SUB | NIEDERSÄCHSISCHE STAATS- UND UNIVERSITÄTSBIBLIOTHEK GÖTTINGEN

Werk

Titel: Mathematische Zeitschrift
Verlag: Springer
Jahr: 1982
Kollektion: Mathematica
Digitalisiert: Niedersächsische Staats- und Universitätsbibliothek Göttingen
Werk Id: PPN266833020_0180
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN266833020_0180

Übergeordnetes Werk

Werk Id: PPN266833020
PURL: http://resolver.sub.uni-goettingen.de/purl?PPN266833020

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 Zeitschrift

Band 180 · 1982

Herausgegeben von / Edited by

H. Heyer, Tübingen

O. H. Kegel, Freiburg

W. Klingenberg, Bonn

C. C. Moore, Berkeley

A. Pfluger, Zürich

H. H. Schaefer, Tübingen

L. Smith, Göttingen

W. v. Wahl, Bayreuth

Wissenschaftlicher Beirat / Advisory Board

T. tom Dieck, Göttingen · B. Eckmann, Zürich

E. Heinz, Göttingen · K. H. Hofmann, New Orleans

U. Krengel, Göttingen · W. Magnus, New York

G. Pickert, Gießen · K. W. Roggenkamp, Stuttgart

H. Salzmann, Tübingen · K. Zeller, Tübingen

Springer-Verlag
Berlin · Heidelberg · New York

"Mathematische Zeitschrift" was founded in 1918 and edited by *L. Lichtenstein* in cooperation with *K. Knopp, E. Schmidt* and *I. Schur;* after Lichtenstein's death, 1933, it was edited by *K. Knopp.* The Editorial Committee was increased to include *E. Kamke* and *F.K. Schmidt* in 1933, *R. Nevanlinna* in 1936 and *H. Wielandt* in 1950. The latter became Managing Editor in 1952. He was succeeded in 1973 by *H. Salzmann* and 1977 by *H. Heyer.*

Die **„Mathematische Zeitschrift"** wurde im Jahre 1918 von *L. Lichtenstein* unter der Mitwirkung von *K. Knopp, E. Schmidt* und *I. Schur* gegründet und herausgegeben. Nach dem Tode Lichtensteins übernahm *K. Knopp* 1933 die Herausgabe. Die Schriftleitung ergänzte sich 1933 durch *E. Kamke* und *F.K. Schmidt,* 1936 durch *R. Nevanlinna* und 1950 durch *H. Wielandt,* der 1952 die Herausgabe übernahm. 1973 ging die Federführung an *H. Salzmann* und 1977 an *H. Heyer* über.

Manuscripts may be sent to / Manuskripte nehmen entgegen:

H. Heyer, Mathematisches Institut der Universität, Auf der Morgenstelle 10, D-7400 Tübingen, Federal Republic of Germany

O.H. Kegel, Mathematisches Institut der Universität, Albertstr. 23b, D-7800 Freiburg, Federal Republic of Germany

W. Klingenberg, Mathematisches Institut der Universität, Wegelerstr. 10, D-5300 Bonn, Federal Republic of Germany

C.C. Moore, Department of Mathematics, University of California, Berkeley, CA 94720, USA

A. Pfluger, Eidgen. Technische Hochschule, Mathematisches Seminar, CH-8000 Zürich, Schweiz

H.H. Schaefer, Mathematisches Institut der Universität, Auf der Morgenstelle 10, D-7400 Tübingen, Federal Republic of Germany

L. Smith, Mathematisches Institut der Universität, Bunsenstr. 3–5, D-3400 Göttingen, Federal Republic of Germany

W. v. Wahl, Mathematisches Institut der Universität, Postfach 3008, D-8580 Bayreuth, Federal Republic of Germany

The exclusive copyright for all languages and countries, including the right for photomechanical and any other reproductions, also in microform, is transferred to the publisher.
Authors of this journal can benefit from library and photocopy fees collected by VG WORT if certain conditions are met. If an author lives in the Federal Republic of Germany or in West Berlin it is recommended that he contacts Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2, for detailed information.

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, bleiben 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.
Autoren dieser Zeitschrift können unter gewissen Voraussetzungen in die Individualausschüttung von Mitteln aus der Bibliothektantieme und dem Fotokopieraufkommen mit einbezogen werden. Genaue Informationen erteilt die Verwertungsgesellschaft WORT, Abteilung Wissenschaft, Goethestraße 49, D-8000 München 2.

Springer-Verlag/Berlin Heidelberg New York
Printers: Universitätsdruckerei H. Stürtz AG, Würzburg
Printed in Germany – © Springer Verlag GmbH & Co. KG Berlin Heidelberg 1982

Inhalt/Contents

Bier, T., Schwardmann, U.: Räume normierter Bilinearformen und Clifford-strukturen 203

Borell, C.: Positivity Improving Operators and Hypercontractivity . . . 225

Bunce, L.J.: The Theory and Structure of Dual *JB*-Algebras 525

Cambern, M., Greim, P.: The Dual of a Space of Vector Measures . . . 373

Cameron, P.J., Neumann, P.M., Teague, D.N.: On the Degrees of Primitive Permutation Groups 141

Christensen, J.P.R., Ressel, P.: Positive Definite Kernels on the Complex Hilbert Sphere 193

Craven, T.C.: On the Prime Ideal of an Ordering of Higher Level . . . 553

DeFranza, J., Fleming, D.J.: The Banach Algebra of Conservative Triangular Matrices 463

Dodds, P.G., Schep, A.R.: Compact Integral Operators on Banach Function Spaces 249

Dotti Miatello, I.: Ricci Curvature of Left Invariant Metrics on Solvable Unimodular Lie Groups 257

Ecker, K.: Estimates for Evolutionary Surfaces of Prescribed Mean Curvature 179

Eklof, P.C.: Homological Dimension and Stationary Sets 1

Fleming, D.J., s. DeFranza, J. 463

Frigerio, A., Verri, M.: Long-Time Asymptotic Properties of Dynamical Semigroups on W^*-algebras 275

Goodearl, K.R., Warfield, R.B. Jr.: Primitivity in Differential Operator Rings 503

Gotzmann, G.: Einige einfach-zusammenhängende Hilbertschemata . . . 291

Greim, P., s. Cambern, M. 373

Harman, G.: Primes in Short Intervals 335

Hofbauer, F., Keller, G.: Ergodic Properties of Invariant Measures for Piecewise Monotonic Transformations 119

Howie, J.: On Locally Indicable Groups 445

Ihringer, T.: On Groupoids Having a Linear Congruence Class Geometry 395
Innami, N.: Differentiability of Busemann Functions and Total Excess 235
Jensen, H.E.: Second-Dual Liminal C^*-Algebras 567
Kąkol, J.: A Note on Metrizable Linear Topologies Strictly Finer than a Given Metrizable Complete Linear Topology 287
Kaniuth, E.: Weak Containment and Tensor Products of Group Representations 107
Kato, T.: Superconvexity of the Spectral Radius, and Convexity of the Spectral Bound and the Type 265
Keller, G., s. Hofbauer, F. 119
Kirwan, W.E., Schober, G.: New Inequalities from Old Ones 19
Kolster, M.: Quadratic Forms and Artin's Reciprocity Law 81
Kosarew, S.: Das Modulproblem für holomorphe Einbettungen mit konkaver Umgebungsstruktur 307
Kowalski, O., Vanhecke, L.: Ball-Homogeneous and Disk-Homogeneous Riemannian Manifolds 429
Neumann, P.M., s. Cameron, P.J., et al. 141
Opfer, G., Schober, G.: On Convergence and Quasiregularity of Interpolating Complex Planar Splines 469
Pedersen, E.K.: Geometrically Defined Transfers, Comparisons 535
Platte, E.: Nicht-ausgeartete reguläre Überlagerungen 413
Ressel, P., s. Christensen, J.P.R. 193
Reyes, A.: Some Properties of a Regular Sequence in Hilbert Space . . . 331
Rocha-Caridi, A., Wallach, N.R.: Projective Modules over Graded Lie Algebras. I 151
Röhl, F.: On the Isomorphism Problem for Integral Group-Rings of Circle-Groups 419
Ronan, M.A.: Locally Truncated Buildings and M_{24} 489
Ruscheweyh, S., Wirths, K.-J.: On Extreme Bloch Functions with Prescribed Critical Points 91
Sauvigny, F.: Flächen vorgeschriebener mittlerer Krümmung mit eineindeutiger Projektion auf eine Ebene 41
Schep, A.R., s. Dodds, P.G. 249
Schober, G., s. Kirwan, W.E. 19
Schober, G., s. Opfer, G. 469
Schwardmann, U., s. Bier, T. 203
Sebastiani, M.: Variation du nombre de Lefschetz dans une famille d'automorphismes analytiques 483
Shen, C.L.: The Gap Phenomena of Yang-Mills Fields over the Complete Manifold 69

Steen, G. Van: Galois Coverings of the Non-Archimedean Projective Line 217
Teague, D.N., s. Cameron, P.J., et al. 141
Teixeira, M.A.: On Topological Stability of Divergent Diagrams of Folds 361
Toppila, S. Winkler, J.: On the Spherical Derivative of a Meromorphic Function with a Valiron Deficient Value 545
Vanhecke, L., s. Kowalski, O. 429
Verri, M., s. Frigerio, A. 275
Wallach, N.R., s. Rocha-Caridi, A. 151
Warfield, R.B. Jr., s. Goodearl, K.R. 503
Weber, H.: Unabhängige Topologien, Zerlegung von Ringtopologien . . 379
Weidmann, J.: Absolut stetiges Spektrum bei Sturm-Liouville-Operatoren und Dirac-Systemen 423
Werner, H.: Rationale Interpolation von $|x|$ in äquidistanten Punkten . . 11
Werner, M.: Zu einer Aussage über die kettengeometrischen Erzeugenden 79
Winkler, J., s. Toppila, S. 545
Wirths, K.-J., s. Ruscheweyh, S. 91
Zimmermann, B.: Das Nielsensche Realisierungsproblem für hinreichend große 3-Mannigfaltigkeiten 349

Covered by Zentralblatt für Mathematik and
Current Mathematical Publications

Mathematische Zeitschrift

Band 180 Heft 1 1982

Herausgegeben von / Edited by
H. Heyer, Tübingen
O. H. Kegel, Freiburg
W. Klingenberg, Bonn
C. C. Moore, Berkeley
A. Pfluger, Zürich
H. H. Schaefer, Tübingen
L. Smith, Göttingen
W. v. Wahl, Bayreuth

Wissenschaftlicher Beirat / Advisory Board
T. tom Dieck, Göttingen · B. Eckmann, Zürich
E. Heinz, Göttingen · K. H. Hofmann, New Orleans
U. Krengel, Göttingen · W. Magnus, New York
G. Pickert, Gießen · K. W. Roggenkamp, Stuttgart
H. Salzmann, Tübingen · K. Zeller, Tübingen

8 Z Va+. 527

Springer-Verlag
Berlin Heidelberg New York

Math. Z. ISSN 0025–5874 MAZEAX 180 (1) 1–140 (1982) April 1982

Niedersächsische Staats- u.
Universitätsbibliothek

Mathematische Zeitschrift

Die „**Mathematische Zeitschrift**" wurde im Jahre 1918 von *L. Lichtenstein* unter der Mitwirkung von *K. Knopp, E. Schmidt* und *I. Schur* gegründet und herausgegeben. Nach dem Tode Lichtensteins übernahm *K. Knopp* 1933 die Herausgabe. Die Schriftleitung ergänzte sich 1933 durch *E. Kamke* und *F. K. Schmidt*, 1936 durch *R. Nevanlinna* und 1950 durch *H. Wielandt*, der 1952 die Herausgabe übernahm. 1973 ging die Federführung an *H. Salzmann* und 1977 an *H. Heyer* über.

"**Mathematische Zeitschrift**" was founded in 1918 and edited by *L. Lichtenstein* in cooperation with *K. Knopp, E. Schmidt* and *I. Schur;* after Lichtenstein's death, 1933, it was edited by *K. Knopp.* The Editorial Committee was increased to include *E. Kamke* and *F. K. Schmidt* in 1933, *R. Nevanlinna* in 1936 and *H. Wielandt* in 1950. The latter became Managing Editor in 1952. He was succeeded in 1973 by *H. Salzmann* and 1977 by *H. Heyer.*

Geschäftsführender Herausgeber / Managing Editor

H. Heyer — Mathematisches Institut der Universität, Auf der Morgenstelle 10
D-7400 Tübingen, Federal Republic of Germany

Manuskripte nehmen außerdem entgegen / Manuscripts may also be sent to:

O. H. Kegel — Mathematisches Institut der Universität, Albertstr. 23b
D-7800 Freiburg, Federal Republic of Germany

W. Klingenberg — Mathematisches Institut der Universität, Wegelerstr. 10
D-5300 Bonn, Federal Republic of Germany

C. C. Moore — Department of Mathematics, University of California,
Berkeley, CA 94720, USA

A. Pfluger — Eidgen. Technische Hochschule, Mathematisches Seminar
CH-8000 Zürich, Schweiz

H. H. Schaefer — Mathematisches Institut der Universität, Auf der Morgenstelle 10
D-7400 Tübingen, Federal Republic of Germany

L. Smith — Mathematisches Institut der Universität, Bunsenstr. 3–5
D-3400 Göttingen, Federal Republic of Germany

W. v. Wahl — Mathematisches Institut der Universität, Postfach 3008
D-8580 Bayreuth, Federal Republic of Germany

Math. Z. 180, 1–9 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Homological Dimension and Stationary Sets

Paul C. Eklof*

Mathematics Department, Bedford College, University of London, Regent's Park, London NW1 4NS, England
Current address: Mathematics Department, University of California, Irvine, CA 92717, U.S.A.

Introduction

Throughout this paper R will denote a ring with 1, and all modules referred to will be left R-modules. We let $gld(R)$ denote the left global dimension of R, and, if M is a module, $pd(M)$ the projective dimension of M (see e.g. [8]).

A *continuous chain* of modules is a sequence $\{M_\nu | \nu < \alpha\}$ such that for all ν, M_ν is a submodule of $M_{\nu+1}$, and for all limit ordinals σ, $M_\sigma = \bigcup_{\nu<\sigma} M_\nu$. Auslander proved the following result which he used as a lemma in proving the Global Dimension Theorem.

Auslander's Lemma ([1]). *Let $\{M_\nu | \nu < \alpha\}$ be a continuous chain of modules such that $pd(M_0) \leqq n$ and for all $\nu+1<\alpha$, $pd(M_{\nu+1}/M_\nu) \leqq n$. Then $pd(\bigcup_{\nu<\alpha} M_\nu) \leqq n$.*

In this paper we shall prove a converse to a strong version of Auslander's Lemma, thereby giving a *lower* bound to $pd(\bigcup_{\nu<\alpha} M_\nu)$ in terms of the projective dimension of quotients M_μ/M_ν $(\nu<\mu<\alpha)$ of modules in the chain (Theorem 1.6). The converse says, roughly, that $pd(M) \geqq n$ provided "sufficiently many" of the quotients M_μ/M_ν have projective dimension $\geqq n$; the precise meaning of "sufficiently many" is given in terms of the notion of a *stationary* set. (Set-theoretic terminology is defined in Sect. 0.)

For this converse we need to impose some additional assumptions, namely, that α is a regular uncountable cardinal (e.g. $\alpha = \aleph_{k+1}$ for any k); that every M_ν is generated by fewer than α elements; and, in addition, "coherence" assumptions. For example, here is a very useful – though not the most general – version of the main result.

Main Theorem (version III). *Let R be a left coherent ring, and let M be an R-module such that every finitely generated submodule of M is finitely presented. Suppose M is the union of a continuous chain $\{M_\nu | \nu < \aleph_{k+1}\}$ of R-modules such*

* Partially supported by National Science Foundation Grant MCS-8003591

0025-5874/82/0180/0001/$01.80

that for all ν, M_ν is generated by $\leqq \aleph_k$ elements and $pd(M_\nu) \leqq n-1$. Then $pd(M) \geqq n$ if and only if $\{\nu < \aleph_{k+1} |$ there exists $\mu > \nu$ such that $pd(M_\mu/M_\nu) \geqq n\}$ is stationary in $\aleph_{k+1}$.

The necessity of the condition is a strong version of (the contrapositive of) Auslander's Lemma. It is only this strengthened version of Auslander's Lemma - for whose expression the notion of stationary set is essential - which has a converse. (See Remark 1.8(5).)

In Sect. 0 we review the set-theoretic prerequisites required for what follows. In Sect. 1 we prove the Main Theorem. Finally in Sect. 2 we indicate how the Main Theorem can be used to give new expositions of proofs of results of Osofsky; connections with Osofsky's work are discussed.

I would like to thank László Fuchs for asking me about a converse to Auslander's Lemma, and Mike Prest for helpful conversations in the course of this work.

0. Set-Theoretic Preliminaries

We shall assume the reader has some familiarity with ordinal and cardinal numbers. We shall adopt the convention that every ordinal number is the set of its predecessors, and that a cardinal number is an ordinal which is not equipotent with any of its predecessors. We shall use interchangeably the notations ω_α and $\aleph_\alpha$ for the αth infinite cardinal. Thus, in particular, $\omega = \omega_0 = \aleph_0 = \{n: n \text{ is a finite ordinal}\}$. If κ is a cardinal, κ^+ denotes the successor cardinal of κ i.e. if $\kappa = \aleph_\alpha$, $\kappa^+ = \aleph_{\alpha+1}$.

For any set X, let $|X|$ denote the cardinality of X. If κ is an infinite cardinal, the *cofinality of* κ, denoted $cf(\kappa)$, is $\min\{|X|: X \subseteq \kappa$ is cofinal in X i.e. $\sup X = \kappa\}$. An infinite cardinal κ is called *regular* if $cf(\kappa) = \kappa$. In particular, every successor cardinal is regular.

Let κ be a regular uncountable cardinal. A subset $\mathscr{C}$ of κ is called *closed and unbounded* (or *cub*, for short) if

(i) $\mathscr{C}$ is closed (or continuous), i.e. for all $X \subseteq \mathscr{C}$, if $\sup X \in \kappa$, then $\sup X \in \mathscr{C}$; and (ii) $\mathscr{C}$ is unbounded, i.e. $\sup \mathscr{C} = \kappa$.

For example, the following are cubs in κ: κ itself; for every $\alpha < \kappa$, $\{\nu < \kappa: \nu > \alpha\}$; the set of all limit ordinals $< \kappa$; the set of ordinals $< \kappa$ which are limits of limit ordinals, etc.

A function $\phi: \kappa \to \kappa$ is called *normal* if it is strictly increasing and continuous (i.e. for all $X \subseteq \kappa$, if $\sup X \in \kappa$, then $f(\sup X) = \sup f(X)$). It is easy to see that the range of a normal function is a cub in κ, and, conversely, that every cub in κ is the range of a unique normal function, namely the function which "enumerates" the cub in increasing order. For example, the set of all limit ordinals $< \kappa$ is the range of the normal function $\phi: \kappa \to \kappa$ defined by $\phi(\alpha) = \omega\alpha$.

Subsets of κ which are cubs, or, more generally, which contain a cub, should be thought of as being very large subsets of κ, containing "almost all" of κ. The following lemma, which is crucial, implies that the intersection of two very large subsets of κ is also very large.

0.1. **Lemma.** *If $\mathscr{C}_0$ and $\mathscr{C}_1$ are cubs in κ then $\mathscr{C}_0 \cap \mathscr{C}_1$ is a cub in κ.*

Proof. Clearly $\mathscr{C}_0 \cap \mathscr{C}_1$ is closed since $\mathscr{C}_0$ and $\mathscr{C}_1$ are. To prove that it is unbounded we must show that for every $\mu < \kappa$ there exists $\tau \in \mathscr{C}_0 \cap \mathscr{C}_1$ such that $\tau \geqq \mu$. Let $\tau_0 = \mu$; if τ_{n-1} has been defined and n is even (resp. odd) choose $\tau_n \in \mathscr{C}_0$ (resp. $\mathscr{C}_1$) such that $\tau_n \geqq \tau_{n-1}$; this is possible since $\mathscr{C}_0$ (resp. C_1) is unbounded. Let $\tau = \sup\{\tau_n | n \in \omega\}$; then $\tau \in \mathscr{C}_0$ since $\tau = \sup\{\tau_{2m} : m \in \omega\}$ and $\mathscr{C}_0$ is closed; and $\tau \in \mathscr{C}_1$ since $\tau = \sup\{\tau_{2m+1} : m \in \omega\}$. □

(In fact, the intersection of fewer than κ cubs in κ is a cub in κ.)

