mathrm{Im} z < \pi\}$. Hence ϕ is a Pick function and

admits, according to [5, Sect. 2, Theorem 1], a unique canonical representation of the form

$$\phi(\zeta)=\alpha\zeta+\beta+\int\left(\frac{1}{\lambda-\zeta}-\frac{\lambda}{\lambda^2+1}\right)d\mu(\lambda),\qquad \mathrm{Im}\,\zeta>0,$$

where $\alpha\geqq 0$, $\beta\in\mathbb{R}$ and μ is a positive Borel measure for which $\int\frac{1}{\lambda^2+1}d\mu(\lambda)$ is finite. We conclude from [5, Sect. 2, Lemma 2], that the measure μ puts no mass in $\mathbb{R}_+$. We have

$$\mathrm{Im}\,\phi(\zeta)=\alpha\,\mathrm{Im}\,\zeta+\int\frac{\mathrm{Im}\,\zeta}{|\lambda-\zeta|^2}d\mu(\lambda).$$

Since

$$0\leqq\mathrm{Im}\,\phi(\zeta)<\pi\quad\text{for}\quad\mathrm{Im}\,\zeta>0$$

it follows that $\alpha=0$. The boundedness of $\mathrm{Im}\,\phi(\zeta)$ furthermore entails that the measure μ is absolutely continuous with respect to Lebesgue measure. According to [5, Sect. 2, Lemma 1], we have for $a<b$

$$\mu(b)-\mu(a)=\lim_{\eta\to 0}\frac{1}{\pi}\int_a^b\mathrm{Im}\,\phi(x+i\eta)\,dx.$$

Thus $d\mu(\lambda)=h(\lambda)d\lambda$, where $h:]-\infty,0]\to[0,1]$ is a measurable function. If, on the other hand, a function $\phi:\mathbb{R}_+\to\mathbb{R}$ is given on the form

$$\phi(x)=\beta+\int_{-\infty}^{0}\left(\frac{1}{\lambda-x}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda,\qquad x\in\mathbb{R}_+,$$

where $\beta\in\mathbb{R}$ and $h:]-\infty,0]\to[0,1]$ is a measurable function non-equivalent to 0, then ϕ admits an analytic continuation into the upper half plane which maps the upper half plane into the strip $\{z\in\mathbb{C}|0<\mathrm{Im}\,z<\pi\}$. Hence the function $F(x)=\phi(\exp x)$, $x\in\mathbb{R}$ belongs to the class $\mathscr{E}$ by virtue of Theorem 2.3 and the proof is complete. Q.E.D.

The expression for a function $F\in\mathscr{E}$ derived in Theorem 2.4 remains valid for the analytic continuation of F into the strip $\{z\in\mathbb{C}|0<\mathrm{Im}\,z<\pi\}$. We obtain by computation that $\beta=\mathrm{Re}\,F(i\frac{\pi}{2})$. An operator monotone function defined on the whole real line is linear. No such phenomenon occurs for exponential ordering monotone functions as can be seen from Theorem 2.4.

3. The Functional Equation

Theorem 3.1. *Let F be a function in $\mathscr{E}$ with canonical representation as given by Theorem 2.4. Then F satisfies $F(-x)=-F(x)$ $\forall x\in\mathbb{R}$ if and only if $\beta=0$ and $h(\lambda^{-1})=h(\lambda)$ for almost all $\lambda\in]-\infty,0[$.*

Proof. Assume first that $F(-x)=-F(x)$ $\forall x\in\mathbb{R}$. The integral representation of F given in Theorem 2.4 shows that F allows an analytic continuation into the strip $\{z\in\mathbb{C}|-\pi<\mathrm{Im}\,z<\pi\}$. The function $-F(-\zeta)$ is analytic in the same strip and

coincides with F on the real line. Hence $F(-\zeta)=-F(\zeta)$ for $-\pi<\operatorname{Im}\zeta<\pi$. Setting $\zeta=i\theta$ for $\theta\in]0,\pi[$ in the integral representation for F, we obtain

$$F(\zeta)+F(-\zeta)=2\beta+\int_{-\infty}^{0}\left(\frac{1}{\lambda-\exp i\theta}+\frac{1}{\lambda-\exp(-i\theta)}-\frac{2\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda=0$$

for $\theta\in]0,\pi[$. The expression in the parenthesis is calculated to be $\dfrac{-2(1-\lambda^2)\cos\theta}{(\lambda^2-2\lambda\cos\theta+1)(\lambda^2+1)}$ which vanishes for $\theta=\frac{\pi}{2}$, hence $\beta=0$. Putting $t=\cos\theta$, we conclude that

$$\int_{-\infty}^{0}\frac{1-\lambda^2}{(\lambda^2-2t\lambda+1)(\lambda^2+1)}h(\lambda)\,d\lambda=0\qquad\forall t\in]-1,1[.$$

We split the range of integration at the point $\lambda=-1$ and make the substitution $\lambda\to\lambda^{-1}$ in the first term. We obtain

$$\int_{0}^{-1}\frac{(1-\lambda^{-2})h(\lambda^{-1})}{(\lambda^{-2}-2t\lambda^{-1}+1)(\lambda^{-2}+1)}\frac{(-1)}{\lambda^2}d\lambda+\int_{-1}^{0}\frac{(1-\lambda^2)h(\lambda)}{(\lambda^2-2t\lambda+1)(\lambda^2+1)}d\lambda=0$$

for $t\in]-1,1[$. This can be written as

$$\int_{-1}^{0}\frac{1+\lambda^2}{\lambda^2-2t\lambda+1}\cdot\frac{1-\lambda^2}{(1+\lambda^2)^2}(h(\lambda)-h(\lambda^{-1}))\,d\lambda=0\qquad\forall t\in]-1,1[.$$

We make the substitution $y=\dfrac{\lambda}{\lambda^2+1}$ and observe that

$$\frac{dy}{d\lambda}=\frac{1-\lambda^2}{(1+\lambda^2)^2}>0\quad\text{for}\quad\lambda\in]-1,0].$$

Furthermore

$$\frac{1+\lambda^2}{\lambda^2-2t\lambda+1}=\frac{1}{1-2ty}=\sum_{n=0}^{\infty}(2ty)^n\quad\text{for}\quad y\in[-\tfrac{1}{2},0]\quad\text{and}\quad|t|<1.$$

Consequently

$$\sum_{n=0}^{\infty}2^n\int_{-1/2}^{0}y^n(h(\lambda(y))-h(\lambda(y)^{-1}))\,dy\,t^n=0\qquad\forall t\in]-1,1[.$$

Hence

$$\int_{-1/2}^{0}y^n(h(\lambda(y))-h(\lambda(y)^{-1}))\,dy=0\quad\text{for}\quad n=0,1,2,\ldots.$$

We conclude that $h(\lambda(y))=h(\lambda(y)^{-1})$ for almost all $y\in[-\frac{1}{2},0[$. Since the mapping $[-\frac{1}{2},0]\ni y\to\lambda(y)\in[-1,0]$ preserves null-sets, the desired conclusion follows. Next we prove that the condition is sufficient. Let

$$F(x)=\int_{-\infty}^{0}\left(\frac{1}{\lambda-\exp x}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda\qquad\forall x\in\mathbb{R},$$

where $h(\lambda)=h(\lambda^{-1})$ for almost all $\lambda\in]-\infty,0[$. We have

$$F(-x)=\int_{-\infty}^{0}\left(\frac{1}{\lambda-\exp(-x)}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda \qquad \forall x\in\mathbb{R}.$$

Making the substitution $\lambda\to\lambda^{-1}$ we obtain

$$\begin{aligned}F(-x)&=\int_{0}^{-\infty}\left(\frac{1}{\lambda^{-1}-\exp(-x)}-\frac{\lambda^{-1}}{\lambda^{-2}+1}\right)h(\lambda^{-1})\frac{(-1)}{\lambda^2}\,d\lambda\\&=-\int_{-\infty}^{0}\left(\frac{\lambda^{-1}}{\lambda\exp(-x)-1}+\frac{\lambda^{-1}}{\lambda^2+1}\right)h(\lambda)\,d\lambda \qquad \forall x\in\mathbb{R}.\end{aligned}$$

The expression in the parenthesis is calculated to be

$$\frac{1+\lambda\exp x}{(\lambda-\exp x)(\lambda^2+1)}=\frac{1}{\lambda-\exp x}-\frac{\lambda}{\lambda^2+1}$$

and the proof is complete. Q.E.D.

Corollary 3.2. *A function $F:\mathbb{R}\to\mathbb{R}$ is in the class $\mathcal{E}$ and satisfies $F(-x)=-F(x)$ $\forall x\in\mathbb{R}$ if and only if F admits a canonical representation*

$$F(x)=\int_{-1}^{0}\left(\frac{1}{\lambda-\exp x}+\frac{\exp x}{1-\lambda\exp x}\right)h(\lambda)\,d\lambda \qquad \forall x\in\mathbb{R},$$

where $h:[-1,0]\to[0,1]$ is a measurable function and $d\lambda$ denotes Lebesgue measure. The equivalence class containing h is uniquely determined by F.

Proof. Let F be as assumed. According to Theorem 3.1 we have

$$F(x)=\int_{-\infty}^{0}\left(\frac{1}{\lambda-\exp x}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda \qquad \forall x\in\mathbb{R},$$

where $h:]-\infty,0]\to[0,1]$ is a measurable function and $h(\lambda)=h(\lambda^{-1})$ for almost all $\lambda\in]-\infty,0[$. We split the range of integration at the point $\lambda=-1$, make the substitution $\lambda\to\lambda^{-1}$ in the first term and obtain

$$\begin{aligned}F(x)&=\int_{0}^{-1}\left(\frac{1}{\lambda^{-1}-\exp x}-\frac{\lambda^{-1}}{\lambda^{-2}+1}\right)h(\lambda^{-1})\frac{(-1)}{\lambda^2}\,d\lambda+\int_{-1}^{0}\left(\frac{1}{\lambda-\exp x}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda\\&=\int_{-1}^{0}\left(\frac{\lambda^{-1}}{1-\lambda\exp x}-\frac{\lambda^{-1}}{1+\lambda^2}+\frac{1}{\lambda-\exp x}-\frac{\lambda}{\lambda^2+1}\right)h(\lambda)\,d\lambda \qquad \forall x\in\mathbb{R}.\end{aligned}$$

The expression in the parenthesis is calculated to be

$$\frac{1}{\lambda-\exp x}+\frac{\exp x}{1-\lambda\exp x}.$$

On the other hand, if F has the indicated form, then we can extend h to $]-\infty,0[$ by the requirement that $h(\lambda^{-1})=h(\lambda)$ for $\lambda\in[-1,0[$. Following the steps above in reverse order back to the representation given in Theorem 2.4, and using Theorem 3.1, we conclude that $F\in\mathcal{E}$ and $F(-x)=-F(x)$ for $x\in\mathbb{R}$. Q.E.D.

We are now ready to prove Theorem 1.1. The bijection $\Phi: \mathscr{E} \to \mathscr{P}$ has the property that $\Phi(F)(x^{-1}) = \Phi(F)(x)^{-1}\ \forall x \in \mathbb{R}_+$ if and only if $F(-x) = -F(x)\ \forall x \in \mathbb{R}$. The operator monotone functions mapping $\mathbb{R}_+$ into R_+ and satisfying $f(x^{-1}) = f(x)^{-1}\ \forall x \in \mathbb{R}_+$ thus appear as the image under Φ of the functions characterized in Corollary 3.2. This proves Theorem 1.1. Note that Φ maps the functions $F(x) = \alpha x : \mathbb{R} \to \mathbb{R}$, $\alpha \in [0, 1]$ onto the functions $f(x) = x^{\alpha} : \mathbb{R}_+ \to \mathbb{R}_+$.

References

1. Anderson, W.N., Jr., Duffin, R.J.: Series and parallel addition of matrices. J. Math. Anal. Appl. **26**, 576–594 (1969)
2. Anderson, W.N., Jr., Schreiber, M.: The infimum of two projections. Acta Sci. Math. **33**, 165–168 (1972)
3. Anderson, W.N., Jr., Trapp, G.E.: Shorted operators. II. SIAM J. Appl. Math. **28**, 60–71 (1975)
4. Bendat, J., Sherman, S.: Monotone and convex operator functions. Trans. Am. Math. Soc. **79**, 58–71 (1955)
5. Donoghue, W.: Monotone matrix functions and analytic continuation. Berlin, Heidelberg, New York: Springer 1974
6. Fujii, J.I.: Arithmetico-geometric mean of operators. Math. Japon **23**, 667–669 (1979)
7. Fujii, J.I.: On geometric and harmonic means of operators. Math. Japon **24**, 203–207 (1979)
8. Fujii, J.I., Fujii, M.: Some remarks on operator means. Math. Japon **24**, 335–339 (1979)
9. Fujii, J.I.: Initial conditions on operator monotone functions. Math. Japon **24**, 459–462 (1979)
10. Kubo, I., Ando, T.: Means of positive linear operators. Math. Ann. **246**, 205–224 (1980)
11. Löwner, K.: Über monotone Matrixfunktionen. Math. Z. **38**, 177–216 (1934)
12. Nishio, K., Ando, T.: Characterizations of operations derived from network connections. J. Math. Anal. Appl. **53**, 539–549 (1976)
13. Pusz, W., Woronowicz, S.L.: Functional calculus for sesquilinear forms and the purification map. Rep. Math. Phys. **8**, 159–170 (1975)

Received July 21, 1980

Math. Ann. 256, 37–42 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Problème de Levi et morphisme localement de Stein

B. Jennane

Département de Mathématique, Université de Rabat, Rabat, Marocco

En 1977 Skoda [9] apporta, à l'aide d'un contre-exemple, une réponse négative à la question de savoir si un fibré holomorphe à base et à fibre de Stein était de Stein. Auparavant, cette question qui était posée depuis 1953 par Serre [8] fut généralisée et l'on se demandait si un espace analytique complexe X pour lequel il existe un morphisme localement de Stein $\pi : X \to \Omega$ à base Ω de Stein était de Stein (voir aussi le contre-exemple de Fornaess [3]). Rappelons que π surjectif est dit localement de Stein si tout point de Ω possède un voisinage ouvert dont l'image inverse par π est de Stein. Dans [5] l'auteur a montré que si la base Ω est de Stein, alors les groupes de cohomologie $H^p(X, \mathscr{F})$ sont nuls pour $p \geqq 2$ et pour tout faisceau cohérent $\mathscr{F}$ au-dessus de X.

Il a aussi démontré dans [6] que $H^1(X, \mathscr{F})$ est nul si et seulement si il est séparé. Dans cet article on montre que la nullité du groupe $H^1(X, \Theta_X)$ est suffisante pour que X soit de Stein, Θ_X étant le faisceau des germes de fonctions holomorphes. Nous obtenons alors le théorème suivant:

Théorème. *Soit X et Ω deux espaces analytiques complexes de dimensions bornées et tels qu'il existe un morphisme localement de Stein $\pi : X \to \Omega$. On suppose Ω de Stein. Alors les assertions suivantes sont équivalentes:*

i) *X est de Stein.*
ii) *$H^1(X, \Theta_X)$ est séparé.*
iii) *$H^1(X, \Theta_X)$ est de dimension finie.*
iv) *$H^1(X, \Theta_X) = 0$.*

Markoe [7] a montré exactement le même résultat pour X réunion croissante d'espaces de Stein au lieu d'être un espace muni d'un morphisme localement de Stein à base de Stein. Sa démonstration utilise des arguments qui relèvent de la topologie algébrique. La nôtre est purement géométrique et est une adaptation de la démonstration du théorème 2.1 de Fornaess et Narasimhan [3]. Dans leur démonstration ces auteurs n'utilisent guère l'hypothèse que l'ouvert, dont ils montrent la propriété qu'il est de Stein, est localement de Stein. Ceci permet d'adapter facilement la démonstration à notre situation. De plus un raisonnement par récurrence sur la dimension de Ω et la nullité des groupes $H^2(X, \mathscr{F})$ nous

0025-5831/81/0256/0037/$01.20

permettent de nous passer de toute hypothèse du type X réduit où certaines hypersurfaces sont de Stein.

Démonstration du théorème. Les équivalences ii)⇔iii)⇔iv) ont été démontrées dans [7]. Il suffit donc de voir que iv) implique i). Supposons alors qu'en plus des hypothèses du théorème nous avons $H^1(X, \Theta_X)=0$.

Réduction du problème au cas où X est réduit

Soit $n(\Theta_X)$ le nilradical de l'espace (X, Θ_X). Par définition, nous avons

$$\mathrm{red}X=(X, \Theta_{\mathrm{red}X}) \quad \text{avec} \quad \Theta_{\mathrm{red}X}=\Theta_{X/n(\Theta_X)}$$

et $n(\Theta_X)$ est un faisceau cohérent. De la suite exacte

$$0\to n(\Theta_X)\to \Theta_X\to \Theta_{X/n(\Theta_X)}\to 0$$

on déduit la suite longue de cohomologie

$$\to H^1(X, \Theta_X)\to H^1(X, \Theta_{\mathrm{red}X})\to H^2(X, n(\Theta_X))\to .$$

D'après [5], $H^2(X, n(\Theta_X))$ est nul. Si donc $H^1(X, \Theta_X)=0$, nous avons aussi $H^1(X, \Theta_{\mathrm{red}X})=0$. Ainsi les hypothèses faites sur X se transmettent à red$X$. Comme X est de Stein si et seulement si red$X$ l'est [4], on peut supposer dans le théorème que X est réduit.

Seconde réduction

Supposons que l'on sache démontrer l'assertion:

(A) «Le théorème est vrai sous l'hypothèse supplémentaire que pour toute $f\in\Gamma(\Omega, \Theta_\Omega)$ non constante sur aucune composante irréductible de Ω, l'hypersurface dans X, $\{f\circ\pi=0\}$ est de Stein».

On peut alors démontrer le théorème par récurrence sur $\dim\Omega$ de la façon suivante: si $\dim\Omega=1$, $\{f\circ\pi=0\}$ est une réunion discrète de fibres de π, lesquelles sont de Stein.

Supposons le théorème vrai pour $\dim\Omega\leqq n-1$ et considérons le cas $\dim\Omega=n$. L'espace $Y=\{x\in X|(f\circ\pi)(x)=0\}$ est tel que l'application

$$y \xrightarrow{\pi/Y} \{x\in\Omega|f(x)=0\}$$

est un morphisme surjectif localement de Stein, de base un espace de Stein.

De plus l'exactitude de la suite

$$H^1(X, \Theta_X)\to H^1(X, \Theta_{X/(f\circ\pi)})\to H^2(X, (f\circ\pi))$$

[où $(f\circ\pi)$ est le faisceau d'idéaux engendré par $f\circ\pi$] montre que

$$H^1(Y, \Theta_Y)=H^1(X, \Theta_{X/(f\circ\pi)})=0$$

car $H^2(X, (f\circ\pi))=0$ d'après [5]. L'hypothèse de récurrence implique par conséquent que Y est de Stein et l'assertion (A) permet alors de conclure que X est de Stein.

La démonstration de l'assertion (A) sera calquée sur celle du théorème 2.1 de [3]. Nous supposons donc que $H^1(X, \Theta_X)=0$ et que pour toute fonction $f\in\Gamma(\Omega, \Theta_\Omega)$ une constante sur aucune composante irréductible de Ω, l'hypersurface $\{f\circ\pi=0\}\subset X$ est de Stein, et nous devons montrer que X est alors de Stein. En fait, il suffit de montrer que X est holomorphiquement convexe: Si A est un sous-ensemble analytique compact convexe de X, Ω étant de Stein, $\pi(A)$ est réduit à un point. A est alors contenue dans une fibre, donc de dimension zéro. D'après le théorème de Grauert [1], X est alors de Stein si et seulement si il est holomorphiquement convexe.

Lemme 1. *Toute fonction holomorphe sur* $Y=\{x\in X|(f\circ\pi)(x)=0\}$ *s'étend en une fonction holomorphe dans* X.

Ce lemme n'est rien d'autre que le lemme 2.2 de [3].

Pour la suite nous adopterons les notations suivantes. Soit f une fonction holomorphe dans Ω qui n'est constante sur aucune composante irréductible. Pour $t\in\mathbb{C}$, on pose

$$Y_t=\{y\in X|(f\circ\pi)(y)=t\}.$$

Si $K\subset X$ est un compact, l'enveloppe holomorphiquement convexe de $K\cap Y_t$ est la même dans (X, Θ_X) et dans (Y_t, Θ_{Y_t}) d'après le lemme 1. Nous la désignons par $K(t)$

$$K(t)=\widehat{K\cap Y_t}=(K\cap Y_t)\hat{}_{Y_t}.$$

Lemme 2. *Pour tout compact* $K\subset X$, *il existe* $\varepsilon>0$ *et un compact* $L\subset X$ *tels que, si* $|t|<\varepsilon$, *nous avons* $K(t)\subset L$.

(Voir lemme 2.4 de [3].)

Lemme 3. *Soit* Z *un espace de Stein,* K *un compact dans* Z. *Soit* $U\Subset Z$ *un domaine de Runge relativement compact dans* Z, $K\subset U$ *et soit* $L=\bar{U}$.

Si $\mathcal{J}$ *et* $\mathcal{F}$ *sont deux faisceaux analytiques cohérents sur* Z *tels qu'il existe un* Θ_Z*-morphisme* $\alpha:\mathcal{J}\to\mathcal{F}$ *surjectif, alors il existe* $c>0$ *tel que pour tout* $s\in\Gamma(Z, \mathcal{F})$ *et tout* $\tau>0$, *on peut trouver* $s'\in\Gamma(Z, \mathcal{J})$ *vérifiant*

$$\alpha(s')=s \quad et \quad \|s'\|_K\leqq c\|s\|_L+\tau.$$

Les normes sont ici définies de la façon usuelle à partir de recouvrements localement fini de Z par des ouverts de Stein U_i et des représentations

$$\Theta_Z^{p_i}\to\mathcal{J}\to 0 \quad \text{et} \quad \Theta_Z^{q_i}\to\mathcal{F}\to 0 \quad \text{dans} \quad U_i.$$

Ce lemme n'est rien d'autre que le lemme 2.3.2 de [3].

Lemme 4. *Pour tout compact* $K\subset X$, *il existe un compact* $L\subset X$, $K\subset L$ *et* $\varepsilon>0$ *tels que l'on ait la propriété suivante: pour* $|t|<\varepsilon$, *il existe* $M_t>0$ *tel que pour tout* $\tau>0$ *et tout* $g\in\Gamma(Y_t, \Theta_{Y_t})$, *on peut trouver* $G\in\Gamma(X, \Theta_X)$ *vérifiant*

$$G|_{Y_t}=g \quad et \quad \|G\|_K\leqq M_t\|g\|_{L\cap Y_t}+\tau.$$

Preuve. a) Soit $(\Omega_i)_{i\in\mathbb{N}}$ un recouvrement localement fini de Ω tel que $U_i=\pi^{-1}(\Omega_i)$ soit de Stein pour tout $i\in N$. On suppose que pour un entier p nous avons

$K \subset \bigcup_{i=1}^{p} U_i$. Soit $K_j \subset U_j, j=1,\ldots,p$ des compacts vérifiant $\bigcup_{v=1}^{p} K_j \supset K$. La nullité du groupe de cohomologie $H^1(X, \Theta_X)$ se traduit par le fait que l'application cobord

$$\delta = c^0(\mathscr{U}, \Theta_X) \to Z^1(\mathscr{U}, \Theta_X)$$

où $\mathscr{U} = (U_i)_{i \in \mathbb{N}}$, est surjective. C'est aussi une application linéaire continue entre espaces de Fréchet. D'après le théorème de l'application ouverte, on peut trouver des compacts $K'_j, j=1,\ldots,p$, $K_j \subset K'_j \subset U_j$ et une constante $c>0$ tels que pour tout $\{g_{ij}\} \in Z^1(\mathscr{U}, \Theta_X)$ et tout $\sigma > 0$, il existe $\{g_i\} \in C^0(\mathscr{U}, \Theta_X)$ vérifiant

$$g_{ij} = g_i - g_j \quad \text{dans} \quad U_{ij}$$

et

$$\max_{1 \leqq i \leqq p} \|g_i\|_{K_i} \leqq c \max_{1 \leqq i, j \leqq p} \|g_{ij}\|_{K'_i \cap K'_j} + \sigma .$$

b) Soit $\mathscr{J}(t)$ le faisceau d'idéaux engendré par $(f-t)$. La multiplication par $(f-t)$ définit un morphisme surjectif

$$\Theta_X \to \mathscr{J}(t).$$

D'après le lemme 3, et du fait que X est réduit, pour tous $1 \leqq i, j \leqq p$ il existe un compact $E_{ij} \subset U_{ij}$ et une constante c'_t tels que pour tout $s \in \Gamma(U_{ij}, \Theta_X)$

$$\|s\|_{K'_i \cap K'_j} \leqq c'_t \|(f-t) \cdot s\|_{E_{ij}} .$$

Pour la suite de la démonstration, on va considérer des compacts $E_i \subset U_i$, $i=1,\ldots,p$, tels que $E_{ij} \subset E_i \cap E_j$.

c) Considérons maintenant le morphisme surjectif

$$\Theta_X \to \Theta_{X/\mathscr{J}(t)}$$

au-dessus des ouverts de Stein U_i. Le lemme 3 implique encore l'existence de compacts $L'_j \subset U_i, j=1,\ldots,p$ et de constantes $c''_t > 0$ tels que pour tout $\lambda > 0$ et tout $\tilde{g} \in \Gamma(Y_t, \Theta_{X/\mathscr{J}(t)})$, il existe $h_i \in \Gamma(U_i, \Theta_X)$, $i \in \mathbb{N}$ dont l'image dans $\Gamma(U_i, \Theta_{X/\mathscr{J}(t)})$ est $\tilde{g}|_{U_i \cap Y_t}$ et

$$\left.\begin{array}{l} \|h_j\|_{E_j} \leqq c''_t \|\tilde{g}\|_{L'_j \cap Y_t} + \lambda \\ |t| \text{ est assez petit.} \end{array}\right\} \text{lorsque } j=1,\ldots,p \text{ et}$$

d) Considérons enfin le morphisme de faisceaux

$$\Theta_{X/(f-t)} \to \Theta_{Y_t}$$

au-dessus de l'espace de Stein Y_t. D'après le lemme 2, il existe $\varepsilon > 0$ et un compact $L \subset X$ tels que

$$\left(\bigcup_{j=1}^{p} L'_j \cap Y_t\right)^{\wedge}_{Y_t} \subset L \quad \text{pour} \quad |t| < \varepsilon$$

et d'après la construction de L (voir démonstration du lemme 2), $L \cap Y_t$ contient un domaine de Runge dans Y_t, qui lui-même contient $\bigcup_{j=1}^{p} L'_j \cap Y_t$. Du lemme 3 on

déduit qu'il existe alors $c_t'''>0$ tel que tout $g\in\Gamma(Y_t,\Theta_{Y_t})$ possède une extension $\tilde{g}\in\Gamma(Y_t,\Theta_{X/(f-t)})$ vérifiant

$$\|\tilde{g}\|_{\cup L_j\cap Y_t}\leqq c_t'''\|g\|_{L\cap Y_t}\quad\text{pour}\quad|t|<\varepsilon.$$

e) Venons-en maintenant à la démonstration du lemme 4. Soit $g\in\Gamma(Y_t,\Theta_{Y_t})$.

D'après d), g s'étend en $\tilde{g}\in\Gamma(Y_t,\Theta_{X/(f-t)})$ et d'après c) il existe $h_i\in\Gamma(U_i,O_X)$ qui induit $\tilde{g}|_{U_i\cap Y_t}$. On utilise alors b) et le fait que X est réduit et codim $Y_t>0$ pour avoir $h_i-h_j=(f-t)\cdot g_{ij}$ où $\{g_{ij}\}\in Z^1(\mathscr{U},\Theta_X)$ et pour $|t|$ assez petit et $i,j\in\{1,\ldots,p\}$

$$\begin{aligned}\|g_{ik}\|_{K_i\cap K_j}&\leqq\|g_{ij}\|_{E_{ij}}\leqq c_t'\|h_i-h_j\|_{E_{ij}}\\&\leqq 2c_t'\max_{1\leqq i\leqq p}\|h_i\|_{E_i}\leqq 2c_t'c_t''c_t'''\|g\|_{L\cap Y_t}+\lambda\mu.\end{aligned}$$

On pose $c_t=2c_t'c_t''c_t'''$.

D'après a), il existe $g_i\in\Gamma(U_i,\Theta_X)$ tel que $g_i-g_j=g_{ij}$ des U_{ij} et

$$\begin{aligned}\|g_i\|_{K_i}&\leqq c\max\|g_{ij}\|_{K_i\cap K_j}+\sigma\qquad i=1,\ldots,p\\&\leqq c\cdot c_t\|g\|_{L\cap Y_t}+\lambda c\mu+\sigma.\end{aligned}$$

On définit la fonction $G\in\Gamma(X,\Theta_X)$ par $G=h_i-(f-t)g_i$ dans U_i, nous avons $G|_{Y_t}=g$ et

$$\begin{aligned}\|G\|_K&\leqq\max_{1<i<p}(\|h_i\|_{E_i}+M\|g_i\|_{K_i})\quad\text{où}\quad M=\sup_K|f-t|\\&\leqq(c_t''c_t'''+Mcc_t)\|g\|_{L\cap Y_t}+M\lambda c\mu+\pi\sigma\\&\leqq M_t\|g\|_{L\cap Y_t}+\tau.\end{aligned}$$

Démonstration du théorème. Nous allons utiliser le lemme 4 pour montrer que X est holomorphiquement convexe. Soit K un compact dans X. Rappelons que nous avons un morphisme de Stein $\pi:X\to\Omega$ et que Ω est de Stein. L'enveloppe holomorphiquement convexe de $\pi(\hat{K})$ dans Ω contient $\pi(K)$, où $\hat{K}$ est l'enveloppe holomorphiquement convexe de K dans X.

En effet, si $x\notin\widehat{\pi(K)}$, il existe $f\in\Gamma(\Omega,\Theta_\Omega)$ tel que

$$|f(x)|>\sup_{\pi(K)}|f|$$

donc pour tout $y\in\pi^{-1}(x)$, $|(f\circ\pi)(y)|>\sup_K|f\circ\pi|$ ce qui veut dire que $\pi^{-1}(x)\cap\hat{K}=\emptyset$ et qu'alors $x\notin\pi(\hat{K})$.

$\widehat{\pi(K)}$ est compact: il en est de même pour $\pi(\hat{K})$.

Soit maintenant $x_0\in\Omega$ et soit $f\in\Gamma(\Omega,\Theta_\Omega)$ tels que $f(x_0)=0$. On suppose f non constante dans aucune composante irréductible de Ω et on pose encore

$$Y_t(x_0)=\{x\in X|(f\circ\pi)(x)=t\}\quad\text{pour}\quad t\in\mathbb{C}.$$

Soit $L\subset X$ un compact et $\varepsilon>0$ du lemme 4. Pour $|t|<\varepsilon$ nous avons

$$\hat{K}\cap Y_t\subset(L\cap Y_t)^{\hat{}}_{Y_t}=\text{enveloppe d'holomorphie de } L\cap Y_t \text{ dans } Y_t. \qquad (*)$$

En effet, soit $q\in Y_t$, $q\notin(L\cap Y_t)^{\hat{}}_{Y_t}$. On peut trouver $g\in\Gamma(Y_t,\Theta_{Y_t})$ tel que

$$|g(q)|>1\quad\text{et}\quad\|g\|_{L\cap Y_t}<1/2M_t$$

g s'étend à X en une fonction $G \in \Gamma(X, \Theta_X)$ vérifiant, d'après le lemme 4

$$\|G\|_K \leqq M_t \|g\|_{L \cap Y_t} + \tfrac{1}{2} < 1$$

puisque $|G(q)| = |g(q)| > 1$, on a $q \notin \hat{K}$. Ce qui démontre l'inclusion. Maintenant d'après le lemme 2 appliqué à L, si ε est assez petit, $\bigcup_{|t|<\varepsilon} (L \cap Y_t)^{\wedge}_{Y_t}$ est relativement compact, donc à fortiori $\bigcup_{|t|<\varepsilon} \hat{K} \cap Y_t$ d'après l'inclusion (*). $\pi(\hat{K})$ étant compact, on peut trouver $\varepsilon > 0$ et $x_0, \ldots, x_r \in \Omega$ tels que

$$\hat{K} \subset \bigcup_{i=0}^{r} \bigcup_{|t|<\varepsilon} Y_t(x_i)$$

ce qui montre que $\hat{K}$ est compact, et X est holomorphiquement convexe. C.Q.F.D.

Remerciement. Je remercie H. Skoda pour l'aide qu'il continue à m'apporter et l'intérêt qu'il porte à mon travail et je lui exprime toute ma reconnaissance.

Bibliographie

1. Docquier, F., Grauert, H.: Levisches Problem und Rungescher Satz für Teilgebiete Steinscher Mannigfaltigkeit. Math. Ann. **140**, 94 (1960)
2. Fornaess, J.E.: 2-dimensional counterexamples to generalizations of the Levi problem. Math. Ann. **230**, 169–173 (1977)
3. Fornaess, J.E., Narasimhan, R.: The Levi problem on complex spaces with singularities. Math. Ann. **248**, 47–72 (1980)
4. Grauert, H.: Ein Theorem der analytischen Garbentheorie und die Modulräume komplexer Strukturen. Publ. Math. I.H.E.S. **5**, 233–292 (1960)
5. Jennane, B.: Groupes de cohomologie d'un fibré à base et à fibre de Stein. Inventiones math. **54**, 75–79 (1979)
6. Jennane, B.: Groupes de cohomologie d'un fibré à base et à fibre de Stein. Séminaire Lelong-Skoda, 1978–79. Lecture Notes in Mathematics, Vol. 822. Berlin, Heidelberg, New York: Springer 1980
7. Markoe, A.: Runge families on inductive limits of Stein spaces. Ann. Inst. Fourier Grenoble **27**, 117–127 (1977)
8. Serre, J.P.: Quelques problèmes globaux relatifs aux variétés de Stein. Colloque sur les fonctions de plusieurs variables, Bruxelles, 1953
9. Skoda, H.: Fibrés holomorphes à base et à fibre de Stein. Inventiones math. **43**, 97–107 (1977)

Reçu le 13 octobre 1980

Math. Ann. 256, 43–50 (1981)

Mathematische Annalen
© Springer-Verlag 1981

On Ergodic Properties of Inner Functions

Ch. Pommerenke

Fachbereich 3, Mathematik der Technischen Universität, Strasse des 17. Juni, D-1000 Berlin 12

1. Introduction and Results

1.1.

An analytic function f in the unit disk $\mathbb{D}$ is called an *inner function* if $|f(z)|<1$ for $z\in\mathbb{D}$ and if the angular limits satisfy $|f(\zeta)|=1$ for almost all $\zeta\in\partial\mathbb{D}$. Thus f can be considered as a measurable mapping of the unit circle $\partial\mathbb{D}$ onto itself. It preserves the Lebesgue measure mes if and only if $f(0)=0$.

Let now $f_n=f\circ\ldots\circ f$ denote the iterates of the inner function f and let $z_n=f_n(0)$. The following dichotomy is partly due to Neuwirth [5] and completely to Aaronson [1]: If

$$g_n(z)\equiv\frac{f_n(z)-z_n}{1-\bar{z}_n f_n(z)}\to 0\ (n\to\infty)\quad\text{for}\quad z\in\mathbb{D} \tag{1.1}$$

then f is exact [10, p. 111] and thus ergodic; if (1.1) does not hold then f is not even ergodic.

The condition (1.1) is equivalent [9, Chap. VI; 7, Theorem 1] to

$$(z_{n+1}-z_n)/(1-\bar{z}_n z_{n+1})\to 0\quad(n\to\infty) \tag{1.2}$$

and is thus satisfied, in particular, if f has a fixed point in $\mathbb{D}$. On the other hand, if f has a fixed point ζ on $\partial\mathbb{D}$ with angular derivative $|f'(\zeta)|<1$ then (1.2) does not hold; in the case $|f'(\zeta)|=1$, condition (1.2) may or may not hold.

We shall somewhat strengthen the positive part of the Aaronson-Neuwirth and generalize it to arbitrary sequences of inner functions (instead of the sequence of iterates).

Theorem 1. *Let f_n be inner functions for $n=1,2,\ldots$ and let $z_n=f_n(0)$. If (1.1) holds then*

$$\sup_A\sup_E\left|\frac{\operatorname{mes}[A\cap f_n^{-1}(E)]}{\operatorname{mes} f_n^{-1}(E)}-\frac{\operatorname{mes} A}{2\pi}\right|\to 0\quad(n\to\infty), \tag{1.3}$$

where A ranges over all arcs on $\partial\mathbb{D}$ and E over all measurable sets with $\operatorname{mes} E>0$.

0025-5831/81/0256/0043/$01.60

The uniform mixing condition (1.3) is rather strong. Consider for instance $f_n(0)=0$. Then $\mathrm{mes} f_n^{-1}(E)=\mathrm{mes} E$, and it follows from (1.3) that

$$\mathrm{mes}[A\cap f_n^{-1}(E)]\to\frac{1}{2\pi}\mathrm{mes}\,A\,\mathrm{mes}\,E \quad (n\to\infty) \tag{1.4}$$

for every measurable set E and every arc A, and it is easy to see that (1.4) then holds for every measurable set E. Thus (1.3) implies strong mixing [10, p. 38] in the measure preserving case.

In particular the theorem can be applied to the iterates of f. Let (1.1) hold. If B is of the form $B=f_n^{-1}(E_n)$ for every n with suitable Borel sets E_n then, by (1.3),

$$\mathrm{mes}(A\cap B)=\frac{1}{2\pi}\mathrm{mes}\,A\,\mathrm{mes}\,B$$

for every arc A on $\partial\mathbb{D}$. It easily follows that $\mathrm{mes}\,B=0$ or $=2\pi$. Thus f is exact which is Aaronson's result.

A somewhat more general case is that $f_n=\varphi_1\circ\varphi_2\circ\ldots\circ\varphi_n$ where the inner functions φ_n tend to an inner function f satisfying (1.2) with $z_n=f\circ\ldots\circ f(0)$. It follows from [8, Theorem 4] that (1.1) with $z_n=f_n(0)$ holds so that (1.3) is again satisfied, by Theorem 1.

1.2.

Let φ_n be inner functions with $\varphi_n(0)=0$ for $n=1,2,\ldots$ and let $f_n=\varphi_n\circ\ldots\circ\varphi_2\circ\varphi_1$. We consider partitions $\mathscr{A}$ of $\partial\mathbb{D}$ into finitely many sets $A_1,\ldots,A_N$ and write

$$A_{k_0\ldots k_n}=A_{k_0}\cap f_1^{-1}(A_{k_1})\cap\ldots\cap f_n^{-1}(A_{k_n}). \tag{1.5}$$

We define the *lower entropy* $\underline{h}=\underline{h}[(\varphi_n)]$ of (φ_n) by

$$\underline{h}=\sup_{\mathscr{A}}\liminf_{n\to\infty}\frac{1}{n+1}\sum_{k_0=1}^{N}\ldots\sum_{k_n=1}^{N}\mathrm{mes}\,A_{k_0\ldots k_n}\log\frac{2\pi}{\mathrm{mes}\,A_{k_0\ldots k_n}}. \tag{1.6}$$

For the special case $\varphi_n\equiv f$, the limit exists and $\underline{h}$ is the entropy $h(f)$ of f [10, p. 73].

Theorem 2. *Let φ_n be inner functions with*

$$\varphi_n(0)=0, \quad |\varphi_n'(0)|\leqq a<1 \quad (n=1,2,\ldots). \tag{1.7}$$

Then the lower entropy $\underline{h}$ of (φ_n) satisfies

$$\underline{h}\geqq c(1-a), \tag{1.8}$$

where $c>0$ is an absolute constant.

Specializing to the case $\varphi_n\equiv f$ we immediately obtain:

Corollary. *An inner function f with $f(0)=0$ has an entropy*

$$h(f)\geqq c(1-|f'(0)|). \tag{1.9}$$

For example, the function z^p $(p=1,2,\ldots)$ has entropy $\log p$ [10, pp. 14, 168]. Since measure-theoretic entropy is bounded by topological entropy [3, 10, p. 156]

it is easy to show [10, p. 184] that

$$|f'(z)| \leqq M\,(z \in \mathbb{D}) \Rightarrow h(f) \leqq \log M\,. \tag{1.10}$$

The author does not know whether every inner function has finite entropy.

1.3.

Finally we prove an equidistribution result for the pre-images of a point in $\mathbb{D}$ (instead of a set on $\partial\mathbb{D}$). For $b \in \mathbb{D}$ and $A \subset \partial\mathbb{D}$, we define

$$\sigma_n(b, A) = \sum_{\zeta} (1 - |\zeta|^2), \tag{1.11}$$

where we sum over all $\zeta = |\zeta| e^{i\theta}$ with $f_n(\zeta) = b$ and $e^{i\theta} \in A$.

Theorem 3. *Let f_n be inner functions. Let* (1.1) *be satisfied and let the set of limit points of $z_n = f_n(0)$ have zero (logarithmic) capacity. If $b \in \mathbb{D}$ does not lie in a certain exceptional set of zero capacity, then*

$$\frac{\sigma_n(b, A)}{\sigma_n(b, \partial\mathbb{D})} \to \frac{\operatorname{mes} A}{2\pi} \quad (n \to \infty) \tag{1.12}$$

for every arc A on $\partial\mathbb{D}$.

If the f_n are the iterates of an inner function f, it follows that (1.12) holds if f has a fixed point in $\mathbb{D}$ or if at least (1.2) is satisfied.

2. Uniform Mixing

We need the following fact used also by Neuwirth [5, p. 361].

Lemma 1. *Let g be an inner function and let $E \subset \partial\mathbb{D}$ be measurable. Then*

$$\frac{1}{2\pi} \int_{g^{-1}(E)} \frac{1 - |z|^2}{|\zeta - z|^2} |d\zeta| = \frac{1}{2\pi} \int_E \frac{1 - |g(z)|^2}{|\zeta - g(z)|^2} |d\zeta| \quad (z \in \mathbb{D}). \tag{2.1}$$

Proof. For $z \in \mathbb{D}$, we consider

$$w(z) = \frac{1}{2\pi} \int_E \frac{\zeta + z}{\zeta - z} |d\zeta|, \qquad \operatorname{Re} w(z) = \frac{1}{2\pi} \int_E \frac{1 - |z|^2}{|\zeta - z|^2} |d\zeta|. \tag{2.2}$$

Then w and $h = w \circ g$ are analytic and of bounded real part in $\mathbb{D}$. For almost all $\sigma \in \partial\mathbb{D}$, the radial limits $h(\sigma)$ and $g(\sigma)$ exist so that $w(z) \to h(\sigma)$ as $z \to g(\sigma)$ along $\{g(r\sigma) : 0 \leqq r < 1\}$. Since w is a normal function it follows from the Lehto- Virtanen theorem [4, 6, p. 268] that w has the radial limit $h(\sigma)$ at $g(\sigma)$.

If χ_E denotes the characteristic function of E then $\operatorname{Re} w(\zeta) = \chi_E(\zeta)$ for almost all $\zeta \in \partial\mathbb{D}$, and it follows from Löwner's lemma that

$$\operatorname{Re} w(g(\sigma)) = \operatorname{Re} h(\sigma) = \chi_{g^{-1}(E)}(\sigma) \text{ for almost all } \sigma \in \partial\mathbb{D}.$$

Hence the functions on both sides of (2.1) have the same radial limit almost everywhere, and since they are bounded and harmonic it follows that they are identical.

Proof of Theorem 1. Let $\varepsilon>0$ and determine $r<1$ such that

$$(1-r)^2/|e^{i\varepsilon}-r|^2 \leqq \varepsilon. \tag{2.3}$$

If A is an arc on $\partial\mathbb{D}$ of length α, let $A^{\pm}$ be arcs of lengths $\alpha\pm 2\varepsilon$ and the same midpoint (possibly $A^-=\emptyset$ or $A^+=\partial\mathbb{D}$). We consider the harmonic measure

$$\omega(A^{\pm}, r\zeta)=\frac{1}{2\pi}\int_{A^{\pm}}\frac{1-r^2}{|z-r\zeta|^2}|dz| \quad (|\zeta|=1). \tag{2.4}$$

If $\zeta\in A$ then, by (2.3),

$$1-\omega(A^+, r\zeta)=\frac{1}{2\pi}\int_{\partial\mathbb{D}\setminus A^+}\frac{1-r^2}{|z-r\zeta|^2}|dz|\leqq\varepsilon.$$

Together with a similar argument for A^-, this shows that

$$\omega(A^+, r\zeta)\geqq 1-\varepsilon \quad (\zeta\in A), \qquad \omega(A^-, r\zeta)\leqq\varepsilon \quad (\zeta\notin A). \tag{2.5}$$

Let $E\subset\partial\mathbb{D}$ and $\operatorname{mes}E>0$. We set

$$E_n=f_n^{-1}(E), \qquad E_n^*=\left\{\frac{z-z_n}{1-\bar{z}_n z}: z\in E\right\}. \tag{2.6}$$

Since f_n is an inner function we obtain from Lemma 1 that, for $z\in\mathbb{D}$,

$$\begin{aligned}\int_{E_n}\frac{1-|z|^2}{|\zeta-z|^2}|d\zeta| &= \int_E\frac{1-|f_n(z)|^2}{|\zeta-f_n(z)|^2}|d\zeta| \\ &= \int_{E_n^*}\frac{1-|g_n(z)|^2}{|\zeta-g_n(z)|^2}|d\zeta|,\end{aligned} \tag{2.7}$$

where g_n is defined by (1.1). Since $g_n(0)=0$ it follows that

$$\operatorname{mes}E_n=\operatorname{mes}E_n^*. \tag{2.8}$$

We see from (2.4) and (2.7) that

$$\begin{aligned}\int_{E_n}\omega(A^{\pm}, r\zeta)|d\zeta| &= \frac{1}{2\pi}\int_{A^{\pm}}\left(\int_{E_n}\frac{1-r^2}{|\zeta-rz|^2}|d\zeta|\right)|dz| \\ &= \frac{1}{2\pi}\int_{A^{\pm}}\left(\int_{E_n^*}\frac{1-|g_n(rz)|^2}{|\zeta-g_n(rz)|^2}|d\zeta|\right)|dz|.\end{aligned} \tag{2.9}$$

Since $g_n(rz)\to 0$ as $n\to\infty$ uniformly on $z\in\partial\mathbb{D}$ by (1.1), it follows that there exists $n_0=n_0(\varepsilon)$, independent of A and E, such that, for $n\geqq n_0$,

$$\left|\int_{E_n}\omega(A^{\pm}, r\zeta)|d\zeta|-\frac{1}{2\pi}(\alpha\pm 2\varepsilon)\operatorname{mes}E_n^*\right|<3\varepsilon\operatorname{mes}E_n^*.$$

Since $\operatorname{mes}E_n^*=\operatorname{mes}E_n$ by (2.8), we deduce therefore from (2.5) that

$$(1-\varepsilon)\operatorname{mes}(A\cap E_n)\leqq\int_{A\cap E_n}\omega(A^+, r\zeta)|d\zeta|<\left(\frac{\alpha}{2\pi}+4\varepsilon\right)\operatorname{mes}E_n$$

and also that

$$\operatorname{mes}(A\cap E_n)\geqq \int_{E_n}\omega(A^-,r\zeta)|d\zeta| - \int_{(\partial\mathbb{D}\setminus A)\cap E_n}\omega(A^-,r\zeta)|d\zeta|$$
$$> \left(\frac{\alpha}{2\pi}-4\varepsilon\right)\operatorname{mes}E_n.$$

Hence we see that, for $n\geqq n_0(\varepsilon)$,

$$\left|\operatorname{mes}(A\cap E_n)-\frac{\alpha}{2\pi}\operatorname{mes}E_n\right|<5\varepsilon\operatorname{mes}E_n,$$

and (1.3) follows because of (2.6) and (2.8).

3. The Lower Entropy

We prove first an estimate on composition; the bounds are certainly not sharp.

Lemma 2. *Let the functions φ_n be analytic in $\mathbb{D}$ with $|\varphi_n(z)|<1$. Let*

$$\varphi_n(0)=0,\quad |\varphi_n'(0)|\leqq a,\quad \tfrac{1}{2}\leqq a<1 \tag{3.1}$$

for $n=1,2,\ldots$ and define $f_n=\varphi_n\circ\ldots\circ\varphi_2\circ\varphi_1$. Then

$$|f_n(z)|<a^n(1-|z|)^{-13}\quad(|z|<1). \tag{3.2}$$

Proof. We consider the function

$$\psi(z)=z\frac{a+z}{1+az}\quad(z\in\mathbb{D}). \tag{3.3}$$

It follows from (3.1) that

$$|\varphi_n(z)|\leqq|z|\frac{|\varphi_n'(0)|+|z|}{1+|\varphi_n'(0)||z|}\leqq\psi(|z|).$$

If ψ_n denotes the n-th iterate of ψ, induction shows that

$$|f_n(z)|\leqq\psi_n(|z|)\quad\text{for}\quad z\in\mathbb{D},\quad n=1,2,\ldots. \tag{3.4}$$

It is known [9, Sect. 37] that, as $n\to\infty$,

$$g_n(z)\equiv a^{-n}\psi_n(z)\to g(z)=z+\ldots\quad(z\in\mathbb{D}), \tag{3.5}$$

where $g(\psi(z))=ag(z)$. If $0\leqq x<1$ then, by (3.3),

$$g_{n+1}(x)=\frac{\psi(\psi_n(x))}{a^{n+1}}=g_n(x)\frac{1+a^{n-1}g_n(x)}{1+a^{n+1}g_n(x)}\geqq g_n(x).$$

Hence we see that $g_n(x)\leqq g(x)$ for $0\leqq x<1$ and $n=1,2,\ldots$.

Since $a\geqq 1/2$ it is easy to verify that ψ is univalent in $\{|z|<1/5\}$. Since $|\psi(z)|\leqq|z|$ it follows that ψ_n and thus g is univalent in $\{|z|<1/5\}$. Hence the Koebe distortion theorem [6, p. 21] shows that $|g(z)|<1$ for $|z|\leqq 1/8$.

We define $x_0=1/8$ and $x_{n+1}=\psi^{-1}(x_n)>x_n$ for $n=0,1,\ldots$. We show now by induction that

$$g(x)<(1-x)^{-13} \quad (0\leqq x\leqq x_n). \tag{3.6}$$

This is true for $n=0$. Let (3.6) hold for n and let $x_0\leqq x\leqq x_{n+1}$. Then $0<\psi(x)\leqq\psi(x_{n+1})=x_n$ and thus

$$g(x)=\frac{1}{a}g(\psi(x))<\frac{1}{a}(1-\psi(x))^{-13}$$
$$\leqq\frac{1}{a}\left(\frac{8+a}{9}\right)^{13}(1-x)^{-13}\leqq(1-x)^{-13}$$

because $x\geqq 1/8$ and $a<1$. This proves (3.6), and it follows from (3.4) that, for $z\in\mathbb{D}$,

$$|f_n(z)|\leqq a^n g_n(|z|)\leqq a^n g(|z|)<a^n(1-|z|)^{-13}.$$

We deduce a quantitative version of Theorem 1 for our particular case.

Lemma 3. *Let φ_n be inner functions satisfying* (1.7) *and let $f_n=\varphi_n\circ\ldots\circ\varphi_1$. Then, for some absolute constant K,*

$$\left|\frac{\operatorname{mes}[A\cap f_n^{-1}(E)]}{\operatorname{mes}E}-\frac{\operatorname{mes}A}{2\pi}\right|\leqq K\exp\left[-\frac{1-a}{84}n\right] \tag{3.7}$$

for all arcs $A\subset\partial\mathbb{D}$ and all sets $E\subset\partial\mathbb{D}$ with $\operatorname{mes}E>0$.

Proof. We assume first that $1/2\leqq a<1$. We choose

$$r=r_n=1-\tfrac{1}{20}a^{n/14}, \qquad \varepsilon=\varepsilon_n=a^{n/42}.$$

Then (2.3) is satisfied. We obtain from Lemma 2 that

$$|f_n(z)|\leqq a^n(1-r_n)^{-13}=K_1a^{n/14} \quad (|z|=r_n) \tag{3.10}$$

with some absolute constant K_1.

We proceed now as in the proof of Theorem 1. We see from (2.9) (with $g_n=f_n$ and $E_n^*=E$) and from (3.10) that

$$\left|\int_{E_n}\omega(A^{\pm},r\zeta)|d\zeta|-\frac{\alpha\pm 2\varepsilon_n}{2\pi}\operatorname{mes}E\right|<K_1a^{n/14}\operatorname{mes}E$$

because $\operatorname{mes}E_n=\operatorname{mes}E$ in our case when $f_n(0)=0$. We deduce that

$$\left|\frac{\operatorname{mes}[A\cap E_n]}{\operatorname{mes}E}-\frac{\alpha}{2\pi}\right|<2\varepsilon_n+K_1a^{n/14}\leqq K\exp\left[-\frac{1-a}{42}n\right]$$

for $a\geqq 1/2$, and (3.7) follows for $0\leqq a<1$.

Our next lemma generalizes a well-known fact [10, p. 83] about entropy to the lower entropy defined by (1.6).

Lemma 4. *Let $q\geqq 1$ and define a new sequence by*

$$\psi_m=\varphi_{mq}\circ\ldots\circ\varphi_{mq-q+1} \quad (m=1,2,\ldots). \tag{3.11}$$

Then

$$\underline{h}[(\varphi_n)] \geqq q^{-1}\underline{h}[(\psi_m)].$$

Proof. We set $g_m = \psi_m \circ \ldots \circ \psi_2 \circ \psi_1$. Given n we determine m from $mq \leqq n < (m+1)q$. Since $f_{mq} = g_m$ we obtain from (1.5) that

$$\operatorname{mes} A_{k_0 \ldots k_n} \leqq \operatorname{mes} B_{l_0 \ldots l_m}, \qquad B_{l_0 \ldots l_m} \equiv A_{k_0} \cap g_1^{-1}(A_{k_q}) \cap \ldots \cap g_m^{-1}(A_{k_{mq}}),$$

where $l_\mu \equiv k_{\mu q}$. Since $A_1, \ldots, A_N$ is a partition it follows that

$$\sum_{k_0=1}^{N} \ldots \sum_{k_n=1}^{N} \operatorname{mes} A_{k_0 \ldots k_n} \log \frac{2\pi}{\operatorname{mes} A_{k_0 \ldots k_n}} \geqq \sum_{l_0=1}^{N} \ldots \sum_{l_m=1}^{N} \operatorname{mes} B_{l_0 \ldots l_m} \log \frac{2\pi}{\operatorname{mes} B_{l_0 \ldots l_m}},$$

and the assertion follows because $n/m \to q$ as $n \to \infty$.

Proof of Theorem 2. Let

$$q = \left[\frac{84}{1-a}\log(4K)\right] + 1 \tag{3.12}$$

and let ψ_m $(m = 1, 2, \ldots)$ be defined by (3.11). Then $\psi_m(0) = 0$ and $|\psi_m'(0)| \leqq a^q$ by (1.7). We write $g_m = \psi_m \circ \ldots \circ \psi_1$.

Let $A_1 = \{|z| = 1, \operatorname{Im} z \geqq 0\}$, $A_2 = \{|z| = 1, \operatorname{Im} z < 0\}$. It follows from Lemma 3 and from (3.12) that

$$\operatorname{mes}[A_k \cap \psi_m^{-1}(E)] \leqq \tfrac{3}{4} \operatorname{mes} E \quad (k = 1, 2; m = 1, 2, \ldots) \tag{3.13}$$

for all measurable sets E on $\partial\mathbb{D}$. Writing

$$A_{k_0} \cap g_1^{-1}(A_{k_1}) \cap \ldots \cap g_m^{-1}(A_{k_m}) = A_{k_0} \cap \psi_1^{-1}(E)$$

with $E = A_{k_1} \cap \psi_2^{-1}(A_{k_2}) \cap \ldots \cap (\psi_m \circ \ldots \circ \psi_2)^{-1}(A_{k_m})$, we obtain from (3.13) by induction that

$$\operatorname{mes}[A_{k_0} \cap \ldots \cap g_m^{-1}(A_{k_m})] \leqq \pi(3/4)^m \quad (m = 1, 2, \ldots).$$

Hence $\underline{h}[(\psi_m)] \geqq \log(4/3)$ by (1.6), and we obtain from (3.12) and Lemma 4 that

$$\underline{h}[(\varphi_n)] \geqq q^{-1} \log(4/3) \geqq c(1-a).$$

4. Equidistribution

Proof of Theorem 3. (a) Since f_n is an inner function there exists, by a theorem of Frostman [2, p. 37], a set B_n of zero capacity such that, for $b \in \mathbb{D} \setminus B_n$,

$$\left|\frac{f_n(z) - b}{1 - \bar{b} f_n(z)}\right| = \prod_{\zeta \in Z_n(b)} \left|\frac{z - \zeta}{1 - \bar{\zeta} z}\right| \quad (z \in \mathbb{D}), \tag{4.1}$$

where $Z_n(b)$ denotes the set of zeros of $f_n - b$. The union B of all B_n together with all limit points of (z_n) has zero capacity.

Let now $b \in \mathbb{D} \setminus B$. Since b is not a limit point of (z_n) it follows from (1.1) that

$$\varrho_n = \min\{|\zeta| : \zeta \in Z_n(b)\} \to 1 \quad \text{as} \quad n \to \infty. \tag{4.2}$$

Furthermore, with $\zeta = \varrho e^{i\theta}$,

$$2\log\left|\frac{1-\bar{\zeta}z}{z-\zeta}\right| \geqq \frac{(1-\varrho^2)(1-|z|^2)}{|1-\varrho e^{-i\theta}z|^2}$$

$$\geqq (1-\varrho^2)\frac{1-|z|^2}{|e^{i\theta}-z|^2} - (1-\varrho^2)\frac{6(1-\varrho_n)}{(1-|z|)^3}$$

and also $2\log 1/|\zeta| \leqq (1-\varrho^2)/\varrho_n^2$. Summing over $\zeta \in Z_n(b)$ we obtain from (4.1) and (1.11) that

$$\frac{\log\left|\frac{1-\bar{b}f_n(z)}{f_n(z)-b}\right|}{\log\left|\frac{1-\bar{b}z_n}{z_n-b}\right|} \geqq \frac{\varrho_n^2}{\sigma_n(b,\partial\mathbb{D})}\sum_{\zeta\in Z_n(b)}\frac{(1-\varrho^2)(1-|z|^2)}{|e^{i\theta}-z|^2} - \frac{6\varrho_n^2(1-\varrho_n)}{(1-|z|)^3}. \tag{4.3}$$

If we write $a_n = (z_n-b)/(1-\bar{b}z_n)$ and define g_n by (1.1) then

$$\left|\frac{1-\bar{b}f_n(z)}{f_n(z)-b}\right| = \left|\frac{1-\bar{a}_n g_n(z)}{g_n(z)-a_n}\right| \sim |a_n| \quad (n\to\infty) \tag{4.4}$$

because a_n is bounded away from zero.

If the probability measure μ_n on $\partial\mathbb{D}$ is defined [see (1.11)] by

$$\mu_n(A) = \sigma_n(b,A)/\sigma_n(b,\partial\mathbb{D})$$

then we obtain from (4.2), (4.3), and (4.4) that

$$\limsup_{n\to\infty}\int_{\partial\mathbb{D}}\frac{1-|z|^2}{|t-z|^2}\,d\mu_n(t) \leqq 1 \quad \text{for} \quad z\in\mathbb{D}.$$

It follows that μ_n converges weakly to the normalized Lebesgue measure on $\partial\mathbb{D}$, and this is equivalent to our assertion (1.12).

References

1. Aaronson, J.: A remark on the exactness of inner functions (to appear)
2. Collingwood, E.F., Lohwater, A.J.: The theory of cluster sets. Cambridge: Cambridge University Press 1966
3. Goodwyn, L.W.: Topological entropy bounds measure-theoretic entropy. Proc. Am. Math. Soc. **23**, 679–688 (1969)
4. Lehto, O., Virtanen, K.I.: Boundary behaviour and normal meromorphic functions. Acta Math. **97**, 47–65 (1957)
5. Neuwirth, J.H.: Ergodicity of some mappings of the circle and the line. Israel J. Math. **31**, 359–367 (1978)
6. Pommerenke, Ch.: Univalent functions. Göttingen: Vandenhoeck and Ruprecht 1975
7. Pommerenke, Ch.: On the iteration of analytic functions in a halfplane. I. J. London Math. Soc. **19**, 439–447 (1979)
8. Pommerenke, Ch.: On asymptotic iteration of analytic functions in the disk. Analysis (to appear)
9. Valiron, G.: Fonctions analytiques. Paris: Presses Universitaires de France 1954
10. Walters, P.: Ergodic theory – Introductory lectures. Lecture Notes in Mathematics, Vol. 458. Berlin, Heidelberg, New York: Springer 1975

Received November 18, 1980

Math. Ann. 256, 51–62 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Fermatquadrupel

Werner Meyer[1] und Wolfram Neutsch[2]

1 Mathematisches Institut der Universität, Beringstrasse 1, D-5300 Bonn, Bundesrepublik Deutschland
2 Institut für Astrophysik und extraterrestrische Forschung, Auf dem Hügel 71, D-5300 Bonn, Bundesrepublik Deutschland

1. Einführung

Bei der Berechnung der Picardzahl $\varrho(N)$ der Fermatfläche $F_N (N \in \mathbb{N})$ im dreidimensionalen komplexen projektiven Raum mit der Gleichung $z_1^N + z_2^N + z_3^N + z_4^N = 0$ spielen die in der vorliegenden Untersuchung betrachteten "Fermatquadrupel" (4) eine besondere Rolle. Nach Hirzebruch u. Zagier [1] ist $\varrho(N) = 1 + (\# \mathfrak{F}_N)$. Diese Anzahl wurde für Primzahlen N durch Parry [3] in einem allgemeineren Zusammenhang ermittelt. Unter Benutzung des Nichtverschwindens Dirichletscher L-Reihen an der Stelle 1 bewies er, daß in diesem Falle lediglich die sogenannten "trivialen" Lösungen der Form $(a, N-a, b, N-b)$ auftreten. Diese Quadrupel entsprechen den Geraden auf der Fermatfläche F_N. In dieser Arbeit konstruieren wir einen elementaren Beweis für die genannte Aussage.

Neben den trivialen Lösungen fanden wir drei weitere Familien von Fermatquadrupeln („Standardquadrupel"), außerdem für 22 Werte von $N \leqq 614$ noch zusätzliche teilerfremde Lösungen („Ausnahmequadrupel"). Die letzte teilerfremde Ausnahmelösung im betrachteten Bereich erhielten wir für $N = 180$.

Die Verteilung der bekannten Ausnahmelösungen (vgl. Tabelle 1) läßt vermuten, daß außer den genannten keine weiteren Fermatquadrupel existieren. Für den Spezialfall $N = 2p$, p Primzahl, wird die Vermutung in dieser Arbeit bewiesen. Verfeinerungen der benutzten Verfahren sollten es ermöglichen, auch z.B. die Fälle $N = 3p$ und $N =$ Primzahlpotenz zu behandeln.

Aufgrund des Zusammenhangs mit der Picardzahl stellt sich die Frage, welche Kurven auf F_N zu den nichttrivialen Fermatquadrupeln gehören. Dieses Problem bleibt einer späteren Untersuchung vorbehalten.

2. Eine Vermutung über die Fermatquadrupel

Im folgenden sei stets: N eine natürliche Zahl,

$$R_N = [0, N) \cap \mathbb{Z}, \tag{1}$$

$$G_N = \{t \in R_N ; ggT(t, N) = 1\} \tag{2}$$

0025-5831/81/0256/0051/$02.40

und $\mathfrak{G}_N$ sei die multiplikative Gruppe aller zu N teilerfremden Reste. Für $k\in\mathbb{Z}$ bezeichne

$$\bar{k}=k-\left[\frac{k}{N}\right]N \tag{3}$$

die kleinste nichtnegative mod$N$ zu k kongruente Zahl.

Ein 4-Tupel $L=(a_1\ a_2,\ a_3,\ a_4)\in R_N^4$ heiße *N-Fermatquadrupel*, wenn gilt:

$$\bigwedge_{t\in G_N}\sum_{i=1}^{4}B_1\left(\frac{ta_i}{N}\right)=0 \quad \text{und} \quad a_i\neq 0 \tag{4}$$

wobei

$$B_1(x)=\begin{cases}x-[x]-\frac{1}{2} & \text{für} \quad x\in\mathbb{R}\setminus\mathbb{Z}\\ 0 & \text{für} \quad x\in\mathbb{Z}\end{cases} \tag{5}$$

das erste Bernoulli„polynom" ist. Die Menge aller N-Fermatquadrupel werde mit $\mathfrak{F}_N$ bezeichnet.

Die Bedingung (4) ist gleichwertig mit

$$\bigwedge_{t\in G_N}\sum_{i=1}^{4}\overline{ta_i}=2N \tag{6}$$

und mit

$$\bigwedge_{t\in G_N}\sum_{i=1}^{4}\left[\frac{ta_i}{N}\right]=2t-2. \tag{7}$$

Aus dieser Definition folgt für alle $k\in\mathbb{N}$

$$L=(a_1,a_2,a_3,a_4)\in\mathfrak{F}_N \Leftrightarrow kL:=(ka_1,ka_2,ka_3,ka_4)\in\mathfrak{F}_{kN}. \tag{8}$$

Daher genügt es, nur die Menge

$$\mathfrak{F}_N^*=\{(a_1,a_2,a_3,a_4)\in\mathfrak{F}_N;ggT(a_1,a_2,a_3,a_4,N)=1\} \tag{9}$$

zu betrachten. Es ist leicht zu sehen, daß jeder Teiler von a_1, a_2, a_3, a_4 auch N teilt. Daher ist auch

$$\mathfrak{F}_N^*=\{(a_1,a_2,a_3,a_4)\in\mathfrak{F}_N;ggT(a_1,a_2,a_3,a_4)=1\}. \tag{10}$$

Mit $L=(a_1,a_2,a_3,a_4)\in\mathfrak{F}_N^*$ und $t\in\mathfrak{G}_N$ ist auch

$$\overline{tL}=(\overline{ta_1},\overline{ta_2},\overline{ta_3},\overline{ta_4})\in\mathfrak{F}_N^* \tag{11}$$

und umgekehrt.

Ferner gilt offenbar:

$$\bigwedge_{\pi\in S_4}(a_{\pi(1)},a_{\pi(2)},a_{\pi(3)},a_{\pi(4)})\in\mathfrak{F}_N^* \Leftrightarrow a_1,a_2,a_3,a_4)\in\mathfrak{F}_N^*. \tag{12}$$

Damit operiert die Gruppe $\mathfrak{G}_N\times S_4$ in natürlicher Weise auf $\mathfrak{F}_N^*$ und auf $\mathfrak{F}_N$. Ist $a_i+a_j=N$ für ein Paar (i,j) mit $i\neq j$, so gilt dies auch für $\overline{ta_i}+\overline{ta_j}$ mit $t\in G_N$. Daraus ergibt sich sofort, daß alle 4-Tupel der Gestalt $(a, N\text{–}a, b, N\text{–}b)$ und ihre Permutationen N-Fermatquadrupel sind. Diese nennen wir *triviale Fermatquadrupel*.

Es ist leicht zu beweisen, daß die folgenden 4-Tupel:

$$\text{für } N=2K\colon\ L_1=(1, K, K+1, 2K-2), \tag{13}$$

$$\text{für } N=2K\colon\ L_2=(1, K+1, K+2, 2K-4), \tag{14}$$

$$\text{für } N=3K\colon\ L_3=(1, K+1, 2K+1, 3K-3) \tag{15}$$

und ihre Bilder unter $\mathfrak{G}_N \times S_4$ ebenfalls N-Fermatquadrupel sind. Wir nennen sie *Standardquadrupel 1., 2., 3. Art.* Für $N \geqq 12$ sind sie stets nichttrivial und paarweise verschieden voneinander.

Alle Elemente aus $\mathfrak{F}_N^*$, die weder trivial noch Standardquadrupel sind, heißen *Ausnahmequadrupel.* Die Menge der Ausnahmequadrupel zu N sei $\mathfrak{A}_N$.

Mit Hilfe der Rechenanlagen des Regionalen Hochschulrechenzentrums Bonn und des Max-Planck-Instituts für Radioastronomie Bonn wurden alle Fermatquadrupel für $N \leqq 614$ ermittelt. Hierbei fanden sich die in Tabelle 1 aufgeführten Ausnahmequadrupel.

Aus jedem Transitivitätsbereich der Gruppe $\mathfrak{G}_N \times S_4$ auf $\mathfrak{F}_N^*$ wurde jeweils genau eine Lösung angegeben.

Eine genauere Betrachtung der Tabelle 1 führt zu folgender

Vermutung 1. Außer den genannten existieren keine weiteren Ausnahmequadrupel.

Standardquadrupel treten nur auf, wenn $2|N$ oder $3|N$ ist. Die Tabelle 1 zeigt, daß dies auch für die Ausnahmequadrupel mit $N \leqq 614$ richtig ist.

Dies führt zu der

Vermutung 2. Sei $ggT(N, 6)=1$, $L \in \mathfrak{F}_N \Rightarrow L$ ist eine triviale Lösung.

3. Vorbereitende Bemerkungen

In diesem Abschnitt stellen wir einige häufiger benötigte Definitionen und Gleichungen zusammen.

Seien $N \in \mathbb{N}$, $x \in \mathbb{C}$ eine N-te Einheitswurzel; $a \in R_N$. Dann ist:

$$\sum_{t \in R_N} x^t = \begin{cases} 0 & \text{für } x \neq 1 \\ N & \text{für } x = 1, \end{cases} \tag{16}$$

$$\sum_{t \in R_N} t x^t = \begin{cases} \dfrac{N}{x-1} & \text{für } x \neq 1 \\ \dfrac{N(N-1)}{2} & \text{für } x = 1. \end{cases} \tag{17}$$

Wir setzen

$$T(N; x, a) = \sum_{t \in R_N} x^t B_1\left(\frac{ta}{N}\right), \tag{18}$$

$$S(N; x, a) = \sum_{t \in G_N} x^t B_1\left(\frac{ta}{N}\right). \tag{19}$$

Tabelle 1. Tabelle der bekannten Ausnahmequadrupel. Aus jedem Transitivitätsbereich der Gruppe $\mathfrak{G}_N \times S_4$ ist genau ein Vertreter $L=(a, b, c, d)$ aufgeführt; in der letzten Spalte steht die Länge l des L enthaltenden $\mathfrak{G}_N$-Transitivitätsbereiches

N	a	b	c	d	l
12	1	4	9	10	4
	1	6	8	9	4
14	1	7	9	11	2
15	1	6	10	13	8
18	1	6	14	15	6
	1	7	12	16	6
	1	9	12	14	6
20	1	4	17	18	8
	1	6	16	17	8
	1	9	13	17	2
	1	10	12	17	8
21	1	4	18	19	12
24	1	6	19	22	8
	1	8	17	22	8
	1	8	19	20	4
	1	11	17	19	2
	1	12	16	19	4
	1	12	17	18	4
	1	13	16	18	4
	2	9	16	21	4
28	1	9	21	25	4
	1	15	18	22	6
30	1	4	27	28	8
	1	7	25	27	8
	1	8	25	26	8
	1	10	24	25	8
	1	11	24	24	2
	1	12	20	27	8
	1	12	23	24	8
	1	15	17	27	8
	1	15	19	25	4
	1	15	20	24	8
	1	16	21	22	8
	1	17	19	23	2
30	1	18	20	21	8
	1	19	20	20	2
	2	15	21	22	4
36	1	19	24	28	6
	2	9	28	33	12
40	1	21	24	34	8
	1	21	26	32	8
42	1	6	37	40	12
	1	6	38	39	12
	1	8	37	38	12
	1	12	33	38	12
	1	13	32	38	6
	1	15	31	37	12
	1	15	32	36	12
	1	16	30	37	12
	1	16	33	34	12
	1	18	32	33	12
	1	19	24	40	12
	1	21	24	38	12
	1	21	25	37	4
	1	21	29	33	6
	1	21	30	32	12
	1	24	29	30	6
48	1	25	32	38	8
	1	25	34	36	8
60	1	12	49	58	16
	1	15	49	55	8
	1	17	49	53	4
	1	20	41	58	16
	1	20	49	50	8
	1	23	47	49	4
	1	24	41	54	8
	1	24	46	49	16
	1	25	45	49	8
	1	27	41	51	8
60	1	29	41	49	4
	1	30	40	49	8
	1	30	41	48	8
	1	31	34	54	8
	1	31	38	50	8
	1	31	40	48	8
	1	31	42	46	8
	1	36	41	42	8
	2	15	48	55	16
	2	21	40	57	16
	2	21	46	51	8
	2	25	38	55	4
	3	32	33	52	4
66	1	25	44	62	20
	2	39	45	46	10
72	3	16	57	68	12
78	1	32	61	62	24
	1	39	55	61	8
84	1	29	63	75	12
	1	43	48	76	12
	1	43	50	74	12
	1	43	58	66	12
	1	43	60	64	12
	2	33	58	75	6
90	3	20	72	85	24
120	1	49	83	107	8
	1	61	80	98	16
	1	61	82	96	16
	2	25	98	115	16
	4	25	96	115	16
156	1	79	110	122	24
180	3	40	147	170	24

Das letzte bekannte Ausnahmequadrupel tritt somit für $N=180$ auf. Die Anzahl der verschiedenen $\mathfrak{G}_N \times S_4$- bzw. S_4-Transitivitätsbereiche ist 101 bzw. 920

Dann gelten folgende Beziehungen:

$$T(N; x, a) = \sum_{d|N} \sum_{\substack{0 \leqq t < \frac{N}{d} \\ \left(t, \frac{N}{d}\right)=1}} x^{dt} B_1\left(\frac{dta}{N}\right) = \sum_{d|N} S\left(\frac{N}{d}; x^d, a\right). \tag{20}$$

Die Möbiussche Umkehrformel liefert

$$S(N; x, a) = \sum_{d|N} \mu(d) T\left(\frac{N}{d}; x^d, a\right). \tag{21}$$

Für die Berechnung von $T(N;x,a)$ und $S(N;x,a)$ sind folgende Formeln nützlich:

Sei $d=ggT(a,N)$, $N=dn$, $a=d\alpha$, dann ist

$$T(N;x,a)=\begin{cases}0 & \text{für} \quad x^n\neq 1\\ d\cdot T(n;x,\alpha) & \text{für} \quad x^n=1\end{cases} \tag{22}$$

außerdem für $\varepsilon\in\mathbb{Z}$, $ggT(\varepsilon,N)=1$:

$$T(N;x,a)=T(N;x^{\varepsilon},\overline{a\varepsilon}), \tag{23}$$

$$S(N;x,a)=S(N;x^{\varepsilon},\overline{a\varepsilon}), \tag{24}$$

$$T(N;x,1)=\begin{cases}\frac{1}{2}\frac{x+1}{x-1}; & x\neq 1\\ 0 & x=1.\end{cases} \tag{25}$$

Die Gln. (22), (23), (25) lassen sich zusammenfassen zu: Sei $d=ggT(N,a)$, $ab\equiv d\,(\mathrm{mod}\,N)$, $ggT(b,N)=1$, $n=N/d$. Dann ist

$$T(N;x,a)=\begin{cases}\frac{d}{2}\frac{x^b+1}{x^b-1} & \text{für} \quad x^n=1; \quad x\neq 1\\ 0 & \text{für} \quad x=1 \quad \text{oder} \quad x^n\neq 1\end{cases} \tag{26}$$

also auch

$$T(N;x,a)=\frac{d}{2i}\operatorname{ctg}\frac{tb\pi}{n} \quad \text{für} \quad x=e^{\frac{2\pi it}{n}}\neq 1. \tag{27}$$

4. Beweis der Vermutung für einige Spezialfälle

Die Bedingung (4), daß $L=(a_1,a_2,a_3,a_4)$ ein Fermatquadrupel zu N ist, impliziert für jede N-te Einheitswurzel x:

$$\sum_{i=1}^{4} S(N;x,a_i)=0. \tag{28}$$

Als erstes untersuchen wir den Fall: $N\in\mathbb{P}=\{p\in\mathbb{N};\ p\,\text{prim}\}$.

Satz 1 (Parry). *Sei N eine Primzahl. Dann ist jedes N-Fermatquadrupel trivial.*

Beweis. Für $N\in\mathbb{P}$ und eine N-te Einheitswurzel x ist $\mathfrak{F}_N=\mathfrak{F}_N^*$ und $S(N;x,a)=T(N;x,a)$. Sei $L=(a_1,a_2,a_3,a_4)\in\mathfrak{F}_N^*$ und seien $b_i\in G_N$ mit $a_ib_i\equiv 1(N)$. Dann geht (28) unter Benutzung von (26) über in

$$\frac{1}{2}\sum_{i=1}^{4}\frac{x^{b_i}+1}{x^{b_i}-1}=0 \quad \text{für alle } x; \quad x^N=1\neq x. \tag{29}$$

Nach Beseitigung der Nenner wird daraus

$$\bigwedge_{x^N=1\neq x} 2-x^{b_1}-x^{b_2}-x^{b_3}-x^{b_4}$$
$$+x^{b_1+b_2+b_3}+x^{b_1+b_2+b_4}+x^{b_1+b_3+b_4}+x^{b_2+b_3+b_4}-2x^{b_1+b_2+b_3+b_4}=0. \tag{30}$$

Die letzte Gleichung gilt auch für $x=1$. Daraus folgt, daß die linke Seite als Polynom in x ein Vielfaches von (x^N-1) ist. Das bedeutet: Nach Reduktion der Exponenten $\bmod N$ auf R_N müssen sich alle Glieder paarweise zu 0 ergänzen. Wegen $0<b_i<N$ kann sich der erste Summand nur gegen den letzten wegheben:

$$\bigwedge_{x^N=1} 2-2x^{b_1+b_2+b_3+b_4}=0 \tag{31}$$

also

$$b_1+b_2+b_3+b_4\equiv 0 \pmod N. \tag{32}$$

Daraus folgt weiter

$$\bigwedge_{x^N=1} -x^{b_1}-x^{b_2}-x^{b_3}-x^{b_4}+x^{N-b_1}+x^{N-b_2}+x^{N-b_3}+x^{N-b_4}=0. \tag{33}$$

Da der Grad dieses Polynoms kleiner als N ist, muß die linke Seite von (33) identisch verschwinden, d. h.

$$\{b_1,b_2,b_3,b_4\}=\{N-b_1,N-b_2,N-b_3,N-b_4\}. \tag{34}$$

Dasselbe gilt für die zu den $b_i \pmod N$ reziproken $a_i \Rightarrow L$ ist triviales Fermatquadrupel, w.z.b.w.

Bemerkung. 1. Satz 1 ist ein Sonderfall eines Satzes von Parry. Im Unterschied zum obigen elementaren Beweis benutzt Parry das Nichtverschwinden gewisser Dirichletscher L-Reihen an der Stelle 1.

2. Satz 1 läßt sich auch [ausgehend von (27)] durch Abschätzungen der Kotangensfunktion herleiten.

Satz 2. *Sei* $p\in\mathbb{P}$, $N=2p$ *und* L *ein Ausnahmequadrupel zu* N. *Dann gilt:* $p=7$ *und* $L\sim(1,7,9,11) \bmod \mathfrak{G}_N\times S_4$.

Zum Beweis benötigen wir den folgenden

Hilfssatz. *Sei* $p\in\mathbb{P}$, $0<\mu<2p$, $0<\sin\frac{\mu\pi}{p}\leqq 3\sin\frac{\pi}{p}$. *Dann ist* $0<\mu<4$ *oder* $0<p-\mu<4$.

Beweis. Wegen $0<\sin\frac{\mu\pi}{p}$ und der Symmetrie des Sinus dürfen wir $0<\mu\leqq\frac{p}{2}$ annehmen. Für $p\geqq 5$ ist $\sin\frac{4\pi}{p}\Big/\sin\frac{\pi}{p}=4\cos\frac{\pi}{p}\cos\frac{2\pi}{p}>4\cos^2\frac{2\pi}{p}$. Weiter ist für $p\geqq 13$: $4\cos^2\frac{2\pi}{p}>3$, also $3\sin\frac{\pi}{p}<\sin\frac{4\pi}{p}$. Die letzte Ungleichung gilt auch für $p=11$. Wegen der Monotonie des Sinus in $\left[0,\frac{\pi}{2}\right]$ folgt daraus die Behauptung. Für $p\leqq 7$ ist die Aussage des Hilfssatzes trivial.

Beweis des Satzes. Für $p=2$ ist Satz 2 trivial.
Sei $p\geqq 3$ und $L=(\alpha,\beta,\gamma,\delta)$. Wir unterscheiden zwei Fälle:

1. $2\nmid\alpha\beta\gamma\delta$

oder

2. $2\mid\alpha\beta\gamma\delta$.

Fall 1. Sei zunächst $p \nmid \alpha\beta\gamma\delta$ und seien a, b, c, d definiert durch

$$a\alpha \equiv b\beta \equiv c\gamma \equiv d\delta \equiv 1 \pmod{N}. \tag{35}$$

Die Formeln (21) und (26) ergeben für alle $t \in \mathbb{Z}$ mit $p \nmid t$:

$$S(N; x, \alpha) = \frac{1}{2i} \sin^{-1}\left(\frac{2ta\pi}{p}\right) \tag{36}$$

wobei $\sin^{-1}(x) = (\sin(x))^{-1}$ ist und

$$x = e^{\frac{2\pi i t}{p}} \tag{37}$$

gesetzt wurde. Die Bedingung (28) für Fermatquadrupel liefert dann:

$$\sin^{-1}\left(\frac{2ta\pi}{p}\right) + \sin^{-1}\left(\frac{2tb\pi}{p}\right) + \sin^{-1}\left(\frac{2tc\pi}{p}\right) + \sin^{-1}\left(\frac{2td\pi}{p}\right) = 0. \tag{38}$$

Mit $t = (p-\alpha)/2$ erhalten wir

$$\sin^{-1}\left(\frac{b^*\pi}{p}\right) + \sin^{-1}\left(\frac{c^*\pi}{p}\right) + \sin^{-1}\left(\frac{d^*\pi}{p}\right) = \sin^{-1}\left(\frac{\pi}{p}\right) \tag{39}$$

wobei

$$0 < b^*, c^*, d^* < 2p$$

und

$$b^* \equiv -\alpha b; \quad c^* \equiv -\alpha c; \quad d^* \equiv -\alpha d \pmod{2p} \tag{40}$$

ist. O.B.d.A. sei

$$\sin^{-1}\left(\frac{b^*\pi}{p}\right) \geqq \sin^{-1}\left(\frac{c^*\pi}{p}\right) \geqq \sin^{-1}\left(\frac{d^*\pi}{p}\right) \tag{41}$$

und somit

$$3 \sin^{-1}\left(\frac{b^*\pi}{p}\right) \geqq \sin^{-1}\frac{\pi}{p} > 0. \tag{42}$$

Aus dem Hilfssatz folgt daher

$$0 < b^* < 4 \quad \text{oder} \quad 0 < p - b^* < 4. \tag{43}$$

Wegen (40) ist b^* ungerade, also

$$b^* = 3 \quad \text{oder} \quad b^* = p - 2 \tag{44}$$

da $b^* = 1$ offensichtlich nur triviale Lösungen liefert.

Sei $b^* = 3$. Dann ergeben sich für $p \leqq 7$ nur triviale Fermatquadrupel. Für $p \geqq 11$ erhalten wir die Ungleichungen

$$2 \sin^{-1}\frac{c^*\pi}{p} \geqq \sin^{-1}\frac{c^*\pi}{p} + \sin^{-1}\frac{d^*\pi}{p} = \sin^{-1}\frac{\pi}{p} - \sin^{-1}\frac{3\pi}{p} > 2 \sin^{-1}\frac{4\pi}{p}, \tag{45}$$

$$0 < c^* < 4 \quad \text{oder} \quad 0 < p - c^* < 4 \Rightarrow c^* \in \{1, 3, p-2\}. \tag{46}$$

$c^* \in \{1, 3\}$ führt auf triviale Quadrupel.

Somit ist in jedem Fall eine der Zahlen b^*, c^*, d^* gleich $p-2$ und wegen (35) und (40) gilt

$$\alpha \equiv (p+2)\beta, (p+2)\gamma \quad \text{oder} \quad (p+2)\delta \pmod{2p}. \tag{47}$$

Entsprechende Kongruenzen gelten auch für β, γ und δ.

Daher bewirkt die Multiplikation mit $p+2$ eine Permutation von L.

Ist jedoch $p|\alpha\beta\gamma\delta$, so dürfen wir o.B.d.A. $p=\delta$ und $p\nmid\alpha\beta\gamma$ voraussetzen. Wenn man alle Terme wegläßt, die d oder d^* enthalten, liefern die gleichen Überlegungen wie oben, daß die Multiplikation mit $p+2$ eine Permutation von α, β, γ (und somit von L) bewirkt.

Sei k die Ordnung der durch Multiplikation mit $p+2$ entstehenden Permutation von L. Dann ist

$$1 \leqq k \leqq 4 \quad \text{und} \quad (p+2)^k \equiv 1 \pmod{2p} \tag{48}$$

also auch

$$2^k \equiv 1 \pmod{p} \tag{49}$$

woraus $p \leqq 7$ folgt.

Für $p=3$ ist $(p+2) \equiv -1 \bmod 2p$ und für $p=5$: $(p+2)^2 \equiv -1 \bmod 2p$, dies widerspricht der Nichttrivialität von L.

Demnach ist $p=7$, $N=14$, $p+2=9$, $k=3$ und

$$L \sim (\bar{1}, \bar{9}, \overline{9^2}, \bar{7}) = (1, 9, 11, 7) \mod \mathfrak{G}_{14} \times S_4. \tag{50}$$

Damit ist Fall 1 abgeschlossen.

Fall 2. Wir dürfen o.B.d.A. annehmen, daß $2\nmid\alpha\beta$, $2|\gamma$, $2|\delta$ gilt.

Sei zunächst $p|\beta$, also $p=\beta$. Wir definieren a, c und d durch:

$$0<a<2p, \quad 0<c,d<p; \quad a\alpha \equiv 1, \quad c\gamma \equiv d\delta \equiv 2 \pmod{2p}. \tag{51}$$

Formeln (21) und (26) ergeben damit für alle t mit $p\nmid t$:

$$\sin^{-1}\frac{2ta\pi}{p} + \operatorname{ctg}\frac{tc\pi}{p} + \operatorname{ctg}\frac{td\pi}{p} = 0. \tag{52}$$

Wir setzen

$$t = \frac{p+\alpha}{2}; \quad c^* \equiv tc, \quad d^* \equiv td \pmod{p}; \quad 0<c^*, d^*<p \tag{53}$$

und erhalten

$$\operatorname{ctg}\frac{c^*\pi}{p} + \operatorname{ctg}\frac{d^*\pi}{p} = \sin^{-1}\frac{\pi}{p}. \tag{54}$$

Sei etwa

$$\operatorname{ctg}\frac{c^*\pi}{p} \geqq \operatorname{ctg}\frac{d^*\pi}{p}. \tag{55}$$

Dann folgt

$$\operatorname{ctg}\frac{c^*\pi}{p} \geqq \tfrac{1}{2}\sin^{-1}\frac{\pi}{p} > \sin^{-1}\frac{2\pi}{p} > \operatorname{ctg}\frac{2\pi}{p} > 0 \tag{56}$$

und

$$0 < c^* < 2\,; \quad c^* = 1\,; \quad c \equiv 2a, \quad \gamma \equiv \alpha \pmod{p} \tag{57}$$

sowie

$$\operatorname{ctg}\frac{d^*\pi}{p} = \sin^{-1}\frac{\pi}{p} - \operatorname{ctg}\frac{\pi}{p} = \frac{1-\cos\frac{\pi}{p}}{\sin\frac{\pi}{p}} = \operatorname{ctg}\frac{\left(\frac{p-1}{2}\right)\pi}{p} \tag{58}$$

folglich

$$d^* = \frac{p-1}{2}\,; \quad d \equiv -a, \quad \delta \equiv -2\alpha \pmod{p} \tag{59}$$

und

$$L \sim (1, p, p+1, 2p-2) \tag{60}$$

d.h. L ist eine Standardlösung 1. Art.

Daher kann β – und aus Symmetriegründen auch α – nicht durch p teilbar sein. Anstelle von (54) erhalten wir

$$\sin^{-1}\frac{b^*\pi}{p} + \operatorname{ctg}\frac{c^*\pi}{p} + \operatorname{ctg}\frac{d^*\pi}{p} = \sin^{-1}\frac{\pi}{p} \tag{61}$$

wobei neben (51) und (53) noch

$$b\beta \equiv 1, \quad b^* \equiv -\alpha b \pmod{2p}; \quad 0 < b, b^* < 2p \tag{62}$$

gilt.

Nun sind zwei Möglichkeiten zu unterscheiden:

Falls

$$\sin^{-1}\frac{b^*\pi}{p} \geqq \operatorname{ctg}\frac{c^*\pi}{p} \quad \text{und} \quad \sin^{-1}\frac{b^*\pi}{p} \geqq \operatorname{ctg}\frac{d^*\pi}{p} \tag{63}$$

ist, erhalten wir analog zu Fall 1, (42)–(46),

$$\begin{gathered} b^* = p-2\,; \quad b \equiv 2a, \quad \alpha \equiv 2\beta \pmod{p}\,; \\ \alpha \equiv p + 2\beta \pmod{2p}\,. \end{gathered} \tag{64}$$

Ist hingegen z.B.

$$\operatorname{ctg}\frac{c^*\pi}{p} \geqq \sin^{-1}\frac{b^*\pi}{p} \quad \text{und} \quad \operatorname{ctg}\frac{c^*\pi}{p} \geqq \operatorname{ctg}\frac{d^*\pi}{p} \tag{65}$$

so ergibt sich

$$\operatorname{ctg}\frac{c^*\pi}{p} > \tfrac{1}{3}\sin^{-1}\frac{\pi}{p} > \sin^{-1}\frac{3\pi}{p} > \operatorname{ctg}\frac{3\pi}{p}\,;$$

$$0<c^*<3\,; \quad c^*\in\{1,2\}, \tag{66}$$

$$\alpha\equiv p+\gamma \quad \text{oder} \quad \alpha\equiv p+2\gamma \pmod{2p}. \tag{67}$$

Insgesamt sind 5 Kongruenzen für α:

$$\begin{aligned}&\alpha\equiv p+2\beta\,; \quad \alpha\equiv p+\gamma\,; \quad \alpha\equiv p+\delta\,;\\ &\alpha\equiv p+2\gamma\,; \quad \alpha\equiv p+2\delta \pmod{2p}\end{aligned} \tag{68}$$

und ebensoviele für β:

$$\begin{aligned}&\beta\equiv p+2\alpha\,; \quad \beta\equiv p+\gamma\,; \quad \beta\equiv p+\delta\,;\\ &\beta\equiv p+2\gamma\,; \quad \beta\equiv p+2\delta \pmod{2p}\end{aligned} \tag{69}$$

möglich.

Wäre $\alpha=\beta$, so könnten wir durch Multiplikation mit einem geeigneten Element aus $\mathfrak{G}_N$ eine Lösung L' der Form

$$L'=(\alpha',\beta',\gamma',\delta')=(2p-1,2p-1,\gamma',\delta') \tag{70}$$

finden. Daraus folgt aber wegen (6):

$$\gamma'=\delta'=1 \tag{71}$$

d.h.: L' und damit auch L sind triviale Fermatquadrupel.

Angenommen, es sei $\gamma=\delta$. Dann dürfen wir o.B.d.A. annehmen:

$$\gamma=\delta=2p-2\,; \quad \alpha+\beta=4 \tag{72}$$

somit

$$\alpha=1\,; \quad \beta=3 \tag{73}$$

oder umgekehrt. Die letzten Gleichungen sind offenbar für $p>5$ mit den Kongruenzen (68) nicht verträglich. Für $p=3$ ist $L=(1,3,4,4)$. Das ist nicht zulässig, da $p\nmid\alpha\beta\gamma\delta$ ist. Es bleibt nur:

$$p=5\,; \quad L=(1,3,8,8)\sim(7,1,6,6) \pmod{\mathfrak{G}_{10}}. \tag{74}$$

Dies ist jedoch im Gegensatz zur Voraussetzung ein Standardquadrupel 2. Art. Da α und β ungerade, γ und δ gerade sind, haben wir gezeigt: α, β, γ, δ sind paarweise verschieden voneinander (und von N–α, N–β, N–γ, N–δ). Insbesondere ist $p\geqq 5$.

Sei nun $\beta\equiv p+2\alpha \pmod{2p}$. Falls $\alpha\equiv p+2\beta \pmod{2p}$ wäre, erhielten wir $\beta\equiv 4\beta \pmod{p}$, also $p=3$.

Falls $\alpha\equiv p+\gamma \pmod{2p}$ wäre, entstünde

$$aL=a(\alpha,\overline{p+2\alpha},\overline{p+\alpha},\delta)=(1,p+2,p+1,\overline{a\delta})\,; \quad \overline{a\delta}=2p-4 \tag{75}$$

also ein Standardquadrupel 2. Art. Sei schließlich $\alpha\equiv p+2\gamma \pmod{2p}$. Dann ist

$$L=(\overline{p+2\gamma},\overline{p+4\gamma},\gamma,\overline{-7\gamma}). \tag{76}$$

Falls $p \equiv 3 \pmod 4$ ist, setzen wir

$$\gamma = \frac{p+1}{2} \tag{77}$$

und erhalten

$$L = \left(1, p+2, \frac{p+1}{2}, \frac{5p-7}{2}\right). \tag{78}$$

Wegen $\delta < 2p$ folgt $p \leqq 5$, also $p = 3$, was bereits ausgeschlossen wurde. Folglich ist $p \equiv 1 \pmod 4$. Jetzt wählen wir

$$\gamma = \frac{p+3}{2}. \tag{79}$$

Mit $p = 5$ ergibt sich der Widerspruch

$$\alpha = 3, \quad \beta = 1, \quad \gamma = 4; \quad \delta = 12 > 2p. \tag{80}$$

Für $p > 5$ hingegen ist

$$L = \left(3, p+6, \frac{p+3}{2}, \frac{5p-21}{2}\right) \tag{81}$$

und wegen $\delta < 2p$ finden wir $p < 21$; $p \in \{13, 17\}$.

Durch Multiplikation von L mit $(p-2)$ ersehen wir, daß L kein Fermatquadrupel ist.

Wegen der Symmetrie in γ und δ können die beiden restlichen Möglichkeiten in (68) auch nicht eintreten. Also ist $\beta \not\equiv p + 2\alpha$ und aus dem nämlichen Grunde $\alpha \not\equiv p + 2\beta$. (68) und (69) reduzieren sich damit auf

$$\alpha \equiv p+\gamma; \quad \alpha \equiv p+\delta; \quad \alpha \equiv p+2\gamma \quad \text{oder} \quad \alpha \equiv p+2\delta \pmod{2p} \tag{82}$$

sowie

$$\beta \equiv p+\gamma; \quad \beta \equiv p+\delta; \quad \beta \equiv p+2\gamma \quad \text{oder} \quad \beta \equiv p+2\delta \pmod{2p}. \tag{83}$$

Da unsere Voraussetzungen in α und β bzw. γ und δ symmetrisch sind und wegen $\alpha \neq \beta$, $\gamma \neq \delta$ ersehen wir, daß nur 4 der 16 Kombinationen von (82) und (83) wesentlich verschieden sind. Diese lauten:

$$\alpha \equiv p+\gamma; \quad \beta \equiv p+\delta \pmod{2p}, \tag{84}$$

$$\alpha \equiv p+2\gamma; \quad \beta \equiv p+2\delta \pmod{2p}, \tag{85}$$

$$\alpha \equiv p+\gamma; \quad \beta \equiv p+2\delta \pmod{2p}, \tag{86}$$

$$\alpha \equiv p+\gamma; \quad \beta \equiv p+2\gamma \pmod{2p}. \tag{87}$$

Die letzte Gleichung führt – wie man mit $\gamma = p+1$ sofort sieht – auf ein Standardquadrupel 2. Art.

In den ersten beiden Fällen folgte mit Hilfe der Beziehung (6) und $p > 3$:

$$\gamma + \delta \equiv 0 \pmod p \tag{88}$$

und – weil γ und δ gerade sind

$$\gamma+\delta\equiv 0 \pmod{2p}; \quad \gamma+\delta=2p=N \tag{89}$$

d.h., L wäre ein triviales Fermatquadrupel.

Daher bleibt nur noch (86) zu behandeln.

Wir dürfen o.B.d.A. für δ einen beliebigen geradzahligen Wert in $(0, 2p)$ vorschreiben.

Sei zuerst $p\equiv 1 \pmod 3$. Dann wählen wir

$$\delta=\frac{4p+2}{3} \Rightarrow \beta=\frac{5p+4}{3}; \quad \beta+\delta=3p+2; \quad \alpha+\gamma=p-2. \tag{90}$$

Für $p\equiv 2 \pmod 3$ und $p>5$ hingegen setzen wir

$$\delta=\frac{4p+4}{3} \Rightarrow \beta=\frac{5p+8}{3}; \quad \beta+\delta=3p+4; \quad \alpha+\gamma=p-4. \tag{91}$$

Beide Aussagen sind mit $|\alpha-\gamma|=p$ unvereinbar.

Mit $p=5$ und $\delta=8$ erhalten wir endlich eine Standardlösung 2. Art, und zwar:

$$\delta=8; \quad \beta=1; \quad \alpha+\gamma=11; \quad |\alpha-\gamma|=5 \Rightarrow \alpha=3, \quad \gamma=8 \tag{92}$$

Damit ist Satz 2 vollständig bewiesen.

Literatur

1. Hirzebruch, F., Zagier, D.: The Atiyah-Singer theorem and elementary number theory. Mathematics Lecture Series 3. Boston: Publish or Perish 1974
2. Katsura, T., Oda, T., Yui, N.: The arithmetic of Fermat surfaces over finite fields (Vorabdruck)
3. Parry, C.: Private Mitteilung
4. Tate, J.T.: Algebraic cycles and poles of the zeta functions. In: Arithmetical and algebraic geometry, pp. 93–110. New York: Harper and Row 1965

Eingegangen am 24. November 1980

Math. Ann. 256, 63–66 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Non existence de relèvement pour certaines mesures finiement additives et retractés de $\beta\mathbb{N}$

Michel Talagrand

Équipe d'Analyse, Tour 46, Université Paris VI, 4, Place Jussieu, F-75230 Paris Cedex 05, France

Soit $\mathfrak{P}(\mathbb{N})$ l'ensemble des parties de $\mathbb{N}$. Une mesure positive finiement addive sur $\mathbb{N}$ est une application m de $\mathfrak{P}(\mathbb{N})$ dans $\mathbb{R}^+$ telle que $A\cap B=\emptyset\Rightarrow m(A\cup B)=m(A)+m(B)$. Un *relèvement* ϱ de m est un homorphisme d'algèbres de Boole de $\mathfrak{P}(\mathbb{N})$ dans $\mathfrak{P}(\mathbb{N})$ tel que pour tout A on ait $m(A\Delta\varrho(A))=0$ et $m(A)=0\Rightarrow\varrho(A)=\emptyset$. Les faits suivants sont expliqués en détail dans [1]. On les rappelle brièvement. Puisque $\mathfrak{P}(\mathbb{N})$ est l'algèbre des ouverts-fermés de $\beta\mathbb{N}$, ϱ définit une application continue θ de $\beta\mathbb{N}$ dans $\beta\mathbb{N}$. On peut identifier de façon naturelle m à une mesure de Radon sur $\beta\mathbb{N}$, et on appelle $S(m)$ le fermé de $\beta\mathbb{N}$ support de cette mesure. Pour toute partie A de $\mathbb{N}$, désignons par $\bar{A}$ son adhérence dans $\beta\mathbb{N}$. Pour $x\in\beta\mathbb{N}$, on a $\theta(x)\in\bar{A}\Leftrightarrow x\in\varrho(A)$. Pour $n\in\mathbb{N}$, on a donc $\theta(n)\in\bar{A}\Rightarrow n\in\varrho(A)\Rightarrow m(A)>0$, ce qui montre que $\theta(n)\in S(m)$. On a donc $\theta(\beta\mathbb{N})\subset S(m)$. Réciproquement, si $x\in S(m)$, pour tout $A\subset\mathbb{N}$, on a $x\in\bar{A}\Rightarrow x\in\overline{\varrho(A)}$ puisque $\overline{A\Delta\varrho(A)}\cap S(m)=0$, étant donné que $m(A\Delta\varrho(A))=0$. On a donc $x\in\bar{A}\Rightarrow\theta(x)\in\bar{A}$, i.e. $x=\theta(x)$. On a donc montré que θ est une *rétraction* de $S(m)$. [Réciproquement, on peut montrer que toute rétraction de $S(m)$ permet de définir un relèvement de m.]

Il est montré dans [1] que $S(m)$ n'est pas nécessairement séparable. Dans ce cas, la non-existence de ϱ est évidente. Il n'en est pas de même quand $S(m)$ est séparable, par exemple si la mesure de Radon associée est atomique et portée par $\beta\mathbb{N}\backslash\mathbb{N}$. Le but de ce travail est de prouver que sous l'hypothèse du Continu (HC) on a

Théorème (HC). *Il existe un fermé séparable S de $\beta\mathbb{N}\backslash\mathbb{N}$ qui n'est pas un retract de $\beta\mathbb{N}$. Ainsi aucune mesure finiement additive sur $\mathbb{N}$ qui admet S comme support n'admet de relèvement.*

La contruction est délicate. Elle utilise deux lemmes.

Pour $A\subset\mathbb{N}$, on pose $\mathbb{N}\backslash A=A^c$. On dira qu'une suite $(\mathscr{F}_n)$ d'ultrafiltre sur une algèbre $\mathscr{A}$ est *dense en soi* si elle satisfait la condition suivante:
(D) $\forall A\in\mathscr{A}$, l'ensemble $\{n; A\in\mathscr{F}_n\}$ est soit vide, soit infini.

Lemme 1. *Soit $\mathscr{A}$ une sous-algèbre de $\mathfrak{P}(\mathbb{N})$, et $(\mathscr{F}_n)$ une suite dense en soi d'ultrafiltres sur $\mathscr{A}$. Soit $\mathscr{B}\subset\mathfrak{P}(\mathbb{N})$. Alors il existe pour tout n un ultrafiltre $\mathscr{G}_n$ sur*

0025-5831/81/0256/0063/$01.00

l'algèbre $\mathscr{A}'$ engendrée par $\mathscr{A}$ et $\mathscr{B}$, plus fin que $\mathscr{F}_n$, et tel que la suite $(\mathscr{G}_n)$ soit dense en soi.

Preuve. Soit $\mathscr{H}$ un filtre sur $\mathscr{A}'$, dont la restriction à $\mathscr{A}$ soit égale à l'intersection des $\mathscr{F}_n$. D'après le lemme de Zorn, on peut supposer $\mathscr{H}$ maximal pour cette propriété. Pour tout n soit $\mathscr{G}_n$ un ultrafiltre sur $\mathscr{A}'$ plus fin que $\mathscr{H}$ et que $\mathscr{F}_n$. Soit $B\in\mathscr{A}'$ qui contient un $\mathscr{G}_{n_0}$. Alors la maximalité de $\mathscr{H}$ montre que la restriction à $\mathscr{A}$ du filtre $\mathscr{H}\wedge B^c$ engendré par $\mathscr{H}$ et B^c ne contient pas l'ùn des $\mathscr{F}_p$, soit $\mathscr{F}_{p_0}$. Il existe donc $A\in\mathscr{A}$, avec $A\in\mathscr{F}_p$ et $A^c\in\mathscr{H}\wedge B^c$. Pour tout p tel que $A\in\mathscr{F}_p$, on a donc $B\in\mathscr{G}_p$. La preuve est terminée.

Voici maintenant le point crucial.

Lemme 2. *Soit $\mathscr{A}$ une sous-algèbre dénombrable de $\mathfrak{P}(\mathbb{N})$, contenant les parties finies de $\mathbb{N}$. Soit $\mathscr{F}_n$ une suite d'ultrafiltres sur $\mathscr{A}$, dense en soi. Soit $\mathscr{H}_n$ des filtres sur $\mathscr{A}$ tels que $\mathscr{H}_n\not\subset\mathscr{F}_1$ et que pour tout n, $\mathscr{H}_n\subset\mathscr{F}_p$ pour au plus un p. Supposons de plus que pour tout p, l'ensemble $I_p=\{n\in N;\mathscr{H}_n=\mathscr{H}_p\}$ appartienne à $\mathscr{A}$, et que les $\mathscr{F}_p$ et les $\mathscr{H}_p$ contiennent les complémentaires des parties finies de $\mathbb{N}$. Alors il existe une partie $X\subset\mathbb{N}$ et une suite d'ultrafiltes $\mathscr{G}_n$ sur l'algèbre $\mathscr{A}'$ engendrée par $\mathscr{A}$ et X, plus fins que $\mathscr{F}_n$ dense en soi, avec $X\in\mathscr{G}_1$, et de sorte que pour $a\in X$ on ait $X^c\in\mathscr{H}_a$ (dans cette écriture, on identifie $\mathscr{H}_a$ et le filtre qu'il engendre sur $\mathbb{N}$).*

Preuve. Soit (A_n) une énumération des éléments infinis de $\mathscr{A}$, telle que chaque ensemble soit ènuméré une infinité de fois, et que $A_1\in\mathscr{F}_1$.

On va construire une partie infinie K de $\mathbb{N}$, pour $p\in K$, un entier $m_p\in\mathbb{N}$, un entier a_p, une partie $H_p\in\mathscr{A}$ et pour $p\in\mathbb{N}$ un entier ℓ_p, un entier b_p, de sorte que les conditions suivantes soient vérifiées pour tout n, où l'on pose $K_n=K\cap[1\ n]$.

(1) $1\in K$, $m_1=1$

(2) Les entiers m_p, ℓ_q pour $p\in K_n$, $q\in\mathbb{N}$, $q\leqq n$, sont deux à deux distincts

(3) Les entiers a_p, b_q, pour $p\in K_n$, $q\in\mathbb{N}$, $q\leqq n$, sont deux à deux distincts.

(4) S il existe k avec $A_n\in\mathscr{F}_k$, on a $A_n\in\mathscr{F}_{\ell_n}$, $b_n\in A_n$.

(5) On a $n\in K_n$ si et seulement s'il existe $p\in K_{n-1}$ tel que $A_n\in\mathscr{F}_{m_p}$.

(6) Si $n\in K_n$, $a_n\in A_n\setminus\bigcup_{i\in K_{n-1}}H_{a_i}$.

(7) Si $n\in K_n$, $\mathscr{H}_{a_n}\not\subset\mathscr{F}_{m_i}$ pour $i\in K_{n-1}$.

(8) Si $n\in K_n$, on a $H_{a_n}\in\mathscr{H}_{a_n}$, $H_{a_n}\notin\mathscr{F}_{m_i}$, pour $i\in K_{n-1}$, $a_i\notin H_{a_n}$ pour $i\in K_n$.

(9) Si $n\in K_n$, $A_n\setminus\bigcup_{i\in K_n}H_{a_i}\in\mathscr{F}_{m_n}$.

(10) Si $n\in K_n$, $\mathscr{F}_{m_n}$ ne contient aucun $\mathscr{H}_k$ pour $I_k\in\mathscr{F}_{m_i}$, $i\in K_n$.

(11) Si $n\in K_n$, et si $I_k\in\mathscr{F}_{m_n}$, on a $\mathscr{H}_k\not\subset\mathscr{F}_{m_i}$ pour $i\in K_{n-1}$.

Pour données initiales on prends a_1 quelconque dans A_1, puis $H_{a_1}\in\mathscr{H}_{a_1}$ avec $H_{a_1}\notin\mathscr{F}_1$, ce qui est possible puisque par hypothèse $\mathscr{H}_{a_1}\not\subset\mathscr{F}_1$, et on prends b_1 quelconque $b_1\neq a_1$.

Supposons la construction effectuée au rang n. Pour b_n on prend un entier $\in A_{n+1}$, différent de ceux déjà construits. S'il existe k avec $A_{n+1}\in\mathscr{F}_k$, on prends pour ℓ_{n+1} un entier quelconque différent de ceux déjà construits et tel que $A_{n+1}\in\mathscr{F}_{\ell_{n+1}}$, s'il n'existe pas de $p\in K_n$ tel que $A_{n+1}\in\mathscr{F}_{m_p}$, la construction est terminée. Sion soit p un tel entier.

Première étape. S'il existe $k\in\mathbb{N}$ avec $I_k\in\mathscr{F}_{m_p}$, d'après (10) et (11), on a $\mathscr{H}_k\not\subset\mathscr{F}_{m_i}$ pour $i\in K_n$. On prends pour a_{n+1} un point quelconque de $(I_k\cap A_{n+1})\setminus\bigcup_{i\in K_n}H_{a_i}$ plus grand

que les a_i et b_i déjà construits. [Cet ensemble est infini car il appartient à $\mathscr{F}_{m_p}$ puisque d'après (9), $H_{a_i} \notin \mathscr{F}_{m_p}$ pour $i \leqq p$ et que d'après (8) $H_{a_i} \notin \mathscr{F}_{m_p}$ *pour* $i > p$.] La condition (7) est alors satisfaite.

Dans le cas contraire, $\mathscr{F}_{m_p}$ est porté par le complémentaire de chaque I_k, et comme précédemment est porté par $B = A_{n+1} \setminus \bigcup_{i \in K_n} H_{a_i}$. Pour $a \in B$, $\mathscr{H}_a$ prends donc une infinité de valeurs distinctes. Par hypothèse il existe alors $a_{n+1} \in B$ tel que (7) soit satisfaite.

D'après (7), il existe $H_{a_{n+1}} \in \mathscr{H}_{a_{n+1}}$ avec $H_{a_{n+1}} \in \mathscr{F}_{m_i}$ pour $i \in K_n$, $a_i \notin H_{a_{n+1}}$ pour $i \in K_{n+1}$. *Reste à construire* m_{n+1} de sorte que (9), (10), (11) soient vérifiées. Soit $C = A_{n+1} \setminus \bigcup_{i \in K_{n+1}} H_{a_i}$. Cet ensemble porte $\mathscr{F}_{m_p}$.

Deuxième étape. S'il existe un k avec $I_k \notin \mathscr{F}_{m_p}$, il suffit alors de prendre m_{n+1} assez grand tel que $I_k \cap C \in \mathscr{F}_{m_{n+1}}$ pour que les conditions (9) et (10) soient satisfaites. La condition (11) est automatiquement vérifiée, d'après (10) et (11), puisque le seul $I_r \in \mathscr{F}_{m_{n+1}}$ est I_k, qui appartient à $\mathscr{F}_{m_p}$.

Dans le cas contraire, si E est l'ensemble des k tels qu'il existe $i \in K_{n-1}$ avec $\mathscr{H}_k \subset \mathscr{F}_{m_i}$, $\mathscr{F}_{m_p}$ est porté par $C \setminus E$. Soit r supérieur à tous les m_i, ℓ_i déjà construits, tel que $C \setminus E \in \mathscr{F}_r$, et que $\mathscr{F}_r$ ne contiennc aucun $\mathscr{H}_k$ pour $I_k \in \mathscr{F}_{m_i}$, $i \in K_n$. Si $\mathscr{F}_n$ ne contient aucun I_ℓ, on pose $r = m_{n+1}$, ce qui termine la construction. Si $\mathscr{F}_r$ contient un I_ℓ, on prend pour m_{n+1} un entier m plus grand que r, tel que $I_\ell \cap (C \setminus E) \in \mathscr{F}_m$ et que $\mathscr{F}_m$ ne contienne aucun $\mathscr{H}_k$ pour $I_k \in \mathscr{F}_{m_i}$, $i \in K_n$, et ne contienne pas $\mathscr{H}_\ell$. Ceci assure que (10) est vérifiée, (11) l'étant automatiquement.

La construction est terminée. Il est clair que K est infini, puisque $A_n \in \mathscr{F}_1 \Rightarrow n \in K$. Posons $X = \{a_i ; i \in K\}$. Pour $i \in K$, on a $H_{a_i} \in \mathscr{H}_{a_i}$ et $H_{a_i} \cap X = \emptyset$ d'après (6) et (8), donc pour $a \in X$ on a $X^c \in \mathscr{H}_a$. Pour tout i, et tout $n \geqq i$, on a $A_n \in \mathscr{F}_{m_i} \Rightarrow n \in K \Rightarrow a_n \in A_n$. Ainsi $A_n \cap X \neq \emptyset$. Ceci montre que l'on peut considérer l'ultrafiltre $\mathscr{G}_{m_i}$ sur $\mathscr{A}'$ engendré par $\mathscr{F}_{m_i}$ et X. Pour tout $n \geqq k$, si $A_n \in \mathscr{F}_k$, on a $b_n \in X^c$; on peut donc en particulier lorsque m n'est pas de la forme m_i, considérer l'ultrafiltre $\mathscr{G}_m$ sur $\mathscr{A}'$ engendré par $\mathscr{F}_m$ et X^c. Il est clair que $X \in \mathscr{G}_1$. Pour conclure, il ne reste qu'à vérifier que la suite $\mathscr{G}_n$ est dense en soi.

S'il existe m tel que $A_n \cap X \in \mathscr{G}_m$, m est de la forme m_i. Pour tout $q \geqq i$ tel que $A_q = A_n$, on a $A_q \in \mathscr{F}_{m_i}$, donc $q \in K$, donc $A_n \in \mathscr{F}_{m_q}$, donc $A_n \cap X \in \mathscr{G}_{m_q}$, ce qui montre que $A_n \cap X$ appartient à une infinité de $\mathscr{G}_m$.

S'il existe m tel que $A_n \cap X^c \in \mathscr{G}_m$, alors $A_n \in \mathscr{F}_m$. Pour tout $q \geqq m$ tel que $A_q = A_n$, on a $A_n \in \mathscr{F}_{\ell_q}$, et puisque ℓ_q n'est pas de la forme m_i, on a $A_n \cap X^c \in \mathscr{G}_q$, ce qui montre que $A_n \cap X^c$ appartient à une infinité de $\mathscr{G}_q$. Le lemme est démontré.

Preuve du théorème. Soit $(J_\alpha)_{\alpha < \Omega}$ une énumération de parties infinies de $\mathbb{N}$, où Ω désigne le premier ordinal non dénombrable.

Soient (p_n), (q_n) deux parties denses de l'ensemble $K = \{0, 1\}^{\mathbb{N}}$, où les p_i et les q_j sont deux à deux distincts. Posons $\mathscr{A} = \{I \subset \mathbb{N}$; il existe un ouvert fermé U de K avec $n \in I \Leftrightarrow p_n \in U\}$. Soit $\mathscr{A}^0$ engendrée par $\mathscr{A}$ et les parties finies de $\mathbb{N}$. Pour tout n soit $\mathscr{F}'_n$ l'ultrafiltre sur $\mathscr{A}$ avec $I \in \mathscr{F}'_n \Leftrightarrow q_n \in \overline{\{p_n ; n \in I\}}$. Il est facile de vérifier que le filtre $\mathscr{F}^0_n$ qu'il engendre sur $\mathscr{A}^0$ est encore un ultrafiltre, et que la suite $(\mathscr{F}^0_n)$ est dense en soi.

Pour $\alpha < \Omega$, on va construire par induction une suite croissante d'algèbres dénombrables $\mathscr{A}^\alpha \subset \mathfrak{P}(\mathbb{N})$, et des suites croissantes d'ultrafiltre $(\mathscr{F}^\alpha_n)_{\alpha < \Omega}$ sur $\mathscr{A}^\alpha$.

[On identifie dans la suite un filtre sur une sous-algèbre de $\mathfrak{P}(\mathbb{N})$ avec le filtre qu'il engendre dans $\mathfrak{P}(\mathbb{N})$.] Si $\mathscr{A}^\alpha$ est construite, soit $(\mathscr{H}(\alpha,\beta))_{\beta<\Omega}$ une énumération $\mathscr{H}_{(\alpha,\beta)}=(\mathscr{H}(\alpha,\beta,n))_{n\in\mathbb{N}}$ des suites $(\mathscr{H}_n)$ de filtres sur $\mathscr{A}^\alpha$ qui possèdent les propriétés suivantes:

(12) Pour tout n, $\mathscr{H}_n$ contient les complémentaires des parties finies de $\mathbb{N}$.

(13) Pour tout n, $\mathscr{H}_n \not\subset \mathscr{F}_1^\alpha$.

(14) Pour tout n, il existe au plus un p avec $\mathscr{H}_n \subset \mathscr{F}_p^\alpha$.

(15) Pour tout p, l'ensemble $\{n\in\mathbb{N}; \mathscr{H}_n=\mathscr{H}_p\}$ appartient à $\mathscr{A}^\alpha$.

Soit $\alpha\to(\alpha_1,\alpha_2)$ une surjection de Ω sur Ω^2 telle que $\alpha_1\leqq\alpha$ pour $\alpha\in\Omega$. On va effectuer la construction de sorte que les conditions suivantes soient vérifiées pour tout α:

(16) La suite $(\mathscr{F}_n^\alpha)_n$ est dense en soi.

(17) $J_\alpha\in\mathscr{A}^{\alpha+1}$.

(18) Il existe $X_\alpha\in\mathscr{A}^{\alpha+1}$ tel que $X_\alpha\in\mathscr{F}_1^\alpha$, et que pour $a\in X_\alpha$, on ait $X_\alpha^c\in\mathscr{H}(\alpha_1,\alpha_2,a)$.

Supposons la construction effectuée pour tout ordinal $\beta<\alpha$. Si α est limite, on pose $\mathscr{A}^\alpha=\bigcup_{\beta<\alpha}\mathscr{A}_\beta$, $\mathscr{F}_n^\alpha=\bigcup_{\beta<\alpha}\mathscr{F}_n^\beta$. Si $\alpha=\gamma+1$, on utilise successivement les lemmes 1 et 2, ce qui termine la construction.

La condition (17) montre que $\mathfrak{P}(\mathbb{N})=\bigcup_{\alpha<\Omega}\mathscr{A}^\alpha$. Soit $\mathscr{F}_n=\bigcup_{\alpha<\Omega}\mathscr{F}_n^\alpha$. La suite $\mathscr{F}_n$ est dense en soi. Soit S son adhérence dans $\beta\mathbb{N}$. On va montrer que S n'est pas rétract. Ceci repose sur le fait suivant:

Fait. *Soit $(\mathscr{K}_n)$ une suite de $\beta\mathbb{N}\setminus\mathbb{N}$ telle que $\mathscr{F}_1=\lim_{n\in\mathscr{F}_1}\mathscr{K}_n$. Alors $\{n;\mathscr{K}_n=\mathscr{F}_1\}\in\mathscr{F}_1$.*

Preuve. Sinon on peut supposer que pour tout n on a $\mathscr{K}_n\neq\mathscr{F}_1$. Il existe $\gamma<\Omega$ tel que pour $p\in\mathbb{N}$, $I_p=\{n;\mathscr{K}_n=\mathscr{K}_p\}\in\mathscr{A}^\gamma$ et que si $\mathscr{K}_n$ désigne la restriction de $\mathscr{K}_n$ à $\mathscr{A}^\gamma$, on ait $\mathscr{K}_n\neq\mathscr{F}_1^\gamma$ pour tout n, et que pour tout $n\notin I_p$ on ait $\mathscr{H}_n\neq\mathscr{H}_p$.

Il existe donc $\beta<\Omega$ tel que $\mathscr{H}_n=\mathscr{H}(\gamma,\beta,n)$. Soit $\alpha<\Omega$ tel que $(\alpha_1,\alpha_2)=(\gamma,\beta)$. La condition (18) montre que $X_\alpha\in\mathscr{F}_1$, mais pour $a\in X_\alpha$, on a $X_\alpha^c\in\mathscr{H}(\gamma,\beta,n)=\mathscr{H}_n$, et donc $X_\alpha^c\in\mathscr{K}_n$. Ceci contredit le fait que $\mathscr{F}_1=\lim_{n\in\mathscr{F}_1}\mathscr{K}_n$. Le fait est démontré.

Si S est un rétract de $\beta\mathbb{N}$, il existe une suite $\mathscr{K}_n$ de S telle que pour tout k on ait $\mathscr{F}_k=\lim_{n\in\mathscr{F}_k}\mathscr{K}_n$. On en déduit donc que $I_1=\{n;\mathscr{K}_n=\mathscr{F}_1\}\in\mathscr{F}_1$. Mais puisque la suite $\mathscr{F}_n$ est dense en soi, il existe p avec $I_1\in\mathscr{F}_p\neq\mathscr{F}_1$. On a donc $\mathscr{F}_1=\lim_{n\in\mathscr{F}_p}\mathscr{K}_n$, ce qui est absurde puisque $\mathscr{F}_1\neq\mathscr{F}_p$.

Le théorème est démontré.

Bibliographie

1. Maharam, D.: Finitely additive measures on the integers. Sankhya Ser. **38**, 44–59 (1976)

Reçu le 26 novembre 1980

Math. Ann. 256, 67–84 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Universal Families of C^∞-Functions on D^2

Klaus Jänich

Fakultät für Mathematik der Universität, D-8400 Regensburg, Federal Republic of Germany

Introduction

The present paper brings the series [4–7] to a certain conclusion (see Sect. 6 at the end of the paper), by constructing the "universal $\mathscr{K}$-families" announced in [6, p. 114]. The original motivation had been the following. If stable families of C^∞-functions are considered with a view of their possible use as gradient models in the sense of Thom [10], then the *bifurcation set* is of particular importance. The C^∞ singularity theory gives information about the local structure of the bifurcation set. But not every set of this local behavior can be realized as bifurcation set, and thus the question arises, of which nature the additional *global* properties of bifurcation sets might be and how they could be studied. To this end it seems desirable to find "universal" families, from which all others could be induced, just like fibre bundles can be induced from universal fibre bundles. The algebraic-topological data of the "universal bifurcation set" would then, hopefully, play a similar rôle for bifurcation sets as, say, the cohomology ring of Graßmann manifolds does for vector bundles. — Section 2 of [6] gives a scheme for the construction of such universal families, and in [4–7] the first steps are taken. While [4] and [5] are essentially concerned with questions that are "global" with respect to the basic but local with respect to the fibre, actual families of functions of some simple kind are studied in [6], which then leads to the construction of universal "contact preserving" families in [7]. These contact preserving families, although very special, are sufficiently general to serve as start of the induction process, which we are now going to carry out. — Various aspects of this process have already been treated in the exam papers [2, 3, 9] of my students. They all have helped to clarify the concepts.

1. Differentiable Stratifications of Differentiable Targets

To separate the generalities from the details of the concrete situation can make a presentation clearer – in the present paper I am doubly justified to proceed this way, since I will have to apply the generalities *twice*, upon two quite different cases,

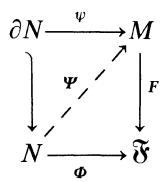
0025-5831/81/0256/0067/$03.60

and they may in fact be useful in other situations. This as an apology, should one be necessary, for the following definitions.

Definition 1. By a *differentiable target* $\mathfrak{F}$ we mean a set $\mathfrak{F}$ together with a differentiability notion for germs $(\mathbb{R}^n,0)\to\mathfrak{F}$, for all $n\in\mathbb{N}$, of which we only require that the composition of differentiable germs $(\mathbb{R}^k,0)\xrightarrow{C^\infty}(\mathbb{R}^n,0)\to\mathfrak{F}$ is again differentiable.

In the applications, $\mathfrak{F}$ will always be some set of C^∞-functions $D^2\to\mathbb{R}$. – Note that we now can speak of differentiable maps from differentiable manifolds M with or without boundary into a differentiable target $\mathfrak{F}$.

Definition 2. The differentiable maps $M\to\mathfrak{F}$ shall be called $\mathfrak{F}$-*objects* over M. Two $\mathfrak{F}$-objects over a manifold N with boundary are said to be $\mathfrak{F}$-*isotopic*, if there is an $\mathfrak{F}$-object H over $N\times[0,1]$ such that H_0 and H_1 are the two given $\mathfrak{F}$-objects and $H_t|\partial N$ does not depend on $t\in[0,1]$. An $\mathfrak{F}$-object $F:M\to\mathfrak{F}$ over a manifold without boundary shall be called $\mathfrak{F}$-*universal*, if for each manifold N with boundary, $\mathfrak{F}$-object $\Phi:N\to\mathfrak{F}$ and differentiable map $\psi:\partial N\to M$ with $F\circ\psi=\Phi|\partial N$ there is an extension of ψ to a differentiable map $\Psi:N\to M$,



such that $F\circ\dot{\Psi}$ and Φ are $\mathfrak{F}$-isotopic.

One may think of a universal $\mathfrak{F}$-object $F:M\to\mathfrak{F}$ as of something like a weak homotopy equivalence between M and $\mathfrak{F}$. – Next we shall define the notion of a differentiable decomposition of a differentiable target $\mathfrak{F}$. Although briefly written as $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$, it will in fact consist of three data: (1) a decomposition into a family $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ of pairwise disjoint subsets, (2) a vector bundle like structure $E_\alpha\to\mathfrak{F}_\alpha$ for each $\alpha\in A$, corresponding, intuitively, to the normal bundle of $\mathfrak{F}_\alpha$ in $\mathfrak{F}$, and (3) a linear map $\delta\Phi_x:T_xM\to E_{\Phi(x)}$ into the fibre of E_α at $\Phi(x)\in\mathfrak{F}_\alpha$, for each $\mathfrak{F}$-object Φ over M and $x\in M$, as if given, intuitively, by the differential of Φ at x. – To formulate (2), we introduce the following terminology.

Definition 3. By a k-dimensional *vector bundle target* over the differentiable target $\mathfrak{F}$ we mean a surjective map $E\to\mathfrak{F}$ together with a k-dimensional linear structure on each fibre and a differentiable manifold structure on Φ^*E for each $\mathfrak{F}$-object $\Phi:M\to\mathfrak{F}$, such that $\Phi^*E\to M$ becomes a differentiable vector bundle over M and such that for any differentiable $N\xrightarrow{\Psi}M\xrightarrow{\Phi}\mathfrak{F}$ the canonical map $\Psi^*\Phi^*E\to(\Phi\circ\Psi)^*E$ is a differentiable isomorphism of vector bundles over N.

Definition 4. A *differentiable decomposition* of a differentiable target $\mathfrak{F}$ consists of the following three data:

(1) a decomposition of $\mathfrak{F}$ into a family $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ of pairwise disjoint subsets,

(2) a vector bundle target $E_\alpha\to\mathfrak{F}_\alpha$ for each $\alpha\in A$, and

(3) for each $\mathfrak{F}$-object $\Phi: M \to \mathfrak{F}$ and $x \in M$ a linear map $\delta\Phi_x: T_x M \to E_{\Phi(x)}$ into the fibre of E_α at $\Phi(x) \in \mathfrak{F}_\alpha$, such that for differentiable $N \xrightarrow{\Psi} M \xrightarrow{\Phi} \mathfrak{F}$ the chain rule holds: $\delta(\Phi \circ \Psi)_x = \delta\Phi_{\Psi(x)} \circ d\Psi_x$.

These data constitute a differentiable decomposition of $\mathfrak{F}$, if they satisfy the "Submanifold Axiom" and the "Tubular Extension Axiom" given below.

An $\mathfrak{F}$-object Φ over M shall be called *transversal* to $\mathfrak{F}_\alpha$, if $\delta\Phi_x$ is surjective for all $x \in \Phi^{-1}(\mathfrak{F}_\alpha)$. By a *tubular extension* of a differentiable $\varphi: M \to \mathfrak{F}_\alpha$ we mean the germ at $M \subset \varphi^* E_\alpha$ of an $\mathfrak{F}$-object Φ over an open neighborhood of the zero section of $\varphi^* E_\alpha$ such that $\Phi | M = \varphi$ and $\delta\Phi_x$, for each $x \in M$, is the canonical map $T_x(\varphi^* E_\alpha) \to E_{\varphi(x)}$. – With this terminology, we can formulate the two axioms as follows:

Submanifold Axiom. If $\partial M = \emptyset$ and $\Phi: M \to \mathfrak{F}$ is transversal to $\mathfrak{F}_\alpha$, then $M_\alpha := \Phi^{-1}(\mathfrak{F}_\alpha)$ is a submanifold of M, and $\delta\Phi_{M_\alpha}: TM | M_\alpha \to (\Phi | M_\alpha)^* E_\alpha$ induces an isomorphism of the normal bundle μ_α of M_α in M onto $(\Phi | M_\alpha)^* E_\alpha$.

Tubular Extension Axiom. If $\varphi: M \to \mathfrak{F}_\alpha$ is differentiable, then any tubular extension of $\varphi | \partial M$ can be extended to a tubular extension of φ itself.

It should be remarked, that this notion of a "differentiable decomposition" is still very general. For instance, a differentiable manifold $X =: \mathfrak{F}$ together with just any decomposition $\{X_\alpha\}_{\alpha \in A}$ into submanifolds defines in a canonical way a differentiable decomposition in our technical sense.

Definition 5. Let $\{\mathfrak{F}_\alpha\}_{\alpha \in A}$ be a differentiable decomposition of $\mathfrak{F}$. An $\mathfrak{F}$-object Φ over M shall be called *A-transversal* or an *A-object*, if both Φ and $\Phi | \partial M$ are transversal to $\mathfrak{F}_\alpha$ for all $\alpha \in A$. Two A-objects over M are said to be *A-isotopic*, if there is an $\mathfrak{F}$-object H over $M \times [0,1]$ such that (i) H_0 and H_1 are the two given A-objects, (ii) $H_t | \partial M$ does not depend on $t \in [0,1]$, (iii) each H_t is A-transversal, and (iv) $\alpha(H_t(x))$, the index of the subset to wich $H_t(x)$ belongs, does not depend on $t \in [0,1]$.

Thus, such an "A-isotopy" H could be called a "strata-preserving transversal isotopy".

Definition 6. An A-object $F: M \to \mathfrak{F}$ over a manifold without boundary shall be called *A-universal*, if for each A-object $\Phi: N \to \mathfrak{F}$ and differentiable map $\psi: \partial N \to M$ with $F \circ \psi = \Phi | \partial N$, there is an extension $\Psi: N \to M$ of ψ such that $F \circ \Psi$ and Φ are A-isotopic A-objects.

The purpose of our "induction process" will be to construct, in certain cases, an A-universal object from given $\mathfrak{F}_\alpha$-universal objects for the individual $\alpha \in A$. We shall now introduce the additional axioms which will make such an induction possible.

Definition 7. A differentiable decomposition $\{\mathfrak{F}_\alpha\}_{\alpha \in A}$ of a differentiable target $\mathfrak{F}$ shall be called a *differentiable stratification* of $\mathfrak{F}$, if it satisfies the following additional axioms:

(1) *Finiteness*: A is finite.

(2) *Connectedness*: Each $\mathfrak{F}_\alpha$ is differentiably path connected.

(3) *Openness of Transversality*: If Φ is an $\mathfrak{F}$-object over M, $x \in M$ and $\delta\Phi_x$ surjective, then there is an open neighborhood U of x in M such that $\Phi|U$ is A-transversal.

(4) *Tubular Homogeneity*: (see below).

(5) *Tubular Local Triviality*: (see below).

To formulate the axioms (4) and (5) concisely and suggestively, we must improve our terminology. Let us first agree about the use of the words *strata preserving*. While dealing with $\{\mathfrak{F}_\alpha\}_{\alpha \in A}$ we have, at numerous occasions, to consider sets X which are canonically decomposed into families $\{X_\alpha\}_{\alpha \in A}$ of subsets. We then call X_α a "stratum", and a map $X \to Y$ is strata preserving, if it maps each X_α into the corresponding Y_α. If X is a differentiable manifold, a vector field on X is called strata preserving, if each flow line locally stays within a stratum, which in the case where the X_α are submanifolds means, that the field is tangent to the strata. – Next, let E be a differentiable vector bundle over a manifold M. By a *local C^∞-trivialization* of E at $x \in M$ we mean a diffeomorphism $h : \Omega \to U \times \Omega_x$ as follows: U is an open neighborhood of x in M, the Ω is an open neighborhood of the zero section in $E|U$ and $\Omega_x := \Omega \cap E_x$; and h is the identity on the zero section and maps fibres into fibres:

$$\begin{array}{ccc} & U & \\ \cap & & \cap \\ \Omega & \xrightarrow{h} & U \times \Omega_x \\ & \searrow \quad \swarrow & \\ & U & \end{array}$$

Finally, let E be a differentiable vector bundle over M with a Riemannian metric, let $\varepsilon : M \to \mathbb{R}$ be a positive C^∞ function and $D_\varepsilon E$ the disc bundle of radius ε. A vector field v on $D_\varepsilon E \backslash 0$ shall be called an *inwards vector field*, if it is tangent to the fibres and if its radial component at e is $-e$, which means $\langle e, v(e) \rangle = -\langle e, e \rangle$, *for all* $e \in D_\varepsilon E \backslash 0$.

Tubular Homogeneity Axiom: Let a representative of a tubular extension of an $\mathfrak{F}_\alpha$-object $\varphi : M \to \mathfrak{F}_\alpha$ and a Riemannian metric on $\varphi^* E_\alpha$ be given. Then if the positive differentiable $\varepsilon : M \to \mathbb{R}$ is sufficiently small, there is a strata preserving inwards vector field on $D_\varepsilon \varphi^* E_\alpha \backslash 0$.

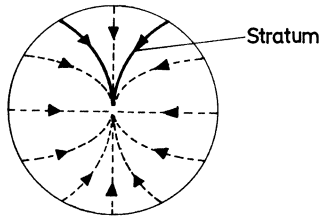


Fig. 1

Tubular Local Triviality Axiom. Let a representative of a tabular extension of an $\mathfrak{F}_\alpha$-object $\varphi : M \to \mathfrak{F}_\alpha$ be given. Then for any $x \in M$ there is a strata preserving local C^∞-trivialization of $\varphi^* E_\alpha$ at x.

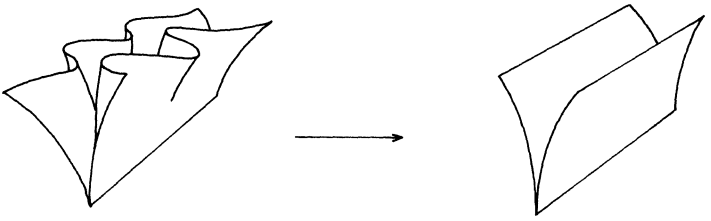


Fig. 2

2. The Inductive Construction of Universal Objects

Theorem 1 ("*Criterion*"). *Let $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ be a differentiable stratification of a differentiable target $\mathfrak{F}$. Then an A-object F over a manifold M without boundary is A-universal if and only if its restrictions $F_\alpha: M_\alpha \to \mathfrak{F}_\alpha$ to the submanifolds $M_\alpha := F^{-1}(\mathfrak{F}_\alpha)$ are $\mathfrak{F}_\alpha$-universal for all $\alpha\in A$.*

Theorem 2 ("*Construction*"). *Let $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ be a differentiable stratification of a differentiable target $\mathfrak{F}$. If for each $\alpha\in A$ there exists an $\mathfrak{F}_\alpha$-universal object, then there also exists an A-universal object. Moreover, if there are $\mathfrak{F}_\alpha$-objects over compact manifolds M_α with boundary, which are $\mathfrak{F}_\alpha$-universal over $M_\alpha \backslash \partial M_\alpha$ for all $\alpha\in A$, then there is an A-object over a compact manifold M with boundary, which is A-universal over $M\backslash\partial M$.*

Before we turn to the proofs, let us add one remark about the applicability of Theorem 2 in the case where we are interested not in an A-universal, but in an $\mathfrak{F}$-universal object:

Definition (Transversality Property). We say that the differentiable stratification $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ has the *transversality property*, if any $\mathfrak{F}$-object Φ over N with an A-transversal restriction $\Phi|\partial N$ is $\mathfrak{F}$-isotopic to an A-object over N.

Note. If $\{\mathfrak{F}_\alpha\}_{\alpha\in A}$ has the transversality property, then an A-universal object is also $\mathfrak{F}$-universal.

Proof. Let F over M be A-universal and let a differentiable map $\psi:\partial N\to M$ and an $\mathfrak{F}$-object Φ over N with $F\circ\psi = \Phi|\partial N$ be given. Applying a collar argument and the ordinary transversality theorem, we see that without loss of generality ψ may be thought to be transversal to the strata M_α. But then $F\circ\psi = \Phi|\partial N$ is A-transversal, hence by the assumed transversality property, Φ itself may be supposed to be an A-object, and the existence of the desired Ψ follows from the A-universality of F. q.e.d.

Now to the proofs of the theorems. We shall simply write α-object, α-isotopic ... instead of $\mathfrak{F}_\alpha$-object, $\mathfrak{F}_\alpha$-isotopic etc. – Let us first take care of the trivial "$\Rightarrow$" part of Theorem 1. Given an A-universal F over M, a differentiable map $\psi:\partial N\to M_\alpha$ and an α-object Φ over N with $F_\alpha\circ\psi = \Phi|\partial N$, we have to find an extension $\Psi: N\to M_\alpha$ of ψ such that $F_\alpha\circ\Psi$ and Φ are α-isotopic. – To denote map germs at the zero section of maps from open neighborhoods of the zero section of a vector bundle E over X into some Y, we shall write $(E,X)\to Y$. In particular,

consider the vector bundle Φ^*E_α over N. We shall first extend $\psi:\partial N\to M_\alpha$ to a germ $\tilde{\psi}:(\Phi^*E_\alpha|\partial N,\partial N)\to M$ as follows: Let μ_α denote the normal bundle of M_α in M, and choose a germ $\tau_\alpha:(\mu_\alpha,M_\alpha)\to M$ of a tubular map. Since δF induces a bundle isomorphism $\mu_\alpha\cong F_\alpha^*E_\alpha$ and since $\Phi^*E_\alpha|\partial N$ is canonically the same as $\psi^*F_\alpha^*E_\alpha$, we have a canonical bundle map from $\Phi^*E_\alpha|\partial N$ to μ_α over ψ, and we define $\tilde{\psi}$ to be its composition with τ_α. But then $F\circ\tilde{\psi}$ is in fact a tubular extension of $F\circ\psi=\Phi|\partial N$ and hence can be extended to a tubular extension $\tilde{\Phi}$ of Φ. It follows from the openness of transversality that a suitable representative of $\tilde{\Phi}$ will be A-transversal, and using the A-universality of F, we find an extension $\tilde{\Psi}:(\Phi^*E_\alpha,N)\to M$ of $\tilde{\psi}$ such that its restriction $\Psi:=\tilde{\Psi}|N$ has the desired properties. q.e.d.

For the other parts of the proofs we need the *subordination* relation among the $\alpha\in A$: If $f\in\mathfrak{F}_\alpha$ and $\tau:(E_f,0)\to\mathfrak{F}$ a tubular extension of $\{f\}\subset\mathfrak{F}_\alpha$, then we say that $\alpha'\in A$ is subordinate to α (written $\alpha'<\alpha$) with respect to τ, if each representative of $\tau|E_f\backslash 0$ hits $\mathfrak{F}_{\alpha'}$. It follows from the tubular extension and local triviality axioms and from the connectedness of the $\mathfrak{F}_\alpha$, that in this way a relation $<$ ("subordinate") is well defined in $A\times A$. By the submanifold axiom, α cannot be subordinate to itself. Moreover, if $\alpha'<\alpha$ and an $\mathfrak{F}$-object Φ is transversal to $\mathfrak{F}_\alpha$ at some x with $\Phi(x)\in\mathfrak{F}_\alpha$, then each representative of its germ at x hits $\mathfrak{F}_{\alpha'}$, and thus it follows from the openness of transversality that the relation is transitive and hence defines a partial ordering on A.

The point of this ordering is the following: If $\beta\in A$ is maximal with respect to subordination, then for any A-object Φ over M the submanifold $\Phi^{-1}(\mathfrak{F}_\beta)$ is *closed*. With this fact in mind, one easily checks that $\{\mathfrak{F}_\alpha\}_{\alpha\in A\backslash\beta}$ is canonically a differentiable stratification of $\mathfrak{F}\backslash\mathfrak{F}_\beta$, and since it has one stratum less, we can apply induction hypotheses upon it.

Proof of Theorem 1. Let β be maximal in A and $A':=A\backslash\beta$. Let F be an A-object over M such that F_α is α-universal for each $\alpha\in A$. We have to show that F is A-universal. Put $M':=M\backslash M_\beta$ and $F':=F|M'$. By induction hypothesis, we may assume that F' is A'-universal. – Now let a differentiable map $\psi:\partial N\to M$ and an A-object Φ over N with $F\circ\psi=\Phi|\partial N$ be given. Since F_β is β-universal, we can find an extension $\Psi_\beta:N_\beta\to M_\beta$ of $\psi_\beta:=\psi|\partial N_\beta$ and a β-isotopy h between $\Phi_\beta:=\Phi|N_\beta$ and $F_\beta\circ\Psi_\beta$. This h is a β-object over the manifold $B:=N_\beta\times[0,1]$ with boundary $\partial B=N_\beta\times 0\cup\partial N_\beta\times[0,1]\cup N_\beta\times 1$. The first step in our proof will be the choice of a certain tubular extension of $h|\partial B$.

We introduce the following terminology. If μ is the normal bundle of a submanifold $Y_0\subset Y$ and $f:X\to Y_0$ a differentiable map, then by a *tubular germ for* $Y_0\subset Y$ *along* f we mean an extension of f to a differentiable germ $(f^*\mu,X)\to Y$, such that its differential induces the canonical map $f^*\mu\to\mu$ over f. In the case where f is the identity on Y_0 we simply speak of a tubular germ for $Y_0\subset Y$. – Returning to our present situation, we proceed as follows. (i) We first choose a tubular germ for ∂N_β in ∂N. (ii) We extend it to a tubular germ σ for N_β in N, thus now writing $\partial\sigma$ for the germ under (i). (iii) ψ and $\partial\sigma$ together determine a tubular germ for $M_\beta\subset M$ along ψ_β, which we extend to a tubular germ τ along Ψ_β. – These data (i)–(iii) fix a tubular extension of $h|\partial B$ as follows. Over $N_\beta\times 0$, where h is given by Φ_β, we have the isomorphism between the normal bundle ν_β of N_β in N and $\Phi_\beta^*E_\beta$, which is

induced by $\delta\Phi$, and we define the tubular extension by the composition

$$(\Phi_\beta^* E_\beta, N_\beta) \cong (\nu_\beta, N_\beta) \xrightarrow{\sigma} N \xrightarrow{\Phi} \mathscr{F}.$$

At the other end, h is given as $h_1 = F_\beta \circ \Psi_\beta$. Here δF, restricted to M_β, defines a bundle isomorphism between the normal bundle μ_β of M_β in M and $F_\beta^* E_\beta$, hence also between $\Psi_\beta^* \mu_\beta$ and $\Psi_\beta^* F_\beta^* E_\beta = h_1^* E_\beta$, and we define the tubular extension of h_1 as the composition $(\Psi_\beta^* F_\beta^* E_\beta, N_\beta) \cong (\Psi_\beta^* \mu_\beta, N_\beta) \xrightarrow{\tau} M \xrightarrow{F} \mathfrak{F}$. – These tubular extensions coincide over the boundary ∂N_β, and we complete them in the obvious way to a tubular extension of $h|\partial B$ which over $\partial N_\beta \times [0,1]$ does not depend on $t \in [0,1]$.

That was the first step. In a second step, we will now find an open neighborhood U of N_β in N and an extension of $\psi|\partial U$ to a map $\Psi_U : U \to M$ such that $\Phi|U$ and $F \circ \Psi_U$ are A-isotopic. To this purpose we extend our tubular extension of $h|\partial B$ to a tubular extension $(h^* E_\beta, N_\beta \times [0,1]) \to \mathfrak{F}$ of h and choose a vector field germ v at the zero section on $h^* E_\beta$ with the following properties: (1) v is $\partial/\partial t$ on the zero section itself and over $\partial N_\beta \times [0,1]$. (2) v is a germ "over" $\partial/\partial t$ with respect to the bundle projection $h^* E_\beta \to N_\beta \times [0,1]$ and (3) v is strata preserving. – The possibility to find such a vector field germ follows from the tubular local triviality axiom. – Now we choose representatives of our various tubular germs, tubular extensions and of the vector field germ. Then for a sufficiently small open neighborhood Ω of the zero section in $h_0^* E_\beta = \Phi_\beta^* E_\beta$, the field v will be integrable on the time interval $[0,1]$ for initial values in Ω; the flow $\Omega \times [0,1] \to h^* E_\beta$, composed with the representative of the tubular extension of h, will define an $\mathfrak{F}$-object over $\Omega \times [0,1]$; the composition $\Omega \to h_1^* E_\beta = \Psi_\beta^* F_\beta^* E_\beta \cong \Psi_\beta^* \mu_\beta \xrightarrow{\tau} M$ (where the first map is the flow map at time $t=1$) will define a map $\Omega \to M$, and the composition $\Omega \subset \Phi_\beta^* E_\beta \cong \nu_\beta \xrightarrow{\sigma} N$ will map Ω diffeomorphically onto an open neighborhood U of N_β in N. With respect to this diffeomorphism $\Omega \cong U$, the above $\mathfrak{F}$-object over $\Omega \times [0,1]$ becomes an $\mathfrak{F}$-object H_U over $U \times [0,1]$, the above map $\Omega \to$ M becomes a map $\Psi_U : U \to M$, and if only Ω had been small enough, Ψ_U will extend $\psi|\partial U$ and H_U will be an A-isotopy between $\Phi|U$ and $F \circ \Psi_U$, as one can check, if intuition is not trusted, in a tedious but straightforward manner.

In a third and last step we apply the induction hypothesis (A'-universality of F'). If we slow down H_U radially on a suitable tubular neighborhood of N_β in U, we can obtain an A-isotopy H (constant in t outside the tube) between the original Φ and an A-object over N which not only is given on ∂N by $F \circ \psi$, but also on some open neighborhood U' of N_β by $F \circ \psi_U | U'$. Thus we may, without loss of generality, assume that Φ itself has this property. Then for a suitable tubular neighborhood T of N_β in U', we can apply the induction hypothesis to $\Phi | N \setminus \mathring{T}$ (here we use the tubular homogeneity axiom) to extend $\psi \cup \Psi_U | T$ by a suitable $\Psi' : N \setminus \mathring{T} \to M'$ to a map $\Psi : N \to M$, for which then Φ and $F \circ \Psi$ are A-isotopic as desired. q.e.d.

Proof of Theorem 2. By assumption and induction hypothesis, there exists a universal β-object $F_\beta : M_\beta \to \mathfrak{F}_\beta$ and a universal A'-object $F' : M' \to \mathfrak{F}'$. The idea for the construction of a universal A-object is roughly the following: Take a representative of a tubular extension of F_β on $E := F_\beta^* E_\beta$. For a suitable metric,

this will define an A-object Φ over the disc bundle DE. The restriction of Φ to the sphere bundle SE will be an A'-object, hence up to isotopy given as $F' \circ \Psi$ by a map $\Psi : SE \to M'$, which we shall use to "attach" Φ over DE to F' over M' somehow; and if everything is suitably chosen, we should obtain an A-object which is A-universal by Theorem 1. – To realize this idea, we will apply some technical facts and operations, which I will describe beforehand, in abstracto and modulo certain differential-topological routine manipulations involving collars and corners, which we will take for granted. – In view of the "compact case", we now slightly extend our use of the word "universal" and call an A-object over a manifold M *with* boundary universal, if it is universal over $M \backslash \partial M$.

(1) A-Homotopy Equivalence. Let X and Y be manifolds without boundary and F an A-object over X. A differentiable map $\varrho : Y \to X$ shall be called an *A-homotopy equivalence with regard to F*, if it is transversal to the strata and if a differentiable $\sigma : X \to Y$ exists, such that there are differentiable homotopies H and K between $\varrho \circ \sigma$ and Id_X resp. between $\sigma \circ \varrho$ and Id_Y, such that the H_t and K_t are transversal and strata preserving with regard to F and $F \circ \varrho$ respectively.

Note. If F is A-universal and ϱ an A-homotopy equivalence, then $F \circ \varrho$ is also A-universal.

(2) Thickening by Vector Bundles. Let F be an A-object over a manifold M with boundary and V a differentiable vector bundle over M. Let $\pi : DV \to M$ be the disc bundle with respect to a Riemannian metric on V, and SV the sphere bundle. Then the A-object $F \circ \pi$ over the manifold DV with boundary $DV|\partial M \cup SV$ (corners properly taken care of) shall be called a *thickening of F by V.* – The projection π (restricted to the interior of DV) is an A-homotopy equivalence with regard to $F|M \backslash \partial M$, namely a strong "$A$-deformation retraction" in the obvious sense, and thus a thickening of an A-universal object is again A-universal.

(3) A-Attaching. Let F and G be A-objects over n-dimensional manifolds X and Y with boundary, let $X_0 \subset \partial X$ and $Y_0 \subset \partial Y$ be $(n-1)$-dimensional closed submanifolds with boundary such that ∂X_0 and ∂Y_0 are transversal to the strata, and let finally $\Psi : X_0 \to Y_0$ be a diffeomorphism with $G \circ \Psi = F|X_0$. Then we obtain (after some preparation using strata preserving tubes and collars) an A-object $G \cup_\Psi F$ over the manifold $Y \cup_\Psi X$ with boundary, and we say that $G \cup_\Psi F$ is the result of A-attaching F at G with Ψ.

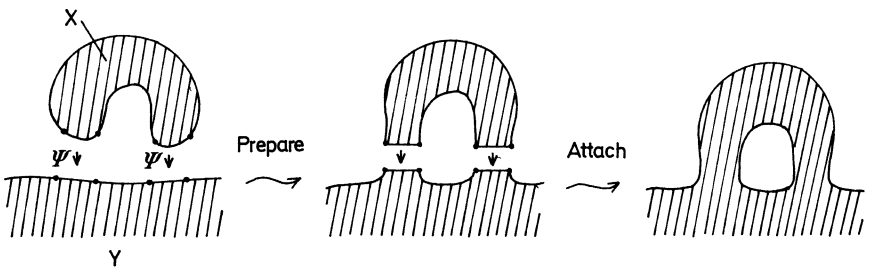


Fig. 3

(4) Bulges. We call attention to a particular case of A-attaching by giving it a special name: Let X be $Y_0 \times [0,1)$ or $Y_0 \times [0,1]$ and let F have the following properties: Each F_t is an A-object, $F_0 = G|Y_0$, and the strata type of $F_t(x)$ does not depend on t (– as if F were an A-isotopy, only that here we have no reason to require that F be constant in t over ∂Y_0). Let the attaching map be the identity $\mathrm{Id}_0 : Y_0 \times 0 \to Y_0$. Then we call $G \cup_{\mathrm{Id}_0} F$ an (open or closed) *bulge* of G. – It follows from the Note in (1), that if G is A-universal, then any bulge of G is also A-universal.

From these general considerations, we now turn to the proof of Theorem 2. We begin by choosing provisionally the following objects and maps:

(a) A β-universal F_β over a manifold M_β with boundary; compact if possible.

(b) A representative of a tubular extension of F_β and a Riemannian metric on $E := F_\beta^* E_\beta$, such that this representative defines an A-object Φ over DE for which a strata preserving inwards vector field on $DE \backslash 0$ exists.

(c) An A'-universal object F' over a manifold M' with boundary; compact if we are in the compact case. (Induction hypothesis.)

(d) A differentiable map $\Psi : SE \to M' \backslash \partial M'$, with $\Psi|\partial SE$ and $\Psi|S\mathring{E} \backslash \partial SE$ both transversal to the strata, such that $F' \circ \Psi$ and $\Phi|SE$ are "weakly A'-isotopic", which means that there is an $\mathfrak{F}'$-object H over $SE \times [0,1]$ which connects $F' \circ \Psi$ and $\Phi|SE$, such that each H_t is A'-transversal and the strata type of $H_t(x)$ does not depend on t.

The existence of such a map Ψ is what we can conclude from the A'-universality of F' (using it first for $\Phi|\partial SE$). – These data are provisional, and we shall improve them. The additional properties which we want them to have are stated in the following.

Claim. To prove Theorem 2, it would suffice to find data (a)–(d) with the additional property that M' and SE are of the same dimension and Ψ is an embedding and $\Psi(SE)$ closed in M'.

Proof of this Claim. First thicken F' by the trivial bundle $M' \times \mathbb{R}$ to an F'' over M'', which then is still A'-universal. $\partial M''$ is the "double" $M' \cup_{\mathrm{Id}_{\partial M'}} M'$ of M', and hence we can use Ψ to define a map $\Psi_0 : SE \to \partial M''$ into one of the two copies of $M' \backslash \partial M'$ in $\partial M''$. Now we attach the A'-object H over $SE \times [0,1]$ [see (d)] to the disjoint union of F'' and Φ by $\Psi_0 : SE \times 0 \to \partial M''$ and by the inclusion $SE \times 1 \to SE \subset DE$.

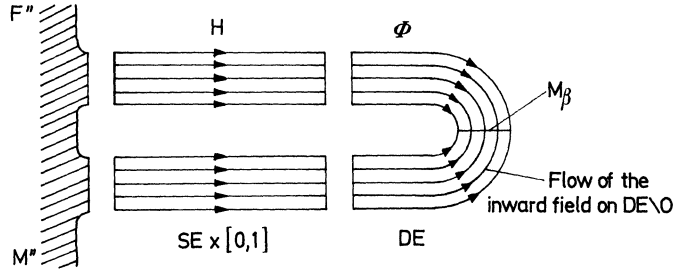


Fig. 4

Then we obtain an A-object F over a manifold M, compact in the compact case. Its restriction to the β-stratum is just the given F_β over $M_\beta \subset DE \subset M$ and hence β-universal. The restriction of F to $M \backslash M_\beta$ however is an (open) bulge of a (closed) bulge of F'', hence A'-universal, and the A-universality of F follows from Theorem 1. – This proves our claim, and what remains to do is the

Improvement of the Data (a)–(d). We first improve the data so that Ψ becomes a closed embedding. In the noncompact case, we simply choose a closed embedding $\chi: SE \to \mathbb{R}^N$ into some euklidian space, leave F_β as it is but replace F' by $F' \circ \pi$ over $M' \times \mathbb{R}^N$ and Ψ by $\Psi \times \chi$. In the compact case, we may assume $\chi(SE) \subset \mathring{D}^N$ and proceed analogously, replacing F' by its thickening by $M' \times \mathbb{R}^N$. – This was a first step, and we will now assume, that (a)–(d) are given with Ψ a closed embedding in M'. Note that this property is not lost if we thicken F' further by a bundle ξ, but let F_β and $\Psi: SE \to M' \subset D\xi$ unchanged.

Let μ denote the vector bundle over SE which is induced by Ψ from the normal bundle of $\Psi(SE) \subset M'$; the normal bundle of SE in M', so to speak. With the help of a suitable strata preserving tubular map for $\Psi(SE) \subset M'$ we can see that a thickening of $\Phi | SE$ by μ is, up to a diffeomorphism, "weakly" A'-isotopic to the restriction of F' to a full dimensional bounded transversal submanifold of $M' \backslash \partial M'$, closed in M'. On the other hand, if μ is of the form $\mu = \pi^* \eta$ for some vector bundle η over M_β, and if we denote the thickening of F_β by η as $\tilde{F}_\beta$ and define $\tilde{\Phi}$ on $D\tilde{E}$ correspondingly, then $\tilde{\Phi} | S\tilde{E}$ becomes the thickening of $\Phi | SE$ by μ, and hence our whole problem of improving the data is now reduced to the question: Can we improve F' by further thickening to the effect that μ becomes $\pi^* \eta$ for some η over M_β?

But this can be achieved by thickening F' by its sufficiently high dimensional stable normal bundle. Because then μ will be a high dimensional stable normal bundle of SE and thus be of the form $\pi^* \eta$ with $\eta = \bar{E} \oplus \nu$, where $\bar{E}$ is complementary to E (i.e. $E \oplus \bar{E}$ trivial) and ν is a stable normal bundle of M_β. This completes the proof of Theorem 2.

3. C^∞ Families of Functions on D^2

Notations and Terminology. $\mathfrak{m}(2)$ is the maximal ideal in the ring of C^∞ germs $(\mathbb{R}^2, 0) \to \mathbb{R}$. The Jacobi ideal of a singularity $\eta \in \mathfrak{m}(2)^2$ shall be denoted by $\mathfrak{J}(\eta) := \left\langle \frac{\partial \eta}{\partial x}, \frac{\partial \eta}{\partial y} \right\rangle$; by the *type* of η we mean its class under orientation preserving right equivalence; its *codimension* is the dimension of $\mathfrak{m}(2)/\mathfrak{J}(\eta)$ as a real vector space. – We shall consider C^∞ maps $F: D^2 \times M \to \mathbb{R}$, where M is a differentiable manifold with boundary, and we denote the variable in D^2 usually by $z = x + iy$ and in M by u. Such a map F will be called a *normal family over* M, if there is an integer $m \leqq 1$ (its "Hopf index") and a neighborhood of $S^1 \times M$ in $D^2 \times M$ on which F is given by $(z, u) \mapsto \mathrm{Re}\, z^{1-m}$ (resp. $|z|^2$ if $m = 1$); and in the same sense we speak of individual *normal functions* $f: D^2 \to \mathbb{R}$; see [6, p. 115]. The "partial germ" of F at $(z, u) \in \mathring{D}^2 \times M$ with respect to the z-coordinate, translated to $0 \mapsto 0$, shall be denoted by $g_1 F(z, u) \in \mathfrak{m}(2)$, it is represented by $v \mapsto F(z+v, u) - F(z, u)$. We say that the family F is *singular at* (z, u) if $g_1 F(z, u) \in \mathfrak{m}(2)^2$, i.e. if the function $F_u := F | D^2 \times u$ is singular at z. The singularity at (z, u) is *versally unfolded by* F, if it has finite

codimension and the differential

$$\delta_{(z,u)}F : T_uM \to \mathcal{m}(2)/\mathcal{J}(g_1F(z,u))$$

of the composed map $z \times M \subset \mathring{D}^2 \times M \xrightarrow{g_1F} \mathcal{m}(2) \xrightarrow{\text{proj}} \mathcal{m}(2)/\mathcal{J}(g_1F(z,u))$ at $u \in M$ is surjective. Its kernel $\ker\delta_{(z,u)}F \subset T_uM$ is then called the *equisingularity subspace* of the singularity.

Definition. A normal family F over a manifold M without boundary is called *b-stable*, if all its singularities are versally unfolded and if for each $u \in M$ the equisingularity subspaces of the singularities of F_u intersect transversally in T_uM, i.e. the codimension of their intersection is the sum of individual codimensions. A normal family over a manifold M *with* boundary is b-stable, if its restrictions to $M \backslash \partial M$ and ∂M are b-stable.

Our goal is to construct "universal" b-stable families, but not in this generality. Further restrictions will be made about types and numbers of the singularities in a family. The principal restriction will be that all singularities must be *simple* in the sense of Arnol'd [1]. As in [7], a finite formal linear combination, with positive integer coefficients, of simple singularity types shall be called a *constellation* C. If $m(\tau)$ denotes the Hopf index of a singularity type τ, we speak of $m(C) := \sum n_i m(\tau_i)$ as of the Hopf index of the constellation $C = \sum n_i \tau_i$. Each normal function $f : D^2 \to \mathbb{R}$ with only simple singularities "has" a well defined constellation $C(f)$, and if F is a b-stable family with only simple singularities, then we denote by $\mathcal{K}(F)$ the set $\{C(F_u) | u \in M\}$ of its constellations. On the set of *all* constellations, we have a natural partial ordering by *subordination*: a constellation C' is subordinate to $C \neq C'$, written $C' < C$, if it occurs in any b-stable family in which C occurs. If a set of constellations with each element contains all subordinate ones, we call it complete. The constellation sets of b-stable families with only simple singularities are complete.

Definition. Let $\mathcal{K}$ be a finite, complete set of constellations of a common Hopf index $m \leqq 1$. A b-stable family F shall be called a *$\mathcal{K}$-family*, if $\mathcal{K}(F) \subset \mathcal{K}$. Two $\mathcal{K}$-families over M are *$\mathcal{K}$-isotopic*, if there is a normal family H over $M \times [0,1]$ such that (i) H_0 and H_1 are the two given families, (ii) $H_t | D^2 \times \partial M$ does not depend on $t \in [0,1]$, (iii) each H_t is a $\mathcal{K}$-family, and (iv) the constellation $C(H_{(u,t)})$ does not depend on $t \in [0,1]$. – A $\mathcal{K}$-family F over a manifold M without boundary is called *$\mathcal{K}$-universal*, if for each $\mathcal{K}$-family Φ over a manifold N with boundary and differentiable map $\psi : \partial N \to M$ with $F \circ (Id_{D^2} \times \psi) = \Phi | D^2 \times \partial N$, there is an extension $\Psi : N \to M$ of ψ such that $F \circ (\mathrm{Id}_{D^2} \times \Psi)$ and Φ are $\mathcal{K}$-isotopic $\mathcal{K}$-families.

The purpose of the present paper is the construction of universal $\mathcal{K}$-families. We now begin to prepare the problem for the application of Theorem 2. First we note, that the set $\mathfrak{F}$ of all normal functions $D^2 \to \mathbb{R}$ has a natural differentiable target structure: a germ $(\mathbb{R}^n, 0) \to \mathfrak{F}$ is differentiable, if it can be represented by a normal family over an open neighborhood of 0 in $\mathbb{R}^n$. For a given constellation C or constellation set $\mathcal{K}$, we have the "subtargets" $\mathfrak{F}_C$ and $\mathfrak{F}_{\mathcal{K}}$ of those normal functions whose constellation is C or belongs to $\mathcal{K}$, respectively. $\mathfrak{F}_C$ need not be connected, and in view of the connectedness axiom for differentiably stratified targets, we introduce the notion of a "subconstellation":

Definition. Two normal functions of constellation C are of the same *sub- or fine constellation*, if they can be connected by a differentiable path in $\mathfrak{F}_C$. For a subconstellation c, we write $\mathfrak{F}_c$ for the set of all normal functions of that subconstellation.

Of course, c "is" nothing but $\mathfrak{F}_c$, but we find it notationally convenient to distinguish between the subconstellation as a property or feature of a function on one hand and the set of functions sharing this feature on the other. – Each $\mathfrak{F}_c$ is subdivided further into the subsets of common *level situation*. As in [7], two elements of $\mathfrak{F}_c$ are of the same level situation L, if there is a "contact preserving" path in $\mathfrak{F}_c$ between them [7, Definition 2].

Let us write $\mathfrak{S}(C)$ and $\mathfrak{S}(\mathscr{K})$ for the sets of subconstellations of functions in $\mathfrak{F}_C$ and $\mathfrak{F}_{\mathscr{K}}$ respectively, and similarly $\Lambda(c)$ and $\Lambda(C)$ for the sets of level situations of functions in $\mathfrak{F}_c$ and $\mathfrak{F}_C$. In [8], kindly written for our purpose, Maier proves the following

Lemma. *If C is a constellation of Hopf index $\leqq 1$, then $\Lambda(C)$ is finite.*

Consequently, $\mathfrak{S}(C)$ and $\Lambda(c)$ for each $c \in \mathfrak{S}(C)$ are finite, and if $\mathscr{K}$ is a finite complete set of constellations of Hopf index $m \leqq 1$, then $\mathfrak{S}(\mathscr{K})$ is finite.

The rest of the paper is now organized as follows. First we will show, that $\{\mathfrak{F}_c\}_{c \in \mathfrak{S}(\mathscr{K})}$, in a natural way, is a differentiable stratification of $\mathfrak{F}_{\mathscr{K}}$ such that the $\mathfrak{S}(\mathscr{K})$-objects are just the $\mathscr{K}$-families. Hence, by a first application of Theorem 2, the original problem of finding universal $\mathscr{K}$-families is reduced to the problem of constructing universal c-objects. Next we show, that $\{\mathfrak{F}_L\}_{L \in \Lambda(c)}$ in a natural way is a differentiable stratification of $\mathfrak{F}_c$ and has the "Transversality Property". Thus according to the Note about the applicability of Theorem 2, the original problem is further reduced by Theorem 2 to the construction of universal L-objects – but such, as we shall see, have already been found in [7]. It can be shown, by the way, that $\{\mathfrak{F}_c\}_{c \in \mathfrak{S}(\mathscr{K})}$ also has the Transversality Property. We are not going to use this fact in the present paper, but it implies that the universal $\mathscr{K}$-families are also $\mathfrak{F}_{\mathscr{K}}$-universal, which may have interesting consequences.

4. The Differentiable Stratification $\{\mathfrak{F}_c\}_{c \in \mathfrak{S}(K)}$

So let $\mathscr{K}$ be a finite complete set of constellations of Hopf index $m \leqq 1$ and let f be a normal function with $C(f) \in \mathscr{K}$. Let $gf(z) \in \mathscr{m}(2)$ denote the germ of f at z, translated to $0 \mapsto 0$. Then we define a $\operatorname{codim} C(f)$-dimensional real vector space E_f by

$$E_f := \bigoplus_{\substack{f \text{ singular} \\ \text{at } z}} \mathscr{m}(2)/\mathscr{J}(gf(z)).$$

With these vector spaces E_f (or, strictly speaking: with $\{f\} \times E_f$) as fibres, we have a surjective map $E_c \to \mathfrak{F}_c$ for each $c \in \mathfrak{S}(\mathscr{K})$, which carries a canonical vector bundle target structure. More precisely, if φ is a c-object over M, then its singular set $S_\varphi := \{(z,u) \in D^2 \times M | \varphi_u \text{ singular at } z\}$ is the disjoint union of subsets $S_1, \ldots, S_n$, corresponding to the different singularity types $\tau_1, \ldots, \tau_n$ occurring in c, and each projection $S_i \to M$ is a finite covering. For sufficiently high r, only depending on c, the Jacobi ideals define *subbundles* $\mathscr{J}_i$ of $S_i \times J_0^r(2)$ with fibres $\mathscr{J}(g_1 \varphi(z,u))/\mathscr{m}(2)^{r+1}$

for $(z,u)\in S_i$, where $J^r_0(2)$ denotes the jet space $\mathscr{m}(2)/\mathscr{m}(2)^{r+1}$; namely because these $\mathscr{J}_i$ are the images of differentiable bundle homomorphisms $S_i\times J^r_0(2)\to S_i\times J^r_0(2)$ of constant corank = codim τ_i. Thus we have quotient vector bundles $S_i\times J^r_0(2)/\mathscr{J}_i$ over S_i, and for sufficiently high r the direct sum $\bigoplus_{i=1}^{n}$ of the direct images of these bundles under $S_i\to M$ will be canonically the same as φ^*E_c. In this way, a vector bundle target structure for $E_c\to\mathfrak{F}_c$ is well defined.

Next, if Φ is an $\mathfrak{F}_{\mathscr{K}}$-object over M, we define $\delta\Phi_u: T_uM\to E_{\Phi(u)}$ as given by $\delta_{(z,u)}\Phi: T_uM\to\mathscr{m}(2)/\mathscr{J}(g_1\Phi(z,u))$ (see p. 77) for each individual summand. Clearly δ, as an actual differential, satisfies the chain rule in the sense of Definition 4, Sect. 1.

Having thus introduced the data, we have now to check the axioms of Definitions 4 and 7 for $\{\mathfrak{F}_c\}_{c\in\mathfrak{S}(\mathscr{K})}$. Finiteness and Connectedness are already known. All other axioms are, in one way or other, concerned with the transversality notion, and hence we shall first describe what the surjectivity of $\delta\Phi_u$ implies in terms of the local unfolding theory. So let Φ be an $\mathfrak{F}_{\mathscr{K}}$-object over a manifold M without boundary, $u\in M$ and $\delta\Phi_u: T_uM\to E_{\Phi(u)}$ surjective, with a kernel of dimension q. Let $z_1,\ldots,z_p\in D^2$ be the points in which Φ_u is singular, let $\eta_i:=g_1\Phi(z_i,u)\in\mathscr{m}(2)$, $i=1,\ldots,p$ be its singularities there, and let $\Phi_1,\ldots,\Phi_p$ denote the unfoldings of these $\eta_1,\ldots,\eta_p$ by Φ, understood as germs at $(0,u)$ on $\mathbb{R}^2\times M$, represented by $(z,v)\to\Phi(z_i+z,v)-\Phi(z_i,u)$. Note that all information about transversality and strata of Φ close to u is contained in these unfoldings. Now we are going to compare $(\Phi_1,\ldots,\Phi_p)$ with a particularly nice set of other unfoldings $(F_1,\ldots,F_p)$ as follows. For each i, choose a miniversal unfolding f_i of η_i over $\mathscr{m}(2)/\mathscr{J}(\eta_i)$, given by a linear section $\mathscr{m}(2)/\mathscr{J}(\eta_i)\to\mathscr{m}(2)$, let π_i denote the projection from $\mathbb{R}^q\times E_{\Phi(u)}$ onto the direct summand $\mathscr{m}(2)/\mathscr{J}(\eta_i)$ and lift f_i to an unfolding $F_i:=\pi_i^*f_i$ over $\mathbb{R}^q\times E_{\Phi(u)}$. To express the relation between $(\Phi_1,\ldots,\Phi_p)$ and $(F_1,\ldots,F_p)$, let us call two unfoldings A and B over M at u *equivalent over* Id, written $\overset{\sim}{\mathrm{Id}}$, if there is a local diffeomorphism $H:(\mathbb{R}^2\times M,(0,u))\to(\mathbb{R}^2\times M,(0,u))$ over M, namely satisfying $\pi\circ H=\pi$, where $\pi:\mathbb{R}^2\times M\to M$ is the projection, and a C^∞ germ $\alpha:(M,u)\to\mathbb{R}$, such that $A=B\circ H+\pi\circ\alpha$. Then from the local unfolding theory, one derives the following

Fact. *There is a local diffeomorphism $h:(M,u)\to(\mathbb{R}^q\times E_{\Phi(u)},0)$ such that $\Phi_i\overset{\sim}{\mathrm{Id}}h^*F_i$ for $i=1,\ldots,p$ simultaneously; and for each such local diffeomorphism, the composition $T_uM\to\mathbb{R}^q\times E_{\Phi(u)}\to E_{\Phi(u)}$ of the differential dh_u with the projection onto $E_{\Phi(u)}$ is the given $\delta\Phi_u$.*

All those stratification axioms, that are concerned with given $\mathfrak{F}_{\mathscr{K}}$-objects and surjective $\delta\Phi_u$ and which state properties only depending on the $\overset{\sim}{\mathrm{Id}}$-classes of $\Phi_1,\ldots,\Phi_p$ can now be checked straightforwardly as consequences of corresponding well known properties of versal unfoldings of simple singularities. And in fact, almost all of the axioms do fall under this heading, namely the Submanifold Axiom, Openness of Transversality, Tubular Homogeneity, and Tubular Local Triviality. It only remains to take care of the Tubular Extension Axiom.

So let φ be a c-object over M, and let a tubular extension ψ of $\varphi|\partial M$ be given. We have to show that it can be extended to a tubular extension of all of φ (see Definition 5). Let us begin with the following observation. If Ψ is just *any*

extension of φ and ψ to a germ at the zero section of a normal family over $E:=\varphi^*E_c$, then it will automatically be the germ of an $\mathfrak{F}_{\mathscr{K}}$-object over E, and thus we can speak of the homomorphisms $E_u \to E_u$ induced by $\delta\Psi_u$ for $u\in M$, let us denote them by $\delta'\Psi_u$. If for each $u\in M$ this $\delta'\Psi_u$ *is* the identity on E_u, then by the openness of transversality, Ψ is the germ of an $\mathfrak{S}(\mathscr{K})$-object and a tubular extension of φ as desired. – This implies, first of all, that the use of partitions of unity on M reduces the problem to a local one with respect to M. To solve this local problem, first consider a boundary point $u_0\in\partial M$. Choose an open neighborhood U of u_0 in M and any extension of $\varphi|U$ and $\psi|(E|\partial U)$ to a germ Ψ at the zero section of a normal family over $E|U$. After making U smaller, if necessary, we may assume that $\delta'\Psi_u: E_u\to E_u$, $u\in U$, defines a bundle isomorphism $f: E|U\to E|U$, which is the identity over ∂U, and then $\Psi\circ f^{-1}$ is an extension of $\psi|(E|\partial U)$ to a tubular extension of $\varphi|U$. – Finally, let u_0 be an interior point of M. We first choose a tubular extension Ψ_0 of the c-object φ_{u_0} over the single point $\{u_0\}$, using unfoldings $F_1,\ldots,F_p$ of the singularities $\eta_1,\ldots,\eta_p$ as above [namely first by linear sections $m(2)/\mathscr{J}(\eta_i)\to m(2)$ over the summands of E_{u_0}, then lifted to E_{u_0} by the projections] and using a partition of unity on $D^2\times E_{u_0}$. Then we extend Ψ_0 and $\varphi|U$, for some open neighborhood U of u_0 in M, to a germ Ψ at U of a normal family over $E|U$ and argue as we just did in the boundary case. – This was the last step in checking the stratification axioms. In addition, we note that obviously for an $\mathfrak{F}_{\mathscr{K}}$-object, $\mathfrak{S}(\mathscr{K})$-transversality means b-stability, and thus we can summarize our considerations as

Proposition 1. $\{\mathfrak{F}_c\}_{c\in\mathfrak{S}(\mathscr{K})}$, *with the data E and δ given as above, is a differentiable stratification of the differentiable target $\mathfrak{F}_{\mathscr{K}}$. Furthermore, the $\mathfrak{S}(\mathscr{K})$-objects are just the $\mathscr{K}$-families, and $\mathfrak{S}(\mathscr{K})$-universality means $\mathscr{K}$-universality.*

5. The Differentiable Stratification $\{\mathfrak{F}_L\}_{L\in\Lambda(c)}$

Let c be a fixed subconstellation of a constellation $\sum n_i\tau_i$ with Hopf index $m\leqq 1$. For a c-function f, let $S_f\subset D^2$ denote its singular set and $V_f:=\mathbb{R}^{S_f}$ the $\sum n_i$-dimensional vector space of mappings $S_f\to\mathbb{R}$. These spaces form, canonically, a vector bundle target $V_c\to\mathfrak{F}_c$, since for any $\mathfrak{F}_c$-object Φ over M the singular set $S_\Phi\subset D^2\times M$ is a differentiable covering over M. The bundle $\Phi^*V_c=:V_\Phi$ shall be called the singular value bundle or the *value bundle* of Φ. To avoid an overloaded notation, we shall occasionally drop the Φ and write V for the value bundle of Φ and V_u for its fibre over $u\in M$, and similarly S and S_u for the singular set and fibre.

The actual singular values of Φ define the *value section* $v_\Phi: M\to V_\Phi$. Since for each simply connected neighborhood U of u in M we have canonical trivializations $S_{\Phi|U}\cong U\times S_u$ and $V_{\Phi|U}\cong U\times V_u$ correspondingly, the value section has a well defined "fibre component" $\bar v: U\to V_u$, and its differential at u is a linear map $d\bar v_u: T_uM\to V_u$. – Next, let us introduce a notation for the various "diagonals" in $V_f=\mathbb{R}^{S_f}$: If X is a finite set and Z a decomposition of it, then we write $\Delta^Z\subset\mathbb{R}^X$ for the subspace of those mappings $X\to\mathbb{R}$, which are constant on the elements of Z. If Z' is finer than Z, written $Z'\leqq Z$, then we have $\Delta^Z\subset\Delta^{Z'}$. Now if f is a c-function and Σ_f the decomposition of S_f into *contact classes* (see [7]), then we define $E_f:=V_f/\Delta^{\Sigma_f}$, and if Φ is a c-object over M and $u\in M$, then we define $\delta\Phi_u: T_uM\to E_{\Phi(u)}$ to be the composition $T_uM\xrightarrow{d\bar v_u} V_u\xrightarrow{\text{proj}} V_u/\Delta^{\Sigma_u}$. These E and δ will later form the structural data of our stratification, but before we can

establish them as such, we must introduce the *level situations* of c into our context. – The level situation $L(f)$ of a c-function f is its class under contact preserving c-isotopy (see [7]). Let Φ be a c-object over M and $u \in M$. If two singularities of $\Phi(u)$ don't have contact, then there is a whole neighborhood of u in M over which the corresponding singularities don't have contact: contact can suddenly be broken but not suddenly be tied. In particular, $\mathfrak{F}_L$-objects must in fact be contact preserving, and this observation implies the

Note. The $\mathfrak{F}_L$-objects are just the L-families in the sense of [7], and thus the universal L-families constructed in [7] are $\mathfrak{F}_L$-universal.

But this also implies, that for an $\mathfrak{F}_L$-object Φ over M the "contact diagonal" $\Delta^{\Sigma\Phi}$ defines a differentiable subbundle of the value bundle V_Φ. Thus the quotient $E_\Phi := V_\Phi / \Delta^{\Sigma\Phi}$ becomes a differentiable vector bundle over M, and in this way we have defined our vector bundle target $E_L \to \mathfrak{F}_L$. Together with δ given as above, we have now introduced the *data*, and it remains to check the axioms and the additional Transversality Property for $\{\mathfrak{F}_L\}_{L \in \Lambda(c)}$. Again, Finiteness and Connectedness are already known.

Openness of Transversality is the easiest to check: Let a c-object Φ over M be given, $u_0 \in M$ and U a simply connected open neighborhood of u_0 in M. We use the canonical isomorphism $S_{\Phi|U} \cong U \times S_{u_0}$ to describe the value section of $\Phi|U$ by its fibre component $\bar{v}: U \to V_{u_0}$ and to understand decompositions of the S_u, for $u \in U$, as decompositions of S_{u_0}. Then Φ is transversal at $u \in U$ if and only if $\bar{v}$ is transversal at u to the contact diagonal $\Delta^{\Sigma u}$. But since contact cannot suddenly be tied, we will have $\Sigma_u \leqq \Sigma_{u_0}$ for all $u \in U$ if U is sufficiently small, hence $\Delta^{\Sigma u} \supset \Delta^{\Sigma u_0}$, and thus if Φ is transversal at u_0, it will be transversal in a whole neighborhood: our system satisfies the Openness of Transversality Axiom.

Now suppose Φ to be transversal at u_0, and denote the composition $U \xrightarrow{\bar{v}} V_{u_0} \to V_{u_0}/\Delta^{\Sigma u_0}$ by $\bar{\bar{v}}: U \to E_{\Phi(u_0)} =: E_0$. Then the differential of $\bar{\bar{v}}$ at u_0 is just the surjective $\delta\Phi_{u_0}$. Thus up to suitable local coordinates and after making U smaller, we may assume the following about U and $\bar{\bar{v}}$:(i) $U = A \times B$, where A is an open ball in some $\mathbb{R}^k$ and B is an open ball around 0 in E_0, (ii) $\bar{\bar{v}}$ is just the projection $A \times B \to B \subset E_0$, and (iii) $\Sigma_u \leqq \Sigma_{u_0}$ for all $u \in U$. – We are now going to study the decomposition of U into the strata $(\Phi|U)^{-1}(\mathfrak{F}_L)$ by way of comparison with the decomposition of U, which is derived from the "diagonals" in E_0. More precisely: If Z is any decomposition of the singular set S_{u_0} over u_0, which is finer than the contact decomposition $(Z \leqq \Sigma_{u_0})$, then it defines a "diagonal subspace" $E^Z := \Delta^Z / \Delta^{\Sigma u_0} \subset E_0$. The connectedness components of $E^Z \setminus \bigcup_{Z < Z' \leqq \Sigma_{u_0}} E^{Z'}$ shall be called *compartments of type* Z. The decomposition of E_0 into compartments induces a corresponding decomposition of U, and if X is a compartment of type Z, we call $A \times (B \cap X)$ a "pseudostratum" of type Z. What, then, is the relation between strata and pseudostrata in U?

To state the basic fact about the relation between contact and singular values (lemma below), we introduce the following terminology: Let f be a c-function, S_f its singular set, $A_f := \{x \in D^2 | f(x) \text{ critical value}\}$ the union of its critical levels. A connectedness component of $A_f \setminus S_f$ is called an *arm* of $z \in S_f$, if z is in its closure. Two singularities are said to have *direct contact*, if they have a common arm.

Lemma (Hendlmeier [3]): *Let φ be a c-object over an open interval $(-\varepsilon, \varepsilon) \subset \mathbb{R}$ and let $t \mapsto z_i(t)$, $i = 1, 2$ be two sections of the singular covering $S_\varphi \to (-\varepsilon, \varepsilon)$ with $\varphi_t(z_1(t))$*

$= \varphi_t(z_2(t))$ *for all* t. *Then if* z_1 *and* z_2 *have direct contact over* $t=0$, *they will keep direct contact over a whole neighborhood of* 0 *in* $(-\varepsilon, \varepsilon)$.

Proof. Let X be the closure of a common arm between $z_1(0)$ and $z_2(0)$. We define a germ at X of a vector field ξ on $D^2 \times (-\varepsilon, \varepsilon)$ over $\frac{\partial}{\partial t}$ satisfying $\nabla_\xi \varphi = \frac{d}{dt}\varphi_t(z_1(t))$ as follows: The condition is convex, hence we only have the "local" task of finding such a germ at each point of X. At the regular points this is no problem, and at z_1 and z_2, the local unfolding theory does provide us with such germs (local triviality of τ-families, see f.i. [5, p. 149]). – A representative of such a germ will, for a sufficiently small open neighborhood Y of X in $D^2 \times 0$ and a sufficiently small $\delta > 0$, be integrable on $Y \times (-\delta, \delta)$, the flow $h: Y \times (-\delta, \delta) \to D^2 \times (-\delta, \delta)$ will satisfy $\varphi_t(h_t(y)) = \varphi_0(y) + \varphi_t(z_1(t))$ for all $(y, t) \in Y \times (-\delta, \delta)$, and $h_t(X)$ will be the closure of a common arm of $z_1(t)$ and $z_2(t)$ for all those t. q.e.d.

Now we return to our strata and pseudostrata. We denote the level situation at u_0 by L_0 and *claim*: (a) The L_0-stratum is contained in the pseudostratum $A \times 0$ of type Σ_{u_0}, and (b) Each pseudostratum is contained in a stratum. – *Proof.* A point u belongs to a pseudostratum of type Z if and only if for each $\sigma \in \Sigma_{u_0}$, the decomposition of σ into "value levels at u", i.e. into the levels of $v(u)|\sigma : \sigma \to \mathbb{R}$, coincides with the decomposition as given by $Z \leqq \Sigma_{u_0}$. In particular, if u is an L_0-point, then $\Sigma_u = \Sigma_{u_0}$ because of our assumption $\Sigma_u \leqq \Sigma_{u_0}$, and since the value is constant on contact classes, the condition is satisfied for $Z = \Sigma_{u_0}$ and (a) follows. To prove (b), we must show that each c-object $\gamma^* \Phi$ over $[0, 1]$, which is induced from Φ by a path γ in a pseudostratum, is contact preserving. Because of $\Sigma_u \leqq \Sigma_{u_0}$, contact can only occur within the contact classes $\sigma \in \Sigma_{u_0}$. But since γ remains in a pseudostratum, the value levels within σ do not depend on $t \in [0, 1]$, and applying the above lemma we conclude: All direct contact and hence *all* contact in $\gamma^* \Phi$ is preserved locally on $[0, 1]$. Since contact cannot suddenly be tied anyway, this implies that the whole contact relation of $\gamma^* \Phi$ is locally constant on $[0, 1]$, hence constant, and (b) is proved. – It follows from (a) and (b), first of all, that the L_0-stratum is $A \times 0$, its actual normal bundle in U and its "abstract normal bundle" $(\Phi | A \times 0)^* E_{L_0}$ both are canonically the same as $A \times E_0$, and δ induces the identity. Thus we have verified the Submanifold Axiom for our system. Moreover, (b) shows that "pseudostrata preserving" implies "strata preserving", and the Tubular Homogeneity and Local Triviality Axioms follow from corresponding obvious properties of the pseudostratification.

Thus it remains to show that our data satisfy the Tubular Extension Axiom and have the Transversality Property (p. 71). As a preparation for both, we shall prove the "Value Realization Lemma" below. To introduce its terminology, consider a c-object φ over M, with value bundle V and value section $v_\varphi = : v : M \to V$. Let $\pi : W \to M$ be some other vector bundle over M. By a *value extension* of φ on W we mean a germ w at $M \subset W$ of a map over M (i.e. fibre preserving) from an open neighborhood of M in W into V, with $w|M = v$. If Φ is an extension of φ to a c-object germ on W at M, then its value bundle germ is canonically the germ of $\pi^* V$, and hence we may and shall consider its value section germ v_Φ as a value extension of v_φ on W. Now if $v_\Phi = w$, we say that Φ *realizes* the value extension w.

Value Realization Lemma. *Let* φ *be a c-object over a manifold* M *with boundary and* w *a value extension of* φ *on a vector bundle* W *over* M. *Then any realization of the value extension* $w|\partial W$ *of* $\varphi|\partial M$ *can be extended to a realization of* w.

Proof. Let ψ be a realization of $w|\partial W$. We have to extend it to a realization Φ of w. We will do this in three constructive steps. The *first step* is to choose the (germs of the) singular set of the future Φ: Let c have the constellation $\Sigma n_i\tau_i$. For each i, we extend the "τ_i-part" $S_{\varphi i}$ of the singular set of φ to a germ S_i at $S_{\varphi i}$ of a 2-codimensional submanifold of $D^2 \times W$, which under $D^2 \times W \to W$ covers an open neighborhood of the zero section and which over the boundary coincides with $S_{\psi i}$. The simply amounts to extending a C^∞ map $S_{\varphi i} \to \mathring{D}^2$ to a map germ on $\pi_i^* W$, where $\pi_i : S_{\varphi i} \to M$ is the projection; with prescribed boundary conditions. – *Second step:* Now we construct the germ on $D^2 \times W$ at $S = \bigcup S_i$ of the future Φ. This means, we extend the function germ f on $D^2 \times M$ at S_φ, as given by $\varphi : D^2 \times M \to \mathbb{R}$, to a germ F on $D^2 \times W$ at S_φ, such that (i) over the boundary, F coincides with the germ given by ψ, (ii) $F|S$ is as prescribed by w in the obvious way, and (iii) S is the germ of the singular set of one and hence any representative of F, and the singularities of F at S_i are of type τ_i. – Why is this possible? Formulated in neutral notation, it amounts to the following problem: Given a vector bundle ξ over a manifold X, a τ-family g over X (see [5]) and an extension h of $g|\partial X$ to the germ at ∂X of a τ-family on $\partial\xi$. Find an extension H of g to a germ at X of a τ-family on ξ with $H|\partial\xi = h$. – Now this problem can be solved by either using the classification of τ-families [5] or else by first solving it on a sufficiently high jet-level, where we have the advantage that the r-jets of type τ form a submanifold of the jet space, and then correcting this crude solution, which means to solve an analogous problem for families with vanishing r-jet, which is reduced to a simple local question by partition of unity. – *Third and last step:* Now we extend $\varphi : D^2 \times M \to \mathbb{R}$ to a function germ Φ on $D^2 \times W$ at $D^2 \times M$, of which we only require, that it behaves normally at $S^1 \times W$, that over the boundary it is given by ψ and that its germ at S_φ is F. Then Φ, without further adjustment, is an extension of ψ to a realization of w as desired. q.e.d.

Now we apply the lemma to verify the Tubular Extension Axiom for $\{\mathfrak{F}_L\}_{L \in \Lambda(c)}$. Let φ be an $\mathfrak{F}_L$-object over M and $E := \varphi^* E_L = V/\Delta^0$, where V is the value bundle and Δ^0 the contact diagonal subbundle of φ. Whether or not an extension Φ of φ to a c-object germ on E at M is a tubular extension, solely depends on the value extension $w : (E, M) \to V$ which it realizes. More precisely: Φ is a tubular extension of φ if and only if the composition $E \to V \to V/\Delta^0$ of the fibre differential of w with the canonical projection is the identity on E. Now let a tubular extension ψ of $\varphi|\partial M$ be given, realizing some value extension w_0 on ∂E. Extend w_0 to a value extension w of v_φ with the property just mentioned, which is a local problem and in fact no problem at all, and apply the Value Realization Lemma to find the desired extension of ψ to a tubular extension Φ of φ.

Thus we have verified the Tubular Extension Axiom, and it remains to show, that our differentiable stratification of $\mathfrak{F}_c$ has the Transversality Property. Let φ be a c-object over M such that $\varphi|\partial M$ is transversal. Choose a Riemannian metric on the value bundle V, a positive function ε on M and an extension of φ to a c-object Φ over the ε-disc bundle $D_\varepsilon V$, which actually realizes the canonical value extension map $D_\varepsilon V \to V$ (translation by the value section $v : M \to V$). All this is possible by the Value Realization Lemma. – Now if $\alpha : M \to D_\varepsilon V$ is a section which vanishes on ∂M, then $\alpha^* \Phi$ is a c-object which is c-isotopic to φ. Thus all we have to do is to find α such that in addition $\alpha^* \Phi$ is transversal. But Φ certainly is transversal, and an object induced from Φ is transversal if and only if the inducing map is transversal to the strata of Φ. Thus the zero section on ∂M (inducing φ) is transversal to the

strata, and hence can be extended, by an ordinary transversality argument, to a section $\alpha: M \to D_\varepsilon V$ transversal to the strata; and thus we have proved that $\{\mathfrak{F}_L\}_{L \in \Lambda(c)}$ has the Transversality Property. – We summarize the present section into

Proposition 2. *$\{\mathfrak{F}_L\}_{L \in \Lambda(c)}$ with data E and δ as defined above, is a differentiable stratification of $\mathfrak{F}_c$ and satisfies the additional Transversality Property.*

6. Conclusion

Well then! Gathering Theorem 2, the Propositions 1 and 2, the two Notes on pp. 71 and 81 and the construction of the universal L-families in [7] all together, we obtain as corollary the

Theorem 3. *If $\mathscr{K}$ is a complete, finite set of constellations of a common Hopf index $m \leqq 1$, then there exists a $\mathscr{K}$-family over a compact manifold M with boundary, which over $M \backslash \partial M$ is $\mathscr{K}$-universal.*

But in fact, we do know more than the mere *existence* of universal $\mathscr{K}$-families. The universal L-families, from which our inductive construction (Proof of Theorem 2) starts, are described very explicitly in [7], following the induction it should therefore be possible to get more specific information about the universal $\mathscr{K}$-families; and the "Criterion" (Theorem 1) might help to detect universal $\mathscr{K}$-families among "naturally" occurring $\mathscr{K}$-families.

Acknowledgement. I wish to thank Th. Bröcker for useful comments.

References

1. Arnol'd, V.I.: Normal forms for functions near degenerate critical points, the Weyl groups of A_k, D_k, E_k, and Lagrangian singularities. Funkcional. Anal. i Prilozen **6**, 3–25 (1972); Functional Anal. Appl. **6**, 254–272 (1972)
2. Dornisch, H.-P.: Ein Universalitätskriterium für $\mathscr{K}$-Familien. Zulassungsarbeit Regensburg 1978
3. Hendlmeier, F.: Lagenstabile Familien konstanter Konstellation. Diplomarbeit Regensburg 1978
4. Jänich, K.: Symmetry properties of singularities of C^∞-functions. Math. Ann. **238**, 147–156 (1978)
5. Jänich, K.: Global families of local unfoldings. Math. Ann. **242**, 147–158 (1979)
6. Jänich, K.: Symmetry properties of certain C^∞-functions on the 2-dimensional disk. Topology **19**, 111–133 (1980)
7. Jänich, K.: Contact preserving families of functions. Topology **19**, 283–290 (1980)
8. Maier, Chr.: Über C^∞-Funktionen mit vorgegebenen Singularitäten auf D^2 (in preparation)
9. Reich, A.: b-stabile Entfaltungen von Familien konstanter Konstellation. Zulassungsarbeit Regensburg 1980
10. Thom, R.: Stabilité structurelle et morphogénèse. New York: Benjamin 1972. (Translated by D. H. Fowler: Structural stability and morphogenesis. New York: Benjamin-Addison Wesley 1975)

Received December 18, 1980

Math. Ann. 256, 85–94 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Some Estimates for Type and Cotype Constants

H. König[1] and L. Tzafriri[2,*]

1 Mathematisches Seminar der Universität, Olshausenstrasse 40–60, D-2300 Kiel
2 Department of Mathematics, Hebrew University, Givat Ram, Jerusalem, Israel

Introduction

The notions of type and cotype have been successfully used in calculations of projections constants and Banach-Mazur distances of finite dimensional spaces to Hilbert spaces. In many cases, these calculations require to know whether the type and the cotype constants of a given n-dimensional space can be computed with only n vectors.

Let X be a Banach space, $1 \leqq p \leqq 2 \leqq q < \infty$ and $n \in N$. We define $a_{p,n}(X)$, $b_{q,n}(X)$, $\alpha_{p,n}(X)$, and $\beta_{q,n}(X)$ as the smallest constants such that for every choice of $x_1, \ldots, x_n$ in X, we have

$$b_{q,n}(X)^{-1}\left(\sum_{j=1}^{n}\|x_j\|^q\right)^{1/q} \leqq \left(\int_0^1 \left\|\sum_{j=1}^{n} r_j(t)x_j\right\|^2 dt\right)^{1/2} \leqq a_{p,n}(X)\left(\sum_{j=1}^{n}\|x_j\|^p\right)^{1/p}$$

$$\beta_{q,n}(X)^{-1}\left(\sum_{j=1}^{n}\|x_j\|^q\right)^{1/q} \leqq \left(\int_0^1 \left\|\sum_{j=1}^{n} g_j(\omega)x_j\right\|^2 dP\right)^{1/2} \leqq \alpha_{p,n}(X)\left(\sum_{j=1}^{n}\|x_j\|^p\right)^{1/p},$$

where $\{r_j\}_{j=1}^{\infty}$ are the Rademacher functions and $\{g_j\}_{j=1}^{\infty}$ is a sequence of independent standard Gaussian variables on a probability space (Ω, P). The space X is said to be of (Rademacher) type p or cotype q if $a_p(X) = \sup_{n\in N} a_{p,n}(X) < \infty$, respectively, $b_q(X) = \sup_{n\in N} b_{q,n}(X) < \infty$. Similarly, X is said to be of Gaussian type p or cotype q if $\alpha_p(X) = \sup_{n\in N} \alpha_{p,n}(X) < \infty$, respectively, $\beta_q(X) = \sup_{n\in N} \beta_{q,n}(X) < \infty$.

The most important result on type and cotype constants of a finite dimensional Banach space was obtained by Tomczak-Jaegermann [11] who showed that, for Gaussian constants of a n-dimensional space X_n, one has

$$\alpha_2(X_n) \leqq 2\alpha_{2,n}(X_n) \quad \text{and} \quad \beta_2(X_n) \leqq 2\beta_{2,n}(X_n).$$

As a corollary, she got that the Banach-Mazur distance of X_n to the n-dimensional Euclidean space satisfies

$$d(X, l_2^n) \leqq Cn^{1/p-1/q},$$

* Research supported in part by NSF Grant MCS 79-03042

0025-5831/81/0256/0085/$02.00

where $C=4\alpha_{p,n}(X_n)\beta_{q,n}(X_n)$ (previously, it was shown in [2] that the Gaussian type 2 and cotype 2 constants of X_n can be computed by using $n(n+1)/2$ vectors). N. Tomczak-Jaegerman's result was extended, by a different argument, to Gaussian type $1<p<2$ and cotype $q>2$ in [5] yielding

$$\alpha_p(X_n)\leqq C_p\alpha_{p,n}(X_n) \quad \text{and} \quad \beta_q(X_n)\leqq C_q\beta_{q,n}(X_n), \tag{1}$$

where C_p and C_q are constants depending only on p, respectively q.

In the first part of this note, we give a simplified proof of the interpolation argument used to prove (1). As in [11], problems on the type and cotype constants are reduced to corresponding questions for (q,r)-absolutely summing operators. Given $1\leqq r\leqq q<\infty$, $n\in N$ and a map $T:X\to Y$, we let $\pi_{q,r}^{(n)}(T)$ be the smallest constant C such that, for any n vectors $x_1,\ldots,x_n$ in X, we have

$$\left(\sum_{i=1}^n \|Tx_i\|^q\right)^{1/q}\leqq C\sup_{\|x^*\|x^*\leqq 1}\left(\sum_{i=1}^n |\langle x^*,x_i\rangle|^r\right)^{1/r}.$$

The operator T is called (q,r)-absolutely summing if $\pi_{q,r}(T)=\sup_{n\in N}\pi_{q,r}^{(n)}(T)<\infty$. Given $S:l_2^n\to X$, we define the ideal norm $l(S)$ by

$$l(S)=\left(\int_\Omega\left\|\sum_{i=1}^n g_i(\omega)Se_i\right\|^2 dP\right)^{1/2},$$

where $\{e_i\}_{i=1}^n$ denotes the unit vector basis of l_2^n and $\{g_i\}_{i=1}^n$ has the same meaning as above.

By arguments similar to those used in [11], one proves the

Lemma. *For every Banach space X*

$$\beta_{q,n}(X)=\sup\{\pi_{q,2}^{(n)}(T);T:l_2^n\to X \quad \textit{with} \quad l(T)\leqq 1\}$$

$$\alpha_{p,n}(X)=\sup\{l(S);S:l_2^n\to X \quad \textit{with} \quad (\pi_{p',n}^{(n)})^*(S)\leqq 1\},$$

where $p'=p/(p-1)$ and $(\pi_{q',2}^{(n)})^*$ denotes the adjoint ideal norm.

Thus, (1) follows immediately if we can prove that

$$\pi_{q,2}(T_n)\leqq C_q\pi_{q,2}^{(n)}(T),$$

for every operator $T_n:X\to Y$ of rank n. This will be done in part one. We will also show that a similar statement holds for $(q,1)$-absolutely summings norms.

In the second part we consider the corresponding problems for the Rademacher type p and cotype q constants (which are normally referred to simply as type and cotype constants). Since the Rademacher and the Gaussian type constants are equivalent we obtain that, for any n-dimensional space X_n,

$$a_p(X_n)\leqq C_p a_{p,n}(X_n).$$

The Gaussian and the Rademacher cotype constants, however, may be inequivalent. For instance, $b_2(l_\infty^n)=n^{1/2}$ while $\beta_2(l_\infty^n)$ behaves asymptotically as $(n/\log(n+1))^{1/2}$ (cf. [2]). In the cotype case, we shall prove that

$$b_q(X_n)\leqq C_q b_{q,n}(X_n)(\log[b_{q,n}(X_n)+1])^{1/2}.$$

We do not know, however, whether the logarithmic term is really necessary. Using a similar method of proof, we also derive a formula for the cotype index q of a Banach space in terms of its type p constant (as well-known [9], if X is of some type $p>1$ then it is also of some cotype $q<\infty$). More exactly, we prove that any Banach space X is of cotype q for

$$q=C_p a_p(X)^{p'}, \tag{2}$$

where C_p depends only on p. This estimate is, in some sense, the best possible. Indeed, if X is the space l_r, $r>2$, then X is of cotype r but not of cotype $q<r$. On the other hand,

$$a_p(l_r)\leqq a_2(l_r)^{2/p'}\leqq B_r^{1/p'},$$

with B_r being the Khintchine constant for the Rademacher r-averages which, by Haagerup [4], is of order of magnitude $(r/e)^{1/2}$, as $r\to\infty$. Thus, the value of q in (2) should satisfy $q\geqq e a_p(X)^{p'}$, in general, and $q\geqq e a_2(X)^2$, in particular. This means that $C_2\geqq e$; our computations, however, yield only the estimate $C_2\leqq 6$ [for spaces X where $a_2(X)$ is relatively large one can get even $C_2\leqq 4$].

For the n-dimensional subspaces X_n of a space X of type 2, formula (2) implies that the relative projection constants $\lambda_X(X_n)$ and the Banach-Mazur distances to l_2^n satisfy

$$\lambda_X(X_n)\leqq 16 a_2(X) n^{1/2-\varepsilon}$$

$$d(X_n, l_2^n)\leqq 16 a_2(X) n^{1/2-\varepsilon},$$

where $\varepsilon=1/6a_2(X)^2$.

1. Gaussian Constants and $(q,2)$ Summing Operators

We begin by stating

Theorem 1. *Let $1<p<2<q<\infty$. Then there is $C_q\leqq C/(q-2)$ such that for all $n\in N$ and all n-dimensional spaces X_n*

$$\alpha_p(X_n)\leqq C_{p'}\alpha_{p,n}(X_n) \quad \text{and} \quad \beta_q(X_n)\leqq C_q\beta_{q,n}(X_n).$$

To prove Theorem 1, it suffices, by the lemma in the introduction, to prove

Proposition 1. *For any $q>2$, there is $C_q\leqq C/(q-2)$ such that, for any $n\in N$ and any rank n operator $T:X\to Y$,*

$$\pi_{q,2}(T)\leqq C_q\pi_{q,2}^{(n)}(T).$$

Proof. (This is a simplified version of the argument in [5].) We recall first that the approximation numbers of an operator $S:X\to Y$ are $\alpha_j(S)=\inf\{\|S-S_j\|;\, S_j:X\to Y$ of rank $<j\}$, $j\in N$. For $0<r\leqq\infty$, let

$$S_r(X,Y)=\left\{S:X\to Y;\, \sigma_r(S)=\left(\sum_{j=1}^{\infty}\alpha_j(S)^r\right)^{1/r}<\infty\right\}$$

$$S_{2,1}(X,Y)=\left\{S:X\to Y;\, \sigma_{2,1}(S)=\sum_{j=1}^{\infty}\alpha_j(S)j^{-1/2}<\infty\right\}.$$

Thus $\sigma_{2,1}(S)$ is the norm of $(\alpha_j(S))_{j\in N}$ in the Lorentz sequence space $l_{2,1}$.

Since $\pi_{q,2}(T)=\sup\{\pi_{q,2}(TA): A: l_2 \to X, \ \|A\| \leqq 1\}$, it suffices to prove Proposition 1 for maps of the form $T: l_2^n \to Y$. We will show that

$$\pi_{q,2}(T) \leqq C_q \sigma_q(T) \leqq C_q \pi_{q,2}^{(n)}(T). \tag{3}$$

Step 1. There is an orthonormal basis $(e_j)_{j=1}^n$ of l_2^n with $\alpha_j(T) \leqq \|Te_j\|$, for all $1 \leqq j \leqq n$. Indeed, if $j<n$ orthonormal vectors have been found, let $Y_j=[e_1, \ldots, e_j]$ and $P_j: l_2^n \to Y_j \subseteq l_2^n$ be the orthogonal projection onto Y_j. Thus, $\alpha_{j+1}(T)$ $\leqq \|T-TP_j\| = \|T|Y_j^{\perp}\|$. Hence, there is $e_{j+1} \in Y_j^{\perp}$ of norm one with $\alpha_{j+1}(T)$ $\leqq \|Te_{j+1}\|$. Since $\operatorname{rank} T \leqq n$, $\alpha_k(T)=0$ for $k>n$. This implies that

$$\sigma_q(T)=\left(\sum_{j=1}^n \alpha_j(T)^q\right)^{1/q} \leqq \left(\sum_{j=1}^n \|Te_j\|^q\right)^{1/q} \leqq \pi_{q,2}^{(n)}(T).$$

Step 2. To prove the left-hand side inequality in (3) by interpolation, we need first the limiting case as $q \to 2$. Take $S \in S_{2,1}(X, Y)$. Choosing $D_j: X \to Y$ of $\operatorname{rank} D_j < 2^j$ with $\|S-D_j\| \leqq 2\alpha_{2^j}(S)$, we can write S as a telescopic sum $S=\sum_{j=0}^{\infty} S_j$, $\|S_j\| \leqq 4\alpha_{2^j}(S)$ with $\operatorname{rank} S_j < 2^{j+2}$ (take $S_j=D_{j+1}-D_j$). Since the $(2,2)$-absolutely summing norm on a m-dimensional space is $\sqrt{m}$ [3], we infer for $M<N$ that

$$\begin{aligned}\pi_2\left(\sum_{j=M+1}^N S_j\right) &\leqq \sum_{j=M+1}^N \pi_2(S_j) \leqq \sum_{j=M+1}^N \|S_j\| 2^{(j+2)/2} \\ &\leqq 16 \sum_{k=2^M}^{2^N} k^{-1/2} \alpha_k(S) \to 0,\end{aligned}$$

as $N>M \to \infty$ since $S \in S_{2,1}(X, Y)$. Thus, S is 2-absolutely summing with $\pi_2(S)$ $\leqq 16\sigma_{2,1}(S)$.

Step 3. We show that the K-functional of the real interpolation theory [1] satisfies the following equivalence

$$K(t, T; S_1(X, Y), S_\infty(X, Y)) \sim K(t, (\alpha_j(T))_{j=1}^\infty; l_1, l_\infty). \tag{4}$$

Here S_∞ denotes all continuous linear maps. By the definition of the K-functional,

$$\begin{aligned}K(t, T; S_1, S_\infty) &= \inf\left\{\sum_{j \in N} \alpha_j(T_t) + t\|T-T_t\|; T_t: X \to Y\right\} \\ &\geqq \inf\left\{\sum_{j=1}^{[t]} (\alpha_j(T_t) + \|T-T_t\|)\right\} \geqq \sum_{j=1}^{[t]} \alpha_j(T) \\ &\sim K(t, (\alpha_j(T))_{j=1}^\infty; l_1, l_\infty);\end{aligned}$$

for the last equivalence see [1]. For $t \geqq 1$, choose $T_t: X \to Y$ of rank $<[t]$ with $\|T-T_t\| \leqq 2\alpha_{[t]}(T)$. Then

$$\alpha_j(T_t) \leqq \alpha_j(T) + \|T-T_t\| \leqq 3\alpha_j(T)$$

for all $j<[t]$ and $\alpha_j(T_t)=0$ for $j \geqq [t]$. Thus,

$$\begin{aligned}K(t, T; S_1, S_\infty) &\leqq \sum_{j=1}^{[t]} \alpha_j(T_t) + t\|T-T_t\| \\ &\leqq 3 \sum_{j=1}^{[t]} (\alpha_j(T) + \alpha_{[t]}(T)) \leqq 6 \sum_{j=1}^{[t]} \alpha_j(T),\end{aligned}$$

proving (4).

Step 4. The Lorentz space $l_{2,1}$ is a real interpolation space: $l_{2,1}=(l_1, l_\infty)_{1/2,1}$ [1]. Hence, (4) implies that

$$S_{2,1}(X, Y)=(S_1(X, Y), S_\infty(X, Y))_{1/2,1}.$$

Let $q>2$, $1/q=(1-\theta)/2$ and $\eta=(1+\theta)/2$. Then, by the reiteration theorem [1] or [7],

$$(S_{2,1}(X, Y), S_\infty(X, Y))_{\theta,q}=(S_1(X, Y), S_\infty(X, Y))_{\eta,q}=S_q(X, Y),$$

where the last inequality follows from (4) and the fact that $(l_1, l_\infty)_{\eta,q}=l_q$. Denoting the class of all $(q, 2)$-absolutely summing operators X and Y by $\Pi_{q,2}(X, Y)$, we have [for $1/q=(1-\theta)/2$]

$$(\Pi_{q,2}(X, Y), \Pi_{\infty,2}(X, Y))_{\theta,q}\subseteq\Pi_{q,2}(X, Y),$$

with $\Pi_{\infty,2}=S_\infty$. This is an easy consequence of

$$l_q(X)=(l_2(X), l_\infty(X))_{\theta,q}.$$

Using these equalities and inclusions and step 2, we get $S_q(X, Y)\subseteq\Pi_{q,2}(X, Y)$. Thus, there is C_q such that, for any $T: X\to Y$, $\pi_{q,2}(T)\leqq C_q\sigma_q(T)$. Actually, C_q depends only on q: the inequality $C_q\leqq C/(q-2)$ can be verified by checking the constants in the reiteration theorem.

Corollary. *For any $q>2$, there is C_q such that, for any $n\in N$ and any rank n operator $T: X\to Y$,*

$$\pi_{q,1}(T)\leqq C_q\pi_{q,1}^{(n)}(T).$$

Proof. It is well-known that

$$\begin{aligned}\pi_{q,1}(T)&=\sup\{\pi_{q,1}(TA); A: l_\infty\to X, \|A\|=1\}\\&\leqq\sup\{\pi_{q,2}(TA); A: l_\infty\to X, \|A\|=1\}.\end{aligned}$$

Thus, by Proposition 1,

$$\pi_{q,1}(T)\leqq C_q\sup\{\pi_{q,2}^{(n)}(TA); A: l_\infty\to X, \|A\|=1\}.$$

By Maurey [8], the $(q, 1)$- and $(q, 2)$-absolutely summing norms are equivalent on $\Pi_{q,1}(l_\infty, X)$ and the argument does not depend on the number of vectors considered. Thus,

$$\pi_{q,1}(T)\leqq D_q\sup\{\pi_{q,1}^{(n)}(TA); \|A\|=1\}\leqq D_q\pi_{q,1}^{(n)}(T).$$

2. Rademacher Type and Cotype Constants

The Gaussian and Rademacher constants of a space X are related by

$$\left.\begin{aligned}C_p^{-1}\alpha_{p,n}(X)\leqq a_{p,n}(X)\leqq C\alpha_{p,n}(X)\\\beta_{q,n}(X)\leqq Cb_{q,n}(X),\end{aligned}\right\}\tag{5}$$

where C is an absolute constant and C_p depends on p only. In fact, the last two inequalities follow from

$$\int_0^1\left\|\sum_{j=1}^n r_j(t)x_j\right\|^2 dt\leqq C^2\int_\Omega\left\|\sum_{j=1}^n g_j(\omega)x_j\right\|^2 dP(\omega),$$

cf. Pisier [10]. For the first inequality, use the symmetry of the g_j's to conclude that

$$\begin{aligned}\int_\Omega \left\| \sum_{j=1}^n g_j(\omega) x_j \right\|^p dP(\omega) &= \int_0^1 \left\| \sum_{j=1}^n r_j(t) g_j(\omega) x_j \right\|^p dP(\omega)\, dt \\ &\leqq \int_\Omega \left(\int_0^1 \left\| \sum_{j=1}^n r_j(t) g_j(\omega) x_j \right\|^2 dt \right)^{p/2} dP(\omega) \\ &\leqq a_{p,n}(X)^p \sum_{j=1}^n \|x_j\|^p .\end{aligned}$$

The equivalence of the Gaussian p- and 2-moments implies the first inequality of (5).

Theorem 2. *Let $1<p<2<q<\infty$. Then there are constants D_p and C_q, depending only on p, respectively q, such that for any $n\in N$ and any n-dimensional space X_n, we have*

$$\left.\begin{aligned} a_p(X_n) &\leqq D_p a_{p,n}(X_n) \\ b_q(X_n) &\leqq C_q b_{q,n}(X_n)[\log(b_{q,n}(X_n)+1)]^{1/2} . \end{aligned}\right\} \tag{6}$$

In fact, one has the stronger estimate

$$b_q(X_n) \leqq C_q \beta_{q,n}(X_n)[\log(b_{q,n}(X_n)+1]^{1/2} . \tag{7}$$

Remark. The last inequality is asymptotically the best possible as is seen by choosing $X_n = l_\infty^n$; whether (6) is so, too, it remains open. The constant C_q can be bounded by $C\sqrt{q}$ as $q\to\infty$.

Proof. The first inequality in (6) is an immediate consequence of Theorem 1 and (5). In order to prove (7), and therefore also the second inequality in (6), we will show that, for any $q>2$, any space X_n with $\dim X_n = n > 2^q b_{q,n}(X_n)^q$ and for any finite sequence $\{z_i\}_{i=1}^m$ in X_n, we have

$$\int_\Omega \left\| \sum_{i=1}^m g_i(\omega) z_i \right\|^2 dP(\omega) \leqq C_q^2 [\log(b_{q,n}(X_n)+1)] \int_0^1 \left\| \sum_{i=1}^m r_i(t) z_i \right\|^2 dt , \tag{8}$$

for a suitable choice of C_q. It would follow then, by using Theorem 1 and (5), that

$$\begin{aligned} b_q(X_n) &\leqq C_q [\log(b_{q,n}(X_n)+1)]^{1/2} \beta_q(X_n) \\ &\leqq C_q' [\log(b_{q,n}(X_n)+1)]^{1/2} \beta_{q,n}(X_n) \\ &\leqq C_q'' [\log(b_{q,n}(X_n)+1)]^{1/2} b_{q,n}(X_n) . \end{aligned}$$

In the case $n \leqq 2^q b_{q,n}(X_n)^q$, the same estimate is obtained in a trivial manner since, necessarily, $b_q(X_n) \leqq n^{1/q}$ and thus, $b_q(X_n) \leqq 2 b_{q,n}(X_n)$.

Hence, it suffices to prove (8). We will use ideas from Maurey and Pisier [9]. Let $\{z_i\}_{i=1}^m$ be an arbitrary sequence in X_n and denote

$$Z = \operatorname{span}\{r_i(t) z_i\}_{i=1}^m \subseteq L_2(X_n) .$$

Then $\{r_i(t) z_i\}_{i=1}^m$ is a 1-unconditional basis of Z. Define γ_j to be the smallest constant such that for any sequence $\{v_i\}_{i=1}^j$ in Z, of pairwise disjoint elements

[relative to the lattice structure inherited from the 1-unconditional basis $\{r_i(t)z_i\}_{i=1}^m$ of Z], one has

$$\inf_{1\leqq i\leqq j}\|v_i\|_Z\leqq\gamma_j\left\|\sum_{i=1}^{j}v_i\right\|_Z$$

Then $1=\gamma_1\geqq\gamma_2\geqq\ldots$ and $\gamma_{ij}\leqq\gamma_i\gamma_j$ for all i and j (see e.g., [7, p. 91]). Let $k=[2^q b_{q,n}(X_n)^q]+1$. In view of the condition imposed on n we have $k\leqq n$ and moreover, since $b_{q,n}(Z)\leqq b_{q,n}(L_2(X_n))=b_{q,n}(X_n)$ we get, for any sequence $\{v_i\}_{i=1}^k$ of pairwise disjoint vectors in Z, that

$$\left(\inf_{1\leqq i\leqq k}\|v_i\|_Z\right)k^{1/q}\leqq\left(\sum_{i=1}^{k}\|v_i\|_Z^q\right)^{1/q}\leqq b_{q,n}(X_n)\left\|\sum_{i=1}^{k}v_i\right\|_Z.$$

Hence, $\gamma_k\leqq b_{q,n}(X_n)k^{-1/q}\leqq 1/2$.

Let $r=2\log_2 k$. Then, with $\gamma_j=0$ for $j>m$, we have

$$\begin{aligned}\left(\sum_{j=1}^{m}\gamma_j^r\right)^{1/r}&\leqq k^{1/r}\left(\sum_{j=0}^{\infty}k^j\gamma_{kj}^r\right)^{1/r}\\&\leqq k^{1/r}\left(\sum_{j=0}^{\infty}k^j\gamma_k^{jr}\right)^{1/r}\\&\leqq k^{1/r}\left(\sum_{j=0}^{\infty}(k/2^r)^j\right)^{1/r}\leqq 2.\end{aligned}$$

Consequently, for any sequence $\{u_j\}_{j=1}^m$ of pairwise disjoint vectors in Z such that $\|u_1\|\geqq\|u_2\|\geqq\ldots\geqq\|u_m\|$, we have

$$\left(\sum_{j=1}^{m}\|u_j\|_Z^r\right)^{1/r}\leqq\left(\sum_{j=1}^{m}\gamma_j^r\right)^{1/r}\left\|\sum_{j=1}^{m}u_j\right\|_Z$$

i.e., Z satisfies a lower r-estimate for disjoint elements. Define $T:l_\infty^m\to Z$ by $\{a_i\}_{i=1}^m\to\sum_{i=1}^{m}a_i r_i(t)z_i$. Then T is a positive map [again relative to the lattice structure on Z inherited from the 1-unconditional basis $\{r_i(t)z_i\}_{i=1}^m$ of Z]. It follows (e.g., cf. [7, pp. 84–85]), that, for any choice of $a^{(j)}\in l_\infty^m$, $1\leqq j\leqq l$,

$$\begin{aligned}\left(\sum_{j=1}^{l}\|Ta^{(j)}\|_Z^r\right)^{1/r}&\leqq 2\left\|\sum_{j=1}^{l}|Ta^{(j)}|\right\|_Z\\&\leqq 2\|T\|\left\|\sum_{j=1}^{l}|a^{(j)}|\right\|_\infty\end{aligned}$$

and thus $\pi_{r,1}(T:l_\infty^m\to Z)\leqq 2\|T\|$. Since T is defined on l_∞^m, one has, by Maurey [8], that

$$\pi_{2r}(T)\leqq\frac{C}{2}\pi_{r,1}(T),$$

for some absolute constant C (independent of r!). Thus,

$$\pi_{2r}(T)\leqq C\|T\|.$$

The Pietsch factorization theorem yields, in this case, the existence of a sequence $\delta_i \geqq 0$ with $\sum_{i=1}^{m} \delta_i = 1$ so that

$$\|T(\{a_i\}_{i=1}^m)\| \leqq \pi_{2r}(T)\left(\sum_{i=1}^m |a_i|^{2r}\delta_i\right)^{1/2r}$$
$$\leqq C\|T\|\left(\sum_{i=1}^m |a_i|^{2r}\delta_i\right)^{1/2r}.$$

Integrating this inequality with $a_i = g_i(\omega)$, we obtain

$$\left[\int_\Omega \int_0^1 \left\|\sum_{i=1}^m g_i(\omega) r_i(t) z_i\right\|^2 dt\, dP(\omega)\right]^{1/2}$$
$$\leqq C\|T\|\left[\int_\Omega \left(\sum_{i=1}^m |g_i(\omega)|^{2r}\delta_i\right)^{2/2r} dP(\omega)\right]^{1/2}$$
$$\leqq C\|T\|\left(\int_\Omega \left(\sum_{i=1}^m |g_i(\omega)|^{2r}\delta_i\right) dP(\omega)\right)^{1/2r}$$
$$\leqq C\|T\|\,\|g_1\|_{2r}.$$

The left-hand side inequality is nothing but the Gaussian 2-average; concerning the right-hand side inequality, notice that $\|T\| = \left(\int_0^1 \left\|\sum_{i=1}^m r_i(t) z_i\right\|^2 dt\right)^{1/2}$, i.e., the Rademacher average, and $\|g_1\|_{2r} \leqq C'\sqrt{r}$. This proves (8) since $r \sim q \log b_{q,n}(X_n)$.

It is well-known that a Banach space of type p with $p>1$ is of cotype q for some $q<\infty$. Similar ideas as in the previous proof can be used to derive a quantitative version of this result:

Theorem 3. *Let $p>1$, $p'=p/(p-1)$ and X be a Banach space. If $\dim X = 2^n$ and $q = 2+2^{p'}\, a_{p,n}(X)^{p'}$, then $b_q(X) \leqq 2$. In particular, if X is of type $p>1$ then it is of cotype q, for $q = 2+2^{p'}\, a_p(X)^{p'}$.*

Proof. Clearly, it suffices to prove the first assertion.

Step 1. Let $\dim X = 2^n$ and $\{z_i\}_{i=1}^\infty$ be a dense sequence in X. Let $Z = \operatorname{span}\{r_i(t) z_i\}_{i=1}^\infty$ in $L_p(X)$. Define $\{\gamma_j\}_{j=1}^\infty$ as in the previous proof. Given $m \in N$, choose $\{w_i\}_{i=1}^m$ such that

$$\inf_{1 \leqq i \leqq 2^m} \|w_i\|_Z = 1, \qquad \left\|\sum_{i=1}^{2^m} w_i\right\|_Z = \frac{1}{\gamma_{2^m}}.$$

Since $\{r_i(t) z_i\}_{i=1}^\infty$ is 1-unconditional in Z, we may assume that $\|w_i\|_Z = 1$ for all $i = 1, \ldots, 2^m$, and, moreover,

$$\sup_{1 \leqq i \leqq 2^m} |c_i| \leqq \left\|\sum_{i=1}^{2^m} c_i w_i\right\|_Z \leqq \frac{1}{\gamma_{2^m}} \sup_{1 \leqq i \leqq 2^m} |c_i|,$$

for all scalar sequences $\{c_i\}_{i=1}^{2^m}$. Thus, $\operatorname{span}\{w_i\}_{i=1}^{2^m}$ is $1/\gamma_{2^m}$-isomorphic to $l_\infty^{2^m}$ which contains (in the real case) isometrically l_1^m. Thus, there are vectors $v_1, \ldots, v_m \in Z$ (e.g.,

the Rademacher elements over the $\{w_i\}_{i=1}^{2^m}$) which satisfy

$$\sum_{i=1}^{m} |c_i| \leqq \left\| \sum_{i=1}^{m} c_i v_i \right\|_Z \leqq \frac{1}{\gamma_{2^m}} \sum_{i=1}^{m} |c_i|,$$

for every choice of $\{c_i\}_{i=1}^m$. Using the fact that $a_{p,m}(Z) \leqq a_{p,m}(L_p(X)) = a_{p,m}(X)$, we get, by integrating this inequality with $c_i = r_i(t)$,

$$m \leqq \left(\int_0^1 \left\| \sum_{i=1}^{m} r_i(t) v_i \right\|_Z^p dt \right)^{1/p} \leqq a_{p,m}(X) \frac{1}{\gamma_{2^m}} m^{1/p},$$

i.e.,

$$\gamma_{2^m} \leqq a_{p,m}(X) m^{-1/p}. \tag{9}$$

Step 2. If $a_{p,n}(X) n^{-1/p'} \geqq 1/2$, $q = 2 + 2^{p'} a_{p,n}(X)^{p'} \geqq 2 + n > n$. Since $b_2(l_2^{2^n}) = 1$ and $d(X, l_2^{2^n}) \leqq 2^{n/2}$ it follows that $b_2(X) \leqq 2^{n/2}$. An interpolation argument shows

$$b_q(X) \leqq b_2(X)^{2/q} \leqq 2^{n/q} \leqq 2.$$

Step 3. We want to derive the same conclusion in the case when $a_{p,n}(X) n^{-1/p} < 1/2$. Let $\{y_i\}_{i=1}^l$ be an arbitrary finite sequence in X. We may assume that $\|y_1\| \geqq \|y_2\| \geqq \ldots$. Since the $\{z_i\}_{i=1}^{\infty}$ were dense in X,

$$\|y_j\| \leqq \gamma_j \left(\int_0^1 \left\| \sum_{i=1}^{j} r_i(t) y_i \right\|^p dt \right)^{1/p},$$

for all j. Hence, using the fact that $p \leqq 2$, we get

$$\left(\sum_{j=1}^{l} \|y_j\|^q \right)^{1/q} \leqq \left(\sum_{j=1}^{l} \gamma_j^q \right)^{1/q} \left(\int_0^1 \left\| \sum_{i=1}^{l} r_i(t) y_i \right\|^2 dt \right)^{1/2}$$

and thus, $b_q(X) \leqq \left(\sum_{j=1}^{\infty} \gamma_j^q \right)^{1/q}$.