A subset E of κ is called *stationary* in κ if for every cub $\mathscr{C}$ in κ, $\mathscr{C} \cap E \neq \varnothing$. In particular, by 0.1, every subset of κ containing a cub is stationary.

The following provides examples of stationary sets which are not cubs.

0.2. **Lemma.** *Let $n \geqq 1$. For every $0 \leqq k \leqq n-1$*

$$E_k = \{\sigma < \omega_n : \sigma \text{ is a limit ordinal and } cf(\sigma) = \omega_k\}$$

is a stationary subset of ω_n. If $n > 1$, E_k does not contain a cub.

Proof. If $\phi : \omega_n \to \omega_n$ is any normal function, then $\phi(\omega_k)$ is a limit ordinal of cofinality ω_k. Hence E meets the range of every normal function, i.e. E meets every cub in ω_n. If $n > 1$, E_k does not contain a cub since there exists $l \neq k$ such that $0 \leqq l \leqq n-1$, and E_l is stationary and disjoint from E_k. □

The axiom of choice is needed to show the existence of stationary subjects of ω_1 which do not contain any cub. In fact, the axiom of choice implies that there are 2^κ essentially different stationary subsets of any regular uncountable κ, where E_1 and E_2 are essentially different if $(E_1 - E_2) \cup (E_2 - E_1)$ is stationary.

1. The Converse to Auslander's Lemma

We begin by introducing some terminology (cf. [2]).

1.1. **Definition.** A module M is *κ-generated* if M has a generating set of cardinality $< \kappa$. (Note that in keeping with set theoretic usage, our terminology differs from the usage of some algebraists. Thus e.g. M is κ^+-generated if and only if it is generated by at most κ elements.) If M is κ^+-generated, a *κ-filtration* of M is a continuous chain $\{M_\nu | \nu < \kappa\}$ of κ-generated submodules of M whose union is M.

It is easy to see that every κ^+-generated module has a κ-filtration: simply construct one by transfinite induction, taking unions at limit ordinals, and at stage $\nu+1$ defining $M_{\nu+1}$ to include M_ν and the νth element of M in some predetermined well-ordering of type $\leqq \kappa$.

The following lemma is purely set-theoretic in character, but, for convenience, we formulate it in terms of κ-filtrations of modules. It is proved by a standard set-theoretic argument, in which the regularity and uncountability of κ plays a crucial role.

1.2. **Lemma.** *Let κ be a regular uncountable cardinal. Let $A=\bigcup_{\nu<\kappa} A_\nu$, $B=\bigcup_{\nu<\kappa} B_\nu$ be κ-filtrations of κ^+-generated modules A and B and let $f: A\to B$ be any function (not necessarily a homomorphism). Then there is a normal function $\phi: \kappa\to\kappa$ such that for all $\nu<\kappa$, $f(A_{\phi(\nu)})=f(A)\cap B_{\phi(\nu)}$.*

Proof. We claim that $\mathscr{C}\overset{\text{def}}{=}\{\nu<\kappa | f(A_\nu)=(A)\cap B_\nu\}$ is a cub in κ; if true, then we are done for we take ϕ to be a normal function whose range is $\mathscr{C}$. Clearly $\mathscr{C}$ is closed (since the given κ-filtrations are continuous by definition). We must show that for any $\mu<\kappa$ there exists $\nu\in\mathscr{C}$ such that $\nu\geqq\mu$. Let $\nu_0=\mu$; if ν_n has been defined and n is even choose $\nu_{n+1}\geqq\nu_n$ such that $f(A_{\nu_n})\subseteq B_{\nu_{n+1}}$; if n is odd choose $\nu_{n+1}\geqq\nu_n$ such that $f(A_{\nu_{n+1}})\supseteq f(A)\cap B_{\nu_n}$. (In both cases the choice is possible since A_{ν_n} (resp. B_{ν_n}) is κ-generated and κ is regular.) Let $\nu=\sup\{\nu_n: n\in\omega\}$; then clearly $\nu<\kappa$ and $\nu\in\mathscr{C}$. □

1.3. **Definition.** Say that M is *κ-presented* if there is a short exact sequence

$$0\to K\to F\xrightarrow{\pi} M\to 0$$

such that F is free of rank κ and K is κ-generated. A ring R is left *κ-coherent* if every κ-generated left ideal of R is κ-presented, or, equivalently, if every κ-generated submodule of a free module is κ-presented (cf. [3]). R is called left *coherent* if R is left $\aleph_0$-coherent. (What is called $\aleph_0$-coherent in [3] is called $\aleph_1$-coherent in our terminology.)

1.4. **Theorem.** *Let κ be a regular uncountable cardinal. Suppose R is κ^+-coherent and M is a κ^+-presented R-module. If $pd(M)\leqq n$, then there is a κ-filtration $\{M_\nu|\nu<\kappa\}$ of M such that for all $\nu<\kappa$, $pd(M_\nu)\leqq n$ and for all $\nu<\mu<\kappa$, $pd(M_\mu/M_\nu)\leqq n$.*

Before beginning the proof we note, for ease of reference, the following well-known fact (which follows easily e.g. from the long exact sequence for Ext).

1.5. **Lemma.** *Let $0\to A\to B\to C\to 0$ be a short exact sequence of R-modules. Then $pd(C)\leqq\max\{pd(A)+1,\ pd(B)\}$, and equality holds if $pd(A)\neq pd(B)$. In particular, if B is projective and A is not, $pd(C)=pd(A)+1$.* □

Proof of 1.4. The proof is by induction on $n\in\omega$. Suppose first that $n=0$, i.e. M is projective. One can appeal to a well-known theorem of Kaplansky, but here is a self-contained proof. There is a free module F of rank κ and homomorphisms $\pi: F\to M$ and $\rho: M\to F$ such that $\pi\circ\rho=\mathrm{id}_M$. Then it is possible to construct a κ-filtration $\{M_\nu: \nu<\kappa\}$ of M and a κ-filtration $\{F_\nu: \nu<\kappa\}$ of F such that each F_ν is a direct summand of F and for all $\nu<\kappa$, $\rho(M_\nu)\subseteq F_\nu$ and $\pi(F_\nu)=M_\nu$. (In fact, apply Lemma 1.2 with $A=B=F$, $\{A_\nu|\nu<\kappa\}=\{B_\nu|\nu<\kappa\}$ a κ-filtration by direct summands, and $f=\rho\circ\pi$; if ϕ is as in 1.2, let $F_\nu=A_{\phi(\nu)}$ and $M_\nu=\pi(F_\nu)$.) Clearly for all ν, M_ν is projective. For all $\nu<\mu<\kappa$, $\pi|F_\mu$ induces a homomorphism

$$\bar{\pi}: F_\mu/F_\nu\to M_\mu/M_\nu$$

and $\rho|M$ induces a homomorphism

$$\bar{\rho}: M_\mu/M_\nu \to F_\mu/F_\nu.$$

Clearly $\bar{\pi}\circ\bar{\rho}$ is the identity on M_μ/M_ν, which demonstrates that M_μ/M_ν is projective.

Now suppose that $n>0$ and that (1) has been proved for $n-1$. Let M be a κ^+-generated module of projective dimension n. By hypothesis there is a short exact sequence

$$0\to K\to F\xrightarrow{\pi} M\to 0$$

where F is free of rank κ and K is a κ^+-generated submodule of F. Note that $pd(K)\leqq n-1$ (Lemma 1.5). Moreover since R is κ^+-coherent K is κ^+-presented. Hence by induction, K has a κ-filtration $K=\bigcup_{\nu<\kappa} K_\nu$ such that for all ν, $pd(K_\nu)\leqq n-1$ and for all $\nu<\mu<\kappa$, $pd(K_\mu/K_\nu)\leqq n-1$.

Letting $F=\bigcup_{\nu<\kappa} F_\nu$ be a κ-filtration of F by direct summands and again applying Lemma 1.2 (with $A=K$, $A_\nu=K_\nu$, $B=F$, $B_\nu=F_\nu$ and $f=$inclusion) we obtain a normal function $\phi:\kappa\to\kappa$ such that for all $\nu<\kappa$, $F_{\phi(\nu)}\cap K=K_{\phi(\nu)}$. Let $M_\nu=\pi(F_{\phi(\nu)})$. Then for all $\nu<\mu<\kappa$ we have

$$\begin{array}{ccccc} 0\to & K_{\phi(\mu)} & \to F_{\phi(\mu)} & \to M_\mu & \to 0 \\ & \cup| & \cup| & \cup| & \\ 0\to & K_{\phi(\nu)} & \to F_{\phi(\nu)} & \to M_\nu & \to 0 \end{array}$$

and $F_{\phi(\nu)}\cap K_{\phi(\mu)}=K_{\phi(\nu)}$. By Lemma 1.5 the top row implies $pd(M_\mu)\leqq n$. Moreover this diagram induces an exact sequence

$$0\to K_{\phi(\mu)}/K_{\phi(\nu)}\to F_{\phi(\mu)}/F_{\phi(\nu)}\to M_\mu/M_\nu\to 0.$$

Since $F_{\phi(\mu)}/F_{\phi(\nu)}$ is free and $pd(K_{\phi(\mu)}/K_{\phi(\nu)})\leqq n-1$ we have that $pd(M_\mu/M_\nu)\leqq n$ (Lemma 1.5). This completes the proof. □

1.6. **Main Theorem.** (Version I). *Let κ, R, and M be as in Theorem* 1.4. *Let $\{M_\nu|\nu<\kappa\}$ be any κ-filtration of M. Then $pd(M)\geqq n$ if and only if*

$$S_1=\{\nu<\kappa|pd(M_\nu)\geqq n\}$$

or $S_2=\{\nu<\kappa|$ there exists $\mu>\nu$ such that $pd(M_\mu)\leqq n-1$ and $pd(M_\mu/M_\nu)\geqq n\}$ is stationary in κ.

Proof. ($\Leftarrow$) Suppose that S_1 or S_2 is stationary in κ but $pd(M)\leqq n-1$. By Theorem 1.4, there is a κ-filtration $\{M'_\nu|\nu<\kappa\}$ of M such that for all $\nu<\mu<\kappa$, $pd(M'_\mu/M'_\nu)\leqq n-1$. By Lemma 1.2 (with $A=B=M$, $A_\nu=M_\nu$, $B_\nu=M'_\nu$ and $f=$identity) there is a cub $\mathscr{C}$ (the range of ϕ) such that for all $\nu\in\mathscr{C}$, $M_\nu=M'_\nu$. Suppose that S_1 is stationary in κ; then there is a $\nu\in\mathscr{C}\cap S_1$; but then $pd(M'_\nu)\leqq n-1$, $pd(M_\nu)\geqq n$ and $M_\nu=M'_\nu$, a contradiction.

Now suppose that S_1 is not stationary in κ but S_2 is stationary. Then $\kappa-S_1$ contains a cub so by Lemma 0.1 and the definition of stationary, there

exists $\nu \in \mathscr{C} \cap (\kappa - S_1) \cap S_2$. By definition of S_1 and S_2, $pd(M_\nu) \leqq n-1$ and there exists $\mu > \nu$ such that $pd(M_\mu) \leqq n-1$ and $pd(M_\mu/M_\nu) \geqq n$. Choose $\tau \in \mathscr{C}$ such that $\tau \geqq \mu$. Then $M_\tau/M_\nu = M'_\tau/M'_\nu$ and we shall obtain a contradiction if we prove that $pd(M_\tau/M_\nu) \geqq n$. To this end consider the short exact sequences

$$0 \to M_\mu \to M_\tau \to M_\tau/M_\mu \to 0 \tag{1}$$

$$0 \to M_\mu/M_\nu \to M_\tau/M_\nu \to M_\tau/M_\mu \to 0. \tag{2}$$

By Lemma 1.5 applied to (1), since $pd(M_\mu) \leqq n-1$ and $pd(M_\tau) = pd(M'_\tau) \leqq n-1$, we have $pd(M_\tau/M_\mu) \leqq n$. Now consider (2): by choice of μ, $pd(M_\mu/M_\nu) \geqq n$: if $pd(M_\tau/M_\nu)$ were $\leqq n-1$ we would have $pd(M_\tau/M_\mu) = n+1$ by Lemma 1.5, which is impossible; thus $pd(M_\tau/M_\nu) \geqq n$.

($\Rightarrow$) We shall prove that if S_1 and S_2 are both non-stationary then $pd(M) \leqq n-1$. By assumption there are cubs $\mathscr{C}_1$ and $\mathscr{C}_2$ such that $\mathscr{C}_l \cap S_l = \varnothing$ ($l = 1, 2$). Then $\mathscr{D} = \mathscr{C}_1 \cap \mathscr{C}_2$ is a cub (Lemma 0.1); so it is the range of a normal function $f\colon \kappa \to \kappa$. But for all ν, $pd(M_{f(\nu)}) \leqq n-1$ since $f(\nu) \notin S_1$; and $pd(M_{f(\nu+1)}/M_{f(\nu)}) \leqq n-1$ since $f(\nu) \notin S_2$ and $pd(M_{f(\nu+1)}) \leqq n-1$. Therefore by Auslander's Lemma, $pd(M) \leqq n-1$. □

It is easy to construct examples where one of S_1 and S_2 is stationary but the other is not (see 1.8(5) and (6)).

1.7. **Main Theorem** (version II). *Let κ, R, M, $\{M_\nu \mid \nu < \kappa\}$ be as above. Suppose also that for all ν, $pd(M_\nu) \leqq n-1$. Then $pd(M) \geqq n$ if and only if $\{\nu < \kappa \mid$ there exists $\mu > \nu$ such that $pd(M_\mu/M_\nu) \geqq n\}$ is stationary in κ.* □

In many natural examples one will have $pd(M_{\nu+1}/M_\nu) \geqq n$ whenever there is a $\mu > \nu$ such that $pd(M_\mu/M_\nu) \geqq n$ (cf. proof of Theorem 2.1).

Obviously, version III is an easy consequence of 1.7. The only point that has to be made is that the hypothesis on M implies that M is κ^+-presented; indeed M is the direct limit of its finitely generated submodules, $M_i (i \in I)$. If

$$0 \to K_i \to F_i \xrightarrow{\pi_i} M_i \to 0$$

is a finite presentation of M_i for each $i \in I$, and if we let $F = \bigoplus_{i \in I} F_i$ and $\pi\colon F \to M$ be induced by the π_i, then it is easy to see that F and $\ker(\pi)$ are κ^+-generated. The hypothesis on M in Version III is satisfied if R is a domain which is commutative (or more generally a left and right Ore ring) and M is torsion-free (cf. [11; Ch. II; Prop. 4.6]).

We could also give a version of the Main Theorem in which the hypotheses on M and R are replaced by the more general hypothesis that M is κ^+-resolvable (cf. [9]).

1.8. *Remarks.* (1) The special case of the Main Theorem for $R = \mathbb{Z}$ is given in [2], Lemma 2.1(2).

(2) It is essential in 1.6 that $\{M_\nu : \nu < \kappa\}$ be a κ-filtration and not simply a continuous chain (whose elements might not be κ-generated). Indeed, let $R = \mathbb{Z}$, $\kappa = \omega_1$, and let $K \subseteq F$ be free $\mathbb{Z}$-modules such that $F/K \cong Q$. For every $\mu < \omega_1$, let $K_\mu = K$ and $F_\mu = F$. Define $M'_0 = \bigoplus_{\mu < \omega_1} K_\mu$, $M'_1 = F_0 \oplus \bigoplus_{0 < \mu < \omega_1} K_\mu$, and, in gen-

eral, $M'_\nu = \bigoplus_{\mu<\nu} F_\mu \oplus \bigoplus_{\nu\leqq\mu<\omega_1} K_\mu$. Then for *all* ν, $M'_{\nu+1}/M'_\nu \cong Q$, so $pd(M_\tau/M_\nu)=1$ whenever $\nu<\tau<\omega_1$. But $M=\bigcup_{\nu<\omega_1} M'_\nu$ has projective dimension 0 since $M = \bigoplus_{\mu<\omega_1} F_\mu$.

(3) We have not defined "stationary" or "closed unbounded" for κ a singular cardinal, but the following example shows that we cannot expect a result like 1.6 for singular κ. Let $R=\mathbb{Z}$, $\kappa=\aleph_\omega$. We shall define a continuous chain $\{M'_\nu: \nu<\kappa\}$ of $\mathbb{Z}$-modules, each of cardinality $<\kappa$, such that for *all* $\mu<\nu<\kappa$, $pd(M'_\mu/M'_\nu)=1$ but $M=\bigcup_{\nu<\kappa} M'_\nu$ is free (of rank κ). Let $K\subseteq F$ be as in (1), and for all $\mu<\aleph_\omega$ let $K_\mu=K$, $F_\mu=F$. Define

$$M'_0 = \bigoplus_{\mu<\omega_1} K_\mu \oplus \bigoplus_{n=1}^{\infty} K_{\omega_n}$$

If $\nu\leqq\omega_1$, let

$$M'_\nu = \bigoplus_{\mu<\nu} F_\mu \oplus \bigoplus_{\nu\leqq\mu<\omega_1} K_\mu \oplus \bigoplus_{n=1}^{\infty} K_{\omega_n}.$$

If $\omega_m<\nu\leqq\omega_{m+1}$, let

$$M'_\nu = \bigoplus_{\mu<\nu} F_\mu \oplus \bigoplus_{\nu\leqq\mu<\omega_{m+1}} K_\mu \oplus \bigoplus_{n=m+1}^{\infty} K_{\omega_n}.$$

Thus $|M'_\nu|=\aleph_{m+1}$. For all $\nu<\aleph_\omega$, $M'_{\nu+1}/M'_\nu$ is not free but

$$M = \bigoplus_{\mu<\aleph_\omega} F_\mu.$$

(4) We also cannot expect 1.6 to hold for $\kappa=\omega$. Let $R=\mathbb{Z}$, and $M=\bigoplus_{i<\omega} \mathbb{Z}_i$, where $\mathbb{Z}_i\cong\mathbb{Z}$. Let $M'_n=\bigoplus_{i<n}\mathbb{Z}_i \oplus 2\mathbb{Z}_n$; then for all $n<m<\omega$, $pd(M'_m/M'_n)=1$.

(5) If E is any subset of ω_1 it is easy to construct by transfinite induction an ω_1-filtration $\{M_\nu: \nu<\omega_1\}$ of free abelian groups such that for all ν, there is a μ such that $pd(M_\mu/M_\nu)\geqq 1$ if and only if $pd(M_{\nu+1}/M_\nu)\geqq 1$ if and only if $\nu\in E$ (cf. [2; Thm. 6.3]). It follows that the unstrengthened version of Auslander's Lemma (i.e. $pd(M)\geqq n \Rightarrow \{\nu | pd(M_{\nu+1}/M_\nu)\geqq n\}\neq\varnothing$) has a converse which is false. In particular, if E is non-stationary but non-empty (e.g. E can be unbounded) we have $\{\nu | pd(M_{\nu+1}/M_\nu)\geqq 1\}\neq\varnothing$ but $pd(\bigcup_\nu M_\nu)=0$.

(6) There is an ω_1-filtration $\{M_\nu | \nu<\omega_1\}$ of abelian groups such that M_0 is free, M_1 is non-free and for all $\nu>0$, $M_{\nu+1}/M_\nu$ is free. This shows that we need to consider both S_1 *and* S_2 in 1.6: in this case (for $n=1$), $S_2=\{0\}$ is non-stationary, but $pd(\bigcup_\nu M_\nu)\geqq 1$ (since $S_1=\omega_1-\{0\}$).

2. Applications

The first result giving a lower bound for homological dimension related to cardinality was proved by Kaplansky [4]. Later Pierce [10] and, especially, Osofsky proved results concerning lower bounds for projective dimension. The

original proofs by Osofsky of these results (in [5–8]) were quite different from those below and involved the construction of specific projective resolutions. In a later paper, [9] (of which we were unaware when a first version of this paper was researched and written) Osofsky gives new proofs of her old – as well as new – results by methods which are similar to those used here. However there is no explicit use of stationary sets and no explicit formulation of a converse to Auslander's Lemma. The general theorems from which she derives her specific results ([9; Theorems 4.6, 4.6′, 4.10]) are, consequently, rather technical and – especially in one case – difficult to remember or quote. We believe that our Main Theorem provides an interesting alternative way of looking at the proofs of the Osofsky results.