Let k be the first integer $\leqq n$ such that

$$\gamma_{2^k} \leqq a_{p,n}(X) k^{-1/p} \leqq 1/2.$$

Then

$$\begin{aligned} \left(\sum_{j=1}^{\infty} \gamma_j^q \right)^{1/q} &\leqq 2^{k/q} \left(\sum_{i=0}^{\infty} 2^{ik} \gamma_{2^{ik}}^q \right)^{1/q} \\ &\leqq 2^{k/q} \left(\sum_{i=0}^{\infty} 2^{ik} \gamma_{2^k}^{iq} \right)^{1/q} \\ &\leqq 2^{k/q} \left(\sum_{i=0}^{\infty} 2^{-(q-k)i} \right)^{1/q} \leqq 2^{k/q} 2^{1/q} \leqq 2. \end{aligned}$$

Note here that $k - 1 \leqq 2^{p'} a_{p,n}(X)^{p'}$ and thus, $k + 1 \leqq q$, $q - k \geqq 1$. Hence, $b_q(X) \leqq 2$ in the second case too.

Corollary. *Assume X is a Banach space of type 2. Let $\varepsilon = (6a_2(X)^2)^{-1}$. Then the relative projection constant of any n-dimensional subspace X_n of X and the Banach-*

Mazur distance of X_n to l_2^n can be estimated by

$$\lambda_X(X_n) \leqq 16 a_2(X) n^{1/2-\varepsilon}$$

$$d(X_n, l_2^n) \leqq 16 a_2(X) n^{1/2-\varepsilon}.$$

Proof. It is known [6] that $\max(\lambda_X(X_n), d(X_n, l_2^n)) \leqq 4 a_2(X) \beta_2(X_n)$. Using Tomczak-Jaegerman's result and Hölder's inequality, we get that

$$\beta_2(X_n) \leqq 2\beta_{2,n}(X_n) \leqq 2 b_{2,n}(X_n) \leqq 2 n^{1/2-1/q} b_{q,n}(X_n).$$

Thus, Theorem 3 implies for $q = 6a_2(X)^2 \geqq 2 + 2^2 a_2(X)^2$ that $\beta_2(X_n) \leqq 4 n^{1/2-\varepsilon}$, since $b_{q,n}(X_n) \leqq 2$. □

Remarks. (1) The value of q in Theorem 3 may be slightly improved. Choosing $c = e^{-1/p'}$ instead of $c = 1/2$ in distinguishing the cases $a_{p,n}(X) n^{-1/p'} \gtrless c$, one gets that X is of type q for any

$$q > (p' \log 2) + (ep' \log 2) a_p(X)^{p'}.$$

(2) The existence of an $\varepsilon > 0$ in the corollary is an open problem in spaces of type p for $1 < p < 2$. Remark (1) yields a positive answer for small type p constants and values of p near 2: if

$$a_p(X) < \left(\frac{2(p-1) - (2-p)\log 2}{e(2-p)\log 2} \right)^{1/p'}$$

the corollary holds for X and $\varepsilon = \frac{1}{2} - \left(\frac{1}{p} - \frac{1}{r} \right) (> 0)$, where $r > (p' \log 2) + (ep' \log 2) a_p(X)^{p'}$. This condition makes sense only for $p > 1.56$.

References

1. Bergh, J., Löfström, J.: Interpolation spaces, an introduction. Berlin, Heidelberg, New York: Springer 1976
2. Figiel, T., Lindenstrauss, J., Milman, V.D.: The dimension of almost spherical sections of convex bodies. Acta Math. **139**, 53–94 (1977)
3. Garling, D.J.H., Gordon, Y.: Relations between some constants associated with finite dimensional Banach spaces. Israel J. Math. **9**, 346–361 (1971)
4. Haagerup, U.: Les meilleures constantes de l'inegalite de Khintchine. C.R. Acad. Sci. Paris **286**, 259–262 (1978)
5. König, H.: Type constants and $(q, 2)$-summing norms defined by n vectors. Israel J. Math. **37**, 130–138 (1980)
6. König, H., Retherford, J.R., Tomczak-Jaegermann, N.: Eigenvalues of $(p, 2)$-summing operators and constants associated with normed spaces. J. Funct. Anal. **37**, 88–126 (1980)
7. Lindenstrauss, J., Tzafriri, L.: Classical Banach spaces. II. Function spaces. Ergebnisse der Mathematik, Vol. 92. Berlin, Heidelberg, New York: Springer 1979
8. Maurey, B.: Une nouvelle caracterisation des applications (p, q)-sommantes. Seminaire Maurey-Schwartz 1973–74, Expose XII
9. Maurey, B., Pisier, G.: Series de variables aleatoires vectoriolles independantes et proprietes geometriques des espaces de Banach. Studia Math. **58**, 45–90 (1976)
10. Pisier, G.: Types des espaces normes. C. R. Acad. Sci. Paris **276**, 1673–1674 (1973)
11. Tomczak-Jaegermann, N.: Computing 2-summing norm with few vectors. Ark. Math. **17**, 273–277 (1979)

Received December 20, 1980

Math. Ann. 256, 95–119 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Cuspidal Projections of Space Curves

Ragni Piene

Department of Mathematics, University of Oslo, P.B. 1053, Blindern, Oslo 3, Norway

Table of Contents

1. Introduction 95
2. Projection of Branches 96
3. On the Existence of Cuspidal Projections 99
4. The Tangent Developable and its Singularities 110
5. The Case of Positive Characteristic 116

1. Introduction

Let $C \subset \mathbb{P}_k^3$ be a smooth, proper, algebraic curve over an algebraically closed field k. If the projection $\pi : C \to \mathbb{P}_k^2$ of C from a point $v \in \mathbb{P}_k^3 - C$ is such that $\pi : C \to \pi(C)$ is birational and such that the plane curve $\pi(C)$ has only cuspidal (unibranch) singularities, then we say that π is a *cuspidal projection* of C.

Ferrand showed that a curve that admits a cuspidal projection is a set-theoretical complete intersection, if the base field k has positive characteristic [F1]. He therefore asked: which curves admit a cuspidal projection? What we present below, grew out of an attempt to answer this question. The problem is viewed here mainly as a geometrical one. Therefore we first attack it in the characteristic 0 case; the passage to positive characteristic is deferred until the last section (Sect. 5).

Suppose a cuspidal projection $\pi : C \to \mathbb{P}^2$ exists. Then clearly the centre of projection v has to lie on the tangent developable of C, and v has to be very singular on this surface. Namely, if C has degree d and genus g, then $\pi(C)$ has $\delta = \frac{1}{2}(d-1)(d-2) - g$ cusps (counted properly), because there are no other kinds of singularities (no self-crossings). For example, v could be a δ-multiple point of the developable, or $\pi(C)$ could have cusps of higher order, arising from v lying on tangents to C at points of inflection or hyperosculation.

One expects a "general" curve $C \subset \mathbb{P}^3$ to be such that its tangent developable has no points of multiplicity greater than 3, that it has no points of inflection, and that cusps arising from projection along tangents at points of hyperosculation are not worse than double. Therefore, it is natural to believe that a general curve, with

0025-5831/81/0256/0095/$05.00

$\delta \geqq 4$, does not admit a cuspidal projection. We shall show that this holds for curves of genus 4 and degree 6 – these are canonical curves, and they have $\delta = 6$. Our results can be stated as follows:

Theorem 1. *Every curve with $\delta \leqq 3$ admits a cuspidal projection.*

Theorem 2. *A general canonical curve of genus 4 does not admit a cuspidal projection.*

Note that a canonical curve $C \subset \mathbb{P}^3$ (of genus 4) is the complete intersection of a quadric and a cubic surface. Hence Theorem 2 indicates that there is no relation in general between the property of admitting a cuspidal projection and that of being a (set-theoretical) complete intersection. If fact, the curves in Ferrand's theorem are the set-theoretical intersection of a very special surface – namely the cone of the cuspidal projection – with some other surface, and, as he remarked later [F2], this can only happen in positive characteristic.

Theorem 1 is proved by examining each type of curve satisfying $\delta \leqq 3$. Because of Castelnuovo's bound on the genus of a space curve, there are only three cases: the twisted cubic, and the elliptic and rational quartics. In the characteristic 0 case, we study the possible configurations of the tangents to these curves; in particular, we use Telling's classification of rational quartics [T]. The proof also requires some general facts about the tangent developable and its singularities, and thus links up with classical enumerative geometry [Z] – the necessary material is gathered in Sect. 4. In the course of the proof, the various types of possible cuspidal projections are described.

In order to prove Theorem 1 in positive characteristic, we use the existence of a lifting of the curve and its embedding to characteristic 0. Then we apply the characteristic 0 version of the theorem to the lifted curve, and show that the cuspidal projection "specializes" to the given curve.

Theorem 2 is proved by showing that the presence of certain phenomena, necessary for a cuspidal projection to exist, implies that the canonical curve does not have general moduli. We use the fact that these curves are complete intersections (so that we have a nice parametrization of them), and that they lie on a quadric surface. We do not here pursue the question of generalizing Theorem 2, though it is reasonable to believe that similar methods work to prove the same result for other curves that admit a good parameter space.

For Theorem 2, the characteristic 0 proof works essentially as it is also in positive characteristic.

2. Projection of Branches

Fix a base field k, algebraically closed and of characteristic 0. Let $P = \mathbb{P}_k^3$ denote projective 3-space over k – in the coordinate free way we shall also write $P = \mathbb{P}(V)$, with V a 4-dimensional vector space over k. Let $C_0 \subset P$ be a reduced, closed curve, and let $h: C \to C_0$ denote its normalization. We shall assume that C_0 spans P, i.e., C_0 is not contained in a plane.

For each point $p \in C$ we can choose (affine) coordinates around $h(p) \in P$ such that the branch of C_0 determined by p has a (formal) parametrization at $h(p)$ equal to

$$(+) \qquad \begin{aligned} x &= at^{l_1+1} + \dots \\ y &= bt^{l_2+2} + \dots \\ z &= ct^{l_3+3} + \dots \end{aligned}$$

with $abc \neq 0$ and $0 \leqq l_1 \leqq l_2 \leqq l_3$ [B, p. 184]. Even if $h(p)$ is a singular point of C_0 we shall call the line $y=z=0$ for the tangent to C_0 at p, and the plane $z=0$ for the osculating plane to C_0 at p.

We call $k_i(p) = l_{i+1} - l_i$ the ith stationary index of the branch at the point (or of p, for short); thus $k_0(p)$ is the number of cusps of C_0 at p, $k_1(p)$ the number of flexes (points of inflection), and $k_2(p)$ the number of stalls (points of hyperosculation). A point $p \in C$ with $l_1 = l_2 = l_3 = 0$ is called regular – there are only a finite number of non-regular points. The triple (l_1+1, l_2+2, l_3+3) is called the type of p. Note that $l_3 + 3 \leqq d$, where d is the degree of C. If C maps to a plane curve, the type of p with respect to this map will be a pair (m_1+1, m_2+2).

Let $\pi : P - \{v\} \to \bar{P} \cong \mathbb{P}^2$ denote the projection from a point $v \in P$ onto a plane $\bar{P}$. The rational map $\pi | C_0 : C_0 \to \bar{P}$ is defined on C_0 if $v \notin C_0$, and is in any case defined on C; by abuse of notation, we shall call this morphism π also.

Set $\bar{C} = \pi(C)$. Suppose $p \in C$ is of type (l_1+1, l_2+2, l_3+3). The point $\pi(p)$ on the corresponding branch of $\bar{C}$ will be of type (l_1+1, l_2+2), unless v is on the tangent to C_0 at p.

Suppose v is on the tangent to C_0 at p [but $v \neq h(p)$]. Then the plane curve branch will have type (l_2+2, l_3+3). This cusp will be ordinary, i.e., of type (2, 3), if and only if p was regular on C_0.

For $\bar{p} \in \bar{C}$, we set $\delta_{\bar{p}} = \text{length } \mathcal{O}_{C,\bar{p}} / \mathcal{O}_{\bar{C},\bar{p}}$; $\delta_{\bar{p}}$ is the "number of double points" of $\bar{C}$ at $\bar{p}$. We let $\delta = \sum_{\bar{p} \in \bar{C}} \delta_{\bar{p}}$ denote the "total number of double points" of $\bar{C}$; if $\bar{C}$ has degree d and geometric genus g, we have the formula $\delta = \frac{1}{2}(d-1)(d-2) - g$. If $C_0 \subset \mathbb{P}^3$ has degree d and geometric genus g, we shall call the above number δ for the *number of apparent double* points of C_0.

For $p \in C$, we let $\delta(p)$ denote the "number of double points" of the corresponding *branch* of $\bar{C}$ at $\pi(p)$. In particular, if $\bar{C}$ is unibranch at $\pi(p)$, then $\delta(p) = \delta_{\pi(p)}$ holds.

The number $\delta(p)$ of double points of $\bar{C}$ absorbed by the cusp $\pi(p)$ depends of course on the type, though it is not always determined by it. For example, if $(l_2+2, l_3+3) = 1$, then

$$\delta(p) = \tfrac{1}{2}(l_2+1)(l_3+2).$$

In particular,

$$\delta(p) = \begin{cases} 1 & \text{if } p \text{ is a regular point} \\ 3 & \text{if } p \text{ is an ordinary flex (i.e., } l_1 = 0, l_2 = l_3 = 1). \end{cases}$$

If p is an ordinary stall ($l_1=l_2=0$, $l_3=1$), the type is no longer sufficient to determine $\delta(p)$. However, we have the following:

Lemma 1. *If $p\in C$ has a parametrization*

$$x=at+\dots$$
$$y=b_2t^2+b_3t^3+\dots$$
$$z=c_4t^4+c_5t^5+\dots$$

with $ab_2c_4\neq 0$ and $b_2c_5\neq b_3c_4$, then the projection of C_0 from a point on its tangent at p gives a cusp with $\delta(p)=2$; in fact, the cusp is ramphoid of the 1st type.

Remark. Recall that a ramphoid cusp of the sth type is one which can be put in the form

$$x=t^2$$
$$y=at^4+(\text{even powers of } t)+bt^{2s+3}+(\text{higher powers of } t),$$

with $ab\neq 0$ [S-K, p. 77]. Such a cusp is equivalent to $s+1$ ordinary double points [because it is analytically equivalent to a monomial cusp of type $(2, 2s+3)$].

Proof. The cusp $\pi(p)$ is equivalent to one of the form

$$\bar{y}=y(b_2+b_3t+\dots)^{-1}=t^2$$
$$\bar{z}=z(b_2+b_3t+\dots)^{-1}=(c_4t^4+\dots)(b_2^{-1}-b_2^{-2}b_3t+b_2^{-3}(b_3^2-b_2b_4)t^2+\dots)$$
$$=b_2^{-1}c_4t^4+b_2^{-2}(b_2c_5-b_3c_4)t^5+\dots. \quad \square$$

This lemma will be used in Sect. 3, in the proof of Theorem 2. We shall also need, in the proof of Lemma 7 of Sect. 4, to know how many other tangents a given tangent to C_0 meets. Therefore we need to know what kind of singularity one obtains by intersecting the tangent developable of C_0 with a plane containing a tangent.

Set $T(p)=$ the tangent to C_0 at p, and let $X=\bigcup_{p\in C} T(p)$ denote the tangent developable of C_0. If $p\in C$ has a parametrization (+), then $h(p)\in X$ has a (formal) parametrization

$$x=at^{l_1+1}+\dots+s(a(l_1+1)t^{l_1}+\dots)$$
$$y=bt^{l_2+2}+\dots+s(b(l_2+2)t^{l_2+1}+\dots)$$
$$z=ct^{l_3+3}+\dots+s(c(l_3+3)t^{l_3+2}+\dots).$$

Since $T(p)$ is given by $y=z=0$, a plane H containing $T(p)$ has an equation $\alpha y+\beta z=0$, with $\alpha\neq 0$ if and only if H is not the osculating plane. Suppose it is not. Then $(H\cap X)_{\text{red}}=T(p)\cup D$, where the plane curve D has a singularity of type (l_1+1, l_3+3) at $h(p)$. [Note that $H\cap X$ contains $T(p)$ l_2+1 times.] Moreover, $T(p)$ is the tangent to D at $h(p)$. Thus we have proved the following:

Lemma 2. *The intersection number of D with $T(p)$ at $h(p)$ is given by*

$$i(D, T(p); h(p)) = l_3 + 3 .$$

In Sect. 4 we shall return to a global study of the tangent developable X.

3. On the Existence of Cuspidal Projections

Let $C \subset P$ be a smooth, irreducible curve of degree $r_0 = d$, genus g, and assume C is not contained in a plane. Denote by δ the number of apparent double points of C, i.e., set

$$\delta = \tfrac{1}{2}(d-1)(d-2) - g .$$

Theorem 1. *All curves $C \subset P$ with $\delta \leqq 3$, admit a cuspidal projection.*

Proof. Castelnuovo's bound on the genus of a space curve [B, p. 234; G-H, p. 252] shows that $\delta \leqq 3$ implies $d(d-2) \leqq 12$, if d is even, and $(d-1)^2 \leqq 12$, if d is odd. Therefore there are only three cases to consider:

1) $C \subset P$ is a twisted cubic, i.e., $\delta = 1$, $d = 3$, $g = 0$.
2) $C \subset P$ is an elliptic quartic, i.e., $\delta = 2$, $d = 4$, $g = 1$.
3) $C \subset P$ is a rational quartic, i.e., $\delta = 3$, $d = 4$, $g = 0$.

In each case we shall describe the possible cuspidal projections.

Case 1. The tangent developable X of C has no singularities outside its cuspidal edge C (see Sect. 4). By projecting C from any point on $X - C$, we obtain a plane cubic with one (ordinary) cusp. (Observe that, for degree reasons, any projection of C is necessarily birational onto its image.)

Case 2. Since C has degree 4, if p is a branch of type $(l_1 + 1, l_2 + 2, l_3 + 3)$, we always have $l_3 + 3 \leqq 4$, hence $l_3 \leqq 1$. Therefore $l_2 \leqq 1$; in fact, $l_2 = 0$ for all p (i.e., C has no flexes): Let $p \in C$ be a point of type $(1, l_2 + 2, l_3 + 3)$. Consider the pencil $\{H_\lambda\}$ on C cut out by planes containing the tangent $T(p)$. Its base point divisor is equal to $(l_2 + 2)p$. By the Riemann-Roch theorem, it follows that

$$\begin{aligned} 2 \leqq h^0(H - (l_2+2)p) &= 4 - (l_2+2) + 1 - g + h^1(H - (l_2+2)p) \\ &= 2 - l_2 + h^1(H - (l_2+2)p) . \end{aligned}$$

Since l_2 is 0 or 1, $H - (l_2 + 2)p$ is non-special, hence we get $2 \leqq 2 - l_2$, and thus $l_2 = 0$.

Since $l_3 \leqq 1$, the stalls of C are all ordinary [and their number is then $k_2 = 16$ – see Sect. 4, (1)].

We shall show that C admits two kinds of cuspidal projections – one gives a plane curve with one ramphoid cusp, the other a plane curve with two ordinary cusps.

Suppose $v \in P - C$ is a point such that the projection from v, $\pi : C \to \pi(C)$, is *not* birational. Then necessarily $\deg \pi = 2$ and $\deg \pi(C) = 2$, so C is on a quadric cone with vertex v. Call this cone K. Now we know that C is the base locus of a pencil of quadrics. If C is on at least one smooth quadric, it will be on no more than 4 quadric cones, and K must be one of these. If C is not on a smooth quadric, there is a pencil of quadric cones containing C; we shall see below that this is impossible.

By the Riemann-Hurwitz formula, $\pi : C \to \pi(C)$ has 4 branch points, so K contains the tangents to C at 4 points. Let p be one of these. The tangent plane H

to K along $T(p)$ intersects C only at p, hence with intersection number 4. Hence H is a hyperosculating plane to C, so p is a stall. Each of the 16 stall tangents thus intersects exactly 3 others, in the same point, and these points of intersection are the vertices of four quadric cones containing C. It follows that these four cones are the only quadric cones that contain C.

Let $p \in C$ be a stall, and $v \in T(p)$ any point different from p and different from the vertex of the (unique) quadric cone containing C and $T(p)$. Then the projection π of C from v is birational onto its image, and $\pi(C)$ is a plane, elliptic quartic with one ramphoid cusp (necessarily of the 1st type).

Consider now the *nodal curve* of C (the "double curve" of the tangent developable X, see Sect. 4). Note that it contains no bitangents, since C cannot have any. Assume first that the nodal curve is double (of multiplicity 2) on X. Then the projection of C from any point on it, not on C and different from the 4 vertices of the cones, is birational onto its image – the argument above shows that otherwise all tangents are stall tangents – and the projected curve will have 2 ordinary cusps. That the nodal curve is non empty, follows from the fact that it has degree (see Sect. 4)

$$b = \tfrac{1}{2}(r_1(r_1-1) - r_2 - 3r_0) = \tfrac{1}{2}(8 \cdot 7 - 12 - 3 \cdot 4) = 16.$$

Suppose the nodal curve had a component of multiplicity $\geqq 3$ on X. Then the projection of C from any point of that component could not be birational onto its image (since a plane, elliptic curve cannot have a point of multiplicity $\geqq 3$), hence all the points of the component would be vertices of quadric cones containing C. As we have seen, this is impossible.

(Thus we have shown that the nodal curve is double; moreover, the four vertices of the cones are quadruple points on the nodal curve – this checks with the fact that this curve has a double point cycle of degree $3T = 48 = 4 \cdot 4 \cdot 3$, see Sect. 4.)

Case 3. These curves have been classified by Telling [T]; she considered the various ways of projecting the rational normal quartic in $\mathbb{P}^4$ to $\mathbb{P}^3$. We shall distinguish between two cases: a) the general rational quartic, b) the equianharmonic rational quartic.

Observe first that we need not worry about birationality: Let $v \in P - C$ be a point, $\pi : C \to \pi(C)$ the projection from it. If π is not birational, then necessarily $\deg \pi = 2$, so C lies on a quadric cone – this is impossible since C is rational. (If C were on a quadric cone, it would possess a linear system of degree 2 and dimension 1 – by Riemann-Roch, this is impossible since C has genus 0. See also [B, p. 212; T, p. 41].)

In Case a), $C \subset P$ is a generic (or almost so) projection of the rational normal quartic in $\mathbb{P}^4$. It has only the kind of singularities that is predicted by dimension count – in particular the nodal curve is double. The tangent developable has T points of multiplicity 3, given by (Sect. 4)

$$T = \tfrac{1}{6}(r_1 - 4)((r_1 - 3)(r_1 - 2) - 6g) = 4.$$

Among these triple points are the $d(1,2) = 4$ points where a tangent meets the curve again – hence there are *no* points where 3 distinct tangents intersect [T, p. 55]. It follows that such a C can not be projected onto a tricuspidal quartic.

For the number k_1 of flexes and k_2 of stalls, there are three possibilities (all flexes and stalls are ordinary, since $d=4$).

Case a_1. $k_1=0$, $k_2=4$. (This is the most general C.) The only possible type of cuspidal projection is obtained by projecting C from the (unique, see Sect. 4) point of intersection of a stall tangent with another tangent. This point is not on the curve: It is different from the stall [T, pp. 46, 47]. If it were another point on the curve, by projecting from it we would obtain a plane rational cubic with a ramphoid (double) cusp – this is impossible.

Thus C admits a projection onto a plane rational quartic with one ordinary cusp and one ramphoid cusp (necessarily of the 1st type).

Case a_2: $k_1=1$, $k_2=2$. This curve admits the same type of cuspidal projection as the one above, but also an additional one: Projecting from a point on the flex tangent, one obtains a plane quartic with one cusp, of type (3, 4).

Case a_3. $k_1=2$, $k_2=0$. The projection from a point of one of the two flex tangents is the only type of possible cuspidal projection.

In Case b), $C \subset P$ is the projection of the rational normal quartic from a general point on a certain quadric hypersurface in $\mathbb{P}^4$ (this quadric is the nucleus of the fundamental polarity, see [T, pp. 8, 65]). In this case, the nodal curve is triple, there are no flexes, and hence 4 stalls (the nodal curve is a plane conic through these 4 points). The reason for the name of this curve, is that the 4 stalls form an equianharmonic set on the curve (which means their cross-ratios are equianharmonic).

Also, the $d(1,2)=4$ (Sect. 4) points where a tangent meets the curve again (the so-called Steinerian points of C), are just the stalls [T, p. 66]. The only way of obtaining a cuspidal projection of the equianharmonic quartic, is to project it from a point on its nodal curve, not on C. The projected curve is a tricuspidal quartic (and all tricuspidal quartics are obtainable in this way). □

Curves with $\delta \geqq 4$

It seems natural to believe that a general curve of degree d and genus g, with $\delta = \frac{1}{2}(d-1)(d-2)-g \geqq 4$, does *not* admit a cuspidal projection. In fact, this would follow if we could prove the following:

1) A general curve $C \subset P$ of degree d and genus g has only such singularities that are predicted by a dimension count (in particular, C has no flexes and only ordinary stalls, the nodal curve is double, and there are no points of multiplicity greater than 3 on the tangent developable).

2) For a general $C \subset P$, the projection from a point on a stall tangent gives a ramphoid cusp of the 1st type.

If 1) and 2) hold, then, for a general $C \subset P$, one can obtain only (the equivalent of) 3 or fewer cusps by projection – the types of projections are the ones described in the proof of Theorem 1.

If we want 1) to be true (without being tautological), we must of course be careful about how to define "general". We have seen that for $d=4$ and $g=0$, 1) is true when "general" means "most curves of degree 4 and genus 0". With a similar

definition, 1) fails for $d=4$ and $g=1$ (since these curves all have quadruple points on their tangent developable). So though there is a very natural definition of "general" in this case – namely "the intersection of two general quadrics" – it is not one that makes 1) hold. Moreover, these general curves also have general moduli – hence it would not help to impose that condition. All one can expect is therefore that an ad hoc version of 1) – sufficient for our purposes – would always be true. We shall prove this for the case $d=6$ and $g=4$. These are the canonical curves of genus 4, and they are the complete intersection of a quadric and a cubic surface.

Suppose a curve C lies on a quadric surface Q, and that C is of type (a, b) on Q. If a or b is greater than 3, then C has an infinity of quadri-secants – a phenomenon which is *not* "predicted by a dimension count". But then, curves that lie on a quadric are not usually (for big a or b) considered to be "general". It turns out, however, that the property in 2) is easier to verify for curves that are the (complete) intersection of a quadric with another surface. We shall prove 2) for the case $d=6$ and $g=4$, i.e., for canonical curves of genus 4. Note that a curve C which is the intersection of a general quadric and a general cubic surface, has general moduli (i.e., C is general as a curve of genus 4).

Remark. Cuspidal projections $\pi: C \to \mathbb{P}^2$ such that $\pi(C)$ has only *ordinary* cusps, are (intuitively) harder to produce than arbitrary ones. In fact, in many cases one can ascertain that such projections do not exist, as a consequence of existing bounds on the number of cusps that plane curves (with only ordinary cusps and nodes as singularities) can have. For a discussion of these bounds, we refer to [Zar, p. 221] [note a misprint: the inequality (18) should be strict]; here we give only two examples.

(a) Rational Curves
Suppose $\pi: C \to \mathbb{P}^2$ is a cuspidal projection of a rational curve and that $\pi(C)$ has only ordinary cusps. From the Plücker formulas it follows [Zar, loc. cit.] that the degree d of C must satisfy

$$1/2(d-1)(d-2) \leqq 3/2(d-2),$$

hence $d \leqq 4$. In other words, rational curves of degree $\geqq 5$ do not admit cuspidal projections such that all cusps are ordinary.

(b) Canonical Curves of Genus 5
Suppose $C \subset \mathbb{P}^3$ is the (isomorphic) projection of a canonical curve ($g=5$, $d=8$) in $\mathbb{P}^4$. If there was a cuspidal projection π of C such that $\pi(C)$ had only ordinary cusps, we would obtain a plane curve of degree 8 with 16 cusps – in contradiction with the fact that a plane curve of degree $d=8$ has at most $1/2(d-1)(d-4)+1=15$ cusps [Zar, loc. cit.].

Let $C \subset P$ be a canonical curve of genus 4, so that $\mathcal{O}_C(1)=\Omega_C^1$, the degree of C is $d=2g-2=6$, and C has $\delta=\frac{1}{2}(d-1)(d-2)-g=6$ apparent double points. Moreover, C is the intersection of a quadric and a cubic surface, and if these surfaces are general, the curve C will have general moduli.

Theorem 2. *A general canonical curve $C \subset P$ of genus 4 does not admit a cuspidal projection.*

Proof. First of all, if C is general, its only non-regular points are ordinary stalls: The non-regular points of a canonical curve are its Weierstrass points [B, p. 201; G-H, p. 274], and if C is general, these are normal [G-H, p. 277]. If $p \in C$ is a Weierstrass point, of type $(1, l_2+2, l_3+3)$, then the gap sequence at p is $(1, 2, l_2+3, l_3+4)$, so p is normal iff $l_2=0$ and $l_3=1$.

Now suppose we have established the following three lemmas.

Lemma 3. *Let $C \subset P$ be a general canonical curve, i.e., $C = Q \cap F$ is the intersection of a general quadric Q and a general cubic surface F in P. Let $p \in C$ be a stall and $v \in P - C$ a point on the tangent to C at p. If $\pi : C \to \mathbb{P}^2$ denotes the projection of C from v, then $\pi(p) \in \pi(C) \subset \mathbb{P}^2$ is a ramphoid cusp of the 1st type.*

Lemma 4. *Suppose C is the normalization of*

(i) *a plane sextic curve of genus 4, with 6 ordinary cusps, or*

(ii) *a plane sextic curve of genus 4, with 4 ordinary cusps and 1 ramphoid cusp.*

Then C does not have general moduli.

Lemma 5. *If $C \subset P$ is a general canonical curve of genus 4, then no two stall tangents intersect.*

Then, from Lemma 3, it follows that if $C \subset P$ is general, the only types of cusps one can obtain by projection, are ordinary cusps and ramphoid cusps of the 1st type. Therefore, if $\pi : C \to \mathbb{P}^2$ is a cuspidal projection of C, there are only 4 possibilities: $\pi(C)$ has 6 ordinary cusps, or 4 ordinary and 1 ramphoid cusp, or 2 ordinary and 2 ramphoid cusps, or 3 ramphoid cusps. Lemma 4 rules out the first two possibilities, and Lemma 5 the last two.

Proof of Lemma 3. Since all smooth quadrics are projectively equivalent, we shall fix one; call it Q. Then the curves C are parametrized by cubic surfaces: Let $A = k[F_{ijkl}]$ denote the ring of coefficients of cubic, homogeneous polynomials in 4 variables, and set $\mathscr{F} = \sum F_{ijkl} X_0^i X_1^j X_2^k X_3^l$, the universal cubic. We have a (complete intersection) map

$$\Phi : \mathrm{Proj}(A[X_0, X_1, X_2, X_3]/(Q, \mathscr{F})) \to \mathrm{Proj}(A).$$

Let $U \subset \mathrm{Proj}(A)$ be the open subscheme such that, if $\mathscr{C} = \Phi^{-1}(U)$, $\Phi : \mathscr{C} \to U$ is smooth, of relative dimension 1. Thus we have a family of canonical curves of genus 4, containing the general ones.

The Weierstrass points on the fibres of Φ form an effective, relative divisor W on $\mathscr{C}$ over U. It can be defined as follows: There is a natural homomorphism [P1, Sect. 6]

$$a^{g-1} : \Phi^* \Phi_* \Omega^1_{\mathscr{C}/U} \to \mathscr{P}^{g-1}_{\mathscr{C}/U}(\Omega^1_{\mathscr{C}/U}).$$

Since a^{g-1} is a map between locally free sheaves of rank g, we can take its determinant

$$\det a^{g-1} : \Phi^* \mathscr{M} \to \Lambda^g \mathscr{P}^{g-1}_{\mathscr{C}/U}(\Omega^1_{\mathscr{C}/U}) \cong (\Omega^1_{\mathscr{C}/U})^{\otimes g(g+1)},$$

where we have put $\mathscr{M} = \Lambda^g \Phi_* \Omega^1_{\mathscr{C}/U}$. The corresponding section

$$w_{\mathscr{C}/U} : \mathscr{O}_{\mathscr{C}} \to (\Omega^1_{\mathscr{C}/U})^{\otimes g(g+1)} \otimes \Phi^* \mathscr{M}^{-1}$$

defines the (relative) Weierstrass divisor

$$W = \operatorname{div}(w_{\mathscr{C}/U}) \in \operatorname{Div}^+(\mathscr{C}/U).$$

Replacing U by a smaller open, we may assume that all Weierstrass points of the fibres $\mathscr{C}_u = \Phi^{-1}(u)$ are normal.

For $p \in \mathscr{C}$, set $u = \Phi(p)$; then $p \in \mathscr{C}_u = Q \cap \mathscr{F}_u \subset \mathbb{P}^3$. Set $p_0 = (1,0,0,0)$ and Q_0 the quadric defined by $X_0 X_3 = X_1 X_2$. Define a scheme $\tilde{\mathscr{C}}$ by

$$\tilde{\mathscr{C}} = \{(p, \alpha) \in \mathscr{C} \times P\mathrm{GL}(3) \mid \alpha(p) = p_0, \alpha(Q) = Q_0\}.$$

Let $\varphi : \tilde{\mathscr{C}} \to \mathscr{C}$ denote the projection. Then φ is smooth, of relative dimension 4: The fibres of φ are

$$\varphi^{-1}(p) = \{\beta \in P\mathrm{GL}(3) \mid \beta(p_0) = p_0, \beta(Q_0) = Q_0\}.$$

Consider

$$G = \{\beta \in P\mathrm{GL}(3) \mid \beta(p_0) = p_0\},$$

$$S = \{\text{quadrics through } p_0\}.$$

There is a surjective map $\gamma : G \to S$, given by $\gamma(\beta) = \beta(Q_0)$, and $\gamma^{-1}(Q_0) = \varphi^{-1}(p)$. Now the fibres of γ have dimension 4, since $\dim G = 15 - 3 = 12$, and $\dim S = 9 - 1 = 8$.

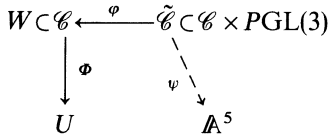


Fig. 1

We shall now define a rational map $\tilde{\mathscr{C}} - - \to \mathbb{A}^5$ (see Fig. 1). Suppose $(p, \alpha) \in \tilde{\mathscr{C}}$. Then $u = \Phi(p)$ corresponds to a homogeneous cubic polynomial

$$F = \sum F_{ijkl} X_0^i X_1^j X_2^k X_3^l,$$

where the F_{ijkl}'s are defined up to multiplication by the same non-zero scalar. Note that $F_{3000} = 0$, since $p \in \mathscr{C}_u = Q \cap F$. In affine coordinates (x, y, z), with $X_0 \neq 0$, the equation of Q becomes $z = xy$, and substituting this in the affine equation of F, we get a polynomial in two variables

$$h(x, y) = \sum h_{ij} x^i y^j,$$

which, together with $z = xy$, determines $\mathscr{C}_u$.

Assume now that $h_{01} \neq 0$ (i.e., $F_{2010} \neq 0$). Then we normalize the h_{ij}'s so that $h_{01} = -1$. Expand y in terms of x:

$$y = a_1 x + a_2 x^2 + \ldots + a_5 x^5 + \ldots,$$

where

$$a_1 = h_{10}$$

$$a_2 = a_1^2 h_{02} + a_1 h_{11} + h_{20}$$

$$a_3 = a_2 h_{11} + 2a_1 a_2 h_{02} + h_{30} + a_1 h_{21} + a_1^2 h_{12} + a_1^3 h_{03}$$

$$a_4 = a_1^2 h_{13} + a_1^2 h_{22} + a_1 h_{31} + 3a_1 a_2 h_{03} + 2a_1 a_2 h_{12} + a_2 h_{21} + a_2^2 h_{02} + a_3 h_{11}$$

$$a_5 = a_4 h_{11} + 2a_1 a_4 h_{02} + 2a_2 a_3 h_{02} + a_3 h_{21} + a_2^2 h_{12} + 2a_1 a_3 h_{12} + 3a_1^2 a_3 h_{03} + a_2 h_{31} + 2a_1 a_2 h_{22} + 3a_1^2 a_2 h_{13} + a_1^2 h_{32} + a_1^3 h_{23}$$

etc.

Thus $p \in \mathscr{C}_u \subset P$ has a parametrization

$$(*) \qquad \begin{aligned} x &= x \\ y &= a_1 x + a_2 x^2 + \dots \\ z &= xy = a_1 x^2 + a_2 x^3 + \dots . \end{aligned}$$

Thus we have a map $\psi : \tilde{\mathscr{C}} - - \to \mathbb{A}^5$, defined for all points (p, α) such that $h_{01} \neq 0$, by $\psi(p, \alpha) = (a_1, a_2, a_3, a_4, a_5)$.

Suppose (p, α) is such that $h_{01} = 0$. Then, since p is a smooth point of $\mathscr{C}_u$, necessarily $h_{10} \neq 0$. We may then assume $h_{10} = -1$ to get a parametrization

$$\begin{aligned} y &= y \\ x &= h_{02} y^2 + (h_{02} + h_{03}) y^3 + \dots \\ z &= h_{02} y^3 + (h_{02} + h_{03}) y^4 + \dots . \end{aligned}$$

Since, by assumption, $\mathscr{C}_u$ has no flexes, we must have $h_{02} \neq 0$, so p is a regular point of $\mathscr{C}_u$. Hence $\psi(p, \alpha)$ is defined whenever $p \in W$, so we get a morphism

$$\psi : \varphi^{-1}(W) \to \mathbb{A}^5 .$$

Suppose ψ is defined at (p, α). The parametrization $(*)$ is equivalent to

$$\begin{aligned} x &= x \\ z &= a_1 x^2 + a_2 x^3 + \dots \\ y &= (a_3 - a_1^{-1} a_2^2) x^3 + (a_4 - a_1^{-1} a_2 a_3) x^4 + (a_5 - a_1^{-1} a_2 a_4) x^5 + \dots . \end{aligned}$$

If $p \in W$, then $a_3 = a_1^{-1} a_2^2$ holds, and – since any stall is ordinary – $a_4 \neq a_1^{-1} a_2 a_3$. According to Lemma 1, the ramphoid cusp obtained by projecting the curve from a point on the stall tangent $T(p)$, will be of the 1st type if

$$a_1 (a_5 - a_1^{-1} a_2 a_4) \neq a_2 (a_4 - a_1^{-1} a_2 a_3)$$

holds.

Set

$$V = \{ a \in \mathbb{A}^5 \mid a_1 \neq 0, a_1 a_4 \neq a_2 a_3, a_2^2 = a_1 a_3 \} .$$

Then $\psi(\varphi^{-1}W)\subset V$. Moreover, we claim that $\psi:\varphi^{-1}W\to V$ is generically surjective: By construction, $\psi:\tilde{\mathscr{C}}--\to\mathbb{A}^5$ factors through the map

$$\tilde{\varphi}:\tilde{\mathscr{C}}\to\tilde{U}=\{u\in U\,|\,p_0\in\mathscr{F}_u\},$$

defined by $\tilde{\varphi}(p,\alpha)=\alpha(\mathscr{F}_{\Phi(p)})$.

Both this map and $U--\to\mathbb{A}^5$ are generically surjective. Since $\psi^{-1}(\psi(\tilde{\mathscr{C}})\cap V)=\varphi^{-1}W$, the claim follows. (If $\varphi^{-1}W$ has more than one irreducible component, ψ is generically surjective on each of them, because of the homogeneous nature of the map ψ.)

Set

$$\begin{aligned}V_1&=\{a\in V\,|\,a_1(a_1a_5-a_2a_4)=a_2(a_1a_4-a_2a_3)\}\\&=\{a\in V\,|\,a_1a_5-a_2a_4=a_2a_4-a_3^2\}.\end{aligned}$$

Because of the independence of the defining equations,

$$\dim V_1=\dim V-1,$$

hence

$$\dim\psi^{-1}(V_1)<\dim\varphi^{-1}W$$

holds.

Since the property $(p,\alpha)\in\psi^{-1}(V_1)$ of the point (p,α) is independent of α, the above inequality implies

$$\dim\varphi(\psi^{-1}(V_1))<\dim W.$$

Set $U_1=\Phi(\varphi(\psi^{-1}V_1))$, and let U_1 denote its closure. The map $\Phi|W:W\to U$ is finite and onto, therefore

$$\dim U_1<\dim U.$$

Hence we have found $U_0=U-U_1$, open and non-empty, with the property that any curve $\mathscr{C}_u$, $u\in U_0$, is such that its stalls satisfy the condition of the lemma. □

Proof of Lemma 4. Suppose $\bar{C}\subset\mathbb{P}^2$ is a plane, irreducible curve of degree d and genus g. Let $N\cong\mathscr{O}_{\bar{C}}(d)$ denote its normal bundle,

$$J=F^1(\Omega^1_{\bar{C}})\subset\mathscr{O}_{\bar{C}}$$

its jacobian ideal, and $\pi:C\to\bar{C}$ its normalization.

It is known [Zar, VIII, Sect. 5; A, for ordinary cusps and nodes] that the space of locally trivial deformations of $\bar{C}$ has dimension

$$r=\dim H^0(\bar{C},J\otimes N).$$

Now we have $J\subset\pi_*(J\mathscr{O}_C)$, hence $r\leqq r'$, where

$$r'=\dim H^0(\bar{C},\pi_*J\mathscr{O}_C\otimes N)=\dim H^0(C,J\mathscr{O}_C\otimes\mathscr{O}_C(d)).$$

Suppose $J\mathscr{O}_C\otimes\mathscr{O}_C(d)$ is non-special. Then, by Riemann-Roch,

$$r'=\deg(J\mathscr{O}_C\otimes\mathscr{O}_C(d))+1-g=d^2-\deg(J\mathscr{O}_C^{-1})+1-g.$$

This gives [P2, 3.9]

$$r' = 3d + 2g - 2 - \deg(I^{-1}) + 1 - g = 3d + g - 1 - \deg(I^{-1}),$$

where $I = F^0(\Omega^1_{C/\bar{C}})$ denotes the ramification ideal of π.

In particular, if $\deg(I^{-1}) < 3d$ holds, then $\deg(J\mathcal{O}_C \otimes \mathcal{O}_C(d)) > 2g-2$ and $H^1(C, J\mathcal{O}_C \otimes \mathcal{O}_C(d)) = 0$.

Apply the above to $d=6$ and $g=4$. In Case i) local computations show $J = \pi_* J\mathcal{O}_C$, hence $r = r'$ and $\deg(I^{-1}) = \#\text{cusps} = 6 < 3 \cdot 6 = 18$. Hence we get $r = 15$. In Case ii) we have $\deg(I^{-1}) = 5$, hence $J\mathcal{O}_C \otimes \mathcal{O}_C(6)$ is still non special, and we get $r \leqq r' = 16$.

Modulo projective transformations, the dimension of the family of locally trivial deformations is equal to $r-8$. Hence plane sextics of genus 4 with 6 ordinary cusps form a family of dimension 7; those with 4 ordinary and 1 ramphoid cusp (if such exist) form a family of dimension $\leqq 8$. Since the moduli space of genus 4 curves has dimension $3g-3=9$, a curve C which is the normalization of either of the above plane curves, cannot have general moduli. □

Remark. For plane sextics with the other two configurations of cusps – 2 ordinary and 2 ramphoid, or 3 ramphoid – we get families of dimensions 9 and 10, respectively. Hence the above method gives no conclusion in those cases.

Proof of Lemma 5. Suppose Q is a smooth quadric, and $C = Q \cap F$ a canonical curve. We observe that if p and q are distinct points of C lying on the same ruling L of Q, then the tangents $T(p)$ and $T(q)$ of C do not intersect. For, suppose they did. Since the tangent planes to Q at p and q intersect in L, this implies $T(p) = T(q) = L$. Hence L is a bitangent to C and has intersection multiplicity at least 4 with C, contrary to the fact that L intersects the cubic surface F, hence also C, in (the equivalent of) three points.

Let $p, q \in C$ be two points not on the same ruling of Q. After a projective transformation, we may assume

$$Q : X_0 X_3 - X_1 X_2 = 0$$
$$p = (1,0,0,0)$$
$$q = (0,0,0,1).$$

The choice of coordinates gives us the coefficients (F_{ijkl}) of a cubic polynomial F (determined modulo Q and up to scalar multiplication). Since $p, q \in C = Q \cap F$, we have $F_{3000} = F_{0003} = 0$. The tangent planes to Q and to F at p and q are given by

$$T_Q(p) : X_3 = 0 \qquad T_F(p) : F_{2100} X_1 + F_{2010} X_2 + F_{2001} X_3 = 0$$
$$T_Q(q) : X_0 = 0 \qquad T_F(q) : F_{1002} X_0 + F_{0102} X_1 + F_{0012} X_2 = 0.$$

The tangents to C at p and q are

$$T(p) : X_3 = 0, F_{2100} X_1 + F_{2010} X_2 = 0$$
$$T(q) : X_0 = 0, F_{0102} X_1 + F_{0012} X_2 = 0.$$

The two tangents intersect if and only if

$$F_{2100}F_{0012}=F_{0102}F_{2010}$$

holds.

Assume $F_{2010}\neq 0$. The computations made in the proof of Lemma 3 show that the condition that p be a stall, is that the following equality holds:

$$\begin{aligned}&F_{2010}^4F_{1200}^2-F_{2010}^3F_{2100}F_{1110}F_{1200}-F_{2010}^3F_{2100}F_{2001}F_{1200}\\&-F_{2010}^4F_{2100}F_{0300}+F_{2010}^3F_{2100}^2F_{0210}+F_{2010}^3F_{2100}^2F_{1101}\\&-F_{2010}^2F_{2100}^3F_{0120}-F_{2010}^2F_{2100}^3F_{1011}+F_{2010}F_{2100}^4F_{0030}\\&+F_{2010}F_{2100}^3F_{1020}F_{1110}+F_{2010}F_{2100}^3F_{1020}F_{2001}-F_{2100}^4F_{1020}=0.\end{aligned}$$

(This is the equation $a_2^2-a_1a_3=0$, where $a_1=h_{10}=-F_{2010}^{-1}F_{2100}, a_2=\ldots.$)

Note that this equation is invariant under the change $F_{ijkl}\rightsquigarrow F_{ikjl}$. Since at least one of F_{2010} and F_{2100} is non-zero, our assumption $F_{2010}\neq 0$ does not mean any loss of generality.

If we change F_{ijkl} to F_{ljki} in the above equation, we obtain the condition that the point q be a stall. These conditions are seen, by inspection, to be independent, i.e., the fact that $T(p)$ and $T(q)$ intersect and p is a stall does not imply that q is a stall; or, if p and q are stalls, then by moving F (and keeping the stalls) we get a curve such that $T(p)$ and $T(q)$ don't intersect. Thus we may assume this to be true for any pair of stalls on a general canonical curve. In fact, similarly to the proof of Lemma 3, one can define

$$\begin{aligned}\mathscr{D}=\{(p,q,\alpha)\in\mathscr{C}\times_U\mathscr{C}\times \mathrm{PGL}(3)\,|\,&p,q \text{ stalls}, \alpha(p)=p_0,\\&\alpha(q)=q_0=(0,0,0,1), \alpha(Q)=Q_0\}\end{aligned}$$

and consider the map $\mathscr{D}\to\{u\in U\,|\,p_0,q_0\in\mathscr{F}_u\}$ and argue as in that proof. □

Remark. By performing the same computations for a curve which is the intersection of two quadrics (an elliptic quartic curve), we find that, if $T(p)$ and $T(q)$ intersect and p is a stall, then necessarily q is a stall too. This is just as expected, since we have already seen [in the proof of Theorem 1, Case 2)] that for such curves, any stall tangent intersects three other stall tangents but no other tangents.

Examples of Cuspidal Projections

Theorem 2 says that general canonical curves of genus 4 do not admit a cuspidal projection. There exist, however, canonical curves that do. Recall the following:

Let $S\subset P$ be a (general) smooth surface of degree d, and suppose $\pi: S\to\mathbb{P}^2$ is the projection of S from a general point v. The *curve of contact* $C\subset S$ of S with respect to π can be defined as the ramification divisor $\Sigma^1(\pi)$ of π. Another description of C is that it is the intersection of S with the 1st polar S_1 of S with respect to v. The curve C has degree $d(d-1)$ and genus $g=\frac{1}{2}d(d-1)(2d-5)+1$; moreover, the number of cusps of $\pi(C)$ (these are ordinary, since π is a generic projection) is equal to the degree $d(d-1)(d-2)$ of the ramification divisor $\Sigma^{1,1}(\pi)=\Sigma^1(\pi|C)$ of $\pi|C: C\to\pi(C)$ [P3, Sect. 5].

Apply this to $d=3$. Then $S:F=0$ is a cubic surface, its 1st polar $S_1:Q=\sum v_i\frac{\partial F}{\partial X_i}=0$ is a quadric, and the intersection of the cubic and quadric (not general as such) is a canonical curve C. The projected curve $\pi(C)$ has 6 ordinary cusps, hence no other singularities, and so $\pi: C\to\mathbb{P}^2$ is a cuspidal projection of C.

It is known that a smooth curve $D=S_1\cap S_2$, where S_i is a surface of degree d_i, is such that all its chords through a general point $v\in P-D$ lie on a cone of degree $(d_1-1)(d_2-1)$ (Valentiner, Noether – see [B, p. 204]). This is seen as follows: The projection from v, $\pi: D\to\mathbb{P}^2$, is birational onto its image $\bar{D}=\pi(D)$. The conductor ideal $\mathbf{C}=\mathscr{H}om_{\bar{D}}(\pi_*\mathcal{O}_D,\mathcal{O}_{\bar{D}})$ of D in $\bar{D}$ satisfies

$$\pi_*\Omega^1_D=\omega_{\bar{D}}\otimes\mathbf{C},$$

where $\omega_{\bar{D}}=\mathcal{O}_{\bar{D}}(d_1d_2-3)$ denotes the dualizing sheaf of $\bar{D}$. Since $\Omega^1_D=\mathcal{O}_D(d_1+d_2-4)$ holds, one sees that

$$H^0(\bar{D},\mathbf{C}\otimes\mathcal{O}_{\bar{D}}(m))\neq 0$$

if and only if $m\geqq(d_1-1)(d_2-1)$. Therefore, the singularities of the plane curve $\bar{D}$ (as defined by the conductor) lie on a curve of degree $(d_1-1)(d_2-1)$.

The following converse is also true (Halphen – see [B, p. 204; G-P, p. 32]): Suppose a curve $D\subset P$ has degree d_1d_2 and does not lie on a surface of degree $<\min(d_1,d_2)$. If the chords to D through a (general) point v lie on a cone of degree $(d_1-1)(d_2-1)$, then D is the complete intersection of two surfaces of degrees d_1 and d_2.

Therefore, a sextic curve $C\subset\mathbb{P}^2$ of genus 4, with 6 cusps, is the projection of a canonical curve $C\subset P$ if and only if the cusps lie on a conic. As Zariski observes [Zar, p. 223], there are 6-cuspidal sextics without this property – they also form a 15-dimensional family.

Remark. It follows from a theorem of Segre [S] that any 6-cuspidal sextic with all its cusps on a conic, can be obtained as the projection of the intersection of a cubic surface with its 1st polar.

Another question one can ask, is the following: given d and g such that there exist space curves of degree d and genus g, does there exist one that admits a cuspidal projection?

Here are some cases where the answer is known to be yes:

1. $g=0$, all d

The rational curves $\mathbb{P}^1_k\xrightarrow{\sim}C\subset\mathbb{P}^3_k$, given by $(u,v)\rightsquigarrow(u^d,u^{d-1}v,uv^{d-1},v^d)$, were shown by Hartshorne (1964, [H]) to be set-theoretical complete intersections if $\operatorname{char}k>0$. Ferrand [F1] observed that they admit a cuspidal projection: the projection of C from a point on one of the two inflectionary tangents gives a plane curve with a monomial cusp of type $(d-1,d)$; hence that plane curve can have no other singularities. (Ferrand went on to prove that any curve which admits a cuspidal projection is a set-theoretical complete intersection in positive characteristic [F1, 2.3].)

2. $g=1$, $d=4$

All curves admit a cuspidal projection (Theorem 1).

3. $g=4$, $d=6$

Examples were given above.

4. $g=1$ or 2, $d=5$ or 6

Suppose $\bar{C}\subset\mathbb{P}^2$ is a curve of degree d and genus g, with $\kappa=\frac{1}{2}(d-1)(d-2)-g$ ordinary cusps (hence no other singularities), and let $\pi: C\to\bar{C}$ denote its normalization. If $d>2g$, then $\mathcal{O}_C(1)=\pi^*\mathcal{O}_{\mathbb{P}^2}(1)|C$ is very ample, hence it embeds C in $\mathbb{P}^N$, where $N=d-g$. By factoring the given projection $\mathbb{P}^N\dashrightarrow\mathbb{P}^2$ generically via a $\mathbb{P}^3$, we obtain C embedded in $\mathbb{P}^3$ and such that the projection $\pi: C\to\bar{C}\subset\mathbb{P}^2$ is cuspidal (see the Remark below).

Since there exist plane curves (because $d\leqq 6$, see [Zar, p. 222]) with

$$\begin{aligned} g=1 \quad & d=5 \quad \kappa=5 \\ & d=6 \quad \kappa=9 \\ g=2 \quad & d=5 \quad \kappa=4 \\ & d=6 \quad \kappa=8, \end{aligned}$$

the above method applies to these cases.

[There also exist plane curves of genus 3 and degree 5 (resp. 6), with 3 (resp. 7) cusps, but here $d>2g$ no longer holds.]

Remark. That a cuspidal projection $\pi: C\to\mathbb{P}^2$ of a smooth curve $C\subset\mathbb{P}^N$ factors via an embedding $C\subset\mathbb{P}^3$, can be seen as follows.

Let $L\cong\mathbb{P}^{N-3}\subset\mathbb{P}^N$ be the centre of the projection π. By assumption, $L\cap C=\emptyset$, so in particular L contains no tangent to C. If $Y=\mathbb{P}(\mathscr{P}^1_C(1))\to X\subset\mathbb{P}^N$ denotes the tangent developable of C (see Sect. 4), it follows that $X\cap L$ is finite. Hence a general $L'\cong\mathbb{P}^{N-4}\subset L$ satisfies $L'\cap X=\emptyset$. Take π' to be the projection of C from such a L'. Since $\pi(C)$ has only cuspidal singularities and $\pi': C\to\mathbb{P}^3$ is unramified, it follows that π' is an embedding.

5. $d>2g$, $6\nmid d$

The following result is due to Ephraim and Kulkarni [E-K, Corollary 3.9]: For any nonnegative integer g and any d, not divisible by 6, and such that $d>2g$, there exists a plane curve $C\subset\mathbb{P}^2$ with only cuspidal singularities. Proceeding as in 4. above, we can thus show the existence of space curves of degree d and genus g (in the above range) admitting a cuspidal projection.

4. The Tangent Developable and its Singularities

Let $h: C\to C_0\subset P$ be as in the beginning of Sect. 2. Recall that we can describe the tangent developable of C_0 in the following way [P1]: Let $\mathscr{P}^m_C(1)$ denote the $(m+1)$-bundle of principal parts of order m of the line bundle $\mathcal{O}_C(1)=h^*\mathcal{O}_P(1)$. There are canonical maps

$$a^m: H^0(P,\mathcal{O}_P(1))_C=V_C\to\mathscr{P}^m_C(1).$$

The cokernel of a^1 is isomorphic to $\Omega^1_{C/P} \otimes \mathcal{O}_C(1)$. Hence, since C is a smooth curve, the image $\mathscr{P}^1 = \mathrm{Im}(a^1)$ is a rank 2 bundle on C. Setting $Y = \mathbb{P}(\mathscr{P}^1)$ we get a closed embedding $Y \hookrightarrow C \times P$. The tangent developable $X \subset P$ is then the image of Y under the projection onto the second factor. Note that the map $f : Y \to X$ is finite (this is true in arbitrary characteristic provided the curve C_0 is not strange) and birational, hence the singularities of X are resolved by normalization.

For $m = 2, 3$ the homomorphisms a^m are also generically surjective, since C_0 spans P. Set $\mathscr{P}^m = \mathrm{Im}(a^m)$. The bundle $\mathscr{P}^2$ represents the osculating planes of C_0, while $\mathscr{P}^3$ is isomorphic to V_C.

Let r_0 denote the degree of C_0 and g its (geometric) genus. The rank r_1 of C_0, defined as the number of tangents to C_0 meeting a given (general) line, is equal also to the degree of the tangent developable X. The class r_2 of C_0 is defined as the number of osculating planes containing a given (general) point.

Set $k_i = \sum_{p \in C} k_i(p)$, $i = 1, 2, 3$. We shall use repeatedly the following formulas [P1, 3.2]:

$$
\begin{aligned}
r_1 &= 2r_0 + 2g - 2 - k_0 \\
r_2 &= 3(r_0 + 2g - 2) - 2k_0 - k_1 \\
k_2 &= 4(r_0 + 3g - 3) - 3k_0 - 2k_1 .
\end{aligned} \tag{1}
$$

The rank and class also have interpretations in terms of dual varieties. Recall that the dual variety $\check{Z} \subset \check{\mathbb{P}}^n$ of a variety $Z \subset \mathbb{P}^n$ is defined as the closure of the set of hyperplanes tangent to Z at smooth points. The dual variety $\check{C} \subset \check{P}$ of $C_0 \subset P$ is thus a ruled surface, in fact it is the tangent developable X^* of the curve $C^* \subset \check{P}$, whose points are the osculating planes of C_0. We shall call C^* the dual curve of C_0. Since $\mathrm{char}\, k = 0$, biduality holds: The dual variety of $\check{C} = X^*$ is the curve C_0 (which is also the dual curve of C^*), and the dual variety of C^* is the tangent developable X of C_0. One of the characterizations of a developable surface is just that it is a ruled surface whose tangent planes are constant along a generator, i.e., whose dual variety is a curve.

We have

$$
\begin{aligned}
r_1 &= \text{degree of } X = \text{degree of } X^* \\
r_2 &= \text{degree of } C^* = \text{rank of } X .
\end{aligned}
$$

For the stationary indices k_i^* of C^* we have the duality

$$
k_i^* = k_{2-i}, \qquad i = 0, 1, 2 .
$$

Remarks. 1) For the duality between C_0 and C^*, see e.g. [P1]. Moreover, let $G = \mathrm{Grass}_2(V)$ denote the Grassmann variety of lines in P and $G^* = \mathrm{Grass}_2(V^\vee)$ the Grassmannian of lines in $\check{P}$. Then X can be considered as a curve $\Gamma \subset G$; in fact, $C \to G$, defined by the quotient $V_C \to \mathscr{P}^1$, is the 1st associated map of C_0. Likewise, X^* corresponds to a curve $\Gamma^* \subset G^*$, and the curves Γ and Γ^* are equal under the natural identification $G \cong G^*$ [P1, 5.2]. The fact that the ruled surface X (resp. X^*) is developable, is reflected on the curve Γ (resp. Γ^*) by the fact that its tangents are all contained in G (resp. G^*).

2) A proof, along the same lines, of the equality $\check{X}=C^*$ can be obtained using the functoriality of the bundles of principal parts [P1, Sect. 6]: If $q: Y=\mathbb{P}(\mathscr{P}^1)\to C$ denotes the structure map, one shows that $q^*a^1: q^*V_C\to q^*\mathscr{P}^2$ factors through $a^1_Y: V_Y\to\mathscr{P}^1_Y(1)$, where $\mathscr{P}^1_Y(1)$ denotes the principal parts of order 1 of the line bundle $\mathscr{O}_Y(1)=f^*(\mathscr{O}_P(1)|_X)$. Since a^1_Y is generically surjective, $\mathrm{Im}(a^1)$ and $q^*\mathscr{P}^2$ are isomorphic as quotients of V_Y, and hence the dual map of $X\subset P$ is defined on the normalization Y of X and has for image $C^*\subset\check{P}$.

3) Since the dual map of X is defined on Y, it follows that the Nash transformation of X is equal to the normalization $f: Y\to X$. Now Y is smooth, hence the singular points of X are such that the local Euler obstruction is equal to the multiplicity [L-T, 6.1.4].

Let us examine the singularities of the tangent developable X of C_0. Since X has smooth normalization, it is of the type studied in [P3] – we shall use results and notations from there.

The surface X has degree $\mu_0=r_1$, rank $\mu_1=r_2$, and class $\mu_2=0$. Its cuspidal edge consists of the curve C_0 and the tangents at flexes (taken with proper multiplicities) of C_0, hence has degree

$$c=r_0+k_1\,.$$

This can (also) be seen as follows: The cuspidal edge is the image of the ramification divisor R of $f: Y\to X$, hence it is defined by the ideal $F^0(\Omega^1_{Y/X})=F^0(\mathrm{Coker}\,a^1_Y: V_Y\to\mathscr{P}^1_Y(1))$. A local study of a^1_Y gives the support of R; moreover, it follows that the rational equivalence class of R is given by

$$[R]=(c_1(\mathscr{P}^1_Y(1))-c_1(q^*\mathscr{P}^2))\cap[Y]$$

since $\mathrm{Im}(a^1_Y)=q^*\mathscr{P}^2$. Using the standard exact sequences

$$0\to q^*\Omega^1_C\to\Omega^1_Y\to\Omega^1_{Y/C}\to 0 \tag{2}$$

and

$$0\to\Omega^1_{Y/C}\otimes\mathscr{O}_Y(1)\to q^*\mathscr{P}^1\to\mathscr{O}_Y(1)\to 0\,, \tag{3}$$

one obtains

$$[R]=(q^*(c_1(\Omega^1_C)+c_1(\mathscr{P}^1)-c_1(\mathscr{P}^2))+c_1(\mathscr{O}_Y(1)))\cap[Y]\,.$$

This gives

$$c=\deg f_*[R]=2g-2+r_1-r_2+r_1=r_2-2r_1+r_0+k_1+2r_1-r_2=r_0+k_1\,.$$

In addition to the cuspidal edge, the developable X has a double (or higher multiple) curve, called the *nodal curve* of C_0. It consists of points that are on more than one tangent to C_0. Eventual bitangents to C_0 are part of the nodal curve.

Let b denote the degree of the nodal curve. If the nodal curve is double and the flexes of C_0 are ordinary, we get from [P3, Theorem 4]:

$$\begin{aligned}2b=\mu_0(\mu_0-1)-\mu_1-3c&=r_1(r_1-1)-r_2-3(r_0+k_1)\\&=r_1(r_1-4)-k_0-2k_1\,.\end{aligned}$$

If the nodal curve consists of curves D_j, where D_j is ordinary j-multiple, then the degrees b_j of D_j satisfy [P3, Sect. 4]

$$\sum j(j-1)b_j = r_1(r_1-1) - r_2 - 3(r_0+k_1)$$

(still assuming the flexes to be ordinary).

The cuspidal edge and the nodal curve may themselves be singular, and also they have points of intersection – the points on X of higher multiplicity are found among these. In particular, if the nodal curve is double and the flexes ordinary, X will have a finite number of points of multiplicity $\geqq 3$. In fact, one can define a triple point cycle $[M_3]$ on Y, with respect to $f: Y \to X$ [K, Chap. V], and the number

$$T = \tfrac{1}{6} \deg f_*[M_3]$$

will be called the (total) number of triple points of X. Note that this number is well defined even if the nodal curve is not double and the flexes are not ordinary, since $[M_3]$ is defined as the double point cycle of the map from the double point scheme of f to Y (see [K, loc. cit.]). As in [P3, Sect. 5] we shall apply Kleiman's triple point formula [K, V. 82] to compute T.

Proposition 1. *The total number T of triple points of the tangent developable X of C_0 is given by*

$$T = \tfrac{1}{6}(r_1-4)((r_1-3)(r_1-2) - 6g).$$

Remarks. 1) Suppose $S \subset P$ is a ruled (non-developable) surface of degree m and genus g, with ordinary singularities, i.e., a double curve, a finite number of pinch points, and a finite number t of (ordinary) triple points. Then t is given by [K, V. 84]

$$t = \tfrac{1}{6}(m-4)((m-3)(m-2) - 6g).$$

Now t is also equal to the number of tritangent planes to S: since the dual variety $\check{S} \subset \check{P}$ is a ruled surface which is also of degree m and genus g, the number $\check{t}$ of triple points of $\check{S}$ has the same expression as t, and the triple points of $\check{S}$ are just the tritangent planes of S.

Similarly, the tritangent planes to the developable X correspond to the triple points of the dual variety of C_0, which is the tangent developable X^* of $C^* \subset \check{P}$. Since C^* has the same rank and genus as C_0, the number $\check{T}$ of triple points of X^* is equal to T.

2) The observations made in 1) show that the total number of triple points on a ruled surface (developable or not) is equal to its number of tritangent planes, and that this number is determined by its genus and its degree viewed as a curve Γ in the Grassmann variety G of lines in P. Note that a triple point of the ruled surface corresponds to a Schubert cycle $\sigma_{2,0}$ (consisting of all lines through the point) on G which is trisecant to Γ. A tritangent plane corresponds to a Schubert cycle $\sigma_{1,1}$ (lines in the plane) which is trisecant to Γ. Now the $\sigma_{2,0}$-planes in G corresponds to the $\sigma_{1,1}$-planes in G^*; since $\Gamma = \Gamma^*$ under the identification $G = G^*$, the equality between the number of triple points and the number of tritangent planes is justified.

Proof. With notations as in [P3, Sect. 5], let $\Gamma_c \subset Y$ denote the reduced ramification locus of $f: Y \to X$, and $\Gamma_b \subset Y$ the reduced inverse image via f of the nodal curve. Set $i = (\Gamma_b, \Gamma_c)$, and let $\mathcal{N} = f^*\Omega^1_P - \Omega^1_Y$ denote the virtual conormal bundle of $f: Y \to P$.

We shall apply Kleiman's triple point formula [K, V. 82] – as in [P3, p. 226] – to the map f. This we can do, because the double point scheme of f has the expected dimension, namely 1 [K, p. 387]. Assume now that the nodal curve is double and the flexes of C_0 are ordinary. Then the formula yields

$$3T = \tfrac{1}{2}\deg f_*[M_3] = (\Gamma_b^2) + 4(\Gamma_c^2) + 4i - r_1(b+c) + c_2(\mathcal{N}).$$

Lemma 6. (i) $(\Gamma_c^2) = 2 - 2g + k_0 + 2k_1$.
(ii) $(\Gamma_b^2) = (r_1 - 1)(r_1^2 - 7r_1 - 2k_0 - 4k_1) + 9r_1 + 6k_0 + 12k_1$.
(iii) $i = (r_1 - 4)(r_0 + k_1) - k_0 - 2k_1$.
(iv) $c_2(\mathcal{N}) = 2(r_1 + 2g - 2)$.

Proof of Lemma 6. (i) Write $\Gamma_c = C \cup \bigcup_i F_i$, where $\{f(F_i)\}$ are the flex tangents to C_0, and $C \subset Y$ is the section of $q: Y \to C$ given by the quotient $\mathcal{P}^1 \to \mathcal{O}_C(1)$. If $\mathcal{N}_{C/Y}$ denotes the conormal bundle of C in Y, then $(C^2) = -\deg \mathcal{N}_{C/Y}$. From the exact sequences (2), (3), and

$$0 \to \mathcal{N}_{C/Y} \to \Omega^1_Y|_C \to \Omega^1_C \to 0,$$

and the fact $q^*\Omega^1_C|_C = \Omega^1_C$, it follows that

$$\deg \mathcal{N}_{C/Y} = \deg \Omega^1_{Y/C} = \deg \mathcal{P}^1 - 2\deg \mathcal{O}_Y(1) = r_1 - 2r_0 = 2g - 2 - k_0.$$

The formula follows from this, since we have $(C, F_i) = 1$ and $(F_i, F_j) = 0$.

(ii) Set $H = c_1(f^*\mathcal{O}_X(1)) \cap [Y]$ and $\check{H} = c_1(\check{f}^*\mathcal{O}_{\check{P}}(1)) \cap [Y]$, where $f: Y \to \check{P}$ denotes the dual map of X [recall that $\check{f}(Y) = C^*$, the dual curve of C_0]. Then we have [P3, Sect. 5]

$$[\Gamma_b] = (r_1 - 1)H - 3[\Gamma_c] - \check{H}.$$

Since we have $(H^2) = \mu_0 = r_1$, $(H, \check{H}) = \mu_1 = r_2$, and $(\check{H}^2) = \mu_2 = 0$ this and (i) imply

$$(\Gamma_b^2) = r_1(r_1 - 1)^2 + 9(2k_1 + k_0 - 2g + 2) - 6(r_1 - 1)(r_0 + k_1) - 2r_2(r_1 - 1) + 6r_2.$$

The stated formula is a rewriting of this one, using the relations (1) between the r_i's and k_i's stated in the beginning of Sect. 4.

(iii) The expressions for $[\Gamma_b]$ and $[\Gamma_c]$ given above give

$$i = (\Gamma_b, \Gamma_c) = (r_1 - 1)(r_0 + k_0) - 3(2 - 2g + k_0 + 2k_1) - \sigma,$$

where $\sigma = (\Gamma_c, \check{H})$ is the *class of immersion* of $f(\Gamma_c)$ in X, i.e., σ is the degree of $\check{f}(\Gamma_c)$ in $\check{P}$. But $\check{f}(\Gamma_c) = C^*$, hence $\sigma = r_2$. The rest is again formal manipulations.

(iv) By the definition of $\mathcal{N}$,

$$\begin{aligned} c_2(\mathcal{N}) &= c_2(f^*\Omega^1_P - \Omega^1_Y) \\ &= c_2(f^*\Omega^1_P) - c_1(f^*\Omega^1_P)c_1(\Omega^1_Y) + c_1(\Omega^1_Y)^2 - c_2(\Omega^1_Y) \\ &= 6c_1(\mathcal{O}_Y(1))^2 + 4c_1(\mathcal{O}_Y(1))c_1(\Omega^1_Y) + c_1(\Omega^1_Y)^2 - c_2(\Omega^1_Y). \end{aligned}$$

Now $q: Y \to C$ is a ruling, so $c_1(\Omega^1_Y)^2 = 2c_2(\Omega^1_Y)$ and

$$\begin{aligned} c_1(\Omega^1_Y) &= c_1(q^*\Omega^1_C) + c_1(\Omega^1_{Y/C}) \\ &= c_1(q^*\Omega^1_C) + c_1(q^*\mathscr{P}^1) - 2c_1(\mathscr{O}_Y(1)), \\ c_1(\Omega^1_Y)^2 &= -4(2g-2). \end{aligned}$$

Hence,

$$c_2(\mathscr{N}) = 6r_1 + 4(2g-2+r_1-2r_1) - 2(2g-2) = 2(r_1+2g-2). \quad \square$$

To finish the proof of the proposition, use the lemma on the terms in the triple point formula to obtain

$$3T = \tfrac{1}{2}r_1((r_1-1)(r_1-2) - 3r_0 - 3r_2 + 22 - 3k_1) - 10r_0 + 2r_2 + 6k_0 + 2k_1 - 2(2g-2),$$

and then apply (1) to eliminate r_0, r_2, and k_0.

Though we assumed for simplicity the nodal curve to be double and the flexes to be ordinary, the same formula is valid in the general case: Replace Γ_b by $\Sigma_j (j-1)\Gamma_{b_j}$, if $f(\Gamma_{b_j})$ is j-multiple on X. If the flexes are not ordinary, each F_i will count with a multiplicity m_i in the double point cycle of f. Adjusting the expressions of the lemma accordingly, and plugging them into the triple point formula, one observes that the "bad" terms disappear, and one is left with exactly the same expression. $\square$

Assume now that the nodal curve is double and the flexes ordinary. The points of multiplicity $\geqq 3$ of X contains: the t points where three distinct tangents to C_0 meet, the γ intersections (outside C_0) of the nodal curve with the flex tangents, the $d(1,2)$ points where a tangent to C_0 cuts C_0 again, the k_0 cusps and the k_1 flexes of C_0.

Zeuthen gives the following formulas for t and $d(1,2)$ [Z]:

$$\begin{aligned} t &= \tfrac{1}{6}r_1(r_1^2 - 3r_1 - 9r_0 - 3r_2 - 9k_1) + \tfrac{1}{3}(39r_0 - 29r_1 + 21r_2 + 39k_1), \\ d(1,2) &= r_1(r_0 - 6) + 4r_0 - 6k_0 - 2k_1. \end{aligned}$$

[For a proof when $k_0 = k_1 = 0$, see [G-H, p. 294]. A formula for $d(1,2)$ including the flexes – counted twice – is given by Le Barz [LB], also under the assumption that C_0 is smooth.]

Lemma 7. $\gamma = (r_1 - 6)k_1$.

Proof. Let H be a plane containing a tangent $T(p)$ to C_0, and assume H is not osculating. Then $H \cap X = D \cup (l_2 - l_1 + 1)T(p)$, where the plane curve D has degree $r_1 - (l_2 - l_1 + 1)$ and a singularity of type $(l_1 + 1, l_3 + 3)$ at $h(p)$. By Lemma 2, D and $T(p)$ have intersection number $l_3 + 3$ at $h(p)$. The number of intersections of $T(p)$ and X outside $h(p)$ is thus equal to

$$\begin{aligned} \gamma(p) &= r_1 - (l_2 - l_1 + 1) - l_3 - 3 = r_1 - 4 - l_3 - l_2 + l_1 \\ &= r_1 - 4 - k_2(p) - 2k_1(p) - k_0(p). \end{aligned}$$

In particular, the tangent $T(p)$ meets

$$\left.\begin{matrix} r_1-4 \\ r_1-5 \\ \\ r_1-6 \end{matrix}\right\} \text{ other tangents if } \begin{cases} p \text{ is a regular point} \\ p \text{ is an ordinary cusp} \\ \quad \text{or an ordinary stall} \\ p \text{ is an ordinary flex.} \end{cases} \square$$

Remarks. 1) With the above formulas, we obtain the following equality:

$$T = t + d(1,2) + \gamma + k_0 + 2k_1 .$$

Why the flexes, as opposed to the other triple points of X, appear with multiplicity 2, remains a mystery....

2) The stalls of C_0 are also points where the nodal curve intersects the cuspidal edge, but these points are singular on neither curve and are of multiplicity 2 on X (they are in fact "cuspidal" Whitney umbrellas).

In [P1, 5.1(ii)] it is shown that the dual (plane) curve of a generic plane projection of $C_0 \subset P$, is equal to the corresponding plane section of the dual variety $X^* \subset \check{P}$ of C_0. Going through that proof, one verifies that the genericity assumption is in fact unnecessary. Therefore, the results on the nature of the singularities of plane sections of the developable (Sect. 2), could also have been obtained by duality from the knowledge of the singularities of plane projections. Note that if p is a point of type (l_1+1, l_2+2, l_3+3), then the corresponding point on the dual curve C is of type $(l_3-l_2+1, l_3-l_1+2, l_3+3)$ [B, p. 184].

5. The Case of Positive Characteristic

Suppose now that k is an algebraically closed field of characteristic $p>0$. We shall see that Theorems 1 and 2 also hold in this case.

Proof of Theorem 1. Case i) is obvious. In order to prove Cases ii) and iii) we shall use a lifting argument to reduce to the characteristic 0 case (so that we don't need to show e.g. that Telling's classification of rational quartics is valid in positive characteristic).

Proposition 2. *Let k be an algebraically closed field of characteristic $p>0$ and $C \subset \mathbb{P}^3_k$ a smooth, proper curve such that $\mathcal{O}_C(1)$ is non-special. Then $C \subset \mathbb{P}^3_k$ is the reduction* (mod p) *of a curve $\mathscr{C}_{\bar{K}} \subset \mathbb{P}^3_{\bar{K}}$ over an algebraically closed field $\bar{K}$ of characteristic* 0. *Suppose $\mathscr{C}_{\bar{K}}$ admits a cuspidal projection. If the induced projection of C is birational onto its image, then it is also cuspidal.*

Proof. Let W denote the Witt ring of k, and K its field of fractions. It is known [G, Corollary 4, p. 14] that C can be lifted to W: There exists a smooth curve $\mathscr{C} \to \mathrm{Spec}(W)$ such that $\mathscr{C}_k = C$.

Since, by assumption, $H^1(C, \mathcal{O}_C(1)) = 0$, we can also lift the given embedding of C to an embedding $\mathscr{C} \subset \mathbb{P}^3_W$: If $s \in H^0(C, \mathcal{O}_C(1))$ is a non-zero section, then the subscheme $\mathrm{div}(s) \subset C$ can be lifted to $\mathscr{C}$, since $\mathscr{C} \to \mathrm{Spec}\, W$ is flat, to give a relative

divisor $S \in \mathrm{Div}^+(\mathscr{C}/W)$. Hence we get an associated invertible sheaf $\mathscr{L}$ on $\mathscr{C}$ such that $\mathscr{L}_k \cong \mathscr{O}_C(1)$.

Multiplication by the uniformizing parameter p of W gives an exact sequence

$$0 \to \mathscr{L} \xrightarrow{p} \mathscr{L} \to \mathscr{O}_C(1) \to 0,$$

hence also

$$0 \to H^0(\mathscr{C}, \mathscr{L}) \xrightarrow{p} H^0(\mathscr{C}, \mathscr{L}) \xrightarrow{\alpha} H^0(C, \mathscr{O}_C(1)) \to H^1(\mathscr{C}, \mathscr{L}) \xrightarrow{p} H^1(\mathscr{C}, \mathscr{L}) \to 0.$$

Therefore $H^1(\mathscr{C}, \mathscr{L}) = 0$ and $H^0(\mathscr{C}, \mathscr{L}) \xrightarrow{\alpha} H^0(C, \mathscr{O}_C(1))$ is surjective.

By assumption, there is a 4-dimensional k-vector space $V \subset H^0(C, \mathscr{O}_C(1))$ so that the diagram

$$\begin{array}{ccc} H^0(C, \mathscr{O}_C(1))_C & \longrightarrow & \mathscr{O}_C(1) \\ \cup & \nearrow & \\ V_C & & \end{array}$$

corresponds to

$$\begin{array}{ccc} C & \xrightarrow{j} & \mathbb{P}(H^0(C, \mathscr{O}_C(1))) \\ & \searrow^{i} & \downarrow \\ & & \mathbb{P}(V) \cong \mathbb{P}^3_k. \end{array}$$

Since α is surjective, we can choose a lifting $\mathscr{V}$ of V,

$$\begin{array}{ccc} H^0(\mathscr{C}, \mathscr{L})_{\mathscr{C}} & \longrightarrow & \mathscr{L} \\ \cup & \nearrow & \\ \mathscr{V}_{\mathscr{C}} & & \end{array}$$

and get rational maps

$$\begin{array}{ccc} \mathscr{C} & \overset{j}{\dashrightarrow} & \mathbb{P}(H^0(\mathscr{C}, \mathscr{L})) \\ & \searrow^{i} & \downarrow \\ & & \mathbb{P}(\mathscr{V}) \cong \mathbb{P}^3_W \end{array}$$

specializing to i and j. Consequently i and j are defined on all of $\mathscr{C}$. Moreover, since i and j give embeddings of the closed fibre, they are embeddings [the singular loci of $i(\mathscr{C})$ and $j(\mathscr{C})$ being closed].

Let $\bar{K}$ be an algebraic closure of K, and $\bar{W}$ the valuation ring in $\bar{K}$ extending W. Set $\bar{\mathscr{C}} = \mathscr{C} \times \mathrm{Spec}(\bar{W})$ and $\bar{\mathscr{L}} = \mathscr{L} \times_W \bar{W}$, and let $\bar{i} : \bar{\mathscr{C}} \hookrightarrow \mathbb{P}^3_{\bar{W}}$ denote the embedding obtained from i by base change.

Suppose $\bar{\pi}_{\bar{K}} : \bar{\mathscr{C}}_{\bar{K}} \to \mathbb{P}^2_{\bar{K}}$ is a cuspidal projection of $\bar{i}_{\bar{K}} : \bar{\mathscr{C}}_{\bar{K}} \hookrightarrow \mathbb{P}^3_{\bar{K}}$. Let $\bar{\pi} : \bar{\mathscr{C}} \to \mathbb{P}^2_{\bar{W}}$ denote the corresponding projection of $\bar{\mathscr{C}}$. Assume $\pi = \bar{\pi}_k : C \to \mathbb{P}^2_k$ is birational onto its image. Clearly π has at least as many ramification points as $\bar{\pi}_{\bar{K}}$, and since the total number δ of singularities of $\bar{\pi}_{\bar{K}}(\bar{\mathscr{C}}_{\bar{K}})$ and $\pi(C)$ is the same, π could only be non-cuspidal if a cusp of $\bar{\pi}_{\bar{K}}(\bar{\mathscr{C}}_{\bar{K}})$ (with δ-value greater than 1) split into two or more singular points (at least one of which non-unibranch) of $\pi(C)$. But since $\bar{W}$ is henselian, this cannot happen: the singular subscheme of $\bar{\pi}(\bar{\mathscr{C}})$, as defined by the

conductor say, is finite over Spec $\bar{W}$, hence splits into a disjoint union of local components; hence there are at least as many points in the open fibre as in the closed. □

The above proposition applies to the Cases 2) and 3) of Theorem 1: in both cases $\mathcal{O}_C(1)$ is non-special; hence we can lift to characteristic 0. In Case 3), any cuspidal projection of $\bar{\mathcal{C}}_{\bar{K}}$ induces a projection of C which is necessarily birational onto its image, since C is rational. In Case 2), one argues as in the characteristic 0 case that there can't be an infinity of non birational projections of C; since, by the characteristic 0 proof, there are infinitely many centres giving cuspidal projections of $\bar{\mathcal{C}}_{\bar{K}}$, the proof is completed. □

Proof of Theorem 2. The argument showing that a general canonical curve $C \subset \mathbb{P}^3_k$ has no flexes and only ordinary stalls, works also in the positive characteristic case.

To prove Lemma 3, one first observes that, even in characteristic 2 [SGA7], any smooth quadric is projectively equivalent to one of the form $Q_0 : X_0X_3 = X_1X_2$. To be able to use the same definition of Weierstrass points, one has to be careful, however: it could happen that the map $a^{g-1} = a^3$ is nowhere surjective (all points of C are stalls). To show that a^3 is generically surjective for a general C, it suffices to show that a general C has at least one regular (see Sect. 2) point. Fixing Q_0 and the point $(1,0,0,0)$, one sees – computing a parametrization as in the characteristic 0 case – that regularity means avoiding two hypersurfaces in the parameter space of the cubic polynomials. Hence there is an open, dense U such that the stalls of the curves $\mathcal{C}_u$, $u \in U$, are defined by the relative divisor $W = \mathrm{div}(w_{\mathcal{C}/U})$ as in the characteristic 0 case, and the rest of the proof of Lemma 3 goes through as before.

The deformation argument given to prove Lemma 4, holds in positive characteristic as well. Though the jacobian ideal in this case might be "too small", it is still true that $H^0(\bar{C}, J \otimes N)$ gives the locally trivial deformations, and the equality (conductor)$\cdot I = J\mathcal{O}_C$ [P2, p. 261] is still valid (if J is "too small", then so is I). Note that the dimensions r and r' might be different from their values in characteristic 0; in the present case, since all singularities are of multiplicity 2, the only difference occurs in characteristic 2. Then $\deg(I^{-1}) = 6 \cdot 2 = 12$ in Case i), and $\deg(I^{-1}) = 4 \cdot 2 + 6 = 14$ in Case ii). Hence $r \leqq r'$, where $r' = 9$ in Case i) and $r' = 7$ in Case ii) (in char 2, $J \neq \pi_* J\mathcal{O}_{\tilde{C}}$ also at an ordinary cusp).

Finally, one observes that Lemma 5 works as it stands (again using [SGA7] if $\mathrm{char}\, k = 2$). □

Acknowledgement. I would like to thank Robin Hartshorne for suggesting looking at the canonical curves of genus 4 in this context.

References

[A] Artin, M.: Deformation theory. Course at M.I.T., 1974

[B] Baker, H.F.: Principles of geometry, Vol. V. Cambridge: Cambridge University Press 1939

[E-K] Ephraim, R., Kulkarni, R.S.: Combinatorial embeddings of surfaces as algebraic curves. Indiana Univ. Math. J. **29**, 21–30 (1980)

[F1] Ferrand, D.: Set-theoretical complete intersections in characteristic $p > 0$. In: Algebraic geometry, Copenhagen 1978. Lønsted, K. (ed.). Lecture Notes in Mathematics, Vol. 732, pp. 82–89. Berlin, Heidelberg, New York: Springer 1979

[F2] Ferrand, D.: Commentaires à Set-theoretical complete intersections in characteristic $p>\Delta$. Preprint (1979)
[G-H] Griffiths, P., Harris, J.: Principles of algebraic geometry. New York: Wiley 1978
[G] Grothendieck, A.: Géométrie formelle et géométrie algébrique. Séminaire Bourbaki, Exp. 182. Vol. 1958/59
[G-P] Gruson, L., Peskine, C.: Genre des courves de l'espace projectif. In: Algebraic geometry, Tromsø 1977. Olson, L. (ed.) Lecture Notes in Mathematics, Vol. 687, pp. 31–59. Berlin, Heidelberg, New York: Springer 1978
[H] Hartshorne, R.: Complete intersections in characteristic $p>0$. Am. J. Math. **101**, 380–383 (1979)
[K] Kleiman, S.L.: The enumerative theory of singularities. In: Real and complex singularities, Oslo 1976, pp. 297–396. Holm, P. (ed.). Groningen: Sijthoff and Noordhoff 1978
[L-T] Lê, D.T., Teissier, B.: Variétés polaires locales et classes de Chern des variétés singulières. Preprint, École Polytechnique 1979
[LB] LeBarz, P.: Validité de certaines formules de géométrie énumerative. C. R. Acad. Sci. Paris **289**, 755–758 (1979)
[P1] Piene, R.: Numerical characters of a curve in projective n-space. In: Real and complex singularities, Oslo 1976, pp. 475–496. Holm, P. (ed.). Groningen: Sijthoff and Noordhoff 1978
[P2] Piene, R.: Polar classes of singular varieties. Ann. Scie. École Norm. Sup. **11**, 247–276 (1978)
[P3] Piene, R.: Some formulas for a surface in $\mathbb{P}^3$. In: Algebraic geometry, Tromsø 1977, Olson, L. (ed.). Lecture Notes in Mathematics, Vol. 687, pp. 196–235. Berlin, Heidelberg, New York: Springer 1978
[S] Segre, B.: Sulla caratterizziazione delle curve di diramazionc dei piani multipli generali. Mem. Accad. Ital. Mat. **1**, 5–31 (1930)
[SGA7] Deligne, P.: Quadriques. In: Groupes de monodromie en géométrie algébrique (SGA7II). Deligne, P., Katz, N. (ed.). Lecture Notes in Mathematics, Vol. 340, pp. 62–82. Berlin, Heidelberg, New York: Springer 1973
[S-K] Semple, J.G., Kneebone, G.T.: Algebraic curves. Oxford: Clarendon Press 1959
[T] Telling, H.G.: The rational quartic curve in space of three and four dimensions. Cambridge Tracts No. 34. Cambridge: Cambridge University Press 1936
[Zar] Zariski, O.: Algebraic surfaces. 2nd ed. Berlin, Heidelberg, New York: Springer 1971
[Z] Zeuthen, H.G.: Sur les singularités ordinaires d'une courbe gauche et d'une surface développable. Ann. Mat. Ser. 2, **3**, 175–217 (1869)

Received July 25, 1980; in revised form December 29, 1980

Note added in proof. A special case of 1) and 2) of Sect. 3 (p. 101) has been proved by G. Sacchiero ("Su una classe di curve che non ammettono proiezioni cuspidali", to appear). He shows that, if $\operatorname{char} k=0$, and if $C\subset\mathbb{P}^3_k$, of degree d and genus g, is a generic projection of a normal, non-special curve in $\mathbb{P}^{d-g}$ ($d-g\geqq 4$, $\delta\geqq 4$), then C does not admit a cuspidal projection.

Math. Ann. 256, 121–132 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Finite p-Periodic Quotients of General Linear Groups

B. Bürgisser

Forschungsinstitut für Mathematik, ETH Zentrum, CH-8092 Zürich, Switzerland

1. Introduction

1.1.

For a group G of virtually finite (cohomological) dimension let $\hat{H}^i(G, A)$, $i \in \mathbb{Z}$, be the i-th Farrell-Tate cohomology with coefficients in the G-module A (see [7]). The groups $\hat{H}^i(G, A)$ are always torsion-groups; they agree with the usual cohomology $H^i(G, A)$ for dimensions i greater than the virtual (cohomological) dimension of G.

The group G is called p-periodic for a prime number p, if there exists $m_p \in \mathbb{N}$ such that the functors $\hat{H}^i(G, -, p)$ and $\hat{H}^{i+m_p}(G, -, p)$ are naturally equivalent for all $i \in \mathbb{Z}$, where $\hat{H}^i(G, A, p)$ denotes the p-primary component of $\hat{H}^i(G, A)$. The smallest such number m_p is called the p-period of G. If G is p-periodic, then every subgroup U of G is p-periodic and the p-period of U divides the p-period of G. If G is a finite group, then $\hat{H}^i(G, A)$ is the usual Tate cohomology and the above notion of p-periodicity is the same as in [5, Chap. XII].

1.2.

In this paper we investigate – for arithmetic and S-arithmetic groups $G = \mathrm{GL}(n, O_S(k))$ – whether there exist finite p-periodic quotients G/N with N torsion-free. We note that, in that case, G itself is p-periodic; more precisely [3, Theorem 3.14]: Let G be virtually of type FP and N a normal subgroup of finite index in G without p-torsion. If G/N is p-periodic with p-period μ_p, then G is p-periodic with p-period m_p, where m_p divides μ_p.

1.3.

Finite p-periodic quotients G/N (N torsion-free) also appear in topological applications. If such a quotient of G exists, then – under certain finiteness assumptions on N – the group G operates freely on a fibre bundle $\mathbb{R}^s \tilde{\times} M$ where M is a smooth manifold of dimension m ($= p$-period of G/N) with $H^*(M, \mathbb{Z}_{(p)}) \cong H^*(S^m, \mathbb{Z}_{(p)})$. Here S^m denotes the m-sphere and $\mathbb{Z}_{(p)}$ the localization of

0025-5831/81/0256/0121/$02.40

$\mathbb{Z}$ in $p\mathbb{Z}$. This is a result of C. B. Thomas. He proved similar, accordingly stronger statements if G/N is p-periodic for several prime numbers p at once.

1.4.

Let k be an algebraic number field with discriminant $\Delta(k)$, R a Dedekind ring contained in k and $G=\mathrm{GL}(n,R)$. In several cases G is known to be virtually of type FP. If in particular $R=O(k)$, the ring of algebraic integers in k, we obtain the following result:

Theorem. *Let be $n\geqq 2$ and $G=\mathrm{GL}(n,O(k))$. Then there exists a torsion-free normal subgroup N of G of finite index with the following properties: G/N is p-periodic with p-period $2(p-1)$ for all prime numbers p with $\frac{n}{2}+1<p\leqq n+1$ and $p\nmid\Delta(k)$; moreover, given $c\in\mathbb{N}$, one can choose N such that G/N has no p-torsion for all prime numbers p with $n+1<p\leqq c$ and $p\nmid\Delta(k)$.*

We will take for this normal subgroup N the principal congruence subgroup of level P, where P is a certain principal prime ideal of $O(k)$. It follows from [10, Sect. 1] that N is of type FL.

The result above is proved in Sect. 7 (Theorem 7.4) under less restrictions on p and in a more general version dealing with the S-arithmetic group $\mathrm{GL}(n,O_S(k))$.

Taking into account 1.2 one gets from the Theorem above a result about p-periodicity of $G=\mathrm{GL}(n,O(k))$, which has been proved in [4, Theorem 1.1] using completely different methods.

We recall that for prime numbers $p\leqq n+1$ the group $G=\mathrm{GL}(n,O(k))$ contains p-torsion. It follows that for $p\leqq\frac{n}{2}+1$ there cannot exist any torsion-free normal subgroup N of G of finite index such that G/N is p-periodic.

The proof of Theorem 7.4 uses several preliminaries given in Sects. 1–6. Many of these belong to algebraic number theory, with the objective of constructing the appropriate prime ideal P of $O(k)$.

In Sect. 8 we look closer at non-Galois extensions k of $\mathbb{Q}$ of dimension 3; and in Sect. 9 we consider the case of $k=\mathbb{Q}(\theta)$, where θ is a root of x^r-a (r odd, $a\in\mathbb{Z}$). In both cases we get statements analogous to those in the above theorem also for prime numbers p dividing $\Delta(k)$ (Theorems 8.1 and 9.1).

1.5.

Completely analogous results are obtained for the special linear group $\mathrm{SL}(n,O(k))$:

Theorem. *Let be $n\geqq 2$ and $G=\mathrm{SL}(n,O(k))$. Then there exists a torsion-free normal subgroup N of G of finite index with the following properties: G/N is p-periodic with p-period $2(p-1)$ for all prime numbers p with $\frac{n}{2}+1<p\leqq n+1$ and $p\nmid\Delta(k)$; moreover, given $c\in\mathbb{N}$, one can choose N such that G/N has no p-torsion for all prime numbers p with $n+1<p\leqq c$ and $p\nmid\Delta(k)$.*

For $k=\mathbb{Q}$ one obtains a result concerning $\mathrm{SL}(n,\mathbb{Z})$ which is already contained in the authors Thesis [3, Sect. 4].

1.6.

We add that a recent result of Carter [6] together with our Theorem 7.2 yield the following statement about the reduced projective class group of the ring $\mathbb{Z}G$, written $\tilde{K}_0(\mathbb{Z}G)$, where $G=\mathrm{GL}(n, O(k))$: For any prime number p with $\frac{n}{2}+1<p \leqq n+1$ and $p \nmid \Delta(k)$ there exists an element in $\tilde{K}_0(\mathbb{Z}G)$ of order $\frac{1}{2}(p-1)$. We will discuss some refinements of this result elsewhere.

2. p-Periodicity of $\mathrm{SL}(n, F_q)$ and $\mathrm{GL}(n, F_q)$

Let be $n \in \mathbb{N}$; for a prime number q the field with q elements is denoted by F_q.

Lemma 2.1. *Let be $n \geqq 2$ and p a prime number with $\frac{n}{2}+1<p \leqq n+1$. Let q be a prime number such that $q \neq p$ and $p^2 \nmid |\mathrm{GL}(n, F_q)|$. Then $\mathrm{SL}(n, F_q)$ and $\mathrm{GL}(n, F_q)$ are both p-periodic with p-period $2(p-1)$.*

Proof. Since $p^2 \nmid |\mathrm{GL}(n, F_q)|$, both groups $\mathrm{SL}(n, F_q)$ and $\mathrm{GL}(n, F_q)$ are p-periodic. That the p-period of $\mathrm{SL}(n, F_q)$ is exactly $2(p-1)$ has been proved in [3, Theorem 4.7] using a result of Swan [11]. Since the p-period of $\mathrm{GL}(n, F_q)$ is according to the assumption at most $2(p-1)$, the assertion follows.

3. Principal Congruence Subgroups

Let R be a Dedekind ring whose quotient field k is a number field; $[k:\mathbb{Q}]=d<\infty$. We remark that such a ring R contains $O(k)$ (= ring of algebraic integers in k). Let I be an ideal of R; we denote by $N(I)$ the absolute norm of I; i.e. $N(I)=|R/I|$.

The principal congruence subgroup of $\mathrm{GL}(n, R)$ of level I, written $\mathrm{GL}(n, R, I)$, is the kernel of the canonical homomorphism $\mathrm{GL}(n, R) \to \mathrm{GL}(n, R/I)$. Hence $\mathrm{GL}(n, R, I)$ is a normal subgroup of $\mathrm{GL}(n, R)$, of finite index if $I \neq (0)$. If $N(I)>2^d$, then $\mathrm{GL}(n, R, I)$ is torsion-free; the proof is essentially the same as in the case $R=O(k)$ (see [3, Lemma 5.8]). We summarize:

Lemma 3.1. *Let R be a Dedekind ring, whose quotient field k is a number field, $[k:\mathbb{Q}]=d$. Let $I \neq (0)$ be an ideal in R such that $N(I)>2^d$. Then the principal congruence subgroup $\mathrm{GL}(n, R, I)$ is a torsion-free normal subgroup of $\mathrm{GL}(n, R)$ of finite index.*

4. p-Periodicity of $\mathrm{GL}(n, R)/N$

For a prime number p we denote by $(\mathbb{Z}/p^2\mathbb{Z})^*$ the cyclic group of relatively prime residue classes of $\mathbb{Z}$ modulo p^2.

Theorem 4.1. *Let R be a Dedekind ring, whose quotient field k is a number field, $[k:\mathbb{Q}]=d$; let be $n \geqq 2$ and M a set of prime numbers. Assume that a prime ideal P of R exists such that $N(P)$ is a prime number $>2^d$ and $N(P)+p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p \in M$. Then there exists a torsion-free normal subgroup N of $\mathrm{GL}(n, R)$ of finite index with following properties:*

(a) *the p-Sylow subgroup of* $\mathrm{GL}(n,R)/N$ *is cyclic of order p and the p-period of* $\mathrm{GL}(n,R)/N$ *is* $2(p-1)$ *for all* $p\in M$ *with* $\frac{n}{2}+1<p\leqq n+1$.

(b) $\mathrm{GL}(n,R)/N$ *has no p-torsion for all* $p\in M$ *with* $p>n+1$.

Proof. The principal congruence subgroup $N=\mathrm{GL}(n,R,P)$ is by Lemma 3.1 a torsion-free normal subgroup of $\mathrm{GL}(n,R)$ of finite index. The field R/P has $N(P)=q$ elements, where q is a prime number. For $p\in M$ the residue class $q+p^2\mathbb{Z}$ has order $p(p-1)$ in $(\mathbb{Z}/p^2\mathbb{Z})^*$. Therefore one has

$$p\nmid q^r-1 \quad \text{for} \quad 0<r<p-1 \quad \text{and} \quad p-1<r<2(p-1)$$

and $p^2\nmid q^{p-1}-1$; moreover $p\mid q^{p-1}-1$.

(a) If $\frac{n}{2}+1<p\leqq n+1$, we conclude from the considerations above that

$$p^2\nmid|\mathrm{GL}(n,F_q)|=q^{\frac{n(n-1)}{2}}\prod_{k=1}^{n}(q^k-1).$$

Using Lemma 2.1 we get: $\mathrm{SL}(n,F_q)$ and $\mathrm{GL}(n,F_q)$ are both p-periodic with p-period $2(p-1)$. Let ψ be the canonical homomorphism $\mathrm{GL}(n,R)\to\mathrm{GL}(n,R/P)$. Since the restriction $\psi|_{\mathrm{SL}(n,R)}:\mathrm{SL}(n,R)\to\mathrm{SL}(n,R/P)$ is surjective [2, p. 267], one gets

$$\mathrm{SL}(n,F_q)\cong\mathrm{SL}(n,R/P)\subset\mathrm{Im}\,\psi\subset\mathrm{GL}(n,R/P)\cong\mathrm{GL}(n,F_q).$$

Therefore $\mathrm{GL}(n,R)/\mathrm{GL}(n,R,P)\cong\mathrm{Im}\,\psi$ is p-periodic with p-period $2(p-1)$, and the p-Sylow subgroup is cyclic of order p.

(b) If $p>n+1$, the same considerations show that $p\nmid|\mathrm{GL}(n,F_q)|$ and therefore $p\nmid|\mathrm{GL}(n,R)/\mathrm{GL}(n,R,P)|$.

We add a remark concerning the case $n=1$: The group $\mathrm{GL}(1,R)$ is the group of units in R, written $U(R)$. If R is the ring of algebraic integers in k, then one concludes from the well known structure of $U(R)$ (Dirichlet's Unit Theorem): there exists a torsion-free normal subgroup N of $U(R)$ of finite index such that $U(R)/N$ is periodic with period 2.

Remark 4.2. The group $N'=\mathrm{GL}(n,R,I)\cap\mathrm{SL}(n,R)$ is a torsion-free normal subgroup of $\mathrm{SL}(n,R)$ of finite index, and $\mathrm{SL}(n,R)/N'\cong\mathrm{SL}(n,F_q)$. The proof above shows that Theorem 4.1 is valid for $\mathrm{SL}(n,R)$ as well.

The following theorem is a special case of our Theorem 7.4 below. Nevertheless we formulate and prove it here, because its proof is much simpler than that of Theorem 7.4. We recall that k denotes a number field, $\Delta(k)$ the discriminant of k and $O(k)$ the ring of algebraic integers in k.

Theorem 4.3. *Let be* $n\geqq 2$, $G=\mathrm{GL}(n,O(k))$ *and assume that k is a Galois extension of* $\mathbb{Q}$ *with abelian Galois group. Then there exists a torsion-free normal subgroup N of G of finite index with the following properties: The p-Sylow subgroup of G/N is cyclic or order p and the p-period of G/N is* $2(p-1)$ *for all prime numbers p with* $\frac{n}{2}+1<p\leqq n+1$ *and* $p\nmid\Delta(k)$*; moreover, given* $c\in\mathbb{N}$, *one can choose N such that G/N has no p-torsion for all prime numbers p with* $n+1<p\leqq c$ *and* $p\nmid\Delta(k)$.

Proof. The assumption implies that $k \subset \mathbb{Q}(\xi_r)$ for a certain primitive r-th root of unity ξ_r [8, Theorem V.5.9]. If one chooses r as small as possible (r is then called the conductor of k), then a prime number divides r if and only if it divides $\Delta(k)$. Let M be the following set of prime numbers p:

$$M = \left\{ p \,\middle|\, \frac{n}{2} + 1 < p \leqq c \text{ and } p \nmid \Delta(k) \right\}.$$

For $p \in M$ let $s(p) + p^2\mathbb{Z}$ be a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$. According to the Chinese Remainder Theorem there is a residue class $x + r \cdot \prod_{p \in M} p^2\mathbb{Z}$ such that $x \equiv 1 (\mathrm{mod}\, r)$ and $x \equiv s(p) (\mathrm{mod}\, p^2)$ for all $p \in M$. Since x is relatively prime to $r \cdot \prod_{p \in M} p^2$, Dirichlet's theorem tells us that there exists a prime number $q \in x + r \cdot \prod_{p \in M} p^2\mathbb{Z}$ with $q > 2^d$ (as above $d = [k : \mathbb{Q}]$). It follows that $q + p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p \in M$. Since $q \equiv 1 (\mathrm{mod}\, r)$, $q\mathbb{Z}$ splits completely in $\mathbb{Z}[\xi_r]$; hence $q\mathbb{Z}$ splits completely in $O(k)$. This means $q \cdot O(k) = \prod_{i=1}^{d} Q_i$, where the Q_i are different prime ideals in $O(k)$. Therefore $N(Q_i) = q$ for $1 \leqq i \leqq d$ and the assertion follows from Theorem 4.1.

5. Background Material from Algebraic Number Theory

5.1.

Let R be a Dedekind ring with quotient field k (not necessarily a number field) and P a prime ideal in R. Let k_P be the completion of k with respect to the P-adic valuation and R_P the valuation ring in k_P. Then R_P is a discrete valuation ring with maximal ideal $\hat{P} = P \cdot R_P$ (we consider k as subfield of k_P). Since $P \subset \hat{P}$ and $R \subset R_P$ one has for any $r \in \mathbb{N}$ a ring homomorphism $\mathbf{i}_P : R/P^r \to R_P/\hat{P}^r$, $\mathbf{i}_P(x + P^r) = x + \hat{P}^r$. It is well known that $\mathbf{i}_P$ is an isomorphism [8, Lemma I.3.11, Corollary II.2.7]. We use the following notation: Let I be an ideal in R and $I = \prod_{i=1}^{h} Q_i^{e_i}$ its prime ideal factorization in R. Then we denote by $(R/I)^*$ the group of those residue classes of R modulo I, which are relatively prime to I; i.e.

$$(R/I)^* = \{x + I \mid x \in R,\ x \notin Q_i \text{ for } 1 \leqq i \leqq h\}.$$

It is easy to see that the restriction of the above isomorphism $\mathbf{i}_P$ induces an isomorphism $(R/P^r)^* \to (R_P/\hat{P}^r)^*$. We summarize:

Lemma 5.1. *With the notation above, the canonical map $\mathbf{i}_P : R/P^r \to R_P/\hat{P}^r$ is an isomorphism, which maps $(R/P^r)^*$ onto $(R_P/\hat{P}^r)^*$.*

5.2.

We consider the following situation: let $O(k)$ be the ring of algebraic integers in a number field k, p a prime number and $p \cdot O(k) = \prod_{i=1}^{h} P_i^{e_i}$, where the P_i are different prime ideals in $O(k)$. The relative degree of P_i over $p\mathbb{Z}$ is denoted by f_i; i.e. $f_i = [O(k)/P_i : \mathbb{Z}/p\mathbb{Z}]$. We choose an Index i, $1 \leqq i \leqq h$, and write for short $P = P_i$, $e = e_i$, $f = f_i$. We recall the following properties of the completions:

(i) $O(k)_P$ is the integral closure of $\mathbb{Z}_p$ in k_p;

(ii) $\hat{p} \cdot O(k)_P = \hat{P}^e$;

(iii) $[O(k)_P/\hat{P}:\mathbb{Z}_p/\hat{p}]=f$;

(iv) $[k_P:\mathbb{Q}_p]=e\cdot f$, where $\mathbb{Z}_p$ denotes the valuation ring of the p-adic completion $\mathbb{Q}_p$ and $\hat{p}$ the maximal ideal in $\mathbb{Z}_p$.

We consider the norm map $N_P=N_{k_P/\mathbb{Q}_p}:k_P\to\mathbb{Q}_p$. Because of property (i) above, one has $N_P(O(k)_P)\subset\mathbb{Z}_p$; and an easy calculation using property (ii) shows that N_P induces for every $r\in\mathbb{N}$ a map

$$\mathbf{N}_P:O(k)_P/\hat{P}^{re}\to\mathbb{Z}_p/\hat{p}^r,\qquad \mathbf{N}_P(x+\hat{P}^{re})=N_P(x)+\hat{p}^r.$$

The situation considered in 5.1 [for $R=O(k)$] and 5.2 may be illustrated by the following commutative diagram:

$$(*)\qquad \begin{array}{ccccc} O(k) & \xrightarrow{i_P} & O(k)_P & \xrightarrow{N_P} & \mathbb{Z}_p \\ {\scriptstyle\pi_P}\downarrow & & {\scriptstyle\hat{\pi}_P}\downarrow & & \downarrow{\scriptstyle\hat{\pi}_p} \\ O(k)/P^{re} & \xrightarrow[\mathbf{i}_P]{\cong} & O(k)_P/\hat{P}^{re} & \xrightarrow[\mathbf{N}_P]{} & \mathbb{Z}_p/\hat{p}^r \end{array}$$

where the vertical maps are the projections, and i_P the inclusion.

6. Elements Whose Norm Lie in a Prescribed Residue Class

Let as before k be a number field with discriminant $\Delta(k)$, $O(k)$ the ring of algebraic integers in k and $[k:\mathbb{Q}]=d$, and $N=N_{k/\mathbb{Q}}:k\to\mathbb{Q}$ the norm map.

Definition 6.1. We denote by $\delta(k)$ the product of those prime numbers p for which $p\cdot O(k)$ has only ramified (prime ideal) factors.

Since $p\cdot O(k)$ has a ramified factor if and only if $p|\Delta(k)$, it follows:

Remark 6.2. $\delta(k)|\Delta(k)$.

If k is Galois over $\mathbb{Q}$, then $\delta(k)$ and $\Delta(k)$ contain the same prime factors. Otherwise it is possible that $\delta(k)$ contains fewer different prime numbers than $\Delta(k)$ (see [4, Example 5.3]).

Lemma 6.3. *Let p be a prime number not dividing $\delta(k)$ and $s\in\mathbb{Z}$ be relatively prime to p. Then there exists, for any $r\in\mathbb{N}$, a residue class $C\in(O(k)/p^r\cdot O(k))^*$ such that $N(x)\equiv s(\mathrm{mod}\,p^r)$ for all $x\in C$.*

Proof. Let be $p\cdot O(k)=\prod_{i=1}^{h}P_i^{e_i}$, the prime ideal factorization of $p\cdot O(k)$ in $O(k)$. Since $p\nmid\delta(k)$, one of the ramification indices e_i is 1; say $e_1=1$. According to Lemma 6.4 below there exists a unit $y\in O(k)_{P_1}$ such that $N_{P_1}(y)=s$; clearly $y+P_1^r\in(O(k)_{P_1}/\hat{P}_1^r)^*$. Lemma 5.1 shows that there is a residue class $z+P_1^r\in(O(k)/P_1^r)^*$ such that $\mathbf{i}_{P_1}(z+P_1^r)=y+\hat{P}_1^r$. The Chinese Remainder Theorem tells us that

$$\eta:O(k)/p^r\cdot O(k)\to\prod_{i=1}^{h}O(k)/P_i^{re_i},$$

$$\eta(x+p^r\cdot O(k))=(x+P_1^r,x+P_2^{re_2},\ldots,x+P_h^{re_h})$$

is an isomorphism, which maps $(O(k)/p^r\cdot O(k))^*$ onto $\prod_{i=1}^{h}(O(k)/P_i^{re_i})^*$. Let be

$$C=\eta^{-1}(z+P_1^r,1+P_2^{re_2},\ldots,1+P_h^{re_h}).$$

Then $C \in (O(k)/p^r \cdot O(k))^*$, and for any $x \in C$ one gets using diagram (*) of Sect. 5:

$$\hat{\pi}_p N(x) = \prod_{i=1}^{h} \hat{\pi}_p N_{P_i}(x) = \prod_{i=1}^{h} \mathbf{N}_{P_i} \mathbf{i}_{P_i} \pi_{P_i}(x) = s + \hat{p}^r,$$

since

$$\mathbf{N}_{P_1} \mathbf{i}_{P_1} \pi_{P_1}(x) = \mathbf{N}_{P_1}(y + \hat{P}_1^r) = s + \hat{p}^r$$

and

$$\mathbf{N}_{P_i} \mathbf{i}_{P_i} \pi_{P_i}(x) = \mathbf{N}_{P_i} \mathbf{i}_{P_1}(1 + P_i^{re_i}) = 1 + \hat{p}^r$$

for $2 \leqq i \leqq h$. Taking into account the commutative diagram

$$\begin{array}{ccc} \mathbb{Z} & \xrightarrow{i_p} & \mathbb{Z}_p \\ {\scriptstyle \pi_p}\downarrow & & \downarrow{\scriptstyle \hat{\pi}_p} \\ \mathbb{Z}/p^r\mathbb{Z} & \xrightarrow[\mathbf{i}_p]{\cong} & \mathbb{Z}/\hat{p}^r \end{array}$$

one concludes $\mathbf{i}_p \pi_p N(x) = s + \hat{p}^r$. Since $\mathbf{i}_p \pi_p(s) = s + \hat{p}^r$ and since $\mathbf{i}_p$ is injective, it follows that $N(x) \equiv s (\mathrm{mod}\, p^r)$.

We now formulate and prove the lemma used above

Lemma 6.4. *Let P be an unramified prime ideal in $O(k)$ containing the prime number p and let $s \in \mathbb{Z}$ be relatively prime to p. Then there exists a unit $y \in O(k)_P$ such that $N_P(y) = s$.*

Proof. Since P is unramified, k_P is an unramified extension of $\mathbb{Q}_P$ [see properties (ii)–(iv) in 5.2]. Moreover s is a unit in $\mathbb{Z}_p$, because s is relatively prime to p. In this situation a theorem [1, p. 131] tells us that there is a unit $y \in O(k)_P$ such that $N_P(y) = s$.

The assertion of Lemma 6.3 is also valid under the following assumption about the dimension $d = [k : \mathbb{Q}]$:

Lemma 6.5. *Let p be a prime number with $gcd(d, p(p-1)) = 1$, and $s \in \mathbb{Z}$ relatively prime to p. Then there exists a residue class $C \in (O(k)/p^r \cdot O(k))^*$ such that $N(x) \equiv s (\mathrm{mod}\, p^r)$ for all $x \in C$.*

Proof. We consider the homomorphism $\mu : (\mathbb{Z}/p^r\mathbb{Z})^* \to (\mathbb{Z}/p^r\mathbb{Z})^*$, $\mu([t]) = [t]^d$. The group $(\mathbb{Z}/p^r\mathbb{Z})^*$ has order $\varphi(p^r) = p^{r-1}(p-1)$ (φ = Euler function). The assumption implies that $gcd(d, \varphi(p^r)) = 1$, hence μ is an isomorphism. It follows that there is a residue class $t + p^r\mathbb{Z} \in (\mathbb{Z}/p^r\mathbb{Z})^*$ such that

$$\mu(t + p^r\mathbb{Z}) = s + p^r\mathbb{Z}.$$

One has $N(t) = t^d \equiv s (\mathrm{mod}\, p^r)$. Let be $C = t + p^r \cdot O(k)$. Then $C \in (O(k)/p^r O(k))^*$ and it is easy to see that for $x \in C$ one has $N(x) \equiv N(t) (\mathrm{mod}\, p^r)$.

Using the Chinese Remainder Theorem one gets from the Lemmata 6.3 and 6.5:

Theorem 6.6. *Let M be a finite set of prime numbers p with $p \nmid \delta(k)$ or $gcd(d, p(p-1)) = 1$. For $p \in M$ let $s(p) \in \mathbb{Z}$ be relatively prime to p. Then there is a residue class $C \in (O(k) / \prod_{p \in M} p^r \cdot O(k))^*$ such that $N(x) \equiv s(p) (\mathrm{mod}\, p^r)$ for all $x \in C$ and all $p \in M$.*

We add the following remark: let I be a non-trivial ideal in $O(k)$ and C any residue class of $O(k)$ modulo I. Then there are infinitely many $x \in C$ with $N(x) > 0$. For choose an element $y \in I$ with $N(y) > 0$ and $x_0 \in C$. Then $x_0 + ny \in C$ for all $n \in \mathbb{Z}$ and $N(x_0 + ny) \to \infty$ for $n \to \infty$.

7. Prime Ideals Whose Norm Lie in a Prescribed Residue Class

7.1.

We use the following terminology (see [8]): A prime p of a number field k is an equivalence class of valuations on k. A prime is called infinite, if it contains an archimedian valuation; otherwise it is called finite. A finite prime may therefore be identified with a prime ideal in $O(k)$. An infinite prime p is called real, if the completion k_p is the field $\mathbb{R}$; the other infinite primes are called complex.

A modulus for k is a formal product $m = \prod_p p^{r(p)}$ taken over all primes of k in which $r(p)$ is a nonnegative integer and $r(p) > 0$ for only a finite number of p. Furthermore $r(p) \leqq 1$ if p is a real prime, and $r(p) = 0$ if p is complex. One may consider a modulus m as product $m_0 m_\infty$, where m_0 is the product of the finite primes and m_∞ the product of infinite primes appearing with positive exponent in m.

We recall: for a real prime p of k let $\sigma : k \to k_p$ be the canonical embedding of k into the real field k_p. Then for $x, y \in k^*$ the congruence $x \equiv y (\mathrm{mod}\, p)$ means $\sigma\left(\frac{x}{y}\right) > 0$. For a finite prime p the congruence $x \equiv y (\mathrm{mod}\, p^r)$ means $\frac{x}{y} \in O(k)_{(p)}$ [= localization of $O(k)$ in p] and $\frac{x}{y} - 1 \in p^r \cdot O(k)_{(p)}$. One writes $x \equiv y (\mathrm{mod}\, m)$ for a modulus $m = \prod_p p^{r(p)}$, if $x \equiv y (\mathrm{mod}\, p^{r(p)})$ for all primes p with $r(p) > 0$.

Furthermore we use the notation (m a modulus): $k_m = \left\{ \frac{x}{y} \middle| x,\ y \in O(k),\ x \cdot O(k) \right.$ and $\left. y \cdot O(k)$ relatively prime to $m_0 \right\}$

$$k_{m,1} = \{ y \in k_m | y \equiv 1 (\mathrm{mod}\, m) \}.$$

There is a canonical map $k_{m,1} \subset k_m \xrightarrow{i} I^m$, $i(x) = x \cdot O(k)$. Here I^m denotes the part of the ideal group generated by all prime ideals not dividing m_0. The cosets of I^m modulo $i(k_{m,1})$ are called ray classes mod m.

Lemma 7.1. *Let be $h \in \mathbb{N}$ and $x \in O(k)$ such that $x \cdot O(k)$ is relatively prime to $h \cdot O(k)$ and $N(x) > 0$. Let J be an ideal of $O(k)$ lying in the ray class $i(x) \cdot i(k_{m,1})$, where $m = m_0 m_\infty$, $m_0 = h \cdot O(k)$ and $m_\infty =$ product of all the real primes of k. Then $N(J) \equiv N(x) (\mathrm{mod}\, h)$.*

Proof. We remark that $i(x) \in I^m$, because $x \cdot O(k)$ is relatively prime to $h \cdot O(k)$.

Since $J \in i(x) \cdot i(k_{m,1})$, there exists $y \in k_{m,1}$ such that $J = x \cdot y \cdot O(k)$. Then $xy \equiv x (\mathrm{mod}\, m_0)$. Let be $\sigma_i : k \to \bar{\mathbb{Q}}$ $(1 \leqq i \leqq d = [k : \mathbb{Q}])$ the embeddings of k in the algebraic closure $\bar{\mathbb{Q}}$ of $\mathbb{Q}$. Assume that $\sigma_1, \ldots, \sigma_u$ are the real embeddings and

$\sigma_{u+1}, \ldots, \sigma_{u+2v=d}$ the complex one, where $\sigma_{u+v+j} = \bar{\sigma}_{u+j}$ for $1 \leqq j \leqq v$. The field $L = \prod_{i=1}^{d} \sigma_i(k)$ is a finite dimensional Galois extension of $\mathbb{Q}$ and $\sigma_i(O(k)) \subset O(L)$ for $1 \leqq i \leqq d$. It follows that $\sigma_i(xy) \equiv \sigma_i(x) \pmod{h \cdot O(L)}$ for $1 \leqq i \leqq d$ and therefore $N(xy) \equiv N(x) \pmod{h \cdot O(L)}$, which implies $N(xy) \equiv N(x) \pmod{h}$. Since $J = xy \cdot O(k)$, one has that $N(J) = |N(xy)|$. But

$$N(xy) = N(x) \cdot N(y) > 0.$$

For $N(x) > 0$ according to the assumption and

$$N(y) = \prod_{i=1}^{u} \sigma_i(y) \cdot \prod_{i=u+1}^{u+v} \sigma_i(y) \cdot \bar{\sigma}_i(y) > 0$$

[the first product is positive, because $y \equiv 1 \pmod{m_\infty}$]. It follows that

$$N(J) = N(xy) \equiv N(x) \pmod{h}.$$

Theorem 7.2. *Let k be an algebraic number field and $d = [k : \mathbb{Q}]$. Let M be a finite set of prime numbers p with $p \nmid \delta(k)$ or $\gcd(d, p(p-1)) = 1$. Then there exists a principal prime ideal P of $O(k)$ such that $N(P)$ is a prime number $> 2^d$ and $N(P) + p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p \in M$.*

Proof. We choose for every $p \in M$ an integer $s(p)$ such that $s(p) + p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$. According to Theorem 6.6 (for $r = 2$) there is a residue class $C \in (O(k) / \prod_{p \in M} p^2 \cdot O(k))^*$ such that $N(x) \equiv s(p) \pmod{p^2}$ for all $x \in C$ and all $p \in M$. Let m be the modulus $m = m_0 m_\infty$, where $m_0 = \prod_{p \in M} p^2 \cdot O(k)$ and $m_\infty =$ product of all real primes of k. We choose $x \in C$ such that $N(x) > 0$ (cf. the remark following Theorem 6.6). Then $x \cdot O(k)$ is relatively prime to $\prod_{p \in M} p^2 \cdot O(k)$. The set of prime ideals of $O(k)$ which lie in the ray class $H = i(x) \cdot i(k_{m,1})$ has (Dirichlet) density $[I^m : i(k_{m,1})]^{-1}$ (see [8, Theorem V.10.3]). It follows that there exist infinitely many prime ideals $P \in H$ of relative degree one (over $P \cap \mathbb{Z}$) with $N(P) > 2^d$. Then P is a principal ideal, and $N(P)$ is a prime number. Using Lemma 7.1 one gets

$$N(P) \equiv N(x) \equiv s(p) \pmod{p^2},$$

hence $N(P) + p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p \in M$.

Remark. Using the Tchebotarev Density Theorem there is an easy proof of the following weaker version of Theorem 7.2: If $p \nmid \delta(k)$, then there exists a prime ideal P of $O(k)$ such that $N(P)$ is a prime number $> 2^d$ and $N(P) + p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$.

The theorem on $\mathrm{GL}(n, O(k))$ stated in the introduction now follows by combining Theorems 7.2 and 4.1. Note the assertion about prime numbers made in 6.2. For $\mathrm{SL}(n, O(k))$ we need Remark 4.2 as well.

7.2.

We consider now a more general type of rings R: Let S be a finite set of finite primes of k. Put $O_S(k) = \bigcap_{p \notin S} O(k)_{(p)}$, where the intersection is taken over all finite

primes not in S [and $O(k)_{(p)}$ = valuation ring of p]. The ring $O_S(k)$ is a Dedekind ring with quotient field k. Note that if $S=\emptyset$, then $O_S(k)=O(k)$. We now generalize Theorem 7.2.

Lemma 7.3. *Let S be a finite set of finite primes of k and M a finite set of prime numbers p with $p\nmid\delta(k)$ or $gcd(d, p(p-1))=1$. Then there exists a prime ideal P of $O_S(k)$ such that $N(P)$ is a prime number $>2^d$ and $N(P)+p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p\in M$.*

Proof. The proof of Theorem 7.2 shows that there exists infinitely many prime ideals Q of $O(k)$ such that $N(Q)$ is a prime number $>2^d$ and $N(Q)+p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p\in M$. But almost all of these (i.e. all except finitely many) generate a prime ideal $P=Q\cdot O_S(k)$ in $O_S(k)$, and

$$N(P)=|O_S(k)/Q\cdot O_S(k)|=|O(k)/Q|=N(Q).$$

Theorem 4.1 and Lemma 7.3 yield

Theorem 7.4. *Let be $n\geqq 2$, k a number field, $d=[k:\mathbb{Q}]$, and let G be the S-arithmetic group $G=\mathrm{GL}(n, O_S(k))$. Then there exists a torsion-free normal subgroup N of G of finite index with the following properties: The p-Sylow subgroup of G/N is cyclic of order p and the p-period of G/N is $2(p-1)$ for all prime numbers p with $\frac{n}{2}+1<p\leqq n+1$ and ($p\nmid\delta(k)$ or $gcd(d, p(p-1))=1$); moreover, given $c\in\mathbb{N}$, one can choose N such that G/N has no p-torsion for all prime numbers p with $n+1<p\leqq c$ and ($p\nmid\delta(k)$ or $gcd(d, p(p-1))=1$).*

For this normal subgroup N of G one can take the principal congruence subgroup $\mathrm{GL}(n, O_S(k), P)$, where P is a certain principal prime ideal of $O_S(k)$ (see the proofs of Theorem 4.1 and Lemma 7.3).

We recall that $G=\mathrm{GL}(n, O_S(k))$ is virtually of type FP [9, Theorem 4]. Therefore the result above together with the theorem quoted in the introduction (1.2) yields information about the p-periodicity of $\mathrm{GL}(n, O_S(k))$ (cf. [4, Theorem 3.4]).

8. Number Fields of Dimension 3

Theorem 8.1. *Let k be a number field of dimension 3 (over $\mathbb{Q}$) such that $\Delta(k)$ is not a square in $\mathbb{Z}$. Let be $n\geqq 2$, p a prime number with $\frac{n}{2}+1<p\leqq n+1$ and $G=\mathrm{GL}(n, O(k))$. Then there exists a torsion-free normal subgroup N of G of finite index, such that the p-Sylow subgroup of G/N is cyclic of order p and the p-period of G/N is $2(p-1)$.*

We will give a proof of this assertion without using the results of Sects. 5–7. Nevertheless we note that for prime numbers $p\nmid\delta(k)$ this theorem is a consequence of Theorem 7.4. We need the following lemma, in which $(-)$ denotes the Legendre symbol, as used in the theory of quadratic residues.

Lemma 8.2. *Let p_0 be a prime number >2 and $s+p_0^2\mathbb{Z}$ a generator in $(\mathbb{Z}/p_0^2\mathbb{Z})^*$. If $u\in\mathbb{Z}$ is not a square, then there exist infinitely many prime numbers q such that $q\equiv s(\mathrm{mod}\, p_0^2)$ and $\left(\frac{u}{q}\right)=-1$.*

Proof. This is a straightforward calculation using the Chinese Remainder Theorem, the multiplicativity of the Legendre symbol and the law of quadratic reciprocity.

Proof of Theorem 8.1. There is an algebraic integer θ such that $k=\mathbb{Q}(\theta)$. Let $f(x)\in\mathbb{Z}[x]$ be the irreducible polynomial of θ over $\mathbb{Q}$, and $\theta_1=\theta$, θ_2, θ_3 the (complex) roots of $f(x)$. The discriminants satisfie

$$\Delta(f)=\prod_{1\leqq i<j\leqq 3}(\theta_i-\theta_j)^2=h^2\cdot\Delta(k) \text{ for a certain } h\in\mathbb{Z}.$$

Since $\Delta(k)$ is not a square, nor is $\Delta(f)$ a square. According to Lemma 8.2 there is a prime number $q>8$ such that $q\nmid\Delta(f)$ and $\left(\frac{\Delta(f)}{q}\right)=-1$ and $q+p^2\mathbb{Z}$ generates $(\mathbb{Z}/p^2\mathbb{Z})^*$. Hence $q\cdot O(k)=\prod_{i=1}^{r}P_i$, where the P_i are different prime ideals in $O(k)$ and $1\leqq r\leqq 3$. Reducing f modulo q one gets a polynomial $[f]\in\mathbb{Z}/q\mathbb{Z}[x]$. Since $q\nmid\Delta(f)$, it follows that $[f](x)=\prod_{i=1}^{r}f_i(x)$, where the $f_i(x)$ are different irreducible polynomials in $\mathbb{Z}/q\mathbb{Z}[x]$ (with leading coefficient one). According to Stickelberger's theorem, $r\equiv 1(\mathrm{mod}\,2)$ if and only if $\Delta([f])$ is a square in $\mathbb{Z}/q\mathbb{Z}$. But $\Delta([f])$ is not a square, because $\left(\frac{\Delta(f)}{q}\right)=-1$. Hence $r=2$ and $q\cdot O(k)=P_1\cdot P_2$. Therefore P_1 or P_2 has relative degree one over $q\mathbb{Z}$, and the assertion follows now from Theorem 4.1.

From the proof it is clear that Theorem 8.1 is valid also for the S-arithmetic groups $\mathrm{GL}(n, O_S(k))$ (cf. Sect. 7.2).

Concerning the assumption in Theorem 8.1 we remark that the discriminant $\Delta(k)$ of a number field k of dimension 3 is a square in $\mathbb{Z}$ if and only if k is Galois over $\mathbb{Q}$.

9. $k=\mathbb{Q}(\theta)$, θ a Root of x^r-a

Theorem 9.1. *Let θ be a root of $f(x)=x^r-a$, where $r\in\mathbb{N}$ and $a\in\mathbb{Z}$; assume that r is odd and $f(x)$ irreducible over $\mathbb{Q}$. Let be $n\geqq 2$ and $G=\mathrm{GL}(n, O(\mathbb{Q}(\theta)))$. Then there exists a torsion-free normal subgroup N of G of finite index with the following properties: The p-Sylow subgroup of G/N is cyclic of order p and the p-period of G/N is $2(p-1)$ for all prime numbers p with $\frac{n}{2}+1<p\leqq n+1$; moreover, given $c\in\mathbb{N}$, one can choose N such that G/N has no p-torsion for all prime numbers p with $n+1<p\leqq c$.*

Proof. Let M and M' be the following sets of prime numbers:

$$M=\left\{p\,\middle|\,\frac{n}{2}+1<p\leqq c\right\} \quad\text{and}\quad M'=\{p|p|r\}\backslash M.$$

For every $p \in M$ choose a generator $s(p)+p^2\mathbb{Z}$ of $(\mathbb{Z}/p^2\mathbb{Z})^*$. According to the Chinese Remainder Theorem there is a residue class $x+\prod_{p\in M} p^2 \cdot \prod_{p\in M'} p\cdot\mathbb{Z}$ such that $x \equiv s(p) (\mathrm{mod}\, p^2)$ for all $p \in M$ and $x \equiv 2 (\mathrm{mod}\, p)$ for all $p \in M'$. Since r is odd, x is relatively prime to $\prod_{p\in M} p^2 \cdot \prod_{p\in M'} p$. Dirichlet's theorem shows now that there exists a prime number $q > 2^r$ which lie in $x+\prod_{p\in M} p^2 \cdot \prod_{p\in M'} p\cdot\mathbb{Z}$ and which does not divide $\Delta(f) = \pm r^r a^{r-1}$. Then $q+p^2\mathbb{Z}$ is a generator in $(\mathbb{Z}/p^2\mathbb{Z})^*$ for all $p \in M$. We show now that $\mu : (\mathbb{Z}/q\mathbb{Z})^* \to (\mathbb{Z}/q\mathbb{Z})^*$, $\mu(c) = c^r$, is an isomorphism. Assume that there is an element $c \in \ker \mu$ of prime order p; then $p|r$ and $p|q-1$. If $p \in M'$, then $q \equiv 2 (\mathrm{mod}\, p)$, hence $p \nmid q-1$, a contradiction. If $p \in M$, then $q^p = ((q-1)+1)^p \equiv 1 (\mathrm{mod}\, p^2)$, which contradicts $q \equiv s(p) (\mathrm{mod}\, p^2)$. We have therefore proved $\ker \mu = \{1\}$. It follows that $[f](x)$ has a root in $\mathbb{Z}/q\mathbb{Z}$, where $[f](x)$ denotes the polynomial obtained by reducing the coefficients of $f(x) = x^r - a$ modulo q. This implies, since $q \nmid \Delta(f)$, that $q \cdot O(k)$ is divided by a prime ideal P of $O(k)$ of relative degree one over $q\mathbb{Z}$. Hence $N(P) = q$ and the assertion follows from Theorem 4.1.

From this proof it is clear that the S-arithmetic version of Theorem 9.1 is also true (cf. Sect. 7.2).

Acknowledgement. The present paper grew out of certain considerations in my Doctoral Thesis [3] written under the direction of Prof. B. Eckmann. I would like to thank him for his stimulating interest and many helpful conversations.

References

1. Artin, E.: Algebraic numbers and algebraic functions. I. Princeton University, New York University 1950–1951
2. Bass, H.: Algebraic k-theory. New York, Amsterdam: Benjamin 1968
3. Bürgisser, B.: Gruppen virtuell endlicher Dimension und Periodizität der Cohomologie. Diss. ETH, Nr. 6425 (1979)
4. Bürgisser, B.: On the p-periodicity of arithmetic subgroups of general linear groups. Comment. Math. Helv. (to appear)
5. Cartan, H., Eilenberg, S.: Homological algebra. Princeton: Princeton University Press 1956
6. Carter, D.W.: Projective module groups of $SL_n(\mathbb{Z})$ and $GL_n(\mathbb{Z})$. J. Pure Appl. Algebra **16**, 49–54 (1980)
7. Farrell, F.T.: An extension of Tate cohomology to a class of infinite groups. J. Pure Appl. Algebra **10**, 153–161 (1977)
8. Janusz, G.J.: Algebraic number fields. Pure and Applied Mathematics, Vol. 55. New York: Academic Press 1973
9. Serre, J.P.: Cohomologie des groupes discrets. Prospects in Math. Ann. Math. Study **70**, 77–169 (1971)
10. Serre, J.P.: Arithmetic groups. Proceedings of a symposium, held at Durham in September 1977, on homological and combinatorial techniques in group theory. Cambridge: Cambridge University Press 1979
11. Swan, R.G.: The p-period of a finite group. Illinois J. Math. **4**, 341–346 (1960)

Received October 28, 1980; in revised form January 6, 1981

Math. Ann. 256, 133–143 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Die Modulräume stabiler 3-Bündel auf $\mathbb{P}_3$ mit den Chernklassen $c_1=0$, $c_3=c_2^2-c_2$

Heinz Spindler

Mathematisches Institut der Universität, Bunsenstraße 3–5, D-3400 Göttingen, Bundesrepublik Deutschland

Einleitung

In [3] ist gezeigt worden, daß die zweite und dritte Chernklasse stabiler holomorpher Vektorbündel E vom Rang 3 auf dem komplex-projektiven Raum den Ungleichungen $c_2(E)\geqq 2$, $c_3(E)\leqq c_2(E)^2-c_2(E)$ genügen, wenn $c_1(E)=0$ gilt. In dieser Note wird gezeigt, daß die Modulräume $\mathbf{M}_n$ der stabilen 3-Bündel auf $\mathbb{P}_3$ mit den extremen Chernklassen $c_1=0$, $c_2=n$, $c_3=n^2-n$ ($n\geqq 2$) glatt und irreduzibel sind. Im Fall $n\neq 3$ wird auch die Rationalität von $\mathbf{M}_n$ bewiesen.

1

Es sei E ein stabiles holomorphes Vektorbündel vom Rang 3 auf $\mathbb{P}_3$ mit den Chernklassen $c_i=c_i(E)$. Es sei $c_1=0$. Nach [3] gilt dann $c_2\geqq 2$.

Die Riemann-Rochsche Formel für E lautet in diesem Fall

$$\chi(E)=\tfrac{1}{2}c_3-2c_2+3.$$

Da E stabil ist, gilt zum anderen

$$\chi(E)=h^2(E)-h^1(E)\leqq h^2(E).$$

Weiter gilt für allgemeine Ebenen $H\subset\mathbb{P}_3$ die Abschätzung

$$h^2(E)\leqq\binom{h^1(E|H)}{2}$$

(vgl. Elencwajg u. Forster [1] und Schneider [2]).

Ist H eine allgemeine Ebene in $\mathbb{P}_3$, so ist $h^2(E|H)=0$, und nach der Riemann-Rochschen Formel für $E|H$ gilt dann

$$h^1(E|H)=c_2-3+h^0(E|H).$$

In [3] wurde nun gezeigt, daß für allgemeine Ebenen $H\subset\mathbb{P}_3$ $h^0(E|H)\leqq 1$ gilt. Es folgt damit insgesamt die wichtige Ungleichung

$$\chi(E)\leqq h^2(E)\leqq\binom{c_2-2}{2}, \tag{1}$$

0025-5831/81/0256/0133/$02.20

d. h. insbesondere $c_3 \leqq c_2^2 - c_2$. Ist sogar $h^0(E|H)=0$ für allgemeine Ebenen in $\mathbb{P}_3$, so erhält man

$$c_2 \geqq 3 \quad \text{und} \quad \chi(E) \leqq h^2(E) \leqq \binom{c_2-3}{2}. \tag{2}$$

Genau im Fall $c_2=3$ führt dies zu derselben Abschätzung wie (1), während für $c_2 \geqq 4$ immer $\binom{c_2-3}{2} < \binom{c_2-2}{2}$ gilt. Aus (1) und (2) ergibt sich insbesondere, daß alle stabilen 3-Bündel auf $\mathbb{P}_3$ mit $c_1=0$, $c_2=2$ oder $c_2 \geqq 4$ und

$$\binom{c_2-3}{2} < \chi(E) \leqq \binom{c_2-2}{2}$$

auf allen Ebenen in $\mathbb{P}_3$ instabil sind.

Besonders einfach ist nun die Klassifikation derjenigen stabilen 3-Bündel E auf $\mathbb{P}_3$ mit $c_1=0$ und festem $c_2=n$, für die die Euler-Poincaré-Charakteristik $\chi(E)$ maximal wird, für die also

$$\chi(E) = \binom{n-2}{2} \quad (\text{d. h. } c_3 = n^2 - n)$$

gilt. Nach (1) gilt für solche Bündel

$$h^0(E)=0, \quad h^1(E)=0, \quad h^2(E)=\binom{n-2}{2}, \quad h^3(E)=0.$$

Wir setzen nun weiter voraus, daß für allgemeine Ebenen H in $\mathbb{P}_3$ die Gleichung

$$h^0(E|H)=1$$

gilt, was nach (2) im Fall $n \neq 3$ automatisch erfüllt ist. Aus der exakten Sequenz

$$0 = H^0(E) \to H^0(E|H) \to H^1(E(-1)) \to H^1(E) = 0$$

folgt dann $h^1(E(-1))=1$. Hieraus entnimmt man, daß $h^0(E|H)=1$ für alle Ebenen $H \subset \mathbb{P}_3$ gilt.

Da E stabil ist, ergibt sich mit Hilfe der Riemann-Rochschen Formel die folgende Tabelle für die Dimensionen $h^q(E(p))$, $0 \leqq q \leqq 3$, $-3 \leqq p \leqq 0$:

$h^q(E(p))$

0	0	0	0
$\binom{n+1}{2}+\delta$	$\binom{n}{2}+\varepsilon$	$\binom{n-1}{2}$	$\binom{n-2}{2}$
δ	ε	1	0
0	0	0	0

(q ↑, p →) (3)

$(\delta \leqq \varepsilon \leqq 1)$

Nach dem Satz von Beilinson (vgl. [4]) gibt es nun eine Spektralsequenz

$$E_1^{pq} = H^q(E(p)) \otimes \Omega^{-p}_{\mathbb{P}_3}(-p) \Rightarrow E^{p+q} = \begin{cases} E & \text{für } p+q=0 \\ 0 & \text{sonst.} \end{cases}$$

Es ist also $E_\infty^{pq}=0$ für $p+q \neq 0$ und $E_\infty^{p,-p} = F^p/F^{p+1}$, wobei

$$0 = F^1 \subset F^0 \subset F^{-1} \subset F^{-2} \subset F^{-3} = E$$

eine Filtrierung von E durch kohärente Garben F^i ist. Ist auch $E_\infty^{0,0}=0$ und $E_\infty^{-3,3}=0$, so bekommt man E als Extension

$$0 \to E_\infty^{-1,1} \to E \to E_\infty^{-2,2} \to 0.$$

Mit Hilfe dieser Spektralsequenz zeigen wir zunächst

Satz 1. *Es sei $c_2 = n \geqq 2$. Ist E ein stabiles 3-Bündel auf $\mathbb{P}_3$ mit $c_1=0$, $c_3 = c_2^2 - c_2$, so gilt*

$$h^1(E(-2)) = 0.$$

Beweis. Ist $n=3$ und $h^0(E|H)=0$ für allgemeine Ebenen H in $\mathbb{P}_3$, so folgt leicht

$$h^1(E(-2)) = h^1(E(-1)) = h^0(E|H) = 0.$$

Sei jetzt also $h^0(E|H)=1$ für alle Ebenen $H \subset \mathbb{P}_3$. Es ist nur zu zeigen, daß in der Tabelle (3) die Fälle a) $\delta=\varepsilon=1$ und b) $\delta=0$, $\varepsilon=1$ unmöglich sind.

zu a): Annahme: $\delta=\varepsilon=1$. Man erhält folgendes E_1-Diagramm:

0	0	0	0
$\mathcal{O}(-1)^{\oplus\binom{n+1}{2}+1}$	$T(-2)^{\oplus\binom{n}{2}+1}$	$\Omega(1)^{\oplus\binom{n-1}{2}}$	$\mathcal{O}^{\oplus\binom{n-2}{2}}$
$\mathcal{O}(-1)$	$T(-2)$	$\Omega(1)$	0
0	0	0	0

q ↑, p →

Und hieraus als E_2-Diagramm

0	0	0	0
$\ker d_1^{-3,2}$	$\dfrac{\ker d_1^{-2,2}}{\operatorname{Im} d_1^{-3,2}}$	$\dfrac{\ker d_1^{-1,2}}{\operatorname{Im} d_1^{-2,2}}$	$\operatorname{coker} d_1^{-1,2}$
$\ker d_1^{-3,1}$	$\dfrac{\ker d_1^{-2,1}}{\operatorname{Im} d_1^{-3,1}}$	$\operatorname{coker} d_1^{-2,1}$	0
0	0	0	0

q ↑, p →

Es folgt $E_\infty = E_3$. Nach dem Satz von Beilinson ist

$$E_\infty^{-3,2} = \ker d_2^{-3,2} = 0, \quad E_\infty^{-3,1} = \ker d_1^{-3,1} = 0,$$

$$E_\infty^{-2,1} = \frac{\ker d_1^{-2,1}}{\operatorname{Im} d_1^{-3,1}} = 0$$

und E als Extension

$$0 \to E_\infty^{-1,1} \to E \to E_\infty^{-2,2} \to 0$$

gegeben. Da

$$0 \to \mathcal{O}(-1) \xrightarrow{d_1^{-3,1}} T(-2) \xrightarrow{d_1^{-2,1}} \Omega(1) \to E_2^{-1,1} \to 0 \qquad (*)$$

exakt ist, gilt $\operatorname{rg}(E_2^{-1,1}) = 1$.

Da $E_\infty^{-1,1} = E_2^{-1,1}/\operatorname{Im} d_2^{-3,2}$ als Untergarbe von E torsionsfrei ist, und weil $d_2^{-3,2}$ ein Monomorphismus ist, kann nur

$$E_2^{-3,2} = \ker d_1^{-3,2} = 0 \quad \text{oder} \quad \cong E_2^{-1,1}$$

gelten.

Im ersten Fall, wäre $E_\infty^{-1,1} = E_2^{-1,1}$, also $E_2^{-1,1} \subset E$. Da E stabil ist, müßte $c_1(E_2^{-1,1}) < 0$ gelten im Widerspruch zu $(*)$. Im zweiten Fall wäre $E_2^{-1,1}$ eine Untergarbe von $E_1^{-3,2}$, also wieder $c_1(E_2^{-1,1}) < 0$ im Widerspruch zu $(*)$. Der Fall a) ist also unmöglich.

zu b): Annahme: $\delta = 0$, $\varepsilon = 1$. Als E_2-Diagramm erhalten wir:

E_2^{pq}

0	0	0	0
$\ker d_1^{-3,2}$	$\dfrac{\ker d_1^{-2,2}}{\operatorname{Im} d_1^{-3,2}}$	*	*
0	$\ker d_1^{-2,1}$	$\operatorname{coker} d_1^{-2,1}$	0
0	0	0	0

q ↑, p →

Nach dem Satz von Beilinson sind

$$d_2^{-3,2} : \ker d_1^{-3,2} \to \operatorname{coker} d_1^{-2,1} \quad \text{und} \quad d_1^{-2,1} : T(-2) \to \Omega(1)$$

Monomorphismen. $\operatorname{coker} d_1^{-2,1}$ ist also eine Torsionsgarbe und somit auch $\ker d_1^{-3,2}$. Da aber $\ker d_1^{-3,2} \subset \mathcal{O}(-1)^{\oplus \binom{n+1}{2}}$ auch torsionsfrei ist, folgt

$$\ker d_1^{-3,2} = 0.$$

Hieraus folgt: $E_\infty^{-1,1} = \operatorname{coker} d_1^{-2,1}$. Da aber $E_\infty^{-1,1}$ Untergarbe von E ist, ist auch $\operatorname{coker} d_1^{-2,1}$, torsionsfrei, also Null. $d_1^{-2,1} : T(-2) \to \Omega(1)$ ist also ein Isomorphismus, was unmöglich ist. Damit ist Satz 1 bewiesen.

Satz 2. *Es sei $n \geqq 2$. E sei ein stabiles 3-Bündel auf $\mathbb{P}_3$ mit den Chernklassen $c_1 = 0$, $c_2 = n$, $c_3 = n^2 - n$, und es gelte $h^0(E|H) = 1$ für allgemeine Ebenen H in $\mathbb{P}_3$. Dann gibt es eine durch E eindeutig bestimmte Ebene $H_0 \subset \mathbb{P}_3$ und eine bis auf Multiplikation mit $\lambda \in \mathbb{C}^*$ eindeutig bestimmte Extension*

$$0 \to \Omega^1_{\mathbb{P}_3}(1) \to E \to \mathcal{O}_{H_0}(1-n) \to 0.$$

Beweis. Die Spektralsequenz $E_1^{pq} = H^q(E(p)) \otimes \Omega^{-p}(-p)$ hat das folgende E_1-Diagramm:

E_1^{pq}

	0	0	0	0
	$\mathcal{O}(-1)^{\oplus\binom{n+1}{2}}$	$T(-2)^{\oplus\binom{n}{2}}$	$\Omega(1)^{\oplus\binom{n-1}{2}}$	$\mathcal{O}^{\oplus\binom{n-2}{2}}$
	0	0	$\Omega(1)$	0
	0	0	0	0

q ↑, p →

Im E_2-Diagramm kann nur das Differential

$$d_2^{-3,2} : E_2^{-3,2} \to E_2^{-1,1} = \Omega(1)$$

auftreten. Nach dem Satz von Beilinson ist dann $0 = E_\infty^{-3,2} = \ker d_2^{-3,2}$, $d_2^{-3,2}$ also ein Monomorphismus. $E_\infty^{-1,1} = \operatorname{coker} d_2^{-3,2}$ ist Untergarbe von E, also insbesondere torsionsfrei.

Behauptung. $E_2^{-3,2} = 0$.

Beweis. Da $d_2^{-3,2}$ injektiv ist, ist $\operatorname{rg} E_2^{-3,2} \leqq 3$. Wäre $\operatorname{rg} E_2^{-3,2} = 3$, so wäre $\operatorname{coker} d_2^{-3,2}$ vom Rang Null und damit überhaupt 0. Es wäre also $E_2^{-3,2} \cong \Omega(1)$ und somit

$$\Omega(1) \subset E_1^{-3,2} = \mathcal{O}(-1)^{\oplus\binom{n+1}{2}},$$

was unmöglich ist. Wäre $0 < \operatorname{rg} E_2^{-3,2} < 3$, so wäre auch

$$0 < \operatorname{rg}(\operatorname{coker} d_2^{-3,2}) < 3$$

und somit, da E stabil ist: $c_1(\operatorname{coker} d_2^{-3,2}) < 0$.

Andererseits ist aber

$$c_1(\operatorname{coker} d_2^{-3,2}) = c_1(\Omega(1)) - c_1(E_2^{-3,2}) \geqq 0,$$

weil $c_1(E_2^{-3,2}) < 0$ ist $\left[E_2^{-3,2}\right.$ ist Untergarbe von $\left.\mathcal{O}(-1)^{\oplus\binom{n+1}{2}}\right]$. Es muß also $\operatorname{rg} E_2^{-3,2} = 0$ gelten. Damit ist die Behauptung bewiesen. Aus dem Satz von Beilinson folgt: Der Komplex

$$E_1^{p,2} : 0 \to \mathcal{O}(-1)^{\oplus\binom{n+1}{2}} \xrightarrow[a]{} T(-2)^{\oplus\binom{n}{2}} \xrightarrow[b]{} \Omega(1)^{\oplus\binom{n-1}{2}} \xrightarrow[c]{} \mathcal{O}^{\oplus\binom{n-2}{2}} \to 0$$

ist bis auf die Stelle $E_1^{-2,2} = T(-2)^{\oplus\binom{n}{2}}$ exakt. Es gilt

$$E_\infty^{-2,2} = E_2^{-2,2} = \frac{\ker b}{\operatorname{im} a}, \quad E_\infty^{-1,1} = E_1^{-1,1} = \Omega(1), \quad E_\infty^{0,0} = E_\infty^{-3,3} = 0,$$

und somit haben wir eine Extension

$$0 \to \Omega(1) \to E \to E_\infty^{-2,2} \to 0.$$

Es bleibt nur noch folgendes zu zeigen:

Behauptung. $H_0 = \operatorname{supp} E_\infty^{-2,2}$ ist eine Ebene in $\mathbb{P}_3$ und es gilt

$$E_\infty^{-2,2} \cong \mathcal{O}_{H_0}(1-n).$$

Beweis. Da $\operatorname{im} b = \ker c$ ein Unterbündel von $\Omega(1)^{\oplus\binom{n-1}{2}}$ ist, ist $\ker b$ ein Unterbündel von $T(-2)^{\oplus\binom{n}{2}}$ vom Rang $\binom{n+1}{2}$. Wir erhalten die exakte Sequenz

$$0 \to \mathcal{O}(-1)^{\oplus\binom{n+1}{2}} \xrightarrow{\tilde{a}} \ker b \to E_\infty^{-2,2} \to 0.$$

$\operatorname{supp} E_\infty^{-2,2} = \{\det \tilde{a} = 0\}$ ist also ein Divisor vom Grad

$$c_1(\ker b) - c_1\left(\mathcal{O}(-1)^{\oplus\binom{n+1}{2}}\right) = 1,$$

d. h. $\operatorname{supp} E_\infty^{-2,2}$ ist eine Ebene $H_0 \subset \mathbb{P}_3$. Insbesondere ist $E_\infty^{-2,2}$ in natürlicher Weise ein $\mathcal{O}_{H_0}$-Modul. Um zunächst zu sehen, daß $E_\infty^{-2,2}$ als $\mathcal{O}_{H_0}$-Modul invertierbar ist, zeigen wir, daß für alle Punkte $x \in H_0$

$$\dim_{\mathbb{C}} E_\infty^{-2,2}(x) = 1$$

gilt. Dazu wähle man eine Gerade $L \subset \mathbb{P}_3$ mit $L \cap H_0 = x$. Schränkt man nun die Filtrierung

$$\mathcal{O}(-1)^{\oplus\binom{n+1}{2}} \subset \ker b \subset T(-2)^{\oplus\binom{n}{2}}$$

auf L ein, so erhält man die Filtrierung

$$\mathcal{O}_L(-1)^{\oplus\binom{n+1}{2}} \subset (\ker b)|L = \oplus \mathcal{O}_L(a_i) \subset \mathcal{O}_L^{\oplus\binom{n}{2}} \oplus \mathcal{O}_L(-1)^{\oplus 2\binom{n}{2}}$$

und somit $-1 \leqq a_i \leqq 0$ für alle i. Da $c_1(\ker b) = -\binom{n+1}{2} + 1$ ist, kann nur

$$(\ker b)|L \cong \mathcal{O}_L \oplus \mathcal{O}_L(-1)^{\oplus\binom{n+1}{2}-1}$$

gelten. Aus der exakten Sequenz

$$0 \to \mathcal{O}_L(-1)^{\oplus\binom{n+1}{2}} \to \mathcal{O}_L(-1)^{\oplus\binom{n+1}{2}-1} \oplus \mathcal{O}_L \to E_\infty^{-2,2}|L \to 0$$

folgt dann unmittelbar $\dim_{\mathbb{C}} E_\infty^{-2,2}(x) = 1$.

Es gilt also $E_\infty^{-2,2} = \mathcal{O}_{H_0}(a)$ mit einer wohlbestimmten ganzen Zahl a. Aus den Sequenzen

$$0 \to \mathcal{O}(a-1) \to \mathcal{O}(a) \to \mathcal{O}_{H_0}(a) \to 0,$$
$$0 \to \Omega(1) \to E \to \mathcal{O}_{H_0}(a) \to 0$$

errechnet man ($h = c_1(\mathcal{O}(1))$:

$$c(\mathcal{O}_{H_0}(a)) = \frac{1+ah}{1-(1-a)h} = 1 + h + (1-a)h^2 + (1-a)^2 h^3$$

und

$$c(\mathcal{O}_{H_0}(a)) = \frac{1+nh^2+(n^2-n)h^3}{c(\Omega(1))} = 1 + h + nh^2 + n^2 h^3.$$

Es folgt $a = 1-n$. Damit ist Satz 2 vollständig bewiesen.

Wir können nun umgekehrt zeigen:

Satz 3. *Für jedes $n \geqq 2$ gibt es stabile 3-Bündel auf $\mathbb{P}_3$ mit den Chernklassen $c_1 = 0$, $c_2 = n$, $c_3 = n^2 - n$, die auf allen Ebenen in $\mathbb{P}_3$ instabil sind.*

Beweis. Sei $n \geqq 2$ und $H_0 \subset \mathbb{P}_3$ eine Ebene.

$\mathrm{Ext}^1_{st}(\mathcal{O}_{H_0}(1-n), \Omega^1_{\mathbb{P}_3}(1))$ sei die Menge aller Extensionen

$$\xi = [0 \to \Omega(1) \to E \to \mathcal{O}_{H_0}(1-n) \to 0] \in \mathrm{Ext}^1(\mathcal{O}_{H_0}(1-n), \Omega^1_{\mathbb{P}_3}(1)),$$

für die E ein stabiles 3-Bündel auf $\mathbb{P}_3$ ist.

Wir identifizieren auf kanonische Weise die Garbe

$$\mathcal{E}xt^1(\mathcal{O}_{H_0}(1-n), \mathcal{O}_{\mathbb{P}_3}) \text{ mit } \mathcal{O}_{H_0}(n).$$

Da $\mathcal{H}om(\mathcal{O}_{H_0}(1-n), \Omega^1_{\mathbb{P}_3}(1)) = 0$ ist, ist der kanonische Homomorphismus

$$\mathrm{Ext}^1(\mathcal{O}_{H_0}(1-n), \Omega^1_{\mathbb{P}_3}(1)) \to H^0(\mathcal{E}xt^1(\mathcal{O}_{H_0}(1-n), \Omega^1_{\mathbb{P}_3}(1))$$
$$\simeq \mathrm{Hom}(T_{\mathbb{P}_3}(-1), \mathcal{O}_{H_0}(n)) \text{ ein Isomorphismus.}$$

Einer beliebigen Extension $\xi = [0 \to \Omega(1) \to E \to \mathcal{O}_{H_0}(1-n) \to 0]$ wird durch Dualisieren der Homomorphismus

$$\varphi : T(-1) \to \mathcal{O}_{H_0}(n)$$

der exakten Sequenz

$$0 \longrightarrow E^\vee \longrightarrow T(-1) \xrightarrow{\varphi} \mathcal{E}xt^1(\mathcal{O}_{H_0}(1-n), \mathcal{O}_{\mathbb{P}_3}) \longrightarrow \mathcal{E}xt^1(E, \mathcal{O}_{\mathbb{P}_3}) \longrightarrow 0$$
$$\| \wr$$
$$\mathcal{O}_{H_0}(n)$$

zugeordnet. Da E eine lokalfreie Auflösung der Länge 1 besitzt, ist E genau dann lokalfrei, wenn $\mathcal{E}xt^1(E, \mathcal{O}_{\mathbb{P}_3}) = 0$ ist, d. h. wenn φ ein Epimorphismus ist. Es ist automatisch $h^0(E) = 0$, und die weitere zur Stabilität von E erforderliche Bedingung $h^0(E^\vee) = 0$ ist gleichbedeutend mit der Injektivität der Abbildung

$$H^0\varphi : H^0(T_{\mathbb{P}_3}(-1)) \to H^0(\mathcal{O}_{H_0}(n)).$$

$\varphi: T_{\mathbb{P}_3}(-1) \to \mathcal{O}_{H_0}(n)$ ist eindeutig bestimmt durch

$$\varphi': T_{H_0}(-1) \oplus \mathcal{O}_{H_0} \to \mathcal{O}_{H_0}(n),$$

und es ist in kanonischer Weise $H^0\varphi = H^0\varphi'$. Damit erhalten wir eine Bijektion

$$\begin{aligned} \mathrm{Ext}^1_{st}(\mathcal{O}_{H_0}(1-n), \Omega_{\mathbb{P}_3}(1)) \xrightarrow{\sim} \{\varphi \in \mathrm{Hom}(T_{H_0}(-1) \oplus \mathcal{O}_{H_0}, \mathcal{O}_{H_0}(n)) \mid \\ \text{(i) } \varphi \text{ ist epimorph} \\ \text{(ii) } H^0\varphi \text{ ist monomorph}\}. \end{aligned}$$

Da (i) und (ii) Zariski-offene Eigenschaften sind, brauchen wir nur zu zeigen, daß es Homomorphismen $\varphi: T_{H_0}(-1) \oplus \mathcal{O}_{H_0} \to \mathcal{O}_{H_0}(n)$ gibt, so daß

a) die Bedingung (i),

b) die Bedingung (ii) erfüllt ist.

Dann ist $\mathrm{Ext}^1_{st}(\mathcal{O}_{H_0}(1-n), \Omega(1)) \neq \emptyset$ und Satz 3 bewiesen.

zu a): Da $\Omega_{H_0}(n+1) \oplus \mathcal{O}_{H_0}(n)$ global erzeugt ist, gibt es nach Serre (vgl. [4]) einen nullstellenfreien Schnitt in diesem Bündel und somit auch einen Epimorphismus

$$T_{H_0}(-1) \oplus \mathcal{O}_{H_0} \twoheadrightarrow \mathcal{O}_{H_0}(n).$$

zu b): Da $h^0(T_{H_0}(-1)) = 3 < h^0(\mathcal{O}_{H_0}(n))$ ist, wenn $n \geqq 2$ ist, genügt es, einen Homomorphismus

$$\psi: T_{H_0}(-1) \to \mathcal{O}_{H_0}(n)$$

zu finden, so daß $H^0\psi$ injektiv ist.

Seien x, y, z homogene Koordinaten von H_0. Wir betrachten nun z. B. $f_1 = yzh$, $f_2 = xzh$, $f_3 = -2xyh \in H^0(\mathcal{O}_{H_0}(n))$, wobei $h = z^{n-2}$ ist. f_1, f_2, f_3 sind linear unabhängig in $H^0(\mathcal{O}_{H_0}(n))$ und definieren einen Homomorphismus

$$f: \mathcal{O}_{H_0}^{\oplus 3} \to \mathcal{O}_{H_0}(n).$$

Auf den Fasern über $(x:y:z) \in H_0$ ist

$$f(x:y:z)(a,b,c) = ayzh + bxzh - 2cxyh.$$

Hieraus ergibt sich für $(a,b,c) = (\lambda x, \lambda y, \lambda z) \in \mathcal{O}_{H_0}(-1)(x:y:z)$

$$f(x:y:z)(\lambda x, \lambda y, \lambda z) = 0.$$

f induziert also einen Homomorphismus

$$\psi: T_{H_0}(-1) \to \mathcal{O}_{H_0}(n)$$

mit injektiver Abbildung

$$\begin{aligned} H^0\psi: \mathbb{C}^3 = H^0(T_{H_0}(-1)) &\to H^0(\mathcal{O}_{H_0}(n)), \\ (a,b,c) &\mapsto af_1 + bf_2 + cf_3. \end{aligned}$$

Damit ist Satz 3 bewiesen.

2

Zur Abkürzung bezeichnen wir mit $\mathfrak{M}_n$ die Menge aller stabilen Vektorbündel E vom Rang 3 auf $\mathbb{P}_3$ mit den Chernklassen $c_1=0$, $c_2=n$, $c_3=n^2-n$.

$\mathfrak{M}_n' \subset \mathfrak{M}_n$ sei die Teilmenge der Bündel E, für die gilt:

$$h^0(E|H)=1 \text{ für alle Ebenen } H \subset \mathbb{P}_3.$$

Wie wir gesehen haben ist $\mathfrak{M}_n'$ genau dann nicht-leer, wenn $n \geqq 2$ ist, und $\mathfrak{M}_n'=\mathfrak{M}_n$, wenn $n \neq 3$ ist.

Satz 4. *Es sei $E \in \mathfrak{M}_n'$. Dann gilt:*

$$h^0(\mathscr{E}nd\, E)=1$$

$$h^1(\mathscr{E}nd\, E)=\begin{cases}\frac{1}{2}n(3n+7)+3, & \text{falls } n\neq 3\\ 28, & \text{falls } n=3\end{cases}$$

$$h^2(\mathscr{E}nd\, E)=\begin{cases}\frac{1}{2}n(3n-17)+11, & \text{falls } n\neq 3\\ 0, & \text{falls } n=3.\end{cases}$$

Beweis. E sei durch die Extension

$$0\to\Omega(1)\to E\to\mathcal{O}_{H_0}(1-n)\to 0 \tag{4}$$

gegeben ($H_0 \subset \mathbb{P}_3$ Ebene). Hieraus erhält man weitere exakte Sequenzen:

$$0\to E^\vee\to T(-1)\to\mathcal{O}_{H_0}(n)\to 0, \tag{5}$$

$$0\to E^\vee\otimes\Omega(1)\to\mathscr{E}nd\,\Omega\to\Omega(1)\otimes\mathcal{O}_{H_0}(n)\to 0, \tag{6}$$

$$0\to E^\vee\otimes\Omega(1)\to\mathscr{E}nd\, E\to E_0(1-n)\to 0, \tag{7}$$

wobei $E_0 := E|H_0$ ist. Schränkt man (4) auf H_0 ein, so ergibt sich

$$0\to\mathcal{O}_{H_0}(-n)\to\Omega(1)\otimes\mathcal{O}_{H_0}\to E_0\to\mathcal{O}_{H_0}(1-n)\to 0. \tag{8}$$

Es folgt:

$$\begin{aligned}\chi(\mathscr{E}nd\, E)&=\chi(E^\vee\otimes\Omega(1))+\chi(E_0(1-n))\\ &=\chi(\mathscr{E}nd\,\Omega)-\chi(\Omega(1)\otimes\mathcal{O}_{H_0}(n))+\chi(E_0(n-4))\\ &=1-h^0(\Omega(1)\otimes\mathcal{O}_{H_0}(n))+\chi(E_0(n-4)).\end{aligned}$$

Tensoriert man (8) mit $\mathcal{O}(n-4)$, so läßt sich $\chi(E_0(n-4))$ berechnen, und man erhält

$$\chi(\mathscr{E}nd\, E)=-1-h^0(\Omega(1)\otimes\mathcal{O}_{H_0}(n))+h^0(\Omega(1)\otimes\mathcal{O}_{H_0}(n-4))-\begin{cases}0 & (n\neq 3)\\ 1 & (n=3).\end{cases}$$

Aus (6) folgt $h^2(E^\vee\otimes\Omega(1))=h^3(E^\vee\otimes\Omega(1))=0$, und aus (7), (8) somit mittels Serre-Dualität

$$h^2(\mathscr{E}nd\, E)=h^0(E_0(n-4))=h^0(\Omega(1)\otimes\mathcal{O}_{H_0}(n-4)).$$

Da E stabil ist, gilt $h^0(\mathscr{E}nd\, E)=1$, $h^3(\mathscr{E}nd\, E)=0$, also

$$h^1(\mathscr{E}nd\, E)=h^0(\Omega(1)\otimes\mathcal{O}_{H_0}(n))+\begin{cases}2 & \text{falls } n\neq 3\\ 3 & \text{falls } n=3.\end{cases}$$

Auswerten ergibt die Behauptung des Satzes.

Satz 5. *Ist* $E\in\mathfrak{M}_3\setminus\mathfrak{M}_3'$, *so existiert eine Auflösung*

$$0\to\mathcal{O}(-1)^{\oplus 6}\to T(-2)^{\oplus 3}\to E\to 0 \tag{9}$$

von E, *und es gilt* $h^1(\mathcal{E}nd\,E)=28$, $h^2(\mathcal{E}nd\,E)=0$.

Beweis. Der erste Teil des Satzes folgt sofort aus der Beilinson-Spektralsequenz. Es ist $h^q(E(k))=0$ für $q>0$, $k\geqq 0$, und somit folgt aus der Sequenz

$$0\to E\otimes\Omega(2)\to E(1)^{\oplus 4}\to E(2)\to 0$$

$h^2(E\otimes\Omega(2))=0$ und aus der Sequenz

$$0\to\mathcal{E}nd\,E\to E\otimes\Omega(2)^{\oplus 3}\to E(1)^{\oplus 6}\to 0$$

dann auch $h^2(\mathcal{E}nd\,E)=0$. Da E stabil ist, folgt

$$h^1(\mathcal{E}nd\,E)=1-\chi(\mathcal{E}nd\,E).$$

Man errechnet (L Gerade, H Ebene, $L\subset H\subset\mathbb{P}_3$):

$$\begin{aligned}\chi(\mathcal{E}nd\,E)&=\chi(E\otimes\Omega(2)^{\oplus 3})-\chi(E(1)^{\oplus 6})\\&=\chi(E(1)^{\oplus 12})-\chi(E(2)^{\oplus 3})-\chi(E(1)^{\oplus 6})=6\chi(E(1))-3\chi(E(2))\\&=3\chi(E(1))-3\chi(E(2)|H)=-3\chi(E(2)|L)=-27,\end{aligned}$$

also $h^1(\mathcal{E}nd\,E)=28$.

Aus den Sätzen 4 und 5 folgt mittels Deformationstheorie:

Satz 6. *Die Modulräume* $\mathbf{M}_n$ *der Isomorphieklassen stabiler 3-Bündel auf* $\mathbb{P}_3$ *mit den Chernklassen* $c_1=0$, $c_2=n$, $c_3=n^2-n$ $(n\geqq 2)$, *sind glatt und reindimensional. Es gilt*

$$\dim\mathbf{M}_n=\begin{cases}\frac{1}{2}n(3n+7)+3, & \text{falls}\quad n\neq 3\\ 28 & \text{falls}\quad n=3.\end{cases}$$

$\mathbf{M}_n'\subset\mathbf{M}_n$ sei der abgeschlossene reduzierte komplexe Unterraum der Isomorphieklassen von Bündeln $E\in\mathfrak{M}_n'$.

Satz 7. $\mathbf{M}_n$ *ist irreduzibel und rational, falls* $n\neq 3$ *ist.* $\mathbf{M}_3'$ *ist eine irreduzible glatte rationale Hyperfläche in* $\mathbf{M}_3$.

Beweis. Sei $\mathbb{P}_3^*$ der projektive Raum der Ebenen in $\mathbb{P}_3$, $\mathbb{F}\subset\mathbb{P}_3\times\mathbb{P}_3^*$ sei die Inzidenzrelation "$x\in H$". $p:\mathbb{P}_3\times\mathbb{P}_3^*\to\mathbb{P}_3$ und $q:\mathbb{P}_3\times\mathbb{P}_3^*\to\mathbb{P}_3^*$ seien die Projektionen.

Die relative Ext-Garbe

$$\mathcal{E}=\mathcal{E}xt_q^1(\mathcal{O}_{\mathbb{F}}\otimes p^*\mathcal{O}_{\mathbb{P}_3}(1-n),\ p^*\Omega^1_{\mathbb{P}_3}(1))$$

ist dann ein Vektorbündel auf $\mathbb{P}_3^*$ vom Rang

$$r=\dim_{\mathbb{C}}\operatorname{Ext}^1(\mathcal{O}_H(1-n),\Omega^1_{\mathbb{P}_3}(1))=h^0(\Omega^1_{\mathbb{P}_3}(1)\otimes\mathcal{O}_H(n))=\tfrac{1}{2}n(3n+7)+1$$

($H\subset\mathbb{P}_3$ Ebene). Indem man nun E die zugehörige Extension (Satz 2) zuordnet, erhält man eine offene Einbettung $\mathbf{M}_n'\subset\mathbb{P}(\mathcal{E}^\vee)$ von $\mathbf{M}_n'$ in das projektive Bündel der Geraden in $\mathcal{E}$. $\mathbf{M}_n'$ ist also irreduzibel, glatt und rational. Da $\mathbf{M}_n'=\mathbf{M}_n$ gilt, falls $n\neq 3$ ist, ist der Satz bewiesen.

Mit Hilfe der zweiten Spektralsequenz von Beilinson (vgl. [4, S. 245]) kann man zeigen, daß jedes Bündel $E \in \mathfrak{M}_3 \backslash \mathfrak{M}'_3$ eine Auflösung der Gestalt

$$0 \to \mathcal{O}_{\mathbb{P}_3}(-2)^{\oplus 3} \to \mathcal{O}_{\mathbb{P}_3}(-1)^{\oplus 6} \to E \to 0 \tag{10}$$

besitzt.

$$\tilde{\mathbf{M}}_3 = \{A \in \mathrm{Hom}(\mathcal{O}_{\mathbb{P}_3}(-2)^{\oplus 3}, \mathcal{O}_{\mathbb{P}}(-1)^{\oplus 6}) \mid \text{(a) } A(x) \text{ ist injektiv für alle } x \in \mathbb{P}_3, \text{ (b) } H^0(A^\vee): H^0(\mathcal{O}_{\mathbb{P}_3}(1)^{\oplus 6}) \to H^0(\mathcal{O}_{\mathbb{P}}(2)^{\oplus 3}) \text{ ist injektiv}\}$$

ist eine nicht-leere Zariskioffene Teilmenge des Vektorraumes

$$\mathrm{Hom}(\mathcal{O}(-2)^{\oplus 3}, \mathcal{O}(-1)^{\oplus 6}) \simeq \mathrm{Hom}_{\mathbb{C}}(\mathbb{C}^3, \mathbb{C}^6) \otimes H^0(\mathcal{O}(1))$$

und durch $A \mapsto \mathrm{coker}\, A$ wird eine surjektive holomorphe Abbildung $f: \tilde{\mathbf{M}}_3 \to \mathbf{M}_3 \backslash \mathbf{M}'_3$ induziert. Da $\tilde{\mathbf{M}}_3$ irreduzibel ist, ist auch $\mathbf{M}_3 \backslash \mathbf{M}'_3$ und damit auch $\mathbf{M}_3$ irreduzibel.

Literatur

1. Elencwajg, G., Forster, O.: Bounding cohomology groups of vector bundles on $\mathbb{P}_n$. Math. Ann. **246**, 251–270 (1980)
2. Schneider, M.: Chernklassen semi-stabiler Vektorraumbündel vom Rang 3 auf dem komplex-projektiven Raum. Crelles J. **315**, 211–220 (1980)
3. Spindler, H.: Ein Satz über die Einschränkung holomorpher Vektorbündel auf $\mathbb{P}_n$ mit $c_1 = 0$ auf Hyperebenen. Preprint, Göttingen 1980 (erscheint in Crelles J.)
4. Okonek, Ch., Schneider, M., Spindler, H.: Vector bundles on complex projective spaces. Progress in Mathematics, Vol. 3. Boston: Birkhäuser 1980

Eingegangen am 8. Januar 1981

Math. Ann. 256, 145–156 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Valuations and Semi-Valuations of Graded Domains

David F. Anderson[1] and Jack Ohm[2]

[1] Department of Mathematics, University of Tennessee, Knoxville, TN 37916, USA
[2] Department of Mathematics, Louisiana State University, Baton Rouge, LA 70803, USA

Let w be a valuation of a field L having value group C_w and residue field $k_w = R_w/m_w$, where R_w, m_w is the valuation ring of w, and let $h: R_w \to k_w$ be the canonical homomorphism. If u is a valuation of k_w having value group A_u, then the valuation v of L whose ring is $h^{-1}(R_u)$ is called the "composite" of w and u. All valuations of rank greater than one occur (non-trivially) as such composites. Moreover, the associated value groups form a lexicographically exact sequence $0 \to A_u \to B_v \to C_w \to 0$. These concepts were generalized in [10, Sect. 3] to the case that R_w, m_w is an arbitrary quasi-local domain and R_u is any subring of k_w.

The ring theory for a special construction of this type, namely the case that w is a valuation and R_u is a subfield of k_w, has been studied recently by a number of people; and, in particular, the work of [2] led to the question of whether one can have an exact sequence of semi-value groups $0 \to A_u \to B_v \to C_w \to 0$ corresponding to this construction which does not split. Here the group C_w should be totally ordered and the group A_u trivially ordered, i.e., $A_u^+ = 0$.

The main result of the present paper answers this question by establishing that any exact sequence $0 \to A \to B \to C \to 0$ with C totally ordered and A an arbitrary semi-value group can be realized as such a lexicographically exact semi-value group sequence. Another way of expressing this is that any lexicographic extension of a semi-value group by a totally ordered group is again a semi-value group.

The proof of this result involves finding an appropriate valuation w which extends a given semi-valuation u. Our first efforts at defining valuations which produce non-split semi-value sequences resulted in ad hoc constructions involving gradings, and it was only after we realized that the gradings of twisted group rings yield appropriate valuations that these ideas fell into place.

Section 1 contains a discussion of how one can define valuations by means of gradings and a description of the residue fields of these valuations. Section 2 generalizes these considerations to valuations which extend a prescribed valuation or semi-valuation. In Sect. 3 these ideas are applied to twisted group rings, and the main result is proved. Section 4 is somewhat of an afterthought and is devoted to putting into proper perspective the relationship between twisted group rings and their associated semi-valuations.

0025-5831/81/0256/0145/$02.40

0. Terminology and Review

If D is a domain, we use D^* to denote the set of non-zero elements of D. The group of units of D will be denoted by $U(D)$. The letter $\mathbb{Z}$ denotes the additive group of integers, and $\mathbb{Z}^+$ the additive monoid of non-negative integers. All rings, groups, and monoids will be commutative.

(a) *Gradings* (cf. [9, Sect. 2.11], [6]). We shall work only with graded domains $D=\bigoplus_{\gamma\in\Gamma} D_\gamma$, where Γ is a cancellative, torsionless monoid. The cancellative property allows Γ to be embedded in a group hull $\langle\Gamma\rangle$, and the torsionless property is equivalent to $\langle\Gamma\rangle$ being torsion-free. Moreover, a cancellative monoid Γ is torsionless if and only if Γ can be given a total ordering which is compatible with the monoid operation, and this is the form in which we shall use this property.

The elements of D_γ are called *homogeneous* of $\deg\gamma$, and it turns out that 1 is homogeneous of $\deg 0$. Each $d\in D$ has a unique representation $d=\sum_{\gamma\in\Gamma} d_\gamma$, where $d_\gamma\in D_\gamma$ and all but finitely many $d_\gamma=0$. The element d_γ is called the homogeneous component of d of $\deg\gamma$. We shall use H to denote the set of non-zero homogeneous elements of D; H is closed under multiplication.

The monoid ring $k[\Gamma]$, where k is a field, consists of all formal sums $\sum_{\gamma\in\Gamma}\alpha_\gamma X^\gamma$, $\alpha_\gamma\in k$ and all but finitely many $\alpha_\gamma=0$. Addition is defined by adding coefficients, and multiplication is defined distributively using $X^\gamma\cdot X^{\gamma'}=X^{\gamma+\gamma'}$. An element α of the coefficient field k will be identified with αX^0. It is easily seen that $k[\Gamma]$ is a domain if and only if Γ is cancellative and torsionless.

(b) *Valuations and semi-valuations* (cf. [10, 6]). A *semi-valuation* of a domain D is a map u of D^* into an ordered group A such that for all $x, y\in D^*$,

(i) $u(xy)=u(x)+u(y)$,
(ii) $u(x+y)\geqq\inf_{A_u}\{u(x), u(y)\}$ whenever $x+y\in D^*$,
(iii) $u(-1)=0$.

Here A_u denotes the ordered subgroup of A generated by $u(D^*)$ and is called the *semi-value group* of u. The terminology of (ii) means $u(x+y)\geqq a$ for all $a\in A_u$ such that $a\leqq u(x), u(y)$. Sometimes it is convenient to regard u as being defined on all of D by defining $u(0)=+\infty$. If A_u is totally ordered, then u is called a *valuation*.

Note that if K is the quotient field of D, then u extends uniquely to a semi-valuation of K, which will also be denoted by u; and then $A_u=u(K^*)$. The set $R_u=\{x\in K^*|u(x)\geqq 0\}\cup\{0\}$ forms a subring of K, called the *semi-valuation ring* of u. R_u has quotient field K if and only if the semi-value group A_u is filtered, and in this case A_u is also called the *group of divisibility* of R_u.

(c) *Ordered groups* (cf. [10, 6]). We shall find it convenient to write our short exact sequences of groups in the form $1\to A\xrightarrow{\alpha} B\xrightarrow{\beta} C\to 0$ with 1, A, and B multiplicative and C and 0 additive. If A, B, and C are ordered groups with non-negative elements A^+, B^+, and C^+ respectively, then the sequence is called *lexicographically exact* (or *lex*, for short) if $B^+=\{b\in B|\beta(b)>0 \text{ or } b\in\alpha(A^+)\}$; when this is the case, we shall also say that B is a *lex extension* of A by C. It will sometimes simplify notation to avoid explicit reference to the map α and merly regard it as inclusion, and we shall not hesitate to do this when it can be done without causing confusion.

A homomorphism $\alpha: A \to B$ of ordered groups is called a *V-homomorphism* if for any $a_0, a_1, \ldots, a_n \in A$, $a_0 \geqq \inf_A\{a_1, \ldots, a_n\}$ implies that $\alpha(a_0) \geqq \inf_B\{\alpha(a_1), \ldots, \alpha(a_n)\}$. If A and B are lattice ordered groups, a V-homomorphism is merely one which preserves infs. An injective V-homomorphism is called a *V-embedding*, and A is called a *V-subgroup* of B if the inclusion map is a V-homomorphism.

1. Valuations Defined by Gradings

Fix throughout Sect. 1 a graded domain $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$, where Γ is a cancellative, torsionless monoid, and a monoid homomorphism $\phi: \Gamma \to C$ of Γ into a totally ordered additive group C. It is straightforward to check that a valuation $v: D^* \to C$ is defined by

$$v(d_{\gamma_1} + \ldots + d_{\gamma_t}) = \inf_{i=1,\ldots,t} \{\phi(\gamma_i)\}, \tag{1.1}$$

where $d_{\gamma_i} \in D^*_{\gamma_i}$ and $\gamma_i \neq \gamma_j$ for $i \neq j$. We shall call v the *inf valuation* of D with respect to the Γ-grading and the homomorphism ϕ. If ϕ is injective and C is generated by $\phi(\Gamma)$, in which case the homomorphism is equivalent to the inclusion map of Γ into its group hull $\langle\Gamma\rangle$, we shall drop the reference to ϕ.

One can reduce to the case that ϕ is injective by thickening the grading: Let $D'_c = \oplus\{D_\gamma | \phi(\gamma) = c\}$. Then $D = \bigoplus_{c \in \phi(\Gamma)} D'_c$, and the inf valuation with respect to this new grading and the inclusion homomorphism $\phi(\Gamma) \subset C$ clearly coincides with the original valuation v. If ϕ is injective, then $\phi(\gamma_i) \neq \phi(\gamma_j)$ for $\gamma_i \neq \gamma_j$; so in this case (1.1) is the only way of defining a valuation v of D^* into C so that $v(d_{\gamma_i}) = \phi(\gamma_i)$ for every homogeneous element d_{γ_i} of D^*.

Examples 1.2. Let k be a field, let C be a totally ordered group, and let C^+ be the monoid of non-negative elements of C.

(a) Let $\phi: C^+ \to C$ be the inclusion homomorphism and let $D = k[C^+]$ be the monoid ring of C^+ over k. Here $D = \bigoplus_{c \in C^+} kX^c$, and the inf valuation of D with respect to the C^+-grading is defined by

$$v(\alpha_1 X^{c_1} + \ldots + \alpha_t X^{c_t}) = \inf_{i=1,\ldots,t} \{c_i\},$$

where $\alpha_i \in k^*$, and $c_1, \ldots, c_t$ are distinct elements of C^+.

(b) Let Γ be the free monoid on any set $C_0 \subset C$, $\Gamma = \bigoplus_{c \in C_0} \mathbb{Z}^+_c$, and let $\phi: \Gamma \to C$ be defined by $\phi((\ldots, n_c, \ldots)) = \sum_{c \in C_0} n_c c$. Then the monoid ring $D = k[\Gamma]$ is just the polynomial ring over k in the variables $\{X_c\}_{c \in C_0}$. If we think of D as being this polynomial ring, then the inf valuation of D with respect to the Γ-grading and the homomorphism ϕ is given by

$$v(\alpha X^{n_1}_{c_1} \ldots X^{n_t}_{c_t}) = n_1 c_1 + \ldots + n_t c_t,$$

and

$$v(M_1 + \ldots + M_r) = \inf_{i=1,\ldots,r} \{v(M_i)\},$$

where $\alpha \in k^*$, $n_i \in \mathbb{Z}^+$, and $c_i \in C_0$, and the M_i are non-zero monomials of distinct degree.

(b') If we thicken the Γ-grading in (b) by defining $D'_c = \oplus \{kX_{c_1}^{n_1} \ldots X_{c_t}^{n_t} | n_1 c_1 + \ldots + n_t c_t = c\}$, then $D = \bigoplus_{c \in \phi(\Gamma)} D'_c$ is a grading for which the inclusion homomorphism $\phi(\Gamma) \subset C$ yields the same inf valuation v as in (b). (The homomorphism ϕ will play a more essential role in Sect. 2.)

Examples (a) and (b) yield two different proofs that every totally ordered group C is a value group. Krull originally used the group ring construction $k[C]$ to prove this. However, it is clear that $k[C]$ and $k[C^+]$ both define the same inf valuation on their quotient field.

We want next to describe the residue field k_v of v. If $H = \{\Gamma$-homogeneous elements of $D^*\}$, *then* D_H is a $\langle \Gamma \rangle$-graded domain:

$$D_H = \bigoplus_{\gamma \in \langle \Gamma \rangle} (D_H)_\gamma, \text{ where } (D_H)_\gamma = \{d_1/d_2 | d_1, d_2 \text{ are homogeneous elements of } D \text{ with } d_2 \neq 0 \text{ and } \deg d_1 - \deg d_2 = \gamma\}.$$

In particular, $(D_H)_0$ is a field; in examples (a) and (b), it is k. [More generally, whenever $D = k[\Gamma]$, $(D_H)_0 = k$.]

Proposition 1.3. *If ϕ is injective, then $R_v = (D_H)_0 + m_v$ (and thus k_v may be identified with $(D_H)_0$).*

Proof. The inclusion $\supset$ is clear, so let $0 \neq r \in R_v$. Then $r = a/b$, with $a, b \in D^*$ and $v(a) \geqq v(b)$. If $v(a) > v(b)$, then $r \in m_v$; so we may assume equality holds. We can write $a = a_{\alpha_1} + \ldots + a_{\alpha_t}$ and $b = b_{\beta_1} + \ldots + b_{\beta_s}$, with $\alpha_i, \beta_j \in \Gamma$ and $\alpha_1 < \ldots < \alpha_t$, $\beta_1 < \ldots < \beta_s$. Then $v(a) = \alpha_1 = \beta_1 = v(b)$; so $a/b = (a_{\alpha_1}/b_{\beta_1}) + y$, where

$$\begin{aligned} y &= (a_{\alpha_1} + \ldots + a_{\alpha_t})/(b_{\beta_1} + \ldots + b_{\beta_s}) - (a_{\alpha_1}/b_{\beta_1}) \\ &= [b_{\beta_1}(a_{\alpha_2} + \ldots + a_{\alpha_t}) - a_{\alpha_1}(b_{\beta_2} + \ldots + b_{\beta_s})]/b_{\beta_1}(b_{\beta_1} + \ldots + b_{\beta_s}). \end{aligned}$$

It follows that $v(y) > 0$. Hence $a/b \in (D_H)_0 + m_v$. Q.E.D.

Now let us consider the case that ϕ is no longer assumed injective. Let k be the field $(D_H)_0$ formed with respect to the Γ-grading. As noted above, we may thicken the Γ-grading to a Γ'-grading, where $\Gamma' = \phi(\Gamma)$; and both the Γ and Γ'-gradings give rise to the same valuation v. If H' is the set of non-zero Γ'-homogeneous elements of D, we see by Proposition 1.3 that $k_v = (D_{H'})_0$ and $R_v = k_v + m_v$. The next proposition shows how k and k_v are related.

Proposition 1.4. *$k_v = k(\{d_1/d_2 | d_1, d_2$ are Γ-homogeneous elements of D^* with $v(d_1/d_2) = 0\})$.*

Proof. The inclusion $\supset$ is immediate from the definitions. To see $\subset$, let $0 \neq \xi \in k_v$. Then $v(\xi) = 0$, and we may write $\xi = (A_1 + \ldots + A_t)/B$, where the A_i are Γ-homogeneous elements of D^* of different degrees and B is an element of D^*. Then $v(A_1 + \ldots + A_t) = \inf_{i=1,\ldots,t} \{v(A_i)\} = v(B)$; so $\xi = (A_1/B) + \ldots + (A_t/B)$, with $v(A_i/B) \geqq 0$. Since $R_v = k_v + m_v$ is a direct sum, it suffices to prove that ξ is $\equiv \bmod m_v$ to an element of $k(\{\})$. Thus it suffices to prove that each A_i/B is $\equiv \bmod m_v$ to an element of $k(\{\})$; and since this is immediate if $v(A_i/B) > 0$, we may assume that $v(A_i/B) = 0$. But then $B/A_i \in R_v$; and it suffices to prove $B/A_i \equiv \bmod m_v$ to an

element of $k(\{\})$, since if $B/A_i = a + y$, with $a \in k(\{\})$ and $y \in m_v$, then $A_i/B = 1/(a+y) = 1/a - (y/a)/(a+y) \in k(\{\}) + m_v$.

Now write $B = B_1 + \ldots + B_r$, where the B_j are Γ-homogeneous elements of D^* of distinct degree. Then $B/A_i = (B_1/A_i) + \ldots + (B_r/A_i)$; and since $0 = v(B/A_i)$ $= \inf_{j=1,\ldots,r} \{v(B_j/A_i)\}$, either $B_j/A_i \in m_v$ or B_j/A_i is, by definition, in $k(\{\})$. Q.E.D.

Let us now look at a specific example.

Example 1.5. Let C be a totally ordered abelian group with the property

(1.5.1) There are an infinite set of distinct elements $c_1, c_2, \ldots$ of C and positive integers n_i such that $c_i = n_i c_{i+1}$.

For example, one could take C to be the subgroup of the additive group of rational numbers generated by 1, 1/2, 1/4, ..., $1/2^i$, ..., i.e., $C = \{n/2^i | i \in \mathbb{Z}^+ \text{ and } n \in \mathbb{Z}\}$. Let $D = k[X_{c_1}, X_{c_2}, \ldots]$ be the polynomial ring over a field k in the indeterminates $\{X_{c_i}\}$. As in Example 1.2(b), this D may be regarded as being $k[\Gamma]$, where Γ is the free monoid on $\{c_i\}$; and we can define a valuation $w: D^* \to C$ by $w(X_{c_i}) = c_i$ and $w(A_1 + \ldots + A_t) = \inf_{i=1,\ldots,t} \{w(A_i)\}$ if $A_1, \ldots, A_t$ are Γ-homogeneous elements of D^* of distinct degree.

Now let $Z_1 = X_{c_1}/(X_{c_2})^{n_1}$, $Z_2 = X_{c_2}/(X_{c_3})^{n_2}$, ..., $Z_i = X_{c_i}/(X_{c_{i+1}})^{n_i}$, ...

Lemma. *$k_w = k(Z_1, Z_2, \ldots)$, and the Z_i are algebraically independent over k.*

Proof. By Proposition 1.4, for the first assertion it is only necessary to check that $\xi \in k(Z_1, Z_2, \ldots)$ whenever $w(\xi) = 0$ and ξ is a quotient of non-zero monomials in the X_{c_i}. But

$$\begin{aligned}
X_{c_1} &= X_{c_2}^{n_1} Z_1 = (X_{c_3})^{n_1 n_2} Z_1 Z_2^{n_1} = (X_{c_4})^{n_1 n_2 n_3} Z_1 Z_2^{n_1} Z_3^{n_1 n_2} = \ldots \\
X_{c_2} &= \quad = (X_{c_3})^{n_2} Z_2 \quad = (X_{c_4})^{n_2 n_3} Z_2 Z_3^{n_2} \quad = \ldots \\
X_{c_3} &= \quad \quad = (X_{c_4})^{n_3} Z_3 \quad = \ldots . \\
&\vdots \qquad\qquad \vdots
\end{aligned}$$

Thus we may rewrite ξ in the form

$$\xi = \alpha Z_1^{i_1} \ldots Z_r^{i_r} (X_{c_r})^i / Z_1^{j_1} \ldots Z_r^{j_r} (X_{c_r})^j$$

where $\alpha \in k^*$ and $i_l, j_l \in \mathbb{Z}^+$. Since $w(Z_i) = 0$ and $w(\xi) = 0$, we conclude that $i = j$, and hence that $\xi \in k(Z_1, Z_2, \ldots)$.

For the second assertion of the lemma, we must show that $A_1 + \ldots + A_s \neq 0$ whenever the A_i are non-zero monomials of distinct degree in the Z_i. Thus, it will suffice to show that if $A = Z_1^{i_1} \ldots Z_t^{i_t}$ and $B = Z_1^{j_1} \ldots Z_t^{j_t}$, with $i_l, j_l \in \mathbb{Z}^+$ and $(i_1, \ldots, i_t) \neq (j_1, \ldots, j_t)$, then A and B are also (fractional) power products of distinct degree in the X_{c_i}.

Choose r minimal such that $i_r \neq j_r$. Then X_{c_r} occurs to the power $i_r - i_{r-1} n_{r-1}$ in A and $j_r - j_{r-1} n_{r-1}$ in B. Since $i_{r-1} = j_{r-1}$, these powers are distinct. Q.E.D.

One intriguing aspect of this example is that the groups C having a set of elements satisfying (1.5.1) are known to be exactly the groups for which $k[C]$ is not a UFD (cf. [4, p. 78]).

Problem. Characterize the groups C for which the valuation w defined as in 1.5 yields a residue field k_w which is an infinite pure transcendental extension of k.

2. Extensions of Valuations and Semi-Valuations

The examples of 1.2 should really be thought of in the more general setting of extended valuations, or in the yet more general setting of extensions of semi-valuations. Thus, suppose we are given a valuation u of a field k into a totally ordered group B, $u:k^*\to B$, and a homomorphism ϕ of a cancellative, torsionless monoid Γ into B. (In the examples of 1.2, u is the 0 valuation.)

Proposition 2.1. u can be extended to a valuation v of $k[\Gamma]$ into B by defining

$$v(\alpha_1 X^{\gamma_1}+\ldots+\alpha_t X^{\gamma_t})=\inf_{i=1,\ldots,t}\{u(\alpha_i)+\phi(\gamma_i)\}, \tag{2.1.1}$$

where $\gamma_i\neq\gamma_j$ for $i\neq j$ and $\alpha_i\in k^*$.

Proof. Let f and g be non-zero elements of $k[\Gamma]$, and write $f=\sum_{\gamma\in\Gamma}\alpha(\gamma)X^{\gamma}$, $g=\sum_{\gamma\in\Gamma}\beta(\gamma)X^{\gamma}$, where $\alpha(\gamma)$, $\beta(\gamma)\in k$. If $b\in B$ is such that $b\leqq v(f)$, $v(g)$, then $b\leqq u(\alpha(\gamma))+\phi(\gamma)$, $u(\beta(\gamma))+\phi(\gamma)$ for all γ (where $u(0)$ is taken to be $+\infty$). Thus $b-\phi(\gamma)\leqq u(\alpha(\gamma))$, $u(\beta(\gamma))$, and hence $b-\phi(\gamma)\leqq u(\alpha(\gamma)+\beta(\gamma))$. But then $b\leqq u(\alpha(\gamma)+\beta(\gamma))+\phi(\gamma)$ (for all γ), and it follows that $b\leqq v(f+g)$. Thus $v(f+g)\geqq\inf\{v(f), v(g)\}$. (We have fashioned this part of the proof so as to make it also applicable to the proof of Theorem 3.4.)

Consider next $f\cdot g$. We can write $f=\alpha_1X^{\gamma_1}+\ldots+\alpha_tX^{\gamma_t}+(\ldots)$, $g=\beta_1X^{\delta_1}+\ldots+\beta_sX^{\delta_s}+(\ldots)$, where $\alpha_i,\ \beta_i\in k^*$; $v(f)=u(\alpha_1)+\phi(\gamma_1)=\ldots=u(\alpha_t)+\phi(\gamma_t)$, $v(g)=u(\beta_1)+\phi(\delta_1)=\ldots=u(\beta_s)+\phi(\delta_s)$; and $\gamma_1<\ldots<\gamma_t$, $\delta_1<\ldots<\delta_s$.

Then $fg=\alpha_1\beta_1X^{\gamma_1+\delta_1}+\ldots$, and every other term $T=\alpha_i\beta_jX^{\gamma_i+\delta_j}$, i or $j\neq 1$, of fg either has $v(T)>v(\alpha_1\beta_1X^{\gamma_1+\delta_1})$ or has $\gamma_i+\delta_j>\gamma_1+\delta_1$. Thus, the coefficient c of $X^{\gamma_1+\delta_1}$ in fg has the form $c=(\alpha_1\beta_1+\ldots+\alpha_i\beta_j+\ldots)$, with $u(\alpha_i\beta_j)>u(\alpha_1\beta_1)$ for i or $j\neq 1$. Therefore $u(c)=u(\alpha_1\beta_1)$; so $u(\alpha_1\beta_1)+\phi(\gamma_1+\delta_1)$ must occur as one of the values used in computing $v(fg)$. Thus $v(fg)=u(\alpha_1\beta_1)+\phi(\gamma_1+\delta_1)=[u(\alpha_1)+\phi(\gamma_1)]+[u(\beta_1)+\phi(\delta_1)]=v(f)+v(g)$. Q.E.D.

We shall call the extension of u to $k[\Gamma]$ defined in 2.1.1 the *inf extension* with respect to ϕ. If ϕ is injective, a more general result can be found in [10, 6.1]; in fact, all of the ideas of this section are taken from [10, Sects. 5, 6], with the added generality here that $\phi:\Gamma\to B$ is allowed to be an arbitrary homomorphism.

Next we shall generalize Proposition 2.1 to extensions of a semi-valuation into a lattice ordered group.

Theorem 2.2. *Let $u:k^*\to A_u$ be a semi-valuation of a field k having semi-value group A_u, and suppose A_u is a V-subgroup of a lattice ordered group B. Let $\phi:\Gamma\to B$ again be a monoid homomorphism. Then u can be extended to a semi-valuation v of $k[\Gamma]$ into B by defining v as in Proposition 2.1.*

Sketch of Proof (cf. [10, Sects. 5, 6]). By a theorem of Lorenzen, there is a V-embedding $l:B\to\prod B_i$ of B into a product of totally ordered groups B_i. If p_i denotes the projection map into B_i, then $u_i:k^*\to B_i$, with $u_i=p_ilu$, is a valuation. Using Proposition 2.1 and $\phi_i:\Gamma\to B_i$ defined by $\phi_i=pil\phi$, we can extend each u_i to a valuation v_i of $k[\Gamma]$. Then πv_i is a semi-valuation of $k[\Gamma]$; and by the way it is defined, $\pi v_i=v$. Q.E.D.

Applications 2.3. Theorem 2.2 yields two different constructions for proving the Kaplansky-Jaffard theorem that every lattice ordered group is a semi-value group:

(i) [10, Sect. 6]. Let u be the 0 map, let $\Gamma = \oplus \mathbb{Z}^+_{b_i}$ be the free monoid on a set of elements $\{b_i\}$ of B which generate B as a group, and define $\phi: \Gamma \to B$ by $\phi(\ldots, 1_{b_i}, \ldots) = b_i$. Then the semi-valuation defined by Theorem 2.2, extended to the quotient field of $k[\Gamma]$, is onto.

(ii) [6, p. 78]. Let u be the 0 map, let $\Gamma = B$, and let ϕ be the identity map. Then the semi-valuation v of Theorem 2.2 is onto B. (This is the *construction* of Jaffard, but his proof that v is a semi-valuation is different from that of Theorem 2.2 since it does not use Lorenzen's embedding.)

Note that the extended valuations defined in Proposition 2.1 are not valuations defined by gradings, in the sense of Sect. 1. However, if one takes the initial valuation u to be the 0 valuation, as in 2.3, then the resulting version of Theorem 2.2 is valid for graded domains rather than just for domains of the form $k[\Gamma]$; in fact, a slight modification of the proof of Theorem 2.2 yields the following:

Theorem 2.4. *Let $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$ be a graded domain and $\phi: \Gamma \to B$ be a homomorphism of the monoid Γ into a lattice ordered group B. Then the map $v: D^* \to B$ defined by* (1.1) *(where the inf is to be computed in B) is a semi-valuation.*

If each D_γ is non-zero and $\langle \phi(\Gamma) \rangle$, the group generated by $\phi(\Gamma)$, is B, then the usual proof ([10, p. 590] or [5]) shows that the semi-value ring R_v of the semi-valuation v of Theorem 2.4 is Bezout, i.e., finitely generated ideals are principal:

Let a, c be elements of R_v^* which do not differ by a unit of R_v. Since $B = \langle \phi(\Gamma) \rangle$, there are $\gamma_i \in \Gamma$ so that $v(a) = \phi(\gamma_1) - \phi(\gamma_3)$ and $v(c) = \phi(\gamma_2) - \phi(\gamma_3)$. Choose $d_{\gamma_i} \in D^*_{\gamma_i}$. Then $a' = d_{\gamma_1}/d_{\gamma_3}$, $c' = d_{\gamma_2}/d_{\gamma_3}$ differ from a, c respectively by units of R_v. Since $v(a'+c') = \inf\{\phi(\gamma_1), \phi(\gamma_2)\} - \phi(\gamma_3) = \inf\{\phi(\gamma_1) - \phi(\gamma_3), \phi(\gamma_2) - \phi(\gamma_3)\} = \inf\{v(a'), v(c')\}$, it follows that $(a'+c')R_v = (a', c')R_v = (a, c)R_v$. Q.E.D.

One would also like to find generalizations of Proposition 2.1 and Theorem 2.2 to arbitrary graded domains, but it is not clear in general how to extend a valuation u of the domain D_0 to one of $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$. However, if one passes to D_H, then the problem can be solved; and this is what we shall take up next.

3. Valuations of Twisted Group Rings

Fix once again a graded domain $D = \bigoplus_{\gamma \in \Gamma} D_\gamma$, where Γ is a torsionless, cancellative monoid with group hull $\langle \Gamma \rangle = G$; and $H = \{\Gamma\text{-homogeneous elements of } D^*\}$. We shall also assume now that $D_\gamma \neq 0$ for all $\gamma \in \Gamma$; this condition can always be attained by replacing Γ by a submonoid. The quotient ring D_H has a G-grading:

$D_H = \bigoplus_{g \in G} D'_g$, where $D'_g = \{a/b | a,\ 0 \neq b$ are homogeneous elements of D with $\deg a - \deg b = g\}$. The domain D_H has the property that every non-zero homogeneous element is a unit. Conversely, if D has the property that every non-zero homogeneous element is a unit, then $D = D_H$, $\Gamma = G$, and $D_g = D'_g$ for all $g \in G$. Graded domains with this property have been studied (for example, superficially in [1, Sect. 3], and as part of a general theory in [12, Kap. II]).

Definition 3.1. A G-graded domain $D=\bigoplus_{g\in G} D_g$ is called a *twisted group ring* of G over a field K if K is a subring of D and for every $g\in G$, there is a non-zero X^g in D_g such that $D_g=KX^g$.

It follows from the definition that $D_g\neq 0$ for all $g\in G$. Moreover, $K=D_0$; for, $1\in D_0$ implies $1=\alpha X^0$, $\alpha\in K^*$, and thus $X^0=1/\alpha\cdot 1\in K$ and $D_0=KX^0=K$.

Proposition 3.2 (cf. [12, p. 40]). *Let $D=\bigoplus_{g\in G} D_g$ be a G-graded domain, G a torsion-free group, such that every D_g is non-zero. If D has the property that every non-zero homogeneous element is a unit, then D_0 is a field and D is a twisted group ring of G over D_0.*

Proof. If $d_g\in D_g$, $d_h\in D_h$, and $d_gd_h=1$, then $g+h=0$ since $1\in D_0$. Thus $g=-h$; so if every non-zero homogeneous element is a unit, then D_0 is a field.

Since $D_g\neq 0$, we may choose $X^g\neq 0$ in D_g. Then $1/X^g\in D_{-g}$ as we have just seen; and thus for $d_g\in D_g$, $d_g(1/X^g)\in D_0$. Hence, $d_g\in D_0X^g$ and $D_g=D_0X^g$. Q.E.D.

Thus, if we consider only domains $D_H=\bigoplus_{g\in G} D'_g$ such that each D'_g is non-zero, then we are dealing with exactly the twisted group rings of G over a field.

Suppose now we are given a lexicographically exact (lex) sequence (e): $1\to A_u\to B\to C\to 0$ of ordered groups, with C totally ordered and A_u the semi-value group of a semi-valuation u of a field K. (Note: A_u and B will be written multiplicatively and C additively.) We want to extend u to a semi-valuation v having semi-value group B. First recall that (e) "lifts along u" to a short exact sequence (e'): $1\to K^*\to B'\to C\to 0$ making the following diagram commutative [3, p. 108], [7, p. 75]:

$$\begin{array}{lccccccc} (e') & 1 & \longrightarrow & K^* & \longrightarrow & B' & \longrightarrow & C\longrightarrow 0 \\ & & & \downarrow u & & \phi\downarrow & & \| \\ (e) & 1 & \longrightarrow & A_u & \longrightarrow & B & \longrightarrow & C\longrightarrow 0 \quad \text{lex} \end{array} \tag{3.3}$$

(where C is totally ordered).

We can associate a twisted group ring $K^{e'}[C]$ to the exact sequence (e') as follows: For each $c\in C$ choose a preimage $X^c\in B'$. Then $X^{c_1}\cdot X^{c_2}=\tau(c_1,c_2)X^{c_1+c_2}$ for some $\tau(c_1,c_2)\in K^*$. Now form $K^{e'}[C]=\bigoplus_{c\in C} KX^c$, where multiplication is defined distributively by $X^{c_1}X^{c_2}=\tau(c_1,c_2)X^{c_1+c_2}$.

Define a map $v:K^{e'}[C]^*\to B$ by

$$v(\alpha_1X^{c_1}+\ldots+\alpha_tX^{c_t})=\inf_{i=1,\ldots,t} \hat{B}\{u(\alpha_i)\phi(X^{c_i})\},$$

where $\alpha_i\in K^*$ and $c_i\neq c_j$ for $i\neq j$.

Note that if we arrange the exponents so that $c_1<\ldots<c_t$, then $v(\alpha_1X^{c_1}+\ldots+\alpha_tX^{c_t})=u(\alpha_1)\phi(X^{c_1})$ because of the lexicographic ordering.

Theorem 3.4. *v is a semi-valuation (which extends u and has semi-value group B).*

Proof. The proof that $v(f+g)\geqq \inf_B\{v(f),v(g)\}$ for $f, g\in K^{e'}[C]^*$ is the same as in the proof of Proposition 2.1. Consider then fg, where $f=\alpha_1X^{c_1}+\ldots+\alpha_tX^{c_t}$,

$g=\beta_1 X^{d_1}+\ldots+\beta_s X^{d_s}$, with α_i, $\beta_j \in K^*$ and $c_1<\ldots<c_t$, $d_1<\ldots<d_s$. Then $fg=\alpha_1\beta_1\tau(c_1,d_1)X^{c_1+d_1}+$ (terms with exponent $>c_1+d_1$). Thus

$$v(fg)=u(\alpha_1\beta_1\tau)\phi(X^{c_1+d_1})=u(\alpha_1)u(\beta_1)[u(\tau)\phi(X^{c_1+d_1})]=u(\alpha_1)u(\beta_1)[\phi(\tau X^{c_1+d_1})]$$

$$=u(\alpha_1)u(\beta_1)[\phi(X^{c_1}X^{d_1})]=u(\alpha_1)u(\beta_1)\phi(X^{c_1})\phi(X^{d_1})=v(f)v(g). \quad \text{Q.E.D.}$$

Corollary 3.5. *Any lexicographic extension of a semi-value group by a totally ordered group is a semi-value group.*

The "totally ordered" hypothesis of this corollary cannot be weakened to "lattice ordered" [10, p. 587].

There is a certain uniqueness in the way the above v is defined, in that there is only one extension of u to a semi-valuation v of $K^{e'}[C]^* \to B$ with the property that $v(X^c)=\phi(X^c)$. For, such a v has the property that $v(\alpha_1 X^{c_1}+\ldots+\alpha_t X^{c_t})=v(\alpha_1 X^{c_1})$ whenever $c_1<\ldots<c_t$; cf. [10, 6.2].

Application 3.6. The present investigation arose with the question of whether there are non-split lex semi-value sequences $1\to A_u\to B_v\to C_w\to 0$ with A_u trivially ordered (i.e., $A_u^+=0$) and C_w totally ordered. Lex semi-value sequences of this type are exactly those of the form $1\to K^*/k^*\to B_v\to C_w\to 0$, where C_w is value group of some valuation w of a field L and k is a subfield of the residue field K of w [10, Sect. 4]. If one uses the group ring $K[C_w]$ to define the valuation w, then the resulting sequence splits [10, 3.5]. For an explicit example of such a sequence which does not split, proceed as follows: Let C be any torsion-free group which is not free, and totally order C. There is a short exact sequence $1\to F_2\to F_1\to C\to 0$ with F_1, F_2 free groups, and this sequence does not split since C, being non-free, cannot be a subgroub of a free group. Put the trivial order on F_2 and the lexicographic order on F_1. Let $k=\mathbb{Z}/2\mathbb{Z}$, and let $K=k(\{t_i\})$, where $\{t_i\}$ is a set of indeterminates whose cardinality equals the rank of F_2. Then $K^*/k^*\cong K^*\cong$ free group on the irreducible elements of $k[\{t_i\}]\cong F_2$. Thus $1\to F_2\to F_1\to C\to 0$ is a non-split semi-value sequence of the desired type.

These remarks apply, in particular, to the example worked out in Example 1.5. We have there a domain $D=k[X_{c_1},X_{c_2},\ldots]$ and an inf valuation w defined by $w(X_{c_i})=c_i$, and we know that the corresponding inf valuation of $D_{H'}$ is also w. If we assume, in addition, that the group C is generated by $c_1,c_2,\ldots$, then by Propositions 3.2 and 1.3, $D_{H'}$ is a twisted group ring of C over $(D_{H'})_0=k_w$. We showed in Example 1.5 that $k_w=k(Z_1,Z_2\ldots)$, an infinite pure transcendental extension of k; so if k is chosen to be $\mathbb{Z}/2\mathbb{Z}$ as above and $u:k_w^*\to k_w^*/k^*$ to be the identity semi-valuation, then the resulting semi-value group $A_u=k_w^*$ is free. The canonical map $v:L^*\to(L^*/1+m_w)=B_v$ is a semi-valuation since $1+m_w=U(k+m_w)$, and this v may be seen to be equivalent to a semi-valuation defined as in Theorem 3.4 by means of the twisted group ring $D_{H'}$. Viewed from this perspective, the homomorphism v is easily seen to split, so that B_v is actually a subgroup of L^*. But L is a pure transcendental extension of k, from whence it follows that L^*, and hence also B_v, is free. Therefore the corresponding lex semi-value sequence $1\to A_u\to B_v\to C\to 0$ (cf. Sect. 4) is a free resolution of C as above. (The second author intends to enlarge on the ideas behind this example in a future paper).

Another proof that $1\to A_u\to B_v\to C\to 0$ does not split runs as follows: If the sequence splits, then $D_{H'}\cong(D_{H'})_0[C]$ by remarks in Sect. 4. However, this is impossible because D, and hence $D_{H'}$, is a UFD, while $(D_{H'})_0[C]$ is not a UFD [4].

Remark 3.7. The following was proved in [10, 4.1]: Suppose $1\to A\to B\xrightarrow{\beta} C\to 0$ is lex, with C totally ordered. If v is a semi-valuation of a field L having semi-value group B and ring R_v, then $w=\beta v$ is a valuation and A is the semi-value group of a semi-valuation u of the residue field k_w of w. Moreover, if R_w contains a copy of k_w, then $R_v=R_u+m_w$.

One consequence is that Corollary 3.5 may be restated: If $1\to A\to B\to C\to 0$ is lex with C totally ordered, then B is a semi-value group if and only if A is a semi-value group.

Another application has been given by Mott [8], who has applied this observation to the case that B is lattice ordered, defining v to be the inf map of the group ring of B (over some field), which is known to be a semi-valuation by the Kaplansky-Jaffard theorem (cf. 2.3). He thereby establishes the existence of lex semi-value sequences of the above type for B lattice ordered. In addition, he proves that the resulting R_w contains a copy of k_w, which is also the form of the valuation rings R_w resulting from our construction.

4. Twisted Group Rings

We shall conclude by sketching the relationship between twisted group rings, short exact sequences, and the semi-valuations defined in Sect. 3. Most of these facts are well-known, but the valuation-theoretic approach is probably new.

Let K be a field and C a torsion-free group. A *K-twisting* of C is a map $\tau: C\times C\to K^*$ such that

(i) $\tau(a,b)=\tau(b,a)$, and

(ii) $\tau(a,b)\tau(a+b,c)=\tau(b,c)\tau(a,b+c)$ for all $a,b,c\in C$.

The elements $\tau(a,b)\in K^*$ are called the *twisting coefficients* of τ.

Two such twistings are called *equivalent*, $\tau\sim\tau'$, (called "diagonally equivalent" in [11, p. 15]) if there is a function $\delta: C\to K^*$ such that

$$\tau(a,b)=[\delta(a)\delta(b)/\delta(a+b)]\tau'(a,b) \quad \text{for all} \quad a,b\in C.$$

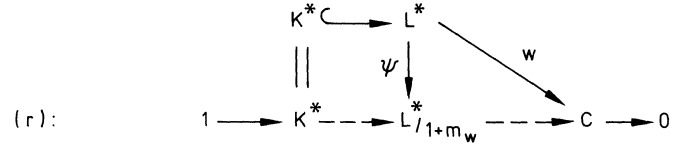
The *twisted group ring* of C over K with respect to the twisting τ, denoted $K^\tau[C]$, is defined as follows: $K^\tau[C]=\bigoplus_{c\in C} KX^c$, the K-vector space with basis $\{X^c\}_{c\in C}$, with multiplication defined distributively via $X^aX^b=\tau(a,b)X^{a+b}$. Condition (ii) above insures that the multiplication is associative and condition (i) that it is commutative. For example, if all the twisting coefficients are 1, then we get just the group ring $K[C]$. One proves $K^\tau[C]$ is a domain when C is torsion free by the same proof as for $K[C]$, using the fact that a torsion-free group can be totally ordered. The identity element of $K^\tau[C]$ is easily seen to be $(1/\tau(0,0))X^0$, and the map $K\to K^\tau[C]$ defined by taking $\alpha\in K$ to $\alpha(1/\tau(0,0))X^0$ identifies K with KX^0.

For twisting τ and τ', $\tau\sim\tau'$ if and only if $K^\tau[C]\cong K^{\tau'}[C]$, where $\cong$ denotes an isomorphism of graded K-algebras, i.e., homogeneous elements of deg c correspond to homogeneous elements of deg c. The proof is routine, the idea being

that given δ, one maps X^c to $\delta(c)X^c$; and conversely, given an isomorphism taking X^c to αX^c, one defines δ by $\delta(c)=\alpha$.

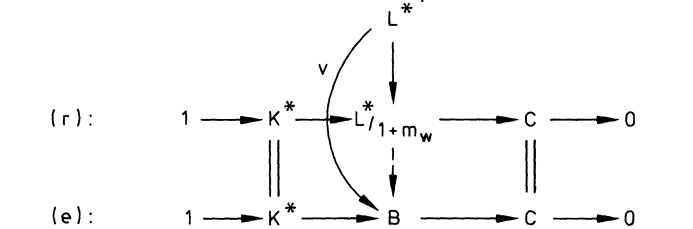
Given any short exact sequence of the form (e): $1\to K^*\to B\xrightarrow{\beta} C\to 0$, choose a set of C-preimages in B, $\{b^c\}_{c\in C}$. Such a set of preimages gives rise to a twisting τ by defining $\tau(c,c')\in K^*$ via $b^c b^{c'}=\tau(c,c')b^{c+c'}$. This τ satisfies (i) since B is abelian and (ii) since B is associative. (Such a function τ arising from a short exact sequence is more commonly called a "factor set", cf. [3, p. 210], [7, pp. 69 and 111]; we felt the "twisting" terminology of [11, p. 14] was better suited to the present discussion.) If τ_1 is given by another set of C-preimages $\{b_1^c\}_{c\in C}$, then $b_1^c=\delta(c)b^c$ for some $\delta(c)\in K^*$, since $\beta(b_1^c)=\beta(b^c)$ implies $b_1^c/b^c\in K^*$. It is trivial to check that this δ is an equivalence function, so $\tau_1\sim\tau$. More generally, the sets of C-preimages for any exact sequence (e_1): $1\to K^*\to B_1\to C\to 0$ which is equivalent to (e) (cf. [7, p. 64]) yield twistings equivalent to τ. Thus, we have defined a map Φ: Ext$(C,K^*)\to\mathcal{G}=\{$graded K-algebra isomorphism classes of twisted group rings $K^\tau[C]\}$. The $K^\tau[C]$ corresponding to (e) will sometimes be written $K^e[C]$ when we want to emphasize the short exact sequence rather than the twisting.

We shall next verify that Φ has left and right inverses. Given a twisted group ring $K^\tau[C]$, we can associate to it a "canonical" extension as follows: Let L be the quotient field of $K^\tau[C]$ and let $w: K^\tau[C]\to C$ be the inf valuation (with respect to some fixed total order of C) defined by $w(X^c)=c$. Then $C\cong L^*/(K^*+m_w)$ by Proposition 1.3, so there are unique homomorphisms which make the bottom row (r) of the following commutative diagram exact:



Here ψ is the canonical homomorphism. On the other hand, if we choose the set of C-preimages in $L^*/1+m_w$ consisting of the elements $\{\psi(X^c)\}_{c\in C}$, then the twisting coefficients given by this set are the same as the twisting coefficients defining τ; so the twisted group ring corresponding to the exact sequence (r) is just $K^\tau[C]$. We have thus defined a map Φ': $\mathcal{G}\to$ Ext(C,K^*) such that $\Phi\circ\Phi'=$ identity.

Finally, let us check that $\Phi'\circ\Phi=$ identity also. Suppose we are given a short exact sequence (e): $1\to K^*\to B\to C\to 0$ with associated twisted group ring $K^e[C]$. Let $v: K^e[C]^*\to B$ be the multiplicative monoid homomorphism defined as in Theorem 3.4 by $v(\alpha_1 X^{c_1}+\ldots+\alpha_t X^{c_t})=\alpha_1 X^{c_1}$ if $c_1<\ldots<c_t$; v extends uniquely to a group homomorphism $v: L^*\to B$. Since $1+m_w\subset\ker v$, this homomorphism factors through $L^*/1+m_w$ to give the commutative diagram



The homomorphism ———→ is then automatically an isomorphism, so $(r)\sim(e)$.

Thus $\Phi' \circ \Phi = \text{identity} = \Phi \circ \Phi'$; so Φ is one-to-one and onto. Instead of using a valuation to get the canonical sequence (r) as we have done, one could also establish the existence of a short exact sequence corresponding to a given twisting by means of the observation that the B occurring in (e) above is just the unit group of $K^e[C]$; cf. [3-II, p. 321]. If char $K=2$, this observation also shows that $U(K^e[C]) \cong L^*/1+m_w = \text{group of divisibility of the domain } k+m_w$, where $k=\mathbb{Z}/2\mathbb{Z}$.

One further fact that follows from the above discussion is that $K^e[C]$ is isomorphic as a graded K-algebra to the group ring $K[C]$ if and only if the sequence (e) splits. A special case of this appears in [1, Theorem 3.1].

References

1. Anderson, D.F.: Graded Krull domains. Comm. Algebra **7**, 79–106 (1979)
2. Anderson, D.F., Dobbs, D.: Pairs of rings with the same prime ideals. Canad. J. Math. **32**, 362–384 (1980)
3. Fuchs, L.: Infinite Abelian groups, Vol. I. New York: Academic Press 1970; Vol. II, 1973
4. Gilmer, R., Parker, T.: Divisibility properties in semigroup rings. Michigan Math. J. **21**, 65–86 (1974)
5. Heinzer, W.: J-Noetherian integral domains with 1 in the stable range. Proc. Am. Math. Soc. **19**, 1369–1372 (1968)
6. Jaffard, P.: Les systèmes d'idéaux. Paris: Dunod 1960
7. MacLane, S.: Homology. New York: Academic Press 1963
8. Mott, J.L.: Nonsplitting sequences of value groups. Proc. Am. Math. Soc. **44**, 39–42 (1974)
9. Northcott, D.: Lessons on rings, modules, and multiplicities. Cambridge: Cambridge University Press 1968
10. Ohm, J.: Semi-valuations and groups of divisibility. Canad. J. Math. **21**, 576–591 (1969)
11. Passman, D.: The algebraic structure of group rings. New York: Wiley-Interscience 1977
12. Schiffels, G.: Graduierte Ringe und Moduln. Bonn. Math. Schr. **11**, 1–122 (1960)

Received March 1, 1979

Math. Ann. 256, 157–189 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Réduction algébrique d'un morphisme faiblement Kählérien propre et applications

F. Campana

Départment de Mathématiques, Université Nancy 1, F-54037 Nancy Cedex, France

Introduction

Les espaces analytiques considérés ici sont réduits et dénombrables à l'infini. Si X est un tel espace, $\mathscr{C}(X)$ désigne l'espace analytique des cycles compacts de dimension pure de X. Si X est compact et irréductible, $a(X)$ désigne la dimension algébrique de X.

Dans la première partie, on étudie la réduction algébrique d'un morphisme faiblement Kählérien[1] propre $\varphi : X \to S$: le théorème 1 assure l'existence d'une telle réduction algébrique, qui est une factorisation méromorphe de φ induisant sur la fibre générale de φ une réduction algébrique[1] de celle-ci. On montre ensuite que si la fibre générale de φ est algébrique, le morphisme φ est algébrique, c'est à dire (localement sur S) S-biméromorphe à un morphisme projectif.

Dans la seconde partie, en itérant la réduction algébrique du morphisme constant $c : X \to \{x\}$, pour X espace faiblement Kählérien compact, on obtient un diagramme d'application méromorphes, défini à équivalence biméromorphe près, dont la première flèche est une réduction algébrique de X. Ce diagramme conduit à de nouveaux invariants biméromorphes de X.

Dans la troisième partie, on applique les résultats précédents pour montrer l'existence d'une application méromorphe surjective $\psi : X \dashrightarrow X'$ dont la fibre générale est irréductible de dimension algébrique zéro, et telle que tout sous-espace compact et irréductible Y de X tel que $a(Y) = 0$ soit contenu dans l'une des fibres de ψ. Enfin, le théorème 4 montre que toute famille analytique pleine de cycles génériquement irréductibles de X, est la famille des fibres d'une application méromorphe multiforme si l'un des cycles est de dimension algébrique zéro.

Définitions, notations

Les espaces analytiques considérés ici sont réduits et dénombrables à l'infini. Si X est un espace analytique, on note $\mathscr{C}(X)$ [resp. $\mathscr{C}_p(X)$ si $p \in \mathbb{N}$] l'espace analytique des cycles (resp. des cycles de dimension pure p) de X (voir [B1]). L'espace analytique

1 Voir définition ci-dessous

0025-5831/81/0256/0157/$06.60

$\mathscr{C}(X)$ est dénombrable à l'infini ([B2, corollaire 3 du théorème 2]). On désignera par $\mathscr{C}^*(X)$ [resp. $\mathscr{C}_p^*(X)$] l'ouvert de Zariski de $\mathscr{C}(X)$ (voir appendice 2) constitué des cycles de X dont le support est irréductible (resp. et de dimension p) et de multiplicité 1. On note $\overline{\mathscr{C}^*(X)}$ [resp. $\overline{\mathscr{C}_p^*(X)}$] l'adhérence de $\mathscr{C}^*(X)$ [resp. $\mathscr{C}_p^*(X)$] dans $\mathscr{C}(X)$.

Si $\varphi : X \to S$ est un morphisme d'espaces analytiques, on notera $\mathscr{C}(X/S)$ le sous-espace analytique fermé de $\mathscr{C}(X)$ (voir [F1, Proposition 3.2]) constitué des cycles relatifs de X sur S (pour φ), et $\varphi_* : \mathscr{C}(X/S) \to S$ le morphisme d'image directe par φ. Rappelons que les éléments de $\mathscr{C}(X/S)$ sont les cycles Z de X contenus dans une fibre de φ, et l'on a alors $\varphi_*(Z) = \varphi(|Z|)$ où $|Z|$ est le support du cycle Z. On notera aussi respectivement $\mathscr{C}_p(X/S)$, $\mathscr{C}_p^*(X/S)$ et $\overline{\mathscr{C}_p^*(X/S)}$ l'intersection de $\mathscr{C}(X/S)$ avec $\mathscr{C}_p(X)$, l'intersection de $\mathscr{C}_p(X/S)$ avec $\mathscr{C}^*(X)$, et l'adhérence dans $\mathscr{C}(X)$ de $\mathscr{C}_p^*(X/S)$.

Si $(Z_s)_{s\in S}$ est une famille analytique de cycles de X paramétrée par un espace analytique (réduit) S, on notera $Z_S \subset S \times X$ le graphe de la famille $(Z_s)_{s\in S}$, $p_X : Z_S \to X$ et $p_S : Z_S \to S$ les projections naturelles. On notera aussi $X_S = p_X(Z_S)$ la réunion des supports des cycles de la famille $(Z_s)_{s\in S}$, et $K_S : X \dashrightarrow \mathscr{C}(S)$ l'application méromorphe de Kodaira (voir appendice 1) de la famille $(Z_s)_{s\in S}$ lorsque Z_S est X-propre. Si $j : S \to \mathscr{C}(X)$ est le morphisme classifiant la famille $(Z_s)_{s\in S}$, on notera, lorsqu'il n'y a pas d'ambiguité, S^* l'ouvert de Zariski de S égal à l'image réciproque par j de $\mathscr{C}^*(X)$, et $\overline{S^*}$ l'adhérence dans S de S^*. La notation $\overline{S^*} = S$ signifie donc que, pour s générique dans S, le cycle Z_s est de support irréductible et de multiplicité un.