2.1. **Theorem** ([5]). *Let R be a (left) valuation ring and M a (left) ideal of R which is $\aleph_{m+1}$-generated but not $\aleph_m$-generated. Then $pd(M)=m+1$.*

Proof. The proof is by induction on m. The proof that $pd(M)\leqq m+1$ uses Auslander's Lemma (see e.g. [8, Prop. 2.43]). We shall concentrate on the proof of the opposite inequality, which uses the converse to Auslander's Lemma. For $m=0$ the result is easy since M is clearly not projective. Suppose that $m=k+1$; by transfinite induction we can choose a strictly increasing chain $\{M_\nu|\nu<\aleph_{k+1}\}$ of submodules of M whose union is M and is such that for all ν, $M_{\nu+1}=Ra_\nu$ for some $a_\nu\in M$. (This is possible since R is a valuation ring and hence the left ideals of R are linearly ordered by inclusion). By induction $pd(M_\nu)\leqq m$ for all ν. We claim that for all ν in $E_k\overset{\text{def}}{=}\{\nu<\aleph_{k+1}|\nu$ is a limit ordinal of cofinality $\aleph_k\}$, $pd(M_{\nu+1}/M_\nu)\geqq k+2$. If true, then since R is left coherent and E_k is stationary in $\aleph_{k+1}$ (cf. Lemma 0.2), the Main Theorem (version III) implies the desired conclusion. But if $\nu\in E_k$, then clearly M_ν is generated by $\aleph_k$ elements and no fewer, so by induction $pd(M_\nu)\geqq k+1$. But $M_{\nu+1}=Ra_\nu$ is free, so $pd(M_{\nu+1}/M_\nu)\geqq k+2$ (cf. Lemma 1.5). This concludes the proof. □

The following Theorem 2.2 is the key technical result which enables one to derive results about the global dimension of direct products of rings and free Boolean rings. For example, if $2^{\aleph_0}=\omega_n<\omega_\omega$, then the global dimension of a countable product of fields is $n+1$ (see [7]).

Before stating Theorem 2.2 we need a definition.

Definition. ([7; p. 641]). A family $A=\{e_i|i<\mu\}$ of idempotents of a ring R is *nice* if: (i) the elements of A commute with each other; and (ii) for all finite subsets B_1, B_2 of A, if $B_1\cap B_2=\varnothing$, then $\prod_{i\in B_1} e_i \prod_{j\in B_2}(1-e_j)\neq 0$.

For example, a countable product of fields contains a nice family of idempotents of size $2^{\aleph_0}$ (see [8, Prop. 2.42 and Thm. 2.51]).

2.2. **Theorem** (cf. [7]). *Let R be left coherent. Suppose $A=\{e_i|i<\omega_n\}$ is a nice family of idempotents of R. Let $M=\sum_{i<\omega_n} Re_i$. Then $pd(M)=n$.*

Proof. The proof is by induction on n. If $n=0$, M is projective since it equals $\bigoplus_{n\in\omega} Rc_n$, where $c_n=e_n\prod_{i=0}^{n-1}(1-e_i)$.

Suppose that $n\geqq 1$. Let $M_\nu=\sum_{i<\nu} Re_i$; this defines an ω_n-filtration of M. By induction $pd(M_\nu)\leqq n-1$. Let σ be such that $cf(\sigma)=\omega_{n-1}$. We claim that $pd(M_{\sigma+1}/M_\sigma)=n$. If true, then, just as in the argument above, we can conclude the proof. To prove the claim we use the short exact sequence

$$0\to\sum_{i<\sigma} Re_\sigma e_i\to Re_\sigma\to M_{\sigma+1}/M_\sigma\to 0$$

which is derived from the isomorphism of $M_{\sigma+1}/M_\sigma$ with $Re_\sigma/\sum_{i<\sigma} Re_\sigma e_i$. Since $\{e_\sigma e_i: i<\sigma\}$ is a nice family of idempotents, $pd(\sum_{i<\sigma} Re_\sigma e_i)=n-1$ by induction, so for $n>1$ we immediately obtain by Lemma 1.5 that $pd(M_{\sigma+1}/M_\sigma)=n$.

(For $n=1$ a different argument is needed, as in [8; Prop. 2.24]). □

Other results can be obtained similarly, including results of [6] and a version of Theorem 4.10 of [9].

References

1. Auslander, M.: On the dimension of modules and algebras. III. Global dimension. Nagoya Math. J. **9**, 67–77 (1955)
2. Eklof, P.C.: Set Theoretic Methods in Homological Algebra and Abelian Groups. Montreal: Les Presses de l'Université de Montréal 1980
3. Jensen, C.U.: Homological dimensions of $\aleph_0$-coherent rings. Math. Scand. **20**, 55–60 (1967)
5. Kaplansky, I.: The homological dimension of a quotient field. Nagoya Math. J. **27**, 139–142 (1966)
5. Osofsky, B.: Global dimension of valuation rings. Trans. Amer. Math. Soc. **127**, 136–149 (1967)
6. Osofsky, B.: Homological dimension and the continuum hypothesis. Trans. Amer. Math. Soc. **132**, 217–230 (1968)
7. Osofsky, B.: Homological dimension and cardinality. Trans. Amer. Math. Soc. **151**, 641–649 (1970)
8. Osofsky, B.: Homological dimensions of modules. Amer. Math. Soc. Regional Conference Series **12**. Providence, Rhode Island: American Mathematic Society 1973
9. Osofsky, B.: Projective dimension of "nice" directed unions. J. Pure Appl. Algebra **13**, 179–219 (1978)
10. Pierce, R.: The global dimension of Bodean rings. J. Algebra **7**, 91–99 (1967)
11. Stenström, B.: Rings of quotients. Berlin-Heidelberg-New York: Springer 1975

Received July 20, 1981

Math. Z. 180, 11–17 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Rationale Interpolation von $|x|$ in äquidistanten Punkten

Helmut Werner

Institut für Angewandte Mathematik der Universität, Wegelerstraße 6, D-5300 Bonn, Bundesrepublik Deutschland

Einleitung

Zahlreiche klassische Fragestellungen der Approximationstheorie sind verknüpft mit der Funktion $|x|$, genommen etwa auf $I:=[-1,+1]$. Denn sie stellt das Modell einer Lipschitz-stetigen aber nicht differenzierbaren Funktion dar. Man weiß, daß die Ordnung der Konvergenz der gleichmäßig besten Approximation mit Polynomen $2n$-ten Grades durch $O(1/n)$ gegeben ist. Das optimal approximierende Polynom interpoliert die Funktion in Punkten, die sich an den Intervallenden zusammendrängen und in der Umgebung des Knickes verhältnismäßig weit auseinanderliegen.

Newman behandelte die beste Approximation von $|x|$ im genannten Intervall mit rationalen Funktionen des Zähler- und Nennergrades n. Vjaceslavov verfeinerte diese Untersuchungen. Das Ergebnis ist, daß die Approximationsgüte wie $O(\exp(-\pi\sqrt{n}))$ konvergiert. Die Interpolationspunkte häufen sich in diesem Fall in der Umgebung der Knickstelle.

Für die bei einer Funktion, die man nicht genauer kennt, naheliegende und aus Gründen der Arbeitsersparnis praktische äquidistante Verteilung von Interpolationsknoten erhält man für $|x|$, wie bereits Bernstein zeigte, jedoch Interpolationspolynome, die für $x\in I\setminus\{-1,0,+1\}$ Polynomwerte liefern, die für $n\to\infty$ divergieren.

Es liegt nahe, mit rationaler Interpolation die gleiche Aufgabe zu lösen und nach den Konvergenzverhältnissen zu fragen. Es zeigt sich, wie hier bewiesen wird, daß bereits bei äquidistanter rationaler Interpolation mit im Zähler und Nenner gleichem geraden Polynomgrad n eine gleichmäßige Konvergenz mit $O\left(\frac{1}{n\cdot\log n}\right)$ erzielt wird, d.h. also eine bessere als für die beste Approximation mit Polynomen. Die maximalen Fehler treten in der Nähe der Knickstelle $x=0$ auf. Die Interpolationen haben also in gewisser Weise ein lokales Verhalten, denn in abgeschlossenen Teilintervallen, die die Knickstelle nicht enthalten, ist die Konvergenz sogar exponentiell, wie die folgenden Überlegungen zeigen.

0025-5874/82/0180/0011/$01.40

Dieses Beispiel unterstreicht die bekannte Erfahrung, daß rationale Funktionen besser geeignet sind als Polynome, Funktionen mit raschen Änderungen der Tangentenrichtung bzw. mit Knicken darzustellen.

Konstruktion der Interpolierenden

Zu interpolieren seien die Daten $(x_\nu, |x_\nu|)$ für $x_\nu := \nu/n$, $\nu = -n, -n+1, \ldots, n-1, n$.

Die gegebenen Interpolationsdaten sind symmetrisch bezüglich $x=0$, d.h. $f(x_\nu)=f(x_{-\nu})$ und $x_{-\nu}=-x_\nu$, $\nu=1,\ldots,n$. Ferner ist $x_0=0$ und $f(0)=0$. Also wird man mit einer symmetrischen rationalen Interpolierenden ansetzen. Sie muß den Faktor x wegen der Nullstelle bei $x=0$ enthalten. Dies zusammen führt auf den Ansatz

$$r_n(x) := x \cdot \frac{p(x)-p(-x)}{p(x)+p(-x)}$$

mit einem Polynom p vom Grade n, und wir werden sehen, $r_n(x)$ hat für gerades n den Zähler- und Nennergrad n (vgl. auch Newman). Offenbar gilt $r_n(x)=r_n(-x)$, so daß es genügt, die Interpolationsbedingungen in x_ν für $\nu=1,\ldots,n$ zu erfüllen. Aus der Gleichung

$$r_n(x_\nu)=x_\nu \quad \text{folgt} \quad 1=\frac{p(x_\nu)-p(-x_\nu)}{p(x_\nu)+p(-x_\nu)},$$

also $p(-x_\nu)=0$.
Das Polynom p hat mithin die Gestalt

$$p(x) := \prod_{\nu=1}^{n} (x+\nu/n),$$

so daß $p(x)+p(-x)$ vom Grade $2\cdot[n/2]$ und

$$p(x)-p(-x) \quad \text{vom Grade } 2\cdot[(n-1)/2]+1$$

ist. Damit ist die Interpolierende explizit angegeben. (Sie ist bekanntlich eindeutig.)

Konvergenzaussagen

Es kommt nun darauf an, die Abweichung zwischen $|x|$ und $r_n(x)$ abzuschätzen.

Satz 1. *In* $[-1, +1]$ *gilt*

$$\varepsilon_n(x) := |x|-r_n(x)=|x|-x\cdot\frac{p(x)-p(-x)}{p(x)+p(-x)}=0\left(\frac{1}{n\cdot\log n}\right)$$

mit

$$p(x) := \prod_{\nu=1}^{n} (x+\nu/n).$$

und für jedes feste $\gamma\in(0,1)$ *gilt sogar*

$$\varepsilon_n(x)=0(\exp(-\gamma n/2)) \quad \text{in } \gamma\leqq|x|\leqq 1.$$

Beweis. Im folgenden lassen wir zur Vereinfachung der Formeln den eigentlich erforderlichen Index n bei $z(x)$ und $p(x)$ fort.

Sei

$$z(x):=\frac{p(-x)}{p(x)}=\prod_{\nu=1}^{n}\left(\frac{\nu/n-x}{\nu/n+x}\right).$$

Diese Funktion muß für $x\in[0,1]$ untersucht werden.

Offensichtlich gilt

$$z(0)=1 \text{ und } |z(x)|<1 \text{ in } (0,1].$$

Für $x\in(1/n,1]$ kann man schärfer abschätzen. Für $x>\nu/n$ gilt nämlich

$$0<\frac{|\nu/n-x|}{\nu/n+x}<1-\nu/nx\leqq\exp(-\nu/nx)$$

und für $x<\nu/n$ analog

$$0<\frac{\nu/n-x}{\nu/n+x}<1-nx/\nu\leqq\exp(-nx/\nu).$$

Damit folgt

$$\begin{aligned}|z(x)|&=\prod_{\nu\leqq nx}|(\ldots)|\cdot\prod_{\nu>nx}(\ldots)\\&\leqq\exp\Big(-\sum_{\nu\leqq nx}\nu/nx-\sum_{\nu>nx}nx/\nu\Big).\end{aligned}$$

Nun kann man abschätzen, indem man einmal die arithmetische Reihe $\Sigma\nu$, zum zweiten das Integral über $1/t$ benutzt.

$$\sum_{\nu\leqq nx}\nu/nx\geqq\frac{nx-1}{2},$$

$$\begin{aligned}\sum_{\nu>nx}nx/\nu&\geqq nx\cdot\int_{nx+2}^{n}\frac{dt}{t}\qquad\left(\text{falls } x\leqq\frac{n-2}{n}\text{ ist}\right)\\&\geqq nx\log(n/(nx+2)).\end{aligned}$$

Es gilt also

$$\sum_{\nu\leqq nx}\frac{\nu}{nx}+\sum_{\nu>nx}\frac{nx}{\nu}\geqq\frac{nx-1}{2}-nx\cdot\log\left(x+\frac{2}{n}\right)=:\phi(x)\qquad\text{für } x\in\left(\frac{1}{n},1-\frac{2}{n}\right]$$

und andererseits

$$\sum_{\nu \leqq nx} \frac{\nu}{nx} + \sum_{\nu > nx} \frac{nx}{\nu} \geqq \frac{nx-1}{2} \geqq \frac{n\gamma - 1}{2}$$

gleichmäßig in $\gamma \leqq x \leqq 1$ für jedes feste $\gamma \in (0,1)$ und $n > \frac{1}{\gamma}$.

Damit folgt für $z(x)$ die Aussage

$$|z(x)| \leqq \sqrt{e} \cdot \exp\left(-\frac{n\gamma}{2}\right) \quad \text{für } x \in [\gamma, 1].$$

Nun gilt in $\left(\frac{1}{n}, \gamma\right)$ für $\gamma \leqq 1 - \frac{2}{n}$ die Abschätzung

$$\phi'(x) = \frac{n}{2} - nx \frac{1}{x + \frac{2}{n}} - n \log\left(x + \frac{2}{n}\right)$$

$$\geqq \frac{n}{2} - n - n \log\left(\gamma + \frac{2}{n}\right) = n\left(\left|\log\left(\gamma + \frac{2}{n}\right)\right| - \frac{1}{2}\right).$$

Wählt man $\gamma \leqq 0.5$, so folgt

$$|\log \gamma| \geqq |\log(0.5)| = \log 2 \geqq 0.693 > 0.5$$

und für hinreichend große n (genauer $n \geqq 19$) ist auch $\left|\log\left(\gamma + \frac{2}{n}\right)\right| > 0.5$, mithin $\phi'(x)$ positiv, d.h.

$$\phi(x) \geqq \phi\left(\frac{1}{n}\right) = \log\left(\frac{n}{3}\right) \quad \text{für } x \in \left[\frac{1}{n}, \frac{1}{2}\right],$$

$$\geqq \frac{n}{4} - \frac{1}{2} \geqq \log\left(\frac{n}{3}\right) \quad \text{für } x \in \left[\frac{1}{2}, 1\right]$$

und

$$|z(x)| \leqq \exp\left(\frac{1 - nx}{2} + nx \log\left(x + \frac{2}{n}\right)\right) \leqq \frac{3}{n} \quad \text{für } x \in \left(\frac{1}{n}, 1\right],$$

n hinreichend groß.

Nach dieser Vorbereitung erhält man die Konvergenzaussage durch folgende Überlegungen.

Es sei im folgenden stets $n \geqq 19$, so daß man auf Grund der beiden vorangegangenen Abschätzungen, $|z(x)|$ für $x \in (\frac{1}{2}, 1]$ durch $\frac{1}{2}$ abschätzen kann.

Ferner haben wir für positive x-Werte die Darstellung

$$\varepsilon_n(x) = |x| - r_n(x)$$

$$= x - x \frac{p(x) - p(-x)}{p(x) + p(-x)} = 2x \cdot \frac{p(-x)}{p(x) + p(-x)}$$

$$= 2 \cdot x \cdot \frac{z(x)}{1 + z(x)}.$$

Sei nun γ festgelegt, $\gamma\in(0,1)$, ferner $n\geqq\max\left(19,\frac{1}{\gamma}\right)$. Dann gilt in $x\in[\gamma,1]\subset\left[\frac{1}{n},1\right]$ die Abschätzung

$$|\varepsilon_n(x)|=2x\cdot\frac{|z(x)|}{1+z(x)}\leqq 4\cdot|z(x)|$$
$$\leqq 4\cdot\sqrt{e}\cdot\exp\left(-\frac{n\cdot\gamma}{2}\right),$$

also die zweite Behauptung des Satzes.

Zum Nachweis der Ersten sind nur noch die Punkte in der Nähe des Nullpunktes, etwa in $(0,\frac{1}{2})$ zu untersuchen.

Betrachten wir zunächst den Fall $x\in\left(0,\frac{1}{n}\right)$.

Dort ist $z(x)>0$, also $\varepsilon_n(x)$ positiv und

$$\varepsilon_n(x)\leqq 2x\cdot z(x).$$

Für $g(x):=x\cdot|z(x)|$ gilt in diesem Falle

$$g(x)=x\cdot\prod_{\nu=1}^{n}\left(\frac{\frac{\nu}{n}-x}{\frac{\nu}{n}+x}\right)\leqq x\cdot\prod\left(1-\frac{x\cdot n}{\nu}\right)$$
$$\leqq x\cdot\exp\left(-x\cdot n\cdot\sum_{\nu=1}^{n}\frac{1}{\nu}\right)$$
$$\leqq x\cdot\exp(-x\cdot n\cdot\log n)=:\psi(x).$$

Man sieht leicht, daß $\psi(x)$ bei $x=\frac{1}{n\cdot\log n}$ sein Maximum annimmt, so daß also

$$0<\varepsilon_n(x)\leqq 2x\cdot z(x)\leqq 2\cdot\psi\left(\frac{1}{n\cdot\log n}\right)=\frac{2\cdot\exp(-1)}{n\cdot\log n}$$

in $\left(0,\frac{1}{n}\right)$ gilt.

Sei nun $x\in\left(\frac{1}{n},\frac{1}{2}\right]$. Es ist dann

$$\frac{g(x)}{g\left(x-\frac{1}{n}\right)}=\frac{x}{x-\frac{1}{n}}\cdot\prod_{\nu=1}^{n}\left|\frac{\frac{\nu}{n}-x}{\frac{\nu}{n}+x}:\frac{\frac{\nu+1}{n}-x}{\frac{\nu-1}{n}+x}\right|$$

$$=\frac{x}{x-\frac{1}{n}}\cdot\frac{x-\frac{1}{n}}{\frac{n+1}{n}-x}\cdot\frac{x}{1+x}=\frac{x^2}{1+x}\cdot\frac{1}{1+\frac{1}{n}-x}<\frac{x^2}{1-x^2}.$$

Da nun in $\left(0, \frac{1}{n}\right)$ für $g(x)$ die Abschätzung

$$g(x) \leqq \frac{1}{e \cdot n \cdot \log n}$$

richtig ist, gilt sie, wie man durch Induktion nach ν sieht, auch in $\left(\frac{\nu}{n}, \frac{\nu+1}{n}\right)$ für jedes $\nu < \frac{n}{2}$.

Damit ist für $(0, \frac{1}{2})$ die Abschätzung gefunden,

$$|\varepsilon_n(x)| \leqq 4 \cdot g(x) \leqq \frac{4}{e \cdot n \cdot \log n}.$$

Dies verifiziert die erste Konvergenzaussage des Satzes.

Die Konvergenzordnung dieser Fehlerabschätzung läßt sich nicht verbessern.

Zur Verifikation setze man $x^* = x^*(n) = \frac{1}{n \cdot \log n}$. Dann gilt für $n \geqq 3$ zunächst $x^*(n) \in \left(0, \frac{1}{n}\right)$, also ist $z(x^*(n))$ positiv und man kann abschätzen

$$\begin{aligned} n \cdot \log n \cdot |\varepsilon_n(x^*(n))| &= \frac{\varepsilon_n(x^*)}{x^*} \\ &= \frac{2z(x^*)}{1+z(x^*)} \geqq z(x^*). \end{aligned}$$

Man benötigt also nur noch eine untere Schranke für $z(x^*(n))$. Mit $c := \log n$ gilt

$$\begin{aligned} 1 < \frac{1}{z(x^*)} &= \prod_{\nu=1}^{n} \left(\frac{\frac{\nu}{n} + \frac{1}{nc}}{\frac{\nu}{n} - \frac{1}{nc}} \right) = \prod \left(1 + \frac{2}{\nu c - 1}\right) \\ &\leqq \exp\left(\sum_{\nu=1}^{n} \frac{2}{\nu \cdot c - 1} \right). \end{aligned}$$

Die Summe kann man mit Hilfe eines Integrals majorisieren

$$\begin{aligned} \sum_1^n \frac{2}{\nu c - 1} &\leqq \frac{2}{c-1} + \int_1^n \frac{2\,dt}{tc-1} = \frac{2}{c-1} + \frac{2}{c} \log\left(\frac{n \cdot c - 1}{c-1}\right) \\ &\leqq \frac{2}{c}\left[\log n + \log\left(\frac{1 - \frac{1}{nc}}{1 - \frac{1}{c}}\right)\right] + \frac{2}{c-1} = 2 + O\left(\frac{1}{c}\right). \end{aligned}$$

Für große Werte von n gilt also mit einem $c_0 > 0$ die Abschätzung

$$\varepsilon_n(x^*(n)) \geqq \frac{z(x^*(n))}{n \cdot \log n} \geqq \frac{c_0}{n \cdot \log n}.$$

Damit ist die genaue Konvergenzordnung von $\|\varepsilon_n\|$ für $n \to \infty$ ermittelt.

Literatur

Bernstein, S.N.: Sur la meilleure approximation de |x| par des polynomes de degrés donnés.. Acta Math. **37**, 1–57 (1913)

Meinardus, G.: Approximation of Functions: Theory and Numerical Methods. Springer Tracts in Natural Philosophy 13. Berlin-Heidelberg-New York: Springer 1967

Natanson, I.P.: Konstruktive Funktionentheorie. Berlin: Adademie Verlag 1955

Newman, D.J.: Rational Approximation to |x|. Michigan Math. J. **11**, 11–14 (1964)

Rivlin, F.J.: The Chebyshev Polynomials. New York-London-Sydney-Toronto: Wiley 1974

Vjaceslavov, N.S.: On the Uniform Approximation of |x| by Rational Functions. Soviet Math. Dokl. **16**, 100–104 (1975)

Eingegangen am 5. Oktober 1981

Math. Z. 180, 19–40 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

New Inequalities from Old Ones*

William E. Kirwan[1] and Glenn Schober[2]

[1] Department of Mathematics, University of Maryland, College Park, Maryland 20742, U.S.A.
[2] Department of Mathematics, Indiana University, Bloomington, Indiana 47405, U.S.A.

1. Introduction

The purpose of this paper is to illustrate a principle which is not new, but deserves to be better known. Basically, it transforms sharp inequalities in the theory of univalent functions into new sharp inequalities by means of subordination. It is surprising how much one can accomplish by means of this elementary technique.

Let S denote the familiar class of univalent, analytic functions f in the disk $\mathbb{D}=\{|z|<1\}$, with the normalization

$$f(z)=z+a_2 z^2+\ldots, \quad z\in\mathbb{D}.$$

Similarly, Σ denotes the class of univalent, analytic functions f in $\tilde{\mathbb{D}}=\{|z|>1\}$, with the normalization

$$f(z)=z+b_0+b_1 z^{-1}+\ldots, \quad z\in\tilde{\mathbb{D}}.$$

The principle that we shall discuss applies as well to many other families, but these will suffice to demonstrate it.

The next section discusses the principle (Basic Lemma), and the following sections apply it. While some of the applications are new, for some of the others the proofs are new.

2. The Basic Lemma

We shall discuss the principle first in its simplest form.

Let λ be a real-valued functional defined on S, and assume $F\in S$ is a function that provides its maximum:

$$\lambda(F)=\max_{f\in S}\lambda(f).$$

Let G be any analytic, univalent function in $\mathbb{D}$ such that F is subordinate to G, i.e., $F\prec G$. Thus there is an analytic (and univalent) function $\omega\colon \mathbb{D}\to\mathbb{D}$ such that $\omega(0)=0$ and

$$F=G\circ\omega.$$

* This work was supported in part by grants from the National Science Foundation

0025-5874/82/0180/0019/$04.40

The function G does not have the normalization of S. However, the function

$$H=\frac{1}{c}G, \qquad c=G'(0)=1/\omega'(0),$$

is back in the family S.

Associated with λ and ω, define the functional $\tilde{\lambda}$ on S by

$$\tilde{\lambda}(f)=\lambda(cf\circ\omega).$$

If $f\in S$, then $cf\circ\omega\in S$, and it is evident that

$$\tilde{\lambda}(f)=\lambda(cf\circ\omega)\leqq\lambda(F)=\tilde{\lambda}(H).$$

Thus H provides a maximum for the new functional $\tilde{\lambda}$, and moreover, the maxima of λ and $\tilde{\lambda}$ are equal. We restate this elementary principle using the notation above.

Basic Lemma. *If F provides the maximum of a functional $\lambda: S\to\mathbb{R}$, then H provides the maximum of the associated functional $\tilde{\lambda}$. In addition, the two maxima are equal.*

In other words, the sharp inequality

$$\lambda(f)\leqq M, \qquad f\in S,$$

implies the sharp inequality

$$\tilde{\lambda}(f)\leqq M, \qquad f\in S.$$

This idea is certainly not new. Its roots go back at least to Löwner [13] and Jenkins [12], and it has been used more explicitly by Garabedian and Schiffer [6], Pfluger [15], and others. However, we believe that it should be applied more extensively. It should be as well known as the more familiar elementary transformations.

Very little changes if we consider the class Σ. One interchanges $\mathbb{D}$ with $\tilde{\mathbb{D}}$ and the origin with infinity. In particular, in this case $\omega: \tilde{\mathbb{D}}\to\tilde{\mathbb{D}}$, $\omega(\infty)=\infty$, and $c=1/\omega'(\infty)$.

We shall proceed directly to applications. Section 3 considers the class S, and Sects. 5 and 6 the class Σ. In Sect. 4 we shall describe properties of support points for both S and Σ. Finally, Sect. 7 will illustrate some constrained problems.

3. Applications in the Class S

Many extremal functions for the class S, e.g., to linear problems, map $\mathbb{D}$ onto the complement of a single analytic arc. In this case, using the notation of the previous section, F is subordinate to only a one parameter family of functions

G. Furthermore, if one starts with the *Koebe function*

$$k(z)=\frac{z}{(1-z)^2}=\sum_{n=1}^{\infty} nz^n,$$

then the normalized functions H all reduce to the Koebe function itself. Nevertheless, the functional may change so that the Basic Lemma may still produce significant results. We shall illustrate this special case first.

Basic Lemma for S and k. *If the Koebe function k provides the maximum of a functional $\lambda: S\to\mathbb{R}$, then it provides the maximum of the associated functional $\tilde{\lambda}$ and the two maxima are equal.*

The same statement is true, of course, for rotations $k_\phi(z)=e^{-i\phi}k(e^{i\phi}z)$ of the Koebe function.

For our first application let us recall a result of Golusin [7], proved by means of the Löwner differential equation.

Theorem 1 (Golusin). *If $f\in S$ and $0\leqq r<1$, then*

$$|f(r)|+|f(-r)|\leqq 2r\frac{1+r^2}{(1-r^2)^2}. \tag{3.1}$$

Since the right side of (3.1) equals $|k(r)|+|k(-r)|$, the inequality (3.1) is sharp for the Koebe function. In the notation of Sect. 2, let $F=k$. Then F is subordinate to $G=ck$ for $c\geqq 1$, and

$$\omega=(ck)^{-1}\circ k. \tag{3.2}$$

The functional associated with

$$\lambda(f)=|f(r)|+|f(-r)|$$

is

$$\tilde{\lambda}(f)=|cf(\omega(r))|+|cf(\omega(-r))|,$$

and the Basic Lemma implies the sharp inequality

$$|cf(\omega(r))|+|cf(\omega(-r))|\leqq|k(r)|+|k(-r)|. \tag{3.3}$$

Let

$$r_1=\omega(r) \quad \text{and} \quad -r_2=\omega(-r). \tag{3.4}$$

Then

$$ck(r_1)=k(r) \quad \text{and} \quad ck(-r_2)=k(-r). \tag{3.5}$$

Thus, after dividing by c, inequality (3.3) becomes

$$|f(r_1)|+|f(-r_2)|\leqq|k(r_1)|+|k(-r_2)|.$$

One verifies easily that all values r_1 and r_2 with $0\leqq r_2\leqq r_1<1$ are possible by varying r and c. Thus we have obtained in an elementary fashion the following sharp conclusion.

Theorem 2 (Jenkins). *If $f \in S$ and $0 \leqq r_2 \leqq r_1 < 1$, then*

$$|f(r_1)| + |f(-r_2)| \leqq \frac{r_1}{(1-r_1)^2} + \frac{r_2}{(1+r_2)^2}.$$

Remarks. As indicated, this theorem was first proved by Jenkins [9] several years after Golusin's work. His proof involved a nontrivial symmetrization argument.

We should point out that the Basic Lemma does not give "something for nothing". One must begin with a substantial result (e.g., Theorem 1) in order to obtain a substantial result (e.g., Theorem 2) in this elementary fashion.

It is interesting to observe, as Golusin did, that (3.1) implies

$$1 + \operatorname{Re}\{a_3 r^2\} + O(r^4) = \frac{1}{2r} \operatorname{Re}\{f(r) - f(-r)\}$$

$$\leqq \frac{1}{2r} [|f(r)| + |f(-r)|] \leqq 1 + 3r^2 + O(r^4).$$

As $r \to 0$, it follows that $\operatorname{Re} a_3 \leqq 3$.

For our second application we recall another result of Golusin [7], obtained through the Löwner differential equation. Incidentally, this result also implies $\operatorname{Re} a_3 \leqq 3$ in a similar fashion.

Theorem 3 (Golusin). *If $f \in S$ and $0 \leqq r < 1$, then*

$$|f'(r)| + |f'(-r)| \leqq \frac{2 + 12r^2 + 2r^4}{(1-r^2)^3}. \tag{3.6}$$

Since the right side of (3.6) equals $k'(r) + k'(-r)$, also this inequality is sharp for the Koebe function. Using the same subordination (3.2), we find that the functional associated with

$$\lambda(f) = |f'(r)| + |f'(-r)|$$

is

$$\tilde{\lambda}(f) = |cf'(\omega(r))\,\omega'(r)| + |cf'(\omega(-r))\,\omega'(-r)|.$$

Thus the Basic Lemma implies the sharp inequality

$$|cf'(\omega(r))\,\omega'(r)| + |cf'(\omega(-r))\,\omega'(-r)| \leqq k'(r) + k'(-r). \tag{3.7}$$

In order to simplify this inequality we differentiate (3.2) and use the notation (3.4):

$$\omega'(r) = \frac{k'(r)}{ck'(r_1)}, \qquad \omega'(-r) = \frac{k'(-r)}{ck'(-r_2)}.$$

In addition, one has

$$\frac{k'(r)}{k'(-r)} = \left[\frac{k(r)}{k(-r)}\right]^2 = \left[\frac{k(r_1)}{k(-r_2)}\right]^2$$

with the help of (3.5). Consequently, (3.7) becomes

$$|f'(r_1)|\frac{k(r_1)^2}{k'(r_1)}+|f'(-r_2)|\frac{k(-r_2)^2}{k'(-r_2)}\leqq k(r_1)^2+k(-r_2)^2.$$

In summary, we have the following sharp result, which appears to be new.

Theorem 4. *If $f\in S$ and $0\leqq r_2\leqq r_1<1$, then*

$$\frac{r_1^2|f'(r_1)|}{1-r_1^2}+\frac{r_2^2|f'(-r_2)|}{1-r_2^2}\leqq\frac{r_1^2}{(1-r_1)^4}+\frac{r_2^2}{(1+r_2)^4}. \tag{3.8}$$

Since the elementary inequality $|r_2f'(-r_2)/f(-r_2)|\geqq(1-r_2)/(1+r_2)$ is sharp also for the Koebe function, inequality (3.8) implies the following sharp corollary.

Corollary. *If $f\in S$ and $0\leqq r_2\leqq r_1<1$, then*

$$\frac{r_1^2|f'(r_1)|}{1-r_1^2}+\frac{r_2|f(-r_2)|}{(1+r_2)^2}\leqq\frac{r_1^2}{(1-r_1)^4}+\frac{r_2^2}{(1+r_2)^4}.$$

By rotation one can extend the inequalities of Theorems 1 to 4 and the Corollary to two points on any diameter.

For our final application of this section it will be convenient to write the function (3.2) as

$$\omega(z)=1+\frac{c(1-z)^2}{2z}\left[1-\sqrt{1+\frac{4z}{c(1-z)^2}}\right]$$
$$=\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k+1}\binom{2k}{k}\left[\frac{z}{c(1-z)^2}\right]^k=\sum_{k=1}^{\infty}\omega_k(c)\,z^k$$

where

$$\omega_k(c)=\sum_{j=1}^{k}\frac{(-1)^{j+1}}{j+1}\binom{2j}{j}\binom{j+k-1}{2j-1}c^{-j}.$$

In particular,

$$\begin{aligned}
&\omega_1(c)=c^{-1}\\
&\omega_2(c)=2c^{-2}(c-1)\\
&\omega_3(c)=c^{-3}(c-1)(3c-5)\\
&\omega_4(c)=2c^{-4}(c-1)(2c^2-8c+7)\\
&\omega_5(c)=c^{-5}(c-1)(5c^3-35c^2+70c-42)\\
&\omega_6(c)=2c^{-6}(c-1)(3c^4-32c^3+108c^2-144c+66).
\end{aligned} \tag{3.9}$$

If $f(z)=z+\sum_{n=2}^{\infty}a_nz^n$ belongs to S, then

$$cf\circ\omega(z)=c\,\omega(z)+c\sum_{n=2}^{\infty}a_n\omega(z)^n=z+\sum_{n=2}^{\infty}\alpha_n(c)\,z^n$$

where

$$\alpha_n(c)=c\sum_{k_1+\ldots+k_\nu=n}a_\nu\omega_{k_1}(c)\ldots\omega_{k_\nu}(c). \tag{3.10}$$

Using this notation, the following is an immediate consequence of the Basic Lemma.

Theorem 5. *Let $p(a_2, \ldots, a_n)$ be a polynomial such that the Koebe function provides a maximum for* $\operatorname{Re} p$ *over S. Then*

$$\operatorname{Re} p(\alpha_2(c), \alpha_3(c), \ldots, \alpha_n(c)) \leqq \operatorname{Re} p(2, 3, \ldots, n) \tag{3.11}$$

for all $c \geqq 1$, and this inequality is sharp (for each c).

As an example, let us consider specifically the polynomial $p(a_2, \ldots, a_n) = a_n$. Since the Bieberbach Conjecture is known to be true for $2 \leqq n \leqq 6$, Theorem 5 implies

$$\operatorname{Re} \alpha_n(c) \leqq n \quad \text{for } 2 \leqq n \leqq 6 \text{ and } c \geqq 1. \tag{3.12}$$

By substituting the formulas (3.9) and (3.10) into (3.12), we obtain after some calculations the following sharp inequalities for $c \geqq 1$.

$$\begin{aligned}
n=2&: \operatorname{Re} a_2 \leqq 2\\
n=3&: \operatorname{Re} a_3 + 4(c-1)\operatorname{Re} a_2 \leqq 8c-5\\
n=4&: \operatorname{Re} a_4 + 6(c-1)\operatorname{Re} a_3 + 2(c-1)(5c-7)\operatorname{Re} a_2 \leqq 20c^2-30c+14\\
n=5&: \operatorname{Re} a_5 + 8(c-1)\operatorname{Re} a_4 + 3(c-1)(7c-9)\operatorname{Re} a_3\\
&\quad + 4(c-1)(5c^2-16c+12)\operatorname{Re} a_2 \leqq 40c^3-105c^2+112c-42\\
n=6&: \operatorname{Re} a_6 + 10(c-1)\operatorname{Re} a_5 + 4(c-1)(9c-11)\operatorname{Re} a_4\\
&\quad + 2(c-1)(20c^2-80c+55)\operatorname{Re} a_3\\
&\quad + (c-1)(35c^3-189c^2+315c-165)\operatorname{Re} a_2\\
&\leqq 70c^4-280c^3+504c^2-420c+132.
\end{aligned}$$

The inequalities for $n=2$ and $n=3$ are trivial ones. The inequality for $n=4$ is significant for $1<c<\frac{7}{5}$ (cf. [4, 5]). For $n=5$ the inequality is of interest for $1<c<2$. For $n=6$ the inequality is nontrivial for $1<c<2+\frac{1}{2}\sqrt{5}$.

Inequalities such as (3.11) were first observed by Friedland and Schiffer [4, 5] by more profound methods. Of course, replacing $f(z)$ by $e^{-i\phi} f(e^{i\phi} z)$ leads to slightly more general inequalities.

If $f \in S$, then its inverse function f^{-1} has an expansion

$$f^{-1}(w) = w + \sum_{n=2}^{\infty} \gamma_n w^n$$

in a neighborhood of $w=0$. Löwner [13] proved that the coefficients γ_n are dominated by the corresponding coefficients of $-k^{-1}(-w)$. Since each γ_n is a polynomial in $a_2, \ldots, a_n$, one can proceed as in Theorem 5. However, the resulting inequalities all turn out to be trivial.

4. Support Points

A *support point* of S is a function $f \in S$ that solves a nontrivial linear extremal problem over S, i.e.,

$$\operatorname{Re} L(f) = \max_{S} \operatorname{Re} L$$

where L is a continuous linear functional on the space of all analytic functions in $\mathbb{D}$ endowed with the topology of locally uniform convergence.

It is known (see [15, 1, 17]) that each support point of S maps $\mathbb{D}$ onto the complement of a single analytic arc Γ extending to ∞. If one truncates the arc Γ to obtain an arc Γ_t, then the Basic Lemma implies that the mappings g_t onto $\mathbb{C}\setminus\Gamma_t$ with $g_t(0)=0$ have the property that $h_t=\dfrac{1}{g_t'(0)}g_t$ is again a support point of S. This uses the observation that if $\lambda=\operatorname{Re}L$ where L is linear, then $\hat{\lambda}=\operatorname{Re}L_t$ where L_t is linear. Since Γ has an asymptote at ∞, it follows that each support point of S is arcwise connected within the set of support points to a rotation of the Koebe function. Consequently, the set of support points of S is arcwise connected. These observations are due to Pfluger [15]. In this section we shall make similar observations for the class Σ and its relatives.

Since truncating and normalizing the boundary slit Γ of a support point leads again to a support point, an interesting, natural question is to what extent can one lengthen Γ and still generate a support point. This question has been pursued by Brickman (unpublished).

The slit Γ of a support point of S has the $\pi/4$-property; that is, the tangent to Γ at each point makes an angle at most $\pi/4$ with the radial direction. That $\pi/4$ is sharp has recently been shown by Pearce [14] by showing that the functions

$$f_{\xi,\eta}(z)=\frac{z-\frac{1}{2}(\xi+\eta)z^2}{(1-\eta z)^2}$$

with $|\xi|=|\eta|=1$ and $0<|\arg(-\xi/\eta)|\leqq\pi/4$ are support points of S. The angle $\pi/4$ is assumed at the finite tip of Γ when $\arg(-\xi/\eta)=\pm\pi/4$. In fact, that the functions $f_{\xi,\eta}$ are support points for $0<|\arg(-\xi/\eta)|<\pi/4$ could be deduced from the Basic Lemma and the cases $\arg(-\xi/\eta)=\pm\pi/4$.

Support points for Σ are defined just as for S, with $\tilde{\mathbb{D}}$ in place of $\mathbb{D}$. Although the family Σ is not compact, it has many support points. However, not every continuous linear functional L has the property that $\operatorname{Re}L$ assumes a maximum on Σ. Nevertheless, the associated linear functional M defined by

$$M(h)=L(h)-L(1)\frac{1}{2\pi}\int_0^{2\pi}h(\operatorname{Re}^{i\theta})\,d\theta \qquad (R>1)$$

does have the property that $\operatorname{Re}M$ assumes a maximum on Σ. In fact, $\operatorname{Re}L$ can assume a maximum on Σ if and only if $L(1)=0$.

Each support point f of Σ maps $\tilde{\mathbb{D}}$ onto the complement of a compact set Γ which lies along the trajectories of the quadratic differential

$$L\left(\frac{1}{f-w}\right)dw^2 \tag{4.1}$$

(see [17, Corollary 10.15]). The expression $Q(w)=L\left(\dfrac{1}{f-w}\right)$ can be extended to an analytic function in a neighborhood of Γ and does not vanish identically (otherwise, L would be trivial on Σ). Therefore $Q(w)$ has at most finitely many zeros on Γ. The only points at which Γ can split or fail to be analytic are at

these zeros. It follows that the continuum Γ consists of finitely many analytic arcs.