Un espace analytique est dit faiblement Kählérien (resp. algébrique) s'il existe un morphisme surjectif $\varphi : X' \to X$ dans lequel X' est une variété Kählérienne compacte (resp. une variété algébrique projective[2]). En particulier, X est compact.

La classe des espaces faiblement Kählériens (resp. algébriques) est stable par image par un morphisme, sous-ensembles analytiques fermés, image réciproque par un morphisme algébrique (voir définition ci-après) et surtout [C2, corollaire 3]:

Théorème[3]. *Si X est un espace analytique faiblement Kählérien (resp. algébrique) les composantes irréductibles de $\mathscr{C}(X)$ sont des espaces faiblement Kählériens (resp. algébriques).*

I. Décomposition algébrique d'un morphisme faiblement Kählérien propre et applications

A. Réduction algebrique d'un morphisme faiblement Kählerien propre

Définition 1. Soit X un espace analytique. Un sous-ensemble E de X est dit *analytiquement maigre* dans X s'il est contenu dans une réunion dénombrable de sous-espaces analytiques fermés de X d'intérieur vide.

Un sous-ensemble E de X qui est analytiquement maigre dans X est donc maigre dans X.

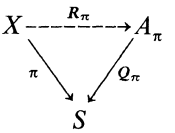
---

2 Dans ce dernier cas on peut donc toujours supposer que φ est une modification. Il serait intéressant de savoir si tout espace faiblement Kählérien a, de manière analogue, une désingularisation qui est une variété Kählérienne

3 Résultat obtenu indépendamment par Fujiki

Soient $\varphi : X \dashrightarrow Y$ une application méromorphe, et X_1 un sous-espace analytique fermé de X. On dira que la restriction de φ à X_1 est une application méromorphe si la restriction de $m_\varphi : X_\varphi \to X$, la projection sur X du graphe $X_\varphi \subset X \times Y$ de φ, à $(m_\varphi)^{-1}(X_1)$ est une modification de X_1.

Définition 2. Soit $\pi : X \to S$ un morphisme propre et surjectif d'espaces analytiques irréductibles. Une application méromorphe au-dessus de S:



est appelée une réduction algébrique de π si:

- Q_π est un morphisme dont les fibres sont des espaces algébriques.
- Il existe un sous-ensemble analytiquement maigre M_π de S tel que, pour tout s de $S - M_\pi$, et toute composante irréductible X_s^0 de $X_s = \pi^{-1}(s)$, la restriction de Q_π à X_s^0 est une application méromorphe qui est une réduction algébrique de X_s^0 (voir [U, p. 24]).

On note $a(X)$ la dimension algébrique d'un espace analytique compact et irréductible (voir [U, p. 24]).

Soit $\pi : X \to S$ un morphisme propre et surjectif d'espaces analytiques irréductibles. Pour tout s de S, on posera: $a(s, \pi) = \sup\{a(X_s^0)$, lorsque X_s^0 décrit l'ensemble, fini, des composantes irréductibles de $X_s = \pi^{-1}(s)\}$ et $a(\pi) = \inf_{s \in S}\{a(s, \pi)\}$.

Si π admet une réduction algébrique, alors on a: $a(\pi) = \dim_{\mathbb{C}}(A_\pi) - \dim_{\mathbb{C}}(S)$, d'après le lemme 4 de la démonstration du théorème 1 ci-dessous.

Dans [F1, p. 32, Definition 4.1], Fujiki a défini la notion de morphisme Kählérien, qui est la notion relative de la notion d'espace (fortement) Kählérien.

Nous dirons ici qu'un morphisme $\pi : X \to S$ est faiblement Kählérien s'il existe un morphisme propre $\pi : Y \to S$ qui est un «Kähler morphism» au sens de [F1, Definition 4.1], et un morphisme surjectif $\varphi : Y \to X$ au-dessus de S. Dans [F1], un tel morphisme est appelé: morphisme de la catégorie $\mathscr{C}/S$.

Fujiki a démontré dans [F1, Théorème 4.5, p. 33 et Proposition 4.8, p. 36] que, si $\pi : X \to S$ est un morphisme faiblement Kählérien au sens ci-dessus, alors les composantes irréductibles de $\mathscr{C}(X/S)$ sont S-propres pour π_*, le morphisme d'image directe par π.

Ce résultat est une version relative du fait que les composantes irréductibles de l'espace des cycles d'un espace faiblement Kählérien sont compactes.

Théorème 1. *Soit $\pi : X \to S$ un morphisme faiblement Kählérien*[4] *propre et surjectif d'espaces analytiques irréductibles. Alors:*

1. *Il existe une réduction algébrique de π, et l'ensemble des s de S tels que $a(s) > a(\pi)$ est analytiquement maigre dans S.*

4 Cette hypothèse est utilisée de manière essentielle dans la démonstration ci-dessous. Il n'est pas certain cependant qu'elle soit nécessaire

2. *Pour tout s de S, et toute composante irréductible* X_s^0 *de* $X_s = \pi^{-1}(s)$, *on a l'inégalité:*

$$a(X_s^0) \geqq a(\pi) + \{\dim_{\mathbb{C}}(X_s^0) - (\dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S))\}.$$

3. *Si* π *est géométriquement plat (voir* [B3]*), pour tout entier k, l'ensemble* S_k *des s tels que* $a(s) \geqq k$ *est une réunion dénombrable de sous-espaces analytiques fermés de S.*

Remarque. L'exemple des familles de tores montre qu'il n'est pas possible de faire mieux, en général.

Ce résultat admet les conséquences immédiates suivantes:

Corollaire 1. *Soit* $\pi : X \to S$ *un morphisme faiblement Kählérien propre et surjectif d'espaces analytiques irréductibles. S'il existe un* $s_0 \in S$, *et une composante irréductible* $X_{s_0}^0$ *de* $X_{s_0} = \pi^{-1}(s_0)$ *tels que* $a(X_{s_0}^0) = 0$, *il existe alors un sous-ensemble S' de S dont le complémentaire est analytiquement maigre dans S, et tel que, pour tout s de S' et toute composante irréductible* X_s^0 *de* $X_s = \pi^{-1}(s)$ *on ait:* $a(X_s^0) = 0$.

Corollaire 2. *Soit* $\pi : X \to S$ *un morphisme faiblement Kählérien propre et surjectif d'espaces analytiques irréductibles.*

S'il existe un sous-ensemble S' de S, qui n'est pas analytiquement maigre dans S, et tel que pour tout s de S' il existe une composante irréductible X_s^0 *de* $X_s = \pi^{-1}(s)$ *qui est un espace algébrique, alors pour tout s de S, l'espace analytique* $X_s = \pi^{-1}(s)$ *est algébrique.*

C'est-à-dire que le morphisme π est à fibres algébriques. Ceci entraîne, d'après le corollaire 1 du théorème 2 ci-dessous que π est, localement sur S, un morphisme algébrique, dans le sens défini ci-dessous.

Corollaire 3. *Soit X un espace analytique faiblement Kählérien compact. Le sous-ensemble* $\mathscr{C}_A(X)$ *de* $\mathscr{C}(X)$ *constitué des cycles de X dont le support est un espace algébrique est une réunion dénombrable de sous-espaces analytiques fermés de* $\mathscr{C}(X)$.

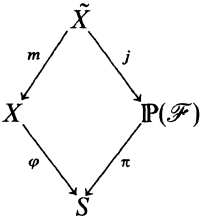
Démonstration du corollaire 3. Soit $Z_X \subset \mathscr{C}(X) \times X$ le graphe de la famille universelle de cycles de X. Puisque X est un espace faiblement Kählérien, le morphisme de première projection $p_{\mathscr{C}(X)} : Z_X \to \mathscr{C}(X)$ est faiblement Kählérien. D'autre part, il est propre, surjectif et géométriquement plat. L'ensemble des $c \in \mathscr{C}(X)$ tels que *l'une* des composantes irréductibles de $|Z_c|$ soit algébrique est donc une réunion au plus dénombrable de sous-espaces analytiques fermés de $\mathscr{C}(X)$.

Pour chacun de ces sous-espaces, on restreint la famille Z_X, à ce sous-espace, et on considère séparément chacune des composantes irréductibles du graphe de la famille ainsi restreinte, dont les fibres ne sont pas toutes algébriques. On applique encore le théorème 1, puis on recommence l'opération. La réunion des sous-espaces analytiques fermés de $\mathscr{C}(X)$ ainsi obtenus est dénombrable et coïncide avec $\mathscr{C}_A(X)$. CQFD

Le théorème 2 ci-dessous décrit la structure des morphismes propres faiblement Kählériens[5] à fibres algébriques: ils sont, localement sur la base, algébriques au sens suivant:

5 La conclusion du théorème 2 ne subsiste pas sans cette hypothèse (voir les remarques suivant le théorème 2 de [C5]). Il est cependant probable que la conclusion du corollaire subsiste sans cette hypothèse

Définition. Un morphisme $\varphi : X \to S$ est dit algébrique s'il est propre, et s'il existe un faisceau analytique cohérent $\mathscr{F}$ sur S, et un diagramme commutatif:



dans lequel:

- m est une modification propre.
- $\mathbb{P}(\mathscr{F}) \xrightarrow{\pi} S$ le projectifié du fibré linéaire associé au faisceau $\mathscr{F}$ (voir [F, p. 51]).
- $j : \tilde{X} \to j(\tilde{X})$ un morphisme biméromorphe de $\tilde{X}$ sur son image par j [qui est analytique fermée et S-propre dans $\mathbb{P}(\mathscr{F})$].

Les morphismes algébriques possèdent des propriétés de stabilité, versions relatives des propriétés de stabilité, énoncées cidessus pour les espaces algébriques (voir [F2]). En particulier:

Proposition 0. *Soient $f : X \to Y$ et $g : Y \to Z$ des morphismes algébriques. Alors $(g \circ f)$ est un morphisme algébrique.*

Démonstration (voir [F2]). Un morphisme algébrique est évidemment faiblement Kählérien et à fibres algébriques. Réciproquement:

Théorème 2. *Soit $\pi : X \to S$ un morphisme propre, dans lequel X est irréductible. Pour que π soit un morphisme algébrique (au sens ci-dessus), il faut et il suffit que les trois conditions suivantes soient réalisées:*

1. *Les fibres de π sont des espaces algébriques.*
2. *Il existe un sous-espace analytique fermé Y de X tel que $\pi(Y) = \pi(X)$ et tel que la restriction de π à Y soit un morphisme algébrique (donc propre).*
3. *Le morphisme π est faiblement Kählérien.*

Ce résultat est une version relative du théorème 2 du chapitre III. On peut lui appliquer les remarques qui suivent l'énoncé de ce dernier théorème, et énoncer un théorème 2′ analogue au théorème 2′ du chapitre III.

Enfin la démonstration du théorème 2 ci-dessus est la même que celle du théorème 2 du chapitre III de [C4] en remplaçant le mot «compact» par S-propre, et en utilisant la version relative du théorème d'algébricité, ainsi que la proposition ci-dessus.

Corollaire. *Soit $\pi : X \to S$ un morphisme propre et faiblement Kählérien à fibres algébriques. Pour tout s de S, il existe un voisinage ouvert U de s dans S tel que la restriction π_U de π à $X_U = \pi^{-1}(U)$, $\pi_U : X_U \to U$, soit un morphisme algébrique.*

Démonstration du corollaire. Soit s un point de S. Il existe alors un voisinage U de s dans S et une section méromorphe multiforme $\zeta : U \dashrightarrow X_U$ de π_U telle que, pour

toute composante irréductible X_U^0 de $X_U = \pi^{-1}(U)$, il existe une composante irréductible $\zeta^0(U)$ de $\zeta(U)$ telle que:

- $\zeta^0(U)$ est contenue dans X_U^0,
- $\zeta^0(U)$ est propre sur U par π_U,
- $\pi(\zeta^0(U)) = \pi(X_U^0)$.

Mais la proposition 0 précédente montre que la restriction de π à $\zeta^0(U)$ est un morphisme algébrique puisqu'une modification propre et un morphisme propre et fini sont des morphismes algébriques. Le théorème 2 montre alors que la restriction de π à X_U^0 est un morphisme algébrique. Le morphisme π_U est donc aussi algébrique. CQFD

Démonstration du théorème 1. Esquissons la démonstration du théorème 1: on peut supposer X et S irréductibles, et φ géométriquement plat. Si $d = \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S)$, alors l'espace analytique $\mathscr{C}_{d-1}(X/S)$ des diviseurs relatifs de X sur S a des composantes irréductibles S-propres dont les fibres sur S sont algébriques, par le lemme 1 ci-dessous.

D'autre part, le lemme 2 ci-dessous montre que, pour chaque s de S, il existe un sous-espace analytique compact Γ_s de $\mathscr{C}_{d-1}(X/S)$ tel que l'application méromorphe de Kodaira K_{Γ_s} (voir appendice 1) associée à Γ_s induise sur $X_s = \pi^{-1}(s)$ une réduction algébrique de X_s. Le théorème de Baire montre alors, puisque $\mathscr{C}_{d-1}(X/S)$ est dénombrable à l'infini, l'existence d'une composante irréductible Γ de $\mathscr{C}_{d-1}(X/S)$ telle que l'application méromorphe de Kodaira associée à Γ fournisse une réduction algébrique de π.

Ce raisonnement permet en fait seulement de montrer que $S' = S - S_{a(\pi)+1}$ n'est pas maigre, mais une étude plus fine montre que S' est une intersection dénombrable d'ouverts de Zariski denses de S; cette étude est basée sur les résultats préliminaires suivants:

Lemme 1. *Soit X un espace analytique compact (resp. normal) de dimension pure d. Toute composante irréductible (resp. connexe) de $\mathscr{C}_{d-1}(X)$ est compacte et algébrique (resp. algébrique projective).*

Démonstration. La proposition 3 de l'appendice 2 montre que l'on peut supposer X irréductible, et qu'il suffit d'établir le lemme pour les composantes irréductibles de $\overline{\mathscr{C}^*_{d-1}(X)}$.

Considérons le diagramme de réduction algébrique:

$$\begin{array}{ccc} X' & \xrightarrow{m} & X \\ {\scriptstyle q}\downarrow & & \\ A' & & \end{array}$$

dans lequel A' est une variété algébrique projective, et q un morphisme géométriquement plat.

Remarquons que, puisque m est une modification propre, lorsque X est normal, le lieu d'indétermination de l'application méromorphe $m^{-1}: X \dashrightarrow X'$ est de codimension au moins 2. (Voir, par exemple, [U, p. 16, Theorem 2.5].)

Il en résulte donc que $m_* : \mathscr{C}_{d-1}(X') \to \mathscr{C}_{d-1}(X)$ est un isomorphisme, d'après [B1, Théorème 6] et [B2, Proposition 6], si X est normal, et est une modification finie dans tous les cas. Soit Γ une composante irréductible de $\overline{\mathscr{C}^*_{d-1}(X')}$. Soit γ un point de Γ^*, et D_γ le diviseur correspondant. Deux cas sont possibles puisque q est géométriquement plat: $|q_*(D_\gamma)| = A'$, ou $q_*(D_\gamma)$ est un diviseur de A'.

Comme on peut supposer que X' est lisse, dans le premier cas, D_γ est un diviseur non-polaire de X' et il est donc isolé dans $\overline{\mathscr{C}^*_{d-1}(X')}$ par le résultat de finitude de [F–F]. Alors $\Gamma = \{\gamma\}$, et la conclusion du lemme 1 est vraie pour Γ. Dans la suite, on note $\overline{\mathscr{C}^*_{d-1}(X')'}$ la réunion des composantes irréductibles de $\overline{\mathscr{C}^*_{d-1}(X)}$ qui ne sont pas constituées de diviseurs non-polaires.

Dans le second cas, $|D_\gamma| = q^{-1}(q|D_\gamma|)$. On peut vérifier, à l'aide de [B2, Proposition 6], que l'on peut définir, puisque q est géométriquement plat, une application d'image réciproque de cycles $q^* : \overline{\mathscr{C}^*_{d-1}(A)} \to \overline{\mathscr{C}^*_{d-1}(X')'}$ qui est un inverse à q_*.

Ceci montre que $q_* : \overline{\mathscr{C}^*_{d-1}(X')'} \to \overline{\mathscr{C}^*_{d-1}(A')}$ est un isomorphisme analytique.

Ceci établit donc le lemme 1, d'après [B1, chapitre 4].

Lemme 2. *Soit X un espace analytique compact et irréductible de dimension d. Il existe une composante irréductible Γ de $\overline{\mathscr{C}^*_{d-1}(X)}$ telle que K_Γ soit une réduction algébrique de X, où K_Γ est l'application de Kodaira associée à Γ (voir appendice 1).*

Démonstration. Si

$$\begin{array}{ccc} X' & \xrightarrow{\ m\ } & X \\ {\scriptstyle q}\downarrow & & \\ A' & & \end{array}$$

est un diagramme de réduction algébrique de X jouissant des propriétés énoncées dans la démonstration du lemme 1, on voit en utilisant l'isomorphisme $q_* : \overline{\mathscr{C}^*_{d-1}(X')'} \to \overline{\mathscr{C}^*_{d-1}(A')}$ qu'il s'agit de montrer l'existence d'une famille diviseurs génériquement irréductibles de A' et qui sépare presque les points de A'.

- Si $\dim_{\mathbb{C}}(A') = 0$, il n'y a rien à démontrer.
- Si $\dim_{\mathbb{C}}(A') = 1$, la famille des points de A' sépare les points de A'.

Si $\dim_{\mathbb{C}}(A') \geqq 2$, et si $\mathscr{D}$ est un fibré en droites très ample sur A', la famille des diviseurs associés aux sections de $\mathscr{D}$ satisfait la condition, d'après le théorème de Bertini. CQFD

Construction de l'application T. Nous allons définir une applications T qui associe, à un sous-espace analytique fermé de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant les propriétés (1) et (2) énoncées ci-dessous, un sous-espace analytique fermé de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant également ces propriétés.

La propriété fondamentale de l'application T est exprimée dans le lemme 3 ci-dessous.

L'intérêt ici de l'application T est qu'elle permet, appliquée de manière répétée à un sous-espace analytique fermé convenable de $\overline{\mathscr{C}^*_{d-1}(X/S)}$, de construire un espace analytique Σ, réunion disjointe d'une famille dénombrable de sous-espaces analytiques fermés irréductibles et S-propres de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant les

propriétés exprimées dans les Propositions 1 et 2 ci-dessous, et qui entraînent immédiatement le théorème 1.

On suppose ici que le morphisme π est géométriquement plat, et l'on note: $Z \subset \overline{\mathscr{C}^*_{d-1}(X/S)} \underset{S}{\times} X$ le graphe de la famille universelle des diviseurs relatifs de X sur S pour π, avec $d = \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S)$.

On note aussi $\pi_* : \overline{\mathscr{C}^*_{d-1}(X/S)} \to S$ le morphisme d'image directe par π et $p : Z \to S$ [resp. $q : Z \to \overline{\mathscr{C}^*_{d-1}(X/S)}$] les restrictions à Z des projections de $\overline{\mathscr{C}^*_{d-1}(X/S)} \times S$ sur ses facteurs.

Soit $\pi_* : \sigma \to S$ un sous-espace analytique fermé irréductible S-propre de $\overline{\mathscr{C}^*_{d-1}(X/S)}$, où l'on a posé $d = \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S)$, et tel que:

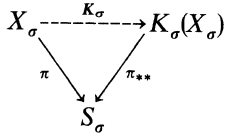
$$\overline{\sigma^*} = \sigma . \tag{1}$$

$$\dim_{\mathbb{C}}(\sigma) \geqq 1 + \dim_{\mathbb{C}}(\pi_*(\sigma)) . \tag{2}$$

On a alors aussi: $\dim_{\mathbb{C}}(X_\sigma) \geqq d + \dim_{\mathbb{C}}(\pi(X_\sigma))$, où $X_\sigma = p(q^{-1}(\sigma))$, c'est-à-dire l'ensemble des points de X par lesquels passe un cycle de la famille σ. (Remarquer que, puisque σ est S-propre, X_σ est un sous-espace analytique fermé S-propre de X.)

On va alors introduire de la façon suivante un sous-espace analytique fermé et S-propre $T(\sigma)$ de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ jouissant aussi de ces propriétés:

Soit



l'application méromorphe de Kodaira associée à σ (voir remarque finale de l'appendice 1) où $S_\sigma = \pi_*(\sigma) = \pi(X_\sigma)$.

Soit $I_\sigma \subset X_\sigma$ le lieu d'indétermination de l'application méromorphe K_σ; c'est un sous-espace analytique fermé de X_σ.

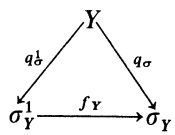
Soit $\hat{X}_\sigma$ le sous-ensemble analytique fermé de X_σ constitué de l'ensemble des points x de X_σ tels que $\dim_{\mathbb{C}}(\pi^{-1}(\pi(x)) \cap I_\sigma) = d$ au point x.

Soient $Z_\sigma \subset \sigma \underset{S}{\times} X$ le graphe de la famille universelle de cycles de X paramétrée par σ, et $q_\sigma : Z_\sigma \to \sigma$ et $p_\sigma : Z_\sigma \to X$ les restrictions à Z_σ de q et p.

Pour toute composante irréductible Y de $p_\sigma^{-1}(\hat{X}_\sigma)$ telle que l'on ait l'inégalité:

$$\dim_{\mathbb{C}}(q_\sigma(Y)) \geqq 1 + \dim_{\mathbb{C}}(\pi_*(q_\sigma(Y))) , \tag{3}$$

notons:



la factorisation de Stein irréductible (voir appendice 2) de la restriction de q_σ à Y, où $\sigma_Y = q_\sigma(Y)$. Rappelons que σ^1_Y est un sous-espace analytique irréductible de

$\overline{\mathscr{C}^*_{d-1}(X/S)}$ tel que $\overline{(\sigma^1_Y)^*}=\sigma^1_Y$, que f_Y est un morphisme propre fini et surjectif, et que les fibres de l'application méromorphe localement bornée q^1_σ sont, génériquement sur σ^1_Y, irréductibles.

On vérifie facilement que σ^1_Y est S-propre, et que l'on a aussi: $\dim_{\mathbb{C}}(\sigma^1_Y)\geqq 1+\dim_{\mathbb{C}}(\pi_*(\sigma^1_Y))$.

Soit $T(\sigma)$ l'espace analytique fermé réunion disjointe des σ^1_Y, pour tous les Y définis ci-dessus: $T(\sigma)$ est encore S-propre, vérifie $\overline{T(\sigma)}^*=T(\sigma)$ et pour chacune de ses composantes irréductibles, on à l'inégalité (2).

Remarquons que si σ est irréductible, alors on a l'inégalité: $\dim_{\mathbb{C}}(T(\sigma))<\dim_{\mathbb{C}}(\sigma)$. En effet, ceci résulte des inégalités suivantes (avec les notations introduites ci-dessous):

- $\dim_{\mathbb{C}}(Y)<\dim_{\mathbb{C}}(Z_\sigma)$, pour toute composante irréductible Y de $p^{-1}(\hat{X}_\sigma)$ telle que l'on ait:

$$\dim_{\mathbb{C}}(q_\sigma(Y))\geqq 1+\dim_{\mathbb{C}}(\pi_*(q_\sigma(Y))) \tag{3}$$

- $\dim_{\mathbb{C}}(Y)=d+\dim_{\mathbb{C}}(\sigma_Y)$,
- $\dim_{\mathbb{C}}(Z_\sigma)=d+\dim_{\mathbb{C}}(\sigma)$,
- $\dim_{\mathbb{C}}(T(\sigma))=\sup_Y\{\dim_{\mathbb{C}}(\sigma^1_Y)\}$,
- $\dim_{\mathbb{C}}(\sigma^1_Y)=\dim_{\mathbb{C}}(\sigma_Y)$, pour toute composante irréductible Y de $p_\sigma^{-1}(X_\sigma)$ vérifiant l'inégalité (3) ci-dessus.

Lorsque σ est une réunion disjointe de sous-espaces analytiques fermés irréductibles de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant les propriétés (1) et (2), on définit $T(\sigma)$ comme la réunion disjointe des $T(\sigma^0)$ lorsque σ^0 décrit l'ensemble des composantes irréductibles de σ.

Enfin, l'application T jouit de la propriété suivante, où l'on note T^l la $l^{\text{ième}}$ itérée de T:

Lemme 3. *Soient σ un sous-espace analytique irréductible fermé de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant les propriétés* (1) *et* (2) *énoncées ci-dessus, s un élément de $\pi_*(\sigma)$, et X^0_s une composante irréductible de $X_s=\pi^{-1}(s)$ contenue dans X_σ.*

L'ensemble L des entiers l tels que X^0_s est contenue dans $X_{T^l(\sigma)}$ est majoré, et l'on note k le plus grand élément de L.

Alors, pour toute composante irréductible σ' de $T^k(\sigma)$, le lieu d'indétermination de l'application méromorphe $K_{\sigma'}$ ne contient pas X^0_s, et définit donc, par restriction, une application méromorphe sur X^0_s.

Démonstration. Le fait que L est majoré résulte de la remarque précédent le théorème, puisque $T^l(\sigma)=\emptyset$ si σ est supérieur à $\dim_{\mathbb{C}}(\sigma)$.

La seconde assertion du lemme résulte immédiatement de la définition de T, et en particulier du fait que, s'il existait une composante irréductible σ' de $T^k(\sigma)$ telle que $\hat{X}_{\sigma'}$ contienne X^0_s, il existerait une composante irréductible Y' de $p_{\sigma'}^{-1}(X_{\sigma'})$ vérifiant l'inégalité (3), et k ne majorerait pas L. CQFD

Construction de l'espace analytique Σ. On suppose encore ici que π est géométriquement plat (voir [B3]), et que $d=\dim_{\mathbb{C}}(X)-\dim_{\mathbb{C}}(S)$.

Soit $\hat{\Sigma}$ le sous-espace analytique fermé de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ constitué des points σ en lesquels la dimension de $\{\pi_*^{-1}(\pi_*(\sigma))\}$ est au moins égale à 1, où $\pi_*:\overline{\mathscr{C}^*_{d-1}(X/S)}\to S$

est le morphisme d'image directe par π, qui, à un diviseur relatif de X sur S, associe l'image de son support par π.

Pour toute composante irréductible $\hat{\Sigma}^0$ de $\hat{\Sigma}$, $\hat{\Sigma}^0$ est S-propre, et l'on a donc: $\dim_{\mathbb{C}}(\hat{\Sigma}^0) \geqq 1 + \dim_{\mathbb{C}}(\pi_*(\hat{\Sigma}^0))$.

Soit alors Σ_0 la réunion disjointe des composantes irréductibles $\hat{\Sigma}^0$ de $\hat{\Sigma}$ telles que $(\hat{\Sigma}^0)^* = \hat{\Sigma}^0$; Σ_0 est donc une réunion dénombrable de sous-espace analytiques fermés de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ qui vérifient les conditions (1) et (2) énoncées dans la construction de T.

On pose alors: Σ est la réunion disjointe, lorsque n décrit $\mathbb{N}$, des espaces analytiques $T^n(\Sigma_0)$ avec $T^0(\Sigma_0) = \Sigma_0$ et $T^{n+1}(\Sigma_0) = T(T^n(\Sigma_0))$. L'espace analytique Σ est donc une réunion disjointe dénombrable de sous-espaces analytiques irréductibles fermés S-propres de $\overline{\mathscr{C}^*_{d-1}(X/S)}$ possédant les propriétés (1) et (2).

Remarquons maintenant que, pour toute composante irréductible Σ^0 de Σ, le morphisme $\pi_* : \Sigma^0 \to S$ est à fibres algébriques d'après le lemme 1, puisque, π étant géométriquement plat, pour tout s de S, $X_s = \pi^{-1}(s)$ est un sous-espace analytique compact de X de dimension pure d, et que $\{\pi_*^{-1}(s) \cap \Sigma^0\}$ est un sous-espace analytique compact de $\mathscr{C}_{d-1}(X_s)$ puisque Σ^0 est S-propre. La remarque finale de l'appendice 1 montre alors que l'on a un diagramme commutatif:

$$\begin{array}{ccc} X_{\Sigma^0} & \xrightarrow{K_{\Sigma^0}} & K_{\Sigma^0}(X_{\Sigma^0}) \subset \mathscr{C}(\Sigma^0/S) \\ & {\scriptstyle\pi}\searrow \quad \swarrow{\scriptstyle\pi_{**}} & \\ & S & \end{array}$$

dans lequel $\pi_{**} : \mathscr{C}(\Sigma^0/S) \to S$ est le morphisme d'image directe par π_*. Il en résulte alors que $K_{\Sigma^0}(X_{\Sigma^0})$ est un sous-espace analytique fermé et S-propre par π_{**} de $\mathscr{C}(\Sigma^0/S)$, et que ses fibres sur S sont algébriques par les propriétés de stabilité des espaces algébriques rappelées ci-dessus.

Proposition 1. *Les notations étant celles de l'énoncé du théorème et π étant géométriquement plat, soient s un point de S, et X_s^0 une composante irréductible de $X_s = \pi^{-1}(s)$.*

Il existe alors une composante irréductible Σ^0 de Σ telle que:

1. *X_s^0 est contenue dans X_{Σ^0}, c'est-à-dire que par tout point de X_s^0 passe un cycle de la famille Σ^0.*

2. *Le lieu d'indétermination de l'application méromorphe de Kodaira $K_{\Sigma^0} : X_{\Sigma^0} \dashrightarrow \mathscr{C}(\Sigma^0)$ associée à Σ^0 ne contient pas X_s^0. La restriction de K_{Σ^0} à X_s^0 définit donc une application méromorphe.*

3. *La restriction de K_{Σ^0} à X_s^0 est une réduction algébrique de X_s^0.*

Démonstration. Le lemme 2 montre l'existence d'une composante irréductible γ de $\overline{\mathscr{C}^*_{d-1}(X_s^0)}$ telle que $K_\gamma : X_s^0 \dashrightarrow \mathscr{C}(\gamma)$ soit une réduction algébrique de X_s^0, puisque π est géométriquement plat et X_s^0 donc de dimension d. En particulier, $\overline{\gamma^*} = \gamma$ et $\dim_{\mathbb{C}}(\gamma) \geqq 1$.

Il existe donc une composante irréductible Σ_0^0 de Σ_0 qui contient γ.

Le lemme 3 montre alors l'existence d'un entier k et d'une composante irréductible Σ^0 de $T^k(\Sigma_0^0)$ tels que Σ^0 possède les deux premières propriétés de l'énoncé du lemme 3.

Par construction de Σ, on voit donc que Σ^0 est une composante irréductible de Σ.

Il reste à démontrer que la restriction de K_{Σ^0} à X_s^0 est une réduction algébrique de X_s^0.

Puisque $\overline{\gamma^*}=\gamma$, la construction de T montre que γ est contenue dans Σ^0 [plus précisément: γ est une composante irréductible de $\pi_*^{-1}(s)\cap\Sigma^0$].

Soit $\bar{\gamma}$ la réunion des composantes irréductibles de $p_{\Sigma^0}^{-1}(X_s^0)$ qui se projettent sur X_s^0 par p: ces composantes irréductibles sont toutes de même dimension, puisque I_{Σ^0}, le lieu d'indétermination de K_{Σ^0}, ne contient pas X_s^0. Notons $(p+d)$ cette dimension. Il est alors facile de vérifier que, si $\gamma_0=\gamma,\gamma_1,\ldots,\gamma_h$ sont les composantes irréductibles de $\bar{\gamma}$, la restriction de K_{Σ^0} est équivalente, au sens de la définition du chapitre 0, Par. 3 de [C4], à l'application méromorphe produit:

$$(K_{\gamma_0}\times K_{\gamma_1}\times\ldots\times K_{\gamma_h}):X_s^0\dashrightarrow\prod_{i=1}^{h}\mathscr{C}(\gamma_i).$$

En effet, l'image de cette application méromorphe est munie d'une application d'addition de cycles:

$$\prod_{i=0}^{h}\mathscr{C}_p(\gamma_i)\to\mathscr{C}_p(\bar{\gamma}),$$

et la composée de cette application d'addition avec l'application méromorphe produit des (K_{γ_i}) est la restriction de K_{Σ^0} à X_s^0.

On a donc une factorisation méromorphe:

$$\begin{array}{ccc} & \overset{K_{\Sigma^0|X_s^0}}{\dashrightarrow} & \mathscr{C}_p(\bar{\gamma})\supset K_{\Sigma^0}(X_s^0) \\ X_s^0 & & \Big\downarrow\Phi \\ & \underset{K_\gamma}{\dashrightarrow} & \mathscr{C}_p(\gamma)\supset K_\gamma(X_s^0) \end{array}$$

Comme $K_{\Sigma^0}(X_s^0)$ est algébrique, et que K_γ est une réduction algébrique de X_s^0, Φ est biméromorphe, et la restriction de K_{Σ^0} à X_s^0 est aussi une réduction algébrique de X_s^0, ce qui établit la proposition 1.

Pour toute composante irréductible Σ^0 de Σ, on a: $\overline{(\Sigma^0)^*}=\Sigma^0$; on en déduit que le graphe $Z_{\Sigma^0}\subset\Sigma^0\underset{S}{\times}X$ de la famille Σ^0 est irréductible. Donc, $X_{\Sigma^0}=p(Z_{\Sigma^0})$ et $K_{\Sigma^0}(X_{\Sigma^0})$ sont aussi irréductibles.

On peut donc poser:

$$a(K_{\Sigma^0})=\dim_{\mathbb{C}}(K_{\Sigma^0}(X_{\Sigma^0}))-\dim_{\mathbb{C}}(\pi(X_{\Sigma^0})).$$

Remarquons que $a(K_{\Sigma^0})$ est la dimension générique, sur $\pi(X_{\Sigma^0})$, de la fibre de $K_{\Sigma^0}(X_{\Sigma^0})$; puisque cette fibre est algébrique, c'est aussi la dimension algébrique de cette fibre. Remarquons aussi que l'on a toujours: $a(K_{\Sigma^0})\leqq d$.

On a alors:

Proposition 2. *Pour tout s de S et toute composante irréductible X_s^0 de $X_s=\pi^{-1}(s)$, on a l'égalité:*

$a(X_s^0)=\sup\{a(K_{\Sigma^0})$, *lorsque Σ^0 décrit l'ensemble des composantes irréductibles de Σ telles que X_{Σ^0} contient X_s^0}$\}$.

Démonstration. Il résulte immédiatement de ce qui précède que $a(X_s^0) \leqq \sup\{a(K_{\Sigma^0}),$ où K_{Σ^0} est une composante irréductible de Σ telle que $X_s^0 \subset X_{\Sigma^0}\}$.

En effet, puisque π est géométriquement plat, si Σ^0 est une composante irréductible de Σ telle que $X_s^0 \subset X_{\Sigma^0}$ et telle que la restriction de K_{Σ^0} à X_s^0 est définie et est une réduction algébrique de X_s^0, la fibre au-dessus de $\pi(X_s^0) = \{s\}$ du graphe de K_{Σ^0} est une modification de X_s^0 et est donc aussi de dimension pure d.

Le rang générique de K_{Σ^0} est égal à

$$\dim_{\mathbb{C}}(X_{\Sigma^0}) - \dim_{\mathbb{C}}(K_{\Sigma^0}(X_{\Sigma^0})) = d - a(K_{\Sigma^0}).$$

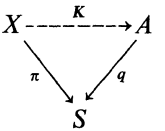
Mais d'après ce qui précède, puisque $\dim_{\mathbb{C}}(K_{\Sigma^0}(X_s^0) \geqq a(K_{\Sigma^0})$ et que l'on a nécessairement $d - a(X_s^0) = \dim_{\mathbb{C}}(X_s^0) - \dim_{\mathbb{C}}(K_{\Sigma^0}(X_s^0)) \geqq d - a(K_{\Sigma^0})$ on a l'égalité:

$$a(X_s^0) = a(K_{\Sigma^0}).$$

L'inégalité ci-dessus est donc établie.

L'inégalité opposée: $a(X_s^0 \geqq \mathrm{Sup}\{a(K_{\Sigma^0}),$ où Σ^0 est une composante irréductible de Σ telle que $X_s^0 \subset X_{\Sigma^0}\}$ résulte immédiatement du lemme suivant:

Lemme 4. *Soit*

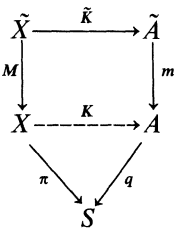


un diagramme dans lequel X, A et S sont des espaces analytiques irréductibles, π et q des morphismes propres et surjectifs, et K une application méromorphe surjective.

Si les fibres de q sont toutes des espaces algébriques, alors, pour tout s, et toute composante irréductible X_s^0 de $X_s = \pi^{-1}(s)$, on a l'inégalité:

$$a(X_s^0) \geqq (\dim_{\mathbb{C}}(A) - \dim_{\mathbb{C}}(S)) + (\dim_{\mathbb{C}}(X_s^0) - \dim_{\mathbb{C}}(X) + \dim_{\mathbb{C}}(S)) \geqq \dim_{\mathbb{C}}(A) - \dim_{\mathbb{C}}(S).$$

Démonstration. Le théorème d'aplatissement géométrique de [B3], appliqué au morphisme surjectif $\bar{K}: \bar{X} \to A$, où $\bar{X}$ est le graphe dans $(X \times A)$ de $\bar{K}$, et K la restriction de la seconde projection, montre qu'il existe un diagramme:



dans lequel les morphismes sont propres et surjectifs, M et m sont des modifications et où $\tilde{K}$ est géométriquement plat.

Puisque m est une modification, et que q est à fibres algébriques, $(q \circ m)$ est aussi à fibres algébriques.

Soit $\tilde{X}_s^0$ une composante irréductible de $(\pi \circ M)^{-1}(s)$; $\tilde{K}(\tilde{X}_s^0)$ est algébrique, d'après ce qui précède.

Puisque $\tilde{K}$ est géométriquement plat, on a donc l'égalité:

$$\begin{aligned}\dim_{\mathbb{C}}(\tilde{X}_s^0) &= \dim_{\mathbb{C}}(\tilde{K}(\tilde{X}_s^0)) + \dim_{\mathbb{C}}(\tilde{X}) - \dim_{\mathbb{C}}(\tilde{A}) \\ &= a(\tilde{K}(\tilde{X}_s^0)) + \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(A).\end{aligned}$$

Comme $a(\tilde{K}(\tilde{X}_s^0)) \leqq a(\tilde{X}_s^0)$ on a aussi:

$$\dim_{\mathbb{C}}(\tilde{X}_s^0) - a(\tilde{X}_s^0) \quad \leqq \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(A).$$

Soit maintenant X_s^0 une composante irréductible de $\pi^{-1}(s)$. Il existe une composante $\tilde{X}_s^0$ de $(\pi \circ M)^{-1}(s)$ telle que $M(\tilde{X}_s^0) = X_s^0$.

Mais on a, par le théorème 2 du chapitre 0, Par. 6 de [C4]

$$\dim_{\mathbb{C}}(X_s^0) - a(X_s^0) \leqq \dim_{\mathbb{C}}(\tilde{X}_s^0) - a(\tilde{X}_s^0).$$

C'est-à-dire que:

$$\begin{aligned}a(X_s^0) &\geqq \dim_{\mathbb{C}}(X_s^0) - \dim_{\mathbb{C}}(X) + \dim_{\mathbb{C}}(A) \\ &= \dim_{\mathbb{C}}(A) - \dim_{\mathbb{C}}(S)) + (\dim_{\mathbb{C}}(X_s^0) - \dim_{\mathbb{C}}(X) + \dim_{\mathbb{C}}(S)).\end{aligned}$$

Comme $\dim_{\mathbb{C}}(X_s^0) \geqq \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S)$, on a bien l'inégalité: $a(X_s^0) \geqq \dim(A) - \dim_{\mathbb{C}}(S)$ cd qui établit le lemme 4. CQFD

On est maintenant en mesure de démontrer la propriété 3 de l'énoncé du théorème 1.

L'égalité $a(X_s^0) = \sup\{a(K_{\Sigma^0}),$ où Σ^0 est une composante irréductible de Σ telle que $X_{\Sigma^0} \supset X_s^0\}$ est donc établie pour tout s et toute composante irréductible de $\pi^{-1}(s) = X_s$.

Il résulte donc de la définition de $a(s, \pi)$ que, pour tout s de S, $a(s, \pi) = \sup\{a(K_{\Sigma^0}),$ où Σ^0 est une composante irréductible de Σ telle que $\pi_*(\Sigma^0)$ contient $s\}$.

Soit alors Σ_k la réunion des composantes irréductibles Σ^0 de Σ telles que $a(\Sigma^0) \geqq k$, ceci pour un entier k.

Ce qui précède montre que, avec les notations de l'énoncé du théorème 1, S_k est égal à $\pi_*(\Sigma_k)$.

Comme Σ_k a, au plus, un ensemble dénombrable de composantes irréductibles, et que pour chaque composante irréductible Σ^0 de Σ_k, $\pi_*(\Sigma^0)$ est un sous-ensemble analytique fermé de S, S_k est bien une réunion au plus dénombrable de sous-espaces analytiques fermés de S.

La propriété 3 est donc établie. Passons à la propriété 1:

Soit k_0 *le plus grand des entiers* k tels que $S_k = S$. On a alors $k_0 = a(\pi)$, et comme S_{k_0+i} est une réunion au plus dénombrable de sous-espaces de sous-espaces analytiques fermés de S dont aucun n'est égal à S, S_{k_0+i} est analytiquement maigre dans S pour $i \in \mathbb{N}^*$.

Montrons maintenant la propriété 1: puisque Σ est dénombrable à l'infini, il existe, d'après la première assertion du lemme 3, une composante irréductible Σ^0 de Σ telle que, pour tout s d'un ensemble S' de S qui n'est pas maigre dans S, il existe une composante irréductible X_s^0 de $X_s = \pi^{-1}(s)$ telle que: $X_{\Sigma^0} \supset X_s^0$ et

(a) $X_s^0 \subset X_{\Sigma^0} = X$.

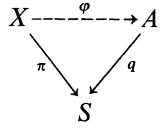
(b) La restriction de K_{Σ^0} à X_s^0 définit une réduction algébrique de X_s^0.

De la propriété 3, il résulte que $a(K_{\Sigma^0}) = a(\pi)$, et de la propriété (b) il résulte que la fibre générique de K_{Σ^0} est irréductible, puisque S' n'est pas maigre dans S.

La propriété 3 montre alors que les propriétés (a) et (b) sont en fait vraies pour toute composante irréductible X_s^0 de $\pi^{-1}(s)$, pour tout s d'un S'' dont le complémentaire dans S est analytiquement maigre dans S. Ceci établit la propriété 1 dans le cas où π n'est pas nécessairement géométriquement plat.

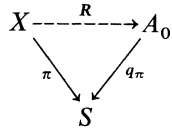
La propriété 2 se déduit immédiatement du lemme 4 appliqué à une réduction algébrique de π. CQFD

Proposition 3. *Soit*

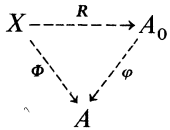


un diagramme commutatif dans lequel π et q sont des morphismes propres et surjectifs, π est faiblement Kählérien, q à fibres algébriques, X irréductible et φ une application méromorphe surjective.

Si

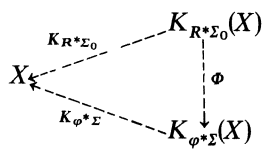


est une réduction algébrique de π, alors il existe une application méromorphe surjective $\Phi : A_0 \dashrightarrow A$ rendant commutatif le diagramme:



Démonstration. En utilisant le théorème d'aplatissement géométrique, on peut supposer, quitte à modifier X, A et A_0, que φ et R sont des morphismes géométriquement plats. Posons: $\Delta = \dim(A) - \dim(S)$ [resp. $\Delta_0 = \dim(A_0 - \dim_{\mathbb{C}}(S)$]. Puisque les fibres de A sur S sont algébriques, il existe, d'après la démonstration du théorème 1, une composante irréductible Σ (resp. Σ_0) de $\overline{\mathscr{C}^*_{\Delta-1}(A/S)}$ [resp. $\overline{\mathscr{C}^*_{\Delta_0-1}(A_0/S)}$] telle que K_Σ (resp. K_{Σ_0}) sépare presque les points de A (resp. A_0). Σ et Σ_0 sont S-propres.

Puisque R et φ sont géométriquement plats, on peut construire des morhismes d'image réciproque $R^* : \Sigma_0 \to \mathscr{C}_{d-1}(X/S)$ et $\varphi^* : \Sigma \to \mathscr{C}_{d-1}(X/S)$, où $d = \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(S)$. Il est facile de vérifier que l'on peut construire des applications de Kodaira $K_{R^*\Sigma_0}$ et $K_{\varphi^*\Sigma}$ respectivement associées aux familles $(R^*\Sigma_0)$ et $(\varphi^*\Sigma)$ sur X, que l'on a une factorisation:

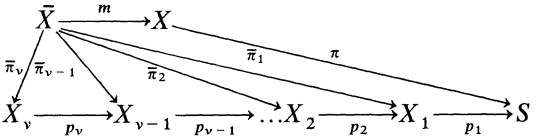


et que $K_{R^*\Sigma_0}(X)$ [resp. $K_{\varphi^*\Sigma}(X)$] est biméromorphe à A_0 (resp. A). Ceci établit la proposition.

De la propriété 1 de l'énoncé du théorème 1, on déduit immédiatement le:

Théorème 2. *Soit $\pi : X \to S$ un morphisme surjectif, propre et faiblement Kählérien d'espaces analytiques irréductibles.*

Il existe alors une modification $m : \bar{X} \to X$ et un diagramme commutatif:



dans lequel les p_i sont géométriquement plats et à fibres algébriques pour $i = 1, \ldots, \nu-1, \nu$, dans lequel $\bar{\pi}_{i+1}$ est une réduction algébrique de $\bar{\pi}_i$ pour $i = 0, \ldots, \nu-1$, avec $\bar{\pi}_0 = \pi \circ m$, et dans lequel $a(\bar{\pi}_\nu) = 0$, c'est-à-dire que la fibre générale de $\bar{\pi}_\nu$ a des composantes irréductibles de dimension algébrique zéro. Parmi ces diagrammes, il en existe un dont proviennent tous les autres.

La construction de tels diagrammes se fait en itérant les réductions algébriques des π_i.

On appellera un tel diagramme une *décomposition algébrique de π.*

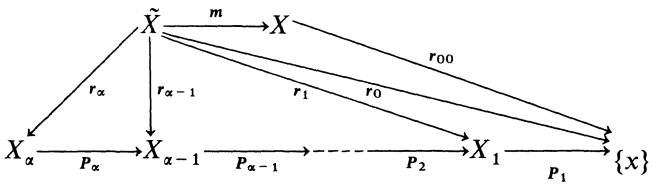
Remarque. Deux cas se produisent:

- π_ν est une modification. On dira alors que π est de type A.
- π_ν n'est pas une modification. Ceci signifie donc que $a(\pi_\nu) = 0$ et que $\dim_{\mathbb{C}}(X) \geqq 2 + \dim_{\mathbb{C}}(X_\nu)$: la fibre «générale» (c'est-à-dire dans le complémentaire d'un ensemble analytiquement maigre de X_ν) de π_ν est de dimension algébrique 0. On dira que π est de type B.

B. Decomposition algebrique d'un espace Kählérien. Applications

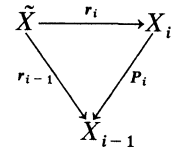
Lorsque X est un espace analytique faiblement Kählérien compact, on peut appliquer le théorème 1 à tout morphisme constant $r_{00} : X \to \{x\}$, où $\{x\}$ est l'espace analytique réduit à un point. On obtient le:

Théorème 1'. *Soit X un espace analytique faiblement Kählérien compact et irréductible. Il existe alors une modification propre $m : \tilde{X} \to X$ et un diagramme commutatif de morphismes surjectifs:*



dans lequel:

● *Pour tout* $i=1,\ldots,\alpha$, *le diagramme:*



est une réduction algébrique du morphisme

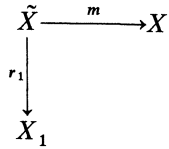
$$r_{i-1}:\tilde{X}\to X_{i-1}\ ;$$

● *l'une des fibres de* r_α *a une composante irréductible de dimension algébrique zéro (voir corollaire du théorème* 1*).*

Un tel diagramme, unique à équivalence biméromorphe près pour X *donné, est appelé une décomposition algébrique de* X.

Remarques. (Les démonstrations des remarques 4 et 5 sont ci-dessous.)

1. Le diagramme



est donc une réduction algébrique de X.

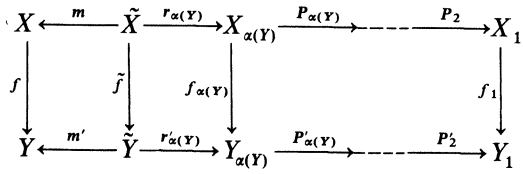
2. La suite d'entiers:

$$a(X)=a_1(X)=\dim_{\mathbb{C}}(X_1)\leqq a_2(X)=\dim_{\mathbb{C}}(X_2)\leqq\ldots\leqq a_\alpha(X)$$
$$=\dim_{\mathbb{C}}(X_\alpha)\leqq\dim_{\mathbb{C}}(X)$$

ne dépend que de la classe de biméromorphie de X, ainsi que l'entier $\alpha(X)=\alpha$.

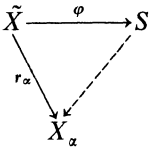
3. La décomposition algébrique de X est «duale», en échangeant sous-espaces et morphismes surjectifs, de la codécomposition algébrique de X obtenue en itérant la coréduction algébrique de X décrite dans [C4].

4. Soit $f:X\to Y$ un morphisme surjectif d'espaces analytiques faiblement Kählériens compacts et irréductibles. Alors $\alpha(X)\geqq\alpha(Y)$, et il existe un diagramme commutatif de morphismes surjectifs, dans lequel les lignes sont des diagrammes de décomposition algébrique de X et Y respectivement:



5. Si $\varphi:\tilde{X}\to S$ est un morphisme surjectif dont l'une des fibres a une composante irréductible de dimension algébrique zéro, il existe une application méromorphe

surjective $\psi : S \to X_\alpha$ rendant communatitif le diagramme:



6. Deux cas se présentent:
- $a_\alpha = \dim_{\mathbb{C}}(X)$: on dira alors que X est de type A.
- $\dim_{\mathbb{C}}(X) \geqq a_\alpha + 2$: on dira alors que X est de type B; voir.

7. La classe des espaces analytiques faiblement Kählériens compacts et irréductibles de type A est stable par les opérations usuelles: image par un morphisme, sous-espaces, produits, image réciproque par un morphisme algébrique. De plus, les composantes irréductibles de l'espace des cycles d'un tel espace sont aussi de type A (démonstration analogue à celle du corollaire 3 du théorème de [C1]).

Proposition 4. *Soient X un espace analytique faiblement Kählérien compact et irréductible et $f : X \to Y$ un morphisme surjectif.*

Alors Y est aussi un espace analytique faiblement Kählérien, $\alpha(X) \geqq \alpha(Y)$, et on a un diagramme commutatif:

$$\begin{array}{ccccccccccccc}
X & \xleftarrow{m} & X & \xrightarrow{r_{\alpha(X)}} & X_{\alpha(X)} & \xrightarrow{p_{\alpha(X)}} & \ldots X_{\alpha(Y)} & \xrightarrow{P_{\alpha(Y)}} & \ldots \to X_2 & \xrightarrow{P_2} & X_1 & \xrightarrow{P_1} & \{x\} \\
\downarrow f & & \downarrow f & & & & \downarrow f_{\alpha(Y)} & & & & \downarrow f_2 & & \downarrow f_1 & & \downarrow f_0 \\
Y & \xleftarrow{m'} & Y & & \xrightarrow{r'_{\alpha(Y)}} & & Y_{\alpha(Y)} & \xrightarrow{P'_{\alpha(Y)}} & \ldots \to Y_2 & \xrightarrow{P'_2} & Y_1 & \xrightarrow{P'_1} & \{y\}
\end{array}$$

dans lequel:

$$X \xleftarrow{m} \bar{X} \xrightarrow{r_{\alpha(X)}} X_{\alpha(X)} \to \ldots \to X_2 \xrightarrow{p_2} X_1 \xrightarrow{p_1} \{x\}$$

$$Y \xleftarrow{m'} \bar{Y} \xrightarrow{r'_{\alpha(Y)}} Y_{\alpha(Y)} \xrightarrow{p} \ldots \longrightarrow Y_2 \xrightarrow{p'_2} Y_1 \xrightarrow{p'_2} \{y\}$$

sont des diagrammes de décomposition algébrique de X et Y respectivement.

Démonstration. f_0 existe. Montrons l'existence de f_{i+1} si f_i existe, pour $i \leqq \alpha(Y) - 1$. On a donc un diagramme

$$\begin{array}{ccccc}
X & \dashrightarrow^{r_{i+1}} & X_{i+1} & \xrightarrow{p_{i+1}} & X_i \\
\downarrow f & & & & \downarrow f_i \\
Y & \dashrightarrow^{r'_{i+1}} & Y_{i+1} & \xrightarrow{p'_{i+1}} & Y_i
\end{array}$$

et il faut montrer l'existence d'une application méromorphe surjective $f_{i+1}: X_{i+1} \dashrightarrow Y_{i+1}$ laissant commutatif le diagramme écrit ci-dessus.

Considérons l'application méromorphe produit

$$K = (p_{i+1} \circ r_{i+1}) \times (r'_{i+1} \circ f): X \to X_i \underset{Y_i}{\times} Y_{i+1}.$$

Désignons par L son image. La projection $L \to X_i$ est à fibres algébriques, puisque $p'_{i+1}: Y_{i+1} \to Y_i$ est à fibres algébriques.

L'existence de f_{i+1} résulte alors simplement du fait que

$$\begin{array}{ccc} & & X_{i+1} \\ r_{i+1}: X & \overset{r_i}{\dashrightarrow} & \downarrow p_{i+1} \\ & & X_i \end{array}$$

est une réduction algébrique de $r_i: X \dashrightarrow X_i$, par la proposition 3.

La proposition est donc établie.

Théorème 3. *Soient X un espace analytique faiblement Kählérien compact et irréductible, et Y un sous-ensemble analytique compact et irréductible de dimension algébrique zéro de X.*

Il existe alors un diagramme commutatif de morphismes surjectifs:

$$\begin{array}{ccc} \tilde{X} & \xrightarrow{m} & X \\ \downarrow{\scriptstyle \varphi} & & \\ S & & \end{array}$$

et un sous-espace analytique compact et irréductible $\tilde{Y}$ de $\tilde{X}$ tels que:

- *m est une modification propre.*
- *φ est un morphisme géométriquement plat.*
- *$a(\tilde{Y}) = 0$.*
- *$m(\tilde{Y}) = Y$.*
- *$\tilde{Y}$ est contenu dans une fibre de φ.*
- *Il existe un osus-ensemble S' de S, qui est intersection dénombrable d'ouverts de Zariski denses de S, et tel que, pour tout s de S', $\tilde{X}_s = \varphi^{-1}(s)$ est irréductible et de dimension algébrique zéro.*

Remarques. 1. Compte tenu de la remarque 5 suivant le théorème 1', ce résultat signifie donc que les sous-espaces compacts et irréductibles de dimension algébrique zéro de X sont contenus dans les fibres de l'application méromorphe $(r_{a(X)} \circ m^{-1}): X \dashrightarrow X_{a(X)}$. En particulier, $\dim_{\mathbb{C}}(Y) \leqq \dim_{\mathbb{C}}(X) - \dim_{\mathbb{C}}(X_{a(X)})$.

2. Ce corollaire signifie que X contient des sous-espaces analytiques compacts et irréductibles de dimension algébrique zéro non triviaux (i.e.: non réduits à des points) si et seulement s'il est de type B.

Si X est de type B, et si $a(X) > 0$, alors X contient une infinité non dénombrable de tels sous-espaces, et la réunion de ceux-ci remplit au moins une intersection dénombrable d'ouverts de Zariski denses de X.

Démonstration du théorème 3. Prenons pour $m:\tilde{X}\to X$ la modification d'un diagramme de décomposition algébrique de X (avec les notations du théorème 2'). L'existence d'un $\tilde{Y}$ tel que $a(\tilde{Y})=0$ et $m(\tilde{Y})=Y$ résulte du lemme 1 ci-dessous.

Dans le diagramme de décomposition algébrique de X, avec les notations du théorème, on a :

$r_1(\tilde{Y})$ est réduit à un point, puisque c'est un espace algébrique et que $a(r_1(\tilde{Y}))=0$.

Soit alors i le plus grand des entiers j tels que $r_j(\tilde{Y})$ est réduit à un point. On a $i\leqq\alpha$.

Si $i<\alpha$, comme $r_i(\tilde{Y})$ est un point, $r_{i+1}(\tilde{Y})$ est contenu dans la fibre de p_{i+1} au-dessus de $r_i(Y)$. Mais cette fibre est algébrique. Donc $i=\alpha$ et $r_\alpha(\tilde{Y})$ est un point.

Ce qui signifie que $\tilde{Y}$ est contenu dans un fibre de r_α. On a donc, en fait, établi que Y est contenu dans l'une des fibres de r_α : la famille des fibres de r_α ne dépend pas du choix de m. On en déduit en particulier que $\dim_{\mathbb{C}}(Y)\leqq\dim_{\mathbb{C}}(X)-\dim_{\mathbb{C}}(X_\alpha)$. CQFD

Lemme 1. *Soit $\varphi:X\to Y$ un morphisme surjectif d'espaces analytiques compacts et irréductibles. Si $a(Y)=0$, X est réunion de sous-espaces analytiques irréductibles Z tels que $\varphi(Z)=Y$ et $a(Z)=0$.*

Démonstration. Par récurrence sur $\dim_{\mathbb{C}}(X)-\dim_{\mathbb{C}}(Y)=\delta(\varphi)$.

Si $\delta=0$, alors $a(X)=0$ et le résultat est démontré.

Si la propriété est vraie pour $\delta\leqq\delta_0$, 2 cas de présentent :

- $a(X)=0$, le résultat est alors vrai.
- $a(X)\geqq 1$ Soit $f:X\dashrightarrow\mathbb{P}_1$ une application méromorphe surjective. Le lemme 2 ci-dessous montre que :

$$f\times\varphi:X\dashrightarrow\mathbb{P}_1\times Y \quad \text{est surjective.}$$

Soient $x_0\in X$ un élément de X, et X_{x_0} une fibre *de* f contenant x_0. La restriction φ_{x_0} de φ à X_{x_0} est surjective, et comme $\delta(\varphi_{x_0})=\delta(\varphi)-1$, l'hypothèse de récurrence montre l'existence d'un Z contenu dans X_{x_0}, contenant x_0 et tel que $\varphi(Z)=Y$ et $a(Z)=0$. CQFD

Lemme 2. *Soient $f:X\dashrightarrow A$ et $g:X\dashrightarrow Y$ des applications méromorphes surjectives d'espaces analytiques compacts et irréductibles. Si A est un espace algébrique, et si Y est de dimension algébrique zéro, alors l'application méromorphe produit $\varphi=f\times g:X\dashrightarrow A\times Y$ est surjective.*

Démonstration. Le lemme 3 ci-dessous, appliqué à l'image de φ, notée Z, donne :

$$\dim_{\mathbb{C}}(Z)=a(Z)+\dim_{\mathbb{C}}(Y),$$

mais

$$a(Z)\geqq a(A)=\dim_{\mathbb{C}}(A).$$

Donc

$$\dim_{\mathbb{C}}(Z)\geqq\dim_{\mathbb{C}}(A)+\dim_{\mathbb{C}}(Y)=\dim_{\mathbb{C}}(A\times Y). \quad \text{CQFD}$$

Lemme 3. *Soient A et Y des espaces analytiques compacts et irréductibles, et Z un sous-espace analytique compact et irréductible de $A \times Y$ tel que la projection naturelle $p: Z \to Y$ soit surjective.*

Si $a(Y)=0$ et si A est un espace algébrique, alors $\dim_{\mathbb{C}}(Z)=a(Z)+\dim_{\mathbb{C}}(Y)$ où $a(Z)$ est la dimension algébrique de Z.

Démonstration. On a les inégalités suivantes (voir [U, théorème 3.8], ou [C4, chapitre 0, Par. 5, théorème 2]):

- $a(A \times Y)=a(A)+a(Y)=\dim_{\mathbb{C}}(A)+a(Y)$.
- $a(Z) \geqq a(A \times Y)+\dim_{\mathbb{C}}(Z)-\dim_{\mathbb{C}}(A)-\dim_{\mathbb{C}}(Y)$ (puisque Z est un sous-espace de $A \times Y$).
- $a(Z) \leqq a(Y)+\dim_{\mathbb{C}}(Z)-\dim_{\mathbb{C}}(Y)$ puisque $p: Z \to Y$ est surjective. On en déduit le résultat. CQFD

II. Familles de cycles de dimension algébrique zero

Le théorème ci-dessous montre que l'on peut «algébriser» les cycles d'une famille compacte de cycles d'un espace compact et irréductible en relevant cette famille par un morphisme propre et surjectif $p: \tilde{X} \to X$ tel que $\dim_{\mathbb{C}}(\tilde{X})=\dim_{\mathbb{C}}(X)$ puis en prenant l'image directe des cycles de la famille relevée par un morphisme $q: \tilde{X} \to Y$ pour lequel ces cycles relevées sont saturés.

Des exemples simples montrent qu'un tel morphisme q n'existe pas toujours si $\tilde{X}=X$ ou si p est une modification.

Ce résultat implique (corollaire 1) que les familles pleines[6] dans X de cycles sont presque les fibres d'une application méromorphe, si le support de l'un des cycles a une composante irréductible de dimension algébrique zéro.

Théorème. *Soient X un espace analytique compact et irréductible, et S un sous-espace analytique compact et irréductible de $\mathscr{C}(X)$ tel que:*

- $\overline{S^*}=S$,
- *la famille S est pleine dans X.*

Il existe alors un diagramme d'espaces analytiques compacts et irréductibles et de morphismes surjectifs:

$$\begin{array}{ccc} X_0 & \xrightarrow{p} & X \\ {\scriptstyle q}\downarrow & & \\ Y & & \end{array}$$

et un sous-espace analytique compact et irréductible S_0 de $\mathscr{C}(X_0)$ tels que:

1. $\dim_{\mathbb{C}}(X_0)=\dim_{\mathbb{C}}(X)$.
2. $\overline{S_0^*}=S_0$. *On note Z_{s_0} le cycle de X_0 d'indice $s_0 \in S_0$.*
3. *Si $p_*: S_0 \dashrightarrow \mathscr{C}(X)$ est l'application méromorphe d'image directe par p, alors $p_*(S_0)=S$ et p_* est biméromorphe de S_0 sur S. (Pour la notion générale d'image directe, voir [C1].)*

6 C'est-à-dire telles que la réunion des supports des cycles de la famille est égale à X (voir [C4, chapitre III])

4. *Le morphisme q est géométriquement plat, et pour tout s_0 de S_0, on a: $q(|Z_{s_0}|)$ est un sous-espace algébrique de Y, et $|Z_{s_0}| = q^{-1}(q(|Z_{s_0}|))$. On note $q_*: S_0 \dashrightarrow \mathscr{C}(Y)$ l'application d'image directe par q.*

De plus, $q_*: S_0 \dashrightarrow \mathscr{C}(Y)$ est génériquement injective, et le graphe $Z_Y \subset q_*(S_0) \times Y$ de la famille universelle de cycles de Y paramétrée par $q_*(S_0)$ est donc algébrique sur Y et sur $q_*(S_0)$ pour les projections induites par les projections de $q_*(S_0) \times Y$ sur ses facteurs.

Remarques. 1. On ne peut pas toujours prendre pour q une modification de X : par exemple, si T est un tore complexe et si $X = \mathrm{Sym}^2(T)$, soient $\pi: T \times T \to X$ la projection naturelle, et $p_1: T \times T \to T$ la première projection.

Le morphisme produit $(p_1 \times \pi): T \times T \to T \times X$ est un plongement dont l'image est le graphe d'une famille analytique de cycles de X paramétrée par T. Le morphisme $j: T \to \mathscr{C}_2(X)$ est un plongement.

Si T est de dimension algébrique zéro, les cycles de cette famille sont aussi de dimension algébrique zéro.

S'il existait une application méromorphe $q: X \dashrightarrow Y$ telle que l'image directe par q des cycles de cette famille soient algébriques, et que ces cycles soient saturés pour q, ceci entrainerait que les cycles de cette famille seraient les fibres de q. Ceci est impossible, puisque par chaque point de X passent deux cycles de cette famille.

2. La démonstration du théorème montre que, pour que l'on puisse prendre pour q une modification de X, il suffit que les fibres du morphisme $p_X: Z_S \subset S \times X \to X$ soient, génériquement sur X, irréductibles, c'est à dire que, pour x générique dans X, l'ensemble des s tels que $x \in |Z_s|$ soit un sous-ensemble analytique fermé irréductible de S.

3. Le théorème généralise le corollaire 2 du théorème d'algébricité de [C2] (théorème de Ueno sur la structure des sous-ensembles analytiques fermés irréductibles X d'un tore T): en effet, si l'on prend pour $S \subset \mathscr{C}(T)$ la famille des translatés $-\tau + X$ de X par les éléments τ de T, pour tout t de T, l'ensemble des $\tau \in T$ tels que $t \in (-\tau + X)$ s'identifie à X, et l'ensemble des $s \in S$ correspondants est irréductible, et l'on peut même prendre pour des raisons d'homogénéité, p égale à l'identité de T. On retrouve ainsi le résultat de Ueno (voir [0, Theorem 10.9, p. 120]).

Corollaire. *Soient X un espace analytique compact et irréductible, et S un sous-espace analytique compact et irréductible de $\mathscr{C}(X)$ tel que:*

- $\overline{S^*} = S$.
- *La famille S est pleine dans X.*
- *Il existe $s_1 \in S^*$ tel que $a(|Z_{s_1}|) = 0$.*

Si $Z_S \subset S \times X$ est le graphe de la famille universelle de cycles de X paramétrée par S, alors $\dim_{\mathbb{C}}(Z_S) = \dim_{\mathbb{C}}(X)$; c'est à dire qu'il existe un ouvert de Zariski dense X' de X, et un entier k tels que, pour tout x de X', l'ensemble des $s \in S$ tels que $x \in |Z_s|$ est fini et possède exactement k éléments.

Si, de plus, X est faiblement Kählérien, alors S est une composante irréductible de $\mathscr{C}(X)$.

Remarques. 1. Montrons comment l'égalité $\dim_{\mathbb{C}}(Z_S) = \dim_{\mathbb{C}}(X)$ entraîne l'existence de k et X' avec les propriétés énoncées:

Soit $p_X : Z_S \to X$ la projection de Z_S sur X induite par la seconde projection de $S \times X$: p_X est surjective puisque S est une famille pleine, et pour tout x de X, $p_X^{-1}(x)$ est l'ensemble des $s \in S$ tels que $x \in |Z_S|$.

Si $\dim_{\mathbb{C}}(Z_S) = \dim(X)$, alors il existe un ouvert de Zariski dense X' de X tel que $P'_X : Z'_S = p_X^{-1}(X') \to X'$ soit un revêtement non ramifié de degré k.

2. On n'a pas toujours $k=1$: l'exemple donné dans la remarque 1 suivant l'énoncé du théorème précédent est tel que $k=2$.

3. Le corollaire signifie donc que dans la situation décrite dans les hypothèses, on a un diagramme de morphismes surjectifs et d' espaces compacts :

$$\begin{array}{ccc} Z_S & \xrightarrow{p_X} & X \\ {\scriptstyle p_S}\downarrow & & \\ S & & \end{array}$$

dans lequel $\dim_{\mathbb{C}}(Z_S) = \dim_{\mathbb{C}}(X)$ et où les fibres de p_S sont les cycles de la famille S : les cycles de la famille S sont donc «presque» les fibres d'une application méromorphe.

Démonstration du théorème. Quelques remarques préliminaires sont nécessaires : soient U, V et W des espaces analytiques compacts et irréductibles, tels que W est un sous-ensemble analytique compact de $U \times V$, et tels que $p_u(W) = U$, où $p_u : U \times V \to U$ est la première projection.

Il existe alors un ouvert de Zariski dense U' de U tel que $W' = p_u^{-1}(U')$, réduit dans $U' \times V$, soit le graphe d'une famille analytique de cycles de V paramétrée par U'. Soit $K_{W'} : U' \to \mathscr{C}(V)$ le morphisme associé. Le théorème d'aplatissement géométrique montre que $K_{W'}$ se prolonge en une application méromorphe $K_W : U \dashrightarrow \mathscr{C}(V)$, que l'on appellera encore l'application méromorphe de Kodaira *de W dans $U \times V$.*

On vérifie aisément les assertions suivantes :

1. L'image de K_W est un sous-ensemble analytique compact U_1 de $\mathscr{C}(V)$. Si $W_1 \subset U_1 \times V$ est le graphe de la famille universelle de cycles de V paramétrés par U_1, et $K_W \times \mathrm{id}_V : U \times V \dashrightarrow U_1 \times V$ l'application méromorphe produit, alors $W_1 = (K_W \times \mathrm{id}_V)(W)$, et W est l'unique composante irréductible de $U \underset{U_1}{\times} W_1$ dont l'image par $(K_W \times \mathrm{id}_V)$ est W_1. Si $K_W : U \dashrightarrow K_W(U)$ est biméromorphe, alors $(K_W \times \mathrm{id}_V) : W \dashrightarrow W_1$ est biméromorphe.

2. Si le diagramme suivant commute, dans lequel les morphismes sont surjectifs et les espaces analytiques compacts et irréductibles, et si $K_{W_0} : U \dashrightarrow \mathscr{C}(V_0)$ induit une application biméromorphe de U sur $K_{W_0}(U)$, alors $K_W : U \dashrightarrow \mathscr{C}(V)$ induit une application biméromorphe de U sur $K_W(U)$:

$$\begin{array}{ccccc} W_0 & \hookrightarrow & U \times V_0 & \xrightarrow{P_{V_0}} & V_0 \\ {\scriptstyle (\mathrm{id}_U \times p)|_{W_0}}\downarrow & & {\scriptstyle (\mathrm{id}_U \times p)}\downarrow & & {\scriptstyle p}\downarrow \\ W & \hookrightarrow & U \times V & \xrightarrow{i_V} & V \end{array}$$

Réciproquement, si W_0 est la seule composante irréductible de $\left(W \underset{V}{\times} V_0\right)$ dont l'image par $((\mathrm{id}_U) \times p)$ est W, et si $K_W : U \longrightarrow K_W(U)$ est biméromorphe, alors $K_{W_0} : U \longrightarrow K_{W_0}(U)$ est aussi biméromorphe.

Démontrons maintenant le théorème:

Soit:

$$\begin{array}{ccc} Z_S \subset S \times X & \xrightarrow{p_X} & X \\ {\scriptstyle p_S}\downarrow & & \\ S & & \end{array}$$

le graphe de la famille universelle de cycles de X paramétrée par S.

Première étape. On peut supposer que la fibre générique du morphisme $p_X : Z_S \to X$ est irréductible. En effet:

Soit

$$\begin{array}{ccc} Z_1 & \xrightarrow{p_X^1} & X_1 \\ {\scriptstyle \lambda}\downarrow & & \downarrow{\scriptstyle q_1} \\ Z_S & \xrightarrow{p_X} & X \end{array}$$

le diagramme de factorisation de Stein irréductible du morphisme $p_X : Z_S \to X$. Donc, λ est une modification finie, et q_1 un morphisme propre et fini.

Soit $Z_2 \subset S \times X_1$ l'image de Z_1 par le morphisme produit $(p_S \circ \lambda) \times p_X^1 : Z_1 \to S \times X_1$.

Considérons alors:

● $K : S \dashrightarrow \mathscr{C}(X_1)$ l'application méromorphe de Kodaira de Z_2 dans $S \times X_1$: elle induit une application biméromorphe de S sur $S_1 = K(S)$,d'après la propriété 2 ci-dessus. De plus, on a : $\overline{S_1^*} = S_1$ puisque les fibres de $(p_S \circ \lambda)$ sont génériquement (sur S) irréductibles, et que les fibres de Z_2 sur S sont les images par $(p_S \circ \lambda) \times p_X^1$ des fibres de $(p_S \circ \lambda)$.

● $Z_3 \subset S_1 \times X_1$ le graphe de la famille universelle de cycles de X_1 paramétrée par S. La propriété un ci-dessus montre alors que $K \times \mathrm{id}_{X_1} : Z_2 \dashrightarrow Z_3$ est biméromorphe.

De plus, la fibre générique du morphisme $p_{X_1} : Z_3 \longrightarrow X_1$ est irréductible, $\dim_{\mathbb{C}}(X_1) = \dim_{\mathbb{C}}(X)$ puisque q_1 est fini, et l'image directe par q_1 de la famille universelle de cycles de X_1 paramétrée par S_1 est la famille universelle de cycles de X paramétrée par S. Il suffit donc de démontrer le théorème pour la famille S_1. *On suppose donc dans la suite que la fibre générique du morphisme $p_X : Z_S \longrightarrow X$ est irréductible.*

On note alors $K : X \dashrightarrow \mathscr{C}(S)$ l'application méromorphe de Kodaira associée à Z_S dans $S \times X$, et $Y_0 \subset \mathscr{C}(S)$ son image. (Donc $\overline{Y_0^*} = Y_0$.)

Deuxième étape. On peut supposer aussi qu'il existe un morphisme surjectif et géométriquement plat $q : X \longrightarrow Y$ équivalent à l'application méromorphe K (au

sens de la définition du chapitre 0, Par. 3 de [C4]). En effet, il suffit pour cela de considérer le graphe $\bar{X} \subset X \times Y$ de l'application méromorphe K, puis de construire, grâce au théorème d'aplatissement géométrique de [B3] un diagramme commutatif:

$$\begin{array}{ccc} X_0 & \xrightarrow{q} & Y \\ {\scriptstyle M}\downarrow & & \downarrow{\scriptstyle m} \\ \bar{X} & \xrightarrow{\bar{K}} & Y_0 \end{array}$$

dans lequel:

- m et M sont des modifications.
- $\bar{K}:\bar{X} \longrightarrow Y_0$ est la restriction à $\bar{X}$ de la projection naturelle de $X \times Y_0$ sur Y_0.
- q est géométriquement plat.

On note $\mu:\bar{X} \longrightarrow X$ la restriction à $\bar{X}$ de la projection naturelle de $X \times Y_0$ sur X.

Soient alors:

- $Z_0 \subset S \times X_0$ l'unique composante irréductible de l'image réciproque de $Z \subset S \times X$ par le morphisme produit $(\mathrm{id}_S \times (\mu \circ M)):S \times X_0 \longrightarrow S \times X$.
- $k:S \dashrightarrow \mathscr{C}(X_0)$ l'application de Kodaira de Z_0 dans $S \times X_0$, S_0 l'image de S par k et $Z_{S_0} \subset S_0 \times X_0$ le graphe de la famille universelle de cycles de X_0 paramétrée par S_0.

Troisième étape. Montrons maintenant que la conclusion du théorème est satisfaite avec la donnée de S_0 et du diagramme:

$$\begin{array}{ccc} X_0 & \xrightarrow{p=\mu\circ M} & X \\ {\scriptstyle q}\downarrow & & \\ Y & & \end{array}$$

Ceci va résulter du lemme suivant:

Lemme. *Avec les notations ci-dessus, si $Z_Y \subset S \times Y$ est l'image de $Z_{S_0} \subset S_0 \times X_0$ par l'application analytique produit $((\mathrm{id}_{S0}) \times q):S_0 \times X_0 \longrightarrow S_0 \times Y$, et si $q_*:S_0 \dashrightarrow \mathscr{C}(Y)$ est l'application méromorphe localement bornée d'image directe par q, on a:*

1. $Z_{S_0} = (\mathrm{id}_{S_0} \times q)^{-1}(Z_Y)$.
2. *$q_*:S_0 \dashrightarrow \mathscr{C}(Y)$ est aussi l'application méromorphe de Kodaira de Z_Y dans $S_0 \times Y$ et elle induit une application biméromorphe de S_0 sur $q_*(S_0)$. (En d'autres termes, sa restriction à un ouvert de Zariski dense convenable de S_0 est analytique et injective.)*

Démonstration. La propriété universelle de l'espace des cycles (de S, ici) montre, en particulier, avec les notations introduites ci-dessus, que

$$(k \times M)(Z_{S_0}) = (\mathrm{id}_S \times K)^{-1}(Z_{Y_0}), \quad \text{où:}$$

- $(k \times M):S_0 \times M_0 \dashrightarrow S \times \bar{X}$ est l'application méromorphe produit.
- $Z_{Y_0} \subset S \times Y_0$ est le graphe de la famille universelle de cycles *de* S paramétrée par $Y_0 = K(X) = \bar{K}(\bar{X})$.

On note $p_S^0:Z_{Y_0} \longrightarrow S$ la projection naturelle.

Puisque k, m et M sont des modifications, on en déduit que Z_{S_0} est l'unique composante irréductible de $(\mathrm{id}_{S_0} \times q)^{-1}(Z_Y)$ qui se surjette sur Z_Y.

Montrons maintenant que $(\mathrm{id}_{S0} \times q)^{-1}(Z_Y)$ est irréductible.

Soient $p_Y : Z_Y \longrightarrow Y$ la restriction à $Z_Y \subset S_0 \times Y$ de la seconde projection.

Notons W le sous-espace analytique compact de $Z_Y \times X$ défini par: $W = (p_Y \times q)^{-1}(\Delta_Y)$, où Δ_Y est la diagonale de $Y \times Y$.

Alors:

1. D'une part, W est le graphe d'une famille analytique de cycles de X_0 paramétrée par Z_Y, et associée au morphisme $p_Y : Z_Y \longrightarrow Y$, composé avec le *morphisme* injectif $j : Y \longrightarrow \mathscr{C}(X)$, qui à $y \in Y$ associe le cycle $q^{-1}(y)$, puisque q est géométriquement plat.

2. D'autre part, l'image de W par le morphisme produit $\pi_S \times \pi_X : Z_Y \times X \longrightarrow S_X$ est égal à $(\mathrm{id}_{S_0} \times q)^{-1}(Z_Y)$, où π_S et π_X sont les projections de $(S_0 \times Y \times X)$ sur ses premier et dernier facteurs respectivement.

3. Le morphisme $\pi_S \times \pi_X : W \longrightarrow (\mathrm{id}_{S_0} \times q)^{-1}(Z_Y)$ est un isomorphisme. De 1., on déduit que la projection naturelle de W sur Z_Y est un morphisme ouvert. Toute composante irréductible de W se surjette donc sur Z_Y par cette projection. Comme il en existe une seule par 3. et par le début de la démonstration, on en déduit que W, ainsi que $(\mathrm{id}_{S_0} \times q)^{-1}(Z_Y)$ sont irréductibles.

La première partie du lemme est donc établie.

Démontrons la seconde:

le fait que q_* coincide avec l'application de Kodaira de Z_Y dans $S_0 \times Y$ est une propriété générale des images directes de cycles.

Montrons que q_* est génériquement injective. En effet, pour tout s de S_0, on a $|q_*(Z_s)| = q(|Z_s|)$, si $(Z_s)_{s \in S_0}$ est la famille universelle de cycles de X_0 paramétrée par S_0. Comme, d'autre part, on a aussi:

$$|Z_s| = q^{-1}(q(|Z_s|)) \quad \text{puisque} \quad Z_{S_0} = (\mathrm{id}_{S_0} \times q)^{-1}(Z_Y),$$

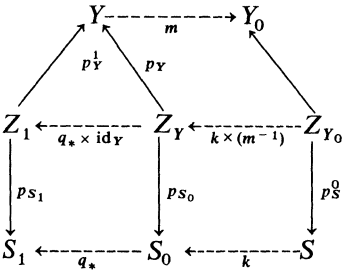
on en déduit que q_* est injective sur l'ouvert de Zariski dense de S_0^* constitué des points s au voisinage desquels q_* est analytique. CQFD

Achevons la démonstration du lemme:

Les propriétés 1, 2, 3 de l'énoncé sont démontrées. Pour démontrer la propriété 4, il suffit maintenant de démontrer que le morphisme $p_S : Z_Y \longrightarrow S_0$ est algébrique.

Soient $S_1 = q_*(S)$ et $Z_1 \subset S_1 \times Y$ le graphe de la famille analytique de cycles de Y paramétrée par S_1, et $p_{S_1} : Z_1 \longrightarrow S_1$, et $p_Y^1 : Z_1 \longrightarrow Y$ les projections naturelles.

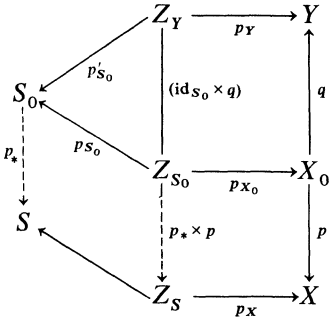
On a donc un diagramme commutatif d'applications méromorphes, dans lequel les flèches horizontales sont biméromorphes.



Le théorème d'algébricité de [C2] montre alors que les morphismes p_Y^1 et p_S^0 sont algébriques, puisque Z_1 (resp. Z_{Y_0}) est le graphe de la famille universelle de cycles de Y (resp. de S) paramétrée par S_1 (resp. Y_0) et que $\overline{S_1^*} = S_1$ (resp. $\overline{Y_0^*} = Y_0$).

Il en résulte immédiatement que les morphismes p_Y et p_{S_0} sont également algébriques. CQFD

Démonstration du corollaire. Le théorème précédent montre qu'il existe un diagramme commutatif de morphismes et applications méromorphes surjectifs:



dans lequel p_Y et p'_{S_0} sont des morphismes algébriques, $\dim_{\mathbb{C}}(X_0) = \dim_{\mathbb{C}}(X)$, p est biméromorphe ainsi que $p_* \times p$, et où l'application de Kodaira $K : S_0 \dashrightarrow \mathscr{C}(Y)$ de Z_Y dans $(S_0 \times Y)$ est biméromorphe de S_0 sur $K(S_0)$.

Soient $\tilde{S} \subset S \times S_0$ le graphe de p_*, et (s_1, s_0) un élément de $\tilde{S}$ tel que $s_1 \in S^*$ et $a(|Z_{s_1}|) = 0$. Alors Z_{s_0} est un cycle de X_0 de même dimension (pure) que $\dim(|Z_{s_1}|)$ et tel que $p(|Z_{s_0}|) = |Z_{s_1}|$. Les composantes irréductibles de $|Z_{s_0}|$ sont donc toutes de dimension algébrique zéro.

Le triangle supérieur gauche du diagramme montre alors, puisque p'_{S_0} est un morphisme algébrique à fibre générique irréductible, que p'_{S_0} est en fait une modification.

Puisque $K : S_0 \dashrightarrow K(S_0)$ est biméromorphe, il en résulte que son image est Y, identifié au sous-espace analytique compact de $\mathscr{C}(Y)$ constitué des points de Y affectés de la multiplicité un, puisque p_Y est surjectif. Donc p_Y est aussi une modification.

Ceci signifie donc que les images directes par q des cycles $(Z_{s_0})_{s_0 \in S_0}$ de X_0 sont les points de Y. Autrement dit: la famille $(Z_{s_0})_{s_0 \in S_0}$ est la famille des fibres du morphisme géométriquement plat $q : X_0 \longrightarrow Y$ et $S_0 = Y$. Il en résulte que Z_{S_0} s'identifie au graphe dans $Y \times X = S_0 \times X$ du morphisme q. Donc, p_{X_0} est un isomorphisme. Comme $(p_* \times p)$ est une modification, on a l'égalité:

- $\dim_{\mathbb{C}}(Z_S) = \dim_{\mathbb{C}}(Z_{S_0})$
- $\dim_{\mathbb{C}}(Z_{S_0}) = \dim_{\mathbb{C}}(X_0)$ résulte de ce qui précède.
- $\dim_{\mathbb{C}}(X_0) = \dim_{\mathbb{C}}(X)$ résulte de l'énoncé du théorème.

Montrons que si X est faiblement Kählérien, S est une composante irréductible de $\mathscr{C}(X)$: si σ est une composante irréductible de $\mathscr{C}(X)$ qui contient S, alors σ est compacte, $\overline{\sigma^*} = \sigma$, σ est pleine dans X et l'un des cycles paramétrés par σ est de dimension algébrique zéro. Si $Z_\sigma \subset \sigma \times X$ est le graphe de la famille universelle de

cycles de X paramétrée par σ, alors Z_σ est irréductible et contient Z_S. Comme $\dim(Z_\sigma)=\dim(X)=\dim(Z_S)$ par le corollaire, on a $Z_S=Z_\sigma$ et donc $S=\sigma$. CQFD

Appendice 1

Application de Kodaira

Soient $(Z_s)_{s\in S}$ une famille analytique de cycles d'un espace analytique X, paramétrée par un espace analytique compact et irréductible S, telle que le graphe $Z_S\subset S\times X$ de la famille $(Z_s)_{s\in S}$ soit de support irréductible. (Cette condition est vérifiée si, par exemple, $\overline{S^*}=S$).

Puisque S est compact, $|Z_S|$ est compact, et son image X_S par la seconde projection de $S\times X$ est donc un sous-espace analytique compact et irréductible de X, et la restriction $p_X:|Z_S|\to X_S$ de cette seconde projection à $|Z_S|$ est un morphisme propre et surjectif. Soit $p_S:|Z_S|\to S$ la restriction à $|Z_S|$ de la première projection de $S\times X$.

Le théorème d'aplatissement géométrique de [B3], appliqué au morphisme $p_X:Z_S\to X_S$, montre que l'on a une application méromorphe $j_{p_X}:X_S\dashrightarrow\mathscr{C}(|Z_S|)$ associée à p_X. Plus précisément, si X'_S est l'ouvert de Zariski dense de X_S constitué des points en lesquels X_S est normal et la fibre de p_X de dimension pure $d=\dim(Z_S)-\dim(X_S)$, alors la restriction $p'_X:Z'_S\to X'_S$ de p_X à $Z_S=p_X^{-1}(X'_S)$ est un morphisme propre et géométriquement plat d'après [B1, chapitre 1, Par. 2, Theoreme 1]; c'est-à-dire que les fibres de p'_X forment une famille analytique de cycles de Z'_S paramétrée par X'_S. Soit $j'_{p_X}:X'_S\to\mathscr{C}(Z_S)$ le morphisme classifiant cette famille. Le théorème d'aplatissement géométrique montre que j'_{p_X} se prolonge en une application méromorphe $j_{p_X}:X_S\dashrightarrow\mathscr{C}(Z_S)$. Le théorème d'image directe montre aussi que l'on a un morphisme d'image directe par p_S,

$$(p_S)_*:j_{p_X}(X_S)\dashrightarrow\mathscr{C}(S).$$

Définition 1. L'application méromorphe composée $(p_S)_*\circ j_{p_X}:X_S\dashrightarrow\mathscr{C}(S)$ est appelée l'application de Kodaira de la famille $(Z_s)_{s\in S}$, et notée K_S.

Définition 2. Dans la situation ci-dessus, on dira que la famille $(Z_s)_{s\in S}$ sépare presque les points de X_S si l'application K_S est une application biméromorphe de X_S sur son image.

Cette notion généralise celle d'amplitude:

Exemple. Si X est une variété algébrique projective et irréductible de dimension au moins égale à 2, et D un fibré en droites très ample sur X, alors si on prend $S=\mathbb{P}(H^0(X,D))$, on a:

- $\overline{S^*}=S$ par le théorème de Bertini.
- $X_S=X$.
- S sépare presque les points de X puisque l'application K_S envoie alors X dans la grassmanienne des hyperplans de S, et est un isomorphisme de X sur son image.

Remarque. On peut donner une version relative de l'application de Kodaira dans la situation suivante: Soient $\varphi:X\to Y$ un morphisme, et $(Z_s)_{s\in S}$ une famille analytique

de cycles de X paramétrée par un espace analytique irréductible S, et telle que $j_Z : S \to \mathscr{C}(X)$ désignant le morphisme associé, on ait :

- $j_Z(S) \subset \mathscr{C}(X/Y)$.
- $\overline{S^*} = S$.
- $\varphi_* : S \to Y$ est propre, où φ_* est le morphisme d'image directe par φ.

Il est alors facile de vérifier que, si $Z_S \subset S \times X$ est le graphe de la famille $(Z_s)_{s \in S}$, Z_S est irréductible et Y-propre pour le composé de φ avec la projection (p_X) de Z_S sur X.

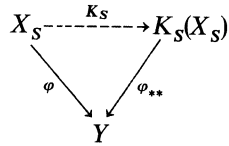
Si $X_S = (p_X)(|Z_S|)$, on peut donc définir une application méromorphe dite application de Kodaira relative $K_S : X_S \dashrightarrow \mathscr{C}(S)$; on vérifie aisément que :

1. $K_S(X_S) \subset \mathscr{C}(S/Y)$ où S est au-dessus de Y par φ_*.

On a donc un morphisme d'image directe par (φ_*), noté $\varphi_{**} : \mathscr{C}(S/Y) \to Y$ que l'on peut restreindre à $K_S(X_S)$.

On note encore φ_{**} cette restriction.

2. L'application méromorphe K_S est au-dessus de Y, c'est-à-dire que le diagramme suivant commute :



Appendice 2

Addition de cycles – factorisation de Stein irréductible

Soient p un entier positif ou nul, et k un entier >0.

L'application $A_k(p) : [\mathscr{C}_p(X)]^k \to \mathscr{C}_p(X)$, qui à k cycles associe leur somme, est analytique [B1].

Posons alors : $\mathscr{C}_p^\infty(X) = \bigcup_{k=1}^{+\infty} [\mathscr{C}_p(X)]^k$ (réunion disjointe et $A(p) : \mathscr{C}_p^\infty(X) \to \mathscr{C}_p(X)$ est le morphisme dont la restriction à chacun des $[\mathscr{C}_p(X)]^k$ est $A_k(p)$. Alors, le théorème 2 de [B2] admet aussi la conséquence suivante :

Proposition 0. *Le morphisme $A(p)$ est propre et fini.*

Démonstration. Soit K un sous-ensemble compact de $\mathscr{C}_p(X)$.

Soit $M = \sup_{c \in K} (\mathrm{vol}_h(c))$. Alors $0 < M < +\infty$.

Soit $\mathrm{Supp} c_K$ le support de K, défini comme suit : si $X_K \subset K \times X$ est la restriction à $K \times X$ du graphe de la famille universelle de cycles paramétrée par $\mathscr{C}_p(X)$, et si p_K, p_X sont les restrictions à X_K des premières et secondes projections, alors $\mathrm{Supp} c_K = p_X(X_K)$ et comme X_K est K-propre et K compact, $\mathrm{Supp} c_K$ est compact.

Si $\gamma \in [\mathscr{C}_p^\infty(X)]$ est dans $(A(p))^{-1}(K)$, alors $\gamma = (Z^1, \ldots, Z^k)$, et comme le support de $A(p)(\gamma)$ est la réunion $|Z^1| \cup \ldots \cup |Z^k|$, on en déduit que $|Z^j| \subset \mathrm{Supp} c_K$ si $j = 1, \ldots, k$.

Comme, de plus, $\mathrm{vol}_h(A(p)(\gamma)) = \sum_{j=1}^{k} \mathrm{vol}_h(Z^j)$ on en déduit donc que $\mathrm{vol}_h(Z^j) \leqq M$ pour $j = 1, \ldots, k$.

Pour conclure, il suffit donc de montrer que k est borné. Soit donc K' l'ensemble des $c \in \mathscr{C}_p(X)$ tels que:

- $\mathrm{vol}_h(c) \leqq M$.
- $\mathrm{supp}(c) \subset \mathrm{Supp}\, c_K$.

K' est compact. Posons alors $m = \inf_{c \in K'} (\mathrm{vol}_h(c))$: la compacité de K' entraîne immédiatement que $m > 0$.

De ce qui précède, on déduit aussi que $\mathrm{vol}_h(Z^j) \geqq m$ pour $j = 1 \ldots k$, et donc que $k \cdot m \leqq \sum_{j=1}^{k} \mathrm{vol}_h(Z^j) \leqq M$.

On a donc: $k \leqq M/m$. Ceci établit la proposition. (La finitude est évidente.)

Le corollaire 3 admet de nombreuses conséquences:

Si $Z = \sum_{i \in I_X} n_i Z_i$ est un cycle de X, nous noterons $\mu t(Z)$ et appellerons multiplicité totale de X le nombre entier $\mu t(Z) = \sum_{i \in I} n_i$.

Pour tout couple d'entiers strictement positifs p et m, notons $\mathscr{C}_p^{(m)}(X)$ le sous-ensemble de $\mathscr{C}_p(X)$ égal à $(\mu t)^{-1}([m, +\infty[)$ où $\mu t : \mathscr{C}_p(X) \to \mathbb{N}$ est l'application définie ci-dessus.

Proposition 1. *$\mathscr{C}_p^{(m)}(X)$ est analytique fermé dans $\mathscr{C}_p(X)$.*

Démonstration. $\mathscr{C}_p^{(m)}(X) = \bigcup_{p=0}^{+\infty} A_m(p)\,[(\mathscr{C}_p(X)^m]$; or, puisque $A_m(p)$ est propre, par la proposition 0, chacun des sous-ensembles de la réunion disjointe du membre de droite est analytique fermé dans $\mathscr{C}_p(X)$.

En particulier $\mathscr{C}_p^{(1)}(X) - \mathscr{C}_p^{(2)}(X) = \mathscr{C}_p^*(X)$ est un ouvert de Zariski de $\mathscr{C}_p(X)$: c'est aussi l'ensemble des cycles de X de support irréductible et de multiplicité 1. On notera $\mathscr{C}^*(X)$ son adhérence dans $\mathscr{C}(X)$. On notera $\mathscr{C}^*(X)$ [resp. $\mathscr{C}^*(X)$] la réunion des $\mathscr{C}_p^*(X)$ [resp. $\mathscr{C}_p^*(X)$] pour $p \in \mathbb{N}$. On a donc:

Proposition 2. *$\mathscr{C}^*(X)$ est un ouvert de Zariski de $\mathscr{C}(X)$.*

Lorsque X est un espace analytique et k un entier, le groupe symétrique d'ordre k, σ_k, agit analytiquement sur X^k par:

$$\sigma(z_1, \ldots, z_k) = (z_{\sigma(1)}, \ldots, z_{\sigma(k)}) \quad \textit{pour} \quad \sigma \in \sigma_k.$$

Rappelons que X^k admet un quotient analytique, noté $\mathrm{Sym}^k(X)$, pour cette action. Conformément à l'usage de [B1], *on notera $\pi_k : X^k \to \mathrm{Sym}^k(X)$ la projection et $\mathrm{Sym}^k(X) \Subset X$ le sous-espace analytique fermé de $\mathrm{Sym}^k(X) \times X$ qui est l'ensemble des $(\pi_k(z_1, \ldots, z_k), \zeta)$ tels que ζ est égal à l'un des z_i pour $i = 1, \ldots, k$.*

Si $\varphi : X^k \to Y$ est un morphisme tel que $\varphi \circ \sigma = \varphi$ pour tout σ de σ_k, il existe un unique morphisme, noté $\mathrm{Sym}^k(\varphi) : \mathrm{Sym}^k(X) \to Y$ tel que $\mathrm{Sym}^k(\varphi) \circ \pi_k = \varphi$.

Proposition 3. *Soit S une composante irréductible de $\mathscr{C}_p(X)$. Il existe alors des entiers l et h, des entiers $(k_1, \ldots, k_h)$ tels que $k_1 + \ldots + k_h = l$, et des composantes irréductibles $(S_1, \ldots, S_h)$ distinctes de $\overline{\mathscr{C}_p^*(X)}$ tels que l'image de $\mathrm{Sym}^{k_1}(S_1) \times \ldots \times \mathrm{Sym}^{k_h}(S_h)$ par*

$$A_S = \left\{ (A_h(p)) \circ \left(\prod_{i=1}^{h} \mathrm{Sym}^{k_i}(A_{k_i}(p)) \right) \right\}$$

soit égale à S, où

$$\prod_{i=1}^{h} \mathrm{Sym}^{k_i}(A_{k_i}(p)) : \left\{\prod_{i=1}^{h} \mathrm{Sym}^{k_i}(S_i)\right\} \to \{\mathscr{C}_p(X)\}^h$$

est le morphisme produit des $\{\mathrm{Sym}^{k_i}(A_{k_i}(p))\} : \mathrm{Sym}^{k_i}(S_i) \to \mathscr{C}_p(X)$.

Les données $(h, k_1, \ldots, k_h, S_1, \ldots, S_h)$ *sont alors définies de manière unique par cette condition, et le morphisme d'addition ainsi défini:*

$\prod_{i=1}^{h} \mathrm{Sym}^{k_i}(S_i) \to S$ *est une modification propre et finie.*

Remarques. 1. Il s'agit d'une décomposition de S en «produit» de composantes irréductibles de $\overline{\mathscr{C}^*(X)}$.

2. Le morphisme $A_S : \mathrm{Sym}^{k_1}(S_1) \times \ldots \times \mathrm{Sym}^{k_h}(S_h) \to S$ n'est pas nécessairement un isomorphisme, comme le montre l'exemple suivant: $X = \mathbb{P}_3$ et S est le composante irréductible de $\mathscr{C}_1(\mathbb{P}_3)$ constituée des cycles Z qui sont somme de deux coniques planes. Lorsque ces deux coniques planes dégénèrent en quatre droites coplanaires, la fibre audessus du cycle correspondant à ces quatre droites a, en général, $\binom{4}{2} = 6$ points.

Démonstration. Soit l le plus grand des entiers k tels que S soit contenu dans $\mathscr{C}_p^{(k)}(X)$. Il est alors clair que $S \cap \{\mathscr{C}_p^{(l)}(X) - \mathscr{C}_p^{(l+1)}(X)\}$ est un ouvert de Zariski dense de S, et puisque la restriction de $\mathrm{Sym}^l(A_l(p))$ à $\mathrm{Sym}^l(\overline{\mathscr{C}^*(X)})$ est une modification finie de son image [qui est $\mathscr{C}_p^{(l)}(X)$] par le lemme ci-dessous, il existe une unique composante irréductible σ de $\mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)})$ telle que:

- $\mathrm{Sym}^l(A_l(p)) : \sigma \to A$ est une modification finie.
- σ est l'unique composante irréductible de $\mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)})$ dont l'image par $\mathrm{Sym}^l(A_l(p))$ est S.

Il existe alors des composantes irréductibles de $\overline{\mathscr{C}_p^*(X)}$, définies à l'ordre près par le lemme ci-dessous, $(S_1, \ldots, S_l)$ telles que la projection du produit $(S_1 \times \ldots \times S_l)$ sur $\mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)})$ soit égale à S.

Il existe alors des entiers h et $(k_1, \ldots, k_h)$ tels que $\left(\sum_{i=1}^{h} k_i\right) = l$ et une partition de l'ensemble $J = [1, \ldots, l]$ en sous-ensembles $J_1, \ldots, J_h$ tels que $S_j = S_{j'}$ si et seulement si j et j' appartiennent au même sous-ensemble J_i de J. Les entiers k_i sont les cardinaux des ensembles J_i pour $i = 1, \ldots, h$.

Toujours d'après le lemme ci-dessous, la projection de $(S_1 \times \ldots \times S_l)$ sur σ se factorise par le morphisme $(S_1 \times \ldots \times S_l) \to \prod_{i=1}^{h} (\mathrm{Sym}^{k_i}(S_i))$, où pour tout i, S_i est la composante irréductible de $\overline{\mathscr{C}_p^*(X)}$ égale à S_j pour tout j de J_i, et la projection du produit $\left(\prod_{i=1}^{l} \mathrm{Sym}^{k_i}(S_i)\right) \to \sigma$ est une modification finie.

On obtient alors la proposition 3 en composant cette projection avec la projection de σ sur S induite par $\mathrm{Sym}^l(A_l(p))$. Il reste à établir le:

Lemme. *Pour tout entier* l *et tout élément* γ *de* $\{\mathscr{C}_p^{(l)}(X) - \mathscr{C}_p^{(l+1)}(X)\}$ *il existe un unique élément* ζ *de* $\mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)})$ *tel que* $\mathrm{Sym}^l(A_l(p))(\zeta) = \gamma$.

Démonstration. L'hypothèse signifie que $\mu t(\gamma) = l$. Il suffit alors de décomposer γ en cycles irréductibles de multiplicité totale égale à 1. CQFD

Soit $(Z_s)_{s\in S}$ une famille analytique de cycles de X paramétrée par un espace analytique S. Nous noterons $j_z : S \to \mathscr{C}(X)$ le morphisme associé. Soit $S^* = j_Z^{-1}(\mathscr{C}^*(X))$; S^* est un ouvert de Zariski de S. On notera $\overline{S^*}$ l'adhérence de S^* dans S. On notera $Z_S \subset S \times X$ le graphe de $(Z_s)_{s\in S}$.

Définition. Nous appellerons diagramme de Stein associé à $(Z_s)_{s\in S}$ la donnée de tout couple $((Z_\sigma)_{\sigma\in\Sigma}, \lambda)$ constitué d'une famille analytique de cycles $(Z_\sigma)_{\sigma\in\Sigma}$ paramétrée par un espace analytique Σ et d'un morphisme surjectif $\lambda : \Sigma \to S$ tels que, dans le diagramme commutatif suivant:

$$\begin{array}{ccc} Z_\Sigma & \xrightarrow{\lambda \times \mathrm{id}_X} & Z_S \\ {\scriptstyle p_\Sigma}\downarrow & & \downarrow{\scriptstyle p_S} \\ \Sigma & \xrightarrow{\lambda} & S \end{array}$$

- $\overline{\Sigma^*} = \Sigma$.
- $(\lambda \times \mathrm{id}_X)(Z_\Sigma) = Z_S$.
- $(\lambda \times \mathrm{id}_S)^{-1}(Z_S)$ est une réunion de composantes irréductibles de Z_Σ.

On dira qu'un diagramme de Stein $((Z_{\sigma'})_{\sigma'\in\Sigma'}, \lambda')$ provient d'un diagramme $((Z_\sigma)_{\sigma\in\Sigma}, \lambda)$ s'il existe un morphisme $\mu : \Sigma' \to \Sigma$ tel que $\lambda \circ \mu = \lambda'$.

Proposition 4. *Soit X un espace analytique; pour toute famille analytique de cycles $(Z_s)_{s\in S}$ paramétrée par un espace analytique S, il existe un unique diagramme de Stein $((Z_\sigma)_{\sigma\in\Sigma_0}, \lambda_0)$ final (c'est à dire dont proviennent les autres).*

Dans le diagramme:

$$\begin{array}{ccc} Z_{\Sigma_0} & \xrightarrow{\lambda_0 \times \mathrm{id}_X} & Z_S \\ {\scriptstyle p_{\Sigma_0}}\downarrow & & \downarrow{\scriptstyle p_S} \\ \Sigma_0 & \xrightarrow{\lambda_0} & S \end{array}$$

le morphisme λ_0 est propre et fini, et le morphisme $(\lambda_0 \times \mathrm{id}_X) : Z_{\Sigma_0} \to Z_S$ est une modification propre et finie.

On appellera le morphisme $\lambda_0 : \Sigma_0 \to S$ *la factorisation de Stein irréductible* de la famille $(Z_s)_{s\in S}$.

La proposition 4 admet la conséquence suivante:

Corollaire 1. *Soient X un espace analytique normal et $\pi : X \to S$ un morphisme propre et surjectif.*

Si

$$\begin{array}{ccc} X & \xrightarrow{\varrho} & \Sigma_0 \\ & {\scriptstyle \pi}\searrow \quad \swarrow{\scriptstyle \lambda_0} & \\ & S & \end{array}$$

est la factorisation de Stein (uselle) du morphisme π, il existe un ouvert de Zariski dense Σ_0^* de Σ_0 tel que les fibres de la restriction $\varrho^*: X^* = \varrho^{-1}(\Sigma_0^*) \to \Sigma_0^*$ de ϱ à X^* sont toutes irréductibles.

Démonstration de la proposition. Supposons S irréductible.

Soit $j_Z: S \to \mathscr{C}_p(X)$ le morphisme associé à la famille analytique de cycles $(Z_s)_{s\in S}$. Pour tout entier k, notons $S^{(k)}$ l'image réciproque de $\mathscr{C}_p^{(k)}(X)$ par j_Z: c'est un sous-espace analytique fermé de S. Il existe donc, puisque $S^{(k)} \supset S^{(k+1)}$, un plus grand entier k tel que $S^{(k)} = S$. Notons l cet entier.

On a alors un diagramme:

$$\begin{array}{ccc} & & S \\ & & \downarrow j_Z \\ \mathrm{Sym}^l(A(p)): \mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)}) & \longrightarrow & \mathscr{C}_p^{(l)}(X) \end{array}$$

Puisque $j_Z(S)$ recontre $\{\mathscr{C}_p^{(l)}(X) - \mathscr{C}_p^{(l+1)}(X)\}$, le lemme de la démonstration de la proposition 3 montre qu'il existe une unique composante irréductible S_1 de $S \underset{\mathscr{C}_p^{(l)}(X)}{\times} \mathrm{Sym}(\overline{\mathscr{C}_p^*(X)})$ qui se projette sur S par la première projection. (Si S n'était pas irréductible, on peut procéder ainsi pour toute composante irréductible de S.)

Soit alors Σ_0 l'image réciproque de S_1 par la projection de $W = S \times (\mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)}) \overset{*}{\times} \overline{\mathscr{C}_p^*(X)})$ sur $S \times \mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)})$.

Notons alors:

- λ_0 la restriction à Σ_0 de la projection de W sur S.
- $j_{\Sigma_0}: \Sigma_0 \to \overline{\mathscr{C}_p^*(X)}$ la restriction à Σ_0 de la projection de W sur $\mathscr{C}_p^*(X)$.
- Z_{Σ_0} le graphe dans $(\Sigma_0 \times X)$ de la famille de cycles de X paramétrée par Σ_0 et associée au morphisme j_{Σ_0}.

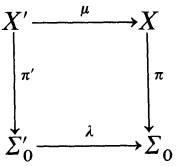
Il est alors clair que le couple constitué de cette famille de cycles et du morphisme λ_0 est un diagramme de Stein associé à la famille $(Z_s)_{s\in S}$, et que le morphisme λ_0 est propre, fini et surjectif. Le fait que $(\lambda_0 \times \mathrm{id}_X): Z_{\Sigma_0} \to Z_S$ soit une modification finie provient du fait que les composantes irréductibles de Σ_0 sont en correspondance bijective avec les composantes irréductibles de Z_S.

Si $((Z_\sigma)_{\sigma\in\Sigma}, \lambda)$ est un diagramme de Stein associé à la famille $(Z_s)_{s\in S}$, et le morphisme $j_{Z_\Sigma}: \Sigma \to \overline{\mathscr{C}^*(X)}$ associé à la famille $(Z_\sigma)_{\sigma\in\Sigma}$, le morphisme produit $(\lambda \times j_{Z_\Sigma}): \Sigma \to S \times \overline{\mathscr{C}_p^*(X)}$ est d'image propre et finie sur S puisque $(\lambda \times \mathrm{id}_X)(Z_\Sigma)$ est égal à la Z_S. Pour toute composante irréductible de S, cette image a un certain degré l sur cette composante irréductible, ce qui permet en fait de définir un morphisme $\Sigma \to S \times \mathrm{Sym}^l(\overline{\mathscr{C}_p^*(X)}) \overset{*}{\times} \overline{\mathscr{C}_p^*(X)}$ dont on peut vérifier que l'image est Σ_0.

Il est facile de vérifier que la famille $(Z_\sigma)_{\sigma\in\Sigma}$ est l'image réciproque par le morphisme $\Sigma \to \Sigma_0$ ainsi construit de la famille $(Z_\sigma)_{\sigma\in\Sigma_0}$.

Démonstration du corollaire. Quitte à restreindre le morphisme π à un ouvert de Zariski dense de Σ_0, on peut supposer que π est géométriquement plat (voir [B3]), et on peut aussi supposer que S est irréductible.

Soit alors $\lambda_0 : \Sigma'_0 \to \Sigma_0$ la factorisation de Stein irréductible de la famille $(X_\sigma)_{\sigma \in \Sigma_0}$ (qui est la famille des fibres de π). Si



est le diagramme associé, comme μ est une modification finie, et que X est normal, c'est un isomorphisme.

Comme les fibres de $(\pi' \circ \mu^{-1})$ sont irréductibles, et que Σ'_0 est un morphisme fini et propre, puisque les fibres de π sont connexes, λ doit être aussi un morphisme de degré un. Ce qui prouve que c'est un isomorphisme au points de Σ_0 où Σ_0 est normal. Les fibres de π sont donc bien génériquement irréductibles. CQFD

Bibliographie

[B1] Barlet, D.: Familles analytiques de cycles. In: Fonctions de plusieurs variables complexes II. Séminaire Norguet. Lecture Notes in Mathematics, Vol. 482, pp. 1–158. Berlin, Heidelberg, New York: Springer 1975

[B2] Barlet, D.: Convexité de l'espace des cycles. Bull. Soc. Math. France **106**, 373–397 (1978)

[B3] Barlet, D.: Majoration de volume et forme géométrique du théorème d'aplatissement. C. R. Acad. Sci. **288**, (1979) et séminaire Lelong (1978–1979)

[C1] Campana, F.: Images directes de cycles par un morphisme. In: Fonctions de plusieurs variables complexes. IV. Séminaire Norguet. Lecture Notes in Mathematics, Vol. 807, pp. 25–65. Berlin, Heidelberg, New York: Springer 1980

[C2] Campana, F.: Algébricité et compacité dans l'espace des cycles. C. R. Acad. Sci. **289** (1979) et Math. Ann. **251**, 7–18 (1980)

[C3] Campana, F.: Diviseurs non-polaires (submitted for publication)

[C4] Campana, F.: Application de l'espace des cycles. Prepublication Université Nancy I (Mai 1980)

[C5] Campana, F.: Coréduction d'un espace analytique faiblement Kählérien compact. Invent. Math. **63**, 187–223 (1981)

[F] Fischer, G.: Complex analytic geometry. In: Lecture Notes in Mathematics, Vol. 538. Berlin, Heidelberg, New York: Springer 1976

[F–F] Fischer, G., Forster, O.: Ein Endlichkeitssatz für Hyperflächen auf kompakten komplexen Räumen. J. Reine Angew. Math. **306**, 88–93 (1979)

[F1] Fujiki, A.: Closedness of the Douady space of a compact Kähler space. Publ. Res. Inst. Math. Sci. **14**, 1–52 (1978)

[F2] Fujiki, A.: On the Douady space of a compact space in the category C (Submitted for publication)

[G] Grothendieck, A.: Techniques de construction en géométrie analytique. Séminaire Cartan (1960/61)

[U] Ueno, K.: Classification theory of algebraic varieties and compact complex spaces. In: Lecture Notes in Mathematics, Vol. 493. Berlin, Heidelberg, New York: Springer 1975

Reçu le 12 juin 1980

Math. Ann. 256, 191–197 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Symmetrische Hilbertsche Modulformen und Modulfunktionen zu $\mathbb{Q}(\sqrt{17})$

C. F. Hermann

Moltkestraße 27, D-6900 Heidelberg, Bundesrepublik Deutschland

In der vorliegenden Arbeit wird die Struktur des Rings der symmetrischen Hilbertschen Modulformen geraden Gewichts zu $\sqrt{17}$ bestimmt. Über die Quotientenbildung von Modulformen erhält man darüber hinaus Erzeugende für den Körper der symmetrischen Modulfunktionen zu $\sqrt{17}$.

Wir verwenden folgende Bezeichnungen:

$$H = \{z \in \mathbb{C}, \operatorname{Im} z > 0\},$$

$$H^2 = H \times H,$$

$$H_2 = \{Z = X + iY \in M_2(\mathbb{C}),\ Z = Z^t,\ Y \text{ ist positiv definit}\}$$

$$\mathfrak{o} = \left\{a + b\frac{1+\sqrt{17}}{2},\ a, b \in \mathbb{Z}\right\}.$$

Die zu $\alpha \in \mathfrak{o}$ konjugierte Zahl bezeichnen wir mit $\bar{\alpha}$. Mit Γ bezeichnen wir die Hilbertsche Modulgruppe zu $\sqrt{17}$, also die Gruppe aller Substitutionen $H^2 \to H^2$

$$(z_1, z_2) \mapsto \left(\frac{\alpha z_1 + \beta}{\gamma z_1 + \delta}, \frac{\bar{\alpha} z_2 + \bar{\beta}}{\bar{\gamma} z_2 + \bar{\delta}}\right)$$

mit $\alpha, \beta, \gamma, \delta \in \mathfrak{o}$ und $\alpha\delta - \beta\gamma = 1$.

Wir können Γ mit der Gruppe $\mathrm{Sl}_2(\mathfrak{o})/\pm 1$ identifizieren. Sei $v : \Gamma \to \mathbb{C}$ ein (abelscher) Charakter und sei $k \in \mathbb{Z}$. Eine auf H^2 holomorphe Funktion f, welche

$$f\left(\frac{\alpha z_1 + \beta}{\gamma z_1 + \delta}, \frac{\bar{\alpha} z_2 + \bar{\beta}}{\bar{\gamma} z_2 + \bar{\delta}}\right) = v(M) \cdot (\gamma z_1 + \delta)^k \cdot (\gamma z_2 + \delta)^k \cdot f(z_1, z_2)$$

erfüllt für alle $M = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix} \in \Gamma$ und alle $(z_1, z_2) \in H^2$, heißt Modulform zu Γ vom Gewicht k und zum Charakter v. Die Gesamtheit der Modulformen zu festem Charakter und Gewicht bilden einen Vektorraum, welchen wir mit $[\Gamma, v, k]$ bezeichnen. Eine Modulform heißt symmetrisch bzw. schiefsymmetrisch, falls $f(z_2, z_1) = f(z_1, z_2)$ bzw. $f(z_2, z_1) = -f(z_1, z_2)$ gilt für alle $(z_1, z_2) \in H^2$.

0025-5831/81/0256/0191/$01.40

Auf H^2 meromorphe Funktionen, welche unter Γ invariant sind, heißen Modulfunktionen.

Wir definieren nun die Thetakonstanten auf H_2.

Für $m_1, m_2, m_1', m_2' \in \{0,1\}$ setzen wir $m=(m_1,m_2)$ und $m'=(m_1',m_2')$. Sei $g=(g_1,g_2)\in\mathbb{Z}^2$ und $Z\in H_2$.

Wir setzen $Z[g+m]=\left(g_1+\frac{m_1}{2}, g_2+\frac{m_2}{2}\right) Z \begin{pmatrix} g_1+\frac{m_1}{2} \\ g_2+\frac{m_2}{2} \end{pmatrix}$ und $(g+m)(m')^t$

$$=\left(g_1+\frac{m_1}{2}, g_2+\frac{m_2}{2}\right)\cdot\begin{pmatrix} m_1' \\ m_2' \end{pmatrix}.$$

Die auf H_2 holomorphe Funktion

$$\Theta_m(Z)=\sum_{g\in\mathbb{Z}^2} e^{\pi i Z[g+m]+(g+m)(m')^t}$$

heißt Thetakonstante der Charakteristik $m=(m_1 m_2 m_1' m_2')$. Falls $(-1)^{(m_1 m_1'+m_2 m_2')}=1$ gilt, heißt m gerade. Man kann zeigen, daß Θ_m genau dann nicht identisch auf H_2 verschwindet, wenn m gerade ist (Igusa [7]).

Wir setzen $A=\begin{pmatrix} \tau & \bar{\tau} \\ \bar{\tau} & -\tau \end{pmatrix}$ mit $\tau=\sqrt{\frac{1}{2\sqrt{17}}+\frac{1}{2}}$ und $\bar{\tau}=\sqrt{\frac{-1}{2\sqrt{17}}+\frac{1}{2}}$. Es gilt $A=A^t=A^{-1}$.

Durch $\psi(\alpha)=A\begin{pmatrix} \alpha & 0 \\ 0 & \bar{\alpha} \end{pmatrix}A$, $\alpha\in\mathfrak{o}$, wird eine Darstellung von $\mathfrak{o}$ durch Matrizen aus $M_2(\mathbb{Z})$ definiert. Sie ist explizit durch

$$\psi\left(a+b\frac{1+\sqrt{17}}{2}\right)=\begin{pmatrix} a+b & 2b \\ 2b & a \end{pmatrix}$$

gegeben.

Wir setzen $\varphi(z_1,z_2)=A\begin{pmatrix} z_1 & 0 \\ 0 & z_2 \end{pmatrix}A$.

Das Paar (φ,ψ) ist die (im wesentlichen) einzige Modulareinbettung für Γ. Modulareinbettungen wurden von Hammond in [3] ausführlich untersucht.

Wir bilden nun die auf H^2 holomorphen Funktionen $\Theta_m(\varphi(z_1,z_2))$.

Da die Gruppe Γ von den speziellen Substitutionen

$$(z_1,z_2)\mapsto\left(-\frac{1}{z_1},-\frac{1}{z_2}\right)$$

und

$$(z_1,z_2)\mapsto(z_1+\alpha, z_2+\bar{\alpha}), \qquad \alpha\in\mathfrak{o},$$

erzeugt wird, kennt man das Transformationsverhalten von $\Theta_m(\varphi(z_1,z_2))$, wenn man weiß, wie sich $\Theta_m(Z)$ unter $Z\mapsto -Z^{-1}$ und $Z\mapsto Z+\psi(\alpha)$, $\alpha\in\mathfrak{o}$ transformiert.

Nach einiger Rechnung läßt sich zeigen:

Satz 1. *Die Funktion* $\Theta(z_1, z_2) = (\Theta_{1111}(\varphi(z_1, z_2)))^2$ *ist eine symmetrische Hilbertsche Modulform vom Gewicht* 1 *und zu einem Charakter* v_0, *welcher durch* $v_0\left(\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)\right) = -1$ *und* $v_0\left(\left(\begin{smallmatrix} 1 & \alpha \\ 0 & 1 \end{smallmatrix}\right)\right) = e^{\frac{\pi i}{2}(\alpha+\bar{\alpha})}$ *gegeben ist.*

Wir bestimmen nun die Nullstellen von $\Theta_{1111}(\varphi(z_1, z_2))$.

Satz 2. *Die Funktion* $\Theta_{1111}(\varphi(z_1, z_2))$ *verschwindet in den Diagonalpunkten* $(z, z) \in H^2$ *von erster Ordnung. Jede Nullstelle von* $\Theta_{1111}(\varphi(z_1, z_2))$ *ist äquivalent modulo* Γ *zu einem Diagonalpunkt.*

Beweis. Wir entnehmen Satz 5.1 in der schon zitierten Arbeit [3] von Hammond, daß die Nullstellen des Produkts

$$\prod_m \Theta_m(\varphi(z_1, z_2))$$

in den schief-hermiteschen Kurven F_1, F_2 und F_4 enthalten sind. Schief-hermitesche Kurven wurden von Hirzebruch in den Arbeiten [4–6] ausführlich untersucht. Wir entnehmen [6, Satz 3.2], daß die Kurven F_1, F_2 und $F_4 \backslash F_1$ modulo Γ irreduzibel sind.

Es ist leicht zu zeigen, daß $\prod_{m \backslash \{(1111)\}} \Theta_m(\varphi(z_1, z_2))$ auf $F_2 \cup F_4 \backslash F_1$ wirklich verschwindet.

In [7] hat Igusa bewiesen, daß in einem gegebenen Punkt $Z \in H_2$ höchstens eine gerade Thetakonstante verschwinden kann. Daher verschwindet $\Theta_{1111}(\varphi(z_1, z_2))$ auf $F_2 \cup F_4 \backslash F_1$ nicht. Die Funktion $\Theta_{1111}(\varphi(z_1, z_2))$ ist schiefsymmetrisch und verschwindet daher in Diagonalpunkten.

Daß die Thetakonstanten in den angegebenen Punkten von erster Ordnung verschwinden, zeigt eine einfache Abschätzung (siehe die Arbeit [2] von Freitag).

Wir konstruieren nun Modulformen vom Gewicht 3.

Dazu setzen wir $(m_1 m_2 m_1' m_2') := \Theta_m(\varphi(z_1, z_2))$ für $m = (m_1 m_2 m_1' m_2')$ und definieren:

$$\begin{aligned}
\eta_1(z_1, z_2) = &-(1100)(0011)(0000) + (1100)(0010)(0001) \\
&+(1001)(0110)(0000) + (1001)(0100)(0010) \\
&+(1000)(0100)(0011) + (1000)(0110)(0001) \\
\eta_2(z_1, z_2) = &\ (1100)(0011)(0000) + (1100)(0010)(0001) \\
&+(1001)(0110)(0000) - (1001)(0100)(0010) \\
&+(1000)(0100)(0011) - (1000)(0110)(0001)
\end{aligned}$$

mit diesen Bezeichnungen gilt:

Satz 3. $\eta_1^2, \eta_2^2 \in [\Gamma, v_0, 3]$, $\eta_1 \cdot \eta_2 \in [\Gamma, v_0^{-1}, 3]$.

Satz 3 läßt sich mit Hilfe der Thetatransformationsformeln nachweisen. Wir verzichten auf rechnerische Details.

Satz 4. *Die Funktion*

$$G_2(z_1, z_2) := \frac{\eta_1^2(z_1, z_2) - 4\eta_2^2(z_1, z_2)}{\Theta(z_1, z_2)}$$

ist eine symmetrische, nicht identisch verschwindende Hilbertsche Modulform zu Γ vom Gewicht 2 und zum trivialen Charakter.

Beweis. $\eta_1^2 - 4\eta_2^2$ ist holomorph auf H^2. In Diagonalpunkten zerfallen die Funktionen $\Theta_m(\varphi(z_1, z_2))$ in Produkte elliptischer Thetakonstanten. Diese sind folgendermaßen definiert:

$$\vartheta_{mm'}(z) = \sum_{g \in \mathbb{Z}} (-1)^{m'} \cdot e^{\pi i z \left(g + \frac{m}{2}\right)^2}.$$

Es gilt:

$$\Theta_{m_1 m_2 m_1' m_2'}(\varphi(z, z)) = \vartheta_{m_1 m_1'}(z) \cdot \vartheta_{m_2 m_2'}(z).$$

Nach einer kleinen Rechnung erhält man:

$$\eta_1^2(z, z) - 4\eta_2^2(z, z) = 0 \quad \text{für alle} \quad t \in H.$$

Nach Satz 2 ist die Funktion $\dfrac{\eta_1^2(z_1, z_2) - 4\eta_2^2(z_1, z_2)}{\Theta_{1111}(\varphi(z_1, z_2))}$ holomorph auf H^2 und schiefsymmetrisch.

Wir folgern $G_2(z_1, z_2) \in [\Gamma, 1, 2]$.

Es gilt $\varphi\left(\dfrac{5+\sqrt{17}}{2} z, \dfrac{5-\sqrt{17}}{2} z\right) = z \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix}$ und

$$\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 2 & 2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}.$$

Die Funktionen $\Theta_m\left(\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix} z \begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}\right)$ lassen sich leicht auf die Funktionen $\Theta_m(z)$ zurückführen.

Unter Verwendung von $\lim\limits_{y \to \infty} \vartheta_{10}(iy) = 0$ und $\lim\limits_{y \to \infty} \vartheta_{00}(iy) = \lim\limits_{y \to \infty} \vartheta_{01}(iy) = 1$ erhalten wir

$$\lim_{y \to \infty} G_2\left(\frac{5+\sqrt{17}}{2} iy, \frac{5-\sqrt{17}}{2} iy\right) = -16.$$

Damit ist Satz 4 bewiesen.

Wir setzen

$$G_4 = \eta_1 \cdot \eta_2 \cdot \Theta$$
$$H_4 = \Theta^4$$
$$G_6 = \eta_1^2 \cdot \Theta^3$$
$$H_6 = \eta_1^3 \cdot \eta_2.$$

Satz 5. *Der Ring der symmetrischen Hilbertschen Modulformen geraden Gewichts zu* $\sqrt{17}$ *wird von* G_2, G_4, H_4, G_6 *und* H_6 *erzeugt.*

Beweis. Sei $f \in [\Gamma, v, 2k]$. Die auf H definierte Funktion $(pf)(z) = f(z, z)$ ist eine elliptische Modulform vom Gewicht $4k$ zu einem gewissen Charakter v_p.

Der Ring der elliptischen Modulformen von durch 4 teilbarem Gewicht wird von pG_2 und pH_6 erzeugt. Sei nun $f_0 \in [\Gamma, 1, 2k]$. Wir können $pf(z) = 0$ für alle $z \in H$ voraussetzen.

Die Funktion $\dfrac{f_0(z_1, z_2)}{\Theta_{1111}(\varphi(z_1, z_2))}$ ist holomorph und schiefsymmetrisch. Wir folgern $\dfrac{f_0}{\Theta} \in [\Gamma, v_0^{-1}, 2k-1]$.

Es gilt $p\left(\dfrac{f_0}{\Theta}\right) = p(\eta_1 \cdot \eta_2) P_1(pG_2, pH_6)$ für ein geeignetes homogenes Polynom P_1. Die Modulform

$$f_1 = \frac{f_0}{\Theta} - \eta_1 \cdot \eta_2 \cdot P_1(G_2, H_6) \in [\Gamma, v_0^{-1}, 2k-1]$$

verschwindet in Diagonalpunkten.

Daher gilt $\dfrac{f_1}{\Theta} \in [\Gamma, v_0^{-2}, 2k-2]$.

$$\frac{f_1}{\Theta} \text{ ist eine Spitzenform.}$$

Es gibt daher ein homogenes Polynom P_2 in G_2 und H_6, so daß

$$f_2 = \frac{f_1}{\Theta} - \eta_1^2 \cdot \eta_2^2 \cdot P_2(G_2, H_6) \text{ in Diagonalpunkten verschwindet.}$$

Es folgt $\dfrac{f_2}{\Theta} \in [\Gamma, v_0, 2k-3]$.

Wir finden ein Polynom $P_3(G_2, H_6)$, so daß

$$f_3 = \frac{f_2}{\Theta} - \eta_1^2 \cdot P_3(G_2, H_6) \text{ in Diagonalpunkten verschwindet.}$$

Wir folgern $f_4 = \dfrac{f_3}{\Theta} \in [\Gamma, 1, 2k-4]$.

Wir können unser Reduktionsverfahren auch auf f_4 anwenden und kommen zu einem Ende, weil es keine Modulformen von negativem Gewicht gibt. Damit ist Satz 5 bewiesen.

Wir bezeichnen den graduierten Polynomring $\mathbb{C}[G_2, G_4, H_4, G_6, H_6]$ mit $\mathfrak{R}$. Das Bild eines Polynoms $P \in \mathfrak{R}$ im Ring der Modulformen bezeichnen wir mit $P(z_1, z_2)$.

Die Polynome

$$P_1 = G_4 G_6 - H_4 H_6$$
$$P_2 = G_4^3 - \tfrac{1}{4} G_6 H_6 + \tfrac{1}{4} G_4 G_6 G_2$$
$$P_3 = G_6^2 - 4 H_4 G_4^2 - G_6 H_4 G_2$$

und

$$P_4 = G_4^4 - \tfrac{1}{4} H_4 H_6 (H_6 - G_4 G_2)$$

werden auf die Nullform abgebildet, d. h. $P_i(z_1, z_2) = 0$ für $i = 1, 2, 3, 4$ und alle $(z_1, z_2) \in H^2$. Dies folgt unmittelbar aus den Definitionen.

Satz 6. *Sei $P \in \mathfrak{R}$ und es gelte $P(z_1, z_2) = 0$ für alle $(z_1, z_2) \in H$. Dann gibt es Polynome $Q_1, Q_2, Q_3 \in \mathfrak{R}$ mit $P = Q_1 P_1 + Q_2 P_2 + Q_3 P_3$.*

Beweis. Sei P ein homogenes Polynom aus $\mathfrak{R}$ vom Grad k. Die Modulform $P(z_1, z_2)$ vom Gewicht k verschwinde identisch auf H^2. Weil zwischen $G_2(z_1, z_2)$, $G_4(z_1, z_2)$ und $H_4(z_1, z_2)$ keine algebraischen (identischen) Relationen bestehen können, dürfen wir $k > 4$ voraussetzen. Es gibt geeignete Polynome Q_i, $i = 1, 2, 3, 4$, so daß

$$P - \sum_{i=1}^{4} Q_i P_i = H_4 R_{k-4} \quad \text{gilt mit} \quad R_{k-4} \in \mathfrak{R}.$$

Es folgt $R_{k-4}(z_1, z_2) = 0$ für alle $(z_1, z_2) \in H^2$. Also liegen die Polynome $\in \mathfrak{R}$, welche auf die Nullform abgebildet werden, in dem von den P_i, $i = 1, 2, 3, 4$ erzeugten Ideal.

Es gilt $P_4 = G_2 P_2 + \frac{1}{4}(H_6 + G_4 G_2) P_1$, also wird dieses Ideal bereits von P_1, P_2 und P_3 erzeugt.

Satz 6 ist damit bewiesen.

Satz 7. *Zwischen den Elementen $P_1, P_2, P_3 \in \mathfrak{R}$ möge die Relation $R_1 P_1 + R_2 P_2 + R_3 P_3 = 0$ bestehen, $R_i \in \mathfrak{R}$, $i = 1, 2, 3$. Dann gibt es ein Element $S \in \mathfrak{R}$ mit*

$$R_1 P_1 + R_2 P_2 + R_3 P_3 = S(G_6 P_1 - 4 H_4 P_2 - G_4 P_3).$$

Der Beweis läßt sich rein algebraisch führen.

Aus den Sätzen 6 und 7 läßt sich auf kombinatorischem Wege eine Dimensionsformel ableiten.

Satz 8. *Es gilt* $\dim[\Gamma, 1, 2k] = \frac{1}{3} k^2 + \frac{1}{2} k + s(k)$

mit

$s(k)$	=	1	1/6	2/3	1/2	2/3	1/6
falls $k \equiv$		0	1	2	3	4	5 mod 6

Satz 9. *Der Körper der symmetrischen Hilbertschen Modulformen zu $\mathbb{Q}(\sqrt{17})$ wird erzeugt von*

$$\frac{\eta_1(z_1, z_2) \cdot \eta_2(z_1, z_2)}{\Theta^3(z_1, z_2)} \quad \textit{und} \quad \frac{\eta_2^2(z_1, z_2)}{\eta_1^2(z_1, z_2)}.$$

Beweis. Bekanntlich läßt sich eine symmetrische Hilbertsche Modulfunktion $j(z_1, z_2)$ als Quotient zweier Modulformen gleichen Gewichts ausdrücken:

$$j(z_1, z_2) = \frac{f(z_1, z_2)}{g(z_1, z_2)}$$

für alle $(z_1, z_2) \in H^2$ und gewisse $f, g \in [\Gamma, 1, 2k]$, $k \in \mathbb{Z}$.

Satz 9 folgt daher aus den Sätzen 5 und 6.

Literatur

1. Busam, R.: Eine Verallgemeinerung gewisser Dimensionsformeln von Shimizu. Invent. Math. **11**, 110–149 (1970)
2. Freitag, E.: Zur Theorie der Modulformen zweiten Grades. Nachr. Akad. Wiss. Göttingen, Math. Phys. Kl. II. 151–157 (1965)
3. Hammond, W.: The modular groups of Hilbert and Siegel. J. Math. **88**, 497–515 (1966)
4. Hirzebruch, F.: Hilbert modular surfaces. Enseignement math. **29**, 183–281 (1973)
5. Hirzebruch, F.: Kurven auf Hilbertschen Modulflächen und Klassenzahlrelationen. In: Lecture Notes in Mathematics, Vol. 442. Berlin, Heidelberg, New York: Springer 1975
6. Hirzebruch, F.: Classification of Hilbert modular surfaces. In: Complex analysis and algebraic geometry, pp. 43–77. Cambridge: Cambridge University Press 1977
7. Igusa, I.: On the graded ring of theta-constants. Am. J. Math. **86**, 219–245 (1964)

Eingegangen am 21. April 1979; revidierte Fassung am 4. August 1980

Math. Ann. 256, 199–214 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Conducteur des représentations du groupe linéaire

H. Jacquet[1], I. I. Piatetski-Shapiro[2], and J. Shalika[3]

[1] Department of Mathematics, Columbia University, New York, NY 10027, USA
[2] Department of Mathematics, The Johns Hopkins University, Baltimore, MD 21218, USA
[3] Department of Mathematics, Yale University, New Haven, CT 06520, USA

1. Introduction et notations

(1.1) Soit F un corps local non archimédien, ψ un caractère additif non trivial de F, π une représentation admissible et irréductible du groupe $G_r = \mathrm{GL}(r, F)$ sur un espace complexe V. Dans [G–J], on a défini des facteurs $L(s, \pi)$ et $\varepsilon(s, \pi, \psi)$ qui sont de la forme:

$$\begin{aligned} L(s,\pi) &= P_\pi(q^{-s})^{-1}, \quad P_\pi \in \mathbb{C}[X], \quad P_\pi(0) = 1; \\ \varepsilon(s,\pi,\psi) &= \varepsilon_\pi(q^{-s},\psi), \quad \varepsilon_\pi(X,\psi) = cX^m; \end{aligned} \tag{1}$$

où l'on désigne par q le module du corps F. Nous ferons toujours l'hypothèse que l'exposant du caractère ψ est 0, c'est à dire que le plus grand idéal de F sur lequel ψ est trivial est l'anneau des entiers $\mathfrak{R}$. Lorsqu'il en est ainsi, l'entier m ne dépend que de π.

Par exemple, si $r=1$, alors π est un caractère de $F^\times$ (non nécessairement de module un) et l'entier m est positif. De plus si $m=0$ le caractère est non ramifié c'est-à-dire de la forme $\pi(x)=|x|^s$. Si au contraire m est strictement positif alors π est ramifié; de plus, si $\mathfrak{P}$ est l'idéal maximal de l'anneau $\mathfrak{R}$, on sait que π est trivial sur le groupe $1+\mathfrak{P}^m$ et non trivial sur les groupes $1+\mathfrak{P}^a$ avec $0<a<m$.

Pour $r>1$, on peut se demander si l'entier m est positif. On peut se demander aussi s'il existe une interprétation de l'entier m analogue à celle rappelée ci-dessus. Nous allons démontrer que tel est bien le cas, pourvu que la représentation π soit «générique» [cf. (1.2)]. Par exemple si $r=2$, alors une représentation générique est simplement une représentation de dimension infinie et le résultat est maintenant classique [C, D, N].

Cette question avait été d'abord soulevée par A. Weil (pour $r>2$) et résolue, tout au moins dans le cas cuspidal, par D. Kazdan et M. Novodvorsky. Peut-être d'ailleurs d'autres mathématiciens ont-ils, avant nous, résolu cette question complètement. Aussi ne prétendons nous à aucune espèce d'originalité. Mais comme beaucoup de mathématiciens se sont posés ou nous ont posé cette question, il ne nous à pas paru tout à fait inutile de publier notre démonstration. Elle est du reste calquée sur celle donnée dans [C] pour le cas $r=2$.

0025-5831/81/0256/0199/$03.20

(1.2) Introduisons maintenant les notations et les définitions dont nous aurons besoin. Nous noterons $\mathfrak{R}$, $\mathfrak{P}$, ω, v, q l'anneau des entiers, l'idéal maximal, une uniformisante, la valuation normalisée et le module du corps F respectivement. Nous poserons $\alpha(x)=|x|=q^{-v(x)}$. En particulier $|\omega|=q^{-1}$.

Soit A_r le sous-groupe des matrices diagonales dans G_r. Si les coefficients diagonaux de la matrice $a\in A_r$ sont les nombres $(a_1, a_2, \ldots, a_r)$ nous écrirons

$$a=\mathrm{diag}(a_1, a_2, \ldots, a_r). \tag{1}$$

Soit N_r le groupe trigonal strict supérieur dans G_r. Un caractère additif non trivial ψ de F étant choisi, nous définirons un caractère $\theta_{\psi,r}$ de N_r par la formule

$$\theta_{\psi,r}(n)=\prod_{i=1}^{r-1}\psi(n_{i,i+1}). \tag{2}$$

Une représentation admissible et irréductible π de G_r est dite *générique* s'il existe un espace $\mathscr{W}$ de fonctions sur G_r se transformant à gauche par $\theta_{\psi,r}$ qui soit invariant par translations à droite, la représentation de G_r sur $\mathscr{W}$ étant équivalente à π. L'espace $\mathscr{W}$ est alors unique et se note $\mathscr{W}(\pi;\psi)$. La condition d'être générique ne dépend pas de ψ. D'ailleurs si ψ_1 est un autre caractère additif non trivial de F, on a $\psi_1(x)=\psi(a_1x)$ avec a_1 dans $F^\times$. Pour W dans $\mathscr{W}(\pi;\psi)$ la fonction W_1 définie par

$$W_1(g)=W(ag), \qquad a=\mathrm{diag}(a_1^{r-1}, a_1^{r-2}, \ldots, a_1, 1),$$

est dans $\mathscr{W}(\pi;\psi_1)$. Enfin si π est générique la représentation contragrediente $\tilde{\pi}$ l'est aussi. De plus, si W est dans $\mathscr{W}(\pi;\psi)$, alors la fonction $\tilde{W}$ définie par

$$\tilde{W}(g)=W(w_r{}^tg^{-1}), \qquad w_r=\begin{pmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{pmatrix}, \tag{3}$$

est dans $\mathscr{W}(\tilde{\pi};\bar{\psi})$, où l'on désigne par $\bar{\psi}$ l'imaginaire conjugué de ψ.

(1.3) Quelques remarques sur l'organisation de cet article. Nous vérifions au Par. 2 des propriétés simples mais essentielles des polynômes $P_\pi(X)$ pour π générique. Au Par. 3 nous étudions les propriétés des fonctions se transformant à gauche par $\theta_{\psi,r}$ qui sont, à droite, fonctions propres de l'algèbre de Hecke. Au Par. 4, imitant la méthode du cas $r=2$, nous définissons le «vecteur essentiel» de l'espace $\mathscr{W}(\pi;\psi)$. Pour cela nous utilisons l'équation fonctionnelle (4.2)(1). Avec les résultats rappelés ou prouvés au Par. 2, c'est la seule propriété des facteurs L et ε dont nous aurons besoin. Les propriétés du vecteur essentiel sont établies au Par. 5: elles découlent aussitôt de l'identité (4.2)(1).

2. Le facteur L des représentations génériques

(2.1) Le but de ce paragraphe est de prouver le résultat suivant:

Proposition. *Soit π une représentation générique irréductible de G_r. Alors les polynômes*

$$P_\pi(X) \text{ et } P_{\tilde{\pi}}(q^{-1}X^{-1})$$

sont premiers entre eux dans l'anneau $\mathbb{C}[X, X^{-1}]$.

(2.2) Nous aurons à utiliser quelques résultats simples sur les représentations génériques et les représentations induites. Soit Q le sous-groupe parabolique standard de type $(r_1, r_2, \ldots, r_n)$ de G_r. Pour chaque entier $i, 1 \leqq i \leqq n$, soit π_i une représentation admissible et irréductible de G_{r_i}. Nour noterons $I(\pi_1, \pi_2, \ldots, \pi_n)$ la représentation de G_r induite par la représentation suivante de Q:

$$\begin{pmatrix} g_1 & & & \\ 0 & g_2 & & * \\ \vdots & & \ddots & \\ 0 & & & g_n \end{pmatrix} \mapsto \pi_1(g_2) \otimes \pi_2(g_2) \otimes \ldots \otimes \pi_n(g_n).$$

Cette représentation admet une suite de composition finie; pour que l'un des quotients irréductibles de la suite de composition soit générique, il faut et il suffit que chacun des π_i le soit, et il y a alors un seul quotient irreductible générique (cf. [R]). Si tous les π_i sont génériques nous noterons $\sigma(\pi_1, \pi_2, \ldots, \pi_n)$ l'unique composant générique de $I(\pi_1, \pi_2, \ldots, \pi_r)$. La «transitivité» de l'opération d'induction montre aussitôt que si les π_i sont génériques, alors:

$$\sigma(\pi_1, \pi_2, \ldots, \pi_n) = \sigma[\sigma(\pi_1, \pi_2, \ldots, \pi_{j-1}), \sigma(\pi_j, \pi_{j+1}, \ldots, \pi_n)]. \tag{1}$$

Comme la représentation contragrédiente d'une représentation générique est elle même générique, la relation évidente

$$I(\pi_1, \pi_2, \ldots, \pi_n)^{\sim} = I(\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_n). \tag{2}$$

entraîne

$$\sigma(\pi_1, \pi_2, \ldots, \pi_n)^{\sim} = \sigma(\tilde{\pi}_1, \tilde{\pi}_2, \ldots, \tilde{\pi}_n). \tag{3}$$

De même, on a

$$I(\pi_1, \pi_2, \ldots, \pi_n) \otimes \mu = I(\pi_1 \otimes \mu, \pi_2 \otimes \mu, \ldots, \pi_n \otimes \mu) \tag{4}$$

pour tout caractère μ de $F^{\times}$. D'où:

$$\sigma(\pi_1, \pi_2, \ldots, \pi_r) \otimes \mu = \sigma(\pi_1 \otimes \mu, \pi_2 \otimes \mu, \ldots, \pi_r \otimes \mu). \tag{5}$$

Enfin toute représentation admissible et irréductible de G_r est un composant d'une représentation induite $I(\pi_1, \pi_2, \ldots, \pi_2)$ où les π_i sont cuspidales («supercuspidal» dans la terminologie d'Harisch-Chandra); de plus, si $I(\pi_1, \pi_2, \ldots, \pi_n)$ et $I(\zeta_1, \zeta_2, \ldots, \zeta_s)$ ont un composant irréductible commun, alors $n = s$ et $(\pi_1, \pi_2, \ldots, \pi_n) = (\zeta_1, \zeta_2, \ldots, \zeta_n)$, à l'ordre près. Il s'en suit que toute représentation générique est de la forme

$$\pi = \sigma(\pi_1, \pi_2, \ldots, \pi_n),$$

où les π_i sont cuspidales et $(\pi_1, \pi_2, \ldots, \pi_n)$ est bien déterminé à l'ordre près.

Par exemple, si Q est le sous-groupe parabolique de type $(1, 1, \ldots, 1)$ dans G_r et si $\pi_1, \pi_2, \ldots, \pi_r$ sont des caractères de $F^{\times}$ tels que $\pi_i(\pi_{i+1})^{-1} = \alpha$, alors on a

$$\pi_i = \mu \alpha^{\frac{1}{2}(r+1)-i}, \quad 1 \leqq i \leqq r,$$

et

$$\sigma(\pi_1, \pi_2, \ldots, \pi_r) = \mu \otimes \sigma_r,$$

où σ_r est la représentation spéciale de Steinberg: cela résulte par exemple de ce que σ_r est de carré intégrable, donc générique (cf. [G–J, Par. 7; J2]).

(2.3) Vérifions maintenant la proposition (2.1) pour une représentation de la forme

$$\pi = \mu \otimes \sigma_r. \tag{1}$$

On a alors $\tilde{\pi} = \mu^{-1} \otimes \sigma_r$. Si μ est ramifié la proposition est évidente car alors

$$P_\pi(X) = 1, \quad P_{\tilde{\pi}}(X) = 1. \tag{2}$$

Si μ n'est pas ramifié, alors $\mu = \alpha^t$ pour un t convenable et [G–J, Par. 7]:

$$P_{\mu \otimes \sigma_r}(X) = P_{\sigma_r}(q^{-t}X) = 1 - q^{-t-\frac{1}{2}(r-1)}X, \tag{3}$$

$$P_{\mu^{-1} \otimes \sigma_r}(X) = 1 - q^{t-\frac{1}{2}(r-1)}X. \tag{4}$$

La proposition est donc encore évidente.

Passons maintenant au cas d'une représentation de la forme

$$\pi = \sigma(\pi_1, \pi_2, \ldots, \pi_r) \tag{5}$$

où les π_i sont des caractères non ramifiés.

Nous pouvons supposer que la suite $(\pi_1, \pi_2, \ldots, \pi_r)$ est de la forme

$$\left.\begin{array}{l} \mu_1^1, \mu_2^1, \ldots, \mu_{r_1}^1, \mu_1^2, \mu_2^2, \ldots, \mu_{r_2}^2, \ldots, \mu_1^m, \mu_2^m, \ldots, \mu_{r_m}^m \\ \mu_k^i (\mu_{k+1}^i)^{-1} = \alpha, \quad \mu_{r_i}^i (\mu_1^j)^{-1} \neq \alpha, \quad \text{si} \quad i \neq j. \end{array}\right\} \tag{6}$$

En posant $\mu_1^i = \alpha^{u_i}$, nous avons donc

$$\sigma(\mu_1^i, \mu_2^i, \ldots, \mu_{r_i}^i) = \sigma_{r_i} \otimes \alpha^{u_i - \frac{r_i - 1}{2}},$$

$$\pi = \sigma\left(\sigma_{r_1} \otimes \alpha^{u_1 - \frac{r_1 - 1}{2}}, \ldots, \sigma_{r_m} \otimes \alpha^{u_m - \frac{r_m - 1}{2}}\right).$$

Il résulte alors de [J2, (2.7.4)] que $P_\pi[X]$ divise le produit

$$\prod_i P_{\sigma_{r_i}}\left(q^{-u_i + \frac{1}{2}(r_i - 1)}X\right).$$

De même $P_{\tilde{\pi}}(X)$ divise le produit

$$\prod_i P_{\sigma_{r_i}}\left(q^{u_i - \frac{1}{2}(r_i - 1)}X\right).$$

Si $P_\pi(X)$ et $P_{\tilde{\pi}}(q^{-1}X^{-1})$ ont un facteur commun (non trivial) dans l'anneau $\mathbb{C}[X, X^{-1}]$, il existe au moins un couple (i,j) tel que les polynômes

$$P_{\sigma_{r_i}}\left(q^{-u_i + \frac{1}{2}(r_i - 1)}X\right) = 1 - q^{-u_i}X,$$

$$P_{\sigma_{r_j}}\left(q^{u_j - \frac{r_j - 1}{2}} q^{-1}X^{-1}\right) = 1 - q^{u_j - r_j}X^{-1},$$

soient identiques, au produit près par une unité. C'est impossible si $i=j$. Si $i\neq j$, cela impose $\alpha^{u_j-r_j}=\alpha^{u_i}$ ou encore $\mu_1^i(\mu_{r_j}^j)^{-1}=\alpha^{-1}$. Or ceci contredit (6). La proposition est donc prouvée dans ce cas.

(2.4) En général, écrivons π sous la forme

$$\pi=\sigma(\pi_1,\pi_2,\ldots,\pi_j,\pi_{j+1},\ldots,\pi_n),$$

où les représentations cuspidales $\pi_1,\pi_2,\ldots,\pi_j$ ne sont pas des caractères non-ramifiés tandis que $\pi_{j+1},\ldots,\pi_n$ sont des caractères non ramifiés de $F^\times$. Alors:

$$\pi=\sigma(\pi_1,\pi_2,\ldots,\pi_j,\zeta) \quad \text{avec} \quad \zeta=\sigma(\pi_{j+1},\ldots,\pi_n).$$

Toujours d'après [J2, (2.7.4)] le polynôme $P_\pi(X)$ divise le produit

$$P_\zeta(X)\prod_{1\leqq i\leqq j}P_{\pi_i}(X)$$

Mais $P_{\pi_i}(X)=1$ [loc. cit. (3.1)]. Donc $P_\pi(X)$ divise en fait $P_\zeta(X)$. De même $P_{\tilde{\pi}}(q^{-1}X^{-1})$ divise $P_{\tilde{\zeta}}(q^{-1}X^{-1})$. D'aprés (2.3) ces polynômes sont premiers entre eux. Le résultat s'en suit.

(2.5) Nous aurons aussi besoin du résultat auxiliaire suivant:

Proposition. *Soit π une représentation irréducible générique de G_r. Le degré de P est au plus r. Pour que π contienne la représentation unité du sous groupe compact maximal $K_r=\mathrm{GL}(r,\mathfrak{R})$, if faut et il suffit que le degré de P_π soit r.*

Démonstration. Ecrivons π sous la forme

$$\pi=\sigma(\pi_1,\pi_2,\ldots,\pi_m)$$

où les π_i sont cuspidales et $m\leqq r$. Une fois de plus $P_\pi(X)$ divise le produit $\prod P_{\pi_i}(X)$. Comme chaque P_{π_i} est de degré au plus un, on voit que P_π est de degré au plus r. Supposons maintenant que P_π soit de degré r. Alors $m=r$ et chaque P_{π_i} doit être de degré un. Cela signifie que chaque π_i est un caractère non ramifié de $F^\times$. Supposons la représentation induite $\zeta=I(\pi_1,\pi_2,\ldots,\pi_r)$ nonirréductible. Alors il existe au moins un couple (i,j) tel que $\pi_i\cdot\pi_j^{-1}=\alpha$ [B–Z, Theorem (4-11)]. On peut supposer que c'est le couple $(1,2)$. Alors on a:

$$\pi=\sigma(\zeta,\pi_3,\ldots,\pi_r), \qquad \zeta=\sigma(\pi_1,\pi_2)=\sigma_2\otimes\mu.$$

A nouveau P_π divise le produit

$$P_\zeta\prod_{i=3}^{r}P_{\pi_i}.$$

Mais comme P_ζ est de degré un [G–J, Par. 7] on obtient une contradiction. Ainsi ξ est irréductible et $\pi=\xi$. Donc π contient la représentation unité de K_r.

En sens inverse, si π est une représentation irréductible qui contient la représentation unité de K, ou comme nous dirons encore, une représentation *non-ramifiée*, alors π est le composant non-ramifié d'une représentation induite

$\xi = I(\pi_1, \pi_2, \ldots, \pi_r)$ où les π_i sont des caractères non-ramifiés de $F^\times$. Alors

$$P_\pi(X) = \prod_{1 \leqq i \leqq r} P_{\pi_i}(X)$$

et chaque P_{π_i} est de degré un [G–J, Par. 6]. Donc P_π est de degré r.

(2.6) *Remarque.* On peut montrer que l'énoncé obtenu en supprimant dans la Proposition (2.5) le mot générique est encore vrai.

3. Fonctions de Whittaker formelles

(3.1) Nous aurons à utiliser des résultats élémentaires sur les fonctions se transformant à gauche par le caractère $\theta = \theta_{\psi,r}$. Nous les énonçons une fois pour toutes.

Pour g dans G_r posons

$$\begin{aligned} &\alpha_i(g) = a_i/a_{i+1}, \qquad 1 \leqq i \leqq r-1 \\ &\alpha_r(g) = a_r \end{aligned} \tag{1}$$

si $g = nak$ avec $n \in N_r$, $k \in K_r$ et

$$a = \mathrm{diag}(a_1, a_2, \ldots, a_r).$$

Soit φ une fonction lisse à droite telle que

$$\varphi(ng) = \theta(n)\varphi(g).$$

Alors il existe une constante C telle que

$$\varphi(g) \neq 0 \Rightarrow |\alpha_i(g)| \leqq C, \qquad 1 \leqq i \leqq r-1. \tag{2}$$

En effet, comme φ n'a qu'un nombre fini de translatées à droite par K_r, il suffit de montrer l'existence d'un C tel que

$$\varphi(a) \neq 0 \Rightarrow |\alpha_i(a)| \leqq C, \qquad 1 \leqq i \leqq r-1.$$

Mais en prenant

$$n = \begin{pmatrix} 1 & & x_1 & & 0 \\ & 1 & & \ddots & \\ & & \ddots & & \\ 0 & & & 1 & x_{r-1} \\ & & & & \end{pmatrix}$$

on doit avoir

$$\varphi(an) = \varphi(ana^{-1}a) = \prod_{i=1}^{r-1} \psi(\alpha_i(a)x_i)\varphi(a).$$

Or $\varphi(a) = \varphi(an)$ si les $|x_i|$ sont assez petits. D'où la conclusion. On voit même que si φ est K_r-invariant alors on peut prendre $C = 1$.

De même on a

$$\varphi\left[\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1_{r-1} & u \\ 0 & 1 \end{pmatrix}\right] = \varphi\left[\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}\right]\psi[(0, \ldots, 0, 1)gu]. \tag{3}$$

On en conclut que la relation

$$\varphi\left[\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 1_{r-1} & u \\ 0 & 1 \end{pmatrix}\right] = \varphi\left[\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}\right], \quad \text{pour tout } \begin{pmatrix} 1_{r-1} & u \\ 0 & 1 \end{pmatrix} \in K_r,$$

a lieu pour tout g, si et seulement si la relation

$$\varphi\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} \neq 0$$

implique que la dernière ligne de g a tous ses coefficients entiers.

(3.2) Nous aurons à utiliser des résultats connus sur les représentations «non-ramifiées» du groupe G_r. Soit H_r l'algèbre de convolution des fonctions bi-invariantes par $K_r = \mathrm{GL}(r, \mathfrak{R})$ et à support compact. On sait que cette algèbre est isomorphe à l'algèbre S_r des polynômes en $X_1, X_1^{-1}, X_2, X_2^{-1}, \ldots, X_r, X_r^{-1}$ qui sont invariants par permutation des variables. Notons $T_1, T_2, \ldots, T_r$ les fonctions symétriques élémentaires des X_i. Ainsi:

$$T_1 = \sum_i X_i, \qquad T_2 = \sum_{i<j} X_i X_j, \ldots, T_r = \prod_i X_i. \tag{1}$$

Alors S_r est encore isomorphe à l'algèbre de polynômes $\mathbb{C}[T_1, T_2, \ldots, T_r, T_r^{-1}]$.

Dans la suite nous adopterons les conventions suivantes pour les mesures de Haar et les mesures invariantes. La mesure de Haar de G_r (et de K_r) est normalisée par la condition $\mathrm{mes}(K_r) = 1$, celle de N_r par $\mathrm{mes}(K_r \cap N_r) = 1$. La mesure invariante de $N_r \backslash G_r$ est le quotient des mesures précédentes.

Soit maintenant $(x_1, x_2, \ldots, x_r)$ une suite de r nombres *non nuls*. Elle détermine un morphisme $S_r \to \mathbb{C}$, à savoir

$$P \mapsto P(x_1, x_2, \ldots, x_r). \tag{2}$$

L'homomorphisme correspondant $\lambda : H_r \to \mathbb{C}$ s'obtient ainsi. Soit $I(x_1, x_2, \ldots, x_r)$ la représentation de G_r induite par le caractère

$$\begin{pmatrix} a_1 & & * \\ & a_2 & \\ & & \ddots \\ 0 & & a_r \end{pmatrix} \mapsto x_1^{v(a_1)} x_2^{v(a_2)} \ldots x_r^{v(a_r)} \tag{3}$$

du groupe trigonal supérieur. Alors l'espace de cette représentation contient un vecteur et un seul fixé par K_r et égal à un sur K_r. Soit f ce vecteur. Pour tout $\varphi \in H$ on a:

$$\int f(gh)\varphi(h)dh = \lambda(\varphi) f(g). \tag{4}$$

(3.3) Soit maintenant ψ un caractère additif non trivial de F d'exposant 0. Pour toute suite de nombres complexes non nuls $(x_1, x_2, \ldots, x_r)$, il existe une fonction W et une seule telle que

$$W(ngk) = \theta(n) W(g), \quad n \in N, \quad k \in K; \tag{1}$$

$$W(e) = 1; \tag{2}$$

pour tout $\varphi \in H$

$$\int W(gh)\varphi(h)dh = \lambda(\varphi)W(g), \tag{3}$$

où λ est le morphisme de H dans $\mathbb{C}$ décrit dans (3.2) (cf. [S]). Nous noterons cette fonction

$$W(g, x_1, x_2, \ldots, x_r; \psi). \tag{4}$$

Il résulte de l'unicité que

$$W(w\,{}^t g^{-1}, x_1, x_2, \ldots, x_r; \psi) = W(g, x_1^{-1}, x_2^{-1}, \ldots, x_r^{-1}; \bar{\psi}). \tag{5}$$

De plus si $q^{-s} = x$, alors

$$W(g, x_1, x_2, \ldots, x_r; \psi)|\det g|^s = W(g, x_1 x, x_2 x, \ldots, x_r x; \psi). \tag{6}$$

La formule explicite donnée en [S] montre qu'il existe pour chaque $g \in G_r$ un élément $W(g, X_1, X_2, \ldots, X_r; \psi)$ de S_r dont $W(g, x_1, x_2, \ldots, x_r; \psi)$ soit la spécialisation. On a ainsi défini une fonction sur G_r à valeurs dans S_r ; d'ailleurs ses valeurs sur un ensemble compact modulo N_r sont dans un sous-espace vectoriel de dimension *finie* de S_r. D'après (3.1), le support de $W(g, X_1, X_2, \ldots, X_r; \psi)$ est contenu dans l'ensemble des g tels que

$$|\alpha_i(g)| \leqq 1, \quad \text{pour} \quad 1 \leqq i \leqq r-1. \tag{7}$$

Si de plus $|\alpha_r(g)| \leqq 1$ alors $W(g, X_1, X_2, \ldots, X_r; \psi)$ est en fait dans $\mathbb{C}[X_1, X_2, \ldots, X_r]$.

(3.4) Nous aurons à considérer des «intégrales» de la forme

$$\int_{N_r \backslash G_r} W(g, X_1, X_2, \ldots, X_r; \psi)\varphi(g)|\det g|^s dg, \tag{1}$$

où φ est une fonction se transformant à gauche par $\theta_{\bar{\psi}, r} = \bar{\theta}_{\psi, r}$. La fonction φ sera supposée lisse à droite, c'est-à-dire invariante à droite par un sous-groupe compact ouvert de G_r. De plus le support de φ sera toujours contenu dans l'ensemble des g tels que

$$|\alpha_r(g)| \leqq C, \tag{2}$$

où C est une constante non nulle. L'intégrale est à interpréter comme série formelle

$$\sum a_n X^n, \quad X = q^{-s}, \tag{3}$$

à coefficients dans S_r, les coefficients étant donnés par l'intégrale ordinaire

$$a_n = \int_{N_r \backslash G_r} W(g, X_1, X_2, \ldots, X_r; \psi)\varphi(g)\mu_n(\det g)dg, \tag{4}$$

où μ_n est la fonction caractéristique de l'ensemble

$$\{x \in F^\times | v(x) = n\} \tag{5}$$

dans $F^\times$. Il est clair que les conditions (3.3)(7), (3.4)(2) et $v(\det g) = n$ impliquent que g est dans un ensemble compact mod N_r : la fonction à intégrer dans (4) est donc à support compact mod N_r et l'intégrale a bien un sens. De plus notons que les relations (3.3)(7) et (3.4)(2) entrainent $v(\det g) \geqq m$, où m est un entier qui dépend de C. Il en résulte que dans la série (3) on a $a_n = 0$ pour $n < m$.

(3.5) **Lemme.** *Gardons les hypothèses de* (3.4) *mais supposons de plus que* φ *soit invariante à droite par* K_r. *Alors la relation*

$$\int W(g,X_1,X_2,\ldots,X_r;\psi)\varphi(g)|\det g|^s dg=0$$

entraine $\varphi=0$.

Démonstration. If suffit de démontrer le lemme lorsque φ a un support compact modulo N_r. L'hypothèse du lemme s'écrit alors:

$$\int_{N_r\backslash G_r} W(g,x_1,x_2,\ldots,x_r;\psi)\varphi(g)dg=0, \tag{1}$$

pour tout $(x_1,x_2,\ldots,x_r)$. Montrons que cela entraine $\varphi=0$. Pour cela considérons la représentation π de G_r induite, au sens de Mackey, par le caractère (de module 1) $\theta_{\bar{\psi},r}$. Décomposons la en somme continue de représentations irréductibles:

$$\pi=\int \pi_x d\mu(x).$$

Soit V l'ensemble des fonctions lisses à droite et support compact modulo N_r dans l'espace de la représentation π. Comme V a une base linéaire dénombrable, il existe, pour presque tout x, une application linéaire A_x de V dans l'espace de la représentation π_x telle que

$$\langle\varphi_1,\varphi_2\rangle=\int\langle A_x\varphi_1,A_x\varphi_2\rangle d\mu(x), \qquad \varphi_1,\varphi_2\in V. \tag{2}$$

De plus, pour presque tout x, on a

$$A_x\pi(g)\varphi=\pi_x(g)A_x\varphi, \qquad g\in G_r, \qquad \varphi\in V.$$

Soit maintenant φ_0 un élément de V invariant sous K_r. Alors $A_x\varphi_0=0$ à moins que π_x ait un vecteur $v_x\neq 0$ invariant sous K_r. Supposons qu'il en soit ainsi; alors il existe une fonction W et une seule se transformant à gauche par $\theta_{\psi,r}$ telle que, pour tout $\varphi\in V$,

$$(A_x\varphi,v_x)=\int_{N_r\backslash G_r}\varphi(g)W(g)dg.$$

Comme v_x est K_r-invariant la fonction W est K_r-invariant à droite. De plus v_x doit être un vecteur propre pour l'algèbre H_r. Il en est donc de même de la fonction W. Par conséquent cette fonction est un multiple d'une fonction $W(x_1,x_2,\ldots,x_r;\psi)$. Si φ_0 satisfait (1) alors $(A_x\varphi_0,v_x)=0$ ou encore $A_x\varphi_0=0$. Ainsi $A_x\varphi_0=0$ pour presque tout x. D'après (2) on a $\varphi_0=0$. Q.E.D.

4. Le vecteur essentiel

(4.1) Nous sommes maintenant en mesure de définir le vecteur essentiel d'une représentation générique π de G_r. Rappelons une fois de plus que l'exposant de ψ est supposé être 0. Pour tout $W\in\mathcal{W}(\pi;\psi)$, posons

$$\Psi(X,W,X_1,X_2,\ldots,X_{r-1})$$
$$=\int_{N_{r-1}\backslash G_{r-1}} W\begin{pmatrix} g & 0\\ 0 & 1\end{pmatrix} W(g,X_1,X_2,\ldots,X_{r-1};\bar{\psi})$$
$$|\det g|^{s-1/2}dg, \qquad X=q^{-s}. \tag{1}$$

La fonction $g \mapsto W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}$ satisfait évidemment les conditions de (3.4) et l'intégrale représente une série de Laurent formelle à coefficients dans S_{r-1}. D'après (3.4)(6) on a, si U est une indéterminée,

$$\Psi(UX, W, X_1, X_2, \ldots, X_{r-1}) = \Psi(X, W, UX_1, UX_2, \ldots, UX_{r-1}). \tag{2}$$

Théorème. *Il existe un élément et un seul W de $\mathscr{W}(\pi;\psi)$ tel que*

(i) $W\left(g\begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix}\right) = W(g)$ *pour* $h \in K_{r-1}$,

(ii) $\Psi(X, W, X_1, X_2, \ldots, X_{r-1}) = \prod_{i=1}^{r-1} P_\pi^{-1}(XX_i)$,

le second membre de (ii) *étant la série formelle en X à coefficients dans S_{r-1} obtenue en développant $P_\pi^{-1}(XX_i)$ suivant les puissances croissantes de X.*

Démonstration de l'unicité. D'après le lemme (3.4) la restriction de W au sous-groupe

$$\left\{ \begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} \middle| \, g \in G_{r-1} \right\} \tag{3}$$

est uniquement déterminée. Or, d'après les résultats de [B–Z, Théorème (4.9)], un élément de $\mathscr{W}(\pi;\psi)$ est toujours déterminé par sa restriction au sous-groupe (3). L'unicité est donc évidente.

(4.2) *Démonstration de l'existence.* Nous nous appuierons sur une identité dans l'espace des séries de Laurent à coefficients dans S_{r-1}. Notons que cet espace peut être regardé comme un module sur l'anneau des polynômes en X, X^{-1} à coefficients dans S_{r-1}. De plus cet espace admet un automorphisme changeant la série $F(X) = \sum a_n X^n$ en la série

$$F(X^{-1}) = \sum a_n X^{-n} = \sum a_{-n} X^n.$$

Cet automorphisme est compatible avec l'automorphisme correspondant de l'algèbre des polynômes en X, X^{-1}. L'identité que nous avons en vue s'écrit, pour tout élément W de $\mathscr{W}(\pi;\psi)$:

$$\Psi(q^{-1}X^{-1}, \tilde{W}, X_1^{-1}, X_2^{-2}, \ldots, X_{r-1}^{-1}) \prod_{\pi=1}^{r-1} P_{\tilde{\pi}}(q^{-1}X^{-1}X_i^{-1})$$
$$= \prod_{i=1}^{r-1} \varepsilon_\pi(XX_i, \psi) \Psi(X, W, X_1, X_2, \ldots, X_{r-1}) \prod_{i=1}^{r-1} P_\pi(XX_i). \tag{1}$$

Dans cette formule $\tilde{W}$ est défini par (1.2)(3). Rappelons que c'est un élément de l'espace $\mathscr{W}(\tilde{\pi};\bar{\psi})$ attaché à la représentation contragrédiente $\tilde{\pi}$ et au caractère $\bar{\psi}$ imaginaire conjugué de ψ.

L'identité (1) est équivalente à la réunion d'une infinité d'identités dans l'anneau S_{r-1}. Pour démontrer (1) il suffit donc de démontrer l'identité analogue obtenue en remplaçant les X_i par des nombres complexes $x_i \neq 0$. On peut même se restreindre au cas où tous les x_i sont de module 1. Alors la représentation induite $\pi' = I(x_1, x_2, \ldots, x_{r-1})$ de G_{r-1} est irréductible et générique et la fonction

$W' = W(x_1, x_2, \ldots, x_{r-1}; \bar{\psi})$ est un élément de l'espace $\mathscr{W}(\pi'; \bar{\psi})$. De même $W(g, x_1^{-1}, x_2^{-1}, \ldots, x_{r-1}^{-1}; \psi)$ n'est autre que la fonction $(W')^{\sim}$. En posant, pour $\mathscr{W}' \in W(\pi'; \bar{\psi})$,

$$\Psi(X, W, W') = \int W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} W'(g) |\det g|^{s-1/2} dg, \qquad X = q^{-s},$$

les résultats énoncés dans [J–P–S2] montrent que

$$\Psi(q^{-1}X^{-1}, \tilde{W}, \tilde{W}') \prod_i P_{\tilde{\pi}}(q^{-1}X^{-1}x_i^{-1})$$
$$= \prod_i \varepsilon_\pi(Xx_i, \psi) \Psi(X, W, W') \prod_i P_\pi(Xx_i), \tag{3}$$

pout tout W'. En prenant $W' = W(x_1, x_2, \ldots, x_{r-1}; \bar{\psi})$ nous obtenons bien la spécialisation de (1).

D'après les remarques de (3.4) la série formelle $\Psi(X, W, X_1, X_2, \ldots, X_{r-1})$ n'a qu'un nombre fini de termes avec un exposant négatif. Il en est donc de même du membre de droite de (1). Au contraire le membre de gauche est une série formelle avec un nombre fini d'exposants positifs. L'identité (1) montre donc que chaque membre est un polynôme en X, X^{-1} à coefficients dans S_{r-1}. D'après (4.1)(2) il existe même un élément $\Xi(W, X_1, X_2, \ldots, X_{r-1})$ de S_{r-1} tel que

$$\Xi(W, XX_1, XX_2, \ldots, XX_{r-1})$$
$$= \Psi(X, W, X_1, X_2, \ldots, X_{r-1}) \prod_{i=1}^{r-1} P_\pi(XX_i). \tag{4}$$

D'une manière imagée, on peut écrire.

$$\Xi(W, X_1, X_2, \ldots, X_{r-1})$$
$$= \int_{N_{r-1} \backslash G_{r-1}} W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} W(g, X_1, X_2, \ldots, X_{r-1}; \bar{\psi}) |\det g|^{-1/2} dg \prod_{i=1}^{r-1} P_\pi(X_i). \tag{5}$$

Soit alors $I(\pi)$ le sous espace vectoriel de S_{r-1} engendré par les $\Xi(W, X_1, X_2, \ldots, X_{r-1})$ avec W dans $\mathscr{W}(\pi; \psi)$. C'est en fait un idéal de l'anneau S_{r-1}. En effet soit P un élément quelconque de S_{r-1} et φ l'élément correspondant de H_{r-1}. Alors

$$\int W(gh, X_1, X_2, \ldots, X_{r-1}) \varphi(h) dh$$
$$= W(g, X_1, X_2, \ldots, X_{r-1}; \bar{\psi}) P(X_1, X_2, \ldots, X_{r-1}).$$

Si W est dans $\mathscr{W}(\pi; \psi)$ la fonction W_1 définie par

$$W_1(g) = \int W\left(g\begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix}\right) \varphi(h^{-1}) |\det h|^{-1/2} dh$$

est encore dans $\mathscr{W}(\pi; \psi)$. En partant de la formule (5) un calcul formel facile à justifier montre que

$$\Xi(W_1, X_1, X_2, \ldots, X_{r-1}) = \Xi(W, X_1, X_2, \ldots, X_{r-1}) \cdot P(X_1, X_2, \ldots, X_{r-1}).$$

D'où notre assertion.

L'idéal $I(\pi)$ contient le produit $\prod P_\pi(X_i)$. En effet d'après les résultats de [G–K, Théorème (5.2)], il existe un élément W de $\mathscr{W}(\pi;\psi)$ tel que

$$W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} = \theta_{\psi,r-1}(n), \quad \text{si} \quad g=nk, \quad n\in N_{r-1}, \quad k\in K_{r-1}; =0 \text{ sinon}.$$

Pour ce W on a évidemment d'après (5):

$$\Xi(W,X_1,X_2,\ldots,X_{r-1}) = \prod_{i=1}^{r-1} P_\pi(X_i).$$

D'où la conclusion.

Finalement l'identité (1) s'écrit aussi:

$$\Xi(\tilde{W},q^{-1}X_1^{-1},q^{-1}X_2^{-1},\ldots,q^{-1}X_{r-1}^{-1}) = \prod_{i=1}^{r-1} \varepsilon_\pi(X_i,\psi)\Xi(W,X_1,X_2,\ldots,X_{r-1}). \tag{7}$$

En échangeant les rôles de π et $\tilde{\pi}$ cela montre que le produit $\prod_{i=1}^{r-1} P_\pi(q^{-1}X_i^{-1})$ appartient aussi à $I(\pi)$. D'après la proposition (2.1) on voit qu'il n'existe pas de suite $(x_1,x_2,\ldots,x_{r-1})$ avec $x_i\neq 0$, sur laquelle tous les éléments de $I(\pi)$ s'annulent. Si J est l'ideal correspondant de l'anneau $\mathbb{C}[T_1,T_2,\ldots,T_{r-1},T_{r-1}^{-1}]$ [notations de (3.1)], cela signifie que les éléments de J n'ont pas de zero commun $(t_1,t_2,\ldots,t_{r-1})$ avec $t_{r-1}\neq 0$. Il en résulte que

$$J=\mathbb{C}[T_1,T_2,\ldots,T_{r-1},T_{r-1}^{-1}]$$

et $I(\pi)=S_{r-1}$.

En particulier il existe un élément W de $\mathscr{W}(\pi;\psi)$ tel que

$$\Xi(W,X_1,X_2,\ldots,X_{r-1})=1.$$

En revenant aux définitions, cela s'écrit

$$\prod_{i=1}^{r-1} P_\pi(XX_i)\Psi(X,W,X_1,X_2,\ldots,X_{r-1})=1.$$

Or $\Psi(X,W,X_1,X_2,\ldots,X_{r-1})$ appartient à l'espace des séries de Laurent n'ayant qu'un nombre fini de termes avec exposant négatif; cet espace est un module *sans torsion* sur l'anneau $S_{r-1}[X,X^{-1}]$. La fonction W satisfait donc bien à la condition (ii) du théorème (4.1). Alors la formule

$$W_1(g)=\int_{K_{r-1}} W\left[g\cdot\begin{pmatrix} k & 0 \\ 0 & 1 \end{pmatrix}\right]dk$$

definit un element de $\mathscr{W}(\pi;\psi)$ qui satisfait à la condition (i). Comme l'intégrale Ψ ne change pas lorsqu'on remplace W par W_1 la fonction W_1 satisfait aussi (ii). Le théorème est donc démontré.

(4.3) *Exemple.* Supposons que π contienne la représentation unité de K_r. Alors π est équivalente à une représentation induite $I(y_1,y_2,\ldots,y_r)$ et

$W = W(y_1, y_2, \ldots, y_r; \psi)$ appartient à $\mathscr{W}(\pi; \psi)$. On a alors

$$P_\pi(X) = \prod_{i=1}^{r} (1 - y_i X).$$

D'après les résultats de [J–P–S2], on a aussi:

$$\Psi(X, W, X_1, X_2, \ldots, X_{r-1}) = \prod_{1 \leqq i \leqq r,\, 1 \leqq j \leqq r-1} (1 - y_i X_j X)^{-1}.$$

Comme W est invariant sous K_r, on voit que W satisfait les conditions du théorème (4.1).

(4.4) *Définition.* Sous les hypothèses du théorème (4.1), on dira que W est le vecteur essentiel de $\mathscr{W}(\pi; \psi)$.

5. Propriétés du vecteur essentiel

(5.1) Soit à nouveau π une représentation irréductible et générique de G_r. Rappelons que ψ désigne un caractère additif d'exposant 0. Soit m l'entier tel que

$$\varepsilon_\pi(X, \psi) = CX^m. \tag{1}$$

Dans ce paragraphe, nous établissons le théorème suivant:

Théorème. (i) *On a* $m \geqq 0$.

(ii) *Pour tout entier* $a > 0$ *soit* $K_r(a)$ *le sous-groupe de* K_r *forme des matrices du type*

$$\begin{pmatrix} h & v \\ u & x \end{pmatrix}, \quad h \in K_{r-1}, \quad x \in \mathfrak{R}^\times, \quad x \equiv 1 \bmod \mathfrak{P}^a, \quad u \equiv 0 \bmod \mathfrak{P}^a.$$

On pose aussi $K_r(0) = K_r$. *Alors le vecteur essentiel est invariant sous* $K_r(m)$ *et la dimension de l'espace des vecteurs fixés par* $K_r(m)$ *est un.*

(iii) *Si* $a \geqq 0$ *et s'il existe dans la représentation* π *un vecteur non nul fixé par* $K_r(a)$, *alors* $a \geqq m$.

(5.2) Pour démontrer le théorème nous introduirons, pour chaque entier $a \in \mathbb{Z}$, le sous-groupe $H(a)$ de G_r formé des matrices du type

$$\begin{pmatrix} h & u \\ 0 & 1 \end{pmatrix},$$

où h est dans K_{r-1} et où les éléments de la matrice colonne u sont dant $\mathfrak{P}^a$. Cela étant, le premier pas est de démontrer que le vecteur essentiel W est invariant sous $H(0)$. Par définition W est invariant sous le sous-groupe

$$\left\{ \begin{pmatrix} h & 0 \\ 0 & 1 \end{pmatrix} \middle| \, h \in K_{r-1} \right\}.$$

If suffit donc de démontrer que W est invariant sous le sous-groupe de $H(0)$ formé des matrices de $H(0)$ pour lesquelles $h = 1_{r-1}$. Comme W est déterminée par sa restriction au sous-groupe (4.1)(3), il suffit, d'après les remarques de (3.1), de prouver le lemme suivant:

Lemme. *Si W est le vecteur essentiel de $\mathcal{W}(\pi;\psi)$ alors la relation*

$$W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} \neq 0, \quad g \in G_{r-1},$$

entraine que la dernière ligne de g est à coefficients entiers.

Démonstration du lemme. Soit u le degré de P_π. On a donc $u \leqq r$ (2.5). Supposons que $u=r$. Alors π contient la représentation unité de K_r (loc. cit.). D'après (4.3) W est alors invariante sous K_r et notre assertion est alors évidente. Supposons maintenant $u \leqq r-1$ et écrivons

$$P_\pi(X) = \prod_{i=1}^{u} (1 - y_i X).$$

Nous distinguerons deux cas: $u<r-1$ et $u=r-1$.

Dans le cas $u<r-1$, on a l'identité (cf. [J–P–S2]):

$$\int_{N_u \backslash G_u} W\left[\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix}, X_1, X_2, \ldots, X_{r-1}; \bar{\psi}\right] W(g, y_1, y_2, \ldots, y_u; \psi) |\det g|^{s-\frac{1}{2}(r-u-1)} dg$$
$$= \prod (1 - y_j X_i X)^{-1}, \quad 1 \leqq j \leqq u, \quad 1 \leqq i \leqq r-1.$$

Définissons maintenant une fonction φ sur G_{r-1} par les conditions suivantes:

$$\varphi(ngk) = \theta_{\psi, r-1}(n)\varphi(g) \quad \text{si} \quad n \in N_{r-1}, \quad k \in K_{r-1};$$

si $a = \operatorname{diag}(a_1, a_2, \ldots, a_{r-1})$ alors

$$\varphi(a) = W(b, y_1, y_2, \ldots, y_u; \psi) \quad \text{avec} \quad b = \operatorname{diag}(a_1, a_2, \ldots, a_u)$$

si $|a_{u+1}| = \ldots |a_{r-1}| = 1$; $=0$ sinon. Il est clair que φ satisfait les conditions de (3.3). On a évidemment

$$\int_{N_{r-1} \backslash G_{r-1}} \varphi(g) W(g, X_1, X_2, \ldots, X_{r-1}; \bar{\psi}) |\det g|^{s+\frac{1}{2}(r-u-1)} dg = \prod P_\pi(X X_i)^{-1}.$$

D'après le lemme (3.4), on a donc

$$W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} = \varphi(g) |\det g|^{\frac{1}{2}(r-u)} \quad .$$

Le lemme en découle aussitôt.

Si $u=r-1$, on a (cf. [J–P–S2])

$$\int_{N_{r-1} \backslash G_{r-1}} W(g, y_1, y_2, \ldots, y_{r-1}; \psi) W(g, X_1, X_2, \ldots, X_{r-1}; \bar{\psi})$$
$$\Phi[(0, 0, \ldots, 0, 1)g] |\det g|^s dg = \prod (1 - y_i X_j X)^{-1},$$

où Φ est la fonction caractéristique de $\mathfrak{R} \times \mathfrak{R} \times \ldots \times \mathfrak{R}$ dans F^{r-1}. En raisonnant comme plus haut on trouve

$$W\begin{pmatrix} g & 0 \\ 0 & 1 \end{pmatrix} = W(g, y_1, y_2, \ldots, y_{r-1}; \psi) \Phi[(0, 0, \ldots, 0, 1)g] |\det g|^{1/2}$$

et, à nouveau, le lemme découle aussitôt de cette relation.

(5.3) Dénotons toujours par W le vecteur essentiel de $\mathscr{W}(\pi;\psi)$. Alors d'après (4.2)(1):

$$\Psi(q^{-1}X^{-1},\tilde{W},X_1^{-1},X_2^{-1},\ldots,X_{r-1}^{-1})\prod_{i=1}^{r-1}P_{\tilde{\pi}}(q^{-1}X^{-1}X_i^{-1})=\prod_{i=1}^{r-1}\varepsilon_{\pi}(XX_i,\psi).$$

Comme $\Psi(X,\tilde{W},X_1,X_2,\ldots,X_{r-1})$ n'a que des termes avec exposants positifs cela s'écrit aussi [cf. (5.1)(1)]:

$$\Psi(X,\tilde{W},X_1,X_2,\ldots,X_{r-1})=C^{r-1}q^{-m(r-1)}X^{-m(r-1)}\prod_{i=1}^{r-1}X_i^{-m}\prod_{i=1}^{r-1}P_{\tilde{\pi}}^{-1}(XX_i).$$

Soit W' le vecteur essentiel de $\mathscr{W}(\tilde{\pi};\bar{\psi})$. Alors

$$\Psi(X,W',X_1,X_2,\ldots,X_{r-1})=\prod_{i=1}^{r-1}P_{\tilde{\pi}}^{-1}(XX_i).$$

Définissons maintenant un élément W_1 de $\mathscr{W}(\tilde{\pi};\bar{\psi})$ par la formule

$$W_1(g)=W'(ga),\qquad a=\mathrm{diag}(\omega^m,\omega^m,\ldots,\omega^m,1).$$

Alors, il y a une constante C' telle que

$$\Psi(X,\tilde{W},X_1,X_2,\ldots,X_{r-1})=C'\Psi(X,W_1,X_1,X_2,\ldots,X_{r-1}).$$

D'après le lemme (3.4) cela implique $\tilde{W}=C'W_1$ sur le sous-groupe (4.2)(3). Donc, d'après une remarque déjà faite, $\tilde{W}=C'W_1$. Comme W' est invariante sous $H(0)$, cela implique que $\tilde{W}$ est invariante sous $H(m)$. En revenant à la définition de $\tilde{W}$ on voit que cela implique l'invariante de W sous le groupe transposé ${}^tH(m)$.

Si on avait $m<0$, les groupes $H(0)$ et ${}^tH(m)$ engendreraient G_r. La fonction W serait donc constante ce qui est absurde. On a donc $m\geqq 0$. Les groupes $H(0)$ et ${}^tH(m)$ engendrent alors $K(m)$ de sorte que W est bien invariante sous $K(m)$.

Soit maintenant u un entier $\geqq 0$ et W_1 un vecteur de $\mathscr{W}(\pi;\psi)$ invariant sous $K(u)$. L'invariance de W_1 sous $H(0)$ implique que

$$W_1\begin{pmatrix} g & 0\\ 0 & 1\end{pmatrix}\neq 0\Rightarrow|\alpha_i(g)|\leqq 1,\qquad 1\leqq i\leqq r-1$$

[cf. (3.1)]. En particulier dans l'intégrale (4.1)(1) pour $\Psi(X,W_1,X_1,X_2,\ldots,X_{r-1})$ on peut supposer que $|\det g|\leqq 1$ et $W(g,X_1,X_2,\ldots,X_{r-1};\bar{\psi})$ est dans $\mathbb{C}[X_1,X_2,\ldots,X_{r-1}]$ [cf. (3.3)]. La série n'a donc que des termes à exposants positifs et ses coefficients sont dans $\mathbb{C}[X_1,X_2,\ldots,X_{r-1}]$. Il en résulte que $\Xi(W_1,X_1,X_2,\ldots,X_{r-1})$ est un polynôme P en $X_1,X_2,\ldots,X_{r-1}$. D'autre part W_1 étant invariante sous ${}^tH(u)$, la fonction W_1 est invariante sous $H(u)$. La fonction W_2 définiee par

$$W_2(g)=\tilde{W}_1(ga),\qquad a=\mathrm{diag}(\omega^{-u},\omega^{-u},\ldots,\omega^{-u},1),$$

est donc invariante sous $H(0)$. Comme plus haut, il en résulte que $\Xi(W_2,X_1,X_2,\ldots,X_{r-1})$ est un élément de $\mathbb{C}[X_1,X_2,\ldots,X_{r-1}]$. Alors il existe un multiple constant Q de $\Xi(W_2,X_1,X_2,\ldots,X_{r-1})$ tel que:

$$\Xi(\tilde{W}_1,X_1,X_2,\ldots,X_{r-1})=(X_1X_2\ldots X_{r-1})^{-u}Q(X_1,X_2,\ldots,X_{r-1}).$$

D'après (4.2)(7) on a alors:

$$(X_1X_2\ldots X_{r-1})^u Q(q^{-1}X_1^{-1}, q^{-1}X_2^{-1}, q^{-1}X_2^{-1}, \ldots, q^{-1}X_{r-1}^{-1})$$
$$= C^{r-1}q^{-u(r-1)}(X_1, X_2, \ldots, X_{r-1})^m P(X_1, X_2, \ldots, X_{r-1}).$$

Si on avait $u<m$, $Q(q^{-1}, X_1^{-1}, \ldots, q^{-1}, X_{r-1}^{-1})$ serait un polynôme en $X_1, X_2, \ldots, X_{r-1}$ divisible par le produit $X_1X_2\ldots X_{r-1}$, ce qui est absurde. On a donc $u \geqq m$. Enfin si $u=m$ la relation ci-dessus montre que P est une constante. D'après l'unicité dans le théorème (4.1) la fonction W_1 est alors un multiple de W. Q.E.D.

(5.4) *Remarque.* Si π est une représentation admissible irréductible de G_r, générique ou non, on peut encore poser $\varepsilon_\pi(X, \psi) = CX^m$. On peut montrer que l'on a encore $m \geqq 0$ et que $m=0$ si et seulement si π contient la représentation unité de K_r. Toutefois, si $m>0$ la dimension de l'espace des vecteurs fixés par $K(m)$ peut être 0, comme le montre l'exemple des représentations de la forme $g \mapsto \chi(\det g)$, où χ est un caractère ramifié de $F^\times$.

Bibliographie

[B–Z] Bernstein, I.N., Zelevensky, A.V.: Induced representations of reductive p-adic groups. I. Ann. Sci. École Norm. Sup. **4**, 441–472 (1977)

[C] Casselman, W.: On some results of Atkin and Lehner. Math. Ann. **206**, 311–318 (1973)

[D] Deligne, P.: Formes modulaires et représentations de GL(2). In: Modular functions of one variable. II. Lecture Notes in Mathematics, Vol. 349. Berlin, Heidelberg, New York: Springer 1973

[G–J] Godement, R., Jacquet, H.: Zeta functions of simple algebras. In: Lecture Notes in Mathematics, Vol. 260. Berlin, Heidelberg, New York: Springer 1972

[G–K] Gelfand, I.M., Kazdan, D.A.: Representations of GL(n, K) where K is a local field. In: Lie groups and their representations, pp. 95–118. New York: Wiley 1975

[J1] Jacquet, H.: Generic representations. In: Non commutative harmonic analysis, Marseille-Luminy 1976. In: Lecture Notes in Mathematics, Vol. 587, pp. 91–100. Berlin, Heidelberg, New York: Springer 1976

[J2] Jacquet, H.: Principal L-functions of the linear group. In: automorphic forms, representations, and L-functions, Part 2. Am. Math. Soc. 63–95 (1979)

[J–P–S1] Jacquet, H., Piatetski-Shapiro, I.I., Shalika, J.: Automorphic forms on GL(3). I. Ann. of Math. **109**, 169–212 (1979)

[J–P–S2] Jacquet, H., Piatetski-Shapiro, I.I., Shalika, J.: Facteurs L et ε du groupe linéaire. C. R. Acad. Sci. Ser. A **289**, 59–61 (1979)

[N] Novodvorsky, M.: On certain stationary groups in infinite dimensional representations of Chevalley groups. Funkcional. Anal. i. Prilozen **5**, 87–88 (1971). Transl. Functional Anal. Appl. **5**, 74–76 (1971)

[R] Rodier, F.: Whittaker models for admissible representations of real algebraic groups. In: Harmonic analysis on homogeneous spaces. Proc. Symp. Pure Math. Am. Math. Soc. 425–430 (1976)

[S] Shintani, T.: On an explicit formula for class-1 "Whittaker functions" on GL$_n$ over p-adic fields, Proc. Japan Acad. **52**, 180–182 (1976)

Reçu le 4 août 1980

Math. Ann.256, 215–232 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Isoparametrische Hyperflächen in Sphären

II. Über die Zerlegung der Sphäre in Ballbündel

Hans Friedrich Münzner

Fachbereich Mathematik, Universität Bremen, D-2800 Bremen, Bundesrepublik Deutschland

Einleitung

Isoparametrische Hyperflächen in Räumen konstanter Krümmung sind geometrisch durch die Eigenschaft gekennzeichnet, daß die Invarianten des Formoperators, also etwa die Hauptkrümmungen, konstant sind. Da sich diese Eigenschaft auf reguläre Parallelflächen überträgt, bilden isoparametrische Familien, d. h. Parallelscharen derartiger Hyperflächen, den eigentlichen Gegenstand der Theorie. Wir setzen im folgenden immer voraus, daß alle Elemente der Familie zusammenhängend sind. Charakteristische Zahlen, die für alle Hyperflächen einer solchen Familie übereinstimmen, sind die Anzahl g der verschiedenen Hauptkrümmungswerte und die Menge ihrer Vielfachheiten, d. h. der Dimensionen der Eigendistributionen des Formoperators. Von Interesse sind nur die Fälle $g \geqq 2$.

Das Hauptergebnis von Teil I [16] dieser Arbeit war, daß jede isoparametrische Familie in der Standardsphäre $S^n \subset \mathbb{R}^{n+1}$ aus Stücken von Niveauflächen einer Funktion $V = F|S^n$ besteht, wobei $F : \mathbb{R}^{n+1} \to \mathbb{R}$ eine Polynomfunktion g-ten Grades ist, die den Differentialgleichungen

$$\Delta F = cr^{2g-2}, \qquad \|dF\|^2 = g^2 r^{2g}$$

$\left(r^2 = \sum_{i=1}^{n+1} x_i^2\right)$ genügt. Im Falle $c = 0$, der immer eintritt, wenn g ungerade ist, haben alle Hauptkrümmungswerte die gleiche Vielfachheit. Im Falle $c \neq 0$, treten die Vielfachheiten

$$m_+ = (g(n-1) - c)g^{-2}, \qquad m_- = (c + g(n-1))g^{-2}$$

auf, die jeweils der Hälfte der Hauptkrümmungswerte zukommen. Insbesondere ist jede Hyperfläche der Familie orientierbar, und zusammen mit einer Orientierung bestimmt sie eindeutig ein derartiges Polynom, das in [16] als ihr Cartansches Polynom bezeichnet wurde; einer Umorientierung entspricht der Übergang von F zu $-F$.

Alle Arbeiten, welche sich auf die algebraische Klassifikation spezieller Cartanscher Polynome einlassen, zeichnen sich durch einen geradezu abschrecken-

0025-5831/81/0256/0215/$03.60

den Rechenaufwand aus (vgl. etwa Cartan [1–4] und Ozeki u. Takeuchi [11]). Dies war seinerzeit mein Motiv, die globale Geometrie bzw. Topologie der Niveauflächen von $V=F|S^n$ ins Spiel zu bringen. Die wesentlichen differentialtopologischen Fakten sind (vgl. Teil I, Abschn. 5):

Die Funktion V besitzt nur die kritischen Werte $+1$ (Maximum) und -1 (Minimum). Die Mengen $M_\varrho = V^{-1}(\varrho)$ sind zusammenhängende Mannigfaltigkeiten für alle $\varrho \in [-1, +1]$, wobei $\dim M_k = n - m_k - 1$ ($k = \pm 1$; die Zahlen $m_1 = m_+$, $m_{-1} = m_-$ sind die möglichen Vielfachheiten der Hauptkrümmungswerte). Mit den Bezeichnungen $M = M_0$,

$$B_+ = B_1 = \{x \in S^n | V(x) \geqq 0\}, \qquad B_- = B_{-1} = \{x \in S^n | V(x) \leqq 0\}$$

folgt

I. Die Sphäre S^n entsteht durch Verklebung der berandeten Mannigfaltigkeiten (B_+, M) und (B_-, M) längs M.

II. B_k hat (als Tubenumgebung von M_k in S^n) die Struktur eines differenzierbaren Ballbündels über der Mannigfaltigkeit M_k ($k = \pm 1$).

Mit Standardargumenten werden im vorliegenden Teil II die Struktur des Kohomologierings $H^*(M, R)$ in Abhängigkeit von dem Zahlentripel (g, m_+, m_-) berechnet. Dabei ist $R = \mathbb{Z}$, falls M_+ und M_- beide orientierbar sind, und $R = \mathbb{Z}_2$ sonst. Unter anderem gilt $g = \frac{1}{2} \dim_R H^*(M, R)$.

Als Hauptergebnis einer feineren Analyse ergibt sich nun, daß g nur die Werte $1, 2, 3, 4, 6$ annehmen kann und daß auch die Zahlen m_+, m_- gewissen Einschränkungen unterliegen.

Die Abschnitte und Formeln sind durchgängig über beide Teile numeriert; die Literatur findet man am Ende von Teil I. Die Einleitung von Teil I enthält darüber hinaus eine kurze Skizze des Inhalts der einzelnen Abschnitte beider Teile.

7. Der Kohomologiemodul $H^*(M)$

Im folgenden werden die Bedingungen I und II der Einleitung vorausgesetzt, wobei die Mannigfaltigkeiten M, M_k ($k = \pm 1$) zusammenhängend und kompakt sind und $\dim M_k = n - 1 - m_k$ mit $m_k \geqq 1$ gelte; die Herkunft dieser topologischen Situation von den isoparametrischen Hyperflächenfamilien spielt keine Rolle bei den Beweisen. Die Indizes $+1$ bzw. -1 werden weiterhin oft abkürzend als $+$ bzw. $-$ notiert.

Wir berechnen die Kohomologiegruppen

$$H^q(M) = H^q(M, R) \quad \text{und} \quad H^q(M_k) = H^q(M_k, R)$$

mit Hilfe der Mayer-Vietoris-Sequenz und den Dualitätssätzen von Alexander und Poincaré, wobei über R die Konventionen der Einleitung erfüllt seien. Anschließend benutzen wir den Satz von Leray-Hirsch zur Beschreibung der Cup-Struktur von $H^*(M)$.

Die Inklusionen $i_k : M_k \to B_k$, $j_k : M \to B_k$ und die Bündelprojektionen $q_k : B_k \to M_k$, $p_k : M \to M_k$ genügen den Beziehungen

$$q_k \circ i_k = 1_{M_k}, \qquad q_k \circ j_k = p_k \quad (k = \pm 1). \tag{18}$$

Da M_k Deformationsretrakt von B_k und $B_k \backslash M$ ist, können wir die entsprechenden Kohomologiegruppen identifizieren

$$H^q(B_k) = H^q(M_k) = H^q(B_k \backslash M). \tag{19}$$

Sei jetzt $1 \leqq q \leqq n-2$. Die Mayer-Vietoris-Sequenz, angewandt auf die exakte Triade $(S^n; B_+, B_-)$, liefert wegen $H^q(S^n) = H^{q+1}(S^n) = 0$

$$H^q(M) \cong H^q(B_+) \oplus H^q(B_-) = H^q(M_+) \oplus H^q(M_-); \tag{20}$$

speziell sind die Homomorphismen

$$j_k^*: H^q(B_k) \to H^q(M), \quad p_k^*: H^q(M_k) \to H^q(M)$$

injektiv.

Der Alexandersche Dualitätssatz ergibt für $1 \leqq q \leqq n-2$

$$H^q(M_k) = H^q(B_k) \cong H_{n-1-q}(B_{-k} \backslash M) = H_{n-1-q}(M_{-k}). \tag{21}$$

Unsere Konvention über den Koeffizientenring erlaubt die Anwendung des Poincaréschen Dualitätssatzes auf die $(n-1-m_{-k})$-dimensionale Mannigfaltigkeit M_{-k}; wir erhalten aus (21)

$$H^q(M_k) \cong H_{n-1-q}(M_{-k}) \cong H^{q-m-k}(M_{-k}). \tag{22}$$

Durch Iteration dieser Formel gewinnen wir

$$H^q(M_k) \cong H^{q-m_+-m_-}(M_k) \quad \text{für} \quad 1+m_{-k} \leqq q \leqq n-2. \tag{23}$$

Wir fassen die Formeln (20), (21) und (23) zu einem Satz zusammen. Dazu sei $\mu = m_+ + m_-$.

Satz 5. *Unter den genannten Voraussetzungen ist $\gamma = 2(n-1)/\mu$ eine ganze Zahl und es gilt für $k = \pm 1$*

$$H^q(M_k) \cong \begin{cases} R \text{ für } q \equiv 0(\mu), & 0 \leqq q < n-1, \\ R \text{ für } q \equiv m_{-k}(\mu), & 0 \leqq q < n-1, \\ 0 \text{ sonst} \end{cases}$$

sowie

$$H^q(M) \cong \begin{cases} R \text{ für } q = 0, \quad n-1, \\ H^q(M_+) \oplus H^q(M_-) \text{ für } 1 \leqq q \leqq n-2. \end{cases}$$

Zusatz. *Offensichtlich gilt $\gamma = \frac{1}{2}\dim_R H^*(M)$. Wird die Zerlegung der Sphäre S^n in Ballbündel von einer isoparametrischen Hyperfläche erzeugt, so ist $\gamma = g$ die Anzahl ihrer Hauptkrümmungswerte.*

Beweis von Satz 5. Da M_k zusammenhängend ist, gilt $H^0(M_k) \cong R$. Die Formel (21) in Verbindung mit der Konvention über R liefert

$$H^{m-k}(M_k) \cong H_{\dim M_{-k}}(M_{-k}) \cong R, \quad H^q(M_k) = 0 \quad \text{für} \quad 0 < q < m_{-k}.$$

Die Formel (23) impliziert zunächst

$$H^q(M_k) = 0 \quad \text{für} \quad m_{-k} < q < \mu$$

und weiterhin die im Satz 5 angegebene Kohomologiestruktur von M_k. Die Gestalt der Gruppen $H^q(M)$ folgt aus (20) und dem Zusammenhang von M.

Zum Beweis der Ganzzahligkeit von γ beachten wir, daß nach dem Bewiesenen

$$\dim M_k \equiv 0(\mu) \quad \text{oder} \quad \dim M_k \equiv m_{-k}(\mu)$$

gelten muß. Also gibt es folgende Fälle:

Fall 1. $\dim M_k \equiv 0(\mu)$ für $k=+1$ und $k=-1$. Dann gilt wegen

$$2(n-1)-\dim M_+ - \dim M_- = \mu,$$

daß γ eine ganze Zahl ist. (Außerdem folgt aus

$$|\dim M_+ - \dim M_-| = |m_- - m_+| < \mu,$$

daß γ ungerade ist.)

Fall 2. O.B.d.A. gelte $\dim M_+ \equiv m_-(\mu)$. Dann gilt $n-1=\dim M_+ + m_+ \equiv 0(\mu)$, und γ ist eine gerade Zahl.

Zugunsten einer einheitlichen Notation wird im folgenden für $2(n-1)/\mu$ der Buchstabe g benutzt.

Die Struktur des Kohomologierings $H^*(M)$ ergibt sich aus dem Satz von Leray-Hirsch, angewandt auf die Sphärenbündel $p_k: M \to M_k$ (vgl. [5], S. 229). Sei $a_k \in H^{m-k}(M)$ $(k=\pm 1)$ je ein Bild eines Erzeugenden von $H^{m-k}(M_k)=R$ unter p_k^*; nach (20) bzw. Satz 5 ist $\{a_+, a_-\}$ ein R-Erzeugendensystem von $H^{m+}(M)$, falls $m_+ = m_-$, und a_k ein R-Erzeugendes von $H^{m-k}(M)$, falls $m_+ \neq m_-$. Ferner sei e das Einselement des Ringes $H^*(M)$ und $G^k = p_k^*(H^*(M_k))$; G^k ist ein Unterring von $H^*(M)$, der isomorph zu $H^*(M_k)$ ist.

Satz 6. *$H^*(M)$, betrachtet als G^k-Modul, ist frei mit der Basis $\{e, a_{-k}\}$ $(k=\pm 1)$.*

Beweis. O.B.d.A. beweisen wir den Satz für $k=+1$. Seien Σ und D die Fasern der Bündel (M, p_+, M_+) und (B_+, q_+, M_+) im Punkt $q \in M_+$, ferner seien $l: \Sigma \to M$ und $\bar{l}: (D, \Sigma) \to (B_+, M)$ die zugehörigen Einbettungen. (D, Σ) hat die topologische Struktur eines berandeten $(m_+ + 1)$-dimensionalen Balles. Satz 6 ist eine unmittelbare Folge des Satzes von Leray-Hirsch, wenn die Bedingung

$$\{l^*(e), l^*(a_-)\} \text{ bilden eine } R\text{-Basis von } H^*(\Sigma) \tag{24}$$

verifiziert ist. Dazu muß gezeigt werden, daß $l^*(a_-)H^{m_+}(\Sigma)$ erzeugt. Nun sind S^n und M_+ und damit das Bündel (B_+, q_+, M_+) R-orientierbar (da es isomorph zu einem Normalenballbündel ist), d. h.

$$\bar{l}^*: H^{m_+ + 1}(B_+, M) \to H^{m_+ + 1}(D, \Sigma)$$

ist ein Isomorphismus. Im kommutativen Diagramm

$$\begin{array}{ccccccccc}
H^{m_+}(B_+) & \xrightarrow{j^*} & H^{m_+}(M) & \xrightarrow{\delta} & H^{m_+ + 1}(B_+, M) & \longrightarrow & H^{m_+ + 1}(B_+) & \xrightarrow{j^*} & H^{m_+ + 1}(M) \\
 & & \downarrow{\scriptstyle l^*} & & \downarrow{\scriptstyle \bar{l}^*} & & & & \\
 & \longrightarrow & H^{m_+}(\Sigma) & \cong & H^{m_+ + 1}(D, \Sigma) & \longrightarrow & & &
\end{array}$$

gehören die Zeilen zu exakten Kohomologiesequenzen. Da nach (20) j_+^* wie p_+^* injektiv ist, ist δ und also mit $\bar{l}^*$ auch l^* surjektiv. Da nach (18)

$$a_+ \in \operatorname{Im} p_+^* = \operatorname{Im} j_+^* \subset \operatorname{Ker} \delta$$

gilt, ist (auch im Falle $m_+ = m_-$) $l^*(a_-)$ ein R-Erzeugendes von $H^{m_+}(\Sigma)$, und damit gilt (24).

Übrigens kann Satz 5 auch aus Satz 6 hergeleitet werden; vgl. hierzu Wang [15].

8. Erzeugende und Relationen in $H^*(M)$

Um Fallunterscheidungen zu vermeiden, wird im folgenden die Graduierung von $H^*(M)$ so verzerrt, daß m_+ und m_- gleichsam eliminiert werden. Wir definieren:

Eine assoziative graduierte Algebra $G = \sum_{i=0}^{\infty} G_i$ mit 1 über R heiße R-Algebra vom Typ (*), wenn die Bedingungen **A–D** erfüllt sind:

A. Es gibt eine natürliche Zahl $g \geqq 2$ so, daß

$$G_i = \begin{cases} R \text{ für } i = 0, g \\ 0 \text{ für } i > g \\ G_i^+ \oplus G_i^- \text{ für } 1 \leqq i \leqq g-1, \end{cases}$$

wobei $G_i^k = R$ für $k = \pm 1$, $1 \leqq i \leqq g-1$.

(Wir setzen wieder $G_i^+ = G_i^1, \ldots$.)

B. $G^k = G_0 + \sum_{i=1}^{g-1} G_i^k$ ist eine Unteralgebra von G.

C. Es sei x_1 (bzw. y_1) ein R-Erzeugendes von G_1^+ (bzw. G_1^-). Dann ist G ein freier G^+-Modul (bzw. G^--Modul) mit der Basis $\{1, y_1\}$ (bzw. $\{1, x_1\}$).

D. Für $u \in G_i$, $w \in G_j$ gilt $uw = \pm wu \in G_{i+j}$.

Für R verwenden wir die Ringe $\mathbb{Z}$ oder $\mathbb{Z}_2$. Nach den Sätzen 5 und 6 ist evident, daß $G = H^*(M)$ eine R-Algebra vom Typ (*) ist; man definiere $G_0 = H^0(M)$, $G_g = H^{n-1}(M)$ und

$$G_{2j+\varepsilon}^k = p_k^*(H^{j\mu+\varepsilon m-\kappa}(M_k)) \quad (k = \pm 1, \varepsilon = 0, 1; 1 \leqq 2j+\varepsilon \leqq g-1, j \in \mathbb{N}).$$

Man beachte, daß dadurch im Falle $m_+ \neq m_-$ zwar

$$G_{2j+1} = H^{j\mu+m_+}(M) \oplus H^{j\mu+m_-}(M)$$

gilt, doch die Graduierungseigenschaften von Addition und Multiplikation dennoch bewahrt werden. Die exakte Form der Kommutativitätseigenschaft **D** im Falle $R = \mathbb{Z}$ hängt natürlich von m_+ und m_- ab.

Wir wählen jetzt in einer kanonisierten Weise Erzeugende x_i (bzw. y_i) von G_i^+ (bzw. G_i^-) und z von G_g:

Hilfssatz 3. *Die x_i, y_i, z lassen sich so wählen, daß für $1 \leqq i \leqq g-2$ gilt*

$$\begin{aligned} x_{i+1} &= \alpha_{i+1} y_{i+1} + y_i x_1 , \\ y_{i+1} &= \beta_{i+1} x_{i+1} + x_i y_1 , \\ z &= x_{g-1} y_1 = \pm y_{g-1} x_1 \end{aligned} \tag{25}$$

mit $\alpha_i, \beta_i \in R$. Nach der Wahl von x_1, y_1 sind die $x_i, y_i, \alpha_i, \beta_i$ durch (25) eindeutig bestimmt.

Beweis. Für beliebige Erzeugende $\bar{x}_{i+1}$, $\bar{y}_{i+1}$, $\bar{x}_i$, $\bar{y}_i$ hat man wegen **A**

$$\bar{y}_i x_1 = \alpha \bar{x}_{i+1} + \beta \bar{y}_{i+1}$$

und wegen **C**

$$\bar{x}_{i+1} = \gamma \bar{y}_{i+1} + \delta \bar{y}_i x_1$$

mit $\alpha, \beta, \gamma, \delta \in R$. Es folgt durch Elimination von $\bar{y}_i x_1$ und Koeffizientenvergleich nach **A**, daß $\delta = \alpha^{-1}$ eine Einheit in R ist. Dies und eine entsprechende Überlegung für $\bar{y}_{i+1}$ zeigen, daß sich ausgehend von x_1 und y_1 die x_i, y_i und z induktiv so bestimmen lassen, daß (25) gilt.

Die Voraussetzung **B** liefert Darstellungen der Form

$$x_i x_1 = a_{i+1} x_{i+1} , \quad y_i y_1 = b_{i+1} y_{i+1} \quad \text{für} \quad i = 1, \ldots, g-2 , \tag{26}$$

und allgemeiner für geeignete s, t:

$$x_s x_t = a_{s,t} x_{s+t} , \quad y_s y_t = b_{s,t} y_{s+t} . \tag{26'}$$

Ferner setzen wir $a_1 = b_1 = 1$, $a_g = b_g = 0$ und

$$x_1 y_1 = \sigma y_1 x_1 . \tag{27}$$

Die Koeffizienten α_i, β_i, a_i, $b_i \in R$ und $\sigma \in \{1, -1\}$ genügen gewissen Relationen:

Hilfssatz 4. *Werden die R-Erzeugenden von G gemäß (25) gewählt, so gilt für $1 \leqq i \leqq g-2$*

$$\alpha_{i+1} = -\sigma a_i , \quad \beta_{i+1} = -\sigma b_i , \tag{28}$$

$$\begin{aligned} \sigma b_i a_{i+2} &= a_{i+1} b_{i+1} - 1 , \\ \sigma a_i b_{i+2} &= a_{i+1} b_{i+1} - 1 . \end{aligned} \tag{29}$$

Beweis. Sei $g > 2$. Für $i = 1$ erhält man aus (25) durch Elimination von $x_1 y_1$ die Beziehung

$$(\sigma + \alpha_2) y_2 = (\sigma \beta_2 + 1) x_2 \in G_2^+ \cap G_2^- = \{0\}$$

und hieraus (28).

Für $1 \leqq i \leqq g-3$ ersetzen wir in der Identität (25)

$$x_{i+2} = \alpha_{i+2} y_{i+2} + y_{i+1} x_1$$

rechts y_{i+1} gemäß (25) durch $\beta_{i+1}x_{i+1}+x_iy_1$ und erhalten bei Beachtung von (26) und (27)

$$(1-\beta_{i+1}a_{i+2})x_{i+2}-\alpha_{i+2}y_{i+2}=\sigma a_{i+1}x_{i+1}y_1\,.$$

Der Vergleich mit der Identität (25)

$$-\beta_{i+2}x_{i+2}+y_{i+2}=x_{i+1}y_1$$

liefert bei Beachtung von $G^+_{i+2}\cap G^-_{i+2}=\{0\}$ die Beziehungen

$$1-\beta_{i+1}a_{i+2}=-\sigma a_{i+1}\beta_{i+2}\,,\qquad \alpha_{i+2}=-\sigma a_{i+1}$$

und aus Symmetriegründen

$$1-\alpha_{i+1}b_{i+2}=-\sigma b_{i+1}\alpha_{i+2}\,,\qquad \beta_{i+2}=-\sigma b_{i+1}\,.$$

Die jeweils zweiten Identitäten beweisen (28) für $2\leqq i\leqq g-2$ und die jeweils ersten Identitäten gehen mit (28) über in (29) für $1\leqq i\leqq g-3$.

Schließlich gilt wegen **B** $x_{g-1}x_1=0=y_{g-1}y_1$, damit wegen (25)

$$\begin{aligned}z=x_{g-1}y_1&=\alpha_{g-1}y_{g-1}y_1+y_{g-2}x_1y_1\\&=\sigma b_{g-1}y_{g-1}x_1\\&=\sigma^2 b_{g-1}a_{g-1}x_{g-1}y_1\,,\end{aligned}$$

also $1=b_{g-1}a_{g-1}$ und damit (29) für $i=g-2$.

Multipliziert man die Identitäten (29) miteinander, so erhält man für die Zahlen $f_i=a_ib_i$ die Rekursionsformel

$$f_{i-1}f_{i+1}=(f_i-1)^2\quad\text{für}\quad i=2,\ldots,g-1\,. \tag{30}$$

Hilfssatz 5. *Für $\mathbb{Z}$-Algebren vom Typ (*) kann $f_2=a_2b_2$ nur die Werte $0,1,2,3$ annehmen. Die zugehörigen eindeutig bestimmten Stücke von Lösungen der Rekursionsformel* (30) *mit $f_1=1$ sind*

f_1	f_2	f_3	f_4	f_5	f_6	f_7
1	0	1				
1	1	0	1			
1	2	1	0	1		
1	3	4	3	1	0	1

Die entsprechenden Möglichkeiten für $\mathbb{Z}_2$-Algebren sind durch die ersten beiden Zeilen der Tabelle gegeben.

Beweis. Wir zeigen, daß die Lösungsfolgen (e_i) von (30) mit $e_1=1$ und $e_2\leqq -1$ (bzw. $e_2\geqq 4$) die Eigenschaft haben, daß die Folge der Beträge $(|e_i|)$ von $i=2$ an streng monoton wachsend sind. Damit kommen sie für $\mathbb{Z}$-Algebren nicht in Frage, weil für diese $f_g=0$ sein muß.

In der Tat gilt für $e_2 \leqq -1$

$$|e_3| = (|e_2|+1)^2 \geqq (\sqrt{|e_2|}+1)^2$$

und für $e_2 \geqq 4$

$$\sqrt{|e_3|} = e_2 - 1 = (\sqrt{e_2}+1)(\sqrt{e_2}-1) \geqq \sqrt{|e_2|}+1 .$$

Tritt aber für ein $k \geqq 2$ die Konstellation $e_{k-1} \neq 0$, $\sqrt{|e_k|} \geqq \sqrt{|e_{k-1}|}+1$ auf, so folgt aus (30)

$$|e_{k+1}| \geqq (|e_k|-1)^2/|e_{k-1}| \geqq (\sqrt{|e_k|}+1)^2 ,$$

d. h. die Folge $|e_k|, |e_{k+1}|, \ldots$ ist streng monoton wachsend.

Die Formel (30) ist symmetrisch in $i+1$ und $i-1$. Dies ist Ausdruck allgemeinerer Dualitätsrelationen:

Hilfssatz 6. *Es gilt für* $i = 1, \ldots, g-1$:

a) $x_i y_{g-i} = \pm z$;

b) $x_i x_{g-1-i} = \pm x_{g-1}$, $y_i y_{g-1-i} = \pm y_{g-1}$;

c) $a_i = \pm a_{g-i}$, $b_i = \pm b_{g-i}$.

d) *Wenn G kommutativ und g gerade ist, gilt in* a) *und* c) *das positive Vorzeichen.*

Beweis zu a). Nach **B** gilt $y_{g-i} y_i = x_{g-i} x_i = 0$. Man erhält daher durch sukzessive Umwandlungen mit Hilfe der Identitäten (25) und der Kommutativitätsbedingung **D**:

$$z = x_{g-1} y_1 = y_{g-2} x_1 y_1 = \pm y_{g-2} x_2 = \pm x_{g-3} y_1 x_2 = \pm x_{g-3} y_3 \cdots$$

$$\pm z = y_{g-1} x_1 = \pm x_{g-2} y_2 = \ldots$$

(die zweite Zeile wird nur benötigt, wenn g ungerade ist).

Beweis zu b). Man hat

$$\pm z = y_i x_{g-i} = y_i y_{g-i-1} x_1 , \qquad \pm z = y_{g-1} x_1 \neq 0$$

und hieraus $y_i y_{g-i-1} = \pm y_{g-1}$ wegen **B**. Analog für die x_i.

Beweis zu c). Es gilt wegen (26)

$$a_i x_{g-1} = \pm a_i x_i x_{g-1-i} = \pm x_{i-1} x_1 x_{g-1-i} = \pm a_{g-i} x_{g-1}$$

und entsprechend für die y_i.

Da die Vorzeichenunbestimmtheit allein durch Vertauschungen entsteht und für gerades g beim Beweis von a) nur die erste Zeile benötigt wird, gilt d).

Für die Algebra G definieren wir eine Invariante $s = s(G)$ als die kleinste natürliche Zahl $\geqq 2$, für welche $f_s = 0$ gilt. Nach Hilfssatz 5 haben wir für $R = \mathbb{Z}$ die Tabelle

f_2	0	1	2	3
s	2	3	4	6 ,

(31)

deren ersten beiden Spalten den Fall $R=\mathbb{Z}_2$ erschöpfen. Gelänge der Nachweis $s=g$, so wäre das Ziel der Arbeit $g\in\{1,2,3,4,6\}$ erreicht. Tatsächlich gilt $s=g$ nicht einmal für die R-Algebren vom Typ (*), die von der globalen Geometrie isoparametrischer Familien stammen. Jedoch gilt:

Hilfssatz 7. *Es gilt:*

a) *Aus* $a_2=b_2=0$ *folgt* $a_{2j}=b_{2j}=0$ *für* $2\leqq 2j\leqq g$.

b) $f_{s+i}=f_i$ *für* $1\leqq i\leqq g-s$ *und* $s\in\{2,3,4,6\}$.

c) s *ist ein Teiler von* g.

Beweis. Der Beweis wird induktiv geführt. Aus der Situation $0<js<g-1$, $f_{js}=0$, $f_{js+1}=1$ werden wir $f_{js+2}=f_2$ folgern, und die Rekursionsformel (30) liefert $f_{js+i}=f_i$ für $i=2,\ldots,\min\{s+1,g-js\}$.

1. Fall. $s=2$, also $f_2=0$. Aus

$$a_{js+1}a_{js+2}x_{js+2}=x_{js}x_1^2=a_{js,2}a_2x_{js+2}$$

und der entsprechenden Gleichung mit y_{js+2} folgt wegen $f_{js+1}=a_{js+1}b_{js+1}=1$ die Aussage a) und ferner

$$f_{js+2}=f_{js+1}f_{js+2}=a_{js,2}b_{js,2}f_2=0.$$

2. Fall. $s\geqq 3$. Dann ergänzen wir die Induktionsvoraussetzung durch die Annahme $f_{js-2}\neq 0$. Aus (29) folgt

$$\sigma a_{js-2}b_{js}=f_{js-1}-1=0,$$
$$\sigma b_{js-2}a_{js}=f_{js-1}-1=0,$$

und damit wegen (28)

$$\alpha_{js+1}=-\sigma a_{js}=0,\qquad \beta_{js+1}=0.$$

Die Anwendung von (25) ergibt

$$x_{js+1}=y_{js}x_1,\qquad y_{js+1}=x_{js}y_1$$

und damit

$$\begin{aligned}a_{js+2}x_{js+2}&=x_{js+1}x_1=y_{js}x_1x_1=a_2y_{js}x_2\\&\equiv a_2y_{js}y_1x_1\equiv a_2b_{js+1}y_{js+1}x_1\equiv a_2b_{js+1}x_{js+2}\bmod G^-.\end{aligned}$$

Es folgt $a_{js+2}=a_2b_{js+1}$ und das Analogon durch Vertauschen von a und b, und hieraus

$$f_{js+2}=f_{js+1}f_2=f_2,$$

wie zu beweisen war.

Um die Beziehung zwischen g und s aufzudecken, ordnen wir G die graduierte Algebra πG, definiert durch $\pi G=\sum(\pi G)_j$ mit $(\pi G)_j=G_{js}$ zu; πG hat die Unteralgebren $\pi G^k=\pi G\cap G^k$.

Hilfssatz 8. *Es sei* $s=2$ *oder* $s=3$ *und* $g>s$. *Dann gilt*

a) πG *ist eine R-Algebra vom Typ* (*).

b) $x_s^2=y_s^2=x_sy_s+y_sx_s=0$, *also* $s(\pi G)=2$.

Ferner gilt für beliebiges $s\in\{2,3,4,6\}$:

c) $g=2^r s$ *mit einer natürlichen Zahl* r.

Beweis von a). Da s ein Teiler von g ist (Hilfssatz 7), treffen die Eigenschaften **A**, **B** und **D** einer R-Algebra vom Typ (*) auf πG zu. Die Eigenschaft **C** ist gleichwertig zu den Kongruenzen

$$x_{(j+1)s}\equiv\pm y_{js}x_s \bmod G^-,\qquad y_{(j+1)s}\equiv\pm x_{js}y_s \bmod G^+, \tag{32}$$

$$z=\pm x_{g-s}y_s=\pm y_{g-s}x_s. \tag{33}$$

(33) folgt aus Hilfssatz 6. Zum Beweis von (32) unterscheiden wir 2 Fälle:

1. Fall. $s=2$. Nach Hilfssatz 7 gilt $f_{2j+1}=1$, also $a_{2j+1}=b_{2j+1}=\pm1$. Durch Umformungen gemäß (25) und (26) erhält man

$$\pm x_{2j+2}\equiv b_{2j+1}x_{2j+2}\equiv b_{2j+1}y_{2j+1}x_1\equiv y_{2j}y_1x_1\equiv y_{2j}x_2 \bmod G^-$$

und eine entsprechende Kongruenz für $x_{2j}y_2$, womit (32) gezeigt ist $(j=1,\ldots,g/2-2)$.

2. Fall. $s=3$. Dann ist nach den Hilfssätzen 5 und 7 $f_{3j+1}=f_{3j+2}=1$, und man hat die Kongruenzenkette:

$$\begin{aligned}\pm x_{3j+3}&\equiv b_{3j+1}b_{3j+2}x_{3j+3}\equiv b_{3j+1}b_{3j+2}y_{3j+2}x_1\\&\equiv y_{3j}y_1y_1x_1\equiv b_2y_{3j}y_2x_1\equiv\pm y_{3j}x_3 \bmod G^-\end{aligned}$$

und die entsprechende $(j=1,\ldots,g/3-2)$.

Beweis zu b). Aus $f_{s-1}=1$ und (28) folgt $\alpha_s=\beta_s=\alpha=\pm1$, und (25) impliziert

$$x_{s-1}y_1=y_s-\alpha x_s=-\alpha(x_s-\alpha y_s)=-\alpha y_{s-1}x_1.$$

Es folgt wegen $f_s=0$:

$$y_s^2+x_s^2-\alpha(y_sx_s+x_sy_s)=\pm x_{s-1}x_1y_{s-1}y_1=\pm f_sx_sy_s=0. \tag{34}$$

Im Fall $R=\mathbb{Z}_2$ gilt $y_sx_s+x_sy_s=0$, und b) folgt durch Trennung der G^k-Bestandteile.

Im Fall $R=\mathbb{Z}$ führen wir die Annahme $x_sy_s+y_sx_s\neq0$, zum Widerspruch. Denn in Verbindung mit (34) und (32) liefert sie

$$x_s^2\equiv x_s^2+y_s^2\equiv2\alpha x_sy_s\equiv\pm2x_{2s} \bmod G^-$$

und eine entsprechende Relation für y_s^2.

Für die $\mathbb{Z}$-Algebra vom Typ (*)$\bar G=\pi G$ würde folgen [vgl. (26′)]

$$\bar f_2=a_{s,s}b_{s,s}=\pm4,$$

und Hilfssatz 5 schließt diese Möglichkeit aus.

Beweis zu c). Für $s=2,3$ erhalten wir durch iteratives Anwenden von π auf G mit einem geeigneten $r\in\mathbb{N}$ eine Algebra π^rG mit $g(\pi^rG)=s(\pi^rG)$ und aus b) folgt dann

$g=2^r s$. Da wir jeder $\mathbb{Z}$-Algebra G vom Typ (*) die $\mathbb{Z}_2$-Algebra $\hat{G}=G\otimes\mathbb{Z}_2$ zuordnen können, für die $s(\hat{G})\in\{2,3\}$, $g(\hat{G})=g(G)$ und $s(\hat{G})=s(G)$ oder $2s(\hat{G})=s(G)$ gilt, gilt c) allgemein.

9. Der orientierbare Fall

Wir kehren zur geometrischen Situation von Abschn. 7 zurück und beweisen

$$g=\dim_{\mathbb{Z}} H^*(M,Z)/2\in\{1,2,3,4,6\}$$

sowie weitere Eigenschaften des Tripels (g,m_+,m_-) unter der Voraussetzung, daß M_+ und M_- orientierbar sind. Wir geben damit Teilbeweise der in Abschn. 0 von Teil I angeführten Sätze A, B, C.

Der Fall $g=1$ ist durch Satz 5 (Abschn. 7) erledigt; M_+ und M_- sind 0-dimensional, also Punkte und M ist eine topologische Sphäre.

Sei also $g\geqq 2$ und M_+, M_- seien orientierbar. Wir interpretieren $H^*(M,\mathbb{Z})$ gemäß Abschn. 8 als $\mathbb{Z}$-Algebra G vom Typ (*). Die Regel $uv=(-1)^{ij}vu$ für $u\in H^i(M)$, $v\in H^j(M)$ entscheidet über die Vertauschungsregel **D** in Abschn. 8 und damit über σ in (27) und s in (31). Es ergeben sich offensichtlich die folgenden Fälle:

α) G ist kommutativ und $s\in\{2,3,4,6\}$. Dieser Fall liegt vor, wenn m_+ und m_- gerade sind.

β) $s=2$, G ist nicht kommutativ, aber πG ist kommutativ. Dieser Fall liegt vor, wenn m_+ und m_- ungerade sind.

γ) $s=2$, πG ist nicht kommutativ, aber $\pi^2 G$ ist kommutativ. Dieser Fall liegt vor, wenn z. B. m_+ ungerade und m_- gerade ist.

Nun gilt:

Satz 7. *Es sei G eine $\mathbb{Z}$-Algebra vom Typ (*). Man hat dann im Fall α) $g=s\in\{2,3,4,6\}$, im Fall β) $g=2$ und im Fall γ) $g=2$ oder $g=4$.*

Beweis. Nach Hilfssatz 8 gilt $g=2^r s$ mit einer natürlichen Zahl r.

Fall α). Wir nehmen an, daß $g\neq s$, also $r\geqq 1$ gilt, und folgern einen Vorzeichenwiderspruch. Für $i=g/2=2^{r-1}s$ gilt nach Hilfssatz 6c) und d) die Beziehung $a_{i-1}=a_{i+1}$ und nach Hilfssatz 7 $f_i=0$, $f_{i+1}=f_{i-1}=1$. Berechnen wir daher a_{i+1} nach Formel (29), so folgt wegen $\sigma=1$

$$a_{i-1}=a_{i+1}=(f_i-1)/b_{i-1}=-a_{i-1}/f_{i-1}=-a_{i-1}\neq 0,$$

und dies ist ein Widerspruch.

Fall β). Wir nehmen $g>s=2$ an. Dann sind x_2, y_2 die Erzeugenden von $(\pi G)_1$. Es folgt $x_2y_2\neq 0$ und wegen der Kommutativität von πG

$$x_2y_2+y_2x_2=2x_2y_2\neq 0,$$

im Widerspruch zu Hilfssatz 8b).

Fall γ). Für $g>2$ wenden wir den Fall β) auf die $\mathbb{Z}$-Algebra von Typ (*) πG an.

Im Hinblick auf die Fallunterscheidungen von Satz C in Abschn. 0 [16] heben wir die folgende topologische Konsequenz hervor:

Folgerung aus Satz 7. *Seien* M, M_k, m_k $(k=\pm 1)$ *wie in Abschn. 7 definiert und es gelte*

$$g=2(n-1)/(m_+ + m_-) \geqq 3.$$

Dann gilt:

a) *Sind beide* m_k *gerade, so sind die* M_k *orientierbar und es gilt* $m_+ = m_-$.

b) *Sind* M_+ *und* M_- *orientierbar und ist* m_+ *ungerade, so gilt* $g=4$ *und* m_- *ist gerade.*

Beweis. Teil b) ergibt sich aus der Kombination der Fälle β) und γ). Für Teil a) benötigen wir

Hilfssatz 9. *Ist* M_k *nicht orientierbar, so gilt* $m_{-k}=1$.

In der Tat, wenn M_k nicht orientierbar ist, gilt $H^1(M_k, \mathbb{Z}_2) \neq 0$ und damit wegen Satz 5 $m_{-k}=1$.

a) folgt jetzt aus Fall α), da $m_+ \neq m_-$ für die zugehörige Algebra $s=2$ zur Folge hat.

10. Struktursätze

Im verbleibenden nichtorientierten Fall sind wir auf Kohomologie mit Koeffizienten in $\mathbb{Z}_2$ angewiesen, und damit entfallen zunächst die bisher benutzten erfolgreichen Vorzeichenargumente. Es gibt dementsprechend auch unendlich viele Isomorphieklassen für $\mathbb{Z}_2$-Algebren vom Typ (*). Sie werden in diesem Paragraphen klassifiziert (Satz 8), ebenso die auf ihnen möglichen Derivationen eines bestimmten Typs (Hilfssatz 11). In Abschn. 11 werden dann diese Derivationen benutzt, um über die Bocksteinsequenz den Homomorphismus $H^*(M, \mathbb{Z}_4) \to H^*(M, \mathbb{Z}_2)$ zu beschreiben und dadurch Vorzeichenargumente einzubringen.

Für die $\mathbb{Z}_2$-Algebra G vom Typ (*) bezeichnen wir mit ϱG den von $\{1, x_1, y_1\}$ erzeugten Unterring und mit $\bar{s}(G)$ den maximalen Grad von Elementen aus ϱG. Wir untersuchen die Rekonstruktion von G aus ϱG durch sukzessive Adjunktionen.

Dazu sei $\tau G = G[x_g]$ gesetzt, wobei die Unbestimmte x_g der einzigen Relation $x_g^2 = 0$ genügen möge. Die Graduierung von G sei so auf τG fortgesetzt, daß x_g den Grad $g=g(G)$ erhält. Dann gilt:

Hilfssatz 10. *τG ist eine $\mathbb{Z}_2$-Algebra vom Typ (*) mit $g(\tau G)=2g$.*

Beweis. Sei z_g das Erzeugende von G_g; wir setzen

$$\begin{aligned} &y_g = x_g + z_g, \quad z_{2g} = x_g y_g, \\ &x_{g+i} = x_g x_i, \quad y_{g+i} = x_g y_i \quad \text{für} \quad 1 \leqq i \leqq g-1, \\ &\tau G^+ = G_0 + \sum_1^{2g-1} \mathbb{Z}_2 x_i, \quad \tau G^- = G_0 + \sum_1^{2g-1} \mathbb{Z}_2 y_i. \end{aligned} \tag{35}$$

Die Eigenschaften **A** und **D** einer Algebra vom Typ (*) sind evident, ebenso, daß $(\tau G)^+$ Unteralgebra von τG ist. Wegen $y_g^2 = x_g^2 + z_g^2 + 2x_g z_g = 0$ ist auch $(\tau G)^-$ eine Unteralgebra. Also gilt **B**. Schließlich übertragen sich die Relationen (25) vermittels (35) von G auf τG, und damit ist auch **C** erfüllt.

Es gilt nun:

Satz 8. *Es sei G eine $\mathbb{Z}_2$-Algebra vom Typ (*). Dann gilt*

a) *ϱG ist $\mathbb{Z}_2$-Algebra vom Typ (*) mit $g(\varrho G) = \bar{s} \in \{2, 3, 4\}$ und ist isomorph zu $\mathbb{Z}_2[x, y]/I_{\bar{s}}$. Die Ideale $I_{\bar{s}}$ haben die Gestalt $I_2 = (x^2, y^2)$, $I_3 = (x^2 + y^2 + xy, y^3)$ $I_4 = (x^4, y^2)$ (bzw. $I_4 = (x^2, y^4)$).*

b) *G ist isomorph zu $\tau^r(\varrho G)$ für eine geeignete natürliche Zahl r.*

Beweis zu a). Wir begründen die Fallunterscheidung und beschränken uns im übrigen auf die Angabe einer $\mathbb{Z}_2$-Basis von ϱG aus Polynomen in x_1, y_1.

$I_{\bar{s}}$ sei der Kern des Homomorphismus $\varphi : \mathbb{Z}_2[x, y] \to \varrho G$ mit $\varphi(x) = x_1$, $\varphi(y) = y_1$. Es gilt $f_s = a_s b_s = 0$; o.B.d.A. gelte $b_s = 0$ und damit:

$$y_{s-1} y_1 = 0. \tag{36}$$

Schließlich setzen wir $z_i = x_i + y_i$ für $i = 1, \ldots, g-1$.

1. Fall. $s = 2$ und $x_1^2 = 0$. Wegen (36) gilt $y_1^2 = 0$ und damit $\bar{s} = 2$. Die $\mathbb{Z}_2$-Basis von ϱG ist

$$1, x_1, y_1, z_2 = x_1 y_1. \tag{37}$$

2. Fall. $s = 2$ und $x_1^2 \neq 0$. Dann gilt $x_1^3 = x_1 x_2 = x_3 \neq 0$, $x_1^4 = x_2^2 = 0$, $y_1^2 = 0$ und damit $\bar{s} = 4$. Die Basis ist

$$\begin{aligned} &1; x_1; y_1; x_i = x_1^i \quad \text{für} \quad i = 2, 3; \\ &y_2 = x_1^2 + x_1 y_1, y_3 = x_1^2 y_1; z_4 = x_1^3 y_1. \end{aligned} \tag{38}$$

3. Fall. $s = 3$. Aus (36) folgt $y_1^3 = 0$, aus (25) $x_1 y_1 = x_1^2 + y_1^2$ und hieraus $x_1^3 = 0$ und $x_1^2 y_1^2 = 0$, also $\bar{s} = 3$. Die Basis ist

$$\begin{aligned} &1, x_1, y_1, x_2 = x_1^2, y_2 = x_1^2 + x_1 y_1 = y_1^2, \\ &z_3 = x_1^2 y_1 = y_1^2 x_1. \end{aligned} \tag{39}$$

Daß sich in allen 3 Fällen eine $\mathbb{Z}_2$-Algebra vom Typ (*) ergibt und das zugehörige Ideal $I_{\bar{s}}$ die angegebene Gestalt hat, folgt leicht aus der Liste der Monome eines festen Grades $\leqq \bar{s}$ von $\mathbb{Z}_2[x, y]$.

Beweis von b). Die obige Fallunterscheidung zeigt, daß $\bar{s} = s$ oder $\bar{s} = 2s$. Sei also $g = 2^r \bar{s}$ und $J = \{2^l \bar{s} | l = 0, \ldots, r-1\}$. Indem wir Hilfssatz 8a) und b) auf πG, $\pi^2 G$ usw. anwenden, erhalten wir

$$x_j^2 = 0 \quad \text{für alle} \quad j \in J.$$

$\tau^r(\varrho G)$ besitzt die Gestalt $(\varrho G)[\{u_j | j \in J\}] = H$ wobei die u_j über ϱG nur der Bedingung $u_j^2 = 0$ genügen. Zu zeigen ist, daß der Ringhomomorphismus $\psi : H \to G$ mit $\psi(u_j) = x_j$ surjektiv ist, da dann wegen der gleichen $\mathbb{Z}_2$-Dimension von H und G die Isomorphieeigenschaft von ψ folgt.

In der Tat, wenden wir Hilfssatz 6b) auf die Algebren $\varrho G, G$ und $\pi^l G$ an, so erhalten wir Aufspaltungen der Form:

$$x_{\bar{s}-1} = x_q x_{\bar{s}-1-q} \quad \text{für} \quad 1 \leqq q < \bar{s}-1,$$

$$x_{g-1} = x_{\bar{s}-1} x_{g-\bar{s}} = x_{\bar{s}-1} x_{\bar{s}} x_{g-2\bar{s}} = \ldots = x_{\bar{s}-1} \prod_{j \in J} x_j .$$

Diese zeigen, daß sich das Element $x_i \in G_i^+$ $(i = 1, \ldots, g-1)$ darstellen läßt in der Form

$$x_i = x_q \prod_K x_j, \qquad q \in \{0, \ldots, \bar{s}-1\}, \qquad K \subset J, \tag{40}$$

und hieraus folgt $G^+ \subset \psi(H)$. Da $y_1 \in \psi(H)$ und der G^+-Modul G von $\{1, y_1\}$ erzeugt wird, folgt $G \subset \psi(H)$, wie zu beweisen war.

Man beachte, daß in $(\varrho G)[x_j | j \in J]$ die y_j von den z_j algebraisch unterschieden werden können:

$$y_{\bar{s}} = z_{\bar{s}} + x_{\bar{s}}, \qquad z_{2j} = x_j y_j = x_j z_j, \qquad y_{2j} = z_{2j} + x_{2j} (j \in J). \tag{41}$$

Wir betrachten jetzt Derivationen $\delta : G \to G$ [also $\mathbb{Z}_2$-lineare Abbildungen mit der Produktregel $\delta(uv) = (\delta u)v + u(\delta v)$], die überdies die Eigenschaften

$$\delta G_i \subset G_{i+1}, \delta G^k \subset G^k \quad (0 \leqq i \leqq g, k = \pm 1) \tag{42}$$

besitzen; offenbar ist $\delta G_0 = \delta G_g = 0$. Es gilt:

Satz 9. a) *δ ist durch das Wertepaar $(\delta x_1, \delta y_1)$ eindeutig bestimmt.*

b) *Für Derivationen von G mit den Eigenschaften* (42) *gibt es genau die folgenden nichttrivialen Möglichkeiten:*

α) $\bar{s} = 2, \quad g \geqq 4: \quad (\delta x_1, \delta y_1) \in G_2^+ \times G_2^- \setminus \{(0,0)\}$.

β) $\bar{s} = 4: \quad \delta x_1 = x_1^2, \quad \delta y_1 = y_1^2$.

γ) $\bar{s} = 3: \quad \delta x_1 = x_1^2, \quad \delta y_1 = y_1^2$.

Beweis von a). Durch $(\delta x_1, \delta y_1)$ ist $\delta | \varrho G$ wegen der Produktregel eindeutig bestimmt. Insbesondere ist auch $\delta z_{\bar{s}}$ und damit nach (42) auch $\delta x_{\bar{s}}$ (als der G^+-Bestandteil von $\delta z_{\bar{s}}$) eindeutig bestimmt. Die Behauptung folgt bei Benutzung von (41) durch vollständige Induktion nach $j \in J$.

Zum Beweis von Teil b) verzichten wir auf die ausführliche Begründung der Existenz gewisser Derivationen δ, da sie im folgenden nicht benutzt wird.

Fall α). Man verifiziert, daß es zu jedem Paar $(\varepsilon x_2, \eta y_2)$ mit $\varepsilon, \eta \in \{0, 1\}$ eine Derivation $\bar{\delta} : \varrho G[x_2] \to \varrho G[x_2]$ mit (42) so gibt, daß $(\delta x_1, \delta y_1) = (\varepsilon x_2, \eta y_2)$ gilt. Insbesondere gilt $\bar{\delta} z_4 = 0$ und man findet wegen (41) eine Fortsetzung δ auf $G = (\varrho G)[x_j | j \in J]$ mit $\delta x_j = 0$ für $j \in J$, $j \geqq 4$.

Fall β). O.B.d.A. sei $x_1^2 = x_2$, $y_1^2 = 0$ [also $I_4 = (x^4, y^2)$]. Die Annahme $\delta y_1 = y_2$ führt wegen (38) und (42) auf

$$\delta z_2 = \delta(x_1 y_1) = \varepsilon x_2 y_1 + x_1 y_2 \equiv x_3 \bmod G^- \qquad (\varepsilon \in \{0, 1\}),$$

und dies widerspricht der Beziehung

$$0 = \delta(x_1^2) = \delta x_2 \equiv \delta z_2 \bmod G^- .$$

Die Existenz eines δ mit $(\delta x_1, \delta y_1)=(x_2,0)$ verifiziert man zunächst für ϱG und dann für $G=(\varrho G)[x_j|j\in J]$ durch triviale Fortsetzung.

Fall γ). Die Annahme $\delta x_1 \neq 0$, $\delta y_1 = 0$ führt wegen (39) auf den Widerspruch:

$$0=\delta(x_1^2+y_1^2)=\delta(x_1y_1)=x_1^2y_1=z_3\neq 0,$$

ebenso die Annahme $\delta x_1 = 0$, $\delta y_1 \neq 0$.

Die Existenzbehauptung wird wie in den vorigen Fällen bewiesen.
Wir beschreiben die topologisch relevanten Möglichkeiten für δ explizit.

Hilfssatz 11. *Sei* $\delta : G\to G$ *eine Derivation mit* (42) *und* $0\leqq i\leqq g$.
a) *Im Fall* $s=2$, $g\geqq 4$, $\delta x_1 = x_2$, $\delta y_1 = 0$ *gilt:*

$$\delta x_i=\begin{cases} x_{i+1} & \text{für } i\equiv 1(4)\\ 0 & \text{sonst;}\end{cases} \qquad \delta y_i=\begin{cases} y_{i+1} & \text{für } i\equiv 2(4)\\ 0 & \text{sonst.}\end{cases}$$

b) *Im Falle* $s=3$, $\delta x_1 = x_2$, $\delta y_1 = y_2$ *gilt:*

$$\delta x_i=\begin{cases} x_{i+1} & \text{für } i\equiv 1(3)\\ 0 & \text{sonst;}\end{cases} \qquad \delta y_i=\begin{cases} y_{i+1} & \text{für } i\equiv 1(3)\\ 0 & \text{sonst.}\end{cases}$$

Beweis zu a). Es gilt $\delta z_2=\delta(x_1y_1)=x_2y_1=y_3$ und damit $\delta x_2=0$, $\delta y_2=y_3$. Man hat also

$$\delta x_1=x_2,\quad \delta x_2=0,\quad \delta x_3=\delta(x_1x_2)=x_2^2=0$$
$$\delta y_1=0,\quad \delta y_2=y_3,\quad \delta y_3=\delta(y_1y_2)=y_1y_3=0,\quad \delta z_4=\delta(x_1y_3)=0.$$

Wegen (41) folgt $\delta x_j=\delta y_j=0$ für $j=2^l4\leqq g/2$ und damit wegen (40) und der entsprechenden Darstellung der y_i die Periodizität mod4.

Beweis zu b). Es gilt wegen (39)

$$\delta x_1=x_2,\quad \delta x_2=\delta x_1^2=0,\quad \delta y_1=y_2,\quad \delta y_2=\delta y_1^2=0$$
$$\delta z_3=\delta(x_1^2y_1)=x_1^2y_1^2=x_1^2(x_1^2+x_1y_1)=0.$$

Die Periodizität mod3 folgt aus (40) und (41).

11. Der nichtorientierbare Fall

Die Voraussetzungen seien dieselben wie in Abschn. 7. Wir behandeln den Fall, daß mindestens eine der Mannigfaltigkeiten M_+, M_- nicht orientierbar ist. Dazu untersuchen wir den Ringhomomorphismus

$$h: H^*(M,\mathbb{Z}_4)\to H^*(M,\mathbb{Z}_2),$$

der durch den Ringhomomorphismus $\mathbb{Z}_4\to\mathbb{Z}_2$ induziert wird. Dieser Homomorphismus ist mit dem Bocksteinhomomorphismus Sq^1 verbunden durch die exakte Sequenz

$$\ldots\xrightarrow{Sq^1} H^q(X,\mathbb{Z}_2)\xrightarrow{i} H^q(X,\mathbb{Z}_4)\xrightarrow{h} H^q(X,\mathbb{Z}_2)\xrightarrow{Sq^1} H^{q+1}(X,\mathbb{Z}), \tag{43}$$

die von der exakten Sequenz $0\to\mathbb{Z}_2\to\mathbb{Z}_4\to\mathbb{Z}_2\to 0$ induziert wird. Sq^1 wirkt als Derivation vom Grad 1 und hat überdies die Eigenschaft $Sq^1u=u^2$ für $u\in H^1(x,\mathbb{Z}_2)$.

Wir berechnen Sq^1 auf $H^*(M,\mathbb{Z}_2)$ mit Hilfe von

Hilfssatz 12. *Der Homomorphismus*

$$Sq^1: H^{m_k}(M_{-k},\mathbb{Z}_2)\to H^{m_k+1}(M_{-k},\mathbb{Z}_2) \tag{44}$$

ist ein Isomorphismus, wenn M_k nicht orientierbar ist, und Null, wenn M_k orientierbar ist $(k=\pm 1)$.

Beweis. Sei M_k nicht orientierbar. Dann folgt aus dem Alexanderschen Dualitätssatz [vgl. Abschn. 7, (21)]

$$H_q(M_{-k},\mathbb{Z})\cong H^{n-1-q}(M_k,\mathbb{Z})\cong\begin{cases}0 & \text{für } 0<q<m_k\\ \mathbb{Z}_2 & \text{für } q=m_k.\end{cases}$$

Durch Einsetzen in das universelle Koeffiziententheorem der Kohomologie erhält man die exakt-direkte Sequenz

$$0\to\mathrm{Ext}(H_{m_k-1}(M_{-k},\mathbb{Z}),\mathbb{Z}_4)\to H^{m_k}(M_{-k},\mathbb{Z}_4)\to\mathrm{Hom}(\mathbb{Z}_2,\mathbb{Z}_4)\to 0$$

und aus ihr $H^{m_k}(M_{-k},\mathbb{Z}_4)=\mathbb{Z}_2$ (im Fall $m_k=1$ beachtet man $H_0(M_{-k},\mathbb{Z})\cong\mathbb{Z}$). Die Sequenz (43) liefert daher für $X=M_{-k}$, $q=m_k$ einen Isomorphismus i, und damit einen Monomorphismus (44). Aus Satz 5 folgt $m_-=1$, und (44) ist ein Isomorphismus. Wenn M_k orientierbar ist, zeigt man in analoger Weise $H_{m_k}(M_{-k},\mathbb{Z})\cong\mathbb{Z}$, $H^{m_k}(M_{-k},\mathbb{Z}_4)=\mathbb{Z}_4$, und (44) ist der Nullhomomorphismus.

Der Derivation Sq^1 auf $H^*(M,\mathbb{Z}_2)$ entspricht eine Derivation $\delta: G\to G$ der zugehörigen $\mathbb{Z}_2$-Algebra vom Typ (*). Diese Derivation besitzt die Eigenschaften (42), denn $\delta G_i\subset G_{i+1}$ verifiziert man leicht durch Fallunterscheidungen und $\delta G^k\subset G^k$ folgt aus der Natürlichkeit von Sq^1, d. h. aus $Sq^1\circ p_k^*=p_k^*\circ Sq^1$; insbesondere entspricht die Wirkung von $\delta|G^k$ der Wirkung von Sq^1 auf $H^*(M_k,\mathbb{Z}_2)$. Aus Hilfssatz 12 entnehmen wir die Wirkung von δ auf die Erzeugenden von G_1. Nach Satz 9 ist damit δ auf ganz G bestimmt. Es gibt im wesentlichen 3 Fälle:

(i) M_+ ist orientierbar, und M_- ist nicht orientierbar, $m_+=1$, $m_->1$. Dann gilt $\bar{s}=2$, $g\geqq 4$ und $\delta x_1\neq 0$, $\delta y_1=0$.

(ii) M_+ ist orientierbar, und M_- ist nicht orientierbar, $m_+=1$, $m_-=1$. Dann gilt $\bar{s}=4$ und $\delta x_1=x_1^2\neq 0$, $\delta y_1=y_1^2=0$.

(iii) Beide M_k sind nicht orientierbar, $m_+=m_-=1$. Dann gilt $\bar{s}=3$, $\delta x_1=x_1^2$, $\delta y_1=y_1^2$.

Alle diese Fälle sind explizit in Hilfssatz 11 beschrieben [(i) und (ii) unter a)].

Wir berechnen jetzt $H^*(M,\mathbb{Z}_4)$. Aus der Bocksteinsequenz (43) sieht man, daß $H^q(M_k,\mathbb{Z}_2)=0$ zur Folge hat $H^q(M_k,\mathbb{Z}_4)=0$, d. h. $H^*(M,\mathbb{Z}_4)$ läßt die gleiche Gradverzerrung zu wie $H^*(M,\mathbb{Z}_2)$. Wir setzen daher wie in Abschn. 8 $\tilde{G}_0=H^0(M,\mathbb{Z}_4)$, $\tilde{G}_{n-1}=H^{n-1}(M,\mathbb{Z}_4)$ und

$$\tilde{G}^k_{2j+\varepsilon}=p_k^*(H^{j\mu+\varepsilon m-\kappa}(M_k,\mathbb{Z}_4))\quad (k=\pm 1;\ \varepsilon=0,1;\ 1\leqq 2j+\varepsilon\leqq g-1;\ j\in\mathbb{N})$$

und $\tilde{G}_i=\tilde{G}_i^+\oplus\tilde{G}_i^-$ für $1\leqq i\leqq g-1$. Damit erhalten wir eine graduierte $\mathbb{Z}_4$-Algebra $\tilde{G}=\bigoplus_0^g\tilde{G}_i$, als Algebra isomorph zu $H^*(M,\mathbb{Z}_4)$, die G durch den Ringhomomorphismus $h:\tilde{G}\to G$ überlagert.

Hilfssatz 11 lehrt, daß Sequenzen $\delta x_i \neq 0$, $\delta x_{i+1} \neq 0$ bzw. $\delta y_i \neq 0$, $\delta y_{i+1} \neq 0$ nicht auftreten. Dies bedeutet in Verbindung mit (43), daß alle $\tilde{G}_i^k$ $(1 \leqq i \leqq g-1,\ k = \pm 1)$ von Null verschieden sind.

Wir fassen die Informationen über $h : \tilde{G} \to G$ zusammen in

Hilfssatz 13. *Es gilt* $\tilde{G}_0 \cong \tilde{G}_g \cong \mathbb{Z}_4$, $h(\tilde{G}_0) = G_0$, $h(\tilde{G}_g) = G_g$. *Für die* $\tilde{G}_i^k$, $1 \leqq i \leqq g-1$, $k = \pm 1$ *gibt es zwei Fälle (modulo Vertauschung von* $\tilde{G}^+$ *und* $\tilde{G}^-$*):*

a) $s(G) = 2$, $\delta x_1 \neq 0$, $\delta y_1 = 0$. *Dann gilt*

$$\tilde{G}_i^+ \cong \begin{cases} \mathbb{Z}_4 & \text{für } i \equiv 0, 3(4) \\ \mathbb{Z}_2 & \text{für } i \equiv 1, 2(4), \end{cases} \qquad \tilde{G}_i^- \cong \begin{cases} \mathbb{Z}_4 & \text{für } i = 0, 1(4) \\ \mathbb{Z}_2 & \text{sonst;} \end{cases}$$

$$h(\tilde{G}_i^+) = \begin{cases} 0 & \text{für } i \equiv 1(4) \\ G_i^+ & \text{sonst,} \end{cases} \qquad h(\tilde{G}_i^-) = \begin{cases} 0 & \text{für } i \equiv 2(4) \\ G_i^- & \text{sonst.} \end{cases}$$

b) $s(G) = 3$, $\delta x_1 = x_1^2 \neq 0$, $\delta y_1 = y_1^2 \neq 0$. *Dann gilt*

$$\tilde{G}_i^k \cong \begin{cases} \mathbb{Z}_4 & \text{für } i \equiv 0(3) \\ \mathbb{Z}_2 & \text{sonst,} \end{cases}$$

$$h(\tilde{G}_i^k) = \begin{cases} 0 & \text{für } i \equiv 1(3) \\ G_i^k & \text{sonst,} \end{cases} \qquad k = \pm 1.$$

Der Beweis entsteht, wie schon angedeutet, durch Einbringen von Hilfssatz 11 in die Sequenz (43) und ist evident bis auf den Fall bei dem (43) auf die exakte Sequenz $0 \to \mathbb{Z}_2 \to H^q(M_k, \mathbb{Z}_4) \to \mathbb{Z}_2 \to 0$ führt (z. B. für $\delta x_2 = \delta x_3 = 0$). Dieser wird erledigt durch

Hilfssatz 14. *Es sei* X *ein topologischer Raum mit* $H^q(X, \mathbb{Z}_2) \cong \mathbb{Z}_2$. *Dann ist* $H^q(X, \mathbb{Z}_4) \ncong \mathbb{Z}_2 \oplus \mathbb{Z}_2$.

Beweis. Nach einem universellen Koeffiziententheorem gilt

$$H^q(X, G) = (H^q(X, \mathbb{Z}) \otimes G) \oplus (H^{q+1}(X, \mathbb{Z}) * G),$$

wobei $G_1 * G_2 = \mathrm{Tor}^1(G_1, G_2)$ gesetzt ist. Also ist für $H^q(X, \mathbb{Z}_2) \cong \mathbb{Z}_2$ eine der Gruppen $H^q(X, \mathbb{Z}) \otimes \mathbb{Z}_2$ oder $H^{q+1}(X, \mathbb{Z}) * \mathbb{Z}_2$ Null. Es folgt, daß auch eine der beiden Gruppen $H^q(X, \mathbb{Z}) \otimes \mathbb{Z}_4$ bzw. $H^{q+1}(X, \mathbb{Z}) * \mathbb{Z}_4$ Null und also $H^q(X, \mathbb{Z}_4)$ jeweils isomorph zur anderen ist. Aber wenn für eine Gruppe G die Gruppe $G \otimes \mathbb{Z}_2$ (bzw. $G * \mathbb{Z}_2$) zyklisch ist, muß offensichtlich auch $G \otimes \mathbb{Z}_4$ (bzw. $G * \mathbb{Z}_4$) zyklisch sein.

Das Hauptresultat dieses Abschnitts ist:

Satz 10. *Die Voraussetzungen aus Abschn. 7 seien erfüllt; ferner sei* M_- *nicht orientierbar. Dann gibt es nur die folgenden Möglichkeiten:*

a) M_+ *ist orientierbar. Dann gilt* $m_+ = 1$, $g = 4$.

b) M_+ *ist nicht orientierbar. Dann gilt* $m_+ = m_- = 1$ *und* $g = 3$ *oder* $g = 6$.

Beweis von a). Nach Hilfssatz 12 liegt einer der Fälle (i) oder (ii) bzw. der Fall a) von Hilfssatz 11 vor. Es folgt $g = 2^{r+1}$, $r \geqq 1$.

Wir nehmen $g > 4$, also $r \geqq 2$ an. Dann ist $d = g/2$ ein Vielfaches von 4. Man hat $x_d + y_d = x_{d-1} y_1$ und nach Hilfssatz 13, Fall a) können Urbilder u_d, v_d, u_{d-1}, v_1 von x_d, y_d, x_{d-1}, y_1 unter h so gewählt werden, daß $u_d + v_d = u_{d-1} v_1$ gilt.

Nach (40) besteht die Darstellung $x_{d-1}=x_3x_c=x_1x_2x_c$ mit $c\equiv 0 \bmod 4$. Damit gilt $x_{d-1}^2=0$, weil hier der Faktor $x_s^2=x_2^2=0$ auftritt (vgl. Hilfssatz 8). Da h nach Hilfssatz 13 $\tilde{G}_{g-2}^+$ isomorph auf G_{g-2}^+ abbildet, folgt $u_{d-1}^2=0$ und damit

$$(u_d+v_d)^2=u_{d-1}^2v_1^2=0.$$

Andererseits entspricht $\tilde{G}_d$ einer Kohomologiegruppe $H^l(M,\mathbb{Z}_4)$, deren Grad $l=2^{r-1}(m_++m_-)$ nach Voraussetzung gerade ist. Es folgt $u_dv_d=v_du_d$, und wegen $h(u_dv_d)=x_dy_d=z$ ist $u_dv_d=w$ Erzeugendes von $\tilde{G}_g\cong H^{n-1}(M,\mathbb{Z}_4)\cong\mathbb{Z}_4$. Wegen $H^{n-1}(M_k,\mathbb{Z}_4)=0$ gilt ferner $u_d^2=0=v_d^2$.

Insgesamt erhalten wir

$$(u_d+v_d)^2=2w\neq 0.$$

Den Widerspruch beweist $g=4$.

Beweis für den Fall b). Man hat nach Hilfssatz 11 bzw. 12 $m_+=m_-=1$, $s(G)=3$ und damit $g=2^r3$ mit $r\geqq 1$. Wir nehmen $g>6$, also $r\geqq 2$ an und setzen $d=2^{r-2}3=g/4$. Nach Hilfssatz 8 besteht die Identität $x_{2d}+y_{2d}=x_dy_d$. Sie ist nach Hilfssatz 13b) Bild unter h von

$$u_{2d}+v_{2d}=u_dv_d,$$

wenn man die Erzeugenden $u_d\in\tilde{G}_d^+$, $v_d\in\tilde{G}_d^-$ usw. geeignet wählt. Die Beziehungen $x_d^2=0$, $y_d^2=0$ implizieren $u_d^2=n_1u_{2d}$, $v_d^2=n_2v_{2d}$ mit geraden Zahlen n_1,n_2 und damit

$$(u_{2d}+v_{2d})^2=0.$$

Andererseits gilt $\tilde{G}_{2d}=H^{2d}(M,\mathbb{Z}_4)\cong\mathbb{Z}_4$ und damit wie bei Teil a)

$$(u_{2d}+v_{2d})^2=2u_{2d}v_{2d}\neq 0.$$

Der Widerspruch vollendet den Beweis des Satzes.

Literatur (Ergänzung zur Literatur von Teil I)

16. Münzner, H.F.: Isoparametrische Hyperflächen in Sphären. I. Math. Ann. **251**, 57–71 (1980)
17. Spanier, E.H.: Algebraic topology. New York: McGraw-Hill 1966

Eingegangen am 8. August 1980

Math. Ann. 256, 233–248 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Perturbations of Selfadjoint Operators with Point Spectra by Restrictions and Selfadjoint Extensions

Konrad Schmüdgen*

Sektion Mathematik und NTZ, Karl-Marx-Universität Leipzig, DDR-7010 Leipzig, German Democratic Republic

1. Introduction

Let $Lu=-u''+qu$, $q(t)$ be a real locally integrable function on $[0,+\infty)$, be a singular Sturm-Liouville operator on the domain of all absolutely continuous functions in $L_2(0,+\infty)$ with $u(0)=u'(0)=0$. Then, L is a closed symmetric operator with deficiency indices either $(1,1)$ or $(2,2)$. Suppose that L has deficiency indices $(1,1)$, i.e. L is in the limit-point case in the classical theory of Weyl [16, 7]. All selfadjoint extensions of L can be parametrized by the boundary condition

$$L_\alpha : \sin\alpha u(0)-\cos\alpha u'(0)=0, \quad \alpha\in[0,2\pi).$$

A well-studied problem is the change of the spectrum of L_α under changes of the boundary condition (see for example [2, 6, 10]). The goal of the present paper is to study this problem in a more general setting. Roughly speaking, we investigate how the spectrum of a selfadjoint operator A with point spectrum is changed if we firstly restrict A to a symmetric operator A_0, say with deficiency indices (k,k), and then extend A_0 again to a selfadjoint operator B. The case $k=1$ has also been considered by Wolf [17] and Donoghue [4].

To be more precise, we need some definitions and notations. Let A and B be unbounded selfadjoint operators in a separable Hilbert space $\mathscr{H}$. Let $k\in\mathbb{N}$ or $k=+\infty$. A is called *related* to B, i.e. $A\sim B$, if there exist a unitary operator U in $\mathscr{H}$ and a dense subspace $\mathscr{D}$ of the domain $D(A)$ such that $A\mathscr{D}\subseteq\mathscr{D}$ and $A\phi=U^{-1}BU\phi$ for all $\phi\in\mathscr{D}$. The operator U is called a *weak interwining operator* for A and B. If the restriction $A\restriction\mathscr{D}$ has deficiency indices (k,k), than we say A is *k-related* to B and we write $A\underset{k}{\sim}B$.

Further, let $(a_n, n\in\mathbb{N})$ and $(b_n, n\in\mathbb{N})$ be unbounded sequences of real numbers. We will denote by $A_{(a_n)}$ the diagonal operator in l_2 acting on the standard orthonormal basis $\{e^k=(\delta_{nk})\}$ by $A_{(a_n)}e^k=a_ke^k$. We shall say the set $\{a_n\}$ is k-related resp. related to the set $\{b_n\}$ if there exist a permutation σ so that

$$A_{(a_n)}\underset{k}{\sim}A_{(\sigma(b_n))} \quad \text{resp.} \quad A_{(a_n)}\sim A_{(\sigma(b_n))}.$$

* *Present address:* Sektion Mathematik, Ernst-Moritz-Arndt-Universität, DDR-2200, Greifswald, German Democratic Republic

0025-5831/81/0256/0233/$03.20

Again we use the notations $\{a_n\} \underset{k}{\sim} \{b_n\}$ and $\{a_n\} \sim \{b_n\}$. In other words, $\{a_n\} \underset{k}{\sim} \{b_n\}$ means that a selfadjoint operator A with point spectrum $\{a_n\}$ can be transformed into a selfadjoint operator B with point spectrum $\{b_n\}$ by restricting A to a symmetric operator with deficiency indices (k, k) on a dense invariant domain and then extending this restriction to B.

The main results of the paper are the following. For an arbitrary semibounded real sequence (a_n) without finite cluster points we give a complete characterization of all sequences (b_n) with $\{a_n\} \underset{1}{\sim} \{b_n\}$ (Theorem 4.1). Some results of the so-called theory of the determination of a second order ordinary differential equation by two spectra ([6], for a survey up to 1964 see [10]) fit in a natural way in our construction. Further, we prove that $\{a_n\} \sim \{b_n\}$ for unbounded real sequences (a_n), (b_n) if and only if $\sup_n a_n = \sup_n b_n = +\infty$ or $\inf_n a_n = \inf_n b_n = -\infty$, that is, iff $\{a_n\}$ and $\{b_n\}$ are unbounded in a common direction (Theorem 4.5). An obvious consequence from this is that the sets $\{a_n\}$ and $\{b_n\}$ of two unbounded real sequences (a_n), (b_n) are always related after at most two steps, i.e. there is a (unbounded) real sequence (c_n) so that $\{a_n\} \sim \{c_n\}$ and $\{c_n\} \sim \{b_n\}$. We conjecture that similar assertions are true for arbitrary unbounded selfadjoint operators in a separable Hilbert space.

Let us add a few comments concerning the relations "$\sim$", "$\underset{k}{\sim}$" and last mentioned results. Clearly, both relations are reflexive and symmetric. But they are not transitive in general (see Theorems 4.1 and 4.5). The relation "$\sim$" and Theorem 4.5 are somewhat in the spirit of Neumann's famous paper about unbounded matrices. In a sense Theorem 4.5 shows the pathology of unbounded operators. However, our investigations have been more motivated by an application given in [15] where the results of this paper are used to construct new irreducible representations of the Heisenberg commutation relation $PQ - QP = -iI$ by essentially selfadjoint operators P, Q on a common dense invariant domain in a Hilbert space. Possible applications of this kind are the main reason for putting the invariance $A\mathcal{D} \subseteq \mathcal{D}$ into the definition of "$A \sim B$". It should be noted that this is not automatically fulfilled because there exist a restriction of a selfadjoint operator A to a closed symmetric densely defined operator A_0 such that the only subset $\mathcal{D} \subseteq \mathcal{D}(A_0)$ with $A_0\mathcal{D} \subseteq \mathcal{D}$ is $\mathcal{D} = \{0\}$ [12]. By Sect. 2, such an operator A_0 has deficiency indices $(+\infty, +\infty)$.

The procedure of restriction of a selfadjoint operator and selfadjoint extension is discussed in Sect. 2. If $\{a_n\} \underset{1}{\sim} \{b_n\}$, then (a_n) resp. (b_n) are the poles resp. zeros of a certain Pick function, that is, an analytic function in the upper half-plane with non-negative imaginary part. The main technical tool which gives two of the necessary and sufficient conditions for $\{a_n\} \underset{1}{\sim} \{b_n\}$ is the asymptotic behaviour of this function along the imaginary axis (Sect. 3). The main results and some of its consequences are proved in Sect. 4.

2. Restriction and Selfadjoint Extension

2.1

Let A be an unbounded selfadjoint operator in a Hilbert space H and $V = (A-i)(A+i)^{-1}$ its Cayley transform. Then, V is a unitary operator in H. Let $\tilde{A}$

be a closed symmetric restriction of A with deficiency spaces $\mathscr{H}_+$ and $V\mathscr{H}_+$ for i resp. $-i$. Let P_+ be the orthogonal projection on the closed subspace $\mathscr{H}_+$. Then, the Cayley transform $\tilde{V}$ of $\tilde{A}$ is a partial isometry from $\mathscr{H}\ominus\mathscr{H}_+$ on $V\mathscr{H}\ominus V\mathscr{H}_+$, i.e. $\tilde{V}=V(I-P_+)$.

Now, conversely, let $\mathscr{H}_+$ be a closed linear subspace of $\mathscr{H}$ and $\tilde{V}:=V(I-P_+)$. Let $\tilde{A}$ the (first of all not necessary densely defined) closed symmetric operator with Cayley transform $\tilde{V}$. Let $\mathscr{D}_\infty(\tilde{A}):=\bigcap_{n=1}^{\infty}\mathscr{D}(\tilde{A}^n)$. Our first aim is to study when the linear subspaces $\mathscr{D}(\tilde{A}^n)$, $n\in\mathbb{N}$, and $\mathscr{D}_\infty(\tilde{A})$ are dense in $\mathscr{H}$. We denote by $\mathscr{H}_n$ the closed linear subspace of $\mathscr{H}$ generated by $\mathscr{H}_+, V^{-1}\mathscr{H}_+, \ldots, V^{-(n-1)}\mathscr{H}_+$ and by P_n the orthogonal projection on $\mathscr{H}_n$. Clearly, $\mathscr{H}_1=\mathscr{H}_+$ and $P_1=P_+$.

Proposition 1. i) *$\mathscr{D}(\tilde{A}^n)$ is dense in H if and only if $\mathscr{D}(A^n)\cap\mathscr{H}_n=\{0\}$. In particular, $\tilde{A}$ is densely defined closed symmetric operator iff $\mathscr{D}(A)\cap\mathscr{H}_+=\{0\}$.*

ii) *Suppose that $\mathscr{H}_+$ is a finite dimensional subspace of $\mathscr{H}$ with $\mathscr{D}(A)\cap\mathscr{H}_+=\{0\}$. Then, $\mathscr{D}_\infty(\tilde{A})$ is dense in $\mathscr{H}$. Moreover, $\mathscr{D}_\infty(\tilde{A})$ is a core for $\tilde{A}^n$ $\forall n\in N$.*

Proof. Ad i): By the von Neumann theory of the Cayley transform we have

$$\mathscr{D}(\tilde{A})=(I-V)(\mathscr{H}\ominus\mathscr{H}_+)\equiv(I-V)(I-P_+)\mathscr{H}$$

and $\tilde{A}\psi=i(I+V)(I-P_+)\eta$ for $\psi=(I-V)(I-P_+)\eta$. Since

$$(I-P_n)\mathscr{H}=\{\eta\in\mathscr{H}:(I+V)^k\eta\perp\mathscr{H}_+\ \ \forall k=0,1,\ldots,n-1\},$$

it follows that $D(\tilde{A}^n)=(I-V)(I-P_n)\mathscr{H}$. The operator $T:=(I-V)^n(I-P_n)$ has a dense range if and only if the null space of

$$T^*=(I-P_n)(I-V)^n(-V)^{-n}$$

is trivial, that is,

$$P_n(I-V)^n(-V)^n\eta=(I-V)^n(-V)^{-n}\eta$$

implies $\eta=0$. Since $\mathscr{D}(A^n)=(I-V)^n\mathscr{H}$, this is equivalent to $\mathscr{D}(A^n)\cap\mathscr{H}_n=\{0\}$.

Ad ii): The main step of the proof is to show that $D(\tilde{A}^{n+1})$ is a core for $\tilde{A}^n$, $n=0,1,\ldots$. [In particular, this implies that $\mathscr{D}(\tilde{A}^{n+1})$ is dense in $\mathscr{H}$ for all $n\in\mathbb{N}$.] Since

$$(I-V)^{-n}(I-V^*)^{-n}=[1/4(A^2+I)]^n=\sum_{r=0}^{n}4^{-r}\binom{n}{r}A^{2r},$$

we have

$$\langle(I-V)^{-n}\phi,(I-V)^{-n}\psi\rangle=\sum_{r=0}^{n}4^{-r}\binom{n}{r}\langle A^r\phi,A^r\psi\rangle=:\langle\phi,\psi\rangle_n$$

on $\mathscr{D}(A^n)$. To prove that $\mathscr{D}(\tilde{A}^{n+1})$ is a core for $\tilde{A}^n$, it is sufficient to show that $\mathscr{D}(\tilde{A}^{n+1})$ is dense in $\mathscr{D}(\tilde{A}^n)$ endowed with the norm $\|\cdot\|_n$. Assume that a vector $\psi\in\mathscr{D}(\tilde{A}^n)$ is $\langle\cdot,\cdot\rangle_n$-orthogonal to $\mathscr{D}(\tilde{A}^{n+1})$. Let $\psi=(I-V)^n\zeta$ where $\zeta\in(I-P_n)\mathscr{H}$. Then,

$$\langle\psi,(I-V)^{n+1}\eta\rangle_n=0=\langle\zeta,(I-V)\eta\rangle$$

for all $\eta \in (I-P_{n+1})\mathscr{H}$. Hence, $(I-V^*)\zeta \in \mathscr{H}_{n+1}$. Since we assumed that $\mathscr{H}_+$ is of finite dimension, $\mathrm{Lin}(\mathscr{H}_+, V^{-1}\mathscr{H}_+, \ldots, V^{-n}\mathscr{H}_+)$ is a *closed* linear subspace of $\mathscr{H}$, that is,

$$\mathscr{H}_{n+1} = \mathrm{Lin}(\mathscr{H}_+, V^{-1}\mathscr{H}_+, \ldots, V^{-n}\mathscr{H}_+).$$

On the other hand, by the formula of the Cayley transform, we have

$$\mathrm{Lin}(\mathscr{H}_+, V^{-1}\mathscr{H}_+, \ldots, V^{-n}\mathscr{H}) = \mathrm{Lin}(\mathscr{H}_+, (A-i)^{-1}\mathscr{H}_+, \ldots, (A-i)^{-n}\mathscr{H}_+).$$

Therefore,

$$(I-V^*)\zeta \equiv -2i(A-i)^{-1}\zeta = \xi^0 + (A-i)^{-1}\xi^1 + \ldots + (A-i)^{-n}\xi^n$$

with certain vectors $\xi^0, \xi^1, \ldots, \xi^n$ of $\mathscr{H}_+$. This equation gives us $\xi^0 \in \mathscr{D}(A)$. Since $\mathscr{D}(A) \cap \mathscr{H}_+ = \{0\}$ by assumption, we get $\xi^0 = 0$ and

$$\zeta = 1/2i[\xi^1 + (A-i)^{-1}\xi^2 + \ldots + (A-i)^{-(n-1)}\xi^n],$$

that is, $\zeta \in \mathscr{H}_n$. Because $\zeta \in (I-P_n)\mathscr{H}$, it follows that $\zeta = 0$ and $\psi = 0$. This completes the proof that $\mathscr{D}(\tilde{A}^{n+1})$ is a core for $\tilde{A}^n$.

In order to show that $\mathscr{D}_\infty(\tilde{A})$ is a core for $\tilde{A}^n$, we have to prove that $\mathscr{D}_\infty(\tilde{A})$ is $\|\cdot\|_n$-dense in $\mathscr{D}(\tilde{A}^n)$. We use a simple topological argument which the author has seen in another context in [13]. Let $\phi_n \in \mathscr{D}(\tilde{A}^n)$ and $\varepsilon > 0$. Since $\mathscr{D}(\tilde{A}^{k+1})$ is $\|\cdot\|_k$-dense in $\mathscr{D}(\tilde{A}^k)$ by the preceding consideration, there are vectors $\phi_k \in \mathscr{D}(\tilde{A}^k)$, $k \geqq n$, so that $\|\phi_k - \phi_{k+1}\|_k < \varepsilon \cdot 2^{-2k-1}$. Because $\|\cdot\|_k \leqq 2\|\cdot\|_{k+1}$, this immediately implies that $(\phi_k, k \geqq n)$ is $a\|\cdot\|_m$-Cauchy sequence for all $m \geqq n$. Since $\mathscr{D}(\tilde{A}^m)$ endowed with $\langle\cdot,\cdot\rangle_m$ is a Hilbert space and the norm $\|\cdot\|_m$ is stronger than the norm of $\mathscr{H}$, there exist a vector $\psi \in \bigcap_{m \geqq n} \mathscr{D}(\tilde{A}^m) \equiv \mathscr{D}_\infty(\tilde{A})$ such that

$$\lim_{k \to +\infty} \|\phi_k - \psi\|_m = 0 \quad \forall m \geqq n.$$

By the triangle inequality, we have $\|\phi_n - \phi_{n+r}\|_n < \varepsilon$ for all $r \in \mathbb{N}$. For $r \to +\infty$ we get $\|\phi_n - \psi\|_n \leqq \varepsilon$ completing the proof. □

2.2

Suppose now that $\mathscr{H}_+$ is a closed linear subspace of $\mathscr{H}$ so that $\mathscr{D}(A) \cap \mathscr{H}_+ = \{0\}$. Recall that the corresponding closed symmetric operator $\tilde{A}$ is densely defined, $D(\tilde{A}) = (I-V)(I-P_+)\mathscr{H}$ and

$$\tilde{A}\psi = i(I+V)(I-P_+)\eta \quad \text{for} \quad \psi = (I-V)(I-P_+)\eta, \quad \eta \in \mathscr{H}.$$

The deficiency spaces for i and $-i$ are $\mathscr{H}_+$ and $V\mathscr{H}_+$. Hence the selfadjoint extensions of A are in one-to-one correspondence to unitary operators W of $\mathscr{H}_+$ onto $\mathscr{H}_+$ and given by the formulaes

$$D(A_W) = \{\psi = (I-V)(I-P_+)\eta + \xi_+ + VW\xi_+, \eta \in \mathscr{H}, \xi_+ \in \mathscr{H}_+\}, \tag{1}$$

$$A_W\psi = i(I+V)(I-P_+)\eta + i\xi_+ - iVW\xi_+. \tag{2}$$

For $\psi = (I-V)(I-P_+)\eta + \xi + VW\xi_+$, $\eta \in \mathscr{H}$, $\xi_+ \in \mathscr{H}_+$, and $\lambda \in \mathbb{C}_1$ it follows that

$$(A_W - \lambda)\psi = (A+i)^{-1}[2i(A-\lambda)(I-P_+)\eta + i(A-\lambda)(I-W)\xi_+ - (\lambda A + I)(I+W)\xi_+]. \tag{3}$$

2.3

Now we shall turn to operators with point spectra and restrictions of deficiency indices (1, 1). For the remainder of this paper let us make the convention that lower indices such as n, k etc. always denote the components of sequences.

Let $(a_n, n\in\mathbb{N})$ be a sequence of distinct real numbers having no finite cluster point and let $A=A_{(a_n)}$ be the corresponding diagonal operator in l_2. Suppose that $(\xi_n)\in l_2$, $\|\xi\|=1$, $(\xi_n)\notin\mathscr{D}(A_{(a_n)})$ and $\xi_n\neq 0\ \forall n\in\mathbb{N}$. Take $\mathscr{H}_+=\mathrm{Lin}\{\xi\}$. Let $W=e^{i\vartheta}\neq -1$, ϑ real, be the unitary operator in $\mathscr{H}_+$ and let A_W be the selfadjoint operator defined by (1) and (2).

Proposition 2. *Then, the spectrum of A_W consists only of eigenvalues of multiplicity one which are precisely the zeros of the meromorphic function*

$$f(z)=\sum_{n=1}^{\infty}\frac{a_n z+1}{a_n-z}|\xi_n|^2-K_\vartheta, \tag{4}$$

where

$$K_\vartheta:=i(1-e^{i\vartheta})(1+e^{i\vartheta})^{-1}. \tag{5}$$

Proof. In the considered case (3) reads to

$$((A_W-\lambda)\psi)_n=\frac{a_n-\lambda}{a_n+i}[2i\zeta_n+i(1-e^{i\vartheta})C\xi_n]-\frac{a_n\lambda+1}{a_n+i}(1+e^{i\vartheta})C\xi_n \tag{7}$$

for $\psi=(I-V)\zeta+C(\xi+e^{i\vartheta}V\xi)$, $\zeta=(I-P_+)\zeta$.

Since all selfadjoint extensions of symmetric operators with finite deficiency indices have the same continuous spectrum [1, p. 322], the spectrum $\sigma(A_W)$ contains only eigenvalues.

First suppose that $f(\lambda)\neq 0$. If $\lambda\notin\sigma(A)$, i.e. $\lambda\neq a_n\ \forall n\in\mathbb{N}$, then (7) implies

$$\begin{aligned}\langle(A+i)(A-\lambda)^{-1}(A_W-\lambda)\psi,\xi\rangle&=C\left[i(1-e^{i\vartheta})-\sum_{n=1}^{\infty}\frac{a_n\lambda+1}{a_n-\lambda}|\xi_n|^2(1+e^{i\vartheta})\right]\\&=-C(1+e^{i\vartheta})f(\lambda).\end{aligned}$$

Thus, $(A_W-\lambda)\psi=0$ leads to $C=0$. By (7), this gives $\zeta_n=0\ \forall n\in N$, that is, $\psi=0$ and λ is not an eigenvalue of A_W. Of course, $\lambda=a_k$ cannot be an eigenvalue by (7).

Now let $f(\lambda)=0$. Put

$$\zeta_n=\frac{1}{2i}\left[\frac{\lambda a_n+1}{a_n-\lambda}(1+e^{i\vartheta})\xi_n-i(1-e^{i\vartheta})\xi_n\right]. \tag{8}$$

Because $f(\lambda)=0$ and $\|\xi\|=1$, we have $\langle\zeta,\xi\rangle=0$, i.e. $\zeta=(I-P_+)\zeta$. Then, $(A_W-\lambda)\psi=0$ for $\psi=(I-V)\zeta+\xi+e^{i\vartheta}V\xi$. Thus, λ is an eigenvalue of A_W. Conversely, $(A_W-\lambda)\psi=0$ for $\psi=(I-V)\zeta+C(\xi+e^{i\vartheta}V\xi)$ and $\langle\zeta,\xi\rangle=0$ leads (up to a constant factor) to the eigenvector defined by means of (8). Therefore, λ is of multiplicity one.

Remark. The assumptions $W=e^{i\vartheta}\neq -1$ and $\xi_n\neq 0\ \forall n\in\mathbb{N}$ are no loss of generality. For $W=-1$ the aelfadjoint extension A_W coincides with A. If $\xi_k=0$, then the k-th components of A_W and A are the same, that is, $A_W e^k=Ae^k=a_k e^k$.

3. Asymptotic Behaviour of Pick Functions

By a Pick function we mean an analytic function in the upper half-plane with non-negative imaginary part. We refer to Donoghue [5] and Lewin [9] for more details about these functions.

3.1

Let (a_n) be a real sequence with $a_n < a_{n+1}\ \forall n \in \mathbb{N}$ and $\lim_{n\to\infty} a_n = +\infty$. Let $(\xi_n) \in l_2$. We consider the Pick function

$$f(z) = \sum_{n=1}^{\infty} \frac{a_n z + 1}{a_n - z} |\xi_n|^2 - K_\vartheta \quad \text{where} \quad K_\vartheta \in \mathbb{C}_1 . \tag{9}$$

Lemma 1. i) *If* $(a_n \xi_n^2) \in l_1$, *then* $f(i\infty) \equiv \lim_{|y|\to+\infty} f(iy) = \sum_{n=1}^{\infty} -a_n|\xi_n|^2 - K_\vartheta$.

If $(a_n \xi_n^2) \notin l_1$, *then* $\lim_{|y|\to+\infty} \mathrm{Re} f(iy) = -\infty$.

ii) *Assume that* $(a_n \xi_n^2) \in l_1$.

If $(a_n \xi_n) \in l_2$, *then* $\lim_{|y|\to+\infty} (f(iy) - f(i\infty))y = \sum_{n=1}^{\infty} i(a_n^2 + 1)|\xi_n|^2$.

If $(a_n \xi_n) \notin l_2$, *then* $\lim_{|y|\to+\infty} \mathrm{Im}(f(iy) - f(i\infty))y = +\infty$.

Proof. Ad i): We have

$$f(iy) = \sum_{n=1}^{\infty} \frac{i a_n y}{a_n - y} |\xi_n|^2 + \sum_{n=1}^{\infty} \frac{1}{a_n - iy} |\xi_n|^2 - K_\vartheta .$$

Since $\dfrac{1}{|a_n - iy|} |\xi_n|^2 \leqq |\xi_n|^2$ for $|y| \geqq 1$ and $(\xi_n) \in l_2$, the second sum converges to zero for $|y| \to +\infty$.

Suppose now that $(a_n \xi_n^2) \in l_1$. Because of $\left| \dfrac{i a_n y}{a_n - iy} \xi_n^2 \right| \leqq |a_n \xi_n^2|$, we can apply again the Lebesgue theorem of dominated convergence which gives

$$\lim_{|y|\to+\infty} f(iy) = \sum_{n=1}^{\infty} -a_n|\xi_n|^2 - K_\vartheta .$$

Next suppose $(a_n \xi_n^2) \notin l_1$. Assume that $\mathrm{Re} f(iy_k) \geqq M$ for a sequence $|y_k| \to +\infty$ and a certain constant M. This implies

$$\mathrm{Re} \sum_{n=1}^{\infty} \frac{i a_n y}{a_n - iy} |\xi_n|^2 = \sum_{n=1}^{\infty} -a_n |\xi_n|^2 \frac{y_k^2}{a_n^2 + y_k^2} \geqq M'$$

with a constant M'. Since $a_n > 0$ for all $n \geqq n_0$, we get

$$\sum_{n=1}^{m} -a_n |\xi_n|^2 \frac{y_k^2}{a_n^2 + y_y^2} \geqq M'$$

for all $m \geqq n_0$, $m, k \in \mathbb{N}$. Taking $k \to +\infty$, it follows that

$$\sum_{n=1}^{m} -a_n|\xi_n|^2 \geqq M' \quad \forall m \geqq n_0, \qquad m \in \mathbb{N}.$$

Since $a_n > 0$ $\forall n \geqq n_0$, this gives $(a_n\xi_n^2) \notin l_1$ which contradicts our assumption. This proves i).

Ad ii): Since $(a_n\xi_n^2) \in l_1$, $f(i\infty)$ exists according to *i*). Hence

$$(f(iy) - f(i\infty))y = \sum_{n=1}^{\infty} (a_n^2 + 1)|\xi_n|^2 \frac{y}{a_n - iy}.$$

In the case $(a_n\xi_n) \in l_2$ the Lebesgue theorem yields

$$\lim_{|y| \to +\infty} (f(iy) - f(i\infty))y = \sum_{n=1}^{\infty} i(a_n^2 + 1)|\xi_n|^2.$$

If $(a_n\xi_n) \notin l_2$, then the same argument as in part i) leads to

$$\lim_{|y| \to +\infty} \operatorname{Im}(f(iy) - f(i\infty))y = +\infty.$$

3.2

Let $(a_n, n \in \mathbb{N})$, $(b_n, n \in \mathbb{N})$ be the real sequences. Suppose that $a_n < b_n < a_{n+1}$ $\forall n \in \mathbb{N}$, $a_n > 0$ for $n \geqq n_0$ and $0 \notin [a_n, b_n]$ $\forall n \in \mathbb{N}$.

$$g(z) = \prod_{n=1}^{\infty} \left(1 - \frac{z}{b_n}\right)\left(1 - \frac{z}{a_n}\right)^{-1}. \tag{10}$$

Then, $g(z)$ is a meromorphic Pick function by a classical result of Krein [9, p. 308].

Lemma 2. i) $g(i\infty) \equiv \lim_{|y| \to +\infty} g(iy) = \prod_{n=1}^{\infty} a_n b_n^{-1}$.

ii) *Suppose that* $g(i\infty) \neq 0$.

Then, $\overline{\lim_{|y| \to +\infty}} |(g(iy) - g(i\infty))y| = +\infty$ *if and only if* $(b_n - a_n) \notin l_1$.

iii) *Suppose* $g(i\infty) = 0$. *In this case,* $\overline{\lim_{|y| \to +\infty}} |(g(iy) - g(i\infty))y| = +\infty$ *if and only if*

$$\overline{\lim_{x \to +\infty}} \left(\log x - \sum_{n=1}^{\infty} \log \frac{b_n^2(a_n^2 + x)}{a_n^2(b_n^2 + x)}\right) = +\infty.$$

Proof. Ad i): First assume that $\prod_{n=1}^{\infty} a_n b_n^{-1} \neq 0$. In that case the assertion is equivalent to

$$\lim_{|y| \to +\infty} g(iy) \prod_{n=1}^{\infty} a_n b_n^{-1} = \lim_{|y| \to +\infty} \lim_{k \to \infty} \prod_{n=1}^{k} \left(1 - \frac{z}{b_n}\right)\left(1 - \frac{z}{a_n}\right)^{-1} b_n a_n^{-1}$$

$$= \lim_{|y| \to +\infty} \prod_{n=1}^{\infty} \frac{b_n - iy}{a_n - iy} = 1.$$

Since $\prod_{n=1}^{\infty} a_n b_n^{-1} \neq 0$, we have $\left(\frac{b_n - a_n}{a_n}\right) \in l_1$. Let $S = \sum_{n=1}^{\infty} \frac{b_n - a_n}{|a_n|}$. A straightforward calculation gives

$$\left| \prod_{n=1}^{\infty} \frac{b_n - iy}{a_n - iy} - 1 \right| \leqq e^S \sum_{n=1}^{\infty} \frac{b_n - a_n}{(a_n^2 + y^2)^{1/2}}. \tag{11}$$

Because the sum tends to zero by the Lebesgue theorem, the assertion follows.

Now we turn to the case $\prod_{n=1}^{\infty} a_n b_n^{-1} = 0$. We have to prove that

$$\lim_{|y| \to +\infty} \log |g(iy)| = \lim_{|y| \to \infty} \sum_{n=1}^{\infty} \log q_n(y) = -\infty$$

where

$$q_n(y) := (b_n^2 + y^2)^{1/2} (a_n^2 + y^2)^{-1/2} a_n b_n^{-1}.$$

Assume the contrary, that is, $\sum_{n=1}^{\infty} \log q_n(y_k) \geqq M$ for a sequence $|y_k| \to +\infty$ and a constant M. Since $|a_n| < |b_n|$ for all $n \geqq n_0$ implies $\log q_n(y) < 0$, we have

$$\sum_{n=1}^{m} \log q_n(y_k) \geqq M' \quad \forall m, k \in \mathbb{N}$$

with a constant M'. For $k \to \infty$ we get

$$\sum_{n=1}^{m} \log a_n b_n^{-1} \geqq M' \quad \forall m \in \mathbb{N},$$

hence $(\log a_n b_n^{-1}) \in l_1$ which contradicts $\prod_{n=1}^{\infty} a_n b_n^{-1} = 0$.

Ad ii): Let $(b_n - a_n) \in l_1$. Using (11) and i), we obtain

$$|(g(iy) - g(i\infty))y| = \left| g(i\infty) y \left(\frac{g(iy)}{g(i\infty)} - 1 \right) \right| = \left| g(i\infty) y \left(\prod_{n=1}^{\infty} \frac{b_n - iy}{a_n - iy} - 1 \right) \right|$$

$$\leqq g(i\infty) e^S \sum_{n=1}^{\infty} \frac{(b_n - a_n)|y|}{(a_n^2 + y^2)^{1/2}} \leqq e^S \sum_{n=1}^{\infty} b_n - a_n. \tag{12}$$

Consequently, the function $(g(iy) - g(i\infty))y$ is bounded.

Conversely, suppose that $(b_n - a_n) \notin l_1$. Assume that $|(g(iy_k) - g(i\infty))y_k| \leqq M$ for a real sequence (y_k) with $|y_k| \to +\infty$ and a constant M. We shall denote by $h_{r,s}(z)$, $r \in \mathbb{N}$, $r \leqq s$, $s \in \mathbb{N}$ or $s = +\infty$, the Pick function $\prod_{n=r}^{s} \frac{b_n - z}{a_n - z}$. Clearly, for $r \leqq s$, r, $s \in \mathbb{N}$,

$$(b_r - a_r) \frac{y_k}{a_r - iy_k} + \sum_{n=r+1}^{s} (b_n - a_n) \frac{y_k}{a_n - iy_k} \prod_{l=r}^{n-1} \frac{b_l - iy_k}{a_l - iy_k} = y_k \left(\prod_{n=r}^{s} \frac{b_n - iy_k}{a_n - iy_k} - 1 \right)$$

implies that

$$\lim_{k\to\infty} |y_k(h_{r,s}(iy_k)-1)| = \sum_{n=r}^{s} (b_n - a_n). \tag{13}$$

Since $\left|\frac{b_n-iy}{a_n-iy}\right| \geqq 1$ for $n \geqq n_0$, $\arg\frac{b_n-iy}{a_n-iy} \gtrless 0$ for $y \gtrless 0$ and $h_{1,\infty}(z)$ is a Pick function, it follows that

$$\left|\prod_{n=n_0}^{m} \frac{b_n-iy}{a_n-iy} - 1\right| \leqq \left|\prod_{n=n_0}^{\infty} \frac{b_n-iy}{a_n-iy} - 1\right| \quad \forall m \geqq n_0, \quad y \in \mathbb{R}_1. \tag{14}$$

By (12) and (14), we have

$$\begin{aligned} |y_k(h_{n_0,m}(iy_k)-1)| &\underset{(14)}{\leqq} |y_k(h_{n_0,\infty}(iy_k)-1)| \\ &\leqq |h_{1,n_0-1}(iy_k)|^{-1}(|y_k(h_{1,\infty}(iy_k)-1)| + |y_k(h_{1,n_0-1}(iy_k)-1)|) \\ &\underset{(12)}{\leqq} |h_{1,n_0-1}(iy_k)|^{-1}(g(i\infty)^{-1}M + |y_k(h_{1,n_0-1}(iy_k)-1)|) \end{aligned}$$

for all $k, m \in \mathbb{N}$, $m \geqq n_0$. Passing to the limit as $k \to \infty$ and applying (13) we obtain

$$\sum_{n=n_0}^{m} b_n - a_n \leqq 1 \cdot (g(i\infty)^{-1}M + \sum_{n=1}^{n_0} (b_n - a_n)) \quad \text{for all} \quad m \in \mathbb{N}, \quad m \geqq n_0.$$

However, this is impossible, because $(b_n - a_n) \notin l_1$.

Ad iii): Since $\overline{\lim_{|y|\to+\infty}} |g(iy)y| = +\infty$ is equivalent to

$$\overline{\lim_{|y|\to+\infty}} y^2 \prod_{n=1}^{\infty} \frac{a_n^2(b_n^2+y^2)}{b_n^2(a_n^2+y^2)} = +\infty,$$

the assertion follows by taking the logarithm.

Remark. If $0 \in [a_k, b_k]$ for $k \in \mathbb{N}$, then some slight modifications are necessary. In the infinite product (10) defining $g(z)$ the factor $\left(1-\frac{z}{b_k}\right)\left(1-\frac{z}{a_k}\right)^{-1}$ must be replaced by $\frac{z-b_k}{z-a_k}$. Then, $g(i\infty) = \prod_{\substack{n=1 \\ n\neq k}}^{\infty} a_n b_n^{-1}$. With these modification the assertions of Lemma 2 remain valid in that case.

4. The Main Results

4.1

Theorem 1. *Let* $(a_n, n \in \mathbb{N})$ *be a real sequence with* $a_n < a_{n+1}$ $\forall n \in \mathbb{N}$ *and* $\lim_{n\to\infty} a_n = +\infty$. *Suppose* $(b_n, n \in \mathbb{N})$ *is a real sequence with* $a_n \neq b_k$ $\forall n, k \in \mathbb{N}$.

Then, $\{a_n\} \underset{1}{\sim} \{b_n\}$ *if and only if possible after new enumeration of* (b_n) *the following conditions are fulfilled.*

(a) $a_n < b_n < a_{n+1}$ $\forall n \in \mathbb{N}$ *(Case I) or* $b_n < a_n < b_{n+1}$ $\forall n \in \mathbb{N}$ *(Case II).*

(b) *If* $\left(\frac{b_n - a_n}{a_n}\right)^1 \in l_1$, *then* $(b_n - a_n) \notin l_1$.

(c) *If* $\left(\frac{b_n - a_n}{a_n}\right)^1 \notin l_1$, *then*

$$\lim_{x \to +\infty} \left(\log x - \sum_{n=1}^{\infty} \log \frac{b_n^2 (a_n^2 + x)}{a_n^2 (b_n^2 + x)} \right)^1 = +\infty$$

in case I and

$$\lim_{x \to +\infty} \left(\log x - \sum_{n=1}^{\infty} \log \frac{a_n^2 (b_n^2 + x)}{b_n^2 (a_n^2 + x)} \right)^1 = +\infty$$

in case II.

Proof. First we prove the necessity part of the theorem. Suppose that $\{a_n\} \underset{1}{\sim} \{b_n\}$, that is, there exist a permutation σ, a unitary operator V in l_2, a dense domain D so that

$$A_{(a_n)} \phi = V^{-1} A_{(\sigma(b_n))} V \phi \quad \forall \phi \in D, \qquad A_{(a_n)} \mathscr{D} \subseteq \mathscr{D} \quad \text{and} \quad A_{(a_n)} \restriction \mathscr{D}$$

has deficiency indices (1, 1). Let $(A_{(a_n)} \restriction \mathscr{D})^* \xi = i\xi$, $\|\xi\| = 1$. If $\xi_k = 0$ for a number $k \in \mathbb{N}$, then it follows immediately from (3) that a_k is an eigenvalue of A_W contradicting the assumption $a_r \neq b_s$ $\forall r, s \in N$. Thus, $\xi_n \neq 0$ $\forall n \in \mathbb{N}$ and we can apply Proposition 2 in Sect. 2. Since (a_n) and $(\sigma(b_n))$ are the poles resp. zeros of the Pick function $f(z)$ defined by (4), both sequences must alternate. Consequently, (a) is satisfied. Let us consider case I. For simplicity we assume that $\sigma(b_n) = b_n$ $\forall n \in \mathbb{N}$ and that $0 \notin [a_n, b_n]$ $\forall n \in \mathbb{N}$ (otherwise, the infinite products must be slightly modified according to the last remark in the preceding section). Then, we have

$$f(z) = c \prod_{n=1}^{\infty} \left(1 - \frac{z}{b_n}\right)\left(1 - \frac{z}{a_n}\right)^{-1} \quad \text{for all } z \in \mathbb{C}_1$$

where c is a certain positive constant [9, p. 308]. Clearly, $\left(\frac{b_n - a_n}{a_n}\right) \in l_1$ is equivalent to $\prod_{n=1}^{\infty} a_n b_n^{-1} \neq 0$. Suppose that $\left(\frac{b_n - a_n}{a_n}\right) \in l_1$. Then, by Lemmas 3.1ii) and 3.2ii), $(a_n \xi_n) \notin l_2$ implies $(b_n - a_n) \notin l_1$. Hence, (b) is fulfilled. Assume now $\left(\frac{b_n - a_n}{a_n}\right) \notin l_1$. Then, condition (c) follows from Lemmas 3.1ii) and 3.2iii). In case II the assertion follows by changing the role of (a_n) and (b_n). [Recall that the relation "$\underset{1}{\sim}$" and conditions (b) and (c) are symmetric.] This completes the first part of the proof.

---

1 Here we make the convention that for $a_n = 0$ or $b_n = 0$ the corresponding term will be set zero

Now we show that (a), (b), (c) are sufficient. As already noted in the proof of the necessity part, it suffices to treat case I. Again we assume that $0 \notin [a_n, b_n]$ $\forall n \in \mathbb{N}$. Since $a_n < b_n < a_{n+1}$ $\forall n \in \mathbb{N}$, the meromorphic function

$$g(z) := \prod_{n=1}^{\infty} \left(1 - \frac{z}{b_n}\right)\left(1 - \frac{z}{a_n}\right)^{-1}$$

has a non-negative imaginary part in the upper half-plane [9, p. 308]. By a theorem of Nevanlinna [1, 5], there exist real numbers $a, b, a \geqq 0$, and a positive Borel measure μ on $\mathbb{R}_1$ with $\int (t^2+1)^{-1} d\mu(t) < +\infty$ such that

$$g(z) = az + b + \int \frac{tz+1}{t-z} \frac{d\mu(t)}{t^2+1}.$$

Since the imaginary part of $g(z)$ is continuous and vanishes in each real intervall outside the poles, $\mu([\alpha, \beta]) = 0$ for all intervalls $[\alpha, \beta]$ which do not contain a pole [5]. Therefore, μ is a sum of point measures with mass at the poles, i.e.

$$g(z) = az + b + \sum_{n=1}^{\infty} \frac{a_n z + 1}{a_n - z} \frac{\mu_n}{a_n^2 + 1}$$

where $\mu_n > 0$ $\forall n \in \mathbb{N}$. Putting $\xi'_n = \mu_n^{1/2}(a_n^2+1)^{-1/2}$, we have

$$\sum_{n=1}^{\infty} |\xi'_n|^2 = \int (t^2+1)^{-1} d\mu(t) < +\infty$$

and

$$g(z) = az + b + \sum_{n=1}^{\infty} \frac{a_n z + 1}{a_n - z} |\xi'_n|^2.$$

Since $|a_n| < |b_n|$ for large n, we have $a \equiv \lim_{|y| \to +\infty} g(iy)(iy)^{-1} = 0$. By construction, $(\xi'_n) \in l_2$ and $\xi'_n \neq 0$ $\forall n \in \mathbb{N}$. Now we choose a real number ϑ so that $b = K_\vartheta \|\xi'\|^2$. Further, in view of Lemmas 3.1 and 3.2 conditions (b) and (c) ensure $(a_n \xi'_n) \notin l_2$. If we take $\mathscr{H}_+ = \mathrm{Lin}\{\xi\}$ and $\xi = \xi' \|\xi\|^{-1}$, then all assumptions of Proposition 2 in Sect. 2 are fulfilled. Hence, the set b_n of all zeros of $g(z)$ coincides with the spectrum of A_W, that is, we have $\{a_n\} \curlywedge_1 \{b_n\}$ completing the proof of the theorem.

Let (a_n), (b_n) be sequences which satisfy the assumptions and the conditions (a), (b), (c) of Theorem 1. Then, $\{a_n\} \curlywedge_1 \{b_n\}$ and by the preceding proof, $\sigma(A_W) = \{b_n, n \in \mathbb{N}\}$. Let $A_W \phi^k = b_k \phi^k$, $\|\phi^k\| = 1$. Clearly, the unitary operator U defined by $U\phi^k = e^k$ fulfills $A_W = U^{-1} A U$ and is a weak interwining operator for $A_{(a_n)}$ and $A_{(b_n)}$. For applications given in [15] we need the following additional fact about U.

Corollary 2. *If $P_E U = U P_E$ for a subset $E \subsetneqq \mathbb{N}$, then $P_E = 0$ or $P_E = I$. Here P_E will denote the projection in l_2 defined by $(P_E \phi)_n = \phi_n$ if $n \in E$ and $(P_E \phi)_n = 0$ if $n \notin E$.*

Proof. Let λ be an eigenvalue of A_W. By Sect. 2, λ is of multiplicity one and a corresponding eigenvector is $\psi=(I-V)\zeta+\xi+e^{i\vartheta}V\xi$ where ζ is defined by (8). A little calculation gives

$$\psi_n=\frac{\xi_n(1+e^{i\vartheta})(a_n-i)}{(a_n-\lambda)}. \tag{15}$$

Recall that $\xi_n\neq 0$ for all $n\in\mathbb{N}$ (because of $a_r\neq b_s$ $\forall r,s\in\mathbb{N}$) and $W\equiv e^{i\vartheta}\neq -1$. Hence, by (15), $\psi_n\neq 0$ $\forall n\in\mathbb{N}$ and $(\phi^k)_n\neq 0$ $\forall k,n\in\mathbb{N}$. This immediately implies the assertion. □

4.2

Remarks. 1. Let us keep on the notations as in the proof of Theorem 1. Suppose $\{a_n\}\underset{1}{\sim}\{b_n\}$. By the considerations in Sect. 2, the sequences (a_n) and (b_n) are (up to a permutation) already determined by the sequences $(|\xi_n|)$ and the real number K_ϑ. Conversely, the preceding proof shows that the sequence $(|\xi_n|)$ and the number K_ϑ, hence $e^{i\vartheta}$, are uniquely determined by the sequences (a_n) and (b_n). Moreover, if (a_n) and (b_n) are given, then all possible deficiency vectors ξ, $\|\xi\|=1$, are precisely $(\xi_n=e^{i\vartheta_n}|\xi_n|)$ where (ϑ_n) is an arbitrary real sequence. Thus, the restriction $A\restriction\mathcal{D}$ is unique up to a unitary equivalence with $A_{(e^{i\vartheta_n})}$. Further, we see that the spectra $\sigma(A_{W'})$ for all unitary operators $W'=e^{i\vartheta'}$, ϑ' real, are completely determined by $\{a_n\}$ and $\{b_n\}$. For Sturm-Liouville operators in the limit-point case this reflects the well-known fact [6, 10] that two spectra determine all other spectra obtained by changing the boundary condition.

2. The assumption $a_n\neq b_k$ $\forall k,n\in\mathbb{N}$ is no loss of generality. In the general case let $\{a_{n_r}, r=1,\ldots,m\}$, $m\in\mathbb{N}$ or $m=+\infty$, be the subset of all terms in (a_n) for which there is a $k_r\in\mathbb{N}$ so that $a_{n_r}=b_{k_r}$. Then, $\{a_n\}\underset{1}{\sim}\{b_n\}$ if and only if the sets $\{a_n\}\backslash\{a_{n_r}\}$ and $\{b_n\}\backslash\{b_{k_r}\}$ are 1-related. To prove this, put $\xi_{n_r}=0$ for all r (see the remark in 2.2).

4.3

Corollary 3. *Let (a_n), (b_n) as in Theorem 1. Suppose that $a_n<b_n<a_{n+1}$ $\forall n\in\mathbb{N}$ and* $\left(\frac{b_n-a_n}{a_n}\right)\notin l_1$.

Let $G=\overline{\lim_{x\to+\infty}}\sum_{n=1}^{\infty}\frac{(b_n^2-a_n^2)x}{(b_n^2+x)(a_n^2+x)}$.

If $G<1$, then $\{a_n\}\underset{1}{\sim}\{b_n\}$. For $G>1$ the sets $\{a_n\}$, $\{b_n\}$ are not 1-related.

In particular, the condition $\left(\frac{b_n^2-a_n^2}{b_n^2a_n^2}\right)\in l_1$ *is necessary for* $\{a_n\}\underset{1}{\sim}\{b_n\}$.

Proof. First we note that $\left(\frac{b_n^2-a_n^2}{b_n^2a_n^2}\right)\notin l_1$ is equivalent to

$$\sum_{n=1}^{\infty}\log\frac{b_n^2(a_n^2+x)}{a_n^2(b_n^2+x)}=+\infty$$

for $x \neq 0$. Hence, if $\left(\frac{b_n^2 - a_n^2}{b_n^2 a_n^2}\right) \notin l_1$, then $G = +\infty$ and $\{a_n\}$, $\{b_n\}$ are not 1-related by Theorem 1 (c).

Assume now $\left(\frac{b_n^2 - a_n^2}{b_n^2 a_n^2}\right) \in l_1$. Then,

$$\frac{d}{dx}\left(\sum_{n=1}^{\infty} \log \frac{b_n^2(a_n^2 + x)}{a_n^2(b_n^2 + x)}\right) = \sum_{n=1}^{\infty} \frac{b_n^2 - a_n^2}{(b_n^2 + x)(a_n^2 + x)}.$$

Using this, the assertion follows from the l'Hospital rule. □

Corollary 4. *Let (a_n), (b_n) be real sequences with $\sup_n a_n = \sup_n b_n = +\infty$. Suppose that $a_{k_1} < b_{m_1} < a_{k_2} < b_{m_2} < \ldots < b_{m_r}$ for natural numbers*

$$n_1 < n_2 < \ldots < n_r, \quad m_1 < m_2 < \ldots < m_r.$$

Then, there exist subsequences $(c_n = a_{k_n}, n \in \mathbb{N})$ and $(d_n = b_{m_n}, n \in \mathbb{N})$ of (a_n) resp. (b_n) so that

$$a_{k_n} < b_{m_n} < a_{k_{n+1}} \quad \forall n \in \mathbb{N}$$

and $\{c_n\} \underset{1}{\sim} \{d_n\}$.

Proof. By induction we choose subsequences $(c_n = a_{k_n})$ and $(d_n = b_{m_n})$ of (a_n) resp. (b_n) satisfying the following conditions:

i) $2 < 2c_n < d_n < c_{n+1}$,

ii) $c_n d_n f_k (c_k^2 + c_n d_n)^{-1} (d_k^2 + c_n d_n)^{-1} \leqq 2^{-k}$,

iii) $c_k d_k f_n (c_n^2 + c_k d_k)^{-1} (d_n^2 + c_k d_k)^{-1} \leqq 2^{-n}$

for all $k, n \in \mathbb{N}$, $k = r+1, \ldots, n$, $n \geqq r+1$, where $f_l := b_l^2 - a_l^2$. Because of $\sup_n a_n$ $\sup_n b_n = +\infty$, such subsequences exist. Now we check that $\{c_n\} \underset{1}{\sim} \{d_n\}$. *i)* implies $\left(\frac{d_n - c_n}{c_n}\right) \notin l_1$. Hence, by the preceding corollary, it is sufficient to show that

$$f(x) := \sum_{n=r+1}^{\infty} \frac{f_n x}{(c_n^2 + x)(d_n^2 + x)} \leqq \frac{1}{2}$$

for large x.

Let $x \geqq c_{r+1} d_{r+1}$. Since the sequence $(c_n d_n, n \geqq r+1)$ is monotonic, we can choose $k \in \mathbb{N}$, $k \geqq r+1$, such that $c_k d_k \leqq x \leqq c_{k+1} d_{k+1}$. Because the function $\frac{x}{(a+x)(b+x)}$, $a > 0$, $b > 0$, $x > 0$, is increasing for $x \leqq \sqrt{ab}$ and decreasing for $x \geqq \sqrt{ab}$, we obtain $f(x) \leqq 1/2$ by splitting the sum into $\sum_{n=r+1}^{k} \ldots + \sum_{n=k+1}^{\infty} \ldots$ and applying ii) and iii). This completes the proof of the corollary.

4.4

Theorem 5. *Suppose* (a_n) *and* (b_n) *are arbitrary unbounded real sequences. Then*

(i) $\{a_n\} \sim \{b_n\}$ *if and only if* $\sup_n a_n = \sup_n b_n = +\infty$ *or* $\inf_n a_n = \inf_n b_n = -\infty$.

(ii) *There exist a real (unbounded) sequence* (c_n) *with* $\{a_n\} \sim \{c_n\}$ *and* $\{c_n\} \sim \{b_n\}$.

Proof. First we prove (i). Assume that $\sup a_n = \sup b_n = +\infty$. By the preceding corollary, there exist subsequences (c_{1n}) of (a_n), (d_{1n}) of (b_n) with $\{c_{1n}\} \underset{1}{\sim} \{d_{1n}\}$, $c_{11} = a_1$, $d_{11} = b_1$, so that the remainder sequences $(a_n)\backslash(c_{1n})$ and $(b_n)\backslash(d_{1n})$ still have infinite suprema. Applying corollary 4 to these remainder sequences and using a simple induction argument, we see that (a_n) and (b_n) are disjoint unions of countably many subsequences (c_{kn}) resp. (d_{kn}) such that $\{c_{kn}\} \underset{1}{\sim} \{d_{kn}\}$ for all $k \in \mathbb{N}$. Since

$$A_{(a_n)} = \sum_k \oplus A_{(c_{kn})}, \qquad A_{(b_n)} = \sum_k \oplus A_{(d_{kn})},$$

this leads to $\{a_n\} \sim \{b_n\}$. To verify this, we only have to check that $\mathscr{D}_\infty(\tilde{A}_{(a_n)})$ is dense in l_2 for the corresponding restriction of $A_{(a_n)}$. Because the restriction $\tilde{A}_{(c_{kn})}$ of $A_{(c_{kn})}$ has deficiency indices (1, 1), the subspace $\sum_k \oplus \mathscr{D}_\infty(\tilde{A}_{(c_{kn})})$ of $\mathscr{D}_\infty(\tilde{A}_{(a_n)})$ is dense in l_2 according to Proposition 2.2. In a similar way the case $\inf_n a_n = \inf_n b_n = -\infty$ will be treated.

Suppose now that the condition in i) is not true, say $a_n \geqq c$, $b_n \leqq d$ for all $n \in \mathbb{N}$. Assume that $\{a_n\} \sim \{b_n\}$, i.e. $A_{(a_n)}\phi = U^{-1} A_{(\sigma(b_n))} U\phi$ on a dense invariant domain $\mathscr{D}$. Then,

$$c\|\phi\|^2 \leqq \langle A_{(a_n)}\phi, \phi\rangle \leqq d\|U\phi\|^2 = d\|\phi\|^2 \quad \forall \phi \in \mathscr{D}.$$

Therefore, $A_{(a_n)} \restriction \mathscr{D}$ is a bounded operator which implies that (a_n) is a bounded sequence. Similarly, (b_n) is bounded which contradicts the assumption of the theorem. This finishes the proof of i).

ii) follows immediately from i) by taking an arbitrary real sequence (c_n) with $\sup_n c_n = +\infty$ and $\inf_n c_n = -\infty$.

Remark. The author conjectures that the assertions of Theorem 5 are also valid in the general case. More precisely, let A and B be unbounded selfadjoint operators in a separable Hilbert space $\mathscr{H}$. Let $\mathscr{S}_A$ resp. $\mathscr{S}_B$ be the intersection of the unit sphere in $\mathscr{H}$ with $\mathscr{D}(A)$ resp. $\mathscr{D}(B)$.

Conjecture. $A \sim B$ *if and only if*

$$\sup_{\phi \in \mathscr{S}_A} \langle A\phi, \phi\rangle = \sup_{\phi \in \mathscr{S}_B} \langle B\phi, \phi\rangle = +\infty$$

or

$$\inf_{\phi \in \mathscr{S}_A} \langle A\phi, \phi\rangle = \inf_{\phi \in \mathscr{S}_B} \langle B\phi, \phi\rangle = -\infty.$$

The necessity part is obvious. An affirmative answer would imply that two arbitrary unbounded selfadjoint operators (in a separable Hilbert space) are related after at most two steps.

4.5

Example. Let $a_n=n$, $b_n=n+\alpha n^\beta$ for $n=1,2,\ldots$, where $\alpha,\beta\in\mathbb{R}_1$, $\alpha>0$. Then, $\{a_n\}\underset{1}{\sim}\{b_n\}$ if and only if $\alpha<1$ and $\beta\leqq 0$ or $\alpha=1$ and $\beta<0$.

Proof. Suppose that $\{a_n\}\underset{1}{\sim}\{b_n\}$. Then the sequences must alternate (at least after new enumeration and leaving out equal terms). This leads to $\alpha<1$, $\beta\leqq 0$ or $\alpha=1$, $\beta<0$.

Now suppose that $\alpha<1$ and $\beta\leqq 0$. Then, (a_n) and (b_n) alternate and $\left(\frac{b_n-a_n}{a_n}\right)\notin l_1$. Let us use the notation

$$H_\gamma=\overline{\lim_{x\to+\infty}}\sum_{n=1}^{\infty}\frac{n^\gamma x}{(n^2+x)^2}\quad\text{for}\quad\gamma\in\mathbb{R}_1.$$

We have

$$G\equiv\overline{\lim_{x\to+\infty}}\sum_{n=1}^{\infty}\frac{(b_n^2-a_n^2)x}{(b_n^2+x)(a_n^2+x)}\leqq 2\alpha H_{\beta+1}+\alpha^2 H_{2\beta}.$$

Using the Euler-Maclaurin sum formula, it can be shown that $H_1\leqq 1/2$ [2]. (In fact, it is $H_1=1/2$.) This immediately implies that $H_{2\beta}=0$ and $G\leqq\alpha<1$. Therefore, by Corollary 3, $\{a_n\}\underset{1}{\sim}\{b_n\}$. In the case $\alpha=1$ and $\beta<0$ we have $H_{\beta+1}=H_{2\beta}=0$, hence $G=0$, and the assertion follows in a similar way.

References

1. Achieser, N.I., Glasmann, I.M.: Lineare Operatoren im Hilbert-Raum. Berlin: Akademie-Verlag 1977
2. Aronszajn, N.: On a problem of Weyl in the theory of singular Strum-Liouville equations. Am. J. Math. **79**, 597–610 (1957)
3. Berg, L.: Über eine Abschätzung von Mathieu, Math. Nachr. **7**, 257–259 (1952)
4. Donoghue, W.F.: On the perturbation of spectra. Comm. Pure Appl. Math. **18**, 559–579 (1965)
5. Donoghue, W.F.: Monotone matrix functions and analytic continuation. Berlin, Heidelberg, New York: Springer 1974
6. Gelfand, I.M., Levitan, B.M.: Determination of a differential equation from its spectral function. Izv. Akad. Nauk SSSR, Ser. Mat. **15**, 309–360 (1951)
7. Jörgens, K., Rellich, F.: Eigenwerttheorie gewöhnlicher Differentialgleichungen. Berlin, Heidelberg, New York: Springer 1976
8. Kato, T.: Perturbation theory for linear operators. Berlin, Heidelberg, New York: Springer 1966
9. Lewin, B.J.: Nullstellenverteilung ganzer Funktionen. Berlin: Akademie-Verlag 1962
10. Levitan, B.M., Gasimov, M.G.: Determination of a differential equation by two spectra. Usp. Mat. Nauk **116**, 3–63 (1964)

11. von Neumann, J.: Zur Theorie der unbeschränkten Matrizen. J. reine angew. Math. **161**, 208–236 (1929)
12. Naimark, M.A.: On the square of a closed symmetric operator. Dokl. Akad. Nauk SSSR **26**, 863–867 (1940); **28**, 207–208 (1940)
13. Richter, P.: Unitary representations of countably infinite dimensional Lie groups. Preprint, Leipzig, 1977
14. Riesz, F., SZ.-Nagy, B.: Vorlesungen über Funktionalanalysis. Berlin: Deutscher Verlag der Wissenschaften 1956
15. Schmüdgen, K.: On the Heisenberg commutation relation $PQ-QP=iI$ (to appear)
16. Weyl, H.: Über gewöhnliche Differentialgleichungen mit Singularitäten. Math. Ann. **68**, 222–269 (1910)
17. Wolf, F.: Perturbation by changes of one-dimensional boundary condition. Indag. Math. **18**, 360–366 (1956)

Received August 12, 1980

Math. Ann. 256, 249–254 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Das formale Prinzip und Modifikationen komplexer Räume

Siegmund Kosarew

Fakultät für Mathematik der Universität, Universitätsstraße 31, D-8400 Regensburg, Bundesrepublik Deutschland

Einleitung

Sei X eine kompakte komplexe Untermannigfaltigkeit der komplexen Mannigfaltigkeit Y. In [3] hat dann Grauert gezeigt, daß die konvergente Umgebungsstruktur von X in Y schon durch die formale Komplettierung $\hat{Y}$ von Y längs X festgelegt wird, falls X exzeptionell in Y ist, d. h. auf einen Punkt niedergeblasen werden kann. Genauer bedeutet dies folgendes:

Für jede weitere holomorphe abgeschlossene Einbettung $\sigma: X \hookrightarrow Z$ von X in eine komplexe Mannigfaltigkeit Z mit $\hat{Y} \cong \hat{Z}$ gibt es, jeweils nach Schrumpfen längs X, eine biholomorphe Abbildung $\varphi: Y \to Z$ mit $\varphi(X) = X$.

In der vorliegenden Arbeit soll diese Aussage auf den Fall komplexer Räume verallgemeinert werden. Wir benutzen dabei die Sprechweise: Für eine (abgeschlossene) holomorphe Einbettung $\tau: X \hookrightarrow Y$ gilt das formale Prinzip, falls man für jede weitere Einbettung $\sigma: X \hookrightarrow Z$, jeden mit τ und σ verträglichen formalen Isomorphismus $\hat{f}: \hat{Y} \to \hat{Z}$ bis zu einer beliebig vorgegebenen Ordnung durch einen konvergenten approximieren kann (nach eventuellem Schrumpfen längs X). Zweckmäßigerweise untersucht man im Hinblick auf obiges Problem die Gültigkeit des formalen Prinzips in Zusammenhang mit Modifikationen f von Y in X

$$\begin{array}{ccc} X & \xrightarrow{\tau} & Y \\ {\scriptstyle f|X}\downarrow & & \downarrow{\scriptstyle f} \\ X^* & \xrightarrow{\tau^*} & Y^* \end{array}$$

f heißt dabei Modifikation von Y längs X, falls f eigentlich ist, außerhalb von X und X^* einen Isomorphismus induziert und X bzw. X^* analytisch dünn [2, 0.43] in Y bzw. Y^* sind. Dabei haben wir $\tau(X)$ mit X identifiziert, was wir in Zukunft stets beibehalten werden. Das Hauptresultat formuliert sich dann folgendermaßen

Theorem. *Sei $f: (Y, X) \to (Y^*, X^*)$ eine Modifikation komplexer Räume, $Y_1^* \subset Y^*$ relativ kompakt und offen, $Y_1 := f^{-1}(Y_1^*)$, $X_1^* := X^* \cap Y_1^*$ und $X_1 := X \cap Y_1$. Dann gilt das formale Prinzip für die Einbettung $X_1 \subset Y_1$ genau dann, wenn es für die Einbettung $X_1^* \subset Y_1^*$ richtig ist.*

0025-5831/81/0256/0249/$01.20

Der Artinsche Approximationssatz liefert sofort das

Corollar 1. *Sei $f:(Y,X)\to(Y^*,X^*)$ eine Modifikation komplexer Räume, $x^*\in X^*$ ein Punkt. Zu jedem formalen Isomorphismus $\hat{f}:\hat{Y}\to\hat{Z}$ und jeder natürlichen Zahl e gibt es eine offene Umgebung W von x^* in Y^*, sowie einen auf $f^{-1}(W)\subset Y$ definierten konvergenten Isomorphismus nach Z, der $\hat{f}$ bis zur Ordnung e approximiert.*

Wir führen ferner die entsprechende Verallgemeinerung von [3, Abschn. 4, Satz 6] an.

Corollar 2. *Ist X eine kompakte, exzeptionelle analytisch dünne Teilmenge von Y, so gilt für die Einbettung $X\subset Y$ das formale Prinzip.*

Beim Beweis des Theorems benötigt man die Sätze von Ancona über formale Dilatationen und Kontraktionen [1], sowie das Chow Lemma [5]. Alle in dieser Arbeit vorkommenden komplexen Räume werden als separiert und von abzählbarer Topologie vorausgesetzt.

1. Einige Hilfssätze über monoidale Transformationen

Wir stellen hier einige Aussagen über monoidale Transformationen komplexer Räume zusammen, die für den Beweis des Theorems benötigt werden und sich nicht explizit in der Literatur finden.

Sei X ein komplexer Raum, $\mathcal{J}\subset\mathcal{O}_X$ ein kohärentes Ideal, Y der durch $\mathcal{J}$ definierte abgeschlossene Unterraum von X. $p:\tilde{X}\to X$ bezeichne die monoidale Transformation von X mit Zentrum Y. Wir setzen $\mathcal{J}:=\mathcal{J}.\mathcal{O}_{\tilde{X}}$.

Satz 1.1. *Sei $e\in\mathbb{N}\setminus\{0\}$ eine natürliche Zahl. Dann induziert jeder X-Automorphismus φ von $\tilde{Y}_e:=(|p^{-1}(Y)|,\mathcal{O}_{\tilde{X}}/\mathcal{J}^{e+1}|_{\tilde{Y}})$ die Identität auf $\tilde{Y}_{e-1}$.*

Beweis. Zunächst gilt

$$\mathcal{O}_{\tilde{Y}_{e-1}}(1)=\mathcal{O}_{\tilde{X}}(1)\otimes\mathcal{O}_{\tilde{X}}\mathcal{O}_{\tilde{Y}_{e-1}}\cong\mathcal{J}/\mathcal{J}^{e+1}$$

und für $\varphi^{\#}:\varphi^{-1}\mathcal{O}_{\tilde{Y}_e}\xrightarrow{\sim}\mathcal{O}_{\tilde{Y}_e}$, $\varphi^{\#}(\varphi^{-1}(\mathcal{J}/\mathcal{J}^{e+1}))=\mathcal{J}/\mathcal{J}^{e+1}$ also

$$\varphi_{e-1}^*\mathcal{O}_{\tilde{Y}_{e-1}}(1)\cong\mathcal{O}_{\tilde{Y}_{e-1}}(1). \qquad (*)$$

Da die Aussage insbesondere lokal in X ist, kann man annehmen, daß $\mathcal{J}=\sum_{i=0}^{n}a_i\mathcal{O}_X$ und $\tilde{X}\subset X\times\mathbb{P}^n$

$$\begin{array}{ccc}\tilde{Y}_{e-1} & \xrightarrow{\varphi_{e-1}} & \tilde{Y}_{e-1}\underset{j}{\subset}\tilde{X}\subset X\times\mathbb{P}^n\\ & \searrow\quad\swarrow & \\ & X & \end{array}$$

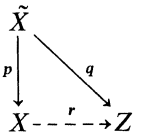
Die Einbettung $j:\tilde{Y}_{e-1}\hookrightarrow X\times\mathbb{P}^n$ wird durch die globalen Schnitte $\xi_0,\ldots,\xi_n\in\Gamma(Y_{e-1},\mathcal{O}_{\tilde{Y}_{e-1}}(1))$ gegeben, wobei $\xi_i\equiv p^*(a_i)\bmod\mathcal{J}^{e+1}$. Die Kommutativität des letzten Diagramms impliziert

$$\xi_i\equiv\varphi_{e-1}^*(p^*(a_i))\bmod\mathcal{J}^{e+1}$$

und somit $\xi_i = \varphi_{e-1}^* \xi_i$. Es folgt daher $j = j \circ \varphi_{e-1}$ und damit auch die Behauptung.

Sei $q : X \to Z$ ein eigentlicher Morphismus komplexer Räume und $Z' \subset Z$ ein abgeschlossener Unterraum definiert durch das kohärente Ideal $\mathcal{J}_{Z'}$ mit $\mathcal{J}_{Z'} \mathcal{O}_{\tilde{X}} = \mathcal{J}$.

Wir betrachten folgendes Diagramm



Es gilt dann folgende Aussage:

Satz 1.2. *Sei $Q \subset Y$ offen und relativ kompakt. Dann existiert ein $l_0 = l_0(Q, p)$ mit folgender Eigenschaft:*

Jeder Morphismus $r : Q_l \to Z'_l$ auf den l-ten infinitesimalen Umgebungen mit $r \circ p_l = q_l$ läßt sich für $l \geqq l_0$ auf eindeutige Weise zu einem konvergenten Morphismus $r' : X' \to Z$ mit $r' \circ p = q$ fortsetzen. Dabei bezeichne $X' \subset X$ eine geeignete offene Teilmenge mit $X' \cap Y = Q$.

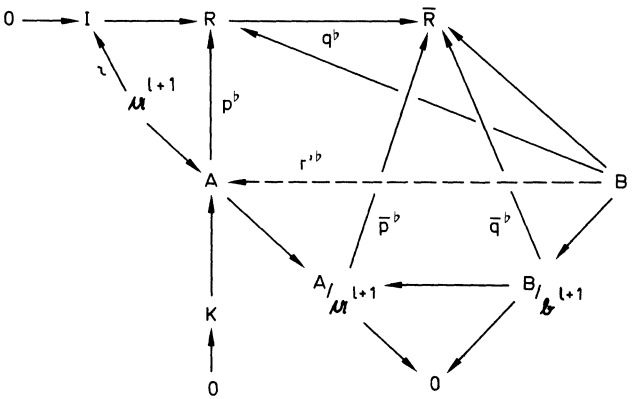
Beweis. Zunächst betrachten wir die exakte Sequenz

$$0 \longrightarrow \mathcal{K} \longrightarrow \mathcal{O}_X \xrightarrow{p^\flat} p_* \mathcal{O}_{\tilde{X}}.$$

Da p außerhalb von Y ein Isomorphismus ist und $\mathcal{K}$ kohärenter $\mathcal{O}_X$-Modul, gilt $\mathcal{J}^{l_1} \mathcal{K} = 0$ und somit nach Artin-Rees auch $\mathcal{J}^{l_0} \cap \mathcal{K} = 0$, jeweils längs Q. Mit Hilfe von [4, EGA III, Théorème 2.3.1] kann man annehmen, daß

$$p^\flat : \mathcal{J}^{l_0} \to p_*(\mathcal{J}^{l_0})$$

längs Q ein Isomorphismus ist. Sei also $l \geqq l_0$ und r wie im Satz. Es genügt die eindeutige Fortsetzung von r lokal um jeden Punkt x von Q zu konstruieren. Wir führen dazu einige Abkürzungen ein: $z := r(x)$, $A := \mathcal{O}_{X,x}$, $B := \mathcal{O}_{Z,z}$, $K := \mathcal{K}_x$, $\mathfrak{a} := \mathcal{J}_x$, $\mathfrak{b} := \mathcal{J}_{Z',z}$, $R := (p_* \mathcal{O}_{\tilde{X}})_x$, $\bar{R} := (p_* \mathcal{O}_{\tilde{X}_l})_x$, $I := (p_*(\mathcal{J}^{l+1}))_x$. Es ist dann folgendes Diagramm auf eindeutige Weise durch den gestrichelten Pfeil zu ergänzen, daß alle Teildiagramme kommutieren.



Da $K \cap \mathfrak{a}^{l+1} = 0$ folgt die Eindeutigkeit. Sei nun $b \in B$. Wir zeigen zuerst $q^\flat(b) =: s \in R$ liegt im Bild von $p^\flat$. Es existiert ein $\bar{a} \in A/\mathfrak{a}^{l+1}$ mit $\bar{p}^\flat(\bar{a}) = \bar{s} \in \bar{R}$, wobei $\bar{s}$ das Bild von s in $\bar{R}$ bezeichne. Also folgt $p^\flat(a) \equiv q^\flat(b) \bmod I$. Wegen der Surjektivität von $\mathfrak{a}^{l+1} \to I$ gilt $p^\flat(a) = q^\flat(b) + p^\flat(a_1)$ mit $a_1 \in \mathfrak{a}^{l+1}$. Man setze jetzt $r'^\flat(b) := a - a_1$. Die Abbildung $r'^\flat$ ist dann wegen $K \cap \mathfrak{a}^{l+1} = 0$ ein wohldefinierter $\mathbb{C}$-Algebrenhomomorphismus. Die Behauptung des Satzes ist damit bewiesen.

Bemerkung 1.3. Insbesondere gilt in dem Falle, wo q die monoidale Transformation von Z mit Zentrum Z' ist, die entsprechende Aussage des Satzes für Isomorphismen, wobei allerdings l_0 noch von Z abhängig ist.

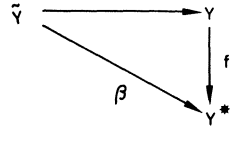
Satz 1.4. *Sei $p : \tilde{X} \to X$ eine eigentliche Modifikation von X längs der analytischen Teilmenge $Y \subset X$ und $\mathscr{I} \subset \mathscr{O}_X$ ein kohärentes Ideal mit $V(\mathscr{I}) = Y$.*

Behauptung. Ist die formale Komplettierung $\hat{p}$ die (formale) monoidale Transformation von $\mathscr{I}\mathscr{O}_{\hat{X}}$, so ist p die monoidale Transformation von $\mathscr{I}$ in der Kategorie der komplexen Räume.

Beweis[1]. Der $\mathscr{O}_{\hat{X}}$-Modul $\mathscr{I}.\mathscr{O}_{\hat{X}}$ ist invertierbar. Da für $a \in p^{-1}(Y)$ der kanonische Morphismus $\mathscr{O}_{\tilde{X},a} \to \mathscr{O}_{\hat{X},a}$ treuflach ist folgt die Invertierbarkeit von $\mathscr{I}.\mathscr{O}_{\tilde{X}}$ als $\mathscr{O}_{\tilde{X}}$-Modul. Ist $q : X' \to X$ die monoidale Transformation von X mit Zentrum $\mathscr{I}$, so existiert ein X-Morphismus $h : \tilde{X} \to X'$, dessen formale Komplettierung $\hat{h}$ ein Isomorphismus ist. Dann muß aber schon h ein Isomorphismus gewesen sein.

2. Beweis des Theorems

Sei $f : (Y, X) \to (Y^*, X^*)$ wie im Theorem vorgegeben. f ist dann insbesondere bimeromorph (vgl. [2, 4.9]). Wir zeigen zunächst eine Richtung des Theorems: Gilt das formale Prinzip für $X_1^* \subset Y_1^*$, so gilt es auch für $X_1 \subset Y_1$. Wegen der Übersichtlichkeit der Schreibweise führen wir für die Beschränkung auf solche relativ kompakten Teilmengen keine neuen Bezeichnungen ein und lassen daher den Index "$_1$" fortan weg. Nach dem Chow Lemma (vgl. [5, Corollary 2] – die Reduziertheit der Basis ist dabei überflüssig) gibt es ein kommutatives Diagramm komplexer Räume



(I)

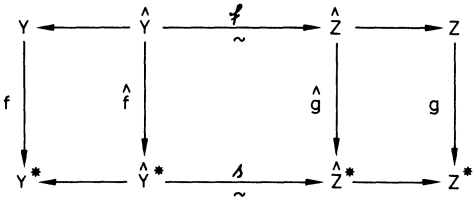
wobei α und β lokal in Y monoidale Transformationen sind mit Zentrum M bzw. M^*, und $M_{\text{red}} = X_{\text{red}}$ bzw. $M^*_{\text{red}} = X^*_{\text{red}}$ gilt.

Sei nun $\sigma : X \hookrightarrow Z$ eine weitere holomorphe Einbettung von X und $\mathfrak{f} : \hat{Y} \to \hat{Z}$ ein formaler Isomorphismus.

---

1 Mit Benutzung einer Anmerkung des Referenten

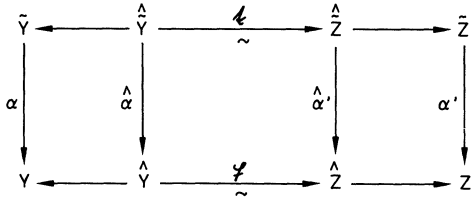
Das Théorème C von [1] liefert die Existenz des folgenden Diagramms



(II)

wobei g eine Modifikation von Z längs X ist und $\mathfrak{s}: \hat{Y}^* \to \hat{Z}^*$ ein formaler Isomorphismus.

Wendet man nun den Satz über die Analytisierung von Dilatationen an (siehe [1, Théorème D]), so ergibt sich



(III)

Dabei ist α' eine Modifikation von Z in X und $\mathfrak{t}: \tilde{\hat{Y}} \to \tilde{\hat{Z}}$ ein formaler Isomorphismus. Wir setzen $\beta' := g \circ \alpha'$. Wegen 1.4 ist (lokal in Z^*) β' die monoidale Transformation mit Zentrum $\mathfrak{s}(M^*)$, entsprechend α' die monoidale Transformation mit Zentrum $\mathfrak{f}(M)$.

Sei nun $e \in \mathbb{N}$ eine natürliche Zahl. Nach Voraussetzung können wir $\mathfrak{s}$ bis zu einer Ordnung e_1 durch einen konvergenten Isomorphismus approximieren (nach eventuellem Schrumpfen längs X^*), welchen man zu einem Isomorphismus $\phi: \tilde{Y} \to \tilde{Z}$ liften kann. Nach 1.1 existiert nun eine Ordnung e_2, so daß $\hat{\phi}$ und $\mathfrak{t}$ auf der e_2-ten infinitesimalen Umgebung von $\alpha^*(X)$ in $\tilde{Y}$ übereinstimmen. Ist e_2 genügend groß, so kann ϕ zu einem Isomorphismus $f: Y \to Z$ (wieder nach Schrumpfen längs X) runtergedrückt werden und zwar auf eindeutige Weise. ϕ und $\mathfrak{f}$ stimmen dabei bis zu einer Ordnung e_3 überein. Ist nun e_1 genügend groß gewählt worden, so kann man erreichen, daß $e_3 \geqq e$ gilt, wie man leicht aus der Konstruktion von e_2 und e_3 erkennt. Diese Richtung des Theorems ist damit bewiesen. Die andere Richtung folgt durch eine analoge Argumentation:

Sei $\mathfrak{s}: \hat{Y}^* \to \hat{Z}^*$ ein formaler Isomorphismus mit $\sigma^* = \mathfrak{s} \circ \tau^*$. Es existieren dann nach dem Satz über die Analytisierung von Dilatationen die Diagramme (II) und (III). Man approximiere jetzt $\mathfrak{f}$ bis zu einer Ordnung e_1 durch einen konvergenten Isomorphismus $f: Y \to Z$, was nach Voraussetzung möglich ist und lifte f zu einem Isomorphismus $\phi: \tilde{Y} \to \tilde{Z}$. Wegen 1.2 kann man auf eindeutige Weise ϕ zu einem Isomorphismus $s: Y^* \to Z^*$ runterdrücken, falls e_1 genügend groß war. s und $\mathfrak{s}$ müssen dann bis zu einer Ordnung e_2 übereinstimmen. Durch geeignete Wahl von e_1 kann man $e_2 \geqq e$ erreichen. Damit ist das Theorem vollständig bewiesen.

Danksagung. Meinen ausdrücklichen Dank möchte ich hier noch Herrn Prof. Dr. Jürgen Bingener für seine stetige Unterstützung und für die zahlreichen Ratschläge aussprechen.

Literatur

1. Ancona, V., Tomassini, G.: Théorèmes d'existence pour les modifications analytiques. Invent. Math. **51**, 271–286 (1979)
2. Fischer, G.: Complex analytic geometry. In: Lecture Notes in Mathematics, Vol. 538. Berlin, Heidelberg, New York: Springer 1976
3. Grauert, H.: Über Modifikationen und exzeptionelle analytische Teilmengen. Math. Ann. **146**, 331–368 (1962)
4. Grothendieck, A., Dieudonné, J.: Elements de géométrie algébrique. Publ. Math. I.H.E.S. **4**, **8**, **11**, **17**, **20**, **24**, **28**, **32** (1960–1967)
5. Hironaka, H.: Flattening in complex-analytic geometry. Am. J. Math. **97**, 503–547 (1975)

Eingegangen am 15. August 1980

Math. Ann. 256, 255–276 (1981)

Mathematische Annalen
© Springer-Verlag 1981

On 4-Manifolds with Singular Torus Actions

Paul Melvin[1]

Department of Mathematics, University of California, Santa Barbara, CA 93106, USA

This paper studies the homeomorphism classification problem for closed oriented 4-manifolds which support effective codimension-two torus actions. With suitable restrictions on the actions a classification is obtained in terms of numerical invariants computable from orbit data. These invariants are related to the fundamental groups and the cohomology rings of the underlying manifolds. They do not distinguish among all 4-manifolds with actions as above, and so the general classification problem remains open.

The focus here is on actions with some singular orbits (orbits of codimension greater than two), refered to as *singular actions*. This continues the investigations of Orlik, Raymond, and Pao. A classification in the presence of fixed points was obtained by Pao in [8] following earlier results of Orlik and Raymond [5, 6] (including a classification in the simply-connected case). Work on the general singular case was begun by Pao in [9]. Much of the present paper was motivated by questions raised there.

The core of the geometry of the paper is in Sect. 2 where various techniques for locally modifying singular actions on 4-manifolds is developed. Section 3 reviews and refines the classification in the fixed point case for use later in the paper.

In the next few sections, various algebraic invariants of a 4-manifold M with a singular torus action are studied. It is known that the *fundamental group* of M is a central cyclic extension of a free product of cyclic groups. Section 4 gives a classification of such groups and consequently a complete set of invariants for $\pi_1 M$. These results are applied in Sect. 5 to clarify the overlap between singular and non-singular actions. In Sect. 6 the *parity* (even or odd) of the intersection form on $H_2 M$ is computed, along with the *characteristic function*, an invariant which measures how much M, if even, deviates from being a spin manifold.

Section 7 applies the methods of Sect. 2 to show that, with suitable restrictions on the action, M is determined by its fundamental group, Euler characteristic, signature, parity and characteristic function. This classification includes, for

1 Partially supported by NSF grant MCS78-01048

0025-5831/81/0256/0255/$04.40

example, any action satisfying one of the following conditions:

(1) The intersection of the isotropy groups of all orbits outside some component of singular orbits is trivial,

(2) the genus of the orbit space is non-zero and $\pi_1 M$ is centerless, or

(3) the intersection of the isotropy groups of some pair of of singular orbits is finite of order $\leqq 2$.

1. Preliminaries

All manifolds in this paper are connected, closed and oriented unless specified otherwise, and all group actions are effective. For convenience we work throughout in the smooth category.

Diffeomorphisms between manifolds are assumed orientation preserving and often will be denoted by $=$.

A manifold M is called a *G-manifold* if it is equipped with an action of a Lie group G. The *orbit space* of the action is denoted by M^*, and the image in M^* of any subset X of M is written X^*. If G is the 2-dimensional torus group

$$T^2 = \{(u, v) \in \mathbb{C}^2 \mid |u| = |v| = 1\}$$

and M is a 4-manifold, then M^* is a 2-manifold, possibly with boundary. Interior points of M^* correspond to *principal orbits* and *E-orbits* (orbits with non-trivial finite cyclic isotropy groups). Boundary points correspond to *singular orbits* (orbits with infinite isotropy), consisting of *C-orbits* (with circle isotropy) and *fixed points*. M is called a *singular T^2-manifold*, or equivalently the action is *singular*, if it has some singular orbits. Otherwise it is called *non-singular*.

We shall often identify a 4-dimensional T^2-manifold by its *weighted orbit space* as defined by Orlik and Raymond [5, 6]. Some comments are in order as our notation differs somewhat from theirs.

The weighted orbit space consists of M^* together with orbit invariants for the C and E-orbits and a characteristic class, defined as follows.

The *orbit invariant* of a C-orbit is the pair (p, q) of relatively prime integers (determined up to sign, i.e. $(p, q) = (-p, -q)$) specifying its isotropy group $G(p, q) \subset T^2$. Here $G(p, q)$ denotes the image of the embedding $S^1 \to T^2$ given by $z \to (z^p, z^q)$. This definition differs from that of Orlik and Raymond in Sect. 1 of [5]: our $G(p, q)$ is their $G(-q, p)$.

For the E-orbits we will use an un-normalized version of the Seifert invariants (Sect. 1.3 in [6]) depending upon the choice of a cross-section

$$q : \partial N^* \to \partial N$$

to the principal T^2-bundle $\partial N \to \partial N^*$, where N denotes an invariant tubular neighborhood of the union of all E-orbits in M. In particular let E_0 be the E-orbit in a component N_0 of N, and let x be a point in ∂N_0. Denote the orbits of x under the restricted actions of $G(1, 0)$ and $G(0, 1)$ by h and k, respectively. Then the image Q_0 of q in ∂N_0 satisfies a homology relation

$$\alpha Q_0 + \beta h + \gamma k = 0$$

in N_0, with $|\alpha|$ equal to the *order* of E_0 (that is the order of its isotropy group). The triple (α, β, γ), determined up to sign, is called the *orbit invariant* of E_0 (associated to q).

The *characteristic class* of M is now defined as the obstruction to extending q over $M^* - N^*$. It is an element (b, c) in

$$H^2(M^*, N^*; \pi_1 T^2) = \begin{cases} 0 & \text{if } M \text{ is singular} \\ \mathbb{Z} \oplus \mathbb{Z} & \text{if } M \text{ is non-singular.} \end{cases}$$

Observe that by altering the section q the identifications of E-orbit invariants

$$(\alpha, \beta, \gamma) = (\alpha, \beta + m\alpha, \gamma + n\alpha) \qquad m, n \in \mathbb{Z}$$

can be made at will in the singular case. In the non-singular case, changing (α, β, γ) to $(\alpha, \beta + m\alpha, \gamma + n\alpha)$ changes the characteristic class (b, c) to $(b - m, c - n)$. In either case, choosing m to be 0 and n to be the product of 1 with all primes which divide β but not γ, one easily obtains the following useful result.

Lemma 1.1. *There is a cross-section q (as above) for which all the associated E-orbit invariants* (α, β, γ) *have* $gcd(\beta, \gamma) = 1$.

Some of the geometric constructions in the next section will involve 4-manifolds with T^3-actions. The reader is referred to Pak [7] for the equivariant classification. As above we associate to any singular orbit with isotropy group $G(p, q, r)$ [defined analogously to $G(p, q)$] the orbit invariant (p, q, r), determined up to sign.

2. Local Modifications

This section will describe various local modifications of singular T^2-actions on 4-manifolds. This is the core of the geometry in this paper. The basic goal is to change a singular action into some sort of canonical form by modifying it on "elementary" pieces of the underlying manifold. Such a program is begun here. A few applications will be given in Sect. 7 using algebraic invariants computed in Sect. 4 and 6.

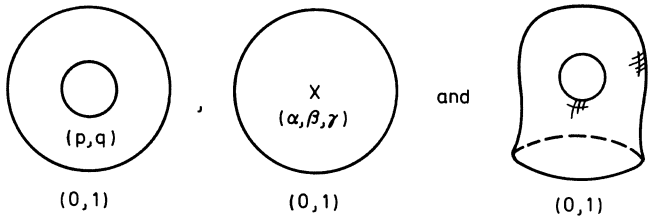
An *equivariant sum* of two T^2-manifolds M_1 and M_2 is formed by removing open tubular neighborhoods of C-orbits of the same isotropy type from each, and identifying the resulting boundaries by an (orientation reversing) equivariant diffeomorphism. To obtain a well-defined operation, the C-orbits along which the sum is taken must be prescribed up to ambient equivariant diffeomorphism. One then has a commutative and associative operation on the category of T^2-manifolds with prescribed C-orbits of a given type. The notation

$$M_1 + M_2$$

(cf. Pao [9]) is reserved for the equivariant sum when the prescribed C-orbits are of type $G(0, 1)$ and are unique up to ambient equivariant diffeomorphism (e.g. when M_1 and M_2 are fixed point free).

In [9] Pao observed (as a consequence of the equivariant classification theorem of Orlik and Raymond [6]) that every singular T^2-manifold is an equivariant sum of certain *elementary* T^2-manifolds.

Definition. Let $Q(p,q)$, $S(\alpha,\beta,\gamma)$ and H be the T^2-manifolds with weighted orbit spaces



respectively. Any T^2-manifold weakly equivalent to some $Q(p,q)$ is called an *elementary T^2-manifold* of *type Q*. Similarly define elementary T^2-manifolds of *type S* and *type H*.

Proposition (Pao). *Every singular T^2-manifold without fixed points is an equivariant sum of elementary T^2-manifolds of types Q, S, and H.*

As is shown below, it is always possible to modify the action on a singular T^2-manifold to eliminate factors of type H. The factors of types Q and S, moreover, are quite well understood. For example, Pao showed that they are all diffeomorphic to the product of a circle with a lens space, and explicitly identified the T^2-manifolds of type Q. One may likewise identify those of type S.

Proposition 2.1. (1) (Pao) $Q(p,q)=S^1\times L(p,q)$.

(2) $S(\alpha,\beta,\gamma)=S^1\times L(\delta,\gamma)$, where $\delta=gcd(\alpha,\beta)$.

Proof. (1) See [9].

(2) It is evident that $S(\alpha,\beta,\gamma)$ supports a T^3-action with weighted orbit space

$(0,0,1)$●——●(α,β,γ).

Reparametrize this action by the automorphism

$$\begin{pmatrix} \beta/\delta & -\alpha/\delta & 0 \\ a & b & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

of T^3, where a and b satisfy $a\alpha+b\beta=\delta$. This new action, with orbit space

$(0,0,1)$●——●$(0,\delta,\gamma)$,

is the product of the S^1 action on S^1 (by multiplication) and the T^2-action on $L(\delta,\gamma)$ with orbit space

$(0,1)$●——●(δ,γ)

(see Sect. 2 in [5]). In particular $S(\alpha,\beta,\gamma)=S^1\times L(\delta,\gamma)$. □

The object of the remainder of this section is to show how to replace certain factors (or combinations of factors) in an equivariant sum decomposition of a singular T^2-manifold, without changing the underlying manifold. Using the proposition above, for example, one might hope to replace $Q(p,q)$ by $Q(p,q')$ for any q' with $qq'\equiv\pm1 \pmod p$ (cf. the classification of lens spaces). Unfortunately

this is not possible in general as the sum operation involves the choice of a framed circle in $Q(p,q)$. Nevertheless much can still be said. The following notation will be useful.

Definition. Let M_1 and M_2 be fixed point free T^2-manifolds which have C-orbits of type $G(0,1)$. Write

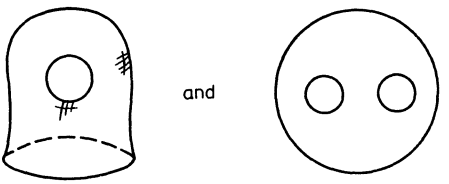
$$M_1 \sim M_2$$

if there is a diffeomorphism $M_1 \to M_2$ which is equivariant near (i.e. on a neighborhood of) some C-orbit of type $G(0,1)$ in M_1.

Observe that if $M_1 \sim M_2$, then $M + M_1$ and $M + M_2$ are diffeomorphic for *any* T^2-manifold M with a prescribed C-orbit of type $G(0,1)$.

Theorem 2.2. $H \sim Q(0,1) + Q(0,1)$.

Proof. Set $P = Q(0,1) + Q(0,1)$. The circle subgroup $G(1,0)$ of T^2 acts freely on both manifolds, H and P. The orbit spaces of these actions are 3-manifolds H_0 and P_0 with $S^1 = G(0,1)$-actions given by the orbit spaces



respectively, that is

$$H_0 = M_{o,1,1,0} \quad \text{and} \quad P_0 = M_{o,0,3,0}$$

in the notation of [10]. In fact

$$H = S^1 \times H_0 \quad \text{and} \quad P = S^1 \times P_0,$$

where the C-orbits in H (or P) are of the form $S^1 \times p$, with p a fixed point in H_0 (or P_0).

Observe that H_0 and P_0 are both diffeomorphic to $S^1 \times S^2 \# S^1 \times S^2$ (by the classification theorem in [10]). Choose a fixed point p in H_0. Any diffeomorphism from H_0 to P_0 is isotopic to one which maps p to a fixed point in P_0, and furthermore is equivariant near p. Crossing with the identity on S^1 gives a diffeomorphism from H to P equivariant near $S^1 \times p$. Thus $H \sim P$. □

An inductive application of this theorem yields

Corollary 2.3. *If a 4-manifold admits any singular T^2-actions, then it admits one with orbit space of genus zero.*

The next theorem deals with factors of types Q and S.

Theorem 2.4. (1) $Q(p,q) \sim Q(p,q')$ *provided* $q \equiv \pm q' \pmod{2p}$.

(2) $S(\alpha,\beta,\gamma) \sim S(\alpha,\beta',\gamma')$ *provided* $\beta = \pm\beta'$ *and* $\gamma \equiv \gamma' \pmod{2\beta}$.

Remark. We suspect that the converse of (1) also holds, but have no proof at present. In contrast (2) may be strengthened. For example one can show that

$S(\alpha,\beta,\gamma)\sim S(\alpha,\beta',\gamma)$ provided $gcd(\alpha,\beta)=gcd(\alpha,\beta')$ ($=\delta$ by definition) and $\beta\equiv\pm\beta'$ (mod $gcd(\alpha,\delta^2)$). We omit the proof.

Proof of 2.4. To prove (1), suppose that $q'=jq+kp$, with $j=\pm1$ and k an even integer.

Extend the T^2-action on $Q(p,q)$ to a T^3-action with weighted orbit space

$$(0,0,1)\bullet\!\!-\!\!-\!\!-\!\!\bullet(0,p,q).$$

The automorphism

$$\begin{pmatrix} j & 0 & 0\\ 0 & 1 & 0\\ 0 & k & j\end{pmatrix}$$

of T^3 alters this action (without changing the underlying manifold). The new orbit space is

$$(0,0,j)=(0,0,1)\bullet\!\!-\!\!-\!\!-\!\!\bullet(0,p,q'),$$

and so there is a diffeomorphism

$$f:Q(p,q)\to Q(p,q')$$

(the identity on the underlying manifold) which preserves C-orbits of type $G(0,1)$. Observe that the automorphism above maps $G(0,1,0)$ to $G(0,1,k)$. As k is even, and there are only two framings for a circle in a 4-manifold, it follows that f is isotopic to a diffeomorphism which is equivariant near some C-orbit of type $G(0,1)$. Thus $Q(p,q)\sim Q(p,q')$.

An analogous argument, starting with the T^3-action

$$(0,0,1)\bullet\!\!-\!\!-\!\!-\!\!\bullet(\alpha,\beta,\gamma)$$

on $S(\alpha,\beta,\gamma)$, establishes (2). □

The next result is a more specialized local modification which will be applied in Sect. 7.

Theorem 2.5. (1) $Q(2,1)+Q(p,q)\sim Q(2,1)+Q(p',q')$ *provided* $(p,q)\equiv(p',q')$ (mod 2).
(2) $Q(2,1)+S(\alpha,\beta,\gamma)\sim Q(2,1)+S(\alpha,\beta',\gamma')$ *provided* $(\beta,\gamma)\equiv(\beta',\gamma')$ or $\alpha\equiv1$ (mod 2).
The proof will be based on two lemmas.

Lemma 2.6. (1) $Q(p,1)+Q(r,s)\sim Q(p,1)+Q(r-ps,s)$.
(2) $Q(p,1)+S(\alpha,\beta,\gamma)\sim Q(p,1)+S(\alpha,\beta-p\gamma,\gamma)$.

Proof. Consider the weak equivalence

$$f:Q(p,1)+Q(r,s)\to Q(-p,1)+Q(r-ps,s)$$

induced by the automorphism

$$\begin{pmatrix} 1 & -p\\ 0 & 1\end{pmatrix}$$

of T^2. Since this automorphism preserves $G(1,0)\subset T^2$, f is $G(1,0)$-equivariant.

Let C be a C-orbit of type $G(0,1)$ in $Q(p,1)+Q(r,s)$. Then $f(C)$ is a C-orbit of type $G(-p,1)$ in $Q=Q(-p,1)+Q(r-ps,s)$. Since the C-orbits in Q of types $G(-p,1)$ and $G(0,1)$ are both principal for the action of $G(1,0)$, there is a $G(1,0)$-equivariant automorphism g of Q carrying $f(C)$ onto a C-orbit of type $G(0,1)$. In fact g can be chosen so that gf is T^2-equivariant near C. Thus

$$Q(p,1)+Q(r,s)\sim Q(-p,1)+Q(r-ps,s).$$

Theorem 2.4 (1) now gives (1). An analogous argument proves (2). □

Lemma. 2.7. *Let M be a T^2-manifold with a prescribed C-orbit C of type $G(0,1)$. If M is a spin manifold and C or its square is null-homotopic in M, then $M+S(\alpha,\beta,\gamma)\sim M+S(\alpha,\beta,-\gamma)$.*

Remarks. (1) The spin condition can be omitted, using the results of Sect. 6.

(2) It is not true in general that $S(\alpha,\beta,\gamma)\sim S(\alpha,\beta,-\gamma)$. For example the two T^2-manifolds $S(3,0,1)+S(3,0,1)$ and $S(3,0,1)+S(3,0,-1)$ are diffeomorphic to $S^1\times(L(3,1)\# L(3,1))$ and $S^1\times(L(3,1)\# L(3,-1))$ respectively (see Sect. 5 in [9]), which are not even homotopy equivalent.

Proof of 2.7. Choose a C-orbit C' of type $G(0,1)$ in $S(\alpha,\beta,\gamma)$. Let N and N' be invariant tubular neighborhoods of C and C' respectively. By definition

$$M+S(\alpha,\beta,\gamma)=(M-\operatorname{int}N)\bigcup_f(S(\alpha,\beta,\gamma)-\operatorname{int}N')$$

with $f:\partial N\to\partial N'$ equivariant.

By the hypothesis on C there is an isotopy h_t of M carrying C onto $-C$. We may assume that the end of this isotopy

$$h_1:M\to M$$

leaves N invariant and is fiber preserving. The restriction of h_t to C gives a map of a Klein bottle into M representing an element κ in $H_2(M;\mathbb{Z}_2)$. Since M is spin, the self-intersection $\kappa\cdot\kappa$ is zero. It follows that h_t can be chosen so that the restriction of h_1 to N is *antiequivariant* [i.e. $h_1(zx)=\bar{z}h_1(x)$ for x in N and z in T^2].

On $S(\alpha,\beta,\gamma)$, with T^3-action as in the proof of Proposition 2.1, consider the weak equivalence

$$g:S(\alpha,\beta,\gamma)\to S(\alpha,-\beta,-\gamma)$$

induced by the automorphism

$$\begin{pmatrix}1&0&0\\0&-1&0\\0&0&-1\end{pmatrix}$$

of T^3. Clearly g is antiequivariant (in the sense above).

Gluing $h_1|(M-\operatorname{int}N)$ and $g|(S(\alpha,\beta,\gamma)-\operatorname{int}N')$ together yields a diffeomorphism

$$M+S(\alpha,\beta,\gamma,)\to M+S(\alpha,-\beta,-\gamma).$$

By general position, h_1 could have been chosen to be the identity near some C-orbit of type $G(0,1)$, and so $M+S(\alpha,\beta,\gamma)\sim M+S(\alpha,-\beta,-\gamma)$. The lemma now follows from Theorem 2.4 (2). □

Proof of 2.5. To prove (1) first use Theorem 2.4 (1) to reduce q so that $0\leqq q\leqq p$. By Lemma 2.6 (1), $Q(2,1)+Q(p,q)\sim Q(2,1)+Q(p-2q,q)$. Set $p'=p-2q$ and apply Theorem 2.4 (1) as above to reduce q, giving

$$Q(2,1)+Q(p,q)\sim Q(2,1)+Q(p',q')$$

with $0\leqq q'\leqq p'$ and $p'<p$. Observe that in the process above the parities of p and q are unchanged, that is $(p,q)\equiv(p',q')\ (\mathrm{mod}\,2)$. The result follows by induction, since p and q are relatively prime.

The proof of (2) is similar. Lemma 2.7 applies since $Q(2,1)=S^1\times\mathbb{R}P^3$ (by Proposition 2.1) is spin and any C-orbit of type $G(0,1)$ is of order two in $\pi_1 Q(2,1)$. Combining this with Theorem 2.4 (2) gives

$$Q(2,1)+S(\alpha,\beta,\gamma)\sim Q(2,1)+S(\alpha,\beta,\gamma')\quad\text{if}\quad\gamma\equiv\pm\gamma'\,(\mathrm{mod}\,2\beta)$$

which is the analogue of Theorem 2.4 (1) in the proof above. The proof now parallels the one above, using (2) of Lemma 2.6 in place of (1), and the fact that β and γ may be assumed relatively prime (Lemma 1.1).

If α is odd a further argument is needed to show that the manifold in question does not depend on β and γ. In particular, using the equivalence displayed above together with Lemma 2.6 (2) we have

$$\begin{aligned}Q(2,1)+Q(\alpha,0,1)&=Q(2,1)+Q(\alpha,\alpha,1)\\&\sim Q(2,1)+Q(\alpha,1,1)\\&\sim Q(2,1)+Q(\alpha,1,\alpha)\\&=Q(2,1)+Q(\alpha,1,0).\end{aligned}$$

This completes the proof of the theorem. □

To conclude this section, we give another consequence of Lemma 2.6.

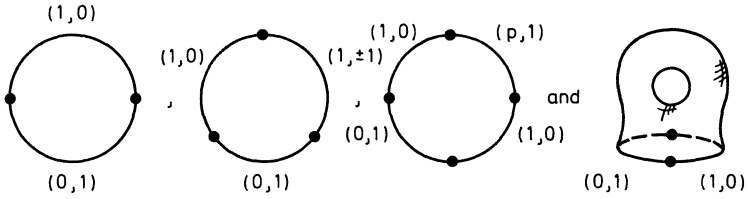
Theorem 2.8. $Q(p,1)+Q(r,1)\sim Q(0,1)+Q(t,1)$, *where* $t=gcd(p,r)$.

Proof. Denote $Q(p,1)+Q(r,1)$ by $M(p,r)$. Lemma 2.6 (1) shows that $M(p,r)\sim M(p,r-p)$. Since $M(p,r)=M(r,p)$, the Euclidean algorithm yields $M(p,r)\sim M(0,t)$, as desired. □

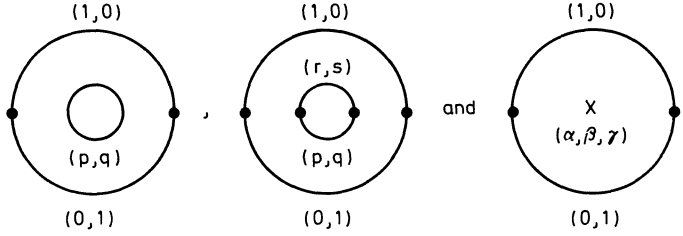
3. The Fixed Point Case Revisited

The topological classification of T^2-manifolds with fixed points, due to Orlik-Raymond and Pao, will be central to our investigation in Sect. 6 of the intersection form of an arbitrary singular T^2-manifold. This section reviews this classification in a form convenient for later reference. The identification of manifolds of type L in Proposition 3.4 is new, and also will be used in Sect. 6.

Definition. Let P_0, $P_\pm$, $P(p)$, and P_1 be the T^2-manifolds with weighted orbit spaces



respectively. Any T^2-manifold weakly equivalent to one of these is called an *elementary T^2-manifold* of *type P*. Similarly define elementary T^2-manifolds of *types R*, *T* and *L* in terms of the T^2-manifolds $R(p,q)$, $T(p,q;r,s)$, and $L(\alpha,\beta,\gamma)$, with orbit spaces



respectively. Set $L_\alpha = L(\alpha,1,0)$ and $L'_\alpha = L(\alpha,1,1)$.

The starting point for the classification of T^2-manifolds with fixed points was the following equivariant classification theorem [5, 6].

Theorem 3.1 (Orlik-Raymond). *Every T^2-manifold with fixed points is an equivariant connected sum of elementary T^2-manifolds of types P, R, T, and L.*

The principal ingredient in the topological classifiction was the identification of the T^2-manifolds of types *P*, *R*, *T*, and *L* [5, 8].

Theorem 3.2. (1) (Orlik-Raymond)

$$P_0 = S^4$$

$$P_\pm = \pm \mathbb{C}P^2$$

$$P(p) = \begin{cases} S^2 \times S^2 & \text{if } p \text{ is even} \\ S^2 \tilde{\times} S^2 (= \mathbb{C}P^2 \# -\mathbb{C}P^2) & \text{if } p \text{ is odd} \end{cases}$$

$$P_1 = 2S^3 \times S^1 \# 2S^2 \times S^2 .$$

(2) (Pao)

$$R(p,q) = \begin{cases} S^3 \times S^1 \# S^2 \times S^2 & \text{if } pq \text{ is even} \\ S^3 \times S^1 \# S^2 \tilde{\times} S^2 & \text{if } pq \text{ is odd} \end{cases}$$

$$T(p,q;r,s) = \begin{cases} S^3 \times S^1 \# 2S^2 \times S^2 & \text{if } pq \text{ and } rs \text{ are even} \\ S^3 \times S^1 \# 2S^2 \tilde{\times} S^2 & \text{if } pq \text{ or } rs \text{ is odd} \end{cases}$$

$$L(\alpha,\beta,\gamma) = L_\alpha \quad \text{or} \quad L'_\alpha .$$

Furthermore, L_α is spin for all α, and L'_α is spin (in fact $L_\alpha = L'_\alpha$) if and only α is odd.

Combining these two theorems with a delicate study of the $\mathbb{Z}_2$-cohomology of L_α and L'_α, Pao completed the topological classification in [8].

Theorem 3.3 (Pao). *Let M be a T^2-manifold with fixed points. If M is even (i.e. every class in H_2M has even self-intersection), then it can be descomposed uniquely as a connected sum of copies of $S^2 \times S^2$, $S^3 \times S^1$, $L_\alpha(\alpha \geqq 2)$ and L'_α (α even and $\geqq 2$). If M is odd, then it can be decomposed uniquely as a connected sum of copies of $\pm\mathbb{C}P^2$, $S^3 \times S^1$, and $L_\alpha(\alpha \geqq 2)$.*

Definition. The connected sum decomposition in the theorem will be called the *normal decomposition* of M.

An explicit identification of the T^2-manifolds of type L will be needed in Sect. 6. In particular, Theorem 3.2 (2) may be strengthened as follows.

Proposition 3.4.

$$L(\alpha,\beta,\gamma)=\begin{cases}L_\alpha & \text{if } \beta\gamma \text{ is even}\\ L'_\alpha & \text{if } \beta\gamma \text{ is odd.}\end{cases}$$

Proof. By Lemma 1.1 we may assume that $gcd(\beta,\gamma)=1$. Observe that

$$L(\alpha,\beta,\gamma)=P_0+S(\alpha,\beta,\gamma)$$

in the notation of Sect. 2. By Theorem 2.4 (2), β can be replaced by its absolute value (if it is negative) and γ can be reduced so that $|\gamma| \leqq \beta$, without altering $gcd(\beta,\gamma)$ or the parity of $\beta\gamma$.

Now the automorphism

$$\begin{pmatrix}-1 & 0\\ 0 & -1\end{pmatrix}$$

of T^2 together with an orientation reversal of the orbit space $L(\alpha,\beta,\gamma)^*$ induces a diffeomorphism between $L(\alpha,\beta,\gamma)$ and $L(\alpha,\gamma,\beta)$. A simple induction yields

$$L(\alpha,\beta,\gamma)=\begin{cases}L(\alpha,1,0) & \text{if } \beta\gamma \text{ is even}\\ L(\alpha,1,1) & \text{if } \beta\gamma \text{ is odd.}\end{cases}\quad\square$$

4. The Fundamental Group

It is known that the fundamental group of any 4-manifold which supports a singular T^2-action is a central cyclic extension of a free product of cyclic groups (see Sect. 3 and 4 in [6] and Sect. 3 in [9]). If it is abelian then it is classified by its invariant factors, easily computable from any presentation. We shall investigate the non-abelian case.

First we give numerical invariants which classify *all* non-abelian groups with the aforementioned structure.

Let G be a non-abelian group. Suppose that there is a central extension

$$1 \to H \xrightarrow{h} G \xrightarrow{q} Q \to 1$$

with

$$H=\mathbb{Z}_m$$

$$Q=*Q_i(i\in I), \quad Q_i=\mathbb{Z}_{n_i}.$$

Assume that $n_i\geqq 2$ for $i\in F\subset I$, and $n_i=0$ otherwise. Set $G_i=q^{-1}Q_i$ (an abelian group since the extension is central) and let k_i be the smallest invariant factor of G_i. That is

$$G_i=\mathbb{Z}_{k_i}\oplus\mathbb{Z}_{mn_i/k_i},\ k_i|mn_i/k_i.$$

Set

$$r=\text{card}(I-F), \text{ the number of } \mathbb{Z} \text{ factors of } Q$$

$$E=\{(n_i,k_i)|i\in F\}.$$

Theorem 4.1. *The non-negative integers m and r and the set E are invariants of G (in the notation above); in particular, m is the order of the center of G. Furthermore, m, r, and E form a complete set of invariants for non-abelian groups which are central cyclic extensions of free products of cyclic groups.*

Proof. Since G is non-abelian, Q is not cyclic and so centerless. Hence $h(H)$ is the entire center of G. It follows that m and the group Q are invariants of G. By the Kurosh subgroup theorem r and the n_i $(i\in F)$ are invariants of Q, and thus of G. Moreover the Q_i $(i\in F)$ are unique up to conjugacy in Q, whence the associated G_i are unique up to conjugacy in G. In particular the groups G_i are invariants of G, yielding the invariance of E.

To prove the second statement we must show that m, r, and E determine G. Evidently G is the free product of the G_i amalgamated along H $(i\in I)$, that is the push out of the set of embeddings

$$h_i: H\to G_i$$

induced by $h: H\to G$.

Thus G is determined by $\{h_i|i\in I\}$ (where h_i is defined up to an automorphism of G_i). As the Q_i are cyclic, an argument using the primary decomposition theorem shows that there is only one embedding of H in G_i with quotient Q_i (up to an automorphism of G_i). Thus h_i is determined by the groups H, Q_i, and G_i, or equivalently by the numerical invariants m, n_i, and k_i.

Hence G is determined by m and $\{(n_i,k_i)|i\in I\}$. Now if $n_i=0$ then $k_i=m$, so r is the number of pairs $(0,m)$ in the set above. The remaining pairs are specified by E. □

The remainder of this section is devoted to computing the invariants defined above for the fundamental group of a singular T^2-manifold M. An explicit presentation of $\pi_1 M$ will be useful.

Lemma 4.2. *Let M be a 4-dimensional singular T^2-manifold whose C-orbits have orbit invariants $(p_1,q_1),\ldots,(p_s,q_s)$ and whose E-orbits have orbit invariants $(\pi_1,\beta_1,\alpha_1),\ldots,(\pi_t,\beta_t,\alpha_t)$. Assume that M has a C-orbit C of type G(0, 1) (that is some*

$(p_i, q_i) = (0, 1)$). *Set* $m = gcd(p_1, \ldots, p_s)$ *and* $r = \operatorname{rank} H_1 M^*$. *Then*

$$\pi_1 M = (h, x_1, \ldots, x_r, y_1, \ldots, y_t | h \text{ is central}, \\ h^m = 1 = y_i^{\alpha_i} h^{\beta_i} (i = 1, \ldots, t)),$$

where h represents the homotopy class of C.

Proof. First assume that M has no fixed points. Let T be an invariant tubular neighborhood of the non-principal orbits in M, and N be the union of all components of T which contain E-orbits. Set $M_0 = \text{closure } (M - T)$. Recall from section 1 that the E-orbit invariants are associated to a particular cross-section $q : \partial N^* \to \partial N$. Extend q across M_0^* to give a product structure $M_0 = M_0^* \times T^2$. Thus

$$M = (M_0^* \times T^2) \cup T.$$

Observe that the image of q is a union of boundary circles $Q_0, \ldots, Q_t$ in a copy of M_0^*. Choose the standard generators h and k for $\pi_1 T^2$, and a basis $x_1, \ldots, x_r, y_1, \ldots, y_t$ for $\pi_1 M_0^*$ with y_i freely homotopic to Q_i. Then Van Kampen's theorem yields a presentation

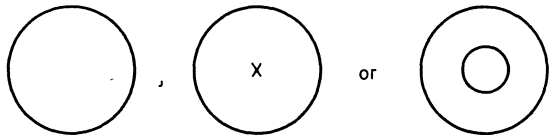
$$(h, k, x_1, \ldots, x_r, y_1, \ldots, y_t | h, k \text{ are central}, \\ h^{p_i} k^{q_i} = 1 \quad (i = 1, \ldots, s), \\ y_i^{\alpha_i} h^{\beta_i} k^{\gamma_i} = 1 \quad (i = 1, \ldots, t))$$

of $\pi_1 M$ (cf. 1.5 in [6]). By hypothesis some $(p_i, q_i) = (0, 1)$, and so $k = 1$. The result follows readily.

If M has fixed points, first remove an open invariant tubular neighborhood of the fixed point set, and then apply the argument above to the resulting (bounded) T^2-manifold. □

We first dispense with the abelian case.

Theorem 4.3. *Let M be a 4-dimensional singular* T^2*-manifold. Then* $\pi_1 M$ *is abelian if and only if M* is one of the following*



with any number of fixed points (that is $r + t \leqq 1$ *in the notation of Lemma 4.2). If there are no fixed points, then* $M = S^1 \times L$ *for some lens space L. If there are fixed points then M is a connected sum of* S^4, $S^3 \times S^1$, *or a manifold of type L (see Sect. 3) with some number of copies of* $\pm \mathbb{C}P^2$ *and* $S^2 \times S^2$.

Proof. Reparametrize the action on M, if necessary, to obtain a C-orbit of type $G(0, 1)$. Then Lemma 4.2 yields a central extension

$$1 \to H \to \pi_1 M \to Q \to 1$$

with $H = \mathbb{Z}_m$ generated by h and

$$Q = \underbrace{\mathbb{Z} * \ldots * \mathbb{Z}}_{r} * \mathbb{Z}_{\alpha_1} * \ldots * \mathbb{Z}_{\alpha_t}.$$

Evidently $\pi_1 M$ is abelian if and only if Q is cyclic, that is $r+t \leqq 1$.

The identification of the resulting manifolds follows from Proposition 2.1 in the fixed point free case and from Theorem 3.3 in the fixed point case. □

We now consider the non-abelian case.

Theorem 4.4. *Let M be a 4-dimensional singular T^2-manifold with $\pi_1 M$ non-abelian. Then $\pi_1 M$ has a cyclic center, generated by any C-orbit in M, with quotient a free product of cyclic groups.*

The complete set of invariants m (the order of the center of $\pi_1 M$), r and E for $\pi_1 M$ given in Theorem 4.1 are computed as follows.

(1) *$m=|H|$, where H is the intersection of all C-orbit isotropy types of M.*

(2) *$r=\operatorname{rank} H_1 M^*$.*

(3) *$E=$ set of pairs $(|Q_i|, |Q_i \cap H|)$, where Q_i ranges over the isotropy types of the E-orbits of M.*

In the notation of Lemma 4.2, $m=gcd(p_1, \ldots, p_s)$ and $E=\{(\alpha_1, gcd(m, \alpha_1, \beta_1)), \ldots, (\alpha_t, gcd(m, \alpha_t, \beta_t))\}$.

Remark. The invariance of m, r and the α_i was established previously by Pao [9].

Proof of 4.4. Reparametrize the action on M so that Lemma 4.2. applies, and adopt the notation there. [It is easy to check that $|H|=gcd(p_1, \ldots, p_s)$ and $(|Q_i|, |Q_i \cap H|) = (\alpha_i, gcd(m, \alpha_i, \beta_i))$, as claimed.] The structure of $\pi_1 M$ follows as in the proof of Theorem 4.3, as does the computation of m, r and the $n_i=\alpha_i$. By definition k_i is the smallest invariant factor of the subgroup G_i of $\pi_1 M$ generated by h and y_i, namely

$$G_i=(h, y_i | h \text{ is central}, h^m=1=y_i^{\alpha_i} h^{\beta_i}).$$

By elementary linear algebra $k_i=gcd(m, \alpha_i, \beta_i)$. □

5. Non-Singular Actions

We digress briefly from the central thread of the paper to discuss the overlap between singular and non-singular T^2-manifolds. At the end of this section we summarize the current state of knowledge about non-singular actions.

Theorem 5.1. *A 4-manifold M supports both a singular and a non-singular T^2-action if and only if it is diffeomorphic to $S^1 \times L$ for some lens space L.*

Proof. Since T^2 acts on any lens space L (Sect. 2 in [5]), one obtains a singular action on $S^1 \times L$ by letting T^2 act trivially on S^1 (cf. Proposition 2.1). To obtain a non-singular action take the product of the standard S^1 action on S^1 with any fixed point free S^1-action on L.

Conversely, suppose that M admits both singular and non-singular T^2-actions. Then any singular action on M is fixed point free, as the number of fixed points is a topological invariant ($=\chi(M)$). Moreover, $\pi_1 M$ must be abelian, for if it were non-abelian it would have a cyclic center (see Theorem 4.4) contradicting Proposition 5.2 below. Thus $M=S^1 \times L$ by Theorem 4.3. □

The diffeomorphism classification of 4-manifolds of the form $S^1 \times L$ is well known, namely

$S^1 \times L = S^1 \times L'$ if and only if $L = \pm L'$.

This can be deduced, using coverings, from Sect. 12.12 in [3].

Proposition 5.2. *Let M be a non-singular T^2-manifold. Then the center of $\pi_1 M$ is not cyclic unless $M = S^3 \times S^1$.*

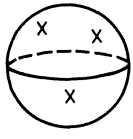
Proof. By 12.15 in [1] M fibers over S^1 with fiber an S^1-manifold N. Furthermore, the monodromy is an equivariant map of finite order. It follows that the center of $\pi_1 M$ is the product of the center of $\pi_1 N$ with $\mathbb{Z}$. This is not cyclic unless $\pi_1 N$ has a trivial center, in which case $N = S^3$ (6.2 in [4]) and so $M = S^3 \times S^1$. □

Non-singular T^2-manifolds have been studied by Zieschang [11], Conner and Raymond [1], and Orlik and Raymond [6]. In particular *injective* T^2-manifolds (that is T^2-manifolds M with $\pi_1 T \to \pi_1 M$ injective for any principal orbit T in M) are fairly well understood. They are determined by their fundamental groups and admit only one weak equivalence class of T^2-actions. Thus their diffeomorphism classification reduces to their weak equivariant classification, perhaps a tractable algebraic problem.

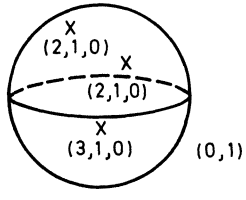
Less is known about *non-injective* T^2-manifolds. Those with abelian fundamental group, however, can be classified using known results. For, any such T^2-manifold has orbit space a sphere with at most two E-orbits (2.5 in [6]) and is thus the union of tubular neighborhoods of two of its orbits. Hence it admits a T^3-action (cf. end of Sect. 1 in [9]). By a theorem of Pak [7] one easily obtains (cf. 12.16 in [1] for partial results)

Theorem 5.3. *A non-singular T^2-manifold with abelian fundamental group is diffeomorphic to T^4 or to $S^1 \times L$ for some lens space L.*

Non-injective T^2-manifolds with non-abelian fundamental group, whose orbit spaces are of the form



with the reciprocal sum of the E-orbit orders greater than one, remain somewhat of a mystery. The manifolds $S^1 \times M$ arise, of course, where M is an analogous Seifert 3-manifold, but there are plenty more. For example,



has first homology $\mathbb{Z} \oplus \mathbb{Z}_2$, which is too small to be of the above form. An analysis of the weak equivalences among such T^2-manifolds should prove fruitful.

6. The Intersection Form

A natural question that arises when studying the integral intersection form of a 4-manifold M is whether or not there exist classes of odd self-intersection. M is called *odd* if there exist such classes, and *even* otherwise. Over $\mathbb{Z}_2$ this question concerns the existence of mod 2 classes of nonzero self-intersection. By the Wu formula such classes exist if and only if the second Stiefel-Whitney class w_2M of M is nonzero. If $w_2M=0$ then M is said to be *spin.*

Observe that any spin 4-manifold is even. The converse fails in general, however, as there may exist mod 2 classes with nonzero self-intersection which do not come from integral classes. For example the T^2-manifold L_2' is not spin (see Theorem 3.2) yet is even since H_2L_2' is finite.

In order to discuss these questions for non-singular T^2-manifolds, some elementary results from surgery theory will be needed.

Let C be a circle in a 4-manifold M. Denote by M' and M'' the two (possibly diffeomorphic) manifolds obtained from M by surgery along C using the two possible framings. The following is well known (see [2]).

Lemma 6.1. *M is spin if and only if M' or M'' is spin.*

The analogous result fails over $\mathbb{Z}$, that is M may be even while both M' and M'' are odd. For example let $M=L_2'$ and C represent the generator of $H_1M=\mathbb{Z}_2$. Then $M'=M''=S^2\tilde{\times}S^2$. The following weaker result, however, will suffice for our purposes.

Lemma 6.2. *If M' or M'' is even then so is M. Conversely, if M is even and C is of infinite or odd finite order in H_1M, then M' or M'' is even.*

Proof. We only sketch the proof as the argument is standard. The first statement follows from general position.

Now suppose M is even. If C is of infinite order then a Mayer-Vietoris argument shows that H_2M' and H_2M'' are generated by classes in $M-C$, which have even self-interaction. Thus M' and M'' are both even. If C is of finite order then there is an additional generator x. Since C is of odd order (by hypothesis) the parity of the self-intersection of x may be changed by changing the framing for the surgery. Thus either M' or M'' is even. □

We now return to the category of T^2-manifolds.

Call a subgroup of T^2 *special* if it is isomorphic to S^1 or to a cyclic group of even order. Two special subgroups are *mod 2 equivalent* if their intersection is again special. (This is an equivalence relation since every special subgroup has a unique element of order two.) Observe that there are three mod 2 equivalence classes of special subgroups. For, the class of a circle subgroup $G(p,q)$ is determined by the mod 2 reduction of the pair (p, q), and the class of a cyclic subgroup of even order is the same as the class of any circle subgroup containing it.

Definition. Let M be a T^2-manifold and K be a mod 2 equivalence class of special subgroups of T^2. Define the *orbit class* of M associated to K to be the set of all orbits of M whose isotropy groups are in K. An orbit class will be called *pure* if it contains no C-orbits, and *vacuous* if it contains no orbits at all.

Note that the orbit class of an E-orbit of even order with orbit invariants (α, β, γ) is determined by the mod 2 reduction of the pair (β, γ).

Theorem 6.3. *Let M be a 4-manifold with a singular T^2-action. Then M is spin if and only if it contains a vacuous orbit class.*

Remark. This result was obtained previously by Pao [9] for the case of fixed point free actions without E-orbits. Our proof is partially modelled on his approach.

Proof of 6.3. First assume that the action has fixed points. Then by Theorem 3.1 M is an equivariant connected sum of elementary T^2-manifolds of types P, R, T, and L.

Observe that it suffices to prove the theorem for each factor in this decomposition $M = \# M_i$ in light of the following two observations:

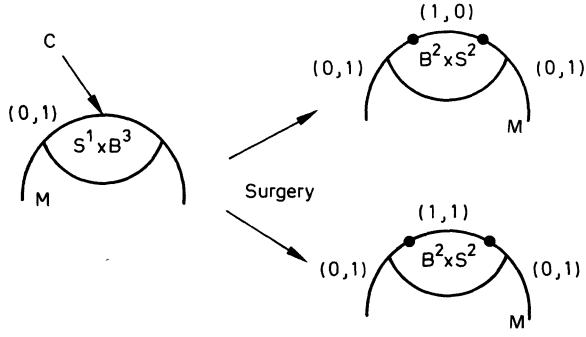
(1) M is spin if and only if every M_i is spin.

(2) M has a vacuous orbit class if and only if every M_i does.

Observation (1) is immediate from the splitting $H_2 M = \oplus H_2 M_i$ (with $\mathbb{Z}_2$-coefficients) and (2) follows from the fact that any invariant 3-sphere in M contains (exactly) two C-orbits in different orbit classes (see Sect. 2 in [5]).

The proof for these elementary factors is simply a matter of verification using Theorem 3.2 and Proposition 3.4.

Now assume that the action is fixed point free. Since it is singular (by hypothesis) there is a C-orbit C. Surgery on C with each of the two possible framings produces two T^2-manifolds M' and M'', each with fixed points, as illustrated in the normalized orbit space diagram below.



For details see Sect. 4 in [9]. Observe that M' and M'' have C-orbits in at most one orbit class distinct from those of M. Moreover, these orbit classes are distinct from one another (i.e. associated to distinct classes of special subgroups). Thus M has a vacuous orbit class if and only if M' or M'' does. The theorem now follows from Lemma 6.1 and the fixed point case. □

Theorem 6.4. *Let M be a 4-manifold with a singular T^2-action. Then M is even if and only if it contains a pure orbit class.*

Proof. The argument is identical to the proof of Theorem 6.3 with the following changes. Substitute the word "even" for "spin" and "pure" for "vacuous". To prove

the result for manifolds of type L observe that they have finite second homology and so are even. Finally replace the last sentence by the following argument:

It follows from the fixed point case that M has a pure orbit class if and only if M' or M'' is even. Thus by Lemma 6.2, M is even if it has a pure orbit class. Conversely assume M is even and suppose that it has no pure orbit class. Then by Theorem 4.4 the center of $\pi_1 M$ is finite cyclic of odd order, generated by C. By Lemma 6.2, M' or M'' is even. This contradiction establishes the theorem. □

The previous two theorems suggest that for an even singular T^2-manifold, the deviation from being spin is measured by the number of E-orbits in its pure orbit classes. This can be made precise.

Definition. To any orbit class K in a T^2-manifold M associate the function which assigns to every integer α the number of E-orbits of order α in K. Order such functions by setting $f<g$ if $f(\alpha_0)<g(\alpha_0)$ for α_0 the smallest integer in the support of $f-g$. Now define the *characteristic function* $c(M)$ of M to be the smallest function associated to a pure orbit class in M (defined to be the zero function if M has no pure orbit classes).

Theorem 6.5. *Let M be a 4-manifold with a singular T^2-action. Then the characteristic function $c(M)$ of M is a topological invariant. If M is even, then $c(M)=0$ if and only if M is spin.*

We first consider the fixed point case, which illuminates the meaning of the characteristic function.

Lemma 6.6. *Let M be a T^2-manifold with fixed points and with characteristic function c. The $c(\alpha)$ is the number of copies of L'_α in the normal decomposition of M.*

Remark. This lemma yields a refinement of Pao's classification Theorem 3.3. Namely, for M as above

$$M=\begin{cases} r(S^3\times S^1)\#\left(r-1+\dfrac{\chi+\sigma}{2}\right)(\mathbb{C}P^2)\#\left(r-1+\dfrac{\chi-\sigma}{2}\right)(-\mathbb{C}P^2)\#\limits_{\alpha} e(\alpha)L_\alpha \\ \qquad\qquad \text{if } M \text{ is odd} \\ r(S^3\times S^1)\#\left(r-1+\dfrac{\chi}{2}\right)(S^2\times S^2)\#\limits_{\alpha}(e(\alpha)-c(\alpha))L_\alpha\#\limits_{\alpha} c(\alpha)L'_\alpha \\ \qquad\qquad \text{if } M \text{ is even,} \end{cases}$$

where $r=\operatorname{rank} H_1 M^*$, $\chi=$Euler characteristic of M, $\sigma=$signature of M, and $e(\alpha)=$the number of E-orbits of order α in M. This is an exercise using additivity of the signature and Theorem 4.4.

Proof of 6.6. The lemma is obvious for M odd (as then $c=0$) so assume that M is even. By Theorem 3.1 M is an equivariant connected sum of elementary T^2-manifolds of types P, R, T, and L. After a suitable reparametrization of the action on M (which does not alter c) we may assume that the manifolds of type L are all of the form $L(\alpha,\beta,\gamma)$. Consequently M has C-orbits of types $G(0,1)$ and $G(1,0)$. Since M is even, Theorem 6.4 shows that the remaining orbit class (consisting of E-orbits with orbit invariants $(\alpha,\beta,\gamma)\equiv(0,1,1)\,(\mathrm{mod}\,2)$) is the unique pure orbit class of M. The lemma now follows from Proposition 3.4. □

Proof of 6.5. The theorem is obvious for odd manifolds, as in the proof of the lemma. Thus we assume that M is even. Moreover, we may assume that there are no fixed points, as the fixed point case follows from the lemma and the uniqueness of the normal decomposition of a T^2-manifold with fixed points.

If $\pi_1 M$ is abelian then it is obvious from Theorem 4.3 that $c(M)$ is the zero function, and therefore topologically invariant. So assume that $\pi_1 M$ is non-abelian.

It follows from Theorem 4.4 that the center Z of $\pi_1 M$ is cyclic, generated by any C-orbit in M. Consider a disjoint collection C of oriented circles $C_0, \ldots, C_k$ in M representing all possible generators of Z (where the trivial group is generated by 1 by convention). Observe that C is unique up to isotopy, and is independent of the particular T^2-action on M.

Now surgery on C in M amounts to surgery on a C-orbit (which kills Z) followed by surgery on k null homotopic circles. With a suitable framing the initial surgery yields an even T^2-manifold M_0 with fixed points. In fact it is evident that the framing can be chosen so that $c(M)=c(M_0)$.

The remaining circles in M_0 to be surgered can all be taken as C-orbits (since they are null homotopic). There is a unique framing on each of them which will preserve the evenness of M_0. Choosing these framings and performing the surgeries thus yields an even T^2-manifold M' with fixed points satisfying

$$c(M)=c(M').$$

Observe that M' can be defined without mention of the T^2-action on M. Namely, consider all the even manifolds which can be obtained from M by surgering C (there are at most two of them). The one with the smallest characteristic function is M'. Thus the topological invariance of $c(M)$ follows from the fixed point case.

The last statement of the theorem is an immediate consequence of Theorem 6.3. □

7. Applications

Here we apply the results of the previous sections to the problem of classifying 4-manifolds which admit singular T^2-actions. Throughout this section M will denote such a manifold.

If the center of $\pi_1 M$ is trivial then $\pi_1 M$ splits as a free product of cyclic groups. The following stable result of Pao [9] suggests that this splitting may be realized geometrically, as it is in the fixed point case (Theorem 3.3). We first give a definition.

Definition. Any 4-manifold which splits as a connected sum of copies of $\pm\mathbb{C}P^2$, $S^2 \times S^2$, $S^3 \times S^1$, L_α and L'_α is called *simple.*

Remark. T^2-manifolds with fixed points are simple. A simple T^2-manifold is classified by its fundamental group, Euler characteristic, signature, parity (even or odd) and characteristic function (cf. remark after Lemma 6.6).

Theorem 7.1 (Pao). *Let M be a singular T^2-manifold for which $\pi_1 M$ has trivial center. Then $M \# S^2 \times S^2$ and $M \# S^2 \tilde{\times} S^2$ are simple.*

In the direction of proving an unstable version of this theorem, Pao established the following result.

Theorem 7.2 (Pao). *Let M be a singular T^2-manifold. If there are two C-orbits in M of types G_1 and G_2 for which*

$$|G_1 \cap G_2| = 1$$

then M is simple.

We will generalize this result in two directions. (The question of whether or not all singular T^2-manifolds whose fundamental groups have trivial center are simple remains open.)

Theorem 7.3. *Let M be a singular T^2-manifold and let T be a component of singular orbits in M. If there are C-orbits in $M - T$ of types $G_1, \ldots, G_k$ for which*

$$|G_1 \cap \ldots \cap G_k| = 1$$

then M is simple.

The proof depends on the following lemma.

Lemma 7.4. *Let C_i be C-orbits of type $G(0,1)$ in T^2-manifolds M_i $(i=0,1)$. Consider the pairwise connected sum $(M, C) = (M_0, C_0) \# (M_1, C_1)$. Then $M_0 + M_1$ is obtained from M by surgering C.*

In particular, if $\pi_1 M_0$ has trivial center then $M_0 + M_1 = M_0 \# M_2$, where M_2 (obtained from M_1 by surgering C_1) is simple.

Proof. Recall that $M_0 + M_1$ is obtained by identifying $M_0 - \operatorname{int} N_0$ and $M_1 - \operatorname{int} N_1$ along their boundaries by an equivariant diffeomorphism where the N_i are invariant tubular neighborhoods of the C_i.

This identification space may be described differently. Let A_0 be a closed subarc of C_0, and A_1 be its image in C_1 under an equivariant diffeomorphism $f : N_0 \to N_1$ extending the identification above. Set $B_i = N_i | A_i$, a tubular neighborhood of A_i. Now $M_0 + M_1$ can be obtained from

$$(M_0 - \operatorname{int} B_0) \bigcup_{f|\partial B_0} (M_1 - \operatorname{int} B_1) = M_0 \# M_1$$

by removing

$$\operatorname{int}\left(\bigcup_i (N_i - \operatorname{int} B_i)\right) = \text{tubular neighborhood of } C_0 \# C_1$$

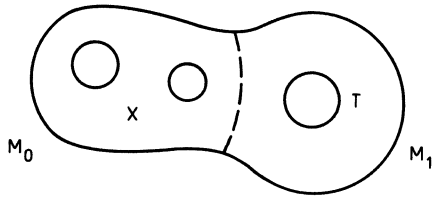
and appropriately identifying the resulting boundary to itself. The effect is to surger $C = C_0 \# C_1$ in $M = M_0 \# M_1$.

Now suppose that $\pi_1 M_0$ has trivial center Z. By Theorem 4.4. C_0 generates Z, and so $C = C_0 \# C_1 \simeq C_1$. Thus if M_2 denotes the 4-manifold obtained from M_1 by (appropriate) equivariant surgery on C_1, then

$$M_0 + M_1 = M_0 \# M_2 .$$

Evidently M_2 has fixed points, and is therefore simple. □

Proof of 7.3. Split M as a sum $M_0 + M_1$ with the given C-orbits in M_0 and T in M_1, as indicated on the orbit space below.



By Theorem 4.4 $\pi_1 M_0$ has trivial center. The lemma yields $M_0 + M_1 = M_0 \# M_2$, with M_2 simple. Evidently M_2^* is not a disc (since $T \subset M_1$) and so M_2 has at least one factor S diffeomorphic to $S^2 \times S^2$ or $S^2 \tilde{\times} S^2$. That is

$$M_2 = S \# M_3$$

with M_3 simple. By Theorem 7.1, $M_0 \# S$ is simple and so $M = M_0 + M_1 = M_0 \# M_2 = (M_0 \# S) \# M_3$ is simple. □

It is often possible to modify the action on a singular T^2-manifold, using the techniques of Sect. 2, so as to satisfy the hypotheses of Theorem 7.3. For example,

Corollary 7.5. *Let M be a singular T^2-manifold for which $\pi_1 M$ has trivial center. If there are three C-orbits in M of types G_1, G_2, and G_3 (lying in distinct components of singular orbits) such that*

$$G \cap G_i = 1 \quad \text{for} \quad i = 1, 2, 3$$

for some circle subgroup G of T^2, then M is simple.

Proof. Reparametrize the action on M so that $G = G(1, 0)$ and $G_1 = G(0, 1)$. The conditions $G \cap G_i = 1$ force $q_i = \pm 1$, where $G_i = G(p_i, q_i)$. By Theorem 2.4(1) arrange that $q_i = 1$ $(i = 2, 3)$. Modify the action further using Theorem 2.8 so that $p_2 = 0$. The resulting action on M has two C-orbits of type $G(0, 1)$ lying in distinct components of singular orbits. The hypotheses of Theorem 7.3 are now clearly satisfied, and so M is simple. □

A similar (and easier) argument using Theorem 2.2 proves

Corollary 7.6. *Let M be a singular T^2-manifold for which $\pi_1 M$ has trivial center. If M^* has non-zero genus, then M is simple.*

We now give another generalization of Theorem 7.2.

Theorem 7.7. *Let M be a singular T^2-manifold for which $\pi_1 M$ has trivial center. If there are two C-orbits in M of types G_1 and G_2 such that*

$$|G_1 \cap G_2| = 2$$

then M is simple.

Proof. The fixed-point case is immediate from Theorem 3.3, so assume that there are no fixed points in M.

Reparametrize the action so that $G_1 = G(0, 1)$. Then $G_2 = G(2, r)$ for some odd r. Since $\pi_1 M$ has trivial center, there must be a C-orbit of type H with $|G_1 \cap H|$ odd.

Thus $H = G(p, q)$ for some odd p and so

$$M = N + Q(2, r) + Q(p, q)$$

for some T^2-manifold N with a prescribed C-orbit of type $G(0, 1)$. By Theorem 2.4(1) arrange that $r = 1$, and then use Theorem 2.5(1) to make $p = 1$. The resulting action satisfies the hypotheses of Theorem 7.2 and so M is simple. □

Now consider the situation when $\pi_1 M$ has non-trivial center. Then M is certainly not simple, as $\pi_1 M$ is not a non-trivial free product. With suitable restrictions on the action on M, however, a classification can be obtained in terms of the same invariants which classify simple manifolds (cf. remark preceeding Theorem 7.1). In particular, we have the following extension of Theorem 7.7.

Theorem 7.8. *Let M be a singular T^2-manifold. If there are two C-orbits in M of types G_1 and G_2 for which*

$$|G_1 \cap G_2| = 2$$

then M is determined by its fundamental group, Euler characteristic, signature, parity and characteristic function c.

Proof. Recall that $\pi_1 M$ has center of order $m = |H|$, where H is the intersection of all the C-orbit types in M (see Theorem 4.4). Since $H \subset G_1 \cap G_2$, $m | 2$. Thus $m = 1$ or 2.

The case $m = 1$ was settled in Theorem 7.7 (this includes the fixed point case) and so we assume $m = 2$. It follows from Theorem 4.4 that every element of the invariant set E of $\pi_1 M$ is a pair whose second component is 1 or 2. Let K denote the subset of E having exactly $c(\alpha)$ pairs $(\alpha, 1)$, for each even α. We will show that

$$M = (r-1)Q(0, 1) + Q(2, 1) \underset{(\alpha, 1) \in K}{+} S(\alpha, 1, 0) \underset{(\alpha, \beta) \in E - K}{+} S(\alpha, \beta, 1), \tag{1}$$

where $r = \operatorname{rank} H_1 M^*$. This will prove the theorem since the manifold on the right is defined only in terms of the fundamental group and characteristic function of M. (Note that the Euler characteristic and signature of M vanish since there are no fixed points.)

As in the proof of Theorem 7.7, reparametrize the action on M so that $G_1 = G(0, 1)$ and $G_2 = (2, r)$. Since $m = 2$, every C-orbit is of type $G(p, q)$ with p even. Thus by Theorem 6.4 M is even and has two pure orbit classes. Using Theorem 2.4(1) arrange that $r = 1$, so that

$$M = Q(2, 1) \underset{i}{+} Q(p_i, q_i) \underset{j}{+} S(\alpha_j, \beta_j, \gamma_j)$$

for suitable $(p_i, q_i) \equiv (0, 1) \pmod 2$ and $(\alpha_j, \beta_j, \gamma_j)$.

By Theorem 2.5 the action can be modified further so that $(p_i, q_i) = (0, 1)$ for all i, $(\beta_j, \gamma_j) = (1, 0)$, $(1, 1)$ or $(2, 1)$ for all j, and $(\beta_j, \gamma_j) = (1, 1)$ if α_j is odd. Thus M admits an action which is either of the form (1) above or of the form

$$M = (r-1)Q(0, 1) + Q(2, 1) \underset{(\alpha, 1) \in K}{+} S(\alpha, 1, 1) \underset{(\alpha, 1) \in E - K}{+} S(\alpha, 1, 0) \underset{(\alpha, 2) \in E - K}{+} S(\alpha, 2, 1), \tag{2}$$

differing from (1) by an interchange of the pure orbit classes. These classes can be switched back by reparametrizing the action with the automorphism

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}$$

of T^2. Unfortunately this also changes the $Q(2,1)$ factor of M to $Q(2,3)$, but that can be restored using Theorem 2.4(1). Thus the action in (2) can be modified to the one in (1), which proves the theorem. □

Remark. Further algebraic invariants are needed for a complete classification of 4-manifolds with singular T^2-actions. For, there exist non-diffeomorphic, singular fixed point free T^2-manifolds with the same fundamental group, parity and characteristic function. Although the only such examples known to the author are provided by the T^2-manifolds whose C and E-orbit types lie in a fixed circle subgroup of T^2 (all diffeomorphic to the product of S^1 with a connected sum of lens spaces, cf. Sect. 5.2 in [9]), we suspect that many others exist.

Acknowledgements. I would like to thank Peter Orlik for introducing me to the subject of transformation groups on 4-manifolds, and the referee for suggesting a short proof of Theorem 4.1.

References

1. Conner, P., Raymond, F.: Holomorphic Seifert fiberings. Proc. of Second Conf. on Compact Transformation Groups, pp. 124–204. Lecture Notes in Mathematics, No. 299. Berlin, Heidelberg, New York: Springer 1972
2. Milnor, J.: A procedure for killing homotopy groups of differentiable manifolds. Proc. Sympos. Pure Math., Vol. III, pp. 39–55. Providence: American Mathematical Society 1961
3. Milnor, J.: Whitehead torsion. Bull. Am. Math. Soc. **72**, 358–426 (1966)
4. Orlik, P.: Seifert manifolds. Lecture Notes in Mathematics, No. 291. Berlin, Heidelberg, New York: Springer 1972
5. Orlik, P., Raymond, F.: Actions of the torus on 4-manifolds. I. Trans. Am. Math. Soc. **152**, 531–559 (1970)
6. Orlik, P., Raymond, F.: Actions of the torus on 4-manifolds. II. Topology **13**, 89–112 (1974)
7. Pak, J.: Actions of torus T^n on $(n+1)$-manifolds M^{n+1}. Pacific J. Math. **44**, 671–674 (1973)
8. Pao, P.: The topological structure of 4-manifolds with effective torus actions. I. Trans. Am. Math. Soc. **227**, 279–317 (1977)
9. Pao, P.: The topological structure of 4-manifolds with effective torus actions. II. Ill. J. Math. **21**, 883–894 (1977)
10. Raymond, F.: Classification of the actions of the circle on 3-manifolds. Trans. Am. Math. Soc. **131**, 51–78 (1968)
11. Zieschang, H.: On toric fiberings over surfaces. Math. Notes **5**, 341–345 (1969)

Received March 3, 1980; in revised form October 15, 1980

Math. Ann. 256, 277–279 (1981)

Mathematische Annalen
© Springer-Verlag 1981

A Banach space that is MLUR but not HR

Mark A. Smith

Department of Mathematics, Miami University, Oxford, OH 45056, USA

In [2], several examples of Banach spaces are given that illustrate the distinctions among various generalizations of uniform rotundity. Nine rotundity notions are considered in that paper. The known implications that exist among these notions are presented in a diagram (see [2, p. 157]). The examples given there show that, with but one possible exception, no other implications exist. The exception is the question, first posed by Anderson [1, p. 31]: does a Banach space have property H whenever it is midpoint locally uniformly rotund?

The purpose of this note is to answer Anderson's question by producing an example of the type described in the title and, hence, to show that the diagram as it appears in [2] is complete.

Although this note may be considered to be an addendum to [2], for the sake of completeness here, the relevant definitions are now stated.

A Banach space B is *midpoint locally uniformly rotund* (MLUR) if whenever x is in B and $\{x_n\}$ and $\{y_n\}$ are sequences in B such that $\|x\|=1$, $\|x_n\|\to 1$, $\|y_n\|\to 1$, and $\|2x-(x_n+y_n)\|\to 0$, then $x_n-y_n\to 0$.

A Banach space B has *property H* (H) if whenever x is in B and $\{x_n\}$ is a sequence in B such that $\|x_n\|\to\|x\|$ and $x_n\to x$ weakly, then $x_n\to x$. If B is H and is also rotund (strictly convex), write B is HR.

The Example. $(c_0, \|\cdot\|_G)$

For each positive integer k, define $V_k: c_0\to\mathbb{R}$ by

$$V_k(x^1, x^2, \ldots)=\sup\{|x^1-x^j|: j\geqq k\}.$$

Let $\{\alpha_n\}_{n=2}^{\infty}$ be a sequence of positive real numbers such that $\sum_{n=2}^{\infty}\alpha_n^2=1$. Define $V: c_0\to\ell^2$ and $T: c_0\to\ell^2$ by

$$V(x)=(\alpha_2 V_2(x), \alpha_3 V_3(x), \ldots) \quad \text{and} \quad T(x)=(x^1, \alpha_2 x^2, \alpha_3 x^3, \ldots)$$

for $x=(x^1, x^2, \ldots)$. For x in c_0, define

$$\|x\|_G=(\|x\|_\infty^2+\|V(x)\|_2^2+\|T(x)\|_2^2)^{1/2},$$

0025–5831/81/0256/0277/$01.00

where $\|\cdot\|_\infty$ and $\|\cdot\|_2$ denote the usual norms on c_0 and ℓ^2 respectively.

Then $\|\cdot\|_G$ is a norm on c_0 since T is a continuous, linear injection and $V(x+y) \leqq V(x)+V(y)$ in ℓ^2 for all x and y in c_0. Moreover, $\|\cdot\|_G$ is equivalent to the usual norm on c_0 since $\|\cdot\|_\infty \leqq \|\cdot\|_G \leqq \sqrt{6}\|\cdot\|_\infty$.

The space $(c_0, \|\cdot\|_G)$ does not have property H. To see this, let $x=e_1$ and $x_n=e_1+e_n$ where $\{e_n\}$ denotes the usual unit vector basis in c_0. Then $\|x_n\|_G \to \|x\|_G = \sqrt{3}$ and $x_n \to x$ weakly, but $x_n \not\to x$.

The space $(c_0, \|\cdot\|_G)$ is MLUR. To prove this, suppose w and $\{z_n\}$ are in c_0 and such that $\|w\|_G = 1$, $\|w+z_n\|_G \to 1$, and $\|w-z_n\|_G \to 1$. By Lemma 13 of [3], it suffices to show $z_n \to 0$. From the assumptions concerning w and $\{z_n\}$, it follows that

$$2(\|w+z_n\|_G^2 + \|w-z_n\|_G^2) - \|2w\|_G^2 \to 0. \tag{1}$$

The following elementary fact will be useful in the remainder of the proof.

Lemma. *If X is a Banach space and $p: X \to \mathbb{R}$ is a semi-norm, then*

$$2(p(x)^2 + p(y)^2) - p(x+y)^2 \geqq |p(x)-p(y)| \geqq 0$$

for all x and y in X.

Now, from (1), the definition of $\|\cdot\|_G$, the fact that T is linear, the fact that each V_k is a semi-norm on c_0, and the Lemma, it follows that

$$2(\|T(w)+T(z_n)\|_2^2 + \|T(w)-T(z_n)\|_2^2) - \|2T(w)\|_2^2 \to 0 \tag{2}$$

and, for each $k \geqq 2$,

$$2(V_k(w+z_n)^2 + V_k(w-z_n)^2) - V_k(2w)^2 \to 0. \tag{3}$$

Since $(\ell^2, \|\cdot\|_2)$ is uniformly rotund, from (2), it follows that $T(z_n) \to 0$ in ℓ^2 and, hence, by the definition of T,

$$z_n \to 0 \quad \text{weakly in} \quad c_0. \tag{4}$$

For the purpose of obtaining a contradiction, suppose $z_n \not\to 0$ in c_0. Then there exist $\varepsilon > 0$ and a subsequence of $\{z_n\}$, called $\{z_n\}$ again, such that $\|z_n\|_\infty > \varepsilon$ for each n. Since w is in c_0, there exists $k \geqq 2$ such that

$$|V_k(w) - |w^1|| < \varepsilon/4 \quad \text{and} \quad |w^i| < \varepsilon/4 \quad \text{for all} \quad i \geqq k. \tag{5}$$

Since the semi-norm V_k is bounded on the unit ball of $(c_0, \|\cdot\|_\infty)$, it follows, from the Lemma and (3), that there exists a subsequence of $\{z_n\}$, called $\{z_n\}$ again, such that

$$\lim_n V_k(w+z_n) = \lim_n V_k(w-z_n) = V_k(w).$$

From this, choose n_0 such that $n \geqq n_0$ implies

$$|V_k(w \pm z_n) - V_k(w)| < \varepsilon/4. \tag{6}$$

Combining (5) and (6), $n \geqq n_0$ implies

$$V_k(w \pm z_n) < |w^1| + \varepsilon/2. \tag{7}$$

By (4), there exists $n_1 \geqq n_0$ such that $n \geqq n_1$ implies

$$|z_n^i| < \varepsilon/8 \quad \text{for} \quad i = 1, 2, \ldots, k-1. \tag{8}$$

Since $\|z_n\|_\infty > \varepsilon$, for each $n \geqq n_1$, choose $i(n)$ such that $|z_n^{i(n)}| \geqq \varepsilon$. Note, by (8), $i(n) \geqq k$ for all $n \geqq n_1$, and hence, by (5), (8), and the definition of V_k,

$$\begin{aligned} V_k(w \pm z_n) &\geqq |w^1 \pm z_n^1 - w^{i(n)} \mp z_n^{i(n)}| \\ &\geqq |w^1 \mp z_n^{i(n)}| - |w^{i(n)} \mp z_n^1| \\ &\geqq |w^1 \mp z_n^{i(n)}| - 3\varepsilon/8 \end{aligned}$$

for all $n \geqq n_1$. This combined with (7) yields, for $n \geqq n_1$,

$$|w^1 \mp z_n^{i(n)}| - |w^1| < 7\varepsilon/8.$$

But this last inequality yields $|z_n^{i(n)}| < 7\varepsilon/8$ for $n \geqq n_1$, contrary to the choice of $i(n)$. Hence, $z_n \to 0$ in c_0 and the proof that $(c_0, \|\cdot\|_G)$ is MLUR is complete.

As a final remark, it should be noted that $(c_0, \|\cdot\|_G)$ is weakly uniformly rotund (WUR) (see [2] for the definition). This follows from the techniques used in [2], the fact that $(\ell^2, \|\cdot\|_2)$ is uniformly rotund, and the definition of T. The following information appends the table presented in [2, p. 160].

	UR	LUR	HR	MLUR	WLUR	WUR	URWC	URED	R
$(c_0, \|\cdot\|_G)$	−	−	−	+	+	+	+	+	+

References

1. Anderson, K.W.: Midpoint local uniform convexity, and other geometric properties of Banach spaces. Dissertation, University of Illinois 1960
2. Smith, M.A.: Some examples concerning rotundity in Banach spaces. Math. Ann. **233**, 155–161 (1978)
3. Smith, M.A., Turett, B.: Rotundity in Lebesgue-Bochner function spaces. Trans. Am. Math. Soc. **257**, 105–118 (1980)

Received October 21, 1980

Math. Ann. 256, 281-287 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Directedness in Ordered Normed Spaces and Operators

Kung-Fu Ng[1,*] and Michel Duhoux[2]

[1] Department of Pure Mathematics, University College of Swansea (Wales), Great Britain
[2] Institut de Mathématique Pure et Appliquée, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

Let $\mathscr{A}$ be a C^*-algebra, and $\mathscr{A}_h$ the real Banach space of all hermitian elements in $\mathscr{A}$ which is partially ordered by the cone $\mathscr{A}_h^+ = \{a^*a; a \in \mathscr{A}\}$. If $\mathscr{A}$ contains an identity e, then e also serves as an order-unit for $\mathscr{A}_h$; indeed the order interval $[-e, e]$ is precisely the unit ball in $\mathscr{A}_h$. If $\mathscr{A}$ does not contain an identity then, by a classical theorem of Segal, it has an approximate identity: a monotone increasing net (e_λ) of elements in $\mathscr{A}_h^+$ of norm less than 1 such that $\lim e_\lambda a = a = \lim a e_\lambda$ for each a in $\mathscr{A}$. It is natural to ask if such (e_λ) has to be an approximate order-unit, that is, by definition, if the gauge of the union $\bigcup_\lambda [-e_\lambda, e_\lambda]$ of order intervals is the norm of $\mathscr{A}_h$. Unfortunately, this is not the case; for example, in the C^*-algebra of all complex null sequences under the pointwise algebraic operations, and conjugation (as involution), if let $e_n = (1, \ldots, 1, 0, 0, \ldots)$, then $(e_n; n = 1, 2, \ldots)$ is an approximate identity but not approximate order-unit. However, conversely, we do have the following result (John Pickford has also obtained this result).

Theorem 1. *If (e_λ) is an approximate order-unit for $\mathscr{A}_h$, then it is also an approximate identity for $\mathscr{A}$.*

Proof. Let $x \in \mathscr{A}_h$ of norm 1, and let $\varepsilon > 0$. By Segal's theorem (cf. [2, 1.7.2]), there exists $u \in \mathscr{A}_h^+$ with $\|u\| < 1$ such that $\|x - xu\| < \varepsilon$. Take λ_0 such that $e_{\lambda_0} \geqq u$ and consider all $\lambda \geqq \lambda_0$. Then $e_\lambda \geqq e_{\lambda_0} \geqq u$, $0 \leqq e - e_\lambda \leqq e - u$ and so

$$0 \leqq x(e - e_\lambda)x \leqq x(e - u)x,$$

where e is an identity adjoint to $\mathscr{A}$. Hence

$$0 \leqq \|x(e - e_\lambda)x\| \leqq \|x(e - u)x\| \leqq \|x - xu\| \cdot \|x\| < \varepsilon.$$

It follows from the C^*-condition on the norm that $\|x\sqrt{e - e_\lambda}\|^2 < \varepsilon$; also, by the same reason, $\|\sqrt{e - e_\lambda}\|^2 < 2$. Hence

$$\|x(e - e_\lambda)\| \leqq \|x\sqrt{e - e_\lambda}\| \cdot \|\sqrt{e - e_\lambda}\| < \sqrt{2\varepsilon}.$$

* *Current address:* Department of Mathematics, University Science Centre, The Chinese University of Hong Kong, Shatin, New Territories, Hong Kong

0025-5831/81/0256/0281/$01.40

This shows that $x=\lim xe_\lambda$, valid for all $x\in\mathscr{A}_h$ of norm 1 and hence for all $x\in\mathscr{A}$. Similarly $x=\lim e_\lambda x$. □

In what follows, E will always be a partially ordered normed space (with the partial ordering defined by a *closed* cone C, not necessarily proper). With a slight abuse, we say that

(a) E is *additive* if the norm is additive on the positive cone C:

$$\|x+y\|=\|x\|+\|y\| \quad (x,y\in C);$$

(b) E is *directed upwards* if, actually, the open unit ball U in E is directed upwards: whenever $x,y\in U$, there exists $u\in U$ such that $u\geqq x,y$.

(a) implies that C is a normal cone and (b) implies that C is generating. It is known [10, p. 85] and easily seen that E is an approximate order-unit normed space if and only if U is directed upwards and order-convex. Certainly the open unit ball in $\mathscr{A}_h$ is order-convex, and its directedness can be seen immediately from the following result of Asimow, Duhoux, and Ng [6] as it is well known that the dual of $\mathscr{A}_h$ is additive (cf. [2, 2.1.6]).

Theorem 2. *Let E^*, E' be respectively the algebraic and Banach duals of E, partially ordered in the usual way by that of E. Then the following statements are equivalent:*

(i) *E is directed upwards;*

(ii) *E' is additive, and is order-convex in E^*.*

Moreover, if E is assumed to be norm-complete, then the second condition in (ii) *can be dropped as it would then be automatical.*

In a recent book [5, pp. 11–13] it is inferred that the second condition in (ii) can be dropped even if E is not complete. Unfortunately this is not correct. For example, let E be the space consisting of all real sequences (t_n) such that $\lim_{N\to+\infty}\sum_{n=1}^{N} t_n$ exists and finite. Thus $x=(\frac{1}{2},-\frac{1}{3},\frac{1}{4},\ldots)$ is in E. Also E is clearly a dense subspace of c_0 and the relative cone in E is dense in the cone of c_0; thus $E'\cong l_1$, that is, E' is additive. But E does not satisfy (ii) as there does *not* exist y in E with the property that $y\geqq\pm x$ (and consequently E' is not order-convex in E^*).

In [3], it was shown that, for Banach spaces, (i) of Theorem 2 implies a formally much stronger version of directedness property: each compact subset of the open unit ball U in E is majorized by an element in U. Again some simple examples show that the implication may not hold for non-complete spaces (e.g. consider the space consisting of finite sequences). In the following we shall identify those partially ordered normed spaces (E,C) with the strong version of directedness property. We shall say that E has the *l_1-majorized property* if whenever $\sum_{n=1}^{+\infty}\|x_n\|<+\infty$ with $x_n\in C$ then, for each $\varepsilon>0$, there exists x in E of norm less than $\sum_{n=1}^{+\infty}\|x_n\|+\varepsilon$ such that $x\geqq x_n$ for all n. Note that this property is satisfied in *each* of the following cases: (a) E is an order-unit normed space, (b) E is norm-complete. For instance, the space of all polynomials on $[0,1]$ satisfies (a) (and so is directed upwards), but not satisfy (b), while c_0 satisfies (b) but not (a).

Following Rogalski [7] and Wickstead [9], we denote by $H(C')$ the space of all positively homogeneous real valued functions h on the dual cone C' in E' such that h is w^*-continuous on the intersection Δ of C' with the unit ball U^0 (and hence on each norm-bounded subset of C' by homogeneity).

Theorem 3. *Let (E, C) be a partially ordered normed space with the open unit ball U. Then the following statements are equivalent:*

(i) *E is directed upwards and has the l_1-majorized property;*

(ii) *if $h \in H(C')$ with $\sup h(\Delta) < 1$, then there exists x in U such that $h(f) \leqq f(x)$ for all $f \in C'$;*

(iii) *for every precompact subset A of U and every $\varepsilon > 0$, A is majorized by an element of $(1+\varepsilon)U$;*

(iv) *each compact subset of U is majorized by an element of U.*

Proof. (i)⇒(ii). Take $c, \varepsilon > 0$ such that $\sup h(\Delta) < c < c+\varepsilon < 1$. Let $V_0 = cU$ and $V_n = \varepsilon 2^{-n} U$, $(n = 1, 2, \ldots)$. For each $x \in E$, define $W(x) = \{f \in C' ; f(x) > 1\}$. Then $W(x)$ is a w^*-open subset of C', increases as x increases and $C' \setminus V_n^0 \subseteqq \bigcup \{W(x) ; x \in V_n\}$ for each n. Let

$$K = \{f \in C' ; h(f) = 1\}.$$

Then $K \subseteqq C' \setminus V_0^0 \subseteqq \bigcup \{W(x) ; x \in V_0\}$. By w^*-continuity of h on norm-bounded subsets of C', $K \cap V_1^0$ is w^*-compact and it follows that there exists a finite subset F of V_0 such that

$$K \cap V_1^0 \subseteqq \bigcup \{W(x) ; x \in F\}.$$

Since U is directed upwards, there exists $x_0 \in V_0$ with $x_0 \geqq 0$ such that x_0 majorizes each element in F; thus $K \cap V_1^0 \subseteqq W(x_0)$. Since

$$K = (K \cap V_1^0) \cup (K \setminus V_1^0) \subseteqq W(x_0) \cup \bigcup_{x \in V_1} W(x),$$

we can repeat the above argument to find $x_1 \in V_1$ with $x_1 \geqq 0$ such that $K \cap V_2^0 \subseteqq W(x_0) \cup W(x_1)$ and, inductively, $x_n \in V_n$ with $x_n \geqq 0$ such that

$$K \cap V_{n+1}^0 \subseteqq W(x_0) \cup \ldots \cup W(x_n).$$

Since $\sum_{n=0}^{+\infty} \|x_n\| < c + \varepsilon < 1$, it follows from (i) that there exists $x \in U$ such that $x \geqq x_n$ for all n. Then $K \cap V_n^0 \subseteqq W(x)$ for all n and so $K \subseteqq W(x)$. By homogeneity of h, it follows that $h(f) \leqq f(x)$ for all $f \in C'$.

(ii)⇒(iii). Let A be a precompact subset of U. Define $h \in H(C')$ by the rule

$$h(f) = \sup f(A), \quad (f \in C').$$

To verify that h is w^*-continuous at the point f_0 in Δ, let $\varepsilon > 0$ and remark that A may be covered by the $\frac{\varepsilon}{3}$-balls about a finite number of points $x_1, \ldots, x_n$ in E. Now considering the w^*-neighborhood

$$W = \left\{ f \in \Delta ; |f(x_i) - f_0(x_i)| \leqq \frac{\varepsilon}{3}, (1 \leqq i \leqq n) \right\}$$

of f_0 in Δ, it is easy to see that $|h(f) - h(f_0)| \leqq \varepsilon$ for every $f \in W$.

Obviously $\sup h(\Delta) \leqq 1$. Consequently, given $\varepsilon > 0$, it follows from (ii) that there exists $x \in (1+\varepsilon)U$ such that $h(f) \leqq f(x)$ for all $f \in C'$ and, since C is closed, A is majorized by x.

(iii)⇒(iv). Remark that every compact subset A of U is contained in rU, with

$$\max\{\|x\| ; x \in A\} < r < 1 .$$

(iv)⇒(i). Let (x_n) be a sequence in C such that $\sum_{n=1}^{+\infty} \|x_n\| < \lambda < +\infty$. Then $\{x_n ; n = 1, 2, \ldots\} \cup \{0\}$ is a compact subset of λU (since $x_n \to 0$) and the result is clear. □

Let E, F be partially ordered normed spaces. Let $L(F, E)$ denote the space of all continuous linear operators from F to E, $P(F, E)$ the space consisting of precompact operators, and $K(F, E)$ of compact operators. All are partially ordered in the usual way: $T \geqq 0$ if and only if $Ty \geqq 0$ for all $y \geqq 0$ in F. It would be convenient to denote the positive cones in E, F by E_+, F_+ which will continue to be assumed closed. Also, to avoid trivialities, we assume that $E \neq E_+ \neq \{0\}$ and so, by the Hahn-Banach theorem, $E' \neq E'_+ \neq \{0\}$ (with similar assumption and result about F). The following theorem shows the strong version of directedness property studied in Theorem 3 has some interesting connection with the corresponding property in $P(F, E)$ and $K(F, E)$. We denote the open unit balls in E, F, $L(F, E)$, $P(F, E)$, and $K(F, E)$ by U_E, U_F, U_L, U_P, and U_K, respectively.

Theorem 4. *The following statements are equivalent:*

(i) *F is additive and each compact subset of U_E is majorized by an element in U_E;*

(ii) *each compact subset of U_P is majorized by an element in U_P;*

(iii) *each compact subset of U_K is majorized by an element in U_K.*

Proof. (i)⇒(ii) and (iii). Let $\mathscr{T}$ be a precompact subset of U_P, and $\varepsilon > 0$. Define

$$A = \{Ty ; T \in \mathscr{T} \text{ and } y \in F, \|y\| \leqq 1\} .$$

Then it is not difficult to verify that A is a precompact subset of U_E and hence majorized by an element $e \in (1+\varepsilon)U_E$ in view of (i) and Theorem 3. Now we use an idea of [4, 9] to show that $\mathscr{T}$ is majorized by an 1-dimensional ranged operator $g \otimes e$ with appropriate $g \in F'$ such that $\|g\| = 1$. Indeed, let q, p denote the norm of F respectively with domains F_+ and F. Then, by Bonsall's theorem (cf. [10, 1.5]) one can find $g \in F'$ such that $q \leqq g$ on F_+ and $g \leqq p$ on F; thus $\|g\| = 1$ and $g = \|\cdot\|$ on F_+. Let $S = g \otimes e$, that is

$$Sy = g(y)e \quad (y \in F) .$$

Then $Ty \leqq e = Sy$ whenever $y \in F_+$ with $\|y\| = 1$, and $T \in \mathscr{T}$. Therefore $T \leqq S$ for all $T \in \mathscr{T}$, and note also that $S \in (1+\varepsilon)U_K$ (in fact, $\|S\| = \|e\|$). It remains to apply Theorem 3 to $P(F, E)$ [or $K(F, E)$].

(ii)⇒(i). Let A be a compact subset of U_E. Take an arbitrary but fixed $y_0 \in F_+$ with $\|y_0\| = 1$. By Hahn-Banach theorem, take $g \in F'$ of norm 1 such that $g(y_0) = 1$. For each $x \in E$, define $T_x \in K(F, E)$ by $T_x = g \otimes x$. Then the linear map $x \to T_x$ is continuous since $\|T_x\| = \|x\|$ and, consequently, $\{T_a ; a \in A\}$ is a compact subset of

U_K. By (ii), there exists $T \in U_P$ such that $T \geqq T_a$ for all $a \in A$. Then $Ty_0 \geqq T_a y_0 = a$ for all a in A and $Ty_0 \in U_E$. This proves one half of (i). To prove that F is additive, let us first note that, since $P(F, E)$ is directed upwards, the dual $P(F, E)'$ is additive by a duality theorem (cf. [10, 9.7]). Pick an arbitrary but fixed $f \in E'$ with $f \geqq 0$, $\|f\| = 1$. Let $y_1, y_2 \in F_+$ and define $\varphi_i \in P(F, E)'$ by

$$\varphi_i(T) = f(Ty_i) \quad (T \in P(F, E)),$$

where $i = 1, 2$. Then $\varphi_i \geqq 0$ and $\|\varphi_i\| = \|y_i\|$ because, by the Hahn-Banach theorem,

$$\{Ty_i ; T \in P(F, E), \|T\| \leqq 1\} = \{x \in E ; \|x\| \leqq \|y_i\|\} = \{Ty_i ; T \in K(F, E), \|T\| \leqq 1\}.$$

[More precisely, given $x \in E$ such that $\|x\| \leqq \|y_i\|$, it suffices to choose $g \in F'$ with $g(y_i) = 1$, $\|g\| = 1/\|y_i\|$ and to define $T = g \otimes x$ to obtain $T \in K(F, E)$ with $\|T\| \leqq 1$ and $Ty_i = x$.] Similarly

$$\|\varphi_1 + \varphi_2\| = \|y_1 + y_2\|.$$

Since

$$\|\varphi_1 + \varphi_2\| = \|\varphi_1\| + \|\varphi_2\|$$

by an earlier remark, it follows that $\|y_1 + y_2\| = \|y_1\| + \|y_2\|$.

Clearly the same proof also gives (iii)⇒(i). □

Corollary 1. *The following statements are equivalent:*

(i) *F is additive and U_E is majorized by an element in E of norm $\leqq 1$;*

(ii) *U_L is majorized by an element in $L(F, E)$ of norm $\leqq 1$.*

Moreover the results continues to hold if $L(F, E)$ is replaced by $P(F, E)$ or $K(F, E)$.

Proof. If (ii) holds then $L(F, E)$ is directed upwards and so $L(F, E)'$ is additive. By the argument given in the theorem, it follows that F is additive since it can be embedded in $L(F, E)'$ positive-isometrically. Moreover, if U_L is majorized by $S \in L(F, E)$, then U_E is majorized by Sy where $y \in F_+$ and $\|y\| = 1$; indeed, for each $x \in U_E$ there exists, by Hahn-Banach theorem, $T \in U_K (\subseteq U_L)$ such that $Ty = x$. As $T \leqq S$, it follows that $x \leqq Sy$ as asserted.

Conversely, suppose (i) holds and U_E is majorized by some element e in E of norm $\leqq 1$. As in the theorem, take $g \in F'$ with $\|g\| = 1$ and $g = \|\cdot\|$ on F_+. Then $g \otimes e$ is of norm $\leqq 1$ and majorizes U_L. □

Combining Theorems 3 and 4, we also have the following:

Corollary 2. *Let E be norm-complete. Then the following statements are equivalent:*

(i) *F is additive and E is directed upwards;*

(ii) *$K(F, E)$ is directed upwards.*

Remark. Unlike in the situation of Corollary 1, $K(F, E)$ cannot be replaced by $L(F, E)$ in Corollary 2 (or Theorem 4): the cone in $L(F, E)$ might not even be generating even if F is base-normed and E is approximate order-unit normed (cf. [9]). It is not difficult to show, as in Theorem 4, that (ii)⇒(i) of Corollary 2 is true even if E is not necessarily norm-complete. The following example (cf. Theorem 1 of [3]) shows, on the other hand, that the completeness assumption in E is essential for the implication (i)⇒(ii). Let E be the normed vector space of all real

"finite" sequences (each is supported by finite number of coordinates). E is directed upwards in the usual ordering. Let A consist of all elements $x=(\alpha_1, \alpha_2, \ldots)$ in E with $|\alpha_n| \leqq \frac{1}{n}$ for all n. Then A is a *symmetric* compact convex subset which is *not* majorized in E. Let $F=E\times\mathbb{R}$, $B=A\times\{1\}$, and $C=\mathbb{R}^+B$ the cone generated by B as a base. Under the gauge of $co(B\cup -B)$, F is a base-normed space and so is additive. The natural projection T from F to E is in $K(F,E)$ as it maps the unit ball $co(B\cup -B)$ precisely onto B. Then there exists no S in $K(F,E)$ such that $S\geqq 0, T$. For otherwise, for $x\in A$,

$$2S(0,1)=S(-x,1)+S(x,1)\geqq 0+T(x,1)=x$$

because $(-x,1)$, $(x,1)\in A\times\{1\}\subseteq C$. This contradicts the fact that A is not majorized in E.

The following result is dual to Theorem 4 and generalizes Theorem 2.

Theorem 5. *The following statements are equivalent:*

(i) *F is directed upwards and E is additive;*

(ii) *$L(F,E)$ is additive and F' is order-convex in F^* (algebraic dual);*

(iii) *$P(F,E)$ is additive and F' is order-convex in F^*;*

(iv) *$K(F,E)$ is additive and F' is order-convex in F^*.*

Moreover the result continues to hold if "directed upwards" and "additive" are respectively replaced by "$(1+\varepsilon)$-generating for each $\varepsilon>0$" and "1-normal".

Proof. (i)⇒(ii). Since F is directed upwards, F' is certainly order-convex in F^*. Let us next note that, since E is additive, E satisfies the following property

$$(R_1)\ \pm x\leqq z \text{ in } E \text{ implies } \|x\|\leqq\|z\|.$$

Indeed, if we write $z+x=\alpha a$, $z-x=\beta b$ with $\|a\|=1=\|b\|$ and scalars $\alpha,\beta\geqq 0$, then

$$\|2x\|\leqq\alpha+\beta=\|\alpha a\|+\|\beta b\|=\|\alpha a+\beta b\|=\|2z\|.$$

Let $0\leqq T_1$, $T_2\in L(F,E)$ and $y_1, y_2\in U_F$. Then, by (i), there exists $y\in U_F$ such that $y\geqq\pm y_1, y_2$. Hence $T_1 y\geqq\pm T_1 y_1$ and it follows from (R_1) that $\|T_1 y\|\geqq\|T_1 y_1\|$. Similarly $\|T_2 y\|\geqq\|T_2 y_2\|$. Consequently, by additivity of E,

$$\|T_1+T_2\|\geqq\|(T_1+T_2)(y)\|=\|T_1 y\|+\|T_2 y\|\geqq\|T_1 y_1\|+\|T_2 y_2\|.$$

That implies that $\|T_1+T_2\|\geqq\|T_1\|+\|T_2\|$ and hence the equality.

It is trivial that (ii)⇒(iii)⇒(iv), and thus it remains only to show that (iv)⇒(i). Let $0\leqq f_1$, $f_2\in F'$ and pick $x\in E$ with $x\geqq 0$ and $\|x\|=1$. Then $0\leqq f_i\otimes x\in K(F,E)$ and it follows from (iv) that

$$\|f_1\otimes x\|+\|f_2\otimes x\|=\|f_1\otimes x+f_2\otimes x\|=\|(f_1+f_2)\otimes x\|,$$

that is $\|f_1+f_2\|=\|f_1\|+\|f_2\|$. This shows that F' is additive and it follows from Theorem 2 that F is directed upwards. Similarly, E can be embedded in $K(F,E)$ positive-isometrically and is additive.

The verification of the last assertion of the theorem is more straightforward and is due to Ellis and Wickstead (see [8]) for complete F. We omit the details. □

Combining the theorem with the two corollaries of Theorem 4, we have the following results proved by Ellis [4] and Wickstead [9] in the special case when F is complete.

Corollary. 1. *$L(F, E)$ is order-unit normed and F' order-convex in F^* if and only if F is base-normed and E is order-unit normed.*

2. *Let E be norm-complete. Then $K(F, E)$ is approximate order-unit normed and F' order-convex in F^* if and only if F is base-normed and E is approximate order-unit normed.*

Finally let us conclude the paper with a simple example of E that is directed upwards but not approximate order-unit normed. Let $l_\infty^2 = \mathbb{R}^2$ with the norm $\|(\alpha, \beta)\| = \max(|\alpha|, |\beta|)$, and $l_1^2 = \mathbb{R}^2$ with the norm $\|(\alpha, \beta)\| = |\alpha| + |\beta|$. Let $L = L(l_\infty^2 m\, l_1^2)$. Then, Ellis [4] has shown that L is additive but not base-normed. Hence L' is directed upwards but not approximate order-unit normed (we note also that L, L' are *not* Banach lattices though they are Banach spaces and vector lattices).

References

1. Asimow, L., Ellis, A.J.: Convexity theory and its applications in functional analysis. New York: Academic Press 1980
2. Dixmier, J.: C^*-algebras (A translation of: Les C^*-algèbres et leurs représentations). Amsterdam: North Holland 1977
3. Duhoux, M., Ng, K.F.: Decomposition of precompact operators in ordered locally convex spaces. J. London Math. Soc. **13**, 387–392 (1976)
4. Ellis, A.J.: Linear operators in partially ordered normed vector spaces. J. London Math. Soc. **41**, 323–332 (1966)
5. Lacey, H.E.: The isometric theory of classical Banach spaces. In: Grundlehren der mathematischen Wissenschaften, Band 208. Berlin, Heidelberg, New York: Springer 1974
6. Ng, K.F., Duhoux, M.: The duality of partially ordered locally convex spaces. J. London Math. Soc. **8**, 201–208 (1973)
7. Rogalski, M.: Cônes engendrés par un compact étoilé ou convexe, applications à l'étude des espaces de Banach ordonnés. Proceedings of NATO, Conference in Swansea on "Facial structure of compact convex sets and applications", pp. 83–87 (1972)
8. Wickstead, A.W.: Spaces of linear operators between partially ordered Banach spaces. Proc. London Math. Soc. **23**, 141–158 (1974)
9. Wickstead, A.W.: Compact subsets of partially ordered Banach spaces. Math. Ann. **212**, 271–284 (1975)
10. Wong, Y.C., Ng, K.F.: Partially ordered topological vector spaces. Oxford: Clarendon Press 1973

Received July 1, 1980; in revised form December 19, 1980

Math. Ann. 256, 289–301 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Zur optimalen Approximation konvexer Hyperflächen durch Polyeder

Rolf Schneider

Mathematisches Institut der Universität, Albertstraße 23b,
7800 Freiburg, Bundesrepublik Deutschland

Im folgenden ist F eine geschlossene konvexe Hyperfläche (Rand einer kompakten, konvexen Menge mit inneren Punkten) im n-dimensionalen euklidischen Raum $\mathbb{R}^n$ ($n \geqq 2$). F kann durch einbeschriebene Polyeder im Sinne der Hausdorffmetrik approximiert werden. Dabei ist die Hausdorffmetrik d erklärt durch

$$d(A,B) := \max \{\sup_{x \in A} \inf_{y \in B} \|x-y\|, \sup_{x \in B} \inf_{y \in A} \|x-y\|\}$$

für kompakte Teilmengen $A, B \subset \mathbb{R}^n$. Um die erreichbare Güte einer solchen Approximation beschreiben zu können, ist die folgende Definition zweckmäßig.

Definition. Zu $\varepsilon > 0$ sei $m(F,\varepsilon)$ die kleinste Zahl m derart, daß es ein der Fläche F einbeschriebenes konvexes Polytop P mit m Ecken gibt mit $d(F, \partial P) \leqq \varepsilon$.

Dabei bezeichnet ∂ den Rand. Ferner sei $\kappa_n = \pi^{n/2}/\Gamma\left(1+\frac{n}{2}\right)$ das Volumen der n-dimensionalen Einheitskugel und ϑ_n die Überdeckungsdichte der Kugel im n-dimensionalen Raum. Für eine Definition und für Abschätzungen dieser Zahl sehe man Rogers [4, S. 16–20]. Explizit bekannt ist nur $\vartheta_2 = 2\pi/\sqrt{27}$.

Gegenstand dieser Arbeit ist ein Beweis des folgenden Satzes.

Satz. *If F eine Hyperfläche der Klasse C^3 mit positiven Hauptkrümmungen, so gilt*

$$\lim_{\varepsilon \to 0+} m(F,\varepsilon)\, \varepsilon^{(n-1)/2} = 2^{(1-n)/2}\, \kappa_{n-1}^{-1}\, \vartheta_{n-1} \int_F \sqrt{K}\, dF,$$

wo K die Gauß-Kronecker-Krümmung von F (Produkt der Hauptkrümmungen) und dF das Oberflächenelement bezeichnet.

Verabredungsgemäß ist dabei F geschlossen, aber der Beweis läßt sich ohne wesentliche Änderungen auch durchführen, wenn F berandet und kompakt ist. In der Definition von $m(F,\varepsilon)$ wäre dann ∂P durch eine lokal konvexe polyedrische Hyperfläche zu ersetzen.

0025-5831/81/0256/0289/$02.60

Speziell im Fall $n=3$ besagt der Satz

$$\lim_{\varepsilon\to 0+} m(F,\varepsilon)\,\varepsilon=\frac{1}{\sqrt{27}}\int_F \sqrt{K}\,dF. \tag{1}$$

Bezeichnen wir für $m\in\mathbb{N}$ mit P_m ein der Fläche F einbeschriebenes Polytop mit m Ecken, für das unter dieser Bedingung der Abstand $d(F,\partial P_m)$ minimal ist, so ist leicht zu sehen, daß (1) äquivalent ist mit

$$\lim_{m\to\infty} m d(F,\partial P_m)=\frac{1}{\sqrt{27}}\int_F \sqrt{K}\,dF. \tag{2}$$

Diese Beziehung ist erstmals von Fejes Tóth [1], [2, S. 149] aufgestellt und plausibel gemacht worden. Dabei betont er allerdings, keinen Anspruch auf Strenge zu erheben. Möchte man dies doch tun, so bemerkt man, daß der für die Ungleichung

$$\liminf_{m\to\infty} m d(F,\partial P_m)\geqq\frac{1}{\sqrt{27}}\int_F \sqrt{K}\,dF$$

angedeutete Beweis ohne Schwierigkeiten präzisiert werden kann, jedoch offenbar keine Ausdehnung auf höhere Dimensionen gestattet. Den anderen Teil des heuristischen Arguments haben wir jedoch schon im dreidimensionalen Raum nicht sicher nachvollziehen können. Es wird daher im folgenden der n-dimensionale Fall nach einer anderen Methode behandelt. Die einfache Grundidee ist insofern wieder Fejes Tóth zuzuschreiben, als in [2] ein enger Zusammenhang dieser Art von Approximation mit Überdeckungsproblemen aufgezeigt wird. Entscheidend für das Folgende ist nun die naheliegende Bemerkung, daß zur optimalen Approximation einer konvexen Fläche durch einbeschriebene Polyeder in denjenigen Gegenden der Fläche, wo sie flacher ist, nur relativ wenige Eckpunkte der approximierenden Polyeder benötigt werden. Dem wird dadurch Rechnung getragen, daß wir die Fläche mit der durch die *zweite* Grundform gegebenen Riemannschen Metrik versehen und dann Überdeckungen von F durch geodätische Kugeln bezüglich dieser Metrik heranziehen.

Von nun an erfülle F die Voraussetzungen des Satzes. Insbesondere ist dann die zweite Grundform der Fläche (bei geeigneter Orientierung) positiv definit, sie definiert also in der Tat auf F eine Riemannsche Metrik. Begriffe, die sich hierauf beziehen, werden mit II gekennzeichnet. Zum Beispiel ist $L_{\mathrm{II}}(C)$ die Länge einer glatten Flächenkurve C auf F in dieser Metrik. Für $x,y\in F$ bezeichnet $\delta_{\mathrm{II}}(x,y)$ den Riemannschen Abstand von x und y, also das Infimum von $L_{\mathrm{II}}(C)$ über alle glatten Kurven C in F, die x und y verbinden. Schließlich ist $\{y\in F\colon \delta_{\mathrm{II}}(x,y)\leqq\rho\}$ für $x\in F$ und $\rho>0$ die II-geodätische Kugel um x mit Radius ρ und x ihr Mittelpunkt.

Im folgenden ist es notwendig, mit speziellen lokalen Parametrisierungen der Fläche F zu arbeiten. Sei $p\in F$ ein Punkt der Fläche, T_p die Tangentialhyperebene an F in p, H eine dazu parallele und F schneidende Hyperebene und H^+ der von H begrenzte abgeschlossene Halbraum, der p enthält. Bilden die äußeren Normalen an F in je zwei Punkten von $F\cap H^+$ stets einen Winkel $\leqq\alpha$

miteinander, wo $0<\alpha<\frac{\pi}{2}$ beliebig gewählt, aber im folgenden festgehalten ist, so wollen wir $F\cap H^+$ eine *Kappe* nennen. Der Abstand von H und T_p heiße die *Höhe* dieser Kappe und p ihr *Zentrum*. Offenbar gibt es ein $h_0>0$ derart, daß zu beliebigem $p\in F$ eine Kappe der Höhe h_0 mit Zentrum p existiert.

Mit Π_p bezeichnen wir die Orthogonalprojektion auf T_p. Für $y\in F$ sei

$$y^p:=\Pi_p y \quad \text{und} \quad h(y,p):=\|y-y^p\|$$

gesetzt. Dabei ist $\|\cdot\|$ die Norm in $\mathbb{R}^n$.

Sei jetzt F_p eine Kappe mit Zentrum p, und sei $D:=\Pi_p F_p$. D ist also eine kompakte konvexe Teilmenge von T_p. Sei e der zu T_p senkrechte Einheitsvektor, der ins Innere von F weist. Es gibt eine auf D definierte Funktion z der Klasse C^3 derart, daß $x+z(x)e\in F$ für $x\in D$ gilt. Wir nennen z die *Darstellungsfunktion* der Kappe F_p. In T_p wählen wir ein kartesisches (rechtwinkliges) Koordinatensystem mit Nullpunkt p und bezeichnen für $x\in T_p$ (und entsprechend für anders bezeichnete Punkte aus T_p) mit $x_1,\dots,x_{n-1}$ die zugehörigen Koordinaten von x. Wir schreiben

$$z_i=\frac{\partial z}{\partial x_i}, \qquad z_{ij}=\frac{\partial^2 z}{\partial x_i\,\partial x_j}, \qquad z_{ijk}=\frac{\partial^3 z}{\partial x_i\,\partial x_j\,\partial x_k}.$$

Fassen wir T_p als Vektorraum mit Nullpunkt p auf, so können wir darin durch

$$\langle v,w\rangle_p:=\sum z_{ij}(p)\,v_i w_j \tag{3}$$

ein Skalarprodukt $\langle\cdot,\cdot\rangle_p$ und damit eine euklidische Metrik erklären. Summationsindizes laufen hier und im folgenden, wenn nichts anderes gesagt ist, von 1 bis $n-1$. Wir setzen

$$\sqrt{\langle v,v\rangle_p}=:\|v\|_p$$

und bezeichnen $\|v-w\|_p$ auch als den *p-Abstand* der Punkte $v,w\in T_p$. Er hängt nicht von der Wahl der orthonormierten Basis in T_p ab.

Wir verwenden nun die Koordinaten $x_1,\dots,x_{n-1}$ des Punktes $x\in D$ als Parameter der Fläche F im Punkt $x+z(x)e$. Mit $b_{ij}(x)$ seien die Koeffizienten der zweiten Grundform (bezüglich des nach innen weisenden Normalenvektors der Fläche) von F in diesem Punkt bezüglich dieses Parametersystems bezeichnet. Es sei $\xi^p_{\min}(x)$ der kleinste und $\xi^p_{\max}(x)$ der größte relative Eigenwert der Matrix $(b_{ij}(x))$ bezüglich der (positiv definiten) Matrix $(z_{ij}(p))$. Auch diese Zahlen sind unabhängig von der Wahl der Basis in T_p.

Hilfssatz 1. *Zu jedem λ mit $0<\lambda<1$ existiert eine Konstante $0<h_\lambda\leqq h_0$ mit folgender Eigenschaft: Für alle $p\in F$ und alle $x,y\in F$ mit $0<h(x,p)\leqq h_\lambda$ und $0<h(y,p)\leqq h_\lambda$ gilt*

$$\lambda\leqq\xi^p_{\min}(x)\leqq\xi^p_{\max}(x)\leqq\lambda^{-1}, \tag{4}$$

$$\sqrt{\lambda}\leqq\frac{\delta_{\mathrm{II}}(x,y)}{\|x^p-y^p\|_p}\leqq\sqrt{\lambda^{-1}}, \tag{5}$$

$$\lambda\leqq\frac{\delta_{\mathrm{II}}(y,p)}{\sqrt{2h(y,p)}}\leqq\lambda^{-1}. \tag{6}$$

Beweis. Zunächst müssen wir die Existenz einer Zahl c zeigen derart, daß mit den obigen Bezeichnungen

$$|z_i(x)| \leqq c, \quad |z_{ij}(x)| \leqq c, \quad |z_{ijk}(x)| \leqq c \tag{7}$$

für alle x aus dem (hinreichend klein gewählten) Definitionsbereich von z und alle i,j,k gilt, wobei c weder von dem betrachteten Punkt p noch von dem in T_p gewählten Koordinatensystem abhängt. Die einzige Schwierigkeit besteht darin, daß zu jedem Punkt p eine andere Darstellungsfunktion z gehört.

Aus Kompaktheitsgründen gibt es endlich viele Kappen F_{p_i} mit Zentren p_i $(i=1,\ldots,\nu)$ und eine Zahl $0<h_1 \leqq h_0$ derart, daß jede Kappe der Höhe h_1 ganz in einer dieser Kappen enthalten ist. In jeder Tangentialhyperebene T_{p_i} wählen wir ein (im folgenden festzuhaltendes) kartesisches Koordinatensystem mit Nullpunkt p_i.

Sei nun $p \in F$ ein beliebiger Punkt und $q \in \{p_1,\ldots,p_\nu\}$ ein Punkt, so daß die Kappe F_p der Höhe h_1 mit Zentrum p ganz in F_q liegt. Wir bezeichnen jetzt zur Unterscheidung den zu T_q senkrechten (ins Innere von F weisenden) Einheitsvektor mit e_q und die in $D_q = \Pi_q F_q$ erklärte Darstellungsfunktion von F_q mit z^q. Analog seien e_p, z^p und $D_p = \Pi_p F_p$ erklärt. Wir wählen in T_p ein beliebiges kartesisches Koordinatensystem mit Nullpunkt p.

Da z^q dreimal stetig differenzierbar und D_q kompakt ist, gibt es eine Konstante c_0 mit

$$|z_i^q(x)| \leqq c_0, \quad |z_{ij}^q(x)| \leqq c_0, \quad |z_{ijk}^q(x)| \leqq c_0 \tag{8}$$

für alle $x \in D_q$. Diese Konstante kann einheitlich für die endlich vielen Kappen $F_{p_1},\ldots,F_{p_\nu}$ gewählt werden.

Für $x \in D_p$ setzen wir

$$y := x + z^p(x)\, e_p \quad \text{und} \quad \varphi(x) = y' := \Pi_q y.$$

Dann ist φ eine injektive Abbildung von D_p in T_q. Sei ψ die Umkehrabbildung. Sie ist wegen

$$x = y - \langle e_p, y-p\rangle\, e_p$$

(wo $\langle \cdot,\cdot \rangle$ das Skalarprodukt in $\mathbb{R}^n$ bezeichnet) und

$$y = y' + z^q(y')\, e_q \tag{9}$$

explizit gegeben durch

$$\psi(y') = y' + z^q(y')\, e_q - \langle e_p, y' + z^q(y')\, e_q - p\rangle\, e_p$$

und somit von der Klasse C^3. Aus (8) folgt, daß die partiellen Ableitungen der Koordinatenfunktionen der Abbildung ψ bis zur dritten Ordnung dem Betrage nach abgeschätzt werden können durch eine Konstante, die nur von c_0 abhängt. Betrachtet man im Definitionsbereich von ψ eine meßbare Menge vom $(n-1)$-dimensionalen Maß μ, so sieht man leicht, daß ihre Bildmenge unter ψ ein Maß $\geqq \mu/\cos\alpha$ hat (vgl. Definition der Kappen). Die Funktionaldeterminante der Abbildung ψ (bezüglich der in T_q und T_p gegebenen kartesischen Koordinatensysteme) ist also dem Betrage nach $\geqq 1/\cos\alpha$. Jetzt folgt, daß die partiellen Ableitungen der Koordinatenfunktionen der Umkehrabbildung φ bis

zur dritten Ordnung dem Betrage nach durch eine Konstante abgeschätzt werden können, die nur von c_0 und α abhängt.

Wegen

$$z^p(x)=\langle e_p, y-p\rangle$$

und (9) ist

$$z^p(x)=\langle e_p, \varphi(x)+z^q(\varphi(x))\,e_q-p\rangle.$$

Die partiellen Ableitungen der Funktion z^p bis zur dritten Ordnung können also ebenfalls abgeschätzt werden durch eine Konstante c, die nur von c_0 und α abhängt. Im folgenden können wir also die Ungleichungen (7) benutzen, wenn nur z Darstellungsfunktion einer Kappe der Höhe $\leqq h_1$ ist.

Sei jetzt $0<\lambda<1$ vorgegeben. Wir betrachten einen beliebigen Punkt $p\in F$ und dazu eine Kappe F_p der Höhe h_1. O.B.d.A. nehmen wir an, daß p der Nullpunkt 0 des $\mathbb{R}^n$ ist. Sei $D=\Pi_p F_p$ und z die auf D definierte Darstellungsfunktion der Kappe F_p. Im übrigen wählen wir alle Bezeichnungen so wie vor der Formulierung des Hilfssatzes. Es ist

$$b_{ij}(x)=\frac{z_{ij}(x)}{\sqrt{1+\sum z_k(x)^2}} \quad \text{für } x\in D.$$

Definieren wir Funktionen h_{ij} durch

$$z_{ij}(x)=z_{ij}(0)+h_{ij}(x) \quad \text{für } x\in D,$$

so gilt nach dem Mittelwertsatz

$$|h_{ij}(x)|\leqq c_1\,\|x\| \quad \text{für } x\in D$$

mit einer Konstanten c_1, die wegen (7) nur von c_0 abhängt. Ferner definieren wir Funktionen k_{ij} durch

$$b_{ij}(x)=z_{ij}(0)+k_{ij}(x) \quad \text{für } x\in D.$$

Wegen $z_k(0)=0$ ergibt sich dann für $x\in D$

$$\begin{aligned}|k_{ij}(x)|&=\left|z_{ij}(0)\left(\frac{1}{\sqrt{1+\sum z_k(x)^2}}-1\right)+\frac{h_{ij}(x)}{\sqrt{1+\sum z_k(x)^2}}\right|\\&\leqq|z_{ij}(0)|\sum z_k(x)^2+c_1\,\|x\|\\&\leqq c_2\,\|x\|\end{aligned}$$

mit einer Konstanten c_2, die abhängt von c_0, c_1 und z.B. dem Durchmesser von F.

Die relativen Eigenwerte der Matrix $(b_{ij}(x))$ bezüglich der Matrix $(z_{ij}(0))$ sind die Nullstellen des Polynoms

$$\xi\mapsto\det((1-\xi)\,z_{ij}(0)+k_{ij}(x)),$$

also die Lösungen einer Gleichung

$$(1-\xi)^{n-1}+\sum_{k=0}^{n-2}(1-\xi)^k\frac{a_k(x)}{\det(z_{ij}(0))}=0,$$

wobei jeder Summand der Summe $a_k(x)$ ein Produkt von k Koeffizienten der Matrix $(z_{ij}(0))$ und $n-1-k$ Koeffizienten der Matrix $(k_{ij}(x))$ ist. Es gilt daher eine Abschätzung

$$\left|\frac{a_k(x)}{\det(z_{ij}(0))}\right| \leqq c_3 \|x\| \quad \text{für } x \in D$$

mit einer Konstanten c_3, abhängend von c_0, c_2 und einer von p unabhängigen positiven unteren Schranke für $\det(z_{ij}(0))$. Eine solche existiert, da $\det(z_{ij}(0))$ die Gauß-Kronecker-Krümmung von F in p ist und da diese Krümmung auf F positiv und stetig ist. Wegen der stetigen Abhängigkeit der Nullstellen eines Polynoms von seinen Koeffizienten gibt es eine positive Konstante r_1 (abhängend von λ und c_3) derart, daß für alle $x \in D$ mit $\|x\| \leqq r_1$ die relativen Eigenwerte von $(b_{ij}(x))$ bezüglich $(z_{ij}(0))$ sämtlich $\geqq \lambda$ und $\leqq \lambda^{-1}$ sind. Für $\|x\| \leqq r_1$ ist also (4) erfüllt.

Insbesondere gilt für alle $x \in D$ mit $\|x\| \leqq r_1$

$$\lambda \sum z_{ij}(0)\alpha_i\alpha_j \leqq \sum b_{ij}(x)\alpha_i\alpha_j \leqq \lambda^{-1}\sum z_{ij}(0)\alpha_i\alpha_j \tag{10}$$

für beliebige $\alpha_1, \ldots, \alpha_{n-1}$.