In contrast to the class S, the slit Γ actually can split for a support point of Σ. This is the case for the second-coefficient problem in Σ. From the point of view of the Basic Lemma the class Σ is a more fertile ground than the class S for constructing new linear inequalities from old ones. This is because in the class S a support point admits only one subordination chain. In the class Σ there may be many. Before giving examples in Sects. 5 and 6, let us note the analogue of one of Pfluger's results.

Theorem 6. *The set of support points of Σ is arcwise connected.*

Proof. Let $f(z)=z+\sum_{n=0}^{\infty} b_n z^{-n}$ be a support point of Σ corresponding to the linear functional L, let $\Gamma=\mathbb{C}\setminus f(\tilde{\mathbb{D}})$, and let $p\in\Gamma$ be a point of analyticity of Γ. Since $L(1)=0$, we may by letting $b_0\to 0$ connect f to a support point for which $b_0=0$. So assume $b_0=0$.

For $0\leqq t\leqq 1$ let Γ_t be subcontinua of Γ that decrease continuously from $\Gamma_0=\Gamma$ to $\Gamma_1=\{p\}$. For $0\leqq t<1$ let $f_t(z)=b(t)z+\sum_{n=0}^{\infty} b_n(t)z^{-n}$ map $\tilde{\mathbb{D}}$ onto $\mathbb{C}\setminus\Gamma_t$. By the Basic Lemma (for Σ) the function $\frac{1}{b(t)}f_t$ is a support point of Σ, and we may again assume that $b_0(t)=0$. Since Γ has a tangent at p, the support points $\frac{1}{b(t)}f_t$ converge as $t\to 1$ to some rotation of

$$\kappa(z)=z+z^{-1}.$$

Since the rotation $\kappa_\phi(z)=e^{-i\phi}\kappa(e^{i\phi}z)$ solves the problem $\max_{\Sigma}\operatorname{Re}\{e^{2i\phi}b_1\}$, it is a support point of Σ. Furthermore, these rotations obviously form an arcwise connected set of support points. Consequently, an arbitrary support point is arcwise connected within the set of support points to κ.

Let $\Sigma'=\{f\in\Sigma: b_0=0\}$ and $\Sigma_0=\{f\in\Sigma: f \text{ omits } 0\}$. Both Σ' and Σ_0 are compact subsets of Σ. Straightforward modifications of the previous proof imply the following.

Corollary. *The sets of support points for Σ' and Σ_0 are arcwise connected.*

Chang, Schiffer, and Schober [2] very recently have used the Basic Lemma for Σ and the fact that $\max_{\Sigma}|b_n|>\frac{2}{n+1}$ for all $n\geqq 3$ (which they prove) to give restrictions on the splitting of Γ for a support point.

5. Linear Combinations of b_2 and b_1

In the following theorem we apply the Basic Lemma to the estimate $\operatorname{Re} b_2\leqq 2/3$ for Σ.

Theorem 7. *If* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+\dots$ *belongs to* Σ *and* $\mu=\frac{2}{3}(\alpha+\beta+\gamma)$ *where* $|\alpha|=|\beta|=|\gamma|=1$ *and* $\alpha\beta\gamma=1$, *then*

$$\operatorname{Re}\{b_2+\mu b_1\}\leqq\tfrac{2}{3}+\tfrac{1}{2}|\mu|^2-\tfrac{1}{12}\operatorname{Re}\{\mu^3\} \tag{5.1}$$

and for each α, β, γ *the inequality is sharp.*

Proof. We will use the result of Schiffer [16] that $\operatorname{Re} b_2\leqq 2/3$ with equality for

$$F(z)=z(1+z^{-3})^{2/3}.$$

The function F maps $\mathbb{D}$ onto the complement of three segments of equal length emanating from the origin in the directions $e^{2\pi ik/3}$, $k=0, 1, 2$.

If α, β, γ are complex numbers of modulus one and if $\alpha\beta\gamma=1$, then the function

$$H(z)=z[(1+\alpha/z)(1+\beta/z)(1+\gamma/z)]^{2/3} \tag{5.2}$$

maps $\mathbb{D}$ onto the complement of three (possibly degenerate) segments of various lengths, but in the same directions as F. Consequently, F is subordinate to cH for a sufficiently small positive value c. Let

$$\omega=(cH)^{-1}\circ F.$$

Then the Basic Lemma implies that the real part of the second coefficient of $cf\circ\omega$ is bounded by 2/3 for every $f\in\Sigma$. Therefore we need to compute some coefficients. If $\mu=\frac{2}{3}(\alpha+\beta+\gamma)$, one verifies easily the following:

$$\begin{aligned}
F(z)&=z+\tfrac{2}{3}z^{-1}+\dots\\
H(z)&=z+\mu+(\bar\mu-\tfrac{1}{4}\mu^2)z^{-1}+(\tfrac{2}{3}-\tfrac{1}{2}|\mu|^2+\tfrac{1}{6}\mu^3)z^{-2}+\dots\\
H^{-1}(w)&=w-\mu-(\bar\mu-\tfrac{1}{4}\mu^2)w^{-1}-(\tfrac{2}{3}+\tfrac{1}{2}|\mu|^2-\tfrac{1}{12}\mu^3)w^{-2}+\dots\\
\omega(z)&=c^{-1}z-\mu-(\bar\mu-\tfrac{1}{4}\mu^2)cz^{-1}-(\tfrac{2}{3}-\tfrac{2}{3}c^{-3}+\tfrac{1}{2}|\mu|^2-\tfrac{1}{12}\mu^3)c^{-2}z^2+\dots\\
f(z)&=z+b_0+b_1z^{-1}+b_2z^{-2}+\dots\\
cf\circ\omega(z)&=z+(b_0-\mu)c+(b_1-\bar\mu+\tfrac{1}{4}\mu^2)c^2z^{-1}\\
&\quad+(b_2+\mu b_1-\tfrac{2}{3}+\tfrac{2}{3}c^{-3}-\tfrac{1}{2}|\mu|^2+\tfrac{1}{12}\mu^3)c^3z^{-2}+\dots.
\end{aligned}$$

Therefore, we have

$$c^3\operatorname{Re}\{b_2+\mu b_1-\tfrac{2}{3}+\tfrac{2}{3}c^{-3}-\tfrac{1}{2}|\mu|^2+\tfrac{1}{12}\mu^3\}\leqq\tfrac{2}{3},$$

and this inequality is equivalent to (5.1). It is sharp for the function H.

Remarks. By replacing $f(z)$ by $e^{-i\phi}f(e^{i\phi}z)$ one obtains an obvious extension of Theorem 7. For example, by replacing $f(z)$ by $-f(-z)$ one finds the same bounds for the functional $|\operatorname{Re} b_2|+\operatorname{Re}\{\mu b_1\}$.

The set M of values μ generated in Theorem 7 fill out a hypocycloid with three cusps at the points $2e^{2\pi ik/3}$, $k=0, 1, 2$, together with its interior. In particular, for $\alpha=\bar\beta$ and $\gamma=1$ one obtains the sharp inequality

$$\operatorname{Re}\{b_2+\mu b_1\}\leqq\tfrac{2}{3}+\tfrac{1}{2}\mu^2-\tfrac{1}{12}\mu^3 \quad \text{for} \quad -\tfrac{2}{3}\leqq\mu\leqq 2. \tag{5.3}$$

This inequality has been derived by Jenkins [11] by means of his General Coefficient Theorem. The special case $\mu=2$ was observed by Garabedian and Schiffer [6] by our method.

The values $\mu=\rho e^{2\pi ik/3}$, $k=0, 1, 2$, lie in the set M for $-\frac{2}{3}\leqq\rho\leqq 2$. However, we may obtain a sharp bound also for $\rho>2$ by using the inequality

$$\begin{aligned}\operatorname{Re}\{b_2+\rho e^{2\pi ik/3}b_1\}&=\operatorname{Re}\{b_2+2e^{2\pi ik/3}b_1\}+(\rho-2)\operatorname{Re}\{e^{2\pi ik/3}b_1\}\\&\leqq 2+(\rho-2)\cdot 1=\rho.\end{aligned}$$

Equality occurs precisely for the functions

$$f(z)=z+b_0+e^{-2\pi ik/3}z^{-1}.$$

Corollary. *If $k=0, 1, 2$ and $\rho\geqq 2$, then*

$$\operatorname{Re}\{b_2+\rho e^{2\pi ik/3}b_1\}\leqq\rho$$

and this inequality is sharp.

To study the linear problem

$$\max_{\Sigma}\operatorname{Re}\{b_2+\mu b_1\} \tag{5.4}$$

for the remaining values of μ not in the set M, one may return to the corresponding quadratic differential (4.1). For this functional it takes the form

$$(w-b_0+\mu)\,dw^2. \tag{5.5}$$

If f is an extremal function, then $\Gamma=\mathbb{C}\setminus f(\hat{\mathbb{D}})$ lies along the trajectories of this quadratic differential. If Γ is not a single analytic arc, then it is necessary that the coefficient $w-b_0+\mu$ vanish at a point of Γ. By a translation we may assume that $w=0$ is such a point. In this case $b_0=\mu$, and the quadratic differential is simply

$$w\,dw^2.$$

The trajectories of this quadratic differential that contain $w=0$ are just the rays in the directions $e^{2\pi ik/3}$, $k=0, 1, 2$. Consequently, f must be one of the functions (5.2) in the proof of Theorem 7, in which case $\mu=b_0$ is restricted to the set M. For μ not in the set M, we conclude that each extremal function for the problem (5.4) maps onto the complement of a *single* analytic arc Γ lying along a trajectory of the quadratic differential (5.5). In particular, Γ does not split.

6. Higher Coefficients in Σ

In this section we shall study how the Basic Lemma applies to some inequalities involving the coefficient b_3 for Σ.

Let us assume that the odd function

$$F(z)=z+B_1z^{-1}+B_3z^{-3}+\ldots$$

is an extremal function and that F is subordinate to the function

$$G(z)=c[z+C_0+C_1 z^{-1}+C_2 z^{-2}+C_3 z^{-3}+\ldots] \tag{6.1}$$

where

$$H(z)=z+C_0+C_1 z^{-1}+C_2 z^{-2}+C_3 z^{-3}+\ldots \tag{6.2}$$

belongs to Σ. Then $F=G\circ\omega$ where

$$\omega(z)=G^{-1}\circ F(z)=c^{-1}z-C_0+(c^{-1}B_1-cC_1)z^{-1}-c^2(C_2+C_0C_1)z^{-2}$$
$$+[c^{-1}B_3+cB_1C_1-c^3(C_3+2C_0C_2+C_0^2C_1+C_1^2)]z^{-3}+\ldots.$$

If

$$f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$$

is an arbitrary function in Σ, then

$$cf\circ\omega(z)=z+c(b_0-C_0)+[c^2(b_1-C_1)+B_1]z^{-1}$$
$$+c^3(b_2+C_0b_1-C_0C_1-C_2)z^{-2}+[c^4(b_3+2C_0b_2+C_0^2b_1$$
$$+C_1b_1-C_0^2C_1-2C_0C_2-C_1^2-C_3)-c^2B_1(b_1-C_1)+B_3]z^{-3}+\ldots. \tag{6.3}$$

First we start with the sharp inequality

$$\operatorname{Re}b_3\geqq-\tfrac{1}{2}-e^{-6} \tag{6.4}$$

proved by Garabedian and Schiffer [6]. This is more convenient than the corresponding similar upper estimate since the extremal functions

$$F(z)=z+4e^{-3}z^{-1}-(\tfrac{1}{2}+e^{-6})z^{-3}+\ldots \tag{6.5}$$

and $-iF(iz)$ have real coefficients. For our considerations it will be no loss of generality to use only (6.5) and to ignore translations. If we apply the Basic Lemma with the functional $\lambda(f)=-\operatorname{Re}b_3$, then we obtain the new inequality

$$-\tilde{\lambda}(f)=-\lambda(cf\circ\omega)$$
$$=\operatorname{Re}\{c^4(b_3+2C_0b_2+C_0^2b_1+C_1b_1-C_0^2C_1-2C_0C_2-C_1^2-C_3)$$
$$-c^2B_1(b_1-C_1)+B_3\}\geqq-\tfrac{1}{2}-e^{-6}$$

with $B_1=4e^{-3}$ and $B_3=-\frac{1}{2}-e^{-6}$. We simplify and summarize.

Theorem 8. *Suppose that the function* (6.5) *is subordinate to the function* (6.1). *If* $f(z)=z+b_0+b_1z^{-1}+b_2z^{-2}+b_3z^{-3}+\ldots$ *is any function in* Σ, *then*

$$\operatorname{Re}\{c^4(b_3+2C_0b_2+C_0^2b_1+C_1b_1)-4c^2e^{-3}b_1\}$$
$$\geqq\operatorname{Re}\{c^4(C_3+2C_0C_2+C_0^2C_1+C_1^2)-4c^2e^{-3}C_1\} \tag{6.6}$$

and this inequality is sharp for the function (6.2).

Jenkins [11] based a proof of the coefficient result (6.4) on the following sharp inequalities.

Theorem 9 (Jenkins). *If* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$ *belongs to* Σ *and* $0\leqq\tau\leqq 1$, *then*

$$\operatorname{Re}\{b_3+\tfrac{1}{2}b_1^2-4\tau b_1\}\geqq -\tfrac{1}{2}-3\tau^2+2\tau^2\log\tau. \tag{6.7}$$

To include the case $\tau=0$, ambiguous statements should be interpreted as the obvious limits.

Ignoring translations, the extremal functions for (6.7) are of the form

$$\begin{aligned} F(z)&=z+(\tau-\tau\log\tau+i\rho)z^{-1}\\ &\quad+[\tfrac{1}{2}(\tau^2-\rho^2-1)-\tau^2\log\tau-\tfrac{1}{2}\tau^2(\log\tau)^2+i\rho\tau(1+\log\tau)]\,z^{-3}+\ldots \end{aligned} \tag{6.8}$$

for any real constant ρ with

$$|\rho|\leqq\sqrt{1-\tau^2}-\tau\arccos\tau. \tag{6.9}$$

If we apply the Basic Lemma in the framework of this section to the functional $\lambda(f)=-\operatorname{Re}\{b_3+\frac{1}{2}b_1^2-4\tau b_1\}$, then we obtain the sharp inequality

$$\begin{aligned} -\tilde{\lambda}(f)&=-\lambda(cf\circ\omega)\\ &=\operatorname{Re}\{c^4(b_3+2C_0b_2+C_0^2b_1+C_1b_1-C_0^2C_1-2C_0C_2-C_1^2-C_3)\\ &\quad -c^2B_1(b_1-C_1)+B_3+\tfrac{1}{2}[c^2(b_1-C_1)+B_1]^2-[c^2(b_1-C_1)+B_1]\}\\ &\geqq -\tfrac{1}{2}-3\tau^2+2\tau^2\log\tau \end{aligned}$$

where B_1 and B_3 are defined by (6.8). Again, we simplify and summarize.

Theorem 10. *Suppose that the function* (6.8) *is subordinate to the function* (6.1) *and that* $0\leqq\tau\leqq 1$. *If* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$ *is any function in* Σ, *then*

$$\begin{aligned} &\operatorname{Re}\{c^4(b_3+2C_0b_2+\tfrac{1}{2}b_1^2+C_0^2b_1)-4c^2\tau b_1\}\\ &\quad\geqq\operatorname{Re}\{c^4(C_3+2C_0C_2+\tfrac{1}{2}C_1^2+C_0^2C_1)-4c^2\tau C_1\} \end{aligned} \tag{6.10}$$

and this inequality is sharp for the function (6.2).

This theorem is a special case of a result of Jenkins [11, Corollary 12]; however, our point of view is different. Although Theorems 8 and 10 were based on explicit inequalities, both inequalities (6.6) and (6.10) are more implicit. Let us seek to determine more explicit values for c, C_0, C_1, C_2, C_3.

Both extremal functions (6.5) and (6.8) for the inequalities (6.4) and (6.7), respectively, map $\mathbb{D}$ onto the complement of arcs which lie on critical trajectories of the quadratic differential

$$(4\tau-w^2)\,dw^2 \tag{6.11}$$

(cf. [11]). For the function (6.8) the τ is the same, and for the function (6.5) the τ is e^{-3}.

In order for one of these extremal functions to be subordinate to a function G, it is necessary that G map $\mathbb{D}$ onto the complement of subarcs that lie on the same trajectories. It will be useful to determine such mappings.

The trajectories of (6.11) are (maximal) analytic arcs $w=w(t)$ which satisfy

$$(4\tau-w(t)^2)\left(\frac{dw}{dt}\right)^2>0. \tag{6.12}$$

The critical trajectories join pairs of the critical points $-2\sqrt{\tau}$, $2\sqrt{\tau}$, ∞. Evidently the real interval $(-2\sqrt{\tau}, 2\sqrt{\tau})$ is such a trajectory. There are four more. One joins $2\sqrt{\tau}$ to ∞ in the first quadrant, and a second is its reflection in the real axis. Their reflections in the imaginary axis provide the remaining two.

Assume that g maps $\hat{\mathbb{D}}$ onto the complement of subarcs of the critical trajectories of (6.11) and that

$$g(z)=cz+c_0+c_1z^{-1}+c_2z^{-2}+c_3z^{-3}+\ldots, \qquad c>0. \tag{6.13}$$

If one parametrizes the boundary arcs by $w=g(e^{i\theta})$, then the expression (6.12) becomes

$$[4\tau-g(e^{i\theta})^2][ie^{i\theta}g'(e^{i\theta})]^2>0.$$

That is,

$$q=(g^2-4\tau)(zg')^2$$

is analytic in $\hat{\mathbb{D}}$, is real and nonnegative on $|z|=1$, and has a pole of order four at infinity. By the Schwarz reflection principle we may represent

$$q(z)=\hat{\rho}\overline{\alpha\beta\gamma\delta}\,z^{-4}(z-r\alpha)(z-\alpha/r)(z-s\beta)(z-\beta/s)(z-t\gamma)(z-\gamma/t)(z-u\delta)(z-\delta/u)$$

where α, β, γ, δ are complex numbers of modulus one and $\hat{\rho}$, r, s, t, u are positive real numbers.

In general, the boundary slit $\Gamma=\mathbb{C}\setminus g(\hat{\mathbb{D}})$ has from two to four tips, which correspond to points where g' vanishes. At these points of $|z|=1$ the function q has at least a double zero. We shall restrict our attention to the situation where Γ has four tips, where Γ has three tips with one at a critical point $\pm 2\sqrt{\tau}$, or where $\Gamma=[-2\sqrt{\tau}, 2\sqrt{\tau}]$. In other words, we shall assume that Γ contains the critical trajectory $[-2\sqrt{\tau}, 2\sqrt{\tau}]$. The other situations can be handled in a similar fashion, but they lead in general to less explicit formulas. We have made this assumption so that $r=s=t=u=1$ and q is a perfect square:

$$[g(z)^2-4\tau][zg'(z)]^2=q(z)=\hat{\rho}\overline{\alpha\beta\gamma\delta}\,z^{-4}(z-\alpha)^2(z-\beta)^2(z-\gamma)^2(z-\delta)^2. \tag{6.14}$$

Furthermore, since Γ contains the segment $[-2\sqrt{\tau}, 2\sqrt{\tau}]$, we have $c\geqq\sqrt{\tau}$.

Next, substitute the expansion (6.13) into (6.14), and compare like powers of z to find

$$\begin{aligned}
c^4&=\hat{\rho}\overline{\alpha\beta\gamma\delta},\\
2c^3c_0&=-2c^4(\alpha+\beta+\gamma+\delta),\\
c^2(c_0^2-4\tau)&=c^4[2(\alpha+\beta+\gamma+\delta)^2-(\alpha^2+\beta^2+\gamma^2+\delta^2)],\\
-2c^2(c_0c_1+cc_2)&=2c^4[\alpha^3+\beta^3+\gamma^3+\delta^3-(\alpha+\beta+\gamma+\delta)\\
&\qquad\cdot(\alpha^2+\beta^2+\gamma^2+\delta^2)-4\alpha\beta\gamma\delta\overline{(\alpha+\beta+\gamma+\delta)}],
\end{aligned}$$

$$-2c[2c^2 c_3 + c(3c_0 c_2 + c_1^2) + c_1(c_0^2 - 4\tau)]$$
$$= \tfrac{1}{2}c^4[(\alpha^2+\beta^2+\gamma^2+\delta^2)^2 - (\alpha^4+\beta^4+\gamma^4+\delta^4)$$
$$+ 8\alpha\beta\gamma\delta|\alpha+\beta+\gamma+\delta|^2]. \tag{6.15}$$

These are useful relations between the coefficients of g and the parameters α, β, γ, δ. In particular, since $c>0$, the first one implies that

$$\alpha\beta\gamma\delta = 1 \quad \text{and} \quad \hat{\rho} = c^4.$$

It will be convenient to introduce the notation

$$\mu_j = \alpha^j + \beta^j + \gamma^j + \delta^j \quad \text{for } j = 1, 2, 3, \ldots.$$

Then the conditions (6.15) imply successively

$$\begin{aligned} c_0 &= -c\mu_1 \\ 4\tau &= c^2(\mu_2 - \mu_1^2) \\ c_2 &= c(\mu_1^3 - \mu_3 + 4\bar{\mu}_1) + (c_1 + 4\tau c^{-1})\mu_1 \\ c_3 &= c(\tfrac{1}{8}\mu_4 + \tfrac{11}{8}\mu_1^4 - \tfrac{3}{2}\mu_1\mu_3 + 5|\mu_1|^2) + 5\tau c^{-1}\mu_1^2 - 2\tau^2 c^{-3} \\ &\quad + c_1(\mu_1^2 + 2\tau c^{-2} - \tfrac{1}{2}c^{-1}c_1). \end{aligned} \tag{6.16}$$

For the latter two to be useful, we must still find a representation for c_1.