Definieren wir eine Funktion ζ durch

$$z(x) = \tfrac{1}{2}\sum z_{ij}(0)x_ix_j + \zeta(x) \quad \text{für } x \in D, \tag{11}$$

so gilt nach dem Taylorschen Satz

$$|\zeta(x)| \leqq c_4 \|x\|^3 \quad \text{für } x \in D \tag{12}$$

mit einer positiven Konstanten c_4, die nur von c_0 abhängt. Wir bemerken nun, daß die Eigenwerte der Matrix $(z_{ij}(0))$ gerade die Hauptkrümmungen von F im Punkt p sind und daher zwischen zwei von p unabhängigen positiven Konstanten eingeschlossen werden können. Es ist also

$$c_5 \|x\|^2 \leqq \tfrac{1}{2}\sum z_{ij}(0)x_ix_j \leqq c_5^{-1}\|x\|^2 \quad \text{für } x \in T_p \tag{13}$$

mit einer Konstanten c_5. Wir setzen

$$r_2 = \frac{c_5}{c_4}(1-\lambda). \tag{14}$$

Für $x \in D$ mit $0 < \|x\| \leqq r_2$ gilt dann wegen (13), (14), $\lambda < 1$ und (12)

$$\begin{aligned}(\lambda^{-1}-1)\tfrac{1}{2}\sum z_{ij}(0)x_ix_j &\geqq (\lambda^{-1}-1)c_5\|x\|^2 \geqq c_4\|x\|^3 \geqq \zeta(x)\\ &= z(x) - \tfrac{1}{2}\sum z_{ij}(0)x_ix_j\end{aligned}$$

und

$$\begin{aligned}(1-\lambda)\tfrac{1}{2}\sum z_{ij}(0)x_ix_j &\geqq (1-\lambda)c_5\|x\|^2 \geqq c_4\|x\|^3 \geqq -\zeta(x)\\ &= \tfrac{1}{2}\sum z_{ij}(0)x_ix_j - z(x),\end{aligned}$$

also

$$2\lambda z(x) \leqq \sum z_{ij}(0)x_ix_j \leqq 2\lambda^{-1}z(x). \tag{15}$$

Wir können schließlich eine positive Konstante c_6 unabhängig von p wählen derart, daß jeder Punkt $x \in T_p$ mit $z(x) \leqq c_6$ der Ungleichung

$$\|x\| \leqq r := \operatorname{Min}\{r_1, r_2\}$$

genügt. Nun wählen wir die positive Konstante $h_\lambda < v_6$ so, daß der p-Abstand je zweier Punkte der Menge

$$M_\lambda := \{v \in T_p : z(v) \leqq h_\lambda\}$$

kleiner ist als der p-Abstand dieser Menge von der Menge

$$M_6 := \{v \in T_p : z(v) = c_6\}.$$

Daß h_λ unabhängig von p gewählt werden kann, leuchtet ein, wenn statt des p-Abstandes der gewöhnliche Abstand genommen wird; wegen der nach (13) gültigen Abschätzung

$$c_5 \|x-p\|^2 \leqq \tfrac{1}{2}\|x-p\|_p^2 \leqq c_5^{-1}\|x-p\|^2$$

ergibt es sich aber auch für den p-Abstand.

Seien nun $x, y \in F$ Punkte mit $h(x,p) \leqq h_\lambda$, $h(y,p) \leqq h_\lambda$. Sei C eine stetig differenzierbare Kurve auf F, die x und y verbindet. Wir nehmen zunächst an, für jeden Punkt q dieser Kurve gelte $h(q,p) \leqq c_6$. Dann gibt es eine stetig differenzierbare Abbildung

$$u\colon [0,1] \to B_r \quad \text{mit } B_r := \{x \in T_p : \|x\| \leqq r\}$$

mit $u(0) = x^p$ und $u(1) = y^p$ derart, daß die parametrisierte Kurve $t \mapsto u(t) + z(u(t))\,e$ eine Parametrisierung von C ist. Daher gilt nach (10)

$$L_{II}(C) = \int_0^1 \sqrt{\sum b_{ij}(u)\,du_i\,du_j} \geqq \sqrt{\lambda} \int_0^1 \sqrt{\sum z_{ij}(0)\,du_i\,du_j}$$
$$= \sqrt{\lambda}\, L_p(u),$$

wo $L_p(u)$ die Länge der parametrisierten Kurve u in der durch (3) gegebenen Metrik ist. Diese Länge ist nicht kleiner als der p-Abstand von x^p und y^p, also ist

$$L_{II}(C) \geqq \sqrt{\lambda}\, \|x^p - y^p\|_p. \tag{16}$$

Wenn wir nun zulassen, daß die Flächenkurve C den Bereich $\{q \in F : h(q,p) \leqq c_6\}$ verläßt, so gibt es eine Teilkurve C' mit einer Parametrisierung $t \mapsto u(t) + z(u(t))\,e$, wobei u innerhalb von B_r den Rand von M_λ mit dem Rand von M_6 verbindet. Nach der obigen Überlegung ist $\sqrt{\lambda^{-1}}\, L_{II}(C')$ nicht kleiner als der p-Abstand der Endpunkte von u, und dieser ist nach der Wahl von h_λ größer als der p-Abstand von x^p und y^p. Die Ungleichung (16) gilt also auch in diesem Fall.

Da C eine beliebige glatte Kurve zwischen x und y war, folgt

$$\delta_{II}(x,y) \geqq \sqrt{\lambda}\, \|x^p - y^p\|_p, \tag{17}$$

also die linke Ungleichung in (5).

Aus (17) ergibt sich insbesondere im Fall $x=p$

$$\delta_{\mathrm{II}}(y,p)\geqq\sqrt{\lambda}\,\|y^p\|_p=\sqrt{\lambda}\sqrt{\sum z_{ij}(0)\,y_i^p y_j^p}\geqq\lambda\sqrt{2z(y^p)}=\lambda\sqrt{2h(y,p)}$$

nach (15). Dies ist die linke Ungleichung in (6).

Andererseits gilt nach (10) und (15)

$$\begin{aligned}\delta_{\mathrm{II}}(x,y)&\leqq\int_0^1\sqrt{\sum b_{ij}((1-t)\,x^p+t\,y^p)(y_i^p-x_i^p)(y_j^p-x_j^p)}\,dt\\&\leqq\sqrt{\lambda^{-1}}\int_0^1\sqrt{\sum z_{ij}(0)(y_i^p-x_i^p)(y_j^p-x_j^p)}\,dt\\&=\sqrt{\lambda^{-1}}\,\|y^p-x^p\|_p,\end{aligned}$$

also die rechte Ungleichung in (5).

Speziell im Fall $x=p$ ergibt sich nach (15)

$$\delta_{\mathrm{II}}(y,p)\leqq\sqrt{\lambda^{-1}}\sqrt{\sum z_{ij}(0)\,y_i^p y_j^p}\leqq\lambda^{-1}\sqrt{2z(y^p)}=\lambda^{-1}\sqrt{2h(y,p)},$$

also die rechte Ungleichung in (6). Damit ist Hilfssatz 1 bewiesen. □

Wir stellen nun einen Zusammenhang des Approximationsproblems mit Überdeckungen von F durch II-geodätische Kugeln her.

Definition. Zu $\rho>0$ sei $k(F,\rho)$ die kleinste Zahl k derart, daß F durch k II-geodätische Kugeln vom Radius ρ überdeckt werden kann.

Hilfssatz 2. *Ist* $0<\lambda<1$ *und dazu* h_λ *gemäß Hilfssatz* 1 *gewählt, so gilt*

$$k(F,\lambda^{-1}\sqrt{2\varepsilon})\leqq m(F,\varepsilon)\leqq k(F,\lambda\sqrt{2\varepsilon}) \tag{18}$$

für $0<\varepsilon<h_\lambda$.

Beweis. Seien λ und h_λ wie angegeben, sei $0<\varepsilon<h_\lambda$. Es gibt ein der Fläche F einbeschriebenes Polytop P mit $m=m(F,\varepsilon)$ Ecken $x_1,\ldots,x_m$, so daß $d(F,\partial P)\leqq\varepsilon$ ist. Sei $p\in F$. Da die zur Tangentialebene T an F in p parallele Hyperebene im Abstand ε, die F trifft, das Polytop P treffen muß, gibt es einen Index i mit $h(x_i,p)\leqq\varepsilon$. Nach Hilfssatz 1 ist

$$\delta_{\mathrm{II}}(x_i,p)\leqq\lambda^{-1}\sqrt{2h(x_i,p)}\leqq\lambda^{-1}\sqrt{2\varepsilon},$$

also ist p enthalten in der II-geodätischen Kugel um x_i mit Radius $\lambda^{-1}\sqrt{2\varepsilon}$. Da $p\in F$ beliebig war, bilden die II-geodätischen Kugeln um $x_1,\ldots,x_m$ mit Radius $\lambda^{-1}\sqrt{2\varepsilon}$ eine Überdeckung von F. Dies beweist die linke Ungleichung in (18).

Nun setzen wir $\rho=\lambda\sqrt{2\varepsilon}$ und überdecken F durch $k=k(F,\rho)$ II-geodätische Kugeln vom Radius ρ. Seien $x_1,\ldots,x_k$ ihre Mittelpunkte, und sei P die konvexe Hülle dieser Punkte. Wir behaupten

$$d(F,\partial P)\leqq\varepsilon. \tag{19}$$

Angenommen, das wäre falsch. Sei $p \in F$ ein Punkt mit maximalem Abstand von P. Ist q der zu p nächste Punkt in P, so ist die zu $q-p$ orthogonale Hyperebene H durch q eine Stützebene an P (siehe McMullen-Shephard [3], S. 33). Da kein Punkt von F größeren Abstand von P hat als der Punkt p, ist die Tangentialebene T an F in p parallel zu H. Der Abstand von H und T ist größer als ε, also gilt $h(x_i, p) > \varepsilon$ für $i = 1, \ldots, k$. Nach Wahl der Punkte $x_1, \ldots, x_k$ gibt es einen Index i mit $\delta_{\mathrm{II}}(x_i, p) \leqq \rho$. Da jede glatte Kurve auf F, die p mit x_i verbindet, wenigstens einen Punkt y mit $h(y, p) = \varepsilon$ enthält, gilt nach Hilfssatz 1

$$\rho \geqq \delta_{\mathrm{II}}(x_i, p) > \delta_{\mathrm{II}}(y, p) \geqq \lambda \sqrt{2h(y,p)} = \lambda \sqrt{2\varepsilon} = \rho.$$

Dieser Widerspruch zeigt die Richtigkeit von (19). Da $k(F, \lambda\sqrt{2\varepsilon})$ die Eckenzahl des Polytops P ist, folgt die rechte Ungleichung in (18). □

Der nächste Hilfssatz bezieht sich auf Überdeckungen von Teilmengen des euklidischen Raumes durch kleine Kugeln. Er dürfte allgemein bekannt sein. Da er jedoch anscheinend nirgends explizit ausgesprochen ist, geben wir einen Beweis.

Für eine beschränkte Menge M im euklidischen Raum $\mathbb{R}^n$ und für $\rho > 0$ bezeichnen wir mit $K(M, \rho)$ die kleinste Zahl k derart, daß M durch k Kugeln vom Radius ρ überdeckt werden kann. V bezeichne das Volumen im $\mathbb{R}^n$.

Hilfssatz 3. *Für jede beschränkte quadrierbare Menge $M \subset \mathbb{R}^n$ gilt*

$$\lim_{\rho \to 0+} K(M, \rho) \kappa_n \rho^n = \vartheta_n V(M).$$

Beweis. Sei $0 < \lambda < 1$ vorgegeben. Wie man leicht sieht, gibt es eine Zahl $\gamma > 0$ derart, daß zu jedem $\beta \geqq \gamma$ im $\mathbb{R}^n$ ein Würfel W der Kantenlänge β existiert, der durch Einheitskugeln überdeckt werden kann, für deren Anzahl μ die Abschätzung

$$\mu \kappa_n \beta^{-n} < \lambda^{-1} \vartheta_n$$

gilt. Die quadrierbare Menge M läßt sich überdecken durch ν Würfel $W_1, \ldots, W_\nu$ einer Kantenlänge α mit

$$\nu \alpha^n < \lambda^{-1} V(M).$$

Sei nun $0 < \rho \leqq \alpha/\gamma$. Wir setzen $\beta := \alpha/\rho$ und bestimmen dazu W und μ wie oben. Jeder Würfel W_i kann überdeckt werden durch μ Kugeln vom Radius $\alpha/\beta = \rho$, also kann M überdeckt werden durch $\nu\mu$ solche Kugeln. Es folgt $K(M, \rho) \leqq \nu\mu$, also

$$K(M, \rho) \kappa_n \rho^n < \lambda^{-2} \vartheta_n V(M) \tag{20}$$

für $\rho \leqq \alpha/\gamma$.

Andererseits können wir in die Menge M Würfel $W_1, \ldots, W_\kappa$ einlagern derart, daß sie paarweise einen euklidischen Abstand $> 4\delta$ mit einem positiven δ haben und daß

$$\sum_{i=1}^{\kappa} V(W_i) > \lambda V(M)$$

ist. Sei $0<\rho<\delta$ und M überdeckt durch $K(M,\rho)$ Kugeln vom Radius ρ. Sei μ_i die Anzahl dieser Kugeln, die Punkte mit W_i gemeinsam haben. Verschiedene Würfel werden wegen $\rho\leqq\delta$ nicht von derselben Kugel getroffen, daher ist

$$K(M,\rho)\geqq\sum_{i=1}^{\kappa}\mu_i.$$

Ferner gilt

$$\mu_i\kappa_n\rho^n\geqq\vartheta_n V(W_i),$$

denn mit einem Würfel W_i, der dies nicht erfüllt, und seinen überdeckenden Kugeln könnte man eine Kugelüberdeckung des Raumes $\mathbb{R}^n$ einer Dichte $<\vartheta_n$ konstruieren. Es folgt

$$K(M,\rho)\kappa_n\rho^n>\lambda\vartheta_n V(M) \tag{21}$$

für $\rho\leqq\delta$.

Aus (20) und (21) ergibt sich

$$\begin{aligned}\lambda\vartheta_n V(M)&\leqq\liminf_{\rho\to 0+}K(M,\rho)\kappa_n\rho^n\\&\leqq\limsup_{\rho\to 0+}K(M,\rho)\kappa_n\rho^n\leqq\lambda^{-2}\vartheta_n V(M).\end{aligned}$$

Da $\lambda<1$ hier beliebig war, folgt die Behauptung von Hilfssatz 3. □

Wir kehren nun wieder zur Fläche F zurück und bezeichnen für eine offene Teilmenge $M\subset F$ mit $A_{\mathrm{II}}(M)$ ihr $(n-1)$-dimensionales Volumen bezüglich der durch die zweite Grundform gegebenen Riemannschen Metrik. Zur Abkürzung sei $A_{\mathrm{II}}(F)=A_{\mathrm{II}}$ gesetzt.

Hilfssatz 4. $\lim_{\rho\to 0+}k(F,\rho)\kappa_{n-1}\rho^{n-1}=\vartheta_{n-1}A_{\mathrm{II}}$.

Beweis. (Eine entsprechende Aussage dürfte allgemein für kompakte Riemannsche Räume gelten; es ist jedoch bequem, die im vorliegenden Spezialfall gegebenen Vereinfachungen auszunutzen.)

Sei $0<\lambda<1$ gegeben. Dazu sei die Konstante h_λ gemäß Hilfssatz 1 bestimmt. Sei jetzt zunächst $p\in F$ ein fester Punkt. Sei F_p die Kappe der Höhe $h_\lambda/2$ mit Zentrum p und $D:=\Pi_p F_p$; z sei die auf D definierte Darstellungsfunktion von F_p. Auch die weiteren Bezeichnungen seien wie früher gewählt. Ferner nehmen wir o.B.d.A. an, daß p der Nullpunkt des $\mathbb{R}^n$ ist. In dem Unterraum T_p führen wir die durch das Skalarprodukt (3) gegebene euklidische Metrik ein. Für $M'\subset D$ bezeichne $A_p(M')$ das $(n-1)$-dimensionale Volumen von M' bezüglich dieser Metrik.

Sei nun $\varnothing\neq M\subset F_p$ eine offene Teilmenge und $M'=\Pi_p M$ quadrierbar bezüglich T_p. Wir behaupten, daß

$$\lambda^{(n-1)/2}\leqq\frac{A_{\mathrm{II}}(M)}{A_p(M')}\leqq\lambda^{(1-n)/2} \tag{22}$$

ist. Für die relativen Eigenwerte $\xi_1,\ldots,\xi_{n-1}$ der Matrix $(b_{ij}(x))$ mit $x\in D$ gilt nach Hilfssatz 1

$$\lambda\leqq\xi_i\leqq\lambda^{-1}\qquad\text{für } i=1,\ldots,n-1.$$

Aus

$$\frac{\det(b_{ij}(x))}{\det(z_{ij}(0))} = \xi_1 \cdots \xi_{n-1},$$

$$A_{II}(M) = \int_{M'} \sqrt{\det(b_{ij}(x))}\, dx_1 \ldots dx_{n-1},$$

$$A_p(M') = \int_{M'} \sqrt{\det(z_{ij}(0))}\, dx_1 \ldots dx_{n-1}$$

folgt die Behauptung (22).

Nach Hilfssatz 3 gibt es eine Zahl $c_7 > 0$ (abhängend von M') mit

$$\lambda < \frac{K(M', \rho)\, \kappa_{n-1} \rho^{n-1}}{\vartheta_{n-1} A_p(M')} < \lambda^{-1} \tag{23}$$

für $0 < \rho \leqq c_7$. Hierbei bezieht sich $K(M', \rho)$ auf die durch (3) in T_p gegebene Metrik!

Sei $0 < \rho \leqq c_7$, und sei $K(M', \rho) = \mu$ gesetzt. Wir überdecken M' durch Kugeln (bezüglich (3)) $B'_1, \ldots, B'_\mu$ vom Radius ρ. Sei x'^s der Mittelpunkt der Kugel B'_s, sei $x^s := x'^s + z(x'^s)e$ und B_s in F die II-geodätische Kugel um x^s mit Radius $\sqrt{\lambda^{-1}}\rho$. Sei $y \in M$ ein beliebiger Punkt und $y' = \Pi_p y$. Es gibt eine Kugel B'_s mit $y' \in B'_s$. Nach Hilfssatz 1 gilt

$$\delta_{II}(y, x^s) \leqq \sqrt{\lambda^{-1}}\, \|y' - x'^s\|_p \leqq \sqrt{\lambda^{-1}}\, \rho,$$

also $y \in B_s$. Somit bilden die Kugeln $B_1, \ldots, B_\mu$ eine Überdeckung von M.

Umgekehrt seien (immer noch für $\rho \leqq c_7$) $B_1, \ldots, B_\nu$ II-geodätische Kugeln vom Radius $\sqrt{\lambda}\rho$ in F, die M überdecken. O.B.d.A. (notfalls verkleinere man c_7) liegen diese Kugeln in der Kappe der Höhe h_λ mit Zentrum p, so daß die Abschätzungen aus Hilfssatz 1 herangezogen werden können. Sei x'^s das Bild unter Π_p des Mittelpunktes x^s von B_s. Sei $y' \in M'$ ein beliebiger Punkt und $y := y' + z(y')e$. Es gibt ein s mit $y \in B_s$. Nach Hilfssatz 1 gilt

$$\sqrt{\lambda}\rho \geqq \delta_{II}(y, x^s) \geqq \sqrt{\lambda}\, \|y' - x'^s\|_p.$$

Der Punkt y' ist also enthalten in der Kugel B'_s (bezüglich der durch (3) gegebenen Metrik) mit Mittelpunkt x'^s und Radius ρ. Somit bilden die Kugeln $B_1, \ldots, B_\nu$ eine Überdeckung von M'. Es ist also

$$\nu \geqq K(M', \rho). \tag{24}$$

Wir können nun in F endlich viele Punkte $p_1, \ldots, p_m$ und offene Mengen $M_1, \ldots, M_m$ finden, so daß folgendes gilt: $p_i \in M_i$, für $y \in M_i$ gilt $h(y, p_i) \leqq h_\lambda/2$ ($i = 1, \ldots, m$), die Mengen $M_1, \ldots, M_m$ sind paarweise disjunkt, ihre abgeschlossenen Hüllen überdecken F, insbesondere ist

$$\sum_{i=1}^{m} A_{II}(M_i) = A_{II}, \tag{25}$$

und jede Menge $\Pi_{p_i} M_i$ is quadrierbar bezüglich T_p.

Wir können die oben gewählte Zahl c_7 so wählen, daß sie für alle Mengen $M_1, \ldots, M_m$ dasselbe leistet wie für M. Sei jetzt $0<\rho\leqq c_7$. Die (erste) obige Konstruktion ergibt dann für jede Menge M_i eine Überdeckung durch μ_i II-geodätische Kugeln vom Radius $\sqrt{\lambda^{-1}}\rho$. Nach (23) (wo $M'=\Pi_p M_i$ und $K(M',\rho)=\mu_i$ zu setzen ist) und (22) gilt

$$\mu_i \kappa_{n-1}\rho^{n-1}<\lambda^{-(n+1)/2}\vartheta_{n-1}A_{\mathrm{II}}(M_i).$$

Insgesamt überdecken diese Kugeln ganz F. Daher ist

$$k(F,\sqrt{\lambda^{-1}}\rho)\kappa_{n-1}\rho^{n-1}\leqq\sum_{i=1}^{m}\mu_i\kappa_{n-1}\rho^{n-1}\leqq\lambda^{-(n+1)/2}\vartheta_{n-1}A_{\mathrm{II}} \tag{26}$$

für $\rho\leqq c_7$, wobei zuletzt (25) benutzt wurde.

Wir ändern nun die oben getroffene Wahl der Mengen $M_1,\ldots,M_m$ dahingehend ab, daß je zwei der Mengen einen II-geodätischen Abstand $\geqq 4\delta$ mit einem $\delta>0$ haben sollen und daß

$$\sum_{i=1}^{m}A_{\mathrm{II}}(M_i)>\lambda^{-1}A_{\mathrm{II}} \tag{27}$$

sein soll, während die übrigen noch sinnvollen Forderungen bestehen bleiben. Wir können die oben gewählte Zahl c_7 auch für diese Mengen gemeinsam wählen.

Nun sei $0<\rho\leqq c_8:=\sqrt{\lambda^{-1}}\,\mathrm{Min}\{c_7,\delta\}$. Es gibt eine Überdeckung von F durch $k(F,\sqrt{\lambda}\rho)$ II-geodätische Kugeln vom Radius $\sqrt{\lambda}\rho$. Seien $B_1^i,\ldots,B_{\nu_i}^i$ diejenigen unter diesen Kugeln, die Punkte mit M_i gemeinsam haben. Sie überdecken M_i, also gilt nach (23), (24) und (22)

$$\nu_i\kappa_{n-1}\rho^{n-1}>\lambda^{(n+1)/2}\vartheta_{n-1}A_{\mathrm{II}}(M_i).$$

Andererseits ist $B_r^i\cap B_s^j=\varnothing$ für $i\neq j$ wegen $\sqrt{\lambda}\rho<\delta$ und daher

$$\sum_{i=1}^{m}\nu_i\leqq k(F,\sqrt{\lambda}\rho),$$

also

$$k(F,\sqrt{\lambda}\rho)\kappa_{n-1}\rho^{n-1}>\lambda^{(n-1)/2}\vartheta_{n-1}A_{\mathrm{II}} \tag{28}$$

für $\rho\leqq c_8$, wobei zuletzt (27) benutzt wurde.

Aus (26) und (28) folgt nun

$$\begin{aligned}\lambda^{n-1}\vartheta_{n-1}A_{\mathrm{II}}&\leqq\liminf_{\rho\to 0+}k(F,\rho)\kappa_{n-1}\rho^{n-1}\\&\leqq\limsup_{\rho\to 0+}k(F,\rho)\kappa_{n-1}\rho^{n-1}\leqq\lambda^{-n}\vartheta_{n-1}A_{\mathrm{II}}.\end{aligned}$$

Mit $\lambda\to 1$ folgt die Behauptung von Hilfssatz 4. □

Nun können wir den Beweis des Satzes zu Ende führen. Aus den Hilfssätzen 2 und 4 folgt unmittelbar

$$\begin{aligned}
&\lambda^{n-1}\, 2^{(1-n)/2}\, \kappa_{n-1}^{-1}\, \vartheta_{n-1}\, A_{\mathrm{II}} \\
&\quad = \lambda^{n-1}\, 2^{(1-n)/2} \lim_{\rho \to 0+} k(F,\rho)\, \rho^{n-1} \\
&\quad \leqq \liminf_{\varepsilon \to 0+} m(F,\varepsilon)\, \varepsilon^{(n-1)/2} \\
&\quad \leqq \limsup_{\varepsilon \to 0+} m(F,\varepsilon)\, \varepsilon^{(n-1)/2} \\
&\quad \leqq \lambda^{1-n}\, 2^{(1-n)/2} \lim_{\rho \to 0+} k(F,\rho)\, \rho^{n-1} \\
&\quad = \lambda^{1-n}\, 2^{(1-n)/2}\, \kappa_{n-1}^{-1}\, \vartheta_{n-1}\, A_{\mathrm{II}}
\end{aligned}$$

für jedes positive $\lambda < 1$. Mit $\lambda \to 1$ folgt

$$\lim_{\varepsilon \to 0+} m(F,\varepsilon)\, \varepsilon^{(n-1)/2} = 2^{(1-n)/2}\, \kappa_{n-1}^{-1}\, \vartheta_{n-1}\, A_{\mathrm{II}}.$$

Bezeichnet (g_{ij}) die erste Grundform der Hyperfläche F, so ist

$$\frac{\det(b_{ij})}{\det(g_{ij})} = K,$$

also

$$\sqrt{\det(b_{ij})}\, dx_1 \ldots dx_{n-1} = \sqrt{K}\, \sqrt{\det(g_{ij})}\, dx_1 \ldots dx_{n-1}$$

und somit

$$A_{\mathrm{II}} = \int \sqrt{K}\, dF.$$

Damit ist alles bewiesen.

Literatur

1. Fejes Tóth, L.: Approximation by polygons and polyhedra. Bull. Amer. Math. Soc. **54**, 431–438 (1948)
2. Fejes Tóth, L.: Lagerungen in der Ebene, auf der Kugel und im Raum. Berlin, Göttingen, Heidelberg: Springer 1953
3. McMullen, P., Shephard, G.C.: Convex polytopes and the upper bound conjecture. Cambridge: University Press 1971
4. Rogers, C.A.: Packing and Covering. Cambridge: University Press 1964

Eingegangen am 11. April 1980

Math. Ann. 256, 303–310 (1981)

Mathematische Annalen
© Springer-Verlag 1981

On the Rationality of Certain Generating Functions

Diane Meuser

Johns Hopkins University, Baltimore, MD 21218, USA

In this paper we prove the rationality of a generating function associated with r arbitrary polynomials in n variables with coefficients in the ring of integers of any finite algebraic extension of $\mathbb{Q}_p$.

Let p denote an arbitrary prime number, $\mathbb{Q}_p$ the Hensel p-adic field, and K any finite algebraic extension of $\mathbb{Q}_p$; let R denote the ring of integers of K, i.e., the maximal compact subring of K, P the unique maximal ideal of R, and $q = \mathrm{card}(R/P)$. Let r, n denote arbitrary positive integers and $f_1(x), \ldots, f_r(x)$ polynomials in n variables $x_1, \ldots, x_n$ with coefficients in R; put $X = K^n$, $X^0 = R^n = R \times \ldots \times R$. We shall use x and ξ to denote variable points in X and X^0, respectively; we write $f(x) = (f_1(x), \ldots, f_r(x))$. Then the following power series:

$$P(t) = \sum_{e=0}^{\infty} N_e (q^{-n} t)^e, \qquad N_e = \mathrm{card}\{\xi \bmod (P^e)^{(n)};\ f(\xi) \equiv 0 \bmod (P^e)^{(r)}\}$$

has $1 + t + t^2 + \ldots$ as a dominant series. In this paper we shall show that $P(t)$ represents a rational function of t.

In the special case where $r = 1$ the rationality of $P(t)$ was stated as a conjecture by Borewicz and Šafarevič [3, p. 63]; and was settled by Igusa [5, 6]. The proof in the general case is similar to the second proof [6] in that it depends upon the rationality of a "complex power" or local zeta function associated with $f(x)$. A similar complex power in the archimedean case has been investigated by Atiyah [1] and Bernstein-Gel'fand [2].

1. The Rationality of Complex Powers

For every a in $K^x = K - \{0\}$, we shall denote by $|a|_K$ the modulus of the multiplication by a in K; we have

$$aR = P^{\mathrm{ord}(a)}, \qquad |a|_K = q^{-\mathrm{ord}(a)};$$

we put $|0|_K = 0$. We fix an element π of order 1. Also for $u = (u_1, \ldots, u_r)$ in K^r we put

0025-5831/81/0256/0303/$01.60

$$|u|_K = \max\{|u_1|_K, \ldots, |u_r|_K\}.$$

We drop the assumption that the coefficients of $f_1(x), \ldots, f_r(x)$ are in R. Therefore these are arbitrary polynomials in $x_1, \ldots, x_n$ with coefficients in K; for our purpose we may assume that none of them is the constant 0. We take a complex variable s in the right-half plane (defined by $\mathrm{Re}(s)>0$) and a locally constant function Φ on X with compact support. We denote by $|dx|_K$ the Haar measure on X such that $m(X^0)$, the measure of X^0, is 1. Then

$$Z_\Phi(s) = \int_X \Phi(x) \cdot |f(x)|_K^s \cdot |dx|_K$$

defines a holomorphic function in the right-half plane. Before we prove the rationality of this we first prove a lemma.

Lemma. *Let n, p denote positive integers, $p>n$. Let A_i, B_i, $1 \leqq i \leqq p$ be non-negative integers such that for any i, A_i and B_i are not both zero. Let $f_i(k_1, \ldots, k_n)$, $n+1 \leqq i \leqq p$ be linear homogeneous functions with integer coefficients in $k_1, \ldots, k_n$. Consider the series*

$$\sum_{k_1, \ldots, k_n \geqq 0} q^{-\sum_{i=1}^{n} (A_i + sB_i) k_i} \, q^{-\sum_{i=n+1}^{p} (A_i + sB_i) f_i(k_1, \ldots, k_n)},$$

where in addition $k_1, \ldots, k_n$ are integers satisfying a system of linear inequalities with integer coefficients, and there exist constants D_i, $n+1 \leqq i \leqq p$, such that for all $k_1, \ldots, k_n$ satisfying these inequalities we have $f_i(k_1, \ldots, k_n) \geqq D_i$. Then the above series represents a rational function of q^{-s}.

Proof. The proof is by induction on n, i.e., we reduce a series of the above form to a series of similar form in $k_1, \ldots, k_{n-1}$.

For $n=1$ a system of linear inequalities in k_1 can be reduced to $M_1 \leqq k_1 \leqq M_2$ for some integers M_1, $M_2 \geqq 0$ (M_2 may be infinite). Each of the $f_i(k_1)$, $2 \leqq i \leqq p$, we can write as $F_i k_1$ for some integers F_i, and we must have $F_i k_1 \geqq D_i$ for some integers D_i. Now if $M_2 < \infty$ we have a finite sum which is clearly a rational of q^{-s}. If $M_2 = \infty$ then F_i must be non-negative for all i since otherwise we have $k_1 \leqq D_i/F_i$, a contradiction. So the series becomes

$$\sum_{M_1 \leqq k_1} q^{-\left\{\left(A_1 + \sum_{i=2}^{p} A_i F_i\right) + s\left(B_1 + \sum_{i=2}^{p} B_i F_i\right)\right\} k_1}.$$

This is a geometric series so it represents a rational function of q^{-s}.

We now proceed with the induction. We write the system of inequalities in $k_1, \ldots, k_n$ as

$$\sum_{i=1}^{n} C_{ji} k_i \geqq \gamma_j \qquad j = 1, \ldots, t$$

for some t, where the C_{ji} and γ_j are integers for all i and j. We may assume that for $1 \leqq j \leqq t_1$ we have $C_{jn} > 0$, for $t_1 < j \leqq t_2$ we have $C_{jn} < 0$, and for $t_2 < j \leqq t$ we have $C_{jn} = 0$. We rewrite this system as

$$C_{jn}k_n \geqq -\sum_{i=1}^{n-1} C_{ji}k_i + \gamma_j \qquad 1 \leqq j \leqq t_1 \tag{1a}$$

$$-C_{jn}k_n \leqq \sum_{i=1}^{n-1} C_{ji}k_i - \gamma_j \qquad t_1 < j \leqq t_2 \tag{1b}$$

$$\sum_{i=1}^{n-1} C_{ji}k_i \geqq \gamma_j \qquad t_2 < j \leqq t. \tag{2}$$

Consider inequalities (1a) and (1b). Multiply each inequality by

$$\prod_{\substack{l=1 \\ l \neq j}}^{t_1} C_{ln} \prod_{\substack{l=t_1+1 \\ l \neq j}}^{t_2} (-C_{ln}).$$

Now each inequality in (1a) and (1b) has the same coefficient of k_n, namely $\prod_{l=1}^{t_1} C_{ln} \prod_{l=t_1+1}^{t_2} (-C_{ln})$ which we denote by N. So (1a) and (1b) become

$$Nk_n \geqq (N/C_{jn}) \left(-\sum_{i=1}^{n-1} C_{ji}k_i + \gamma_j\right) \qquad 1 \leqq j \leqq t_1 \tag{3a}$$

$$Nk_n \leqq (N/(-C_{jn})) \left(\sum_{i=1}^{n-1} C_{ji}k_i - \gamma_j\right) \qquad t_1 < j \leqq t_2. \tag{3b}$$

Now for $1 \leqq j_1 \leqq t_1$, $t_1 < j_2 \leqq t_2$ we let $I(j_1, j_2) = \{(k_1, \ldots, k_{n-1})$, k_i non-negative integers, satisfying (2), such that

$$(N/C_{j_1 n}) \left(-\sum_{i=1}^{n-1} C_{j_1 i}k_i + \gamma_{j_1}\right) > (N/C_{jn}) \left(-\sum_{i=1}^{n-1} C_{ji}k_i + \gamma_j\right) \qquad 1 \leqq j < j_1$$

$$(N/C_{j_1 n}) \left(-\sum_{i=1}^{n-1} C_{j_1 i}k_i + \gamma_{j_1}\right) \geqq (N/C_{jn}) \left(-\sum_{i=1}^{n-1} C_{ji}k_i + \gamma_j\right) \qquad j_1 < j \leqq t_1$$

$$(N/(-C_{j_2 n})) \left(\sum_{i=1}^{n-1} C_{j_2 i}k_i - \gamma_{j_2}\right) < (N/(-C_{jn})) \left(\sum_{i=1}^{n-1} C_{ji}k_i - \gamma_j\right) \qquad t_1 < j < j_2$$

$$(N/(-C_{j_2 n})) \left(\sum_{i=1}^{n-1} C_{j_2 i}k_i - \gamma_{j_2}\right) \leqq (N/(-C_{jn})) \left(\sum_{i=1}^{n-1} C_{ji}k_i - \gamma_j\right) \qquad j_2 < j \leqq t_2\Big\}.$$

So $I(j_1, j_2)$ is defined by linear inequalities with integer coefficients and for $(k_1, \ldots, k_{n-1})$ belonging to $I(j_1, j_2)$, in order that $(k_1, \ldots, k_n)$ satisfy our original inequalities, k_n is determined by

$$k_n \geqq \max\left\{-\left[\left(-\sum_{i=1}^{n-1} C_{j_1 i}k_i + \gamma_{j_1}\right) \Big/ (-C_{j_1 n})\right], 0\right\} \tag{4a}$$

and

$$k_n \leqq \left[\left(\sum_{i=1}^{n-1} C_{j_2 i}k_i - \gamma_{j_2}\right) \Big/ (-C_{j_2 n})\right] \tag{4b}$$

where [ ] denotes the greatest integer function. Let $J(k_1,\dots,k_{n-1})$ be the set of k_n satisfying (4a) and (4b). Let $f_i(k_1,\dots,k_n)=\sum_{l=1}^{n} F_{il}k_l$ where the F_{il} are integers. We can rewrite our original series as

$$\sum_{\substack{1\leqq j_1\leqq t_1\\ t_1<j_2\leqq t_2}}\sum_{I(j_1,j_2)} q^{-\sum_{i=1}^{n-1}(A_i+sB_i)k_i}\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left(\sum_{l=1}^{n-1}F_{il}k_l\right)}$$

$$\cdot\sum_{J(k_1,\dots,k_{n-1})} q^{-(A_n+sB_n)k_n}\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)F_{in}k_n}.$$

So it is enough to show that the sum over $I(j_1,j_2)$ for any j_1,j_2 is a series of the proper form in $n-1$ variables.

Fix j_1 and j_2. For $0\leqq c_i<C_{j_1n}$, $0\leqq d_i<(-C_{j_2n})$, $1\leqq i\leqq n-1$, we let $I(c_1,\dots,c_{n-1},d_1,\dots,d_{n-1})$ denote the set of $(k_1,\dots,k_{n-1})$ in $I(j_1,j_2)$ such that $k_i\equiv c_i \bmod C_{j_1n}$ and $k_i\equiv d_i \bmod C_{j_2n}$, $1\leqq i\leqq n-1$. Then

$$I(j_1,j_2)=\coprod_{\substack{0\leqq c_i<C_{j_1n}\\ 0\leqq d_i<(-C_{j_2n})}} I(c_1,\dots,c_{n-1},d_1,\dots,d_{n-1})$$

so it is enough to show that the sum over $I(c_1,\dots,c_{n-1},d_1,\dots,d_{n-1})$ for any $c_1,\dots,c_{n-1},d_1,\dots,d_{n-1}$ is a series of the proper form in $n-1$ variables.

Fix c_i and d_i for $1\leqq i\leqq n-1$ and denote $I(c_1,\dots,c_{n-1},d_1,\dots,d_{n-1})$ by I. We may assume that I is not empty. Now let δ_1 be the integer such that

$$\sum_{i=1}^{n-1} C_{j_1i}c_i-\gamma_{j_1}\equiv\delta_1 \bmod C_{j_1n}$$

and $0\leqq\delta_1<C_{j_1n}$; then for $k_i\equiv c_i \bmod C_{j_1n}$ we have

$$-\left[\left(-\sum_{i=1}^{n-1} C_{j_1i}k_i+\gamma_{j_1}\right)\Big/(-C_{j_1n})\right]=\left(-\sum_{i=1}^{n-1} C_{j_1i}k_i+\gamma_{j_1}+\delta_1\right)\Big/C_{j_1n},$$

which we denote by $M_1(k_1,\dots,k_{n-1})$. Let δ_2 be the integer such that

$$\sum_{i=1}^{n-1} C_{j_2i}d_i-\gamma_{j_2}\equiv\delta_2 \bmod C_{j_2n}$$

and $0\leqq\delta_2<(-C_{j_2n})$; then for $k_i\equiv d_i \bmod C_{j_2n}$ we have

$$\left[\left(\sum_{i=1}^{n-1} C_{j_2i}k_i-\gamma_{j_2}\right)\Big/(-C_{j_2n})\right]=\left(\sum_{i=1}^{n-1} C_{j_2i}k_i-\gamma_{j_2}-\delta_2\right)\Big/(-C_{j_2n}),$$

which we denote by $M_2(k_1,\dots,k_{n-1})$. Inequalities (4a) and (4b) then become

$$k_n\geqq\max\{M_1(k_1,\dots,k_{n-1}),0\} \tag{5a}$$

and

$$k_n\leqq M_2(k_1,\dots,k_{n-1}). \tag{5b}$$

Now divide I into two subsets

$$I_1=\{(k_1,\ldots,k_{n-1})\in I;\ M_1(k_1,\ldots,k_{n-1})<0\}$$
$$I_2=\{(k_1,\ldots,k_{n-1})\in I;\ M_1(k_1,\ldots,k_{n-1})\geqq 0\}. \tag{6}$$

Splitting up I this way it is enough to show that the sum over each I_j, $j=1,2$, is a series of the proper form.

In order that the set of k_n in $J(k_1,\ldots,k_{n-1})$ is not empty for $(k_1,\ldots,k_{n-1})$ in I_1, we must have $M_2(k_1,\ldots,k_{n-1})\geqq 0$. So let I_1^* be the set of $(k_1,\ldots,k_{n-1})$ in I_1 which in addition satisfy this inequality. Similarly, in order that the set of k_n in $J(k_1,\ldots,k_{n-1})$ is not empty for $(k_1,\ldots,k_{n-1})$ in I_2 we must have $M_1(k_1,\ldots,k_{n-1})\leqq M_2(k_1,\ldots,k_{n-1})$. So let I_2^* be the set of $(k_1,\ldots,k_{n-1})$ in I_2 which satisfy this inequality. So we must examine

$$\sum_{I_j^*} q^{-\sum_{i=1}^{n-1}(A_i+sB_i)k_i}\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left(\sum_{l=1}^{n-1}F_{il}k_l\right)}$$
$$\cdot \sum_{J(k_1,\ldots,k_{n-1})} q^{-(A_n+sB_n)k_n}\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)F_{in}k_n}$$

for $j=1,2$.

We consider the case where $j=2$ as the other case can be dealt with in a similar manner. We sum over $J(k_1,\ldots,k_{n-1})$ getting

$$\left(1-q^{-\left\{(A_n+sB_n)+\sum_{i=n+1}^{p}(A_i+sB_i)F_{in}\right\}}\right)^{-1}$$
$$\cdot\left\{\sum_{I_2^*} q^{-\sum_{i=1}^{n-1}(A_i+sB_i)k_i}\, q^{-(A_n+sB_n)M_1(k_1,\ldots,k_{n-1})}\right.$$
$$\left.\cdot\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left\{\sum_{j=1}^{n-1}F_{ij}k_j+F_{in}M_1(k_1,\ldots,k_{n-1})\right\}}\right.$$
$$-\sum_{I_2^*} q^{-\sum_{i=1}^{n-1}(A_i+sB_i)k_i}\, q^{-(A_n+sB_n)(M_2(k_1,\ldots,k_{n-1})+1)}$$
$$\left.\cdot\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left\{\sum_{j=1}^{n-1}F_{ij}k_j+F_{in}(M_2(k_1,\ldots,k_{n-1})+1)\right\}}\right\}.$$

We show that the first series in brackets is of the proper form as the other series can be dealt with in a similar manner.

The solutions to $k_i\equiv c_i \bmod C_{j_1 n}$ and $k_i\equiv d_i \bmod C_{j_2 n}$ are determined modulo the least common multiple of $C_{j_1 n}$ and $C_{j_2 n}$, which we henceforth denote by m. If we denote by z_i the smallest non-negative solution, then we can write $k_i=z_i+mk_i'$ where k_i' is a non-negative integer. Denote by I_2^{**} the set of $(k_1',\ldots,k_{n-1}')$ corresponding to the $(k_1,\ldots,k_{n-1})$ in I_2^*. Substituting this equation into the inequalities defining I_2^* we see that I_2^{**} is defined by linear inequalities with integer coefficients. Let $C'_{j_1 n}=m/C_{j_1 n}$. Substituting $k_i=z_i+mk_i'$ into $M_1(k_1,\ldots,k_{n-1})$ we get

$$-C'_{j_1 n}\sum_{i=1}^{n-1} C_{j_1 i}k'_i+\left(-\sum_{i=1}^{n-1} C_{j_1 i}z_i+\gamma_{j_1}+\delta_1\right)\Big/C_{j_1 n}.$$

The second term is an integer by our choice of δ_1. We denote it by M'_1.

Our series becomes

$$q^{-\sum_{i=1}^{n-1}(A_i+sB_i)z_i}\, q^{-(A_n+sB_n)M'_1}\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left\{\sum_{j=1}^{n-1}F_{ij}z_j+F_{in}M'_1\right\}}$$

$$\cdot\sum_{I_2^{**}} q^{-\sum_{i=1}^{n-1}(A_i+sB_i)mk'_i}\, q^{-(A_n+sB_n)C'_{j_1 n}\left(-\sum_{i=1}^{n-1}C_{j_1 i}k'_i\right)}$$

$$\cdot\, q^{-\sum_{i=n+1}^{p}(A_i+sB_i)\left\{\sum_{j=1}^{n-1}(mF_{ij}+F_{in}C'_{j_1 n}(-C_{j_1 j}))k'_j\right\}}.$$

So we now examine the above series. By (6) we have for $(k_1,\ldots,k_{n-1})$ in I_2^* that $M_1(k_1,\ldots,k_{n-1})\geqq 0$; or in terms of k'_i

$$C'_{j_1 n}\left(-\sum_{i=1}^{n-1} C_{j_1 i}k'_i\right)\geqq -M'_1,$$

so the linear function associated to (A_n+sB_n) satisfies the required condition.

Now for $(k_1,\ldots,k_{n-1})$ in I_2^* we have $(k_1,\ldots,k_{n-1},M_1(k_1,\ldots,k_{n-1}))$ satisfying our original linear inequalities, hence

$$\sum_{j=1}^{n-1} F_{ij}k_j+F_{in}M_1(k_1,\ldots,k_{n-1})\geqq D_i.$$

Putting this in terms of k'_i we get

$$\sum_{j=1}^{n-1}(F_{ij}m+F_{in}C'_{j_1 n}(-C_{j_1 j}))k'_j\geqq D_i-\sum_{j=1}^{n-1}F_{ij}z_j-F_{in}M'_1$$

so the linear factors associated to (A_i+sB_i) for $n+1\leqq i\leqq p$ satisfy the required conditions. Hence the induction is complete, which proves the lemma. q.e.d.

We are now ready to prove the theorem on the rationality of complex powers.

Theorem 1. *The complex power $Z_\Phi(s)$ represents a rational function of q^{-s}.*

Proof. We apply Hironaka's main theorem [4, p. 176] to the product $f_1\ldots f_r$; we then get a K-analytic manifold Y and a proper K-analytic map h: $Y\to X$ with the following properties: for each b in Y there exist local coordinates $(y_1,\ldots,y_n)$ centered at b such that, locally around b, we have

$$f_1\ldots f_r\circ h=\varepsilon\cdot\prod_{i=1}^{n} y_i^{N_i}$$

$$h^*(dx_1\wedge\ldots\wedge dx_n)=\varepsilon'\prod_{i=1}^{n} y_i^{\nu_i-1}\,dy_1\wedge\ldots\wedge dy_n$$

in which $\varepsilon, \varepsilon'$ are invertible K-analytic functions and $N_1, \ldots, N_n \geqq 0$; $v_1, \ldots, v_n \geqq 1$. Since the ring of germs of K-analytic functions around b is a unique factorization ring, we have

$$f_j \circ h = \varepsilon_j \prod_{i=1}^{n} y_i^{N_{ji}}$$

in which ε_j is an invertible K-analytic function and $N_{j1}, \ldots, N_{jn} \geqq 0$ for $1 \leqq j \leqq r$. All these are valid in some open neighborhood, say U of b in Y. By making U smaller if necessary, we may assume that $|\varepsilon_j|_K$ for $1 \leqq j \leqq r$, $|\varepsilon'|_K$, and $\Phi \circ h$ are all constants on U; we may also assume that U is compact. Since the preimage under h of the support of Φ is compact, it can be covered by a finite number of open sets such as U. We call them $U_1, U_2, U_3, \ldots$, and pass to $U_1, U_2 - U_1, U_3 - (U_1 \cup U_2), \ldots$; in this way we may assume the covering is disjoint. Then we get

$$Z_\Phi(s) = \sum_U \text{const.} \int_U \max_{1 \leqq j \leqq r} \left\{ \left| \varepsilon_j \prod_{i=1}^{n} y_i^{N_{ji}} \right|_K \right\}^s \left| \varepsilon' \cdot \prod_{i=1}^{n} y_i^{v_i - 1} \right|_K |dy|_K.$$

We identify each U with its image in the y-space, which is a compact open subset of K^n. Thus each U can be written as a finite disjoint union of cosets mod $(P^{e_0})^{(n)}$ for all e_0 sufficiently large. Writing a coset $a + (P^{e_0})^{(n)}$ in the form $(a_1 + P^{e_0}) \times \ldots \times (a_n + P^{e_0})$ and using the fact that y_i in $a_i + P^{e_0}$ implies $|y_i|_k = |a_i|_k$ if a_i is not in P^{e_0}, the general case reduces to showing

$$\int_{(P^{e_0})^{(n)}} \max_{1 \leqq j \leqq r} \left\{ \left| \varepsilon_j \cdot \prod_{i=1}^{n} y_i^{N_{ji}} \right|_K \right\}^s \cdot \left| \varepsilon' \cdot \prod_{i=1}^{n} y_i^{v_i - 1} \right|_K |dy|_K$$

represents a rational function of q^{-s}. But this equals

$$\sum_{k_1, \ldots, k_n \geqq e_0} \int_{\pi^{k_1} U \times \ldots \times \pi^{k_n} U} \max_{1 \leqq j \leqq r} \left\{ \left| \varepsilon_j \cdot \prod_{i=1}^{n} y_i^{N_{ji}} \right|_K \right\}^s \cdot \left| \varepsilon' \cdot \prod_{i=1}^{n} y_i^{v_i - 1} \right|_K |dy|_K.$$

Making the change of variables $y_i = \pi^{k_i} u_i$ for all i this becomes

$$\sum_{k_1, \ldots, k_n \geqq e_0} q^{-\sum_{i=1}^{n} k_i} \int_{U^{(n)}} \max_{1 \leqq j \leqq r} \left\{ \left| \varepsilon_j \cdot \prod_{i=1}^{n} \pi^{N_{ji} k_i} u_i^{N_{ji}} \right|_K \right\}^s \left| \varepsilon' \cdot \prod_{i=1}^{n} \pi^{(v_i - 1) k_i} u_i^{v_i - 1} \right|_K |du|_K.$$

Evaluating the absolute values in the above integral, and using the fact that $m(U^{(n)}) = (1 - q^{-1})^n$, we get

$$(1 - q^{-1})^n q^{-\operatorname{ord} \varepsilon'} \sum_{k_1, \ldots, k_n \geqq e_0} q^{-s \min_{1 \leqq j \leqq r} \left\{ \operatorname{ord} \varepsilon_j + \sum_{i=1}^{n} N_{ji} k_i \right\}} q^{-\sum_{i=1}^{n} v_i k_i}. \tag{7}$$

Let

$$I_j = \left\{ \begin{array}{l} (k_1, \ldots, k_n);\ k_i \geqq e_0 \text{ for all } i, \text{ and} \\ \operatorname{ord} \varepsilon_j + \sum_{i=1}^{n} N_{ji} k_i < \operatorname{ord} \varepsilon_l + \sum_{i=1}^{n} N_{li} k_i \quad \text{for } 1 \leqq l < j \\ \operatorname{ord} \varepsilon_j + \sum_{i=1}^{n} N_{ji} k_i \leqq \operatorname{ord} \varepsilon_{lj} + \sum_{i=1}^{n} N_{li} k_i \quad \text{for } j < l \leqq r \end{array} \right\}.$$

Then (7) equals

$$(1-q^{-1})^n q^{-\operatorname{ord}\varepsilon'} \sum_{j=1}^{r} q^{-s \operatorname{ord}\varepsilon_j} \sum_{I_j} q^{-\sum_{i=1}^{n}(v_i+sN_{ji})k}.$$

So it is enough to show that

$$\sum_{I_j} q^{-\sum_{i=1}^{n}(v_i+sN_{ji})k_i}$$

is a rational function of q^{-s}. This is a series of the type covered by the lemma since I_j is defined by a system of linear inequalities with integer coefficients, and for each i, v_i and N_{ji} are non-negative integers, not both zero. Hence it represents a rational function of q^{-s}, which completes the proof. q.e.d.

2. The Rationality of $P(t)$

We consider the special case where the coefficients of $f_1(x), \ldots, f_r(x)$ are in R and Φ is the characteristic function of $X^0 = R^{(n)}$. Then we have the following theorem relating the generating function in $P(t)$ and the so-specialized complex power $Z_\Phi(s)$. The proof of this is entirely similar to the case $r=1$ in [6].

Theorem 2. *Let $t=q^{-s}$. Then $(1-t)\,P(t)=1-t\cdot Z_\Phi(s)$; hence $P(t)$ is also a rational function of t.*

Proof. In this case $f(x)=(f_1(x), \ldots, f_r(x))$ defines a mapping from $R^{(n)}$ to $R^{(r)}$. Decompose $R^{(n)}$ as

$$R^{(n)} = f^{-1}(0) \coprod \left(\coprod_{e=0}^{\infty} f^{-1}((P^e)^{(r)} - (P^{e+1})^{(r)}) \right).$$

But the measure of $f^{-1}((P^e)^{(r)})$ is N_e-times the measure of $(P^e)^{(n)}$ which is q^{-ne}. Therefore

$$\begin{aligned} Z_\Phi(s) &= \sum_{e=0}^{\infty} (N_e q^{-ne} - N_{e+1} q^{-n(e+1)})\, t^e \\ &= P(t) - t^{-1}(P(t)-1); \end{aligned}$$

which can be written as above. q.e.d.

References

1. Atiyah, M.F.: Resolution of singularities and division of distributions. Comm. Pure Appl. Math. **23**, 145–150 (1970)
2. Bernstein, I.N., Gel'fand, S.I.: Meromorphic property of the function P^λ. Functional Anal. Appl. **3**, 68–69 (1969)
3. Borewicz, S.E., Safarevic, I.R.: Zahlentheorie. Basel, Stuttgart: Birkhaeuser 1966
4. Hironaka, H.: Resolution of singularities of an algebraic variety over a field of characteristic zero. Ann. Math. **79**, 109–326 (1964)
5. Igusa, J.-I.: Complex powers and asymptotic expansions *I*. J. reine angew. Math. **268/269**, 110–130 (1974); ibid. **278/279**, (1975)
6. Igusa, J.-I.: Some observations on higher degree characters. Am. J. Math. **99**, 393–417 (1977)

Received May 18, 1980

Math. Ann. 256, 311–321 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Abelian Theorems for Distributional H-Transform

V.G. Joshi and R.K. Saxena

Department of Mathematics, Shivaji University,
Vidyanagar, 416004 Kolhapur, India

1. Introduction

The object of this paper is to present a new distributional integral transform associated with the product of an exponential function and Fox's H-function, as its kernel. Some Abelian theorems of initial and final value types of this new transform are discussed.

The results obtained in this paper are of general character and include as special cases number of known and unknown results. However, a few particular cases of interest given by Zemanian [13, 14], Mishra [8], Saxena [10] and others are mentioned here.

2. The H-function

Fox [3, p. 408 (1)], introduced H-function in the form of Mellin-Barnes type integral

$$F(z)=\frac{1}{2\pi i}\int_T \chi(\xi)\, z^{-\xi}\, d\xi \tag{2.1}$$

where

$$\chi(\xi)=\frac{\prod_{j=1}^{m}\Gamma(b_j+\xi c_j)\prod_{j=1}^{n}\Gamma(a_j-\xi e_j)}{\prod_{j=m+1}^{q}\Gamma(b_j-\xi c_j)\prod_{j=n+1}^{p}\Gamma(a_j+\xi e_j)} \tag{2.2}$$

and for a suitable contour T, the Mellin transform of $F(z)$ is evidently the coefficient of $z^{-\xi}$ in the integral (2.1).

Braaksma [2, p. 239 (1.1)] has considered the H-function in the form

$$H(z)=\frac{1}{2\pi i}\int_T \psi(\xi)\, z^{\xi}\, d\xi \tag{2.3}$$

0025-5831/81/0256/0311/$02.20

where

$$\psi(\xi)=\frac{\prod_{j=1}^{m}\Gamma(b_j-\beta_j\xi)\prod_{j=1}^{n}\Gamma(1-a_j+\alpha_j\xi)}{\prod_{j=m+1}^{q}\Gamma(1-b_j+\beta_j\xi)\prod_{j=n+1}^{p}\Gamma(a_j-\alpha_j\xi)}. \tag{2.4}$$

It can be seen that (2.3) can be obtained from (2.1) by proper substitutions. In our present work we shall use the definition of the H-function as given in (2.3) where $z\neq 0$ and

$$z^{\xi}=\exp\{\xi(\operatorname{Log}|z|+i\arg z)\}$$

in which $\operatorname{Log}|z|$ denotes the natural logarithm of $|z|$ and $|\arg z|$ is not necessarily the principal value. $H(z)$ in general is a multi-valued function of z. However, it is single-valued on the Riemann surface of $\log z$. Further an empty product is interpreted as unity, p, q, n and m are integers such that

$$0\leqq n\leqq p, \quad 1\leqq m\leqq q, \tag{2.5}$$

α_j $(j=1,2,\ldots,p)$, β_j $(j=1,2,\ldots,q)$ are complex numbers such that no pole of $\Gamma(b_h-\beta_h\xi)$ $(h=1,2,\ldots,m)$ coincides with any pole of $\Gamma(1-a_j+\alpha_j\xi)$ $(j=1,2,\ldots,n)$ i.e.,

$$\alpha_j(b_h+\nu)\neq\beta_h(a_j-1-\lambda) \tag{2.6}$$

$(\nu,\lambda=0,1,2,\ldots,\ h=1,2,\ldots,m,\ j=1,2,\ldots,n)$.

T is the contour in the complex ξ-plane such that the points

$$\xi=\frac{b_h+\nu}{\beta_h} \quad (\nu=0,1,\ldots,\ h=1,2,\ldots,m) \tag{2.7}$$

which are the poles of $\Gamma(b_h-\beta_h\xi)$ $(h=1,2,\ldots,m)$ lie to the right and the points

$$\xi=\frac{a_j-1-\nu}{\alpha_j} \quad (j=1,2,\ldots,n,\ \nu=0,1,\ldots) \tag{2.8}$$

which are the poles of $\Gamma(1-a_j+\alpha_j\xi)$ $(j=1,2,\ldots,n)$ lie to the left of T. Contour T runs from $\xi=\infty-iK$ to $\xi=\infty+iK$, where K is constant such that

$$K>\frac{\operatorname{Im}(b_j)}{\beta_j} \quad (j=1,2,\ldots,m).$$

The conditions for the contour T can be fulfilled on account of (2.6). The contour T may also be taken to run from $\xi=\sigma_1-i\infty$ to $\xi=\sigma_1+i\infty$ (σ_1 real) provided the conditions (2.6) are satisfied.

Recently Gupta and Jain [4] have denoted (2.3) symbolically as

$$H_{p,q}^{m,n}\left[z\left|\begin{matrix}(a_1,\alpha_1),\ldots,(a_p,\alpha_p)\\(b_1,\beta_1),\ldots,(b_q,\beta_q)\end{matrix}\right.\right] \tag{2.9}$$

or, in more compact form by

$$H_{p,q}^{m,n}\left[z\left|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right.\right] \quad \text{or} \quad H_{p,q}^{m,n}(z), \tag{2.10}$$

where $\{(f_r,d_r)\}$ stands for the set of parameters (f_1,d_1), $(f_2,d_2),\ldots,(f_r,d_r)$. A detailed account of the analytic continuation and asymptotic expansions of H-function as defined in (2.3) have been given by Braaksma [2]. However, we mention here a few results which will be useful in our present investigations.

According to Braaksma [2], the H-function makes sense and defines an analytic function of z in the following two cases:

(i) for every $z\neq 0$, if λ is positive, where

$$\lambda=\sum_{j=1}^{q}\beta_j-\sum_{j=1}^{p}\alpha_j \tag{2.11}$$

(ii) if $\lambda=0$ and

$$0<|z|<\beta^{-1}, \tag{2.12}$$

where

$$\beta=\prod_{j=1}^{p}(\alpha_j)^{\alpha_j}\prod_{j=1}^{q}(\beta_j)^{-\beta_j}. \tag{2.13}$$

From [6] it follows that

$$H_{p,q}^{m,n}(z)=O|z|^{e} \quad \text{for small } z, \tag{2.14}$$

where $\lambda\geqq 0$ and $e=\min \operatorname{Re}(b_h/\beta_n)$ $(h=1,2,\ldots,m)$. Also from [6] it is seen

$$H_{p,q}^{m,n}(z)=O|z|^{\eta} \quad \text{for large } z, \tag{2.15}$$

where $\lambda>0$ and $\eta=\min \operatorname{Re}[(a_j-1)/\alpha_j]$ $(j=1,2,\ldots,n)$ and the following conditions are satisfied.

$$\phi\equiv\left[\sum_{1}^{n}(\alpha_j)-\sum_{n+1}^{p}(\alpha_j)+\sum_{1}^{m}(\beta_j)-\sum_{m+1}^{q}(\beta_j)\right]>0, \tag{2.16}$$

$$|\arg z|<\tfrac{1}{2}\phi\pi. \tag{2.17}$$

3. Distributional H-Transform

Let $f(t)$ be a locally integrable function satisfying the following set of conditions A:

(i) $f(t)=0$ for $-\infty<t<T$

(ii) there exists a real number c such that the function

$$f(t)\exp(-wct)\,H_{p,q}^{m,n}\left[\rho(ct)^{\mu}\left|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right.\right]$$

is absolutely integrable over $-\infty<t<\infty$, w, ρ and μ being positive numbers.

Let s be a complex number $s=\sigma+i\tau$. The H-transformation is an operation $\mathscr{H}$ that assigns a function $F(s)$ of the complex variable s to each locally integrable function $f(t)$ satisfying condition A. $\mathscr{H}$ is defined by

$$\mathscr{H}[f(t)] \triangleq F(s) \triangleq \int_{-\infty}^{\infty} \exp(-wst) H_{p,q}^{m,n}\left[\rho(st)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] f(t)\,dt, \tag{3.1}$$

where only sectors with exponentially small behaviour of H-function at infinity should be considered. The function $F(s)$ is called the H-transform of the function $f(t)$. The first of the conditions A allows us to write (3.1) as

$$\mathscr{H}[f(t)] \triangleq F(s) \triangleq \int_{T}^{\infty} \exp(-wst) H_{p,q}^{m,n}\left[\rho(st)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] f(t)\,dt, \tag{3.2}$$

(3.2) is known as the right sided H-transform since the lower limit on the integral is finite.

4. An Initial Value Theorem for the H-Transform

Theorem 1. *If* (i) $f(t)\to 0$ *as* $t\to\infty$ (ii) $f(t)/t^{\eta}$ is absolutely integrable over $0\leqq t<\infty$ *and* (iii) *if there exist a complex number* α *and a real number* η $(\eta>-1)$ such that

$$\lim_{t\to 0+} \frac{f(t)}{t^{\eta}} = \alpha \tag{4.1}$$

then

$$\lim_{\sigma\to\infty} \frac{(w^{\sigma})^{\eta+1} F(\sigma)}{H(\rho/w^{\mu})} = \alpha \tag{4.2}$$

where

$$H(\rho/w^{\mu}) = H_{p+1,q}^{m,n+1}\left[\frac{\rho}{w^{\mu}}\middle|\begin{matrix}(-\eta,\mu)\,\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] \tag{4.3}$$

and

$$F(\sigma) \triangleq \int_{0}^{\infty} \exp(-w\sigma t) H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] f(t)\,dt. \tag{4.4}$$

Proof. By using the result given by Gupta and Jain [4] and the identity [6]

$$H_{0,1}^{1,0}[z|(0,1)] = \exp(-z),$$

we get

$$\int_{0}^{\infty} t^{\eta} \exp(-w\sigma t) H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] dt$$

$$= \frac{1}{(w\sigma)^{\eta+1}} H_{p+1,q}^{m,n+1}\left[\frac{\rho}{w^{\mu}}\middle|\begin{matrix}(-\eta,\mu)\,\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right] \tag{4.5}$$

provided

$$\operatorname{Re}\left(\eta+1+\mu\frac{b_j}{\beta_j}\right)>0 \qquad (j=1,2,\ldots,m),$$

$$\operatorname{Re}\left[(\eta+1)+\mu\frac{a_k-1}{\alpha_k}\right]<0 \qquad (k=1,2,\ldots,n),$$

$\mu>0, |\arg \rho s^{\mu}|<\frac{1}{2}\phi\pi$ where ϕ is defined in (2.16) and $|\arg ws|<\frac{\pi}{2}$.

Let $\sigma>0$ and $y>0$. We may write

$$\begin{aligned}
&|(w\sigma)^{\eta+1}F(\sigma)-\alpha H(\rho/w^{\mu})| \\
&\quad=\left|(w\sigma)^{\eta+1}\int_0^{\infty}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]f(t)\,dt\right. \\
&\qquad\left.-\alpha(w\sigma)^{\eta+1}\int_0^{\infty}t^{\eta}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]dt\right| \qquad (4.6)
\end{aligned}$$

{by using (4.3), (4.4) and (4.5)}

$$\begin{aligned}
&=\left|(w\sigma)^{\eta+1}\int_0^{\infty}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]\{f(t)-\alpha t^{\eta}\}\,dt\right| \\
&=\left|(w\sigma)^{\eta+1}\int_0^{\infty}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]t^{\eta}\left\{\frac{f(t)}{t^{\eta}}-\alpha\right\}dt\right| \\
&\leqq(w\sigma)^{\eta+1}\int_0^{y}\left|t^{\eta}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]\right|\left|\frac{f(t)}{t^{\eta}}-\alpha\right|dt \\
&\quad+(w\sigma)^{\eta+1}\int_y^{\infty}\left|t^{\eta}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]\right|\left|\frac{f(t)}{t^{\eta}}-\alpha\right|dt \\
&=I_1+I_2. \qquad (4.7)
\end{aligned}$$

Now

$$\begin{aligned}
I_1&\leqq w^{\eta+1}\int_0^{y}\left|(\sigma t)^{\eta}\exp(-w\sigma t)\,H_{p,q}^{m,n}\left[\rho(\sigma t)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]d(\sigma t)\right|\sup_{0\leqq t\leqq y}\left|\frac{f(t)}{t^{\eta}}-\alpha\right| \\
&\leqq w^{\eta+1}\int_0^{\sigma y}\left|z^{\eta}\exp(-wz)\,H_{p,q}^{m,n}\left[\rho z^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]\right|dz\sup_{0\leqq t\leqq y}\left|\frac{f(t)}{t^{\eta}}-\alpha\right|,
\end{aligned}$$

where $\sigma t=z$.

In view of (2.14), the integral I_1 is convergent, if $\operatorname{Re}(\eta+\mu e)>0$ where e is defined in (2.14). Since from (4.1) $\lim_{t\to 0+}\dfrac{f(t)}{t^{\eta}}=\alpha$ we have, for $\varepsilon>0$ a pre-assigned number, and choosing y sufficiently small such that

$$\sup_{0\leqq t\leqq y}\left|\frac{f(t)}{t^{\eta}}-\alpha\right|<\varepsilon_1$$

where ε_1 is very small number therefore $I_1 < \varepsilon/2$. Again

$$I_2 \leqq (w\sigma)^{\eta+1} \int_y^\infty \left| t^\eta \exp(-w\sigma t) H_{p,q}^{m,n}\left[\rho(\sigma t)^\mu \middle| \begin{matrix}\{(a_p, \alpha_p)\}\\ \{(b_q, \beta_q)\}\end{matrix}\right]\right| dt \sup_{y \leqq t \leqq \infty} \left| \frac{f(t)}{t^\eta} - \alpha \right|.$$

The function

$$t^\eta H_{p,q}^{m,n}\left[\rho(\sigma t)^\mu \middle| \begin{matrix}\{(a_p, \alpha_p)\}\\ \{(b_q, \beta_q)\}\end{matrix}\right]$$

is bounded on $0 < t < \infty$, provided $\mathrm{Re}(\eta + \mu e) > 0$ where e is defined in (2.14). After fixing y such that $I_1 < \varepsilon/2$, we can choose σ very large so that $I_2 < \varepsilon/2$. From (4.6) and (4.7), we get

$$|(w\sigma)^{\eta+1} F(\sigma) - \alpha H(\rho/w^\mu)| < \varepsilon \tag{4.8}$$

for large σ. Since ε is arbitrary number, we conclude (4.2) from (4.8). The following Lemma will be used in extending this result to distributions.

Lemma. *Let $f(t)$ be H-transformable distribution with its support in $y \leqq t < \infty$ where $y > 0$. Let a number C be such that $\exp(-wct) f(t)$ is a distribution of slow growth* [1]. *Then for $c+1 < \sigma < \infty$,*

$$|F(\sigma)| \leqq M \sigma^\gamma \exp[-w(\sigma - c)\tau], \tag{4.9}$$

where M is constant and τ is any real number satisfying $0 < \tau < y$.

Proof. For $\sigma > c$,

$$F(\sigma) \triangleq \left\langle f(t), \exp(-w\sigma t) H_{p,q}^{m,n}\left[\rho(\sigma t)^\mu \middle| \begin{matrix}\{(a_p, \alpha_p)\}\\ \{(b_q, \beta_q)\}\end{matrix}\right]\right\rangle$$

$$\triangleq \langle \exp(-wct) f(t), \lambda(t) \exp[-w(\sigma - c)t] H_{p,q}^{m,n}[\rho(\sigma t)^\mu] \rangle.$$

We are free to choose $\lambda(t)$ such that its support is contained in $x \leqq t < \infty$, where x is any fixed number satisfying $0 < x < y$. Moreover a distribution of slow growth satisfies the boundedness condition. We therefore have

$$|F(\sigma)| \leqq C \sup_{x \leqq t < \infty} \left| (1+t^2)^r \frac{d^r}{dt^r} \left\{ \lambda(t) \exp[-w(\sigma - c)t] H_{p,q}^{m,n}\left[\rho(\sigma t)^\mu \middle| \begin{matrix}\{(a_p, \alpha_p)\}\\ \{(b_q, \beta_q)\}\end{matrix}\right]\right\}\right|$$

$$\leqq C \sup_{x \leqq t < \infty} \left| (1+t^2)^r \sum_{k=0}^{r} \binom{r}{k} \lambda^{r-k}(t) \frac{d^k}{dt^k} \{\exp[-w(\sigma - c)t] H_{p,q}^{m,n}[\rho(\sigma t)^\mu]\}\right|$$

$$\leqq C \sup_{x \leqq t < \infty} \left| (1+t^2)^r \sum_{k=0}^{r} \binom{r}{k} \lambda^{r-k}(t) \sum_{j=0}^{k} \binom{k}{j} \right.$$

$$\left. \cdot [w(\sigma - c)]^{k-j} \exp[-w(\sigma - c)t] \frac{d^j}{dt^j} H_{p,q}^{m,n}[\rho(\sigma t)^\mu] \right|$$

by using Leibnitz's rule for the n^{th} derivative of the product of two functions.

$$\therefore |F(\sigma)| \leqq C \sup_{x \leqq t < \infty} \left| (1+t^2)^r \sum_{k=0}^{r} \binom{r}{k} \lambda^{r-k}(t) \right.$$

$$\cdot \sum_{j=0}^{k} \binom{k}{j} [w(\sigma-c)]^{k-j} \exp[-w(\sigma-c)(t-x) - w(\sigma-c)x]$$

$$\left. \cdot \frac{d^j}{dt^j} H_{p,q}^{m,n}[\rho(\sigma t)^\mu] \right|. \tag{4.10}$$

If τ is a fixed number $0 < \tau < x$ then

$$w(\sigma-c)^{k-j} \cdot \exp[-w(\sigma-c)]\, t \leqq B \exp[-w(\sigma-c)\tau] \tag{4.11}$$

for each $j = 1, 2, \ldots, k$ where B is constant independent of j. Also since $\sigma > c + 1$.

$$\exp[-w(\sigma-c)(t-x)] \leqq \exp[-w(t-x)]. \tag{4.12}$$

By the differential formulae [6] and the asymptotic expansions of H function as given in Sect. 2,

$$\frac{d^j}{dt^j} H_{p,q}^{m,n}\left[\rho(\sigma t)^\mu \middle| \begin{matrix}(a_p, \alpha_p)\\ (b_q, \beta_q)\end{matrix}\right] \tag{4.13}$$

is bounded and continuous function on $0 < t < \infty$. On using (4.11), (4.12) and (4.13), the result (4.10) can be written as

$$|F(\sigma)| \leqq CB \exp[-w(\sigma-c)\tau] \sup_{x \leqq t < \infty} |(1+t^2)^r \sum_{k=0}^{r} \binom{r}{k}$$

$$\cdot \lambda^{r-k}(t) \sum_{j=0}^{k} \binom{k}{j} \sigma^j \rho^{j/\mu} \exp[-w(t-x)] \frac{d^j}{dt^j} H_{p,q}^{m,n}\left[z^\mu \middle| \begin{matrix}\{(a_p, \alpha_p)\}\\ \{(b_q, \beta_q)\}\end{matrix}\right] \tag{4.14}$$

where $\sigma t = z/\rho^{1/\mu}$,

$$|F(\sigma)| \leqq M \sigma^r \exp[-w(\sigma-c)\tau] \tag{4.15}$$

since the function on the right side of (4.14) is bounded. Here M is sufficiently large number and the numbers C and r depend only on f.

Theorem 2. *Let $f(t)$ be H-transformable distribution having its support in $0 \leqq t < \infty$ and assume that over some neighbourhood of the origin, $f(t)$ is a regular distribution corresponding to a Lebesque integrable function $h(t)$. Further assume that there exists a complex number α and a real number η ($\eta > -1$) such that*

$$\lim_{t \to 0^+} \frac{h(t)}{t^\eta} = \alpha \tag{4.16}$$

then

$$\lim_{\sigma \to \infty} \frac{(w\sigma)^{\eta+1} F(\sigma)}{H(\rho/w^\mu)} = \alpha \tag{4.17}$$

where $H(\rho/w^\mu)$ and $F(\sigma)$ are defined in (4.3) *and* (4.4) *respectively.*

Proof. Let the neighbourhood over which $f(t)$ is a regular distribution be $-\infty<t<T$ $(T>0)$. Let $0<y<T$. Then $f(t)$ can be decomposed as

$$f(t)=f_1(t)+f_2(t) \tag{4.18}$$

where the supports of $f_1(t)$ and $f_2(t)$ are contained in the respective intervals $0\leqq t\leqq y$ and $y<t<\infty$. Thus $f_1(t)$ is a regular distribution that corresponds to $h(t)$ for $0<t<y$. Let

$$F_1(s)\triangleq\mathscr{H}[f_1(t)]$$

and

$$F_2(s)\triangleq\mathscr{H}[f_2(t)].$$

By using (4.9) we find that

$$(w\sigma)^{\eta+1}F_2(\sigma)<Mw^{\eta+1}\sigma^r\sigma^{\eta+1}\exp[-w(\sigma-c)\tau]\to 0 \quad \text{as} \quad \sigma\to\infty,$$

so

$$\lim_{\sigma\to\infty}\frac{(w\sigma)^{\eta+1}F_2(\sigma)}{H(\rho/w^{\mu})}=0.$$

Therefore

$$\lim_{\sigma\to\infty}\frac{(w\sigma)^{\eta+1}F(\sigma)}{H(\rho/w^{\mu})}=\lim_{\sigma\to\infty}\frac{(w\sigma)^{\eta+1}[F_1(\sigma)+F_2(\sigma)]}{H(\rho/w^{\mu})}$$

$$=\lim_{\sigma\to\infty}\frac{(w\sigma)^{\eta+1}F_1(\sigma)}{H(\rho/w^{\mu})}=\alpha \quad \text{(by Theorem 1).}$$

Since the ordinary H-transform of Lebesque integrable function is identical to the distributional H-transform of the corresponding regular distribution we can use Theorem 1.

5. A Final Value Theorem for the H-Theorem

Theorem 3. *Let $f(t)$ be a measurable function on $0<t<\infty$. Let there exist a real number c such that*

$$\exp(-wct)H_{p,q}^{m,n}\left[\rho(ct)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]$$

is absolutely convergent over $-\infty<t<\infty$. Further if there exist a complex number α and a real number η $(\eta>-1)$ such that

$$\lim_{t\to\infty}\frac{f(t)}{t^{\eta}}=\alpha \tag{5.1}$$

then

$$\lim_{\sigma\to 0+}\frac{(w\sigma)^{\eta+1}F(\sigma)}{H(\rho/w^{\mu})}=\alpha \tag{5.2}$$

where $H(\rho/w^{\mu})$ and $F(\sigma)$ are defined in (4.3) *and* (4.4) *respectively.*

Proof. From (5.1) it is clear that $f(t)$ is a function of slow growth. Let σ and y be positive quantities. By proceeding on the same lines as in Theorem 1, we obtain

$$\begin{aligned}
&|(w\sigma)^{\eta+1}F(\sigma)-\alpha H(\rho/w^{\mu})| \\
&\leqq (w\sigma)^{\eta+1}\int_0^y |t^{\eta}\exp(-w\sigma t)\,H_{p,q}^{m,n}[\rho(\sigma t)^{\mu}]|dt \sup_{0<t<y}\left|\frac{f(t)}{t^{\eta}}-\alpha\right| \\
&\quad + w^{\eta+1}\int_{\sigma y}^{\infty} |z^{\eta}\exp(-wz)\,H_{p,q}^{m,n}[\rho z^{\mu}]|dz \sup_{0\leqq t<\infty}\left|\frac{f(t)}{t^{\eta}}-\alpha\right| \qquad (5.3)
\end{aligned}$$

where $\sigma t=z$. Let $\varepsilon>0$ be arbitrarily chosen. By virtue of (5.1) y can be chosen so large that the second term on the right hand of (5.3) which is independent of σ can be made less than $\varepsilon/2$. Having fixed y in this way we then choose σ so small that the first term is also less than $\varepsilon/2$. From (5.3) we get

$$|(w\sigma)^{\eta+1}F(\sigma)-\alpha H(\rho/w^{\mu})|<\varepsilon \quad \text{for small } \sigma.$$

Since ε is arbitrary, we conclude that

$$\lim_{\sigma\to 0^+}\frac{(w\sigma)^{\eta+1}F(\sigma)}{H(\rho/w^{\mu})}=\alpha.$$

This completes the proof of Theorem 3.

Theorem 4. *Let $f(t)$ be H-transformable distribution in D'_R i.e. in the space of distributions whose supports are bounded on left. Assume that over some semi-infinite interval $\tau\leqq t<\infty$, $f(t)$ is a regular distribution corresponding to the Lebesque integrable function $h(t)$. Let $h(t)$ satisfy the following condition.*

$$h(t)\exp(-wct)H_{p,q}^{m,n}\left[\rho(ct)^{\mu}\middle|\begin{matrix}\{(a_p,\alpha_p)\}\\ \{(b_q,\beta_q)\}\end{matrix}\right]$$

is integrable over $-\infty<t<\infty$. Further assume that there exists a complex number α and a real number η $(\eta>-1)$ such that

$$\lim_{t\to\infty}\frac{h(t)}{t^{\eta}}=\alpha, \qquad (5.4)$$

then

$$\lim_{\sigma\to 0+}\frac{(w\sigma)^{\eta+1}F(\sigma)}{H(\rho/w^{\mu})}=\alpha. \qquad (5.5)$$

Proof. Let $f(t)$ be decomposed into

$$f(t)=f_1(t)+f_2(t)$$

where $f_2(t)$ has its support contained in $y\leqq t<\infty$ $(y>\tau)$ and $f_1(t)$ has a bounded support that does not extend to the right of $t=y$. Since $F_1(s)\triangleq\mathscr{H}[f_1(t)]$ is an entire function of s so that

$$\lim_{\sigma\to 0+}(w\sigma)^{\eta+1}F_1(\sigma)=0. \qquad (5.6)$$

Therefore with $F_2(s) \triangleq \mathscr{H}[f_2(t)] \triangleq \mathscr{H}[h(t), 1_+(t-y)]$ where $1_+(t)$ is a Heaviside unit step function. Hence in view of (5.6) we have

$$\lim_{\sigma \to 0+} (w\sigma)^{\eta+1} F(\sigma) = \lim_{\sigma \to 0+} (w\sigma)^{\eta+1} F_2(\sigma) = \alpha H\left(\frac{\rho}{\omega^{\mu}}\right)$$

by virtue of (5.2). This completes the proof of (5.5).

6. Particular Cases

1) Setting $w=\rho=1/2$, $\mu=1$, $m=q=1$, $n=p=0$, $b_1=0$, $\beta_1=1$, we obtain the Abelian theorems for Laplace transform given by Zemanian [13].

2) With $w=1$, $\rho=2$, $\mu=1$, $m=q=2$, $n=p=1$, $a_1=1$, $\alpha_1=1$, $b_1=\frac{1}{2}+\mu$, $b_2=\frac{1}{2}-\mu$, $\beta_1=\beta_2=1$, the results of this paper yield the Abelian theorems for K-transform given by Zemanian [14].

3) Putting $\rho=2$, $\mu=1$, $\alpha_i=1$ $(i=1,2,\ldots,p)$ $\beta_j=1$ $(j=1,\ldots,q)$, we obtain the results given by Mishra [8] for Meijer-Laplace transform.

4) On putting $\rho=2$, $\mu=1$, the results given by Saxena [10] are obtained.

5) On putting $\rho=2$, $w=\mu=n=p=1$, $m=q=2$, $a_1=k+\frac{3}{4}$, $b_1=\frac{1}{4}+m$, $b_2=\frac{1}{4}-m$, $\alpha_1=\beta_1=\beta_2=1$, we obtain the Abelian theorems for Whittaker transform [11]. The Abelian theorems for this transform have been discussed by Moharir and Saxena [9] in non-distributional form.

6) By choosing $w=\rho=n=p=1$, $m=q=2$, $a_1=k+v+\frac{1}{2}$, $\mu=1$, $b_1=2v$, $b_2=0$, $\alpha_1=\beta_1=\beta_2=1$, the results of this paper get reduced to the Abelian theorems for Verma transform [12] given by Akhuri [1].

7) With $w=\rho=\mu=n=p=1$, $m=q=2$, $a_1=1$, $b_1=m-k$, $b_2=-m-k$, $\alpha_1=\beta_1=\beta_2=1$, we obtain the corresponding results for Meijer transformation [7].

8) By putting $w=\rho=\mu=n=p=\alpha_1=\beta_1=\beta_2=1$, $m=q=2$, $a_1=k+\lambda+1/2$, $b_1=\lambda+m$, $b_2=\lambda-m$, we get the corresponding results for generalized Laplace transform given by Mainra [5].

Note

In view of definition of the H-function given in Sect. 2 and (4.3), it is obvious that w can not be given the value zero as the H-function is defined for non-zero argument. Therefore the special cases with $w=0$ as discussed by Saxena [10] and Mishra [8] are not included here.

References

1. Akhuri, S.K.: Ph. D. Thesis, Ranchi University, Ranchi (1970)
2. Braaksma, B.L.J.: Asymptotic expansion and analytic continuation for the class of Barnes integrals. Composito Math. **15**, 239-341 (1964)
3. Fox, C.: The G and H functions as Symmetrical Fourier Kernels. Trans. Am. Math. Soc. **98**, 395-429 (1961)

4. Gupta, K.C., Jain, U.C.: The H-function-II. Proc. Nat. Akad. Sci. India, Sect. A**36**, 661-674 (1966)
5. Mainra, V.P.: A new generalization of Laplace transform. Bull. Calcutta Math. Soc. **53**, 23-31 (1961)
6. Mathai, A.M., Saxena, R.K.: The H-function with applications in statistics and other disciplines. New York: Wiley 1978
7. Meijer, C.S.: Eine neue Erweiterung der Laplace Transformation I. Proc. Nederl. Akad. Wet. Amsterdam **44**, 729-737 (1941)
8. Mishra, O.P.: Some Abelian theorems for the distributional Meijer-Laplace transformation. Indian J. Pure Appl. Math. **3**, 241-247 (1972)
9. Moharir, S.K., Saxena, R.K.: Abelian theorems for the generalized Whittaker transform. J. Shivaji Univ. **18**, 7-11 (1979)
10. Saxena, R.K. (Jodhpur): Abelian theorems for H-transform. Acta Mexicana Ci. Tecn. **7**, 66-76 (1973)
11. Varma, R.S.: A generalization of Laplace transform. Current Sci. **16**, 17-18 (1947)
12. Varma, R.S.: On a generalization of Laplace integral. Proc. Nat. Akad. Sci. India, Sect. A**20**, 209-216 (1951)
13. Zemanian, A.H.: Distribution theory and transform analysis. New York: Mc-Graw Hill 1965
14. Zemanian, A.H.: Some Abelian theorems for Hankel and K-transformations. Siam J. Appl. Math. **14**, 1255-1265 (1966)

Received August 4, 1980

Math. Ann. 256, 323–340 (1981)

Mathematische Annalen
© Springer-Verlag 1981

On the Poles of Andrianov L-Functions

Takayuki Oda

Department of Mathematics, Hokkaido University, Sapporo 060, Japan

Introduction

In [1–3], Andrianov proved functional equations for certain L-functions $Z_F(s)$ attached to Siegel modular cusp forms F of weight k and of degree 2, which are common eigenfunctions of all Hecke operators. Moreover he proved that $\Psi_F(s)=(2\pi)^{-s}\Gamma(s)\Gamma(s-k+2)Z_F(s)$ is holomorphic when k is odd, and $\Psi_F(s)$ is holomorphic except possible simple poles at $s=k$ and $s=k-2$ when k is even. In this paper, we investigate the latter case more precisely, applying the results of [13]. To state the outline of our results, we recall the results of [13].

Let Q be a quadratic form on $\mathbf{Z}^n$, with values in even integers $2\mathbf{Z}$, and of signature $(2+,(n-2)-)$. Let Γ_Q be the group of the proper units of Q, which is a discrete subgroup of the identity component G of the real orthogonal group $\mathrm{SO}(Q)$ for Q. Let D be a bounded symmetric domain attached to Q, and let $S_k(\Gamma_Q)$ be the space of holomorphic cusp forms on D of weight k with respect to Γ_Q. Assume that n is odd and let $S_{2k-(n-4)}(\Gamma_0(N'),\chi_Q)$ be the space of elliptic modular cusp forms of half integral weight $\frac{1}{2}\{2k-(n-4)\}$ with respect to $\Gamma_0(N')$ with multiplicator χ_Q, where N' is the least common multiple of 4 and the level of Q, and χ_Q is a multiplicator defined in [13].

In [13], we considered a certain theta series $\theta_k(z,x)$ with spherical functions attached to the quadratic form Q, which has an automorphic property in the variables $(z,x)\in H\times D$ with respect to $\Gamma_0(N')\times\Gamma_Q$, where H is the complex upper half plane. Then, considering the pairings with respect to the Petersson metrics, we put

$$\iota(f)(x)=\int_{\Gamma_0(N')\backslash H}\overline{\theta_k(z,x)}f(z)v^{k-(n-4)/2}\frac{dudv}{v^2}\qquad(z=u+\sqrt{-1}v),$$

and

$$\varrho(F)(z)=\int_{\Gamma_Q\backslash D}\theta_k(z,x)F(x)\eta(x)^k dx,$$

where dx is the invariant measure on D, and $\eta(x)$ is a certain function of x which is written as a quadratic function of the imaginary parts of the coordinates of x, if D

0025-5831/81/0256/0323/$03.60

is realized as a tube domain. It was shown that the linear mappings $\iota: S_{2k-(n-4)}(\Gamma_0(N'), \chi_Q) \to S_k(\Gamma_Q)$ and $\varrho: S_k(\Gamma_Q) \to S_{2k-(n-4)}(\Gamma_0(N'), \chi_Q)$ can be defined. And clearly from the definitions, these two mappings are mutually adjoint for the Petersson metrics on $S_{2k-(n-4)}(\Gamma_0(N'), \chi_Q)$ and $S_k(\Gamma_Q)$.

Let $I_k(\Gamma_Q)$ be the image of ι, and let $P_k(\Gamma_Q)$ be the kernel of ϱ. By the adjointness of ι and ϱ, we have a direct sum decomposition of $S_k(\Gamma_Q)$:

$$S_k(\Gamma_Q) = I_k(\Gamma_Q) \oplus P_k(\Gamma_Q),$$

and $I_k(\Gamma_Q)$ and $P_k(\Gamma_Q)$ are mutually orthogonal for the Petersson metric on $S_k(\Gamma_Q)$.

Let Q be a quadratic form of 5 variables defined in Sect. 1 of this paper. Then $S_k(\Gamma_Q)$ coincides with the space $S_k(\mathrm{Sp}(2;\mathbf{Z}))$ of the Siegel cusp forms of weight k of degree 2. Let $S_k(\mathrm{Sp}(2;\mathbf{Z})) = I_k(\mathrm{Sp}(2;\mathbf{Z})) \oplus P_k(\mathrm{Sp}(2;\mathbf{Z}))$ be the corresponding decomposition of $S_k(\mathrm{Sp}(2;\mathbf{Z}))$. Then our main result is the following theorem.

Theorem. *The spaces $I_k(\mathrm{Sp}(2;\mathbf{Z}))$ and $P_k(\mathrm{Sp}(2;\mathbf{Z}))$ are stable under the action of the Hecke operators. Moreover, if F is a common eigenfunction of all Hecke operators, then $\Psi_F(s)$ has no poles at $s=k$ and $s=k-2$, or has simple poles at $s=k$ and $s=k-2$, according as $F \in P_k(\mathrm{Sp}(2;\mathbf{Z}))$ or $F \in I_k(\mathrm{Sp}(2;\mathbf{Z}))$ (see, Proposition* (3.1), *Corollary* (3.3), *Theorem* (4.1), *and Theorem* (4.5)).

In the last section, we show that our space $I_k(\mathrm{Sp}(2;\mathbf{Z}))$ is identical with a certain subspace M_k of Siegel cusp forms studied by Maass [9–11], and Andrianov [4].

The referee kindly pointed out to the author the paper of Evdokimov [17], in which similar results are obtained by different method, considering the Maass spaces M_k themselves, in place of our I_k.

1. Definitions and Preliminary Results

In this section, we recall some results of [13], which we need in later sections.

We take as Q of [13], the 5×5 symmetric matrix given by

$$Q = \begin{pmatrix} & & & & 4 \\ & & -2 & & \\ & 4 & & & \\ 4 & & & & \end{pmatrix},$$

which is even integral, and of signature $(2+, 3-)$. Since

$$4Q^{-1} = \begin{pmatrix} & & & & 1 \\ & & & 1 & \\ & & -2 & & \\ & 1 & & & \\ 1 & & & & \end{pmatrix},$$

the level of Q is 4.

Let L be the module of integral row vectors of size 5. For any subring A of the complex number field $\mathbf{C}$, we put $L_A = L \otimes_{\mathbf{Z}} A$. The dual module of L_A over A is denoted by L_A^*.

For any vectors $\ell=(a, b, c, d, e)$ and $\ell'=(a', b', c', d', e')$ of $L_A(a, b, \ldots, d', e'\in A)$ we put

$$Q(\ell)=\ell Q{}^t\ell=8ae+8bd-2c^2,$$

and

$$B(\ell, \ell')=4ae'+4a'e+4bd'+4b'd-2cc'.$$

Then $Q(*)$ is a quadratic form on L_A, and $B(*, *)$ is the symmetric bilinear form associated with Q. Note that the values of $Q(*)$ on L are even integers.

Let G be the identity component of the real orthogonal group $\{g\in M_5(\mathbf{R})|{}^tgQg=Q\}$ for Q. Put $\Gamma_Q=M_5(\mathbf{Z})\cap G$. Then Γ_Q is a discrete subgroup of G. We sometimes omit the subscript Q of Γ_Q, and write $\Gamma=\Gamma_Q$. G acts on $L_{\mathbf{R}}$ by

$$g(v)=v^tg \quad \text{for} \quad v\in L_{\mathbf{R}} \quad \text{and} \quad g\in G,$$

and preserves the quadratic form $Q(*)$.

Let us define the classical domain attached to G. By transport of the structure, the vector space $L_{\mathbf{C}}^*$ dual to $L_{\mathbf{C}}$ is equipped with the inverse quadratic form $Q^*(*)$, defined by

$$Q^*(v^*)=\tfrac{1}{2}(\alpha\varepsilon+\beta\delta-\gamma^2)$$

for $v^*=(\alpha, \beta, \gamma, \delta, \varepsilon)$ of $L_{\mathbf{C}}^*$. Then the symmetric bilinear form $B^*(*, *)$ associated to Q^* is given by

$$B^*(v^*, v^{*\prime})=\tfrac{1}{4}(\alpha\varepsilon'+\alpha'\varepsilon+\beta\delta'+\beta'\delta-2\gamma\gamma')$$

for $v^*=(\alpha, \beta, \gamma, \delta, \varepsilon)$ and $v^{*\prime}=(\alpha', \beta', \gamma', \delta', \varepsilon')$ of $L_{\mathbf{C}}^*$.

Let $\tilde{D}^*$ be the subset of $L_{\mathbf{C}}^*$ defined by

$$\tilde{D}^*=\{\xi=(\xi_0, \xi_1, \xi_2, \xi_3, \xi_4)\in L_{\mathbf{C}}^*|Q(\xi)=0, B(\xi, \bar{\xi})>0\},$$

where $\bar{\xi}=(\bar{\xi}_0, \bar{\xi}_1, \bar{\xi}_2, \bar{\xi}_3, \bar{\xi}_4)$ and $\bar{\xi}_i$ is the complex conjugate of ξ_i. Consider the canonical projection

$$\Pi: L_{\mathbf{C}}^*-\{0\}\to\mathbb{P}^4$$

of $L_{\mathbf{C}}^*-\{0\}$ to the complex projective space $\mathbb{P}^4$ of dimension 4. Note that $\tilde{D}^*$ does not contain the origin $\{0\}$ of $L_{\mathbf{C}}^*$. Let D_H^* be the image of $\tilde{D}^*$ by Π. Then the restriction of Π to $\tilde{D}^*$ defines a $\mathbf{C}^*$-bundle over D_H^*. It is easy to check that D_H^* is the union of two connected components D_H^+ and D_H^-.

The group G acts on $\tilde{D}^*$ by

$$g(\xi)=\xi g^{-1} \quad \text{for} \quad \xi\in\tilde{D}^* \quad \text{and for} \quad g\in G,$$

and resultingly on D_H^* through Π.

G acts transitively on each of the two components of D_H^*, and either D_H^+ or D_H^- is isomorphic to the quotient G/K of G by a maximal compact subgroup K of G.

Let us realize $D_H^{\pm}$ as a tube domain in $\mathbf{C}^3$. Let $\mathcal{Q}$ be the quadric in $\mathbb{P}^4$ defined by

$$\mathcal{Q}=\{\xi=(\xi_0:\xi_1:\xi_2:\xi_3:\xi_4)\in\mathbb{P}^4|Q(\xi)=0\},$$

and let $\mathscr{H}$ be the hyperplane in $\mathbb{P}^4$ defined by the equation

$$\xi_4=0.$$

A simple calculation shows that D_H^* does not intersect with $\mathscr{H}$. Thus D_H^* is contained in $\mathscr{Q}-\mathscr{Q}\cap\mathscr{H}$. Moreover the mapping

$$\mu: z=(z_1,z_2,z_3)\in\mathbf{C}^3\mapsto(z_2^2-z_1z_3: z_1: z_2: z_3: 1)\in\mathscr{Q}$$

defines a biholomorphic isomorphism of $\mathbf{C}^3$ to $\mathscr{Q}-\mathscr{Q}\cap\mathscr{H}$. Let D^* and $D^{\pm}$ be the inverse images $\mu^{-1}(D_H^*)$ and $\mu^{-1}(D_H^{\pm})$ of D_H^* and D_H^-, respectively. Then D^* is given by

$$D^*=\{z=(z_1,z_2,z_3)\in\mathbf{C}^3|\operatorname{Im}z_1\operatorname{Im}z_3-(\operatorname{Im}z_2)^2>0\},$$

and by

$$D^{\pm}=\{z=(z_1,z_2,z_3)\in D^*|\pm(\operatorname{Im}z_1+\operatorname{Im}z_3)>0\}.$$

Note that D^+ and D^- are equal to the Siegel upper and lower half spaces of degree 2, as a pair.

Next we define an automorphy factor on $G\times D^*$. We define an action of G on D^* by transport of the structure through the isomorphism μ. Let $\tilde{\mu}$ be the mapping defined by

$$\tilde{\mu}: z=(z_1,z_2,z_3)\in\mathbf{C}^3\mapsto(z_2^2-z_1z_3,z_1,z_2,z_3,1)\in\tilde{D}^*.$$

By definition, the vector $g(\tilde{\mu}(z))=\tilde{\mu}(z)g^{-1}$ is a non-zero scalar multiple of the vector $\tilde{\mu}(g(z))$ for any $g\in G$ and any $z\in D^*$. We denote by $j(g,z)$ this scalar depending only upon g and z. Then $j(g,z)$ is an automorphy factor, which is real analytic in g and holomorphic in z. In fact,

$$(gg')(\tilde{\mu}(z))=\tilde{\mu}(z)(gg')^{-1}=j(gg',z)\tilde{\mu}((gg')(z)),$$

and

$$\begin{aligned}(gg')(\tilde{\mu}(z))&=\tilde{\mu}(z)g'^{-1}g^{-1}=j(g',z)\tilde{\mu}(g'(z))^{-1}\\&=j(g',z)j(g,g'(z))\tilde{\mu}((gg')(z)).\end{aligned}$$

Hence $j(gg',z)=j(g,g'(z))j(g',z)$.

Now we define automorphic forms. From now on we fix one of two connected components D^+ and D^-, and denote it by D.

Definition (1.1). A holomorphic function $F(z)$ on D is called a holomorphic modular form of weight k with respect to Γ, if it satisfies

$$F(\gamma(z))=j(\gamma,z)^kF(z)\quad\text{for any}\quad\gamma\in\Gamma.$$

Here we observe that our concept of modular form is equivalent to that of Siegel modular forms of degree 2.

First we consider a free $\mathbf{Z}$-module H^1 with generators e_1, e_2, e_3, e_4:

$$H^1=\mathbf{Z}e_1\oplus\mathbf{Z}e_2\oplus\mathbf{Z}e_3\oplus\mathbf{Z}e_4,$$

and its exterior products H^i over $\mathbf{Z}$:

$$H^i = \overset{i}{\wedge} H^1 \ (i = 0, 1, 2, 3, 4).$$

We denote by tr the mapping

$$n(e_1 \wedge e_2 \wedge e_3 \wedge e_4) \mapsto n$$

of H^4 to $\mathbf{Z}$, and define a "polarization" vector p of H^2 by $p = e_1 \wedge e_2 + e_3 \wedge e_4$. Consider a skew symmetric bilinear form $\langle\ ,\ \rangle_p$ on H^1 defined by

$$H^1 \times H^1 \to H^2 \xrightarrow{\wedge p} H^4 \xrightarrow{\text{tr}} \mathbf{Z}$$

$$(a, b) \mapsto a \wedge b \mapsto a \wedge b \wedge p \to \text{tr}(a \wedge b \wedge p).$$

Then for any subring A of $\mathbf{C}$, we denote by $\mathrm{Sp}(2; A)$ the group of linear automorphisms of $H \otimes_{\mathbf{Z}} A$ over A which preserve $\langle\ ,\ \rangle_p$. We represent the elements of $\mathrm{Sp}(2; A)$ as matrices of size 4 with respect to the basis (e_1, e_2, e_3, e_4).

Now consider a symmetric bilinear form:

$$H^2 \times H^2 \to H^4 \xrightarrow{\text{tr}} \mathbf{Z}$$

$$(m, n) \mapsto m \wedge n \mapsto \text{tr}(m \wedge n).$$

Let L^* be the orthogonal complement of $\mathbf{Z}p$ in H^2 with respect to $(\ ,\ )$. Then L^* is a free module over $\mathbf{Z}$ with a basis $(e_1 \wedge e_3, e_1 \wedge e_4, e_1 \wedge e_2 - e_3 \wedge e_4, e_2 \wedge e_3, e_4 \wedge e_2)$. And the restriction $B^*(\ ,\ )$ of the bilinear form $(\ ,\ )$ to L^* is represented by the matrix $4Q^{-1}$ with respect to this basis.

Let $O(L_A^*)$ and $O(L_A)$ be the orthogonal groups for $B^*(\ ,\ )$ and $B(\ ,\ )$, respectively. Then we define an isomorphism

$$O(L_A^*) \to O(L_A)$$
$$g^* \mapsto g$$

by means of the equality

$$\langle g(\ell), g^*(\ell^*) \rangle = \langle \ell, \ell^* \rangle$$

for any ℓ of L_A and any ℓ^* of L_A^*. Here $\langle\ ,\ \rangle$ is the canonical paring $L_A \times L_A^* \to A$.

Consider the composition of this mapping with a natural mapping $\mathrm{Sp}(2; A) \to O(L_A^*)$. Then it is easy to check that this yield a double covering $\sigma : \mathrm{Sp}(2; \mathbf{R}) \to G$, whose kernel is equal to the center $\{\pm 1_4\}$ of $\mathrm{Sp}(2; \mathbf{R})$. Recalling the standard action of $\mathrm{Sp}(2; \mathbf{R})$ on D, it is also easy to check that

$$g_+(M_z) = (AM_z + B)(CM_z + D)^{-1} = M_{\sigma(g_+)(z)},$$
$$\det(CM_z + D) = j(\sigma(g_+), z),$$

and

$$\sigma(\mathrm{Sp}(2; \mathbf{Z})) \subset \Gamma_Q,$$

for any $g_+ = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ of $\mathrm{Sp}(2; \mathbf{R})$, where $M_z = \begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix}$ for $z = (z_1, z_2, z_3)$. By a direct

computation or by the maximality of $\mathrm{Sp}(2;\mathbf{Z})$ we can check that $\sigma(\mathrm{Sp}(2;\mathbf{Z}))=\Gamma_Q$. Thus Γ_Q is isomorphic to $\mathrm{Sp}(2;\mathbf{Z})/\{\pm 1_4\}$. So we sometimes refer to our modular forms as Siegel modular forms of degree 2.

For any vector ℓ of L, we denote by G_ℓ the stabilizer of ℓ in G defined by

$$G_\ell=\{g\in G|g(\ell)=\ell\},$$

and denote by Γ_ℓ the intersection $G_\ell\cap\Gamma$. Then Γ_ℓ is a discrete subgroup of G_ℓ. If $Q(\ell)>0$, G_ℓ is isomorphic to the proper Lorentz group $\mathrm{SO}_0(1,3)$. Assume $Q(\ell)>0$ from now on.

Choose a point x of D. Let X_ℓ be the G_ℓ-orbit $G_\ell x$ of the point x in D, and K_x the isotropy subgroup of G at x. Then X_ℓ is isomorphic to the quotient $G_\ell/G_\ell\cap K_x$, and $G_\ell\cap K_x$ is a maximal compact subgroup of G_ℓ. Thus X_ℓ is a symmetric space of BD type on which Γ_ℓ acts.

For any point $z=(z_1, z_2, z_3)$ of D and for any vector $\ell=(a, b, c, d, e)$ of L with $Q(\ell)>0$, we put

$$\lambda(z,\ell)=a(z_2^2-z_1z_3)+bz_1+cz_2+dz_3+e.$$

It is easy to check that $\lambda(z,\ell)\neq 0$ on D.

We consider the holomorphic 3-form on D

$$\frac{dz_1\wedge dz_2\wedge dz_3}{\lambda(z,\ell)^3}$$

and its restriction on X_ℓ. Since X_ℓ is a *totally real* submanifold in D, this 3-form defines on X_ℓ a Γ_ℓ-invariant measure.

Definition (1.2). For any holomorphic cusp form $F(z)$ of weight k with respect to Γ, we define a period integral $I_\ell(F)$ by

$$I_\ell(F)=\int_{\Gamma_\ell\backslash X_\ell} F(z)\lambda(z,\ell)^{(k-3)}dz_1\wedge dz_2\wedge dz_3.$$

Remark (1.3). As shown in the proof of Sublemma of Sect. 4 of [13], this integral does not depend on the choice of the point x i.e. the choice of G_ℓ-orbit X_ℓ, and we have an equality

$$I_{\gamma\ell}(F)=I_\ell(F)\quad\text{for any}\quad\gamma\in\Gamma.$$

Now we quote Corollary of Theorem 1 of [13], which plays a key role in this paper. Keeping the notations of [13], we denote by $S_{2k-1}(\Gamma_0(4), \chi_Q)$ the space of the elliptic modular cusp forms of half integral weight $k-\frac{1}{2}$ with respect to $\Gamma_0(4)$ (4 is the level of Q) with the multiplicator χ_Q determined by Q.

Theorem (1.4). (Corollary of Theorem 1 of [13].) *Assume that $k\geqq 3$. Let $F(z)$ be a Siegel modular cusp form of weight k. Let H be the complex upper half plane. For $\tau\in H$ we consider the sum*

$$\varrho(F)(\tau)=\sum_{\substack{\ell\in L \bmod \Gamma,\\ Q(\ell)>0}} I_\ell(F)\exp[\pi iQ(\ell)\tau],$$

where the summation runs over a complete representative system of Γ-equivalence classes of vectors ℓ of L with $Q(\ell)>0$. Then this sum converges uniformly on any compact subset of H, and defines an element of $S_{2k-1}(\Gamma_0(4), \chi_Q)$.

For any integer n, we put

$$I(n,F)=\sum_{\substack{Q(\ell)=2n,\\ \ell\in L \bmod \Gamma}} I_\ell(F).$$

Then the Fourier expansion of $\varrho(F)$ at $i\infty$ is given by

$$\varrho(F)(\tau)=\sum_{n=1}^{\infty} I(n;F)\exp[2\pi i n\tau].$$

Definition (1.5). We denote by $P_k(\Gamma)=P_k(\Gamma_Q)=P_k(\mathrm{Sp}(n;\mathbf{Z}))$ the linear subspace of Siegel cusp forms F of weight k satisfying $\varrho(F)=0$.

Proposition (1.6). *For any Siegel cusp form $F(z)$ of weight k, the following statements are equivalent.*

(i) $F\in P_k(\Gamma)$.
(ii) $I(n;F)=0$ *for all positive integer* n.
(iii) $I_\ell(F)=0$ *for all* $\ell\in L$ *with* $Q(\ell)>0$.
(iv) $I_\ell(F)=0$ *for all primitive* $\ell\in L$ *with* $Q(\ell)>0$.

Proof. The equivalences (i)⇔(ii) and (iii)⇔(iv) are trivial by definition. The implication (iii)⇒(ii) is also trivial. The only non-trivial part (ii)⇒(iii) is postponed until the end of the next section, because we require a result on the reduction of vectors of L with respect to Γ.

2. Reduction of vectors of L with respect to Γ

We prove here that any vector of L is transformed to a vector of a canonical form by the elements of Γ, and find a representative system of the Γ-equivalence classes of the vectors of L.

We call a vector $\ell=(a, b, c, d, e)$ of L is *primitive*, if the integers a, b, c, d, e have no common divisor except ± 1. By definition, the zero vector is not primitive.

Proposition (2.1). (Humbert) *Any primitive vector $\ell=(a, b, c, d, e)$ of L is Γ-equivalent to the vector*

$$\left(0,\frac{q}{4},0,1,0\right), \text{ if } q=4ae+4bd-c^2\equiv 0 \bmod 4,$$

or to the vector

$$\left(0,\frac{q+1}{4},1,1,0\right), \text{ if } q=4ae+4bd-c^2\equiv -1 \bmod 4.$$

For any vector $\ell\neq 0$ of L, there exist a unique primitive vector ℓ_0 and a unique positive integer d_ℓ, such that $\ell=d_\ell\ell_0$. Thus the following corollary follows immediately from the above proposition.

Corollary (2.2). *Let $\ell = d_\ell \ell_0$ be a non-zero vector of L, with ℓ_0 primitive and d_ℓ positive. Put $q_0 = q/d_\ell^2$. Then ℓ is Γ-equivalent to the vector*

$$\left(0, \frac{q_0 d_\ell}{4}, 0, d_\ell, 0\right)$$

or to the vector

$$\left(0, \frac{(q_0+1)d_\ell}{4}, d_\ell, d_\ell, 0\right),$$

according as $q_0 \equiv 0 \bmod 4$ or $q_0 \equiv -1 \bmod 4$.

And these vectors constitutes a complete representative system of Γ-equivalence classes of elements of $L - \{0\}$.

Proof of Proposition (2.1). A proof is found in Humbert [6], and it is not difficult. So we omit it.

Proof of Proposition (1.6). We show (ii)⇒(iii). We proceed by induction with respect to $Q(\ell)$. If ℓ is not primitive, then $\ell = d\ell'$ with $d \geqq 2$. Clearly, $G_\ell = G_{\ell'}$ and $\Gamma_\ell = \Gamma_{\ell'}$. Therefore $I_\ell(F) = d^{k-3} I_{\ell'}(F)$ for any cusp form F. Since $Q(\ell) = d_\ell^2 Q(\ell') > Q(\ell')$, we have done in this case. Assume that ℓ is primitive. Put $Q(\ell) = 2n$. By Corollary of Proposition (2.1), for any integer n, there exists at most one Γ-equivalence class of primitive vectors ℓ with $Q(\ell) = 2n$. Therefore,

$$I_n(F) = I_\ell(F) + \sum_{\substack{\ell' \text{ imprimitive,} \\ Q(\ell') = 2n}} I_{\ell'}(F) = 0.$$

By the assumption of induction, $I_{\ell'}(F) = 0$. Hence $I_\ell(F) = 0$, which completes the proof.

3. Period Integrals and Hecke Operators

In this section we consider certain relations between the Hecke operators of Γ_Q and the period integrals $I_\ell(F)$, $I(n; F)$, and especially we show that the space $P_k(\Gamma_Q)$ is stable under the action of the Hecke operators of Γ_Q.

Let $\tilde{G}$ be the identity component of the group of similitudes for Q

$$\{g \in M_5(\mathbf{R}) | {}^t g Q g = s(g) Q, \text{ where } s(g) \text{ is a non zero real scalar}\}.$$

Put $\tilde{G} \cap M_5(\mathbf{Z}) = \tilde{\Gamma}$. Then $\tilde{\Gamma}$ is a semigroup contained in the commensurator of Γ in G. Let $\mathscr{H}(\Gamma, \tilde{\Gamma})$ be the Hecke ring with respect to Γ and $\tilde{\Gamma}$. Note that $s(g)$ is a square of an integer, if $g \in \tilde{\Gamma}$. In fact, the relation ${}^t g Q g = s(g) Q$ implies $(\det g)^2 = s(g)^5$, with $\det g$ and $s(g)$ integers. Therefore $s(g)$ is a square of an integer.

Now we note here that our Hecke ring is isomorphic to that of the Siegel modular group of degree 2. Let $\mathrm{SSp}(2; \mathbf{R})$ be the symplectic similitude group

$$\left\{ g_+ \in M_4(\mathbf{R}) | {}^t g_+ J g_+ = s_+(g_+) J, \text{ with } s_+(g_+) \text{ is a real positive number.} \right.$$
$$\left. J = \begin{pmatrix} 0 & -1_2 \\ 1_2 & 0 \end{pmatrix} \right\},$$

and let $\widetilde{\mathrm{Sp}(2,\mathbf{Z})}$ be the intersection $\mathrm{SSp}(2,\mathbf{Z})\cap M_4(\mathbf{Z})$. Since the homomorphism $\sigma:\mathrm{Sp}(2;\mathbf{R})\to G$ is induced by a rational representation of $\mathrm{Sp}(2;\mathbf{R})$ on $L_{\mathbf{R}}=\mathbf{R}^5$ such that the entries of the matrix $\sigma(g_+)$ is a linear combination of 2-minors of the matrix g_+ for $g_+\in\mathrm{Sp}(2;\mathbf{R})$, we can naturally extend it to a homomorphism $\tilde{\sigma}:\mathrm{SSp}(2;\mathbf{R})\to\tilde{G}$ satisfying $s(\tilde{\sigma}(g_+))=s_+(g_+)^2$ for any $g_+\in\mathrm{SSp}(2;\mathbf{R})$ and $\tilde{\sigma}(\mathrm{Sp}(2;\mathbf{Z}))=\tilde{\Gamma}$. Therefore $\tilde{\sigma}$ induces an isomorphism of $\mathscr{H}(\mathrm{Sp}(2;\mathbf{Z}),\widetilde{\mathrm{Sp}(2;\mathbf{Z})})$ to $\mathscr{H}(\Gamma,\tilde{\Gamma})$.

The group $\tilde{G}$ acts on $L_{\mathbf{R}}$ by $g(\ell)=\ell^t g$ for $\ell\in L_{\mathbf{R}}$ and for $g\in\tilde{G}$, and by extension of scalars, on $L_{\mathbf{C}}$. Now we define an action of $\tilde{G}$ on the dual vector space $L_{\mathbf{C}}^*$ by

$$g(v^*)=s(g)v^*g^{-1}=v^*Q^{-1}\cdot{}^tgQ$$

for $v^*\in L_{\mathbf{C}}^*$ and for $g\in G$. Then for the canonical pairing

$$\langle\ ,\ \rangle: L_{\mathbf{C}}\times L_{\mathbf{C}}^*\to\mathbf{C},$$

we have $\langle g(v), g(v^*)\rangle=s(g)\langle v, v^*\rangle$ for $v\in L_{\mathbf{C}}$ and for $v^*\in L_{\mathbf{C}}^*$. The group $\tilde{G}$ acts on $\tilde{D}_H^*$ in $L_{\mathbf{C}}^*$ by restriction, on D_H^* in $\mathbf{P}^4$ on passing to the quotient, and on D^* by transporting the structure by means of the bijection μ. Since $\tilde{G}$ is connected, its action preserves each of the connected components of D^*. Thus $\tilde{G}$ acts on D. We can also extend the automorphy factor $j(g,z)$ on $G\times D$ to an automorphy factor $j(g,z)$ (the same symbol, by an abuse of notation) on $\tilde{G}\times D$, such that

$$g(\tilde{\mu}(z))=j(g,z)\tilde{\mu}(g(z))$$

and

$$\lambda(g(z),g(\ell))=s(g)j(g,z)^{-1}\lambda(z,\ell)$$

for any $g\in\tilde{G}$ for any $z\in D$ and for any $\ell\in L_R$.

Let us consider a subset of $\tilde{\Gamma}$

$$\{\alpha\in\tilde{\Gamma}\,|\,{}^t\alpha Q\alpha=n^2Q\}$$

for any positive integer n. It is known that this subset is a disjoint union of a finite number of double cosets

$$\bigcup_{i=1}^{h}\Gamma\alpha_i\Gamma.$$

Then a formal sum

$$\sum_{i=1}^{h}\Gamma\alpha_i\Gamma$$

defines an element of $\mathscr{H}(\Gamma,\tilde{\Gamma})$. We denote this element by $T(n)$. Then it is the image by $\tilde{\sigma}$ of the usual $T(n)$-operator for the Siegel modular group of degree 2 (cf. [1, 2]).

Now we define an action of $\Gamma\alpha\Gamma$ on modular forms in the manner compatible with the definition of [1, 2]. Let $F(z)$ be a modular form of weight k on D with respect to Γ. We decompose the set $\Gamma\alpha\Gamma$ into a union of left cosets

$$\Gamma\alpha\Gamma=\bigcup_{j=1}^{d}\Gamma\delta_j.$$

Then we put

$$((\Gamma\alpha\Gamma)F)(z)=s(\alpha)^{\frac{2k-3}{2}}\sum_{j=1}^{d} j(\delta_j,z)^{-k}F(\delta_j(z)),$$

where $s(\alpha)^{1/2}$ is the positive square root of $s(\alpha)$.

Proposition (3.1). *The linear subspace $P_k(\Gamma)$ of the space of cusp forms of weight k with respect to Γ defined by Definition* (1.5), *is preserved under the action of the Hecke ring $\mathscr{H}(\Gamma,\tilde{\Gamma})$.*

Proof. Since $\mathscr{H}(\Gamma,\tilde{\Gamma})$ is generated by $T(n)$, and since $F\in P_k(\Gamma)$ is equivalent to

$$I_\ell(F)=0 \quad \text{for all} \quad \ell\in L \quad \text{with} \quad Q(\ell)>0$$

by Proposition (1.6), it suffices to show that the period integrals $I_\ell(T(n)F)$ are linear combinations of $I_\ell(F)$ for any cusp form F. Therefore our proposition is reduced to the following lemma.

Lemma (3.2). *Let*

$$T(n)=\bigcup_{j=1}^{d}\Gamma\delta_j$$

be a presentation of $T(n)$ as a disjoint union of left cosets with respect to Γ. Put $\ell_j=\delta_j(\ell)=\ell^t\delta_j$ for $1\leqq j\leqq d$. Then

$$I_\ell(T(n)F)=\sum_{j=1}^{d}[\Gamma_{\ell_j}:\delta_j\Gamma_\ell\delta_j^{-1}]\cdot I_{\ell_j}(F)$$

for any cusp form F. Here $[\Gamma_{\ell_j}:\delta_j\Gamma_\ell\delta_j^{-1}]$ is the generalized index of the commensurable groups Γ_{ℓ_j} and $\delta_j\Gamma_\ell\delta_j^{-1}$.

Proof. We follow the arguments of Shintani [16]. By definition,

$$I_\ell(T(n)F)=\int_{\Gamma_\ell\backslash X_\ell}(T(n)F)(z)\lambda(z,\ell)^{(k-3)}dz,$$

and

$$T(n)=n^{2k-3}\sum_{j=1}^{d}j(\ell_j,z)^{-k}F(\delta_j(z)).$$

Hence

$$I_\ell(T(n)F)=\sum_{j=1}^{d}n^{2k-3}\int_{\Gamma_\ell\backslash X}j(\ell_j,z)^{-k}F(\delta_j(z))\lambda(z,\ell)^{(k-3)}dz.$$

Since

$$\lambda(\delta_j(z),\delta_j(\ell))=n^2j(\delta_j,z)^{-1}\lambda(z,\ell)$$

and

$$d(\delta_j z)=n^3j(\delta_j,z)^{-3}dz,$$

we have

$$I_\ell(T(p)F) = \sum_{j=1}^{d} \int_{\Gamma_\ell \backslash X_\ell} F(\delta_j(z))\lambda(\delta_j(z), \delta_j(\ell))^{(k-3)} d(\delta_j(z))$$
$$= \sum_{j=1}^{d} \int_{\delta_j \Gamma_\ell \delta_j^{-1} \backslash \delta_j X_\ell} F(z)\lambda(z, \ell_j)^{(k-3)} dz.$$

Since $G_{\ell_j} = \delta_j G_\ell \delta_j^{-1}$ and the integral does not depend upon $x \in D$ such that $X_\ell = G_\ell x$ (cf. Remark (1.3)), we can put $\delta_j X_\ell = X_{\ell_j}$ in the above integral. Moreover we can easily check that $\delta_j \Gamma_\ell \delta_j^{-1}$ is commensurable with Γ_{ℓ_j}. For each j we choose a discrete subgroup Δ_j in G_{ℓ_j} such that the induces $[\Gamma_{\ell_j} : \Delta_j]$ and $[\delta_j \Gamma_\ell \delta_j^{-1} : \Delta_j]$ are finite. Then,

$$I_\ell(T(n)F) = \sum_{j=1}^{d} [\delta_j \Gamma_\ell \delta_j^{-1} : \Delta_j]^{-1} \int_{\Delta_j \backslash X_{\ell_j}} F(z)\lambda(z, \ell_j)^{k-3} dz$$
$$= \sum_{j=1}^{d} \frac{[\Gamma_{\ell_j} : \Delta_j]}{[\delta_j \Gamma_\ell \delta_j^{-1} : \Delta_j]} I_{\ell_j}(F),$$

which shows the lemma.

Corollary (3.3). *The space $I_k(\mathrm{Sp}(2;\mathbf{Z})) = I_k(\Gamma_Q)$ is stable under the action of all Hecke operators in $\mathscr{H}(\mathrm{Sp}(2;\mathbf{Q}), \widetilde{\mathrm{Sp}(2;\mathbf{Q})})$.*

Proof. It follows immediately from Proposition (3.1), because $I_k(\Gamma_Q)$ is the orthogonal complement of $P_k(\Gamma_Q)$ in $S_k(\Gamma_Q)$ with respect to the Petersson metric, and all Hecke operators are self adjoint for this metric.

4. Poles of Andrianov's L-Functions

In this section we prove the main result. First we recall the results of Andrianov [1, 2].

Let $F(z)$ be a Siegel cusp form of degree 2 of weight k. Assume that $F(z)$ is a common eigenfunction of all Hecke operators $T(n)$ in the Hecke ring $\mathscr{H}(\mathrm{Sp}(2;\mathbf{Q}), \widetilde{\mathrm{Sp}(2;\mathbf{Q})})$:

$$T(n)F = \lambda_F(n)F \qquad (n = 1, 2, 3, \ldots).$$

Put

$$Z_F(s) = \prod_{\substack{p \\ \text{prime}}} \{1 - \lambda_F(p)p^{-s} + (\lambda_F(p)^2 - \lambda_F(p^2) - p^{2k-4})p^{-2s} - \lambda_F(p)p^{2k-3-3s}$$
$$+ p^{4k-6-4s}\},$$

for $\mathrm{Re}(s) \gg 0$,
and

$$\Psi_F(s) = (2\pi)^{-2s}\Gamma(s)\Gamma(s-k+2)Z_F(s),$$

where $\Gamma(s)$ is the gamma function.

By Theorem (3.1.1) of [2], $\Psi_F(s)$ is analytically continued to the whole s-plane as a meromorphic function having at most simple poles at $s=k$, $s=k-2$.

By Proposition (3.1), the space $P_k(\mathrm{Sp}(2;\mathbf{Z}))$ is stable under the action of all Hecke operators. Therefore, $P_k(\mathrm{Sp}(2;\mathbf{Z}))$ has a basis consisting of cusp forms which are common eigenfunctions of all Hecke operators.

Theorem (4.1). *Assume that k is even, and $k \geqq 4$. Let $F \in P_k(\mathrm{Sp}(2;\mathbf{Z}))$ be a common eigenfunction of all Hecke operators. Then $\Psi_F(s)$ is an entire function.*

Proof. Recall the formula (3.8.10) of [2], which represents $\Psi_F(s)$ as an integral of Rankin type. To prove that $\Psi_F(s)$ has no poles at $s=k$ and $s=k-2$, it suffices to show that the residues of possible simple poles at $s=k$ and $s=k-2$ vanish. We can show this by noting that these residues are linear combinations of our period integrals $I_\ell(F)$. To be more precise, we rewrite the formula (3.8.10) in view of the following facts:

(i) *For any $\ell=(0, a, b, c, 0) \in L$ with $Q(\ell)>0$, the inverse image $\sigma^{-1}(G_\ell)$ of G_ℓ by $\sigma: \mathrm{Sp}(2;R) \to G$ coincides with the group G_N defined in Theorem* (3.4.4) *of* [2] *with*

$$N=\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

Moreover $\sigma^{-1}(\Gamma_\ell)$ coincides with Γ_N of (3.5.1) *of* [2].

(ii) *For any $\ell=(0, a, b, c, 0)$ of L with $Q(\ell)>0$, put $x=(ic, -ib/2, ia) \in D$ and $X_\ell = G_\ell x$. Then our X_ℓ coincides with the hyperbolic space H_N of Sect.* 3.4 *of* [2] *for*

$$N=\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}.$$

As a result of these coincidences, we have the following:

For any primitive vector $\ell=(0, a, b, c, 0)$ with $D=-1/2Q(\ell)=b^2-4ac<0$, we have an equality

$$I_\ell(F)=\pm(\sqrt{-1})^{(k-2)}|D|^{\frac{(k-3)}{2}} \int_{D_f} F_D(u)du.$$

Here $\pm$ depends upon the choice of the orientation of the chain $\Gamma_\ell \backslash X_\ell$, and $\int_{D_f} F_D(u)du$ is a certain integral defined in Sects. 3.4 *and* 3.6 *of* [2].

Let F be common eigenfunction of all Hecke operators with the Fourier expansion

$$F(z)=\sum_{N>>0} a(N)\exp[2\pi i\,\mathrm{tr}(NZ)],$$

where $Z=\begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix}$ for $z=(z_1, z_2, z_3)$.

Take a number $D=df^2<0$, where d is a discriminant of an imaginary quadratic field and f is an integer. As in [2], we consider a representative system $\{N_1, \ldots, N_h\}$ of $\mathrm{SL}(2;\mathbf{Z})$-equivalence classes of symmetric positive definite semi-integral primitive matrices $N=\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$, which corresponds to a representative

system of equivalence classes of integral primitive positive definite binary quadratic forms $ax^2+bxy+cy^2$ with discriminant D. Let χ be a character of the class group $\{N_1, \ldots, N_h\}$ whose multiplication is defined as the composition of quadratic forms.

Put $a(\chi)=\sum_{i=1}^{h}\chi(N_i)a(N_i)$ and put $\ell=(0, a_i, b_i, c_i, 0)$ for some i $(i=1, \ldots, h)$, where $N_i=\begin{pmatrix} a_i & b_i/2 \\ b_i/2 & c_i \end{pmatrix}$.

Then, in view of the formula (3.8.11) of [2], the formula (3.8.10) of [2] reads as follows.

Formula (4.2).

$$a(\chi)\Psi_F(s)=\pm\frac{1}{s-k}\left\{\sum_{j=1}^{h}\chi(N_j)\right\}\cdot(2\pi)^{-(k-2)}I_\ell(F)$$
$$\mp\frac{1}{s-k+2}\left\{\sum_{j=1}^{h}\chi(N_j)\right\}\cdot(2\pi)^{-(k-2)}I_\ell(F)$$
$$+\textit{ an entire function of } s.$$

By Proposition (1.6), we have $I_\ell(F)=0$ for all $\ell\in L$ with $Q(\ell)>0$, if $F\in P_k(\mathrm{Sp}(2;\mathbf{Z}))$, which complete the proof of our theorem.

Let us investigate the case $F\in I_k(\mathrm{Sp}(2;\mathbf{Z}))$.

Proposition (4.3). *Assume that k is even and $k>5$. Let*

$$F(z)=\sum_{N=\left(\begin{smallmatrix} a & b/2 \\ b/2 & c \end{smallmatrix}\right)} a(N)\exp[2\pi i\,\mathrm{tr}(NZ)]$$

be the Fourier expansion of $F(z)\in I_k(\mathrm{Sp}(2;\mathbf{Z}))$, with $Z=\begin{pmatrix} z_1 & z_2 \\ z_2 & z_3 \end{pmatrix}$ for $z=(z_1, z_2, z_3)$. Then we have $a(N_1)=a(N_2)$, if the vectors $(0, a_1, b_1, c_1, 0)$ and $(0, a_2, b_2, c_2, 0)$ are Γ-equivalent, where

$$N_i=\begin{pmatrix} a_i & b_i/2 \\ b_i/2 & c_i \end{pmatrix} \quad (i=1, 2).$$

Proof. It immediately follows from a formula (8.43) of [13], which represents the Fourier coefficients of $\iota(f)\in I_k(\mathrm{Sp}(2;\mathbf{Z}))$ by that of $f\in S_{2k-1}(\Gamma_0(4), \chi_Q)$ at various cusps of $\Gamma_0(4)$. This formula reads the following.

Formula (4.4). *Assume that k is even and $k>5$. For any $N=\begin{pmatrix} a & b/2 \\ b/2 & c \end{pmatrix}$ with $a>0$ and $\det N>0$, the Fourier coefficients $a(N)$ of $F=\iota(f)$ is given by*

$$a(N)=c\sum_{d|\ell} d^{k-1}\sum_{N_2|(4, n/d^2)}\psi(N_2)N_2^{k-3/2}a^{N_1}_{n/N_2d^2}(f).$$

Here $N_1=4/N_2$, $a_m^{N_1}(f)$ is the Fourier coefficients of f at various cusps, $\psi(N_2)$ is a certain constant depending only on N_2, $\ell=(0, a, b, c, 0)$, $n=Q(\ell)/2$, and c is a constant depending only on the weight k.

Since the divisibility $d|\ell$ and the equality $n=Q(\ell)/2$ are also satisfied by any vector which is Γ-equivalent to ℓ, this formula implies our proposition.

Remark. Theorem 5 and the formula (8.43) is not proved for odd n in [13]. But we can treat the case n odd, completely similarly as the case n even. To guard aganist any suspicion of miscalculation, we indicate here another proof of Proposition (4.3). By the results of [13], $I_k(\mathrm{Sp}(2,\mathbf{Z}))$ is generated by

$$\omega_m(z)=\sum_{\ell\in L, Q(\ell)=2m}\lambda(z,\ell)^{-k}\ (m=1,2,\ldots).$$

Therefore it suffices to show our proposition for $\omega_m(z)$, which follows immediately from Theorem 4 of [13] and the fact that $G_{v_0}(m, v)$ depends only on the Γ-equivalence class of v by Definition (6.69) of [13].

We conclude this section with the following theorem.

Theorem (4.5). *Assume that k is even and $k>5$. Let $F\in I_k(\mathrm{Sp}(2;\mathbf{Z}))$ be a common eigenfunction of all Hecke operators. Then $\Psi_F(s)$ has simple poles at $s=k$ and $s=k-2$.*

Proof. Since $F\neq 0$ and $I_k(\mathrm{Sp}(2;\mathbf{Z}))\cap P_k(\mathrm{Sp}(2;\mathbf{Z}))=\{0\}$, there exists a primitive vector ℓ such that $I_\ell(F)\neq 0$ by Proposition (1.6). By Proposition (2.1) and Remark (1.3), we may assume that ℓ is of the form $(0, a, b, c, 0)$ with $D=b^2-4ac=df^2$. Apply Formula (4.2) in the proof of Theorem (4.1) for this D, and for each

$$N_i=\begin{pmatrix} a_i & b_i/2 \\ b_i/2 & c_i \end{pmatrix}(i=1,\ldots,h) \text{ put } \ell_i=(0, a_i, b_i, c_i, 0).$$

Then every ℓ_i is primitive and $Q(\ell_i)=2(-D)$. Therefore $\ell_i (i=1,\ldots,h)$ are Γ-equivalent by Proposition (2.1). Accordingly, by Proposition (4.3) we have $a(N_i)=a(N_j)$ for any $i, j (i,j=1,\ldots,h)$. Consequently,

$$a(\chi)=\sum\chi(N_i)a(N_i)=\left\{\sum_{i=1}^{h}\chi(N_i)\right\}a(N_j) \text{ for any } j.$$

Take as χ the trivial character in Formula (4.2). Then in this case it reads as follows:

$$\begin{aligned} a(N_j)\Psi_F(s)=&\pm\frac{1}{s-k}(2\pi)^{-(k-2)}I_\ell(F)\\ &\mp\frac{1}{s-k+2}(2\pi)^{-(k-2)}I_\ell(F)\\ &+\textit{an entire function of } s. \end{aligned}$$

Since $I_\ell(F)\neq 0$, $a(N_j)\neq 0$, accordingly $\Psi_F(s)$ has poles at $s=k$ and $s=k-2$ with residues

$$\pm(2\pi)^{-(k-2)}a(N_j)^{-1}I_\ell(F).$$

Thus we have proved the theorem.

Corollary (4.6). *For any primitive vectors* $\ell_1=(0, a_1, b_1, c_1, 0)$ *and* $\ell_2=(0, a_2, b_2, c_2, 0)$ *of* L, *we have*

$$I_{\ell_1}(F)/I_{\ell_2}(F)=a(N_1)/a(N_2),$$

where

$$N_i=\begin{pmatrix} a_i & b_i/2 \\ b_i/2 & c_i \end{pmatrix} \quad (i=1, 2).$$

Especially, for any $F\neq 0$ *of* $I_k(\mathrm{Sp}(2;\mathbf{Z}))$ *which is a common eigenfunction of all Hecke operators, we have*

$$\varrho(F)(\tau)=\text{const} \sum_{\substack{\ell=(0,a,b,c,0) \bmod \Gamma, \\ \ell=d\ell',\ell'=(0,a',b',c',0) \\ \ell' \text{ primitive}}} d^{k-3}a\left(\begin{pmatrix} a' & b'/2 \\ b'/2 & c' \end{pmatrix}\right)\exp[2\pi i Q(\ell)\tau],$$

where τ *is a point of the complex upper half plane* H, *and the summation runs over a representative system of* Γ*-equivalence classes of vectors of the form* $(0, *, *, *, 0)$.

Proof. As shown in the proof of Theorem (4.5), the residue of the pole at $s=k$ of $\psi_F(s)$ is given by

$$(2\pi)^{-(k-2)}a(N)^{-1}I_\ell(F),$$

which is independent of the choice of the primitive N. Hence the former part of our corollary follows. The latter part immediately follows from this and Theorem (1.4).

5. A few Word on the Space $I_k(\mathrm{Sp}(2;\mathbf{Z}))$

In this section, the author refers to a relation between our results and Maass [9–11] and Andrianov [4].

Let $F\neq 0$ be a common eigenfunction of all Hecke operators, which belongs to $I_k(\mathrm{Sp}(2;\mathbf{Z}))$. Then the following theorem is proved by Rallis-Schifman [14] and Kojima [7].

Theorem. $Z_F(s)=\zeta(s-k+1)\zeta(s-k+2)D(s;f)$, *where* f *is an elliptic cusp form of weight* $2k-2$.

Meanwhile a similar result was obtained by Maass [9–11] and Andrianov [4], for Siegel cusp forms belonging to a certain subspace of $S_k(\mathrm{Sp}(2;\mathbf{Z}))$.

Let $M_k(\mathrm{Sp}(2;\mathbf{Z}))$ be this subspace defined by the condition (1) of [9]:

Let

$$F(Z)=\sum_T a(T)\exp[2\pi i\,\mathrm{tr}(TZ)]$$

be the Fourier expansion of Siegel modular forms $F(Z)$. *Then* $M_k(\mathrm{Sp}(2;\mathbf{Z}))$ *consists of Siegel modular forms, which satisfies*

$$a(T)=\sum_{\substack{d|n,m,t \\ d>0}} d^{k-1}a\begin{pmatrix} 1 & t/2d \\ t/2d & nm/d^2 \end{pmatrix}, \quad T=\begin{pmatrix} n & t/2 \\ t/2 & m \end{pmatrix}.$$

Theorem (5.1). $I_k(\mathrm{Sp}(2;\mathbf{Z}))=M_k(\mathrm{Sp}(2;\mathbf{Z}))$, *for* k *even and* $k\geqq 6$.

Proof. We first show the following lemma.

Lemma. $M_k(\mathrm{Sp}(2;\mathbf{Z}))\cap P_k(\mathrm{Sp}(2;\mathbf{Z}))=\{0\}$, *and* $M_k(\mathrm{Sp}(2;\mathbf{Z}))\subset I_k(\mathrm{Sp}(2;\mathbf{Z}))$.

Proof. Both M_k and P_k are stable under the action of Hecke operators by [11, 4] and by Proposition (3.1). Therefore $M_k\cap P_k$ is also stable under Hecke operators. Let $F\neq 0$ be a common eigenfunction of all Hecke operators contained in $M_k\cap P_k$. Then $Z_F(s)$ has no pole at $s=k$ by Theorem (4.1).

Meanwhile the results of [9–11] and [4] imply that

$$Z_F(s)=\zeta(s-k+1)\zeta(s-k+2)D(s;f)$$
$$=\frac{1}{s-k}\zeta(2)D(k;f)+O(1)$$

in the neightbourhood of $s=k$, where $O(\ )$ is the Landau symbol. Let

$$f(\tau)=\sum a_n \exp[2\pi i n\tau]$$

be the Fourier expansion of the elliptic cusp form f of weight $2k-2$. As is well known, the estimate $a_n=O(n^{(2k-2)/2})=O(n^{k-1})$ implies the convergency of the Euler product for $D(s;f)$ for $\mathrm{Re}(s)>k$. By the same reason, the improved Rankin's ([15]) estimate $a_n=O(n^{(2k-2)/2-1/5})=O(n^{k-1-1/5})$ asserts that the Euler product converges absolutly for $\mathrm{Re}(s)>k-\frac{1}{5}$, and accordingly $D(k;f)\neq 0$. Hence $Z_F(s)$ has a pole at $s=k$, contradicting the holomorphy of $Z_F(s)$ at $s=k$. The latter half of the lemma is evident from the direct sum decomposition $S_k=I_k\oplus P_k$.

Now we proceed to the proof of Theorem (5.1). It suffices to show the converse inclusion. Let us show that the Siegel modular forms F whose coefficients are given by Formula (4.4), satisfies the condition (1) of [9] on the Fourier coefficients of F. Let $a(N)$ be Fourier coefficients of F. Then, by Proposition (2.1) the vectors $(0, n, t, m, 0)$ and $(0, 1, t, nm, 0)$ of L are Γ-equivalent if the integers n, t, and m have no common divisor. Therefore by Proposition (4.3) we have

$$a\left(\begin{pmatrix} n & t/2 \\ t/2 & m \end{pmatrix}\right)=a\left(\begin{pmatrix} 1 & t/2 \\ t/2 & m \end{pmatrix}\right),$$

if n, t, and m have no common divisor except ± 1.

Noting this fact we check the condition (1) of [9], $a(N)$ is defined by Formula (4.4). Therefore $I_k\subset M_k$, which completes our proof.

Remark (5.2). Let us consider a quadratic form defined by a symmetric matrix

$$Q'=\begin{pmatrix} & & 4 \\ & -2 & \\ 4 & & \end{pmatrix}.$$

Then the level of Q' is 4, and of signature $(2+, 1-)$. Since Q is a direct sum of Q' and a kernel form, we have $\chi_Q=\chi_{Q'}$. By Shintani [16], $\chi_{Q'}$ is trivial, accordingly χ_Q

is also trivial. Recall the results of [16, 12, 13], and consider a diagram:

$$\mathfrak{S}_{2k-2}(\mathrm{SL}(2;\mathbf{Z})) \underset{\iota'}{\overset{\varrho'}{\rightleftarrows}} S_{2k-1}(\Gamma_0(4)) \underset{\varrho}{\overset{\iota}{\rightleftarrows}} S_k(\mathrm{Sp}(2;\mathbf{Z})),$$

where $\mathfrak{S}_{2k-2}(\mathrm{SL}(2;\mathbf{Z}))$ is the space of elliptic cusp forms of weight $2k-2$. Here (ι', ϱ') is the pair (ι, ϱ) for Q' (see Introduction), and ϱ' (resp. ι') is discussed precisely in Shintani [16] (resp. in Niwa [12]). By the adjointness of (ι, ϱ) and of (ι', ϱ'), we have a decomposition:

$$S_{2k-1}(\Gamma_0(4)) = \mathrm{Ker}(\iota) \oplus \mathrm{Imag}(\varrho) = \mathrm{Ker}(\iota') \oplus \mathrm{Imag}(\varrho').$$

Since $\dim S_{2k-1}(\Gamma_0(4)) > \dim I_k(\mathrm{Sp}(2;\mathbf{Z})) = \dim \mathfrak{S}_{2k-2}(\mathrm{Sp}(2;\mathbf{Z}))$, ι is not injective, and ϱ is not surjective.

Question. $\mathrm{Imag}(\varrho') = \mathrm{Imag}(\varrho)$?

Question. Is the composition $\iota \circ \varrho'$ injective?

Or equivalently, is the composition $\iota' \circ \varrho$ is surjective?

A similar related problem is discussed in Kojima [8].

Remark (5.3). Let F be a common eigenfunction of all Hecke operators, which belongs to $I_k(\mathrm{Sp}(2;\mathbf{Z}) = M_k(\mathrm{Sp}(2;\mathbf{Z}))$. In the course of the proofs, we obtained the following:

Let f be a elliptic cusp form of weight $2k-2$, satisfying

$$\begin{aligned}\Psi_F(s) &= (2\pi)^{-s}\Gamma(s)\Gamma(s-k+2)Z_F(s) \\ &= (2\pi)^{-s}\Gamma(s)\Gamma(s-k+2)\zeta(s-k+1)\zeta(s-k+2)D(s;f).\end{aligned}$$

Comparing with Formula (4.2), we have an equality

$$\pm(2\pi)^{-(k-2)}a(N_j)^{-1}I_\ell(F) = (2\pi)^{-k}\Gamma(k)\Gamma(2)\zeta(2)D(k;f),$$

which is reduced to

$$\pm(2\pi)^2 a(N_j)^{-1}I_\ell(F) = (k-1)!\,\zeta(2)D(k;f),$$

and consequently,

$$D(k;f) = \pm\frac{24}{(k-1)!}a(N_j)^{-1}I_\ell(F).$$

Thus a special value of L-function of an elliptic cusp form is represented by a kind of period integral of a normalized Siegel cusp form.

Following a suggestion by Mazur, we can formulate the following conjecture.

Conjecture (5.4). Let F be a Siegel cusp form of weight k, which is a common eigenfunction of all Hecke operators. First, we expect to attach to F a mixed Hodge structure $H_{\mathbf{Z}}(F)$ of "poids" $2k-3$ and of rank 4 over the ring O_F generated by the eigenvalues, with a polarization $\psi : H_{\mathbf{Z}}(F) \times H_{\mathbf{Z}}(F) \to O_F[-2k+3]$. Here $O_F[-2k+3]$ is the Hodge structure of Tate "poids" $-2k+3$. Moreover we expect that, if F belongs to $P_k(\mathrm{Sp}(2;\mathbf{Z}))$, this Hodge structure is pure, and that, if F

belongs to $I_k(\mathrm{Sp}(2;\mathbf{Z}))$, this Hodge structure has the following "poids" filtration $\{W_i\}$:

$$W_iH_{\mathbf{Q}}(F)=\{0\}, \quad \text{for} \quad i\leqq 2k-5,$$
$$W_jH_{\mathbf{Q}}(F)=H_{\mathbf{Q}}(F), \quad \text{for} \quad j\geqq 2k-2,$$
$$\mathrm{Gr}_W^{2k-4}H_{\mathbf{Q}}(F)\cong K_F[-k+2],\ \mathrm{Gr}_W^{2k-3}H_{\mathbf{Q}}(F)\cong H_{\mathbf{Q}}(f),$$

and

$$\mathrm{Gr}_W^{2k-2}H_{\mathbf{Q}}(F)\cong K_F[-k+1],$$

where f is a certain ellitic cusp form of weight $2k-2$, and $H_{\mathbf{Q}}(f)$ is the pure Hodge structure of "poids" $-2k+3$ attached to f. K_F is the quotient field of O_F.

By the correspondence principle of Deligne [5] between Hodge structures and ℓ-adic cohomologies, we may expect that there is a similar "poids" filtration on the ℓ-adic representation attached to F.

References

1. Andrianov, A.N.: Dirichlet series with Euler product in the theory of Siegel modular forms of genus 2. Trudy Mat. Inst. Steklov. **112**, 73–94 (1971)
2. Andrianov, A.N.: Euler products corresponding to Siegel modular forms of genus 2. Uspehi Mat. Nauk **29**, 43–110 (1974)
3. Andrianov, A.N.: Siegel modular forms and zeta-functions. Trud. Mat. Inst. Steklov **132**, 132–136 (1973)
4. Andrianov, A.N.: Modular descent and the Saito-Kurokawa conjecture. Invent. Math. **53**, 267–280 (1979)
5. Deligne, P.: Theorie de Hodge, I. Actes du Congrès international des Mathématiciens, I, Nice, pp. 425–430 (1970)
6. Humbert, M.G.: Sur les fonctions abeliennes singulières. J. Math. Pures Appl. V. Sér. 5 (1899)
7. Kojima, H.: Siegel modular cusp forms of degree two. Tohoku Math. J. (to appear)
8. Kojima, H.: On construction of Siegel modular forms of degree two (to appear)
9. Maass, H.: Über eine Spezialschar von Modulformen zweiten Grades. Invent. Math. **52**, 95–104 (1979)
10. Maass, H.: Über eine Spezialschar von Modulformen zweiten Grades. II. Invent. Math. **53**, 249–253 (1979)
11. Maass, H.: Über eine Spezialschar von Modulformen zweiten Grades. III. Invent. Math. **53**, 255–265 (1979)
12. Niwa, S.: Modular forms of half integral weight and the integral of certain theta-functions. Nagoya. Math. J. **56**, 147–161 (1974)
13. Oda, T.: On modular forms associated with indefinite quadratic forms of signature $(2, n-2)$. Math. Ann. **231**, 97–144 (1977)
14. Rallis, S., Schiffman, G.: On a relation between $\widetilde{\mathrm{SL}}_2$ cusp forms and cusp forms on the tube domains associated to orthogonal groups. Trans. Am. Math. Soc. (to appear)
15. Rankin, R.A.: Contributions to the theory of Ramanujan's function $\tau(n)$ and similar arithmetical functions. I–III. Proc. Cambridge Phil. Soc. **35**, 351–356, 357–372 (1939), 150–151 (1940)
16. Shintani, T.: On construction of holomorphic cusp forms of half integral weight. Nagoya. Math. J. **58**, 83–126 (1975)
17. Evdokimov, S.A.: A characterization of Maass sapces of paraboloic Siegel modular forms of degree 2. Mat. Sb. **112**, 133–142 (1980)

Received August 27, 1980

Note added in proof. The following two papers have settled the questions on the injectivity of ι and ι' in Remark (5.2):

Kohnen, W.: Modular forms of half-integral weight on $\Gamma_0(4)$. Math. Ann. **248**, 249–266 (1980)

Zagier, D.: Sur la conjecture de Saito-Kurokawa. Preprint

Math. Ann. 256, 341–362 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Footnotes to a Paper of Beniamino Segre

The Number of g^1_d's on a General d-Gonal Curve, and the Unirationality of the Hurwitz Spaces of 4-Gonal and 5-Gonal Curves

Enrico Arbarello[1] and Maurizio Cornalba[2]*

[1] Department of Mathematics, Harvard University, 1 Oxford Street, Cambridge, MA 02138, USA, and Istituto Matematico "Guido Castelnuovo", Universitá di Roma, I-00100 Roma, Italy
[2] Istituto Matematico, Universitá di Pavia, I-27100 Pavia, Italy

0. Introduction

In this note we offer a few comments on Beniamino Segre's paper "Sui moduli delle curve poligonali e sopra un complemento al Teorema di esistenza di Riemann" which appeared in the Mathematische Annalen in 1928.

In his paper, Segre proves two main results. The first one is

(0.1) **Theorem** (B. Segre). *Let d and g be integers such that $2 \leqq d < (g/2)+1$. Let C be the most general curve of genus g that can be expressed as a d-sheeted ramified covering of $\mathbb{P}^1$, i.e. possessing a base-point-free g^1_d. Then C has only finitely many g^1_d's.*

After reviewing the defommation theory of g^1_d's and of mappings of curves into surfaces, in Sect. 2 we prove a strengthening of Segre's result by showing that, in fact, the given g^1_d is *unique*, and more precisely that C does not possess any g^1_h with $h < (g/2)+1$ except those which are composed with the given g^1_d (cf. Theorem 2.6).

After we completed the proof of this result, Ciro Ciliberto brought to our attention a paper of Permutti [5] where a somewhat weaker result appears. Permutti states that C does not have any g^1_h with $d < h < (g/2)+1$, other than those composed with the given g^1_d. However, his proof appears to be incomplete. For example Permutti's argument relies on an analysis of a certain kind of curves lying on a quadric surface, and the main objection to his reasoning is that he assumes these curves to have nodes as their only singularities. Establishing that this is indeed the case is, in fact, the main technical difficulty in the proof.

The second result in Segre's paper is the following

(0.2) **Theorem** (B. Segre). *Let d be an integer such that $d \geqq 3$. Let $f\colon C \to \mathbb{P}^1$ be a general d-sheeted ramified covering of genus g. Let n be an integer such that*

$$n \geqq \frac{g+d+2}{2}. \tag{0.3}$$

* During the preparation of this manuscript the second named author was partially supported by N.S.F. (Grant MCS-78-07348) and by Consiglio Nazionale delle Ricerche

0025-5831/81/0256/0341/$04.40

Then the curve C can be realizing as a plane curve Γ of degree n with an ordinary $(n-d)$-fold point p,

$$\delta = \binom{n-1}{2} - \binom{n-d}{2} - g$$

nodes and no other singularity, in such a way that the d-sheeted covering f is obtained by projecting Γ from p.

This remarkable theorem suggested us the following count of constants. Suppose n is the minimum integer satisfying (0.3), and also assume that we are in one of the following cases

$$\text{(0.4)} \qquad \begin{cases} d \leqq 5, & g \geqq d-1 \\ d \leqq 6, & g \leqq 10 \\ d = 6, & g = 12 \\ d = 7, & g = 7. \end{cases}$$

In these cases, and *only in these cases*, we have

$$\text{(0.5)} \qquad \frac{n(n+3)}{2} - \binom{n-d+1}{2} - 3\delta \geqq 0.$$

This suggests that the δ nodes of the plane curve in question can be taken generically in $\mathbb{P}^2$ and that one could then mimic Severi's proof of the unirationality of the moduli space of curves of genus $g \leqq 10$ to show that the Hurwitz space of d-sheeted, genus g coverings of $\mathbb{P}^1$ is unirational in case one of the four pairs of relations in (0.4) holds.

Unfortunately, even if (0.5) holds it is by no means clear that the δ double points of Segre's curve are in general position. Even worse, as is the case with Severi's argument, it is far from being clear that inequality (0.5) suffices to insure that, given a general $(\delta+1)$-tuple of points $(p, p_1, \ldots, p_\delta)$, there exists an *irreducible* plane curve Γ of degree n having an *ordinary* $(n-d)$-fold point at p, *nodes* at $p_1, \ldots, p_\delta$ and *no other singularity*. Under the additional, and obviously necessary, assumption that

$$\binom{n-1}{2} - \binom{n-d}{2} - \delta \geqq 0,$$

this is what we prove in Sects. 3 and 4, with one exception (cf. Theorem 3.2, Corollary 4.6, Corollary 4.7). The exception is provided by the observation that the only plane sextic passing doubly through nine general points is a cubic counted twice. It seems to us that this result is of independent interest.

At this point the reader might be led to believe that the proof of unirationality for the above mentioned Hurwitz spaces is complete. However, once our plane curve Γ is constructed, it is no longer clear that the d-sheeted covering of $\mathbb{P}^1$ obtained by projecting Γ from p is general. In Sect. 5 we give a deformation-theoretic proof of this fact. As we remark in (5.4), our results imply Severi's theorem about the unirationality of the moduli space of curves of genus $g \leqq 10$.

Finally we want to emphasize that both in the proof of the uniqueness of g^1_d's and in the proof of the existence of plane curves with nodes in general position, a crucial role is played by a technical result about deformations of curves with cusps (Lemma 1.4) which was first stated and proved in [1] and which, to our mind, is the only really non-classical tool we use.

1. Preliminaries

Throughout this paper, the word *variety* shall be used to denote a (possibly non-reduced) algebraic variety over $\mathbb{C}$. Unless otherwise specified, by *smooth curve* we shall mean a smooth, complete, connected, one-dimensional variety. The words *line bundle* are used as a synonim of "invertible sheaf". The symbol $\mathbb{C}[\varepsilon]$ denotes the ring of dual numbers. The tangent sheaf and the canonical sheaf to a smooth variety M are denoted by the symbols Θ_M, K_M (or simply Θ, K, when no confusion is likely).

Throughout this paper, we shall freely use notations and results from [1]. For the convenience of the reader, we now expose some of them.

We are primarily interested in studying the subvarieties

$$\mathscr{M}^1_{g,d} \subset \mathscr{M}_g$$

of the moduli space of curves of genus g whose points correspond to curves that admit at least one linear series of degree d and dimension one ($a\, g^1_d$). For our purposes, it will often be convenient to work, rather than on $\mathscr{M}_g$, on a smooth variety S over which there is a family of smooth curves

$$p: X \to S \tag{1.1}$$

such that the morphism

$$\phi: S \to \mathscr{M}_g$$

$$\phi(s) = \text{isomorphism class of } p^{-1}(s)$$

is a finite dominant morphism and, for every $s \in S$, the Kodaira-Spencer homomorphism

$$T_s(S) \to H^1(p^{-1}(s), \Theta_{p^{-1}(s)})$$

is an isomorphism. We shall also assume that (1.1) *has a section.* Given a smooth curve C of genus g we may always choose (1.1) so that C appears as one of its fibres. For example, one may take S to be an appropriate covering of a small open subset of the moduli space of genus g curves with a level-n structure (for large enough n) which contains a point corresponding to C. A *family of* g^r_d's *on* (1.1) is the datum of:

i) A morphism

$$f: T \to S.$$

ii) An invertible sheaf $\mathscr{L}$ on $X \times_S T$ such that the restriction of $\mathscr{L}$ to any fibre of the projection

$$p': X \times_S T \to T$$

has degree d.

iii) A locally free rank-$(r+1)$ subsheaf

$$\mathscr{F} \subset p'_*(\mathscr{L})$$

such that, for every $t \in T$, the homomorphism

$$\mathscr{F} \otimes_{\mathscr{O}_T} k(t) \to H^0(p'^{-1}(t), \mathscr{L} \otimes \mathscr{O}_{p'^{-1}(t)})$$

is injective.

We shall say that the above family of g^r_d's is *parametrized by* f. Let $(f, \mathscr{N}, \mathscr{G})$ be another family of g^r_d's. We shall say that this family is *isomorphic* to $(f, \mathscr{L}, \mathscr{F})$ if there are an invertible sheaf $\mathscr{Q}$ on T and an isomorphism

$$\mathscr{L} \simeq \mathscr{N} \otimes p'^*(\mathscr{Q})$$

which induces an identification

$$\mathscr{F} = \mathscr{G} \otimes \mathscr{Q}$$

It was proved in [1] that

(1.2) *The functor*

$$f \longmapsto \left\{ \begin{array}{c} \text{isomorphism classes of families of } g^r_d\text{'s on (1.1)} \\ \text{parametrized by } f \end{array} \right\}$$

is representable by a variety over S:

$$\pi: \mathscr{G}^r_d \to S.$$

Instead of studying the varieties $\mathscr{M}^1_{g,d}$ directly we shall study $\mathscr{G}^1_d$ and its projection $\pi(\mathscr{G}^1_d)$.

We shall make frequent use of the basic concepts of Horikawa's theory of deformations of mappings [3]. Let C be a smooth curve and let

$$\varphi: C \to M$$

be a non-constant morphism into a smooth algebraic variety of dimension n. The *normal sheaf* $N = N_\varphi$ is the cokernel of the injective sheaf homomorphism

$$\varphi_*: \Theta_C \to \varphi^*(\Theta_M).$$

To each deformation of φ parametrized by $\mathrm{Spec}(\mathbb{C}[\varepsilon])$ there is associated an element of $H^0(C, N)$ which will be called the *Horikawa class* of the deformation. Taking the Horikawa class sets up a one-to-one correspondence between the vector space $H^0(C, N)$ and the set of equivalence classes of deformations of φ parametrized by $\mathrm{Spec}(\mathbb{C}[\varepsilon])$.

Let Z be the ramification divisor of φ. Then there is an exact sequence [1]

$$0 \to \mathscr{K}_\varphi \to N_\varphi \to N'_\varphi \to 0 \tag{1.3}$$

where $\mathscr{K}_\varphi$ is (non-canonically) isomorphic to $\mathscr{O}_Z$ and N'_φ is *locally free* of rank $n-1$. We shall need the following result (cf. [1], Corollary 6.11 for a more precise statement)

(1.4) **Lemma.** *Let*

$$q: \mathscr{C} \to T$$

be a family of smooth curves parametrized by an irreducible variety T. Let

$$\varphi: \mathscr{C} \to M$$

be a morphism into a smooth algebraic variety M of dimension $n \geqq 2$. *For each* $t \in T$ *set*

$$C_t = \varphi^{-1}(t)$$
$$\varphi_t = \varphi|_{C_t}.$$

Assume that, for general t, φ_t *induces a birational morphism of* C_t *onto* $\varphi(C_t)$. *Let*

$$h_t: T_t(T) \to H^0(C_t, N_{\varphi_t})$$

be the Horikawa homomorphism. Then, for general t,

$$h_t(T_t(T)) \cap H^0(C_t, \mathscr{K}_{\varphi_t}) = \{0\}.$$

Now, let ℓ be a point of $\mathscr{G}_d^1$ corresponding to a smooth curve C and a base-point-free g_d^1 on C. The g_d^1 corresponds to a two-dimensional subspace $W \subset H^0(C, L)$, where L is a degree d line bundle on C. A choice of basis for W induces a d-sheeted ramified covering

$$f: C \to \mathbb{P}^1.$$

It is shown in [1] that there is an exact sequence

$$0 \to \mathbb{C} \to \mathrm{Hom}(W, W) \to H^0(C, N_f) \xrightarrow{\alpha} T_\ell(\mathscr{G}_d^1) \to 0 \tag{1.5}$$

where the homomorphism α is gotten as follows. Each element of $H^0(C, N_f)$ corresponds to an equivalence class of deformations of f parametrized by $\mathrm{Spec}(\mathbb{C}[\varepsilon])$. Any such deformation induces a family of g_d^1's parametrized by a morphism

$$\mathrm{Spec}(\mathbb{C}[\varepsilon]) \to S$$

and hence a morphism of $\mathrm{Spec}(\mathbb{C}[\varepsilon])$ into $\mathscr{G}_d^1$ sending the unique point of $\mathrm{Spec}(\mathbb{C}[\varepsilon])$ to ℓ.

2.

As is well known, the general existence theorem for special divisors [4] implies that every smooth curve of genus g has a g_d^1 when

$$d \geqq g/2 + 1.$$

On the other hand, when

$$2 \leqq d < g/2 + 1 \tag{2.1}$$

a general curve of genus g does not have a g_d^1 (see for example [1]).

In the first part of his paper [6], Segre computes the exact dimension of $\mathscr{M}^1_{g,d}$ when (2.1) holds. Segre's starting point is Severi's inequality [9]

$$\dim \mathscr{M}^1_{g,d} \leqq 2g+2d-5.$$

He then proceeds by explicitly constructing a smooth curve C of genus g and a base-point-free g^1_d on C such that, if ℓ denotes the corresponding point of $\mathscr{G}^1_d$, the differential

$$\pi_*\colon T_\ell(\mathscr{G}^1_d) \to T_{\pi(\ell)}(S)$$

is injective. Since, as is shown, for example, in [1], $\mathscr{G}^1_d$ is smooth and

$$\dim(\mathscr{G}^1_d) = 2g+2d-5, \tag{2.2}$$

this implies that

$$\dim \pi(\mathscr{G}^1_d) = \dim \mathscr{M}^1_{g,d} = 2g+2d-5. \tag{2.3}$$

Moreover, since it is well known that $\mathscr{M}^1_{g,d}$ is irreducible [2], Segre obtains as a corollary of (2.2) and (2.3) that, when C is a curve corresponding to a general point of $\mathscr{M}^1_{g,d}$ on C there are only finitely many g^1_d's. Our goal in this section is to show that, in fact, there is a *unique* g^1_d on C. It is therefore natural to try and compute the dimension of the variety $\mathscr{G}^1_d \times_S \mathscr{G}^1_d$, or more generally $\mathscr{G}^1_d \times_S \mathscr{G}^1_h$, parametrizing triples of the form

$$(\text{smooth curve } C,\ g^1_d \text{ on } C,\ g^1_h \text{ on } C).$$

Our basic remark is

(2.4) **Proposition.** *Suppose $g \geqq 2$. Let d, h be integers such that*

$$2 \leqq d, \quad 2 \leqq h.$$

Let $\mathscr{X}$ be a component of $\mathscr{G}^1_d \times_S \mathscr{G}^1_h$. Assume that a general point of $\mathscr{X}$ corresponds to a smooth curve C with a base-point-free g^1_d and a base-point-free g^1_h. Let

$$f\colon C \to \mathbb{P}^1$$
$$f'\colon C \to \mathbb{P}^1$$

be the corresponding morphisms. Assume moreover that

$$(f,f') = \varphi\colon C \to Q = \mathbb{P}^1 \times \mathbb{P}^1$$

is not composed with an involution. Then

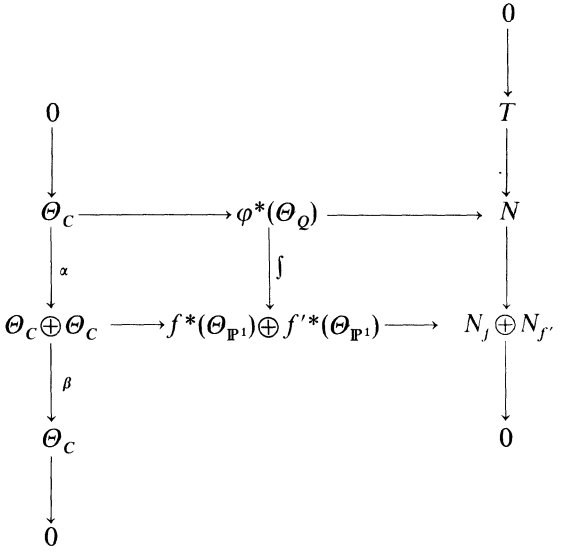
$$\dim \mathscr{X} = g+2d+2h-7$$

and a general point of $\mathscr{X}$ is a reduced point of $\mathscr{G}^1_d \times_S \mathscr{G}^1_h$.

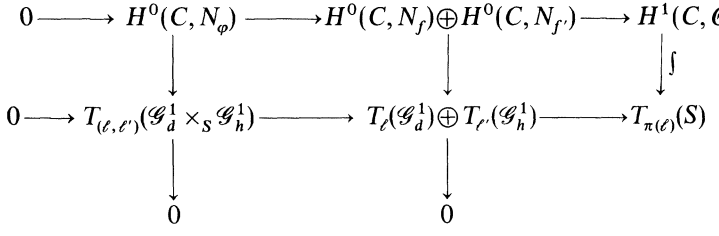
Proof. First of all, notice that every component of $\mathscr{G}^1_d \times_S \mathscr{G}^1_h$ has dimension at least equal to

$$\dim \mathscr{G}^1_d + \dim \mathscr{G}^1_h - \dim S = g+2d+2h-7.$$

Let ℓ, ℓ' be the points of $\mathscr{G}^1_d, \mathscr{G}^1_h$ corresponding to f and f'. There is a diagram of exact sequences



where α is the diagonal homomorphism and β is projection on one of the two factors. The sheaf T is therefore isomorphic to Θ_C and the vertical exact sequence on the right gives rise to a commutative diagram of exact sequences



By (1.5) we get

$$\dim(T_{(\ell,\ell')}(\mathscr{G}^1_d \times_S \mathscr{G}^1_h)) = h^0(C, N_\varphi) - 6.$$

Thus we have to show that

$$h^0(C, N_\varphi) = g + 2d + 2h - 1. \tag{2.5}$$

By the remark at the beginning of the proof and by (1.4), we know that

$$h^0(C, N'_\varphi) \geqq g + 2d + 2h - 7 \geqq g + 1$$

and therefore

$$H^1(C, N_\varphi) = H^1(C, N'_\varphi) = 0.$$

This means that every class in $H^0(C, N_\varphi)$ is unobstructed [3] and, again by (1.4), we conclude that $\mathscr{K}_\varphi = 0$ and $N_\varphi = N'_\varphi$. Thus

$$N_\varphi = \det(\varphi^* \Theta_Q) \otimes K_C$$

has degree

$$\deg(N_\varphi)=2g-2+2d+2h$$

and, from the Riemann-Roch theorem, we get (2.5). Q.E.D.

The condition that φ not be composed with an involution is essential. As an example, consider ramified coverings of smooth curves of type $(2,2)$ in $\mathbb{P}^1 \times \mathbb{P}^1$. The ramification number of these coverings is $r=2g-2$. Hence the corresponding component of $\mathscr{G}^1_{2n} \times_S \mathscr{G}^1_{2n}$ has dimension at least equal to $2g-2$. This number is strictly larger than the postulated one, $g+8n-7$, when g is large enough relative to n.

The main consequence of (2.4) is the following

(2.6) **Theorem.** *Let d, h be integers such that*

$$2 \leqq d < (g/2)+1, \quad 2 \leqq h < (g/2)+1.$$

Let C be a smooth curve corresponding to a general point of $\mathscr{M}^1_{g,d}$. Then C possesses only one g^1_d. Furthermore, any g^1_h on C is composed with the g^1_d.

Proof. Clearly it suffices to prove the last statement. An easy moduli count shows that any g^1_d on C is base-point-free and simple. *Simple* means that the d-sheeted ramified covering

$$f: C \to \mathbb{P}^1$$

associated to the g^1_d is simple, i.e. its ramification points all have index two and no two of them lie over the same point of $\mathbb{P}^1$.

Suppose there is a g^1_h on C; we may assume that it has no base points. Let

$$f': C \to \mathbb{P}^1$$

be the corresponding morphism. Since f is simple, the morphism

$$(f,f')=\varphi: C \to \mathbb{P}^1 \times \mathbb{P}^1$$

is not composed with an involution, unless there is a ramified covering

$$\lambda: \mathbb{P}^1 \to \mathbb{P}^1$$

such that $f'=\lambda \circ f$. If the latter does not occur, Proposition 2.4 applies and, coupled with Segre's theorem 2.3, it implies that

$$2g+2d-5 \leqq g+2d+2h-7,$$

that is

$$2h \geqq g+2,$$

contrary to our assumptions. Q.E.D.

(2.7) **Corollary.** *Let d and C be as in* (2.6). *Denote by L the unique line bundle of degree d on C such that*

$$h^0(C,L) \geqq 2.$$

Let n, m be non-negative integers such that

$$m < d$$
$$nd + m < (g/2) + 1$$

and let $p_1, \ldots, p_m$ be points of C. Then

$$h^0(C, L^n(p_1 + \ldots + p_m)) = n + 1.$$

3.

As we explained in the introduction, it is a classical problem to determine under which conditions there exists an irreducible curve of degree n in $\mathbb{P}^2$ with nodes at δ preassigned general points and no other singularity. Two obvious necessary conditions are

$$\begin{cases} \binom{n+2}{2} - 1 \geqq 3\delta \\ \binom{n-1}{2} \geqq \delta. \end{cases} \tag{3.1}$$

The first one comes from the fact (see Lemma 4.2) that it takes exactly three linearly independent conditions to impose a singularity at a preassigned general point to curves varying in a linear system which has no fixed components and is not composed with a pencil. The second one is just the statement that a plane curve of degree n with more than $\binom{n-1}{2}$ singular points has to be reducible. The following result, to be proved in Sect. 4, shows that, with one exception, conditions (3.1) are also sufficient.

(3.2) **Theorem.** *Let $p_1, \ldots, p_\delta$ be general points in $\mathbb{P}^2$. Let n be a positive integer such that* (3.1) *hold. Then there exists an irreducible plane curve of degree n having nodes at $p_1, \ldots, p_\delta$ and no other singularity, unless*

$$n = 6; \quad \delta = 9.$$

In this case the only plane curve of degree n passing doubly through $p_1, \ldots, p_\delta$ is a smooth cubic counted twice.

It is perhaps interesting to notice that, in case $\delta = 9$, $n = 6$, there may well be special configurations of δ points $p_1, \ldots, p_\delta$ such that there exists an irreducible sextic having double points at the p_i's. In case $p_1, \ldots, p_9$ lie on a smooth cubic Γ, a condition that the p_i's must satisfy for such a sextic to exist is

$$2(p_1 + \ldots + p_9) \sim 18 p_0,$$

where p_0 is an inflectionary point of Γ.

In Sect. 5 we shall prove the unirationality of $\mathscr{M}^1_{g,d}$, when $d \leqq 5$, using plane curves of degree n with δ nodes in general position and an $(n-d)$-fold point. To be able to carry out this program, we will be forced to prove a general-

ization of Theorem 3.2 concerning curves with nodes on a rational surface (Corollary 4.6). Given a linear system $|D|$ of dimension at least 3δ on a smooth rational surface X, we would like to determine conditions under which there exists an irreducible curve $C \in |D|$ with nodes at preassigned general points $p_1, \ldots, p_\delta$ of X and no other singularity. In this section *we shall work under the assumption that* $|D|$ *contains an irreducible curve having singularities at* $p_1, \ldots, p_\delta$. In the next section we shall address the problem of deciding when such a curve exists.

We will need the following well-known result (cf. [11], for example).

(3.3) **Lemma.** *Let X be a smooth rational surface. Let D be a reduced, irreducible curve on X such that $(K_X \cdot D) < 0$. Then a general member of $|D|$ is smooth.*

If D is a curve on a smooth surface X, we shall denote by $p_a(D)$ its virtual genus. Thus

$$p_a(D) = \tfrac{1}{2}((D \cdot D) + (K_X \cdot D)) + 1.$$

Suppose X is rational. For the convenience of the reader, we recall that, in case D_1, D_2 are effective divisors on X such that each component of $D_1 + D_2$ has negative intersection with K_X, the following hold:

$$\dim |D_1| = (D_1 \cdot D_1) - p_a(D_1) + 1 = p_a(D_1) - (K_X \cdot D_1) - 1 \tag{3.4}$$

$$\dim |D_1 + D_2| = \dim |D_1| + \dim |D_2| + (D_1 \cdot D_2) \tag{3.5}$$

$$p_a(D_1 + D_2) = p_a(D_1) + p_a(D_2) + (D_1 \cdot D_2) - 1. \tag{3.6}$$

The main result of this section is the following.

(3.7) **Proposition.** *Let X be a smooth rational surface. Let $|D|$ be a complete linear system of dimension k on X. Let $\{p_1, \ldots, p_\delta\}$ be a general set of δ points on X. Assume that there is an irreducible element of $|D|$ having singularities at $p_1, \ldots, p_\delta$. Assume moreover that*

$$3\delta \leqq k \tag{3.8}$$

$$\delta \leqq p_a(D) \tag{3.9}$$

$$(K_X \cdot D) < 0. \tag{3.10}$$

Then there exists an irreducible element of $|D|$ having nodes at $p_1, \ldots, p_\delta$ and no other singularity.

Proof. By (3.3) we may assume that D is smooth. This takes care of the case $\delta = 0$. Assume then that $\delta \geqq 1$. Let $\mathrm{Sym}^\delta(X)$ be the δ-fold symmetric product of X. We let

$$\Sigma_\delta \subset |D| \times \mathrm{Sym}^\delta(X)$$

be the variety of pairs $(Y, p_1 + \ldots + p_\delta)$, where $p_1, \ldots, p_\delta$ are distinct and Y is an irreducible element of $|D|$ which is singular at $p_1, \ldots, p_\delta$. By hypothesis there is an irreducible component Σ of Σ_δ which, under projection, maps onto an open

subset of $\mathrm{Sym}^{\delta}(X)$. Let Γ be a general element of Σ. Let

$$\varphi: C \to \Gamma \subset X$$

be its normalization. By hypothesis, C is connected; we denote by p the virtual genus of D and by p' the genus of C. We know that

$$p' \leqq p - \delta. \tag{3.11}$$

We only have to show that equality holds in (3.11) and that the δ double points of Γ are in fact nodes. We begin by noticing that

$$\dim \Sigma \geqq (k - 3\delta) + 2\delta = k - \delta. \tag{3.12}$$

Recalling the exact sequence (1.3) and Lemma 1.4 we have

$$\dim \Sigma \leqq h^0(C, N_\varphi) - h^0(C, \mathscr{K}_\varphi) = h^0(C, N'_\varphi). \tag{3.13}$$

By (3.4), hypothesis (3.10) implies that $k \geqq p$. Therefore, using (3.11) and (3.13) we conclude that

$$h^0(C, N'_\varphi) \geqq k - \delta \geqq p - \delta \geqq p'. \tag{3.14}$$

This implies that

$$H^1(C, N_\varphi) = H^1(C, N'_\varphi) = 0 \tag{3.15}$$

unless $N'_\varphi \simeq K_C$. Assume this is not the case. Then (3.15) means that every class in $H^0(C, N_\varphi)$ is unobstructed [3], and, again by (1.4), that $\mathscr{K}_\varphi = 0$ and N_φ is a line bundle. In particular, φ is unramified, and it now suffices to show that equality holds in (3.11). Taking determinants in the exact bundle sequence

$$0 \to \Theta_C \to \varphi^*(\Theta_X) \to N_\varphi \to 0 \tag{3.16}$$

we get

$$N_\varphi = K_C \otimes \varphi^*(K_X^{-1}).$$

By the Riemann-Roch theorem on C and by (3.14), (3.15) we conclude that

$$k - \delta \leqq h^0(C, N_\varphi) = p' - 1 - (K_X \cdot \Gamma).$$

Formula (3.4) tells us that

$$k = p - (K_X \cdot \Gamma) - 1.$$

Comparing this with the preceding inequality shows that equality must hold in (3.11).

Now suppose $K_C \cong N'_\varphi$, so that (3.14) is a string of equalities. It follows, in particular, that the only singularities of Γ are nodes or ordinary cusps at $p_1, \ldots, p_\delta$ and that $k = p$, so that $(K_X \cdot \Gamma) = -1$. Denote by Z the ramification divisor of the normalization map φ. Then we can replace (3.16) with the exact sequence

$$0 \to \Theta_C(Z) \xrightarrow{\varphi_*} \varphi^*(\Theta_X) \to N'_\varphi \to 0$$

and, again taking determinants, we get

$$N'_{\varphi} = K_C(-Z) \otimes \varphi^*(K_X^{-1}).$$

Since $(K_X \cdot \Gamma) = -1$, we must have

$$\deg(Z) = 1.$$

Put differently

(3.17) *The general member of the linear subsystem of $|D|$ consisting of those elements which are singular at $p_1, \ldots, p_\delta$ has a cusp at one of the p_i's, nodes at the other ones, and no other singularity.*

By the generality of the p_i's this is clearly absurd if $\delta \geqq 2$. When $\delta = 1$, we will arrive at a contradiction by a different argument.

Let D' be a general singular element of $|D|$. Thus the unique singularity of D' is a cusp at $q \in X$ and we may assume that q is not a base point for $|D|$. Choose a three-dimensional linear subsystem $\mathscr{D}$ of $|D|$ containing D' for which q is not a base point (this is possible by (3.8)). This system yields a rational map

$$\psi : X \dashrightarrow \mathbb{P}^3.$$

Since a general element of $\mathscr{D}$ is irreducible, ψ does not map X to a curve. Let Y be the image surface of ψ. In our case (3.17) implies that Y is a non-degenerate surface in projective 3-space whose Gauss map is everywhere degenerate. It is a classical theorem that such a surface must be a cone or a tangential developable (cf. [7], for example). In any case, Y is ruled, hence any section of Y by means of a tangent plane is reducible. This means that every singular element of $\mathscr{D}$ must be reducible, a contradiction. Q.E.D.

4.

In this section we prove a technical result concerning the existence of irreducible curves with δ singular points in general position on a rational surface.

(4.1) **Proposition.** *Let X be a smooth rational surface and let δ be a non-negative integer. Let $|D|$ be a complete linear series on X with the following properties.*

i) *A general element of $|D|$ is connected.*

ii) $p_a(D) \geqq \delta$.

iii) $\dim |D| \geqq 3\delta$.

iv) *If S is a general set of δ points in X, there is an element $D' \in |D|$ which is singular at each point of S and is such that, for every irreducible component D'' of D', $(D'' \cdot K_X) < 0$.*

Let S be a general set of δ points in X, and let Σ be the linear subsystem of $|D|$ consisting of those curves in $|D|$ which are singular at each point of S. Then either Σ contains an irreducible curve or else Σ consists of a single divisor of the form 2Γ where Γ is smooth and irreducible. In the latter case: $\dim |D| = 3\delta$, $(\Gamma \cdot \Gamma) = \delta > 0$, $\dim |\Gamma| = \delta$, $p_a(\Gamma) = 1$.

It should be remarked that the second alternative in the thesis of Proposition 4.1 actually occurs and an example of it is provided by the exceptional case in Theorem 3.2.

In the course of the proof of Proposition 4.1 we shall make use of the following well known fact.

(4.2) **Lemma.** *Let X be a smooth surface and Σ a linear system on X. Assume that Σ has no fixed components and is not composed with a pencil. Let p be a general point of X and denote by Σ' the linear subsystem of Σ consisting of the curves in Σ which are singular at p. Then* $\dim \Sigma' = \dim \Sigma - 3$. *In particular if* $\dim \Sigma \leqq 2$ *then Σ' is empty.*

Proof. In a small open set $U \subset X$, Σ is $\{f=0\}_{f \in V}$, where V is a finite-dimensional vector subspace of $\Gamma(U, \mathcal{O}_X)$. If $\dim \Sigma' \geqq \dim \Sigma - 2$, every $f \in V$ satisfies a differential equation $\xi(f) + gf = 0$, where ξ is a vector field on U. Therefore, the integral curve of ξ through a general point p of U is contained in any curve of Σ through p; by Bertini's theorem, a general such curve is irreducible, hence there is a unique curve of Σ passing through p. Thus Σ is a pencil, a contradiction. Q.E.D.

We now begin the proof of Proposition 4.1. First assume $\delta = 0$. Let C be a general element of $|D|$. Suppose that C has at least two irreducible components. Since C is connected there must be two component C_1, C_2 of C such that $(C_1 \cdot C_2) > 0$. The continuous system of curves of the form $C_1' + C_2'$, where $C_1' \in |C_1|$, $C_2' \in |C_2|$ has dimension equal to $\dim |C_1| + \dim |C_2|$. On the other hand,

$$\begin{aligned} \dim |C_1 + C_2| &= \dim |C_1| + \dim |C_2| + (C_1 \cdot C_2) \\ &> \dim |C_1| + \dim |C_2|, \end{aligned}$$

and we conclude that a general element of $|C_1 + C_2|$ is irreducible, in contradiction with the generality of C. The only remaining possibility is that

$$C = h\Gamma$$

where Γ is irreducible. We want to show that $h = 1$. If this is not the case, $\dim |\Gamma| = 0$. Now since $(\Gamma \cdot K_X) < 0$ we must have

$$(\Gamma \cdot \Gamma) = -1, \quad (\Gamma \cdot K_X) = -1.$$

If $h > 1$, $p_a(D) < 0$, contrary to our hypotheses.

From now on we assume that $\delta \geqq 1$. Let C be a general element of the linear system consisting of those curves in $|D|$ which are singular at each point of S. Notice that by the generality of S and C we have that:

(4.3) *If an irreducible curve C_1 contains exactly r points of S and occurs with multiplicity h as a component of C, then given r distinct points $p_1, \ldots, p_r$ in S there exists a curve $C_2 \in |C_1|$ such that $C_2 \cap S = \{p_1, \ldots, p_r\}$ and C_2 occurs with multiplicity h as a component of C.*

Write

$$C = C_1 + \ldots + C_c$$

where each C_i is irreducible. Regardless of repetitions we shall refer to each C_i as a *component* of C. The proof of (4.1) takes place in several elementary steps.

Step I. Let p be a point of S. Then there are at most two components of C containing p; if there is a component of C which is singular at a point of S then each point of S is contained in only one component of C.

Suppose p belongs to C_1, C_2, C_3 (say); then any other point of S belongs to at least three components of C. Since p is general, $\dim|C_3| \geqq 1$, so that these exists $C_3' \in |C_3|$ which does not pass through p. The curve $C' = C - C_3 + C_3'$ is still singular at each point of S, and this contradicts the generality of C. This proves the first part of Step I. The second part can be proved in exactly the same way.

Step II. Every component of C contains at least one point of S.

Suppose this is not true. Since C is connected there must be two components of C, say C_1 and C_2, such that C_1 does not meet S, C_2 does, and $(C_1 \cdot C_2) > 0$. Therefore

$$\dim|C_1 + C_2| \geqq \dim|C_1| + \dim|C_2| + 1.$$

Let $\{p_1, \ldots, p_3\} = S \cap C_2$. Two cases are possible. Either C_2 is the only component of C through $p_1, \ldots, p_r$ and therefore it is singular at each one of these points, or else there are two components of C passing through each one of the points $p_1, \ldots, p_r$. In the first case the linear system of curves belonging to $|C_1 + C_2|$ and singular at $p_1, \ldots, p_r$ has dimension at least equal to

$$\dim|C_1| + \dim|C_2| - 3r + 1.$$

On the other hand, by Lemma 4.2, the algebraic system of curves $C_1' + C_2'$ where $C_1' \in |C_1|$, $C_2' \in |C_2|$ and C_2' is singular at $p_1, \ldots, p_r$ has dimension at most

$$\dim|C_1| + \dim|C_2| - 3r.$$

Thus the most general element of $|C_1 + C_2|$ is not of the form $C_1' + C_2'$. This contradicts the generality of C. The second case can be handled in the same fashion.

If C is not irreducible, then by Steps I and II, and after renumbering the C_i's, if necessary, we see that only five cases are, a priori, possible.

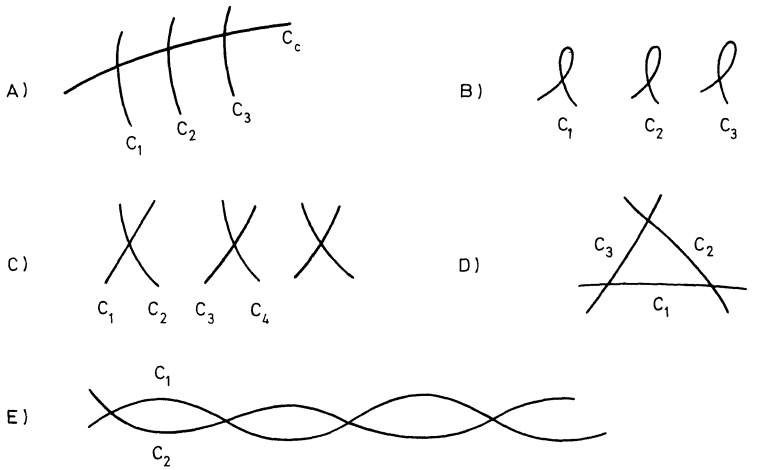
A) $c - 1 = \delta \geqq 2$, $C_c \supset S$, $C_1, \ldots, C_\delta$ *contain only one point of* S.

B) $c = \delta \geqq 2$ *and each* C_i *contains only one point of* S *(and is singular at it).*

C) $c = 2\delta \geqq 4$, *each component of* C *contains only one point of* S *and each point of* S *belongs to exactly two components of* C.

D) $c = \delta = 3$ *and each component of* C *contains exactly two points of* S.

E) $C = C_1 + C_2$ *and* $S \subset (C_1 \cap C_2)$.



Step III. Case A) cannot occur.

Set $C_i \cap S = \{p_i\}$, $i = 1, \ldots, \delta$. By (4.3) we have $|C_1| = \ldots = |C_\delta|$. Suppose $(C_1 \cdot C_1) > 0$. Then

$$\dim |2C_1| \geqq 2 \dim |C_1| + 1.$$

By Bertini's theorem a general element of $|2C_1|$ is irreducible. Moreover $\dim |2C_1| \geqq 3$. This contradicts the generality of S and C. Suppose then $(C_1 \cdot C_1) = 0$. Since $(C_1 \cdot K_X) < 0$, this implies that C_1 is a smooth rational curve and that $\dim |C_1| = 1$. Suppose now that $(C_1 \cdot C_c) \geqq 2$. By (3.5) the dimension of the linear subsystem of $|C_1 + C_c|$ consisting of curves passing doubly through p_1 and simply through $p_2, \ldots, p_\delta$ has dimension at least equal to $\dim |C| - \delta + 1$, whereas the linear system of curves $C_1 + \Gamma$, where Γ belongs to $|C_c|$ and passes through $p_1, \ldots, p_\delta$, has dimension equal to $\dim |C_c| - \delta$. This again is a contradiction. The only remaining possibility is that $(C_1 \cdot C_c) = 1$, $(C_1 \cdot C_1) = 0$ with C_1 a smooth rational curve. Since C_1 moves in a pencil, C_c is a rational curve. Moreover, by (3.3), C_c is smooth. This gives $p_a(C) = 0$, which contradicts ii).

Step IV. Case B) cannot occur.

Let $\{p_i\} = C_i \cap S$. By (4.3) $|C_1| = |C_2| = \ldots = |C_c|$. By (4.2) $\dim |C_1| \geqq 3$. Then there exists an irreducible pencil of curves in $|C_1|$ passing through p_1 and p_2. Let Γ_1 and Γ_2 belong to this pencil; then by Bertini's theorem the pencil generated by $\Gamma_1 + \Gamma_2$ and $C_1 + C_2$ contains an irreducible element, contradicting the generality of C and S.

Step V. Case C) cannot occur.

Set $S = \{p_1, \ldots, p_\delta\}$. We may assume that $p_1 \in C_1 \cap C_2$, $p_2 \in C_3 \cap C_4$ and so on, and moreover that $|C_1| = |C_3| = \ldots$, $|C_2| = |C_4| = \ldots$; set $\Gamma_1 = C_1 + C_2$, $\Gamma_2 = C_3 + C_4$, and so on. The argument used in step IV can now be applied verbatim, replacing C_i with Γ_i throughout, unless $\dim |C_1 + C_2| = 2$, in which case $C_1 = C_2$, $C_3 = C_4$ and $(C_1 \cdot C_1) = 0$. But this would contradict i).

Step VI. Case D) cannot occur.

In this case $C=C_1+C_2+C_3$ and $|C_1|=|C_2|=|C_3|$. Let p_i be the point of S not belonging to C_i. By Bertini's theorem the linear system Σ of curves in $|D|$ having singularities at p_1, p_2, p_3 contains an irreducible member unless $\dim|C_1|=2$. If this happens, then by (3.5)

$$\begin{aligned} 0=\dim\Sigma\geqq\dim|C|-9&=\dim|3C_1|-9\\ &=3\dim|C_1|+3(C_1\cdot C_1)-9\\ &=3(C_1\cdot C_1)-3. \end{aligned}$$

Hence $(C_1\cdot C_1)=(C_1\cdot C_2)=1$. In particular $\dim|3C_1|=\dim|C|=9$. On the other hand

$$p_a(C)=(C\cdot C)-\dim|C|+1=1<\delta$$

contradicting ii).

Step VII. Conclusion of the proof.

Suppose we are in case E). By (3.3) we may suppose that C_1 and C_2 are smooth. First assume that $C_1\neq C_2$. Then the dimension of the continuous system of curves $C_1'+C_2'$, where $C_i'\in|C_i|$ and $C_1'\cap C_2'$ contains S, is equal to

$$\dim|C_1|+\dim|C_2|-2\delta. \tag{4.4}$$

On the other hand, by (3.5),

$$\dim|C_1+C_2|=\dim|C_1|+\dim|C_2|+(C_1\cdot C_2).$$

Thus the linear system of curves in $|D|$ which are singular at each point of S has dimension at least equal to

$$\dim|C_1|+\dim|C_2|+(C_1\cdot C_2)-3\delta. \tag{4.5}$$

Since $(C_1\cdot C_2)\geqq\delta$ and since C is general, comparing (4.4) and (4.5) we get

$$(C_1\cdot C_2)=\delta.$$

Let Σ be the linear system of curves in $|C_2|$ passing through $p_1,\ldots,p_{\delta-1}$. There is a rational dominant map of Σ onto C_1 mapping each element of Σ to the residual intersection with C_1: thus C_1 is rational. For the same reason C_2 is rational. But then

$$p_a(D)=p_a(C_1)+p_a(C_2)+\delta-1=\delta-1,$$

contradicting ii). We now know that $C=2\Gamma$ and that $\dim|\Gamma|=\delta$. Using iii) we get $(\Gamma\cdot\Gamma)\geqq\delta$. In case $(\Gamma\cdot\Gamma)>\delta$ we get a contradiction arguing as in the case $C_1\neq C_2$. In conclusion we have

$$\begin{aligned} &\dim|D|=3\delta\\ &(\Gamma\cdot\Gamma)=\delta\\ &\dim|\Gamma|=\delta\\ &p_a(\Gamma)=(\Gamma\cdot\Gamma)-\dim|\Gamma|+1=1. \end{aligned}$$

Q.E.D.

Combining Propositions 3.7 and 4.1 we get

(4.6) **Corollary.** *Let X be a smooth rational surface and δ a non-negative integer. Let $|D|$ be a complete linear system on X such that:*

i) *a general element of $|D|$ is connected.*

ii) $p_a(D) \geqq \delta$.

iii) $\dim |D| \geqq 3\delta$.

iv) *If $(p_1, \ldots, p_\delta)$ is a general δ-tuple of points of X, there is an element of $|D|$ which is singular at $p_1, \ldots, p_\delta$ and is such that, for every one of its irreducible components D', $(K_X \cdot D') < 0$.*

Then, if $(p_1, \ldots, p_\delta)$ is a general δ-tuple of points of X, one of the following two cases occurs:

a) *There exists an irreducible curve in $|D|$ having nodes at $p_1, \ldots, p_\delta$ as its only singularities.*

b) *We have $\delta > 0$, $\dim |D| = 3\delta$, and the only curve in $|D|$ which is singular at $p_1, \ldots, p_\delta$ is of the form 2Γ, where:*

$$\dim |\Gamma| = (\Gamma \cdot \Gamma) = \delta$$

Γ *is a smooth curve of genus* 1.

We can now prove Theorem (3.2). Notice that all the hypotheses of (4.6) are met by the linear system of curves of degree n on $X = \mathbb{P}^2$. If case b) of Corollary (4.6) occurs, we must have

$$n = 6; \quad \delta = 9,$$

as desired.

As we already announced, in Sect. 5 we shall need the following mild generalization of Theorem (3.2).

(4.7) **Corollary.** *Let n, d, δ be non-negative integers such that*

$$n \geqq d > 0 \tag{4.8}$$

$$\binom{n-1}{2} - \binom{n-d}{2} - \delta \geqq 0 \tag{4.9}$$

$$\frac{n(n+3)}{2} - \binom{n-d+1}{2} - 3\delta \geqq 0. \tag{4.10}$$

Let $(p, q_1, \ldots, q_\delta)$ be a general $(\delta+1)$-tuple of points in $\mathbb{P}^2$. Then there exists an irreducible plane curve of degree n having an ordinary $(n-d)$-fold point at p, nodes at $q_1, \ldots, q_\delta$, and no other singularities, with the exception of the following cases:

$$n = 6, \quad d = 4, \quad \delta = 8;$$

$$n = 6, \quad d = 6, \quad \delta = 9.$$

In fact the only sextic which is singular at nine general points of $\mathbb{P}^2$ is a smooth cubic counted twice.

To prove (4.7) we shall apply (4.6) to the surface X obtained by blowing up $\mathbb{P}^2$ at p, to the linear system $|nH+dE|$, where H is the proper transform of a line in $\mathbb{P}^2$ passing through p and E is the exceptional curve on X, and to the points $p_1, \ldots, p_\delta$ on X corresponding to $q_1, \ldots, q_\delta$. Hypothesis i) in (4.6) is implied by (4.8). Hypotheses ii) and iii) are just (4.9) and (4.10). Hypothesis iv) is automatically satisfied, since $(K_X \cdot C) < 0$ for every curve C on X. If case a) in (4.6) occurs, we are done. If case b) occurs, then $n=2a$, $d=2b$ and

$$1 = p_a(aH+bE) = ab - a - \frac{b^2+b}{2} + 1.$$

Since, by (4.8), we must have $a \geqq b > 0$, this can happen only when $a=3$, $b=2$ or when $a=b=3$. Thus we must have

$$n=6; \quad d=4; \quad \delta = ((3H+2E)\cdot(3H+2E)) = 8$$

or

$$n=6; \quad d=6; \quad \delta = ((3H+3E)\cdot(3H+3E)) = 9$$

This concludes the proof of (4.7).

We have seen that examples of case b) in (4.6) can actually occur. In general every example of case b) can be constructed by taking $D=2\Gamma$, $\delta=(\Gamma\cdot\Gamma)$, where Γ is an irreducible curve of virtual genus equal to one with positive self-intersection on a rational surface X. In fact the genus formula tells us that

$$(K_X \cdot \Gamma) < 0,$$

while

$$\text{(4.11)} \qquad \begin{cases} \dim|\Gamma| = (\Gamma\cdot\Gamma) - p_a(\Gamma) + 1 = \delta \\ \dim|2\Gamma| = 2\dim|\Gamma| + (\Gamma\cdot\Gamma) = 3\delta \\ p_a(2\Gamma) = 2p_a(\Gamma) + (\Gamma\cdot\Gamma) - 1 = \delta + 1. \end{cases}$$

Now, by (4.6), for a general choice of $p_1, \ldots, p_{\delta-1}$, there is an irreducible curve in $|2\Gamma|$ which is singular at $p_1, \ldots, p_{\delta-1}$. Then, by Lemma (4.2) and by (4.11), for a general choice of $p_1, \ldots, p_\delta$, there is a unique element of $|2\Gamma|$ which is singular at $p_1, \ldots, p_\delta$. This element is necessarily equal to $2\Gamma'$, where Γ' is the unique element of $|\Gamma|$ passing through $p_1, \ldots, p_\delta$; by (3.3), Γ' is smooth.

5.

As customary, we denote by $H(d, w)$ the Hurwitz space of d-sheeted simple coverings of $\mathbb{P}^1$ ramified at w points (cf. [2]). It is a classical result that $H(d, w)$ is irreducible, and we shall denote by

$$\text{(5.1)} \qquad h: H(d, w) \to \mathcal{M}_g, \qquad w = 2g + 2d - 2$$

the natural morphism. As one may see by an easy count of parameters, the subvariety $\mathcal{M}^1_{g,d}$ of $\mathcal{M}_g$ introduced in Sect. 1 is the closure of $h(H(d, w))$. In

symbols

$$\overline{h(H(d,w))} = \mathscr{M}^1_{g,d}. \tag{5.2}$$

The main result of this section is

(5.3) **Theorem.** *Let d be an integer such that $d \geqq 3$. Then the Hurwitz space $H(d,w)$, $w = 2g+2d-2$, is unirational in each one of the following cases:*

i) $d \leqq 5, \quad g \geqq d-1$

ii) $d = 6, \quad 10 \geqq g \geqq 5 \quad \text{or} \quad g = 12$

iii) $d = 7, \quad g = 7.$

Observe that the restriction $d \geqq 3$ is really due to our method of proof; it is of course well-known that, for instance, $H(2, 2g+2)$ is unirational.

Our main reason for proving (5.3) is to establish the following

(5.4) **Corollary.** *In each one of cases* i), ii), iii) *in Theorem* (5.3), $\mathscr{M}^1_{g,d}$ *is unirational. In particular $\mathscr{M}_g$ is unirational when $g \leqq 10$.*

Proof. The first part of this corollary follows directly from (5.3). For the second part, observe that, when $g \leqq 5$, the unirationality of $\mathscr{M}_g$ is obvious since a general curve of genus $g \leqq 5$ can be realized as a complete intersection. On the other hand, it is well known that every curve of genus $g \leqq 10$ has a g^1_6 (cf. [4], for example), so that, in case $5 \leqq g \leqq 10$, part ii) of (5.3) applies. Q.E.D.

As a curiosity, the reader should observe that, in case $d = 6$, $g = 10$, the unirational variety $H(6, 20)$ maps via (5.1) onto an open subset of $\mathscr{M}_{10}$ with general fibre consisting of *fourty-two* disjoint copies of $PGL(2, \mathbb{C})$. On the other hand, it follows from (2.6) that in case $d < (g/2)+1$, a general fibre of (5.1) consists of *one* copy of $PGL(2, \mathbb{C})$.

We want to stress that the unirationality of $\mathscr{M}_g$, $g \leqq 10$ (and of $\mathscr{M}^1_{g,3}$ of course) is classically known [10]. It has recently been proved by Sernesi [8], using entirely different methods, that $\mathscr{M}_{12}$ is also unirational. The problem of deciding whether $\mathscr{M}_g$ is unirational is, at the moment, open for $g = 11$ or $g > 12$.

Proof of Theorem (5.3). Let g be an integer such that $g \geqq d-1$. Inspired by Segre's theorem (0.2), we let n be the minimum among the integers m such that

$$m \geqq \frac{g+d+2}{2}.$$

Set

$$\delta = \binom{n-1}{2} - \binom{n-d}{2} - g.$$

Notice, that, in case i), ii) or iii) holds, all the hypotheses of (4.7) are satisfied, and that we are not in one of the exceptional cases mentioned in the statement of this Corollary. Set

$$N = \frac{n(n+3)}{2} - \binom{n-d+1}{2},$$

and fix a point $p \in \mathbb{P}^2$. Let

$$V' \subset \mathbb{P}^N \times \mathrm{Sym}^\delta(\mathbb{P}^2)$$

be the variety of pairs $(Y, q_1 + \ldots + q_\delta)$, where $q_1, \ldots, q_\delta$ are distinct and different from p, and Y is an irreducible plane curve of order n with an ordinary $(n-d)$-fold point at p, nodes at $q_1, \ldots, q_\delta$, and no other singularity. By Corollary 4.7, there is a component V of V' which projects onto an open subset of $\mathrm{Sym}^\delta(\mathbb{P}^2)$. Since the fibres of this projection are linear subspaces of $\mathbb{P}^N$, the variety V is clearly unirational. We want to show that projection from p yields a dominant rational map

$$V \dashrightarrow H(d, w). \tag{5.5}$$

This will prove our theorem.

Let X be the blow-up of $\mathbb{P}^2$ at p. Let

$$\pi: X \to \mathbb{P}^1$$

be the projection from p. Let $\Gamma \subset \mathbb{P}^2$ be an element of V, $\tilde{\Gamma}$ its proper transform, and

$$\varphi: C \to X$$

the normalization map for $\tilde{\Gamma}$. Since

$$N_\varphi \cong K_C \otimes \varphi^*(K_X^{-1}),$$
$$(K_X \cdot \tilde{\Gamma}) < 0,$$

we have

$$H^1(C, N_\varphi) = 0.$$

Letting

$$f = \pi \circ \varphi: C \to \mathbb{P}^1$$

be the d-sheeted ramified covering defined by the pencil of lines in $\mathbb{P}^2$ through p, we obviously have

$$H^1(C, N_f) = 0.$$

Therefore, by Horikawa's theory [3], there are local universal families of deformations of φ and f:

$$\begin{array}{ccc} \mathscr{C} & \xrightarrow{\tilde{\varphi}} & X \\ \downarrow{\scriptstyle \alpha} & & \\ A & & \end{array} \qquad \begin{array}{ccc} \mathscr{C}' & \xrightarrow{\tilde{f}} & \mathbb{P}^1 \\ \downarrow{\scriptstyle \beta} & & \\ B & & \end{array}$$

where

$$C = \alpha^{-1}(a) = \beta^{-1}(b), \qquad \tilde{\varphi}_{|C} = \varphi, \qquad \tilde{f}_{|C} = f,$$

for appropriate $a \in A$, $b \in B$. Composition with π (i.e. projection from p) induces a holomorphic map

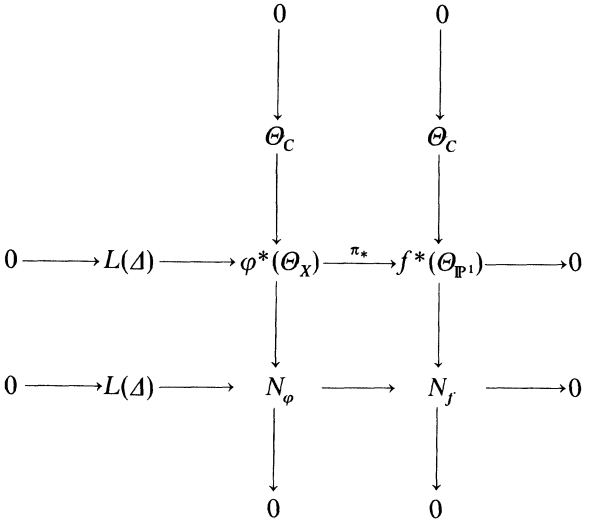
$$\rho: A \to B$$

such that $\rho(a)=b$. The differential of ρ at a can be identified with the homomorphism

$$\rho_*: H^0(C, N_\varphi) \to H^0(C, N_f)$$

induced by the corresponding sheaf homomorphism in the commutative diagram.

(5.6)



where L is the pullback to C of the hyperplane bundle on $\mathbb{P}^2$ and Δ is the pullback of the exceptional divisor on X.

An easy parameter count shows that a general point of B corresponds to a *simple* d-seeted ramified covering of $\mathbb{P}^1$, so that, in order to show that projection from p defines a dominant rational map (5.5), it suffices to show that ρ is open, i.e. that ρ_* is onto. By (5.6) this will follow if we can show that $H^1(C, L(\Delta))=0$ or, which is the same, that

$$|K_C \otimes L^{-1}(-\Delta)| = \emptyset.$$

To do this, it suffices to show that, for a general choice of $q_1, \ldots, q_\delta$, there are no plane curves of degree $n-4$ passing through $q_1, \ldots, q_\delta$ and having an $(n-d)$-fold point at p. In the course of his proof of (0.2), Segre does this for a *special* configuration of points $q_1, \ldots, q_\delta$ (cf. [6], p. 549, line 22, to p. 550, line 23), thereby establishing the result we need.

References

1. Arbarello, E., Cornalba, M.: Su una congettura di Petri. Comment. Math. Helv. **56**, 1-38 (1981)
2. Fulton, W.: Hurwitz schemes and irreducibility of moduli of algebraic curves. Ann. Math. **90**, 542-575 (1969)
3. Horikawa, E.: On deformations of holomorphic maps I. J. Math. Soc. Japan **25**, 372-396 (1973); II, ibid. **26**, 647-667 (1974)
4. Kleiman, S., Laksov, D.: Another proof of the existence of special divisors. Acta Math. **132**, 163-176 (1974)

5. Permutti, R.: Sui moduli delle curve k-gonali. Rend. Mat. ser. V, **9**, 54–58 (1950)
6. Segre, B.: Sui moduli delle curve poligonali e sopra un complemento al teorema di esistenza di Riemann. Math. Ann. **100**, 537–551 (1928)
7. Segre, C.: Preliminari di una teoria delle varietá di spazi. Rend. Circ. Mat. Palermo **30**, 87–121 (1910)
8. Sernesi, E.: L'unirazionalitá della varietá dei moduli delle curve di genere dodici. To appear in Ann. Scuola Norm. Sup. Pisa.
9. Severi, F.: Sul teorema di esistenza di Riemann. Rend. Circ. Mat. Palermo **46**, 105–116 (1922)
10. Severi, F.: Vorlesungen über algebraische Geometrie. Leipzig: 1921
11. Tannenbaum, A.: Families of algebraic curves with nodes. Compositio Math. **41**, 107–126 (1980)

Received October 13, 1980

Math. Ann. 256, 363–386 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Analytic Set-Valued Functions and Spectra

Zbigniew Słodkowski

Department of Mathematics, Polish Academy of Sciences, Sniadeckich 8, PL-00-950 Warsaw, Poland

1. Introduction

Let $F_c(X)$ denote the family of all compact non-void subsets of a topological space X.

Consider an analytic operator-valued function $T(\lambda)$ defined in $G \subset \mathbb{C}$. The behaviour of the set-valued function

$$\lambda \to \sigma(T(\lambda)) : G \to F_c(\mathbb{C}) \tag{1}$$

was studied in classical papers of Rellich, Kato and others; however, they were mainly concerned with isolated eigenvalues of finite multiplicity, which were found to depend analytically on λ. The new progress began with the result of Vesentini [16], that the spectral radius of (1) is a subharmonic function in G. Aupetit [2] has realised the full impact of Vesentini's theorem and shown how to apply it in spectral theory (see also [9, 12, 13]).

Let A be a uniform algebra with Shilov boundary X and $f, g \in A$. Wermer [18] has introduced the following set-valued function

$$\lambda \to \hat{g}(\hat{f}^{-1}(\lambda)) : G \to F_c(\mathbb{C}), \tag{2}$$

where $G = \sigma(f) \setminus f(X)$. Such set-valued functions are useful in studing analytic structures in $\mathfrak{M}_A$. In addition, they have properties similar to those of (1) (see [1, 18, 19]).

Much earlier, in a virtually forgotten paper, Oka [11] introduced the notion of *analytic multi-valued function* (we prefer "*set-valued*" to "multi-valued").

Definition. Let $G \subset \mathbb{C}$ be open. An upper semi-continuous set-valued function

$$\lambda \to K_\lambda : G \to F_c(\mathbb{C}) \tag{3}$$

is said to be *analytic* if the set

$$U = \{(\lambda, z) \in G \times \mathbb{C} : z \notin K_\lambda\} \tag{4}$$

is a domain of holomorphy.

0025-5831/81/0256/0363/$04.80

The results announced by Oka in [11] were proved by Nishino [10] and extended by Yamaguchi [22].

The aim of this paper is to show that these three classes of set-valued functions are practically the same. More specifically, they are locally the same, but more is proved.

The author's interest in this area was aroused by the following question of Aupetit. Assume that G is connected and that given multifunction of type (1) or (2) (we will often say "*multifunction*" instead of "set-valued function") has countable values at each $\lambda \in E$, where $E \subset G$ has positive exterior capacity; does it follow that it has countable values for all λ in G?

The positive answer to this question follows from the following theorem of Oka (Theorem I) and from the author's observation that classes (1) and (2) are contained in (3) (see Corollaries 3.3 and 3.4).

Theorem I (Oka [11], Nishino [10]). *Let $G \subset \mathbb{C}$ be open and connected and*

$$\lambda \to K_\lambda : G \to F_c(\mathbb{C})$$

be an analytic set-valued function. Assume that $E \subset G$ has positive exterior capacity and for each $\lambda \in E$ the set K_λ is countable. Then K_λ is countable for all λ in G.

In fact the author, unaware of Oka's paper, rediscovered Theorem I himself and announced it in [13], together with the following characterization of analytic multifunctions. (Only the condition (vi) has been borrowed recently from Wermer's [21].)

Theorem II. *Let $\lambda \to K_\lambda : G \to F_c(\mathbb{C})$ be upper-semi-continuous. Then the following conditions are equivalent.*

(i) *K_λ is analytic.*

(ii) *For every polynomial $\varrho(\lambda)$ and for every $a, b \in \mathbb{C}$ the function $\lambda \to \max|f_\lambda(K_\lambda)|$, where $f_\lambda(z) = (z - \lambda a - b)^{-1} \exp \varrho(\lambda)$ has local maximum property.*

(v) *For every open $G^* \subset G$ and for every $\psi(\lambda, z)$, plurisubharmonic in a neighbourhood of $\{(\lambda, z) \in G^* \times \mathbb{C} : z \in K_\lambda\}$ the function $\lambda \to \max\{\psi(\lambda, z) : z \in K_\lambda\}$ is subharmonic in G^*.*

(vi) *The algebra of all polynomials in λ, z, restricted to*

$$X = \{(\lambda, z) \in G \times \mathbb{C} : z \in K_\lambda\},$$

is a maximum-modulus algebra.

This is a comprised form of Theorem 3.2, which we prove in Sects. 2 and 3. In Sects. 4–6 we prove the following two theorems. (Below by $\max|K|$ we denote $\max\{|z| : z \in K\}$, provided $K \subset \mathbb{C}$ is compact.)

Theorem III. *Let $G \subset \mathbb{C}$ be open and bounded and $\lambda \to K_\lambda : G \to F_c(\mathbb{C})$ be analytic set-valued function. Assume that $\sup_{\lambda \in G} \max|K_\lambda| < \infty$. Then there is a uniform algebra A and $f, g \in A$ such that the following conditions hold*

(i) *$\sigma(f, g) = \mathrm{Cl}X$, where*

$$X = \{(\lambda, z) : \lambda \in G, z \in K_\lambda\},$$

(ii) $f(\partial_A) \subset \partial G$,
(iii) *A is separable.*

This theorem is a basic step in the proof of the following.

Theorem IV. *Let G and K_λ be the same as in Theorem III. Then there is a separable Hilbert space H and an analytic operator-valued function*

$$\lambda \to T(\lambda) : G \to B(H)$$

such that $K_\lambda = \sigma(T(\lambda))$ for all λ in G.

For more historical background and for related results the reader is referred to Aupetit [2, 24] and Wermer [19–21].

Throughout the paper the abbreviation "Cl" denotes "closure".

2. Pseudoconvexity and Local Maximum Principle

Connections between pseudoconvexity and local maximum properties constitute much of Theorem II (Theorem 3.2). Some of them, however, which seem to be of independent interest, are considered separately in the following theorem.

Theorem 2.1. *Let $U \subset \mathbb{C}^2$ be open. Then the following conditions are equivalent.*

(i) *The set U is pseudoconvex.*

(ii) *The algebra of all polynomials in λ, z, restricted to $X = \mathbb{C}^2 \setminus U$, is a maximum modulus algebra.*

(iii) *For every function $k : N \to \mathbb{C}$ which is a uniform limit of polynomials on a relative neighbourhood N in X, the function $|k|$ has no strict local maximum in N.*

(iv) *For every bounded relative neighbourhood N in X, and for every function $\psi(\lambda, z)$, plurisubharmonic in a neighbourhood of* $\mathrm{Cl}(N)$, *it holds*

$$\max \psi | \mathrm{Cl}(N) \leqq \max \psi | \mathrm{Cl}(N) \setminus N .$$

We say that an algebra A of continuous functions on a locally compact space X is a *maximum modulus algebra* if for every $f \in A$ and for every relatively compact neighbourhood $N \subset X$ it holds

$$\max |f| | \mathrm{Cl}(N) \leqq \max |f| | \mathrm{Cl}(N) \setminus N$$

(cf. [21]).

The implication (i)⇒(ii) is due to Wermer [21, Theorem 1]; his proof is different from ours.

To facilitate the description of the proof we consider two lemmas.

Lemma 2.2. *Assume that $U \subset \mathbb{C}^n$ is not pseudoconvex. Then there exist: a point $u^* \in U$, a compact polynomial $p(\mu)$, a pair of linearly independent vectors $a, b \in \mathbb{C}^n$, and two positive real numbers m, r, such that if we define*

$$T_\alpha = \{u = u^* + \mu b + \alpha a \exp p(\mu) : |\mu| = r\},$$
$$S_\alpha = \{u = u^* + \mu b + \alpha a \exp p(\mu) : |\mu| \leqq r\},$$

for $\alpha\in[0,m]$, then all of them, except S_m, are contained in U, while S_m does intersect $X=\mathbb{C}^n\setminus U$.

Proof. Is actually contained in [17, Sect. 12, 3–5], although this criterion is not stated there explicitely.

Lemma 2.3. *Let X_n, $n=1,2,\ldots$ be a sequence of closed subsets of $\mathbb{C}^2$, such that $X_n\supset X_{n+1}$ for all n. Put $X=$ the intersection of all X_n. Assume that for every X_n and for every strictly plurisubharmonic function ψ the condition* (iv) *of Theorem 2.1 holds. Then it holds for X and for arbitrary plurisubharmonic function ψ.*

Proof. Let N be relatively open in X and bounded. Then there is an open subset V of $\mathbb{C}^2$ such that $V\cap X=N$ and $\mathrm{Cl}(V)\cap X=\mathrm{Cl}(N)$; consequently $\mathrm{Cl}(N)\setminus N=\partial V\cap X$. (We omit the details.) Let ψ be plurisubharmonic in a neighbourhood of $\mathrm{Cl}(N)$; changing V if necessary we may assume that ψ is plurisubharmonic around $\mathrm{Cl}(V)$ too. By [17, Sect. 10.9] there is a sequence of smooth plurisubharmonic functions ψ_n, defined around $\mathrm{Cl}(V)$, such that $\psi_n\geqq\psi_{n+1}$ for $n=1,2,\ldots$ and $\lim\psi_n=\psi$ (pointwise) on $\mathrm{Cl}(V)$. We leave without proof the following

Assertion. Let $\psi_n:F_n\to[-\infty,+\infty)$ and $\psi:F\to[-\infty,+\infty)$ be upper semi-continuous, $\psi=\lim\psi_n$ point-wise on F, where F_n are compact, $F_n\supset F_{n+1}$ and $\bigcap F_n=F$. Then

$$\lim(\max\psi_n|F_n)=\max\psi|F\,.$$

Fix n. Since the functions $\chi_k(\lambda,z)=\psi_n(\lambda,z)+(1/k)(|\lambda|^2+|z|^2)$ are strictly plurisubharmonic, we have $\max\chi_k|\mathrm{Cl}(V)\cap X_n\leqq\max\chi_k|\partial V\cap X_n$. Applying the Assertion to both sides of this inequality we obtain

$$\max\psi_n|\mathrm{Cl}(V)\cap X_n\leqq\max\psi_n|\partial V\cap X\,,$$

for $\chi_k\searrow\psi_n$. Since $\psi_n\searrow\psi$, therefore by application of the assertion to both sides of the new inequality we get $\max\psi|\mathrm{Cl}(N)\leqq\max\psi|\mathrm{Cl}(N)\setminus N$. Q.E.D.

Proof of Theorem 2.1. We shall prove the following implications: (i)⇒(iv)⇒(iii)⇒(i).

(i)⇒(iv). By [7, IX.D.4] there exist strictly pseudoconvex domains [7, IX.A.2] such that $U_n\subset U_{n+1}$ for $n=1,2,\ldots$ and $U=\bigcup_{n=1}^{\infty}U_n$. Put $X_n=\mathbb{C}^2\setminus U_n$. Then $X=\bigcap X_n$ and $X_n\supset X_{n+1}$. Therefore, by Lemma 2.3 it is enough to check condition (iv) for X_n and for strictly plurisubharmonic functions ψ.

Let N be a bounded, relatively open subset of X_n and let ψ be a plurisubharmonic function in a neighbourhood of $\mathrm{Cl}(N)$. Choose (λ^*,z^*) such that $\max\psi|\mathrm{Cl}(N)=\psi(\lambda^*,z^*)$. By maximum principle for plurisubharmonic functions [17, Sect. 11.14], we have $(\lambda^*,z^*)\in\partial N$ (=the topological boundary of N in $\mathbb{C}^2$). Clearly $\partial N\subset(\mathrm{Cl}(N)\setminus N)\cup(N\cap\partial U_n)$. If $(\lambda^*,z^*)\in\mathrm{Cl}(N)\setminus N$ then we are done; we shall show that the case $(\lambda^*,z^*)\in\partial U_n$ is impossible. Indeed, by well-known properties of strictly pseudoconvex domains [7, Corollary IX.A.5, Theorem IX.B.2] there is a (quadratic) polynomial $p(\lambda,z)$ and a ball $B(r)$ around (λ^*,z^*) such that

$$B(r)\cap\mathrm{Cl}(U_n)\cap\{(\lambda,z):p(\lambda,z)=0\}=\{(\lambda^*,z^*)\}\,.$$

The equation $p(\lambda, z)=0$ defines a one-dimensional analytic variety and it is well-known that each point of such a variety has a neighborhood that is a union of several analytic discs. Let D^* be one such disc through (λ^*, z^*), that is

$$(\lambda^*, z^*)\in D^*\subset\{p=0\}.$$

We may also assume that $D^*\subset B(r)$; then $D^*\subset X_n$. Shrinking D^*, if necessary, we may also assume that $D^*\subset N$. Clearly $\max\psi|D^*=\psi(\lambda^*, z^*)$ and by maximum principle for subharmonic function [17, Sect. 9.6], ψ is constant on D^* and so ψ is not strictly plurisubharmonic at (λ^*, z^*), contrary to the assumption.

(iv)⇒(iii) is obvious.

(iii)⇒(ii). Let N be an arbitrary bounded relative neighbourhood in X. Let A denote the uniform closure of polynomials on $\mathrm{Cl}(N)$. Then A is a uniform algebra separating points of $\mathrm{Cl}(N)$. By the Bishop – de Leeuw theorem [6, Sect. 12.10] every function $f\in A$ assumes its maximum modulus at some peak point $x\in\mathrm{Cl}(N)$. Since the meaning of (iii) is that no such peak point belongs to N, we conclude that every polynomial assumes its maximum modulus on $\mathrm{Cl}(N)$ at some point of $\mathrm{Cl}(N)\backslash N$, as desired.

(ii)⇒(i). Suppose that U is not pseudoconvex. Then there exist u^*, $p(\mu)$, a, b, ... etc., as in Lemma 2.2, with properties stated there. Using them we will construct a bounded open set $N\subset\mathbb{C}^2$ and a function $h:\mathrm{Cl}(N)\cap X\to\mathbb{C}$, which is a uniform limit of polynomials on $\mathrm{Cl}(N)\cap X$ and for which $\max|h|>\max\limits_{x\in\partial N\cap X}|h(x)|$. Clearly this will contradict (ii).

Since a, b are linearly independent, we may consider an affine change of coordinates, so that in the new coordinates, say μ, v, the point u^* is $(0,0)$ and

$$T_\alpha=\{(\mu, \alpha\exp p(\mu)):|\mu|=r\},$$

$$S_\alpha=\{(\mu, \alpha\exp p(\mu)):|\mu|\leqq r\},$$

for $\alpha\in[0, m]$. Define $f(\mu, v)=v\exp(-p(\mu))$ and

$$N=\{(\mu, v)\in\mathbb{C}^2:|\mu|<r, |f(\mu, v)-m|<m\}.$$

It is clear that

$$\mathrm{Cl}(N)=\{(\mu, v)\in\mathbb{C}^2:|\mu|\leqq r, |f(\mu, v)-m|\leqq m\}.$$

Observe that if $(\mu, v)\in\mathrm{Cl}(N)$ and $f(\mu, v)-m\in(-\infty, 0)$ then $f(\mu, v)=\alpha$ with $\alpha\in[0, m]$ and $(\mu, v)\in S_\alpha$; consequently $(\mu, v)\notin X$. Thus

$$\mathrm{Cl}(N)\cap X\cap\{(\mu, v)\in\mathbb{C}^2:f(\mu, v)-m\in(-\infty, 0)\}=\emptyset. \tag{2.1}$$

Similarly, if $f(\mu, v)-m=0$ and $(\mu, v)\in\partial N$, then $(\mu, v)\in T_m$. Therefore

$$\partial N\cap X\cap\{(\mu, v)\in\mathbb{C}^2:f(\mu, v)-m\in(-\infty, 0]\}=\emptyset. \tag{2.2}$$

Let $r:\mathbb{C}\backslash(-\infty, 0)\to\mathbb{C}$ denote the unique continuous function such that $r(0)=0$ and $r|\mathbb{C}\backslash(-\infty, 0]$ is a holomorphic brach of $z^{1/2}$. Let Z^* denote the image of $\mathrm{Cl}(N)\cap X$ under $f(\mu, v)-m$. By (2.1) we have $Z^*\subset\mathbb{C}\backslash(-\infty, 0)$. It is geometrically clear that

there is a simply connected compact set Z such that $Z^* \subset Z \subset \mathbb{C}\setminus(-\infty, 0)$. (We omit the proof.) For each positive ε the function $r(z+\varepsilon)$ is holomorphic around Z and, by Runge theorem, this function is a uniform limit of polynomials on Z. It is clear (we omit the details) that $r(z) = \lim_{\varepsilon \to 0} r(z+\varepsilon)$, uniformly on Z (where $\varepsilon > 0$). Summing all this up we conclude that the function $g : \mathrm{Cl}(N) \cap X \to \mathbb{C}$, defined by $g(\mu, v) = r(f(\mu, v) - m)$ is a uniform limit of polynomials on $\mathrm{Cl}(N) \cap X$. The same holds true for $h = \exp(-g)$. Moreover by (2.2)

$$\max |h| \,|\, \partial N \cap X < 1 .$$

On the other hand by Lemma 2.2 there is a point $(\mu', v') \in S_m \cap X$. Since $f(\mu', v') - m = 0$ and $r(Z) \subset \{0\} \cup \{z : \mathrm{Re}\, z > 0\}$, we have

$$\max |h| \,|\, \mathrm{Cl}(N) \cap X \geqq |h(\mu', v')| = 1 . \quad \text{Q.E.D.}$$

3. Characterization of Analytic Set-Valued Functions

In this section we prove Theorem II in its extended form 3.2. Let us remind that a set-valued function $x \to K_x : X \to F_c(Y)$, where X, Y are locally compact spaces, is said to be *upper semi-continuous*, if for every open set $V \subset Y$ the set $\{x \in X : K_x \subset V\}$ is open.

Let $\varphi : G \to [0, \infty)$ be upper semi-continuous. We say conveniently that φ has *local maximum property* if for every open disc D, such that $\mathrm{Cl}(D) \subset G$, we have $\varphi(z) \leqq \max \varphi | \partial D$ for all $z \in D$. The following lemma is due to Cole [1, Sect. 4].

Lemma 3.1. *Let $G \subset \mathbb{C}$ be open and $\varphi : G \to [-\infty, +\infty)$ be upper semi-continuous. Assume that for every polynomial $p(\lambda)$ the function $|\exp p(\lambda)| \varphi(\lambda)$ has local maximum property in G. Then $\log \varphi$ is subharmonic in G.*

Theorem 3.2. *Let $G \subset \mathbb{C}$ be open and let $\lambda \to K_\lambda : G \to F_c(\mathbb{C})$ be an upper semi-continuous set-valued function. Denote*

$$U = \{(\lambda, z) \in G \times \mathbb{C} : z \notin K_\lambda\} ,$$

$$X = \{\lambda, z) \in G \times \mathbb{C} : z \in K_\lambda\} .$$

Then the following conditions are equivalent.

(i) *The multifunction K_λ is analytic, i.e. the set U is pseudoconvex.*

(ii) *For every polynomial $p(\lambda)$ and for every $a, b \in \mathbb{C}$ the function $\lambda \to \max |f_\lambda(K_\lambda)|$, where $f_\lambda(z) = (z - \lambda a - b)^{-1} \exp p(\lambda)$, has local maximum property in $\{\lambda \in G : a\lambda + b \notin K_\lambda\}$.*

(iii) *For every $a, b \in \mathbb{C}$ the function $\lambda \to \log \max \{|z - \lambda a - b|^{-1} : z \in K_\lambda\}$ is subharmonic in $\{\lambda \in G : a\lambda + b \notin K_\lambda\}$.*

(iv) *The function $(\lambda, z) \to -\log \mathrm{dist}(z, K_\lambda) : U \to [-\infty, +\infty)$ is plurisubharmonic.*

(v) *For every G^* open in G and for every $\psi(\lambda, z)$, plurisubharmonic in a neighbourhood of $(G^* \times \mathbb{C}) \cap X$, the function $\lambda \to \max \{\psi(\lambda, z) : z \in K_\lambda\}$ is subharmonic in G^*.*

(vi) *The algebra of all polynomials in λ, z, restricted to X, is a maximum modulus algebra.*

(vii) *For every relatively open and relatively compact* $N \subset X$ *and for every* $\psi(\lambda, z)$, *plurisubharmonic around* $\mathrm{Cl}(N)$, *it holds*

$$\max \psi | \mathrm{Cl}(N) \leqq \max \psi | \mathrm{Cl}(N) \setminus N .$$

Proof. We prove seven inequalities in the following order

$$\text{(i)} \Rightarrow \text{(vi)} \Rightarrow \text{(vii)} \Rightarrow \text{(v)} \Rightarrow \text{(ii)} \Rightarrow \text{(iii)} \Rightarrow \text{(iv)} \Rightarrow \text{(i)} .$$

(i)⇒(vi). Put $X^* = \mathbb{C}^2 \setminus U$. By Theorem 2.1 the restriction, to X^*, of the algebra of all polynomials is a maximum moduls algebra; therefore its restriction to X is also a maximum modulus algebra in view of the fact, that X is relatively open in X^* [for $X = (G \times \mathbb{C}) \cap X^*$; details omitted].

(vi)⇒(vii). In the same way, to have local maximum principle for plurisubharmonic functions on X it is enough to have it on X^*. This in turn is equivalent, by Theorem 2.1, to condition (iii) of Theorem 2.1 being fulfilled on X^*. Suppose that the latter does not hold, i.e. there is $(\lambda^*, z^*) \in X^*$ and a function k, a uniform limit of polynomials on a neighbourhood of (λ^*, z^*) in X^*, such that $|k|$ has strict local maximum at (λ^*, z^*). By (vi) this is not possible when $(\lambda^*, z^*) \in X$. If $(\lambda^*, z^*) \in X^* \setminus X = (\mathbb{C} \setminus G) \times \mathbb{C}$, this is not possible too, for $z \to k(\lambda^*, z)$ is analytic in a plane neighbourhood of z^*. This contradiction proves (vi)⇒(vii).

(vii)⇒(v). Put $\varphi(\lambda) = \max \{\psi(\lambda, z) : z \in K_\lambda\}$. By Lemma 3.1 it is enough to check that for arbitrary polynomial $p(\lambda)$ the function $\tilde{\varphi}(\lambda) = |\exp p(\lambda)| \exp \varphi(\lambda)$ has maximum modulus property in G^*. Take arbitrary $\lambda^* \in G^*$ and arbitrary $\mathrm{disc} D$, with $(\lambda^*, z^*) \in D$, such that $\mathrm{Cl}(D) \subset G^*$. Put $N = X \cap (D \times \mathbb{C})$, and

$$\tilde{\psi}(\lambda) = |\exp p(\lambda) \exp \psi(\lambda, z) .$$

Clearly N is both relatively open and relatively compact in X and

$$\mathrm{Cl}(N) \setminus N \subset (\partial D \times \mathbb{C}) \cap X ;$$

moreover $\tilde{\psi}$ is plurisubharmonic in a neighbourhood of $\mathrm{Cl}(N)$. Using these observations we obtain the following inequalities, which will give the desired conclusion.

$$\begin{aligned} \tilde{\varphi}(\lambda^*) &= \max \{\tilde{\psi}(\lambda^*, z) : z \in K_{\lambda^*}\} \\ &\leqq \max \tilde{\psi} | \mathrm{Cl}(N) \setminus N \quad \text{[by (vii)]} \\ &\leqq \max \tilde{\psi} | (\partial D \times \mathbb{C}) \cap X = \max \tilde{\psi} | \partial D . \end{aligned}$$

(v)⇒(ii). Since the function $\psi(\lambda, z) = |f_\lambda(z)|$ is plurisubharmonic, the function $\lambda \to |f_\lambda(K_\lambda)|$ is subharmonic and hence has local maximum property.

(ii)⇒(iii). This follows directly from Lemma 3.1.

(iii)⇒(iv). Put $\psi(\lambda, z) = -\log \mathrm{dist}(z, K_\lambda)$. Since ψ is upper semi-continuous (we omit the details), therefore ψ is plurisubharmonic if and only if its restriction to every complex line L is subharmonic in $L \cap U$. We consider two cases.

1. $L = \{\lambda\} \times \mathbb{C}$, with λ fixed. Since $\mathbb{C} \setminus K_\lambda$ is pseudoconvex, the function $z \to -\log \mathrm{dist}(z, K_\lambda)$ is subharmonic in $\mathbb{C} \setminus K_\lambda$. Since this function is identical, modulo suitable identifications, with $\psi | L$, we are done.

2. $L=\{(\lambda, a\lambda+b):\lambda\in\mathbb{C}\}$. We have to prove that the function $\lambda\to\psi(\lambda, a\lambda+b)$ is subharmonic in

$$\{\lambda:(\lambda, a\lambda+b)\in U\}=\{\lambda\in G: a\lambda+b\notin K_\lambda\}.$$

But

$$\psi(\lambda, a\lambda+b)=\log\max\{|z-\lambda a-b|^{-1}: z\in K_\lambda\},$$

and so we are done.

(iv)⇒(i). The difficulty of this step is caused by the fact, that the function $\psi(\lambda, z)=-\log\operatorname{dist}(z, K_\lambda)$ does not necessarily tend to $+\infty$ as (λ, z) approaches the boundary of U.

Suppose that U is not pseudoconvex. We will apply Lemma 2.2 to show that this is impossible. Let symbols a, b, u^*, $p(\mu)$, r, m, T_α and S_α have same meaning and properties as in Lemma 2.2. Choose $(\lambda', z')\in\partial U\cap S_m=\partial U\cap(S_m\setminus T_m)$.

Assertion. Let $(\lambda', z')\in S_m\setminus T_m$. Then there is $\delta>0$ and, for every $\alpha\in(m-\delta, m)$, a complex number $z(\alpha)$, such that $(\lambda', z(\alpha))\in S_\alpha\setminus T_\alpha$ and $\lim z(\alpha)=z'$.

Applying this Assertion we consider two cases.

1. $(\{\lambda'\}\times\mathbb{C})\cap U=\emptyset$. Then the assertion implies that S_α meets $\mathbb{C}^2\setminus U$ for $\alpha\in(m-\delta, m)$, which is impossible by Lemma 2.2.

2. If 1. does not hold we have $z'\in K_{\lambda'}$. Since T_m is compact and $T_m\subset U$, there is a neighbourhood $V\supset T_m$, such that $\mathrm{Cl}(V)$ is compact and contained in U. Since ψ is upper semi-continuous, $A=\max\psi|\mathrm{Cl}(N)<+\infty$. As $\lim T_\alpha=T_m$, the set V contains $\bigcup\{T_\alpha:\alpha\in(m-\eta, m)\}$ for some positive η. By maximum principle for plurisubharmonic functions [17, Sect. 14.11] we have $\max\psi|S_\alpha\leqq\max\psi|T_\alpha$, for $\alpha<m$. Finally ψ is bounded by A on the set $\bigcup\{T_\alpha:\alpha\in(m-\eta, m)\}$. However, this is impossible by the assertion. Indeed, we have $\lim\psi(\lambda', z(\alpha))=+\infty$ [for $z(\alpha)\to z'\in K_{\lambda'}$] and $(\lambda', z'(\alpha))\in S_\alpha\setminus T_\alpha$.

It remains to prove the assertion. Denote $u^*=(u_1^*, u_2^*)$, $a=(a_1, a_2)$ and $b=(b_1, b_2)$. Then there is $\mu'\in\mathbb{C}$ with $|\mu'|<r$ such that

$$u_1^*+\mu' b_1+ma_1\exp p(\mu')=\lambda',$$

$$u_2^*+\mu' b_2+ma_2\exp p(\mu')=z'.$$

Let $\varepsilon>0$ be small enough, so that $\mathrm{Cl}(D(\mu', \varepsilon))\subset D(0, r)$ (where D denotes disc of given radius and center) and the equation

$$u_1^*+\mu b_1+ma_1\exp p(\mu)-\lambda'=0$$

has exactly one root in $D(\mu', \varepsilon)$. If $\delta>0$ is small enough and $\alpha\in(m-\delta, m)$, then by Hurwitz's theorem [23, III, Sect. 11] the perturbed equation

$$u_1^*+\mu b_1+\alpha a_1\exp p(\mu)-\lambda'=0$$

has at least one root in $D(\mu', \varepsilon)$. Moreover, we can choose for each $\alpha\in(m-\delta, m)$ one of them, say $\mu(\alpha)$, so that $\lim\mu(\alpha)=\mu'$. Put $z(\alpha)=u_2^*+\mu(\alpha)b_2+\alpha a_2\exp p(\mu(\alpha))$. Then $\lim z(\alpha)=z'$ and $(\lambda', z(\alpha))\in S_\alpha\setminus T_\alpha$. Q.E.D.

Corollary 3.3. *Let $G \subset \mathbb{C}$ be open and X be a Banach space. If $\lambda \to T_\lambda : G \to B(X)$ is an analytic operator-valued function then $\lambda \to \sigma(T_\lambda) : G \to F_c(\mathbb{C})$ is an analytic multifunction.*

Proof. Put $K_\lambda = \sigma(T_\lambda)$. It is enough to check that the condition (iii) of Theorem 3.2 holds. Observe that the function $\log \max\{|z-\lambda a-b|^{-1} : z \in K_\lambda\}$ is subharmonic because it is equal to the logarithm of the spectral radius of an analytic operator-valued function [namely $(T_\lambda - (\lambda a+b)I)^{-1}$]; cf. Vesentini [16, Theorem 1]. Q.E.D.

Corollary 3.4. *Let A be a uniform algebra on X and $f, g \in A$. Then the multifunction $\lambda \to g(f^{-1}(\lambda)) : \sigma(h) \backslash h(X) \to F_c(\mathbb{C})$ is analytic.*

Proof. This is basically Wermer's argument [18]. It is enough to check that $K_\lambda = g(f^{-1}(\lambda))$ fulfills condition (ii) of Theorem 3.2. Let D be a disc such that $\mathrm{Cl}(D) \subset G$ where $G = \{\lambda \in \sigma(h) \backslash h(X) : a\lambda + b \notin K_\lambda\}$. Put $N = h^{-1}(D) \subset M_A$ and let B denote the uniform closure of $A|\mathrm{Cl}(N)$ on $\mathrm{Cl}(N)$. Then B is a uniform algebra with Gelfand space $M_B = \mathrm{Cl}(N)$. Therefore the formula $k = (g - ah - b)^{-1} \exp p(h)$, where $a, b \in \mathbb{C}$ and p is a polynomial, defines an element of B. Denote

$$f_\lambda(z) = (z - \lambda a - b)^{-1} \exp p(\lambda).$$

Then for $\lambda^* \in D$ holds

$$\begin{aligned} \max|f_{\lambda^*}(K_{\lambda^*})| &= \max|k(h^{-1}(\lambda^*))| \\ &\leqq \max|k| \,|\, \partial N \quad \text{(by Rossi's local maximum principle)} \\ &\leqq \max\{\max|k(h^{-1}(\lambda)| : \lambda \in \partial D\} \\ &= \max\{\max|f_\lambda(K_\lambda)| : \lambda \in \partial D\}. \end{aligned}$$

Thus the function $\lambda \to \max|f_\lambda(K_\lambda)|$ has local maximum property. Q.E.D.

4. A Special Set-Valued Selection

For the purpose of the proof of Theorem III we have to extend to several dimensions the notion of analytic set valued function. However, we shall discuss the extended notion only so far as it is necessary for our application.

Definition 4.1. Let G be an open subset of $\mathbb{C}^k$. An upper semi-continuous set-valued function $\lambda \to K_\lambda : G \to F_c(\mathbb{C}^n)$ is said to be *analytic* if and only if for every open subset G' of G, and for every plurisubharmonic function $\psi(\lambda, z)$, defined on a neighbourhood of $\{(\lambda, z) \in G' \times \mathbb{C}^n : z \in K_\lambda\}$, the function

$$\lambda \to \max\{\psi(\lambda, z) : z \in K_\lambda\} : G' \to [-\infty, +\infty)$$

is plurisubharmonic.

Proposition 4.2. (i). *Analyticity of a set-valued function is a local property, i.e. if $\lambda \to K_\lambda : G \to F_c(\mathbb{C}^n)$ and for every $\lambda_0 \in G$ there is a neighbourhood $B(\lambda_0)$ on which K_λ is analytic then K_λ is analytic on G.*

(ii) *If $f : G \to \mathbb{C}^n$, where $G \subset \mathbb{C}^k$, is a single-valued, holomorphic function, then the set-valued function $\lambda \to \{f(\lambda)\} : G \to F_c(\mathbb{C}^n)$ is analytic (in sense of Definition 4.1).*

(iii) *Let* $\lambda \to K_\lambda : G \to F_c(\mathbb{C}^n)$ *be upper semi-continuous. Then* K_λ *is analytic if and only if for every complex line, say* $\lambda = \lambda(t)$, $t \in \mathbb{C}$, *the multifunction*

$$t \to K_{\lambda(t)} : \{t \in \mathbb{C} : \lambda(t) \in G\} \to F_c(\mathbb{C}^n)$$

is analytic.

(iv) *Let* $\lambda \to K_\lambda : G \to F_c(\mathbb{C})$, *where* $G \subset \mathbb{C}^k$, *be an upper semi-continuous set-valued function. Then* K_λ *is analytic if and only if the set* $U = \{(\lambda, z) : \lambda \in G, z \notin K_\lambda\}$ *is locally pseudoconvex at each point* $(\lambda_0, z_0) \in \partial U \cap (G \times \mathbb{C})$.

Proof. (i) and (ii) are trivial.

(iii) Since an upper semi-continuous function is plurisubharmonic if and only if its restriction to each complex line is subharmonic, the "if" part of (iii) follows immediately. As for the "only if" part, by (i) it is enough to show that if $\lambda : \mathbb{C} \to \mathbb{C}^k$ is a $\mathbb{C}$-affine function and $\psi(t, z)$ is a function plurisubharmonic in a neighbourhood of the set $\{0\} \times K_{\lambda(0)} \subset \mathbb{C} \times \mathbb{C}^n$, then

$$\varphi(t) = \max \psi(\{t\} \times K_{\lambda(t)})$$

is subharmonic near 0. Let $\pi : \mathbb{C}^k \times \mathbb{C}^n \to \mathbb{C}^k \times \mathbb{C}^n$ be the orthogonal projection onto $\{\lambda(t) : t \in \mathbb{C}\} \times \mathbb{C}^n$. Now, whenever $(\lambda, z) \in \mathbb{C}^{k+n}$ is such that $\pi(\lambda, z) = (\lambda(t), z)$ for some t and $\psi(t, z)$ is defined, we put $\psi^*(\lambda, z) = \psi(t, z)$. Then ψ^* is plurisubharmonic in a neighbourhood of $\{\lambda(0)\} \times K_{\lambda(0)}$ (details omited) and, by Definition 4.1, the function $\varphi^*(\lambda) = \max \psi^*(\{\lambda\} \times K_\lambda)$ is plurisubharmonic in a neighborhood of $\lambda(0)$. Since $\varphi^*(\lambda(t)) = \varphi(t)$, the subharmonicity of φ follows.

(iv) We prove only the "if" implication (the other will not be used later). Let $\lambda : \mathbb{C} \to \mathbb{C}^k$ be an arbitrary $\mathbb{C}$-affine function. We have to prove that the multifunction $t \to K'_t = K_{\lambda(t)}$ is analytic in

$$G' = \{t \in \mathbb{C} : \lambda(t) \in G\}.$$

Put $U' = \{(t, z) : t \in G, z \notin K'_t\}$. By Theorem 3.2 it is enough to show that U' is pseudoconvex. Since the map

$$(t, z) \to (\lambda(t), z) : U' \to (\lambda(\mathbb{C}) \times \mathbb{C}) \cap U$$

is biholomorphic, it follows (by the assumption on U) that U' is locally pseudoconvex at all its boundary points relative to $G' \times \mathbb{C}$. Therefore if G_n is open and $\mathrm{Cl}(G_n) \subset G'$, then $(G_n \times \mathbb{C}) \cap U'$ is locally pseudoconvex and so pseudoconvex. If in addition $\bigcup_{n=1}^{\infty} G_n = G'$ and $\mathrm{Cl}(G_n) \subset G_{n+1}$, $n = 1, 2, \ldots$, then U' is a union of increasing sequence of pseudoconvex domains [namely $(G_n \times \mathbb{C}) \cap U'$] and so is pseudoconvex itself; cf. [17, Sect. 12.2]. Q.E.D.

Remark. The equivalence (iv) is not true if K_λ are subsets of $\mathbb{C}^n$ with $n \geqq 2$.

The following theorem covers the main technical difficulty of the proof of Theorem III.

Theorem 4.3. *For each positive* δ *there is an analytic set-valued function*

$$(z_1, z_2) \to L_{z_1, z_2} : \{|z_1|^2 + |z_2|^2 > \delta\} \to F_c(\mathbb{C}^2)$$

such that whenever $(z_3, z_4) \in L_{z_1, z_2}$ *then* $z_1 z_3 + z_2 z_4 = 1$.

If z (or A) denotes a vector (or a matrix) then we denote by z^- (or A^-) a vector (matrix) with conjugated terms. Then we may write conveniently

$$M := \{(z_1, z_2, z_3, z_4) \in \mathbb{C}^4 : z_1 z_3 + z_2 z_4 = 1\}$$
$$= \{(z, w) \in \mathbb{C}^2 \times \mathbb{C}^2 : (z, w^-) = 1\},$$

where $(,)$ denotes the standard inner product in $\mathbb{C}^2$. From this representation it follows immediately that M is invariant under the group $\{h^A\}$ of linear maps of $\mathbb{C}^4$ onto itself, where

$$h^A(z, w) = (Az, A^- w)$$

for $(z, w) \in \mathbb{C}^2 \times \mathbb{C}^2$ and A any unitary, 2×2 matrix. This invariance suggests that we should look for a multifunction L_z, satisfying Theorem 4.3, among those with invariant graphs; e.g. of the following form:

Let $\varrho : (\log\delta, \infty) \to \mathbb{R}$ *be smooth. We define*

$$L_{z_1, z_2} = \{w = \|z\|^{-2}(\bar{z}_1, \bar{z}_2) + t(z_2, -z_1) : |t| \leqq \|z\|^{-1} \exp\varrho(\log\|z\|)\} \tag{4.1}$$

for $\|z\| = (|z_1|^2 + |z_2|^2)^{1/2} > \delta$.

Studying such L_z we will make use of the fact that it is, in a sense, determined by its projection on the first coordinate, i.e. by the multifunction

$$z \to K_z : \{z \in \mathbb{C}^2 : \|z\| > \delta\} \to F_c(\mathbb{C}),$$

where

$$K_z = \{\|z\|^{-2}\bar{z}_1 + tz_2 : |t| \leqq \|z\|^{-1} \exp\varrho(\log\|z\|)\}.$$

Put

$$\varphi(z, z_3) = \log|z_2| - \log\|z\| + \varrho(\log\|z\|) - \log|f(z, z_3)|, \tag{4.2}$$

where $z = (z_1, z_2)$, $\|z\| > \delta$, $z_2 \neq 0$ and f is defined by

$$f(z, z_3) = z_3 - \bar{z}_1\|z\|^{-2}.$$

Observe that

$$U' := \{(z, z_3) \in \mathbb{C}^3 : \|z\| > \delta, z_2 \neq 0, z_3 \notin K_z\}$$
$$= \{(z, z_3) \in \mathbb{C}^3 : \|z\| > \delta, z_2 \neq 0, \varphi(z, z_3) < 0\}. \tag{4.3}$$

The following lemma gives conditions, in terms of function ϱ, for local pseudo-convexity of this domain at boundary points of special form.

Lemma 4.4. *Let* $\delta > 0$ *and* $\varrho : (\log\delta, \infty) \to \mathbb{R}$ *be smooth. Assume that* $\varphi(z_1, z_2, z_3)$ *is defined by* (4.2) *and* $x \in (\log\delta, \infty)$. *Then*

(i) *the complex gradient of* φ *at* $(0, e^x, \exp\varrho(x))$ *is proportional to the vector* $(e^{-2x}, \varrho'(x)e^{-x}, -1)$,

(ii) *the inequality*

$$\sum_{j,k=1}^{3} w_j \bar{w}_k \frac{\partial^2 \varphi}{\partial z_j \partial \bar{z}_k}(0, e^x, \exp\varrho(x)) > 0 \tag{4.4}$$

holds for all $(w_1, w_2, w_3) \in \mathbb{C}^3 \setminus \{(0,0,0)\}$ *satisfying*

$$\sum_1^3 w_j \frac{\partial \varphi}{\partial z_j}(0, e^x, \exp \varrho(x)) = 0, \tag{4.5}$$

if and only if the following inequalities hold

$$\varrho''(x) > 0, \tag{4.6}$$

$$\varrho''(x) \cdot ((\varrho'(x) - 1) \exp(2\varrho(x) - 2) - 2) > 2(1 + \varrho'(x))^2. \tag{4.7}$$

The proof of this lemma, purely computational, is postponed to the Appendix.

Proof of Theorem 4.3. Observe first that for every $a \in \mathbb{R}$ there is a smooth function $\varrho : (a, \infty) \to \mathbb{R}$, such that (4.6) and (4.7) are satisfied for all $x > a$. Indeed, one checks that the function

$$\varrho_a(x) = (x-a)^2/2 + 2(x-a) + 5 - a$$

is a good example (details omitted).

The proof will be completed when we show the following

Assertion. Let $\delta > 0$ and $\varrho : (\log \delta, \infty) \to \mathbb{R}$ be a smooth function satisfying (4.6) and (4.7) for all $x > \log \delta$. Assume that

$$z \to L_z : \{z \in \mathbb{C}^2 : \|z\| > \delta\} \to F_c(\mathbb{C}^2)$$

is defined by (4.1). Then L_z is an analytic multifunction.

Observe that each set L_z is a disc with centre $\bar{z}/\|z\|^2$, and that

$$h^A(z, \bar{z}/\|z\|^2) = (Az, (Az)^-/\|Az\|^2)$$

for unitary A. Moreover $(z_2, -z_1)$ is orthogonal to $\bar{z}/\|z\|^2$ and this relation is preserved by h^A, which is a unitary map too. Therefore

$$h^A(\{z\} \times L_z) = \{Az\} \times L_{Az}.$$

Put $M_\delta = \{(z, w) \in M : \|z\| > \delta\}$. Then the set

$$U = \{(z, w) \in M_\delta : w \notin L_z\}$$

is invariant under all h^A. By (4.1) every boundary point of U relative to M_δ is of the form (z, w) with

$$w = (\bar{z}_1, \bar{z}_2)/\|z\|^2 + t(z_2, -z_1),$$

where $|t| = \|z\|^{-1} \exp \varrho(\log \|z\|)$. Let $e^{i\theta} = t/|t|$ and put

$$A = \begin{bmatrix} e^{i\theta} z_2/\|z\| ; & e^{-i\theta} z_1/\|z\| \\ \bar{z}_1/\|z\| ; & \bar{z}_2/\|z\| \end{bmatrix}.$$

One may check by direct computation that A is unitary and

$$\begin{aligned} h^A(z, w) &= (0, \|z\|, e^{-i\theta} t \|z\|, \|z\|^{-1}) \\ &= (0, \|z\|, \exp \varrho(\log \|z\|), \|z\|^{-1}). \end{aligned}$$

Since all $\{h^A\}$ act as biholomorphic maps of M_δ onto itself, and pseudoconvexity is an invariant of biholomorphic mappings, we conclude that in order to know that U is locally pseudoconvex at all its boundary points relative to M_δ, it is enough to check this at each point of the form $(0, e^x, \exp\varrho(x), \exp(-x))$, where $x > \log\delta$.

Now observe that the map

$$(z_1, z_2, z_3) \to (z_1, z_2, z_3, (1 - z_1 z_3)/z_2) : U' \to U \setminus \{z_2 = 0\}, \tag{4.8}$$

where U' is defined by (4.3), is biholomorphic and surjective. By (4.3) and Lemma 4.4 the set U' is locally pseudoconvex at points $(0, e^x, \exp\varrho(x))$; by properties of the map (4.8) the set U is locally pseudoconvex in M_δ at each point $(0, e^x, \exp\varrho(x), \exp(-x))$, $x > \log\delta$, and by the above invariance argument, U is locally pseudoconvex at all its boundary points relative to M_δ.

By Proposition 4.2 (i), analyticity of L_z is a local matter. Pick $z^* = (z_1^*, z_2^*)$ with $\|z^*\| < \delta$. Without loss of generality we may assume (symmetry!) that $z_2^* \neq 0$. Using the fact that the map (4.8) is biholomorphic we conclude that U' is pseudoconvex at all its boundary points relative to $\{\|z\| > \delta, z_2 \neq 0\} \times \mathbb{C}$, and hence by Proposition 4.2 (iv), the multifunction K_z [cf. (4.3)] is analytic in $\{\|z\| > \delta, z_2 \neq 0\}$. Let $\psi(z_1, z_2, z_3, z_4)$ be plurisubharmonic in a neighborhood of the set $\{z^*\} \times L_{z^*}$. We have to prove that the function $\chi(z) = \max \psi(\{z\} \times L_z)$ is plurisubharmonic near z^*. Put

$$\psi'(z, z_3) = \psi(z; z_3, (1 - z_1 z_3)/z_2).$$

Then ψ' is plurisubharmonic and by analyticity of K_z the function $\chi'(z) = \max \psi'(\{z\} \times K_z)$ is plurisubharmonic near z^* and so is χ (for $\chi = \chi'$). Q.E.D.

5. Representing Analytic Multifunctions in Uniform Algebras

For the proof of Theorem III, which we give in this section, we need some additional properties of analytic multifunctions in several dimensions.

Proposition 5.1. *Let $G \subset \mathbb{C}^k$ be open.*

(i) *Assume that $\lambda \to K_\lambda : G \to F_c(\mathbb{C}^n)$ is an analytic multifunction. Then the following multifunctions are analytic:*

$$\lambda \to K_\lambda \times \{\lambda\} : G \to F_c(\mathbb{C}^{n+k}),$$
$$(z, \lambda) \to K_\lambda : \mathbb{C}^m \times G \to F_c(\mathbb{C}^n),$$
$$(z, \lambda) \to \{z\} \times K_\lambda : \mathbb{C}^m \times G \to F_c(\mathbb{C}^{m+n}).$$

(ii) *Assume that $V \subset \mathbb{C}^n$ is open and that multifunctions $\lambda \to K_\lambda : G \to F_c(V)$ and $z \to L_z : V \to F_c(\mathbb{C}^m)$ are analytic. Then the composition of L and K, i.e. the multifunction*

$$\lambda \to R_\lambda = \bigcup \{L_z : z \in K_\lambda\} : G \to F_c(\mathbb{C}^m)$$

is analytic.

(iii) *Let* $\lambda \to K_\lambda : G \to F_c(\mathbb{C}^n)$ *and* $\lambda \to L_\lambda : G \to F_c(\mathbb{C}^m)$ *be analytic multifunctions. Then their pointwise cartesian product*

$$\lambda \to K_\lambda \times L_\lambda : G \to F_c(\mathbb{C}^{n+m})$$

is analytic.

Proof. (i) Since all three cases are proven similarly, we consider only one of them, e.g. $\{z\} \times K_\lambda$. The argument is based on Proposition 4.2 (iii). Let

$$t \to (z(t), \lambda(t)) : \mathbb{C} \to \mathbb{C}^{k+n}$$

be $\mathbb{C}$-affine and let $\psi(t, z, u)$ be plurisubharmonic in a neighbourhood of the set $\{(0, z(0))\} \times K_{\lambda(0)}$; we have to show that the function

$$\varphi(t) = \max \psi(\{(t, z(t)\} \times K_{\lambda(t)})$$

is subharmonic. Put $\psi^*(t, u) = \psi(t, z(t), u)$. Clearly ψ^* is plurisubharmonic near $\{0\} \times K_{\lambda(0)}$, and so $\varphi^*(t) = \max \psi^*(\{t\} \times K_{\lambda(t)})$ is subharmonic. Therefore $\varphi(t)$ is subharmonic near 0, for $\varphi(t) = \varphi^*(t)$.

(ii) Let $G' \subset G$ be open and let $\psi(\lambda, v)$ be plurisubharmonic in H, a neighbourhood of $\{(\lambda, v) : \lambda \in G', v \in R_\lambda\}$. We have to prove that the function

$$\varphi(\lambda) = \max \{\psi(\lambda, v) : v \in R_\lambda\}$$

is plurisubharmonic in G'. Consider the function $\psi'(\lambda, z, v) = \psi(\lambda, v)$ in the set

$$H' = \{(\lambda, z, v) \in \mathbb{C}^{n+m+k} : (\lambda, v) \in H\}.$$

Observe that

$$\begin{aligned} \varphi(\lambda) &= \max \{\psi'(\lambda, z, v) : z \in K_\lambda, v \in L_z\} \\ &= \max \{\chi(\lambda, z) : z \in K_\lambda\}, \end{aligned} \tag{5.1}$$

where the function $\chi(\lambda, z) = \max \{\psi'(\lambda, z, v) : v \in L_z\}$ is defined in

$$W = \{(\lambda, z) : \lambda \in G', \{\lambda\} \times L_z \subset H\}.$$

Since ψ' is plurisubharmonic and the set-valued function $(\lambda, z) \to L_z$ is analytic [by (i)], therefore χ is plurisubharmonic. From this and (5.1) it follows that φ is plurisubharmonic.

(iii) The multifunction $\lambda \to K_\lambda \times L_\lambda$ is analytic, as a composition of the following two multifunctions, that are analytic by (i):

$$\lambda \to K_\lambda \times \{\lambda\} : G \to F_c(\mathbb{C}^n \times G),$$

$$(z, \lambda) \to \{z\} \times L_\lambda : \mathbb{C}^n \times G \to F_c(\mathbb{C}^{n+m}). \quad \text{Q.E.D.}$$

Proof of Theorem III. Let us fix for each $\delta > 0$ an analytic multifunction

$$(z_1, z_2) \to L(z_1, z_2; \delta) : \{z \in \mathbb{C}^2 : \|z\| > \delta\} \to F_c(\mathbb{C}^2),$$

satisfying conditions of Theorem 4.3. Put $X = \mathrm{Cl}\{(\lambda, z) \in \mathbb{C}^2 : \lambda \in G, z \in K_\lambda\}$; by assumptions X is compact. Put $U = \mathbb{C}^2 \setminus X$ and $T = \{1, 2\} \cup U \times \{1, 2\}$. Our aim is to define a compact subset Y of $\mathbb{C}^T$ (considered with topology of point-wise

convergence) and to introduce A as a uniform subalgebra of $C(Y)$ generated by some family of continuous functions. We define Y as the set of all $y=(y_t)_{t\in T}\in\mathbb{C}^T$, such that $(y_1, y_2)\in X$ and for every $(\lambda, z)\in U$ it holds

$$(y_{\lambda,z,1}, y_{\lambda,z,2})\in L(y_1-\lambda, y_2-z; (1/2)\operatorname{dist}(\lambda, z), X).$$

Since the multifunctions $L(\cdot,\cdot;\delta)$ are upper semi-continuous, Y is compact in $\mathbb{C}^T$. Now it is clear that the following formulae define continuous functions on Y:

$$e(y)=1; \quad f(y)=y_1; \quad g(y)=y_2; \quad a_{\lambda,z}(y)=y_{\lambda,z,1}$$

and

$$b_{\lambda,z}(y)=y_{\lambda,z,2} \quad \text{for} \quad (z,\lambda)\in U,$$

where $y=(y_t)_{t\in T}\in Y$. Let A denotes the uniform subalgebra of $C(Y)$ generated by the functions $e, f, g, a_{\lambda,z}, b_{\lambda,z}$ with $(\lambda, z)\in U$.

It remains to show that the algebra A and functions f, g satisfy conditions (i)–(iii) of Theorem III.

(i) Obviously $\sigma(f,g)\supset(f,g)(Y)=X$. To see that, conversely $\sigma(f,g)\subset X$, observe that if $(\lambda, z)\notin X$ then

$$\begin{aligned} &a_{\lambda,z}(y)(f(y)-\lambda)+b_{\lambda,z}(g(y)-z) \\ &=y_{\lambda,z,1}(y_1-\lambda)+y_{\lambda,z,2}(y_2-z)\equiv 1 \text{ on } Y \end{aligned} \tag{5.2}$$

by properties of the multifunctions $L(\cdot,\cdot;\delta)$ and definition of Y. Thus $(\lambda, z)\notin\sigma(f,g)$, as desired.

(ii) Since ∂_A, the Shilov boundary of A, is necessarily contained in Y, it is enough to check that for each $h\in A$ the maximum of $|h|$ on Y can not be assumed at any point of $Y\backslash f^{-1}(\partial G)$. In fact, it is enough to know this for every h belonging to a dense subset of A, e.g. for those of the form

$$w(f, g; a_{\lambda(1),z(1)}, \ldots, a_{\lambda(s),z(s)}; \quad b_{\lambda(s+1),z(s+1)}, \ldots, b_{\lambda(r),z(r)}),$$

where $s<r$, points $(\lambda(i), z(i))$, $i=1, \ldots, r$ vary through U and

$$w(\lambda, z; z_1, \ldots, z_s, z_{s+1}, \ldots, z_r)$$

is an arbitrary complex polynomial of $r+2$ variables. Observe that if for some function h of the above form the global maximum of $|h|$ on Y were assumed at some point of $Y\backslash f^{-1}(\partial G)$, then the following Assertion would have been false:

Assertion. For every function h of the above form the function

$$\lambda\to\log\max|h(f^{-1}(\lambda)|$$

is subharmonic in G.

Proof of the Assertion. Let $(x_1, x_2)\in\mathbb{C}^2$. Denote by $L(x_1, x_2)$ the cartesian product of the sets

$$L(x_1-\lambda(i), x_2-z(i); (1/2)\operatorname{dist}(\lambda(i), z(i)), X)),$$

where $i=1, \ldots, r$, whenever it is defined. By Proposition 5.1 (iii) the function

$$(x_1, x_2)\to L(x_1, x_2)$$

is an analytic set-valued function in a neighbourhood of X. Put

$$R_\lambda = \bigcup \{\{z\} \times L^{\sim}_{\lambda,z} : z \in K_\lambda\}$$

for $\lambda \in G$; clearly $\lambda \to R_\lambda$ is the composition of

$$\lambda \to \{\lambda\} \times K_\lambda \quad \text{and of} \quad (\lambda, z) \to \{z\} \times L^{\sim}_{\lambda,z}.$$

Since these two multifunctions are analytic [Proposition 5.1 (i)], therefore

$$\lambda \to R_\lambda : G \to F_c(\mathbb{C}^{2r+1})$$

is analytic too [Proposition 5.1 (ii)]. Now $\log|w(\cdot)|$ is plurisubharmonic, therefore the function

$$\varphi(\lambda) = \max \log|w(\lambda, z; y_1, y_3, \ldots, y_{2s-1}; y_{2s+2}, \ldots, y_{2r-2}, y_{2r}|,$$

where

$$(z, y_1, y_2, y_3, \ldots, y_{2r-1}, y_{2r}) \in R_\lambda,$$

is subharmonic in G. It follows from the construction of R_λ and h, that

$$\log \max |h(f^{-1}(\lambda)| = \varphi(\lambda) \quad \text{for} \quad \lambda \in G$$

(we omit the details). This proves the Assertion, and (ii).

(iii) However, the algebra A is not separable. To construct a separable one let us consider for every $(\lambda, z) \in U$ the set $U_{\lambda,z}$ of all $(\lambda', z') \in U$ such that

$$(f - \lambda')a_{\lambda,z} + (g - z')b_{\lambda,z}$$

is invertible in A. By (5.2) each $U_{\lambda,z}$ is a neighbourhood of (λ, z). Therefore we can choose a countable family $U_{\lambda(n),z(n)}$, $n = 1, 2, \ldots$, covering U. Define A' as the uniform subalgebra of A generated by the functions $e, f, g, a_{\lambda(n),z(n)}$, $b_{\lambda(n),z(n)}$, $n = 1, 2, \ldots$. Clearly A' is separable and we omit an easy argument that A' and f, g satisfy conditions (i) and (ii). Q.E.D.

6. Proof of Theorem IV

Let us outline first the idea of the proof. Using the algebra A of Theorem III we construct a complex vector boundle with base G and fiber an infinite dimensional Hilbert space, and represent each K_λ, where $\lambda \in G$, as the spectrum of an operator acting in the fiber over λ, so that the operator depends analytically on λ. Since every such bundle is globally analytically trivial, the theorem follows. The proof proceeds by series of lemmas.

Let $T \in B(X)$ (= the algebra of all bounded operators acting in a Banach space X). We remind that $\lambda \notin \sigma_\pi(T)$ *(= the approximate point spectrum)* iff there is a positive ε, such that $\|(T - \lambda I)x\| \geqq \varepsilon \|x\|$ for every $x \in X$.

Lemma 6.1. *Let A be a uniform algebra and X its Shilov boundary. Then there is a separable Hilbert space H and an isometric unital homomorphism $T: A \to B(H)$ such*

that

(i) $\sigma_\pi(T_a) \subset a(X)$ *for every* $a \in A$,

(ii) *for every* $f, g \in A$, *if* $0 \notin \sigma(f) \setminus f(X)$ *then*

$$\sigma(Tg\hat{}) = \{z \in \mathbb{C} : (0, z) \in \sigma(f, g)\},$$

where $(Tg)\tilde{}$ *denotes the operator induced by* Tg *on the quotient space* $H/\mathrm{Im}(Tf)$.

Proof. Since A is separable, the Gelfand space M_A of A is separable too. Choose a dense sequence $(\mu_n)_{n=1}^\infty$ and let m_n denote a fixed probability measure on X representing μ_n. We let $L^2(m_n)$ denote the Hilbert space of (classes of) square summable functions with respect to m_n (on X), with the usual norm. We define $H_n = H^2(m_n)$ as the closure of A in $L^2(m_n)$, and put $H =$ the direct Hilbertian sum of H_n, $n = 1, 2, \ldots$. Now, for any $a \in A$ and $x \in H$ we put $(Ta)x = (ax)_{n=1}^\infty$, where $x = (x_n)_{n=1}^\infty$ and $x_n \in H_n$. Clearly Ta is a bounded operator on H with norm $\leqq \|a\|$, and $T: A \to B(H)$ is a unital homomorphism. To see that $\|Ta\| \geqq \|a\|$ for every $a \in A$, fix $a \in A$ and let $n(k)$, $k = 1, 2, \ldots$ be a sequence such that $\|a\| = \lim |a\hat{}(\mu_{n(k)})|$. Define for each k an element $x^k = (x_n^k)_{n=1}^\infty \in H$, where $x_n^k = e$ if $n = n(k)$ and zero otherwise. Then $\|x^k\| = 1$ and $\|(Ta)x^k\| = |\mu_{n(k)}(a)|$ converges to $\|a\|$; therefore $\|Ta\| \geqq \|a\|$.

Observe that to prove (i) it is enough to show that if $0 \notin a(X)$, i.e. $|a| \geqq \varepsilon$ on X for some positive ε, then $0 \notin \sigma_\pi(Ta)$. The latter holds by the following inequalities:

$$\|(Ta)(x)\|^2 = \sum_1^\infty \int |ax_n|^2 dm_n$$

$$\geqq \sum_1^\infty \int \varepsilon |x_n|^2 dm_n = \varepsilon \|x\|^2,$$

where $x = (x_n)_{n=1}^\infty \in H$.

By (i), if $0 \in \sigma(f) \setminus f(X)$ than range of Tf is a proper, closed subspace of H and the formula $(Tg)\tilde{}(x + (Tf)(H)) = gx + (Tf)(H)$ defines a bounded operator on the quotient space $H/(Tf)(H)$. If $(0, z) \notin \sigma(f, g)$ then $af + b(g - ze) = e$ for some $a, b \in A$. Seeing that T is a homomorphism, one may conclude that $(Tb)\tilde{}((Ta)\tilde{} - zI) = I$, where $(Ta)\tilde{}, (Tb)\tilde{} \in B(H/\mathrm{Im}(Tf))$ are induced by Ta, Tb respectively. Thus $z \notin \sigma((Tg)\tilde{})$.

Conversely, if $(0, z) \in \sigma(f, g)$ then there is a sequence $(n(k))_1^\infty$ of integers, such that

$$(0, z) = \lim_k (\mu_{n(k)}(f), \mu_{n(k)}(g)).$$

Similarly as above we put $x^k = (x_n^k)_{n=1}^\infty$, where $x_n^k = e$ if $n = n(k)$ and zero otherwise. Clearly, elements of the form $a = (a_n)_{n=1}^\infty$, with all a_n's in A, are dense in H. For each such a we have:

$$(a, (Tf - \mu_{n(k)}(f)I)^* x^k) = ((Tf - \mu_{n(k)}(f)I), x^k)$$

$$= \int (f - \mu_{n(k)}(f)e) a_{n(k)} dm_{n(k)}$$

$$= \mu_{n(k)}((f - \mu_{n(k)}(f)e) a_{n(k)}) = 0.$$

Thus $(Tf - \mu_{n(k)}(f)I)^* x^k = 0$ and so $\|(Tf)^* x^k\| \to 0$. In the same way one checks that $\|(Tg - zI)^* x^k\| \to 0$. (We omit this.) These two relations together with the fact that $\|x^k\| = 1$, $k = 1, 2, \ldots$ imply that

$$(Tf)(H) + (Tg - zI)(H) \neq H.$$

(We omit standard argument involving, for example, the open mapping theorem.) Therefore $z \in \sigma((Tg)^{\sim})$. Q.E.D.

The next lemma is due to Taylor [15, 2.1, 2.2].

Lemma 6.2. *Let X, Y, Z be Banach spaces and let $\lambda \to A(\lambda): G \to B(X, Y)$ and*

$$\lambda \to B(\lambda): G \to B(Y, Z)$$

be analytic functions, such that $B(\lambda) \circ A(\lambda) = 0$ for all $\lambda \in G$. Assume that $B(\lambda^*)(Y) = Z$ and $\ker B(\lambda^*) = \operatorname{Im} A(\lambda^*)$ for a given $\lambda^* \in G$. Then there is an open neighbourhood G^* of λ^* such that

(i) $\ker B(\lambda) = \operatorname{Im} A(\lambda)$ *for every* $\lambda \in G^*$,

(ii) *for every analytic function $g: G^* \to Y$ satisfying $B(\lambda) g(\lambda) \equiv 0$ in G and for every $f_0 \in X$ such that $A(\lambda^*) f_0 = g(\lambda^*)$, there is an analytic function $f: G^* \to X$ satisfying $f(\lambda^*) = f_0$ and $A(\lambda) f(\lambda) = g(\lambda)$ in G^*.*

Corollary 6.3. *Let $\lambda \to V(\lambda): G \to B(H)$ be an analytic operator-valued function (H denotes a Hilbert space). Assume that $0 \notin \sigma_\pi(V(\lambda))$ for each $\lambda \in G$. Let $\lambda^* \in G$. Then there is an open neighbourhood $\lambda^* \in G^* \subset G$, a Hilbert space H', and an analytic function*

$$\lambda \to R(\lambda): G^* \to B(H, H'),$$

such that $R(\lambda)(H) = H'$ and $\ker R(\lambda) = \operatorname{Im} V(\lambda)$ for each $\lambda \in G^*$.

Proof. Choose a Hilbert space H' isomorphic to the quotient space $H/V_{\lambda^*}(H)$ and fix a bounded operator $R_0 \in B(H, H')$ such that $\operatorname{Im} V_{\lambda^*} = \ker R_0$ and $\operatorname{Im} R_0 = H'$. Consider Banach spaces $X = Y = B(H, H')$ and $Z = (0)$, and analytic functions

$$\lambda \to A(\lambda): G \to B(X, Y)$$

and $B(\lambda) \equiv 0$ in G, where $A(\lambda)(T) = T \circ V(\lambda)$ for $T \in B(H, H')$. Observe that $A(\lambda)$ is onto [use the fact that $V(\lambda)$ is invertible on $V(\lambda)(H)$] and $A(\lambda^*) R_0 = 0$. We may apply Lemma 6.2 to so defined X, Y, Z and $A(\lambda)$, $B(\lambda)$ and data $f_0 = R_0$ and $g(\lambda) \equiv 0$. We obtain that there is a neighbourhood G' of λ^* and an analytic function

$$\lambda \to R(\lambda): G' \to B(H, H')$$

such that $A(\lambda) R(\lambda) = 0$ and $R(\lambda^*) = R_0$. Thus, by our construction, we have that $\ker R(\lambda^*) = \operatorname{Im} V(\lambda^*)$ and $\operatorname{Im} R(\lambda^*) = H'$. Moreover $R(\lambda) \circ V(\lambda) \equiv 0$ in G'. Therefore we may apply Lemma 6.2 (i) to $X = Y = H$, $Z = H'$ and to $A(\lambda)$, $B(\lambda)$ equal respectively $V(\lambda)$ and $R(\lambda)$, and we conclude that $\operatorname{Im} V(\lambda) = \ker R(\lambda)$ for λ near λ^*. It is well known that $\operatorname{Im} R(\lambda) = H'$ for λ close enough to λ^* [if $\operatorname{Im} R(\lambda^*) = H'$]. Q.E.D.

Corollary 6.4. *Let X, Y be Banach spaces and let $A(\lambda)$, $B(\lambda) \in B(X, Y)$ be bounded linear operators depending analytically on $\lambda \in G$. Assume that $B(\lambda)$ is onto and $\ker A(\lambda)$ contains $\ker B(\lambda)$. Put*

$$\tilde{A}(\lambda)(x + \ker B(\lambda)) = A(\lambda)(x) \quad \text{and} \quad \tilde{B}(\lambda)(x + \ker B(\lambda)) = B(\lambda)(x)$$

for $x \in X$. Then $\tilde{A}(\lambda)$ and $\tilde{B}(\lambda)$ are bounded linear operators from $X/\ker B(\lambda)$ to Y and the composition

$$\lambda \to \tilde{A}(\lambda) \circ (\tilde{B}(\lambda))^{-1}: G \to B(Y)$$

is an analytic function.

Proof. Denote $C(\lambda)=A\tilde{}(\lambda)\circ(B\tilde{}(\lambda))^{-1}$. Observe first, that $C(\lambda)$ is well defined since, by the open mapping theorem, $B\tilde{}(\lambda)$ is a homomorphism.

By a theorem of Hille [5, Sects. 3.10 and 3.11] it suffices to show that for every $y\in Y$ and $y^*\in Y^*$ and for every $\lambda^*\in G$ the function

$$\lambda\to\langle y^*, C(\lambda)y\rangle : G\to\mathbb{C}$$

is analytic in a neighbourhood of λ^*. Fix $\lambda^*\in G$ and $y\in Y$. By Lemma 6.2 (ii) there is a neighbourhood G^* of λ^* and an analytic function

$$\lambda\to x(\lambda): G^*\to X,$$

such that $B(\lambda)x(\lambda)\equiv y$ in G^*. Observe that

$$\begin{aligned} C(\lambda)y &= A\tilde{}(\lambda)\circ(B\tilde{}(\lambda))^{-1}y = A\tilde{}(\lambda)\,(x(\lambda)+\ker B(\lambda)) \\ &= A(\lambda)x(\lambda), \end{aligned}$$

and hence the function $\langle y^*, C(\lambda)y\rangle = \langle y^*, A(\lambda)x(\lambda)\rangle$ is analytic. Q.E.D.

Proof of Theorem IV. Given $\lambda\to K_\lambda : G\to F_c(\mathbb{C})$ satisfying assumptions of the Theorem, let A and f, g denote a uniform algebra and a pair of its elements, such that condition (i)–(iii) of Theorem III are satisfied. Let H and $T: A\to B(H)$ be a Hilbert space and a representation satisfying Lemma 6.1.

Put $V(\lambda)=Tf-\lambda I\in B(H)$ and $S(\lambda)=Tg\in B(H)$. Then it is obvious that $V(\lambda)$ and $S(\lambda)$ satisfy the assumptions of the following

Assertion. Let $V(\lambda)$ and $S(\lambda)$ be $B(H)$-valued analytic functions in $G\subset\mathbb{C}$, where H is a separable Hilbert space. Assume that $0\notin\sigma_\pi(V(\lambda))$ and $V(\lambda)$ commutes with $S(\lambda)$ for each $\lambda\in G$. Let

$$S\tilde{}(\lambda)\in B(H/V(\lambda)(H))$$

denote the operator induced in a natural way by $S(\lambda)$. Then there is a separable Hilbert space H_1, and an analytic, $B(H_1)$-valued function $T(\lambda)$ in G, such that $\sigma(T(\lambda))=\sigma(S\tilde{}(\lambda))$ for each $\lambda\in G$.

Since, with $V(\lambda)$ and $S(\lambda)$ defined as above, $\sigma(S\tilde{}(\lambda))=K_\lambda$, therefore the proof will be finished, if we show the Assertion. Assume for a while that G is connected. By Corollary 6.3 there is an open covering $\{G_j\}_{j=1}^{\infty}$ of G, a Hilbert space H', and analytic functions

$$\lambda\to R_j(\lambda): G_j\to B(H, H'), \qquad j=1,2,\ldots$$

such that for each j and for each $\lambda\in G_j$ it holds $\ker R_j(\lambda)=\operatorname{Im}V(\lambda)$ and $\operatorname{Im}R_j(\lambda)=H'$. [We may take H' independent on j, because for connected G the dimension of $H/V(\lambda)(H)$ is constant.] Since

$$\ker R_j(\lambda)=\ker R_k(\lambda)=\operatorname{Im}V(\lambda) \quad \text{for} \quad \lambda\in G_j\cap G_k,$$

by Corollary 6.4 the function

$$\lambda\to R_{jk}(\lambda)=R_j\tilde{}(\lambda)\circ(R_k\tilde{}(\lambda))^{-1}: G_j\cap G_k\to B(H') \tag{6.1}$$

is analytic $[\tilde{R_j}(\lambda): H/V_\lambda(H) \to H'$ is canonically induced by $R_j(\lambda)]$. Put for every $\lambda \in G_j$

$$T_j(\lambda) = \tilde{R_j}(\lambda) \circ \tilde{S}(\lambda) \circ (\tilde{R_j}(\lambda))^{-1} \in B(H');$$

here $\tilde{S}(\lambda)$ is defined as in the Assertion.

We show that $T_j(\lambda)$ is analytic in G_j. Write

$$T_j(\lambda) = (R_j(\lambda) \circ S(\lambda))\tilde{} \circ (\tilde{R_j}(\lambda))^{-1}, \tag{6.2}$$

where $(R_j(\lambda) \circ S(\lambda))\tilde{}: H/V(\lambda)H \to H'$ is induced in a natural way by $R_j(\lambda) \circ S(\lambda)$. Observe that $\ker R_j(\lambda) \circ S(\lambda) \supset \ker R_j(\lambda)$, for, if $x \in \ker R_j(\lambda) = \operatorname{Im} V(\lambda)$ then $x = V(\lambda)h$ and we have

$$R_j(\lambda)S(\lambda)x = R_j(\lambda)S(\lambda)V(\lambda)x = R_j(\lambda)V(\lambda)S(\lambda)x = 0.$$

Therefore by (6.2) and Corollary 6.4 the function $T_j(\lambda)$ is analytic.

Observe now that if $\lambda \in G_j \cap G_k \cap G_r$ then $R_{jk}(\lambda)R_{kr}(\lambda) = R_{jr}(\lambda)$. Thus the covering $\{G_j\}$ and the family $R_{jk}(\lambda)$ gives an analytic vector bundle with base space G and fiber H', cf. [4]. Let us make for a while an additional assumption that H' is infinite dimensional. By a theorem of Bungart [4], every complex analytic vector bundle over a Stein space with fiber an infinite dimensional Hilbert space is globally analytically trivial. Thus for every j there is an analytic, $B(H')$-valued function $U_j(\lambda)$ in G_j, with all $U_j(\lambda)$ invertible, such that

$$R_{jk}(\lambda) = U_j(\lambda) \circ (U_k(\lambda))^{-1} \quad \text{for} \quad \lambda \in G_j \cap G_k.$$

By (6.1) we have $(U_j(\lambda))^{-1}\tilde{R_j}(\lambda) = (U_k(\lambda))^{-1}\tilde{R_k}(\lambda)$ for every $\lambda \in G_j \cap G_k$; therefore the formula

$$T(\lambda) = U_j(\lambda)^{-1}T_j(\lambda)U_j(\lambda), \quad \text{for} \quad \lambda \in G_j,$$

defines $T(\lambda)$ independently on the choice of j, and $T(\lambda)$ is an analytic, $B(H')$-valued function on G. Moreover, $T(\lambda)$ is similar to $\tilde{S}(\lambda)$ via $U_j(\lambda)^{-1}\tilde{R_j}(\lambda)$, and so $\sigma(T(\lambda)) = \sigma(\tilde{S}(\lambda))$. This settles the case of infinite dimensional H'.

When H' is finite dimensional we take an infinite dimensional separable Hilbert space H_0 and put $H_1 = H' \bar{\otimes} H_0$ (= the algebraic tensor product completed with respect to the natural inner product). Now if we replace the operators $R_j(\lambda)$ defined above by operators

$$\tilde{R_j}(\lambda) \otimes \mathrm{Id}_{H_0} : (H/V(\lambda)H) \bar{\otimes} H_0 \to H' \bar{\otimes} H_0$$

and $\tilde{S}(\lambda)$ by

$$\tilde{S}(\lambda) \otimes \mathrm{Id}_{H_0} \in B((H/V(\lambda)(H)) \bar{\otimes} H_0),$$

and proceed then in the same way as above, then all will work without changes. Eventually we would obtain an analytic operator valued function $\lambda \to T(\lambda) : G \to B(H_1)$ such that $\sigma(T(\lambda)) = \sigma(\tilde{S}(\lambda) \otimes \mathrm{Id}_{H_0}) = \sigma(S(\lambda))$.

In case of disconnected G we treat each component separately. Since we arrive at infinite dimensional separable Hilbert space H_1 in any case, this space may be considered as independent on the component. Q.E.D.

Appendix

In this section we prove Lemma 4.4.

The following lemma collects the formulae used in our computations.

Lemma. (i) *Let $U \subset \mathbb{C}^n$ be open and functions $h: U \to (a,b)$ and $\gamma: (a,b) \to \mathbb{R}$ be smooth. Then*

$$\frac{\partial}{\partial z_j}\gamma(h(z)) = \gamma'(h(z))\frac{\partial}{\partial z_j}h(z), \qquad z \in U, \tag{A.1}$$

$$\frac{\partial^2}{\partial z_j \partial \bar{z}_k}\gamma(h(z)) = \gamma'(h(z))\frac{\partial^2}{\partial z_j \partial \bar{z}_k}h(z) + \gamma''(h(z))\frac{\partial h(z)}{\partial z_j}\cdot\frac{\partial h(z)}{\partial \bar{z}_k}. \tag{A.2}$$

(ii) *If $f: U \to \mathbb{C}$, where $U \subset \mathbb{C}^n$ is smooth and does not vanish in U, then*

$$\frac{\partial}{\partial z_j}\log|f| = \frac{1}{2|f|^2}\frac{\partial}{\partial z_j}(f\bar{f}), \tag{A.3}$$

$$\frac{\partial^2}{\partial z_j \partial \bar{z}_k}\log|f| = -\frac{1}{2|f|^4}\frac{\partial}{\partial z_j}(f\bar{f})\frac{\partial}{\partial \bar{z}_k}(f\bar{f}) + \frac{1}{2|f|^2}\frac{\partial^2}{\partial z_j \partial \bar{z}_k}(f\bar{f}). \tag{A.4}$$

We omit the proof.

Proof of Lemma 4.4. Throughout the proof z denotes the vector (z_1, z_2) and $a(x)$ denotes the point $(0, e^x, \exp\varrho(x))$; (z,z) always denotes the inner product $z_1\bar{z}_1 + z_2\bar{z}_2$.

(i) For sufficiently smooth $h: U \to \mathbb{C}$, where $U \subset \mathbb{C}^n$, $\operatorname{grad} h$ denotes the vector $\left(\frac{\partial h}{\partial z_1}, \ldots, \frac{\partial h}{\partial z_n}\right)$. Applying (A.1) first to $\gamma(x) = (1/2)\log x$ and $h(z; z_3) = z_2\bar{z}_2$, next to $\gamma(x) = (1/2)\log x$ and $h(z, z_3) = (z,z)$, and finally to $\gamma(x) = \varrho((1/2)\log x)$ and $h(z; z_3) = (z,z)$ we get

$$\operatorname{grad}\log|z_2| = (0, 1/2z_2, 0),$$

$$\operatorname{grad}\log\|z\| = (\bar{z}_1/2\|z\|^2, \bar{z}_2/2\|z\|^2, 0),$$

$$\operatorname{grad}\varrho(\log\|z\|) = (1/2)\|z\|^{-2}\varrho'(\log\|z\|)\cdot(\bar{z}_1, \bar{z}_2, 0).$$

Evaluating these formulae at $(z; z_3) = a(x)$ we obtain that

$$\operatorname{grad}(\log|z_2| - \log\|z\| + \varrho(\log\|z\|))(a(x)) = (0, \varrho'(x)e^{\pi x}/2, 0). \tag{A.5}$$

To compute $\operatorname{grad}\log|f|$ observe first that

$$f\bar{f} = z_3\bar{z}_3 - (z_1 z_3 + \bar{z}_1\bar{z}_3)(z,z)^{-1} + z_1\bar{z}_1(z,z)^{-2}.$$

Differentiating this equation we get

$$\frac{\partial}{\partial z_1}(f\bar{f}) = -z_3(z,z)^{-1} + (\bar{z}_1 + z_1\bar{z}_1z_3 + \bar{z}_1^2\bar{z}_3)(z,z)^{-2} - 2\bar{z}_1^2z_1(z,z)^{-3}, \tag{A.6}$$

$$\frac{\partial}{\partial z_2}(f\bar{f}) = (z_1\bar{z}_2z_3 + \bar{z}_1\bar{z}_2\bar{z}_3)(z,z)^{-2} - 2z_1\bar{z}_1\bar{z}_2(z,z)^{-3}, \tag{A.7}$$

$$\frac{\partial}{\partial z_3}(f\bar{f}) = \bar{z}_3 - z_1(z,z)^{-1}. \tag{A.8}$$

Using this and (A.3) we check that

$$(\operatorname{grad}\log|f|)(a(x)) = (1/2)\exp(-\varrho(x))\cdot(e^{-2x}, 0, 1),$$

and by (A.5) we get

$$(\operatorname{grad}\varphi)(a(x)) = (1/2)\exp(-\varrho(x))\cdot(e^{-2x}, \varrho'(x)\exp\varrho(x), -1),$$

as desired.

(ii) We compute first the complex Hessian of φ at (x), i.e. the following Hermitian matrix:

$$\left[\frac{\partial^2\varphi}{\partial z_j\partial\bar{z}_j}\right]_{j,k=1}^{3}.$$

Applying (A.2) first to $\gamma(x) = (1/2)\log x$ and $h(z; z_3) = z_2\bar{z}_2$, next to $\gamma(x) = (1/2)\log x$ and $h(z; z_3) = (z, z)$, and finally to $\gamma(x) = \varrho((1/2)\log x)$ and $h(z; z_3) = (z, z)$ we get

Hessian $\log|z_2| = (0)$,

$$\frac{\partial^2}{\partial z_j\partial\bar{z}_k}\log z = \begin{cases} -z_j\bar{z}_k/2\|z\|^4 + \delta_{jk}/2\|z\|^2 & \text{if } j,k \leqq 2, \\ 0 & \text{otherwise,} \end{cases}$$

where δ_{jk} denotes the Kronecker symbol, and

$$\frac{\partial^2}{\partial z_j\partial\bar{z}_k}\varrho(\log\|z\|) = \delta_{jk}\varrho'(\log\|z\|)/2\|z\|^2 + (\varrho''(\log\|z\|) - 2\varrho'(\log\|z\|))\bar{z}_jz_k/4\|z\|^4,$$

if $j,k \leqq 2$; otherwise $\dfrac{\partial^2}{\partial z_j\partial\bar{z}_k}\varrho(\log\|z\|) = 0$.

Substituting $(z; z_3) = a(x)$ we obtain

$$\text{Hessian}(\log|z_2| - \log\|z\| + \varrho(\log\|z\|))$$

at

$$a(x) = (1/2)e^{-2x}\begin{bmatrix} \varrho'(x) - 1; & 0; & 0 \\ 0; & \varrho''(x)/2; & 0 \\ 0; & 0; & 0 \end{bmatrix}. \tag{A.9}$$

To obtain Hessian of φ at $a(x)$ we compute first that of $f\bar{f}$. To simplify our notation we call a z_1-term any monomial in which z_1 or $\bar{z}_1$ appears multiplicatively. Clearly, evaluation of any z_1-term at $a(x)$ gives 0. Differentiating (A.6)–(A.8) we get

$$\frac{\partial^2}{\partial z_1 \partial \bar{z}_1} f\bar{f} = (z,z)^{\pi 2} + z_1\text{-terms},$$

$$\frac{\partial^2}{\partial z_1 \partial \bar{z}_2} f\bar{f} = z_2 z_3 (z,z)^{-2} + z_1\text{-terms},$$

$$\frac{\partial^2}{\partial z_3 \partial \bar{z}_3} f\bar{f} = 1 + z_1\text{-terms},$$

while the derivatives,

$$\frac{\partial^2}{\partial z_1 \partial \bar{z}_3} f\bar{f}; \quad \frac{\partial^2}{\partial z_2 \partial \bar{z}_2} f\bar{f}, \text{ and } \frac{\partial^2}{\partial z_2 \partial \bar{z}_3} f\bar{f},$$

consist exclusively of z_1-terms. Substituting $(z; z_3) = a(x)$ we get

$$(\text{Hessian } f\bar{f})(a(x)) = \begin{bmatrix} \exp(-4x); & \exp(-3x+\varrho(x)); & 0 \\ \exp(-3x+\varrho(x)); & 0; & 0 \\ 0; & 0; & 1 \end{bmatrix}. \tag{A.10}$$

By (A.6)–(A.8), and since $f\bar{f}$ is real, we have

$$(\text{grad } f\bar{f})(a(x)) = (\exp\varrho(x)) \cdot (-e^{-2x}, 0, 1), \tag{A.11}$$

$$(\text{grad}' f\bar{f})(a(x)) = \left(\frac{\partial}{\partial \bar{z}_1} f\bar{f}, \frac{\partial}{\partial \bar{z}_2} f\bar{f}, \frac{\partial}{\partial \bar{z}_3} f\bar{f}\right) \quad \text{at} \quad a(x)$$

$$= (\exp\varrho(x)) \cdot (-e^{-2x}, 0, 1). \tag{A.12}$$

Using (A.4) and (A.11)–(A.13) we get

$$(\text{Hessian} \log|f|)(a(x))$$

$$= (1/2)\exp(-2\varrho(x) - 2x) \begin{bmatrix} 0; & \exp(-x+\varrho(x); & 1 \\ \exp(-x+\varrho(x)); & 0; & 0 \\ 1; & 0; & 0 \end{bmatrix}.$$

This, together with (A.9), gives

$$(\text{Hessian}\, \varphi)(a(x))$$

$$= (1/2)\exp(-2x - 2\varrho(x)) \begin{bmatrix} (\varrho'(x)-1)\exp(2\varrho(x); & -\exp(-x+\varrho(x)); & -1 \\ -\exp(-x+\varrho(x)); & (\tfrac{1}{2})\varrho''(x)\exp(2\varrho(x)); & 0 \\ -1; & 0; & 0 \end{bmatrix}. \tag{A.13}$$

Now we consider the Levi condition at $a(x)$. By Lemma 4.4(i) Eq. (4.5) implies that $w_3 = w_1 e^{-2x} + w_2 e^{-x} \varrho'(x) \exp \varrho(x)$. Substituting this and (A.13) to (4.4), we get a quadratic form in $w_1, w_2, \bar{w}_1, \bar{w}_2$. Using standard criteria we check that this form is positive definite if and only if (4.6) and (4.7) hold. Q.E.D.

Acknowledgements. The author is grateful to Professor Bernard Aupetit, who introduced him into this area and attracted his attention to many interesting problems, and to Professor John Wermer for encouragement. More than half of the paper was written during the author's stay at the Universities of Toronto and Guelph in 1979. The author wishes to thank the members of the respective departments especially Professors P. Rosenthal, C. Davis, J. Holbrook, and P. Fischer for their friendly hospitality.

References

1. Aupetit, B., Wermer, J.: Capacity and uniform algebras. J. Functional Analysis **28**, 386–400 (1978)
2. Aupetit, B.: Propriétés spectrales des algebres de Banach. Lecture Notes in Mathematics, Vol. 735. Berlin, Heidelberg, New York: Springer 1979
3. Basener, R.F.: A condition for analytic structure. Proc. Am. Math. Soc. **36**, 156–160 (1972)
4. Bungart, L.: On analytic fiber bundles. I. Topology **7**, 55–68 (1968)
5. Hille, E., Phillips, R.S.: Functional analysis and semi-groups. Am. Math. Soc. Colloquium Publications, Vol. XXXI, Providence 1957
6. Gamelin, T.W.: Uniform algebras. Englewood Cliffs: Prentice-Hall 1969
7. Gunning, R.C., Rossi, H.: Analytic functions of several complex variables. Englewood Cliffs: Prentice-Hall 1965
8. Kirchberg, E.: Banach algebras whose elements have at most countable spectra, Parts, I, II. Preprints P-36/79, P-43/79, Akademie der Wissenschaften der DDR, Zentralinstitut für Mathematik und Mechanik (to appear in Studia Math.)
9. Kumagai, D.: Subharmonic functions and uniform algebras. Proc. Am. Math. Soc. **78**, 23–29 (1980)
10. Nishino, T.: Sur les ensembles pseudoconcaves. J. Math. Kyoto Univ. **1–2**, 225–245 (1962)
11. Oka, K.: Note sur les familles de fonctions analytiques multiformes etc. J. Sci. Hiroshima Univ. Ser. A **4**, 93–98 (1934)
12. Senitchkin, V.N.: Subharmonicheskie Funktii i Analitichoskaya strukture v prostranstve maximalnich idealov ravnomernoi algebrii. Mat. Sb. **108**, 115–133 (1979)
13. Słodkowski, Z.: Characterization of spectral multifunctions. Notices Am. Math. Soc. Abstract 770-B 25 (1979)
14. Słodkowski, Z.: On subharmonicity of the capacity of the spectrum. Proc. Am. Math. Soc. **81**, 243–249 (1981)
15. Taylor, J.L.: A joint spectrum for several commuting operators. J. Functional Analysis **6**, 172–191 (1970)
16. Vesentini, E.: On the subharmonicity of the spectral radius. Boll. Un. Mat. Ital. **4**, 427–429 (1968)
17. Vladimirov, V.S.: Methods of the theory of functions of many complex variables. Cambridge, Mass.: M.I.T. Press 1966
18. Wermer, J.: Subharmonicity and hulls. Pacific J. Math. **58**, 283–290 (1975)
19. Wermer, J.: Capacity and uniform algebras. In: Proceedings of Symposia in Pure Mathematics, Vol. XXXV, Part I, 1979
20. Wermer, J.: Potential theory and function algebras. Expository paper submitted to Texas Tech. Math. Series, June 1980
21. Wermer, J.: Maximum modulus algebras and singularity sets. Proc. of the Royal Soc. of Edinburgh **86 A**. 327–331 (1980)
22. Yamaguchi, H.: Sur une uniformité des surfaces constantes d'une fonction entière de deux variables complexes. J. Math. Kyoto Univ. **13**, 417–433 (1973)
23. Saks, S., Zygmund, A.: Analytic functions, Second edition. Warszawa: PWN – Polish Scientific 1965
24. Aupetit, B.: Analytic multivalued functions in Banach algebras and uniform algebras. (to appear)

Received November 6, 1980

Math. Ann. 256, 387–390 (1981)

Mathematische Annalen
© Springer-Verlag 1981

A Generalization of Schwarz Lemma

Hajime Tsuji

Department of Mathematics, Tokyo Metropolitan University, Setagaya Tokyo 158, Japan

0. Introduction

Let D be the open unit disk in $\mathbb{C}$ and let $f: D \to D$ be a holomorphic mapping such that $f(0)=0$.

Then the classical Schwarz lemma states that

$$|f(z)| \leqq |z|$$

and

$$|f'(0)| \leqq 1 .$$

This lemma has been generalized in many ways [1, 4, 5]. And Wu considered a normal family of holomorphic mappings from a complex manifold into itself [4].

In this paper we generalize the classical Schwarz lemma for analytic self-mappings of hyperbolic manifolds and Stein manifolds.

1. A General Schwarz Lemma for Hyperbolic Manifolds

In this section we shall consider holomorphic self-mappings of hyperbolic manifolds. The definition and basic properties of hyperbolic manifolds can be seen in [1].

The following theorem is the main result of this section.

Theorem 1.1. *Let M be a connected hyperbolic manifold and let d be the Kobayashi metric of M. And let $\varphi: M \to M$ be a holomorphic mapping such that $\varphi(M) \subset\subset M$ (where $A \subset\subset B$ means that $\bar{A}$ is a compact subset of B). Let φ^k denote the k-th iteration of φ. Then*

(1) $N = \bigcap_{k=1}^{\infty} \varphi^k(M)$ *is a nonempty compact analytic subset of M.*

(2) *There exists a positive integer k_0 such that $\varphi^{k_0}|N = \mathrm{id}_N$.*

(3) *Let F be the normal bundle of* $\mathrm{Reg}N$. *Then $d\varphi^{k_0}$ acts naturally on F and all the eigenvalues of $d\varphi^{k_0}$ have absolute values less than* 1 *on each fibre of F.*

0025-5831/81/0256/0387/$01.00

(4) $\{\varphi^k\}_{k=1}^{\infty}$ *has a subsequence which converges uniformly on* M *to a holomorphic mapping* ψ *such that*

(i) $\psi(M)=N$,

(ii) $\psi|N=\mathrm{id}_N$.

Proof. We divide the proof of this theorem into several steps.

Step 1. $\{\varphi^k\}_{k=1}^{\infty}$ has a subsequence which converges uniformly on M with respect to the metric d.

Proof. Since $\{\varphi^k\}_{k=1}^{\infty}$ is equicontinuous sequence, we can find a subsequence $\{\varphi^{k_j}\}_{j=1}^{\infty}$ which converges uniformly on $\varphi(M)\subset\subset M$ by Ascoli's theorem and the assumption $\varphi(M)\subset\subset M$. It is easy to verify that $\{\varphi^{k_j+1}\}_{j=1}^{\infty}$ is a desired subsequence.

Step 2. N is a nonempty compact subset of M.

Proof. It is easy to verify that $\overline{\varphi^{k+1}(M)}\subset\varphi^k(\overline{\varphi(M)})\subset\varphi^k(M)$. So we obtain that $\bigcap_{k=1}^{\infty}\overline{\varphi^{k+1}(M)}\subset N\subset\bigcap_{k=1}^{\infty}\overline{\varphi^k(M)}$.

Then $N=\bigcap_{k=1}^{\infty}\overline{\varphi^k(M)}$. This implies that N is a nonempty compact subset of M.

For the next step we need the following lemma.

Lemma 1. *Let* (X,d) *be a compact metric space and let* $f:X\to X$ *be a surjective distance decreasing mapping. Then* f *is an isometry.*

Proof. Let f^k be the k-th iteration of f. Using Ascoli's theorem and equicontinuity of f^k, we can find a subsequence $\{f^{k_j}\}_{j=1}^{\infty}$ which converges uniformly on X. Let h be $\lim_{j\to\infty} f^{k_j}$. We claim that h is surjective. Let q be an arbitrary point of X and choose points p_j such that $f^{k_j}(p_j)=q$. Since X is compact, chosing a subsequence if necessary we may assume $\{p_j\}_{j=1}^{\infty}$ converges to a point $p\in X$. Since $\{f^{k_j}\}_{j=1}^{\infty}$ converges uniformly on X, we conclude that $h(p)=q$. Hence h is surjective. So to prove this lemma it is sufficient to verify that $d(h(p),h(q))=d(f(h(p)),f(h(q))$ holds for all $p, q\in X$. Choose arbitrary two points $p, q\in X$. Since $\{d(f^k(p),f^k(q))\}_{k=1}^{\infty}$ is monotone decreasing, we can find the following equality,

$$d(f(h(p)),f(h(q)))=\lim_{j\to\infty} d(f^{k_j+1}(p),f^{k_j+1}(q))$$

$$=\lim_{k\to\infty} d(f^k(p),f^k(q)).$$

Similarly we obtain the equality,

$$d(h(p),h(q))=\lim_{k\to\infty} d(f^k(p),f^k(q)).$$

Then we have $d(h(p),h(q))=d(f(h(p)),f(h(q)))$ and the lemma is proved.

Step 3. $\varphi|N$ is an isometry of N.

Proof. It is trivial that $\varphi(N)\subset N$. Since φ is distance decreasing mapping, it is sufficient to prove that $\varphi|N$ is surjective. Choose an arbitrary point q of N and let p_j be the point of $\varphi^j(M)$ such that $\varphi(p_j)=q$. Since all p_j are contained in the

compact set $\overline{\varphi(M)}$, we can find a subsequence $\{p_{j_k}\}_{k=1}^{\infty}$ which converges to a point p. By the compactness of N it is easy to verify that $p \in N$. It is trivial to see that $\varphi(p)=q$.

Let $I(N)$ be the isometry group of N. It is well-known that $I(N)$ is a compact topological group with respect to the compact-open topology (see [1, p. 70]). For the next step we need an easy lemma.

Lemma 2. *Let G be a compact topological group and let a be an element of G. Then we can find a subsequence of $\{a^j\}_{j=1}^{\infty}$ which converges to the identity element e of G.*

Proof. It is sufficient to prove that e is a cluster point of $\{a^j\}_{j=1}^{\infty}$. Suppose not. Then since $\{a^j\}_{j=1}^{\infty}$ is relatively compact in G, we can find finite positive integers $j_1, j_2, \ldots, j_p$ and e's neighbourhood V such that $\{a^j\}_{j=1}^{\infty} \subset a^{j_1}V \cup a^{j_2}V \cup \ldots \cup a^{j_p}V$ and $V \cap \{a^j\}_{j=1}^{\infty} = \emptyset$. By the definition of V this implies that $\{a^j\}_{j=1}^{\infty}$ consists of finite points. Then we can find a positive integer k_0 such that $a^{k_0}=e$. This is the contradiction.

Step 4. N is a compact analytic subset of M.

Proof. Let a be the element of $I(N)$ defined by $\varphi|N$. Then by Lemma 2 we can find a subsequence $\{a^{j_k}\}_{j=1}^{\infty}$ of $\{a^j\}_{j=1}^{\infty}$ which converges to the identity element e. Taking a subsequence as in Step 1 we may assume that $\{\varphi^{j_k}\}_{k=1}^{\infty}$ converges uniformly on every compact subset of M. Let ψ^* denote $\lim_{k\to\infty} \varphi^{j_k}$. It is easy to verify that $\psi^*(M)=N$. Then N is the fixed points set of the holomorphic mapping ψ^*. This implies that N is a compact analytic subset of M.

Step 5. There exists a positive integer k_0 such that

$$\varphi^{k_0}|N = \mathrm{id}_N$$

Proof. Since $\varphi|N \in \mathrm{Aut}(N)$, it is sufficient to observe that $\mathrm{Aut}(N)$ is a finite group. Finiteness of $\mathrm{Aut}(N)$ follows from the fact that N is a compact hyperbolic complex space [2].

Now we return to the proof of Theorem 1.1.

(1) and (2) have already proved. We can find a subsequence of $\{\varphi^{mk_0}\}_{m=1}^{\infty}$ which converges uniformly on M in the same way as in step 1. The limit of this sequence ψ satisfies (4). (3) follows from (1), (2), and (4). Q.E.D.

2. A General Schwarz Lemma for Stein Manifolds

We consider analytic self-mappings of Stein manifolds in this section.

Theorem 2.1. *Let (M, g) be a connected Stein manifold with complete hermitian metric g. Let $\varphi: M \to M$ be a holomorphic mapping such that $\varphi(M) \subset\subset M$. Then there exists a point 0 of M which satisfies the following properties.*

(1) *0 is the unique fixed point of φ.*

(2) *All the eigenvalues of $d\varphi: TM_0 \to TM_0$ have absolute values less than 1.*

(3) *$\{\varphi^k\}_{k=1}^{\infty}$ converges uniformly to the constant mapping 0 (i.e. $0(p)=p$ for all $p \in M$).*

Proof. Since $\varphi(M) \subset\subset M$, we can find an open domain V such that $\varphi(M) \subset\subset V \subset\subset M$. It is easy to verify that V is a hyperbolic manifold. Now we

apply Theorem 1.1. to the holomorphic mapping, $\varphi|V: V \to V$. Then we obtain Theorem 2.1. Q.E.D.

As immediate consequences of the above theorem we obtain the following corollaries.

Corollary 2.2. *Let M be a connected Stein manifold. Then the following two conditions are equivalent.*

(1) *M is biholomorphic to a bounded domain in $\mathbb{C}^n (n = \dim_{\mathbb{C}} M)$.*

(2) *There exists an into biholomorphic mapping $\varphi: M \to M$ such that $\varphi(M) \subset\subset M$.*

Corollary 2.3. *Let M be a connected Stein manifold and $\varphi: M \to M$ be a holomorphic mapping such that $\varphi(M) \subset\subset M$. Then there exists a positive integer j_0 such that φ^{j_0} is homotopic to a constant mapping.*

Corollary 2.4. *Let D be a bounded domain with a connected smooth boundary in a Stein manifold. If there exists a holomorphic vector field Z defined on some neighbourhood of ∂D such that Z is transversal everywhere to the boundary, then D is biholomorphic to a bounded domain in $\mathbb{C}^n (n = \dim_{\mathbb{C}} M)$ and D is contractible.*

Proof. By the holomorphic extention theorem in [3] we can extend Z to a holomorphic vector field defined in a neighbourhood of $\bar{D}$. Considering the flow of Z or $-Z$, we can find an into biholomorphic mapping $\varphi: D \to D$ such that $\varphi(D) \subset\subset D$. By Corollary 2.2 we find that D is biholomorphic to a bounded domain in $\mathbb{C}^n (n = \dim_{\mathbb{C}} M)$. Since φ is homotopic to the id_D, Corollary 2.3. implies that D is contractible. Q.E.D.

Using Corollary 2.3, we can give a simple proof to the following classical result (see [1, p. 14]).

Corollary 2.5. *Let D_1, D_2 be two annuli, i.e. $D_j = \{z \in \mathbb{C} | r_j < |z| < 1/r_j\}$ for some $r_j < 1 (j = 1, 2)$, respectively. And let $f: D_1 \to D_2$ be a holomorphic mapping.*

Then we have $|\deg f| \leqq M_2/M_1$ where $M_j (j = 1, 2)$ are moduluses of $D_j (j = 1, 2)$, respectively.

Proof. Let m be $|\deg f|$ and let $D_2^* = \{z \in \mathbb{C} | r_2^{1/m} < |z| < 1/r_2^{1/m}\}$. Then there exists a holomorphic mapping $g: D_1 \to D_2^*$ such that $f = (g)^{\deg f}$. It is easy to verify g is a homotopy equivalence between D_1 and D_2^*. Then Corollary 2.3. implies $1 \leqq M_2^*/M_1$ where M_2^* is the modulus of D_2^* and this implies that $|\deg f| \leqq M_2/M_1$. Q.E.D.

Acknowledgement. The author would like to express his heartly thanks to Professor T. Ochiai and Professor I. Wakabayashi for their valuable comments.

References

1. Kobayashi, S.: Hyperbolic manifolds and holomorphic mappings. New York, Basel: Dekker 1970
2. Kobayashi, S.: Intrinsic distances, measures and geometric function theiry. Bull. Am. Math. Soc. **82**, 357–416 (1976)
3. Shiga, K.: On holomorphic extention from the boundary. Nagoya Math. J. **42**, 55–66 (1971)
4. Wu, H.: Normal families of holomorphic mappings. Acta Math. **119**, 193–233 (1967)
5. Yau, S.T.: A general Schwarz lemma for Kaehler manifolds. Am. J. Math. **100**, 197–203 (1978)

Received November 17, 1980

Math. Ann. 256, 391–399 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Factorization of a Birational Morphism Between 4-Folds*

Mina Teicher

Department of Mathematical Sciences, Tel-Aviv University, Ramat-Aviv, Tel-Aviv, Israel

The aim of this paper is to prove a certain result on birational morphisms which is a step toward proving, if possible, the following general conjecture of Hironaka: Every birational correspondence between nonsingular algebraic varieties factors into a finite product of inverse monoidal transformations with nonsingular centers followed by monoidal transformations with nonsingular centers, [4, Chap. 0, question (F'), p. 149], [10, p. 287].

We shall first recall what has already been proven. Let $f: X \to X'$ be a proper birational morphism between nonsingular algebraic schemes. If the exceptional divisor of f is irreducible and collapses to a nonsingular subvariety, Moishezon [6, p. 139, Theorem 1] proved in 1966 that f itself is a monoidal transformation with nonsingular center.

For schemes of dimension 2 Zariski [11, p. 46, Theorem II. 1.2] proved that f splits into a finite sequence of monoidal transformations, i.e. f is necessarily projective.

For higher dimensions this stronger result cannot hold since Hironaka [3] constructed an example of a nonprojective proper birational morphism between 3-folds.

In 1976 Schaps [8] considered the case where the exceptional divisor of f has 2 nonsingular components collapsing to a nonsingular subvariety, and proved that for $\dim X = 3$ f splits into a product of two monoidal transformations with nonsingular centers.

Preliminary cases that have already been proven deal with birational correspondences that are, in fact, morphisms. This is of interest since for varieties with ground field of characteristic zero any birational correspondence can be converted to a morphism by monoidal transformations with nonsingular centers. Here, too, we will assume that the correspondence is a morphism but we need place no restriction on the characteristic.

The main result of this paper is the following:

* This paper is part of the author's doctoral dissertation prepared at Tel-Aviv University under the supervision of Prof. I. Piatetski-Shapiro

0025-5831/81/0256/0391/$01.80

Theorem. *Let f be a proper birational morphism of nonsingular algebraic varieties of dimension 4. Assume that f is an isomorphism on $X-D$, $D=D_1\cup D_2$, $f(D)=D'$ where D_1, D_2, $D_1\cap D_2$, D' are all nonsingular irreducible varieties, $D_i\neq D_1\cap D_2$, $i=1,2$. Then f can be factored by two monoidal transformations with nonsingular centers.*

The existence of irreducible subvarieties of codimension strictly greater than 2 which are not points creates a whole range of new difficulties in the factorization of birational morphism of algebraic varieties of dim 4.

We blow up X' along D' to get $\phi:\bar{X}\to X'$ and $\bar{f}:X\to\bar{X}$ the induced birational correspondence. Let Δ be the set on which $\bar{f}^{-1}$ is not well defined. Δ can be of codimension 2, 3 or 4. The problem arises in $\operatorname{codim}_X\Delta=3$. Then we need new methods to show first, that $\bar{f}|_{D_1\cap D_2}$ is well defined, and, secondly, that in spite of the fact that $D_1\cap D_2$ collapses, the image, which equals Δ, is nonsingular and thus can be taken as the center of a second monoidal transformation.

1. Preliminaries

In this section we introduce some notations and cite or prove results concerning varieties of arbitrary dimension which will be used later.

Let $f:V\to W$ be a birational correspondence of algebraic varieties. There is a maximal open set in W on which f^{-1} is an isomorphism. Denote by $\underline{S_f(W)}$ the complementary reduced algebraic set and by $\underline{S_f(V)}$ the preimage of $\underline{S_f(W)}$.

Zariski's Main Theorem (ZMT) ([7, p. 415, Proposition 1], [9, p. 24], [2 III, 4.4 Vol. 11, p. 135]). *If $f:V\to W$ is a birational morphism of algebraic varieties with the base W nonsingular of dimension n, then $\underline{S_f(V)}$ is of pure dimension $n-1$ and if E is a component of $\underline{S_f(V)}$ then* $\dim \overline{f(E)}\leqq n-2$.

Let $f_i:V_i\to V$, $i=1,2$, be birational correspondences, $f_i|_{U_i}:U_i\xrightarrow{\sim}U$ isomorphisms of open dense sets. Then the closure in $V_1\underset{V}{\times}V_2$ of $U_1\underset{U}{\times}U_2$ will be called the *graph* of f_1 and f_2.

Lemma 1 [8]. *Let V, V_1, V_2 be nonsingular algebraic varieties of dimension n, $f_i:V_i\to V$ proper. Let $\Lambda\subset V_1\underset{V}{\times}V_2$ be the graph, $p_i:\Lambda\to V_i$ be the projections, which are proper, and $g=p_2p_1^{-1}$ the birational correspondence. If $g(x)$ is a single point for all x in V_1 then g is a morphism.*

Let A be a closed subscheme of the noetherian scheme V and I the coherent ideal sheaf of A on V. Then a pair (A,ϕ) consisting of the subscheme A and a morphism of schemes $\phi:V'\to V$ will be called the *monoidal transformation of V with center A* (or the blowing-up of V along A) if the morphism ϕ satisfies the following conditions:

1) The ideal sheaf $\phi^{-1}(I)\cdot O_{V'}$ on V' is locally everywhere generated by a single element of $O_{V'}$ which is not a zero-divisor.

2) If $g:V''\to V$ is any morphism of algebraic varieties having the property 1), then there exist a unique morphism of algebraic varieties $h:V''\to V'$ such that $g=\phi\circ h$.

This property is called the universal mapping property. This definition is adapted from the definition in Hironaka and Rossi [5, p. 315] for the complex case.

Let $f: V \to W$ be rational, $F \subset V$, $H \subset W$. Denote by $\underline{f[F]}$ the strict image of F i.e. $\overline{f(U)}$ for $U \subset F$ open dense subset of F on which f is well defined and by $\underline{f^{-1}[H]}$ the strict preimage of H, when f is birational.

We shall use a generalized Hironaka's test curve argument. It appears in $2l$ (local), $2g$ (global) and Lemma 3 in [8].

In the remaining part of this section we shall consider the following situation: $f: X \to X'$ is a proper birational morphism of nonsingular algebraic varieties of $\dim n$. $D' = S_f(X')$ and $D = S_f(X)$ are the subvarieties of "bad points" defined earlier, and suppose $D = D_1 \cup D_2$ where D', D_1, D_2, $D_1 \cap D_2$ are all nonsingular irreducible varieties, $D_1 \cap D_2 \subsetneqq D_i$ $i = 1, 2$. From now on we shall refer to such a map as a *birational morphism of nonsingular varieties with two collapsing components in normal position.* By ZMT $S_f(X)$ which equals $D_1 \cup D_2$ is of pure dimension $n-1$. Therefore $\dim D_1 = \dim D_2 = n-1$ and thus D_1 and D_2 are divisors. Let r denote $\dim D'$, and let $\phi: \bar{X} \to X'$ be the blowing up of X' along D', $\bar{f}: X \to \bar{X}$ the induced birational correspondence. (For definition see Lemma 1.) Let $\bar{D} = \phi^{-1}(D')$, and let Δ be the set on which $\bar{f}^{-1}$ is not well-defined. $l = \dim \Delta$.

The following 4 lemmas will be used directly in the proof of the main result.

Lemma 2. *Let f be a birational morphism of nonsingular algebraic varieties with two collapsing components in normal position. Let $z'_1, \ldots, z'_n$ be local coordinates in X' such that $z'_{r+1} = \ldots = z'_n = 0$ are local equations for D'. Let $\varepsilon' = dz'_1 \wedge \ldots \wedge dz'_n$ and $\varepsilon = f^*(\varepsilon')$. Then,* $\mathrm{Div}(\varepsilon) = (n-r-1)D_1 + (2n-r-l-2)D_2$.

Proof. Following the notation [1], for a morphism $h: Y \to Z$ of smooth algebraic varieties, we let $E_h = \dim(h^*(w_z) \otimes w_y^{-1})$.

Let $\bar{\phi}: \bar{\bar{X}} \to \bar{X} - (\mathrm{Sing}\, \Delta)$ be the blowing up of the smooth points of Δ, and let $\bar{\phi}^*(\bar{D}) = \bar{\Delta} + \bar{\bar{D}}$, with $\bar{\bar{D}} = \bar{\phi}^{-1}[\bar{D}]$. As shown in Lemma 3 of [8], $\bar{\bar{D}}$ and $\bar{\Delta}$ are generically isomorphic to D_1 and D_2. To establish the multiplicity of D_i in $\mathrm{Div}(\varepsilon)$ (or more precisely, of $D_i \cap f^{-1}(U')$, with U' a local coordinate neighborhood), it suffices to calculate $E_{\phi \circ \bar{\phi}}$. By the well known formula for the exceptional divisor of a blowing up, $E_\phi = (\mathrm{codim}\, D' - 1)\bar{D} = (n-r-1)\bar{D}$ and similarly $E_{\bar{\phi}} = (n-l-1)\bar{\Delta}$. Thus by (2.4) [1]

$$\begin{aligned} E_{\phi \circ \bar{\phi}} &= E_{\bar{\phi}} + \bar{\phi}^*(E_\phi) \\ &= (n-l-1)\bar{\Delta} + (n-r-1)(\bar{\Delta} + \bar{\bar{D}}) \\ &= (2n-l-r-2)\bar{\Delta} + (n-r-1)\bar{\bar{D}}. \end{aligned}$$

Lemma 3. *Let f be a rational morphism of nonsingular algebraic varieties, D the exceptional divisor $f(D) = D'$. Let $x \in D$, and let $z'_1, \ldots, z'_n$ be local coordinates in a neighborhood of $f(x)$ such that $z'_{r+1} = \ldots = z'_n = 0$ are local equations for D'. Denote $z_i = f^*(z'_i)$. There is a neighborhood of x in which $z_i = q_i t$ for $r < i$ where t is*

a local equation for D. If there exists a k such that $q_k(x) \neq 0$, *then* $\bar{f}$ *is well-defined in a neighborhood of x.*

Proof. By ZMT D is of pure dimension $n-1$ so there is always a local equation for D. Let t be the local equation. Then t divides z_i, for $i>r$, since z_i, for $i>r$ vanishes on D.

First, suppose there exists a k such that $q_k(x) \neq 0$. Then q_k doesn't vanish in a neighborhood of x, U_x. In that neighborhood $z_i = z_k \frac{q_i}{q_k}, r<i$. Thus, in U_x the lifting of the ideal of D' which equals $\langle z_{r+1}, \ldots, z_n \rangle$ is generated by a single element. Therefore by the universal mapping property of blowing-up there exists a $g: U_x \to \bar{X}$ such that $\phi g = f|_{U_x}$. Since $\bar{f}$ is a birational map with the largest domain of definition from all g's that satisfy $\phi g = f$ we get the result that $\bar{f}$ is well-defined in U_x, i.e., $\bar{f}$ is well-defined in a neighborhood of x.

Lemma 4. *Let f be a birational morphism of nonsingular algebraic varieties with two collapsing components in normal positions. Then* $\bar{f}$ *is well-defined in a neighborhood of* $D_1 - D_2$. *Furthermore,* $D_1 - D_2$ *and* $\bar{D} - \Delta$ *are isomorphic via* $\bar{f}$.

Proof. Let $x_0 \in D_1 - D_2$. $D_1 - D_2$ is nonsingular, so there exist local coordinates centered at x_0 of the form, $v_1, \ldots, v_{n-1}, t$, where t is a local equation for D. Let $z'_1, \ldots, z'_n$ be local coordinates centered at $f(x_0)$, such that $z'_{r+1} = , \ldots, = z'_n = 0$ are local equations for D'. Let $z_i = f^*(z'_i)$.

Let q_i $r<i$ satisfy $z_i = q_i t$. We want to prove that there is a k such that $q_k(x_0) \neq 0$ and then using Lemma 3 we will get that $\bar{f}$ is well-defined in a neighborhood of x_0. Since x_0 was an arbitrary point of $D_1 - D_2$, we will get the result, that $\bar{f}$ is well-defined in a neighborhood of $D_1 - D_2$.

Let us denote the matrix $\frac{\partial z_1, \ldots, z_n}{\partial v_1 \ldots v_{n-1} t}$ by M. Then $\varepsilon = \det M \, dv_1 \wedge \ldots \wedge dv_{n-1} \wedge dt$. From Lemma 2, $t^{n-r-1} | \det M$ and if $x_0 \in D_1 - D_2$, $\frac{\det M}{t^{n-r-1}}$ does not vanish in x_0.

$$M = \begin{bmatrix} \frac{\partial z_1}{\partial v_1} & \cdots & \frac{\partial z_1}{\partial v_{n-1}} & \frac{\partial z_1}{\partial t} \\ \vdots & & \vdots & \vdots \\ \frac{\partial z_r}{\partial v_1} & \cdots & \frac{\partial z_r}{\partial v_{n-1}} & \frac{\partial z_r}{\partial t} \\ t\frac{\partial q_{r+1}}{\partial v_1} & \cdots & t\frac{\partial q_{r+1}}{\partial v_{n-1}} & q_{r+1} + t\frac{\partial q_{r+1}}{\partial t} \\ \vdots & & \vdots & \vdots \\ t\frac{\partial q_n}{\partial v_1} & \cdots & t\frac{\partial q_n}{\partial v_{n-1}} & q_n + t\frac{\partial q_n}{\partial t} \end{bmatrix}.$$

So we calculate and get, for suitable determinants $A_{n,i}$ and $A_{n,r}$, that $\det M = t^{n-r-1}\left(t A_{n,r} + \sum_{i=r+1}^{n} q_i A_{n,i}\right)$. If $q_i(x_0)$ equals zero for every i then $\frac{\det M}{t^{n-r-1}}$ van-

ishes at x_0, contradiction. Hence there exists a k such that $q_k(x_0) \neq 0$. As we explained before this completes the proof.

For the second claim of the lemma we recall that by Lemma 2g [8] $\bar{f}$ is an isomorphism almost everywhere on D_1 so $D_1 - D_2$ does not collapse. $\bar{f}$ is well-defined in $X - D_2$ and the only divisor that might collapse there is $D_1 - D_2$ (since outside of D $\bar{f}$ is an isomorphism). So by ZMT $\bar{f}$ is an isomorphism in $X - D_2$. Since $\bar{f}^{-1}$ is not well-defined in Δ, $\bar{f}(X - D_2) \subset \bar{X} - \Delta$. In particular $D_1 - D_2$ is isomorphic to its image by $\bar{f}$ and $\bar{f}(D_1 - D_2) \subset (\bar{X} - \Delta) \cap \bar{D} = \bar{D} - \Delta$. By [8] Lemma 2g $\bar{f}^{-1}(\bar{D} - \Delta) \subset D_1 - D_2$. So $\bar{f}(D_1 - D_2) = \bar{D} - \Delta$ and $D_1 - D_2 \simeq \bar{D} - \Delta$ via $\bar{f}$.

Lemma 5. *Let $f: X \to X'$ be a proper birational morphism of nonsingular algebraic varieties with two collapsing components in normal position. Assume that $\bar{f}$ is well defined in a neighborhood of D_1 such that $D_1 - D_2 \simeq \bar{D} - \Delta$ and Δ is nonsingular. Then $\bar{f}: X \to \bar{X}$ is a well-defined monoidal transformation with nonsingular center.*

Proof. Let $U \supset D_1$ be the neighborhood on which $\bar{f}$ is well-defined. We blow up $\bar{X}$ along Δ to get $\bar{\bar{\phi}}: \bar{\bar{X}} \to \bar{X}$, $\bar{\bar{f}}$ the induced birational correspondence. Let $\bar{\bar{D}} = \bar{\bar{\phi}}^{-1}[D]$ and $\bar{\bar{\Delta}} = \bar{\bar{\phi}}^{-1}(\Delta)$. Δ is nonsingular so by Lemma 3 [8], $S_{\bar{\bar{f}}}(\bar{\bar{X}}) \subset \bar{\bar{D}} \cap \bar{\bar{\Delta}}$. Let $\bar{\bar{x}} \in \bar{\bar{D}} \cap \bar{\bar{\Delta}}$, $\bar{\bar{x}} = \bar{\bar{\phi}}(\bar{\bar{x}})$ and $\bar{\bar{\Gamma}} \subset \bar{\bar{D}}$ be a nonsingular curve through $\bar{\bar{x}}$ in general position transversal to $\bar{\bar{\Delta}}$. Let $\bar{\Gamma} = \bar{\bar{\phi}}(\bar{\bar{\Gamma}})$. Since $\bar{f}(D_1) = \bar{D}$ and $\bar{f}|_{D_1}: D_1 \to \bar{D}$ is proper there exists a point $x \in \bar{f}^{-1}[\bar{\Gamma}] \cap \bar{f}^{-1}(\bar{x}) \cap D_1$. x then belongs to $\bar{\bar{f}}^{-1}[\bar{\bar{\Gamma}}] \cap \bar{\bar{f}}^{-1}(\bar{\bar{x}}) \cap D_1$. $\bar{x} \in \Delta$ so $x \in D_2$, hence $x \in D_2 \cap U$, where $D_2 \cap U$ is an irreducible divisor of U which is included in $(\bar{f}|_U)^{-1}(\Delta)$. We are, therefore, in the situation of Lemma 2 l [8], for $\bar{f}|_U$, so $\bar{\bar{f}}$ is well-defined in a neighborhood U_1 of x and $\bar{\bar{f}}(x) = \bar{\bar{x}}$. By Lemma 2g, 3 [8], $\bar{\bar{f}}$ is an isomorphism almost everywhere on $D_1 \cup D_2$. Therefore no divisor through x can collapse and by ZMT for $\bar{\bar{f}}|_{U_1}$, $\bar{\bar{x}} \notin S_{\bar{\bar{f}}}(\bar{\bar{X}})$. Hence $S_{\bar{\bar{f}}}(\bar{\bar{X}}) = \phi$, thus $\bar{\bar{f}}(x)$ is an single point for all $x \in X$. So by Lemma 1 $\bar{\bar{f}}$ is well-defined everywhere. Furthermore, $\bar{\bar{f}}$ is an isomorphism.

Finally, since $\bar{\bar{f}}$ is well defined, so is $\bar{f}$ and since $\bar{\bar{f}}$ is an isomorphism $\bar{f}$ is a monoidal transformation with nonsingular center.

2. The Main Result

Theorem. *Let $f: X \to X'$ be a proper birational morphism of nonsingular algebraic varieties of dimension 4. Assume that f is an isomorphism on $X - D$, $D = D_1 \cup D_2$, $f(D) = D'$ where D_1, D_2, $D_1 \cap D_2$, D' are all nonsingular irreducible varieties, $D_i \neq D_1 \cap D_2$, $i = 1, 2$. Then f can be factored by two monoidal transformations with nonsingular centers.*

Proof. By ZMT $S_f(X)$ which equals $D_1 \cup D_2$ is of pure dimension 4-1 therefore $\dim D_1 = \dim D_2 = 3$. D' is connected and f is proper so $f^{-1}(D')$ is connected thus $D_1 \cap D_2$ is not empty. From codimension considerations and the fact that $D_1 \cap D_2 \subsetneqq D_2$ we get that $\dim D_1 \cap D_2 = 2$.

By ZMT $\operatorname{Codim} S_f(X') \geqq 2$ thus $\dim D' \leqq 4 - 2 = 2$ hence $r = \dim D' = 0, 1, 2$.

We blow up X' along D' and get $\phi:\bar{X}\to X'$. Let $\bar{f}:X\to\bar{X}$ be the induced birational correspondence ($\bar{f}=p_2 p_1^{-1}$ for $p_1:\Delta\to X$, $p_2:\Delta\to\bar{X}$ and $\Delta\subset X\underset{X'}{\times}\bar{X}$ the graph). Let $\bar{D}=\phi^{-1}(D')$ and Δ be the set on which $\bar{f}^{-1}$ is not well-defined.

From Lemmas 2g, 3 [8], we get, renumbering if necessary, $D_1\not\subset S_{\bar{f}}(X)$, $\bar{f}[D_1]=\bar{D}$ and $\Delta=S_{\bar{f}}(\bar{X})=\bar{f}[D_2]$ irreducible. $D_1\not\subset S_{\bar{f}}(X)$ so $\Delta\subsetneqq\bar{D}$, but $\bar{D}$ is irreducible of dimension 3 so $l=\dim\Delta=0,1,2$. We shall first deal with the easy case where $l=0$, i.e. Δ is a point. The complication arises when $l=1$ i.e. Δ is a curve.

2.1. Δ Is a Point

Since outside of $\Delta\,\bar{f}^{-1}$ is an isomorphism, the fact that Δ is a point implies immediately that $\bar{f}(x)$ is a single point for all $x\in X$ and then by Lemma 1, without using test curve methods, we get that $\bar{f}$ is a well defined morphism.

Δ is a point, hence a nonsingular subvariety. The conditions of Lemma 5 are fulfilled, so $\bar{f}$ is a monoidal transformation with nonsingular center, $f=\phi\bar{f}$ and the theorem is proved for this case.

2.2. Δ Is not a Point

Hence l equals 1 or 2.

By Lemma 4, proved with differential arguments and test curve methods, we get that $\bar{f}$ is well-defined in a neighborhood of D_1-D_2, and $D_1-D_2\simeq\bar{D}-\Delta$ via $\bar{f}$.

We shall first prove, that $\bar{f}$ is well-defined in a neighborhood of $D_1\cap D_2$, and then we shall prove that Δ is nonsingular. From Lemma 5 we then conclude, that $\bar{f}$ is well-defined on the whole of X, and it is a monoidal transformation with nonsingular center. $f=\phi\bar{f}$, and the theorem is proved.

We shall prove now that $\bar{f}$ is well defined in a neighborhood of $D_1\cap D_2$.

Let $x_0\in D_1\cap D_2$. Assume to the contrary that $\bar{f}$ is not well-defined at a neighborhood of x_0. Let $m=m_{X,x_0}$ be the maximal ideal of the local ring at x_0. $D_1\cap D_2$ is nonsingular, thus there exist local coordinates centered at x_0: v_1,v_2,t_1,t_2, where t_i is a local equation for D_i. Let $z'_1,\ldots,z'_4$ be local coordinates centered at $f(x_0)$ such that $z'_{r+1}=\ldots=z'_4=0$ are local equations for D'. $z_i=f^*(z'_i)$. Let q_i, $r<i$, satisfy $z_i=q_i t$, $t=t_1t_2$. Since by assumption $\bar{f}$ is not well-defined at x_0 we have by Lemma 3 $q_{r+1}(x_0)=\ldots=q_4(x_0)=0$, so $q_{r+1},\ldots,q_4\in m$.

Let $\varepsilon'=dz'_1\wedge\ldots\wedge dz'_4$ and $\varepsilon=f^*(\varepsilon')$, so $\varepsilon=dz_1\wedge\ldots\wedge dz_4$. Let

$$M=\frac{\partial(z_1,\ldots,z_4)}{\partial(v_1,v_2,t_1,t_2)}$$

then $\varepsilon=\det M\,dv_1\wedge dv_2\wedge dt_1\,dt_2$. By Lemma 2 $\mathrm{Div}(\varepsilon)=(3-r)D_1+(6-r-l)D_2$, hence, $\det M=t_1^{3-r}t_2^{6-r-l}\,\alpha$ where $\alpha\notin m$. Since, t_1,t_2 are local parameters, they belong to $m-m^2$, so $\det M\in m^{9-2r-l}-m^{10-2r-l}$.

On the other hand,

$$M=\begin{bmatrix} \frac{\partial z_1}{\partial v_1} & \frac{\partial z_1}{\partial v_2} & \frac{\partial z_1}{\partial t_1} & \frac{\partial z_1}{\partial t_2} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial z_r}{\partial v_1} & \frac{\partial z_r}{\partial v_2} & \frac{\partial z_r}{\partial t_1} & \frac{\partial z_r}{\partial t_2} \\ t_1 t_2 \frac{\partial q_{r+1}}{\partial v_1} & t_1 t_2 \frac{\partial q_{r+1}}{\partial v_2} & t_2 q_{r+1}+t_1 t_2 \frac{\partial q_{r+1}}{\partial t_1} & t_1 q_{r+1}+t_1 t_2 \frac{\partial q_{r+1}}{\partial t_2} \\ \vdots & \vdots & \vdots & \vdots \\ t_1 t_2 \frac{\partial q_4}{\partial v_1} & t_1 t_2 \frac{\partial q_4}{\partial v_2} & t_2 q_4+t_1 t_2 \frac{\partial q_4}{\partial t_1} & t_1 q_4+t_1 t_2 \frac{\partial q_4}{\partial t_2} \end{bmatrix}.$$

Consider the last $4-r$ rows of the matrix M. $t_1, t_2 \in m$, $q_{r+1}, \ldots, q_{in} \in m$ (by Lemma 3). So, every coefficient of the last $4-r$ rows belong to m^2, and thus every minor of the last $4-r$ rows belongs to m^{8-2r}. In particular, $\det M \in m^{8-2r}$.

For $l=2$ we get an immediate contradiction, since here, $10-2r-l=8-2r$ and $\det M \notin m^{10-2r-l}$.

For $l=1$ we treat separately the case where $r=0$ or 1 and the case where $r=2$.

Case 1. $r=0$ *or* 1

Here, $10-2r-l=9-2r$; so $\det M \notin m^{9-2r}$.

Since $q_i \in m$, for $i>r$, we can write:

$$q_i=v_i \frac{\partial q_i}{\partial v_1}+v_2 \frac{\partial q_i}{\partial v_2}+t_1 \frac{\partial q_i}{\partial t_1}+t_2 \frac{\partial q_i}{\partial t_2}+m_i, \qquad m_i \in m^2, \ i>r.$$

Substituting that in M, we see that modulo m^3 the columns of the last $4-r$ rows of M are linear combinations of the columns of $\frac{\partial(q_{r+1}, \ldots, q_4)}{\partial(v_1, v_2, t_1, t_2)}$ with coefficients in m^2.

Since $\det M \notin m^{9-2r}$ there exists a minor of the last $4-r$ rows of M which does not belong to m^{9-2r}. So, there exists a minor of the matrix $\frac{\partial(q_{r+1}, \ldots, q_4)}{\partial(v_1, v_2, t_1, t_2)}$ which does not belong to m, and therefore the rank of this matrix is $4-r$. Thus, $q_{r+1}, \ldots, q_4$ form a part of a system of local parameters at x_0. So $V(q_{r+1}, \ldots, q_4)$ is of $\dim r$.

Let A be a component of the set on which $\bar{f}$ is not well-defined, such that $x_0 \in A$. $A \subset D$, so by Lemma 3 the intersection of A with the open set on which q_i are defined is included in $V(q_{r+1}, \ldots, q_4)$. $V(q_{r+1}, \ldots, q_4)$ is of dimension r, so $\dim A \leqq r$. Here $r=0$ or 1, so $\dim A \leqq 1$.

Recall that $\bar{f}=p_2 p_1^{-1}$ for $\Lambda \subset X \underset{X'}{\times} \bar{X}$ the graph and $p_1: \Lambda \to X$, $p_2: \Lambda \to \bar{X}$ the projections. The set on which $\bar{f}$ is not well-defined is the set on which p_1^{-1} is

not an isomorphism. So A is an irreducible component of $S_{p_1}(X)$. By ZMT there exists an irreducible divisor Z in Δ which collapses by p_1 to A. Since outside of Δ $\bar{f}^{-1}$ is an isomorphism, $p_2(Z) \subset \Delta$. So $Z \subset A \underset{X'}{\times} \Delta$ where $\dim A$, $\dim \Delta \leqq 1$. From dimension considerations we get a contradiction.

Remark. We can get a contradiction here in a different way.

Let $g: Y \to X$ be the blowing up of X along A. Let P be the fibre of g above x_0. For every point p of P there is a j such that q_j, q_i/q_j, q_k/q_j induces local coordinates at p. Therefore the lifting of the ideal of D' by fg which equals $\langle g^*(t\,q_2), g^*(t\,q_3), g^*(t\,q_4)\rangle$ is locally everywhere generated by a single element on P. By the universal mapping property there is a well-defined morphism ψ from a neighborhood $\hat{U}$ of P to $\bar{X}$, such that $fg = \phi\psi$. ψ is well-defined on the open subset $Y - g^{-1}(D)$ and coincides with $\bar{f}g$ there, so it coincides with $\bar{f}g$ where ever both are well-defined. Denote $E = g^{-1}(D_2)$, then $Y - E \xrightarrow{\sim}_{g} X - D_2 \xrightarrow{\sim}_{\bar{f}} \bar{X} - \Delta$, so $\psi(E \cap \hat{U}) \subset \Delta$, so $\psi(P) \subseteq \Delta$.

Now let $\bar{P} = \phi^{-1}(f(x_0))$. For every $\bar{p}$ in $\bar{P}$ there is a j such that z'_j, z'_i/z'_j, z'_k/z'_j induces local coordinates at $\bar{p}$. Since the same is true for q_2, q_3, q_4 on P and ψ is an isomorphism almost everywhere, $\psi(P)$ must equal $\bar{P}$. But Δ is a curve and $\bar{P} \simeq P^2$ so we got a contradiction.

Case 2. $r = 2$

Here, $10 - 2r - l = 10 - 4 - 1 = 5$, so $\det M \notin m^5$.

For this case we shall prove that $f(p_2) = f(x_0) = p\,t$. (It is true in general for Δ a curve.)

Let x be a point of D_2 on which $\bar{f}$ is not well-defined, then $\bar{f}(x) = p_2 p_1^{-1}(x)$ is not a single point. p_1 is regular and p_2 is closed hence $p_2 p_1^{-1}(x)$ is closed. p_1 is proper and p_2 is regular hence $p_2 p_1^{-1}(x)$ is connected. $p_2 p_1^{-1}(x)$ is included in Δ since outside of Δ $\bar{f}^{-1}$ is an isomorphism. We are in the case where $\dim \Delta = 1$. Since $\Delta = \bar{f}[D_2]$, it is irreducible. Combining all these facts, we get that $p_2 p_1^{-1}(x) = \Delta$, i.e. $\bar{f}(x) = \Delta$. Now, $\phi(\Delta) = \phi\,\bar{f}(x) = f(x)$, so $\phi(\Delta)$ is a single point, and since $\bar{f}$ is not well-defined at x_0, this single point is actually $f(x_0)$. Furthermore, the image under f of each point x of D_2 on which $\bar{f}$ is not well-defined equals that single point, namely, $f(x_0)$. For x, a point of D_2 on which $\bar{f}$ is well-defined $\bar{f}(x) \subset \bar{f}[D_2] = \Delta$, so we have $f(x) = \phi\bar{f}(x) \subset \phi(\Delta) = f(x_0)$. Therefore we get $f(D_2) = f(x_0)$.

Now, for z'_i the local coordinates centered at $f(x_0)$ we have $z'_i(f(x_0)) = 0$ for $i \leqq r$. Hence $z_i(D_2) = f^*(z'_i)(D_2) = z'_i(f(D_2)) = z'_i(f(x_0)) = 0$ for $i \leqq r$. Therefore z_i for $i \leqq r$ vanishes on D_2 so $t_2 | z_i$ for $i \leqq r$. Here, $r = 2$ and let $z_i = t_2 q_i$ for $i \leqq 2$.

So

$$\frac{\partial(z_1, z_2)}{\partial(v_1, v_2, t_1, t_2)} = \begin{bmatrix} t_2\dfrac{\partial q_1}{\partial v_1} & t_2\dfrac{\partial q_1}{\partial v_2} & t_2\dfrac{\partial q_1}{\partial t_1} & q_1 + t_2\dfrac{\partial q_1}{\partial t_2} \\ t_2\dfrac{\partial q_2}{\partial v_1} & t_2\dfrac{\partial q_2}{\partial v_2} & t_2\dfrac{\partial q_2}{\partial t_1} & q_2 + t_2\dfrac{\partial q_2}{\partial t_2} \end{bmatrix}.$$

Since $t_2 \in m$ every minor of order 2 of the matrix $\dfrac{\partial(z_1, z_2)}{\partial(v_1, v_2, t_1, t_2)}$ belongs to m. The rows of this matrix are the first two rows of M. For the last $4 - r$ rows of

M we showed that every minor belongs to m^{8-2r}. Here, $r=2$, so every minor of the last 2 rows of M belongs to m^4. So $\det M \in m^5$ contradiction.

Since our assumption that $\bar{f}$ is not well-defined at a neighborhood of x_0 led to a contradiction we get that $\bar{f}$ is well-defined in a neighborhood of x_0. x_0 was an arbitrary point of $D_1 \cap D_2$ so $\bar{f}$ is well-defined in a neighborhood of $D_1 \cap D_2$.

We shall now prove that Δ is nonsingular.

Combining what we have just proved with Lemma 4 we get that $\bar{f}$ is well defined in a neighborhood of D_1. A priori by the construction of the birational correspondence $\bar{f}(D_1) \subset \bar{D}$. $\bar{D}$ equals $\bar{f}[D_1]$ (by Lemma 2g [8]), which is included in $\bar{f}(D_1)$ so $\bar{f}(D_1)=\bar{D}$. $D_1 - D_2$ is isomorphic to $\bar{D}-\Delta$ via $\bar{f}$ (by Lemma 4), so $\bar{f}(D_1 \cap D_2)$ must equal Δ. $\bar{f}|_{D_1}: D_1 \to \bar{D}$ is proper since f is proper. It is birational since $D_1 - D_2 \simeq \bar{D} - \Delta$.

For $l=2$ $\bar{f}|_{D_1}: D_1 \to \bar{D}$ is a birational morphism with $D_1 - D_2 \simeq \bar{D} - \Delta$ and $\bar{f}(D_1 \cap D_2) = \Delta$ where $D_1 \cap D_2$ and Δ are of the same dimension, so no divisor in D_1 can collapse. By ZMT $\bar{f}|_{D_1}: D_1 \to \bar{D}$ is an isomorphism and thus $\Delta \simeq D_1 \cap D_2$. $D_1 \cap D_2$ is nonsingular, therefore Δ is nonsingular.

For $l=2$ $\bar{f}|_{D_1}: D_1 \to \bar{D}$ collapses the surface $D_1 \cap D_2$ to the curve Δ. We wish to show that Δ is nonsingular. Since this is a local question we can pass to the Henselization of $\bar{D}$ at a point $\bar{x}$ of Δ. Since $\bar{f}|_{D_1}^{-1}(\bar{x}) \leqq 1$, we apply Theorem 3.1 of Danilov [1] to obtain that Δ is nonsingular at $\bar{x}$.

As we explained in the begining of 2.2 the theorem is proved.

Acknowledgements. I would like to express my gratitude to Prof. I. Piatetski-Shapiro, Prof. H. Hironaka, Dr. M. Schaps, and Dr. O. Gaber whose assistance enabled me to accomplish this paper.

References

1. Danilov, V.I.: Factorization of birational morphisms, (Russian) Izv. Akad. Nauk. **44**, 2 (1980)
2. Grothendieck, A., Dieudonne, J.: Elements de geométrie algebrique. Publ. Math. I.H.E.S. **4**, **8**, **11**, **17**, **20**, **24**, **28**, **32** (1960-1967)
3. Hironaka, H.: On the theory of birational blowing-up. Thesis, Harvard University, 1960
4. Hironaka, H.: Resolutions of singularities of an algebraic variety over a field of characteristic zero. Ann. Math. **79**, 109-326 (1964)
5. Hironaka, H., Rossi, H.: On the equivalence of imbeddings of exceptional complex spaces. Math. Ann. **156**, 313-333 (1964)
6. Moishezon, B.: On n-dimensional compact complex varieties with n algebraically independent meromorphic functions. Am. Math. Soc. Transl. Ser. 2, **63**, 51-177 (1967)
7. Mumford, D.: Introduction to algebraic geometry. Harvard University
8. Schaps, M.: Birational morphism factorizable by two monoidal transformations. Math. Ann. **222**, 23-28 (1976)
9. Shafarevitch, I.: Lectures on minimal models and birational transformation of two-dimensional scheme. Bombay: Tata Institute of Foundamental Research 1966
10. Shafarevitch, I.: Basic algebraic geometry. Berlin, Heidelberg, New York: Springer 1974
11. Zariski, O.: Introduction to the problem of minimal models in the theory of algebraic surfaces. Publ. Math. Soc. Japan 1958

Received March 17, 1980; in revised form February 10, 1981

Math. Ann. 256, 401–418 (1981)

Mathematische Annalen
© Springer-Verlag 1981

The Geometry of Representations of A_m

S. Abeasis,[1,*] A. Del Fra,[1,*] and H. Kraft[2,**]

[1] Istituto Matematico "Guido Castelnuovo", Università degli Studi di Roma, I-00100 Roma, Italy
[2] Sonderforschungsbereich Theoretische Mathematik, Mathematisches Institut der Universität, Beringstrasse 4, D-5300 Bonn 1, Federal Republic of Germany

Introduction

0.1. Let $V_1, V_2, \ldots, V_m$ be finite dimensional vector spaces over some algebraically closed field k of characteristic zero. Consider the space

$$L := \mathrm{Hom}(V_1, V_2) \times \mathrm{Hom}(V_2, V_3) \times \ldots \times \mathrm{Hom}(V_{m-1}, V_m)$$

whose elements

$$A: V_1 \xrightarrow{A_1} V_2 \xrightarrow{A_2} V_3 \longrightarrow \ldots \longrightarrow V_{m-1} \xrightarrow{A_{m-1}} V_m$$

can be viewed as representations of a certain fixed dimension of the Dynkin quiver A_m (oriented in the obvious way; cf. [2]). The group

$$G := \mathrm{GL}(V_1) \times \mathrm{GL}(V_2) \times \ldots \times \mathrm{GL}(V_m)$$

acts in a natural way on L and the orbits are just the equivalence classes of these representations: in particular there are only a finite number of orbits in L. It is an interesting task to describe the *degenerations* of these orbits and to determine the *singularities* in the closure $\bar{O}$ of a given orbit O. The first problem has been solved in [1] (cf. also 2.2 and 9.1). If $A \in L$ is a *complex* (i.e. $A_{i+1}A_i = 0$ for all i), Kempf has shown in [8] that the closure $\bar{O}_A$ of its orbit O_A has *rational singularities* (cf. 1.4 for definitions). We will prove this result in general.

0.2. Theorem. *For any $A \in L$ the closure $\bar{O}_A$ of its orbit is normal, Cohen-Macaulay with rational singularities.*

Our method is similar to the one developed in [13] and is based on the first fundamental theorem of invariant theory (1.2) (compare also with [10]). To the

* Both authors belong to the group GNSAGA of CNR
** This work has been done during the last author was a guest at the Forschungsinstitut für Mathematik (ETH Zürich) and at the University of Rome

0025-5831/81/0256/0401/$03.60

element $A=(A_1,\ldots,A_m)\in L$ we associate two other elements (5.4, basic construction):

$$\tilde{A}\in\tilde{L}:=\mathrm{Hom}(V_1,U_1)\times\mathrm{Hom}(U_1,V_2)\times\ldots\times\mathrm{Hom}(U_{m-1},V_m)$$

$$\tilde{A}:V_1\xrightarrow{\bar{A}_1}U_1\overset{I_1}{\hookrightarrow}V_2\xrightarrow{\bar{A}_2}U_2\overset{I_2}{\hookrightarrow}\ldots\ldots\xrightarrow{\bar{A}_{m-1}}U_{m-1}\xrightarrow{I_{m-1}}V_m$$

where $U_i:=\mathrm{Im}\,A_i\subset V_{i+1}$ and $A_i=I_i\bar{A}_i:V_i\twoheadrightarrow U_i\hookrightarrow V_{i+1}$ is the canonical decomposition, and

$$A'\in L':=\mathrm{Hom}(U_1,U_2)\times\mathrm{Hom}(U_2,U_3)\times\ldots\times\mathrm{Hom}(U_{m-2},U_{m-1})$$

$$A':U_1\xrightarrow{A'_1}U_2\xrightarrow{A'_2}U_3\longrightarrow\ldots\xrightarrow{A'_{m-2}}U_{m-1}$$

$A'_i=\bar{A}_{i+1}I_i$. Now we consider the two quotient maps (5.1, first fundamental theorem)

$$\begin{array}{ccc}\tilde{L} & \xrightarrow{\pi} & L'\\ {\scriptstyle\varrho}\downarrow & & \\ L & & \end{array}$$

defined by "composing":

$$\pi(B_1,C_1,B_2,C_2,\ldots)=(B_2C_1,B_3C_2,\ldots),$$

$$\varrho(B_1,C_1,B_2,C_2,\ldots)=(C_1B_1,C_2B_2,\ldots).$$

By construction we have $\pi(\tilde{A})=A'$, $\varrho(\tilde{A})=A$, and we will show (Proposition 5.5)

a) $\pi^{-1}(\bar{O}_{A'})=\bar{O}_{\tilde{A}}$ and $\mathrm{codim}_{\tilde{L}}\bar{O}_{\tilde{A}}=\mathrm{codim}_{L'}\bar{O}_{A'}$,

b) $\varrho(\bar{O}_{\tilde{A}})=\bar{O}_A$.

By induction (on the length m of the quiver) we may assume that $\bar{O}_{A'}$ is Cohen-Macaulay with rational singularities. This implies by a) that $\bar{O}_{\tilde{A}}$ is Cohen-Macaulay. Using a natural desingularization of $\bar{O}_{\tilde{A}}$ (Section 6) and a result of Kempf (Corollary 1.4) we are able to prove that $\bar{O}_{\tilde{A}}$ has rational singularities too (Proposition 7.1). Since ϱ is a quotient map and $\varrho(\bar{O}_{\tilde{A}})=\bar{O}_A$ by b), we get from Boutot's theorem (1.3b) that $\bar{O}_A$ is normal, Cohen-Macaulay with rational singularities.

0.3. Our second main theorem deals with the singularities arising in the closure $\bar{O}$ of an orbit $O\subsetneq L$. We describe the type (up to smooth equivalence) of the singularity of $\bar{O}$ in a *minimal degeneration* O', i.e. O' is an open orbit in $\bar{O}-O$. Using the description of the degenerations given in [1] we associate in a purely combinatorial way to each minimal degeneration $O'\subset\bar{O}$ a pair (p,q) of natural numbers, called the *index of the minimal degeneration* (9.1).

Theorem. *Let $O'\subset\bar{O}$ be a minimal degeneration of index (p,q). Then O' is of codimension $p+q-1$ in $\bar{O}$ and the singularity of $\bar{O}$ in O' is smoothly equivalent to the isolated singularity of the determinantal variety $D_{p,q}:=\{A\in M_{p,q}\mid \mathrm{rk}\,A\leqq 1\}$ in the origine.*

Our method here is similar to the one in [11]. We describe the orbits in L by a *diagram of boxes* (2.3), the rows representing the indecomposable factors of the corresponding representation. Given two orbits O_υ and O_λ, $O_\upsilon \subset \bar{O}_\lambda$, with associated diagram of boxes υ and λ, we give some reduction procedures which enable us to remove (under certain conditions) common rows (10.2) and common columns (10.1) to obtain a new pair of orbits $O_{\upsilon'}$ and $O_{\lambda'}$ in some smaller space L' with the same type of singularity of $\bar{O}_{\lambda'}$ in $O_{\upsilon'}$ as $\bar{O}_\lambda$ in O_υ. In case of a minimal degeneration of index (p, q) this procedure will end up in the pair $O_{\lambda'} = D_{p,q} - \{0\}$, $O_{\upsilon'} = \{0\}$, proving the claim (10.4).

1. Notations and Preliminaries

1.1. *Notations.* We always work over an algebraically closed field k of characteristic zero. For finite dimensional k-vector spaces $V_1, V_2, \ldots, V_m$ we define

$$L(V_1, V_2, \ldots, V_m):$$

$$= \mathrm{Hom}_k(V_1, V_2) \times \mathrm{Hom}_k(V_2, V_3) \times \ldots \times \mathrm{Hom}_k(V_{m-1}, V_m)$$

and

$$G(V_1, V_2, \ldots, V_m) := \mathrm{GL}(V_1) \times \mathrm{GL}(V_2) \times \ldots \times \mathrm{GL}(V_m).$$

This group acts in a natural way on $L(V_1, \ldots, V_m)$. For any $A \in L(V_1, \ldots, V_m)$ we denote by O_A its orbit in $L(V_1, \ldots, V_m)$.

1.2. *Quotients, First Fundamental Theorem.* If a reductive group G acts regularly on an affine variety Z a regular map $\pi: Z \to X$ from Z into some affine variety X will be called a *quotient with respect to* G if π induces a surjective map from the regular functions on X onto the G-invariant functions on Z.

The *first fundamental theorem* of invariant theory tells us (cf. [15, II.6, Theorem 2.6A] or [14, Par. 3, Théorème 3]) that *the map*

$$\pi: \mathrm{Hom}(U, V) \times \mathrm{Hom}(V, W) \to \mathrm{Hom}(U, W)$$

given by $(A, B) \mapsto BA$ *is a quotient with respect to* $\mathrm{GL}(V)$.

1.3. *Some Properties of Quotients.* Let $\pi: Z \to X$ be a quotient map with respect to a reductive group G.

a) *If* $Y \subseteq Z$ *is* G*-stable and closed then* $\pi(Y) \subseteq X$ *is closed and* $\pi|_Y: Y \to X$ *is a quotient* [12, Chap. I, Sect. 2].

In particular we remark that if in addition Y is *normal* $\pi(Y)$ *is normal* too.

b) Boutot's Theorem: *If* Z *has rational singularities* (1.4) $\pi(Z)$ *has rational singularities too* (cf. [6])

1.4. *Rational Singularities.* Let X be an irreducible variety and $\varphi: Y \to X$ a *resolution of singularities*, i.e. Y is smooth and φ is birational and proper. One says that X has *rational singularities* if a) X *is normal and* b) *the direct images* $R^i\varphi_* \mathcal{O}_Y$ *of*

the structure sheaf $\mathcal{O}_Y$ vanish for $i>0$. (One can show that this definition does not depend on the resolution φ.) One has the following result due to Kempf:

Proposition (cf. [7, p. 50, Proposition]). *Let $\varphi: Y \to X$ be a resolution of singularities, $n := \dim X$. The following statements are equivalent:*

(i) *X has rational singularities.*

(ii) *X is normal and Cohen-Macaulay and for every n-form ω defined on the smooth part of X the form $\varphi^*\omega$ extends to the whole variety Y.*

From this we immediately deduce the following criterion, which will be very useful for our task.

Corollary. *Let X be an irreducible Cohen-Macaulay variety and $\varphi: Y \to X$ a resolution of singularities. Assume that there is a closed subset $X' \subset X$ such that X has rational singularities in $X - X'$ and $\varphi^{-1}(X')$ has codimension at least 2 in Y. Then X has rational singularities.*

(In fact X is normal in $X - X'$, hence normal by Serre's criterion [4, IV, Théorème 5.8.6]. Given any n-form ω defined outside the singular locus of X, we can first extend $\varphi^*\omega$ to $\varphi^{-1}(X - X')$ since $X - X'$ has rational singularities, and then to the whole variety Y, since $\varphi^{-1}(X')$ has codimension at least 2.)

2. Degenerations of Orbits in L

2.1. We recall here some facts proved in [1] (cf. also [2]). To each element $A = (A_1, \ldots, A_{m-1}) \in L := L(V_1, \ldots, V_m)$ we associate an indexed set of integers $n^A = \{n^A_{(i,j)}\}_{1 \leqq i \leqq j \leqq m}$ defined by

$$n^A_{(i,j)} := \begin{cases} \operatorname{rk}(A_{j-1} \ldots A_{i+1} A_i) & \text{for } i<j \\ \dim V_i & \text{for } i=j. \end{cases}$$

It is clear that n^A depends only on the orbit O_A.

2.2. Proposition [1, Proposition 3.1]. *For $A, B \in L$ we have $O_B \subseteq \bar{O}_A$ if and only if $n^B_{(i,j)} \leqq n^A_{(i,j)}$ for all $1 \leqq i \leqq j \leqq n$.*

In particular the orbit O_A is determined by n^A.

2.3. We also introduce the following indexed set of integers $\lambda^A = (\lambda^A_{(i,j)})_{1 \leqq i \leqq j \leqq m}$:

$$\lambda^A_{(i,j)} := n^A_{(i,j)} - n^A_{(i-1,j)} - n^A_{(i,j+1)} + n^A_{(i-1,j+1)}.$$

(We put $n^A_{(r,s)} = 0$ if $r<1$ or $s>m$.)

These numbers have the following interpretation (cf. [1]): *Each representation $A \in L$ has a unique decomposition into a direct sum $A = \oplus E_\tau$ of indecomposable representations E_τ which are of the form*

$$E_{(i,j)}: \underset{1}{0} \longrightarrow \underset{2}{0} \longrightarrow \ldots \longrightarrow \underset{i}{k} \xrightarrow{\mathrm{id}} k \xrightarrow{\mathrm{id}} \ldots \xrightarrow{\mathrm{id}} \underset{j}{k} \longrightarrow 0 \longrightarrow \ldots \longrightarrow \underset{m}{0}$$

for some $1 \leqq i \leqq j \leqq m$ (cf. [2, 2.2]). Now $\lambda^A_{(i,j)}$ *is the number of indecomposable factors E_τ of A of type $E_{(i,j)}$.*

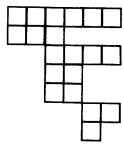
Clearly n^A is determined by λ^A:

$$n^A_{(i,j)} = \sum_{r \leqq i \leqq j \leqq s} \lambda^A_{(r,s)}.$$

It is convenient to represent λ^A as a "*diagram of boxes*", each row starting at i and ending at j stands for one indecomposable factor of type $E_{(i,j)}$.

E.g. the following diagram represents λ^A for A isomorphic to

$$E_{(1,6)} \oplus E_{(1,3)} \oplus E_{(3,6)} \oplus E_{(3,4)} \oplus E_{(3,4)} \oplus E_{(5,6)} \oplus E_{(5,5)}:$$



2.4. Conversely any indexed set $\lambda = (\lambda_{(i,j)})_{1 \leqq i \leqq j \leqq m}$ of natural numbers determines an orbit in $L(V_1, V_2, \ldots, V_m)$ provided $\dim V_i = \hat{\lambda}_i := \sum_{r \leqq i \leqq s} \lambda_{(r,s)}$ (= # boxes in the i^{th} column of λ). We will shortly call such an indexed set a *diagram*, define $\dim \lambda := (\hat{\lambda}_1, \ldots \hat{\lambda}_m) \in \mathbb{N}^m$ to be its *dimension* and denote by O_λ the *corresponding orbit*. For any $A \in O_\lambda$ we have by definition

$$n^A_{(i,j)} = n^\lambda_{(i,j)} := \sum_{r \leqq i \leqq j \leqq s} \lambda_{(r,s)}.$$

In this way we get a bijection between the orbits in $L(V_1, \ldots, V_m)$ and the diagrams of dimension $(\dim V_1, \ldots, \dim V_m)$.

If O_ν, O_λ are orbits in $L(V_1, \ldots, V_m)$, *we say* $\nu \leqq \lambda$ *if* $O_\nu \subseteq O_\lambda$ i.e. if $n^\nu_{(i,j)} \leqq n^\lambda_{(i,j)}$ for all (i,j) (cf. Proposition 2.2). This define an *ordering* on the set of all diagrams of a fixed dimension.

3. Dimension Formula for Stabilizers

3.1. For any $A \in L := L(V_1, \ldots, V_m)$ we denote by $\operatorname{Stab} A$ the *stabilizer* of A in the group $G := G(V_1, \ldots, V_m)$:

$$\operatorname{Stab} A = \{g \in G | gA = A\}.$$

It is easy to see that its Lie algebra $\operatorname{Lie} \operatorname{Stab} A \subset \operatorname{End}(V_1) \times \ldots \times \operatorname{End}(V_m)$ is the *endomorphism ring* of A:

$$\operatorname{Lie} \operatorname{Stab} A = \operatorname{End}_G A.$$

Furthermore we have

$$\operatorname{Hom}_G(E_{(r,s)}, E_{(i,j)}) \cong \begin{cases} k & i \leqq r \leqq j \leqq s \\ 0 & \text{otherwise} \end{cases}$$

for the indecomposable representations $E_{(i,j)}$ (2.3). Using the decomposition of A into indecomposable factors this proves the first part of the following lemma; the second follows by direct computation from the definition of $\lambda^A_{(i,j)}$ in 2.3.

3.2. Lemma.

$$\dim \mathrm{Stab}\, A = \sum_{1 \leqq i \leqq r \leqq j \leqq s \leqq m} \lambda^A_{(i,j)} \cdot \lambda^A_{(r,s)}$$

$$= \sum_{1 \leqq i \leqq j \leqq m} (n^A_{(i,j)} - n^A_{(i,j+1)})(n^A_{(i,j)} - n^A_{(i-1,j)}) .$$

4. An Induction Lemma

4.1. Let $W_1, W_2, \ldots, W_n$ be finite dimensional vector spaces, $n > 2$, and fix an integer $t \in \{2, 3, \ldots, n-1\}$. Consider the map

$$\pi := \pi_t : L := L(W_1, \ldots, W_n) \to L' := L(W_1, \ldots, \hat{W}_t, \ldots, W_n)$$

given by $A = (A_1, \ldots, A_{n-1}) \mapsto (A_1, \ldots, A_t A_{t-1}, \ldots, A_{n-1})$ i.e. by composing at the t^{th} point. This is a *quotient map* with respect to $\mathrm{GL}(W_t)$ (first fundamental theorem, cf. 1.2).

4.2. Proposition. *Assume* $\dim W_{t-1}, \dim W_{t+1} \leqq \dim W_t$.

a) *The map* π *is smooth in* $L^0 := \{A = (A_1, \ldots, A_{n-1}) \in L | A_{t-1}$ or A_t has maximal rank$\}$.

For any $A = (A_1, \ldots, A_{n-1}) \in L$ *such that* A_{t-1} *and* A_t *have maximal rank we have*

b) $\dim \mathrm{Stab}\, A = \dim \mathrm{Stab}\, \pi(A) + (\dim W_t - \mathrm{rk}\, A_t) \cdot (\dim W_t - \mathrm{rk}\, A_{t-1})$ *and*

c) $\pi^{-1}(\pi(\bar{O}_A)) = \bar{O}_A$ *and* $\mathrm{codim}_L \bar{O}_A = \mathrm{codim}_{L'} \bar{O}_{\pi(A)}$.

Proof. a) The tangent map of π in $A \in L$ is given by

$$(d\pi)_A(X_1, \ldots, X_{n-1}) = (X_1, \ldots, X_t A_{t-1} + A_t X_{t-1}, \ldots, X_{n-1}) .$$

Hence $(d\pi)_A$ is surjective if either A_{t-1} is injective or A_t is surjective, i.e. if $A \in L^0$.

b) It is easy to describe the decomposition of $\pi(A)$ into indecomposable factors from the decomposition of A. For this let us label the quiver corresponding to L' by

$$\overset{\cdot}{1} \to \overset{\cdot}{2} \to \ldots \to \overset{\cdot}{t-1} \to \overset{\cdot}{t+1} \to \ldots \to \overset{\cdot}{n}$$

and define for $i \in \{1, 2, \ldots, n\}$ the integers i' and i'' by $i' = i'' = i$ for $i \neq t$ and $t' = t+1$, $t'' = t-1$. Then the factors of A of type $E_{(t,t)}$ disappear and a factor of type $E_{(i,j)}$ of A with $(i,j) \neq (t,t)$ becomes a factor of type $E_{(i',j'')}$ of $\pi(A)$. Since by assumption A_{t-1} is injective and A_t is surjective, we don't have factors of type $E_{r,t-1}$ or $E_{t+1,j}$ in A, and therefore $\lambda^A_{(r,t-1)} = \lambda^A_{(t+1,j)} = 0$. It follows that

$$\sum_{\substack{i \leqq r \leqq j \leqq s \\ (r,j) \neq (t,t)}} \lambda^A_{(i,j)} \lambda^A_{(r,s)} = \sum_{i' \leqq r' \leqq j'' \leqq s''} \lambda^{\pi(A)}_{(i',j'')} \lambda^{\pi(A)}_{(r',s'')} .$$

Hence from the dimension formula 3.2 we get

$$\dim \mathrm{Stab}\, A = \dim \mathrm{Stab}\, \pi(A) + \sum_{\substack{1 \leqq i \leqq t \\ t \leqq s \leqq n}} \lambda^A_{(i,t)} \cdot \lambda^A_{(t,s)} .$$

The last sum is equal to $\left(\sum_{i\leqq t}\lambda^A_{(i,t)}\right)\cdot\left(\sum_{t\leqq s}\lambda^A_{(t,s)}\right)$ and it is easy to see that

$$\sum_{i\leqq t}\lambda^A_{(i,t)}=\dim \operatorname{Ker} A_t=\dim W_t-\operatorname{rk} A_t$$

and

$$\sum_{t\leqq s}\lambda^A_{(t,s)}=\operatorname{codim}\operatorname{Im} A_{t-1}=\dim W_t-\operatorname{rk} A_{t-1}.$$

c) We have $\pi(O_A)=O_{\pi(A)}$, hence $\pi(\bar O_A)=\bar O_{\pi(A)}$ by 1.3a). Furthermore any $C\in L$ satisfies $n^C_{(i,j)}=n^{\pi(C)}_{(i,j)}$ for $i\neq t$, $j\neq t$, $i\leqq j$, and by assumption $n^A_{(i,t)}=n^A_{(i,t-1)}$, $n^A_{(t,j)}=n^A_{(t+1,j)}$ for $i<t$ and $j>t$. Now for $B\in L$ with $\pi(B)\in\bar O_{\pi(A)}$ we get for all (i,j) $n^{\pi(B)}_{(i,j)}\leqq n^{\pi(A)}_{(i,j)}$ by 2.2. Hence $n^B_{(i,j)}\leqq n^A_{(i,j)}$ for $i\neq t$ and $j\neq t$ and therefore

$$n^B_{(i,t)}\leqq n^B_{(i,t-1)}=n^{\pi(B)}_{(i,t-1)}\leqq n^{\pi(A)}_{(i,t-1)}=n^A_{(i,t-1)}=n^A_{(i,t)}$$

for $i<t$ and

$$n^B_{(t,j)}\leqq n^B_{(t+1,j)}=n^{\pi(B)}_{(t+1,j)}\leqq n^{\pi(A)}_{(t+1,j)}=n^A_{(t+1,j)}=n^A_{(t,j)}$$

for $j>t$. This proves $B\in\bar O_A$ by 2.2, i.e. $\pi^{-1}(\pi(\bar O_A))=\bar O_A$. From b) we get (setting $d_i:=\dim W_i$)

$$\begin{aligned}\operatorname{codim}_L\bar O_A&=\dim L-\dim G+\dim \operatorname{Stab} A\\&=\dim L-\dim G+\dim\operatorname{Stab}\pi(A)\\&\quad+(d_t-d_{t+1})(d_t-d_{t-1})\\&=(\dim L-d_td_{t-1}-d_{t+1}d_t+d_{t+1}d_{t-1})\\&\quad-(\dim G-d_t^2)+\dim\operatorname{Stab}\pi(A)\\&=\dim L'-\dim G'+\dim\operatorname{Stab}\pi(A)\\&=\operatorname{codim}_{L'}\bar O_{\pi(A)}.\quad\text{qed}\end{aligned}$$

4.3. *Remark.* It is immediate from the proof above that the proposition holds also for $t=1$ and $t=n$, i.e. for the two projections

$$\pi_1: L(W_1,\ldots,W_n)\longrightarrow L(W_2,\ldots,W_n),$$
$$\pi_n: L(W_1,\ldots,W_n)\longrightarrow L(W_1,\ldots,W_{n-1}).$$

(We set $W_i=0$ for $i<1$ or $i>n$ and $A_i=0$ if $i<1$ or $i>n-1$.)

5. The Basic Construction

5.1. Let $V_1,V_2,\ldots,V_m$ and $U_1,U_2,\ldots,U_{m-1}$ be finite dimensional vector spaces. Consider the following two maps π and ϱ

$$\begin{array}{ccc}\tilde L:=L(V_1,U_1,V_2,U_2,\ldots,U_{m-1},V_m) & \xrightarrow{\pi} & L':=L(U_1,\ldots,U_{m-1})\\ \Big\downarrow{\scriptstyle\varrho} & & \\ L:=L(V_1,\ldots,V_m) & & \end{array}$$

given by

$$\pi(B_1, C_1, B_2, C_2, \ldots, B_{m-1}, C_{m-1}) = (B_2C_1, B_3C_2, \ldots, B_{m-1}C_{m-2}),$$
$$\varrho(B_1, C_1, B_2, C_2, \ldots, B_{m-1}, C_{m-1}) = (C_1B_1, C_2B_2, \ldots, C_{m-1}B_{m-1}).$$

Define

$$G := G(V_1, \ldots, V_m),$$
$$G' := G(U_1, \ldots, U_{m-1})$$

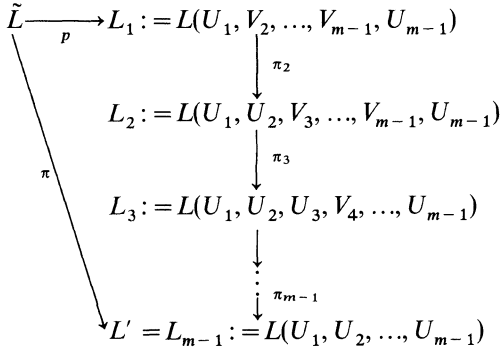
and

$$\tilde{G} := G(V_1, U_1, V_2, U_2, \ldots, U_{m-1}, V_m).$$

The group $\tilde{G}$ is canonically isomorphic to $G \times G'$, the maps π and ϱ are equivariant with respect to the projections $\tilde{G} \to G'$ and $\tilde{G} \to G$ and are *quotient maps* with respect to G and G' respectively (cf. first fundamental theorem 1.2).

5.2. We remark that the map π is a composition $\pi = \pi_{m-1} \cdot \pi_{m-2} \cdot \ldots \cdot \pi_2 \cdot p$ of the following form:

(*)



where p is the canonical projection and π_i is the quotient with respect to $\mathrm{GL}(V_i)$, i.e. it is of the form given in 4.1.

5.3. Consider the following two subsets of $\tilde{L}$:

$$\tilde{L}^0 := \{(B_1, C_1, \ldots, B_{m-1}, C_{m-1}) |$$

for each $i = 1, \ldots, m-2$ either C_i is injective or B_{i+1} is surjective$\}$

$$\tilde{L}^{\max} := \{(B_1, C_1, \ldots, B_{m-1}, C_{m-1}) |$$

all B_i are surjective and all C_i are injective$\}$.

Proposition. *Assume* $\dim U_i \leqq \dim V_i$, $\dim V_{i+1}$ *for* $i = 1, \ldots, m-1$. *Then* $\tilde{L}^0 \supset \tilde{L}^{\max}$ *are not empty and*

a) π *is smooth in* $\tilde{L}^0$,

b) $\varrho|_{\tilde{L}^{\max}} : \tilde{L}^{\max} \longrightarrow \varrho(\tilde{L}^{\max})$ *is a fibration with typical fibre* $G(U_1, \ldots, U_{m-1})$.

(Here fibration means locally trivial in the étale topology.)

roof. The first assertion is clear. For a) consider the decomposition of π given in .2. The assumption implies that each L_i in (*) satisfies the condition of roposition 4.2 and that we can apply 4.2a) to the image of $\tilde{L}^0$ in L_i. Since p is learly smooth this implies that π is smooth in $\tilde{L}^0$. For b) consider the group $I:=\prod_{i=1}^{m-1}(\mathrm{GL}(V_i)\times\mathrm{GL}(U_i)\times\mathrm{GL}(V_{i+1}))$ which operates in a natural way on $\tilde{L}$ nd on L:

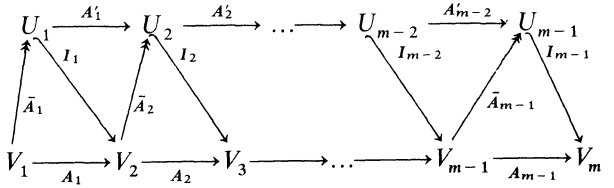
$$(g_1,h_1,g_1',g_2,h_2,g_2',\ldots)(B_1,C_1,B_2,C_2,\ldots)$$
$$=(h_1B_1g_1^{-1},g_1'C_1h_1^{-1},h_2B_2g_2^{-1},g_2'C_2h_2^{-1},\ldots),$$
$$(g_1,h_1,g_1',g_2,h_2,g_2',\ldots)(D_1,D_2,\ldots)=(g_1'D_1g_1^{-1},g_2'D_2g_2^{-1},\ldots).$$

Now clearly $\tilde{L}^{\max}$ is an orbit under H and ϱ is H-equivariant. This implies that the nap $\varrho|_{\tilde{L}^{\max}}$ is of the form $H/H_1\to H/H_2$ with two subgroups $H_1\subset H_2\subset H$, hence ocally trivial in the étale topology. Furthermore it is obvious that for each $C\in\tilde{L}^{\max}$ he fibre $\varrho^{-1}(\varrho(C))$ is the $G(U_1,\ldots,U_{m-1})$ orbit of C and the stabilizer of C in $G(U_1,\ldots,U_{m-1})$ is trivial. qed

5.4. Now let us start with a fixed element $A\in L$,

$$A: V_1\xrightarrow{A_1}V_2\xrightarrow{A_2}\ldots\longrightarrow V_{m-1}\xrightarrow{A_{m-1}}V_m.$$

Defining $U_i:=\mathrm{Im}\,A_i\subset V_i$, $i=1,\ldots,m-1$, we get the following basic construction:



Here $A_i=I_i\cdot\bar{A}_i$ is the canonical factorization and $A_i':=\bar{A}_{i+1}I_i$. In this way we have associated to $A\in L$ two elements $\tilde{A}\in\tilde{L}$ and $A'\in L'$:

$$\tilde{A}: V_1\xrightarrow{\bar{A}_1}U_1\xrightarrow{I_1}V_2\xrightarrow{\bar{A}_2}\ldots\longrightarrow V_{m-1}\xrightarrow{\bar{A}_{m-1}}U_{m-1}\xrightarrow{I_{m-1}}V_m,$$

$$A': U_1\xrightarrow{A_1'}U_2\xrightarrow{A_2'}U_3\longrightarrow\ldots\longrightarrow U_{m-2}\xrightarrow{A_{m-2}'}U_{m-1},$$

such that $\pi(\tilde{A})=A'$ and $\varrho(\tilde{A})=A$. Let us denote by $O_{\tilde{A}}$, $O_{A'}$ and O_A the corresponding orbits in $\tilde{L}$, L' and L respectively.

Remark. It follows from the construction that the diagram of A' is obtained from the diagram of A by *removing the last box of each row*:

$$\lambda^{A'}_{(i,j)}=\lambda^{A}_{(i,j+1)}.$$

5.5. The main induction step for the proof of Theorem 0.2 is the following result.

Proposition. *Set* $N_A := \pi^{-1}(\bar{O}_{A'})$. *Then*

(i) $O_{\tilde{A}} = \varrho^{-1}(O_A) \subset \tilde{L}^{\max}$ *and* $\pi(O_{\tilde{A}}) = O_{A'}$,

(ii) $\bar{O}_{\tilde{A}} = N_A$ *and* $\varrho(N_A) = \bar{O}_A$,

(iii) $\operatorname{codim}_{\tilde{L}} N_A = \operatorname{codim}_{L'} \bar{O}_{A'}$.

Proof. Since $\pi(\tilde{A}) = A'$ and $\varrho(\tilde{A}) = A$ we get $\pi(O_{\tilde{A}}) = O_{A'}$ and $\varrho(O_{\tilde{A}}) = O_A$. Furthermore $\tilde{A} \in \tilde{L}^{\max}$ and $\varrho^{-1}(A)$ is an orbit under $G(U_1, \ldots, U_{m-1})$ by construction (cf. proof of Proposition 5.3), hence $\varrho^{-1}(O_A) = O_{\tilde{A}}$. This proves i). For ii) and iii) we use the decomposition (*) of π (5.2) and remark that each L_i in (*) satisfies the assumption of Proposition 4.2. Since p is smooth and $\bar{A}_1$ and I_{m-1} are of maximal rank we first get $\operatorname{codim}_{\tilde{L}} \bar{O}_{\tilde{A}} = \operatorname{codim}_{L_1} p(\bar{O}_{\tilde{A}})$ and $p^{-1}(p(\bar{O}_{\tilde{A}})) = \bar{O}_{\tilde{A}}$. Now the image of $\tilde{A}$ in each L_i satisfies the assumption of 4.2c). Hence we get by induction $\bar{O}_{\tilde{A}} = \pi^{-1}(\pi(\bar{O}_{\tilde{A}})) = N_A$ and $\operatorname{codim}_{\tilde{L}} N_A = \operatorname{codim}_{L'} \bar{O}_{A'}$. qed

5.6. *Remark.* The construction of the orbit $O_{A'}$ from the orbit O_A does not depend on the special choice of the vector spaces U_i, but only on their dimensions. More precisely we can formulate the Proposition 5.5 in the following way:

If O_λ *is an orbit in* $L(V_1, \ldots, V_m)$ *with associated diagram* λ *and* $U_1, \ldots, U_{m-1}$ *are vector spaces of dimension* $\dim U_i = n^{\lambda}_{(i,i+1)}$ *and if we denote by* λ' *the diagram obtained from* λ *by removing the last box in each row* (i.e. $\lambda'_{(i,j)} = \lambda_{(i,j+1)}$), *then we have*

(i) $\varrho^{-1}(O_\lambda)$ *is an orbit in* $\tilde{L}$ *contained in* $\tilde{L}^{\max}$, $O_{\lambda'} = \pi(\varrho^{-1}(O_\lambda))$,

(ii) $\varrho^{-1}(O_\lambda) = \pi^{-1}(\bar{O}_{\lambda'})$,

(iii) $\operatorname{codim}_{\tilde{L}} \pi^{-1}(\bar{O}_{\lambda'}) = \operatorname{codim}_{L'} \bar{O}_{\lambda'}$.

6. A Resolution of Singularities

6.1. We fix an element $C \in L := L(W_1, W_2, \ldots, W_n)$, $C = (C_1, C_2, \ldots, C_{n-1})$.

For each $i = 1, 2, \ldots, n$ consider the *flag variety* $\mathscr{F}_i$ of flags $F_i = (F_i^0 = W_i \supseteq F_i^1 \supseteq F_i^2 \supseteq \ldots \supseteq F_i^{i-1})$ of nationality $(n^C_{(i,i)}, n^C_{(i-1,i)}, n^C_{(i-2,i)}, \ldots, n^C_{(1,i)})$ (i.e. $\dim F_i^t = n^C_{(i-t,i)}$).

E.g. $\mathscr{F}_1$ is a point and $\mathscr{F}_2$ is a grassmanian. For $B_j : W_j \to W_{j+1}$, $F_j \in \mathscr{F}_j$ and $F_{j+1} \in \mathscr{F}_{j+1}$ we shortly write $B_j F_j \subseteq F_{j+1}$ if $B_j F_j^t \subset F_{j+1}^{t+1}$ for all $t = 0, 1, \ldots, j-1$. Of course $G := G(W_1, \ldots, W_n)$ acts in the obvious way on $\mathscr{F}_C := \mathscr{F}_1 \times \mathscr{F}_2 \times \ldots \times \mathscr{F}_n$. *Define* $Y := Y_C \subset L \times \mathscr{F}_C$ *to be the subvariety of pairs* (B, F), $B = (B_1, \ldots, B_{n-1})$, $F = (F_1, \ldots, F_n)$ *satisfying*

$$B_j F_j \subseteq F_{j+1}, \quad j = 1, \ldots, n-1.$$

It is clear that Y_C is a *closed and G-stable subvariety.* Let us consider the two maps

$$\begin{array}{ccc} Y_C & \xrightarrow{\varphi} & L \\ {\scriptstyle p}\downarrow & & \\ \mathscr{F}_C & & \end{array}$$

induced by the two projections of $L \times \mathscr{F}_C$. It is easy to see that $p: Y_C \to \mathscr{F}_C$ is a *subvectorbundle* of the trivial bundle $pr: L \times \mathscr{F}_C \to \mathscr{F}_C$. In particular Y_C is a *smooth* variety. Furthermore Φ is *proper*, since $Y_C \subset L \times \mathscr{F}_C$ is closed and $\mathscr{F}_C$ is projective.

6.2. Lemma. *$\Phi(Y_C) = \bar{O}_C$ and the induced map $\varphi: Y_C \to \bar{O}_C$ is proper and birational*, i.e. φ is a resolution of singularities (1.4).

Proof. The definition of Y_C implies that for each $(B, F) \in Y_C$, $B := (B_1, \ldots, B_{n-1})$, $F := (F_1, \ldots, F_n)$ we have $B_{j-1} \ldots B_i(W_i) \subset F_j^{j-i}$ for $j > i$, hence $n^B_{(i,j)} \leqq \dim F_i^{j-i} = n^C_{(i,j)}$. Therefore $B \in \bar{O}_C$ by Proposition 2.2 which implies $\Phi(Y_C) \subseteqq \bar{O}_C$. Furthermore $\varphi^{-1}(C) = \{(C, F_C)\}$, $F_C = (F_1, \ldots, F_n)$ defined by $F_i^t := C_{i-1} \ldots C_{i-t}(W_{i-t})$ for $t < i$ $(F_1 := W_1)$. In particular $C \in \Phi(Y_C)$ and hence $\Phi(Y_C) = \bar{O}_C$ (since Y_C is G-stable and closed and Φ is proper), and φ induces an isomorphism $\varphi^{-1}(O_C) \xrightarrow{\sim} O_C$ and so φ is birational. qed

7. Proof of the First Main Theorem

7.1. Now we go back to the basic construction 5.4. The proof of the theorem 0.2 follows immediately by induction from the following result.

Proposition (Notations of Sect. 5). *If $\bar{O}_A$, has rational singularities* (1.4), *then $N_A = \pi^{-1}(\bar{O}_{A'})$ and $\bar{O}_A$ have rational singularities too.*

Proof. Since $\bar{O}_A = \varrho(N_A)$ [Proposition 5.5ii and since ϱ is a quotient map, the first assertion implies the second by Boutot's theorem (1.3b)]. From Lemma 6.2 we get a desingularization of $N_A = \bar{O}_{\tilde{A}}$ (5.5ii):

$$\varphi: Y_{\tilde{A}} \longrightarrow N_A = \bar{O}_{\tilde{A}},$$

$$Y_{\tilde{A}} \subset \tilde{L} \times \mathscr{F}_{\tilde{A}},\ \tilde{L} := L(V_1, U_1, V_2, U_2, \ldots, U_{m-1}, V_m),$$

$\mathscr{F}_{\tilde{A}} := \mathscr{F}_1 \times \mathscr{F}_1' \times \mathscr{F}_2 \times \mathscr{F}_2' \times \ldots \times \mathscr{F}_{m-1}' \times \mathscr{F}_m$ (in obvious notations). The codimension result 5.4 iii) implies that the schematic inverse image $\pi^{-1}(\bar{O}_{A'})$ is Cohen-Macaulay [4, IV, Proposition 15.4.2c')⇒a)]. Since the map π is smooth in $N_A^0 := N_A \cap \hat{L}^0$ (5.3a) and since $\bar{O}_{A'}$ has rational singularities, $\pi^{-1}(\bar{O}_{A'})$ is a reduced Cohen-Macaulay variety, being smooth in the dense open set $O_{\tilde{A}} \subset N_A^0$ (5.5i, ii), [4, IV, 5.8.5], with rational singularities in N_A^0. The following lemma enables us to apply Corollary 1.4 which implies the claim. qed

7.2. Lemma. $\operatorname{codim}_{Y_{\tilde{A}}} \varphi^{-1}(N_A - N_A^0) \geqq 2$.

Proof. Let $C := N_A - N_A^0$. It is enough to prove that for each $F \in \mathscr{F}_{\tilde{A}}$ the set $\varphi^{-1}(C) \cap p^{-1}(F)$ has codimension at least 2 in $p^{-1}(F)$ ($p: Y_{\tilde{A}} \to \mathscr{F}_{\tilde{A}}$ is the projection 7.1). The fibre $p^{-1}(F)$ is a product $H_1 \times K_1 \times H_2 \times \ldots \times H_{m-1} \times K_{m-1}$ of subspaces $H_i \subset \operatorname{Hom}(V_i, U_i)$, $K_i \subset \operatorname{Hom}(U_i, V_{i+1})$. By construction $\varphi^{-1}(C) \cap p^{-1}(F)$ is a union of subvarieties $D_j \subset H_1 \times K_1 \times \ldots \times H_{m-1} \times K_{m-1}$, $j = 1, \ldots, m-1$, D_j defined by the condition that the element in H_j is not surjective and the element in K_j is not injective. Since the generic element of H_j is surjective and the generic element of K_j is injective the subvariety D_j is of codimension greater or equal to 2. qed

8. Determinantal Singularities

8.1. *Definition* (cf. [5] 1.7). Consider two varieties X, Y and two points $x\in X$, $y\in Y$. Then the singularity of X in x is called *smoothly equivalent* to the singularity of Y in y if there is a variety Z, a point $z\in Z$ and two maps

$$\begin{array}{ccc} Z & \xrightarrow{\varphi} & X \\ {\scriptstyle\psi}\downarrow & & \\ Y & & \end{array}$$

such that $\varphi(z)=x$, $\psi(z)=y$, and φ and ψ are smooth in z.

This clearly defines an *equivalence relation* between pointed varieties (X,x). We denote the equivalence class of (X,x) by $\mathrm{Sing}(X,x)$.

Assume that an algebraic group G acts regularly on the variety X. Then $\mathrm{Sing}(X,x)=\mathrm{Sing}(X,x')$ if x and x' belong to the same orbit O. In this case we denote the equivalence class also by $\mathrm{Sing}(X,O)$.

8.2. For two vector spaces U, V of dimension p, q respectively consider the orbit $O_{\min}\subset L(U,V)$ of maps of rank 1. Then $\bar{O}_{\min}=O_{\min}\cup\{0\}$ and $0\in\bar{O}_{\min}$ is an isolated rational singularity (Theorem 0.2, cf. [9, Sect. 2]). We denote its equivalence class by $d_{p,q}$:

$$\mathrm{Sing}(\bar{O}_{\min},0)=d_{p,q}.$$

We have $\dim\bar{O}_{\min}=p+q-1$ *and* $\bar{O}_{\min}$ *is smooth if and only if* $p=1$ *or* $q=1$ [i.e. $\bar{O}_{\min}=L(U,V)$].

Assuming $p,q>1$, one gets for any (r,s): $d_{r,s}=d_{p,q}$ if and only if $(r,s)=(p,q)$ or $(r,s)=(q,p)$. [Since both are represented by isolated singularities one has $p+q-1=r+s-1$. Since $L(U,V)$ is the Zariskitangentspace of $\bar{O}_{\min}$ in 0 we must also have $p\cdot q=r\cdot s$.]

8.3. It is easy to describe a resolution of singularities of $\bar{O}_{\min}$.

Choose a basis $e_1,\ldots,e_p$ of U and a basis $f_1,\ldots,f_q$ of V and consider the one dimensional subspace

$$M:=\{A\in L(U,V)\,|\,Ae_1\subset kf_1, Ae_i=0 \text{ for } i=2,\ldots,p\}.$$

Then the stabilizer P of M in $G:=G(U,V)$ is the parabolic $P=P_{e_1}\times P_{f_1}\subset \mathrm{GL}(U)\times\mathrm{GL}(V)=G$, where P_{e_1}, P_{f_1} are the stabilizers of the lines ke_1, kf_1.

Now $G/P=\mathbb{P}^{p-1}\times\mathbb{P}^{q-1}$, the associated bundle

$$G\times^P M\longrightarrow G/P$$

is the line bundle $\mathcal{O}_{\mathbb{P}^{p-1}\times\mathbb{P}^{q-1}}(-1)$ (given by the Segre embedding) and the canonical map

$$\varphi: G\times^P M\longrightarrow\bar{O}_{\min}$$

is a resolution of singularities [with $\varphi^{-1}(0)=$ zero section of the line bundle]. Hence we obtain the singularity $d_{p,q}$ *by "collapsing" the line bundle* $\mathcal{O}_{\mathbb{P}^{p-1}\times\mathbb{P}^{q-1}}(-1)$ (cf. [9]).

8.4. Let us recall that an orbit O' is called a *minimal degeneration* of the orbit $O \subsetneqq L$ if O' is open in $\bar{O}-O$, or equivalently if for each orbit O'' such that $O' \subsetneqq \bar{O}'' \subsetneqq \bar{O}$ we have either $O''=O$ or $O''=O'$. In the following Sect. 9 we will associate to any minimal degeneration a pair (p,q) of natural numbers called the *index of the minimal degeneration*; it can easily be read off the diagrams of the orbits.

Now we can reformulate our second main theorem (cf. 0.3).

Theorem. *Let $O' \subset \bar{O}$ be a minimal degeneration of orbits in $L(V_1, \ldots, V_m)$ with index (p,q). Then*

$$\operatorname{codim}_{\bar{O}} O' = p+q-1 \quad \text{and} \quad \operatorname{Sing}(\bar{O}, O') = d_{p,q}.$$

The proof will be given in section 10; it uses two general reduction results for singularities in closures of orbits $O \subset L(V_1, \ldots, V_m)$ (10.1, 10.2). From 8.2 we obtain the following corollary:

Corollary. *If $O' \subset \bar{O}$ is a minimal degeneration with $\operatorname{codim}_{\bar{O}} O' \leqq 2$, then $\bar{O}$ is smooth in O'.*

In particular the singular locus of $\bar{O}$ can be strictly smaller than the boundary $\partial O = \bar{O}-O$, even if ∂O has codimension $\geqq 2$.

9. The Index of a Minimal Degeneration

In order to define the index of a minimal degeneration $O' \subset \bar{O}$ we have to recall the main result on degenerations of orbits in $L=L(V_1, \ldots, V_m)$ proved in [1].

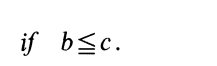
9.1. Proposition. *Let $O' \subset \bar{O}$ be a minimal degeneration, $\nu < \lambda$ the associated diagrams. Then there are natural numbers $a<b$, $b-1 \leqq c<d$ such that*

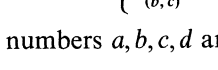
$$\nu_{(i,j)} = \begin{cases} \lambda_{(i,j)} & \text{for } (i,j) \neq (a,d),(a,c),(b,d),(b,c) \\ \lambda_{(a,c)}+1 & \text{for } (i,j)=(a,c) \\ \lambda_{(b,d)}+1 & \text{for } (i,j)=(b,d) \\ \lambda_{(a,d)}-1 & \text{for } (i,j)=(a,d) \\ \lambda_{(b,c)}-1 & \text{for } (i,j)=(b,c) \quad \text{if } b \leqq c. \end{cases}$$

Clearly the numbers a, b, c, d are uniquely determined by the two diagrams $\nu < \lambda$.

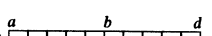
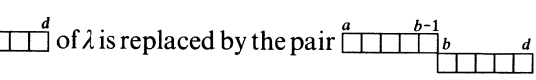
Definition. With the notations of the proposition above the pair $(\nu_{(a,c)}, \nu_{(b,d)})$ is called *the index of the minimal degeneration.*

9.2. In more geometric terms Proposition 9.1 means that for a minimal degeneration $O_\nu \subset \bar{O}_\lambda$ the diagram ν is obtained from the diagram λ performing one of the following two operations:

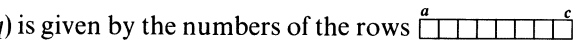
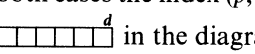
a) *a shift*, i.e. a pair of rows of λ of the form 

is replaced by the pair 

,

b) *a cut*, i.e. a row  of λ is replaced by the pair 

(here $c=b-1$),
and all other rows remain unchanged.

In both cases the index (p, q) is given by the numbers of the rows  and  in the diagram v.

Clearly a cut or a shift always define a degeneration. But in general such a degeneration is not minimal, as one easily sees from examples. The following lemma gives additional conditions for a minimal degeneration showing that certain rows cannot appear in λ and v. (The proof is purely combinatorial and left to the reader.)

Lemma. *Let $O_v \subset \bar{O}_\lambda$ be a minimal degeneration and consider the numbers $a<b$, $b-1 \leqq c<d$ given in Proposition* 9.1. *Then*

$$\lambda_{(i,j)}=v_{(i,j)}=0 \quad \text{for} \quad a \leqq i \leqq b, c \leqq j \leqq d \quad \text{and}$$
$$(i,j) \neq (a,c),(a,d),(b,c),(b,d).$$

9.3. We now can prove the codimension formula given in Theorems 0.3 and 8.4.

Proposition. *Let $O' \subset \bar{O}$ be a minimal degeneration with index (p, q). Then*

$$\operatorname{codim}_{\bar{O}} O' = p+q-1.$$

Proof. In order to simplify the notations let us denote by $[a, b]$ the diagram consisting in one row starting at a and ending at b. Furthermore define for two diagrams α and β:

$$h(\alpha,\beta)=\sum_{i \leqq r \leqq j \leqq s} \alpha_{(i,j)} \cdot \beta_{(r,s)}$$

From 3.2 we have for $A \in O_\lambda$ that $\dim \operatorname{Stab} A = h(\lambda, \lambda)$, hence

$$\operatorname{codim}_{\bar{O}_\lambda} O_v = h(v,v)-h(\lambda,\lambda).$$

Now one easily checks the following equations:

$$h([a,c]+[b,d],[i,j])-h([a,d]+[b,c],[i,j])=\begin{cases}1 & \text{for} \quad a \leqq i<b, c \leqq j<d \\ 0 & \text{otherwise}\end{cases}$$

$$h([i,j],[a,c]+[b,d])-h([i,j],[a,d]+[b,c])=\begin{cases}1 & \text{for} \quad a<i \leqq b, c<j \leqq d \\ 0 & \text{otherwise}.\end{cases}$$

($\alpha+\beta$ denotes the diagram consisting of the union of the rows of α and β.)

Since $v<\lambda$ is a minimal degeneration, there are numbers $a<b$, $b-1 \leqq c<d$ satisfying the equations of Proposition 9.1, i.e. $\lambda=[a,d]+[b,c]+\delta$, $v=[a,c]+[b,d]+\delta$ for a suitable diagram δ (delet $[b,c]$ if $b>c$). By definition of the index (p, q) we have $\delta=(p-1)[a,c]+(q-1)[b,d]+\delta'$, δ' a diagram with $\delta'_{(a,c)}=\delta'_{(b,d)}=0$. Furthermore it follows from Lemma 9.2 that

$$\delta'_{(i,j)}=0 \quad \text{for} \quad a \leqq i \leqq b, \quad c \leqq j \leqq d, \quad (i,j) \neq (a,d),(b,c).$$

Hence we find

$$\begin{aligned} &h(v,v)-h(\lambda,\lambda)\\ &\quad=h([a,c]+[b,d],[a,c]+[b,d])-h([a,d]+[b,c],[a,d]+[b,c])\\ &\qquad+h([a,c]+[b,d],\delta)-h([a,d]+[b,c],\delta)\\ &\qquad+h(\delta,[a,c]+[b,d])-h(\delta,[a,d]+[b,c])\\ &\quad=1+(p-1)+(q-1). \quad \text{qed} \end{aligned}$$

10. Some Reduction Results, Proof of the Second Main Theorem

Let O_v, O_λ be orbits in $L(V_1,\ldots,V_m)$, $O_v\subset\bar{O}_\lambda$, $v\leqq\lambda$ the associated diagrams (2.4). In this section we will describe some procedures in order to obtain from the pair $v\leqq\lambda$ a new pair $v'\leqq\lambda'$ of diagrams with less boxes such that

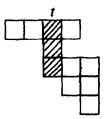
$$\mathrm{Sing}(\bar{O}_\lambda,O_v)=\mathrm{Sing}(\bar{O}_{\lambda'},O_{v'}).$$

(Compare [11] where the same is done for conjugacy classes of matrices.)

10.1. For the first reduction result let us consider the quotient map

$$\pi:=\pi_t:L(V_1,\ldots,V_m)\longrightarrow L(V_1,\ldots,\hat{V}_t,\ldots,V_m)$$

for some fixed $t\in\{1,2,\ldots,m\}$, given by $(A_1,\ldots,A_{m-1})\mapsto(A_1,\ldots,A_tA_{t-1},\ldots,A_m)$ (4.2, 4.3; we always set $V_i=0$ if $i<1$ or $i>m$ and $A_i=0$ if $i<1$ or $i>m-1$). If $O_\lambda\subset L(V_1,\ldots,V_m)$ is an orbit with associated diagram λ the diagram λ' of the orbit $\pi(O_\lambda)$ is obtained from λ by removing the t^{th} column (and pushing the two parts together).

E.g. $\lambda=$  gives $\lambda'=$  .

Proposition. *Assume* $\dim V_{t+1}$, $\dim V_{t-1}\leqq\dim V_t$. *Let* $A,B\in L(V_1,\ldots,V_m)$, $B\in\bar{O}_A$, *such that* A_{t-1} *and* A_t *have maximal rank and either* B_{t-1} *is injective or* B_t *is surjective. Then*

$$\mathrm{Sing}(\bar{O}_A,O_B)=\mathrm{Sing}(\bar{O}_{\pi(A)},O_{\pi(B)}).$$

Proof. The claim follows immediately from Proposition 4.2 (and Remark 4.3). In fact we have by assumption O_A, $O_B\subset L^0:=\{C\in L(V_1,\ldots,V_m)|C_{t-1}$ or C_t has maximal rank$\}$ and $\bar{O}_A=\pi^{-1}(\pi(\bar{O}_A))=\pi^{-1}(\bar{O}_{\pi(A)})$ (4.2c). Hence π induces a smooth map

$$\bar{\pi}:\bar{O}_A\cap L^0\longrightarrow\bar{O}_{\pi(A)}$$

[4.2a)] with $\bar{\pi}(O_B)=O_{\pi(B)}$ and so by Definition 8.1

$$\mathrm{Sing}(\bar{O}_A,O_B)=\mathrm{Sing}(\bar{O}_{\pi(A)},O_{\pi(B)}). \quad \text{qed}$$

10.2. Our second reduction result is obtained in the following way. For $A \in L(V_1, \ldots, V_m)$, $C \in L(W_1, \ldots, W_m)$ the sum $A \oplus B$ is defined using the canonical embedding

$$L(V_1, \ldots, V_m) \oplus L(W_1, \ldots, W_m) \hookrightarrow L(V_1 \oplus W_1, \ldots, V_m \oplus W_m).$$

If λ and δ are the associated diagrams to O_A and O_C the diagram of $O_{A \oplus C}$ is $\lambda + \delta$, given by

$$(\lambda + \delta)_{(i,j)} := \lambda_{(i,j)} + \delta_{(i,j)}$$

i.e. $\lambda + \delta$ consists in the union of the rows of λ and δ.

Proposition. *Let* $A, B \in L(V_1, \ldots, V_m)$, $C \in L(W_1, \ldots, W_m)$.

a) *We have* $O_B \subset \bar{O}_A$ *if and only if* $O_{B \oplus C} \subset \bar{O}_{A \oplus C}$. *Furthermore if* $O_{B \oplus C} \subset \bar{O}_{A \oplus C}$ *is a minimal degeneration then* $O_B \subset \bar{O}_A$ *is a minimal degeneration.*

b) *If* $O_B \subset \bar{O}_A$ *and if* $\operatorname{codim}_{\bar{O}_A} O_B = \operatorname{codim}_{\bar{O}_{A \oplus C}} O_{B \oplus C}$, *then*

$$\operatorname{Sing}(\bar{O}_A, O_B) = \operatorname{Sing}(\bar{O}_{A \oplus C}, O_{B \oplus C}).$$

Proof. a) follows from 2.2 since $n^{A \oplus C}_{(i,j)} = n^A_{(i,j)} + n^C_{(i,j)}$ for all (i,j).

For b) consider the subgroup $H = G(V_1, \ldots, V_m) \times G(W_1, \ldots, W_m) \subset G := G(V_1 \oplus W_1, \ldots, V_m \oplus W_m)$, operating on $E := L(V_1, \ldots, V_m) \oplus L(W_1, \ldots, W_m)$.

Then $O := O_A \times O_C$ and $O' := O_B \times O_C$ are orbits under H in E and $\bar{O} = \bar{O}_A \times \bar{O}_C$. Hence $\operatorname{Sing}(\bar{O}, O') = \operatorname{Sing}(\bar{O}_A, O_B)$ and $\operatorname{codim}_{\bar{O}} O' = \operatorname{codim}_{\bar{O}_A} O_B$. Furthermore we have the decomposition

$$\operatorname{Lie} G = \bigoplus_{i=1}^{m} \operatorname{End}(V_i \oplus W_i) = \operatorname{Lie} H \oplus N$$

with $N := \bigoplus_{i=1}^{m} (\operatorname{Hom}(V_i, W_i) \oplus \operatorname{Hom}(W_i, V_i))$. Now under the operation of $\operatorname{Lie} G$ on $L(V_1 \oplus W_1, \ldots, V_m \oplus W_m)$ we clearly have $NE \cap E = 0$. Since $\bar{O}_{A \oplus C}$ is normal (0.2) the claim follows from Lemma 10.3 below. qed

Remark. In terms of diagrams $\nu \leqq \lambda$ the proposition says that we can remove some common rows of λ and ν without changing the type of singularity, provided the codimension does not change. More precisely, *if* $\lambda = \lambda' + \delta$, $\nu = \nu' + \delta$ *with some diagram* δ *such that* $\operatorname{codim}_{\bar{O}_\lambda} O_\nu = \operatorname{codim}_{\bar{O}_{\lambda'}} O_{\nu'}$, *then* $\operatorname{Sing}(\bar{O}_\lambda, O_\nu) = \operatorname{Sing}(\bar{O}_{\lambda'}, O_{\nu'})$.

10.3. Lemma [11, Proposition 4.2]. *Let* G *be an algebraic group,* $H \subset G$ *a subgroup,* L *a* G*-module and* $E \subset L$ *an* H*-stable subspace. Consider elements* $x \in E$ *and* $y \in \overline{Hx}$ *and assume that*

(i) *there is a decomposition* $\operatorname{Lie} G = \operatorname{Lie} H \oplus N$ *such that* $Ny \cap E = 0$,

(ii) $\operatorname{codim}_{\overline{Hx}} \mathrm{Hy} = \operatorname{codim}_{\overline{Gx}} Gy$,

(iii) $\overline{Gx}$ *is normal in* y.

Then $\operatorname{Sing}(\overline{Gx}, y) = \operatorname{Sing}(\overline{Hx}, y)$.

10.4. Now we can prove our main result on singularities in closures of orbits (Theorems 0.3 and 8.4). Let $O_\nu \subset \bar{O}_\lambda$ be a minimal degeneration of index (p, q). By Proposition 9.1 there are numbers a, b, c, d with $a < b$ and $b - 1 \leqq c < d$ and a

diagram δ such that

$$\lambda=[a,d]+[b,c]+(p-1)[a,c]+(q-1)[b,d]+\delta$$
$$\upsilon=p[a,c]+q[b,d]+\delta$$

(cf. proof of Proposition 9.3; delet $[b,c]$ if $c=b-1$). Using Proposition 9.3 our second reduction result 10.2 implies that we may assume $\delta=\emptyset$ (cf. Remark 10.2). If $c\geqq b$ it follows from the first reduction result 10.1 that

$$\mathrm{Sing}(\bar{O}_\lambda, O_\upsilon)=\mathrm{Sing}(\bar{O}_{\lambda'}, O_{\upsilon'}),$$

where λ', υ' are obtained from λ, υ by removing the c^{th} column, i.e

$$\lambda'=[a,d-1]+[b,c-1]+(p-1)[a,c-1]+(q-1)[b,d-1]$$
$$\upsilon'=p[a,c-1]+q[b,d-1].$$

Hence by induction we may suppose that $c=b-1$, i.e

$$\lambda=[a,d]+(p-1)[a,b-1]+(q-1)[b,d]$$

$$\upsilon=p[a,b-1]+q[b,d].$$

Finally if $a<b-1$ (or $b<d$) we can again apply Proposition 10.1, this time to the a^{th} column (or the d^{th} column), ending up at the pair

$$\lambda=[1,2]+(p-1)[1,1]+(q-1)[2,2]$$
$$\upsilon=p[1,1]+q[2,2],$$

i.e. $O_\lambda=O_{\min}\subset M_{q,p}$, $O_\upsilon=\{0\}$. qed

References

1. Abeasis, S., Del Fra, A.: Degenerations for the representations of an equioriented quiver of type $\mathcal{A}_m$. Boll. Un. Mat. It. (Suppl.) Algebra e Geometria, 157–171 (1980)
2. Gabriel, P.: Unzerlegbare Darstellungen. I. manuscripta math. **6**, 71–103 (1972)
3. Gabriel, P.: Indecomposable representations. II. Ist. Nazionale di Alta Mat. Symposia Mat. **XI**, 81–104 (1973)
4. Grothendieck, A., Dieudonné, J.: EGA 0-IV. Publ. Math. de l'I.H.E.S. **11**, **20**, **24**, **28**, **32**, Paris (1961–1967)
5. Hesselink, W.: Singularities in the nilpotent scheme of a classical group. Trans. Am. Math. Soc. **222**, 1–32 (1976)
6. Hochster, M.: Proceedings of the International Congress of Mathematicians. Helsinki, 1978
7. Kempf, G., Knudsen, F., Mumford, D., Saint-Donat, B.: Toroidal embeddings. I. Springer Lecture Notes, Vol. 339. Berlin, Heidelberg, New York: Springer 1973
8. Kempf, G.: Images of homogeneous vector bundles and varieties of complexes. Bull. Am. Math. Soc. **81**, 900–901 (1975)
9. Kempf, G.: On the collapsing of homogeneous bundles. Invent. Math. **37**, 229–239 (1976)
10. Kraft, H., Procesi, C.: Closures of conjugacy classes of matrices are normal. Invent. Math. **53**, 227–247 (1979)

11. Kraft, H., Procesi, C.: Minimal singularities in GL_n. Invent. Math. **62**, 503–515 (1981)
12. Mumford, D.: Geometric invariant theory. Ergebnisse der Mathematik, Vol. 34. Berlin, Heidelberg, New York: Springer 1970
13. Procesi, C., Kraft, H.: Classi coniugate in $GL(n, \mathbb{C})$. Rend. Sem. Mat. Univ. Padova **59**, 209–222 (1978)
14. Vust, Th.: Sur la théorie des invariants des groupes classiques. Ann. Inst. Fourier (Grenoble) **26**, 1–31 (1976)
15. Weyl, H.: The classical groups, their invariants and representations. Princeton Mathematical Series, Vol. 1, Princeton: Princeton University Press 1946

Received April 1, 1981

Math. Ann. 256, 419–427 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Classical Godeaux Surface in Characteristic *P*

William E. Lang*

Department of Mathematics, University of California, Berkeley, CA 94720, USA

A classical Godeaux surface is a nonsingular minimal projective surface X satisfying $K_X^2=1$, $p_a=p_g=0$, and $\mathrm{Pic}^\tau(X)=Z/5$. (By analogy with Enriques surfaces in characteristic two [2], there should exist non-classical Godeaux surfaces in characteristic five, satisfying $K_X^2=1$, $p_a=0$, $p_g=1$, $\mathrm{Pic}^\tau(X)=\mu_5$ (singular Godeaux surfaces) or α_5 (supersingular Godeaux surfaces). Miranda has recently given an elegant construction of singular Godeaux surfaces.) The purpose of this paper is to give a construction of classical Godeaux surfaces which works in all characteristics. The method, following Godeaux, is to divide a hypersurface Y in P^3 of degree 5 by μ_5, the group scheme of fifth roots of unity. Therefore our method leads to simply connected surfaces of general type with $p_g=0$ in characteristic five, answering a question of Blass. The principal technical difficulty is to show that Y can be chosen so that Y/μ_5 is nonsingular, and we accomplish this using invariant theory and Bertini's theorem.

In the second part of the paper, we show that, roughly speaking, all classical Godeaux surfaces are obtained by the Godeaux method. More precisely, if X' is the canonical model of X, then it is easy to see that there is a natural μ_5-covering Y' of X', and we show that Y' is a quintic hypersurface in P^3, and that X' is the quotient of Y' by the Godeaux action of μ_5 on P^3. The methods used in this part are very close to those used by Miyaoka [8], Reid [15], and Dolgacev [3], but somewhat simpler. In the third part, we give some miscellaneous results and conjectures on classical Godeaux surfaces.

Our notations are standard, except that we use K_X for the dualizing invertible sheaf of a Gorenstein scheme X, and the canonical model of such a scheme is defined as $\mathrm{Proj} \oplus H^0(X, nK_X)$.

1. Construction of Classical Godeaux Surfaces

Fix k, an algebraically closed field of arbitrary characteristic. Let $x_1, \ldots, x_4$ be homogeneous coordinates on P_k^3. Define the Godeaux action of μ_5 on P^3 by

* NSF Postdoctoral Fellow

0025-5831/81/0256/0419/$01.80

$\lambda(x_i)=\lambda^i x_i$ [11]. (Up to linear automorphism, this is the unique linear action of μ_5 on P^3 with zero-dimensional fixed point set.) If Y is an invariant hypersurface not passing through the finite set of fixed points, then μ_5 acts freely on Y, and Y is a μ_5-torsor over $X=Y/\mu_5$.

Theorem 1. *If Y is a sufficiently general invariant quintic hypersurface, then $X=Y/\mu_5$ is a smooth minimal surface such that $K_X^2=1$, $p_a(X)=p_g(X)=0$, and* $\mathrm{Pic}^\tau(X)=Z/5$.

Proof. Since part of the proof uses lifting to characteristic zero, we first recall the classical argument, which works for $\mathrm{char}\, k\neq 5$. In this case, μ_5 is an etale group scheme, and therefore X is smooth if and only if Y is smooth. To show that the general invariant quintic is smooth, we need only exhibit one example, and the classical choice is the Fermat quintic, defined by $x_1^5+x_2^5+x_3^5+x_4^5=0$. One checks easily there is no invariant 2-form on Y, and therefore $p_g(X)=0$. Since $\chi(O_X)=\chi(O_Y)$, and $K_X^2=1/5K_Y^2$, we find that $\chi(O_X)=K_X^2=1$. X is minimal since K_X is ample. (K_X is ample since $f^*K_X=O_Y(1)$ is ample on Y [5, 1, 4.4].) To see that $\mathrm{Pic}^\tau(X)=Z/5$, we observe that since $H^1(O_X)=0$, $\mathrm{Pic}^\tau(X)$ is discrete and finite. By the Lefschetz theorem, Y is simply connected and therefore $\pi_1(X)=Z/5$. If G is an arbitrary finite commutative group scheme, then $\mathrm{Hom}\,(G, \mathrm{Pic}^\tau(X))\simeq H^1_{f1}(X, G')$, where G' is the Cartier dual of G [13, Prop. 6.2.1] and the latter group classifies G-torsors over X. From this, we see that if l is a prime not equal to $p=\mathrm{char}\, k$, then the l-primary part of $\mathrm{Pic}^\tau(X)$ is $Z/5$. If the p-primary part of $\mathrm{Pic}^\tau(X)$ is non-trivial, then X has a non-trivial μ_p-covering. But this implies that $H^0(X, \Omega_X^1)\neq 0$ [2], and therefore, since Y is etale over X, $H^0(Y, \Omega_Y^1)\neq 0$. But this is a contradiction, since Y is a smooth hypersurface in P^3.

In the case $p=5$, we must use a different argument to prove that the general X is smooth. The principal step is contained in the following proposition.

Proposition 1. *The invariant quintics define an embedding of P^3/μ_5 into P^{11}.*

Proof. We know that P^3/μ_5 is isomorphic to $\mathrm{Proj}\, R^G(G=\mu_5)$, where R is the homogeneous coordimate ring of P^3. (Cf. [10, Sect. 12] for the construction of X/G, where G is a finite group scheme. Then the proof of the isomorphism $P^3/G\simeq \mathrm{Proj}\, R^G$ proceeds as in [6, Chap. 4.1], using the fact that in our case G acts linearly on P^3.) Therefore, to prove the proposition, it is enough to show that the graded ring S such that $S_n=H^0(P^3, O_{P^3}(5n))^G$ is generated by elements of degree 1.

It is easy to see that the space of invariants of given degree is spanned by invariant monic monomials. And a monic monomial $x_1^a x_2^b x_3^c x_4^d$ is invariant if and only if $a+2b+3c+4d\equiv 0 \pmod 5$. So we can compute the dimension of the space of invariants of degree n by finding the number of non-negative integral solutions to $a+2b+3c+4d\equiv 0 \pmod 5$, $a+b+c+d=n$. This is independent of the ground field, so we may lift our action to characteristic zero in an obvious way and use Molien's theorem from characteristic zero invariant theory to compute the dimension of the space of invariants of degree n. Let s_n be this dimension.

Molien's Theorem. *Let G be a finite group acting on a vector space V over the complex numbers. Let s_n be the dimension of* $\mathrm{Symm}^n(V)^G$. *Then there is an equality of formal power series*

$$\sum_{n=0}^{\infty} s_n t^n = |G|^{-1} \sum_{g \in G} (\det |I - \sigma(g)\, t|)^{-1}$$

where I is the identity on V and $\sigma(g)$ is the linear transformation induced by g in degree 1.

Proof. See [17, 4.1.3].

In our case, we get $\sum_{n=0}^{\infty} s_n t^n = 1/5\,[(1-t)^{-4} + 4(1-\lambda t)^{-1} \ldots (1-\lambda^4 t)^{-1}]$ $= 1/5\,[(1-t)^{-4} + 4(1-t)(1-t^5)^{-1}]$. ($\lambda$ denotes a non-trivial fifth root of unity.) The coefficient of t^5 in this series is 12, and the coefficient of t^{10} is 58. We need notation for the 12 invariant monomials in degree 5. They are x_1^5, x_2^5, x_3^5, x_4^5, call these a_1, a_2, a_3, a_4; $x_1^3 x_3 x_4$, $x_2^3 x_1 x_3$, $x_3^3 x_2 x_4$, $x_4^3 x_1 x_2$, call these b_1, b_2, b_3, b_4; $x_1 x_3^2 x_4^2$, $x_2 x_1^2 x_3^2$, $x_3 x_2^2 x_4^2$, $x_4 x_1^2 x_2^2$, call these c_1, c_2, c_3, c_4.

We show that the invariant monic monomials in degree 10 are all products of those in degree 5. First, count those which are divisible by one of the a_i. There are $12 \times 4 - 6 = 42$ of these. Among those which are not divisible by any a_i, there are 6 products of distinct elements of group B, 4 squares of elements of group C, and 6 products of elements of group B and C. One checks easily that these are all distinct, and we get 58, the correct number.

Finally, we show that any monic invariant monomial of degree $5n$, $n \geqq 3$, is divisible by one of the a_i. In degree 15, a monic monomial is either divisible by some a_i or is of the form $x_i^3 \prod_{j \neq i} x_j^4$. But none of the latter are invariant. In degree $n \geqq 20$, any monomial is divisible by some a_i. This completes the proof of Proposition 1.

Since P^3/μ_5 is smooth away from the fixed points, a general hyperplane H in P^{11} will miss the images of the fixed points, and by Bertini's theorem, $P^3/\mu_5 \cap H = X$ will be a smooth surface. If we let Y be the quintic corresponding to H, then μ_5 acts freely on Y, and $Y/\mu_5 = X$.

Now for the computation of numerical invariants in characteristic five. First, I claim that for general choice of Y, $\mathrm{Pic}^\tau(X) = Z/5$. For this, note that if Y has a smooth point (which is clearly true for general Y), then Y is a non-trivial μ_5-torsor over X. Therefore, we get an injection $Z/5 \to \mathrm{Pic}^\tau(X)$. But the whole construction lifts over $W(k)$. Let $\mathscr{X}$ be such a smooth lifting of X. Then by a result of Raynaud [14, Th. 1], $\mathrm{Pic}^\tau(\mathscr{X})$ is flat over $W(k)$. Let $\mathscr{X}_{\bar{\eta}}$ be the geometric generic fibre of $\mathscr{X}$. Since $\mathrm{Pic}^\tau(\mathscr{X}_{\bar{\eta}}) = Z/5$, $\mathrm{Pic}^\tau(X) = Z/5$ also.

This implies that $H^1(X, O_X) = 0$. But $\chi(O_X) = 1/5\,\chi(O_Y)$, so $\chi(O_X) = 1$. Therefore $p_g(X) = 0$. Since the map $Y \to X$ is a homeomorphism for the etale topology, and Y is simply connected, X is simply connected.

We know [10, Sect. 12] that invertible sheaves on X correspond to G-linearized invertible sheaves on Y. (If $\mathscr{L}$ is an invertible sheaf on X, then the corresponding sheaf on Y is $f^*\mathscr{L}$.) We know also [2, Prop. 9] that $f^* K_X \simeq K_Y \simeq O_Y(1)$. This shows that $K_X^2 = 1/5\, K_Y^2 = 1$, and that K_X is ample.

Therefore X is minimal. It also shows that the embedding of X into P^{10} given above is obtained by means of the linear system $|5K|$. This completes the proof of Theorem 1.

2. Characterization of Classical Godeaux Surfaces

Lemma (Artin). *Let X be a minimal surface of general type, and let $\mathscr{L}=O_X(D)$ be an invertible sheaf on X. If $(D\cdot E)=0$ for all smooth rational curves E such that $E^2=-2$ (-2 curves), and if $f: X\to X'$ is the map of X to its canonical model (obtained by contracting the -2 curves to rational double points [9]), then there is an invertible sheaf $\mathscr{L}'$ on X' such that $\mathscr{L}=f^*\mathscr{L}'$.*

Proof. See [1, Cor. 2.6].

Let X be a classical Godeaux surface, X' its canonical model. If we choose a generator F of $\mathrm{Pic}^\tau(X)$, this descends to F' on X' by the lemma, and we get a cartesian diagram

$$\begin{array}{ccc} Y & \longrightarrow & Y' \\ \downarrow & & \downarrow \\ X & \longrightarrow & X' \end{array}$$

where the vertical arrows are μ_5-torsors induced by F and F', and the horizontal arrows are birational. (Note that if X contains no -2 curves, or equivalently if K_X is ample, then $X=X'$ and $Y=Y'$.)

Theorem 2. *In the situation of the preceding paragraph, Y' is embedded in P^3 as a quintic hypersurface by $|K_{Y'}|$, the top horizontal arrow is defined by $|K_Y|$, and K_Y is base point free. Furthermore, after a linear automorphism of P^3, the map $Y'\to P^3$ is μ_5-equivariant, where μ_5 acts on P^3 by the Godeaux action.*

Proof. We know that X and X' are Gorenstein. Therefore Y and Y' are Gorenstein by [4, V 9.6]. All maps pull dualizing sheaves back to dualizing sheaves. (This is true for the vertical arrows by [2, Prop. 9], for the bottom horizontal arrow by [1], and the top horizontal arrow by commutativity of the diagram.) Now $K_{X'}$ is ample on X', and therefore since $X'\to Y'$ is a finite morphism, $K_{Y'}$ is ample on Y'. Since μ_5 is linearly reductive, we know that if $\mathscr{L}$ is a line bundle on X, then $H^i(Y, f^*\mathscr{L})\simeq\oplus H^i(X,\mathscr{L}\otimes D)$, where D runs over $\mathrm{Pic}^\tau(X)$ (cf. [3] or [11]), and a similar statement is true for X' and Y'. (This key remark is due to Reid [15].) Therefore $H^0(Y', nK_{Y'})\simeq\oplus H^0(X', nK_{X'}+D')\simeq\oplus H^0(X, nK_X+D)\simeq H^0(Y, nK_Y)$, and therefore Y' is the canonical model of Y. Therefore, we will only need to study the canonical ring of Y.

Using Reid's remark, we compute $H^0(Y, nK_Y)$ for all $n\geqq 1$ by computing the cohomology of $nK+D$ on X. By Riemann-Roch, $\chi(X, nK+D)=1/2n(n-1)+1$. If $n\geqq 2$, then $h^2(nK+D)=h^0((1-n)K-D)=0$. If $n=1$, then $h^2(K+D)=h^0(-D)$, which is 1 if $D=0$, and 0 otherwise.

Next, we compute $h^1(nK+D)=h^1((1-n)K-D)=h^1(-(n-1)K-D)$. If $n\geqq 2$, this vanishes by the Ramanujan-Kodaira vanishing theorem (valid in

characteristic p since $H^1(X, O_X)=0$ [12]). The case $n=1$ requires additional argument. If $D=0$ then $h^1(K_X)=0$ by Serre duality. If $D\neq 0$, then $h^1(K+D)\neq 0$ implies $h^0(K+D)\geqq 2$ and therefore $h^0(2K+2D)\geqq 3$. (If D is a divisor on a non-singular surface, then $h^0(D)\geqq 2$ implies $h^0(2D)\geqq 3$. This follows from the well-known fact that for divisors D and E, the morphism $P(H^0(D)) \times P(H^0(E)) \to P(H^0(D+E))$ induced by multiplication has finite fibres.) But since we already know $h^1(2K+2D)=h^2(2K+2D)=0$, $h^0(2K+2D)=2$. This contradiction shows that $h^1(nK+D)=0$ for $n\geqq 1$, $D\in \mathrm{Pic}^\tau(X)$.

Combining these results, we see that $h^0(Y, nK_Y)=4$ if $n=1$, and $1/2(5n(n-1))+5$ if $n\geqq 2$. Therefore K_Y defines a rational map from Y into P^3. We show that this is a morphism by the argument of [3].

If $D\in \mathrm{Pic}^\tau(X)$, $D\neq 0$, let E_D be the unique curve in $|K+D|$.

Lemma. (a) *If $D\neq D'$ are non-trivial elements of* $\mathrm{Pic}^\tau(X)$, *then E_D and $E_{D'}$ have no common components.* (b) *If D, D', D'' are three distinct non-trivial elements of* $\mathrm{Pic}^\tau(X)$, *then there are no points P such that $P\in E_D$, $E_{D'}$, and $E_{D''}$.*

Proof. (a) If E_D or $E_{D'}$ is irreducible, the result is obvious. Suppose that $E_D=C+\sum n_i C_i$, $E_{D'}=C'+\sum n_i' C_i'$, are decomposition into irreducible components, where C and C' are chosen such that $K\cdot C=K\cdot C'=1$. (Therefore the C_i and $C_{i'}$ are -2 curves.) Let F be the common part of E_D and $E_{D'}$. If $F\neq 0$, then $F^2<0$ and is even, since F is a positive combination of -2 curves. Therefore $(E_D-F)^2=E_D^2-2E_{D'}F+F^2=1+F^2\leqq -1$. ($E_D$ is numerically equivalent to K_X.) But $(E_D-F)^2=(E_D-F)(E_{D'}-F)\geqq 0$ since E_D-F and $E_{D'}-F$ have no common components.

(b) Since $K_X^2=1$, each two E_D and $E_{D'}$, $D\neq D'$, meet transversally at a non-singular point for both curves. If three distinct E_D, $E_{D'}$, $E_{D''}$ meet at a point, then $O_X(E_D-E_{D'})$ restricted to $E_{D''}$ is the structure sheaf of $E_{D''}$. If this is the case, then we have an exact sequence

$$0\to O_X(E_D-E_{D'}-E_{D''})\to O_X(E_D-E_{D'})\to O_{E_{D''}}\to 0$$

and taking cohomology, we get

$$H^0(X, O_X(E_D-E_{D'}))\to H^0(E_{D''}, O_{E_{D''}})\to H^1(X, O_X(E_D-E_{D'}-E_{D''})).$$

But the third term is $H^1(-K-D''')(D'''\in \mathrm{Pic}^\tau(X))$, which is zero by Ramanujan-Kodaira, and the first term is zero since $E_D-E_{D'}$ is a non-trivial torsion divisor. Therefore $H^0(E_{D''}, O_{E_{D''}})=0$. This contradiction completes the proof of the lemma.

Since all non-trivial $D\in \mathrm{Pic}^\tau(X)$, $f^*E_D\in |K_Y|$, the lemma shows that K_Y is without base points, and defines a morphism from Y into P^3. The map must be birational onto a quintic hypersurface Z since $K_Y^2=5$. We get an injection $H^0(Z, O_Z(n))\to H^0(Y, nK_Y)$. But we know that $h^0(Z, O_Z(n))=\binom{n+3}{3}-\binom{n-2}{3}$, which equals 4 if $n=1$, and $1/2(5n(n-1))+5$ if $n\geqq 2$. Since this is the same as $h^0(Y, nK_Y)$, we have an isomorphism for all n. Therefore $Z=Y'$ and Theorem 2 is proved. (We leave the last assertion to the reader.)

3. Further Comments

Proposition 2. *Let X be a classical Godeaux surface and X' its canonical model. Then X' is embedded into P^{10} by $|5K_{X'}|$ and is cut out by quadrics and cubics.*

Proof. By Theorem 2, X' is the quotient of a quintic hypersurface in P^3 by the Godeaux action of μ_5 on P^3. Therefore, by the proof of Theorem 1, X' is embedded into P^{10} by $5K_{X'}$ as a hyperplane section of $P^3/\mu_5 \subset P^{11}$.

For the second assertion, it is enough to show that P^3/μ_5 is cut out in P^{11} by quadrics and cubics. Since the homogeneous coordinate ring of P^3/μ_5 is a subring of $k[x_1, \ldots, x_4]$ (generated by the invariant quintic monomials a_i, b_i, c_i; $i=1, \ldots, 4$), it is clear that the ideal of relations is generated by monomial relations of the form $\prod_{i=1}^{4} a_i^{e_i} b_i^{f_i} c_i^{g_i} - \prod_{i=1}^{4} a_i^{j_i} b_i^{k_i} c_i^{p_i} = 0$, where e_i, f_i, g_i, j_i, k_i, p_i are non-negative integers such that $\sum e_i + f_i + g_i = \sum j_i + k_i + p_i$. However, clearly all such relations are consequences of those of the form $m = \prod_{i=1}^{4} a_i^{h_i} e$, where m is an arbitrary monic monomial, the h_i are non-negative integers, and e is a monomial not divisible by any a_i (that is, any monomial can be written in this standard form). Since any invariant monomial of degree $5n$, $n \geqq 3$, is divisible by some a_i, we see that any monomial can be put into standard form by using only quadric and cubic relations, and that the e appearing in the standard form has degree $\leqq 10$. The proposition is now clear.

Corollary 1 (of Theorem 1). *There are no smooth invariant quintic hypersurfaces in characteristic five.*

Proof. If there were smooth invariant quintics, then the general invariant quintic would be smooth, and we could find a smooth quintic Y and a smooth classical Godeaux surface X with a μ_5 covering $Y \to X$. But the etale Euler number of X is 11, and that of Y is 55, by Noether's formula. This contradicts the fact that Y and X are homeomorphic for the etale topology.

Proposition 3. *If X is a non-ruled surface such that $c_2(X) \geqq 0$ and such that X has a holomorphic 1-form with only isolated zeroes, then $H^0(X, \theta_X) = 0$.*

Proof. Use the section of Ω_X^1 to form an exact sequence

$$0 \to O_X \to \Omega_X^1 \to I_Z K_X \to 0, \tag{1}$$

where I_Z is the ideal sheaf of a non-empty closed subscheme Z of condimension 2. We may identify θ_X with $\Omega_X^1 \otimes K_X^{-1}$. Then we get an exact sequence

$$0 \to H^0(X, K_X^{-1}) \to H^0(X, \Omega_X^1 \otimes K_X^{-1}) \to H^0(I_Z).$$

But $H^0(K_X^{-1}) = 0$ since K_X^n is effective for some n and $H^0(I_Z) = 0$ since Z is non-empty.

Corollary 2. *If Y is a general invariant quintic hypersurface, then the corresponding Godeaux surface has no vector field.*

Proof. If char $k \neq 5$, then Y is etale over X, and since Y does not have a vector field, neither does X. If char $k=5$, then the singularities of Y lie over the zeroes of the 1-form corresponding to the μ_5-covering [2]. Since it is easily checked that Y has only isolated singularities in general, the corollary follows from Proposition 3. Incidentally, one can prove easily by the method of [7] that $h^0(\Omega^1_X)=1$ if X is an arbitrary classical Godeaux surface.

Proposition 4. *The 1-form on a general classical Godeaux surface has only simple zeroes, and therefore the general invariant quintic has only 11 rational double points as singularities.*

Proof. Let $h=(h_1, \ldots, h_{12})$ be homogeneous coordinates on the P^{11} parametrizing invariant quintics such that h corresponds to the invariant quintic hypersurface H_h defined by $h_1a_1+h_2a_2+\ldots+h_4a_4+h_5b_1+\ldots+h_8b_4+h_9c_1+\ldots+h_{12}c_4$. Now consider the subset $\mathscr{I}$ of $P^3 \times P^{11}$ defined as follows: $\mathscr{I}=\{x, h) \mid x$ is a singular point of $H_h\}$. In coordinates, $\mathscr{I}$ is defined by the matrix equation

$$\begin{bmatrix} a_1 \ldots a_4 & b_1 \ldots b_4 & c_1 \ldots c_4 \\ \frac{\partial a_1}{\partial x_1} \cdots \frac{\partial a_4}{\partial x_1} & \frac{\partial b_1}{\partial x_1} \cdots \frac{\partial b_4}{\partial x_1} & \frac{\partial c_1}{\partial x_1} \cdots \frac{\partial c_4}{\partial x_1} \\ \frac{\partial a_1}{\partial x_4} \cdots \frac{\partial a_4}{\partial x_4} & \frac{\partial b_1}{\partial x_4} \cdots \frac{\partial b_4}{\partial x_4} & \frac{\partial c_1}{\partial x_4} \cdots \frac{\partial c_4}{\partial x_4} \end{bmatrix} \begin{bmatrix} h_1 \\ \vdots \\ \\ \vdots \\ h_{12} \end{bmatrix} = 0.$$

Call the 5×12 matrix appearing in this equation J. By writing J out explicitly, one sees that J has rank 1 at the four fixed points of the Godeaux action, and at every other point of P^3, J has rank 3. (For instance, the 3×3 minor of J formed from rows 1, 4, 5 and colums 1, 16, 12 is

$$\begin{bmatrix} x_1^5 & x_2^3 x_1 x_3 & x_4 x_1^2 x_2^2 \\ 0 & x_2^3 x_1 & 0 \\ 0 & 0 & x_1^2 x_2^2 \end{bmatrix}$$

which has determinant $x_1^8 x_2^5$. Similarly, we can find 3×3 minors with determinants $x_i^a x_j^b$ for all $1\ i,j\ 4$, $i \neq j$.) Therefore, if we let $\mathscr{I}_0 = \mathscr{I} - pr_1^{-1}$ (fixed points), $\mathscr{I}_0$ is irreducible of dimension 11. Under the second projection, $\mathscr{I} - \mathscr{I}_0$ projects to a proper subvariety of P^{11}, while $\mathscr{I}$ projects to all of P^{11}, by Corollary 1.

Lemma. *If Y is an invariant quintic hypersurface covering a classical Godeaux surface X with covering map f: $Y \to X$ such that the 1-form ω on X fixed by the Cartier operator has only isolated zeroes, then Y has a double point with distinct tangents at X if and only if ω has a simple zero at $f(x)$. Furthermore, in this case, the singularity of Y is a rational singularity analytically isomorphic to $z^5 = xy$.*

Proof. First, note that f is a homeomorphism for the Zariski topology. We can find an open covering U_i of X and elements g_i of $\Gamma(U_i, O_X^*)$ such that $\omega = dg_i/g_i$

on U_i and such that $f^{-1}(U_i)$ is isomorphic to the subscheme of $U_i \times A^1$ defined by $z^5 = f_i$ [2, Sect. 3]. The lemma follows easily.

Now consider the subscheme $\mathscr{W} \subset \mathscr{I}$ defined as follows: $\mathscr{W} = \{(x, h) \mid H_h$ has a singularity at x worse than $z^2 = xy\}$. It is enough to show that $\mathscr{W} \cap \mathscr{I}_0$ is properly contained in $\mathscr{I}_0$, since then $\mathscr{W}$ projects under pr_2 to a proper subvariety of P^{11}. Then if we take any point of h of P^{11} such that $h \notin pr_2(\mathscr{W}) \cup pr_2(\mathscr{I} - \mathscr{I}_0)$ such that H_h covers a smooth hyperplane section of P^3/μ_5, H_h will have only rational double points as singularities, and there will be 11 of them, since the number of singularities is the same as the number of zeroes of ω (which are all simple), and this number is $c_2(\Omega^1_X) = e(X) = 11$.

To show $\mathscr{W} \cap \mathscr{I} \subseteq \mathscr{I}_0$, consider the fibre of $\mathscr{I}$ over the point (1, 1, 0, 0). This is the P^8 of hypersurfaces of the form $h_1 x_1^5 - h_1 x_2^5 + h_3 x_3^5 + h_4 x_4^5 + h_5 x_1^3 x_3 x_4 + h_7 x_3^3 x_2 x_4 + h_8 x_4^3 x_1 x_2 + h_9 x_1 x_3^2 x_4^2 + h_{10} x_2 x_1^2 x_3^2 + h_{11} x_3 x_2^2 x_4^2$. Dehomogenize with respect to x_2 and find the Taylor expansion about $(x_1, x_3, x_4) = (1, 0, 0)$. We find that the quadratic term is $h_5 z_3 x_4 + h_{10} x_3^2$, so far general choice of coefficients, we have a double point with distinct tangents, and we are finished.

Proposition 5. *For a general classical Godeaux surface with K_X ample, $H^2(X, \theta_X) = 0$.*

Proof. We know that $H^2(X, \theta_X)$ is dual to $H^0(\Omega^1_X \otimes K_X)$. If the 1-form has isolated zeroes, we have an exact sequence

$$0 \to H^0(K_X) \to H^0(\Omega^1_X \otimes K_X) \to H^0(I_Z \otimes K_X^2) \to H^1(K_X).$$

But $H^0(K_X) = H^1(K_X) = 0$. Therefore $H^0(\Omega^1_X \otimes K_X) \simeq H^0(I_Z \otimes K_X^2)$. Pulling back to the covering quintic Y, we see that to prove $H^2(X, \theta_X) = 0$, it is enough to show that no invariant quadric passes through the singular locus of Y.

To show this, consider the linear system of invariant quintics which are singular at the points (1, 0, 0, 1) and (0, 1, 1, 0). These quintics are defined by polynomials of the form $f_1(x_1^5 - x_4^5) + f_2(x_2^5 - x_3^5) + g_1 x_1 x_3^2 + g_2 x_2 x_1^2 x_3^2 + g_3 x_3 x_2^2 x_4^2 + g_4 x_4 x_1^2 x_2^2$. Since the space of invariant quadrics is spanned by $x_1 x_4$ and $x_2 x_3$, the singularities of a member of our 5-dimensional linear system are not contained in any invariant quadric. Therefore, it is enough to find an element of our linear system with only isolated singularities.

For this, let Y be a general member of our linear system with equation h. Let C be defined by $\dfrac{\partial h}{\partial x_1} = \ldots = \dfrac{\partial h}{\partial x_4} = 0$. C only depends on the g_i, and for general choice of the g_i, all components of C have dimension $\leqq 1$, and C does not contain the line l defined by $x_1 = x_4$ and $x_2 = x_3$ as a component. Fix g_i so that this is satisfied, and for each 1-dimensional component V of C, choose a point on $V - l$. Then choose f_i so that Y does not pass through any of these points. Since Sing $Y = C \cap Y$, we see that Y constructed in this manner has isolated singularities not contained in an invariant quadric. Therefore, the general invariant quintic has isolated singularities not contained in an invariant quadric, and therefore, for a general classical Godeaux surface, $H^2(X, \theta_X) = 0$.

Question 1. Are classical Godeaux surfaces with K_X ample in characteristic five unirational?

Remark. We know that such a surface lifts to characteristic zero, and since $p_g = 0$, it is therefore supersingular ($\rho = B_2$). It is also simply connected. Therefore, Question 1 is simply a special case of a question of Shioda [16].

Acknowledgements. I wish to thank P. Blass for inspiring the present paper. I. Dolgacev for letting me study his preprint [3] before publication, J. Roberts, A. Ogus, and R. Hartshorne for stimulating conversations, and D. Richman for his help with the combinatorics.

References

1. Artin, M.: Some numerical criteria for contractability of curves on algebraic surfaces. Am. J. Math. **84**, 485-496 (1962)
2. Bombieri, E., Mumford, D.: Enriques' classification of surfaces in char. p, III. Invent. Math. **35**, 197-232 (1976)
3. Dolgacev, I.: Algebraic surfaces with $q = p_g = 0$ (to appear)
4. Hartshorne, R.: Residues and duality. Lecture Notes in Mathematics, Vol. 20. Berlin, Heidelberg, New York: Springer 1966
5. Hartshorne, R.: Ample subvarieties of algebraic varieties. Lecture Notes in Mathematics, Vol. 156. Berlin, Heidelberg, New York: Springer 1970
6. Knudson, D.: Algebraic spaces. Lecture Notes in Mathematics, Vol. 203. Berlin Heidelberg New York: Springer 1971
7. Lang, W.E.: Two theorems on de Rham cohomology. Compositio Math. **40**, 417-423 (1980)
8. Miyaoka, Y.: Tricanonical maps of numerical Godeaux surfaces. Invent. Math. **34**, 99-111 (1976)
9. Mumford, D.: The canonical ring of an algebraic surface. Ann. Math. **76**, 612-615 (1962)
10. Mumford, D.: Abelian varieties. Bombay: Oxford Press 1970
11. Peters, C.A.M.: On two types of surfaces of general type with vanishing geometric genus. Invent. Math. **32**, 33-47 (1976)
12. Ramanujan, C.P.: Remarks on the Kodaira vanishing theorem. J. Indian Math. Soc. **36**, 41-51 (1972); J. Indian Math. Soc. Suppl. **38**, 121-124 (1974)
13. Raynaud, M.: Specialization du foncteur de Picard. Publ. Math. I.H.E.S. **38**, 27-76 (1970)
14. Raynaud, M.: "p-Torsion" du schema de Picard. Asterisque **64**, 87-148 (1979)
15. Reid, M.: Surfaces with $P_g = 0$, $K^2 = 1$. J. Fac. Sci. Univ. Tokyo Sect. 1 A, **25**, 75-92 (1978)
16. Shioda, T.: On unirationality of supersingular surfaces. Math. Ann. **225**, 155-159 (1977)
17. Springer, T.A.: Invariant theory. Lecture Notes in Mathematics, Vol. 585. Berlin, Heidelberg, New York: Springer 1977

Received September 30, 1980

Math. Ann. 256, 429–437 (1981)

Mathematische Annalen

© Springer-Verlag 1981

On the De Rham Invariant of a Fibered $(4r+1)$-Dimensional Orientable Manifold

J.C. Alexander*

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

The de Rham invariant is the tangential Stiefel-Whitney number $w_2 w_{4r-1}$. After stating the result, we will discuss this number. It depends on the action of the Bockstein β on the middle-dimensional homology or cohomology of the manifold. Let $\mathbb{Z}_2 = \mathbb{Z}/2\mathbb{Z}$. For any odd-dimensional orientable topological manifold M (of dimensions $2s+1$), define

$$b(M) = \operatorname{rank}_{\mathbb{Z}_2}(\beta: H^s(M; \mathbb{Z}_2) \to H^{s+1}(M; \mathbb{Z}_2)).$$

Lusztig et al. [4] have shown that

$$w_2 w_{4r-1} = w_2 w_{4r-1}(M) \equiv b(M) \quad \text{modulo } 2.$$

Kreck [3] has a formula for determining this number if M fibers over a circle. The purpose of this note is to give a formula if M fibers over a connected surface B. Let $\chi(B)$ be the Euler characteristic of B and let F (of dimension $4r-1=2s+1$) be the fiber. Recall M is orientable.

Theorem. $w_2 w_{4r-1} \equiv \chi(B) \cdot b(F)$ modulo 2.

Note that the number is totally insensitive to the action of the fundamental group of B on F (so long as M is orientable). Also note that for any surface of odd Euler characteristic, and any r, there exist M with $w_2 w_{4r-1} \neq 0$. For example let M, defined over the projective plane, be the projective bundle of the Whitney sum of the canonical bundle and a $(4r-1)$-dimensional trivial bundle. Examples over other surfaces of odd Euler characteristic can be constructed via a bundle connected sum operation.

Before proving the result, we indicate why such a formula might be expected. For any manifold M, the "top", or coarsest, Stiefel-Whitney number is w_n. It is coarse in that it depends on very little information about M; it is of course the Euler characteristic modulo 2. If the dimension of M is odd, it is zero. If the dimension of M is $2s$, it depends only on the top Wu class v_s of M; $w_n = v_s^2$. This is another sense in which w_n is the "top" number.

* Partially supported by NSF and IPST, University of Maryland

0025-5831/81/0256/0429/$01.80

If the dimension of M is $4r+1$, the top possibly nonzero Wu class is v_{2r}. The number $w_2 w_{4r-1} = v_{2r} \cup \beta v_{2r}$ [4]. Thus $w_2 w_{4r-1}$ is the top Stiefel-Whitney number. Since $b(M)$ can be computed from the rational and $\mathbb{Z}_2$ Betti numbers of M, the number $w_2 w_{4r-1}$ is also rather coarse, at least if M is orientable.

The geometric significance of w_n and the Euler characteristic are are well-known. The number $w_2 w_{4r-1}$ also has geometric significance [1–3, 5, 6]. DeRham studied this invariant in connection with the linking form on $H_{2r}(M; \mathbb{Z})$ in his thesis [7, Sect. 19]; for this reason, it was christened the deRham invariant in [5].

There is another analogy between w_n and $w_2 w_{4r-1}$. Let V be a finite-dimensional vector space over $\mathbb{Z}_2$ with a symmetric bilinear form $\langle\ ,\ \rangle$ to $\mathbb{Z}_2$. Define the radical of V, denoted Rad V, as the subspace of all $v \in V$ with $\langle v, v' \rangle = 0$ for all $v' \in V$. Then $V/\text{Rad}\, V$ inherits a non-singular form. A characteristic element of V is any v such that $\langle v, v' \rangle = \langle v', v' \rangle$ for all $v' \in V$. A characteristic element always exists and is unique up to the radical. In particular, $\langle v, v \rangle \in \mathbb{Z}_2$, for v a characteristic element is a well-defined invariant of V. It is not hard to see this invariant is nothing but the dimension of $V/\text{Rad}\, V$ modulo 2.

If M has even dimension $2s$, we can consider the form

$$\langle v, v' \rangle = (v \cup v')[M]$$

on $H^s(M; \mathbb{Z}_2)$. The Wu class v_s is a characteristic element and thus w_n is the invariant described above. If M is orientable and has dimension $4r+1$, the Wu class v_{2r} is a characteristic element for the form

$$\langle v, v' \rangle = (v \cup \beta v')[M]$$

on $H^{2r}(M; \mathbb{Z}_2)$. Thus $w_2 w_{4r-1}$ is the invariant described above. We use a relative version of this characterization to prove the present result.

Since $w_2 w_{4r-1}$ is so coarse, one could expect formulae for computing it when M is obtained by geometric constructions from other manifolds. The number w_n is easy to compute for the total space of a fibration – it is multiplicative. Kreck [3] has shown how to compute $w_2 w_{4r-1}$ if M (orientable) fibers over a circle with fiber F. To wit: let V be the free quotient of $H^{2r}(F; Z)$ and let $T: V \to V$ be the action of the fundamental group of the base on V. Define

$$r_\infty = \text{rank}_{\mathbb{Q}}((1-T): V \otimes \mathbb{Q} \to V \otimes Q)$$
$$r_2 = \text{rank}_{\mathbb{Z}_2}((1-T): V \otimes \mathbb{Z}_2 \to V \otimes \mathbb{Z}_2).$$

Kreck's formula is:

$$w_2 w_{4r-1} \equiv r_2 - r_\infty \quad \text{modulo } 2.$$

The theorem of this note is the computation if M fibers over a surface.

The result was proved for even $\chi(B)$ in [1], and we assume that result. Indeed the proof is closely modelled on the proof of [1], and where they are identical, we do not do all the details.

First we do some technical algebra. Thus let V be a finite-dimensional vector space over Z_2 with a symmetric bilinear form to Z_2. We denote this form, or its matrix representation with respect to some basis by $\mathfrak{a}$. Let $H_1, \ldots, H_n, K_1, \ldots, K_n, J: V \to V$ be form-preserving automorphisms of V (or their matrices). We use tilde to denote transposes. Thus $\tilde{H}_i \mathfrak{a} H_i = \mathfrak{a}$, etc. Denote

$$R_p = H_p^{-1} K_p H_p K_p^{-1}$$

(thus K_p^{-1} operates first), and

$$Q_p = R_p R_{p-1} \ldots R_1 \qquad (Q_0 = I = \text{Identity}).$$

Suppose

$$Q_n = J^2.$$

Let U be the direct sum of $2n+1$ copies of V as a vector space (written as block column vectors). We now give U a form. Let α_{pq} be the block matrix:

$$\alpha_{pp} = \begin{pmatrix} \tilde{H}_p \mathfrak{a} + \mathfrak{a} & \tilde{H}_p \mathfrak{a} \\ \tilde{K}_p \mathfrak{a} + \mathfrak{a} + \tilde{K}_p \mathfrak{a} H_p & \tilde{K}_p \mathfrak{a} + \mathfrak{a} \end{pmatrix},$$

$$\alpha_{pq} = \begin{pmatrix} \mathfrak{a} + \tilde{H}_p \mathfrak{a} + \mathfrak{a} H_q + \tilde{H}_p \mathfrak{a} H_q & \mathfrak{a} + \tilde{H}_p \mathfrak{a} + \mathfrak{a} K_q + \tilde{H}_p \mathfrak{a} K_q \\ \mathfrak{a} + \tilde{K}_p \mathfrak{a} + \mathfrak{a} H_q + \tilde{K}_p \mathfrak{a} H_q & \mathfrak{a} + \tilde{K}_p \mathfrak{a} + \mathfrak{a} K_q + \tilde{K}_p \mathfrak{a} K_q \end{pmatrix}$$

$$\alpha_{pq} = 0 \quad \text{if } p > q. \qquad \text{if } p < q,$$

Let $\mathfrak{A}$ be the block matrix

$$\mathfrak{A} = \begin{pmatrix} (\alpha_{pq})_{p,q=1,\ldots,n} & 0 \\ 0 & \mathfrak{a} J \end{pmatrix}.$$

Then $\mathfrak{A}$ is the matrix of a form on U.

Let γ_p be the block matrix

$$\gamma_p = \begin{pmatrix} K_p^{-1} Q_{p-1} + Q_p \\ K_p^{-1} Q_{p-1} + H_p K_p^{-1} Q_{p-1} \end{pmatrix}.$$

Let $G: V \to U$ be the map

$$G = \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_n \\ l + J \end{pmatrix}$$

and $G^*: U \to V$ be the map

$$G^* = (1 + H_1 \quad 1 + K_1 \quad 1 + H_2 \quad 1 + K_2 \ldots 1 + H_n \quad 1 + K_n \quad 1 + J).$$

Let $W_0 = \text{image } G$, $W = \text{kernel } G^*$.

The following technical result contains the algebra of the proof. Fix $n = 0, 1, 2, \ldots$.

Lemma. a) $\mathfrak{A}$ *is non-singular,*
b) G, G^* *are transposes with respect to* $\mathfrak{a}, \mathfrak{A}$,
c) $W_0 \subset W$,
d) $\mathfrak{A} \mid W \otimes W$ *is symmetric,*
e) $W_0 = Rad\, W$,
f) $\dim(W/W_0) \equiv \dim V$ modulo 2.

Proof. a) Note that $\mathfrak{A}$ is upper triangular in block form. We reduce each α_{pp} to an obviously non-singular form by column operations. This will prove (a). Recall $\mathfrak{a} = \tilde{H}_p \mathfrak{a} H_p$. Thus

$$\alpha_{pp} = \begin{pmatrix} \tilde{H}_p \mathfrak{a} + \tilde{H}_p \mathfrak{a} H_p & \tilde{H}_p \mathfrak{a} \\ \tilde{K}_p \mathfrak{a} + \mathfrak{a} + \tilde{K}_p \mathfrak{a} H_p & \tilde{K}_p \mathfrak{a} + \mathfrak{a} \end{pmatrix}.$$

Multiply the second column by $1 + H_p$ on the right and add to the first column. The result is

$$\begin{pmatrix} 0 & \tilde{H}_p \mathfrak{a} \\ \mathfrak{a} H_p & \tilde{K}_p \mathfrak{a} + \mathfrak{a} \end{pmatrix}$$

which is non-singular.

b) We must show

$$\mathfrak{a} G^* = \tilde{G} \mathfrak{A}.$$

This is by induction together with the following straightforward computations, which are left to the reader:

$$\tilde{\gamma}_p \alpha_{pp} = (\tilde{Q}_{p-1} \mathfrak{a} + \tilde{Q}_{p-1} \mathfrak{a} H_p \quad \tilde{Q}_{p-1} \mathfrak{a} + \tilde{Q}_{p-1} \mathfrak{a} K_p),$$

$$\tilde{\gamma}_p \alpha_{pq} = \begin{pmatrix} \tilde{Q}_{p-1} \mathfrak{a} + \tilde{Q}_{p-1} \mathfrak{a} H_q & \tilde{Q}_{p-1} \mathfrak{a} + \tilde{Q}_{p-1} \mathfrak{a} K_q \\ + \tilde{Q}_p \mathfrak{a} + \tilde{Q}_p \mathfrak{a} H_q & + \tilde{Q}_p \mathfrak{a} + \tilde{Q}_p \mathfrak{a} K_q \end{pmatrix}.$$

c) We need to show $G^* G = 0$. By induction we calculate

$$(1 + H_1 \quad 1 + K_1 \ldots 1 + H_k \quad 1 + K_k) \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_k \end{pmatrix} = 1 + Q_k.$$

Therefore $G^* G = 1 + Q_n + 1 + J^2 = 0$.

d) Let

$$G_0^* = (0 \ldots 0 \quad 1 + J),$$
$$G_1^* = G^* + G_0^*.$$

Note that

$$\mathfrak{A} + \tilde{\mathfrak{A}} = \tilde{G}_0^* \mathfrak{a} G_0^* + \tilde{G}_1^* \mathfrak{a} G_1^*.$$

On W, $G_0^* = G_1^*$, thus the right hand side is 0 on W. Thus $\mathfrak{A}$ is symmetric on W.

e) These last two results are formal. Since $\mathfrak{a}, \mathfrak{A}$ are non-singular and G, G^* are transposes, W and U/W_0 are dually paired by $\mathfrak{A}$ [1, Lemma 4.1]. Thus W/W_0 is dually paired to itself.

f) Consider

$$0 \to V \xrightarrow{G} U \xrightarrow{G^*} V \to 0$$

as a chain complex. Its Euler characteristic $\equiv \dim U \equiv \dim V$ modulo 2. Its homology, which has the same Euler characteristic, consists of

$$\textit{kern}\, G,\ W/W_0,\ \textit{cokern}\, G^*.$$

However, kern G and cokern G^* have a dual pairing induced by $\mathfrak{a}$, hence have the same dimensions. Thus $\dim W/W_0 \equiv \dim V$ modulo 2. Thus the lemma is proved.

We turn to the geometry. Any surface B with odd Euler characteristic is the connected sum of tori $T_1, \ldots, T_n$ and a projective plane (cross cap) P. It can be represented as a cell with identifications on the boundary. The cell is pictured in Fig. 1; in Fig. 2 it is represented immersed in R^3, with the identifications made.