Rewrite the equation (6.14) in the form

$$(w^2 - 4\tau)\,dw^2 = q(z)\left(\frac{dz}{z}\right)^2$$

and integrate:

$$\int^w \sqrt{\omega^2 - 4\tau}\,d\omega = \int^z \sqrt{q(\zeta)}\,\frac{d\zeta}{\zeta}, \qquad w = g(z). \tag{6.17}$$

In this case both sides are elementary. If we choose

$$W = \tfrac{1}{2}w\sqrt{w^2-4\tau} - 2\tau \log\frac{w+\sqrt{w^2-4\tau}}{2\sqrt{\tau}} \tag{6.18}$$

for the left side, then the right side is

$$Z = \tfrac{1}{2}c^2(z^2 - z^{-2}) - c^2(\mu_1 z - \bar{\mu}_1 z^{-1}) - 2\tau\log z + K \tag{6.19}$$

where K is the constant of integration. Thus equation (6.17) has become

$$W(g(z)) = Z(z). \tag{6.20}$$

First, let us study the mapping (6.18). If we delete the segment $[-2\sqrt{\tau}, 2\sqrt{\tau}]$ from the plane, then $\sqrt{w^2-4\tau}$ will be single valued, and we specify the branch for which $\operatorname{Im} w \cdot \operatorname{Im}\sqrt{w^2-4\tau} \geqq 0$. However, the logarithm is still multivalued. So for each integer n let H_n^+ and H_n^- denote the upper and lower half planes, respectively. Attach each H_n^+ to H_n^- along the negative real axis from

$-\infty$ to $-2\sqrt{\tau}$, and attach each H_n^- to H_{n+1}^+ along the positive real axis from $2\sqrt{\tau}$ to $+\infty$. Then the mapping (6.18) is single valued on the resulting surface R_w.

The mapping (6.18) carries the halfplanes H_n^+, H_n^- onto the exteriors of semi-infinite strips. We index them so that the mapping carries H_n^+ onto the domain Ω_n^+ which is the complement of the semi-infinite strip

$$\{W: 0 \leqq \operatorname{Re} W < \infty,\ -(4n+2)\pi\tau \leqq \operatorname{Im} W \leqq -4n\pi\tau\}$$

and H_n^- onto the domain Ω_n^- which is the complement of the semi-infinite strip

$$\{W: 0 \leqq \operatorname{Re} W < \infty,\ -(4n+4)\pi\tau \leqq \operatorname{Im} W \leqq -(4n+2)\pi\tau\}.$$

The image surface R is obtained by gluing the lower semi-infinite line of each $\partial\Omega_n^+$ to the top line of $\partial\Omega_n^-$ and the lower line of $\partial\Omega_n^-$ to the top line of $\partial\Omega_{n+1}^+$.

The critical trajectory $[-2\sqrt{\tau}, 2\sqrt{\tau}]$ on the boundary of each H_n^+ and H_n^- corresponds to the vertical segment of $\partial\Omega_n^+$ and $\partial\Omega_n^-$. In addition, the critical trajectories on R_w correspond to the points of the surface R that lie above the imaginary axis.

Next we shall consider the mapping (6.19). We wish to impose conditions on the parameters in order that (6.20) should admit a univalent mapping g. A first observation is that (6.19) must be purely imaginary when $|z|=1$. This implies that the integration constant K is purely imaginary; thus

$$K = 2i\sigma\tau, \quad \sigma \text{ real.}$$

Let R_z denote that subset of the Riemann surface for the logarithm which lies above $\hat{\mathbb{D}}$. We want (6.19) to map R_z into R. This will occur if the boundary of R is contained in the image of ∂R_z under (6.19). The periodicity properties of the mappings (6.18) and (6.19) will then permit one to define a mapping g of the desired form.

It is easy to describe the required property geometrically. Namely, as z traverses the boundary of R_z, the points $Z(z)$ should at least cover the boundary of R. To describe this property more analytically, observe that α, β, γ, δ lie below the critical points of the mapping Z of ∂R_z. Therefore we require that the mapping Z of ∂R_z should pass through consecutive integer multiples of $-2\pi i\tau$ between critical points.

It is no loss of generality to assume that the points α, β, γ, δ are arranged in counterclockwise order on $|z|=1$. Since we may also adjust the branch of the logarithm, it is no loss of generality to require

$$\frac{1}{i}Z(\alpha) \geqq 0, \quad \frac{1}{i}Z(\beta) \leqq -2\pi\tau, \quad \frac{1}{i}Z(\gamma) \geqq -2\pi\tau, \quad \frac{1}{i}Z(\delta) \leqq -4\pi\tau. \tag{6.21}$$

In terms of our parameters the conditions (6.21) become

$$\begin{aligned}
&c^2 \operatorname{Im}\{\alpha(\alpha-2\mu_1)\} - 2\tau(\arg\alpha - \sigma) \geqq 0\\
&c^2 \operatorname{Im}\{\beta(\beta-2\mu_1)\} - 2\tau(\arg\beta - \sigma) \leqq -2\pi\tau\\
&c^2 \operatorname{Im}\{\gamma(\gamma-2\mu_1)\} - 2\tau(\arg\gamma - \sigma) \geqq -2\pi\tau\\
&c^2 \operatorname{Im}\{\delta(\delta-2\mu_1)\} - 2\tau(\arg\delta - \sigma) \leqq -4\pi\tau
\end{aligned} \tag{6.22}$$

where the arguments are chosen to satisfy $\arg\alpha \leqq \arg\beta \leqq \arg\gamma \leqq \arg\delta \leqq \arg\alpha + 2\pi$.

Finally, we need an expression for the coefficient c_1. This is obtained by comparing the expansions of (6.18) and (6.19) near ∞. Substituting (6.13) into (6.18), we find

$$\begin{aligned} W(g(z)) &= \tfrac{1}{2}g(z)^2 - \tau - 2\tau \log(g(z)/\sqrt{\tau}) + \dots \\ &= \tfrac{1}{2}c^2 z^2 + cc_0 z + \tfrac{1}{2}(c_0^2 + 2cc_1) - \tau + \tau \log(c^{-2}\tau) - 2\tau \log z + \dots . \end{aligned}$$

Comparing the constant term with (6.19), one obtains the relation

$$K = 2i\sigma\tau = \tfrac{1}{2}c_0^2 + cc_1 - \tau + \tau \log(c^{-2}\tau)$$

up to an integer multiple of $4\pi i\tau$, which is caused by the possible presence of different branches of the logarithm. However, this constant can be absorbed into σ since it induces in the conditions (6.22) only a translation of the argument through a multiple of the period 2π. Thus we have

$$\begin{aligned} c_1 &= c^{-1}\tau[1 - \log(c^{-2}\tau) + 2i\sigma] - \tfrac{1}{2}c^{-1}c_0^2 \\ &= c^{-1}\tau[1 - \log(c^{-2}\tau) + 2i\sigma] - \tfrac{1}{2}c\mu_1^2 \end{aligned} \tag{6.23}$$

in general.

Examples. The mappings (6.8) may be obtained by choosing

$$\alpha = \exp\{\tfrac{1}{2}i \arccos\tau\} \quad \text{where} \quad \arg\alpha = \tfrac{1}{2}\arccos\tau \in [0, \pi/4]$$

and then

$$\beta = -\bar{\alpha}, \quad \gamma = -\alpha, \quad \delta = \bar{\alpha}, \quad \text{and} \quad c = 1.$$

In this case

$$\begin{aligned} \frac{1}{i}Z(\alpha) &= \sqrt{1-\tau^2} - \tau \arccos\tau + \rho \\ \frac{1}{i}Z(-\bar{\alpha}) &= -\sqrt{1-\tau^2} + \tau \arccos\tau + \rho - 2\pi\tau \\ \frac{1}{i}Z(-\alpha) &= \sqrt{1-\tau^2} - \tau \arccos\tau + \rho - 2\pi\tau \\ \frac{1}{i}Z(\bar{\alpha}) &= -\sqrt{1-\tau^2} + \tau \arccos\tau + \rho - 4\pi\tau \end{aligned} \tag{6.24}$$

where we have replaced $2\sigma\tau$ by ρ. Then the conditions (6.22) reduce to the restriction (6.9) on ρ.

In order to obtain mappings to which (6.8) is subordinate, we may use the general parameters $\alpha, \beta, \gamma, \delta, c, \sigma$ and restrict them by (6.21) for univalence and by (6.24) to shorten the slit. That is,

$$0 \leqq \frac{1}{i}Z(\alpha) \leqq \sqrt{1-\tau^2} - \tau \arccos\tau + \rho,$$

$$-\sqrt{1-\tau^2} + \tau \arccos\tau + \rho - 2\pi\tau \leqq \frac{1}{i}Z(\beta) \leqq -2\pi\tau,$$

$$-2\pi\tau \leqq \frac{1}{i} Z(\gamma) \leqq \sqrt{1-\tau^2} - \tau \arccos \tau + \rho - 2\pi\tau,$$

$$-\sqrt{1-\tau^2} + \tau \arccos \tau + \rho - 4\pi\tau \leqq \frac{1}{i} Z(\delta) \leqq -4\pi\tau.$$

This simplifies slightly in the spirit of (6.22) to

$$\begin{aligned}
&0 \leqq c^2 \operatorname{Im}\{\alpha(\alpha - 2\mu_1)\} - 2\tau(\arg\alpha - \sigma) \leqq \sqrt{1-\tau^2} - \tau\arccos\tau + \rho \\
&0 \leqq c^2 \operatorname{Im}\{\beta(2\mu_1 - \beta)\} + 2\tau(\arg\beta - \sigma - \pi) \leqq \sqrt{1-\tau^2} - \tau\arccos\tau - \rho \\
&0 \leqq c^2 \operatorname{Im}\{\gamma(\gamma - 2\mu_1)\} - 2\tau(\arg\gamma - \sigma - \pi) \leqq \sqrt{1-\tau^2} - \tau\arccos\tau + \rho \\
&0 \leqq c^2 \operatorname{Im}\{\delta(2\mu_1 - \delta)\} + 2\tau(\arg\delta - \sigma - 2\pi) \leqq \sqrt{1-\tau^2} - \tau\arccos\tau - \rho
\end{aligned} \tag{6.25}$$

where ρ satisfies (6.9).

Under the conditions (6.25) the mappings (6.13) are admissible for Theorem 10. If we substitute into (6.10) the coefficients $C_n = \frac{1}{c} c_n$ where the c_n are obtained from (6.15) and (6.23), we have the following consequence of Theorem 10.

Theorem 11. *Let* $f(z) = z + b_0 + b_1 z^{-1} + b_2 z^{-2} + b_3 z^{-3} + \ldots$ *belong to* Σ. *Suppose that* $0 \leqq \tau \leqq 1$, *that* $\sqrt{\tau} \leqq c \leqq 1$, *and that* ρ *and* σ *are real with* $|\rho| \leqq \sqrt{1-\tau^2} - \tau \arccos \tau$. *If* $|\alpha| = |\beta| = |\gamma| = |\delta| = 1$ *with* $\alpha\beta\gamma\delta = 1$, $4\tau = c^2(\mu_2 - \mu_1^2)$, *and* $\arg\alpha \leqq \arg\beta \leqq \arg\gamma \leqq \arg\delta \leqq \arg\alpha + 2\pi$, *and if* (6.25) *is satisfied, then*

$$\begin{aligned}
&\operatorname{Re}\{b_3 - 2\mu_1 b_2 + \tfrac{1}{2} b_1^2 + (\mu_1^2 - 4c^{-2}\tau) b_1\} \\
&\quad \geqq -3|\mu_1|^2 - 2c^{-4}\tau^2[2 - \log(c^{-2}\tau)] \\
&\quad\quad + \tfrac{1}{8} \operatorname{Re}\{\mu_4 - 5\mu_1^4 + 4\mu_1\mu_3 - 16c^{-2}\tau\mu_1^2\}
\end{aligned} \tag{6.26}$$

where $\mu_j = \alpha^j + \beta^j + \gamma^j + \delta^j$ *for* $1 \leqq j \leqq 4$. *For each admissible choice of parameters, this inequality is sharp.*

The inequality (6.26) has some remarkable properties. First, it is independent of ρ and σ. However, restricting ρ and σ (e.g., $\rho = \sigma = 0$) would limit the possibilities for α, β, γ, δ and consequently the region of variability for the μ_j.

Nevertheless, suppose that $\rho = 0$, and denote $t = c^{-2}\tau$. Then the conditions (6.25) require that the expressions

$$\begin{aligned}
&\operatorname{Im}\{\alpha(\alpha - 2\mu_1)\} - 2t(\arg\alpha - \sigma) \\
&\operatorname{Im}\{\beta(2\mu_1 - \beta)\} + 2t(\arg\beta - \sigma - \pi) \\
&\operatorname{Im}\{\gamma(\gamma - 2\mu_1)\} - 2t(\arg\gamma - \sigma - \pi) \\
&\operatorname{Im}\{\delta(2\mu_1 - \delta)\} + 2t(\arg\delta - \sigma - 2\pi)
\end{aligned} \tag{6.27}$$

lie in the interval $[0, \sqrt{c^{-4} - t^2} - t \arccos(c^2 t)]$. Since the right endpoint of this interval is unbounded as c tends to 0, any choice of parameters α, β, γ, δ, σ for which the expressions (6.27) are nonnegative will lie in the interval for

some c. Therefore we have the following more pleasing consequence of Theorem 11.

Theorem 12. *Let* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$ *belong to* Σ, *and suppose* $0\leqq t\leqq 1$. *Let* $|\alpha|=|\beta|=|\gamma|=|\delta|=1$ *where* $\alpha\beta\gamma\delta=1$, $\arg\alpha\leqq\arg\beta\leqq\arg\gamma\leqq\arg\delta\leqq\arg\alpha+2\pi$, $4t=\mu_2-\mu_1^2$, *and* $\mu_j=\alpha^j+\beta^j+\gamma^j+\delta^j$ *for* $1\leqq j\leqq 4$. *Assume that the expressions* (6.27) *are nonnegative for some real* σ. *Then*

$$\begin{aligned}\operatorname{Re}\{b_3-2\mu_1 b_2+\tfrac{1}{2}b_1^2+(\mu_1^2-4t)b_1\}&\geqq -3|\mu_1|^2-2t^2(2-\log t)\\&\quad+\tfrac{1}{8}\operatorname{Re}\{\mu_4-5\mu_1^4+4\mu_1\mu_3-16t\mu_1^2\}.\end{aligned}\tag{6.28}$$

For each admissible choice of parameters this inequality is sharp.

The inequalities (6.26) and (6.28) simplify slightly if $\gamma=\bar{\beta}$ and $\delta=\bar{\alpha}$; in particular, such a choice makes the μ_j's real. If, in addition, $\alpha=-\bar{\beta}=\exp\{\frac{1}{2}i\arccos t\}$, then the inequalities (6.28) reduce to the original ones in Theorem 9. By replacing $f(z)$ by $e^{-i\phi}f(e^{i\phi}z)$ one obtains obvious generalizations of Theorems 11 and 12.

Let us now consider Theorem 8. Since the function (6.5) is a special case of the functions (6.8), corresponding to the choices

$$\tau=e^{-3}\quad\text{and}\quad\rho=0,\tag{6.29}$$

we may obtain admissible mappings for Theorem 8 by specializing the construction for Theorem 11. If we substitute into (6.6) the coefficients $C_n=\frac{1}{c}c_n$ where the c_n are obtained from (6.15) and (6.23) with the choices (6.29), we arrive at the following consequence of Theorem 8.

Theorem 13. *Let* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$ *belong to* Σ. *Suppose that* $e^{-3/2}\leqq c\leqq 1$ *and that* σ *is real. Let* $|\alpha|=|\beta|=|\gamma|=|\delta|=1$ *where* $\alpha\beta\gamma\delta=1$, $\arg\alpha\leqq\arg\beta\leqq\arg\gamma\leqq\arg\delta\leqq\arg\alpha+2\pi$, $4e^{-3}=c^2(\mu_2-\mu_1^2)$, *and* $\mu_j=\alpha^j+\beta^j+\gamma^j+\delta^j$ *for* $1\leqq j\leqq 4$. *Assume that the expressions*

$$\begin{aligned}&c^2\operatorname{Im}\{\alpha(\alpha-2\mu_1)\}-2e^{-3}(\arg\alpha-\sigma)\\&c^2\operatorname{Im}\{\beta(2\mu_1-\beta)\}+2e^{-3}(\arg\beta-\sigma-\pi)\\&c^2\operatorname{Im}\{\gamma(\gamma-2\mu_1)\}-2e^{-3}(\arg\gamma-\sigma-\pi)\\&c^2\operatorname{Im}\{\delta(2\mu_1-\delta)\}+2e^{-3}(\arg\delta-\sigma-2\pi)\end{aligned}\tag{6.30}$$

all lie in the interval $[0,\sqrt{1-e^{-6}}-e^{-3}\arccos(e^{-3})]$. *Then*

$$\begin{aligned}\operatorname{Re}\{b_3-2\mu_1 b_2+[\tfrac{1}{2}\mu_1^2+2e^{-3}c^{-2}(\log c+i\sigma)]b_1\}&\\\geqq -3|\mu_1|^2-2e^{-6}c^{-4}[1-2\log c-(\log c)^2+\sigma^2]&\\+\operatorname{Re}\{\tfrac{1}{8}\mu_4-\tfrac{1}{2}\mu_1^4+\tfrac{1}{2}\mu_1\mu_3-e^{-3}c^{-2}(4+\log c+i\sigma)\mu_1^2\}.&\end{aligned}\tag{6.31}$$

For each admissible choice of parameters, this inequality is sharp.

Corollary. *Let* $f(z)=z+b_0+b_1 z^{-1}+b_2 z^{-2}+b_3 z^{-3}+\ldots$ *belong to* Σ, *and assume that* c *and* σ *are real constants that satisfy*

$$2|\sigma|\leqq\sqrt{e^6c^4-1}-\arccos(e^{-3}c^{-2})\leqq\sqrt{e^6-1}-\arccos(e^{-3})-2|\sigma|.\tag{6.32}$$

Then

$$\operatorname{Re}\{b_3+2e^{-3}c^{-2}(\log c+i\sigma)b_1\} \geqq -\tfrac{1}{2}-e^{-6}c^{-4}[1-4\log c-2(\log c)^2+2\sigma^2]. \tag{6.33}$$

For each admissible c and σ this inequality is sharp.

Proof. In Theorem 13 choose $\alpha=\exp\{\frac{1}{2}i\arccos(e^{-3}c^{-2})\}$, $\beta=-\bar{\alpha}$, $\gamma=-\alpha$, and $\delta=\bar{\alpha}$. Then $\mu_1=\mu_3=0$, $\mu_2=4e^{-3}c^{-2}$, and $\mu_4=8e^{-6}c^{-4}-4$. The restriction on the expressions (6.30) reduces to (6.32), and the inequality (6.31) becomes (6.33).

The choice $c=1$, $\sigma=0$ corresponds to the inequality (6.4), from which Theorem 13 is derived. At the other extreme, the choice $c=e^{-3/2}$, $\sigma=0$ gives the inequality

$$\operatorname{Re}\{3b_1-b_3\}\leqq 3, \tag{6.34}$$

which was first found by Garabedian and Schiffer [6] by a similar method. Inequality (5.3) for $-f(-z)$ and $\mu=2$ and inequality (6.34) show that

$$\operatorname{Re}\{nb_1-b_n\}\leqq n \tag{6.35}$$

is true for $n=2$ and $n=3$. Equality occurs, of course, for $f(z)=z+b_0+z^{-1}$. It would appear that (6.35) is an attractive conjecture for $n\geqq 4$.