Let b_0 be the base point. The identified edges $h_1, k_1, h_2, k_2, \ldots, h_n, k_n, j$ of course represent generators of $\pi_1(B, b_0)$. Suppose M fibers over B, $\pi: M \to B$, and let $F = \pi^{-1}(b_0)$. We use the letters $h_1, k_1, \ldots$ to denote the action of the generators in $\pi_1(B, b_0)$ on F. Let b_1 be a point in the interior of the 2-cell of B and let $F_1 = \pi^{-1}(b_1)$ with inclusions $i_1: F_1 \to M$ and $j_1: (M, \phi) \to (M, F_1)$. All homology and cohomology is with Z_2 coefficients.

As in [1], we consider the bilinear pairing

$$H^{2r}(M, F_1) \otimes H^{2r}(H, F_1) \to H^{4r+1}(M, F_1) = Z_2$$

given by

$$x \otimes x' \to x \cup \beta x'.$$

This is symmetric. Let v be a characteristic element. As in [1], $j_1^* v = v_{2r} \in H^{2r}(M; Z_2)$. Therefore

$$\begin{aligned} w_2 w_{4r-1} &= v_{2r} \cup \beta v_{2r} = j_1^* v \cup \beta j_1^* v \\ &= \dim (H^{2r}(M; F_1)/\operatorname{Rad} H^{2r}(M, F_1)) \qquad \text{modulo 2.} \end{aligned}$$

Let B_0 be a small open disk around b_1 in B and let $\tilde{M} = \pi^{-1}(B - B_0)$. By excision $H^*(M, F_1) = H^*(\tilde{M}, \partial \tilde{M})$. Since Lefshetz duality respects the action of β, we can dualize the problem and investigate intersection products in $H_*(\tilde{M})$. In particular, we consider the intersection product

$$x \otimes x' \to x \cdot \beta x'$$

on $H_{2r+1}(\tilde{M})$ and determine

$$\dim (H_{2r+1}(\tilde{M})/\operatorname{Rad} H_{2r+1}(\tilde{M})).$$

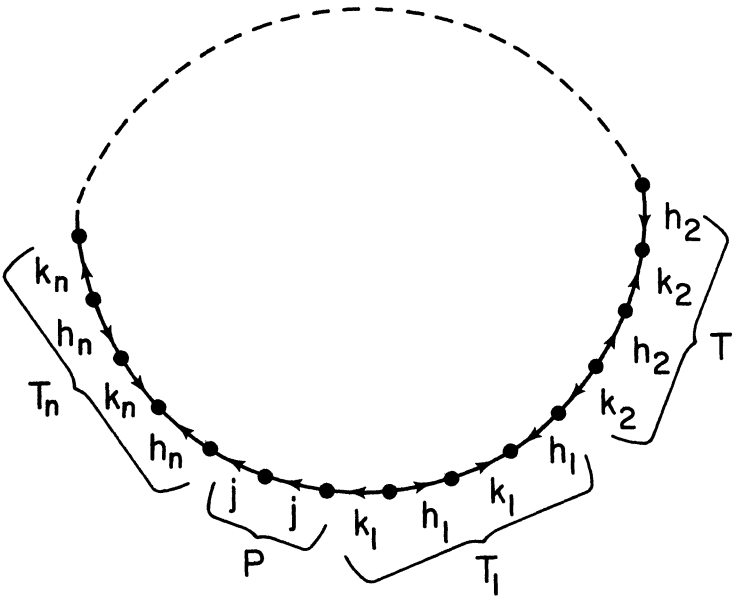

h_2
k_2
h_2
k_2
T
k_n
h_n
T_n
k_n
h_n
h_1
k_1
j
j
k_1
h_1
P
T_1

Fig. 1

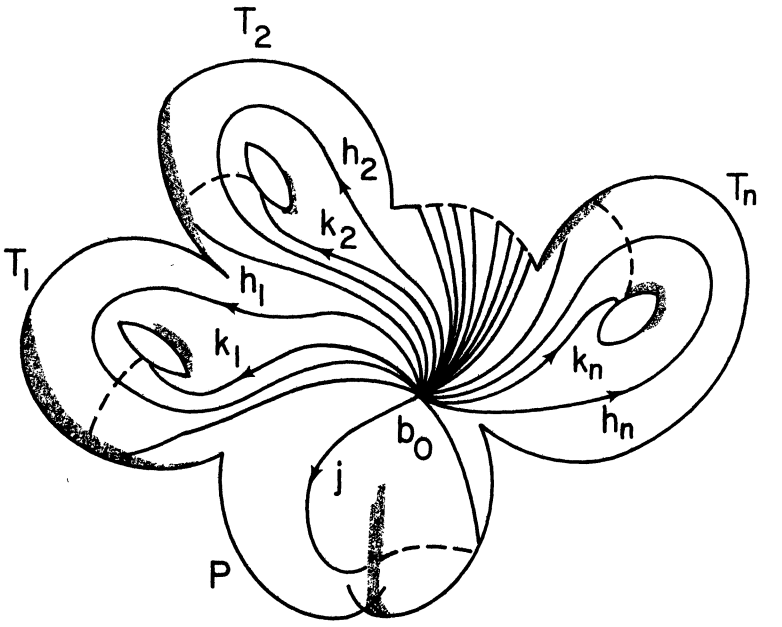

T_2
h_2
k_2
T_n
T_1
h_1
k_1
k_n
h_n
b_0
j
P

Fig. 2

(Remark: Then seem to be no obstructions to proving the result for M a Poincaré duality space; however, it is probably necessary to remain in cohomology and compute in a cellular cochain complex of M.)

Let B_1 be the 1-skeleton of B and let $M_1=\pi^{-1}(B_1)$. Since B_1 is a deformation retract of $B-B_0$, we have that M_1 is a deformation retract of $\tilde{M}$. Thus to determine the intersection product in $\tilde{M}$, we move B_1 to B_1' in general position to B_1, as shown in Fig. 3. Then $M_1'=\pi^{-1}(B_1')$ is in general position to M_1 in M and we can use M_1, M_1' to compute intersection products in $\tilde{M}$ in terms of those in F.

In particular, note that $b_0'\notin M_1$. Thus if $i\colon F\to\tilde{M}$ is the inclusion, for any $x\in$image $i_*\colon H_*(F)\to H_*(\tilde{M})$, we have $x\cdot x'=0$ for all $x'\in H_*(M)$. Therefore, the intersection product induces a well-defined product on $H_*(\tilde{M})/$image i_*.

Consider the exact sequence of the pair (M_1, F):

$$H_{2r+1}(F)\xrightarrow{i_*}H_{2r+1}(M_1)\xrightarrow{j_*}H_{2r+1}(M_1,F)\xrightarrow{\partial}H_{2r}(F).$$

By excision and suspension

$$H_{2r+1}(M_1,F)\xleftarrow[\approx]{\Sigma}\bigoplus_{2n+1}H_{2r}(F).$$

There is a summand of $H_{2r}(F)$ for each 1-cell in B_1. The map Σ is defined as follows. Suppose e is some 1-cell and $a\in H_{2r}(F)$ with a representing cycle $\bar{a}\in C_{2r}(F)$. Let $\bar{e}$ be a lift of e to M_1. Then $\bar{a}\times\bar{e}\in C_{2r+1}(M_1)$ and it is a cycle in $C_{2r+1}(M_1, F)$. The class a in the summand $H_{2r}(F)$ associated to e is sent by Σ to the homology class of $\bar{a}\times\bar{e}$ in $H_{2r+1}(M_1, F)$.

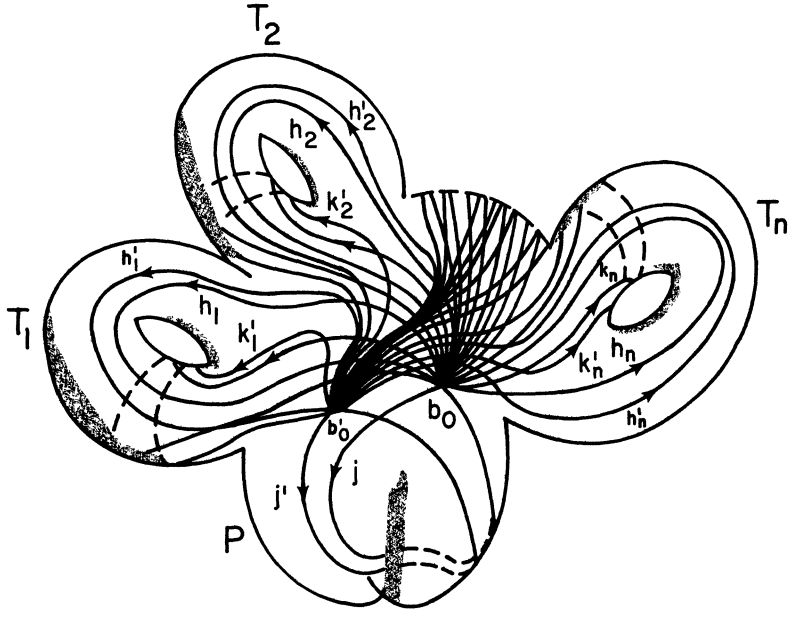


Fig. 3

Denote $H_{2r}(F)=V'$ and $\bigoplus_{2n+1} H_{2r}(F)=U'$. Then the following diagram commutes.

$$\begin{array}{ccccccc} 0 & \longrightarrow & H_{2r+1}(M_1)/\text{image}\, i_* & \xrightarrow{j_*} & H_{2r+1}(M_1, F) & \xrightarrow{\partial} & H_{2r}(F) \\ & & & & \uparrow \Sigma \approx & & \| \\ 0 & \longrightarrow & W' & \longrightarrow & U' & \xrightarrow{G^*} & V' \end{array}$$

where

$$G^*=(1+h_1 \quad 1+k_1 \quad 1+h_2 \quad 1+k_2 \ldots 1+h_n \quad 1+k_n \quad 1+j).$$

Moreover $j_*^{-1}\Sigma\colon \ker G^* \xrightarrow{\approx} H_{2r+1}(\tilde{M})/\text{image}\, i_*$ can be described as follows. Let $\alpha=(\alpha_i)$ be a $(2n+1)$-tuple with $G^*\alpha=0$. Let $\bar{\alpha}=(\bar{\alpha}_i)$ be a $(2n+1)$-tuple of representing chains. Form

$$\Sigma\bar{\alpha}=(\bar{\alpha}_i \times \bar{e}_i)\in C_{2r+1}(M_1)$$

as above. Then $\partial\Sigma\bar{\alpha}$ bounds in $C_{2r}(F)$, say $\partial\Sigma\bar{\alpha}=\partial\bar{\phi}$ for $\bar{\phi}\in C_{2r+1}(F)$. Then $\Sigma\bar{\alpha}+\bar{\phi}$ is a cycle in $C_{2r+1}(M_1)$. A different choice of $\bar{\phi}$ changes the homology class of $\Sigma\bar{\alpha}+\bar{\phi}$ by an element of image i_*. Thus $\Sigma\bar{\alpha}+\bar{\phi}$ represents $j_*^{-1}\Sigma\alpha$ in $H_{2r+1}(M_1)/\text{image}\, i_*$ (or $H_{2r+1}(\tilde{M})/\text{image}\, i_*$). Recall the intersection pairing in which we are interested is defined on this group.

Let $V=H_{2r}(F)/\ker\beta=V'/\ker\beta$ and $U=\bigoplus_{2n+1} V$. Then G^* is still well-defined. Let $\mathfrak{a}$ on V be the form $x\otimes x'\mapsto x\cdot\beta x'$. As explained in [1], a (non-symmetric) intersection pairing on $H_*(M_1, F)$ can be read off from Fig. 3, and $\mathfrak{a}$ induces a pairing on $H_*(M_1, F)/\ker\beta$. Then Σ induces

$$\Sigma\colon U\to H_*(M_1, F)/\ker\beta.$$

Study of Fig. 3 shows if U is given the form $\mathfrak{A}$, then Σ is form preserving and since $\mathfrak{A}$ is non-singular, Σ is an isomorphism. Therefore, $j^{-1}\Sigma$ induces a form-preserving isomorphism from $W\approx\ker G^*$ to

$$(H_{2r+1}(\tilde{M})/\text{image}\, i_*)/\ker(\beta\colon H_{2r+1}(\tilde{M})/\text{image}\, i_*\to H_{2r}(\tilde{M})/\text{image}\, i_*).$$

Hence $W/W_0\approx H_{2r+1}(\tilde{M})/\text{Rad}\, H_{2r+1}(\tilde{M})$.

Thus the lemma applies and we find for v a characteristic element of W, that

$$v\cdot\beta v\equiv\dim(W/W_0)\equiv\dim V=\dim(H_{2r+1}(F)/\ker\beta) \text{ modulo } 2.$$

Since $v_{2r}\cdot\beta v_{2r}=v\cdot\beta v$, the theorem is proved.

References

1. Alexander, J.C., Kahn, S.M.: Characteristic number obstructions to fibering oriented and complex manifolds over surfaces. Topology **19**, 265–282 (1980)

2. Chess, D.: A Poincaré-Hopf type theorem for the deRham invariant. Bull. Am. Math. Soc. **3**, 1031–1036 (1980)
3. Kreck, M.: Bordismusgruppen von Diffeomorphismen, Habilitationschrift, Universität Bonn (1976), (correction, private correspondence to W.D. Neumann, Jan. 3, 1979)
4. Lusztig, G., Milnor, J.W., Peterson, F.P.: Semi-characteristics and cobordism. Topology **8**, 357–360 (1969)
5. Morgan, J.: A product formula for surgery obstructions. Mem. Am. Math. Soc. **14** (1975)
6. Morgan, J., Sullivan, D.: The transversality characteristic class and linking cycles in surgery theory. Ann. Math. **99**, 463–544 (1974)
7. deRham, G.: Sur l'analysis situs de variétés a n dimensions. J. Math. Pures Appl. **10**, 115–200 (1931)

Received November 14, 1980

Math. Ann. 256, 439–447 (1981)

Mathematische Annalen
© Springer-Verlag 1981

The Last Possible Place of Unitarity for Certain Highest Weight Modules

Hans Plesner Jakobsen*

Mathematics Institute, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen ϕ, Denmark

Introduction

The modules in question are those corresponding to holomorphically induced representations on Hermitian symmetric spaces of the non-compact type. Specifically let $\mathfrak{g}$ be a simple Lie algebra over $\mathbb{R}$ and let $\mathfrak{g}=\mathfrak{k}+\mathfrak{p}$ be a Cartan decomposition. By assumption $\mathfrak{k}$ has a non-trivial center $\eta=\mathbb{R}\cdot h_0$ and $\mathfrak{k}=\mathfrak{k}_1\oplus\mathbb{R}\cdot h_0$ where $\mathfrak{k}_1=[\mathfrak{k},\mathfrak{k}]$. The modules W_Λ considered are determined by a pair (Λ_0,λ) where Λ_0 is $\mathfrak{k}_1$-dominant and integral and $\lambda\in\mathbb{R}$. That is, $\Lambda=(\Lambda_0,\lambda)$ determines a finite-dimensional $\mathcal{U}(\mathfrak{k}^{\mathbb{C}})$-module V_Λ and W_Λ is the irreducible quotient of $\mathcal{U}(\mathfrak{g}^{\mathbb{C}})\underset{\mathcal{U}(\mathfrak{k}^{\mathbb{C}}+\mathfrak{p}^+)}{\otimes}V_\Lambda$, where $\mathfrak{p}^+=\{z\in\mathfrak{p}^{\mathbb{C}}|[h_0,z]=iz\}$.

W_Λ may be represented as a space of V_Λ-valued polynomials on $\mathfrak{p}^+$ and the $\mathfrak{g}$-invariant Hermitian form on W_Λ, restricted to the space of first order polynomials, depends linearly on λ. For λ sufficiently negative W_Λ is unitarizable and thus, for Λ_0 fixed, the smallest λ such that W_Λ does not contain all first order polynomials determines the last possible place at which W_Λ can be unitarizable.

This philosophy was used in [5] to prove a conjecture of Kashiwara and Vergne, in the case of $\mathrm{SU}(p,q)$, by means of geometrical methods.

In this article our methods are algebraic. Specifically, the main tool is Bernstein-Gelfand-Gelfand [1, Th. 7.6.23] applied in a manner similar to, and motivated by, Shapovalov's in [8]. The technical side of this article, in fact, consists of adapting a theorem of Shapovalov to our situation.

As a corollary we prove the Kashiwara-Vergne conjecture for $\mathrm{Mp}(n,\mathbb{R})$.

Finally we mention that after we had realized that [8] could be applied, but before having completed the proof, a preprint of an article by Enright and Parthasarathy [2] was channelled to us. In their article a criterion for unitarizability is developed and applied to give a proof of the same conjecture. Though their criterion leads to the same result, the philosophy behind it, and in particular the proof of it, is quite different from ours.

* This research was partly supported by a grant from the NSF

0025-5831/81/0256/0439/$01.80

1. Notation

Let $\mathfrak{g}$ be a simple Lie algebra over $\mathbb{R}$ and $\mathfrak{g}=\mathfrak{k}+\mathfrak{p}$ a Cartan decomposition of $\mathfrak{g}$. We assume that $\mathfrak{k}$ has a nonempty center η; in this case $\eta=\mathbb{R}\cdot h_0$ for an $h_0\in\eta$ whose eigenvalues under the adjoint action on $\mathfrak{p}^{\mathbb{C}}$ are $\pm i$. Let

$$\mathfrak{p}^+=\{z\in\mathfrak{p}^{\mathbb{C}}|[h_0,z]=iz\},$$

and

$$\mathfrak{p}^-=\{z\in\mathfrak{p}^{\mathbb{C}}|[h_0,z]=-iz\}.$$

Let $\mathfrak{k}_1=[\mathfrak{k},\mathfrak{k}]$ and let $\mathfrak{h}$ be a maximal abelian subalgebra of $\mathfrak{k}$. Then $\mathfrak{k}=\mathfrak{k}_1\oplus\mathbb{R}\cdot h_0$, $\mathfrak{h}=(\mathfrak{h}\cap\mathfrak{k}_1)\oplus\mathbb{R}\cdot h_0$, $(\mathfrak{h}\cap\mathfrak{k}_1)^{\mathbb{C}}$ is a Cartan subalgebra of $\mathfrak{k}_1^{\mathbb{C}}$, and $\mathfrak{h}^{\mathbb{C}}$ is a Cartan subalgebra of $\mathfrak{g}^{\mathbb{C}}$. We let σ denote the conjugation in $\mathfrak{g}^{\mathbb{C}}$ relative to the real form $\mathfrak{g}$ of $\mathfrak{g}^{\mathbb{C}}$. The sets of compact and non-compact roots of $\mathfrak{g}^{\mathbb{C}}$ relative to $\mathfrak{h}^{\mathbb{C}}$ are denoted Δ_c and Δ_n, respectively. $\Delta=\Delta_c\cup\Delta_n$. We choose an ordering of Δ such that

$$\mathfrak{p}^+=\sum_{\alpha\in\Delta_n^+}\mathfrak{g}^{\alpha},$$

and set

$$\mathfrak{g}^+=\sum_{\alpha\in\Delta^+}\mathfrak{g}^{\alpha},$$

$$\mathfrak{g}^-=\sum_{\alpha\in\Delta^-}\mathfrak{g}^{\alpha},$$

and

$$\varrho=\tfrac{1}{2}\sum_{\alpha\in\Delta^+}\alpha.$$

Throughout β denotes the unique simple non-compact root. For $\gamma\in\Delta$ let H_γ be the unique element of $i\mathfrak{h}\cap[(\mathfrak{g}^{\mathbb{C}})^\gamma,(\mathfrak{g}^{\mathbb{C}})^{-\gamma}]$ for which $\gamma(H_\gamma)=2$. Then for all γ_1 in Δ

$$\langle\gamma_1,\gamma\rangle=\frac{2(\gamma_1,\gamma)}{(\gamma,\gamma)}=\gamma_1(H_\gamma),\tag{1.1}$$

where $(\cdot,\cdot)$ is the bilinear form on $(\mathfrak{h}^{\mathbb{C}})^*$ obtained from the Killing form of $\mathfrak{g}^{\mathbb{C}}$. For $\alpha\in\Delta_n^+$ choose $x_\alpha\in(\mathfrak{g}^{\mathbb{C}})^\alpha$ such that

$$[x_\alpha,x_\alpha^\sigma]=H_\alpha.\tag{1.2}$$

Following the notation of [7] we let γ_r denote the highest root. Then $\gamma_r\in\Delta_n^+$, and $H_{\gamma_r}\notin[\mathfrak{h}\cap\mathfrak{k}_1]^{\mathbb{C}}$.

Finally we let $u\to u^*$ be the antilinear antiautomorphism of $\mathcal{U}(\mathfrak{g}^{\mathbb{C}})$ that extends the map $x\to -x^\sigma$ of $\mathfrak{g}^{\mathbb{C}}$.

2. Modules

Corresponding to the decomposition

$$\mathcal{U}(\mathfrak{g}^{\mathbb{C}})=(\mathcal{U}(\mathfrak{g}^{\mathbb{C}})\mathfrak{g}^+\oplus\mathfrak{g}^-\mathcal{U}(\mathfrak{g}^{\mathbb{C}}))\oplus\mathcal{U}(\mathfrak{h}^{\mathbb{C}})$$

we let, for $u\in\mathscr{U}(\mathfrak{g}^{\mathbb{C}})$, $\gamma(u)$ denote the unique element of $\mathscr{U}(\mathfrak{h}^{\mathbb{C}})$ for which $u-\gamma(u)$ is in $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})\mathfrak{g}^{+}\oplus\mathfrak{g}^{-}\mathscr{U}(\mathfrak{g}^{\mathbb{C}})$.

Let $\chi\in(\mathfrak{h}^{\mathbb{C}})^{*}$. The Verma module M_{χ} of highest weight $\chi-\varrho$ is defined to be $M_{\chi}=\mathscr{U}(\mathfrak{g}^{\mathbb{C}})/I_{\chi-\varrho}$, where $I_{\chi-\varrho}$ is the left ideal generated by the elements $(H-\chi(H)+\varrho(H))$, $H\in\mathfrak{h}^{\mathbb{C}}$, and $\mathfrak{g}^{+}$. We denote the image of 1 in M_{χ} by $1_{\chi-\varrho}$.

If Λ_0 is a dominant integral weight of $\mathfrak{k}_1$ and if $\lambda\in\mathbb{R}$ we denote by $\Lambda=(\Lambda_0,\lambda)$ the linear functional on $\mathfrak{h}^{\mathbb{C}}$ given by

$$\Lambda|_{(\mathfrak{h}\cap\mathfrak{k}_1)^{\mathbb{C}}}=\Lambda_0,\qquad \Lambda(H_{\gamma_r})=\lambda.$$

Further we let V_{Λ} denote the irreducible finite-dimensional $\mathscr{U}(\mathfrak{k}^{\mathbb{C}})$-module of highest weight Λ. As $\mathscr{U}(\mathfrak{k}_1^{\mathbb{C}})$-modules, clearly $V_{\Lambda}=V_{\Lambda_0}$.

The sesquilinear form B_{Λ} on $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})$,

$$B_{\Lambda}(u,v)=\Lambda(\gamma(v^{*}u)) \tag{2.2}$$

is $\mathfrak{g}$-invariant. We let N_{Λ} denote the kernel of B_{Λ};

$$N_{\Lambda}=\{u\in\mathscr{U}(\mathfrak{g}^{\mathbb{C}})\mid\forall v\in\mathscr{U}(\mathfrak{g}^{\mathbb{C}}):\Lambda(\gamma(v^{*}u))=0\}, \tag{2.3}$$

and set

$$N_{\Lambda}(\mathfrak{k})=N_{\Lambda}\cap\mathscr{U}(\mathfrak{k}^{\mathbb{C}}). \tag{2.4}$$

Let $J_{\Lambda}=I_{\Lambda}+\mathscr{U}(\mathfrak{g}^{\mathbb{C}})N_{\Lambda}(\mathfrak{k})$. Since

$$\mathscr{U}(\mathfrak{g}^{\mathbb{C}})=\mathscr{U}(\mathfrak{p}^{-})\mathscr{U}(\mathfrak{k}^{\mathbb{C}})\mathscr{U}(\mathfrak{p}^{+}),$$

and

$$M_{\Lambda+\varrho}=\mathscr{U}(\mathfrak{p}^{-})\mathscr{U}(\mathfrak{k}^{\mathbb{C}})1_{\Lambda}$$

we have

Lemma 2.1. $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})\underset{\mathscr{U}(\mathfrak{k}^{\mathbb{C}}+\mathfrak{p}^{+})}{\otimes}V_{\Lambda}=\mathscr{U}(\mathfrak{g}^{\mathbb{C}})/J_{\Lambda}$.

Definition 2.2. Let $u\in M_{\chi}$. Corresponding to the decomposition $M_{\chi}=\mathscr{U}(\mathfrak{p}^{-})\mathscr{U}(\mathfrak{k}^{\mathbb{C}})1_{\chi-\varrho}$ we define $P^{-}(u)$ to be the projection of u onto $\mathscr{U}(\mathfrak{p}^{-})1_{\chi-\varrho}$.

The unique irreducible quotient W_{Λ} of $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})\underset{\mathscr{U}(\mathfrak{k}^{\mathbb{C}}+\mathfrak{p}^{+})}{\otimes}V_{\Lambda}$ is given as

$$W_{\Lambda}=\mathscr{U}(\mathfrak{g}^{\mathbb{C}})/N_{\Lambda}. \tag{2.5}$$

Any $\mathfrak{g}$-invariant Hermitian form on W_{Λ} is proportional to B_{Λ}.

For further background information we refer to [7; Sect. 1] and [4; Sect. 2].

The situations in which $J_{\Lambda}\neq N_{\Lambda}$ are quite interesting. In point of fact, by looking at a special case where these two ideals are different, we can describe a simple condition that must be fulfilled in order to have B_{Λ} define a positive-definite inner product on W_{Λ}. For Λ_0 fixed this will be a description of the biggest λ for which W_{Λ} possibly can be unitarizable:

Let Λ_0 be fixed. The module $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})\underset{\mathscr{U}(\mathfrak{k}^{\mathbb{C}}\oplus\mathfrak{p}^{+})}{\otimes}V_{\Lambda}$ may be thought of as a space of V_{Λ}-valued polynomials on $\mathfrak{p}^{+}$. Consider in particular the space $\mathscr{P}_1$ of first order polynomials. Under the action of $\mathscr{U}(\mathfrak{k}^{\mathbb{C}})$, $\mathscr{P}_1$ breaks into a finite sum of irreducible

subspaces; $\mathscr{P}_1 = \bigoplus_{i=1}^{n} S_i$. For each i, B_Λ determines a Hermitian $\mathfrak{k}$-invariant form on S_i, and it is clear that this form is proportional to a fixed non-trivial such by a polynomial $(a_i\lambda + b_i)$ where $a_i \neq 0$. Let $\lambda_0 = \min_{i=1,\ldots,n}\left(-\frac{b_i}{a_i}\right)$. Then: W_Λ unitarizable $\Rightarrow \lambda \leqq \lambda_0$.

We shall describe λ_0 more precisely:

3. Bernstein, Gelfand and Gelfand

The celebrated theorem of Bernstein, Gelfand and Gelfand [1, Th. 7.6.23] describes the situations in which one Verma module can be imbedded into another. A special case is the following: If χ in $(\mathfrak{h}^{\mathbb{C}})^*$ satisfies $\chi(H_\alpha) = 1$ for $\alpha \in \Delta^+$, then $M_{\chi - \alpha} \subset M_\chi$. As was shown, among other things, by Shapovalov in [8], this leads to the following.

Proposition 3.1 ([8]). i) *Let $\alpha \in \Delta^+$, $\chi \in (\mathfrak{h}^{\mathbb{C}})^*$, and assume $\chi(H_\alpha) = 1$. Then there exists an element $\Theta_\alpha(\chi) \in \mathscr{U}(\mathfrak{g}^-)^{-\alpha}$ such that*

$$\forall \gamma \in \Delta^+ : [(\mathfrak{g}^{\mathbb{C}})^\gamma, \Theta_\alpha(\chi)] \in I_{\chi - \varrho}.$$

ii) *If $\alpha \in \Delta^+$ is simple, $\Theta_\alpha(\chi) = x_{-\alpha}$, where $x_{-\alpha} = x_\alpha^\sigma$.*

iii) *If $\alpha \in \Delta^+$ has $\varrho(H_\alpha) > 1$ let $\varepsilon \in \Delta^+$ be a simple root such that $\alpha_1 = s_\varepsilon(\alpha) \in \Delta^+$ has $\varrho(H_{\alpha_1}) < \varrho(H_\alpha)$. If χ is integral and $\chi(H_\varepsilon) < 0$, $\Theta_\alpha(\chi)$ is determined by the equation*

$$\Theta_\alpha(\chi) x_{-\varepsilon}^q = x_{-\varepsilon}^{q+p} \Theta_{\alpha_1}(\psi) \tag{3.1}$$

where $\psi = s_\varepsilon(\chi)$, $q = -\chi(H_\varepsilon)$, and $p = \alpha(H_\varepsilon)$.

iv) *In any basis of $\mathscr{U}(\mathfrak{g}^-)$ the coefficients of $\Theta_\alpha(\chi)$ depend polynomially on χ.*

It follows that (3.1) determines $\Theta_\alpha(\chi)$ for any $\chi \in (\mathfrak{h}^{\mathbb{C}})^*$ with $\chi(\alpha) = 1$.

Proposition 3.1 can be applied to the present situation. However, to do so needs some preparation.

4. Concerning $\mathfrak{p}^- \otimes V_{\Lambda_0}$

Proposition 4.1. *Let $\mathbb{C} \cdot v_0$ be the subspace of highest weight in V_{Λ_0}. Any non-zero $\mathscr{U}(\mathfrak{k}_1^{\mathbb{C}})$-invariant subspace of $\mathfrak{p}^- \otimes V_{\Lambda_0}$ contains elements with non-zero coefficients in $\mathfrak{p}^- \otimes \mathbb{C} \cdot v_0$.*

Proof. Let M be an invariant subspace and assume M is perpendicular to $\mathfrak{p}^- \otimes \mathbb{C} \cdot v_0$. Then M is perpendicular to $\mathscr{U}(\mathfrak{k}_1^{\mathbb{C}}) \cdot (\mathfrak{p}^- \otimes \mathbb{C} \cdot v_0) = \mathfrak{p}^- \otimes V_{\Lambda_0}$. □

Corollary 4.2. *A highest weight of $\mathfrak{p}^- \otimes V_{\Lambda_0}$ is of the form $\Lambda_0 - \alpha$ for some $\alpha \in \Delta_n^+$.*

For $\alpha \in \Delta_n^+$ we write $z_{-\alpha}$ for the element x_α^σ defined by (1.2).

Corollary 4.3. *If a highest weight $\Lambda_0 - \alpha$ occurs in $\mathfrak{p}^- \otimes V_{\Lambda_0}$, there are elements $q_{-\mu}$ of $\mathscr{U}(\mathfrak{k}_1^{\mathbb{C}})^{-\mu}$, $\mu \in \Delta_c^+$, such that the corresponding highest weight vector is given as*

$$z_{-\alpha} \otimes v_0 + \sum_{\mu \in \Delta_c^+} z_{-\alpha+\mu} \otimes q_{-\mu} v_0.$$

Corollary 4.4. *There are no multiplicities.*

5. A Special Case in $\mathfrak{g}=\mathfrak{so}(2n-1,2)$

As we shall see below there are two cases that demand special treatment. One of these is a particular non-compact positive root in the root system for $\mathfrak{g}=\mathfrak{so}(2n-1,2)$:

Following the notation of [2] let $e_1,\ldots,e_n$ be the standard orthonormal basis of $\mathbb{R}^n$. The sets of positive roots are

$$\Delta_c^+=\{e_i\pm e_j|2\leqq i<j\leqq n\}\cup\{e_i|2\leqq i\leqq n\},$$

and

$$\Delta_n^+=\{e_1\pm e_j|2\leqq j\leqq n\}\cup\{e_1\}.$$

If $\Lambda=(\lambda_1,\ldots,\lambda_n)$, Λ is $\mathfrak{k}_1$-integral if and only if $\{\lambda_2,\ldots,\lambda_n\}\subseteq\mathbb{Z}$ or $\mathbb{Z}+\frac{1}{2}$. $\varrho=(2n-1/2,2n-3/2,\ldots,\frac{1}{2})$ and Λ is Δ_c^+ dominant if and only if $\lambda_2\geqq\ldots\geqq\lambda_n\geqq 0$. Observe that $\forall\mu\in\Delta_c^+: s_\mu(e_1)=e_1$, i.e. e_1 is Δ_c^+ isolated from the rest of Δ_n^+.

We want to compute $P^-(\Theta_{e_1}(\chi))$ and observe that

$$e_1=s_{e_1-e_2}s_{e_2-e_3}\ldots s_{e_{n-1}-e_n}(e_n).$$

By using one of the standard forms of $\mathfrak{so}(2n-1,2)$ it is easy to see that there are non-zero elements k_{ij}, m_s, and z_t of $\mathfrak{g}^{\mathbb{C}}$ such that

$$\begin{aligned} k_{ij}&\in\mathfrak{g}^{e_j-e_i};\quad 2\leqq i<j\leqq n,\\ m_s&\in\mathfrak{g}^{-e_s};\quad 2\leqq s\leqq n,\\ z_t&\in\mathfrak{g}^{-e_1+e_t};\quad 2\leqq t\leqq n,\end{aligned}$$

and

$$z_1\in\mathfrak{g}^{-e_1}$$

and that they can be normalized is a manner, insignificant for the result, such that

$$\begin{aligned}&[k_{ij},k_{jr}]=k_{ir};\quad [k_{ij},m_n]=m_i\\ &[z_t,k_{ij}]=\delta_{i,t}z_j;\quad [z_t,m_s]=-\delta_{s,t}z_1.\end{aligned}$$

It is then straightforward to compute $P^-(\Theta_{e_1}(\chi))$ (or $\Theta_{e_1}(\chi)$ for that matter). In fact, one needs only pay attention to the terms m_s in the various expressions. We omit the details and give the result:

Proposition 5.1. *Let $\chi_1=\frac{1}{2}$. Up to a non-zero constant*

$$P^-(\Theta_{e_1}(\chi))=\left(\prod_{i=2}^{n}(\chi_i-\chi_1)\right)z_1.$$

Proposition 5.2. *Let $\chi=\Lambda+\varrho$ and assume $\chi_1=\frac{1}{2}$. Then $P^-(\Theta_{e_1}(\chi))\neq 0$ if and only if Λ_0-e_1 is a highest weight for the $\mathcal{U}(\mathfrak{k}_1^{\mathbb{C}})$-module $\mathfrak{p}^-\otimes V_{\Lambda_0}$.*

Proof. If $P^-(\Theta_{e_1}(\chi))\neq 0$ it is clear that Λ_0-e_1 is a highest weight. Suppose $P^-(\Theta_{e_1}(\chi))=0$. Since $\lambda_1=1-n$ this implies that $\lambda_n=0$. But $e_1=(e_1-e_n)+e_n$ and e_n is simple, so if $\lambda_n=0$, Λ_0-e_1 cannot be a highest weight (cf. Corollary 4.3 and the proof of Proposition 7.3 below). □

6. Concerning the Root System

Since $\mathfrak{p}^-$ is a $\mathscr{U}(\mathfrak{l}_1^{\mathbb{C}})$-module with highest weight $-\beta$, the following is obvious.

Lemma 6.1. *Let $\alpha \in \Delta$. The coefficient to β in α is $1, 0,$ or -1.*

Proposition 6.2. *Root strings are of length at most* 3.

Proof. It is clear that strings through roots of equal length are of length at most 2. Equally obvious is the fact that a string $\beta + i\alpha$, $\beta + (i+1)\alpha$, ..., trough β will have either $i = 1$ and $\alpha \in \Delta_c^+$, or $\alpha \in \Delta_n^+$ and $i \in \{-1, -2\}$. In the last case, by Lemma 6.1, the string has length at most 3. Let l and s abbreviate long and short. A hypothetical string of length 4 will then either be of the form slls or lssl. Since roots of equal length are conjugate under the Weyl group the first case, however, is ruled out because β is long. The only case to be examined, then, is the string β, $\beta + \alpha$, $\beta + 2\alpha$, $\beta + 3\alpha$ with $\alpha \in \Delta_c^+$. In this case $\langle \beta, \alpha \rangle = -3$ and thus $\langle \alpha, \beta \rangle = -1$. Hence $s_\beta(\beta + 3\alpha) = 2\beta + 3\alpha$. Contradiction. □

Proposition 6.3. *Let $\varepsilon_1, \ldots, \varepsilon_t$ denote the set of positive simple roots in Δ_c^+, let $\alpha \in \Delta_n^+$, and assume $\varrho(H_\alpha) > 1$. If $\forall i : \alpha(H_{\varepsilon_i}) \leqq 0$ then either $\mathfrak{g} = \mathfrak{so}(2n-1, 2)$ and $\forall i : \varepsilon_i(H_\alpha) = 0$ or $\mathfrak{g} = \mathfrak{sp}(n, \mathbb{R})$ and $\alpha = \beta + \varepsilon_i$ for a unique ε_i.*

Proof. The assumptions on α imply that $\langle \alpha, \beta \rangle > 0$. Then, by Lemma 6.1, $\langle \alpha, \beta \rangle > = 1$. Write $\alpha = \alpha_1 + \beta$ with $\alpha_1 \in \Delta_c^+$ and assume $-q = \langle \alpha, \alpha_1 \rangle < 0$. Then $s_{\alpha_1}(\alpha) = \alpha + q\alpha_1 \in \Delta_n^+$ and we have got a string β, $\beta + \alpha_1$, $\beta + (q+1)\alpha_1$. By Proposition 6.2 $q = 1$ and thus $\beta + \alpha_1$ and $\beta + 2\alpha_1$ are of equal length. However, this implies that $\beta + 3\alpha_1$ should be a root and thus $\langle \alpha, \alpha_1 \rangle = 0$. This fact combined with $\langle \alpha, \beta \rangle = 1$ easily implies that $(\alpha, \alpha) = \frac{1}{2}(\beta, \beta)$. In other words, we are in a Hermitian symmetric space with two root lengths, that is, $\mathfrak{g} = \mathfrak{so}(2n-1, 2)$ or $\mathfrak{g} = \mathfrak{sp}(n, \mathbb{R})$. The rest is then standard. □

Finally we shall also need the following elementary fact about general root systems:

Lemma 6.4. *Let α_0 be a positive root with $\varrho(H_{\alpha_0}) > 1$. Let $\varepsilon_1, \varepsilon_2, \ldots$* be simple roots and define $\alpha_j = s_{\varepsilon_j}(\alpha_{j-1})$ *for $j = 1, 2, \ldots$. If for all $j : \alpha_j \in \Delta^+$ and $\varrho(H_{\alpha_j}) < \varrho(H_{\alpha_{j-1}})$, then $s_{\varepsilon_1} \ldots s_{\varepsilon_{j-1}}(\varepsilon_j) \in \Delta^+$ for $j = 2, \ldots$.*

Proof. By assumption $\langle \alpha_{j-1}, \varepsilon_j \rangle$ is positive. Thus $\langle s_{\varepsilon_{j-1}}(\alpha_{j-2}), \varepsilon_j \rangle > 0$. Since also $\langle \alpha_{j-2}, \varepsilon_{j-1} \rangle > 0$ it follows that $\varepsilon_{j-1} \neq \varepsilon_j$. Hence [cf. 3, p. 50] $s_{\varepsilon_{j-1}}(\varepsilon_j) \in \Delta^+$ and $\langle \alpha_{j-2}, s_{\varepsilon_{j-1}}(\varepsilon_j) \rangle > 0$. Now write $\alpha_{j-2} = s_{\varepsilon_{j-2}}(\alpha_{j-3})$ and proceed analogously. □

7. The General Case

Proposition 7.1. *Let α be a positive non-compact root with $\varrho(H_\alpha) > 1$ and let $\chi \in (\mathfrak{h}^{\mathbb{C}})^*$ be integral with $\chi(H_\alpha) = 1$. Assume the existence of a simple compact root ε such that $\varrho(H_{\alpha_1}) < \varrho(H_\alpha)$ for $\alpha_1 = s_\varepsilon(\alpha)$, and such that $\chi(H_\varepsilon) < 0$. Let $p = \alpha(H_\varepsilon)$, $q = -\chi(H_\varepsilon)$, and $\psi = s_\varepsilon(\chi)$ (cf. Sect. 3). If $d_\alpha(\chi)$ is defined by the equation $P^-(\Theta_\alpha(\chi)) = d_\alpha(\chi) z_{-\alpha}$, and $d_{\alpha_1}(\psi)$ analogously,*

$$d_\alpha(\chi) = c(\alpha, \alpha_1) \binom{q+p}{p} d_{\alpha_1}(\psi) \tag{7.1}$$

where $c(\alpha, \alpha_1)$ is a non-zero constant.

Proof. $\Theta_{\alpha_1}(\psi)$ is a sum of terms of the form $z_{-\alpha_i}u_{-\mu_i}$ where $u_{-\mu_i}\in\mathscr{U}(\mathfrak{f}_1^{\mathbb{C}})^{-\mu_i}$ and $\alpha_i+\mu_i=\alpha_1$. Obviously

$$x_{-\varepsilon}^{q+p}(z_{-\alpha_i}u_{-\mu_i})=\sum_{s=0}^{q+p}\binom{q+p}{s}z_{-\alpha_i-s\varepsilon}u_{-\mu_i-(q+p-s)\varepsilon}$$

for some elements $u_{-\mu_i-(q+p-s)\varepsilon}\in\mathscr{U}(\mathfrak{f}_1^{\mathbb{C}})^{-\mu_i-(q+p-s)\varepsilon}$. To obtain a term of the form $z_{-\alpha}u_{-q\cdot\varepsilon}$ we need $\mu_i=(s-p)\varepsilon$. Thus $s\geqq p$. Assume $s>p$. Since $\alpha=\alpha_i+s\varepsilon$ and $\alpha_1=\alpha-p\varepsilon=\alpha_i+(s-p)\varepsilon$ we have a string from α_i to $\alpha_i+s\varepsilon$. Hence, by Proposition 6.2, $s=2$, $p=1$. But α_1 and α are of equal length and we reach the contradiction that $\langle\alpha_i,\varepsilon\rangle=-3$. Hence $s=p$, and $\mu_i=0$. Finally, since $-\alpha_1-s\varepsilon$ is a root for $s=0,1,\ldots,p$, $[\mathfrak{g}^{-\alpha_1-s\varepsilon},\mathfrak{g}^{-\varepsilon}]=\mathfrak{g}^{-\alpha_1-(s+1)\varepsilon}$ for $s=0,\ldots,p-1$, so

$$x_{-\varepsilon}^{q+p}z_{-\alpha_1}=\sum_{s=0}^{p}c_s\binom{q+p}{s}z_{-\alpha_1-s\varepsilon}x_{-\varepsilon}^{q+p-s}$$

for some non-zero constants c_s, and (7.1) follows. □

Analogously it follows that

Proposition 7.2. *If $\alpha=\beta+\varepsilon$ with ε simple and compact, and if $s_\beta(\alpha)=\varepsilon$, then*

$$P^-(\Theta_\alpha(\chi))=c(\alpha)\chi(H_\beta)z_{-\alpha},$$

where $c(\alpha)$ is a non-zero constant.

Proposition 7.3. *Let Λ_0 be $\mathfrak{k}_1$-dominant and integral and put $\chi=\Lambda+\varrho$. Let $\alpha\in\Delta_n^+$ and assume $\chi(H_\alpha)=1$. Then $P^-(\Theta_\alpha(\chi))\neq 0$ if and only if $\Lambda_0-\alpha$ is a highest weight in the $\mathscr{U}(\mathfrak{f}_1^{\mathbb{C}})$-module $\mathfrak{p}^-\otimes V_{\Lambda_0}$.*

Proof. If $P^-(\Theta_\alpha(\chi))\neq 0$, $\Theta_\alpha(\chi)$ does not belong to the ideal generated by $N_\Lambda(\mathfrak{k})$ and thus projects onto a non-zero element of $\mathscr{U}(\mathfrak{g}^{\mathbb{C}})/J_\Lambda$. It then follows from Proposition 3.1 *i*) and Lemma 2.1 that $\Lambda_0-\alpha$ is a highest weight in $\mathfrak{p}^-\otimes V_{\Lambda_0}$.

Now consider the converse, and assume that $\varepsilon_1,\ldots,\varepsilon_j$ are simple compact roots such that if

$$\alpha_i=s_{\varepsilon_i}(\alpha_{i-1});\qquad \alpha_0=\alpha,\quad\text{then}\quad \varrho(H_{\alpha_i})<\varrho(H_{\alpha_{i-1}})$$

for $i=1,\ldots,j$. Analogously let $\chi_0=\chi$ and $\chi_i=s_{\varepsilon_i}(\chi_{i-1})$. By (7.1) there exists a non-zero constant $K_j(\alpha)$ such that

$$d_\alpha(\chi)=K_j(\alpha)d_{\alpha_j}(\chi_j)\prod_{i=0}^{j-1}\binom{q_i+p_i}{p_i}$$

where $q_i=-\chi_i(\varepsilon_{i+1})$ and $p_i=\alpha_i(H_{\varepsilon_{i+1}})$. We observe that $\chi_j(\alpha_j)=1$. (It can also be shown that $p_0=p_1=\ldots=p_{j-1}$.) From Lemma 6.4, the assumption that $\Lambda_0-\alpha$ is dominant, and (2.1) we conclude:

$$\begin{aligned}-q_i=\chi_i(H_{\varepsilon_{i+1}})&=(s_{\varepsilon_i}\ldots s_{\varepsilon_1}(\Lambda+\varrho))(H_{\varepsilon_{i+1}})\\&\geqq 1+(s_{\varepsilon_i}\ldots s_{\varepsilon_1}(\Lambda_0))(H_{\varepsilon_{i+1}})\\&\geqq 1+(s_{\varepsilon_i}\ldots s_{\varepsilon_1}(\Lambda_0-\alpha))(H_{\varepsilon_{i+1}})+\alpha_i(H_{\varepsilon_{i+1}})\\&\geqq 1+p_i.\end{aligned}$$

Thus, for $r=1,\ldots,p_i$, $q_i+r\leqq(r-1)-p_i\leqq -1$. Thus $\binom{q_i+p_i}{p_i}\neq 0$ for $i=0,\ldots,j-1$.

Suppose that $\alpha_j=\beta+\varepsilon$ for some simple compact root ε and that $s_\beta(\alpha_j)=\varepsilon$. Then, by Proposition 7.2, $d_{\alpha_j}(\chi_j)=c\chi_j(H_\beta)$, with $c\neq 0$. For the first time we now use the assumption that $\Lambda_0-\alpha$ not only is dominant, but also is a highest weight on $\mathfrak{p}^-\otimes V_{\Lambda_0}$: Suppose $\chi_j(H_\beta)=0$. Since $\beta=\alpha_j-\varepsilon$ and α_j and ε are of equal length, $\chi_j(H_\varepsilon)=1$ ($\chi_j(H_{\alpha_j})=1$). Let $\varepsilon_0=s_{\varepsilon_1}\ldots s_{\varepsilon_j}(\varepsilon)$ and $\gamma=s_{\varepsilon_1}\ldots s_{\varepsilon_j}(\beta)$. Then $\alpha=\gamma+\varepsilon_0$ and $(\Lambda+\varrho)(H_{\varepsilon_0})=1$. Since ε_0 is compact it follows that ε_0 is simple and $\Lambda_0(H_{\varepsilon_0})=0$. Let $\mathbb{C}\cdot v_0$ denote the subspace of highest weight in V_{Λ_0}. According to Corollary 4.3, if $\Lambda_0-\alpha$ is a highest weight the corresponding subspace is given as $\mathbb{C}(z_{-\alpha}\otimes v_0+\sum_{\mu\in\Delta_c^+} z_{-\alpha+\mu}\otimes q_{-\mu}v_0)$. But ε_0 is simple and $x_{-\varepsilon_0}v_0=0$. Thus, the coefficient in $\mathfrak{p}^-\otimes v_0$ of $x_{\varepsilon_0}(z_{-\alpha}\otimes v_0+\sum_{\mu\in\Delta_c^+} z_{-\alpha+\mu}\otimes q_{-\mu}v_0)$ is given as $[x_{\varepsilon_0},z_{-\alpha}]\otimes v_0$ $=Kz_{-\gamma}\otimes v_0$ with $K\neq 0$. It follows from Proposition 6.3 and Proposition 5.2 that this contradiction completes the proof. □

8. A Criterion

If Λ_0 is $\mathfrak{k}_1$-dominant and integral and if $\alpha\in\Delta_n^+$ the equation $(\Lambda+\varrho)(H_\alpha)=1$, with Λ_0 fixed, has a unique solution in λ. It follows from Proposition 3.1 that for this $\lambda=\lambda_\alpha$, $\Theta_\alpha(\Lambda+\varrho)\in N_\Lambda$. Thus, by Proposition 7.3, we arrive at the following criterion:

Proposition 8.1. *Let $\Lambda_0-\alpha_1,\ldots,\Lambda_0-\alpha_t$ be the set of highest weights in the $\mathcal{U}(\mathfrak{k}_1^{\mathbb{C}})$-module $\mathfrak{p}^-\otimes V_{\Lambda_0}$; $\alpha_1,\ldots,\alpha_t\in\Delta_n^+$. Let, for $i=1,\ldots,t$, λ_i be determined by the equation $((\Lambda_0,\lambda_i)+\varrho)(H_{\alpha_i})=1$ and let $\lambda_0=\min\{\lambda_1,\ldots,\lambda_t\}$. If $\Lambda=(\Lambda_0,\lambda)$, $\lambda>\lambda_0$, then W_Λ is not unitarizable.*

Remark. With the exception of some easily handled cases in $\mathfrak{so}(2n-1,2)$ and $\mathfrak{sp}(n,\mathbb{R})$, $\Lambda_0-\alpha$ is a highest weight in $\mathfrak{p}^-\otimes V_{\Lambda_0}$ if and only if $\Lambda_0-\alpha$ is dominant.

9. Mp($n,\mathbb{R}$)

Let $e_1,\ldots,e_n$ denote the standard orthonormal basis of $\mathbb{R}^n$. Then

$$\Delta_c^+=\{e_i-e_j|1\leqq i<j\leqq n\},$$

and

$$\Delta_n^+=\{e_i+e_j|1\leqq i<j\leqq n\}\cup\{2e_j|1\leqq j\leqq n\}.$$

$\Lambda=(\lambda_1,\lambda_2,\ldots,\lambda_n)$ is $\mathfrak{k}_1$-integral and dominant if and only if $\lambda_1\geqq\lambda_2\geqq\ldots\geqq\lambda_n$ and $\lambda_i-\lambda_j\in\mathbb{Z}$. Moreover, $\varrho=(n,n-1,\ldots,1)$. Also observe that $\lambda=\lambda_1$.

Let a,b be non-negative integers with $a\geqq b$. Define $\Lambda_{a,b}=(0,\ldots,0,1,\ldots,1,2,\ldots,2)$ where the string of non-zero integers has length a and the string of 2's has length b.

Proposition 9.1. *Let Λ be $\mathfrak{k}_1$-integral and dominant and let a,b be the largest possible integers such that $\Lambda+\Lambda_{a,b}$ is $\mathfrak{k}_1$-dominant. Then $\lambda=-\left(\frac{a+b}{2}\right)$ is the last possible place of unitarity.*

Proof. First observe that it follows from the proof of Proposition 7.3 that the only case in which an $\alpha \in \Delta_n^+$ can satisfy $\Lambda_0 - \alpha$ dominant but $\Lambda_0 - \alpha$ not a highest weight in $\mathfrak{p}^- \otimes V_{\Lambda_0}$ is when $\alpha = e_i + e_{i+1}$, $i = 1, \ldots, n-1$, and $\Lambda_0(H_{e_i - e_{i+1}}) = 0$. It is then straightforward to see that for $\alpha = e_{a+1} + e_{b+1}$, $\Lambda_0 - \alpha$ is a highest weight in $\mathfrak{p}^- \otimes V_{\Lambda_0}$ and that the λ determined by $(\Lambda + \varrho)(H_\alpha) = 1$ is the last possible place of unitarity. Finally, $\lambda_{a+1} = \lambda$ and $\lambda_{b+1} = \lambda - 1$. (This proof also covers the case $a = b$.) □

By direct comparison with [6; Th. 6.9 and Th. 6.13] we then conclude:

Proposition 9.2. *The Kashiwara-Vergne conjecture is true for* $Mp(n, \mathbb{R})$.

References

1. Dixmier, J.: Algebrès enveloppantes. Paris: Gauthier-Villars 1972
2. Enright, T.J., Parthasarathy, R.: A proof of a conjecture of Kashiwara and Vergne. Preprint, 1980
3. Humphreys, J.E.: Introduction to Lie algebras and representation theory. Berlin, Heidelberg, New York: Springer 1972
4. Jakobsen, H.P., Vergne, M.: Restrictions and expansions of holomorphic representations. J. Functional Analysis **34**, 29–53 (1979)
5. Jakobsen, H.P.: On singular holomorphic representations. Invent Math. **62**, 67–78 (1980)
6. Kashiwara, M., Vergne, M.: On the Segal-Shale-Weil representation and harmonic polynomials. Invent. Math. **44**, 1–47 (1978)
7. Rossi, H., Vergne, M.: Analytic continuation of the holomorphic discrete series of a semi-simple Lie group. Acta Math. **136**, 1–59 (1976)
8. Shapovalov, N.N.: On a bilinar form on the universal enveloping algebra of a complex semi-simple Lie algebra. Functional Analysis Appl. **6**, 307–312 (1972)

Received December 17, 1980

Math. Ann. 256, 449–452 (1981)

Mathematische Annalen
© Springer-Verlag 1981

A Note on 2-Fold Branched Covering Spaces of S^3

Taizo Kanenobu

Department of Mathematics, Kobe University, Kobe 657, Japan

1. Introduction

If a knot $K \subset S^3$ is amphicheiral, then the 2-fold covering space of S^3 branched over K is symmetric (has an orientation reversing diffeomorphism). In this note we show that the converse is not true; it solves Problem 1.23 of [2], which is suggested by Montesinos [4].

Theorem. *There exist non-amphicheiral knots whose 2-fold branched covering spaces are symmetric.*

In Sect. 2, we will construct a composite knot satisfying the properties of Theorem, and in Sect. 3, a prime knot.

2. A Composite Knot

Let K_1 and K_2 be two knots. If there exists an orientation preserving homeomorphism of S^3 which maps K_1 onto K_2, then we will write $K_1 \approx K_2$. We will denote by r an orientation reversing homeomorphism of S^3.

Lemma. *Let K_1 and K_2 be two prime knots. If $K_1 \# K_2$ is amphicheiral and $K_1 \not\approx r(K_2)$, then both K_1 and K_2 are amphicheiral.*

Proof. If $K_1 \# K_2$ is amphicheiral, then $K_1 \# K_2 \approx r(K_1 \# K_2) = r(K_1) \# r(K_2)$. From the unique decomposition theorem of knots [8] and $K_1 \not\approx r(K_2)$ it follows that $K_1 \approx r(K_1)$ and $K_2 \approx r(K_2)$. □

Let K_1 be the pretzel knot (see [6] or [11]) of type $(p_1, \ldots, p_i, \ldots, p_j, \ldots, p_n)$, where n is an odd integer $\geqq 5$ and $p_1, p_2, \ldots, p_n$ are distinct odd integers $\geqq 3$. Let K_2 be the pretzel knot of type $(p_1, \ldots, p_j, \ldots, p_i, \ldots, p_n)$, where $i \neq j$. By the theorem in Sect. 2 of [4], the 2-fold covering spaces of S^3 branched over K_1 and K_2 are the same Seifert fiber space $(O, o; 0|0; (p_1, 1), (p_2, 1), \ldots, (p_n, 1))$ [9], or the manifold with the surgery presentation (see [7, Chap. 9]) of Fig. 1.

This manifold is prime by Theorem 7.1 and Lemma 10.2 of [13]. Hence by the main result of [14], K_1 and K_2 are prime knots. By the classification theorem of

0025-5831/81/0256/0449/$01.00

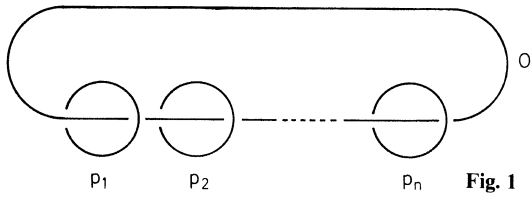


Fig. 1

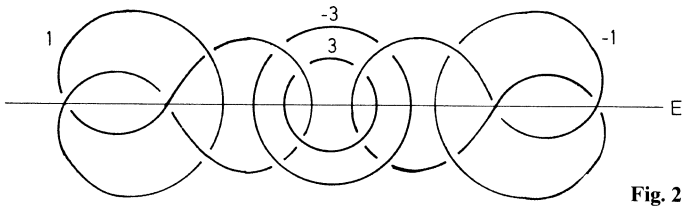


Fig. 2

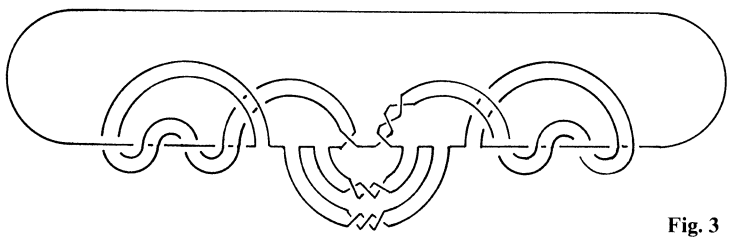


Fig. 3

pretzel knots [6, p. 56], we have $K_1 \not\approx K_2$. Since the signatures of K_1 and K_2 are both $n-1$ ($\neq 0$) [6, p. 71], K_1 and K_2 are non-amphicheiral [7, p. 217]. It follows from Lemma that $K_1 \# r(K_2)$ is non-amphicheiral.

On the other hand, the 2-fold covering space of S^3 branched over $K_1 \# r(K_2)$ is clearly symmetric. Therefore $K_1 \# r(K_2)$ is a desired composite knot.

Remark. We can also construct such a composite knot by using the examples of prime knots in [1] and [10] instead of pretzel knots.

3. A Prime Knot

Let M be the manifold which is obtained by Dehn surgery in the link L according to the diagram of Fig. 2. Because the figure-eight knot, which is a sublink of L, is amphicheiral, it is easily verified that M is symmetric. The link L is strongly-invertible because the symmetry with respect to the axis E leaves L invariant. Using the method of [5], it can be shown that M is a 2-fold covering space branched over the knot K of Fig. 3. Since the signature of K is 16, K is non-amphicheiral. This K is our desired prime knot; the primeness will be proved in Sect. 4.

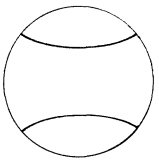


Fig. 4

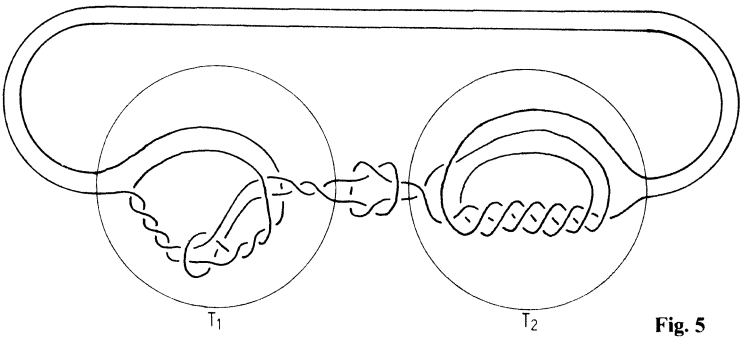


Fig. 5

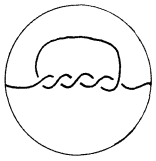


Fig. 6

4. Proof of the Primeness of K

We will prove the primeness of K by the method of Kirby and Lickorish [3]. A *tangle* is a set of disjoint two arcs properly embedded in a 3-ball. The tangle as shown in Fig. 4 is called a *trivial tangle*. A tangle T in a 3-ball B is *prime* if it has the following properties.

(i) Any 2-sphere in B, which meets T transversely in two points, bounds in B a ball meeting T in an unknotted spanning arc.

(ii) The arcs of T cannot be separated by a disc properly embedded in B.

We can see that the knot K is constructed by the tangles T_1 and T_2 as shown in Fig. 5. To prove the primeness of K, it suffices to show that the tangles T_1 and T_2 are prime. If T_i $(i=1,2)$ did not satisfy property (i) or property (ii), then there would be a ball meeting T_i in a knotted spanning arc as shown in Fig. 6. Hence any knot constructed by adding a trivial tangle to T_i would have the torus knot of type (2, 5) as a factor of its prime decomposition. On the other hand, a trivial tangle can be added to T_i to create the knot K_i; K_1 is the pretzel knot of type $(-2,3,7)$, and K_2 is the torus knot of type (3, 7). Let $\Delta_i(t)$ and $\Delta(t)$ be the Alexander polynomials of K_i and the torus knot of type (2, 5) respectively, then $|\Delta_i(-1)|=1$ must be divisible by $|\Delta(-1)|=5$. This contradiction establishes the primeness of K.

Remark. K is concordant to $K_1 \# K_2$.

5. Question

Using the method of Viro [12], we can construct presumably non-amphicheiral knots; they are prime and the 2-fold covering spaces of S^3 branched over them are symmetric. The knot of Fig. 7 is one of them. The 2-fold branched covering space is the manifold as shown in Fig. 8. Is this knot non-amphicheiral?

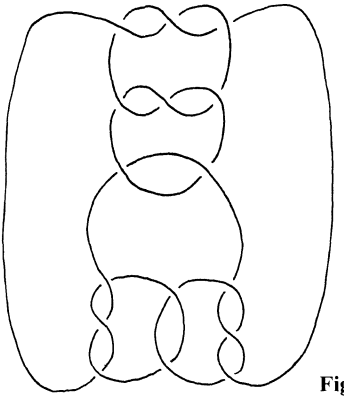


Fig. 7

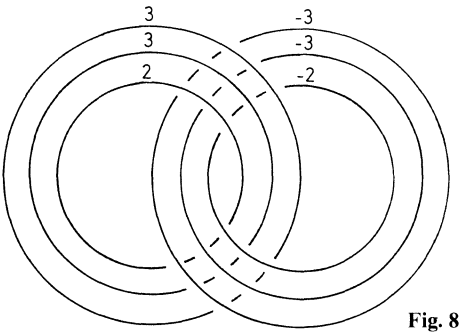


Fig. 8

References

1. Birman, J.S., González-Acuña, F., Montesinos, J.M.: Heegaard splittings of prime 3-manifolds are not unique. Michigan Math. J. **23**, 97–103 (1976)
2. Kirby, R.: Problems in low-dimensional manifold theory. Proc. Symp. Pure Math. XXXII, Part 2, pp. 273–312. Providence R.I.: A.M.S. 1978
3. Kirby, R., Lickorish, W.B.R.: Prime knots and concordance. Math. Proc. Cambridge Philos. Soc. **86**, 437–441 (1979)
4. Montesinos, J.M.: Variedades de Seifert que son recubridores ciclicos ramificados de dos hojas. Bol. Soc. Mat. Méxicana **18**, 1–32 (1973)
5. Montesinos, J.M.: Surgery on links and double branched covers of S^3. Ann. Math. Stud. **84**, 227–259 (1975)
6. Parris, R.L.: Pretzel knots. Princeton University, Ph. D. Thesis 1978
7. Rolfsen, D.: Knots and links. Mathematics Lecture Series 7, Berkeley: Publish or Perish Inc. 1976
8. Schubert, H.: Die eindeutige Zerlegbarkeit eines Knotens in Primknoten. S.-B. Heidelberger Akad. Wiss. Math.-Natur Kl. **3**, 57–104 (1949)
9. Seifert, H.: Topologie dreidimensionaler gefaserter Räume. Acta Math. **60**, 147–238 (1933)
10. Takahashi, M.: Two knots with the same 2-fold branched covering space. Yokohama Math. J. **25**, 91–99 (1977)
11. Trotter, H.F.: Some knots spanned by more than one unknotted surface of minimal genus. Ann. Math. Stud. **84**, 51–62 (1975)
12. Viro, O.Ya.: Nonprojecting isotopies and knots with homeomorphic coverings. J. Soviet Math. **12**, 86–96 (1979)
13. Waldhausen, F.: Eine Klasse von 3-dimensionalen Mannigfaltigkeiten II. Inventiones Math. **4**, 87–117 (1967)
14. Waldhausen, F.: Über Involutionen der 3-Sphäre. Topology **8**, 81–91 (1969)

Received December 22, 1980

Math. Ann. 256, 453–463 (1981)

Mathematische Annalen
© Springer-Verlag 1981

On the Normal Bundles of Smooth Rational Space Curves

D. Eisenbud [1] and A. Van de Ven [2]

1 Department of Mathematics, Brandeis University, Waltham, MA 02254, USA
2 Mathematisch Instituut, Wassenaarseweg 80, Leiden, The Netherlands

I. Introduction

In this note we consider smooth rational curves C of degree n in three-dimensional projective space $\mathbb{P}_3$ (over a closed field of characteristic 0). To avoid trivial exceptions we shall always assume that $n \geqq 4$ (this does not hold however for certain auxiliary curves we shall consider). Let $N = N_C$ be the normal bundle of C in $\mathbb{P}_3$. Since $\deg c_1(\mathbb{P}_3) = 4$, and $\deg c_1(\mathbb{P}_1) = 2$, we have that $\deg c_1(N) = 4n-2$. By a well-known theorem of Grothendieck the bundle N is a direct sum of two line bundles. Hence $N \simeq O_C(2n-1-a) \oplus O_C(2n-1+a)$ for some non-negative $a = a(C)$, which is uniquely determined by C. The question we would like to answer is an obvious one: which values of a occur? We shall show (Theorem 4 below) that a value a occurs if and only if $0 \leqq a \leqq n-4$.

Since for every smooth space curve the normal bundle is generated by global sections we have in any case that $N_C \simeq O_C(m_1) \oplus O_C(m_2)$, with $m_1, m_2 \geqq 0$, therefore $H^1(C, N) = 0$. It follows [K, p. 150] that C represents a smooth point on the Chow variety $\mathrm{Ch}(3, 1, n)$ of effective cycles of dimension 1 and degree n in $\mathbb{P}_3$. Since the set of all smooth rational curves with a fixed degree is obviously connected, we see that the smooth C's represent a smooth, irreducible, $4n$-dimensional (Zariski-)open subset S of $\mathrm{Ch}(3, 1, n)$.

In a forthcoming paper [E-V] we shall prove the following

Theorem. *Let S be the smooth, irreducible $4n$-dimensional (Zariski-) open subset of $\mathrm{Ch}(3, 1, n)$, which parametrises the smooth rational space curves of degree n in $\mathbb{P}_3$. Then there exists a stratification of S by non-empty, locally-closed subsets S_i, $0 \leqq i \leqq n-4$ $\left(\text{i.e. } S = \bigcup_{i=0}^{n-4} S_i \text{ and } S_{i+1} \subset \bar{S}_i\right)$, such that $N_C \simeq O_C(2n-1-i) \oplus O_C(2n-1+i)$ if and only if the point representing C is contained in S_i. For $1 \leqq i \leqq n-4$, the set S_i is irreducible of dimension $4n-2i+1$.*

In particular this implies that for a general curve C the normal bundle N_C is balanced, that is, $N_C \simeq O_C(2n-1) \oplus O_C(2n-1)$ (see also [Ha, Corollary 3]).

0025-5831/81/0256/0453/$02.20

After proving Theorem 4 in Sect. III we turn in Sect. IV to some special topics. If C is a smooth rational curve of degree n, then we prove that C lies on a smooth quadric if and only if the tangent bundle of $\mathbb{P}_3$, restricted to C is isomorphic to $O_C(2n-2)\oplus O_C(n+1)\oplus O_C(n+1)$. On the other hand, if C lies on a smooth quadric, the normal bundle of C is isomorphic to $O_C(2n-1)$ $\oplus O_C(2n-1)$, the same as if C were a general rational curve. (This last fact was also noted by Harris [Ha] as a corollary of a more general result. It will also appear, with a different proof, in the paper [Hu] of Hulek.)

After a first draft of this paper was written, we obtained a preprint of the paper [G-S] by Ghione and Sacchiero, in which the normal bundles of curves with at most ordinary singularities are studied. Ghione and Sacchiero prove a theorem similar to our Theorem 4, except that they allow curves with ordinary singularities and correspondingly must weaken the bound on a to $0\leqq a\leqq n-3$. Since they show that curves with $a=n-3$ lie on quadric cones, they also get the bound $0\leqq a\leqq n-4$ for smooth curves. By contrast, their proof, by example, for the existence of singular curves with all possible values of a, does not, as far as we see, give a systematic production of smooth curves of the given types. The main part of our paper is to give a method for systematically constructing such smooth examples.

II. Some Preliminaries

1. Some Classical Formulae for Space Curves

Let D be a curve in $\mathbb{P}_3$, which is not contained in a plane, let $\tilde{D}$ be the normalisation of D, and $h\colon \tilde{D}\to D$ the canonical projection. Given any point $p\in\tilde{D}$, there exist affine coordinates (ξ_1,ξ_2,ξ_3) in a neighborhood of $h(p)$ $=(0,0,0)$, such that the branch of D, determined by p has a formal parametrisation

$$\begin{aligned}\xi_1&=t^{l_0(p)+1}+(\text{terms of higher order})\\ \xi_2&=t^{l_0(p)+l_1(p)+2}+(\text{terms of higher order})\\ \xi_3&=t^{l_0(p)+l_1(p)+l_2(p)+3}+(\text{terms of higher order})\end{aligned}$$

in a neighborhood of $h(p)$ $(l_0,l_1,l_2\geqq 0)$. The numbers l_0, l_1, and l_2 are independent of the parametrisation thus chosen. We put $l_0(D)=\sum_{p\in\tilde{D}} l_0(p)$ and $l_1(D)$ $=\sum_{p\in\tilde{D}} l_1(p)$. Then there are (compare [P, 3.2]) classical formulae for the degree $\deg(F_D)$ of the tangent (developable) surface F_D of D (that is the surface formed by the tangent lines to D) as well as for the class $r(D)$ of D (the number of osculation planes, passing through a general point of $\mathbb{P}_3$):

$$\begin{aligned}\deg(F_D)&=2\deg(D)+2\Pi(D)-2-l_0(D)\\ r(D)&=3(\deg D+2\Pi(D)-2)-2l_0(D)-l_1(D).\end{aligned}$$

Here $\Pi(D)$ denotes the geometric genus of D (the genus of $\tilde{D}$).

2. Rational Curves with Certain l_0, l_1

Let A_{norm} be a smooth rational curve of degree d in $\mathbb{P}_d$, $d \geqq 4$. We denote by F the tangent surface of A_{norm} and by G the osculation variety of A_{norm} (the union of the osculating two-planes). We have $F \subset G$. If P is a $(d-4)$-dimensional linear subspace of $\mathbb{P}_d$ with $P \cap A_{\text{norm}} = \emptyset$, then A_{norm} projects from P to a rational curve A, of degree d, which spans the $\mathbb{P}_3$ of all $\mathbb{P}_{d-3}$'s passing through P. The following elementary facts can be verified immediately:

(i) if $P \cap G$ (and hence $P \cap F = \emptyset$, then A is a rational curve with $l_0(A) = l_1(A) = 0$;

(ii) if $P \cap F = \emptyset$ and $P \cap G$ consists of one point x, which is smooth on G, such that P and the tangent space to G at x are independent, then $l_0(A) = 0$, $l_1(A) = 1$;

(iii) if $P \cap G$ consists of one point x, contained in F, such that x is smooth on F, and such that P and the tangent space to F at x are independent, then $l_0(A) = 1$, $l_1(A) = 0$.

Since $l_0(A) = l_1(A) = 0$ also holds for a smooth cubic in $\mathbb{P}_3$, we have

Proposition 1. (i) *For every $d \geqq 3$ there are rational curves A of degree d, spanning $\mathbb{P}_3$, with $l_0(A) = l_1(A) = 0$;*

(ii) *for every $d \geqq 4$ there are rational curves A of degree d, spanning $\mathbb{P}_3$, with $l_0(A) = 0$, $l_1(A) = 1$, and also with $l_0(A) = 1$, $l_1(A) = 0$.*

3. A Property of Very Ample Divisors on a Surface

Proposition 2. *Let X be a smooth surface, Y a smooth variety and $f: X \to Y$ a finite map, which maps X birationally onto its image. If D is a very ample divisor on X, and $E \in |D|$ sufficiently general, then $f(E)$ is smooth (and $f|E$ an isomorphism).*

Proof. We have to show the following: if $E \in |D|$ is sufficiently general, then (i) $f|E$ is of maximal rank in every point of E and (ii) $f|E$ is one-to-one.

As to (i) the divisor of E first of all has to avoid all points on X, where f has rank 0. This is easily possible since there is only a finite number of such points. Secondly, E must not be tangent to the kernel of df at any point of X where f has rank 1. But the space of E's which is tangent to the kernel of df in one of these points has codimension 2 in $|D|$; and since the union of all these points consists of finitely many curves minus a finite number of points (where f has rank 0), we see that a general E is not tangent to any of the kernels mentioned before.

To prove (ii), let Z be the closure of the set of points $z \in f(X)$, for which $f^{-1}(z)$ consists of at least two points. Since f is birational, Z is a proper subset of $f(X)$, and therefore consists of finitely many points and curves. A general E certainly does not meet the inverse image of the union of the points. Hence it is enough to show the following: given an irreducible curve $F \subset f(X)$, such that $f: f^{-1}(F) \to F$ is generically k-to-1, with $k \geqq 2$, then the general curve E does not contain any two different points on the same fibre of $f|f^{-1}(F)$. But this is

clear, for the space of elements in $|D|$, containing two different points of such a fibre has codimension two at least.

4. A Property of the Tangent Surface

Proposition 3. *Let C be a space curve (possibly singular), which is not contained in any plane. If F is the tangent surface of C, and $\tilde{F}$ its desingularisation, then the projection from $\tilde{F}$ onto F is a finite map.*

Proof. Let $\tilde{C}$ be the normalisation of C. We can extend the Gauss map, which is a priori defined in all but a finite number of points on $\tilde{C}$, to a regular map $g\colon \tilde{C}\to G$, where G is the Grassmann variety of lines in $\mathbb{P}_3$. Let E be the pull back on $\tilde{C}$ of the $\mathbb{P}_1$-bundle, associated to the universal subbundle on G. The surface E is contained in $G\times\mathbb{P}_3$ in a natural way, and the image of its projection into $\mathbb{P}_3$ is nothing but F. This projection is birational (in general two tangent lines to C don't meet) and finite (the tangent lines to C can't all pass through a fixed point). Hence $E=\tilde{F}$ and the projection from E into $\mathbb{P}_3$ is the projection from $\tilde{F}$ onto F.

III. Proof of the Main Result

This section is dedicated to the proof of

Theorem 4. *Given any integer $n\geqq 4$, there exist smooth rational curves C of degree n in $\mathbb{P}_3$ with normal bundle isomorphic to $O_C(2n-1-a)\oplus O_C(2n-1+a)$ if and only if $|a|\leqq n-4$.*

Proof. Let C be a curve as mentioned in the theorem, and let $N_1\subset N_C$ be isomorphic to $O_C(2n-1+a)$, with $a\geqq 0$. Using the canonical projection from $T_{\mathbb{P}_3}|C$ onto N we obtain from N_1 a subbundle V_1 of $T_{\mathbb{P}_3}|C$. This bundle has rank 2 and $\deg(c_1(V))=2n+a+1$. By a theorem of Nakano [S, p. 265] there exists an exact sequence

$$0\to O_C\otimes O_{\mathbb{P}_3}(-1)\to g^*(U)\to V_1\otimes O_{\mathbb{P}_3}(-1)\to 0,$$

where $g\colon C\to\mathbb{P}_3^*$ is the Gauss map of V_1, and U the universal subbundle on $\mathbb{P}_3^*$. Hence $\deg c_1(g^*(U))=-n+1+a$, and $g^*(O_{\mathbb{P}_3^*}(1))=O_C(n-a-1)$. Since C is not a plane curve, g is not constant and we find already $a\leqq n-2$. If $a=n-2$, then $g(C)$ is a line, in other words, all tangents to C meet the same line, which is impossible. For, projecting C from a general point of this line onto a plane would yield a plane curve, the tangents of which pass through a fixed point; this would mean that this projection is a line and C a plane curve. And in the case $a=n-3$ the image $g(C)$ would be either a line or a conic. The first case can be excluded in the same way as before. If $g(C)$ is a conic, then all the tangents to C are contained in tangent planes to a quadratic cone Q. They can't all pass through the vertex v, hence we can project from v (almost all of) these tangents onto a general plane H, and find that all projections are tangent to the conic of intersection $Q\cap H$. This implies that $C\subset Q$. But on Q there are no smooth rational curves of degree $\geqq 4$.

Now let $a \in Z$, with $0 \leqq a \leqq n-4$. We shall construct a smooth rational curve C of degree $n\,(\geqq 4)$ with $a(C)=a$. We distinguish between three cases:

(1) $n-a \equiv 1(3)$. By Proposition 1 there exists a rational curve A of degree $d = \dfrac{n-a+5}{3}$, spanning $\mathbb{P}_3$, with $l_0(A)=l_1(A)=0$. Let F be the tangent surface of A, let $\tilde{F}$ be its desingularisation and $p\colon \tilde{F} \to F$ the projection. We know by II.4 that $\tilde{F}$ is a $\mathbb{P}_1$-bundle over $\mathbb{P}_1$. Let $\tilde{A}$ be the curve of tangent points on $\tilde{F}$, i.e. the intersection of $\tilde{A}$ with a fibre $\tilde{V}$ consists of all those points, where $p(\tilde{V})$ is tangent to A. Since for a general $\tilde{V}$ the line $p(\tilde{V})$ is tangent to A in exactly one point, $\tilde{A}$ is a section in $\tilde{F}$ (hence $\tilde{A}$ can be identified with the normalisation of A in an obvious way and our notation remains reasonably consistent). Then $p^*(O_F(1)) = [\tilde{A} + \lambda \tilde{V}]$ for some $\lambda \in Z$. Using the fact that $\deg F = 2d-2$ (see II.1) we find

$$(\tilde{A}+\lambda\tilde{V})^2 = 2d-2$$
$$\tilde{A}(\tilde{A}+\lambda\tilde{V}) = d,$$

hence $\tilde{A}^2 = 2$ and $\lambda = d-2$. Therefore $\tilde{F}$ is either isomorphic to $\mathbb{P}_1 \times \mathbb{P}_1$ or Σ_2*, and $\tilde{A}+k\tilde{V}$ is very ample as soon as $k \geqq 1$. So if we take $k=n-d$, we have that $\tilde{A}+k\tilde{V}$ is very ample and $(\tilde{A}+k\tilde{V})(\tilde{A}+(d-2)\tilde{V}) = n$. In other words, a general curve $\tilde{C}$ in $|\tilde{A}+k\tilde{V}|$ projects onto a curve C of degree n in $\mathbb{P}_3$. This curve is smooth by Proposition 2, for p is finite by Proposition 3. We also know that $p|\tilde{C}\colon \tilde{C} \to C$ is an isomorphism, and we may furthermore assume that F is smooth in a general point of C.

We define a map $a\colon C \to A$ as follows: for any point $c \in C$ we set $a(c)$ equal to the image in A of the point in $\tilde{A}$ lying on the fiber of $\tilde{F}$ through $(p|C)^{-1}(c)$. Less formally, we may describe this map by saying that for general c the tangent line to A in $a(c)$ meets C in c.

Next we define $f\colon C \to \mathbb{P}_3^*$ by setting $f(c)$ equal to the osculating plane of A in $a(C)$. By elementary differential geometry we see that for general c, $f(c)$ is the tangent plane to F at c; hence for any c, $f(c)$ contains the tangent line to C at c. In this way, f defines a 2-sub-bundle V_1 of $T_{\mathbb{P}_3}|C$ which contains T_C as a sub-bundle. Using the Nakano sequence and the fact that $r(A) = 3d-6$ (by II.1), we find that $\deg(c_1(V_1)) = 2n+a+1$. Thus V_1 projects onto a 1-dimensional subbundle of N with degree $2n-1+a$.

(2) $n-a \equiv 0(3)$. In this case we proceed exactly in the same way as in case (1), starting from a curve A of degree $d = \dfrac{n-a+6}{3}$.

(3) $n-a \equiv 2(3)$. Again we proceed as before, now taking $d = \dfrac{n-a+7}{3}$.

We shall show in $[E-V]$ that the geometric situation described above is in fact typical; *all* curves in S_i can be obtained by the preceding construction, as divisors on developable surfaces or on cones over curves of suitable class.

---

* This is true for every smooth rational curve, as follows from the Nakano sequence

IV. Some Special Results

1. Curves on Quadrics

From many points of view the simplest space curves are those that lie on a quadric surface. Since there are no smooth rational curves of degree $\geqq 4$ on a quadratic cone, we restrict our attention to the smooth quadrics, which are isomorphic to $\mathbb{P}_1 \times \mathbb{P}_1$. The smooth rational curves of degree n on such a surface are the graphs of $(n-1)$-to-1 maps from $\mathbb{P}_1$ to $\mathbb{P}_1$; they lie in the divisor class $n-1$, 1. Since they have infinitely many $(n-1)$-secants, whereas the general rational curve of degree 5 has only one quadrisecant, and the general rational curve of degree $n \geqq 6$ has no $(n-1)$-secants at all, they are very special.

We shall show first that this specialness reflects itself in the restriction of the bundle $T_{\mathbb{P}_3}$ to the curve:

Proposition 5. *A smooth rational space-curve C of degree $n \geqq 3$ is contained in a smooth quadric if and only if $T_{\mathbb{P}_3}|C \simeq O_C(2n-2) \oplus O_C(n+1) \oplus O_C(n+1)$.*

Proof. If C is contained in a smooth quadric Q, then, by the adjunction formula, C is of type $(n-1, 1)$. The pencil of lines on Q, which intersect C in one point, gives rise to 1-dimensional subbundle L of $T_{\mathbb{P}_3}|C$, which has degree $2n-2$ by Nakano's theorem. Let L_p be the 1-subbundle of $T_{\mathbb{P}_3}|C$, which comes from joining the points of C with a fixed point $p \in C$ (in p itself you have to take the tangent line to C). Again by Nakano's theorem we have that $\deg(c_1(L_p)) = n+1$. Given any point $x \in C$ and a 2-dimensional subspace H of $T_{\mathbb{P}_3}(x)$, we can find a bundle L_p, such that $L_p(x) \not\subset H$. Thus a direct sum of bundles of the form L_p will map onto $T_{\mathbb{P}_3}|C$. Therefore, given any decomposition $T_{\mathbb{P}_3}|C \simeq O_C(a_1) \oplus O_C(a_2) \oplus O_C(a_3)$, each of a_1, a_2, and a_3 must be at least $n+1$. On the other hand the existence of L implies that at least one of the a_i's is $2n-2$ or more. Since $a_1 + a_2 + a_3 = 4n$, this leaves us with only the possibility, $T_{\mathbb{P}_3}|C \simeq O_C(2n-2) \oplus O_C(n+1) \oplus O_C(n+1)$, where the first summand is identical with L as soon as $n \geqq 4$.

Conversely, if $T_{\mathbb{P}_3}|C$ contains a direct summand $L \simeq O_C(2n-2)$, then L gives rise to a surface S by way of the Gauss map. The surface S has degree 1 or 2; since C is not contained in a plane, S is a quadric. Further, S can't be a cone if $n \geqq 4$, for there are no smooth rational curves of degree $\geqq 4$ on a quadratic cone. Therefore, if $d \geqq 4$, the curve C is contained in a quadric, such that the Gauss images of the fibres of L form one of the line systems on the quadric. Since every space cubic is contained in a smooth quadric, we have proved the proposition.

On the other hand, the normal bundle of a space-curve on a smooth quadric is the same as that of the general curve:

Proposition 6. *A smooth rational space-curve C of degree $n \geqq 3$ which is contained in a smooth quadric has normal bundle*

$$N_C \simeq O(2n-1) \oplus O(2n-1).$$

This result was first noted by Harris [Ha]. In [H], Hulek analyses the normal bundles of all smooth curves on quadrics in $\mathbb{P}_3$, and in particular obtains a new proof of our result.

We shall give a geometric proof, and then an algebraic proof, which makes the result a consequence of local duality.

Proof of Proposition 6. Since every smooth rational curve of degree 4 has a balanced normal bundle (Theorem 4), we can assume that $n \geqq 5$.

Again, let L be the 1-subbundle of $T_{\mathbb{P}_3}|C$, given by the pencil of lines on the quadric Q, which intersect C in one point. Since such a line is never tangent to C, we obtain, by projecting L into N, a 1-subbundle L' of N with $\deg(c_1(L')) = 2n-2$. Consequently, either N is balanced, or $N \simeq O_C(2n) \oplus O_C(2n-2)$. All we have to do is to exclude the second possibility.

Suppose that N has a 1-subbundle $M \simeq O_C(2n)$ (such a bundle is necessarily unique). Then $N = L' \oplus M$, and the inverse image of M in $T_{\mathbb{P}_3}|C$ is a 2-bundle V with $\deg(c_1(V)) = 2n+2$. It follows from Proposition 5 that $V \simeq O_C(n+1) \oplus O_C(n+1)$. So $T_{\mathbb{P}_3}|C$ has a family $\{L_t\}_{t \in P_1}$ of 1-subbundles isomorphic to $O_C(n+1)$. By way of its Gauss map each of these surfaces yields an irreducible ruled surface W_t. The degree of W_t is either $n-1$ or at most $\frac{1}{2}(n-1)$. In the second case W_t has to be identical with Q, for if an irreducible surface of degree $\leqq \frac{1}{2}(n-1)$ and an irreducible quadric have a curve of degree n in common, they must coincide. But W_t can't be identical with Q, for in a general point $c \in C$ the line $g_{L_t}(c)$ is not contained in Q. Therefore all surfaces W_t are of degree $n-1$. They intersect Q in a divisor of type $(n-1, n-1)$, the support of which consists of at most two irreducible curves [a line $g_{L_t}(c)$, which is not contained in Q, intersects Q in c and one other point]. Hence $Q \cap W_t = C \cup (n-2)R_t$, where R_t is a line on Q. In this way we obtain a $1-1$-correspondence between the subbundles L_t (or the surfaces W_t) and the pencil of lines on Q, for which $R_t C = n-1$. We take t general, such that in particular R_t and C meet transversally in $n-1$ points. The Gauss map g_{L_t} gives a map $f\colon C \to R_t$. This map can't be constant; otherwise W_t would be a cone with vertex $v \in R_t$, and there would always be a point $x \in C \cap R_t$, $x \neq v$, such that $g_{L_t}(x) = R_t$, which is impossible (L' and M have no fibre in common). So f is a map of degree $n-2$. For every point $y \in C \cap R_t$ we have $f(y) = y$, otherwise $g_{L_t}(y)$ would again be R_t. But we claim more, namely that $g_{L_t}(y)$ is tangent to C at y. This can be seen in the following way. For a general point $x \in C$ the plane through x and R_t is certainly not the plane $V(x)$, hence $L_t(x)$ is the intersection of these two planes. Therefore $g_{L_t}(y)$ will be the intersection of the limit positions of these two planes if x approaches y, provided of course that these limit positions are different. The limit position of the first plane is the plane through R_t and the tangent line to C at y. The second plane contains this tangent line, but certainly not R_t. Hence $L_t(y) = T_C(y)$.

Now take the other coordinate z on Q as a coordinate on both C and R_t. Then f can be seen as a map from $P_1(z)$ onto itself of degree $n-2$. We may assume that $f(\infty) \neq \infty$. An easy calculation shows that $f'(y) = 0$ for every $y \in C \cap R_t$. Hence $f \equiv \dfrac{p_1(z)}{p_2(z)}$, where $p_1(z)$ and $p_2(z)$ are (inhomogeneous) poly-

nomials of degree $n-2$ in z, without common roots. Let z_0 be a value of z, corresponding to a point of $R_t \cap C$. We have $p_1(z_0)-z_0 p_2(z_0)=0$ and also $p_1'(z_0)p_2(z_0)-p_2'(z_0)p_1(z_0)=0$, hence $p_1'(z_0)-z_0 p_2'(z_0)=0$. But this means that this last equation, which is of degree at most $n-2$ would have $n-1$ different roots, since by assumption C and R_t have $n-1$ different points in common. This contradiction shows that our assumption, namely the existence of a 1-subbundle L of N with $L \simeq O_C(2n)$ is impossible. Hence N_C is always balanced if $C \subset Q$.

The idea behind this proof can be used to study the normal bundle of any smooth space curve. Let C be such a curve and $L \subset N_C$ a 1-dimensional subbundle. Let V be the inverse image of L in $T_{\mathbb{P}_3}|C$. If we take a general line $R \subset \mathbb{P}_3$ [such that R is not contained in any plane $g_V(x)$, $x \in C$], then V gives rise to a map $f\colon C \to R$ by taking for $f(x)$ the point $g_V(x) \in R$. The degree of f can be expressed in terms of $\deg c_1(L)$. This map has the following property: if $y \in C$, such that the tangent line to C at y intersects R in z, then $f(y)=z$. Moreover, the limiting position of the line through x and $f(x)$ if x approaches y is the tangent line to C at y. Thus the existence of special 1-subbundles of N leads to the existence of special correspondences. A certain converse also holds, though it should be stated with care. We hope to return to this point at another time.

Sketch of an Algebraic Proof of Proposition 6. Suppose C is contained in the quadric $X_0X_2-X_1X_3=0$. Realising C as the graph of an $(n-1)$-to-1 map, we may write it parametrically as

$$\mathbb{P}_1 \ni (s,t) \to (p_0,p_1,p_2,p_3) \in \mathbb{P}_3,$$

where each p_i is a form of degree n in 2 variables, and $p_0=sp$, $p_1=sq$, $p_2=tp$, $p_3=tq$ for some relatively prime forms p and q.

In this situation the normal bundle of C is determined by the following commutative diagram with exact rows and columns:

$$\begin{array}{ccccccccc}
 & & 0 & & 0 & & & & \\
 & & \downarrow & & \downarrow & & & & \\
 & & O_C & \xrightarrow{(n-1)} & O_C & & & & \\
 & & \downarrow{\scriptstyle (s,t)} & & \downarrow{\scriptstyle (p_0,p_1,p_2,p_3)} & & & & \\
 & & O_C^2(1) & \xrightarrow{\mathscr{J}} & O_C^4(n) & & & & \\
 & & \downarrow{\scriptstyle \binom{t}{-s}} & & \downarrow{\scriptstyle \kappa} & & & & \\
0 & \longrightarrow & T_C = O(2) & \xrightarrow{\phi} & T_{\mathbb{P}_3}|C & \longrightarrow & N_C & \longrightarrow & 0, \\
 & & \downarrow & & \downarrow & & & & \\
 & & 0 & & 0 & & & &
\end{array}$$

where

$$\mathscr{J} = \frac{\partial(p_0, p_1, p_2, p_3)}{\partial(s,t)}$$

is the Jacobian matrix.

Applying the functor $F = \sum_v H^0(C, O_C(v))$, and writing $S(k)$ for the free $\mathbb{C}[s,t]$-module whose generator has degree $-k$, we may turn the computation of N into a problem on graded modules. One sees easily that $F(\kappa)$ is the map

$$S(-n-1) \oplus S(-n-1) \oplus S(-2n:2) \to S^4(-n)$$

given by the matrix

$$\begin{pmatrix} t & 0 & \frac{\partial q}{\partial s} \\ -s & 0 & \frac{\partial q}{\partial t} \\ 0 & t & \frac{-\partial p}{\partial s} \\ 0 & -s & \frac{-\partial p}{\partial t} \end{pmatrix}$$

and a simple calculation then shows that

$$F(\phi): S(-n-1) \oplus S(-n-1) \oplus S(-2n+2) \to S(-2)$$

is given by the matrix

$$\left(p, q, \frac{1}{t}\left(p\frac{\partial q}{\partial s} - q\frac{\partial p}{\partial s}\right) + J\right),$$

where $J = \det \begin{pmatrix} \frac{\partial p}{\partial s} & \frac{\partial p}{\partial t} \\ \frac{\partial q}{\partial s} & \frac{\partial q}{\partial t} \end{pmatrix}$, and the expression $\left(p\frac{\partial q}{\partial s} - q\frac{\partial p}{\partial s}\right)$ is formally divisible by t.

Changing coordinates we may assume $p = tr$ for some form r of degree $n-2$, and then the above matrix can be rewritten $(p, q, J_0 + J)$, where

$$J_0 = \det \begin{pmatrix} \frac{\partial q}{\partial s} & q \\ \frac{\partial r}{\partial s} & r \end{pmatrix}.$$

We now note that $J = (1-n)J_0$; since both J and J_0 are multilinear in r and q, it suffices to check this for monomials, where it is evident. Thus the above matrix has the form

$$F(\phi)=\left(p, q, \frac{(2-n)}{(1-n)}J\right).$$

By the local duality theorem (see for example [G-H, 5.1]) J generates the unique minimal ideal of $\mathbb{C}[s,t]/(p,q)$ and it follows that $\ker F(\phi)=S(-2n+1)\oplus S(-2n+1)$.

Translating this into the language of bundles, we find

$$N=O_C(2n-1)\oplus O_C(2n-1)$$

as required.

2. Realising the Splitting of the Normal Bundle

Let C be a smooth rational curve of degree 3. Fixing $p\in C$, we can construct a 2-bundle V_p on C as follows. If $x\in C$, $x\neq p$, then we take for $V_p(x)$ the 2-dimensional linear subspace of $T_{\mathbb{P}_3}|C(x)$, determined by the well-defined plane through p and the tangent line to C at x; if $x=p$ we use instead the osculation plane to C at p. The bundle V_p gives a 1-subbundle $L_p\subset N=N_{C/\mathbb{P}_3}$, and $N\cong L_p\oplus L_q$ if $p\neq q$. In fact, $N\cong O_C(5)\oplus O_C(5)$ and the pencil of 1-subbundles of N, which are isomorphic to $O_C(5)$ is exactly the pencil $\{L_p\}$, $p\in C$.

All this is easy to verify.

We can identify the projective bundle $P=P_C$ of N with the bundle attaching to each point $x\in C$ the projective line of planes, passing through the tangent line of C at x. Then a splitting of N consists of two sections in this bundle, which don't meet. In other words, a splitting consists of two fields of planes H_1, H_2 along C, such that both $H_1(x)$ and $H_2(x)$ always contain the tangent line to C at x, but $H_1(x)\neq H_2(x)$ for all $x\in C$. In this light the preceding construction for a cubic can be interpreted in the following way. We start by taking a fixed point $p\in C$. If $x\neq p$ we take for $H_p(x)$ the tangent plane to the quadratic cone over C with vertex p; for $x=p$ we take $H_p(p)=\lim_{x\to p} H_p(x)$. If we carry out this construction for two different points of C, we obtain a splitting of N.

Given any 1-subbundle of N, in other words a section s in P, then for k large enough there always exist surfaces S of degree k (containing C and singular in a finite number of points on C) such that the field of tangent planes to S along C (plus limiting positions) yields the section s. This is a consequence of the fact that the natural homomorphism

$$\Gamma(C,\mathscr{I}(k))\to\Gamma(C,\mathscr{I}(R)/\mathscr{I}^2(k))=(C,N^*(k))$$

is surjective for $k\geqq k_0$ (here $\mathscr{I}$ denotes the ideal sheaf of C).

So given a smooth rational curve C we can always realise the splitting of N_C using surfaces of sufficiently high degree. Is it possible to do so in a simple way, using surfaces of a degree as low as possible, as was done before with quadratic cones for a cubic? We don't know a general answer to this (admittedly not very precise) question, but we know an attractive answer in several special cases. As an example we treat the case of quartics. By the Main

Theorem the normal bundle is always isomorphic to $O_C(7) \oplus O_C(7)$ for such a curve C.

Let C_{norm} be a smooth rational quartic in $\mathbb{P}_4$. We always can find a point $c \in \mathbb{P}_4 (c \notin C_{\text{norm}})$ and linear embedding of $\mathbb{P}_3$ in $\mathbb{P}_4$, such that the projection f from c onto $\mathbb{P}_3$ maps C_{norm} isomorphically onto C. Now C is contained in a smooth quadric as a curve of type $(3, 1)$. The planes through c which meet C_{norm} in three points correspond by f to the lines on Q which meet C in three points. Let V be such a plane, meeting C_{norm} in three different points p_1, p_2, and p_3. Since no three of the four points p_1, p_2, p_3 and c are collinear, there exists a smooth conic $K \subset V$ passing through p_1, p_2, and p_3, such that its tangents in p_1 and p_2 meet in c.

We consider the uniquely determined linear correspondence between the points of K and those of C_{norm}, for which p_1, p_2, and p_3 correspond to themselves. The union of the lines joining corresponding points (plus limit positions) is a scroll, i.e. a ruled surface of degree 3, isomorphic to $\mathbb{P}_2$ with one point blown up; one checks this most easily by showing that the degree of the surface is 3 and checking the possibilities. The lines on R determine a 1-subbundle of $T_{\mathbb{P}_4} | C_{\text{norm}}$, which is isomorphic to $O_C(5)$ by Nakano's theorem. Projecting from c we obtain a 1-subbundle of $T_{\mathbb{P}_3} | C$, which is tangent to C in at least two points, namely $f(p_1)$ and $f(p_2)$. Further projection into N_C gives a 1-subbundle N_1, of degree 7 or more. Since N_C does not contain any 1-subbundles of degree 8 or more, we find that $N_1 \cong O_C(7)$. Its inverse image in $T_{\mathbb{P}_3} | C$ in a general point $x \in C$ can be obtained as f [tangent plane to R at $f^{-1}(x)$]. On the other hand, $f(R)$ is the unique Steiner surface in $\mathbb{P}_3$, which has $f(V)$ as its double line and $f(p_1)$, $f(p_2)$ as its pinch points [G-H, p. 629]. Thus the tangent planes along C of two of these Steiner surfaces give a splitting of N_C. To obtain all 1-subbundles of N_C, isomorphic to $O_C(7)$ one has to consider some limit cases.

Acknowledgements. Both authors are grateful to the S.F.B. "Theoretische Mathematik" in Bonn, FRG, for support and hospitality during the preparation of this paper.

References

[B] Brieskorn, E.: Über holomorphe $\mathbb{P}_n$-Bündel über $\mathbb{P}_1$. Math. Ann. **157**, 343–357 (1965)

[E-V] Eisenbud, D., Van de Ven, A.: On the family of smooth rational space curves with given degree and normal bundle.

[G-H] Griffiths, P.A., Harris, J.: Principles of algebraic geometry. New York: Wiley 1978

[G-S] Ghione, F., Sacchiero, G.: Normal bundles of rational curves in $\mathbb{P}_3$ (preprint)

[Ha] Harris, J.: A note on the normal bundles of space curves (unpublished)

[Hu] Hulek, K.: The normal bundle of a curve on a quadric (preprint, May 1981)

[K] Kodaira, K.: A theorem of completeness of characteristic systems for analytic families of compact submanifolds of complex manifolds. Ann. Math. **75**, 146–162 (1962)

[P] Piene, R.: Numerical characters of a curve in projective n-space. In: Real and complex singularities. Oslo 1976. Groningen: Sijthoff en Noordhoff

[S] Simonis, J.: A class of indecomparable algebraic vector bundles. Math. Ann. **192**, 262–278 (1971)

Received December 23, 1980

Math. Ann. 256, 465–473 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Zur Arithmetik von Diedergruppenerweiterungen

Hans-Jochen Bartels

Mathematisches Institut der Universität, Bunsenstraße 3–5, D-3400 Göttingen, Bundesrepublik Deutschland

Einleitung

Kriterien dafür, wann bei abelschen und auch einigen nicht abelschen galoisschen Zahlkörpererweiterungen K/k der Hassesche Normensatz gilt, sind von verschiedenen Autoren angegeben worden (vgl. [5, 6, 10] und die dort zitierte Literatur). Für nicht galoissche Erweiterungen K/k ist dagegen vergleichsweise wenig bekannt [vgl. 6]. Einen Beitrag zu dieser Frage liefert die vorliegende Note. Genauer wird folgendes gezeigt:

Satz 1. *k sei ein endlicher algebraischer Zahlkörper oder ein algebraischer Funktionenkörper einer Variablen über einem endlichen Konstantenkörper. $K \supset k$ sei eine endliche separable Erweiterung von k vom Grad m, deren galoissche Hülle L über k als Galoisgruppe $Gal(L/k)$ die Diedergruppe der Ordnung $2 \cdot m$ hat. Dann gilt für K/k der Hassesche Normensatz.*

Dabei sagt man, daß für eine endliche Erweiterung globaler Körper K/k der Hassesche Normensatz gilt, wenn

$$k^* \cap N_{K/k}(I_K) = N_{K/k}(K^*)$$

ist, I_K die Idelegruppe von K. Der Beweis von Satz 1 benutzt die Berechnung der Galoiskohomologie algebraischer Tori nach Nakayama und Tate [15]. Als Anwendungen von Satz 1 werden im letzten Abschnitt dieser Note angegeben:

a) Ist K maximaler kommutativer Teilkörper eines zentralen Schiefkörpers D über dem Zahlkörper k, so gilt für K/k der Hassesche Normensatz, wenn $[D: k] < 36$ ist (Absatz 3, Satz 2).

b) Für die Normeinsgruppen von Schiefkörpern D mit Zentrum k und kleinen Grades über k gilt ein Lokal-Global-Prinzip für die Konjugationsklassen (Absatz 3, Satz 3).

1.

Im folgenden bezeichne k einen globalen Körper, d.h. einen endlichen algebraischen Zahlkörper oder einen algebraischen Funktionenkörper einer Variablen

0025-5831/81/0256/0465/$01.80

über einem endlichen Konstantenkörper. K sei eine endliche separable Erweiterung von k etwa vom Grad m. Die galoissche Hülle L von K über k sei eine Diedergruppenerweiterung von k vom Grad $2 \cdot m$. Es sei $G = \mathrm{Gal}(L/k) = \langle s, t \rangle$ mit $s^2 = t^m = 1$, $sts^{-1} = t^{-1}$.

Man hat zunächst das

Lemma 1. *Ist m ungerade, dann gilt sowohl für K/k als auch für L/k der Hassesche Normensatz.*

Beweis. Die Bedingung $(m, 2) = 1$ besagt, daß alle p-Sylowgruppen G_p von G zyklisch sind. Nach Chebotarev gibt es daher für jede Primzahl p unendlich viele Primstellen v von L, deren Zerlegungsgruppe G_v über k G_p enthält. Nach [5, Theorem 2.5] („Übergang zu den p-Sylowgruppen") gilt daher für L/k der Hassesche Normensatz. Mit [6, Proposition 1] folgt dann auch die Gültigkeit des Normensatzes für K/k. □

Das folgende Lemma 2 reduziert den Beweis von Satz 1 auf den Fall $m = 2^{\mu+1}$, $\mu \in \mathbb{N}$:

Lemma 2. *Gilt Satz 1 für $m = 2^{\mu+1}$, $\mu \in \mathbb{N}$ so auch für beliebige m.*

Beweis. Ist $m = 2^{\mu} \cdot m'$ mit $\mu \in \mathbb{N}$ und $(m', 2) = 1$, $m' > 1$, so läßt sich K als Kompositum zweier Körper K_1, K_2 mit $[K_1 : k] = 2^{\mu}$, $[K_2 : k] = m'$ schreiben.

Falls $\mu = 1$ ist, gilt für K_1/k der Normensatz. Falls $\mu > 1$ ist, sind die galoisschen Hüllen von K_1 bzw. K_2 über k Diedergruppenerweiterungen vom Grad $2^{\mu+1}$ bzw. $2 \cdot m'$ über k. Nach Voraussetzung (bzw. nach Lemma 1) gilt dann für K_1 (bzw. K_2) der Hassesche Normensatz über k. Damit ist er dann auch für K/k richtig [vgl. 6, Proposition 2]. □

2.

Zum Beweis von Satz 1 für den Fall $m = 2^{\mu+1}$ benötigen wir einige allgemeine Vorbereitungen: $K \supset k$ sei eine zunächst noch beliebige separable Erweiterung globaler Körper. Es bezeichnen dann G_m die multiplikative Gruppe (als algebraische Gruppe über k) und $R_{K/k} G_m$ das durch $R_{K/k} G_m(A) = G_m(A \otimes_k K)$ für kommutative k-Algebren A definierte Gruppenschema. ${}^1R_{K/k} G_m$ sei wie üblich der Kern der Normabbildung $N_{K/k} : R_{K/k} G_m \to G_m$. Etwa nach [13, S. 70] gilt für K/k der Hassesche Normensatz genau dann, wenn für den algebraischen Torus $T = {}^1R_{K/k} G_m$ das Hasse-Prinzip gilt, d.h. wenn für beliebige endliche galoissche Erweiterungen L/k die natürliche Abbildung

$$\omega_L : H^1(L/k, T(L)) \to H^1(L/k, T(A_L))$$

injektiv ist, dabei ist A_L der zu L gehörige Adelering. Um die Injektivität aller Abbildungen ω_L nachzuweisen, genügt es, die Injektivität von ω_L für einen galoisschen Zerfällungskörper L von T nachzuprüfen [7, S. 46/47]. Ein Zerfällungskörper für $T = {}^1R_{K/k} G_m$ ist die galoissche Hülle L von K/k. L bezeichne im folgenden stets die galoissche Hülle von K/k und es sei $G := \mathrm{Gal}(L/k)$, $H := \mathrm{Gal}(L/K) \leqq G$. Der G-Linksmodul $X^*(T) := \mathrm{Hom}(G_m, T)$ berechnet sich etwa

nach [13] sowie [7, S. 11] zu

$$X^*(T)=\mathbb{Z}[G/H]^0, \tag{1}$$

dabei bezeichnet $\mathbb{Z}[G/H]^0$ den Kern der Augmentation $\mathbb{Z}[G/H]\to\mathbb{Z}$. Nach [8, S. 409] (vgl. auch [9] S. 188, Fußnote 4) ist die Injektivität von ω_L gleichbedeutend mit der Surjektivität von

$$\Phi: \bigoplus_v \hat{H}^{-2}(G_v, X^*(T))\to H^{2}(G, X^*(T)).$$

dabei bezeichnet G_v die Zerlegungsgruppe der Primstelle v von L und Φ ist die direkte Summe der lokalen Verlagerungsabbildungen

$$\Phi_v=\mathrm{Cor}_v: \hat{H}^{-2}(G_v, X^*(T))\to\hat{H}^{2}(G, X^*(T)).$$

Ist jetzt wie in Abschnitt 1 G Diedergruppe der Ordnung $2m$ und gilt überdies $m=2^{\mu+1}$, $H=\langle s\rangle$, so ist für eine geeignete Primstelle v von L $G_v=\langle s\rangle=H$ und wegen $\hat{H}^{-2}(\cdot, X^*(T))=H_1(\cdot, X^*(T))$ genügt es daher zum Beweis von Satz 1, für $m=2^{\mu+1}$ das folgende Lemma nachzuprüfen:

Lemma 3. Cor: $H_1(H, \mathbb{Z}[G/H]^0)\to H_1(G, \mathbb{Z}[G/H]^0)$ *ist surjektiv.*

Beweis. Der Kern $\mathbb{Z}[H\backslash G]^0$ der Augmentation $\mathbb{Z}[H\backslash G]\to\mathbb{Z}$ ist ein G-Rechtsmodul. Wendet man auf

$$0\to\mathbb{Z}[H\backslash G]^0\to\mathbb{Z}[H\backslash G]\to\mathbb{Z}\to 0$$

den Funktor $\mathrm{Tor}^G_*(\cdot, X^*(T))$ an, so ergibt sich wegen $\mathrm{Tor}^G_i(\mathbb{Z}[H\backslash G], X^*(T))\cong\mathrm{Tor}^H_i(\mathbb{Z}, X^*(T))$ (Lemma von Shapiro, vgl. [3, S. 99] eine exakte Sequenz

$$H_1(H, \mathbb{Z}[G/H]^0)\xrightarrow{\mathrm{Cor}} H_1(G, \mathbb{Z}[G/H]^0)\xrightarrow{\delta}\mathbb{Z}[H\backslash G]^0\underset{\mathbb{Z}[G]}{\otimes}\mathbb{Z}[G/H]^0.$$

Hieraus entnimmt man: Cor ist sicher dann surjektiv, wenn $\delta=0$ ist. Wegen der Endlichkeit von $H_1(G, \mathbb{Z}[G/H]^0)$ genügt es hierzu nachzuprüfen, daß $\mathbb{Z}[H\backslash G]^0\underset{\mathbb{Z}[G]}{\otimes}\mathbb{Z}[G/H]^0$ keine Torsion besitzt. Hier gilt genauer der

Hilfssatz. $\mathbb{Z}[H\backslash G]^0\underset{\mathbb{Z}[G]}{\otimes}\mathbb{Z}[G/H]^0$ *ist freier* $\mathbb{Z}$*-Modul vom Rang* $m/2$.

Beweis des Hilfssatzes. $H-tH$ ist erzeugendes Element des $\mathbb{Z}[G]$-Linksmoduls $\mathbb{Z}[G/H]^0$ und man rechnet nach:

$$\mathbb{Z}[G/H]^0=\frac{\mathbb{Z}[G]}{(\mathbb{Z}[G](1+t+\ldots+t^{m-1})+\mathbb{Z}[G](1+ts))}$$

Entsprechend ist

$$\mathbb{Z}[H\backslash G]^0=\frac{\mathbb{Z}[G]}{((1+t+\ldots+t^{m-1})\mathbb{Z}[G]+(1+st)\mathbb{Z}[G])}$$

Für das Tensorprodukt ergibt sich damit wegen

$$\mathbb{Z}[G](1+t+\ldots+t^{m-1})=(1+t+\ldots+t^{m-1})\mathbb{Z}[G]$$

folgende Präsentation:

$$\mathbb{Z}[H\backslash G]^0 \underset{\mathbb{Z}[G]}{\otimes} \mathbb{Z}[G/H]^0=\frac{\mathbb{Z}[G]}{(\mathbb{Z}[G](1+t+\ldots+t^{m-1})+\mathbb{Z}[G](1+ts)+(1+st)\mathbb{Z}[G])}.$$

Nun ist

$$\begin{aligned}\mathbb{Z}[G](1+t+\ldots+t^{m-1}) &= \mathbb{Z}(1+t+\ldots+t^{m-1})+\mathbb{Z}(s+st+\ldots+st^{m-1})\\ \mathbb{Z}[G](1+ts) &= \sum_i \mathbb{Z}(t^i+st^{m-i-1}) \qquad (2)\\ (1+st)\mathbb{Z}[G] &= \sum_i \mathbb{Z}(t^i+st^{i+1}).\end{aligned}$$

Wählt man als $\mathbb{Z}$-Basis von $\mathbb{Z}[G]$

$$\{1,t,\ldots,t^{m-1},1+st^{m-1},t+st^{m-2},\ldots,t^{m-1}+s\},$$

so lassen sich mit den Relationen (2) zunächst die letzten m Basiselemente eliminieren, und für die ersten m Basiselemente verbleiben als Relationen:

$$1+t+\ldots+t^{m-1},\ t^{i-1}-t^{m-i-1} \qquad i=1,\ldots,\frac{m}{2}-1.$$

Damit läßt sich auch noch die Hälfte der t^i $i=0,\ldots,m-1$ eliminieren und es folgt die Behauptung des Hilfssatzes. □

Wie zu Beginn des Abschnittes 2 bereits bemerkt wurde, folgt aus dem Lemma 3 die Gültigkeit des Hasse-Prinzips für den Torus ${}^1R_{K/k}G_m$ und daher der Hassesche Normensatz für die Erweiterung K/k. Zusammen mit Lemma 2 folgt damit:

Satz 1. *k sei ein endlicher algebraischer Zahlkörper oder ein algebraischer Funktionenkörper einer Variablen über einem endlichen Konstantenkörper. $K\supset k$ sei eine endliche separable Erweiterung von k vom Grad m, deren galoissche Hülle L über k als Galoisgruppe $Gal(L/k)$ die Diedergruppe der Ordnung $2\cdot m$ hat. Dann gilt für K/k der Hassesche Normensatz.*

Bemerkung 1. In [6, Theorem 5] wird unter anderem gezeigt, daß mit den oben gewählten Bezeichnungen im Fall $m=4$ für K/k der Hassesche Normensatz gilt, allerdings unter der zusätzlichen Voraussetzung, daß für wenigstens eine Primstelle v von L die Zerlegungsgruppe $G_v=G$ ist. Der oben bewiesene Satz 1 zeigt, daß diese einschränkende Voraussetzung im Fall $G=D_4$, Diedergruppe der Ordnung 8, überflüssig ist.

Bemerkung 2. Sind L,K,k wie in Satz 1 etwa mit $m=4$, so gilt für L/k im allgemeinen kein Hassescher Normensatz. Ist nämlich L/k unverzweigt (zur Existenz solcher unverzweigter Erweiterungen [vgl. 14, Lemma 2.7], [1, S. 75],

so kann nach [5] Theorem 4.1 für L/k im Falle $\mathrm{char}\, k \neq 2$ kein Hassescher Normensatz gelten. In diesem speziellen Fall lassen sich über den sogenannten Zahlknoten

$$k^* \cap N_{L/K}(I_L)/N_{L/k}(L^*)$$

übrigens noch genauere Aussagen gewinnen, vgl. hierzu etwa [10, S. 43/44].

3.

In diesem Abschnitt werden einige Anwendungen von Satz 1 angegeben. Dabei beschränken wir uns auf Zahlkörper. Wie bereits in [2] ausgeführt wurde, hängt die Gültigkeit des Hasseschen Normensatzes eng mit der Frage nach einem Lokal-Global-Prinzip für die Konjugationsklassen in den einfachzusammenhängenden halbeinfachen Gruppen vom Typ ${}^1A_n$ zusammen. Nennt man wie in [14] einen Oberkörper K des Zahlkörpers k k-adäquat, wenn K maximaler kommutativer Teilkörper eines zentralen Schiefkörpers D über k ist, so hat man zunächst den

Satz 2. *Ist K k-adäquat vom Grad $[K: k] < 6$, so gilt für K/k der Hassesche Normensatz.*

Zum Beweis von Satz 2 benötigen wir ein Lemma, das unter allgemeineren Voraussetzungen richtig ist. Deswegen bezeichne K/k eine zunächst noch beliebige endliche Zahlkörpererweiterung und L sei die galoissche Hülle von K/k. $G = \mathrm{Gal}(L/k)$ und $H = \mathrm{Gal}(L/K)$ haben die gleiche Bedeutung wie zu Beginn von Abschnitt 2. Für Primzahlen p mit $p|[L: k]$ bezeichne $H_p \leqq H$ eine p-Sylowgruppe von H und $G_p \leqq G$ eine p-Sylowgruppe von G, die H_p enthält. $L_p \subseteq \tilde{L}_p \subseteq L$ seien die zu G_p bzw. H_p gehörigen Unterkörper von L. Ist dann v Primstelle von L, so bezeichne $(G_p)_v$ die Zerlegungsgruppe von v über L_p. Mit diesen Bezeichnungen gilt folgendes Lemma, das die Ergebnisse von [14, Proposition 2.5] auf den Fall einer nicht notwendig galoisschen Erweiterung K/k verallgemeinert:

Lemma 4. *Es seien K k-adäquat, L die galoissche Hülle von K/k und $p|[L: k]$. Dann existieren zwei Primstellen v_1, v_2 von L, deren Einschränkungen auf L_p nicht äquivalent sind und für die $G_p = (G_p)_{v_i} \cdot H_p$ ist für $i = 1, 2$.*

Der Beweis des Lemmas folgt im wesentlichen dem in [14] für den Fall $H = \{1\}$ durchgeführten Beweis. Es sei D zentraler Schiefkörper über k, so daß $K \subseteq D$ maximaler kommutativer Teilkörper ist. Bezeichnet $\mathrm{Br}(k) = H^2(k, G_m)$ die Brauergruppe von k, so sei für Oberkörper M von k $\mathrm{Br}(M/k)$ die Untergruppe von $\mathrm{Br}(k)$ der durch M zerfällten Algebrenklassen. Die Klasse $[D]$ des Schiefkörpers D in $\mathrm{Br}(k)$ ist dann ein Element der Ordnung $[K: k]$ in $\mathrm{Br}(K/k) \subseteq \mathrm{Br}(L/k)$. Ist

$$j\colon \mathrm{Br}(L/k) \to \mathrm{Br}(L/L_p)$$

der durch $[A] \mapsto [A \otimes_k L_p]$ definierte Homomorphismus, so hat man eine exakte Sequenz

$$0 \to \mathrm{Br}(L_p/k) \to \mathrm{Br}(L/k) \xrightarrow{\ j\ } \mathrm{Br}(L/L_p)$$

und gelangt so zu einem Element $j([D])$ in $\mathrm{Br}(L/L_p)$. $j([D])$ liegt wegen $\tilde{L}_p \supset K$ bereits in $\mathrm{Br}(\tilde{L}_p/L_p)$ und hat die Ordnung $[\tilde{L}_p: L_p]$. Einerseits ist nämlich die Ordnung von $j([D])$ ein gemeinsamer Teiler von $[K: k]$ und $[L: L_p]$, d.h. ein Teiler von $[\tilde{L}_p: L_p]$, andererseits enthält Ker $j = \mathrm{Br}(L_p/k)$ aber auch nur Elemente einer Ordnung s mit $s|[L_p: k]$, so daß die Ordnung von $j([D])$ nicht echt kleiner als $[\tilde{L}_p: L_p]$ sein kann, da $[L_p: k]$ und $[\tilde{L}_p: L_p]$ prim sind. Nach [14, Proposition 2.2] gibt es daher einen Schiefkörper C mit Zentrum L_p, der $\tilde{L}_p$ als maximalen kommutativen Teilkörper enthält, die Klasse $[C]$ in $\mathrm{Br}(\tilde{L}_p/L_p)$ hat die Ordnung $[\tilde{L}_p: L_p]$. Bezeichnet für eine Primstelle v von L_p $(L_p)_v$ die Komplettierung von L_p bezüglich v, so ist bekanntlich

$$\mathrm{Br}(\tilde{L}_p/L_p) \subseteq \bigoplus_v \mathrm{Br}((L_p)_v).$$

Die Klasse $[C]$ schreibt sich daher als eine endliche Summe $[C] = c_{v_1} + \ldots + c_{v_m}$ mit $m \in \mathbb{N}$, $c_{v_i} \in \mathrm{Br}((L_p)_{v_i})$. $[C]$ hat die Ordnung $p^r := [\tilde{L}_p: L_p]$, also hat auch mindestens eines der c_{v_i} die Ordnung p^r. Da die Summe der Invarianten der c_{v_i} in $\mathbb{Q}/\mathbb{Z}$ Null ergibt, gibt es sogar mindestens zwei Primstellen etwa v_1, v_2, für die c_{v_i} die Ordnung p^r hat. Dann gilt aber für jede Fortsetzung w_i von v_i $(i = 1, 2)$ auf $\tilde{L}_p$:

$$p^r | [(\tilde{L}_p)_{w_i}: (L_p)_{v_i}] \qquad i = 1, 2,$$

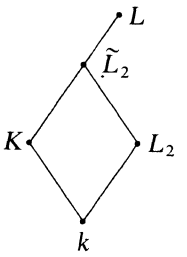
d.h. es ist für $i = 1, 2$ $[(\tilde{L}_p)_{w_i}: (L_p)_{v_i}] = [\tilde{L}_p: L_p]$. Hieraus kann man schließen, daß über v_i $(i = 1, 2)$ nur eine Primstelle von $\tilde{L}_p$ liegt, und das ist gerade die Behauptung von Lemma 4. □

Nun zum *Beweis von Satz 2*. Ist K/k irgendeine Erweiterung von Zahlkörpern vom Grad 2, 3, 5 oder allgemeiner vom Primzahlgrad, so gilt für K/k der Hassesche Normensatz etwa nach [2, Lemma 4]. Hat K/k den Grad 4, so unterscheiden wir mehrere Fälle:

a) K/k ist k-adäquat und galoissch
b) K/k ist k-adäquat, nicht galoissch,
   $G = D_4$, die Diedergruppe der Ordnung 8.
c) K/k ist k-adäquat, nicht galoissch,
   $G = A_4$, die alternierende Gruppe der Ordnung 12.
d) K/k ist k-adäquat, nicht galoissch,
   $G = S_4$, die symmetrische Gruppe der Ordnung 24.

Im Fall a) gilt der Normensatz nach [5] Theorem 3.1, der Fall b) ist mit Satz 1 erledigt. Im Fall c) hat H die Ordnung 3, daher ist $H_2 = \{1\}$. Nach Lemma 4 existiert also eine Primstelle v von L, für die $G_2 = (G_2)_v$, also $G_v \supseteq G_2$ gilt, wenn G_v die Zerlegungsgruppe von v über k bezeichnet. Andererseits ist nach Chebotarev auch G_3 in einer Zerlegungsgruppe G_w von w über k enthalten für eine passende Primstelle w von L. Nach [5, Theorem 2.5] gilt daher für L/k der Hassesche Normensatz und damit nach [6, Proposition 1] auch für K/k.

Im Fall d) hat H die Ordnung 6 und ist isomorph einer symmetrischen Gruppe S_3 der Ordnung 6. Es ist $\tilde{L}_3 = L_3$ und $\tilde{L}_2$ hat den Grad 4 über L_2. Man rechnet nach, daß die galoissche Hülle von $\tilde{L}_2$ über L_2 L ist, und daß L/L_2 eine Diedergruppenerweiterung ist (vgl. das Diagramm).



Nach Satz 1 gilt daher für $\tilde{L}_2/L_2$ der Hassesche Normsatz. Ist daher $\alpha \in k$ lokal überall Norm bei der Erweiterung K/k, so existiert zunächst eine Zahl $\beta \in \tilde{L}_2$ mit $N_{\tilde{L}_2/L_2}(\beta)=\alpha$. Sind dann $a, b \in \mathbb{Z}$ derart, daß $4 \cdot a + 3 \cdot b = 1$ ist, so gilt

$$N_{K/k}(\alpha^a N_{\tilde{L}_2/K}(\beta^b)) = \alpha^{4 \cdot a} N_{\tilde{L}_2/k}(\beta^b) = \alpha^{4a} \cdot N_{L_2/k}(\alpha^b) = \alpha^{4a+3b} = \alpha.$$

Damit gilt also auch im Fall d) der Normensatz. □

Ist D ein Schiefkörper mit Zentrum k vom Grad n^2 über k, so bezeichne im folgenden $\mathfrak{G}$ die zugehörige Normeinsgruppe, genauer ist für kommutative k-Algebren A

$$\mathfrak{G}(A) = \{x \in D \otimes_k A \mid RN_{D/k}(x) = 1\}, \tag{3}$$

dabei ist $RN_{D/k}$ die reduzierte Norm auf D. Im Hinblick auf die bereits in [2, Abschnitt 2] diskutierte Frage nach einem Lokal-Global-Prinzip für die Konjugationsklassen in $\mathfrak{G}(k)$ gilt der

Satz 3. *Ist k ein Zahlkörper und D ein Schiefkörper mit Zentrum k vom Grad $[D: k] = n^2 < 36$ und bezeichnet $\mathfrak{G}$ die zu D vermöge (3) gehörige Normeinsgruppe, so gilt für die Konjugationsklassen von $\mathfrak{G}(k)$ ein Lokal-Global-Prinzip, d.h.: Sind $x, y \in \mathfrak{G}(k)$ in $\mathfrak{G}(k_v)$ konjugiert für alle Primstellen v von k so auch schon in $\mathfrak{G}(k)$.*

Beweis. Nach [2] Satz 2 genügt es, folgendes zu zeigen: Bei beliebigem $y \in \mathfrak{G}(k)$ gilt für den Zentralisator $\mathfrak{G}_y$ von y in $\mathfrak{G}$ das Hasse-Prinzip, d.h. die kanonische Abbildung

$$\omega: H^1(k, \mathfrak{G}) \to \prod_v H^1(k_v, \mathfrak{G}_y)$$

ist injektiv. Da D ein Schiefkörper ist, ist für $y \in \mathfrak{G}(k) \subseteq D$ $E := k[y] \subseteq D$ ein Körper und der Zentralisator von y in D ist eine zentrale einfache E-Algebra Z_y. Für den Zentralisator $\mathfrak{G}_y$ von y in $\mathfrak{G}$ hat man eine exakte Sequenz algebraischer Gruppen

$$1 \to \mathfrak{G}_y \to \tilde{\mathfrak{G}}_y \xrightarrow{\varphi} G_m \to 1, \tag{4}$$

dabei ist für kommutative k-Algebren A $\tilde{\mathfrak{G}}_y(A) = (Z_y \otimes_k A)^* = G_m(Z_y \otimes_k \cdot A)$ und φ wird durch die reduzierte Norm von D nach k induziert. Aus der kurzen exakten Sequenz (4) ergeben sich in der Kohomologie exakte Sequenzen, die die Zeilen des folgenden kommutativen Diagramms ausmachen:

$$\begin{array}{ccccccccc} 1 \to & \mathfrak{G}_y(k) & \to & Z_y^* & \xrightarrow{\varphi} & k^* \to & H^1(k, \mathfrak{G}_y) & \to & H^1(k, \tilde{\mathfrak{G}}_y) = \{1\} \\ & \downarrow & & \downarrow & & \downarrow & \downarrow & & \downarrow \\ 1 \to & \prod_v \mathfrak{G}_y(k_v) & \to & \prod_v (Z_y \otimes_k k_v)^* & \to & \prod_v k_v^* \to & \prod_v H^1(k_v, \mathfrak{G}_y) & \to & \prod_v H^1(k_v, \tilde{\mathfrak{G}}_y) = \{1\} \end{array} \tag{5}$$

Nach [4, S. 28, (3.6)] ist hierbei

$$\varphi(Z_y^*) = N_{E/k}(RN_{Z_y/E}(Z_y^*)).$$

Es ist

$$Z_y \otimes_k k_v \cong \bigoplus_{w \mid v} A_w$$

w Primstelle von E

mit zentralen einfachen E_w-Algebren A_w vom Grad $[Z_y : E]$ über E_w, E_w ist dabei die Komplettierung von E an der Primstelle w. Unter Benutzung von [4, S. 28, (3.6)] berechnet sich die Abbildung

$$\varphi_v : (Z_y \otimes_k k_v)^* = (\bigoplus_{w \mid v} A_w)^* \to k_v^*$$

folgendermaßen:

$$\varphi_v((a_w)) = \prod_{w \mid v} N_{E_w/k_v}(RN_{A_w/E_w}(a_w)).$$

Um die Injektivität von $H^1(k, \mathfrak{G}_y) \to \prod_v H^1(k_v, \mathfrak{G}_y)$ nachzuweisen, hat man nach (5) die Injektivität der Abbildung

$$k^*/N_{E/k}(RN_{Z_y/E}(Z_y^*)) \to \prod_v k_v^*/\varphi_v((Z_y \otimes_k k_v)^*)$$

nachzuprüfen. Ist dazu $\alpha \in k^*$, so daß $\alpha \in \varphi_v((Z_y \otimes_k k_v)^*)$ für alle Primstellen v von k ist, so ist zunächst α lokal überall Norm bei der Erweiterung E/k, und weil E k-adäquat vom Grad $\leqq 5$ ist [vgl. 14, Proposition 2.2], gilt nach Satz 2 für E/k der Normensatz. Damit existiert jedenfalls ein $\beta \in E^*$ mit $N_{E/k}(\beta) = \alpha$. Es ist zu zeigen, daß $\beta \in RN_{Z_y/E}(Z_y^*)$ gewählt werden kann. Nach dem Normensatz von Hasse und Schilling für einfache Algebren liegt $\beta \in E^*$ genau dann in $RN_{Z_y/k}(Z_y^*)$, wenn β an allen den reellen Primstellen w von E positiv ist, an denen die E-Algebra Z_y verzweigt ist. Sei daher w_1 eine reelle verzweigte Primstelle von E, v bezeichne die Einschränkung von w_1 auf k. Sind dann $w_1, \ldots, w_r$ die sämtlichen über v liegenden Primstellen von E, und sind $w_1, \ldots, w_{\rho_1}$ reell, $w_{\rho_1+1}, \ldots, w_{\rho_1+\rho_2}$ komplex und etwa $w_1, \ldots, w_{\nu_1}$ reell und verzweigt ($\nu_1 \leqq \rho_1$), so sei

$$\gamma^{(v)} = (\gamma_1^{(v)}, \ldots, \gamma_r^{(v)}) \in E \otimes_k k_v \cong \bigoplus_{i=1}^{r} E_{w_i}$$

folgendermaßen gewählt:

$$\gamma_i^{(v)} := \begin{cases} -1 & \text{falls } \beta < 0 \text{ an der Stelle } w_i,\ i \leqq \nu_1 \\ 1 & \text{falls } \beta > 0 \text{ an der Stelle } w_i,\ i \leqq \nu_1, \end{cases} \tag{6}$$

Für $i > v_1$ seien die $\gamma_i \in \{\pm 1\}$ so, daß $\prod_{i=1}^{r} \gamma_i^{(v)} = 1$ wird. Falls $v_1 = r$ ist, gilt mit der Festsetzung (6) automatisch $\prod_{i=1}^{r} \gamma_i^{(v)} = 1$; andernfalls wäre nämlich $N_{E/k}(\beta) = \alpha$ an der reellen Stelle v von k negativ, was wegen $\alpha \in \varphi_v((Z_y \otimes_k k_v)^*)$ bei $v_1 = r$ nicht sein kann. Damit existiert also ein $\gamma^{(v)} \in ({}^1R_{E/k} G_m)(k_v)$ mit $\beta \cdot \gamma^{(v)} > 0$ an allen $w \mid v$, die reell und verzweigt sind. Ein solches $\gamma^{(v)}$ verschaffe man sich für alle reellen Primstellen v von k. Wendet man dann den schwachen Approximationssatz auf den Torus $T = {}^1R_{E/k} G_m$ bezüglich der reellen Stellen von k an [vgl. 11, S. 45, 5.1], so findet man ein $\gamma \in ({}^1R_{E/k} G_m)(k)$ mit: $\beta\gamma > 0$ für alle reellen verzweigten Primstellen von E; und da $N_{E/k}(\beta\gamma) = N_{E/k}(\beta) = \alpha$ ist, beweist das die Behauptung. □

Bemerkung 3. Der oben zitierte schwache Approximationssatz für Tori ist in [11] ohne Beweis angegeben. Dieser Satz ergibt sich als Folgerung aus den Resultaten von Nakayama und Tate [15], vgl. dazu insbesondere [12, Chap. III, Theorem 6b].

Bemerkung 4. Die Voraussetzung $[D: k] < 36$ ist beim Beweis von Satz 3 nur bei der Heranziehung von Satz 2 benötigt worden. Würde Satz 2 ohne jede Gradeinschränkung richtig sein, so würde also auch Satz 3 ohne jede Voraussetzung über den Grad von D gelten.

Danksagung. Für die Verkürzung des ursprünglichen Beweises von Satz 1 danke ich dem Referenten.

Literatur

1. Artin, E., Tate, J.: Class field theory. New York: Benjamin 1967
2. Bartels, H.-J.: Zur Arithmetik von Konjugationsklassen in algebraischen Gruppen. J. Algebra **70**, 179–199 (1981)
3. Cassels, J.W.S., Fröhlich, A.: Algebraic number theory. London: Academic Press 1967
4. Draxl, P., Kneser, M.: SK_1 von Schiefkörpern. Berlin, Heidelberg, New York: Springer 1980
5. Gurak, S.: On the Hasse norm principle. J. reine angew. Math. **299/300**, 16–27 (1978)
6. Gurak, S.: The Hasse norm principle in non-abelian extensions. J. reine angew. Math. **303/304**, 314–318 (1978)
7. Harder, G.: Über die Galois-Kohomologie der Tori. Dissertation, Hamburg 1963
8. Harder, G.: Über die Galoiskohomologie halbeinfacher Matrizengruppen II. Math. Z. **92**, 396–415 (1966)
9. Harder, G.: Halbeinfache Gruppenschemata über Dedekindringen. Invent. Math. **4**, 165–191 (1967)
10. Heider, F.-P.: Zur Theorie der zahlentheoretischen Knoten. Dissertation, Köln 1978
11. Kneser, M.: Schwache Approximation in algebraischen Gruppen, Colloque sur la théorie des groupes algébriques. CBRM, Brüssel 1962
12. Kneser, M.: Lectures on Galois cohomology of classical groups. Tata Institute of Fundamental Research, Bombay 1969
13. Ono, T.: On the Tamagawa number of algebraic tori. Ann. Math. **78**, 47–73 (1963)
14. Schacher, M.: Subfields of division rings I. J. Algebra **9**, 451–477 (1968)
15. Tate, J.: The cohomology groups of tori in finite galois extensions of number fields. Nagoya Math. J. **27**, 709–719 (1966)

Eingegangen am 22. September 1980; revidierte Fassung am 5. Februar 1981

Math. Ann. 256, 475–495 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Projective Degenerations of Enriques' Surfaces

Jayant Shah*

Department of Mathematics, Northeastern University, Boston, MA02115, USA

1. Introduction

This paper is essentially another version of Horikawa's analysis [1]. The purpose here is to obtain a complete list of ("minimal") projective degenerations analogous to the list obtained by Morrison [2] by the analytic method of Kulikov, Persson and Pinkham. (Morrison's list is reproduced in the appendix.) Recall that a nonsingular projective surface, F, is called an Enriques' surface if $\dim H^1(F, o_F) = \dim H^2(F, o_F) = 0$ and $2K_F = 0$ where K_F denotes the canonical divisor class of F. An Enriques' surfaces may be obtained as the quotient of a hyperelliptic K3 surface of degree 4 by an involution [1]. We may study degenerations of an Enriques' surface by considering degenerations of the corresponding K3 surface. Degenerations of K3 surfaces of degree 4 are described in [5]. We repeat the analysis keeping track of the involution.

Let $S = \operatorname{Spec} \mathbb{C}[\![t]\!]$. Let 0 denote the closed point of S and let $*$ denote the generic point. By a family of surfaces over S, we mean a flat projective morphism $\mathfrak{f}: X \to S$ such that the generic geometric fiber of $\mathfrak{f}$ is a nonsingular connected surface. We let X_0 denote the special fiber of $\mathfrak{f}$ over 0 and let X_* denote the generic fiber of $\mathfrak{f}$. Let $\bar{X}_*$ denote the geometric generic fiber of $\mathfrak{f}$. A family of surfaces, $\mathfrak{f}': X' \to S$, is called a (projective) modification of the family $\mathfrak{f}: X \to S$ if there exists a map $\rho: S \to S$ such that X'_* and the pullback of X_* via ρ are isomorphic over the generic point of S. If there is an involution $\iota: X \to X$ over S, the modification $\mathfrak{f}': X' \to S$ is called an ι-modification if there is an involution $\iota': X' \to X'$ over S such that the pull-backs of ι and ι' to X'_* agree over $*$.

Let Σ_0 denote a nonsingular quadric in $\mathbb{P}_3$. Let Σ_2^0 denote a cone in $\mathbb{P}_3$ over a nonsingular conic. A hyperelliptic K3 surface of degree 4 is birationally a double cover of Σ_0 or Σ_2^0. Since a double cover of Σ_2^0 cannot specialize to a double cover of Σ_0, a sufficiently general Enriques' surface is the quotient of a double cover of Σ_0. Let T be a K3 surface which is a double cover of Σ_0,

* Partially supported by NSF Grant No. MCS79-03720

0025-5831/81/0256/0475/$04.20

endowed with an involution ι without fixed points such that T/ι is an Enriques' surface. The involution ι induces an involution I of $\mathbb{P}_3$ which preserves Σ_0. The involution ι may be recovered by applying I and interchanging the sheets of T over Σ_0 [1]. Let G be the stabilizer of Σ_0 in the group $\mathrm{PGL}_I(4)$ of automorphisms of $\mathbb{P}_3$ which commute with I.

Consider a family of double covers of Σ_0, $\mathfrak{f}\colon X\to S$, ramified over an I-invariant family of curves on Σ_0, such that X_* is a K3 surface. Let ι be the involution of X obtained by applying I and interchanging the sheets. It follows from the geometric invariant theory (see Proposition 2.1 in [4]) that there exists a family of double covers of Σ_0, $\mathfrak{f}'\colon X'\to S$, ramified over an I-invariant family of curves on Σ_0, such that it is an ι-modification of the family $\mathfrak{f}\colon X\to S$ and such that the special fiber X'_0 is semistable under the action of G. Moreover, X'_0 is unique if it belongs to a minimal orbit. The quotient of X' by the involution gives us a family of Enriques' surfaces whose special fiber is unique in the above sense. The trouble with this method of obtaining degenerations is that the special fiber may have significant limit singularities [3] in which case there may exist an ι-modification $\mathfrak{f}''\colon X''\to S$ such that X''_0 is birationally a K3 surface and the quotient of X''_0 by the involution is birationally an Enriques' surface. Consequently, if X'_0 has significant limit singularities, we deform the family under the action of a one-parameter subgroup of $\mathrm{PGL}_I(4)$ so as to obtain an ι-modification $\mathfrak{g}\colon Y\to S$ such that Y_0 is a double cover of Σ_2^0. Let G' denote the stabilizer of Σ_2^0 in $\mathrm{PGL}_I(4)$. We modify the family further under the action of G' to obtain an ι-modification $\mathfrak{g}'\colon Y'\to S$ such that Y'_0 is a double cover of Σ_2^0 and has only insignificant limit singularities.

Horikawa's goal in [1] is to analyze the image of the period map, for which one considers all families of Enriques' surfaces over the unit disk Δ. Horikawa applies Borel extension theorem to conclude that it is enough to analyze the special families considered in [1]. However, Borel's theorem requires that the period map be locally liftable and it is not clear that this is true in the case of Enriques' surfaces. In [4], we show how to calculate the monodromy of any family of surfaces $X\to\Delta$ in which the central fiber has only insignificant limit singularities of types listed in [3]. Thus, we recover Horikawa's Main Theorem without resorting to Borel extension theorem.

2. Terminology

2.1. Quadruple Points of Embedding Dimension 5

These singularities arise when a hyperelliptic K3 surface with a fixed-point-free involution degenerates into a (singular) surface on which the involution acquires fixed points. We will be interested in singularities which are quotients of rational double points, simple elliptic double points and non-normal limits of such singularities. Let $R'=\mathbb{C}[\![x,y]\!]$. Let I be the involution of R' which sends each variable to its negative. Let f be an I-invariant element of R'. Let $R=R'[z]/(z^2-f)$. Let ι be the involution of R which sends $x\mapsto -x$, $y\mapsto -y$, $z\mapsto -z$. Let $R_0=$the ring of invariants in R. R_0 is a Cohen-Macaulay ring of multiplicity 4 and embedding dimension 5.

Let $P=\operatorname{Spec} R$,
$P_0=\operatorname{Spec} R_0$,
$\tilde{P}=$ the disjoint union of the normalized components of P,
$\tilde{R}=$ the affine ring of $\tilde{P}$,
$\tilde{P}_0=$ the disjoint union of the normalized components of P_0,
$\tilde{R}_0=$ the affine ring of $\tilde{P}_0$,
$\Lambda=$ the singular locus of P,
$\tilde{\Lambda}=$ the inverse image of Λ in $\tilde{P}$,
$\Lambda_0=$ the singular locus of P_0,
$\tilde{\Lambda}_0=$ the inverse image of Λ_0 in $\tilde{P}_0$.

The involution ι extends to an involution $\tilde{\iota}$ of $\tilde{P}$ and $\tilde{P}_0\simeq\tilde{P}/\tilde{\iota}$. We will need the following cases.

A_n/ι: $f=y^2+x^n$, n even, so that P is a rational double point of type A_n. P_0 is a rational quadruple point.

A_∞/ι: $f=y^2$ and P is a point of type A_∞. The morphism $\tilde{P}\to P$ is given on the ring level by the homomorphism

$$\mathbb{C}[\![x,y,z]\!]/(z^2-x^2)\to\mathbb{C}[\![x,y]\!]\oplus\mathbb{C}[\![x,v]\!]\simeq\tilde{R}$$

sending $x\mapsto(x,x)$, $y\mapsto(y,y)$ and $z\mapsto(x,-x)$. $\tilde{\iota}$ is induced by the involution I on each factor of $\tilde{R}$. Therefore, $\tilde{P}_0$ is the disjoint union of two rational double points of type A_2.

$\tilde{E}_7/\iota$: $f=xy(x+y)(x+ay)$, $a\in\mathbb{C}$, $a\neq 0$ or 1, so that P is a simple elliptic double point of type $\tilde{E}_7$. P_0 is a rational quadruple point.

$T_{2,4,\infty}/\iota$: $f=y^2(x^2+y^2)$; the singularity of P is of type $T_{2,4,\infty}$.

$$\tilde{R}=\mathbb{C}[\![z/y,x,y]\!]/((z/y)^2-(x^2+y^2));$$

$\tilde{P}$ is a rational double point of type A_2. The involution $\tilde{\iota}$ sends $x\mapsto -x$, $y\mapsto -y$ and leaves z/y invariant. Therefore,

$$\begin{aligned}\tilde{R}_0&=\mathbb{C}[\![z/y,x^2,xy,y^2]\!]/((z/y)^2-(x^2+y^2))\\&\simeq\mathbb{C}[\![u,v,w]\!]/(w^2+u^2+v^4).\end{aligned}$$

$\tilde{P}_0$ is a rational double point of type A_4. Λ_0 and $\tilde{\Lambda}_0$ are nonsingular and connected.

$T_{2,\infty,\infty}/\iota$: $f=x^2y^2$; the singularity at P is of type $T_{2,\infty,\infty}$. $\tilde{R}\simeq R'\oplus R'$. The homomorphism $R\to\tilde{R}$ sends $x\mapsto(x,x)$, $y\mapsto(y,y)$ and $z\mapsto(xy,-xy)$. The involution $\tilde{\iota}$ interchanges the factors and sends $x\mapsto -x$, $y\mapsto -y$ in each factor. $\tilde{R}_0\simeq\mathbb{C}[\![u,v]\!]$ such that the inclusion $\tilde{R}_0\hookrightarrow\tilde{R}$ is given by sending $u\mapsto(x,-x)$ and $v\mapsto(y,-y)$. Λ_0 and $\tilde{\Lambda}_0$ are isomorphic to $\operatorname{Spec}\mathbb{C}[\![\alpha,\beta]\!]/(\alpha\beta)$.

2.2. *Types of (Possibly Singular) Surfaces*

We recall some definitions from [5]. Let V be a projective (possibly singular) surface over $\mathbb{C}$. Assume that V has only insignificant limit singularities. Let $h^{p,q}(V)=$ the dimension of the (p,q)-component of the mixed Hodge structure on $H^2(V,\mathbb{Q})$. Then, V is called *a surface of Type I* if $h^{0,0}(V)=h^{1,0}(V)=h^{0,1}(V)$

$=0$ (that is, the mixed Hodge structure on $H^2(V, \mathbb{Q})$ is a pure Hodge structure). V is called *a surface of Type II* if $h^{0,0}(V)=0$, but $h^{1,0}(V)\neq 0$. It is called *a surface of Type III* if $h^{0,0}(V)\neq 0$.

A singular projective surface, V, is called a *(singular) K3 surface* if it can be deformed into a nonsingular K3 surface. V is called a *(singular) Enriques' surface* if it can be deformed into a nonsingular Enriques' surface. Let V be an Enriques' surface. It is called *a surface of Type I*, II* or III** if it is the quotient by an involution of a K3 surface of Type I, II or III, respectively. V is said to be *of Class* α if the involution is without fixed points. V is said to be *of Class* β if the involution has fixed points.

2.3. Chains and Cycles of Rational Curves

Let n be an integer greater than 1. Let C be a reduced, connected, projective curve which is the union of n nonsingular rational curves such that the singularities of C consist of ordinary double points. C is called *a chain of n rational curves* if its dual graph is homeomorphic to the unit interval. C is called *a cycle of n rational curves* if its dual graph is homeomorphic to a circle. A rational curve whose singular locus consists of one ordinary double point is called *a cycle of one rational curve.*

3. Statement of the Main Theorem

We will prove

Main Theorem. *Let* $\mathfrak{f}: X\to S$ *be a family of double covers of* Σ_0, *endowed with an involution* ι *over S, such that the involution has no fixed points on the generic geometric fiber* $\bar{X}_*$ *and the quotient of* $\bar{X}_*$ *by the involution is a nonsingular Enriques surface. Then, there exists a projective* ι*-modification* $\mathfrak{g}: Y\to S$ *such that the special fiber* Y_0 *is a double cover of* Σ_0 *or* Σ_2^0 *and has only insignificant limit singularities. Let F be the quotient of* Y_0 *by the involution. Then, F is one of the surfaces listed below where we have used the following notation:*

σ = *the image in F of the fixed points of the involution. (Cardinality of* $\sigma\leqq 4$ *and each point in* σ *is a quadruple point with embedding dimension equal to* 5.)

δ = *the singular locus of F.*

δ_0 = *the set of isolated singularities of F not in* σ.

δ_1 = *the union of the one-dimensional components of* δ.

$\tilde{F}$ = *the disjoint union of the normalized components of F.*

$\tilde{\delta}_1$ = *the inverse image of* δ_1 *in* $\tilde{F}$.

The list is as follows.

*Surfaces of Type I**

Class α; *type I*-*α: δ *consists of rational double points. Desingularization of F is a nonsingular Enriques' surface.*

Class β; *type I*-*β: *F is rational.* σ *consists of points of type* A_{2n}/ι, $1\leqq n\leqq 4$. $\delta=\sigma$ *plus some rational double points.*

*Surfaces of Type II**

Class α:

Type II-α_1: F is birationally ruled over an elliptic curve. δ consists of a simple elliptic double point and some rational double points.*

Type II-α_2: F is birationally ruled over an elliptic curve. δ consists of a nonsingular rational curve. F has 4 simple pinch points on δ_1 so that $\tilde{\delta}_1$ is a nonsingular elliptic curve.*

Type II-α_3: F is rational. δ consists of a nonsingular elliptic curve and some rational double points. F has no pinch points on δ_1 and $\tilde{\delta}_1$ is a nonsingular elliptic curve.*

Class β:

Type II-β_1: F is rational. δ consists of σ and some rational double points. σ consists of two rational quadruple points of type $\tilde{E}_7/\iota$.*

Type II-β_2: F is rational. δ_0 contains only rational double points. δ_1 is nonsingular and consists of one or two rational curves. On each component of δ_1, F has two simple pinch points and two points of type A_∞/ι; thus, $\tilde{\delta}_1$ is nonsingular and consists of rational curves. σ may have rational quadruple points of type A_2/ι outside of δ_1.*

Type II-β_3: F is the union of two rational components which intersect in δ. σ consists of 4 points of type A_∞/ι. δ is a nonsingular rational curve. F has no pinch points on δ away from σ.*

*Surfaces of Type III**

Class α:

Type III-α: F is rational. δ_0 contains only rational double points. δ_1 is a cycle of n rational curves where $n=1$ or 2. The singularity of F at the singular points of δ_1 is of the type $T_{2,\infty,\infty}$. F has no pinch points on the nonsingular part of δ_1. $\tilde{\delta}_1$ is a cycle of $2n$ rational curves.*

Class β:

Type III-β_1: F is rational. δ_0 contains only rational double points. δ_1 is a nonsingular rational curve. σ consists of two points of type A_∞/ι. F has one double pinch point on δ_1 away from σ so that $\tilde{\delta}_1$ is a chain of two rational curves.*

Type III-β_2: F is the union of two rational components which intersect in δ. δ is a chain of 2 rational curves and the singularity of F at the singular point of δ is of the type $T_{2,\infty,\infty}$. Each component of δ contains two points of σ which are of the type A_∞/ι. F has no other pinch points on δ.*

Type III-β_3: F is rational. δ_0 contains only rational double points. δ_1 is a nonsingular rational curve. It contains two points of σ of the type $T_{2,4,\infty}/\iota$. In addition, σ may contain two rational quadruple points of type A_2/ι. F has no pinch points on δ_1 away from σ. $\tilde{\delta}_1$ is a nonsingular rational curve.*

Type III-β_4: F is rational. δ_0 contains only rational double points. δ_1 is a cycle of $2n$ rational curves where $n \leqq 2$. σ equals the set of singular points of δ_1 and each point in σ is of the type $T_{2,\infty,\infty}/\iota$. F has no pinch points on δ_1 away from σ. $\tilde{\delta}_1$ is a cycle of $2n$ rational curves.*

Remark 3.1. Horikawa's Main Theorem in [1] concerning the image of the period map follows from the above theorem. Argue as in the proof of Corollary 6.2 in [4] and then apply Horikawa's Theorem 2.2 in [1].

Remark 3.2. We have not shown that the quadruple points listed in §2.1 are insignificant limit singularities. However, we have the following.

Proposition 3.3. *Let* $\mathfrak{g}$: $Y \to S$ *be the family of surfaces as described in the conclusion of the Main Theorem. Let* Z *be the quotient of* Y *by the involution and let* $Z \to S$ *denote the family of surfaces thus obtained. Let* $Z_\rho \to S$ *be a new family obtained by pulling back the family* $Z \to S$ *via a map* ρ: $S \to S$. *Let* $Z' \to Z_\rho$ *be a resolution of singularities. Then every exceptional divisor in* Z' *is birationally ruled.*

Proof. Let $Y_\rho \to S$ be the family of surfaces obtained by pulling back the family $\mathfrak{g}$: $Y \to S$ via ρ. Let $Y' = Y_\rho \underset{Z_\rho}{\times} Z'$. Let $Y'' \to Y'$ be a resolution of singularities. Let π'': $Y'' \to Y_\rho$ be the canonical morphism. Y'' is a resolution of singularities of Y_ρ also. Since the singularities of Y_ρ are insignificant limit singularities [3], every exceptional divisor E in Y'' (i.e. $\dim \pi''(E) < \dim(E)$) is birationally ruled. Since every exceptional divisor in Z' is dominated by an exceptional divisor in Y'', all the exceptional divisors in Z' are birationally ruled. Q.E.D.

4. Representation of Enriques' Surfaces

We recall Horikawa's description [1]. Let F be a nonsingular Enriques' surface. Let T be its universal covering. T is a K3 surface. T is birationally equivalent to a branched double cover of Σ_0 or Σ_2^0 such that the singularities of the double cover consist of rational double points. Let $\{x_0, x_1, x_2, x_3\}$ be a basis of $H^0(\mathbb{P}_3, o_{\mathbb{P}_3}(1))$. Let I be the involution which keeps x_0 and x_3 fixed and sends x_1 and x_2 to $-x_1$ and $-x_2$ respectively. Let

$$\mathbb{W}^+ = \text{the space spanned by } x_0,\ x_3 \text{ and}$$

$$\mathbb{W}^- = \text{the space spanned by } x_1,\ x_2.$$

Then

$$(H^0(\mathbb{P}_3, o_{\mathbb{P}_3}(2)))^I = S^2(\mathbb{W}^+) \oplus S^2(\mathbb{W}^-)$$

and

$$(H^0(\mathbb{P}_3, o_{\mathbb{P}_3}(4)))^I = S^4(\mathbb{W}^+) \oplus S^2(\mathbb{W}^+) \otimes S^2(\mathbb{W}^-) \oplus S^4(\mathbb{W}^-).$$

For $i \geqq 1$, let $\mathbb{V}_i = \mathbb{C} \cdot x_0^i \oplus \mathbb{C} \cdot x_3^i$. Let $\mathbb{B} = \mathbb{V}_4 \oplus \mathbb{V}_2 \otimes S^2(\mathbb{W}^-) \oplus S^4(\mathbb{W}^-)$.

Let Q be the quadric defined by the equation $q = x_0 x_3 - q'' = 0$ where q'' is a homogeneous polynomial of degree 2 in variables x_1, x_2. q is invariant under I and we have the isomorphism $\mathbb{B} \simeq (H^0(Q, o_Q(4)))^I$. Choose coordinates so that

T is a double cover of Q. Then, T is ramified over a curve B defined by an equation $f=0$ where $f\in\mathbb{B}$. B does not pass through the fixed points of I. The covering transformation ι of T over F is obtained by applying I and interchanging the sheets of T over Q.

5. Geometric Stability of Double Covers of Σ_0

Let $\{u_0, u_1\}$ be the coordinates on the first factor of $\mathbb{P}_1\times\mathbb{P}_1$ and let $\{v_0, v_1\}$ be the coordinates on the second factor. Let $\mathbb{P}_1\times\mathbb{P}_1\to\mathbb{P}_3$ be the Segre embedding; let Σ_0 denote the image. Let $x_0=u_0v_0$, $x_1=u_0v_1$, $x_2=u_1v_0$, $x_3=u_1v_1$. Let $u=u_1/u_0$ and $v=v_1/v_0$. Let I_1 be the involution which keeps u_0 fixed and sends $u_1\mapsto -u_1$. Let I_2 be the involution which keeps v_0 fixed and sends $v_1\mapsto -v_1$. Let $I=$the involution $I_1\otimes I_2$.

Let γ denote the automorphism of $\mathbb{P}_1\times\mathbb{P}_1$ which interchanges the factors; let $\langle\gamma\rangle$ be the group generated by γ. For $i=1,2$, let G_i be the subgroup of PGL(2) which commutes with I_i; G_i is the stabilizer of the set of fixed points of I_i. Let G_m denote the one-dimensional multiplicative group. Then, G_1 and G_2 are isomorphic to the semidirect product $G_m\cdot(\mathbb{Z}/2)$ where G_m acts via the transformations $\{u_0\mapsto au_0,\ u_1\mapsto u_1/a\}$ and $\{v_0\mapsto av_0,\ v_1\mapsto v_1/a\}$ where $a\in\mathbb{C}$. $\mathbb{Z}/2$ is generated by the involution which interchanges u_0, u_1 as an element of G_1 and interchanges v_0, v_1 as an element of G_2. Let G be the stabilizer of Σ_0 in the group $\mathrm{PGL}_I(4)$ of automorphisms of $\mathbb{P}_3$ which commute with I. $G\simeq(G_1\times G_2)\cdot\langle\gamma\rangle$ and hence, G acts on $H^0(\mathbb{P}_3, o_{\mathbb{P}_3}(1))$, $\mathbb{B}$ and $|\mathbb{B}|$.

Let Δ denote the set of fixed points of I on Σ_0.

$$\Delta=\{(0,0),(0,\infty),(\infty,0),(\infty,\infty)\}.$$

The four lines, $u=0$, $u=\infty$, $v=0$, $v=\infty$, are called the edges. Let B be a curve on Σ_0 corresponding to a point in $|\mathbb{B}|$. An easy calculation gives the two propositions below (cf. [5]):

Proposition 5.1. *B is unstable if and only if either B has a point of multiplicity >4 (which must necessarily be in Δ) or it has a quadruple point in Δ with an edge as a tangent of multiplicity $\geqq 3$ at that point.*

Proposition 5.2. *Suppose that B is semistable.*

(a) B is strictly semistable (i.e. semistable, but not stable) if and only if either (i) an edge is a component of B with multiplicity 2 or (ii) B has a quadruple point in Δ.

(b) B is strictly semistable and belongs to a minimal orbit if and only if B is one of the following curves:

(i) $B=2L_1+2L_2+B_0$ where L_1 and L_2 are skew edges and the components of B_0 are mutually disjoint lines, none which is an edge.

(ii) B has two quadruple points, P and P', in Δ and B consists of four I-invariant curves of bidegree $(1,1)$, each of which passes through the points P and P'. Moreover, B does not contain an edge as a component with multiplicity 2.

(iii) B contains each edge with multiplicity 2.

Theorem 5.3. *Let B be a curve on Σ_0 corresponding to a point in $|\mathbb{B}|^{ss}$ belonging to a minimal orbit. Let T be the double cover of Σ_0 ramified over B. Let $F = T/\iota$ where ι is the involution on T defined by applying I and interchanging the sheets of T. Let σ, δ, δ_0, δ_1, $\tilde{\delta}_1$ be as in the statement of the Main Theorem. Then we have the following cases:*

*Surfaces of Type I**

B is reduced and does not have either consecutive triple points or points of multiplicity $\geqq 4$. F belongs to Class α if B does not meet Δ and it belongs to Class β otherwise.

*Surfaces of Type II**

B reduced.

(1) Singularities of B consist of two infinitely near triple points, P and $I(P)$. $B = C + I(C) + L + I(L)$ where C is a twisted cubic and L is a line such that C, $I(C)$ and L are mutually tangent at P. B does not meet Δ. F is a surface of type II^*-α_1.

(2) B has two quadruple points, P and P'. B consists of four distinct curves of bidegree $(1,1)$, each passing through P and P'. Either none of the points, P and P', is in Δ and F is of type II^*-α_1 or both the points, P and P', are in Δ and F is of type II^*-β_1.

B not reduced. B contains a double component, but has no component of multiplicity >2. Let C be an irreducible double component of maximal degree.

(3) C has bidegree $(2,2)$ and the equation of C is I-invariant. C is a nonsingular elliptic curve not meeting Δ. F is of type II^*-α_3.

(4) C has bidegree $(2,2)$ and the equation of C is anti-invariant under I. C is a nonsingular elliptic curve containing Δ. F is of type II^*-β_3.

(5) C is a twisted cubic. $B = 2C + L + I(L)$ where L is a line which is not an edge such that $C \cap L$ consists of two distinct points. C is I-invariant and contains exactly two points of Δ which are connected by an edge. F is of type II^*-β_2.

(6) C has bidegree $(1,1)$ and is I-invariant. C passes through two points of Δ not connected by an edge. $B = 2C + B_0$ such that $B_0 \cap C$ is disjoint from Δ and consists of 4 distinct points. F is of type II^*-β_2.

(7) C is a line. B consists of lines; $B = 2C + 2C' + L + I(L) + L' + I(L')$ where C and C' are skew, L and L' are skew and neither L nor L' is an edge. If C is not an edge, $C' = I(C)$ and F is of type II^*-α_2. If C is an edge, C' is also an edge and F is of type II^*-β_2.

*Surfaces of Type III**

B has a double component, but has no component of multiplicity >2. Let C be an irreducible double component of maximal degree.

(1) C is a twisted cubic. $B = 2C + L + I(L)$ where L is a line which is not an edge such that L is tangent to C. C contains exactly 2 points of Δ which are connected by an edge. F is of type III^*-β_1.

(2) C has bidegree $(1,1)$ and does not meet Δ. C is not I-invariant and $B = 2C + 2I(C)$. C and $I(C)$ intersect transversely in two distinct points. F is of type III-α.*

(3) C is an I-invariant curve of bidegree $(1,1)$, passing through two points of Δ not connected by an edge. $B = 2C + B_0$ such that $B_0 \cap C$ is disjoint from Δ.

(3a) B_0 is reduced. Either B_0 is tangent to C at two points, P and $I(P)$ or $B_0 = C' + I(C')$ such that C' and $I(C')$ intersect transversely in two points which lie on C. F is of type III-β_1.*

(3b) $B_0 = 2C'$ where C' is I-invariant and passes through two points of Δ such that $C \cap C'$ consists of two distinct points. F is of type III-β_2.*

(4) B has two quadruple points, P and P', in Δ and consists of I-invariant curves of bidegree $(1,1)$, each of which passes through P and P'. $B = 2C + C' + C''$ where C' and C'' are distinct from C.

(4a) C' and C'' are distinct. F is of type III-β_3.*

(4b) $C' = C''$. F is of type III-β_4.*

(5) B_{reduced} consists of 4 distinct lines, L_1, L_2, L_3, L_4, such that either none of the lines is an edge, $L_2 = I(L_1)$, $L_4 = I(L_3)$ and B does not meet Δ or each of the lines is an edge. F is of type III-α if B does not meet Δ and it is of type III*-β_4 otherwise.*

Surfaces with Significant Limit Singularities

B has a component C of multiplicity >2. $B = 3C + C'$ where C and C' are I-invariant curves of bidegree $(1,1)$ and C is nonsingular. B has two quadruple points contained in Δ and C, C' pass through these points. F has a cuspidal curve if $C \neq C'$ and it has a tacnodal curve if $C = C'$.

Proof. Let ε = the singular locus of T,

ε_1 = the union of the one-dimensional components of ε,
$\tilde{T}$ = the union of the normalized components of T,
$\tilde{\varepsilon}_1$ = the inverse image of ε_1 in $\tilde{T}$.

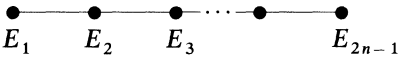
The involution ι lifts to an involution $\tilde{\iota}$ of $\tilde{T}$ and $\tilde{F} \simeq \tilde{T}/\tilde{\iota}$.

Below, in checking the geometry of the curves B, the following observation is useful. Let C be an I-invariant curve on Σ_0 of bidegree $(2n, 2n)$, $n \geqq 1$, defined by an equation $f = 0$ where f is an element of $H^0(\Sigma_0, o_{\Sigma_0}(2n))$. Every point in Δ is contained in an invariant, open, affine subset of Σ_0 which is isomorphic to $\operatorname{Spec} \mathbb{C}[x, y]$ with involution defined by the transformation $x \mapsto -x$, $y \mapsto -y$. Moreover, if $\bar{f}$ is the image of f in $\mathbb{C}[x, y]$, then, $\bar{f}$ is invariant (respectively, anti-invariant) if f is invariant (respectively, anti-invariant). It follows that at every point in Δ, C has even multiplicity ($\geqq 0$) if f is invariant and odd multiplicity if f is anti-invariant. Note also that any I-invariant curve C intersects an edge in an I-invariant divisor. If the degree of the divisor is odd, it must contain a point of Δ.

A. Suppose that B is Reduced

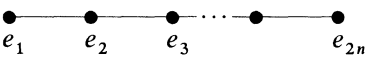
A-1. T has only rational double points as singularities if and only if B has neither consecutive triple points nor points of multiplicity $\geqq 4$. Such a curve is

stable. If T has only rational double points, it is a surface of Type I. (The type of a singular K3 surface is determined by the method described in [4].) B does not meet Δ if and only if ι has no fixed points. Therefore, if T has only rational double points and B does not meet Δ, F has only rational double points as singularities and the minimal desingularization of F is a nonsingular Enriques' surface. Suppose now that T has only rational double points as singularities and B meets Δ. Clearly, σ consists of rational quadruple points of type A_{2n}/ι, $n \leqq 4$. Let p be a point in σ. Let P be the point in T above p. $\hat{o}_{T,P} \simeq \mathbb{C}[\![x, y, z]\!]/(yz + x^{2n})$ and the involution acts by sending each variable to its negative. Let $T' \to T$ be the minimal resolution of singularities. T' is a K3 surface. The dual graph of the exceptional divisor E above P is of the form



$$E_i \cdot E_i = -2.$$

The involution ι extends to an involution ι' of T'. It is easy to check that ι' acts on E_i trivially if i is odd and nontrivially if i is even. In particular, ι' has no isolated fixed points and T'/ι' is nonsingular. Let $F' = T'/\iota'$. F' is a resolution of singularities of F. Let e be the reduced image of E. The dual graph of e is of the form



$$e_i \cdot e_i = \begin{cases} -4 & \text{if } i \text{ is odd} \\ -1 & \text{if } i \text{ is even} \end{cases}$$

T' is a double cover of F' ramified over $\bigcup_{i \text{ odd}} e_i$. Let K'_T be the canonical divisor class of T' and let K'_F be the canonical divisor class of F'. Let $\tau\colon T' \to F'$ be the canonical morphism. We have the formula $0 = K'_T = \tau^* K'_F + \bigcup_{i \text{ odd}} [E_i]$ where $[E_i]$ denotes the divisor class of E_i. Therefore, for all positive integers r, $\tau^*(rK'_F) = -r \bigcup_{i \text{ odd}} [E_i]$. It follows that $\dim(H^0(F', rK'_F)) = 0$. Since F has only rational singularities, $\chi(F', o_{F'}) = \chi(F, o_F) = 1$. Therefore, $\dim(H^1(F', o_{F'})) = 0$. Hence, F is a rational surface.

A-2. B has consecutive triple points.
By Lemma 4.2 in [1], B is of the form as claimed in the theorem and is stable. B does not meet Δ. The singular locus ε of T consists of two simple elliptic double points and hence T is birationally ruled over an elliptic curve. Since ι has no fixed points, δ consists of a single simple elliptic double point. Being dominated by a ruled surface, F is birationally ruled. Let F' be a desingularization of F. $\chi(F', o_{F'}) = \chi(F, o_F) - 1 = 0$. Therefore, $\dim(H^1(F', o_{F'})) = 0$ and F is birationally ruled over an elliptic curve.

A-3. B has quadruple points, none of which is in Δ.
By Lemma 4.10 in [1], B has two quadruple points, P and $I(P)$ and $B = C + I(C) + C' + I(C')$ where C and C' are curves of bidegree $(1, 1)$, passing through the points P and $I(P)$. B does not meet Δ and is stable. ε consists of two simple elliptic double points and some rational double points. Therefore this case is similar to Case A-2.

A-4. B has a quadruple point in Δ.
B is strictly semistable. Since B belongs to a minimal orbit, B has two quadruple points, P and P', in Δ and $B=C_1+C_2+C_3+C_4$ where each C_i is a $(1,1)$-curve, passing through the points P and P'. B may have ordinary double points at points in Δ other than P and P'. T is again a surface of the same type as in Case A-3. Therefore, F must be birationally ruled. Since F has only rational singularities, $\chi(F', o_{F'})=\chi(F, o_F)=1$ where F' is a desingularization of F. Hence, F is a rational surface.

B. Suppose that B has a multiple component. Let C be the component of maximal degree among the irreducible components of B with the greatest multiplicity.

B-1. C has bidegree $(2,2)$. $B=2C$. Since C is irreducible, C can have at most one singular point which then would have to be in Δ. But then, B would be strictly semistable and reducible. Therefore, C must be nonsingular and B must be stable.

B-1a. The equation of C is I-invariant.
C cannot meet Δ as it would have a singular point there. T consists of two copies of Σ_0, intersecting transversely in ε which is a nonsingular elliptic curve. The involution $\tilde{\imath}$ interchanges the sheets of $\tilde{T}$ and hence, $\tilde{F}\simeq\Sigma_0$. Clearly, δ and $\tilde{\delta}$ are nonsingular elliptic curves and F has no pinch points on δ.

B-1b. The equation of C is anti-invariant under I.
T is the same type of surface as in Case B-1a above. But, C must contain Δ. Therefore, σ consists of 4 points of type $A_\infty/\imath$. Above each point in σ, F has two distinct points. Therefore, $\tilde{\imath}$ must leave each sheet of $\tilde{T}$ invariant. Hence, $\tilde{F}$ consists of two rational components intersecting in δ. Since $\delta=\varepsilon/\imath$, δ is a nonsingular rational curve. F has no pinch points on δ away from σ.

B-2. C is a twisted cubic. $B=2C+L+L'$ where L and L' are lines. There are two skew edges on each of which C cuts out a divisor of degree 1. Therefore, C must pass through exactly two points, P and P', in Δ. Since C must cut out divisors of degree 2 on the remaining edges, P and P' must be connected by an edge. L cannot be an edge because, otherwise, L must equal L' to produce a point of even multiplicity in Δ and B would not be minimal. Hence, $L'=I(L)$. B is thus stable.

B-2a. $L\cap C=$two distinct points. T and hence F are rational surfaces. ε is a nonsingular rational curve and T has four simple pinch points on ε. $\imath$ has two fixed points on ε and hence δ is a nonsingular rational curve. σ consists of two points of type $A_\infty/\imath$. F has two simple pinch points on δ away from σ and $\tilde{\delta}$ is a nonsingular rational curve.

B-2b. L is tangent to C. This case is similar to the case above except that now T has two double pinch points on ε. Therefore, $\tilde{\varepsilon}$ is a cycle of two rational curves and $\tilde{\delta}$ is a chain of two rational curves.

B-3. C is an I-invariant double component of bidegree $(1,1)$. $B=2C+B'$. Since the equation of $2C$ is I-invariant, the equation of B' must be I-invariant. C must cut every edge in an I-invariant point and hence, C must pass through exactly two points of Δ and these points cannot lie on the same edge.

B-3a. B' intersects C outside of Δ. B' may however meet Δ. B is stable.

B-3a_1. $C\cap B'$ consists of four distinct points. This case is similar to Case B-2a.

B-3a_2. B' is reduced and $C\cap B'$ consists of two points, each with multiplicity 2. B' is either tangent to C at two points or $B'=C'+I(C')$ such that C' and $I(C')$ intersect on C at two points. This case is similar to Case B-2b.

B-3a_3. $B'=2C'$ where C' is an I-invariant nonsingular curve of bidegree $(1,1)$. T consists of two copies of Σ_0 intersecting in a cycle of two rational curves. $\delta=\varepsilon/\iota$ and consists of a chain of two rational curves. Each component of δ has two points of type A_∞/ι. $\tilde{\iota}$ leaves each sheet of $\tilde{T}$ invariant, as in Case 3-1b. Hence, F consists of two rational components.

B-3b. B' and C meet in Δ. B has a quadruple point in Δ and hence, is strictly semistable. $B'=C'+C''$ where C' and C'' are I-invariant curves of bidegree $(1,1)$ and $C\cap C'=C\cap C''\subset\Delta$.

B-3b_1. B' is reduced. T and hence F are rational surfaces. δ_1 is a nonsingular rational curve. F has two points of type $T_{2,4,\infty}/\iota$ on δ_1 and no other pinch points. $\tilde{\delta}_1$ is a nonsingular rational curve. σ may have one or two points of type A_2/ι.

B-3b_2. $C'=C''$. ε consists of a cycle of two rational curves. T consists of two copies of Σ_0 intersecting in ε. ε has two fixed points of I. Hence, δ also is a cycle of two rational curves. Since $\tilde{F}$ has a single point above each point in σ, $\tilde{\iota}$ must interchange the sheets of $\tilde{T}$. Therefore, $\tilde{F}\simeq\Sigma_0$ and $\tilde{\delta}$ is a cycle of two rational curves.

B-4. C is a component of multiplicity $\geqq 3$. C must be I-invariant and cannot be an edge. Therefore, C must be a $(1,1)$-curve, passing through two points of Δ. At these points, B must have quadruple points. Therefore, B must be strictly semistable and $B=3C+C'$ where C' is I-invariant and $C\cap C'\subset\Delta$. F has a cuspidal nodal curve if $C\neq C'$ and it has a tacnodal curve if $C=C'$.

B-5. C has bidegree $(1,1)$ and it is not invariant under I. Therefore, $B=2C+2I(C)$ where C does not meet Δ. B is stable. A point in $C\cap I(C)$ cannot be fixed under I and hence, $C\cap I(C)$ consists of two distinct points. T consists of two copies of Σ_0 intersecting in a cycle of two rational curves. ι has no fixed points and $\tilde{\iota}$ must interchange the sheets of $\tilde{T}$. Therefore, δ is a cycle of one rational curve, $\tilde{\delta}$ is a cycle of 2 rational curves and $\tilde{F}\simeq\Sigma_0$.

B-6. C is a line.

B-6a. B does not contain an edge as a double component. B must be of the form $2C+2I(C)+B'$ where the components of B' are mutually skew lines. No component of B' can be an edge since that edge would have to have multiplicity 2 in order to produce points of even multiplicity in Δ. Therefore, $B'=L+I(L)+L'+I(L')$. B does not meet Δ and B is stable.

B-6a$_1$. $L\neq L'$. $\tilde{T}\simeq E\times\mathbb{P}_1$ where E is a nonsingular elliptic curve. Since $\tilde{T}$ is an unramified cover of $\tilde{F}$, $\dim(H^1(\tilde{F}, o_{\tilde{F}}))=\dim(H^1(\tilde{T}, o_{\tilde{T}}))=1$. Therefore, F is birationally ruled over an elliptic curve. δ is a nonsingular rational curve and F has four simple pinch points on it.

B-6a$_2$. $L=L'$. T consists of two copies of Σ_0 intersecting in a cycle of 4 rational curves. $\tilde{\iota}$ interchanges the sheets of $\tilde{T}$. Therefore, $\tilde{F}\simeq\Sigma_0$, δ consists of a cycle of 2 rational curves and $\tilde{\delta}$ is a cycle of 4 rational curves.

B-6b. C is an edge. B is strictly semistable. By the argument given in Case B-6a, $B=2C+2C'+L+I(L)+L'+I(L')$ where C, C' are skew edges and L, L' are skew lines.

B-6b$_1$. Neither L nor L' is an edge. T is of the same type as in Case B-6a$_1$. However, δ consists of two disjoint, nonsingular, rational curves, on each of which F has two simple pinch points. Each component of δ has two points of σ, each of which is a point of type A_∞/ι. $\tilde{\delta}$ consists of two disjoint, nonsingular, rational curves. F must be a ruled surface. Moreover, $\chi(\tilde{F}, o_{\tilde{F}})=\chi(F, o_F)+\chi(\tilde{\delta}, o_{\tilde{\delta}})-\chi(\delta, o_\delta)=1$. Therefore, F is rational.

B-6b$_2$. L is an edge. L' must be an edge also. T is of the same type as in Case B-6a$_2$. δ is a cycle of four rational curves. Each singular point of δ is a point of type $T_{2,\infty,\infty}/\iota$ on F. Above each point of σ, F contains only one point. Therefore, $\tilde{\iota}$ must interchange the sheets of $\tilde{T}$. F is rational and $\tilde{\delta}$ is a cycle of 4 rational curves. Q.E.D.

6. Geometric Stability of Double Covers of Σ_2^0

Let Σ_2^0 be the quadric cone in $\mathbb{P}_3$ defined by the equation $q'=x_0x_3-x_1^2=0$. Let D denote the curve on Σ_2^0 defined by the equation $x_2=0$. Σ_2^0 is invariant under the involution I. Let p_0 be the point $(x_0=1, x_1=x_2=x_3=0)$ and let p_∞ be the point $(x_0=x_1=x_2=0, x_3=1)$. Let V denote the vertex of the cone. The lines on Σ_2^0 defined by the equation $x_1=0$ are called the edges. Let $\Delta'=$the set $\{p_0, p_\infty\}$. The set of fixed points of I on Σ_2^0 consists of p_0, p_∞ and V.

Let $G'=$the stabilizer of Σ_2^0 in $\mathrm{PGL}_I(4)$. There exist subgroups G_u, G_m and G_s of G' such that $G'\simeq G_u\cdot G_m\cdot G_s$ [4, 5]. G_u is the unipotent subgroup and consists of transformations which act trivially on the vector space $\mathbb{A}_1$ spanned by x_0, x_1, x_3 and takes x_2 to x_2+ax_1 for some complex number a. The spaces $\mathbb{C}\cdot x_2$ and $\mathbb{A}_1$ are invariant under G_m and G_s. $G_m\cdot G_s$ is the stabilizer of D in G'. G_s is the subgroup of $\mathrm{Aut}(D)$ which leaves the set $\{p_0, p_\infty\}$ invariant. G_m acts trivially on D and $\mathbb{A}_1$, and its action on $\mathbb{C}\cdot x_2$ determines a character. $\mathrm{Dim}\, G_u=\dim G_m=\dim G_s=1$.

Let $\mathbb{B}_2 = \mathbb{C} \cdot x_1^2 \oplus \mathbb{V}_2$, $\mathbb{B}_3 = x_1 \cdot \mathbb{B}_2$, $\mathbb{B}_4 = \mathbb{C} \cdot x_1^4 \oplus x_1^2 \mathbb{V}_2 \oplus \mathbb{V}_4$. Note that $\mathbb{B} = \mathbb{C} \cdot x_2^4 \oplus \mathbb{C} \cdot x_2^3 x_1 \oplus x_2^2 \mathbb{B}_2 \oplus x_2 \mathbb{B}_3 \oplus \mathbb{B}_4$. Let $\theta, \theta', \theta_0, \ldots$ denote elements of $\mathbb{B}_2$, $\phi, \phi', \phi_0, \ldots$ denote elements of $\mathbb{B}_3$, and $\xi, \xi', \xi_0, \ldots$ denote elements of $\mathbb{B}_4$. $a, a', b, \ldots$ will denote complex numbers.

Let B be a curve on Σ_2^0 defined by an equation $f=0$ where $f \in \mathbb{B}$. We will need to consider two cases:

1. B does not pass through the vertex.
2. B has an ordinary double point at the vertex.

Case 1. B does not pass through V.

B is defined by an equation of the form $f = x_2^4 + a x_2^3 x_1 + x_2^2 \theta + x_2 \phi + \xi = 0$. There exists $g_1 \in G_u$ such that $f^{g_1} = x_2^4 + x_2^2 \theta' + x_2 \phi' + \xi'$. Moreover, if $g_2 \in G_m$, $(f^{g_1})^{g_2} = \dfrac{1}{b^4}\{x_2^4 + b^2 x_2^2 \theta' + b^3 x_2 \phi' + b^4 \xi'\}$. Let $\Omega = \mathrm{Symm}(\mathbb{B}_2^* \oplus \mathbb{B}_3^* \oplus \mathbb{B}_4^*)$ where the superscript * denotes the dual of a vector space. Grade Ω by assigning weight i to $\mathbb{B}_i^*$. Then, modulo the action of G_s, B is determined by a point in $\mathrm{Proj}\,\Omega$. (See [4] for more details.)

For $2 \leqq i \leqq 4$, let p_{r_i}: $\mathrm{Proj}\,\Omega \cdots \to \mathbb{B}_i$ be the G_s-invariant rational map defined by the canonical projection. If $w \in \mathrm{Proj}\,\Omega$, let $p_{r_i}(w)$ denote the empty set if p_{r_i} is not defined at w. If $\alpha \in |\mathbb{B}_i|$, let $\bar{\alpha}$ denote its restriction to D; let $v_0(\bar{\alpha}) =$ the multiplicity of p_0 in the divisor $\bar{\alpha}$ and $v_\infty(\bar{\alpha}) =$ the multiplicity of p_∞ in $\bar{\alpha}$.

The following proposition is easy to check (cf. Proposition 4.2 in [4]).

Proposition 6.1. *Let $w \in \mathrm{Proj}\,\Omega$. w is unstable (under the action of G_s) if and only if either $v_0(\overline{p_{r_i}(w)}) \geqq i$ for $2 \leqq i \leqq 4$ or $v_\infty(\overline{p_{r_i}(w)}) \geqq i$ for $2 \leqq i \leqq 4$. w is strictly semistable if and only if either, for $2 \leqq i \leqq 4$, $v_0(\overline{p_{r_i}(w)}) = i$ if $p_{r_i}(w)$ is not empty or, $2 \leqq i \leqq 4$, $v_\infty(\overline{p_{r_i}(w)}) = i$ if $p_{r_i}(w)$ is not empty. w is strictly semistable and belongs to a minimal orbit if and only if for $2 \leqq i \leqq 4$, $v_0(\overline{p_{r_i}(w)}) = v_\infty(\overline{p_{r_i}(w)}) = i$ if $p_{r_i}(w)$ is not empty.*

Theorem 6.2. *Let B be a curve on Σ_2^0 corresponding to a point in $(\mathrm{Proj}\,\Omega)^{ss}$ belonging to a minimal orbit. Let T be the double cover of Σ_2^0 ramified over B. Let ι be the involution of T obtained by applying I and interchanging the sheets of T. Let $F = T/\iota$. Let σ, δ, δ_0 $\delta_1, \tilde{\delta}_1$, be as in the statement of the Main Theorem. Then we have the following cases:*

*Surfaces of Type I**

B is reduced and does not have either consecutive triple points or points of multiplicity $\geqq 4$. F belongs to Class α if B does not meet Δ' and it belongs to Class β otherwise.

*Surface of Type II**

(1) B is reduced and has two quadruple points, P and P'. B consists of 4 distinct curves of degree 2, each of which passes through P and P'. Either none of the

points, P and P', is in Δ' and F is of type II-α_1 or both the points, P and P', are in Δ' and F is of type II*-β_1.*

Cases when B is not reduced. B contains a double component, but has no component of multiplicity >2. Let C be an irreducible double component of maximal degree.

(2) Degree of $C=4$. $B=2C$. C is a nonsingular elliptic curve not meeting Δ'. F is of type II-α_3.*

(3) Degree of $C=2$ and C is a nonsingular I-invariant curve containing Δ'. $B=2C+B'$ such that B' does not meet Δ' and $C\cap B'$ consists of 4 distinct points. F is of type II-β_2.*

*Surface of Type III**

B has a double component, C, of degree 2.

(1) C does not meet Δ' and is not I-invariant. $B=2C+2I(C)$. C and $I(C)$ intersect transversely in two distinct points. F is of type III-α.*

(2) C is I-invariant, $B=2C+B'$ such that B' does not meet Δ'. B' is reduced and $C\cap B'$ consists of two distinct points, each with multiplicity 2. F is of type III-β_1.*

(3) *B has quadruple points at p_0 and p_∞. $B=2C+C'+C''$ where C, C', C'' are I-invariant curves of degree 2, each of which passes through p_0 and p_∞. C' and C'' are distinct from C.*

(3a) C' and C'' are distinct. F is of type III-β_3.*

(3b) $C'=C''$. F is of type III-β_4.*

Surfaces with Significant Limit Singularities

$B=3C+C'$ where C and C' are nonsingular, I-invariant, distinct curves of degree 2, containing Δ'. F has a cuspidal curve.

Proof. Since all the cases are similar to the ones in Theorem 5.3, we analyze only the geometry of the curve B.

A. Suppose that B is reduced. B cannot have consecutive triple points (Lemma 6.3, [1]).

A-1. B has no quadruple points. B is stable.

A-2. B has quadruple points, none of which is in Δ'. By Lemma 6.4 in [1], B has two quadruple points, P and $I(P)$, and $B=C+I(C)+C'+I(C')$ where C and C' are nonsingular curves of degree 2, passing through the points P and $I(P)$. B does not meet Δ' and is stable.

A-3. B has a quadruple point in Δ'. B is strictly semistable. Since B is minimal, B has two quadruple points, p_0 and p_∞, and $B=C_1+C_2+C_3+C_4$ where each C_i is a nonsingular curve of degree 2, containing Δ'.

B. Suppose that B has a multiple component. Let C be the component of maximal degree among the irreducible components of B with the greatest

multiplicity. Note that degree of C cannot be 1 or 3 because such a curve must pass through the vertex V.

B-1. Degree of $C=4$. $B=2C$. If the equation of C were anti-invariant under I, C would intersect each edge at points fixed by I and, in particular, C would pass through V. Therefore, the equation of C must be I-invariant and B does not meet Δ'. B is stable.

B-2. C is an I-invariant double component of degree 2, containing Δ'. $B=2C+B'$ where the equation of B' must be I-invariant. If B' is not reduced, then $B'=2C'$ where C' is I-invariant and hence must contain Δ'.

B-2a. B' intersects C outside of Δ'. B' is reduced and B is stable. $B'\cap C$ is an I-invariant divisor on C, defined by an I-invariant equation. Hence, either $B'\cap C$ consists of 4 distinct points or it equals $2P+2I(P)$.

B-2b. $B'\cap C\subset\Delta'$. B' must have double points at p_0 and p_∞. Therefore, B has quadruple points in Δ' and is strictly semistable. $B'=C'+C''$ where C', C'' are I-invariant curves which may or may not be distinct.

B-3. C is a component of multiplicity $=3$. C must be an I-invariant curve of degree 2 containing Δ'. B must have quadrupole points at p_0 and p_∞. Therefore, $B=3C+C'$ where C' is an I-invariant curve of degree 2 containing Δ'.

B-4. C is a curve of degree 2 which is not invariant under I. Therefore, $B=2C+2I(C)$ and B does not meet Δ'. B is stable.

Case 2. B has an ordinary double point at V.

B is defined by an equation of the form $f=x_2^3x_1+x_2^2\theta+x_2\phi+\xi=0$. There exists $g_1\in G_u$ such that $f^{g_1}=x_2^3x_1+x_2^2\theta'+x_2\phi'+\xi'$ where $\theta'\in\mathbb{V}_2$. Moreover, if $g_2\in G_m$, $(f^{g_1})^{g_2}=\frac{1}{b^3}\{x_2^3x_1+bx_2^2\theta'+b^2x_2\phi'+b^3\xi'\}$. Let $\Omega'=\mathrm{Symm}(\mathbb{V}_2^*\oplus\mathbb{B}_3^*\oplus\mathbb{B}_4^*)$. Grade Ω' by assigning weight 1 to $\mathbb{V}_2^*$, weight 2 to $\mathbb{B}_3^*$ and weight 3 to $\mathbb{B}_4^*$. Then, modulo the action of G_s, B is determined by a point in $\mathrm{Proj}\,\Omega'$. Proposition 6.1 applies here as well and gives the conditions for stability, semistability and minimality of orbits for points of $\mathrm{Proj}\,\Omega'$. The following theorem is proved by an analysis similar to that in the proof of Theorem 6.2.

Theorem 6.3. *Let B be a curve on Σ_2^0 corresponding to a point in $(\mathrm{Proj}\,\Omega')^{ss}$ belonging to a minimal orbit. Let T, ι, F be defined as in Theorem 6.2. Then we have the following cases:*

*Surfaces of Type I**

B is reduced and does not have either consecutive triple points or points of multiplicity $\geqq 4$. F is of type I-β.*

*Surfaces of Type II**

(1) B is reduced and has quadruple points at p_0 and p_∞. B consists of the two edges and 3 distinct, nonsingular, I-invariant curves of degree 2, each of which contains Δ'. F is of type II^*-β_1.

(2) B has a double component, C, which is a nonsingular, I-invariant curve of degree 2 containing Δ'. $B=2C+B'$ such that B' does not meet Δ' and $C\cap B'$ consists of 4 distinct points. F is of type II^*-β_2.

*Surfaces of Type III**

B has double component, C, which is a nonsingular, I-invariant curve of degree 2 containing Δ'.

(1) $B=2C+B'$ where B' does not meet Δ', B' is reduced and $C\cap B'$ consists of two distinct points, each with multiplicity 2. F is of type III^*-β_1.

(2) $B=2C+C'+C''$ where C' is a nonsingular, I-invariant curve of degree 2, containing Δ' and C'' consists of the two edges. C' is distinct from C. F is of type III^*-β_3.

Remark 6.4. The important point about Theorem 6.3 is that there are no semistable surfaces with significant limit singularities.

7. Proof of the Main Theorem

Let B_t be the ramification locus of X. B_t is a family of curves on Σ_0. Let B_0 be the fiber of B_t over the closed point, 0, of S. By Proposition 2.1 in [4], we may assume that B_0 is semistable and belongs to a minimal orbit. Then, the family $\mathfrak{f}\colon X\to S$ is already of the required form unless B_0 contains an I-invariant component, C, of multiplicity $\geqq 3$.

Choose coordinates as in Sect. 5. Since G acts transitively on $|\mathbb{W}^-|$, we may assume that C is defined by the equation $x_1-x_2=0$. Change coordinates in $\mathbb{P}_3$ as follows: $x_1\mapsto x_1+x_2$, $x_2\mapsto x_1-x_2$. Note that the coordinate transformation commutes with the involution I and leaves the vector spaces $\mathbb{W}^+$, $\mathbb{W}^-$, $\mathbb{V}_i$ and $\mathbb{B}$ invariant. Σ_0 is now defined by the equation $x_0x_3-x_1^2+x_2^2=0$. Let Σ_2^0 be the quadric cone defined by the equation $x_0x_3-x_1^2=0$. $\Sigma_0\cap\Sigma_2^0=C$ and C is cut out by the plane $x_2=0$ on Σ_0 and Σ_2^0. The point $(0,0)$ on Σ_0 and the point p_0 on Σ_2^0 coincide. Similarly, the points (∞,∞) and p_∞ coincide.

Let G_0 be the stabilizer of C in G. Since G fixes the vertex V of Σ_2^0, G_0 is a subgroup of G' under which C is invariant. Clearly, $G_0=G\cap G'$ and $G_0\simeq G_s\cdot(\mathbb{Z}/2)$ where the second factor is generated by an element of G_m which sends $x_2\mapsto -x_2$. Let λ_n be the one-parameter subgroup of G_m which sends $x_2\mapsto t^n x_2$.

We use the same notation for elements of $\mathbb{B}_2$, $\mathbb{B}_3$, $\mathbb{B}_4$ and $\mathbb{C}$ as in Sect. 6. Moreover, let $\theta_t, \theta_t', \ldots$ denote elements of $\mathbb{B}_2\otimes\mathbb{C}[\![t]\!]$, $\phi_t, \phi_t', \ldots$ denote elements of $\mathbb{B}_3\otimes\mathbb{C}[\![t]\!]$, $\xi_t, \xi_t', \ldots$ denote elements of $\mathbb{B}_4\otimes\mathbb{C}[\![t]\!]$ and $a_t, a_t', \ldots$ denote elements of $\mathbb{C}[\![t]\!]$.

Case 1. $B_0 = 4C$.
Let $\mathbb{N} = x_2^2 \cdot \mathbb{B}_2 \oplus x_2 \cdot \mathbb{B}_3 \oplus \mathbb{B}_4$ and $\mathbb{N}_t = \mathbb{N} \otimes \mathbb{C}[\![t]\!]$. The following lemma is analogous to (and easier than) Corollary 3.14 or Lemma 5.3 in [5]:

Lemma 7.1 *(standardization of the equation). The family of curves, B_t, is equivalent under the action of G to a family of curves defined by an equation of the form $x_2^4 + t h_t$ where $h_t \in \mathbb{N}_t$.*

Lemma 7.2 *(application of geometric invariant theory). After replacing t by a suitable root of t, the family B_t may be modified under the action of G_0 so that the new family is defined by an equation of the form $f_t = x_2^4 + t^{2n} x_2^2 \theta_t + t^{3n} x_2 \phi_t + t^{4n} \xi_t = 0$ such that $\{\theta_0, \phi_0, \xi_0\} = \lim_{t \to 0} \{\theta_t, \phi_t, \xi_t\} \neq 0$ and $\{\theta_0, \phi_0, \xi_0\}$ determines a point in $(\mathrm{Proj}\,\Omega)^{ss}$, belonging to a minimal orbit.*

(Proof is analogous to that of Lemma 3.16 in [5].)

Proposition 7.3. *Let $\mathfrak{f}: X \to S$ be a family of double covers of Σ_0 as in the Main Theorem such that $B_0 = 4C$. Then, there exists a projective ι-modification $\mathfrak{f}': X' \to S$ such that the special fiber X'_0 is a double cover of Σ_2^0 ramified over a curve defined by an equation of the form $x_2^4 + x_2^2 \theta_0 + x_2 \phi_0 + \xi_0 = 0$ where $\{\theta_0, \phi_0, \xi_0\}$ determines a point in $(\mathrm{Proj}\,\Omega)^{ss}$ belonging to a minimal orbit.*

Proof. Assume that the family $\mathfrak{f}: X \to S$ has already been modified so that the ramification locus B_t is defined by an equation $f_t = 0$ where f_t is of the form given in Lemma 7.2. Modify the trivial family of quadrics, $\Sigma_0 \times S$ under the action of the one-parameter subgroup λ_n. Let W_n be the new family. The geometric generic fiber of W_n is isomorphic to Σ_0 and the special fiber is isomorphic to Σ_2^0. The divisor B_t on $\Sigma_0 \times S$ is modified to a divisor B'_t on W_n defined by the equation $f'_t = x_2^4 + x_2^2 \theta_t + x_2 \phi_t + \xi_t = 0$. The double cover induced by the family $\mathfrak{f}: X \to S$ over the generic fiber of W_n extends to a double cover of W_n, ramified over the divisor B'_t (see Sect. 6 in [4]). Q.E.D.

The family $\mathfrak{f}': X' \to S$ of the proposition above is of the form required by the Main Theorem in all cases except when the special fiber B'_0 of the family B'_t has a triple component. In that case, the equation of B'_0 takes the form $(x_2 + a x_1)^3 (x_2 - 3a x_1) = 0$ where $a \neq 0$. But then, the family of curves B_t must have been defined by an equation of the form

$$f_t = (x_2 + a t^n x_1)^3 (x_2 - 3a t^n x_1) + t^{2n+1} x_2^2 \theta'_t + t^{3n+1} x_2 \phi'_t + t^{4n+1} \xi'_t = 0.$$

Under the action of G, B_t is equivalent to a family of curves defined by the equation $x_2^3 (a'' x_2 + t^n x_1) + t^{2n+1} x_2^2 \theta''_t + t^{3n+1} x_2 \phi''_t + t^{4n+1} \xi''_t = 0$ where $a'' \neq 0$. Thus, it remains to consider

Case 2. $\mathfrak{f}: X \to S$ is a family of double covers of Σ_0 as in the Main Theorem, ramified over a family of curves, B_t, which is defined by an equation of the form

$$a'_t x_2^4 + t^n x_2^3 x_1 + t^{2n+1} x_2^2 \theta'_t + t^{3n+1} x_2 \phi'_t + t^{4n+1} \xi'_t = 0$$

where $n \geqq 0$ and a'_t is a unit if $n > 0$.

Analogous to Lemma 7.1, we have

Lemma 7.4. *The family of curves B_t may be transformed under the action of G so that we may assume that $\theta_t' \in \mathbb{V}_2 \otimes \mathbb{C}[[t]]$ in the above equation.*

Analogous to Lemma 7.2, we have

Lemma 7.5. *The family of curves B_t may be modified under the action of G_0 so that the new family is defined by an equation of the form*

$$f_t = a_t x_2^4 + t^n x_2^3 x_1 + t^{2n+m} x_2^2 \theta_t + t^{3n+2m} x_2 \phi_t + t^{4n+3m} \xi_t = 0$$

such that m is even and positive, $\{\theta_0, \phi_0, \xi_0\} = \lim_{t \to 0} \{\theta_t, \phi_t, \xi_t\} \neq 0$ and $\{\theta_0, \phi_0, \xi_0\}$ determines a point in $(\mathrm{Proj}\, \Omega')^{ss}$ belonging to a minimal orbit.

Finally, analogous to Proposition 7.3, we have

Proposition 7.6. *Let $\mathfrak{f}: X \to S$ be a family of double covers of Σ_0, ramified over a family of curves which is defined by an equation of the form given in Lemma 7.5. Then, there exists an ι-modification $\mathfrak{g}: Y \to S$ such that the special fiber Y_0 is a double cover of Σ_2^0 ramified over a curve corresponding to a point in $(\mathrm{Proj}\, \Omega')^{ss}$ belonging to a minimal orbit.*

Appendix

D. Morrison's list of analytic degenerations of Enriques' surfaces.

Define a *marked surface* (V, D) to be an algebraic surface V together with a reduced effective divisor $D \subset V$ which has local normal crossings.

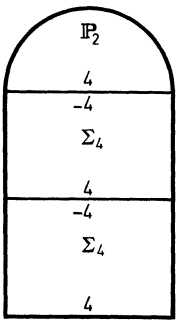
A *semistable degeneration of surfaces* is a proper holomorphic map $\pi: X \to \Delta$ from a three-dimensional complex manifold to a disk such that the fibers $X_t = \pi^{-1}(t)$ are connected and smooth for $t \neq 0$ and the central fiber $X_0 = \pi^{-1}(0)$ is a divisor consisting of smooth components with normal crossings. A theorem of Mumford asserts that after a base change, any degeneration is birational to a semistable one. A semistable degeneration is called *weakly projective* if it is birational to a projective degeneration. Let $\pi: X \to \Delta$ be a semistable degeneration of surfaces. Write $X_0 = \sum V_\alpha$ as the sum of the non-singular components. For $\alpha \neq \beta$, let $C_{\alpha\beta} = V_\alpha \cap V_\beta$. Let $D_\alpha = \bigcup_{\alpha \neq \beta} C_{\alpha\beta}$. D_α is called the marking divisor or the divisor of double curves.

Theorem (D. Morrison). *Let $X \to \Delta$ be a weakly projective, semistable degeneration of Enriques' surfaces. Then, X is birational to a semistable degeneration such that the central fiber is one of the following:*

(iα) *A Smooth Enriques' Surface*

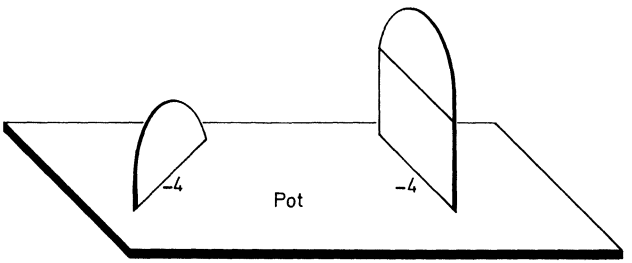
(iβ) *A Flower Pot*

A *flower* is a marked surface $(\mathbb{P}_2, C)$ with $C \subset \mathbb{P}_2$ a conic. A *stalk* is a marked surface $(\Sigma_4, s_0 + s_\infty)$ where Σ_4 is a relatively minimal, rational, ruled surface with a section s_0 of self-intersection 4 and a section s_∞ of self-intersection -4. A *stalk assembly* (of length k) consists of a flower joined to $k-1$ stalks, as shown below.



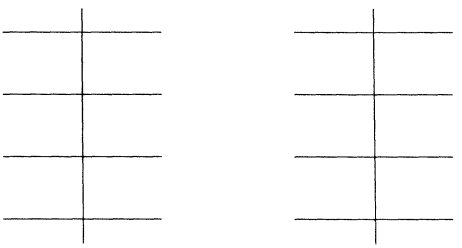
(The numbers are self-intersection numbers.)

A *pot* is a rational surface V, marked by $D=\Sigma C_\alpha$ with each C_α rational, and $C_\alpha \cdot C_\beta = -4\delta_{\alpha\beta}$. A *flower pot* consists of a pot meeting stalk assemblies along its double curves.

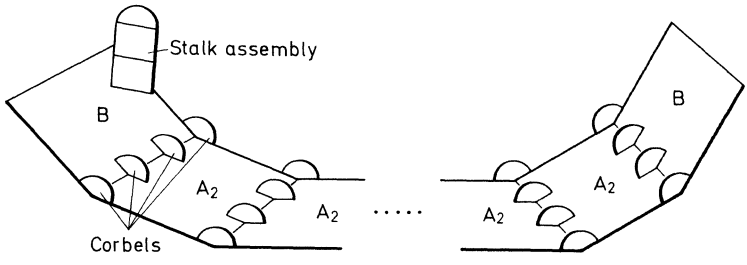


(iiα) *An Elliptic Chain*
The central fiber consists of a chain of elliptic ruled surfaces meeting along elliptic curves and meeting a rational component at one end along an elliptic curve.

(iiβ) *A Rational Chain*
A marked surface (V, D) is called of *type* A_2 if V is rational, ruled, and D is as shown:

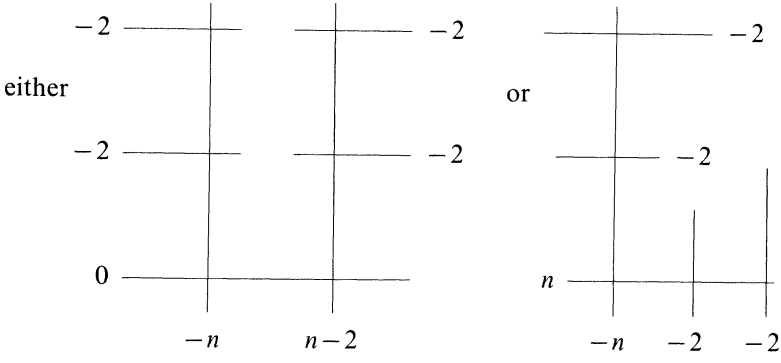


where each curve is rational and has self-intersection $=-2$. A *corbel* is a marked surface $(\mathbb{P}_2, L_1+L_2)$, where L_1 and L_2 are lines on $\mathbb{P}_2$. The central fiber consists of a chain of components of type A_2 with rational components of type B at each end, meeting corbels and stalk assemblies as shown below.



(iiiα) *Polyhedral*
A marked surface (V, D) is called *rationally polygonal* if V is rational and D is a cycle of rational curves. The central fiber consists of rationally polygonal components and its dual complex is a triangulation of $\mathbb{R}\mathbb{P}_2$.

(iiiβ) *Polyhedral With Boundary*
A marked surface (V, D) is called *quasi-polygonal* if V is rational, ruled, and D consists of rational curves whose configuration is



The central fiber consists of rationally polygonal and quasi-polygonal surfaces, together with flowers, stalks and corbels. The dual complex of the central fiber with flowers, stalks and corbels deleted is a triangulation of a disk. The vertices on the boundary correspond to the quasi-polygonal components, while those in the interior of the disk correspond to the rationally polygonal ones.

References

1. Horikawa, E.: On the periods of Enriques' surfaces. II. Math. Ann. **235**, 217–246 (1978)
2. Morrison, D.: Degenerations of Enriques' surfaces. Thesis, Harvard University, 1980
3. Shah, J.: Insignificant limit singularities of surfaces and their mixed Hodge structure. Ann. Math. **109**, 497–536 (1979)
4. Shah, J.: A complete moduli space for K3 surfaces of degree 2. Ann. Math. **112**, 485–510 (1980)
5. Shah, J.: Degenerations of K3 surfaces of degree 4. Trans. Am. Math. Soc. **263**, No. 2, 271–308 (1981)

Received June 23, 1980; in revised form February 9, 1981

Math. Ann. 256, 497–507 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Local and Global Splittings in the State Space of a JB-Algebra

P. J. Stacey

Department of Pure Mathematics, La Trobe University, Bundoora 3083, Victoria, Australia

1. Introduction and Notation

A complete normed real Jordan algebra A with 1 is said to be a JB-*algebra* if and only if, for all $a, b \in A$,

(i) $\|a \circ b\| \leqq \|a\| \, \|b\|$,
(ii) $\|a^2\| = \|a\|^2$,
(iii) $\|a^2\| \leqq \|a^2 + b^2\|$,

where $\circ$ denotes the Jordan product on A. Such an algebra becomes a norm complete order-unit space when the set of all squares is taken as the positive cone A^+ and the identity element is the order unit (see Theorem 2.1 of [3] for details). The cone A^+ gives rise to a dual ordering on A^* with positive cone A^{*+} and the subset $K = \{f \in A^{*+} : f(1) = 1\}$ of A^{*+} is said to be the *state space* of A. The state space of A is a compact convex set in the relativized weak * topology from A^* whose elements are known as *states* of A.

If x is a state on A then face (x) will be used to denote the smallest face of K containing x. The linear span of face (x) in A^* is the order-ideal in A^* generated by x and will be denoted by V_x. The set of all linear operators T on V_x which satisfy $-\lambda 1 \leqq T \leqq \lambda 1$ for some $\lambda \in \mathbb{R}$, where 1 is the identity operator on V_x, is known as the *ideal centre* $O(V_x)$ of V_x.

Properties of the ideal centre $O(V_x)$ of V_x including the theory of central measures were developed by Wils in [23]. In the paper [23] Wils indicated that in certain compact convex sets satisfying an extension axiom "Ext" the local nature of his theory could be removed and he showed that both Choquet simplices and the state spaces of C^*-algebras satisfied this extension axiom. The relation between local and global theories of central measures was further pursued by Edwards in [9] where it was shown that Wils' axiom "Ext" holds if and only if, for each $x \in K$,

$$O(V_x)x = O(A^*)x$$

where $O(A^*)$ denotes the ideal centre of A^*, defined in a similar way to $O(V_x)$. The purpose of the present paper is to show that this condition holds when K is the state space of a JB-algebra. Thus when K is the state space of a JB-algebra each

0025-5831/81/0256/0497/$02.20

order-bounded operator on V_x extends to an order-bounded operator on A^*. The proof involves a case by case consideration of certain summands in the enveloping JBW-algebra A^{**} of A (as discussed in [3, 20, and 14]). For certain of these summands arguments closely parallel to the well-known arguments for C^*-algebras can be used. However for other summands new techniques must be found. The notation is based on that introduced in [3]. In particular, if a, b, c belong to a JB-algebra A, then $\{abc\}$ denotes the Jordan triple product and U_a denotes the mapping $b \mapsto \{aba\}$.

2. The Relevant Structure Theory of JBW-Algebras

A JB-algebra which is the dual of some real Banach space is known as a JBW-*algebra*. Thus in particular the enveloping algebra A^{**} of a JB-algebra A is a JBW-algebra and it will be through the use of the structure of this JBW-algebra that the main theorem of this paper will be proved. In order to state the relevant structure theorem some notation has to be introduced.

If H is a real Hilbert space with inner product $(a, b) \mapsto s(a, b)$ then the *spin factor* $A(H, s)$ is defined to be $\mathbb{R} \oplus H$ equipped with the Jordan product

$$(\lambda, h) \circ (\mu, k) = (\lambda\mu + s(h, k), \lambda k + \mu h).$$

The norm on $A(H, s)$ is defined by

$$\|(\lambda, h)\| = |\lambda| + s(h, h)^{1/2}$$

from which it can be seen that $A(H, s)$ is reflexive and that the dual of $A(H, s)$ is $\mathbb{R} \oplus H$ equipped with the norm

$$\|(\mu, k)\|_{\text{dual}} = \max\{|\mu|, s(k, k)^{1/2}\}.$$

Simple checking shows that the spin factor $A(H, s)$ is a JBW-algebra. Another important example of a JBW-algebra is the set M_3^8 of all self-adjoint 3×3 matrices over the Cayley numbers, equipped with the usual symmetrized matrix product and matrix norm.

Theorem 3.9 of [20], Theorems 6.4 and 6.6 of [22], and Theorem 2 of [21] show that an arbitrary JBW-algebra can be described in terms of spin factors, M_3^8 and associative algebras. More precisely, the following theorem is an immediate corollary of the theorems cited above. The notation $L_\infty(\Omega, \mu, M)$ where Ω is a locally compact Hausdorff space, μ is a Radon measure on Ω and M is a normed space will be used to denote the space of equivalence classes under equality locally μ-almost everywhere of all bounded functions from Ω into M which are strongly μ-measurable (i.e. measurable in the sense of Lusin's theorem).

Theorem 2.1. *Let B be a* JBW*-algebra. Then there exists a family $\{z_\alpha : \alpha \in \Lambda\}$ of orthogonal central idempotents in B such that $\sum_\alpha z_\alpha = 1$ and such that, for each $\alpha \in \Lambda$, $z_\alpha B$ is isomorphic either to*

(i) *the self-adjoint part of a weakly closed real *-algebra of operators on a complex Hilbert space or*

(ii) *an algebra $L_\infty(\Omega_\alpha, \mu_\alpha, A_\alpha)$ where Ω_α is a locally compact Hausdorff space, μ_α is a Radon measure on Ω_α and A_α is either M_3^8 or some spin factor $A(H_\alpha, s_\alpha)$.*

3. Reversible JW-Algebras

In this section it will be shown that the normal state space of the self-adjoint part of a real weakly closed *-algebra of operators satisfies Wils' extension axiom "Ext". This will be done by methods quite closely parallel to those used in C^*-algebra theory.

Throughout this section let $\mathfrak{A}$ be a weakly closed real *-algebra of operators on a complex Hilbert space and let A be the self-adjoint part of $\mathfrak{A}$. In contrast to the complex case a state x on A does not extend uniquely to a state of $\mathfrak{A}$. Nevertheless x can be extended to a state $\tilde{x}$ on $\mathfrak{A}$ by

$$\tilde{x}(a) = x(\tfrac{1}{2}a + \tfrac{1}{2}a^*).$$

Then, by the real analogue of the GNS construction, a real Hilbert space H_x and *-representation π_x of $\mathfrak{A}$ on H_x with cyclic vector ξ_x can be obtained such that

$$\tilde{x}(a) = \langle \pi_x(a)\xi_x, \xi_x \rangle$$

for each $a \in \mathfrak{A}$. As in the complex case (described in Proposition 2.5.1 of [8]) the mapping $\theta : T \mapsto x_T$ defined by

$$(\theta T)(a) = x_T(a) = \langle \pi_x(a) T\xi_x, \xi_x \rangle$$

for all $a \in A$ is a linear order isomorphism from the self-adjoint part of $\pi_x(\mathfrak{A})'$ onto V_x, the order ideal generated by x in A^*. If S is an idempotent in $O(V_x)$ then $\theta^{-1}S\theta$ is an idempotent in $O([\pi_x(\mathfrak{A})']_{sa})$. Hence, as in the complex case described in 4.4 of [23], $(\theta^{-1}S\theta)(1)$ commutes with each self-adjoint element of $\pi_x(\mathfrak{A})'$. The following simple lemma shows that $(\theta^{-1}S\theta)(1)$ commutes with every element of $\pi_x(\mathfrak{A})'$ and hence belongs to the centre of $\pi_x(\mathfrak{A})''$.

Lemma 3.1. *Let B be a *-algebra. Then any self-adjoint idempotent S which commutes with all self-adjoint elements necessarily commutes with all elements.*

Proof. It suffices to show that S commutes with each skew-adjoint element R. So let $R = -R^*$. Then

$$\begin{aligned}(RS - SR)^* &= S^*R^* - R^*S^* \\ &= RS - SR\end{aligned}$$

and so, by hypothesis,

$$S(RS - SR) = (RS - SR)S.$$

Premultiplying by $1 - S$ yields $(1 - S)RS = 0$ and postmultiplying by $1 - S$ yields $SR(1 - S) = 0$. Hence $SR = RS$.

If x is a normal state on A then, by Theorem 2.3 of [20], x is continuous with respect to the weak * topology on A (which is equal to the relativized ultraweak operator topology). It follows that $\tilde{x}$ is continuous with respect to the weak * topology on $\mathfrak{A}$ (which is also equal to the relativized ultraweak operator topology) and hence, by the real analogue of the argument on p. 41 of [17], π_x is continuous with respect to the ultraweak operator topologies. Then, by the real analogue of Proposition 1.16.2 of [17], it follows that $\pi_x(\mathfrak{A})$ is ultraweakly closed and hence that $\pi_x(\mathfrak{A}) = \pi_x(\mathfrak{A})''$.

Theorem 3.2. *Let A be a* JBW*-algebra which is isomorphic to the self-adjoint part of some weakly closed real *-algebra of operators $\mathfrak{A}$. Then, for any normal state x on A,*

$$O(V_x)x = O(A_*)x$$

where V_x denotes the order ideal generated by x in A_*.

Proof. A_* is easily seen to be an order ideal in A^* and the order ideal generated by x in A_* is therefore equal to the order ideal generated by x in A^*. Hence if S is an idempotent in $O(V_x)$ then, by the discussion preceding Lemma 3.1, $(\theta^{-1}S\theta)(1)$ belongs to the centre of $\pi_x(\mathfrak{A})''$ and hence, by the discussion following Lemma 3.1, to the centre of $\pi_x(\mathfrak{A})$. From the real analogue of Proposition 1.10.5 of [17] it follows that there exists a central projection z in $\mathfrak{A}$ such that π_x is a *-isomorphism from $z\mathfrak{A}$ onto $\pi_x(\mathfrak{A})$. Hence there exists a self-adjoint element b in the centre of $\mathfrak{A}$ such that $\pi_x(b) = (\theta^{-1}S\theta)(1) = \theta^{-1}(S(x))$. It is easily checked that the mapping $a \mapsto ba$ is a weak *-continuous order bounded operator on A and hence is of the form R^* where R is an order bounded operator on A_*. Thus there exists $R \in O(A_*)$ such that

$$\theta(\pi_x(R^*1) = \theta(\pi_x(b)) = Sx. \tag{1}$$

However, for each $a \in A$,

$$\begin{aligned}\theta(\pi_x(R^*1))(a) &= \langle \pi_x(a)\pi_x(R^*1)\xi_x, \xi_x\rangle \\ &= \langle \pi_x(aR^*1)\xi_x, \xi_x\rangle \\ &= \tilde{x}(a \cdot R^*1) \\ &= x(ab) \\ &= x(R^*(a)) \\ &= (Rx)(a). \end{aligned} \tag{2}$$

Combining the Eqs. (1) and (2) yields $Sx = Rx$ and so every idempotent S in $O(V_x)$ can be extended to an element R of $O(A_*)$ with the same effect on x. Since $O(V_x)$ is isomorphic to the self-adjoint part of the centre of $\pi_x(\mathfrak{A})'$, which is uniformly generated by its idempotents, it follows that $O(V_x)x = O(A_*)x$.

4. JBW-Algebras of the form $L_\infty(\Omega, \mu, A)$

In the present section it will be shown that the normal state space of a JBW-algebra of the form $L_\infty(\Omega, \mu, A)$, where Ω is a locally compact Hausdorff space, μ is a Radon measure on Ω and A is either M_3^8 or a spin factor $A(H, s)$, satisfies Wils' axiom "Ext". To a large extent the two cases $A = M_3^8$ and $A = A(H, s)$ can be treated together.

By a well-known result valid for a reflexive space A and described for example in the proof of Theorem 1 of [21], $L_\infty(\Omega, \mu, A)_*$ is isomorphic to $L_1(\Omega, \mu) \otimes_\gamma A^*$ where γ denotes the greatest cross norm on $L_1(\Omega, \mu) \otimes A^*$. Hence $L_\infty(\Omega, \mu, A)_*$ can be regarded as the subspace $C(X)_* \otimes_\gamma A^*$ of $C(X)^* \otimes_\gamma A^*$ where X is the hyperstonian compact Hausdorff space such that $C(X)$ is isomorphic to $L_\infty(\Omega, \mu)$.

By Theorem 4.4 of [13], bearing in mind that A is reflexive and hence a Phillips space as defined after Theorem 3.1 of [13], $C(X)^* \otimes_\gamma A^*$ is isometric to $M(X, A^*)$, the space of all weak regular measures of bounded variation on X with values in

A^*. If m is a positive element of $M(X, A^*)$ then a finite scalar-valued measure $\tilde{m}$ is defined on X by

$$\tilde{m}(S) = \|m(S)\| = \langle 1, m(S)\rangle$$

for each Borel subset S of X. In fact $\tilde{m}$ is the scalar variation measure of m, as defined in [6], which takes a particularly simple form in the present situation since the norm of A^* is additive on the positive cone.

Lemma 4.1. *Let A be either a spin factor or M_3^8 and let m be a positive element of $C(X)_* \otimes_\gamma A^*$, where this space is regarded as a subspace of $M(X, A^*)$. Then $\tilde{m}$ is a normal measure.*

Proof. Regarding m as an element of $L_\infty(\Omega, \mu, A)_*$ its restriction to the centre Z of $L_\infty(\Omega, \mu, A)$ will, by duality, be an element of Z_*. This element can be identified with a normal measure n on X by the equation

$$n(f) = m(f \otimes 1),$$

where $f \in C(X)$ and where $f \otimes 1$ denotes the element of $C(X) \otimes A$ such that $(f \otimes 1)(x) = f(x)1$ for each $x \in X$. However, it is easily seen from the definition of $\tilde{m}$ that

$$\tilde{m}(f) = m(f \otimes 1)$$

for each $f \in C(X)$ and hence $\tilde{m}$ is normal.

Lemma 4.2. *Let A be either a spin factor or M_3^8, let m be a positive element of $C(X)_* \otimes_\gamma A^*$ and let X_m be the support of $\tilde{m}$. Then there exists a positive element g in $L_\infty(X_m, \tilde{m}, A^*)$ with the property that, for each $T \in O(V_m)$, there exists $R \in O(V_g)$ such that*

$$R(h)\tilde{m} = T(h\tilde{m})$$

for all $h \in V_g$. Further, if $T^2 = T$ then $R^2 = R$.

Proof. By Theorem 3.2 of [13] there exists a strongly $\tilde{m}$-measurable function θ from X_m to A^* such that

(i) $\|\theta(x)\| = 1$ for all $x \in X_m$ and

(ii) $m(\psi) = \int_X \langle \theta(x), \psi(x)\rangle d\tilde{m}(x)$ for each $\psi \in L_1(X_m, \tilde{m}, A)$.

Let g be the equivalence class in $L_\infty(X_m, \tilde{m}, A^*)$ containing θ. It is clear from (ii) that g is positive and that, in the obvious notation, $m = g\tilde{m}$.

Let h be a positive element in V_g, let $n = h\tilde{m}$, let T be a positive element of $O(V_m)$ and let $X_{T(n)}$ be the support of $\tilde{T}(n)$. Then as above there exists a positive element f in $L_\infty(X_{T(n)}, \tilde{T}(n), A^*)$ such that $T(n) = f\tilde{T}(n)$. However by the choice of T there exists $\lambda \in \mathbb{R}^+$ such that $0 \leqq T(n) \leqq \lambda n$, from which it follows that $0 \leqq \tilde{T}(n) \leqq \lambda\tilde{n}$ and hence that $0 \leqq \tilde{T}(n) \leqq \lambda\|h\|\tilde{m}$ where $\|h\|$ is the function $x \mapsto \|h(x)\| = \langle 1, h(x)\rangle$. By the usual argument for scalar valued measures it follows that there exists a positive $\psi \in L_\infty(X_m, \tilde{m})$ such that $\tilde{T}(n) = \psi\tilde{m}$ and hence that $T(n) = f\psi\tilde{m}$.

Since A^* is a Phillips space there exists a bounded strongly measurable function α from X_m to A^* such that, for all $\theta \in L_1(X_m, \tilde{m})$,

$$\int_{X_{T(n)}} \theta(x) f(x) d T\tilde{n}(x) = \int_{X_m} \theta(x)\alpha(x) d\tilde{m}(x)$$

and hence $f\psi$ can be regarded as an element of $L_\infty(X_m, \tilde{m}, A^*)$ which is zero on $X_m \backslash X_{T(n)}$. Hence

$$T(h\tilde{m}) = f\psi\tilde{m}$$

where $f\psi \in L_\infty(X_m, \tilde{m}, A^*)$. It is easily seen that the element $f\psi$ of $L_\infty(X_m, \tilde{m}, A^*)$ is uniquely determined by h and so R is well-defined by the equation $R(h) = f\psi$. By construction R is linear, order bounded and positive.

Since $O(V_m)$ is positively generated a non-positive element T of $O(V_m)$ also gives rise to an element R of $O(V_g)$ and it is easily seen that if $T^2 = T$ then $R^2 = R$, as required.

Lemma 4.2 enables the original problem to be transferred to a problem about ideal centres in $L_\infty(X_m, \tilde{m}, A^*)$. This in turn will be transferred to a problem about ideal centres in $L_\infty(X_m, \tilde{m}, A)$ where the Jordan structure is available to aid calculations.

Lemma 4.3. *Let $A(H, s)$ be a spin factor. Then the natural vector space identification of $A(H, s)$ with $A(H, s)^*$ is an order isomorphism and a homeomorphism for the norm topologies. Hence $L_\infty(X, \mu, A(H, s))$ is order isomorphic to $L_\infty(X, \mu, A(H, s)^*)$ for any compact Hausdorff space X and any measure μ on X.*

Proof. Recall that $(\lambda, h) \in A(H, s)$ is identified with the linear functional $\theta : (\mu, k) \mapsto \lambda\mu + s(h, k)$ in $A(H, s)^*$. A simple calculation shows that $(\mu, k) \geqq 0$ in $A(H, s)$ if and only if $\mu \geqq \|k\|$. Hence θ will be a positive linear functional if and only if $\lambda\mu + s(h, k) \geqq 0$ whenever $\mu \geqq \|k\|$.

Suppose firstly that $\lambda \geqq \|h\|$. Then, whenever $\mu \geqq \|k\|$,

$$\begin{aligned} \lambda\mu + s(h, k) &\geqq \lambda\mu - \|h\| \, \|k\| \\ &\geqq 0 \end{aligned}$$

and so θ is positive. Conversely, if θ is positive, then

$$\begin{aligned} \lambda\|h\| - \|h\|^2 &= \theta(\|h\|, -h) \\ &\geqq 0. \end{aligned}$$

Thus either $\|h\| = 0$ or $\lambda \geqq \|h\|$. If $\|h\| = 0$ then

$$\lambda = \theta(1, 0) \geqq 0 = \|h\|$$

and so, in this case as well, $\lambda \geqq \|h\|$.

Under the natural map between $A(H, s)$ and $A(H, s)^*$ it is easily seen that the dual norm $\|(\lambda, h)\| = \max\{|\lambda|, \|h\|\}$ is equivalent to the $A(H, s)$ norm $\|(\lambda, h)\| = |\lambda| + \|h\|$, as required.

Lemma 4.4. *There is a vector space isomorphism from M_3^8 onto M_3^{8*} which is also an order isomorphism and a homeomorphism for the norm topologies. Hence $L_\infty(X, \mu, M_3^8)$ is order isomorphic to $L_\infty(X, \mu, M_3^{8*})$ for any compact Hausdorff space X and any measure μ on X.*

Proof. As described on pp. 16 and 17 of [15], every Cayley number a can be written as $a = a_1 + a_2 i + a_3 j + a_4 ij + a_5 l + a_6 il + a_7 jl + a_8 ijl$ where $a_i \in \mathbb{R}$, $i^2 = -1$, $j^2 = -1$, $ij = -ji$, $l^2 = -1$, $li = -il$, $lj = jl$. Hence, if b is another Cayley number, the real part of ab is $a_1 b_1 - a_2 b_2 - a_3 b_3 - a_4 b_4 - a_5 b_5 - a_6 b_6 - a_7 b_7 - a_8 b_8$. It follows

that if $A=(a_{ij})$ and $B=(b_{ij})$ belong to M_3^8 then the trace of $B\circ A$, which is real since $B\circ A$ is Hermitian, is given by

$$\begin{aligned}\mathrm{Tr}(B\circ A)&=\frac{1}{2}\sum_{i,j}(b_{ij}a_{ji}+a_{ji}b_{ij})\\&=\frac{1}{2}\sum_{i,j}(b_{ij}a_{ij}^*+a_{ij}^*b_{ij})\\&=\frac{1}{2}\sum_{i,j}\mathrm{Re}(b_{ij}a_{ij}^*+a_{ij}^*b_{ij})\\&=\sum_{i,j,k}b_{ij}^k a_{ij}^k.\end{aligned}$$

Hence every element of M_3^{8*} is of the form $\theta_B: A\mapsto \mathrm{Tr}(B\circ A)$.