7. Constrained Problems

The underlying principle from Sect. 2 applies just as easily to constrained problems. We shall outline it in its simplest form.

Let λ and Λ be functionals defined on S. Assume that λ is real valued and that $F\in S$ is a function that provides its maximum over all functions $f\in S$ for which $\Lambda(f)=0$:

$$\lambda(F)=\max\{\lambda(f): f\in S \text{ and } \Lambda(f)=0\}.$$

Here Λ may be real or complex valued or, more generally, a vector.

As in Sect. 2 suppose that F is subordinate to a univalent function G, that $H=\dfrac{1}{c}G$ belongs to S, and that $\omega=G^{-1}\circ F$. On S define the functionals $\hat{\lambda}$ and $\hat{\Lambda}$ associated with λ and Λ by

$$\hat{\lambda}(f)=\lambda(cf\circ\omega) \quad \text{and} \quad \hat{\Lambda}(f)=\Lambda(cf\circ\omega).$$

If $f\in S$ satisfies the constraint $\hat{\Lambda}(f)=0$, then $cf\circ\omega$ belongs to S and satisfies the constraint $\Lambda(cf\circ\omega)=0$. Therefore

$$\hat{\lambda}(f)=\lambda(cf\circ\omega)\leqq\lambda(F)=\hat{\lambda}(H).$$

Since H also satisfies the constraint $\hat{\Lambda}(H)=0$, it provides a maximum for the new functional $\hat{\lambda}$ over all functions f in S satisfying the new constraint $\hat{\Lambda}(f)=0$. Furthermore, the two constrained maxima are equal. We summarize this principle using the notation above.

Basic Lemma with Constraint. *If F provides the maximum of a functional λ over all functions $f \in S$ for which $\Lambda(f)=0$, then H provides the maximum of the associated functional $\tilde{\lambda}$ over all functions $f \in S$ for which $\tilde{\Lambda}(f)=0$. In addition, the two maxima are equal.*

The corresponding statement for the class Σ is obvious.

First we shall give a very elementary illustration. Recall the following inequalities of Gronwall [8].

If $f(z)=z+a_2 z^2+\ldots$ belongs to S and $|z|=r<1$, then

$$r/(1+|a_2|r+r^2) \leqq |f(z)| \quad \text{and} \quad (1-r^2)/(1+|a_2|r+r^2)^2 \leqq |f'(z)|. \tag{7.1}$$

We shall demonstrate how the full inequalities (7.1) follow from the special case when $a_2=0$.

We take as given the inequalities

$$r/(1+r^2) \leqq |f(z)| \quad \text{and} \quad (1-r^2)/(1+r^2)^2 \leqq |f'(z)|, \quad |z|=r<1,$$

for functions $f \in S$ with $a_2=0$. That is, for fixed $z=re^{i\phi}$, we are given that the functionals

$$\lambda_1(f)=-|f(re^{i\phi})| \quad \text{and} \quad \lambda_2(f)=-|f'(re^{i\phi})|$$

are bounded above by $-r/(1+r^2)$ and $-(1-r^2)/(1+r^2)^2$, respectively, subject to the constraint that $\Lambda(f)=a_2$ is zero. Equality occurs for the function $F(z)=z/(1+e^{-2i\phi}z^2)$, which is easily seen to be subordinate to $G=cH$ where

$$H(z)=z/(1+te^{i\psi}z+e^{2i\psi}z^2), \qquad c=(1+\tfrac{1}{2}t)\,e^{i(\phi+\psi)},$$
$$0 \leqq t \leqq 2, \quad \text{and} \quad 0 \leqq \psi < 2\pi.$$

Using $\omega=(cH)^{-1}\circ F$, the Basic Lemma with Constraint asserts that the associated functionals

$$\tilde{\lambda}_1(f)=-|cf(\omega(re^{i\phi}))| \quad \text{and} \quad \tilde{\lambda}_2(f)=-|cf'(\omega(re^{i\phi}))\,\omega'(re^{i\phi})|$$

have the same bounds, respectively, subject to the constraint that $\tilde{\Lambda}(f)=(a_2+te^{i\psi})/c$ is zero. That is, if $\zeta=\omega(re^{i\phi})$, then

$$r/(1+r^2) \leqq |cf(\zeta)| \quad \text{and} \quad (1-r^2)/(1+r^2)^2 \leqq |cf'(\zeta)\,\omega'(re^{i\phi})|$$

for all functions f in S with $a_2=-te^{i\psi}$. Since

$$r/(1+r^2)=e^{-i\phi}F(re^{i\phi})=e^{-i\phi}cH(\zeta)=|c|\,|\zeta|/(1+t|\zeta|+|\zeta|^2)$$

and

$$(1-r^2)/(1+r^2)^2=F'(re^{i\phi})=cH'(\zeta)\,\omega'(re^{i\phi})$$
$$=|c\omega'(re^{i\phi})|(1-|\zeta|^2)/(1+t|\zeta|+|\zeta|^2)^2,$$

it follows that

$$|\zeta|/(1+t|\zeta|+|\zeta|^2) \leqq |f(\zeta)| \quad \text{and} \quad (1-|\zeta|^2)/(1+t|\zeta|+|\zeta|^2)^2 \leqq |f'(\zeta)|.$$

By allowing ϕ, ψ, and t to vary, we have these inequalities for any $\zeta \in \mathbb{D}$ and $|a_2|=t$.

Our final application is to a constrained problem in Σ. We shall start with the inequality

$$\operatorname{Re} b_3 \leqq \tfrac{1}{2} \quad \text{whenever} \quad b_1 = 0, \tag{7.2}$$

which has been proved by Jenkins [10] (see also Duren [3]). Actually, Jenkins and Duren proved that $|b_n| \leqq 2/(n+1)$ whenever $b_k = 0$ for $1 \leqq k \leqq n/2$. We could as well apply the Basic Lemma with Constraint to this inequality; however, the inequality (7.2) will be sufficient to indicate its potential.

Theorem 14. *Suppose* $f(z) = z + b_0 + b_1 z^{-1} + b_2 z^{-2} + \ldots$ *belongs to* Σ, *and for* $j = 1, 2$ *let* $\mu_j = \alpha^j + \beta^j + \gamma^j + \delta^j$ *where* $|\alpha| = |\beta| = |\gamma| = |\delta| = 1$ *and* $\alpha\beta\gamma\delta = 1$. *If*

$$b_1 = \tfrac{1}{8}(\mu_1^2 - 2\mu_2),$$

then

$$\operatorname{Re}\{b_3 + \mu_1 b_2\} \leqq \tfrac{1}{2} + \tfrac{1}{4}|\mu_1|^2 - \operatorname{Re}\{\tfrac{1}{32}\mu_2^2 - \tfrac{3}{32}\mu_1^2\mu_2 + \tfrac{5}{128}\mu_1^4\},$$

and this inequality is sharp.

Proof. The inequality (7.2) asserts that the functional $\lambda(f) = \operatorname{Re} b_3$ is bounded by $\frac{1}{2}$ whenever $\Lambda(f) = b_1$ is zero. Equality occurs for the odd function

$$F(z) = z(1 + z^{-4})^{1/2} = z + \tfrac{1}{2}z^{-3} + \ldots.$$

For this function $\mathbb{C} \setminus F(\mathbb{D})$ consists of the intersection of the disk $|w| \leqq \sqrt{2}$ with the coordinate axes. If $|\alpha| = |\beta| = |\gamma| = |\delta| = 1$ and $\alpha\beta\gamma\delta = 1$, then

$$H(z) = z[(1 + \alpha/z)(1 + \beta/z)(1 + \gamma/z)(1 + \delta/z)]^{1/2}$$

belongs to Σ and maps $\mathbb{D}$ onto the complement of a compact subset of the coordinate axes. Therefore $F \prec cH$ for a sufficiently small positive value c.

We may use the formalism at the beginning of Sect. 6 to determine the associated functionals $\tilde{\lambda}$ and $\tilde{\Lambda}$. In particular, using (6.3) we find that

$$\begin{aligned}\tilde{\lambda}(f) = \operatorname{Re}\{&c^4(b_3 + 2C_0 b_2 + C_0^2 b_1 + C_1 b_1 - C_0^2 C_1 - 2C_0 C_2 - C_1^2 - C_3) \\ &- c^2 B_1(b_1 - C_1) + B_3\}\end{aligned} \tag{7.3}$$

and

$$\tilde{\Lambda}(f) = c^2(b_1 - C_1) + B_1 \tag{7.4}$$

where

$$B_1 = 0 \quad \text{and} \quad B_3 = \tfrac{1}{2} \tag{7.5}$$

are coefficients of F and

$$\begin{aligned} C_0 &= \tfrac{1}{2}\mu_1 \\ C_1 &= \tfrac{1}{8}\mu_1^2 - \tfrac{1}{4}\mu_2 \\ C_2 &= \tfrac{1}{2}\bar{\mu}_1 - \tfrac{1}{16}\mu_1^3 + \tfrac{1}{8}\mu_1\mu_2 \\ C_3 &= \tfrac{1}{2} - \tfrac{1}{4}|\mu_1|^2 - \tfrac{1}{32}\mu_2^2 - \tfrac{1}{32}\mu_1^2\mu_2 + \tfrac{3}{128}\mu_1^4 \end{aligned} \tag{7.6}$$

are coefficients of H. Since $\tilde{\lambda}(f)$ is bounded by $\frac{1}{2}$ whenever $\tilde{\Lambda}(f) = 0$, the conclusion of the theorem follows by substituting the coefficients (7.5) and (7.6) into (7.3) and (7.4).

References

1. Brickman, L., Wilken, D.R.: Support points of the set of univalent functions. Proc. Amer. Math. Soc. **42**, 523–528 (1974)
2. Chang, A., Schiffer, M., Schober, G.: On the second variation for univalent functions. J. Analyse Math. (to appear)
3. Duren, P.L.: Coefficients of meromorphic schlicht functions. Proc. Amer. Math. Soc. **28**, 169–172 (1971)
4. Friedland, S., Schiffer, M.: Global results in control theory with applications to univalent functions. Bull. Amer. Math. Soc. **82**, 913–915 (1976)
5. Friedland, S., Schiffer, M.: On coefficient regions of univalent functions. J. Analyse Math. **31**, 125–168 (1977)
6. Garabedian, P.R., Schiffer, M.: A coefficient inequality for schlicht functions. Ann. of Math. (2) **61**, 116–136 (1955)
7. Golusin, G.M.: On distortion theorems in the theory of conformal mapping. Mat. Sb. **18**, 379–390 (1946) [Russian]
8. Gronwall, T.H.: Sur la déformation dans la représentation conforme sous des conditions restrictives. C.R. Acad. Sci. Paris Sér. A **162**, 316–318 (1916)
9. Jenkins, J.A.: Symmetrization results for some conformal invariants. Amer. J. Math. **75**, 510–522 (1953)
10. Jenkins, J.A.: On certain coefficients of univalent functions II. Trans. Amer. Math. Soc. **96**, 534–545 (1960)
11. Jenkins, J.A.: On certain coefficients of univalent functions. In: Analytic Functions. Proceedings of a Conference (Princeton 1957), pp. 159–194. Princeton, N.J.: Princeton University Press 1960
12. Jenkins, J.A.: Univalent Functions and Conformal Mapping. Berlin-Heidelberg-New York: Springer 1965
13. Löwner, K.: Untersuchungen über schlichte konforme Abbildungen des Einheitskreises I. Math. Ann. **89**, 103–121 (1923)
14. Pearce, K.: New support points of S and extreme points of HS. Proc. Amer. Math. Soc. **81**, 425–428 (1981)
15. Pfluger, A.: Lineare Extremalprobleme bei schlichten Funktionen. Ann. Acad. Sci. Fenn. Ser. AI **489** (1971)
16. Schiffer, M.: Sur un problème d'extrémum de la représentation conforme. Bull. Soc. Math. France **66**, 48–55 (1938)
17. Schober, G.: Univalent Functions-Selected Topics. Lecture Notes in Mathematics **478**. Berlin-Heidelberg-New York: Springer 1974

Received October 7, 1981

Math. Z. 180, 41 – 67 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Flächen vorgeschriebener mittlerer Krümmung mit eineindeutiger Projektion auf eine Ebene

Friedrich Sauvigny

Institut für Mathematik der Rheinisch-Westfälischen Technischen Hochschule, Templergraben 55, D-5100 Aachen, Bundesrepublik Deutschland

0. Einleitung

Mit den direkten Methoden der Variationsrechnung ist es möglich, das parametrische Plateauproblem für Flächen vorgeschriebener mittlerer Krümmung =diese nennen wir kurz $H(\ )$-Flächen=in einem Zylinder zu lösen (s. [6, 24, 10, 7, 18]). Hat nun die berandende Jordankurve eine konvexe, eineindeutige Projektion auf eine Ebene, so ist jede sich einspannende Minimalfläche auf diese Ebene eineindeutig projizierbar. Dieser Satz von Radó, dessen Beweis man Kneser [20] verdankt, wurde von Gulliver und Spruck [8] auf den Fall H =const bei C^2-Randkurven verallgemeinert. Der Wunsch, ebenso wie im Fall $H\equiv 0$ (Radó [24], §3) mit obigen Arbeiten [7, 18] einen parametrischen Zugang zum Dirichletproblem der nichtparametrischen $H(x, y, z)$-Flächen-Gleichung zu stetigen Randwerten zu eröffnen, erfordert eine Übertragung von Radó's Satz auf $H(\ )$-Flächen:

Sei

$$B := \{w = u + iv \mid |w| < 1\},\ \mathfrak{e} = (0, 0, 1), \qquad h \geqq 0$$

und

$$\mathfrak{Z}_{2h} := \{(x, y, z) \in \mathbb{R}^3 \mid 4h^2(x^2 + y^2) \leqq 1\}$$

gesetzt. Auf $\mathfrak{Z}_{2h}$ schreiben wir $H = H(x, y, z) \in C^{1+\alpha}(\mathfrak{Z}_{2h})\ (\alpha \in (0, 1))$ mit $\sup_{\mathfrak{Z}_{2h}} |H| \leqq h$ und $\frac{\partial}{\partial z} H \geqq 0$ vor. Wir betrachten nun Jordankurven $\Gamma \subset \mathfrak{Z}_{2h}$ der Klasse $\mathfrak{C}_{2h}$ (s. Definition 5).

Gehen wir nun von einer $H(\ )$-Fläche $\mathfrak{x}(u, v)$: $\bar{B} \to \mathfrak{Z}_{2h}$ zur Kurve Γ aus, so beweisen wir für deren Normale $\mathfrak{X}(u, v)$ in Satz 2 eine Randbedingung

$$\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0 \quad \text{für alle } (u, v) \in \partial B. \tag{1}$$

Unter der Randbedingung (1) ist im allgemeinen nicht zu erwarten, daß $\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0$ für alle $(u, v) \in B$ folgt. Geht man aber von einer, im Sinne von Definition 6, stabilen $H(\ )$-Fläche aus, so ist dieser Schluß richtig! Dieser

0025-5874/82/0180/0041/$05.40

Satz 3 ergibt sich nach eingehender Diskussion der in §1 hergeleiteten Differentialgleichung von $\mathfrak{x}$, zusammen mit der Randbedingung (1). Wir erhalten also (Satz 2 und Satz 3) als zentrales Ergebnis folgenden

Satz. *Unter oben angegebenen Voraussetzungen an $H(x,y,z)$ und Γ gilt: Eine stabile $H(\ )$-Fläche $\mathfrak{x}(u,v)$: $\bar{B}\to\mathfrak{Z}_{2h}$ zu Γ ist eineindeutig auf die x,y-Ebene projizierbar, und ihre nichtparametrische Darstellung $z=\zeta(x,y)\in C^2(\Omega)\cap C^0(\bar{\Omega})$ löst das Dirichletproblem*

$$\begin{aligned} \mathscr{L}(\zeta)&=2H(x,y,\zeta(x,y))\,(1+|\nabla\zeta|^2)^{3/2} && \text{in } \Omega \\ \zeta&=f && \text{auf } \partial\Omega. \end{aligned} \tag{*}$$

Dabei ist der Minimalflächenausdruck zu „$\mathscr{L}$" abgekürzt worden (Definition 7).

Da nach Hilfssatz 8 die von Gulliver und Spruck und Hildebrandt konstruierte Lösung im Sinne von Definition 6 stabil ist, kann man mit diesem Satz das Dirichletproblem (*) zu C^2-Daten sofort lösen. Um aber einen parametrischen Zugang zu diesem Dirichletproblem mit stetigen Daten zu eröffnen (Satz 5), ist in §4 eine Approximation mit projizierbaren Lösungen $\mathfrak{x}^n(u,v)\,(n=1,2,\ldots)$ zu Kurven Γ^n der Klasse $\mathfrak{C}_{2h}$ durchzuführen. Entscheidend ist dabei eine Ausage über die „Abgeschlossenheit" der Klasse der projizierbaren Lösungen, nämlich Hilfssatz 12 (vgl. [13], Theorem 2).

Die Frage nach der Eindeutigkeit des Plateauschen Problems können wir auf unserem Weg nur eingeschränkt beantworten: Unter obigen Voraussetzungen gibt es zu Kurven der Klasse $\mathfrak{C}_{2h}$ genau eine stabile $H(\ )$-Fläche (s. Satz 4).

Die vorliegende Arbeit, die ich am Institut für Mathematik der RWTH Aachen angefertigt habe, wurde 1981 von der Mathematisch-Naturwissenschaftlichen Fakultät der Georg-August-Universität Göttingen als Dissertation angenommen. An dieser Stelle möchte ich meinem Lehrer, Herrn Prof. Dr. E. Heinz, Göttingen, für viele Anregungen und wertvolle Hinweise danken.

1. Eine Differentialgleichung für die Normale einer Fläche vorgeschriebener mittlerer Krümmung

Die Einheitskreisscheibe in der Ebene bezeichnen wir mit B, also

$$B:=\{w=u+iv\in\mathbb{C}\,|\,|w|<1\}.$$

In diesem Abschnitt schreiben wir auf der offenen Menge $\mathfrak{O}\subset\mathbb{R}^3$ die mittlere Krümmung $H=H(x,y,z)$ vor

$$H=H(x,y,z)\colon \mathfrak{O}\to\mathbb{R}\in C^{1+\alpha}(\mathfrak{O}) \qquad (\alpha\in(0,1)).$$

Definition 1. Eine nicht verschwindende Lösung $\mathfrak{x}=\mathfrak{x}(u,v)\colon B\to\mathfrak{O}\in C^2(B)$ des elliptischen Systems

$$\begin{aligned} \Delta\mathfrak{x}&=2H(\mathfrak{x})\,\mathfrak{x}_u\times\mathfrak{x}_v \\ \mathfrak{x}_u\cdot\mathfrak{x}_v&=0=\mathfrak{x}_u^2-\mathfrak{x}_v^2 \end{aligned} \qquad \text{in } B$$

nennen wir *Fläche vorgeschriebener mittlerer Krümmung* $H=H(x,y,z)$, bzw. kurz $H(\ )$*-Fläche.*

Ausgehend von der asymptotischen Entwicklung für die komplexe Ableitung einer $H(\ )$-Fläche kann man feststellen, daß die Verzweigungspunkte (in denen $\mathfrak{x}_u^2=0=\mathfrak{x}_v^2$ gilt) in B isoliert sind, und die Normale stetig in die Verzweigungspunkte fortgesetzt werden kann (s. hierzu Heinz, Hildebrandt [15], S. 231).

Also ist die folgende Definition sinnvoll:

Definition 2. Seien $\{w_n\}_{n=1,2,\ldots}$ die Verzweigungspunkte von $\mathfrak{x}$ in B; $B':=B-\{w_n\}_n$. Dann setzen wir in $(u_0,v_0)\in B$:

$$\mathfrak{X}(u_0,v_0):=\lim_{\substack{(u,v)\to(u_0,v_0)\\(u,v)\in B'}}\frac{\mathfrak{x}_u\times\mathfrak{x}_v(u,v)}{|\mathfrak{x}_u\times\mathfrak{x}_v(u,v)|},$$

und nennen $\mathfrak{X}$ die *Normale* an $\mathfrak{x}$.