The mapping $B\mapsto\theta_B$ is easily seen to be a vector space isomorphism from M_3^8 onto M_3^{8*}, which will be an order isomorphism if and only if $\mathrm{Tr}(B\circ A^2)\geqq 0$ for each $A\in M_3^8$ exactly when $B\geqq 0$. From the remarks on p. 330 of [5] and Theorem 3.8(b) on p. 323 of [5] it can be seen that if $\mathrm{Tr}(B\circ A^2)\geqq 0$ for each $A\in M_3^8$ then $B\geqq 0$. Conversely, if $B=C^2$ then, using the associativity of the trace as described on p. 225 of [5] and Proposition 2.7 of [3],

$$\begin{aligned}\mathrm{Tr}(B\circ A^2)&=\mathrm{Tr}(C^2\circ A^2)\\&=\mathrm{Tr}(U_C(A^2))\\&\geqq 0\end{aligned}$$

for each $A\in M_3^8$ and thus $\theta_B\geqq 0$. Hence $B\mapsto\theta_B$ is an order isomorphism. Since M_3^8 is finite dimensional it is also a homeomorphism for the weak and norm topologies.

Now that the original problem can be transferred to a Jordan algebra $L_\infty(X,\mu,A)$ it is necessary to investigate ideal centres in such an algebra. For undefined terms in the next Lemma, the reader is referred to [10].

Lemma 4.5. *Let A be either a spin factor or M_3^8 and let g be a positive element of $L_\infty(X,\mu,A)$ where X is some compact Hausdorff space and μ is a measure on X. Then the weak * closure of V_g is a quadratic ideal in $L_\infty(X,\mu,A)$.*

Proof. It will firstly be shown, following the method of Theorem 2.3 of [10], that V_g is a quadratic ideal. So let $J=\{f\in L_\infty(X,\mu,A): f^2\in V_g\}$ and note that, as in the proof of Theorem 2.3 of [10], J is a quadratic ideal. It will be shown that $J=V_g$ by showing that the positive cone J^+ of J coincides with the positive cone V_g^+ of V_g. If $f\in V_g^+$ then

$$f^2=U_{f^{1/2}}f\leqq\|f\|f$$

and so $f^2\in V_g$ from which, by definition, it follows that $f\in J^+$. Conversely, if $f\in J^+$ then $f^2\in V_g^+$ and hence $f(x)^2\in V_{g(x)}^+$ for almost all $x\in X$. Since M_3^8 is finite-dimensional and since the state space of a spin factor has no proper faces it follows easily that $V_{g(x)}$ is norm closed for each $x\in X$ and it is also easily checked that $V_{g(x)}$ is a Jordan subalgebra of A. Hence, by spectral theory, $f(x)=[f(x)^2]^{1/2}$ belongs to $V_{g(x)}^+$ for almost all $x\in X$. Thus $f\in V_g^+$ as required.

To see that the weak * closure $\bar{V}_g$ of V_g is also a quadratic ideal note that $\bar{V}_g$ is a quadratic ideal if and only if $\{abc\}\in\bar{V}_g$ for all $a,c\in\bar{V}_g$ and all $b\in L_\infty(X,\mu,A)$ where $\{abc\}$ denotes the Jordan triple product. Thus, since Jordan multiplication in a

JBW-algebra is separately weak * continuous and V_g is a quadratic ideal, $\bar{V}_g$ is also a quadratic ideal.

Lemma 4.6. *Let A be either a spin factor or M_3^8, let g be a positive element of $L_\infty(X, \mu, A)$ where X is some compact Hausdorff space and let T be an element of the ideal centre $O(V_g)$ of V_g. Then there exists z_T in the Jordan centre of $\bar{V}_g$, the weak * closure of V_g, such that $Tf = z_T \circ f$ for all $f \in V_g$.*

Proof. By duality T induces an element T^{**} in $O(V_g^{**})$ and, by Theorem 9.19 of [2], $x_T = T^{**}(e)$ belongs to the Jordan centre of V_g^{**}, where e denotes the identity element of the JBW-algebra V_g^{**}. A simple calculation using Proposition 2.3 of [3] shows that $b \mapsto x_T \circ b$ is an order bounded operator on V_g^{**} which agrees with T^{**} on e and is thus equal to T^{**}. As in C^*-algebra theory the identity map on V_g extends to a unique JBW-homomorphism θ from V_g^{**} onto $\bar{V}_g$ and then, letting $z_T = \theta(x_T)$, it can be seen that z_T is in the Jordan centre of $\bar{V}_g$ and that, for each $f \in V_g$,

$$\begin{aligned} Tf &= T^{**}f \\ &= x_T \circ f \\ &= \theta(x_T \circ f) \\ &= \theta(x_T) \circ \theta(f) \\ &= z_T \circ f, \end{aligned}$$

where the equation $x_T \circ f = \theta(x_T \circ f)$ holds since $Tf \in V_g$.

Lemma 4.7. *Let z be an idempotent in the Jordan centre of $\bar{V}_g$ where g is a positive element of $L_\infty(X, \mu, A)$ for some hyperstonian compact Hausdorff space X. Then there exists an idempotent y in the Jordan centre of $L_\infty(X, \mu, A)$ such that $y \circ f = z \circ f$ for all $f \in \bar{V}_g$.*

Proof. By Lemma 4.5 and by Theorem 2.3 of [11], $\bar{V}_g$ is of the form $U_p L_\infty(X, \mu, A)$ for some unique idempotent p in $L_\infty(X, \mu, A)$. Then, by the main theorem on p. 322 of [12], each element z in the Jordan centre of $\bar{V}_g$ is of the form $U_p y$ for some y in the centre of $L_\infty(X, \mu, A)$ and, by the proof of that theorem, we can choose $y^2 = y$ whenever $z^2 = z$. It thus remains to show that, for all $b \in L_\infty(X, \mu, A)$,

$$(U_p y) \circ (U_p b) = y \circ U_p b.$$

However, since y is central and p is the identity of $U_p L_\infty(X, \mu, A)$

$$\begin{aligned} p \circ (y \circ U_p b) &= L_p L_y U_p(b) \\ &= L_y L_p U_p(b) \\ &= L_y U_p(b) \\ &= y \circ U_p(b). \end{aligned}$$

It follows that $U_p(y \circ U_p(b)) = y \circ U_p(b)$. However, since $y^2 = y$ and y is central, Lemma 2.11 of [3] yields $L_y(U_p(b)) = U_y U_p(b)$ and so $y \circ U_p(b) = U_p U_y U_p(b)$.

An application of Macdonald's Theorem (see p. 41 of [15]) shows that for elements p, y, b of any Jordan algebra,

$$U_p U_y U_{p^2}(b) = U_{U_p y} U_p(b).$$

Hence, in the present situation,

$$y \circ U_p(b) = U_{U_p y} U_p(b).$$

However, an application of Lemma 2.11 of [3] to the algebra $U_p L_\infty(X, \mu, A)$ yields

$$U_{U_p y} U_p(b) = U_p(y) \circ U_p(b),$$

and thus $y \circ U_p(b) = U_p(y) \circ U_p(b)$, as required.

The main theorem of the present section can now be proved.

Theorem 4.8. *Let B be a* JBW*-algebra which is Jordan isomorphic to $L_\infty(\Omega, \mu, A)$ where Ω is a locally compact Hausdorff space, μ is a Radon measure on Ω and A is either M_3^8 or a spin factor $A(H, s)$.*

Then for any normal state m on B,

$$O(V_m)m = O(B_*)m$$

where V_m denotes the order ideal generated by m in B_* and where $O(V_m)$ and $O(B_*)$ denote the ideal centres of V_m and B_*.

Proof. By Lemma 2.1 of [23], Theorem 13 Sect. 9 and Theorem 7 Sect. 11 of [16] and Proposition 1.3.1 of [17] it suffices to show that every idempotent in $O(V_m)$ extends to an element of $O(B_*)$. So let $T = T^2$ be an idempotent in $O(V_m)$. By Lemma 4.2 there exists $R \in O(V_g)$, with $R^2 = R$, such that

$$R(h)\tilde{m} = T(h\tilde{m})$$

for all $h \in V_g$, where g is an element of $L_\infty(X_m, \tilde{m}, A^*)$ such that $m = g\tilde{m}$ and where X_m is the support of $\tilde{m}$.

By Lemmas 4.3 and 4.4 there is an isomorphism θ from $L_\infty(X, \tilde{m}, A^*)$ onto $L_\infty(X_m, \tilde{m}, A)$ defined by $(\theta f)(x) = \phi(f(x))$ for some isomorphism ϕ from A^* onto A. Then, by Lemmas 4.6 and 4.7, there exists an idempotent y in the Jordan centre of $L_\infty(X_m, \tilde{m}, A)$ such that

$$(\theta R \theta^{-1})(f) = y \circ f$$

for all $f \in V_{\theta g} = \theta(V_g)$.

It follows from the special properties of A that y must be of the form $\chi_E \otimes 1$ where E is some $\tilde{m}$-measurable subset of X_m with characteristic function χ_E and hence, for all $x \in X_m$ and $h \in V_g$

$$\begin{aligned}[(\theta R)(h)](x) &= \chi_E(x)(\theta h)(x) \\ &= \chi_E(x)\phi(h(x)) \\ &= \phi[\chi_E(x)h(x)].\end{aligned}$$

Thus

$$\phi[(Rh)(x)] = \phi[\chi_E(x)h(x)]$$

from which it follows that

$$R(h) = \chi_E \cdot h$$

and hence that

$$T(h\tilde{m}) = \chi_E h\tilde{m}$$

where T is the idempotent in $O(V_m)$ chosen at the start of the proof. Since X_m is hyperstonian and $\tilde{m}$ is normal, E can, by Theorem 9, Chap. 3 Sect. 11 of [16], be chosen to be closed and open. Hence $n \mapsto \chi_E n$ is an idempotent in $O(B_*)$ when B_* is considered as an order ideal in $M(X, A^*)$ and so the proof is complete.

5. The Main Results

The results from Sects. 3 and 4 can now be combined with the structure theory from Sect. 2 to yield the main results of this paper.

Theorem 5.1. *Let A be a* JBW*-algebra with state space K. Then, for each $x \in K$,*

$$O(V_x)x = O(A^*)x$$

where V_x denotes the order ideal in A^* generated by x and where $O(V_x)$ and $O(A^*)$ denote the ideal centres of V_x and A^*.

Proof. By Theorem 2.1 the JBW-envelope A^{**} of A contains an orthogonal family $\{z_\alpha : \alpha \in \Lambda\}$ of central idempotents such that $z_\alpha \circ A^{**}$ is either the self-adjoint part of a weakly closed real *-algebra of operators or a Jordan algebra $L_\infty(\Omega_\alpha, \mu_\alpha, A_\alpha)$. However each central idempotent z in A^{**} is of the form P^*1 where P is the idempotent in $O(A^*)$ defined by

$$(Pk)(a) = k(z \circ a)$$

for each $k \in A^{***} \cong A^*$ and $a \in A^{**}$. Thus, for each $\alpha \in \Lambda$, $z_\alpha = P_\alpha^* 1$ for some $P_\alpha \in O(A^*)$.

For each $\alpha \in \Lambda$, each $x \in K$ and each $T \in O(V_x)$ the result of either Theorem 3.2 or Theorem 4.8 can be applied to the element $P_\alpha x$ of the predual $P_\alpha A^*$ of the JBW-algebra $z_\alpha \circ A^{**}$ to yield an element $S_\alpha \in O(P_\alpha A^*)$ such that $TP_\alpha x = S_\alpha P_\alpha x$. It is then easily checked that $S = \sum_{\alpha \in \Lambda} S_\alpha P_\alpha$ is an element of $O(A^*)$ such that $Sx = Tx$, as required.

Corollary 5.2. *Let B be a* JBW*-algebra with normal state space K. Then, for each $x \in K$, $O(V_x)x = O(B_*)x$.*

Proof. It is easily seen by extending the identity map on B to a JBW-homomorphism from B^{**} onto B and then using duality that B_* is of the form PB^* for some idempotent $P \in O(B^*)$. The result thus follows immediately from Theorem 5.1 since V_x is the order ideal generated by x in both B_* and B^*.

Corollary 5.3. *Let A be a* JBW*-algebra with state space K and let $x \in K$. Then the following three statements are equivalent*

(i) $O(V_x)x = \{\lambda x : \lambda \in \mathbb{R}\}$ *(i.e. x is primary)*

(ii) $O(A^*)x = \{\lambda x : \lambda \in \mathbb{R}\}$ *(i.e. x is globally primary)*

(iii) $c(x) \circ A^{**}$ *is a* JBW*-factor, where $c(x)$ is the smallest central idempotent p in A^{**} for which $p(x) = 1$.*

Proof. It follows immediately from Theorem 5.1 that (i) and (ii) are equivalent. It is well-known and easily seen, as in the proof of Theorem 5.1, that the mapping $P \mapsto P^*1$ is an isomorphism from $O(A^*)$ onto the Jordan centre of A^{**} and that $P(x)=x$ if and only if $(P^*1)(x)=1$. Hence the Jordan centre of $c(x)\cdot A^{**}$ is isomorphic to the ideal centre of $P_x A^*$, where P_x is the smallest idempotent in $O(A^*)$ for which $P_x(x)=x$. It follows from Corollary 4 of [9] that (ii) is equivalent to (iii).

Acknowledgements. The author is grateful to the members of the Mathematical Institute, Oxford for hospitality while this work was being done. He is particularly grateful to C. M. Edwards for raising the problem solved in Corollary 3.5, for his help with the solution of that problem along the lines of the theory in Sect. 3 of this paper and for much other helpful advice.

References

1. Alfsen, E.M.: Compact convex sets and boundary integrals. Berlin, Heidelberg, New York: Springer 1971
2. Alfsen, E.M., Shultz, F.W.: Non-commutative spectral theory for affine function spaces on convex sets. Mem. Am. Math. Soc. **172**, (1976)
3. Alfsen, E.M., Shultz, F.W., Størmer, E.: A Gelfand-Neumark theorem for Jordan algebras. Adv. Math. **28**, 11–57 (1978)
4. Bourbaki, N.: Espaces vectoriels topologiques. Hermann: Paris 1955
5. Braun, H., Koecher, M.: Jordan-algebren. Berlin, Heidelberg, New York: Springer 1966
6. Dinculeanu, N.: Mesures vectorielles et opérations linéaires. C. R. Acad. Sci. Paris **246**, 2328–2331 (1958)
7. Dixmier, J.: Les algèbres d'operateurs dans l'espace hilbertien. Paris: Gauthier-Villars 1969
8. Dixmier, J.: Les C^*-algèbres et leurs représentations. Paris: Gauthier-Villars 1964
9. Edwards, C.M.: On global central decomposition for a universal cap. Math. Ann. **213**, 223–230 (1975)
10. Edwards, C.M.: On the facial structure of a JB-algebra. J. London Math. Soc. **19**, 335–344 (1979)
11. Edwards, C.M.: Ideal theory in JB-algebras. J. London Math. Soc. **16**, 507–513 (1977)
12. Edwards, C.M.: On the centres of hereditary JBW-subalgebras of a JBW-algebra. Math. Proc. Cambridge Phil. Soc. **85**, 317–324 (1979)
13. Gil de Lamadrid, J.: Measures and tensors. Trans. Am. Math. Soc. **114**, 98–121 (1965)
14. Hanche-Olsen, H.: A note on the bidual of a JB-algebra. Math. Z. **175**, 29–31 (1980)
15. Jacobson, N.: Structure and representations of Jordan algebras. Am. Math. Soc. Colloq. Publ. **39**. Providence: American Mathematical Society 1968
16. Lacey, H.E.: The isometric theory of classical Banach spaces. Berlin, Heidelberg, New York: Springer 1974
17. Sakai, S.: C^*-algebras and W^*-algebras. Berlin, Heidelberg, New York: Springer 1971
18. Schaeffer, H.H.: Topological vector spaces. Berlin, Heidelberg, New York: Springer 1971
19. Schatten, R.: A theory of cross spaces. Ann. Math. Stud. **26**, (1950)
20. Shultz, F.W.: On normed Jordan algebras which are Banach dual spaces. J. Functional Analysis **31**, 360–376 (1979)
21. Stacey, P.J.: Type I_2 JBW-algebras. Quart. J. Math. Oxford (2), (to appear)
22. Størmer, E.: Jordan algebras of type I. Acta Math. **115**, 165–184 (1966)
23. Wils, W.: The ideal center of partially ordered vector spaces. Acta Math. **127**, 41–77 (1971)

Received August 28, 1980; in revised form February 9, 1981

Math. Ann. 256, 509–516 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Stabilität starkstetiger, positiver Operatorhalbgruppen auf C^*-Algebren

Ulrich Groh und Frank Neubrander

Mathematische Fakultät der Universität, Auf der Morgenstelle 10, D-7400 Tübingen, Bundesrepublik Deutschland

0. Einleitung

Ausgangspunkt der vorliegenden Arbeit ist folgendes, auf A.M. Lyapunov zurückgehende, klassische Resultat aus der Stabilitätstheorie gewöhnlicher Differentialgleichungen [1].

Es sei A eine komplexe $n \times n$-Matrix. Dann sind äquivalent:

1. *Für alle $x_0 \in \mathbb{C}^n$ gilt für die Lösung $T(t)x_0 := e^{At}x_0$ des abstrakten Cauchyproblems $\dot{x} = Ax$, $x(0) = x_0$:*

$$\|T(t)x_0\| \to 0 \quad \text{für } t \to \infty.$$

2. *Für die Spektralschranke $s(A) = \max\{\mathrm{Re}(\lambda): \lambda \in \sigma(A)\}$ gilt $s(A) < 0$.*
3. *Es existiert eine positiv-definite $n \times n$-Matrix Y mit $A^*Y + YA = -I$.*

Auf einem unendlichdimensionalen Banachraum E führt das abstrakte Cauchyproblem zu dem folgenden Ergebnis [7]:

Es sei $(A, D(A))$ ein dicht definierter, linearer Operator auf einem Banachraum E mit nichtleerer Resolventenmenge $\rho(A)$. Dann hat das abstrakte Cauchyproblem $\dot{x} = Ax$, $x(0) = y$ für alle $y \in D(A)$ genau dann eine stetig differenzierbare Lösung u, wenn $(A, D(A))$ der Generator einer starkstetigen Halbgruppe $\{T(t)\}$ ist. Für die Lösung u gilt dann $u(t) = T(t)y$.

Wie in 1. ist man im besonderen an Stabilitätsaussagen für diese Halbgruppe interessiert. Beispiele zeigen aber, daß im allgemeinen aus $s(A) < 0$ nicht auf die Stabilität geschlossen werden kann [8, Sect. 23. 16], [5].

Wir befassen uns in dieser Arbeit mit starkstetigen Halbgruppen positiver Operatoren auf C^*-Algebren und untersuchen, welche Aussagen über die Stabilität solcher Halbgruppen gemacht werden können. Dabei knüpfen wir an Resultate an, die für positive Operatorhalbgruppen auf kommutativen C^*-Algebren in [4] und [5] bewiesen wurden. Einfache Beispiele zeigen, daß bei unseren Aussagen im allgemeinen nicht auf die Positivität verzichtet werden kann.

0025-5831/81/0256/0509/$01.80

1. Vorbemerkungen

Unter einer *starkstetigen Operatorhalbgruppe* $\{T(t)\}_{t\geq 0}$ auf einem Banachraum E verstehen wir eine Familie $\{T(t)\}_{t\geq 0}\subseteq\mathscr{L}(E)$ mit $T(0)=I$, $T(s+t)=T(s)\,T(t)$ $(s,t\geq 0)$ und der Stetigkeit der Abbildung $t\mapsto T(t)$ für die *starke Operatortopologie* auf $\mathscr{L}(E)$. Zur Abkürzung werden wir unter diesen Voraussetzungen von der „Halbgruppe $\{T(t)\}$" sprechen. Für eine Einführung in die wesentlichen Begriffe dieser Theorie, wie z.B. Definition und Eigenschaften des *Generators* $(A, D(A))$, verweisen wir auf [13, Abschn. 3], [4] und die dort angegebene Literatur.

Unter einer *C*-Algebra* $\mathfrak{A}$ verstehen wir eine komplexe Banachalgebra mit Involution $*$ und $\|x^*x\|=\|x\|^2$ für alle $x\in\mathfrak{A}$. Besitzt $\mathfrak{A}$ eine Einheit, so bezeichnen wir diese mit $\mathbb{1}$. Eine C^*-Algebra $\mathfrak{A}$ nennen wir eine *W*-Algebra*, falls es einen Banachraum $\mathfrak{A}_*$ gibt, dessen Dual $(\mathfrak{A}_*)^*$ als Banachraum isomorph zu $\mathfrak{A}$ ist. Es ist $\mathfrak{A}_*$ eindeutig bestimmt [14, III.3.9] und $\mathfrak{A}$ besitzt stets eine Einheit [14, I.10.2]. Wir nennen $\mathfrak{A}_*$ den *Prädual* von $\mathfrak{A}$. Mit $\mathfrak{A}_+:=\{x^*x: x\in\mathfrak{A}\}$ bezeichnen wir den *positiven Kegel* einer C^*-Algebra. Dieser ist erzeugend und normal [12, p. 215] und der duale Kegel $\mathfrak{A}^*{}_+:=\{\varphi\in\mathfrak{A}^*: \varphi(x)\geq 0$ für alle $x\in\mathfrak{A}_+\}$ ist schwach $*$-abgeschlossen und erzeugend. Entsprechend ist der positive Kegel $\mathfrak{A}_*{}^+:=\mathfrak{A}_*\cap\mathfrak{A}^*_+$ im Prädual einer W^*-Algebra erzeugend [14, III.4.2]. Unter einer *von Neumann-Algebra* $\mathfrak{A}$ auf einem Hilbertraum $\mathfrak{H}$ verstehen wir eine selbstadjungierte, in der schwachen Operatortopologie abgeschlossene Unteralgebra von $\mathscr{L}(\mathfrak{H})$ mit $\mathbb{1}_{\mathfrak{H}}\in\mathfrak{A}$. Im besonderen ist jede von Neumann-Algebra eine (abstrakte) W^*-Algebra.

2. Das Problem $s(A)=\omega_0$

Ist $\{T(t)\}$ eine starkstetige Operatorhalbgruppe, so sind die Spektren $\sigma(T(t))$ durch den *Spektralradius* $r(T(t))=e^{\omega_0 t}$ mit festem $\omega_0\in[-\infty,\infty)$ beschränkt [8, p. 457]; [4, 1.1(iv)], während das Spektrum $\sigma(A)$ des Generators $(A, D(A))$ nur rechtsseitig beschränkt ist. Wir nennen $s(A):=\sup\{\mathrm{Re}(\lambda): \lambda\in\sigma(A)\}$ die *Spektralschranke* von A und es gilt $s(A)\leq\omega_0$ [4].

Im folgenden wollen wir zeigen, daß für positive Halbgruppen auf C^*-Algebren und den Prädualen von W^*-Algebren stets $s(A)=\omega_0$ gilt. Dabei nennen wir eine Halbgruppe $\{T(t)\}$ *positiv*, falls $T(t)\,\mathfrak{A}_+\subseteq\mathfrak{A}_+$ für alle $0\leq t\in\mathbb{R}$. Ein einfaches Beispiel zeigt, daß für die gewünschte Schlußfolgerung im allgemeinen auf die Positivität nicht verzichtet werden kann.

2.1. **Satz.** *Sei $\mathfrak{A}$ eine C*-Algebra mit Einheit und sei $\{T(t)\}$ eine starkstetige, positive Operatorhalbgruppe auf $\mathfrak{A}$. Dann gilt $-\infty<s(A)=\omega_0$ und $s(A)\in\sigma(A)$.*

Beweis. Da alle Operatoren $T(t)$ $(t\geq 0)$ positiv sind und unsere Algebra eine Einheit besitzt, existiert zu jedem $t\geq 0$ ein Zustand $\varphi_t\in\mathfrak{A}^*$ mit

$$T(t)^*\varphi_t=r(T(t))\,\varphi_t=e^{\omega_0 t}\varphi_t$$

[6, 2.1]. Für $n\in\mathbb{N}$ sei

$$E_n:=\{\varphi\in\mathfrak{A}^*_+:\ \varphi(\mathbb{1})=1,\ T(2^{-n})^*\varphi=e^{\omega_0 2^{-n}}\varphi\}.$$

Die Folge der E_n, $n\in\mathbb{N}$, ist monoton fallend und es ist wegen der schwach * Kompaktheit der Einheitskugel von $\mathfrak{A}^*$ $\bigcap_{n\in\mathbb{N}} E_n \neq \emptyset$; sei etwa $\varphi\in\bigcap_{n\in\mathbb{N}} E_n$. Aus der Halbgruppeneigenschaft $T(s+t)=T(s)T(t)$ $(s,t\in\mathbb{R}_+)$ und der schwach * Stetigkeit der Abbildung $(t\mapsto T(t))$: $\mathbb{R}_+\to\mathfrak{A}^*$ folgt dann $T(t)^*\varphi=e^{\omega_0 t}$ für alle $t\geqq 0$. Angenommen $\omega_0=-\infty$. Dann gilt $T(t)^*\varphi=0$, also $0=\varphi(T(t)\mathbb{1})$ für alle $t>0$. Aus Stetigkeitsgründen bedeutet dies aber $1=\varphi(\mathbb{1})=0$, ein Widerspruch, der sich nur durch $\omega_0>-\infty$ auflöst. Da $e^{\omega_0 t}$ ein Element des Punktspektrums $P_\sigma(T(t)^*)$ eines jeden Operators $T(t)^*$ $(t\geqq 0)$ ist, folgt aus [4, 1.6 u. 1.7] $\omega_0\in\sigma(A)$ oder $s(A)=\omega_0$. □

2.2. *Bemerkung.* Wie das Beispiel der Translationshalbgruppe auf der C^*-Algebra $C_0(\mathbb{R}_+)$ der im „Unendlichen" verschwindenden stetigen Funktionen auf $\mathbb{R}_+$ zeigt (vgl. unten 3.3), existiert im allgemeinen keine stetige Linearform, welche invariant unter allen $T(t)$ $(t\geqq 0)$ ist, sobald man auf die Existenz einer Einheit verzichtet. □

2.3. **Satz.** *Sei $\mathfrak{A}$ eine W^*-Algebra mit Prädual $\mathfrak{A}_*$ und sei $\{T(t)\}$ eine starkstetige, positive Operatorhalbgruppe auf $\mathfrak{A}_*$. Dann gilt $s(A)=\omega_0$.*

Beweis. Nach [5, Theorem 3] gilt für alle $\lambda>s(A)$ und $\varphi\in\mathfrak{A}_*$

$$R(\lambda,A)=\int_0^\infty e^{-\lambda t}T(s)\varphi\,ds.$$

Für $\varphi\in\mathfrak{A}_*^+$ gilt aber $\|\varphi\|=\varphi(\mathbb{1})$, also folgt hieraus die Existenz des Integrals

$$\int_0^\infty e^{-\lambda t}\|T(s)\|\,ds$$

für alle $\varphi\in\mathfrak{A}_*$ und alle $\lambda>s(A)$. Im besonderen existiert für jede positive Halbgruppe $\{T(t)\}$ mit $s(A)<0$

$$\int_0^\infty \|T(s)\|\,ds,$$

d.h. für eine solche Halbgruppe muß für $t>0$ stets $r(T(t))<1$ sein [10, p. 121].

Angenommen es gibt eine positive Halbgruppe $\{T(t)\}$ mit $s(A)<\omega_0$. Wir betrachten dann die Halbgruppe $S(t):=e^{-\omega_0 t}T(t)$ mit Generator $(B,D(B))$. Für diese Halbgruppe gilt $s(B)=s(A)-\omega_0<0$ aber $r(S(t))=1$ $(t>0)$. Dies ist aber ein Widerspruch zu unseren vorherigen Überlegungen. Daher gilt stets $s(A)=\omega_0$. □

2.4. **Korollar.** *Sei $\mathfrak{A}$ eine C^*-Algebra und sei $\{T(t)\}$ eine starkstetige, positive Operatorhalbgruppe auf dem Dual $\mathfrak{A}^*$ von $\mathfrak{A}$. Dann gilt $s(A)=\omega_0$.*

Beweis. Nach [14, III.2.4] ist der Bidual von $\mathfrak{A}$ eine W^*-Algebra. Es folgt daher die Behauptung sofort aus Satz 2.3. □

Eine Modifikation des Beispiels in [15] zeigt, daß es starkstetige Operatorhalbgruppen auf C^*-Algebren mit $s(A)<\omega_0$ gibt.

2.5. *Beispiel.* Wir betrachten die C^*-Algebra c_0 und zeigen, daß es eine Halbgruppe $\{T(t)\}$ auf c_0 gibt mit $\|T(t)\| = e^t$ $(t \geqq 0)$ und $\sigma(A) = \{in\colon n \in \mathbb{N}\}$. Dies bedeutet $0 = s(A) < \omega_0 = 1$.

Sei $l^\infty(n)$ der Vektorraum $\mathbb{C}^n$ versehen mit der Supremumsnorm. Es ist dann die C^*-Algebra $\mathfrak{A} := \{(x_n)\colon x_n \in l^\infty(n), \lim_n \|x_n\| = 0\}$ versehen mit der Supremumsnorm isomorph zu c_0. Sei für $n \in \mathbb{N}$ $A_n = (a_{i,j}^{(n)})$ die wie folgt definierte $n \times n$-Matrix:

$$a_{i,j}^{(n)} = \begin{cases} 1 & j = i+1 \\ 0 & \text{sonst} \end{cases} \qquad (i,j = 1, 2, \ldots, n).$$

Es gilt dann $\sigma(A_n) = \{0\}$ $(n \in \mathbb{N})$ und für $B_n := in + A_n$ $\sigma(B_n) = \{in\}$ $(n \in \mathbb{N})$. Es ist $R(\lambda, B_n)$ gegeben durch die Matrix

$$(\lambda - in)^{-n} \cdot \begin{pmatrix} (\lambda - in)^{n-1} & (\lambda - in)^{n-2} \ldots & 1 \\ 0 \ddots & \ddots & (\lambda - in)^{n-1} \ldots (\lambda - in) \\ \vdots \; 0 & \ddots & \vdots \\ 0 \ldots & \ddots & (\lambda - in)^{n-1} \end{pmatrix}.$$

Demnach gilt $\|R(\lambda, B_n)\| \leqq 1$ $(n \in \mathbb{N})$. Für $x = (x_n) \in \mathfrak{A}$ und $0 \leqq t \in \mathbb{R}$ sei $T(t)x := (e^{in} e^{tA_n} x_n)$. Da $\|e^{tA_n}\| = \|e^{tA_n} \mathbb{1}_n\|$ ist, gilt $\|T(t)\| = \sup_n \{e^{tA_n}\} = e^t$ $(t \geqq 0)$. Eine leichte Rechnung zeigt nun, daß für $x \in \bigoplus_{n=1}^{\infty} l^\infty(n)$: $\lim \|T(t)x - x\| = 0$ erfüllt ist. Da $\bigoplus_{n=1}^{\infty} l^\infty(n)$ dicht in $\mathfrak{A}$ ist, folgt hieraus die starke Stetigkeit der Halbgruppe $\{T(t)\}$ auf $\mathfrak{A}$. Der Generator B der Halbgruppe $\{T(t)\}$ ist durch $\bigoplus_{n=1}^{\infty} B_n$ gegeben mit $D(B) = \{(x_n) \in \mathfrak{A}\colon (B_n x_n) \in \mathfrak{A}\}$. Wegen $\{in\colon n \in \mathbb{N}\} \subseteq \sigma(B)$ gilt $s(B) \geqq 0$. Für alle $\lambda \notin \{in\colon n \in \mathbb{N}\}$ existiert $R(\lambda, B_n)$ und es ist wegen $\|R(\lambda, B_n)\| \leqq 1$ $R := \bigoplus_{n=1}^{\infty} R(\lambda, B_n)$ ein stetiger Operator auf $\mathfrak{A}$ mit $(\lambda - B) \circ R = I$ auf $\mathfrak{A}$ und $R \circ (\lambda - B) = I$ auf $D(B)$. Daher ist λ ein Element der Resolventenmenge von B und $R(\lambda, B) = R$ oder $0 = s(B) < \omega_0 = 1$. □

3. Stabilität positiver Operatorhalbgruppen auf C^*-Algebren

Eine starkstetige Operatorhalbgruppe $\{T(t)\}$ heiße *stabil*, falls es einen Operator $P \in \mathscr{L}(\mathfrak{A})$ gibt für den in einer der drei Standardoperatortopologien $\lim_{t \to \infty} T(t) = P$ gilt. Es ist dann P eine Projektion auf den Fixraum $F := \{x \in \mathfrak{A}\colon T(t)x = x$ für alle $t \geqq 0\}$ und man kann somit die Zerlegung von $\mathfrak{A}$ in F und $\ker P$ voraussetzen. Da die hier interessierende Fragestellung auf dem abgeschlossenen, $T(t)$-invarianten Teilraum F trivial ist, soll nur untersucht werden, wann $\lim T(t) = 0$ gilt (hier und im weiteren falls nichts anderes erwähnt immer für $t \to \infty$). Es ist daher sinnvoll folgende Sprachregelung zu treffen.

3.1. *Definition.* Sei E ein Banachraum. Eine starkstetige Halbgruppe $\{T(t)\}$ auf E heiße

1. *Gleichmäßig stabil*, falls $\lim \|T(t)\| = 0$ gilt.
2. *Stark stabil*, falls in der starken Operatortopologie $\lim T(t) = 0$ gilt.
3. *Schwach stabil*, falls in der schwachen Operatortopologie $\lim T(t) = 0$ gilt.
4. *Exponentiell stabil*, falls es ein $t_0 \in \mathbb{R}_+$ und ein $\omega \in \mathbb{R}_+$ gibt, so daß für alle $t \geqq t_0$ $\|T(t)\| \leqq e^{-\omega t}$ gilt.

In [2, p. 111] wird gezeigt, daß für einen positiven Operator T auf einer kommutativen C^*-Algebra mit Einheit die Konvergenz der Potenzen $\{T^n\}_{n \in \mathbb{N}}$ in der schwachen Operatortopologie äquivalent zur Normkonvergenz ist. Wir wollen nun diesen „diskreten" Sachverhalt für positive Operatorenhalbgruppen auf nichtkommutative C^*-Algebren mit Einheit übertragen. Ein einfaches Beispiel zeigt, daß auf die Existenz einer Einheit nicht verzichtet werden kann.

3.2. **Satz.** *Sei $\mathfrak{A}$ eine C^*-Algebra mit Einheit und sei $\{T(t)\}$ eine starkstetige, positive Operatorhalbgruppe auf $\mathfrak{A}$. Dann sind äquivalent:*

1. $s(A) < 0$.
2. *Die Halbgruppe $T(t)$ ist gleichmäßig stabil.*
3. *Die Halbgruppe $T(t)$ ist stark stabil.*
4. *Die Halbgruppe $T(t)$ ist schwach stabil.*

Beweis. Wegen Satz 2.1 und $\|T(t)\| \leqq M e^{\omega_0 t}$ genügt es, die Implikation von 4. nach 1. zu zeigen. Zu $0 < t_0 \in \mathbb{R}$ existiert aber ein $0 \neq \varphi \in \mathfrak{A}^*$ mit

$$T(t_0)^* \varphi = r(T(t_0))\, \varphi .$$

Also gilt für alle $0 \neq x \in \mathfrak{A}$

$$\varphi(T(t_0)^n x) = r(T(t_0))^n \varphi(x) \to 0$$

für $n \to \infty$. Daraus folgt $r(T(t_0)) < 1$ und somit wegen $r(T(t_0)) = e^{\omega_0 t}$: $s(A) = \omega_0 < 0$. □

3.3. *Bemerkung.* Ist $\mathfrak{A}$ eine C^*-Algebra ohne Einheit, so gelten die obigen Äquivalenzen im allgemeinen nicht mehr. Als Beispiel dafür betrachten wir die C^*-Algebra $C_0(\mathbb{R}_+)$ und die Halbgruppe $\{T(t)\}$ mit $(T(t)f)(s) = f(s+t)$ ($f \in C_0(\mathbb{R}_+)$, $0 \leqq s, t \in \mathbb{R}$). Dann gilt $\|T(t)\| = 1$ ($t \geqq 0$) aber $\lim_{t \to \infty} \|T(t)f\| = 0$ für alle $f \in C_0(\mathbb{R}_+)$. Entsprechendes gilt für die obige Halbgruppe auch auf $L^1(\mathbb{R}_+)$. Es übertragen sich also die Äquivalenzen von 3.2 nicht auf Präduale von W^*-Algebren. □

4. Stabilität von implementierten Halbgruppen auf von Neumann-Algebren

Ist $\mathfrak{A}$ eine W^*-Algebra und $\{T(t)\}$ eine Operatorhalbgruppe auf $\mathfrak{A}$, so nennen wir diese *schwach*-stetig*, falls die Abbildung $t \mapsto T(t)$ stetig ist, wenn $\mathscr{L}(\mathfrak{A})$ mit der Topologie der punktweisen Konvergenz auf $(\mathfrak{A}, \sigma(\mathfrak{A}, \mathfrak{A}_*))$ versehen ist. Eine Charakterisierung solcher Halbgruppen ist in Analogie zum Satz von Hille-

Yosida möglich [2, 3.1.10]. Gegeben sei nun eine von Neumann-Algebra $\mathfrak{A}$ auf einem Hilbertraum $\mathfrak{H}$ und sei $\{U(t)\}$ eine starkstetige Operatorhalbgruppe auf $\mathfrak{H}$ mit Generator $(B, D(B))$. Für $x \in \mathscr{L}(\mathfrak{H})$ sei $T(t)x := U(t)^* \circ x \circ U(t)$ $(t \geqq 0)$. Falls für $t \geqq 0$ $T(t)\mathfrak{A} \subseteq \mathfrak{A}$ gilt, so nennen wir die auf $\mathfrak{A}$ durch $\{T(t)\}$ induzierte Halbgruppe *implementiert durch* $\{U(t)\}$.

4.1. *Bemerkungen*

1. Unter Benutzung von [14, II.2.6] zeigt eine einfache Rechnung, daß eine implementierte Halbgruppe stets schwach*-stetig ist.
2. Jeder Operator $T(t)$ $(t \geqq 0)$ einer implementierten Halbgruppe ist vollständig positiv und stetig für die schwach*-Topologie auf $\mathfrak{A}$.
3. Ist $\{T(t)\}$ eine implementierte Halbgruppe, so ist die präadjungierte Halbgruppe $\{T(t)_*\}$ schwachstetig auf $\mathfrak{A}_*$ und muß daher nach [2, 3.1.8] sogar starkstetig sein. Aus Satz 2.3 folgt dann, daß für implementierte Halbgruppen stets $s(A) = \omega_0$ gelten muß, da $s(A) = s(A_*)$ und dies auch für die entsprechenden ω_0 gilt.
4. Mit $\{U(t)\}$ ist auch die adjungierte Halbgruppe $\{U(t)^*\}$ auf $\mathfrak{H}$ starkstetig mit Generator $(B^*, D(B^*))$. Für die von $\{U(t)\}$ auf $\mathfrak{A}$ implementierte Halbgruppe $\{T(t)\}$ mit Generator $(A, D(A))$ zeigt eine zu [2, 3.2.55] analoge Überlegung, daß für $x \in \mathfrak{A}$ das folgende äquivalent ist:
   (i) $x \in D(A)$
   (ii) *Für* $\xi \in D(B)$ *gilt* $x(\xi) \in D(B^*)$ *und die lineare Abbildung*

$$(*) \qquad (\xi \mapsto x(B\xi) + B^*(x\xi)) \colon D(B) \to H$$

besitzt eine stetige Fortsetzung auf $\mathfrak{H}$.

Für $x \in D(A)$ ist dann Ax durch die stetige Fortsetzung von $(*)$ gegeben. Wir schreiben daher kurz $Ax = xB + B^*x$. □

Wir wollen im folgenden einige äquivalente Aussagen zur exponentiellen Stabilität einer implementierten Halbgruppe angeben. Es zeigt sich, daß die Lösbarkeit der Operatorgleichung

$$yB + B^*y = -x \qquad (x, y \in \mathfrak{A}_+)$$

eine notwendige und hinreichende Bedingung dafür ist, und daher dies ein unendlichdimensionales Analogon zum klassischen Resultat von Lyapunov darstellt (s. Einleitung).

4.2. **Satz.** *Sei* $\mathfrak{A}$ *eine von Neumann-Algebra auf einem Hilbertraum* $\mathfrak{H}$ *und sei* $\{T(t)\}$ *eine schwach*-stetige Operatorhalbgruppe auf* $\mathfrak{A}$ *mit Generator* $(A, D(A))$, *die von einer auf* $\mathfrak{H}$ *starkstetigen Operatorhalbgruppe* $\{U(t)\}$ *mit Generator* $(B, D(B))$ *implementiert wird. Dann sind die folgenden Aussagen äquivalent:*

1. $s(A) < 0$.
2. *Die Halbgruppe* $\{U(t)\}$ *ist exponentiell stabil.*
3. *Es existiert ein* $0 \leqq x \in D(A)$ *mit* $Ax = -\mathbb{1}$.
4. *Es existiert ein* $0 \leqq x \in D(A)$ *mit* $x(D(B)) \subseteq D(B^*)$ *und* $xB + B^*x = -\mathbb{1}$.
5. *Für alle* $0 \leqq x \in \mathfrak{A}$ *existiert ein* $0 \leqq y \in D(A)$ *mit* $Ay = -x$.

6. *Für alle* $0 \leqq x \in \mathfrak{A}$ *existiert ein* $0 \leqq y \in D(A)$ *mit* $y(D(B)) \subseteq D(B^*)$ *und* $yB + B^* y = -x$.

7. $\int_0^\infty \|U(s)\xi\|^2 ds$ *existiert für alle* $\xi \in \mathfrak{H}$.

8. $\int_0^\infty (T(s)x)\xi | \eta) ds$ *existiert für alle* $\xi, \eta \in \mathfrak{H}$ *und alle* $x \in \mathfrak{A}$.

Beweis. Da $s(A) = \omega_0$ gilt, zeigt eine leichte Rechnung die Äquivalenz von 1. und 2., während die Äquivalenz der Behauptungen 3. und 4. beziehungsweise 5. und 6. sich aus Bemerkung 4.1.4. ergibt.

1.$\Rightarrow$3.: Wegen $\omega_0 = s(A) < 0$ existiert $R(0, A)$ und ist auf Grund der Integraldarstellung [4, 1.1 (iii)] positiv. Daher gilt $R(0, A)\mathbb{1} = x \in D(A)_+$ oder $Ax = -\mathbb{1}$ mit einem $x \in D(A)_+$

3.$\Rightarrow$5.: Sei $0 \leqq x \in D(A)$ mit $Ax = -\mathbb{1}$. Dann gilt

$$T(t)x - x = \int_0^t T(s) A x \, ds = -\int_0^s T(s)\mathbb{1} \, ds \qquad (t \geqq 0)$$

und damit für alle $0 \leqq t \in \mathbb{R}$: $0 \leqq \int_0^t T(s)\mathbb{1} \, ds \leqq x$. Da $\left(\int_0^t T(s)\mathbb{1} \, ds\right)_{t \geqq 0}$ eine monoton wachsende, beschränkte Familie positiver Elemente in $\mathfrak{A}$ ist, existiert $\lim_{t \to \infty} \int_0^t T(s)\mathbb{1} \, ds$ in der schwachen Operatortopologie von $\mathscr{L}(\mathfrak{H})$. Diese ist aber auf den beschränkten Mengen von $\mathfrak{A}$ äquivalent zur $\sigma(\mathfrak{A}, \mathfrak{A}_*)$-Topologie, also existiert in $\mathfrak{A}$ für alle $\varphi \in \mathfrak{A}_*$ $\int_0^\infty \varphi(T(s)\mathbb{1}) ds$. Ist $0 \leqq x \in \mathfrak{A}$ so gilt $0 \leqq x \leqq \|x\| \cdot \mathbb{1}$, d.h. für $0 \leqq \varphi \in \mathfrak{A}_*$ und $0 \leqq s \in \mathbb{R}$ ist $\varphi(T(s)x) \leqq \|x\| \varphi(T(s)\mathbb{1})$ woraus die Existenz von $\int_0^\infty \varphi(T(s)x) ds$ für alle $0 \leqq x \in \mathfrak{A}$, $0 \leqq \varphi \in \mathfrak{A}_*$ folgt. Da die positiven Kegel von $\mathfrak{A}_*$ und $\mathfrak{A}$ erzeugend sind, existiert $\int_0^\infty \varphi(T(s)x) ds$ für alle $x \in \mathfrak{A}$ und $\varphi \in \mathfrak{A}_*$, und es gilt $\varphi(R(0, A)x) = \int_0^\infty \varphi(T(s)x) ds$. Wie man sieht ist $0 \leqq R(0, A)$.

3.$\Rightarrow$7.: Wie der obige Beweis gezeigt hat, existiert $\int_0^\infty T(s)\mathbb{1} \, ds$ in der schwachen Operatortopologie, d.h. $\int_0^\infty \|U(s)\xi\|^2 ds = \int_0^\infty (T(s)\mathbb{1}\,\xi | \xi) ds$ existiert für alle $\xi \in \mathfrak{H}$.

7.$\Rightarrow$8.: Mit Hilfe der „Polarisationsidentität" folgt die Existenz des Integrals $\int_0^\infty (U(s)\xi | U(s)\eta) ds$ für alle $\xi, \eta \in \mathfrak{H}$, d.h. es existiert $\int_0^\infty \psi(T(s)\mathbb{1}) ds$ für alle $\psi \in \mathfrak{A}_\sim$, $\mathfrak{A}_\sim$ der Vektorraum der Vektorzustände auf $\mathfrak{A}$, und es sei $\mathfrak{A}_\sim^+ := \mathfrak{A}_\sim \cap \mathfrak{A}_*^+$. Nach [14, III.4.2 und II.2.6] ist $\mathfrak{A}_\sim^+$ ein erzeugender Kegel in $\mathfrak{A}_\sim$. Wie im Beweis der Implikation von 3. nach 5. schließen wir, daß für alle $x \in \mathfrak{A}$ und $\psi \in \mathfrak{A}_\sim$ $\int_0^\infty \psi(T(s)x) ds$ existiert, also erst recht für alle $\xi, \eta \in \mathfrak{H}$ $\int_0^\infty ((T(s)x)\xi | \eta) ds$.

8.⇒9.: Sei $0 \leqq \varphi \in \mathfrak{A}_*$, $0 \leqq x \in \mathfrak{A}$ und sei $0 < t \in \mathbb{R}$ fest aber beliebig. Da $\mathfrak{A}_\sim^+$ dicht in $\mathfrak{A}^+$ ist [14, 4.10], existiert ein $0 \leqq \psi \in \mathfrak{A}_\sim$ mit

$$\|\varphi - \psi\| \leqq (t\,\|x\| \sup\{\|T(s)\|:\ 0 \leqq s \leqq t\})^{-1}.$$

Dann gilt

$$\int_0^t \varphi(T(s)\,x)\,ds \leqq \|\varphi - \psi\|\,(t\,\|x\| \sup\{\|T(s)\|:\ 0 \leqq s \leqq t\}) + \int_0^t \psi(T(s)\,x)\,ds$$
$$\leqq 1 + \int_0^\infty \psi(T(s)\,x)\,ds.$$

Daraus folgt aber die Existenz von $\int_0^\infty \varphi(T(s)\,x)\,ds$ für alle $\varphi \in \mathfrak{A}_*$ und alle $x \in \mathfrak{A}$.

9.⇒1.: Da $R(0, A)$ existiert, folgt wegen $s(A) \in \sigma(A)$, $s(A) < 0$. □

Literatur

1. Bellman, R.E.: Introduction to matrix analysis. New York: McGraw Hill 1960
2. Bratteli, O., Robinson, D.W.: Operator algebras and quantum statistical mechanics I. Berlin, Heidelberg, New York: Springer 1979
3. Choquet, G., Foias, C.: Solution d'un problème sur les itérés d'un operateur positif sur $C(K)$ et propriétés des moyennes associées. Ann. Inst. Fourier (Grenoble) **25**, 109–129 (1975)
4. Derndinger, R., Nagel, R.: Der Generator stark stetiger Verbandshalbgruppen auf $C(X)$ und dessen Spektrum. Math. Ann. **245**, 159–177 (1979)
5. Greiner, G., Voigt, J., Wolff, M.: On the spectral bound of the generator of semigroups of positive operators. J. Operator Theory **5**, 245–256 (1981)
6. Groh, U.: The peripheral spectrum of Schwarz operators on C^*-algebras. Math. Z. **176**, 311–318 (1981)
7. Hille, E.: Une generalisation du problem de Cauchy. Ann. Inst. Fourier (Grenoble), **4**, 31–48 (1953)
8. Hille, E., Phillips, R.S.: Functional analysis and semigroups. 2nd ed. Amer. Math. Soc. Coll. Publ. Vol. 31, Providence, R.I. 1957
9. Mil'stein, G.N.: Exponential stability of positive semigroups in a linear topological space I, II. Matem. Soviet Math. (Iz. VUZ.) **19**, 35–42, 51–61 (1975)
10. Pazy, A.: Semi-groups of linear operators and applications to partial differential equations. University of Maryland, Lecture Note 10 (1974)
11. Pedersen, G.K.: On the operator equation $HT + TH = 2K$. Indiana Univ. Math. J. **25**, 1029–1034 (1976)
12. Schaefer, H.H.: Topological vector spaces. 4th print. Berlin, Heidelberg, New York: Springer 1980
13. Schaefer, H.H.: Ordnungsstrukturen in der Operatorentheorie. Jber. Deutsch. Math.-Verein. **82**, 33–50 (1980)
14. Takesaki, M.: Theory of operator algebras I. Berlin, Heidelberg, New York: Springer 1979
15. Zabczyk, J.: A note on C_0-semigroups. Bull. Acad. Pol. Sci. **23**, 895–898 (1975)

Eingegangen am 6. September 1980; revidierte Fassung am 13. Februar 1981

Math. Ann. 256, 517–534 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Holomorphic Siegel Modular Forms Associated to SO(n, 1)

Stephen S. Kudla*

Department of Mathematics, University of Maryland, College Park, MD 20742, USA

Introduction

In 1926 Hecke [4] constructed a family of holomorphic elliptic modular forms of weight 1 associated to a real quadratic field. These forms are produced by integrating a certain non-holomorphic theta-series, associated to the norm form of the field, over a fundamental domain for a certain group of units. Hecke's idea was later taken up by Siegel in his celebrated paper [9] on indefinite quadratic forms. Once again the integral of a certain theta-series over a fundamental domain of a group of units of an indefinite quadratic form is an elliptic modular form – in fact an Eisenstein series – but this form is never holomorphic. In his concluding remarks, Siegel poses the problem of finding an exact generalization of Hecke's result for indefinite quadratic forms of arbitrary signature. For forms of signature $(n, 1)$ Siegel constructed one analogue of Hecke's theta-series and observed that its integral is identically zero unless $n=1$ (Hecke's case).

In this paper we will show that there is, in fact, a natural generalization of Hecke's construction for indefinite quadratic forms of signature $(n, 1)$ which produces a holomorphic Siegel modular form of genus n and weight $\frac{1}{2}(n+1)$ as the integral of a non-holomorphic theta-series. For $n=1$ our result coincides with Hecke's, while for each $n>1$ we show by example that our Siegel modular forms can be non-zero.

More precisely, let V be a $\mathbb{Q}$ vector space with a non-degenerate symmetric $\mathbb{Q}$-bilinear form $(\ ,\ )$ of signature $(n, 1)$. For convenience, we assume that $m=n+1$ is even. Let $L_0 \subset V$ be a lattice such that $L_0 \subset L_0^*$ where

$$L_0^* = \{v \in V \mid (v, v') \in \mathbb{Z}, \forall\, v' \in L_0\}$$

is the dual lattice. Let $L = L_0^n$, and let $L^* = (L_0^*)^n$. Let $G = \mathrm{SO}(V)$, let

$$E_L = \{g \in G \mid gL = L \text{ and } g \text{ acts trivially in } L^*/L\}$$

* Research partially supported by National Science Foundation Grant No. MCS78-02817

0025-5831/81/0256/0517/$03.60

be the group of units of L, and let

$$E = E_L^+ = E_L \cap G(\mathbb{R})^0$$

where $G(\mathbb{R})^0$ is the connected component of $G(\mathbb{R})$.

Then for $h \in L^*/L$, which is non-singular as defined in Sect. 2, and for $\tau \in \mathfrak{H}_n$, the Siegel space of genus n, we define

$$\vartheta(\tau, h, L) = \sum_{\substack{X \in V^n \\ X - h \in L \\ (X,X) > 0 \\ \bmod E}} \varepsilon(X) |E_X|^{-1} e_n(\tau(X, X))$$

where $\varepsilon(X) = \pm 1$ is defined in Proposition 2.3.

Theorem. *$\vartheta(\tau, h, L)$ is a holomorphic Siegel modular form of weight $\frac{1}{2}(n+1)$. In particular, if $m = n+1$ is even, and if $(X, Y) \equiv 0(2)$ for all X and Y in L_0, then*

$$\vartheta(\gamma\tau, h, L) = \left(\frac{(-1)^{\frac{1}{2}m} \Delta}{|\det d|}\right) (\operatorname{sgn} \det d)^{\frac{1}{2}(n-1)} \det(c\tau + d)^{\frac{1}{2}m}$$

$$\cdot e_n((h, h)\, b^t a)\, \vartheta(\tau, ha, L)$$

for all $\gamma = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathrm{Sp}(n, \mathbb{Z})$ with $c \equiv 0(N)$ where N and Δ are given in Proposition 1.1.

To prove these facts we construct, in Sect. 1, a certain non-holomorphic theta-series Θ whose transformation law in τ is readily determined and which may be viewed as an E invariant n-form on the symmetric space $D = \mathrm{SO}(n, 1)^0/\mathrm{SO}(n)$. We then obtain the function ϑ by integrating Θ over a fundamental domain for E in D. The necessary integrals are computed in Sect. 2. In Sect. 3, we give an example of a non-vanishing ϑ for each $n \geqq 1$. Finally, we show that for $n = 1$ and h singular, a slight modification of our procedure yields Hecke's Eisenstein series of weight 1.

The family of functions $\vartheta(\tau, h, L)$ is intimately connected with the generating function for intersection numbers constructed in [7]. In fact the construction of the present paper can be used to extend the mapping I_k of [7] to the case where $k > \frac{1}{4}m$ and to prove the non-vanishing of this mapping in certain cases. This will be discussed in detail elsewhere.

Finally, it should be possible to give a representation theoretic interpretation of our result. Roughly speaking we have shown that the trivial representation of $\mathrm{SO}(n, 1)$ is paired with a certain 'holomorphic' representation of $\mathrm{Sp}(n, \mathbb{R})$ under the correspondence determined by the Weil/oscillator representation of the pair $\mathrm{Sp}(n, \mathbb{R}) \times O(n, 1)$, [3, 5, 6].

Notation. For $X \in M_n(\mathbb{C})$, we let $e_n(X) = e^{i\pi \operatorname{tr}(X)}$. We let

$$\mathfrak{H}_n = \{\tau \in M_n(\mathbb{C}) \mid {}^t\tau = \tau \text{ and } v = \operatorname{Im} \tau > 0\}.$$

Finally, we view $X = \{X_1, X_2, \ldots, X_k\} \in V^k$ as a row vector with values in V, and for $X \in V^k$ and $Y \in V^l$ we let $(X, Y) = ((X_i, Y_j)) \in M_{k,l}(\mathbb{Q})$.

1. A Theta-Series and Its Transformation Law

Let V and $(\ ,\ )$ be as in the introduction. Let

$$D=\{Z\in V(\mathbb{R})\,|\,(Z,Z)=-1\}^0$$

be one component of the hyperbolid of two sheets in $V(\mathbb{R})$, and let $G=\mathrm{SO}(V)$ be the special orthogonal group of V viewed as an algebraic group over $\mathbb{Q}$. We will identify the tangent space $T_Z(D)$ to D at Z with the subspace $Z^\perp\subset V(\mathbb{R})$. We fix an orientation of $V(\mathbb{R})$ and choose an orientation of $T_Z(D)$ such that the basis $\{v_1,\ldots,v_n,Z\}$ is properly oriented for $V(\mathbb{R})$ if $\{v_1,\ldots,v_n\}$ is a properly oriented basis for $T_Z(D)=Z^\perp$. Note that the action of $G(\mathbb{R})^0$, the connected component of $G(\mathbb{R})$, preserves this orientation.

For $Z\in D$, the majorant of $(\ ,\ )$ associated to Z is given by

$$(X,X)_Z=(X_1,X_1)-2(X_2,X_2)$$

where $X=X_1+X_2\in V(\mathbb{R})$, with $X_1\in Z^\perp$, and $X_2\in\operatorname{span} Z$. For $\tau=u+iv\in\mathfrak{H}_n$, $Z\in D$, and $X\in V(\mathbb{R})^n$, we let

$$(X,X)_{\tau,Z}=u(X,X)+iv(X,X)_Z$$

so that $(\ ,\ )_{\tau,Z}\colon V(\mathbb{R})^n\times V(\mathbb{R})^n\to M_n(\mathbb{C})$; and we let

$$f_{\tau,Z}(X,W)=\det v^{\frac{1}{2}}\det(X,W)\,e_n((X,X)_{\tau,Z})$$

where $W\in(Z^\perp)^n$. Then $f_{\tau,Z}(\cdot,W)$ is a Schwarz function on $V(\mathbb{R})$, and, for fixed $X\in V(\mathbb{R})^n$, we may view $\omega_X=f_{\tau,\cdot}(X,\cdot\,)$ as an n-form on D, via

$$\omega_{X,Z}(W)=f_{\tau,Z}(X,W).$$

Note that, for $g\in G(\mathbb{R})^0$

$$g^*\omega_X=\omega_{g^{-1}X}.$$

Now choose a lattice $L_0\subset V(\mathbb{Q})$ with $L_0^*\supset L_0$ and let L and L^* be as in the introduction. For $h\in L^*/L$ and $\tau\in\mathfrak{H}_n$, define the theta-series

$$\Theta(\tau,h,Z,W)=\sum_{\substack{X\in V(\mathbb{Q})^n\\ X-h\in L}} f_{\tau,Z}(X,W).$$

We may view $\Omega=\Theta(\tau,h,\cdot,\cdot\,)$ as an n-form on D via

$$\Omega_Z(W)=\Theta(\tau,h,Z,W),$$

and we observe that Ω will be invariant under $E=E_L^+$.

Take $N\in\mathbb{Z}_{>0}$ such that

$$NL_0^*\subset L_0\quad\text{and}\quad N(X,Y)\in 2\mathbb{Z},\qquad\forall\,X,Y\in L_0^*,$$

and let

$$\Gamma_L=\left\{\gamma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\mathrm{Sp}(n,\mathbb{Z})\,\middle|\,\operatorname{tr}(Xa,Xb)\in 2\mathbb{Z}\text{ and }\operatorname{tr}(Yc,Yd)\in 2\mathbb{Z},\ \forall\,X,Y\in L\right\}$$

and

$$\Gamma_0(N)=\{\gamma\in\Gamma_L \mid c\equiv 0(N)\}.$$

Then we have the following transformation law:

Proposition 1.1. i) *For all* $\gamma=\begin{pmatrix}a & b\\ c & d\end{pmatrix}\in\Gamma_0(N)$,

$$\Theta(\gamma\tau, h, Z, W)=\lambda_\gamma\,\mathfrak{g}(b, d; L)\,\det(c\tau+d)^{\frac{1}{2}m}\,e_n((h, h)\,b^t a)\,\Theta(\tau, ha, Z, W)$$

where $\lambda_\gamma^8=1$, and

$$\mathfrak{g}(b, d; L)=|\det d|^{-\frac{1}{2}m}\sum_{X\in L/Ld} e_n((X, X)\,bd^{-1}).$$

ii) *If* $m=n+1$ *is even and* $(X, Y)\equiv 0(2)$, *for all* $X, Y\in L_0$, *then*

$$\Theta(\gamma\tau, h, Z, W)=\chi(d)\,\det(c\tau+d)^{\frac{1}{2}m}\,e_n((h, h)\,b^t a)\,\Theta(\tau, ha, Z, W),$$

where

$$\chi(d)=\left(\frac{(-1)^{\frac{1}{2}m}\Delta}{|\det d|}\right)(\operatorname{sgn}\det d)^{\frac{1}{2}(n-1)}$$

and $\Delta=\det((v_i, v_j))$ *where* $\{v_1, \ldots, v_m\}$ *is a* $\mathbb{Z}$ *basis for* L_0.

Proof. By Proposition 8.3 of [17] we have

$$\Theta(\gamma\tau, h, Z, W)=\varepsilon(\gamma)^{-1}\,c(\gamma\omega^{-1}, \omega)\,\mathfrak{g}(b, d; L)\,\det(c\tau+d)^{\frac{1}{2}m}\cdot e_n((h, h)\,b^t a)\,\Theta(\tau, ha, Z, W)$$

where $(\varepsilon(\gamma)^{-1}\,c(\gamma\omega^{-1}, \omega))^8=1$ and $\mathfrak{g}(b, d; L)$ is as above. If, in addition, $m=n+1$ is even, then

$$\varepsilon(\gamma)^{-1}\,c(\gamma\omega^{-1}, \omega)=(\operatorname{sgn}\det d)^{\frac{1}{2}(n-1)}.$$

On the other hand, the simpler theta-series

$$\theta(\tau, Z)=\sum_{X\in L} e_n((X, X)_{\tau, Z})$$

satisfies

$$\theta(\gamma\tau, Z)=(\operatorname{sgn}\det d)^{\frac{1}{2}(n-1)}\,\mathfrak{g}(b, d; L)\,\det(c\tau+d)^{\frac{1}{2}(n-1)}\,\theta(\tau, Z)$$

for all $\gamma\in\Gamma_0(N)$. Thus

$$\mu(\gamma)=(\operatorname{sgn}\det d)^{\frac{1}{2}(n-1)}\,\mathfrak{g}(b, d; L)$$

defines a character of $\Gamma_0(N)$ which, if we assume that $(X, Y)\equiv 0(2)$ for all X and Y in L_0, is trivial on the subgroup $\Gamma_0'(N)$ generated by the matrices $\begin{pmatrix}a & \\ & a^{-1}\end{pmatrix}$, with $a\in\mathrm{SL}_n(\mathbb{Z})$, $\begin{pmatrix}1 & b\\ & 1\end{pmatrix}$, with ${}^t b=b\in M_n(\mathbb{Z})$, and $\begin{pmatrix}1 & \\ c & 1\end{pmatrix}$, with ${}^t c$

$=c\in NM_n(\mathbb{Z})$. By Lemma 5 of [1], the matrices

$$\gamma_0=\left(\begin{array}{c|c} \begin{matrix}1_{n-1} & \\ & a_0\end{matrix} & \begin{matrix} & b_0\end{matrix} \\ \hline \begin{matrix} & \\ c_0 & \end{matrix} & \begin{matrix}1_{n-1} & \\ & d_0\end{matrix}\end{array}\right)$$

with $\begin{pmatrix}a_0 & b_0\\ c_0 & d_0\end{pmatrix}\in \mathrm{SL}_2(\mathbb{Z})$ and $c_0\equiv 0(N)$ give a set of double coset representatives for $\Gamma_0'(N)\backslash\Gamma_0(N)/\Gamma_0'(N)$. For such γ_0,

$$\begin{aligned}\mathfrak{g}(b,d;L)&=|d_0|^{-\frac{1}{2}m}\sum_{X\in L_0/d_0L_0} e(\tfrac{1}{2}(X,X)b_0d_0^{-1})\\ &=\left(\left(\frac{-1}{d_0}\right)\operatorname{sgn}d_0\right)^{\frac{1}{2}m}\left(\frac{\Delta}{|d_0|}\right)\end{aligned}$$

by Proposition 1.6 of Shintani [10]. Therefore

$$\mu(\gamma)=(\operatorname{sgn}\det d)^{\frac{1}{2}(n-1)}\left(\frac{(-1)^{\frac{1}{2}m}\Delta}{|\det d|}\right)$$

as claimed.

2. The Integral of the Theta-Series

In this section we will compute the integral of the n-form Ω determined by $\Theta(\tau,h,\cdot,\cdot)$ over a fundamental domain for E in D. To do this we must first exclude some singular terms.

Definition. Let $h\in L^*/L$. Then h is non-singular if, for any choice of representative $h'\in L^*$ for h, the following conditions hold:

1) $(h',h')\notin M_n(\mathbb{Z})$.

2) If $M\in Z_{>0}$ is the least integer such that $M(h',h')\in M_n(\mathbb{Z})$, then $M(h',h')$ is invertible modulo M, i.e. $M(h',h')\in \mathrm{GL}_n(\mathbb{Z}/M\mathbb{Z})$.

Observe that, since $(h',L)\subset M_n(\mathbb{Z})$, these conditions are independent of the choice of h'. Moreover, condition 2) implies that $\det(h',h')\neq 0$.

The following result will be proved in the appendix:

Proposition 2.1. *Suppose that $h\in L^*/L$ is non-singular. Then the integral*

$$I(\tau)=\int_{E\backslash D}\Omega$$

is absolutely convergent. Moreover, this integral may be computed term by term.

Now recalling that $g^*\omega_X=\omega_{g^{-1}X}$, where $\omega_X=f_{\tau,\cdot}(X,\cdot)$, as above, we have:

$$I(\tau)=\int_{E\backslash D}\sum_{X\equiv h(L)}\omega_X=\int_{E\backslash D}\sum_{\substack{X\equiv h(L)\\ \bmod E}}\Big(\sum_{g\in E_X\backslash E}\omega_{g^{-1}X}\Big)=\sum_{\substack{X\equiv h(L)\\ \bmod E}}\int_{E_X\backslash D}\omega_X=\sum_{\substack{X\equiv h(L)\\ \bmod E}}I(\tau,X)$$

where we let $E_X=\{g\in E \mid gX=X\}$ and $I(\tau, X)=\int\limits_{E_X\backslash D} \omega_X$.

Observe that $\omega_X=0$ unless $\dim_{\mathbb{R}} \operatorname{span} X=n$.

Lemma 2.2. *If* $\dim_{\mathbb{R}} \operatorname{span} X=n$, *then* E_X *is finite.*

Proof. First suppose that $(\ ,\ )\restriction_{\operatorname{span} X}$ is non-degenerate. Then

$$G_X(\mathbb{R})=\{g\in G(\mathbb{R}) \mid gX=X\}$$
$$\cong \mathrm{SO}((\operatorname{span} X)^{\perp})$$

is compact, and $E_X=E\cap G_X(\mathbb{R})$ is finite. Next, if $(\ ,\ )\restriction_{\operatorname{span} X}$ is degenerate, take $X_0\in\operatorname{span} X$ such that $X_0\neq 0$ but $(X_0, X)=0$. Then choose $Y_0\in V(\mathbb{R})$ such that $(Y_0, Y_0)=0$ and $(Y_0, X_0)=\frac{1}{2}$. Note that $\operatorname{span}\{X_0, Y_0\}^{\perp}\subset\operatorname{span} X$, so that $G_X(\mathbb{R})$ preserves span $\{X_0, Y_0\}$, and the restriction map $G_X(\mathbb{R})\to O(\operatorname{span}\{X_0, Y_0\})$ is injective. But if $g\in O(\operatorname{span}\{X_0, Y_0\})$ with $gX_0=X_0$, then $gY_0=Y_0+aX_0$ and $0=2a=(gY_0, gY_0)$. Thus $G_X(\mathbb{R})=E_X=1_V$.

We now observe that by the definition,

$$(X, X)_{\tau, Z}=\tau(X, X)+2iv(X, Z)(Z, X),$$

and

$$\omega_{X,Z}(W)=e_n(\tau(X, X))\det v^{\frac{1}{2}}\det(X, W)\exp(-2\pi\operatorname{tr}(v(X, Z)(Z, X))).$$

Therefore, we have:

$$I(\tau, X)=|E_X|^{-1}\det v^{\frac{1}{2}}\, e_n(\tau(X, X))\, I_0(v, X)$$

where

$$I_0(v, X)=\int\limits_D \omega_X^0$$

and

$$\omega_{X,Z}^0(W)=\det(X, W)\exp(-2\pi\operatorname{tr}(v(X, Z)(Z, X))).$$

Next we observe that, for $h\in \mathrm{GL}_n(\mathbb{R})$,

$$I_0({}^thvh, X)=\det h^{-1}\, I_0(v, X\,{}^th),$$

so that, writing $v=h\,{}^th$ with $h\in \mathrm{GL}_n^+(\mathbb{R})$, we have

$$I_0(v, X)=\det v^{-\frac{1}{2}} I_0(1_n, X\,{}^th).$$

Therefore it will be sufficient to compute the integrals $I_0(X)=I_0(1_n, X)$ for $X\in V(\mathbb{R})^n$ with $\dim_{\mathbb{R}}\operatorname{span} X=n$. There will be three cases depending on $(\ ,\ )\restriction_{\operatorname{span} X}$.

Proposition 2.3. *Assume that* $(\ ,\ )\restriction_{\operatorname{span} X}$ *is positive definite, and let* $Z_0=D\cap(\operatorname{span} X)^{\perp}$. *Then*

$$I_0(X)=2^{-\frac{1}{2}n}\varepsilon(X)$$

where

$$\varepsilon(X)=\begin{cases}+1 & \text{if } X \text{ is a properly oriented basis for } T_{Z_0}(D),\\ -1 & \text{otherwise.}\end{cases}$$

Proof. For $a\in\mathbb{R}^n$, let

$$Z_0(a)=Z_0+X\cdot a$$

where $Z_0=D\cap(\operatorname{span} X)^{\perp}$ and $X\cdot a=\sum_{k=1}^{n} a_k X_k$. Then $(Z_0(a),\ Z_0(a))=q^2-1$ where $q^2=(X\cdot a, X\cdot a)$; and we have a parameterization of D via

$$\varphi\colon \{a\in\mathbb{R}^n \mid q^2<1\}\to D$$
$$a\mapsto Z(a)=Z_0(a)(1-q^2)^{-\frac{1}{2}}.$$

The corresponding coordinate frame field on D will be given by

$$(d\varphi)\left(\frac{\partial}{\partial a_k}\right)=(1-q^2)^{-\frac{1}{2}}X_k+(1-q^2)^{-3/2}(X\cdot a, X_k)Z_0(a).$$

Thus, setting $W=d\varphi\left(\left\{\frac{\partial}{\partial a_1},\ldots,\frac{\partial}{\partial a_n}\right\}\right)$, we have

$$\begin{aligned}\det(X,W)&=\det((1-q^2)^{-\frac{1}{2}}(X,X)+(1-q^2)^{-3/2}(X,X\cdot a)(X\cdot a,X))\\&=(1-q^2)^{-3n/2}\det((X,X)(1_n-(X\cdot a,X\cdot a))+(X,X)\,{}^t a(X,X))\\&=(1-q^2)^{-3n/2}\det(X,X)\det((1-{}^t a(X,X)\,a)1_n+a\,{}^t a(X,X)).\end{aligned}$$

Writing $(X,X)={}^tTT$ with $T\in\mathrm{GL}_n^+(\mathbb{R})$, and setting $b=Ta$, we obtain for the last factor

$$\begin{aligned}\det((1-{}^t bb)1_n+b\,{}^t b)&=(1-{}^t bb)^{n-1}\\&=(1-q^2)^{n-1}.\end{aligned}$$

Thus

$$\det(X,W)=(1-q^2)^{-1-\frac{1}{2}n}\det(X,X).$$

Similarly, for $Z=Z(a)$,

$$\begin{aligned}\operatorname{tr}((X,Z)(Z,X))&=(Z,X)(X,Z)\\&=(1-q^2)^{-1}(X\cdot a,X)(X,X\cdot a)\\&=(1-q^2)^{-1}\,{}^t a(X,X)^2 a.\end{aligned}$$

Thus we have

$$I_0(X)=\varepsilon(X)\det(X,X)\int_{q^2<1}(1-q^2)^{-1-n/2}\exp(-2\pi(1-q^2)^{-1}\check{Q}(a))\,da$$

where $q^2=Q(a)={}^t a(X,X)\,a$ and $\check{Q}(a)={}^t a(X,X)^2 a$. But an easy calculation shows that

$$\begin{aligned}&\int_{{}^t aQa<1}(1-{}^t aQa)^{-1-n/2}\exp(-2\pi(1-{}^t aQ^2a)\,da\\&\quad=\int_{R^n}\exp(-2\pi\,{}^t bQ^2b)\,db=2^{-n/2}\det Q^{-1}.\end{aligned}$$

Thus $I_0(X)=\varepsilon(X)\,2^{-n/2}$ as claimed.

Proposition 2.4. *Assume that* $(\ ,\ )|_{\mathrm{span}\, X}$ *is indefinite but non-degenerate. Then*

$$I_0(X)=0.$$

Proof. Since $(\ ,\ )|_{\mathrm{span}\, X}$ is indefinite, $\mathrm{span}\, X$ has signature $(n-1,1)$ for $(\ ,\ )$, so that $D_0=D\cap \mathrm{span}\, X$ has dimension $n-1$. Take $X'\in(\mathrm{span}\, X)^{\perp}$ with $(X',X')=1$, and for $a\in\mathbb{R}$, and $Z_0\in D_0$, set

$$Z_0(a)=Z_0+aX'.$$

Then $(Z_0(a),Z_0(a))=a^2-1$, and we obtain a parameterization of D:

$$\varphi\colon D_0\times(-1,1)\to D$$
$$(Z_0,a)\mapsto Z(a)=Z_0(a)(1-a^2)^{-\frac{1}{2}}.$$

Identifying $T_{Z_0}(D_0)\cong Z_0\cap \mathrm{span}\, X$ we have

$$d\varphi(W_0)=(1-a^2)^{-\frac{1}{2}}W_0,$$

and

$$d\varphi\left(\frac{d}{da}\right)=(1-a^2)^{-\frac{1}{2}}X'+a(1-a^2)^{-3/2}Z_0(a).$$

Thus if we take $W=d\varphi\left\{\dfrac{d}{da},W_0\right\}$ for any coordinate frame field W_0 on D_0,

$$\det(X,W)=a(1-a^2)^{-2}\det((X,Z_0),(X,W_0)).$$

Likewise we have

$$\mathrm{tr}((X,Z(a))(Z(a),X))=(Z(a),X)(X,Z(a))$$
$$=(1-a^2)^{-1}(Z_0,X)(X,Z_0).$$

Thus

$$I_0(X)=\int_{-1}^{1}a(1-a^2)^{-1}\Big(\int_{D_0}\mu_X(a)\Big)\,da$$

where

$$\mu_X(a)_{Z_0}(W_0)=\det((X,Z_0),(X,W_0))\exp(-2\pi(1-a^2)^{-1}(Z_0,X)(X,Z_0)).$$

But it is now clear that the integrand is an odd function of a, so that $I_0(X)=0$ as claimed.

Although the condition that h be non-singular excludes the terms associated to X for which $(\ ,\ )|_{\mathrm{span}\, X}$ is degenerate, the 'local' integral $I(\tau,X)$ is still well defined for such X, and we give its value for the sake of completeness.

Proposition 2.5. *Assume that* $(\ ,\ )|_{\mathrm{span}\, X}$ *is degenerate, and take* $X'\in\mathrm{span}\, X$ *so that* $(X',X)=0$. *Take* $X''\in V$ *such that* $(X'',X')=\frac{1}{2}$ *and* $(X'',X'')=0$. *Take any basis* X_0 *for span* $\{X',X''\}^{\perp}$ *and set* $X^{\natural}=\{X'+X'',X_0\}$. *Then*

$$I_0(X)=\varepsilon(X^{\natural})\,\mathrm{sgn}\det(X,X^{\natural})\,2^{-1-n/2}$$

where ε *is as in Proposition* 2.3.

Proof. For $t\in\mathbb{R}_+^\times$, set

$$Z_0(t)=tX''-t^{-1}X',$$

so that $(Z_0(t), Z_0(t))=-1$. Then for $a\in\mathbb{R}^{n-1}$ set

$$Z_0(t, a)=Z_0(t)+X_0\cdot a$$

and

$$Z(t, a)=Z_0(t, a)(1-q^2)^{-\frac{1}{2}}$$

where $q^2=(X_0\cdot a, X_0\cdot a)$. We then have a parameterization

$$\varphi\colon \mathbb{R}_+^\times\times\{a\in\mathbb{R}^{n-1}\mid q^2<1\}\to D$$

$$(t, a)\mapsto Z(t, a).$$

As in the previous cases

$$d\varphi\left(\frac{\partial}{\partial t}\right)=(1-q^2)^{-\frac{1}{2}}(X''+t^{-2}X')$$

and

$$d\varphi\left(\frac{\partial}{\partial a_k}\right)=(1-q^2)^{-\frac{1}{2}}X_{0,k}+(1-q^2)^{-3/2}(X_{0,k}, X_0 a)Z_0(t, a).$$

Setting $W=d\varphi\left(\left\{\frac{\partial}{\partial t}, \frac{\partial}{\partial a}\right\}\right)$ we obtain:

$$\det(X, W)=(1-q^2)^{-n/2}\det((X, X''), (X, X_0)$$
$$+(1-q^2)^{-1}(X, Z_0(t, a))(X_0\cdot a, X_0)).$$

Now

$$(X, Z_0(t, a))=t(X, X'')+(X, X_0\cdot a)$$

so that, adding a suitable multiple of the first column (X, X'') to the remaining columns, we obtain:

$$(1-q^2)^{-\frac{1}{2}n-(n-1)}\det((X, X''), (1-q^2)(X, X_0)+(X, X_0\cdot a)(X_0\cdot a, X_0)).$$

If we write $(X_0, X_0)={}^tTT$ and set $b=Ta$, we obtain:

$$(1-q^2)^{-\frac{1}{2}n-(n-1)}\det((X, X''), (X, X_0))\det\begin{pmatrix}1 & \\ & (1-{}^tbb)1_{n-1}+b\,{}^tb\end{pmatrix}$$
$$=(1-q^2)^{-1-n/2}\det(X, X^\natural).$$

Next, for $Z=Z(t, a)$, we must compute

$$(X, Z)=(1-q^2)^{-\frac{1}{2}}[t(X, X'')+(X, X_0\cdot a)]$$

and

$$(Z, X)(X, Z)=(1-q^2)^{-1}[At^2+2tB(a)+C(a)]$$

where

$$(X^\natural, X)(X, X^\natural)=\begin{pmatrix}A & B\\ {}^tB & C\end{pmatrix},$$

$$B(a)=B\cdot a$$

and

$$C(a) = {}^t a\, C\, a.$$

Thus we obtain:

$$I_0(X) = \varepsilon(X^\natural) \int_0^\infty \int_{q^2<1} \det(X, X^\natural)(1-q^2)^{-1-n/2} \cdot \exp(-2\pi(1-q^2)^{-1}[At^2 + 2B(a)\,t + C(a)])\, da\, dt.$$

Here we note that at the point $t=1$, $a=0$, the coordinate frame for our parameterization is precisely $X^\natural$. Computing the 't' integral first we obtain:

$$\int_0^\infty \exp(-2\pi(1-q^2)^{-1} A[t+B'(a)]^2)\, dt\, \exp(2\pi(1-q^2)^{-1} AB'(a)^2)$$

where $B'(a) = A^{-1}B(a)$. The integral here is

$$\int_0^\infty \exp(-2\pi(1-q^2)^{-1} At^2)\, dt - \int_0^{B'(a)} \exp(-2\pi(1-q^2)^{-1} At^2)\, dt$$
$$= \tfrac{1}{2}(2\pi A)^{-\frac{1}{2}}(1-q^2)^{-\frac{1}{2}}\Gamma(\tfrac{1}{2}) - I(a),$$

where $I(a)$ is the second integral. The contribution to $I_0(X)$ of this second term is:

$$-\varepsilon(X^\natural)\det(X, X^\natural) \int_{q^2<1} (1-q^2)^{-1-n/2} \cdot \exp(2\pi(1-q^2)^{-1}[A^{-1}B(a)^2 - C(a)])\, I(a)\, da.$$

But since $I(-a) = -I(a)$, this integral is zero. The contribution of the first term above is:

$$\varepsilon(X^\natural)\det(X, X^\natural)\tfrac{1}{2}(2A)^{-\frac{1}{2}} \int_{q^2<1} (1-q^2)^{-1-\frac{1}{2}(n-1)} \exp(-2\pi(1-q^2)^{-1}\check{Q}(a))\, da,$$

where

$$\check{Q}(a) = C(a) - A^{-1}B(a)^2$$
$$= {}^t a(C - {}^t BA^{-1}B)\, a.$$

As in the proof of Proposition 2.3, this integral is just

$$2^{-\frac{1}{2}(n-1)}\det(C - {}^t BA^{-1}B)^{-\frac{1}{2}},$$

but then

$$(X^\natural, X)(X, X^\natural) = \begin{pmatrix} A & B \\ {}^t B & C \end{pmatrix}$$
$$= \begin{pmatrix} A & \\ & C - A^{-1}[B] \end{pmatrix} \begin{bmatrix} 1 & A^{-1}B \\ & 1 \end{bmatrix}$$

where, for example, $A^{-1}[B] = {}^t BA^{-1}B$. Thus we obtain:

$$I_0(X) = \varepsilon(X^\natural)\det(X, X^\natural)\tfrac{1}{2}(2A)^{-\frac{1}{2}(n-1)}|\det(X, X^\natural)|^{-1} A^{\frac{1}{2}}$$
$$= 2^{-1-n/2}\varepsilon(X^\natural)\operatorname{sgn}\det(X, X^\natural)$$

as claimed.

Combining these results we obtain the following:

Theorem 2.6. *Assume that $h \in L^*/L$ is non-singular. Then the function*

$$2^{n/2} I(\tau) = \sum_{\substack{X \equiv h(L) \\ (X,X) > 0 \\ \bmod E}} |E_X|^{-1} \varepsilon(X) e_n(\tau(X,X))$$

is a holomorphic Siegel modular form of weight $\frac{1}{2}(n+1)$ whose transformation law is the same as that given in Proposition 1.1 for the function Θ. Here $\varepsilon(X)$ is defined in Proposition 2.2.

Remark. The function $\vartheta(\tau, h, L) = 2^{n/2} I(\tau)$ has rational Fourier coefficients. In fact, if E acts without fixed points on D, then $|E_X| = 1$ for all X for which $\varepsilon(X) \neq 0$. In this case the Fourier coefficients of $\vartheta(\tau, h, L)$ will be rational integers.

3. An Example

In this section we will show that for each $n \geqq 1$ there exists a lattice $L_0 \subset L_0^*$ and a non-singular $h \in L^*/L$ such that the function $\vartheta(\tau, h, L)$ does not vanish identically. The case $n = 1$ was already considered by Hecke [4].

Let $V = \mathbb{Q}^m$, $m = n+1$, and let

$$(X, Y) = {}^tX \begin{pmatrix} 1_n & \\ & -1 \end{pmatrix} Y.$$

We begin with the lattice $L_0 = \mathbb{Z}^m \subset V$ so that $L_0 = L_0^*$. Then for any $M \in \mathbb{Z}_{>0}$, we let

$$(X, Y)_M = M^{-1}(X, Y)$$

and

$$L_{0,M} = ML_0,$$

so that the dual lattice to $L_{0,M}$ with respect to $(\ ,\ )_M$ is again L_0^*. For $h \in L^*/ML$, non-singular, define

$$\begin{aligned} \vartheta(\tau, h, M, L) &= \vartheta(\tau, h, L_{0,M}, (\ ,\ )_M) \\ &= \sum_{\substack{X \equiv h(ML) \\ (X,X) > 0 \\ \bmod E}} |E_X|^{-1} \varepsilon(X) e_n(\tau(X,X)), \end{aligned}$$

where we have

$$E_L(M) = \{g \in \mathrm{SO}(V, (\ ,\ )) \mid gL_0 = L_0 \text{ and } g \text{ acts trivially in } L_0^*/ML_0\}$$

and

$$E = E_L^+(M) = E_L(M) \cap \mathrm{SO}^0(n, 1).$$

We want to show the following:

Theorem 3.1. *Let $h_1 = \{e_1, \ldots, e_n\}$ where the e_i are the standard basis vectors for $V = \mathbb{Q}^m$. Then there exist infinitely many $M \in \mathbb{Z}_{>0}$ such that*

$$\vartheta(\tau, h_1, M, L) \neq 0.$$

Proof. Let

$$\vartheta(\tau, h_1, M, L) = \sum_{\beta > 0} a(\beta)\, e_n(\tau\beta)$$

where $\beta = {}^t\beta \in M^{-1} M_n(\mathbb{Z})$, be the Fourier expansion of ϑ. We will show that, for a suitable choice of M,

$$a(M^{-1} 1_n) \neq 0.$$

First observe that if $X \in L$ with $(X, X) = 1_n$, then there exists $X' \in V(\mathbb{Q})$ such that $(X, X') = 0$ and

$$L_0 = \mathbb{Z}\text{-span of } X + \mathbb{Z} \cdot X'.$$

In fact, if $Y \in L_0$, then

$$Y' = Y - X(X, Y) \in L_0 \cap (\operatorname{span} X)^{\perp},$$

and we may take a primitive vector X' in $L_0 \cap (\operatorname{span} X)^{\perp}$. Moreover, $(X', X') = -1$ since, if we write $X^{\#} = \{X, X'\}$

$$(X', X') = \det(X^{\#}, X^{\#}) = \operatorname{disc} L_0 = -1.$$

Thus there exists $g \in O(V, (\ ,\ )) \cap \mathrm{SL}_m(\mathbb{Z})$ such that $g h_1 = X$, and $g e_m = \pm X'$. This shows that

$$\{X \in L \mid (X, X) = 1_n\} = E_L(1) \cdot h_1.$$

Next observe that if $X \equiv h_1(M)$, and $X = g h_1$ with $g \in E_L(1)$, then $g \in E_L(M)$. To see this write

$$g = \begin{pmatrix} A & b \\ c & d \end{pmatrix}$$

with $A \in M_n(\mathbb{Z})$, $b \in \mathbb{Z}^n$ etc. Then, if we view X as an element of $M_{m,n}(\mathbb{Z})$, we have

$$X = g h_1 = \begin{pmatrix} A \\ c \end{pmatrix}$$

so that $X \equiv h_1(M)$ is equivalent to $A \equiv 1_n(M)$ and $c \equiv 0(M)$, i.e.

$$g \equiv \begin{pmatrix} 1_n & b \\ 0 & d \end{pmatrix} \bmod M.$$

But since $g \in \mathrm{SO}(n, 1)$, we also have $d \equiv \det g \equiv 1(M)$ and ${}^t b A = dc$ so that $b \equiv 0(M)$. Therefore

$$\begin{aligned} \{X \in L \mid (X, X) = 1_n \text{ and } X \equiv h_1 \bmod ML\} \\ = E_L(M) \cdot h_1. \end{aligned}$$

Finally we observe that

$$E = E_L^+(M) = \{g \in E_L(M) \mid \theta(g) = 1\}$$

where $\theta : O(V, (,)) \to \mathbb{Q}^{\times}/(\mathbb{Q}^{\times})^2$ is the spinor norm. But by Millson and Raghunathan [8] there exist infinitely many $M \in \mathbb{Z}_{>0}$ such that $\theta(E_L(M)) = 1$ in

$\mathbb{Q}^\times/(\mathbb{Q}^\times)^2$ and hence $E_L^+(M)=E_L(M)$. For any such M we have

$$a(M^{-1}1_n)=\sum_{\substack{X\equiv h_1(ML)\\(X,X)=1_n\\ \bmod E}}\varepsilon(X)$$
$$=\varepsilon(h_1)\neq 0.$$

This proves the theorem.

We now restrict to the special case $n=1$, and we assume that $V=\mathbb{Q}^2$, and $(\ ,\ )=\begin{pmatrix}&1\\1&\end{pmatrix}$. Let $L^*=L=L_0=\mathbb{Z}^2$; and again, for $M\in\mathbb{Z}_{>0}$, let $L_M=ML$ and $(\ ,\ )_M=M^{-1}(\ ,\ )$. Note that in this case $E=\{1_V\}$.

If we parameterize D by

$$\varphi\colon R_+^\times\to D$$
$$t\mapsto \operatorname{span} M^{\frac12}2^{-\frac12}\begin{pmatrix}t\\-t^{-1}\end{pmatrix}=Z_t$$

so that

$$W_t=d\varphi\left(t\frac{d}{dt}\right)=M^{\frac12}2^{-\frac12}\begin{pmatrix}t\\t^{-1}\end{pmatrix}$$

then for $h=\begin{pmatrix}h_1\\h_2\end{pmatrix}\in\mathbb{Z}^2$,

$$\Theta(\tau,h;Z_t,W_t)=M^{-\frac12}2^{-\frac12}v^{\frac12}$$
$$\cdot\sum_{\substack{x_1\equiv h_1(M)\\x_2\equiv h_2(M)}}(t^{-1}x_1+tx_2)\,e(uM^{-1}x_1x_2+\tfrac12 ivM^{-1}(t^{-2}x_1^2+t^2x_2^2)).$$

We now have the following:

Theorem 3.2. i) *If* $(h,h)=2h_1h_2\neq 0$, *then*

$$2^{\frac12}I(\tau)=\int_{R_+^\times}\Theta(\tau,h;Z_t,W_t)\,t^{-1}\,dt$$
$$=\sum_{x_1x_2>0}\operatorname{sgn}x_1\,e(\tau M^{-1}x_1x_2).$$

ii) *If* $(h,h)=0$, *with* $h_1=0$ *and* $h_2\not\equiv 0(M)$, *then*

$$I_s(\tau)=\int_{R_+^\times}\Theta(\tau,h;Z_t,W_t)\,t^{2s-1}\,dt$$

is absolutely convergent for $\operatorname{Re}s>3/2$. *Moreover, the function* $I_s(\tau)$ *has a holomorphic analytic continuation to the whole s-plane, and*

$$2^{\frac12}I_0(\tau)=\frac12-\left\langle\frac{h_2}{M}\right\rangle+\sum_{\substack{x_1\equiv 0(M)\\x_2\equiv h_2(M)\\x_1x_2>0}}\operatorname{sgn}x_1\,e(\tau M^{-1}x_1x_2).$$

Here $\langle x\rangle\in[0,1)$ *is the fractional part of* x.

Proof. The first part of the theorem is a special case of Theorem 2.6. The absolute convergence of (ii) follows from the estimates of the appendix which shows that $\Theta(\tau, h; Z_t, W_t) = O(t^{-3})$ as $t \to 0$. If we write $\Theta = \Theta_0 + \Theta_1$, where Θ_0 is the sum of the terms in which $x_1 = 0$, then

$$\int_0^\infty \Theta_1(\tau, h; Z_t, W_t)\, t^{2s-1}\, dt$$

is an entire function of s whose value at $s=0$ is

$$2^{-\frac{1}{2}} \sum_{\substack{x_1 \equiv 0 (M) \\ x_2 \equiv h_2 (M) \\ x_1 x_2 > 0}} \operatorname{sgn} x_1 \; e(\tau M^{-1} x_1 x_2).$$

Also we find that, for $\operatorname{Re} s > 3/2$,

$$\int_0^\infty \Theta_0(\tau, h; Z_t, W_t)\, t^{2s-1}\, dt = \tfrac{1}{2}(\tfrac{1}{2}\pi v)^{-s} M^s \pi^{-\frac{1}{2}} \Gamma(s+\tfrac{1}{2}) \sum_{x_2 \equiv h_2 (M)} \operatorname{sgn} x_2 |x_2|^{-2s}.$$

The series on the right side of this equation may be written as

$$M^{-2s}\left[H\left(\frac{h_2}{M}, 2s\right) - H\left(1 - \frac{h_2}{M}, 2s\right)\right]$$

where $H(x, s) = \sum_{n=0}^{\infty} (x+n)^{-s}$ is as in Weil [11, p. 58]. Since $H(x, s)$ has a meromorphic continuation to the whole s-plane with only a simple pole with residue 1 at $s=1$, and since $H(x, 0) = \frac{1}{2} - \langle x \rangle$, our assertions follow.

Remark. The function $2^{\frac{1}{2}} I(\tau)$ is precisely the Eisenstein series of weight 1 constructed by Hecke in [4]. It would be of interest to find a construction of $\vartheta(\tau, h, L)$ for singular h, analogous to the one given above, when $n > 1$.

Appendix

We want to describe the growth of the series $\Theta(\tau, h; Z, W)$ on a fundamental domain for E in D.

Let $G = \mathrm{SO}(V)$, viewed as an algebraic group over $\mathbb{Q}$. For a suitable $\mathbb{Q}$-basis $e_1, \ldots, e_m$ for V we have

$$(\ ,\ ) \sim \begin{pmatrix} & & 1 \\ & Q_0 & \\ 1 & & \end{pmatrix}$$

with ${}^t Q_0 = Q_0 \in M_{n-1}(\mathbb{Q})$, and $Q_0 > 0$. Let P be the stabilizer in G of the line $\mathbb{Q} \cdot e_1$. Let

$$U = \left\{ u_b = \begin{pmatrix} 1 & b & * \\ & 1_{n-1} & * \\ & & 1 \end{pmatrix} \right\}$$

be the unipotent radical of P, and let

$$A=\left\{a_t=\begin{pmatrix} t & & \\ & 1_{n-1} & \\ & & t^{-1}\end{pmatrix}\right\}$$

be a maximal $\mathbb{Q}$-split torus in P. Let K be the stabilizer in G of $Z_0=1/\sqrt{2}\,{}^t(1,0,-1)\in D$.

For $t_0\in\mathbb{R}_+^\times$, let

$$A_{t_0}=\{a_t\in A_{\mathbb{R}}\mid t\geqq t_0\}.$$

Then for a suitable compact subset $\omega\subset U_{\mathbb{R}}$ and for a suitable $t_0\in\mathbb{R}_+^\times$ the Siegel set

$$\mathfrak{S}=\omega\cdot A_{t_0}\cdot K$$

has the property that

$$G(\mathbb{R})=E\cdot C\cdot\mathfrak{S}$$

for some finite set $C\subset G(\mathbb{Q})$ [2, Théorème 13.1].

Fix some orthogonal n-frame $W_0\subset T_{Z_0}(D)$, and for $g\in G(\mathbb{R})$ set

$$\Theta^*(g;h,L)=\det v^{\frac{1}{2}}\sum_{X\equiv h(L)}|\det(g^{-1}X,W_0)|\exp(-\pi\,\mathrm{tr}(v(g^{-1}X,g^{-1}X)_0))$$

where $(\ ,\ )_0=(\ ,\ )_{Z_0}$.

It will then be sufficient to estimate the function $\Theta^*(g_0g;h,L)$ for $g_0\in C$ and $g\in\mathfrak{S}$. Since $\Theta^*(g_0g;h,L)=\Theta^*(g;g_0^{-1}h,g_0^{-1}L)$, it will be sufficient to consider $\Theta^*(g;h,L)$ for $g\in\mathfrak{S}$ and arbitrary L and h. Now take $M\in\mathbb{Z}_{>0}$ such that $L_0\supset M\mathbb{Z}^m$ and hence

$$(M\mathbb{Z}^m)^*\supset L_0^*\supset L_0\supset M\mathbb{Z}^m.$$

Writing $M\mathbb{Z}^m=L_{00}$, we have

$$\Theta^*(g;h,L)=\sum_{h_i\equiv h(L_{00}^n)}\Theta^*(g;h_i,L_{00}^n).$$

Next observe that for $\sigma\in\mathrm{GL}_n(\mathbb{Z})$, if we replace v by ${}^t\sigma v\sigma$ in $\Theta^*(g;h,L_{00}^n)$ we have

$$\Theta^*({}^t\sigma v\sigma,g;h,L_{00}^n)=\Theta^*(v,g;h{}^t\sigma,L_{00}^n).$$

Therefore we may assume that v is reduced. There then exist constants C_1 and C_2 such that

$$C_1\cdot v_0\cdot 1_n\leqq v\leqq C_2\cdot v_0\cdot 1_n$$

where v_0 is the smallest of the diagonal entries of v.

Write $g=uak$ with $u\in\omega$, $a\in A_{t_0}$, and $k\in K$. Then

$$(g^{-1}X,g^{-1}X)_0=((ua)^{-1}X,(ua)^{-1}X)_0$$

and

$$|\det(g^{-1}X,W_0)|=|\det((ua)^{-1}X,W_0)|,$$

since $(g^{-1}X, W_0)=(g^{-1}X, W_0)_0$ and since, for $k\in K$, there exists $k'\in O(n)$ such that $kW_0=W_0k'$.

Now write $u^{-1}X=Y=\{Y_1, \ldots, Y_n\}$, and let

$$W_0=\left\{2^{-\frac{1}{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}, W'\right\}$$

where W' is an orthonormal frame in span$\{e_2, \ldots, e_n\}$. Then writing ${}^tY_k=(y_{1k}, {}^ty'_k, y_{mk})$, we have

$$\operatorname{tr}(a_t^{-1}Y, a_t^{-1}Y)_0=\sum_{k=1}^{n}(t^{-2}y_{1k}^2+Q_0[y'_k]+t^2y_{mk}^2)$$

and

$$\begin{aligned}\det(a_t^{-1}Y, W_0)&=2^{-\frac{1}{2}}\det\left(\left(Y, \begin{pmatrix}t\\0\\t^{-1}\end{pmatrix}\right), (Y, W')\right)\\&=2^{-\frac{1}{2}}\sum_{k=1}^{n}|t^{-1}y_{1k}+ty_{mk}|\,|\alpha_k|\end{aligned}$$

where α_k is the kth minor of the matrix (Y, W'). Note that α_k depends only on the y'_j's. Using these facts we obtain:

$$\begin{aligned}\Theta^*(g; h, L_{00}^n)=\det v^{\frac{1}{2}}2^{-\frac{1}{2}}\sum_{X\equiv h(L_{00}^n)}&\left(\sum_{k=1}^{n}|t^{-1}y_{1k}+ty_{mk}|\,|\alpha_k|\right)\\&\cdot\exp\left(-\pi C_1v_0\sum_{j=1}^{n}(a_t^{-1}Y_j, a_t^{-1}Y_j)_0\right).\end{aligned}$$

Thus it will be sufficient to estimate the two series:

$$I_k=\sum_{X\equiv h(L_{00}^n)}|y_{1k}|\,|\alpha_k|\exp(-\pi C'\operatorname{tr}(a_t^{-1}Y, a_t^{-1}Y)_0)$$

and

$$I'_k=\sum_{X\equiv h(L_{00}^n)}|y_{mk}|\,|\alpha_k|\exp(-\pi C'\operatorname{tr}(a_t^{-1}Y, a_t^{-1}Y)_0),$$

where $C'=C_1v_0$.

Observe that $y_{mk}=x_{mk}$, so that we may write

$$\begin{aligned}I_k=\sum_{x_m\equiv h_m(M\mathbb{Z}^n)}&\exp\left(-\pi C't^2\sum_{j=1}^{n}x_{mj}^2\right)\\&\cdot\sum_{\substack{x_1\equiv h_1\\x'\equiv h'}}|y_{1k}|\,|\alpha_k|\exp(-\pi C'\sum_j(t^{-2}y_{1j}^2+Q_0[y'_j])).\end{aligned}$$

If $h_m\not\equiv 0(M\mathbb{Z})$, we may factor an $\exp(-\frac{1}{2}\pi C't^2)$ out of the series provided we write $\exp(-\frac{1}{2}\pi C't^2\sum_j x_{mj}^2)$ in the first part of the summation. Since the inner sum has at most polynomial growth in t as $t\to\infty$, we see that I_k decays exponentially as $t\to\infty$, i.e. $I_k=O(t^Ne^{-C''t^2})$.

If $h_m \equiv 0(M\mathbb{Z}^n)$, then the same argument shows that the part of I_k in which $x_m \neq 0$ decays exponentially as $t \to \infty$. It remains to consider the sum, $I_{k,0}$, of the terms in which $x_m = 0$. Note that if $x_m = 0$, then $y'_j = x'_j$. But now

$$I_{k,0} = \sum_{x' \equiv h'} |\alpha_k| \exp(-\pi C' \sum_j Q_0[x'_j]) \cdot \sum_{x_1 \equiv h_1} |x_{1k} + b x'_k| \exp(-\pi C' t^{-2} \sum_j (x_{1j} + b x'_j)^2),$$

where $u = u_b$. It is easy to verify that the inner sum is $O(t^{(n-1)+4})$ as $t \to \infty$, uniformly in b since $u_b \in \omega$. Therefore, we obtain:

$$I_k = \begin{cases} O(t^N e^{-C'' t^2}) & \text{if } h_m \not\equiv 0(M\mathbb{Z}^n) \\ O(t^{n+3}) & \text{if } h_m \equiv 0(M\mathbb{Z}^n) \end{cases}$$

as $t \to \infty$.

Applying similar arguments to I'_k we obtain $I'_k = O(t^N e^{-C'' t^2})$ as $t \to \infty$, since the factor $|y_{mk}| = |x_{mk}|$ eliminates the terms with $x_m = 0$.

We have thus proved the following:

Proposition A.

$$\Theta^*(g; h, L^n_{00}) = \begin{cases} O(t^N e^{-C'' t^2}) & \text{if } h_m \not\equiv 0(M) \\ O(t^{n+2}) & \text{if } h_m \equiv 0(M), \end{cases}$$

as $t \to \infty$.

We are now in a position to complete the

Proof of Proposition 2.1. Suppose that $h \in L^*/L$ is non-singular. Then for any $g_0 \in G(\mathbb{Q})$ $g_0^{-1} h \in g_0^{-1} L^*/g_0^{-1} L$ is also non-singular. Also, if

$$L^*_{00} \supset L^*_0 \supset L_0 \supset L_{00},$$

as above, and if $h_i \in (L^*_{00})^n$ with $h_i \equiv h(L)$, then h_i is again non-singular. Finally, if $\sigma \in \mathrm{GL}_n(\mathbb{Z})$, and $h \in (L^*_{00})^n/L^n_{00}$ is non-singular, then so is $h'\sigma$. Thus, if $h \in L^*/L$ is non-singular, then so are all the residue classes $h_i \in (L^*_{00})^n/L^n_{00}$ which are involved in the estimates on $\Theta^*(g; h, L)$ given above. On the other hand, if $h \in (L^*_{00})^n/L^n_{00}$ and $h_m \equiv 0(M)$, then, choosing a representative $\tilde{h} \in (L^*_{00})^n$ with $\tilde{h}_m = 0$, we have

$$(\tilde{h}, \tilde{h},) = {}^t\tilde{h}' Q_0 \tilde{h}'$$

where ${}^t\tilde{h} = (\tilde{h}_1, {}^t\tilde{h}', 0)$; and hence $\mathrm{rank}(\tilde{h}, \tilde{h}) \leqq n-1$. Thus if $h \in L^*/L$ is non-singular, then $\Theta^*(g; h, L)$ is rapidly decreasing on the fundamental set $C \cdot \mathfrak{S}$, and Proposition 2.1 is proved.

Acknowledgements. I would like to thank John Millson for many crucial discussions and for his constant encouragement. I would also like to thank the mathematics department at Yale University for their hospitality during the fall term 1979.

References

1. Andrianov, A.N., Maloletkin, G.N.: Behavior of theta series of degree n under modular substitutions. Math. USSR Izv, **39**, 227–241 (1975)
2. Borel, A.: Introduction aux groupes arithmetiques. Paris: Hermann 1969

3. Gelbart, S.: Holomorphic discrete series for the real symplectic group. Invent. Math. **19**, 49–58 (1973)
4. Hecke, E.: Zur Theorie der elliptischen Modulfunktionen. Math. Ann. **97**, 210–242 (1926)
5. Howe, R.: θ-series and invariant theory. Proc. Symp. Pure Math. **33**, 275–285 (1979)
6. Kashiwara, M., Vergne, M.: On the Segal-Shale-Weil representations and harmonic polynomials. Invent. math. **44**, 1–47 (1978)
7. Kudla, S., Millson, J.: Geodesic cycles and the Weil representation, quotients of hyperbolic space and Siegel modular forms. Preprint (1979)
8. Millson, J., Raghunathan, M.S.: Geometric construction of cohomology for arithmetic groups. I. Papers Dedicated to the Memory of V.K. Patodi. Bangalore: Indian Academy of Sciences
9. Siegel, C.L.: Indefinite quadratische Formen und Funktionen Theorie.I. Math. Ann. **124**, 17–54 (1951)
10. Shintani, T.: On construction of holomorphic cusp forms of half-integral weight. Nagoya Math. J. **58**, 83–126 (1975)
11. Weil, A.: Elliptic functions according to Eisenstein and Kronecker. Ergebnisse der Mathematik, Vol. 88, Berlin, Heidelberg, New York: Springer 1976

Received February 15, 1980

Math. Ann. 256, 535–548 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Concentration Functions and a Class of Non-Compact Groups

Karl H. Hofmann[1]* and Arunava Mukherjea[2]

1 Fachbereich Mathematik AG 1, Technische Hochschule, D-6100 Darmstadt, Federal Republic of Germany
2 Department of Mathematics, University of South Florida, Tampa, FL 33620, USA

Concentration functions in R^n have been studied in various contexts by many mathematicians including Paul Lévy, Kolmogorov, Rogozin, LeCam, Eseen, Sozonov, and many others. Recently, these functions have been studied also in groups by Sozonov, Center-Mukherjea, Bougerol and others. The problem can be stated as follows:

Problem. Suppose S is a locally compact second countable non-compact Hausdorff topological group and P is a probability measure on the Borel sets of S such that the support S_P of P generates S. Then under what conditions on S or P, does the sequence

$$f_n(K) \equiv \sup\{P^n(Kx^{-1}) : x \in S\}$$

converge to zero as $n \to \infty$ for every compact $K \subset S$? Also, if it goes to zero, what is its rate of convergence?

Bougerol [1] has studied the rate of convergence of $f_n(K)$ to zero under assumptions on S as well as P; for example, one of his assumptions is when S is a compactly generated compact extension of an abelian group of rank d and P is an étalée aperiodic (aperiodic means that S_P is not contained in a coset of a closed proper normal subgroup) probability, and then he has shown that there is a positive number t such that

$$-\sup\{P^n(xKy) : x, y \in S\} \leqq t.\, n^{-(d/2)}.$$

Though the rate of convergence is a more interesting (and perhaps more applicable) problem, in this paper we will deal with the more basic (and certainly the more difficult) problem of finding when $f_n(K)$ converges to zero as $n \to \infty$ in the general case. The rate of convergence will be the subject matter of a future paper. In this paper we show that with the popular probabilistic assumption that S_P (the support of P) generates S as a closed semigroup (in which case every point in S is a possible point for the random walk induced by P), the sequence $f_n(K)$ converges to zero in a large class of groups containing all pro-Lie groups and maximally almost periodic (MAP) groups. As usual, a pro-Lie group means a

* Permanent address: Department of Mathematics, Tulane University, New Orleans, LA 70118, USA

0025-5831/81/0256/0535/$02.80

locally compact group of the form $\lim G_j$ for a projective system of Lie groups G_j and proper homomorphisms. (A morphism is called proper if its kernel is compact.) It follows from standard results on locally compact groups that all locally compact almost connected groups (i.e. groups G for which G/G_0 is compact where G_0 is the identity component of G) are pro-Lie groups (see [8]). Trivially, all Lie groups and so, in particular, all discrete groups are pro-Lie groups. It is also relevant to point out that in our results on concentration functions the second countability assumption can be removed for almost connected groups. Aside from the results on concentration functions, the results on the structure theory of locally compact groups in this paper are, we hope, of independent interest (besides being needed for the derivation of our main theorem).

In what follows, we make the assumption

$$S = \overline{\sum_{n=1}^{\infty} S_P^n}. \tag{1}$$

(The bar overhead will always denote closure.)

We now state and prove Theorem 1, which then will lead to our main theorem after some interesting (and necessary) structure theory results. The main theorem follows Corollary 13.

Theorem 1. *Let S be as described in the problem. Suppose that there is a compact subset K such that $f_n(K)$ does not converge to zero. Then there is a closed non-compact normal subgroup H_1 such that the quotient S/H_1 is compact (abelian) and monothetic, a compact nowhere dense subgroup $H \subset H_1$, and an element z with $zH = Hz$ such that for any open set V with $H \subset V$, we have*

(i) $H_1 \subseteqq \text{closure}\left(\bigcup_{n=1}^{\infty} z^{-n} V z^n\right)$ *and* (ii) $S = \overline{\bigcup_{n=1}^{\infty} H_1 z^n}$. □

Proof. Suppose that for some compact subset K, $f_n(K)$ does not converge to zero. Then by a result in [3], there exist elements a_n and x_n in S and probability measures Q_n such that

(2) for each k, weak-$\lim_{n\to\infty} P^{n-k} a_n = x_k^{-1} Q_k$ ($\equiv Q'_k$, say); (here, multiplication means convolution, and the unit mass at an element x is also represented by x).

(3) if w_H is the normed Haar measure of a compact subgroup H of S, then for each k we have: $Q_k w_H = Q_k$;

(4) there is a subsequence (k_i) of positive integers such that Q_{k_i} converge weakly to w_H as $i \to \infty$.

By Proposition 2 in [2], there exists an element z in S such that we have:

$$Q'_k = Q'_{k+1} z = P Q'_{k+1}. \tag{5}$$

Now from (5), we have

$$P Q'_k w_H = Q'_k z w_H$$

and from (2), $Q'_k w_H = Q'_k$. Therefore,

$$Q'_k z w_H = P Q'_k w_H = P Q'_k = Q'_k z$$

and this means that

$$Q_k z w_H z^{-1} = Q_k. \tag{6}$$

Similarly, we have:

$$Q_k z^{-1} w_H z = Q_k. \tag{6'}$$

Using (4), (6), and (6'), we can now write

$$w_H z w_H z^{-1} = w_H \quad \text{and} \quad w_H z^{-1} w_H z = w_H.$$

It follows immediately that

$$zH = Hz. \tag{7}$$

Let us now define the set H_1 by

$$H_1 = \text{closure}\left(\bigcup_{n=1}^{\infty} S_P^n S_P^{-n}\right). \tag{8}$$

Notice that $H_1 = H_1^{-1}$. Now we use our assumption (1) for the first time in this proof. Using this assumption, it can be easily verified that

$$\text{for every } x \text{ in } S, \quad x H_1 x^{-1} \subset H_1. \tag{9}$$

This means that $xH_1 = H_1 x$ for each x in S and therefore,

$$S_P^n S_P^{-n} H_1 = S_P^n H_1 S_P^{-n} \subset H_1.$$

It follows that $H_1 H_1 \subset H_1$, and thus, by previous assertions,

$$H_1 \text{ is a normal subgroup.} \tag{10}$$

Now from (5) and (1), we have for each n

$$P^n Q_k' = Q_k' z^n$$

and therefore,

$$S = S \cdot S_{Q_k'} = \left(\overline{\bigcup_{n=1}^{\infty} S_P^n}\right) \cdot S_{Q_k'} = \overline{\bigcup_{n=1}^{\infty} S_{Q_k'} z^n}.$$

Since $S_{Q_k'} = x_k^{-1} S_{Q_k}$, it is clear that for each positive integer k we have:

$$S = \text{closure}\left(\bigcup_{n=1}^{\infty} S_{Q_k} z^n\right). \tag{11}$$

Let us now denote with $\bar{\beta}$, for a given measure β, the measure given by $\bar{\beta}(B) = \beta(B^{-1})$. Now it follows from the second equality in (5) that for every positive integer n,

$$P^n (Q_k' \bar{Q}_k') \bar{P}^n = Q_k' \bar{Q}_k'$$

so that

$$S_P^n S_P^{-n} \subset \overline{S_{Q_k'} S_{Q_k'}^{-1}}.$$

From (2), it follows that

$$w\text{-}\lim_{n\to\infty} P^{n-k}\bar{P}^{n-k} = Q'_k\bar{Q}'_k$$

and thus, for each non-negative integer k,

$$H_1 = \text{closure}(S_{Q'_k}S^{-1}_{Q'_k}). \tag{12}$$

It follows from (10), (11), and (12) that

$$S = \text{closure}\left(\bigcup_{n=1}^{\infty} H_1 z^n\right). \tag{13}$$

Then it is clear that the quotient group S/H_1 is abelian. By Theorem 3 in [2], we know that $f_n(K)$ converges to zero for any compact set K in non-compact abelian groups and thus, if S/H_1 is non-compact, this property will hold in S/H_1 and therefore also in S (by considering the projection of P under the natural map from S onto S/H_1). Therefore, it is no loss of generality to assume that the quotient S/H_1 is compact. It is also monothetic by (13). The theorem will be proven if we now establish the following:

$$H \subset H_1\,; \tag{14}$$

(15) for any open set V with $V \supset H$,

$$H_1 \subset \text{closure}\left(\bigcup_{n=1}^{\infty} z^{-n}Vz^n\right);$$

(16) H is nowhere dense.

To prove (14), notice that using (3), we have:

$$\text{for each } x \text{ in } S_{Q_k},\quad xHx^{-1} \subset S_{Q_k}x^{-1}.$$

Thus, using the equality (12) we have:

$$xHx^{-1} = (xHx^{-1})(xHx^{-1})^{-1} \subset S_{Q_k}x^{-1}xS^{-1}_{Q_k} \subset H_1$$

so that

$$H \subset x^{-1}H_1x \subset H_1 \quad [\text{using (10)}].$$

Next, we establish (15). We will use the measure $P' \equiv (\bar{Q}'_0Q'_0)^2$. Note that the support of P' contains H_1. This is the reason we are considering P' rather than $\bar{Q}'_0Q'_0$. From (5), $Q'_0 = Q'_nz^n$ and therefore,

$$z^n(\bar{Q}'_0Q'_0)z^{-n} = \bar{Q}'_nQ'_n = \bar{Q}_nQ_n. \tag{17}$$

By (4) and (17), we have:

$$z^{k_i}P'z^{-k_i}(\equiv P'_i, \text{ say}) \to w_H \quad (\text{as } i \to \infty).$$

Using (3) and (7) and the fact that $S_{P'}$ contains the identity, it follows that

$$P'w_H = P' \quad \text{and} \quad H \subset S_{P'}.$$

We now claim that for every open set $V \supset H$,

$$S_{P'} \subset \text{closure}\left(\bigcup_{n=1}^{\infty} z^{-n} V z^n\right). \tag{18}$$

To prove this claim, let $x \in S_{P'}$. Suppose that $x \notin \bigcup_{n=1}^{\infty} z^{-n} V z^n$. Then there is an open set W containing x such that

$$V \cap \left(\bigcup_{n=1}^{\infty} z^n W z^{-n}\right) = \emptyset.$$

Notice that for every positive integer i we have:

$$P_i'\left(\bigcup_{n=1}^{\infty} z^n W z^{-n}\right) \geqq P'(W) > 0.$$

It follows that

$$\limsup P_n'(V) \leqq 1 - P'(W) < 1.$$

This is a contradiction since $P_n' \to w_H$, $H \subset V$ and V is open. This establishes (18). Since $H_1 \subset S_{P'}$ and H_1 is a normal subgroup, it follows from (18) that

$$H_1 \subseteqq \text{closure}\left(\bigcup_{n=1}^{\infty} z^{-n} V z^n\right).$$

This proves (15).

Finally to establish (16), assume that H is not nowhere dense. Then H is open. Using (7) and (15), it follows immediately that $H_1 = H$ implying that H_1 is compact, which possibility we already excluded. This contradiction verifies (16). The proof of the theorem is now complete. □

Now we make the following observations on Theorem 1.

(19) It is not clear whether the element z or one of its powers (In Theorem 1) belongs to H_1. This is relevant since nicer results are possible in case $z^m \in H_1$ for some m. In this context, we recall that z first appears in (5); in fact, any limit point of the sequence $(a_{n+1}^{-1} a_n)$, which is relatively compact, can be taken as z. (See Proposition 2 [2].) A natural question is whether the sequence (a_n) can be replaced by a suitable sequence (b_n) such that for some limit v of $(b_{n+1}^{-1} b_n)$, $v^m \in H_1$ for some m. The next remark answers this.

(20) Assume (2), (3), and (4) above. Suppose also that there is another sequence (b_n) such that $P^{n-k} b_n \to R_k'$ for each $k \geqq 0$. Then the sequence $(b_{n+1}^{-1} b_n)$ is relatively compact as before and for any limit point v of this sequence and any positive integer m, there is a limit point z of $(a_{n+1}^{-1} a_n)$ such that

$$v^m \in H_1 \quad \text{if and only if} \quad z^m \in H_1.$$

Proof of (20). Since $P^{n-k} b_n = P^{n-k} a_n (a_n^{-1} b_n)$, it is clear that the sequence $(a_n^{-1} b_n)$ is relatively compact. Let v be the limit of $b_{n_i+1}^{-1} b_{n_i}$ (as $i \to \infty$). Choose a subsequence $(p_i) \subset (n_i)$ such that for some z and u, as $i \to \infty$, we have:

$$a_{p_i}^{-1} a_{p_i+1} \to z^{-1} \quad \text{and} \quad a_{p_i+1}^{-1} b_{p_i+1} \to u.$$

Then, $z^{-1}uv=\lim a_{p_i}^{-1}b_{p_i}$. Using the first equation in this proof, we have:

$$Q'_k z^{-1}uv=R'_k=Q'_k u.$$

Using (4), $Hz^{-1}uv=Hu$ so that using (7), $uv^m u^{-1}\in Hz^m$. Since $H\subset H_1$ and H_1 is normal, (20) follows. □

(21) In Theorem 1, if we define the set H_2 by

$$H_2=\text{closure}\left(\bigcup_{n=1}^{\infty} S_P^{-n}S_P^n\right),$$

then it follows by an argument similar to the one given for H_1 that H_2 is also a normal subgroup. Also, $H_1\subset H_2$. The reason is the following: By (2), we have:

$$a_n^{-1}\bar{P}^n P^n a_n\to\bar{Q}'_0 Q'_0$$

and therefore,

$$S_{Q'_0}^{-1}S_{Q'_0}\subset\text{closure}\left(\bigcup_{n=1}^{\infty} a_n^{-1}S_P^{-n}S_P^n a_n\right)\subset H_2$$

since H_2 is a normal subgroup. This means that if $x\in S_{Q'_0}$, then

$$H_1\subset x^{-1}H_1 x\subset H_2 \quad [\text{using (12)}].$$

This means that in case H_1 or H_2 is compact and S still satisfying (1) and other properties as in Theorem 1, the sequence $f_n(K)$ must go to 0 as $n\to\infty$ for each compact $K\subset S$. This is immediate from (13) since in this case S/H_1 as well as S/H_2 is abelian. □

Let us pause for a moment and consider the situation of Theorem 1 in a diagram:

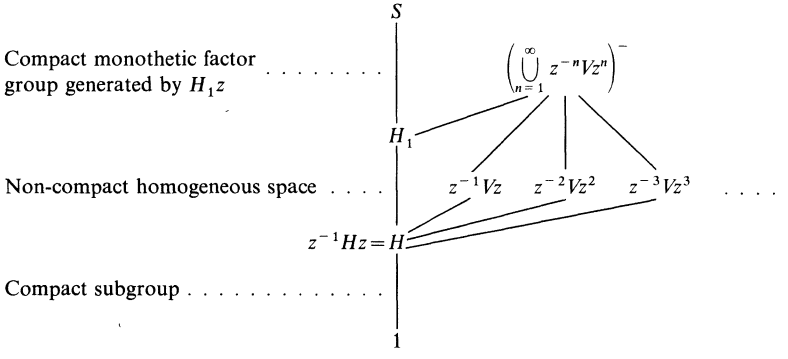


Fig. 1

We do not believe that such a situation is possible at all, and our efforts will now concentrate on presenting evidence for the conjecture that it is impossible. The following definition will allow us to speak more easily about conditions such as the one depicted in Fig. 1.

Definition 2. A locally compact group is called *strange* if it is not compact and contains

(i) a cocompact normal subgroup H_1 (i.e. a closed normal subgroup the quotient with respect to which is compact),

(ii) a compact subgroup H of H_1, and

(iii) an element z

such that for every neighborhood V of H, we have

$$(*) \qquad H_1 \subset \text{closure}\left(\bigcup_{n=1}^{\infty} z^n V z^{-n}\right).$$

A group which is not strange is called *neat.* □

It is clear that in a strange group, the cyclic group generated by z must be infinite discrete. Also, all compact groups are, by definition, neat. We now verify that discrete groups are also neat.

Remark 3. Any discrete group is neat.

Proof. Let G be a discrete group with a cofinite normal subgroup H_1 and a finite subgroup H such that $H_1 = \bigcup_{n=1}^{\infty} z^n H z^{-n}$. Since G/H_1 is finite, there is a natural number m such that $z^m \in H_1$. This means that $z^m = z^n h z^{-n}$ for some $h \in H$ and some integer n. Hence, $z^m \in H$. Then, $H_1 = \bigcup_{k=1}^{m} z^k H z^{-k}$. This implies that H_1 (and therefore, G) is finite. □

From Theorem 1 we derive the following (weaker) conclusion:

Corollary 4. *If S is a non-compact second countable locally compact group and P a probability measure (on the Borel sets of S) whose support S_P satisfies*

$$S = \text{closure}\left(\bigcup_{n=1}^{\infty} S_P^n\right),$$

then $f_n(K) \to 0$ as $n \to \infty$ for all compact subsets K of S unless S is strange. □

We remark that in Corollary 4, second countability can be replaced by almost connectedness. The reason is the following: By [8, p. 58], in any almost connected locally compact group S, there is a compact normal subgroup N such that S/N is second countable. Now the corollary holds for the pushforward measure P_0 on S/N, where $P_0(B) = P(p^{-1}(B))$ with $p(g) = gN$. It is easily verified that $(P^n)_0 = (P_0)^n$ and therefore, the corollary holds also for S since it so happens that S is strange iff S/N is strange. (This last fact is contained in our next lemma.)

Lemma 5. *Let G be a locally compact group and N a closed normal subgroup. If G is strange, then G/N is compact or strange. Conversely, if N is compact and G/N is strange, then G is strange.* □

Proof. First assume that G is strange. Let H_1, H and z be as in Definition 2 and let $p: G \to G/N$ be the quotient homomorphism. If V is any neighborhood of H, then (*) in Definition 2 implies

$$p(H_1)^- \subset \text{closure}\left(\bigcup_{n=1}^{\infty} p(z)^n p(V) p(z)^{-n}\right),$$

since p is a continuous morphism. Now $p(H_1)^-$ is a closed normal subgroup of G/N such that $(G/N)/p(H_1)^-$ is compact. (This follows from the definition of quotient topology and the fact that G/H_1 is compact.) Also, $p(H)$ is a compact subgroup of $p(H_1)^-$ and every neighborhood W of $p(H)$ in G/N is of the form $p(V)$ with the neighborhood $V=p^{-1}(W)$ of H. It follows that G/N is compact or strange. Next, let us assume that N is compact and G/N strange. Then there are K_1, a closed normal subgroup of G/N such that $(G/N)/K_1$ is compact, a compact subgroup $K\subset K_1$, and an element $z_0\in G/N$ such that they satisfy all the properties in the definition of strangeness for G/N. Let p be as above. Then $p^{-1}(K)$ is a compact subgroup of $p^{-1}(K_1)$, a closed normal subgroup of G. It is verified easily that $G/p^{-1}(K_1)$ is compact. Let $z\in p^{-1}(z_0)$. Suppose that there is an open set W containing $p^{-1}(K)$ such that

$$p^{-1}(K_1)\not\subset \text{closure}\left(\bigcup_{n=1}^{\infty} z^n W z^{-n}\right).$$

Since $Np^{-1}(K)N=p^{-1}(K)$ and N is compact, we may assume with no loss of generality that $NWN=W$. Then the set $A=\left(\bigcup_{n=1}^{\infty} z^n W z^{-n}\right)^-$ also satisfies $AN=A$, and thus, $p(A)$ does not contain the subgroup K_1. But notice that $p(W)$ is an open set containing K and

$$K_1\subset\left(\bigcup_{n=1}^{\infty} z_0^n p(W) z_0^{-n}\right)^- \subset p(A)$$

since $p(A)$ is closed in G/N. This contradicts the properties that K_1, K, and z_0 satisfy. This means that G must be strange. □

Lemma 6. *A compact extension of a finite dimensional real vector group is neat.* □

Proof. Let G be such an extension. Then there is a real vector group (finite dimensional) E, a compact group K and a continuous representation $\pi: K\to \text{Aut}(E)$ such that $G=E\times K$ (semidirect with respect to π). We assume that G is strange and our aim then is to derive a contradiction. We proceed through a number of steps.

Step I. If N is the kernel of π, then $\{0\}\times N$ is a compact normal subgroup of G. By Lemma 5, the factor group $G/(\{0\}\times N)$ is strange. We may therefore assume that $N=\{1\}$, i.e. that K operates faithfully on E. By Weyl's trick we may assume now that K is a subgroup of the orthogonal group $O(E)$ of E (w.r.t. a suitable inner product) and that K acts by application on E. Thus, G is a Lie group of euclidean motions of some real vector space.

Step II. Let $E_1\neq E$ be a K-invariant subvector space. Then $E_1\times\{1\}$ is a closed normal subgroup of G, and by Lemma 5 we know that $G/(E_1\times\{1\})$ is strange. Therefore, we will assume that E has no K-invariant subvector spaces other than E and $\{0\}$.

Step III. Let H_1, H and $z=(v,p)\in E\times K$ be as in Definition 2. Every compact subgroup in an almost connected Lie group is contained in a maximal compact subgroup, and all maximal compact subgroups are conjugate. We may, therefore,

assume with no loss of generality that $H \subset \{0\} \times K$. An arbitrary compact neighborhood V of $\{0\} \times K$ is given by

(1) $$V = B \times K, \quad B \text{ a compact ball around 0 in } E.$$

Since K is orthogonal, we have:

(2) $$KB = B.$$

Now setting $p_n = 1 + p + p^2 + \ldots + p^{n-1}$, we have:

(3) $$z^n = (p_n v, p^n).$$

Also, condition (*) of Definition 2 implies that

(4) $$H_1 \subset \left(\bigcup_{n=1}^{\infty} (p_n v, p^n)(B \times K)(-p^{-n} p_n v, p^{-n}) \right)^{-}.$$

The projection V_n of the nth summand in the union into E is now given by

(5) $$V_n = B + p_n(1-K)v \quad [\text{using (2)}].$$

The group $(\{0\} \times K)H_1$ is normalized by K. Thus, if F denotes the projection of H_1 into E, we have:

$$F \times \{1\} = (E \times \{1\}) \cap (\{0\} \times K)H_1,$$

and F is invariant under K, i.e.

(6) $$KF \subset F.$$

From (4), it follows that

(7) $$F \subset B + \left(\bigcup_{n=1}^{\infty} p_n(1-K)v \right)^{-}.$$

Step IV. The component F_0 of 0 in F is a K-invariant vector group. There are two possibilities:

(a) $F_0 = E$, hence $F = E$;

(b) $F_0 = 0$, hence F discrete. But $E/F = G/(\{0\} \times K)H_1$ is compact, hence F is the Z-span of a vector space basis of E. In that case, K acts faithfully on F, which implies that K is finite.

Step V. We now consider the case that $p \neq 1$. Then $1-p \neq 0$ and $(1-p)p_n \in 1-K$ and so using (5),

(8) $$(1-p)V_n = (1-p)B + (1-p)p_n(1-K)v \subset (1-p)B + (1-K)^2 v.$$

This last set is a compact subset C of E which does not depend on n. By (7), we now have:

(9) $$(1-p)F \subset C.$$

But this is a contradiction since $1-p \neq 0$ and $(1-p)F$ is unbounded.

Step VI. This is the last step. Here we complete the proof by showing that the case $p = 1$ also leads to a contradiction.

Let $p=1$. Then $p_n=n$ and thus, $V_n=B+n(1-K)v=B+(1-K)nv$. This means that for all closed balls around 0

$$F\subset B+\left(\bigcup_{n=1}^{\infty}(1-K)nv\right)^- \tag{10}$$

and therefore, also

$$F\subset B+\bigcup_{n=1}^{\infty}(1-K)nv. \tag{11}$$

We now claim that (11) is impossible.

Case (a). $F=E$. The center of gravity of Kv w.r.t. normalized Haar measure on K is 0, because otherwise this center of gravity would span a one-dimensional invariant subspace which would have to be E, and then K is the trivial or the two element group, and it is obvious that the center of gravity of Kv is then 0. Now we know that the center of gravity of $(1-K)v$ is v. Then v has a positive distance from $\cup(1-K)nv$ and for sufficiently small B, this contradicts (11).

Case (b). F is a free lattice subgroup of E which linearly spans E. Hence there is a number d such that any ball with radius d anywhere in E contains at least two points of F. Now let r be the distance from $-v$ to the nearest point of the finite set Kv which is different from $-v$. Then the distance from $2mv$ to the nearest point of $\bigcup_{n=1}^{\infty}(1-K)nv$, which is different from $2mv$, is mr. Choose m so large that $mr>d$. Then the ball around mv with radius d contains at most one point of F (namely v); but this is a contradiction to the choice of d. □

Lemma 7. *A finite extension of a semisimple Lie group is neat.* □

For the proof of this lemma, we need the next result for whose basic idea we are indebted to Joe Wolf (personal communication to K.H.H.).

Lemma 8. *Let $\pi: G\to \mathrm{Gl}(n, C)$ be a representation of a strange locally compact group and let H_1 be as in Definition 2. Then for each h in H_1 the spectrum of $\pi(h)$ is contained in the unit circle.* □

Proof. We denote with W the annulus $\{z\in C: \frac{1}{2}<z<2\}$ in C. Any multiplicative group of complex numbers which is contained in $\bar{W}$ is actually contained in the unit circle $U(1)$. We denote with U the open set in $\mathrm{Gl}(n, C)$ of all g with Spec $g\subset W$. Now let $V=\pi^{-1}(U)$. Since U is invariant under all inner automorphisms of $\mathrm{Gl}(n, C)$, then V is invariant under all inner automorphisms of G. Now let H be as in Definition 2 and $h\in H$. Then $\pi(h)$ lies in a compact subgroup of $\mathrm{Gl}(n, C)$, and thus all eigenvalues of $\pi(h)$ are of absolute value 1. Hence V is a neighborhood of H. Since $z^nVz^{-n}=V$, then the condition (*) of Definition 2 yields $H_1\subset\bar{V}$. Hence, $\pi(H_1)\subset\bar{U}$. If $g\in H_1$, then Spec $\pi(g^n)\subset\bar{W}$ for all integers n. Thus all eigenvalues of $\pi(g)$ must be in $U(1)$. □

Proof of Lemma 7. If G_0 is compact, then G is compact, hence neat. Since the maximal compact normal subgroup of G_0 is characteristic, in view of Lemma 5 it therefore suffices to prove the lemma for G_0 without compact factors. Suppose that G is strange and let H_1 be as in Definition 2. Hence H_1 is cocompact and G_0

has no compact factors, we have: $G_0 \subset H_1$. Now consider the adjoint representation $\mathrm{Ad}: G \to \mathrm{Aut}\, L(G)$ and the injection of $\mathrm{Aut}\, L(G)$ into $\mathrm{Gl}(C \otimes L(G))$. If we denote their composition by π, then Lemma 8 applies and it shows that $\pi(g)$ has its spectrum in the unit circle for all $g \in G_0$. But if $G_0 = KAN$ is an Iwasawa decomposition of G_0, then there is an element $a \in A$ which is semisimple (i.e. for which $1 \otimes \mathrm{Ad}\, a$ is diagonalizable in $C \otimes L(G)$ and for which $\mathrm{Spec}\, \mathrm{Ad}\, a$ is not contained in $U(1)$. This contradiction proves the lemma. □

We observe that the proof actually shows the conclusion of Lemma 7 for all discrete extensions of semisimple Lie groups.

Lemma 9. *Any strange locally compact group is compactly generated.* □

Proof. If H_1, H and z are as in Definition 2, then we take any compact neighborhood U of 1 and a finite set F such that $G = FUH_1$. This is possible since H_1 is cocompact. From (*) in Definition 2, we now have:

$$H_1 \subset \left(\bigcup_{n=1}^{\infty} z^n H U z^{-n} \right)^{-} \subset U \cdot \left(\bigcup_{n=1}^{\infty} z^n H U z^{-n} \right).$$

Thus, $\{z\} \cup H \cup U \cup F$ is a compact generating set for G. □

Now we are ready to bring our fragmentary information of neat groups together. In order to do this in a systematic fashion, we introduce the concept of a neat class of groups in the following:

Definition 10. A class $\mathcal{N}$ of locally compact groups is called *neat* provided it has the following properties:

(i) all discrete and compact groups are in $\mathcal{N}$;
(ii) all compact extensions of finite dimensional real vector groups are in $\mathcal{N}$;
(iii) all finite extensions of semisimple Lie groups are in $\mathcal{N}$;
(iv) if G is a locally compact group with a closed normal subgroup N such that G/N is not compact and G/N is in $\mathcal{N}$, then G is in $\mathcal{N}$;
(v) if a locally compact group G is not in $\mathcal{N}$, then G is compactly generated;
(vi) $\mathcal{N}$ is closed under passage to isomorphic groups. □

It is clear that the above definition is motivated by our results on neat groups. As a consequence of Lemmas 5, 6, 7, and 9, the class of neat groups is a neat class. Now we prove the following result:

Proposition 11. *Any neat class $\mathcal{N}$ contains all*

(i) *Lie groups;*
(ii) *pro-Lie groups;*
(iii) *MAP groups.*

In fact, it contains any group with a non-compact quotient which is one of the above three types.

Proof. The last assertion is immediate from Definition 10(iv). Every compactly generated *MAP*-group is pro-Lie (see [4]). Thus, it follows from Definition 10(v) that $\mathcal{N}$ contains *MAP* groups if it contains pro-Lie groups. Now every pro-Lie group is an extension of a compact group by a Lie group; therefore, it is sufficient to show that any Lie group G is in $\mathcal{N}$. We prove this in three steps.

Step I. If G/G_0 is infinite, then Definition 10(i) and (iv) show that $G \in \mathcal{N}$.

Step II. Suppose that G/G_0 is finite. Let R be the radical of G_0. If G_0/R is not compact, then Definition 10(iii) and (iv) show that $G \in \mathcal{N}$.

Step III. Now suppose that G_0/R is compact. Then G is a compact extension of R. To prove that G is in $\mathcal{N}$, we use induction on the "solvable length" of R. If this length is zero, then G is compact and therefore in $\mathcal{N}$. Now assume that this length is positive. Let R' be the closed commutator subgroup of R. If R/R' is compact, then G/R' is compact. But the solvable length of R' is smaller than that of R and therefore, by induction hypothesis $G \in \mathcal{N}$. If R/R' is not compact, then it contains a vector subgroup modulo which G/R' is compact. Hence by [6], G/R' is a compact extension of a normal vector subgroup. By Definition 10(ii) and (iv), G/R' (and therefore, G) is in $\mathcal{N}$. The induction is complete. The proposition follows. □

The next two results now follow immediately from Proposition 11.

Theorem 12. *Every locally compact group which has a non-compact Lie group quotient is neat. In particular, all pro-Lie groups are neat.*

Corollary 13. *Every locally compact group which has a non-compact MAP group quotient is neat.*

Our main theorem now follows immediately from Corollary 4, Theorem 12, and Corollary 13.

Main Theorem. *Let S be a second countable locally compact group with a non-compact quotient which is pro-Lie or MAP. Let P be a probability measure on the Borel sets of S such that its support S_P satisfies $S = \left(\bigcup_{n=1}^{\infty} S_P^n\right)^-$. Then for all compact subsets K of S, the sequence*

$$f_n(K) = \sup\{P^n(Kx) : x \in S\}$$

converges to zero as n tends to infinity.

We point out that our discussion of the structure of locally compact groups has not fully exhausted all of the information contained in Theorem 1 and its proof. Indeed recall that a probability measure P is called *aperiodic* if it is defined on a locally compact group and its support is not contained in the coset of a closed proper normal subgroup. It is clear from (8) and (10) in the proof of Theorem 1 that if P is aperiodic, then $S = \text{closure}\left(\bigcup_{n=1}^{\infty} S_P^n S_P^{-n}\right)$. As a result, we have the following corollary to Theorem 1:

Corollary 14. *If S is a non-compact second countable locally compact group with an aperiodic probability measure P on the Borel sets of S such that $S = \text{closure}\left(\bigcup_{n=1}^{\infty} S_P^n\right)$, then for all compact subsets K of S we have*

$$\lim_n \sup\{P^n(Kx) : x \in S\} = 0,$$

unless there is a compact nowhere dense subgroup H and an element z in its normalizer such that for every neighborhood V of H, the union $\bigcup_{n=1}^{\infty} z^{-n} V z^n$ is dense in G.

Corollary 14 motivates the following definition which reflects the spirit of Definition 2:

Definition 15. A locally compact group G is called *very strange* if it is not compact and contains a compact subgroup H and an element z in its normalizer such that for every neighborhood V of H the union

$$(**) \qquad \bigcup_{n=1}^{\infty} z^n V z^{-n} \quad \text{is dense in } G.$$

A group which is not very strange is called *rather neat.*

It is clear that a very strange group is strange and also that a neat group is rather neat. Since Lemma 5 is easily seen to hold also for very strange groups, the class of rather neat groups is a neat class. It then follows from Proposition 11 that all pro-Lie groups and all MAP groups are rather neat.

In one regard, our information about very strange groups is better than that about strange groups. Indeed we have the following

Proposition 16. *Let G be a locally compact group with a compact normal subgroup N such that G/N is MAP. (This condition is satisfied in particular if G/N is compact.) If G is very strange, then G/N is compact and N is very strange.*

We do not give the proof here. This means in particular that a cocompact normal subgroup N of a locally compact very strange group is very strange. No such information is available for strange groups.

The strategy for making further progress on the problem of concentration functions on locally compact groups might be the following:

We conjecture that for every locally compact group G which is not compact and every probability measure P [respectively, every aperiodic probability measure P] on Borel sets of G such that $G=\left(\bigcup_{n=1} S_P^n\right)^-$ we have $\lim_{n\to\infty} \sup_{x\in G} P^n(Kx)=0$ for all compact subsets K of G. This conjecture would be true if the following conjecture were true:

Conjecture 17 (resp., 17'). *Every locally compact group G is neat (see Definition 2) respectively, rather neat (see Definition 15).*

If Conjecture 17 fails, then from the existence of a strange (resp. very strange) G we can construct more particular counterexamples. This improves the chance either to find an actual counterexample, or else to rule out the existence of a counterexample and thus prove the conjecture.

A first simple step in this direction is the following observation. If a locally compact group fails to be a neat class $\mathcal{N}$, then G/G_0 fails to be in $\mathcal{N}$, too. For otherwise G/G_0 would have to be compact by Definition 10(iv), and then G would be pro-Lie and therefore in $\mathcal{N}$. Thus if counterexamples to Conjecture 17 exist, then there must exist totally disconnected ones.

However, if a counterexample to Conjecture 17' exists, then one can show that in fact a counterexample G to Conjecture 17' with the following properties must exist, too:

(1) G is zero dimensional without any open or compact normal subgroups.

(2) The subset of all periodic elements (i.e. elements which are contained in a compact subgroup) is dense, but not all elements are periodic.

(3) G is perfect, i.e. G agrees with its own closed commutator group (hence is in particular unimodular).

We omit the proofs of these remarks at this point. These observations indicate, that a counterexample to the main conjecture on concentration functions for aperiodic probability measures might just as well be searched for within the class of locally compact groups with the rather peculiar properties indicated in (1), (2), and (3).

Acknowledgements. We acknowledge several useful remarks of the referee that resulted in a shortening of the proof of Theorem 1. We are also grateful to Eberhard Siebert for his comments that resulted in a slight revision in the statement of Theorem 1 as well as in the proof of Lemma 6.

References

1. Bougerol, P.: Fonctions de concentration sur certains groupes localement compacts. Z. Wahrscheinlichkeitstheorie verw. Gebiete **45**, 135–157 (1978)
2. Center, B., Mukherjea, A.: More on limit theorems for iterates of probability measures on semigroups and groups. Z. Wahrscheinlichkeitstheorie verw. Gebiete **46**, 259–275 (1979)
3. Csiszár, I.: On infinite products of random elements and infinite convolutions of probability measures on locally compact groups. Z. Wahrscheinlichkeitstheorie verw. Gebiete **5**, 279–295 (1966)
4. Heyer, H.: Dualität lokalkompakter Gruppen. Lecture Notes in Mathematics, Vol. 150, pp. 372. Berlin, Heidelberg, New York: Springer 1970
5. Hochschild, G.: The structure of Lie groups. San Francisco: Holden-Day 1971
6. Hofmann, K.H., Mostert, P.S.: Splitting in topological groups. Mem. Am. Math. Soc. **43** (1963) (2nd printing 1972)
7. Montgomery, D., Zippin, L.: Topological transformation groups. New York: Interscience 1964

Received September 23, 1980; in revised form February 18, 1981

Math. Ann. 256, 549–560 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Distributions on Rational Function Fields

Steven Galovich[1] and Michael Rosen[2]

1 Carleton College, Department of Mathematics, Northfield, MN 55057, USA
2 Brown University, Department of Mathematics, Providence, RI 02912, USA

1. Introduction

The notion of a distribution on a projective system of finite abelian groups was first introduced by Barry Mazur. This idea has elegantly unified a number of themes arising in the classical and modern theory of algebraic number fields. Mazur used distributions to provide a new and fruitful approach to the construction of p-adic L-functions. The theory was developed further by Kubert and used extensively in the joint work of Kubert and Lang on modular units. Distributions occur implicitly in the work of Iwasawa and explicitly in the work of Sinnott in connection with the Stickelberger ideal and cyclotomic units.

All the work mentioned above deals with algebraic number fields. The projective system most often considered is $\left\{\frac{1}{m}\mathbf{Z}/\mathbf{Z}\right\}$ where if $n|m$, the map from $\frac{1}{m}\mathbf{Z}/\mathbf{Z}$ to $\frac{1}{n}\mathbf{Z}/\mathbf{Z}$ is multiplication by $\frac{m}{n}$. In two previous papers, [GR1, 2], the authors constructed a theory of units in "cyclotomic" function fields. Let $\mathbf{F}$ be a finite field and $R_T=\mathbf{F}[T]$, the polynomial ring over $\mathbf{F}$. In our work a natural distribution on the system $\left\{\frac{1}{M}R_T/R_T,\ M \text{ monic}\right\}$, played a key role. In this paper we initiate a general theory of distributions on this system. In Sect. 2 we present the relevant definitions and explain how to decompose distributions into χ-distributions where χ denotes a character of $\mathbf{F}^*$. In Sect. 3 we concentrate on ordinary distributions and, under some mild restrictions, construct a universal ordinary distribution. Finally, in Sect. 4, we introduce the notion of a polynomial distribution and provide a complete classification theorem for them.

Our expectation is that the ideas initiated here will play a significant role in the arithmetic theory of Galois extensions of $\mathbf{F}(T)$.

0025-5831/81/0256/0549/$02.40

2. Distributions

Consider the system of abelian groups $X = \left\{\frac{1}{M}R_T/R_T = R_T(M) | M \in R_T,\ M \text{ monic}\right\}$. If $N|M$, let $r_{M/N}: R_T(M) \to R_T(N)$ where $r_{M/N}(A/M) = A/N$ (by abuse of notation we let A/M stand for the coset of A/M modulo R_T). A *distribution* on this projective system is a family of functions

$$\{f_M : R_T(M) \to V | M \in R_T, \text{ monic}\},$$

V an abelian group, subject to the condition that whenever $N|M$

$$f_N(x) = \sum_{r_{M/N}(y) = x} f_M(y)$$

for all $x \in R_T(N)$. Sometimes we will consider such families $\{f_M\}$ which are not defined at the zero elements but satisfy the above relations for $x \neq 0$. These will be called *punctured distributions.*

If $\{f_M\}$ and $\{g_M\}$ are two distributions with image groups V and W respectively, a morphism is a homomorphism $h: V \to W$ such that $h \circ f_M = g_M$ for all monic M. With this definition it is easy to see that distributions on X form a category.

A distribution $\{f_M\}$ is called *real* if $f_M(ax) = f_M(x)$ for all $a \in \mathbf{F}^*$ and all monic M. Distributions of this type occur in [GR1, 2] where it is shown why the word "real" is appropriate.

It will be of importance to generalize the notion of real distribution. Suppose $|\mathbf{F}| = q$ where q is a power of a prime p and let ζ_{q-1} be a complex primitive $(q-1)^{st}$ root of unity. Assume that V is a $\mathbf{Z}[\zeta_{q-1}]$ module. If $\chi: \mathbf{F}^* \to \mathbf{C}^*$ is a complex character of $\mathbf{F}^*$, then we say $\{f_M\}$ is a *χ-distribution* if $f_M(ax) = \chi(a) f_M(x)$ for all $a \in \mathbf{F}^*$ and all monic M. Clearly, a distribution is a χ_0-distribution where χ_0 is the trivial character. Notice that if χ is not trivial, then any χ-distribution $\{f_M\}$ must satisfy $f_M(0/M) = 0$.

If $q-1$ is invertible on V and V is a $\mathbf{Z}[\zeta_{q-1}]$ module then every distribution is a direct sum of χ-distributions, $f_M = \sum f_M^\chi$, where $f_M^\chi(z) = \frac{1}{q-1}\sum \chi(a)^{-1} f_M(ax)$, the sum over all $a \in \mathbf{F}^*$. This decomposition is parallel to the classical decomposition of distributions on the system $\left\{\frac{1}{m}\mathbf{Z}/\mathbf{Z} | m \geqq 1\right\}$ into even and odd distributions.

Suppose that whenever $A/M = B/N$ in $\mathbf{F}(T) = k$ we have $f_M(A/M) = f_N(B/N)$. Such a distribution is called an *ordinary distribution.* In this case the family $\{f_M\}$ can be replaced by a single function $f: k/R_T \to V$ satisfying the relation

$$f(r) = \sum_{Mt = r} f(t)$$

for all monic M. Equivalently, this relation is given by

$$f(r) = \sum_{A \bmod M} f((r+A)/M).$$

We conclude this section by presenting some examples.

Let χ be a non-trivial character of $\mathbf{F}^*$. For $0 \neq A \in R_T$ define $\tilde{\chi}(A)$ to be χ evaluated on the leading coefficient of A. For $A = 0$ define $\tilde{\chi}(0) = 0$. For $A \neq 0$ set

$n(A)=q^{\deg(A)}$. Let V be the group of functions from $\mathbf{C}\to\mathbf{C}$. If $x\in R_T(M)$, there is a unique coset representative A/M such that $\deg(A)<\deg(M)$. We define $\phi_M^\chi(x)=\tilde{\chi}(A)n(A)^{-s}$.

Proposition 2.1. *$\{\phi_M^\chi\}$ is a distribution.*

Proof. Suppose $N|M$ and that either $A=0$ or $\deg(A)<\deg(N)$. We must show $\phi_N^\chi(A/N)=\sum_K \phi_M^\chi\left(\frac{A+KN}{M}\right)$ where the sum is over $K=0$ and those K such that $\deg(K)<\deg(M/N)$.

When $K=0$ we have $\phi^\chi(A/M)=\tilde{\chi}(A)n(A)^{-s}=\phi^\chi(A/N)$. When $K\neq 0$ we have $\deg(A+KN)=\deg(KN)=\deg(K)+\deg(N)$ so that $\phi^\chi\left(\frac{A+KN}{M}\right)=\tilde{\chi}(K)n(K)^{-s}n(N)^{-s}$. Thus we are reduced to showing

$$0=\sum_{K\neq 0}\tilde{\chi}(K)n(K)^{-s}.$$

Since every $K\neq 0$ is uniquely a constant times a monic, this sum equals

$$\sum_{K\text{ monic}}\left(\sum_{a\in\mathbf{F}^*}\chi(a)\right)n(K)^{-s}.$$

The inner sum is zero since χ is a non-trivial character.

We remark that by fixing $s\in\mathbf{C}$ the above distribution reduces to a distribution with values in $\mathbf{C}$. If $s\neq 0$, then these distributions are not ordinary. If $s=0$, then we do get an ordinary distribution and this distribution will play an important role in the next section.

We next present an example of a punctured, ordinary, real distribution. For $A/M\in k/R_T$, $A/M\neq 0$, and $\deg(A)<\deg(M)$, define

$$\phi_0(A/M)=(q-1)(\deg(M)-\deg(A)-1)-1.$$

Proposition 2.2. *$\phi_0: k/R_T-\{0\}\to Z$ is a punctured, ordinary, real distribution.*

Proof. This is Proposition 3.1 of [GR2]. We reproduce the proof here for the convenience of the reader.

With the notation of the last proposition we must show that for $A/N\neq 0$, $\phi_0(A/N)=\sum_K \phi_0\left(\frac{A+KN}{M}\right)$. Let $D=M/N$ and $d=\deg(D)$. Then, using the definition and some trivial manipulation the proof reduces to showing

$$-(q-1)\deg(D)=\sum_{K\neq 0}(q-1)(\deg(D)-\deg(K)-1)-1$$

where the sum is over non-zero polynomials K such that $\deg(K)<\deg(D)$. Since there are $q^n(q-1)$ polynomials of degree n, the right hand side is $(q-1)v$ where

$$\begin{aligned} v&=\sum_{n=0}^{d-1} q^n[(q-1)(d-n-1)-1]\\ &=\sum_{n=0}^{d-1}[(d-n-1)q^{n+1}-(d-n)q^n]\\ &=-d. \end{aligned}$$

This concludes the proof.

We remark that it is not possible to make ϕ_0 into a non-punctured distribution. Suppose, on the contrary, that ϕ_0 can be defined at 0 and remain a distribution. It follows that for all monic M,

$$\phi_0(0) = \sum_{A \bmod M} \phi_0(A/M).$$

It follows that

$$\sum_{\substack{A \bmod M \\ A \neq 0}} \phi_0(A/M) = 0.$$

However, we showed in the course of the proof of Proposition 2.2 that the sum on the left hand side of this last equation is $-(q-1)\deg(M)$. This contradiction shows the impossibility of extending ϕ_0 to a distribution on all of k/R_T.

3. A Universal Ordinary Distribution

Ordinary distributions form a subcategory of the category of all distributions (we always mean distributions on the system X). A *universal ordinary distribution* is by definition an initial object in this subcategory. The following characterization of universal ordinary distributions follows from a theorem of Kubert [K1, 2] (see also [L, Chap. 2] and [KL]). Kubert works on $\mathbf{Q}/\mathbf{Z}$ but his proof applies almost word for word to k/R_T.

Let $\Phi(M)$ denote the order of the group $(R_T/(M))^*$.

Theorem 3.1. *An ordinary distribution $f: k/R_T \to V$ is universal if*

i) *V is generated by $f(k/R_T)$.*

ii) *For each monic M the group generated by $f((1/M)R_T/R_T)$ is free of rank $\Phi(M)$.*

Notice the condition i) is easily achieved if we have a distribution satisfying ii). We simply replace V by the group generated by $f(k/R_T)$.

The construction of a universal ordinary distribution which we shall give depends heavily on our previous work on cyclotomic function fields. Thus, we begin by reviewing the relevant portions of that theory.

For $M \in R_T$ and u a variable, define $u^M = M(\phi + \mu_T)(u)$ where $\phi(u) = u^q$ and $\mu_T(u) = Tu$. In other words, one applies the operator $M(\phi + \mu_T)$ to u. Let $\bar{k}$ be the algebraic closure of k and set

$$\Lambda_M = \{\lambda \in \bar{k} | \lambda^M = 0\}.$$

Then $k(\Lambda_M)$, the subfield of $\bar{k}$ generated by Λ_M, is an abelian extension of k whose Galois group is isomorphic to $(R_T/(M))^*$ under the map $A \leftrightarrow \sigma_A$ where $\sigma_A(\lambda) = \lambda^A$ for all $\lambda \in \Lambda_M$. The analogy with cyclotomic number fields is clear. This construction is due to L. Carlitz. For details the reader is referred to [H1]. The construction has been considerably generalized by Drinfeld and independently by Hayes (see [D, G, H2]).

Let $k(\Lambda_M)^+$ be the subfield of $k(\Lambda_M)$ corresponding to the group $\{\sigma_a | a \in \mathbf{F}^*\}$. The prime at ∞ in k [i.e. $\operatorname{ord}_\infty A = -\deg(A)$] splits completely in $k(\Lambda_M)^+$ and every

prime of $k(\Lambda_M)^+$ above ∞ ramifies totally in $k(\Lambda_M)$. Thus $k(\Lambda_M)^+$ can be thought of as the maximal "real" subfield of $k(\Lambda_M)$. Let O_M and O_M^+ be the integral closures of R_T in $k(\Lambda_M)$ and $k(\Lambda_M)^+$ respectively. We note that by the Dirichlet unit theorem the unit groups in both rings have $\mathbf{Z}$-rank $\dfrac{\Phi(M)}{q-1}-1$.

We now define the cyclotomic units in O_M. Let P'_M be the subgroup of $k(\Lambda_M)^*$ generated by Λ_M and set $C_M = P'_M \cap O_M^*$. C_M is called the group of cyclotomic units of level M. In [GR2] we show that $[O_M^* : C_M]$ is finite. More precisely,

Theorem 3.2. *Let g be the number of distinct monic prime polynomials dividing M. Then $[O_M^* : C_M] = h_M(q-1)^a$ where h_M is the class number of O_M^+ and $a=0$ if $g=1$ and $a = 2^{g-2}+1-g$ if $g>1$.*

In what follows we will only need the fact that $[O_M^* : C_M]$ is finite.

Let χ be a character (complex valued) on $(R_T/(M))^*$. We say that χ is a real character if χ is trivial on $\mathbf{F}^*$.

Proposition 3.3. *Suppose χ is a primitive Dirichlet character on $(R_T/(M))^*$ which is not real. Then $\sum \chi(A) \neq 0$ where the sum is over all monic A of degree less than* $\deg(M)$.

Proof. Let $\bar{h}_M$ be the number of divisor classes of degree zero of $k(\Lambda_M)$. By Theorem 2 of [GR1], $\bar{h}_M$ is a product of terms one of which is $\sum \chi(A)$. Thus $\sum \chi(A) \neq 0$. (In [GR1], M is assumed to be a prime power. However, the proof goes over to the general case with no difficulty.)

The sum which appears in the above proposition is simply the Dirichlet L-series of χ evaluated at 0. The fact that it does not vanish corresponds in the number field setting to the non-vanishing of the generalized Bernoulli number $B_{1,\chi}$ where χ is a primitive *odd* character on $(\mathbf{Z}/m\mathbf{Z})^*$.

We now describe a method of producing new distributions from old via Stickelberger elements. This process is due to Kubert and Lang [KL] (see also [L, Chap. 2, Sect. 8]).

Let G_M be the Galois group of $k(\Lambda_M)/k$, and F be an algebraically closed field of characteristic zero. If $N|M$ we define a map from $F[G_N]$ to $F[G_M]$ by taking $x \in G_N$ to $\sum_{\pi(y)=x} y$ where π is the natural projection from G_M to G_N. Let $\Omega = \varinjlim F[\mathbf{G}_M]$. Since $F[G_M]$ injects into Ω we will think of $F(G_M]$ as a subspace of Ω.

Suppose $f : k/R_T \to F$ is an ordinary distribution. We define a new ordinary distribution, called the *Stickelberger distribution associated to* f, $\mathrm{St}(f) : k/R_T \to \Omega$ as follows: for $A/M \in k$ with M monic and $\deg(A) < \deg(M)$

$$\mathrm{St}(f)(A/M) = \sum_B f(AB/M)\sigma_B^{-1}$$

where the sum is over $B \in (R_T/(M))^*$. It is straightforward to verify that $\mathrm{St}(f)$ is a distribution.

In the last section we produced an ordinary distribution corresponding to each complex character of $\mathbf{F}^*$. We recall the definitions. Suppose $A/M \in k$, M monic, and $\deg(A) < \deg(M)$. If $\chi : \mathbf{F}^* \to \mathbf{C}^*$ is non-trivial then $\phi^\chi(A/M) = 0$ if $A=0$ and is equal

to χ of the leading coefficient of A if $A \neq 0$. Corresponding to the trivial character we define for $A \neq 0$

$$\phi_0(A/M) = (q-1)(\deg(M) - \deg(A) - 1) - 1 .$$

As we saw at the end of Sect. 2, this distribution is "essentially" punctured. This fact will cause some technical complication later.

Proposition 3.4. *Let $\chi : \mathbf{F}^* \to \mathbf{C}^*$ be a non-trivial character. Then, the subgroup of Ω generated by $\mathrm{St}(\phi^\chi)(R_T(M))$ is free of rank $\Phi(M)/(q-1)$.*

Proof. We use [L, Chap. 2, Theorem 8.2] which applies equally well in our function field setting. Let Ψ be a Dirichlet character on $(R_T/M)^*$ and suppose the conductor of Ψ is $N|M$. Define

$$S(\Psi, \phi^\chi) = \sum_B \phi^\chi(B/N)\overline{\Psi(B)}$$

where the sum is over $B \in (R_T/(N))^*$. Theorem 8.2 of [L, Chap. 2] asserts that the F-dimension of the vector space generated by $\mathrm{St}(\phi^\chi)(R_T(M))$ is equal to the number of Dirichlet characters Ψ such that $S(\Psi, \phi^\chi) \neq 0$.

Since we have

$$S(\Psi, \phi^\chi) = \sum_{B \text{monic}} \left(\sum_{a \in \mathbf{F}^*} \chi(a)\overline{\Psi(a)} \right) \overline{\Psi(B)}$$

it follows that $S(\Psi, \phi^\chi) = 0$ unless $\Psi = \chi$ on $\mathbf{F}^*$. If Ψ restricts to χ on $\mathbf{F}^*$, then

$$S(\Psi, \phi^\chi) = (q-1) \sum_{B \text{monic}} \overline{\Psi(B)}$$

which is not zero by Proposition 3.3. Since there are precisely $\Phi(M)/(q-1)$ characters on $(R_T/(M))^*$ which restrict to χ the result follows.

Our plan is to take the sum of the distributions $\mathrm{St}(\phi^\chi)$ and $\mathrm{St}(\phi_0)$ to produce a universal distribution. To accommodate the fact that ϕ_0 is punctured, we are forced to modify the definition of $\mathrm{St}(\phi_0)$. In $F[G_M]$ define $s(G_M)$ to be $\sum_{x \in G_M} x$. If $N|M$ the map from $F[G_N]$ to $F[G_M]$ takes $s(G_N)$ to $s(G_M)$. Define $\Omega' = \varinjlim F[G_M]/(s(G_M))$. There is a natural homomorphism from Ω to Ω' and from now on we assume our Stickelberger distributions have values in Ω'. It is easily seen that Proposition 3.4 remains valid. The reason for introducing Ω' will be apparent from the next lemma.

Lemma 3.5. $\mathrm{St}(\phi_0)$ *extends to an ordinary distribution on all of k/R_T with value zero at zero.*

Proof. We recall that in the course of proving Proposition 2.2 we showed

$$\sum_{\substack{A \bmod M \\ A \neq 0}} \phi_0(A/M) = -(q-1)\deg(M) .$$

It follows that

$$\sum_{\substack{A \bmod M \\ A \neq 0}} \mathrm{St}(\phi_0)(A/M) = -(q-1)\deg(M)s(G_M) = 0$$

(the last equality because we are working in Ω'). It follows that we can define $\mathrm{St}(\phi_0)$ arbitrarily at 0 and get a distribution on k/R_T. We choose this value to be zero.

Proposition 3.6. *The subgroup of Ω' generated by $\mathrm{St}(\phi_0)(R_T(M))$ is free of rank $\Phi(M)/(q-1)-1$.*

Proof. We will need the theory of cyclotomic units in $k(\Lambda_M)$.

Let P be a prime of $k(\Lambda_M)$ lying over the prime ∞ in k. For $B\in(R_T/(M))^*$ let $P_B=\sigma_B^{-1}P$. The set $\{P_B\}$ consists of all the primes at infinity (with repetition). We define $l: k(\Lambda_M)^*\to \mathbf{Z}[G_M]$ by

$$l(\alpha)=\sum \mathrm{ord}_{P_B}(\alpha)\sigma_B^{-1}.$$

Recall that P'_M is the subgroup of $k(\Lambda_M)^*$ generated by the elements of Λ_M. Set $T_M=l(P'_M)$ and E_M equal to the subgroup of Ω' generated by $\mathrm{St}(\phi_0)(R_T(M))$. We claim that $T_M/s(G_M)T_M=E_M$. To see this we need some facts from [GR2]. To begin with there is a $\lambda_M\in\Lambda_M$ such that

$$\mathrm{ord}_P\lambda_M=(q-1)(d-1)-1 \quad \text{where} \quad d=\deg(M).$$

For this choice of λ_M we have $\mathrm{ord}_P(\lambda_M^A)=\phi_0(A/M)$ for all $A\neq 0$ such that $\deg(A)<d$. (See [GR2, Lemma 1.6 and Proposition 1.10]). Since every element of Λ_M has the form λ_M^A and since

$$\begin{aligned} l(\lambda_M^A)&=\sum \mathrm{ord}_P(\lambda_M^A)\sigma_B^{-1}\\ &=\sum \mathrm{ord}_P(\lambda_M^{AB})\sigma_B^{-1}\\ &=\sum \phi_0(AB/M)\sigma_B^{-1}\end{aligned}$$

it follows that $T_M/s(G_M)T_M=E_M$, as asserted.

By Lemma 4.5 of [GR2] we know $[T_{M,0}: l(C_M)]$ is finite where $T_{M,0}$ is the submodule of T_M annihilated by $s(G_M)$. It follows that $\mathrm{rank}\,E_M=\mathrm{rank}\,T_M-1$ $=\mathrm{rank}\,T_{M,0}=\mathrm{rank}\,l(C_M)$. The kernel of l on O_M^* is easily seen to be $\mathbf{F}^*$. Since by Theorem 3.2 $[O_M^*: C_M]$ is finite, it follows that $\mathrm{rank}\,l(C_M)=\mathrm{rank}\,l(O_M^*)$ $=\Phi(M)/(q-1)-1$. This completes the proof.

If χ is the trivial character of $\mathbf{F}^*$ define $\phi^\chi=\phi_0$, and $S: k/R_T\to\Omega'$ by $S=\sum \mathrm{St}(\phi^\chi)$, the sum being over all characters of $\mathbf{F}^*$.

Theorem 3.7. *Let $V\subset\Omega'$ be the subgroup generated by $S(k/R_T)$. Then $S: k/R_T\to V$ is universal ordinary distribution on k/R_T subject to the condition $S(0)=0$.*

Proof. As it is easily seen, the subgroups generated by $\mathrm{St}(\phi^\chi)(R_T(M))$ are independent. It follows from Propositions 3.4 and 3.6 together with the assumption that $S(0)=0$ that the rank of the group generated by $S(T_T(M))$ is $\Phi(M)-1$. The result now follows from Theorem 3.1.

We remark that it is easy to supplement S in such a way that we obtain a universal ordinary distribution with no restriction on the value at 0. Since this is a purely formal process, we omit the details.

4. Polynomial Distributions

To avoid constant repetition we will assume in this section that when we write A/M, A and M are in R_T, M is monic, A is non-zero, and $\deg(A)<\deg(M)$.

Since the distribution ϕ_0 plays such an important role, it is natural to ask if ϕ_0 occurs in some family of distributions. As we shall see, the answer is yes. Notice that $\phi_0(A/M)$ is linear in $\deg(A)$ and $\deg(M)$. In fact any punctured distribution which is linear in $\deg(A)$ and $\deg(M)$ is a multiple of $\phi_0(A/M)$. (Proof: Exercise, or see the proof of Lemma 4.2 below.) This leads to the question of classifying distributions which are polynomial functions of $\deg(A)$ and $\deg(M)$. As it happens, we are able to give a satisfactory answer to this question.

We work with the projective system

$$X=\{R_T(M)|M\in R_T,\ M \text{ monic}\}.$$

Let F be a field of characteristic zero. A punctured distribution $\{f_M : R_T(M)\to F\}$ is called a *polynomial distribution* if there is a polynomial $g(x,y)\in F[x,y]$ such that $f_M(A/M)=g(\deg(A),\deg(M))$. To exhibit examples of such distributions, we first study distributions given by partial zeta functions. The power series expansions of such functions about zero will lead to distributions of the desired type.

Recall that for $A\in R_T$, $A\neq 0$, we have defined $n(A)=q^d$ where $d=\deg(A)$. For $A/M\in R_T(M)$ define the partial zeta distribution $\zeta_M(s, A/M)$ by

$$A/M\to\zeta_M(a, A/M)=\sum_{B\in R_T} n(A+BM)^{-s}.$$

Here s is a complex variable. The sum converges for $\mathrm{Re}(s)>1$.

Proposition 4.1. *The rule $A/M\to\zeta_M(s, A/M)$ defines a punctured distribution on X. If $\deg(A)=k<d=\deg(M)$, then*

$$\zeta_M(s, A/M)=q^{-ks}+q^{-ds}(q-1)/(1-q^{1-s}).$$

Proof. Suppose $M|N$. We must verify

$$\zeta_M(s, A/M)=\sum_K \zeta_M\left(s, \frac{A+KM}{N}\right)$$

where K ranges over all polynomials of degree less than $\deg(N)-\deg(M)$ and 0. This is a straightforward calculation from the definition.

To prove the second assertion note $\deg(A+BM)=\deg(BM)$ for $B\neq 0$. Thus

$$\zeta_M(s, A/M)=n(A)^{-s}+n(M)^{-s}\sum_{\substack{B\in R_T\\ B\neq 0}} n(B)^{-s}.$$

The sum on the right hand side is easily calculated to be $(q-1)/(1-q^{1-s})$ and the result follows.

The formula in the proposition provides an analytic continuation of $\zeta_M(s, A/M)$ to the whole complex plane with a single pole at $s=1$. We may expand this function in a power series about $s=0$. Setting $Q=\ln q$ the resulting series has the form $\sum_{n=0}^{\infty}\frac{(-Q)^n}{n!}g_n(k,d)s^n$ where $g_n(x,y)$ is a polynomial with rational coefficients. This fact can be derived from Taylor's Theorem. Since $A/M\to\zeta_M(s, A/M)$ is a distribution, it follows that $A/M\to g_n(\deg(A),\deg(M))$ is a distribution with rational values. We will investigate the form of $g_n(x,y)$ more closely.

Before doing the general case let's consider $g_0(x,y)$ and $g_1(x,y)$. Substituting $s=0$ in $q^{-ks}+q^{-ds}(q-1)/(1-q^{1-s})$, we find $g_0(x,y)$ is identically zero. To find $g_1(x,y)$ we differentiate with respect to s and then substitute $s=0$. The result is $g_1(x,y)=x-y+\frac{q}{q-1}$. Therefore the distribution $A/M\to g_1(\deg(A),\deg(M))$ is $(-1/(q-1))\phi_0(A/M)$. Thus our old friend ϕ_0 is essentially the linear term in the power series expansion about zero of $\zeta_M(s,A/M)$.

To arrive at a formula for $g_n(x,y)$ we proceed with Eulerian abandon.

$$\zeta_M(s,A/M)=e^{-Qks}+(q-1)e^{-Qds}(1-qe^{-Qs})^{-1}$$

$$=e^{-Qks}+(q-1)\sum_{i=0}^{\infty}q^i e^{-Q(d+i)s}.$$

Thus the coefficient of s^n in this expansion is $g_n(k,d)(-Q)^n/n!$ where $g_n(k,d)=k^n+(q-1)\sum_{i=0}^{\infty}q^i(d+i)^n$. Of course, the sum in the last expression is divergent if we work in $\mathbf{C}$. However, if we work instead in the p-adic number field $\mathbf{Q}_p$, then the sum converges since q is a power of p. So, working in $\mathbf{Q}_p$, define $\beta_j\in\mathbf{Q}_p$ by $\beta_j=\sum_{i=0}^{\infty}q^i i^j$. We then have

$$g_n(x,y)=x^n+(q-1)\sum_{j=0}^{n}\binom{n}{j}\beta_j y^{n-j}. \tag{1}$$

This formula shows $g_n(x,y)$ is a polynomial in $\mathbf{Q}_p[x,y]$ of total degree n. Since $\beta_0=\sum_{i=0}^{\infty}q^i=(1-q)^{-1}$ we see $g_n(s,y)=x^n-y^n+$terms of lower degree.

Of course, we know *a priori* that the coefficients are in fact in $\mathbf{Q}$. To make this evident in the present context and to calculate the β_j, we now show how to obtain a recursion relation for β_j.

Since $(i+1)^m=\sum_{j=0}^{m}\binom{m}{j}i^j$ we have

$$q^{i+1}(i+1)^m=q\sum_{j=0}^{m}\binom{m}{j}q^i i^j.$$

Summing over i we find $\beta_m=q\sum_{j=0}^{m}\binom{m}{j}\beta_j$. Hence, for $m\geqq 1$ we have

$$\beta_m=q(1-q)^{-1}\sum_{j=0}^{m-1}\binom{m}{j}\beta_j. \tag{2}$$

So, for example, $\beta_0=(1-q)^{-1}$, $\beta_1=q(1-q)^{-2}$, and $\beta_2=q(1-1)^{-2}+2q^2(1-q)^{-3}$. It is now clear that in general

$$(1-q)^{m+1}\beta_m\in\mathbf{Z}.$$

For readers who are less than comfortable with this somewhat cavalier derivation, we refer to Ribenboim's book [R, p. 142]. Ribenboim states a lemma of Euler himself which precisely yields our formulas for the numbers β_j, hence for the

polynomials g_n. (In Ribenboim's rendition of Euler's lemma, let $t=-q$ and let $v=-Qs$.)

To summarize, we have produced a sequence of polynomial distributions corresponding to polynomials $g_n(x,y)\in\mathbf{Q}[x,y]$ with $g_n(x,y)=x^n-y^n+$lower degree terms. We next show that by taking linear combinations of these polynomials, we produce *all* polynomial distributions.

Lemma 4.2. *Let $\{f_M\}$ be a punctured distribution with values in a field F. Suppose there is a polynomial $h(u)\in F[u]$ such that $f_M(A/M)=h(\deg(M)-\deg(A))$. Then for some $a\in F$ we have $\{f_M\}$ is the same as $a\phi_0$.*

Proof. We first note that $\{f_M\}$ is in fact an ordinary distribution. Thus we replace $\{f_M\}$ by a single function f. For $k>0$, $k\in\mathbf{Z}$, we have

$$f(1/T^k)=\sum_{\alpha\in F} f\left(\frac{1+\alpha T^k}{T^{k+1}}\right).$$

It follows that

$$h(k)=h(k+1)+(q-1)h(1).$$

Summing from $k=1$ to $k=m-1$ we find

$$h(1)=h(m)+(q-1)(m-1)h(1)$$

and so

$$h(m)=-h(1)((q-1)(m-1)-1)$$

which proves the result.

Lemma 4.3. *Suppose F is a field of characteristic zero and $g(x,y)\in F[x,y]$ gives rise to a polynomial distribution. Then*

$$Dg=\frac{\partial g}{\partial x}+\frac{\partial g}{\partial y}$$

also gives rise to a polynomial distribution.

Proof. Suppose $M|N$. From the usual distribution relation

$$f_M(A/M)=\sum f_N\left(\frac{A+KM}{N}\right)$$

with $k=\deg(A)<\deg(M)=d$ and $K=0$ or $\deg(K)<\deg(N)-\deg(M)=e$ we find

$$g(k,d)=g(k,d+e)+\sum_{r=0}^{e-1}(q-1)q^r g(d+r,d+e).$$

Since this holds for all pairs of positive integers k and d with $k<d$,

$$g(x,y)=g(x,y+e)+\sum_{r=0}^{e-1}(q-1)q^r g(y+r,y+e).$$

Let $g_1=\dfrac{\partial g}{\partial x}$ and $g_2=\dfrac{\partial g}{\partial y}$. Then

$$g_1(x,y)=g_1(x,y+e)$$

and

$$g_2(x,y)=g_2(x,y+e)+\sum_{r=0}^{e-1}(q-1)q^r(g_1(y+r,y+e)+g_2(y+r,y+e)).$$

Adding these two equations shows that $Dg=g_1+g_2$ gives rise to a distribution as asserted.

Lemma 4.4. *Let $\{g_n(x,y)|n=1,2,3,\ldots\}$ be the sequence of polynomial distributions arising from the partial zeta functions. (Note that g_1 and g_2 are not partial derivatives as in the last lemma.) Then*

$$Dg_n=ng_{n-1}.$$

Proof. This follows easily from the Eqs. (1) and (2) derived earlier for $g_n(x,y)$ and the numbers β_m.

Theorem 4.5. *Let F be a field of characteristic zero and suppose $g(x,y)\in F[x,y]$ gives rise to a polynomial distribution on $\{R_T(M)\}$ with values in F. Then there are elements $\alpha_1,\alpha_2,\ldots,\alpha_n\in F$ such that*

$$g=\alpha_1g_1+\alpha_2g_2+\ldots,\alpha_ng_n.$$

Proof. We argue by induction on the total degree of j.

If g has degree 1, then

$$g(x,y)=\alpha x+\beta y+\gamma.$$

By Lemma 4.3, $Dg(x,y)=\alpha+\beta$ is also a distribution. This can happen only if $\alpha+\beta=0$. Thus

$$g(x,y)=\alpha(x-y)+\gamma.$$

The result now follows in this case from Lemma 4.2.

Suppose g has degree n and that the result has been proven for smaller degrees. By Lemma 4.3, Dg defines a polynomial distribution of degree at most $n-1$. By induction there are elements $\beta_1,\beta_2,\ldots,\beta_{n-1}\in F$ such that

$$Dg=\sum_{i=1}^{n-1}\beta_ig_i.$$

Let $h=\sum_{i=2}^{n}\alpha_ig_i$ where $\alpha_i=\beta_{i-1}/(i-1)$ for $2\leqq i\leqq n$. Then

$$D(g-h)=Dg-Dh=0.$$

We have used Lemma 4.4. By elementary calculus it follows that $g-h$ is a function of $x-y$. Thus by Lemma 4.2 there is an $\alpha_1\in F$ such that $g-h=\alpha_1g_1$ and so

$$g=\sum_{i=1}^{n}\alpha_ig_i.$$

The proof is complete.

The polynomial distributions $g_n(k, d)$ suggest a number of tantalizing questions and problems. We describe three areas of future study which seem especially important.

(1) As we remarked in Sect. 2, the distribution $g_1(k, d)$ arises in our work on units and class groups in cyclotomic function fields. For $n \geqq 2$ what is the arithmetic significance of the distributions $g_n(k, d)$? It is natural to suspect that $g_n(k, d)$ is related to units in "class fields" corresponding to elliptic modules of rank n. The possibility of such a connection clearly warrants investigation.

(2) Let P be a fixed monic prime in R_T and let $R_P = \varprojlim (R_T/(P^n))$ be the completion of R_T at P. The distribution $g_n(k, d)$ defines a measure on R_P with values in $\mathbf{Z}_p$, the ring of p-adic integers. Develop the integration theory on R_P with respect to the measures $g_n(k, d)$. In a similar vein, develop integration theory on R_P with respect to the measures $\tilde{\chi}(A)n(A)^{-s}$.

(3) Investigate combinatorial properties of the numbers β_n and the polynomials $\sum_0^n \binom{n}{j} \beta_j y^{n-j} = h_n(y)$. For instance, if

$$f(x) = \sum_0^\infty \frac{\beta_n}{n!} x^n$$

is the exponential generating function of the β_n, then $f(x) = \dfrac{1}{1 - qe^x}$, an expression highly reminiscent of the generating function for the Bernoulli numbers.

Acknowledgement. This work was completed while the authors were on sabbatical leave at the University of California at Berkeley. Both authors are grateful to the Vaughn Foundation Fund for its support

References

[D] Drinfel'd, V.G.: Elliptic modules. Math. USSR Sb. **23**, 561–592 (1974)
[GR1] Galovich, S., Rosen, M.: The class number of cyclotomic function fields. J. Number Theory (to appear)
[GR2] Galovich, S., Rosen, M.: Units and class groups in cyclotomic function fields. J. Number Theory (to appear)
[G] Goss, D.: The algebraist's upper half-plane. Bulletin Amer. Math. Soc. (New Series) **2**, 391–415 (1980)
[H1] Hayes, D.: Explicit class field theory for rational function fields. Trans. Amer. Math. Soc. **189**, 77–91 (1974)
[H2] Hayes, D.: Explicit class field theory in global function fields. Studies in Algebra and Number Theory, G.-C., Rota (ed.), New York: Academic Press 1979
[K1] Kubert, D.: The universal ordinary distribution. Bull. Soc. Math. France **107**, 179–202 (1979)
[K2] Kubert, D.: The $\mathbf{Z}/2\mathbf{Z}$-cohomology of the universal ordinary distribution. Bull. Soc. Math. France **107**, 203–224 (1979)
[KL] Kubert, D., Lang, S.: Modular units. New York: Springer 1980
[L] Lang, S.: Cyclotomic fields. New York: Springer 1978
[R] Ribenboim, P.: 13 Lectures on Fermat's Last Theorem. New York: Springer 1979
[S] Sinnott, W.: On the Stickelberger ideal and the circular units of a cyclotomic field. Ann. of Math. **108**, 107–134 (1978)

Received January 26, 1981; in revised form March 31, 1981

Math. Ann. 256, 561–568 (1981)

Mathematische Annalen
© Springer-Verlag 1981

The Failure of the Topological Frobenius Property for Nilpotent Lie Groups. II

Ole A. Nielsen

Department of Mathematics and Statistics, Queens University, Jeffery Hall, Kingston K7L 3N6, Canada

This paper is a continuation of the author's earlier paper [5] by the same title and the reader is expected to have some familiarity with that paper. Both papers grew out of a desire to understand the topological (or weak) Frobenius reciprocity property – usual called (FP) – for nilpotent Lie groups. (The reader should consult [3] and [5] for the definition of and the known results about (FP).) Felix showed that (FP) does not hold for nilpotent Lie groups [2], and both [5] and the present paper are concerned with the extent to which it fails to hold. More precisely, these two papers are concerned with the following phenomenon: given a connected and simply-connected non-abelian nilpotent Lie group G, there will be closed connected subgroups H of G and irreducible unitary representations ϱ and π of H and G, resp., such that

$$\left.\begin{array}{l}\varrho \text{ is not weakly contained in } \pi|_H \\ \text{but } \pi \text{ is weakly contained in } \operatorname{ind}_H^G \varrho\,.\end{array}\right\} \tag{A}$$

For such a G the author has shown [5] that corresponding to any closed connected non-normal subgroup H of $\hat{G}$ there is a π in G such that (A) holds with ϱ the trivial representation of H and, on the other hand, Felix, Henrichs, and Skudlarek have shown [3] that there is a closed connected one-dimensional subgroup H of G and a ϱ in $\hat{H}$ such that (A) holds with π the trivial representation of G. (At the end of [5] it was remarked that this latter result is actually true for any closed connected non-normal one-dimensional subgroup H of G.) The point of the present paper is to present the following extensions of these results.

Theorem 1. *Let G be a connected and simply-connected nilpotent Lie group and let H be a closed connected one-dimensional non-normal subgroup of G. Then given any ϱ in $\hat{H}$, there will be a π in $\hat{G}$ such that* (A) *holds.*

Theorem 2. *Let G be as in Theorem 1 and let H be a closed connected two-dimensional non-normal subgroup of G. Then given any ϱ in $\hat{H}$ whose kernel contains $H \cap Z$, where Z is the center of G, there will be a π in $\hat{G}$ such that* (A) *holds.*

0025-5831/81/0256/0561/$01.60

Theorem 3. *Let G be as in Theorem 1 and let H a closed connected non-normal subgroup of G. Then there is a ϱ in $\hat{H}$ such that* (A) *holds with π the trivial representation of G.*

One can use Kirillov theory to reduce the proofs of these theorems to certain problems in the geometry of the co-adjoint orbits of nilpotent Lie algebras (this was described in [5] and is due to Kirillov, Moscovici, and Brown). Let H be a closed connected subgroup of a connected and simply-connected nilpotent Lie group G and let $\mathfrak{h}$ denote the subalgebra of the Lie algebra $\mathfrak{g}$ of G corresponding to H. As in [5], let $P(\mathfrak{g};\mathfrak{h})$ denote the subset of $\mathfrak{g}'\times\mathfrak{h}'$ consisting of those pairs (g,h) for which h is the restriction to $\mathfrak{h}$ of some member of the co-adjoint orbit through g, for an element g of $\mathfrak{g}'$ let $Q(\mathfrak{g},g;\mathfrak{h})$ denote the closure in $\mathfrak{h}'$ of $\{h\in\mathfrak{h}':(g,h)\in P(\mathfrak{g};\mathfrak{h})\}$, and for an element h of $\mathfrak{h}'$ let $Q(\mathfrak{g};\mathfrak{h},h)$ denote the closure in $\mathfrak{g}'$ of $\{g\in\mathfrak{g}':(g,h)\in P(\mathfrak{g};\mathfrak{h})\}$. In this notation the reduction just alluded to takes the following form: if π [resp., ϱ] is the element of $\hat{G}$ [resp., $\hat{H}$] corresponding (under the Kirillov correspondence) to a linear form g on $\mathfrak{g}$ [resp., h on $\mathfrak{h}$] then (A) is equivalent to

$$h\notin Q(\mathfrak{g},g;\mathfrak{h}) \quad \text{and} \quad g\in Q(\mathfrak{g};\mathfrak{h},h). \tag{B}$$

So to prove Theorems 1–3 it will be sufficient to prove the following propositions; in each of these propositions $\mathfrak{g}$ denotes a nilpotent Lie algebra and $\mathfrak{h}$ a subalgebra of $\mathfrak{g}$ which is not an ideal in $\mathfrak{g}$.

Proposition 1. *If $\mathfrak{h}$ is one-dimensional then given any linear form h on $\mathfrak{h}$, there will be a linear form g on $\mathfrak{g}$ such that* (B) *holds.*

Proposition 2. *If $\mathfrak{h}$ is two-dimensional then given any linear form h on $\mathfrak{h}$ satisfying $\mathfrak{h}\cap\mathfrak{z}\subset\ker h$, where $\mathfrak{z}$ is the center of $\mathfrak{g}$, there will be a linear form g on $\mathfrak{g}$ such that* (B) *holds.*

Proposition 3. *There is a linear form h on $\mathfrak{h}$ such that* (B) *holds with $g=0$.*

It should be pointed out that some five-dimensional examples will show that Theorem 2 [resp., Proposition 2] becomes false if one deletes from the hypothesis either the condition on the kernel of ϱ [resp., h] or else the condition on the dimension of H [resp., $\mathfrak{h}$]. The remainder of this paper will be devoted to the proofs of these three propositions. But before actually beginning these proofs it will be convenient to record a simple observation which will be used in several inductive arguments and whose proof is essentially contained in [5].

Lemma. *Let $\mathfrak{g}$ be a nilpotent Lie algebra, let $\mathfrak{h}$ be a subalgebra of $\mathfrak{g}$, let $\mathfrak{a}$ be a subspace of the center of $\mathfrak{g}$, and let θ denote the canonical mapping of $\mathfrak{g}$ onto $\mathfrak{g}/\mathfrak{a}$. If two linear forms $\tilde{g}$ and $\tilde{h}$ on $\mathfrak{g}/\mathfrak{a}$ and $\theta(\mathfrak{h})$, resp., satisfy*

$$\tilde{h}\notin Q(\mathfrak{g}/\mathfrak{a},\tilde{g};\theta(\mathfrak{h})) \quad \textit{and} \quad \tilde{g}\in Q(\mathfrak{g}/\mathfrak{a};\theta(\mathfrak{h}),\tilde{h})$$

then the two linear forms $g=\tilde{g}\circ\theta$ and $h=(\tilde{h}\circ\theta)|_{\mathfrak{h}}$ on $\mathfrak{g}$ and $\mathfrak{h}$, resp., satisfy

$$h\notin Q(\mathfrak{g},g;\mathfrak{h}) \quad \textit{and} \quad g\in Q(\mathfrak{g};\mathfrak{h},h).$$

Proof of Proposition 1

Let $\mathfrak{z}$ denote the center of $\mathfrak{g}$, and suppose that $\dim \mathfrak{z}=1$ and that $[\mathfrak{g},\mathfrak{h}]\subset\mathfrak{h}+\mathfrak{z}$. Then there will be a Jordan-Hölder basis $X_1,\ldots,X_n$ for $\mathfrak{g}$, where $n=\dim\mathfrak{g}$, with $X_1\in\mathfrak{z}$ and $X_2\in\mathfrak{h}$ and there will be an element T of $\mathfrak{g}$ satisfying $[T,X_2]=X_1$. Let $f_1,\ldots,f_n$ be the basis for $\mathfrak{g}'$ which is dual to the basis $X_1,\ldots,X_n$, let h be a given element of $\mathfrak{h}'$, and put $\alpha=\langle X_2,h\rangle$. Then

$$\left\langle \exp\left(\frac{1}{\varepsilon}adT\right)X_2,\varepsilon f_1+(\alpha-1)f_2\right\rangle=\alpha,$$

and thus $(\varepsilon f_1+(\alpha-1)f_2,h)\in P(\mathfrak{g};\mathfrak{h})$, for any $\varepsilon>0$ and

$$\langle \exp(adX)X_2,(\alpha-1)f_2\rangle=\alpha-1$$

for any $X\in\mathfrak{g}$, and therefore $(\alpha-1)f_2\in Q(\mathfrak{g};\mathfrak{h},h)$ and $h\notin Q(\mathfrak{g},(\alpha-1)f_2;\mathfrak{h})$.

The proof of the proposition is now easy to give and will be by induction on the dimension of $\mathfrak{g}$. The smallest possible value for $\dim\mathfrak{g}$ is 3 and if $\dim\mathfrak{g}=3$ then $\mathfrak{g}$ must be the Heisenberg algebra [1] and the preceding paragraph applies. Now suppose that $\dim\mathfrak{g}\geqq 4$. If there is a one-dimensional subspace $\mathfrak{a}$ of $\mathfrak{z}$ satisfying $[\mathfrak{g},\mathfrak{h}]\not\subset\mathfrak{h}+\mathfrak{a}$ then the canonical image of $\mathfrak{h}$ in $\mathfrak{g}/\mathfrak{a}$ is not an ideal and one can apply the inductive hypothesis and the lemma to obtain the desired conclusion. And if there is no such $\mathfrak{a}$ then it is easy to see that $\dim\mathfrak{z}=1$ and that $[\mathfrak{g},\mathfrak{h}]\subset\mathfrak{h}+\mathfrak{z}$, and hence that one has the situation considered in the preceeding paragraph.

Proof of Proposition 2

Suppose that $\mathfrak{h}\cap\mathfrak{z}\neq\{0\}$, where $\mathfrak{z}$, as usual, denotes the center of $\mathfrak{g}$, and consider an element h of $\mathfrak{h}'$ which vanishes on $\mathfrak{h}\cap\mathfrak{z}$. Then there is a linear form $\tilde{h}$ on $\theta(\mathfrak{h})$ satisfying $h=\tilde{h}\circ\theta$, where θ is the canonical mapping of $\mathfrak{g}$ onto $\mathfrak{g}/(\mathfrak{h}\cap\mathfrak{z})$, and $\dim\theta(\mathfrak{h})=1$, and so one can apply Proposition 1 to $\mathfrak{g}/(\mathfrak{h}\cap\mathfrak{z})$, $\theta(\mathfrak{h})$, and $\tilde{h}$ and then use the lemma to conclude that a g in $\mathfrak{g}'$ with the desired properties must exist.

In view of what has just been proven one may as well assume that $\mathfrak{h}\cap\mathfrak{z}=\{0\}$ throughout the remainder of the proof. The proof of the proposition under this assumptions will be by induction on the dimension of $\mathfrak{g}$. It is easy to give this inductive argument if one assumes that the statement of the proposition is true whenever $\mathfrak{g}$ and $\mathfrak{h}$ satisfy the following condition:

$$\left.\begin{array}{l}\text{for every one-dimensional subspace } \mathfrak{a} \text{ of } \mathfrak{z}\\ \text{there is a non-zero element } X \text{ of } \mathfrak{h} \text{ with } [\mathfrak{g},X]\subset\mathfrak{a}.\end{array}\right\}\qquad (*)$$

Indeed, as $\mathfrak{h}$ cannot be of co-dimension one in $\mathfrak{g}$ the smallest possible value for $\dim\mathfrak{g}$ is 4 and if $\dim\mathfrak{g}=4$ then (in the notation of [1]) either $\mathfrak{g}=\mathfrak{g}_4$ or $\mathfrak{g}=\mathfrak{g}_3\times\mathfrak{g}_1$; this means that there is a basis $X_1,\ldots,X_4$ for $\mathfrak{g}$ such that the Lie algebra structure on $\mathfrak{g}$ is generated by the brackets $[X_4,X_3]=X_2$ and $[X_4,X_2]=X_1$ if $\mathfrak{g}=\mathfrak{g}_4$ or by the bracket $[X_4,X_3]=X_2$ if $\mathfrak{g}=\mathfrak{g}_3\times\mathfrak{g}_1$. If $Y_1=a_1X_1+\ldots+a_4X_4$ and $Y_2=b_1X_1+\ldots+b_4X_4$ are two elements of $\mathfrak{g}$ which span $\mathfrak{h}$ then, as $\mathfrak{h}$ must be abelian, one must have $a_4b_3=a_3b_4$. But this just means that some non-trivial linear combination of Y_1 and Y_2 is of the form $c_1X_1+c_2X_2$, and hence that $\mathfrak{g}=\mathfrak{g}_4$ and $c_2\neq 0$

(as $\mathfrak{h}\cap\mathfrak{z}=\{0\}$). It is now clear that $\mathfrak{g}$ and $\mathfrak{h}$ satisfy $(*)$ and consequently the conclusion of the proposition. Now suppose that $\dim\mathfrak{g}\geqq 5$ and that $(*)$ does not hold. Then there will be a one-dimensional subspace $\mathfrak{a}$ of $\mathfrak{z}$ such that $[\mathfrak{g},X]\not\subset\mathfrak{a}$ for all non-zero elements X of $\mathfrak{h}$, and hence such that the image of $\mathfrak{h}$ under the canonical mapping θ of $\mathfrak{g}$ onto $\mathfrak{g}/\mathfrak{a}$ has trivial intersection with the center of $\mathfrak{g}/\mathfrak{a}$ (and so is certainly not an ideal in $\mathfrak{g}/\mathfrak{a}$). If h is a given element of $\mathfrak{h}'$ then there will be an element $\tilde{h}$ of $\theta(\mathfrak{h})'$ with $h=\tilde{h}\circ\theta$, and then one can use the lemma and the inductive hypothesis to obtain an element g of $\mathfrak{g}'$ such that g and h satisfy (B).

Now suppose that $\mathfrak{g}$ and $\mathfrak{h}$ satisfy $(*)$ and that $\dim\mathfrak{z}\geqq 2$. If X_1 and X_2 are two linearly independent elements of $\mathfrak{z}$ then there will be non-zero elements X_3 and X_4 of $\mathfrak{h}$ satisfying $[\mathfrak{g},X_3]\subset\mathbf{R}X_1$ and $[\mathfrak{g},X_4]\subset\mathbf{R}X_2$. Now X_3 and X_4 must be linearly independent (as $\mathfrak{h}\cap\mathfrak{z}=\{0\}$), so $\mathfrak{h}=\mathbf{R}X_3+\mathbf{R}X_4$ and $[\mathfrak{g},\mathfrak{h}]\subset\mathbf{R}X_1+\mathbf{R}X_2$, and thus $\mathfrak{z}=\mathbf{R}X_1+\mathbf{R}X_2$. Choose an element X of $\mathfrak{g}$ with $[X,X_3]=X_1$ and write $[X,X_4]=aX_2$ for some real number a. One may assume that $a\neq 0$. Indeed, if $a=0$ and if Y is an element of $\mathfrak{g}$ with $[Y,X_4]=X_2$ then $[Y,X_3]=bX_1$ for some real number b and

$$[(1-b)X+Y,X_3]=X_1$$

and

$$[(1-b)X+Y,X_4]=X_2.$$

It is clear that one may extend $X_1,\ldots,X_4$ to a Jordan-Hölder basis, say $X_1,\ldots,X_n$, for $\mathfrak{g}$, where $n=\dim\mathfrak{g}$. Let $f_1,\ldots,f_n$ denote the corresponding dual basis for $\mathfrak{g}'$, let h be a given element of $\mathfrak{h}'$, and put $\alpha=\langle X_3,h\rangle$, $\beta=\langle X_4,h\rangle$, and

$$g_k=\frac{1}{k}f_1+\frac{1}{ak}f_2+(\alpha-1)f_3+(\beta-1)f_4$$

for $k=1,2,\ldots$. Then

$$\langle\exp(kadX)X_3,g_k\rangle=\alpha$$

and

$$\langle\exp(kadX)X_4,g_k\rangle=\beta,$$

i.e., $(g_k,h)\in P(\mathfrak{g};\mathfrak{h})$, for all k. This shows that

$$g=(\alpha-1)f_3+(\beta-1)f_4\in Q(\mathfrak{g};\mathfrak{h},h),$$

and it is easy to see that $h\notin Q(\mathfrak{g},g;\mathfrak{h})$.

At this point only the case in which $\mathfrak{g}$ and $\mathfrak{h}$ satisfy $(*)$ and $\dim\mathfrak{z}=1$ remains to be disposed of. Now if this is the case then it is a consequence of Kirillov's lemma [4; Lemma 4.1] that there is a Jordan-Hölder basis $X_1,\ldots,X_n$ for $\mathfrak{g}$, where $n=\dim\mathfrak{g}$, with the following properties: $\mathbf{R}X_1=\mathfrak{z}$, $[X_n,X_2]=X_1$, $X_2\in\mathfrak{h}$, and $\ker(adX_2)=\mathbf{R}X_1+\ldots+\mathbf{R}X_{n-1}$. Now as $\mathfrak{h}\subset\ker(adX_2)$ one may evidently assume that $\mathfrak{h}=\mathbf{R}X_2+\mathbf{R}X_p$ for some integer p with $3\leqq p\leqq n-1$ and (since one may replace X_p by X_2+X_p if necessary) that $[X_n,X_p]\neq 0$. Let $f_1,\ldots,f_n$ denote the basis for $\mathfrak{g}'$ which is dual to $X_1,\ldots,X_n$, let h be a given element of $\mathfrak{h}'$, and put $\alpha=\langle X_2,h\rangle$ and $\beta=\langle X_p,h\rangle$. If one puts

$$\xi_{j,i}=\langle adX_n)^jX_p,f_i\rangle$$

for $1 \leqq j \leqq p-1$ and $1 \leqq i \leqq p-j$ then

$$\exp(t\,ad X_n) X_2 = X_2 + tX_1 \tag{1}$$

and

$$\begin{aligned} \exp(t\,ad X_n) X_p &= X_p + \sum_{j=1}^{p-1} \frac{t^j}{j!} (ad X_n)^j X_p \\ &= X_p + \sum_{j=1}^{p-1} \sum_{i=1}^{p-j} \frac{t^j}{j!} \xi_{j,i} X_i \\ &= X_p + \sum_{i=1}^{p-1} \left[\sum_{j=1}^{p-i} \frac{t^j}{j!} \xi_{j,i} \right] X_i \end{aligned} \tag{2}$$

for any real number t.

Suppose (for the moment) that $[X_n, X_p] \in \mathbf{R}X_1 + \mathbf{R}X_2$. Then $\xi_{1,2} = \xi_{2,1}$, and if one puts $\xi = \xi_{1,1}$ and $\eta = \xi_{1,2}$ then

$$\exp(t\,ad X_n) X_p = X_p + (t\xi + \tfrac{1}{2}t^2\eta) X_1 + t\eta X_2$$

for any $t \in \mathbf{R}$. Let k be a positive integer and let λ and ν be two real numbers, and consider the linear form

$$g = \frac{1}{k} f_1 + \lambda f_2 + \nu f_p$$

on $\mathfrak{g}$. This linear form satisfies

$$\langle \exp(k(\alpha - \lambda) ad X_n) X_2, g \rangle = \alpha$$

and

$$\langle \exp(k(\alpha - \lambda) ad X_n) X_p, g \rangle = -\tfrac{1}{2} k\eta(\alpha - \lambda)^2 + (\xi + k\alpha\eta)(\alpha - \lambda) + \nu .$$

Now if $\eta = 0$ and if one takes $\lambda = \alpha - 1$ and $\nu = \beta - \xi$ then

$$\langle \exp(k(\alpha - \lambda) ad X_n) X_p, g \rangle = \beta .$$

This shows that

$$\left(\frac{1}{k} f_1 + (\alpha - 1) f_2 + (\beta - \xi) f_p, h \right) \in P(\mathfrak{g}; \mathfrak{h})$$

for $k = 1, 2, \ldots$, and hence that

$$g = (\alpha - 1) f_2 + (\beta - \xi) f_p \in Q(\mathfrak{g}; \mathfrak{h}, h) .$$

At the same time, $h \notin Q(\mathfrak{g}, g; \mathfrak{h})$ since

$$\langle \exp(ad T) X_2, g \rangle = \alpha - 1$$

for all $T \in \mathfrak{g}$. On the other hand, if $\eta \neq 0$ and if one chooses

$$\lambda = \begin{cases} -\alpha - 2\xi/k\eta & \alpha \neq 0 \\ -[\xi + (\xi^2 + 2k\eta\theta)^{1/2}]/k\eta & \alpha = 0 \end{cases}$$

and

$$v=\begin{cases}\beta & \alpha\neq 0\\ \beta-\theta & \alpha=0,\end{cases}$$

where θ is the real number with $|\theta|=1$ and $\eta\theta>0$, then a simple calculation will show that

$$\langle\exp(k(\alpha-\lambda)adX_n)X_p, g\rangle=\beta.$$

This means that

$$\left(\frac{1}{k}f_1+\lambda f_2+vf_p, h\right)\in P(\mathfrak{g};\mathfrak{h})$$

for $k=1,2,\ldots$, and hence that

$$g=-\alpha f_2+vf_p\in Q(\mathfrak{g};\mathfrak{h},h).$$

Another simple calculation will show that for any T in $\mathfrak{g}$ one has

$$\langle\exp(adT)X_2, g\rangle=-\alpha$$

if $\alpha\neq 0$ and

$$\langle\exp(adT)X_p, g\rangle=\beta-\theta$$

if $\alpha=0$, and hence $h\notin Q(\mathfrak{g},g;\mathfrak{h})$.

Now suppose that $[X_n,X_p]\notin\mathbf{R}X_1+\mathbf{R}X_2$. Then $p>3$ and there is a positive integer r such that

$$(adX_n)^rX_p\notin\mathbf{R}X_1+\mathbf{R}X_2$$

and

$$(adX_n)^{r+1}X_p\in\mathbf{R}X_1+\mathbf{R}X_2, \tag{3}$$

and then, in turn, there is an integer i with $3\leqq i\leqq p-r$ and $\xi_{r,i}\neq 0$. Notice that

$$\xi_{r+2,2}=\xi_{r+3,1}=\xi_{r+3,2}=0 \quad\text{and}\quad \xi_{r+1,2}=\xi_{r+2,1}. \tag{4}$$

Let k be a positive integer and let λ, μ, and v be three real numbers, and consider the linear form

$$g=\frac{1}{k}f_1+\lambda f_2+\mu f_i+vf_p$$

on $\mathfrak{g}$. If one defines two polynomials φ and ψ by

$$\varphi(u,v,w)=\xi_{1,1}(w-v)+\sum_{j=1}^{p-2}\frac{1}{j!}u^j(w-v)^j\left[\frac{1}{j+1}\xi_{j+1,1}(w-v)-\xi_{j,2}v\right]$$

and

$$\psi(u,v)=\sum_{j=1}^{p-i}\frac{1}{j!}\xi_{j,i}u^jv^j$$

then an easy calculation [using (1) and (2)] will show that

$$\langle \exp(k(\alpha-\lambda)adX_n)X_2, g\rangle = \alpha \tag{5}$$

and

$$\langle \exp(k(\alpha-\lambda)adX_n)X_p, g\rangle = v+\varphi(k,\lambda,\alpha)+\mu\psi(k,\alpha-\lambda). \tag{6}$$

It follows from (3) and (4) that $\psi(\cdot, v)$ is of degree r whenever $v\neq 0$ and that $\varphi(\cdot, v, w)$ is of degree at most $r+1$ for any v and w. More precisely, if $\varphi(u, v, w)$ is regarded as a polynomial in u whose coefficients are themselves polynomials in v and w then the coefficient of u^{r+1} is

$$\frac{1}{(r+1)!}(w-v)^{r+1}\left[\frac{1}{r+2}\xi_{r+1,2}(w-v)-\xi_{r+2,1}w\right]$$

$$=\frac{1}{(r+2)!}(w-v)^{r+1}\xi_{r+1,2}[w-(r+3)v].$$

This means that $\varphi(\cdot, \alpha/(r+3), \alpha)$ is a polynomial of degree at most r, and consequently that the limit

$$\lim_{k\to\infty}\frac{\beta-\varphi(k,\alpha/(r+3),\alpha)}{\psi(k,(r+2)\alpha/(r+3))}=\mu_0$$

will exist. Now (5) and (6) imply that the pair

$$\left(\frac{1}{k}f_1+\frac{\alpha}{r+3}+\frac{\beta-\varphi(k,\alpha/(r+3),\alpha)}{\psi(k,(r+2)\alpha/(r+3))}f_i, h\right)$$

belongs to $P(\mathfrak{g};\mathfrak{h})$ for all sufficiently large integers k, and therefore that

$$\frac{\alpha}{r+3}f_2+\mu_0 f_i\in Q(\mathfrak{g};\mathfrak{h},h).$$

On the other hand, it is easy to see that

$$h\notin Q\left(\mathfrak{g},\frac{\alpha}{r+3}f_2+\mu_0 f_i;\mathfrak{h}\right)$$

provided that $\alpha\neq 0$.

Let θ be the real number satisfying $|\theta|=1$ and $\theta\xi_{r,i}>0$. Then $\psi(0,1)=0$ and the leading coefficient of $\psi(\cdot,1)$ has the same sign as θ, so for each positive integer k the equation $\psi(t,1)=\theta k$ has a solution. As

$$\left\langle \exp(tadX_n)X_2, \frac{1}{k}f_i+(\beta-\theta)f_p\right\rangle =0$$

and

$$\left\langle \exp(tadX_n)X_p, \frac{1}{k}f_i+(\beta-\theta)f_p\right\rangle =\beta-\theta+\frac{1}{k}\psi(t,1)$$

[cf. (1) and (2)] for $t\in\mathbf{R}$ and $k=1,2,\ldots$ it follows that

$$\left(\frac{1}{k}f_i+(\beta-\theta)f_p, h\right)\in P(\mathfrak{g};\mathfrak{h})$$

for all such k, and consequently that $(\beta-\theta)f_p \in Q(\mathfrak{g};\mathfrak{h},h)$, provided that $\alpha=0$. Again provided that $\alpha=0$, it is easy to see that $h \notin Q(\mathfrak{g},(\beta-\theta)f_p;\mathfrak{h})$.

Proof of Proposition 3[1]

It is easy to see that as $\mathfrak{h}$ is not an ideal in $\mathfrak{g}$ there must be a linear form g in $\mathfrak{h}^\perp=\{g'\in\mathfrak{g}', g'=0 \text{ on } \mathfrak{h}\}$ and an element X in $\mathfrak{g}$ satisfying $g\circ adX \notin \mathfrak{h}^\perp$. But then, as $\mathfrak{g}$ is nilpotent, there will be a positive integer r with $g\circ(adX)^r \notin \mathfrak{h}^\perp$ and $g\circ(adX)^i \in \mathfrak{h}^\perp$ for $i>r$. Let h denote the restriction to $\mathfrak{h}$ of $-\frac{1}{r!}g\circ(adX)^r$ and put

$$g_n=\frac{1}{n}\sum_{i=0}^{r-1}\frac{1}{i!}n^{i/r}g\circ(adX)^i$$

for $n=1,2,\ldots$. Then $h\notin Q(\mathfrak{g},0;\mathfrak{h})$ as $h\neq 0$ and $g_n\to 0$ in $\mathfrak{g}'$ as $n\to\infty$, so it is only necessary to show that $(g_n,h)\in P(\mathfrak{g};\mathfrak{h})$ for $n=1,2,\ldots$. But this follows from the fact that for any element Y of $\mathfrak{h}$ and any positive integer n one has

$$\begin{aligned}
&\langle \exp(ad(-n^{1/r}X))Y, g_n\rangle \\
&=\frac{1}{n}\sum_{i=0}^{\infty}\frac{1}{i!}n^{i/r}\langle \exp(ad(-n^{1/r}X))Y, g\circ(adX)^i\rangle \\
&\quad -\frac{1}{n}\sum_{i=r}^{\infty}\frac{1}{i!}n^{i/r}\langle \exp(ad(-n^{1/r}X))Y, g\circ(adX)^i\rangle \\
&=\frac{1}{n}\left\langle \sum_{i=0}^{\infty}\frac{1}{i!}n^{i/r}(adX)^i \exp(ad(-n^{1/r}X))Y, g\right\rangle \\
&\quad -\frac{1}{n}\sum_{i=r}^{\infty}\sum_{j=0}^{\infty}\frac{(-1)^j}{i!j!}n^{(i+j)/r}\langle (adX)^jY, g\circ(adX)^i\rangle \\
&=\frac{1}{n}\langle Y,g\rangle-\frac{1}{r!}\langle Y, g\circ(adX)^r\rangle \\
&=\langle Y,h\rangle
\end{aligned}$$

by the choice of g and r.

References

1. Dixmier, J.: Sur la représentations unitaires des groupes de Lie nilpotents. III, Canad. J. Math. **10**, 321–348 (1958)
2. Felix, R.: Über Integralzerlegungen von Darstellungen nilpotenter Liegruppen. Manuscripta Math. **27**, 279–290 (1979)
3. Felix, R., Henrichs, R.W., Skudlarek, H.L.: Topological Frobenius reciprocity for projective limits of Lie groups. Math. Z. **165**, 19–28 (1979)
4. Kirillov, A.A.: Unitary representations of nilpotent Lie groups. Uspehi Mat. Nauk. **17**, 57–110 (1962)
5. Nielsen, O.A.: The failure of the topological Frobenius property for nilpotent Lie groups. Math. Scand. **45**, 305–310 (1979)

Received June 26, 1980; in revised form April 7, 1981

1 This proof is due to Professor R. Felix and replaces the author's much longer original one

Math. Ann. 256, 569–573 (1981)

Mathematische Annalen
© Springer-Verlag 1981

A Gaussian Correlation Inequality for Certain Bodies in $\mathbb{R}^n$

Christer Borell

Department of Mathematics, Chalmers University of Technology, S-41296 Göteborg, Sweden

1. Introduction

Throughout this paper γ stands for the standard Gauss measure in Euclidean $\mathbb{R}^n$, that is,

$$d\gamma(x)=[\exp(-|x|^2/2)]\,dx/(2\pi)^{n/2}.$$

In [2] Pitt shows that

$$\gamma(A\cap B)\geqq\gamma(A)\gamma(B) \tag{1.1}$$

for all barrels $A, B\subseteqq\mathbb{R}^2$. On the other hand, if A and B denote arbitrary barrels in $\mathbb{R}^n$, $n\geqq 3$, then it is still unknown whether (1.1) must be true or not. (See e.g. Dudley [1] and Tong [4, Conjecture 2.3.1].)

The purpose of this paper is to prove one possible multi-dimensional extension of the Pitt correlation inequality. In fact, we will introduce a suitable family $\mathscr{C}$ of bodies in $\mathbb{R}^n$ such that (1.1) holds for all $A, B\in\mathscr{C}$. Any linear isometry in $\mathbb{R}^n$ leaves $\mathscr{C}$ invariant. Moreover, it turns out that $\mathscr{C}$ contains each body in $\mathbb{R}^n$ of the following geometrical shape:

$$\begin{cases} -f(-x')\leqq\dfrac{x_n}{\sqrt{c^2-|x'|^2}}\leqq f(x'),\ |x'|<c,\ x'\in\mathbb{R}^{n-1},\ x_n\in\mathbb{R},\\ c>0,\ f(x')\geqq 1\quad\text{continuous.}\end{cases}$$

2. Cap Symmetrizations

Below we will make frequent use of different cap symmetrizations in $\mathbb{R}^n$. Since these concepts are not quite stable in the mathematical literature, we first specify exact definitions, which will be held fixed throughout the paper.

Let $1\leqq k\leqq n-1$, let U be an $(n-1-k)$-dimensional linear subspace of $\mathbb{R}^n(n\geqq 2)$, and let v be a unit vector in the orthogonal complement $U^\perp$ of U. The *k-dimensional cap symmetrization* $\Sigma(M)=\Sigma(M;U,v)$ of a closed (open) set $M\subseteqq\mathbb{R}^n$ with respect to (U,v) is defined as follows.

0025-5831/81/0256/0569/$01.00

For any $u\in U$ and $r\geqq 0$, let $S^k(u;r)=S^k(u;r;U^\perp)$ be the k-dimensional sphere of centre u and radius r contained in the affine manifold $u+U^\perp$. The set $\Sigma(M)\cap S^k(u;r)=\phi$, if $M\cap S^k(u;r)=\phi$, $=S^k(u;r)$, if $S^k(u;r)\subseteqq M$, and, otherwise, $\Sigma(M)\cap S^k(u;r)$ is a cap of the sphere $S^k(u;r)$, centred at $S^k(u;r)\cap(\{u\}+\mathbb{R}_+v)$, and such that $\Sigma(M)\cap S^k(u;r)$ and $M\cap S^k(u;r)$ have the same k-dimensional surface area measures. Finally, the cap $\Sigma(M)\cap S^k(u;r)$ is closed (open) if M is closed (open).

The set-valued transformation $\Sigma(\cdot\,;U,v)$ maps closed (open) subsets of $\mathbb{R}^n$ into closed (open) subsets of $\mathbb{R}^n$. This, as well as some other well-known results about cap symmetrizations, which we will use below, may be found in the excellent work Sarvas [3].

3. A Gaussian Correlation Inequality in $\mathbb{R}^n$

For simplicity, the $(n-1)$-dimensional cap symmetrization $\Sigma(\cdot\,;\{0\},e_n)$, where $e_n=(0',1)$, is denoted by Σ_+. Set $H_+=\{(x',x_n)\in\mathbb{R}^n;\ x_n\geqq 0\}$ and $H_+^0=\operatorname{int}H_+$. We shall write $E\in\mathscr{E}_+$, if $E\subseteqq H_+$ is closed and

$$(a',a_n)\in\Sigma_+(E)\Rightarrow(a',ta_n)\in\Sigma_+(E),\qquad 0\leqq t\leqq 1.$$

Below σ_r stands for the surface area measure on $S(0;r)=S^{n-1}(0;r;\{0\}^\perp)$ and we set $\bar{B}(0;r)=\{x\in\mathbb{R}^n;\ |x|\leqq r\}$.

Theorem 3.1. *Suppose $E\subseteqq H_+$ is closed and set $r(E)=\sup\{r\geqq 0;\ \bar{B}(0;r)\cap H_+\subseteqq E\}$. For each finite $r\geqq r(E)$, let $\varrho=\varrho(r)$ be the unique solution of the equation*

$$\sigma_r(E)=c_n r^{n-1}\int_0^{\arcsin(\varrho/r)}\sin^{n-2}\theta\,d\theta,\qquad 0\leqq\varrho\leqq(\pi/2)r,$$

where $c_n=2\pi^{(n-1)/2}/\Gamma((n-1)/2)$.

Then $E\in\mathscr{E}_+$ if and only if the function $r\rightsquigarrow\varrho(r)$ is decreasing.

Proof. Let K denote a cap of $S(0;r)$ with spherical radius s. By direct calculation

$$\sigma_r(K)=c_n r^{n-1}\int_0^{s/r}\sin^{n-2}\theta\,d\theta$$

and Theorem 3.1 follows by drawing a figure.

Suppose that $F\subseteqq\mathbb{R}^n$ and $F\cap H_+\in\mathscr{E}_+$. We shall write $F\in\mathscr{F}$ if there exists a linear isometry P in $\mathbb{R}^n$ such that $Pe_n=-e_n$ and $P(F)\cap H_+=F\cap H_+$. Finally, a set $C\subseteqq\mathbb{R}^n$ is said to belong to the class $\mathscr{C}$ if $Q(C)\in\mathscr{F}$ for an appropriate linear isometry Q in $\mathbb{R}^n$.

Theorem 3.2. *Suppose $A,B\in\mathscr{C}$. Then*

$$\gamma(A\cap B)\geqq\gamma(A)\gamma(B).$$

The proof of Theorem 3.2 is based on two lemmas.

Lemma 3.1. *Assume $E\in\mathscr{E}_+$ and set $G=\Sigma_+(H_+^0\setminus E)$. Let $a=(a',a_n)\in G$, $b=(b',b_n)\in H_+$, and suppose $|a|\leqq|b|$, $a_n\leqq b_n$. Then $b\in G$.*

Proof. Consider the function

$$\varphi(r)=\sigma_r(E), \quad r\geqq r(E),$$

where $r(E)<+\infty$ is as in Theorem 3.1 Since the function

$$\sigma_r(\{|x'|\leqq r(E)\}\cap H_+), \quad r\geqq r(E),$$

is decreasing, it follows that φ decreases.

Now, for each $r\geqq r(E)$, let $h(r)\in[0,r]$ be the unique member satisfying

$$\varphi(r)=\sigma_r(\{0<x_n\leqq h(r)\}).$$

We claim that h decreases. This assertion is obvious in the plane since there h equals the function ϱ appearing in Theorem 3.1. Generally, we have

$$\varphi(r)=e_n r\int_0^{h(r)}(r^2-s^2)^{(n-3)/2}ds$$

where e_n is a positive constant. Hence, h decreases for all n as φ does.

Suppose now to the contrary of the statement in Lemma 3.1 that the point $b\notin G$. Then $b_n\leqq h(|b|)$ and since $a\in G$ and G is open, necessarily $h(|a|)<a_n$. However, as h decreases, we now conclude that $|a|>|b|$, which is a contradiction. This completes the proof of Lemma 3.1.

The following simple and well-known lemma needs no proof.

Lemma 3.2. *Let ν be a probability measure in $\mathbb{R}^d$. Then*

$$\nu(\bar{B}(0;r)\cap(R^d\setminus\bar{B}(0;s)))\leqq\nu(\bar{B}(0;r))\nu(R^d\setminus\bar{B}(0;s)).$$

In particular, if $d=1$ and $f\geqq 0$, $-g\leqq 0$ are symmetrically decreasing on $\mathbb{R}$, then

$$\int fg d\nu\leqq\int f d\nu\int g d\nu.$$

Proof of Theorem 3.2. Without loss of generality we may assume that $A\in\mathscr{F}$ and $B=Q(F)$, where Q is a linear isometry in $\mathbb{R}^n$ and $F\in\mathscr{F}$. Let

$$\Sigma(M)=\Sigma_+(M\cap H_+)\cup[-\Sigma_+((-M)\cap H_+)]$$

for $M\subseteqq\mathbb{R}^n$ closed. If $M\subseteqq\mathbb{R}^n$ is open, then we define $\Sigma(M)$ by the same formula except that H_+ is now replaced by H_+^0. Since $\sigma_r(C)=2\sigma_r(C\cap H_+)$, $C=A, F$, we have

$$\sigma_r(A\cap Q(\mathbb{R}^n\setminus F))\leqq\sigma_r(\Sigma(A)\cap\Sigma(\mathbb{R}^n\setminus F)).$$

In fact, the expression in the right-hand side equals the minimum of $\sigma_r(A)$ and $\sigma_r(\mathbb{R}^n\setminus F)=\sigma_r(Q(\mathbb{R}^n\setminus F))$. Hence

$$\gamma(A\cap Q(\mathbb{R}^n\setminus F))\leqq\gamma(\Sigma(A)\cap\Sigma(\mathbb{R}^n\setminus F)).$$

If $M\subseteqq\mathbb{R}^n$, let $M_{x_n}=\{x'\in\mathbb{R}^{n-1};(x',x_n)\in M\}$, $x_n\in\mathbb{R}$. Set $\gamma=\gamma_n$ and $\gamma_n=\gamma_{n-1}\otimes\gamma_1$, respectively. Then, by Lemmas 3.1 and 3.2,

$$\gamma_{n-1}((\Sigma(A)\cap\Sigma(\mathbb{R}^n\setminus F))_{x_n})\leqq\gamma_{n-1}((\Sigma(A))_{x_n})\gamma_{n-1}((\Sigma(\mathbb{R}^n\setminus F))_{x_n})$$

for each $x_n \in \mathbb{R}$. We now apply the same lemmas agains and have

$$\int \gamma_{n-1}((\Sigma(A))_{x_n})\gamma_{n-1}((\Sigma(\mathbb{R}^n \setminus F))_{x_n})d\gamma_1(x_n) \leqq \gamma(\Sigma(A))\gamma(\mathbb{R}^n \setminus F)).$$

Summing up, it follows that

$$\gamma(A \cap (\mathbb{R}^n \setminus B)) \leqq \gamma(A)\gamma(\mathbb{R}^n \setminus B)$$

which is equivalent to the statement in Theorem 3.2.

4. Concrete Elements in $\mathscr{C}$

Finally, in this note, we will pick out some geometrically lucid elements belonging to the class $\mathscr{C}$. Since the way from $\mathscr{E}_+$ to $\mathscr{C}$ is quite simple, we concentrate on the first class below.

Let $c>0$. A closed set $E \subseteqq H_+$ is said to belong to the class $\mathscr{E}_c$ if

$$\bar{B}(0;c) \cap H_+ \subseteqq E \subseteqq \{(x', x_n) \in \mathbb{R}^n ; |x'| \leqq c, x_n \geqq 0\}$$

and

$$(x', x_n) \in E \Rightarrow (x', tx_n) \in E, \quad 0 \leqq t \leqq 1.$$

Theorem 4.1. *Let $\Sigma = \Sigma(\cdot\,; U, v)$ be a k-dimensional cap symmetrization in $\mathbb{R}^n$, $1 \leqq k \leqq n-1$, and suppose*

$$e_n \in U$$

or

$$v = e_n.$$

Then, for any fixed $c>0$, Σ leaves $\mathscr{E}_c$ invariant, that is,

$$E \in \mathscr{E}_c \Rightarrow \Sigma(E) \in \mathscr{E}_c.$$

The proof of Theorem 4.1 depends on the following

Lemma 4.1. *Let $0<r<s$, $0<\theta_0<\pi/2$, and assume the arc*

$$I = \{se^{i\theta} ; |\theta - \theta_0| \leqq \delta\} \subseteqq \mathbb{C}$$

is contained in the strip $T_r = \{z \in \mathbb{C} ; |\mathrm{Re}\, z| \leqq r\}$. Set $p(z) = w$, $z \in S(0;s) \cap T_r$, where $|w| = r$, $\mathrm{Re}\, w = \mathrm{Re}\, z$ and $\mathrm{Im}\, w \geqq 0$. Then the function

$$\varphi(\theta) = \sigma_r(p(e^{i\theta} I)), \quad 0 \leqq \theta \leqq (\pi/2) - \theta_0,$$

decreases.

Lemma 4.1 follows at once by evaluating the derivative φ' explicitly and we omit the details here.

Proof of Theorem 4.1. If the vector $e_n \in U$, then Theorem 4.1 is trivial. Moreover, since cap symmetrizations are continuous from the outside, the case $k=1$, $v=e_n$ is covered by Lemma 4.1 and the principle of mathematical induction.

Let $\mathscr{K}$ be the family of all non-empty compact subsets of $\mathbb{R}^n$. It suffices to show that $\mathscr{E}_c \cap \mathscr{K}$ is invariant under Σ. Note that $\mathscr{E}_c \cap \mathscr{K}$ is a closed subset of $\mathscr{K}$, equipped with the Blaschke topology.

We now proceed by mathematical induction over k. Let $2 \leqq k \leqq n-1$ be fixed and suppose $v = e_n$. We first choose two orthogonal unit vectors a and b belonging to $(U + \mathbb{R}e_n)^{\perp}$ and then introduce the following $(k-1)$-dimensional cap symmetrizations:

$$\Sigma_1 = \Sigma(\cdot\,; U + \mathbb{R}e_n, a),$$
$$\Sigma_2 = \Sigma(\cdot\,; U + \mathbb{R}b, e_n).$$

Assume now that Σ_1 and Σ_2 leave $\mathscr{E}_c \cap \mathscr{K}$ invariant. By applying the key result

$$\Sigma(K) = \lim_{i \to +\infty} (\Sigma_1 \circ \Sigma_2)^i(K), \quad K \in \mathscr{K}, \tag{4.1}$$

it follows at once that Σ leaves $\mathscr{E}_c \cap \mathscr{K}$ invariant.

This completes the proof of Theorem 4.1.

Corollary 4.1. $\mathscr{E}_c \subsetneqq \mathscr{E}_+$, $c > 0$.

Let $C \neq \mathbb{R}^n$ be a barrel in $\mathbb{R}^n$ and set

$$r(C) = \max\{r\,; \bar{B}(0\,; r) \subseteqq C\}.$$

If $C \cap S(0\,; r(C))$ contains at least one $(n-2)$-dimensional sphere, then Corollary 4.1 shows that $C \in \mathscr{C}$. This fact can be proved more directly from the Brunn-Minkowski inequality in spherical geometry. However, our substitute (4.1) is, generally, more illuminating and, especially, it seems to be so in this connection.

References

1. Dudley, R.M.: Gaussian random processes, by I.A. Ibragimov, Yu.A. Rozanov: Book review. Bull. Am. Math. Soc. **2**, 373–378 (1980)
2. Pitt, L.D.: A Gaussian correlation inequality for symmetric convex sets. Ann. Probability **5**, 470–474 (1977)
3. Sarvas, J.: Symmetrization of condensers in n-space. Ann. Acad. Sci. Fenn. **AI 522**, 1–44 (1972)
4. Tong, Y.L.: Probability inequalities in multivariate distributions. New York, London, Toronto, Sydney, San Francisco: Academic Press 1980

Received February 16, 1981

Math. Ann. 256, 574 (1981)

Mathematische Annalen
© Springer-Verlag 1981

Erratum

Uniform Approximation by Meromorphic Functions Having Prescribed Poles

Math. Ann. **243**, 83–93 (1979)

I am indebted to Paul Gauthier for pointing out that the proof of Lemma 13 of the above-cited paper is incorrect as written. The following will correct the proof.

Suppose X is a compact Riemann surface, Y is an open subset of X which we may think of as a possibly disconnected Riemann surface, and Z is a relatively closed subset of Y. Denote by $\partial_X^\infty Z$ the compact set $\bar{Z}-Z=\bar{Z}-Y=\partial Z-Z$, where closures and boundaries are relative to X. Thus, the points of $\partial_X^\infty Z$ are at ∞ as far as the set Z and the surface Y are concerned. In the proof of Lemma 13 the symbol "∂^∞" should be written in place of "∂" wherever the latter occurs.