Somit ist offenbar $\mathfrak{X}\in C^0(B)$ erfüllt. Mit einem Satz der Potentialtheorie ist mit $H\in C^{1+\alpha}(\mathfrak{D})$ dann auch $\mathfrak{x}(u,v)\in C^{3+\alpha}(B)$, also auch $\mathfrak{X}\in C^{2+\alpha}(B')$ richtig. Somit ist in B' die folgende Differentialgleichung für $\mathfrak{X}$ sinnvoll:

Hilfssatz 1. *Unter obigen Voraussetzungen genügt* $\mathfrak{X}(u,v)$ *in* B' *der folgenden Differentialgleichung:*

$$\Delta\mathfrak{X}+2E(2(H(\mathfrak{x}))^2-K)\mathfrak{X}=-2((\nabla H\cdot\mathfrak{x}_u)\mathfrak{x}_u+(\nabla H\cdot\mathfrak{x}_v)\mathfrak{x}_v).$$

Beweis. Für eine Fläche in konformen Parametern haben wir die Ableitungsgleichungen:

$$\mathfrak{X}_u=-\frac{L}{E}\mathfrak{x}_u-\frac{M}{E}\mathfrak{x}_v;\qquad \mathfrak{X}_v=-\frac{M}{E}\mathfrak{x}_u-\frac{N}{E}\mathfrak{x}_v$$

$$\Rightarrow(\mathfrak{X}\times\mathfrak{X}_v)_u-(\mathfrak{X}\times\mathfrak{X}_u)_v=2\mathfrak{X}_u\times\mathfrak{X}_v$$

$$=2\frac{LN-M^2}{E^2}\mathfrak{x}_u\times\mathfrak{x}_v=2EK\mathfrak{X}.$$

Weiter gilt:

$$\mathfrak{X}\times\mathfrak{X}_u=-\mathfrak{X}_v-2H(\mathfrak{x})\mathfrak{x}_v;\quad \mathfrak{X}\times\mathfrak{X}_v=\mathfrak{X}_u+2H(\mathfrak{x})\mathfrak{x}_u.$$

Wegen $(H(\mathfrak{x}))_u=\nabla H\cdot\mathfrak{x}_u$ und analog $(H(\mathfrak{x}))_v=\nabla H\cdot\mathfrak{x}_v$ folgt nun:

$$\begin{aligned}2EK\mathfrak{X}&=(\mathfrak{X}\times\mathfrak{X}_v)_u-(\mathfrak{X}\times\mathfrak{X}_u)_v\\&=\mathfrak{X}_{uu}+2(\nabla H\cdot\mathfrak{x}_u)\mathfrak{x}_u+2H\mathfrak{x}_{uu}+\mathfrak{X}_{vv}+2(\nabla H\cdot\mathfrak{x}_v)\mathfrak{x}_v+2H\mathfrak{x}_{vv}\\&=\Delta\mathfrak{X}+2H\Delta\mathfrak{x}+2((\nabla H\cdot\mathfrak{x}_u)\mathfrak{x}_u+(\nabla H\cdot\mathfrak{x}_v)\mathfrak{x}_v)\\&=\Delta\mathfrak{X}+4EH^2\mathfrak{X}+2((\nabla H\cdot\mathfrak{x}_u)\mathfrak{x}_u+(\nabla H\cdot\mathfrak{x}_v)\mathfrak{x}_v)\quad\text{w.z.b.w.}\end{aligned}$$

Mit den Ableitungsgleichungen beweist man sofort die folgende

Bemerkung. In B' gilt: $\mathfrak{X}_u^2+\mathfrak{X}_v^2=2E(2(H(\mathfrak{x}))^2-K)$.

Nun nehmen wir o.E. an, daß die $H(\ )$-Fläche $\mathfrak{x}$ in 0 einen Verzweigungspunkt hat. Unter $H_2^1(B)$ verstehen wir den Sobolevraum der im schwachen Sinne einmal differenzierbaren Funktionen, welche zusammen mit ihren ersten Ableitungen quadratintegrierbar sind.

Hilfssatz 2. *Für die $H(\ )$-Fläche $\mathfrak{x}(u,v)$ gelte: $\mathfrak{x}_u(0,0)=0$, $\mathfrak{x}_u(u,v)\neq 0$ für alle (u,v) mit $0<u^2+v^2\leqq\varepsilon^2$. Sei $B_\varepsilon:=\{u^2+v^2<\varepsilon^2\}$, so gilt für die Normale $\mathfrak{X}(u,v)$ an $\mathfrak{x}(u,v)$: $\mathfrak{X}(u,v)\in H_2^1(B_\varepsilon)\cap C^0(\bar{B}_\varepsilon)$.*

Außerdem genügt $\mathfrak{X}$ in B_ε dem nichtlinearen System

$$\Delta\mathfrak{X}+|\nabla\mathfrak{X}|^2\,\mathfrak{X}=\mathfrak{F}(u,v) \tag{*}$$

im schwachen Sinne. Dabei ist $|\nabla\mathfrak{X}|^2=\mathfrak{X}_u^2+\mathfrak{X}_v^2$ und

$$\mathfrak{F}(u,v):=-2((\nabla H\cdot\mathfrak{x}_u)\,\mathfrak{x}_u+(\nabla H\cdot\mathfrak{x}_v)\,\mathfrak{x}_v)\in C^\alpha(B_\varepsilon)$$

gesetzt worden.

Beweis. Wir betrachten die folgenden Glättungsfunktionen:

$$\chi(t):[0,\infty)\to[0,2]\in C^\infty([0,\infty))$$

mit

$$\chi(t)=\begin{cases}0 & 0\leqq t\leqq\frac{1}{2}\\ 1 & 1\leqq t<\infty\end{cases}$$

$$\chi(u,v):=\chi(u^2+v^2),\qquad \chi^{(n)}(u,v):=\chi(n^2(u^2+v^2))\qquad \text{für } (u,v)\in\mathbb{R}^2.$$

Wir erklären nun noch: $B_\varepsilon':=B_\varepsilon-\{(0,0)\}$.
Für eine Testfunktion $\Phi(u,v)$: $B_\varepsilon\to\mathbb{R}^3\in C_0^\infty(B_\varepsilon)$ setzen wir dann:

$$\Phi^{(n)}(u,v):=\Phi(u,v)\cdot\chi^{(n)}(u,v)\in C_0^\infty(B_\varepsilon').$$

Da nun das Dirichletintegral invariant unter konformer Abbildung ist, folgt für hinreichend großes $n_0\in\mathbb{N}$:

$$\iint\limits_{B_\varepsilon}|\nabla\chi^{(n)}|^2=\iint\limits_{|w|\leqq\frac{1}{n}}|\nabla\chi^{(n)}|^2=\iint\limits_{B}|\nabla\chi|^2;\qquad n\geqq n_0.$$

Somit folgt die Abschätzung

$$\iint\limits_{B_\varepsilon}|\nabla\Phi^{(n)}|^2\leqq\text{const.}\qquad (n=1,2,\ldots).$$

Zur Abkürzung schreiben wir in diesem Beweis die bei den Integralen auftretenden Differentiale $du\,dv$ nicht auf.

1. Zunächst besitzt $\mathfrak{X}$ schwache Ableitungen

$$D_u\mathfrak{X}:=\begin{cases}\mathfrak{X}_u & \text{in } B_\varepsilon'\\ 0 & \text{in } 0\end{cases}$$

und $D_v\mathfrak{X}$, die quadratintegrierbar sind. Nach Heinz und Hildebrandt [15], 2. Lemma 3 existiert nämlich das Integral:

$$\iint\limits_{B'_\varepsilon}(\mathfrak{X}_u^2+\mathfrak{X}_v^2)\,du\,dv=\iint\limits_{B'_\varepsilon}2E(2H^2-K)\,du\,dv.$$

Zu $\Phi(u,v)\in C_0^\infty(B_\varepsilon)$ betrachten wir die Folge $\Phi^{(n)}:=\Phi\cdot\chi^{(n)}(u,v)$, die $\Phi^{(n)}\to\Phi(n\to\infty)$ f.ü. und $|\Phi^{(n)}|\leqq|\Phi|\cdot 2$ erfüllt.

$$0=\iint\limits_{B_\varepsilon}(\mathfrak{X}\cdot\Phi^{(n)})_u=\iint\limits_{B_\varepsilon}D_u\mathfrak{X}\cdot\Phi^{(n)}+\iint\limits_{B_\varepsilon}\Phi_u^{(n)}\cdot\mathfrak{X},\qquad n\in\mathbb{N}. \tag{n}$$

Mit dem Lebesgueschen Konvergenzsatz sieht man sofort:

$$\lim_{n\to\infty}\iint\limits_{B_\varepsilon}D_u\mathfrak{X}\cdot\Phi^{(n)}=\iint\limits_{B_\varepsilon}D_u\mathfrak{X}\cdot\Phi. \tag{1}$$

Unter Benutzung obiger Abschätzung beweist man leicht:

$$\lim_{n\to\infty}\iint\limits_{B_\varepsilon}\mathfrak{X}\cdot\Phi_u^{(n)}=\iint\limits_{B_\varepsilon}\mathfrak{X}\cdot\Phi_u \tag{2}$$

lim-Bildung in (n) zusammen mit (1) und (2) ergibt:

$$0=\iint\limits_{B_\varepsilon}D_u\mathfrak{X}\cdot\Phi+\iint\limits_{B_\varepsilon}\mathfrak{X}\cdot\Phi_u\qquad\text{für alle }\Phi\in C_0^\infty(B_\varepsilon).$$

Somit ist $D_u\mathfrak{X}$ Distributionsableitung von $\mathfrak{X}$, ebenso wie $D_v\mathfrak{X}$. Also gilt: $\mathfrak{X}\in H_2^1(B_\varepsilon)\cap C^0(\bar{B}_\varepsilon)$.

2. Wir weisen nun nach, daß $\mathfrak{X}$ dem System (*) im schwachen Sinne genügt. Sei $\Phi\in C_0^\infty(B_\varepsilon)$, also $\Phi^{(n)}:=\Phi\cdot\chi^{(n)}\in C_0^\infty(B'_\varepsilon)$. Nach Hilfssatz 1 und der anschließenden Bemerkung gilt nun für $n\in\mathbb{N}$:

$$-\iint\limits_{B_\varepsilon}\nabla\Phi^{(n)}\cdot\nabla\mathfrak{X}+\iint\limits_{B_\varepsilon}|\nabla\mathfrak{X}|^2\,\mathfrak{X}\cdot\Phi^{(n)}=\iint\limits_{B_\varepsilon}\mathfrak{H}\cdot\Phi^{(n)}. \tag{((n))}$$

Wir untersuchen nun das erste Integral in $((n))$.

$$\iint\limits_{B_\varepsilon}\nabla\Phi^{(n)}\cdot\nabla\mathfrak{X}=\iint\limits_{u^2+v^2\leqq\frac{1}{n^2}}\nabla\Phi^{(n)}\cdot\nabla\mathfrak{X}+\iint\limits_{\varepsilon^2\geqq u^2+v^2\geqq\frac{1}{n^2}}\nabla\Phi\cdot\nabla\mathfrak{X}.$$

Nun ist:

$$\Big|\iint\limits_{u^2+v^2\leqq\frac{1}{n^2}}\nabla\Phi^{(n)}\cdot\nabla\mathfrak{X}\Big|$$

$$\leqq\Big(\iint\limits_{B_\varepsilon}|\nabla\Phi^{(n)}|^2\Big)^{1/2}\Big(\iint\limits_{u^2+v^2\leqq\frac{1}{n^2}}|\nabla\mathfrak{X}|^2\Big)^{1/2}\leqq C\cdot\Big(\iint\limits_{u^2+v^2\leqq\frac{1}{n^2}}|\nabla\mathfrak{X}|^2\Big)^{1/2}->0(n->\infty)$$

$$\Rightarrow\lim_{n\to\infty}\iint\limits_{B_\varepsilon}\nabla\Phi^{(n)}\cdot\nabla\mathfrak{X}=\iint\limits_{B_\varepsilon}\nabla\Phi\cdot\nabla\mathfrak{X}.$$

Gehen wir nun in $((n))$ bei den übrigen Summanden mit dem Lebesgueschen Konvergenzsatz zur Grenze über, so erhalten wir:

$$-\iint\limits_{B_\varepsilon} \nabla\Phi\cdot\nabla\mathfrak{X}+\iint\limits_{B_\varepsilon}|\nabla\mathfrak{X}|^2\,\mathfrak{X}\cdot\Phi$$
$$=\iint\limits_{B_\varepsilon}\mathfrak{F}\cdot\Phi \quad \text{für alle } \Phi\in C_0^\infty(B_\varepsilon). \quad \text{w.z.b.w.}$$

Wir fassen nun unsere Ergebnisse zusammen zu

Satz 1. *Sei* $\mathfrak{O}\subset\mathbb{R}^3$ *offen,* $H=H(x,y,z)\in C^{1+\alpha}(\mathfrak{O})$ *vorgeschrieben. Sei* $\mathfrak{x}(u,v)\colon B\to\mathfrak{O}$ *Fläche vorgeschriebener mittlerer Krümmung* $H(x,y,z)$. *Dann gilt für die Normale* $\mathfrak{X}(u,v)$ *an die Fläche* $\mathfrak{x}(u,v)$:

$$\mathfrak{X}(u,v)\in C^{2+\alpha}(B),$$

und sie genügt der folgenden Differentialgleichung in B:

$$\Delta\mathfrak{X}+2(2E(H(\mathfrak{x}))^2-EK-E(\nabla H\cdot\mathfrak{X}))\,\mathfrak{X}=-2E\nabla H.$$

Für den in der Differentialgleichung auftretenden Koeffizienten EK *gilt:*

$$E(u,v)\cdot K(u,v)\in C^{1+\alpha}(B).$$

Beweis. 1. Wegen $H\in C^{1+\alpha}(\mathfrak{O})$ folgt $\mathfrak{x}\in C^{3+\alpha}(B)$, also gilt dann $\mathfrak{X}\in C^{2+\alpha}(B')\cap C^0(B)$. Nach Hilfssatz 2 genügt $\mathfrak{X}$ in der Umgebung eines Verzweigungspunktes im schwachen Sinne dem System

$$\Delta\mathfrak{X}+|\nabla\mathfrak{X}|^2\,\mathfrak{X}=\mathfrak{F}\in C^\alpha(B). \tag{*}$$

Nach einem Regularitätssatz für nichtlineare elliptische Systeme (s. z.B. Tomi [28], dessen Beweis auch unser System (*) erfaßt) gilt nun $\mathfrak{X}\in C^{1+\alpha}(B)$. Mit einem Schluß der Potentialtheorie folgt nun, daß $\mathfrak{X}\in C^{2+\alpha}(B)$ richtig ist.

2. Nach der Bemerkung ist:

$$4EH^2-2EK=\mathfrak{X}_u^2+\mathfrak{X}_v^2\in C^{1+\alpha}(B)$$
$$\Rightarrow E(u,v)\cdot K(u,v)\in C^{1+\alpha}(B).$$

3. Nun gilt die in Hilfssatz 1 angegebene Differentialgleichung in B, und wir beachten:

$$\nabla H=\left(\nabla H\cdot\frac{\mathfrak{x}_u}{|\mathfrak{x}_u|}\right)\frac{\mathfrak{x}_u}{|\mathfrak{x}_u|}$$
$$+\left(\nabla H\cdot\frac{\mathfrak{x}_v}{|\mathfrak{x}_v|}\right)\frac{\mathfrak{x}_v}{|\mathfrak{x}_v|}+(\nabla H\cdot\mathfrak{X})\,\mathfrak{X} \qquad \text{in } B'$$
$$\Rightarrow(\nabla H\cdot\mathfrak{x}_u)\,\mathfrak{x}_u+(\nabla H\cdot\mathfrak{x}_v)\,\mathfrak{x}_v=E\nabla H-E(\nabla H\cdot\mathfrak{X})\,\mathfrak{X} \qquad \text{in } B$$
$$\Rightarrow\Delta\mathfrak{X}+2(2EH^2-EK)\,\mathfrak{X}=-2E\nabla H+2E(\nabla H\cdot\mathfrak{X})\,\mathfrak{X} \qquad \text{in } B$$
$$\Rightarrow\Delta\mathfrak{X}+2(2EH^2-EK-E(\nabla H\cdot\mathfrak{X}))\,\mathfrak{X}=-2E\nabla H \qquad \text{in } B.$$

w.z.b.w.

2. Die Randbedingung - Transversalität an den Rand

Zu $\kappa \geqq 0$ definieren wir die abgeschlossenen Kreiszylinder

$$\mathfrak{Z}_\kappa := \{(x, y, z) \in \mathbb{R}^3 | \kappa^2(x^2 + y^2) \leqq 1\}.$$

Definition 3. Sei $\kappa \geqq 0$, $H = H(x, y, z) \in C^0(\mathfrak{Z}_\kappa)$ vorgeschrieben. Weiter sei $\Gamma \subset \mathfrak{Z}_\kappa$ eine Jordankurve, auf der wir drei verschiedene Punkte $\{\pi_\nu\}_{\nu=0,1,2}$ festlegen. Dann nennen wir eine Lösung

$$\mathfrak{x} = \mathfrak{x}(u, v): \bar{B} \to \mathfrak{Z}_\kappa \in C^2(B) \cap C^0(\bar{B})$$

des elliptischen Systems

$$\begin{aligned} \Delta \mathfrak{x} &= 2H(\mathfrak{x})\, \mathfrak{x}_u \times \mathfrak{x}_v \quad \text{in } B, \\ \mathfrak{x}_u \cdot \mathfrak{x}_v &= 0 = \mathfrak{x}_u^2 - \mathfrak{x}_v^2 \end{aligned}$$

welche die Randbedingungen

$$\begin{aligned} &\mathfrak{x}: \partial B \to \Gamma \qquad \text{topologisch} \\ &\mathfrak{x}(e^{\frac{2\pi i}{3}\nu}) = \pi_\nu \qquad (\nu = 0, 1, 2) \end{aligned}$$

erfüllt, eine *Fläche vorgeschriebener mittlerer Krümmung zu* Γ, $\{\pi_\nu\}_{\nu=0,1,2}$ (kurz $H(\ )$-*Fläche zu* Γ).

Ebenso wie beim Existenznachweis einer parametrischen Lösung im Zylinder, ist auch für uns das folgende Maximumprinzip von zentraler Bedeutung:

Hilfssatz 3 (Maximumprinzip). *Sei* $h > 0$, $H = H(x, y, z) \in C^\alpha(\mathfrak{Z}_{2h})$ *mit* $\sup_{\mathfrak{Z}_{2h}} |H(x, y, z)| \leqq h$ *vorgegeben. Weiter sei* $\mathfrak{x}(u, v): \bar{B} \to \mathfrak{Z}_{2h} \in C^2(B) \cap C^0(\bar{B})$ *eine* $H(\ )$-*Fläche.*

Behauptung. Genügt die Funktion

$$\Phi(u, v) := (x(u, v))^2 + (y(u, v))^2$$

der Bedingung $\Phi(u, v) \not\equiv \dfrac{1}{4h^2}$, *so folgt:*

$$\Phi(u, v) < \frac{1}{4h^2} \quad \textit{für alle } (u, v) \in B.$$

Setzen wir zusätzlich $\mathfrak{x} \in C^1(\bar{B})$ *voraus, so gilt für*

$$(u_0, v_0) \in \partial B \quad \textit{mit } \Phi(u_0, v_0) = \frac{1}{4h^2}:$$

$$\frac{\partial}{\partial \nu} \Phi|_{(u_0, v_0)} > 0.$$

Beweis. Wie von Hildebrandt [18], S. 271, 272 unter Ausnutzung der konformen Parameter der Fläche nachgewiesen wurde, gilt

$$\Delta\Phi \geqq 0 \quad \text{in } B.$$

Das Maximumprinzip für subharmonische Funktionen und das Randpunktlemma von Hopf [19] ergeben nun die Behauptung des Hilfssatzes. w.z.b.w.

Dieses Maximumprinzip legt es nun nahe, die im Falle $H=0$ auftretenden konvexen Grundgebiete zu ersetzen durch:

Definition 4. Sei $\kappa \geqq 0$ vorgegeben. Ein beschränktes, konvexes Gebiet $\Omega \subset \mathbb{R}^2$, dessen Rand $\partial\Omega$ eine Jordankurve bildet, nennen wir *κ-konvex*, wenn es zu jedem Punkt $(x_0, y_0) \in \partial\Omega$ einen Stützkreis $\partial S_{(x_0, y_0)}$ (bzw. im Falle $\kappa = 0$ eine Stützgerade) der Krümmung κ gibt, so daß

$$\Omega \subset S_{(x_0, y_0)} \quad \text{und} \quad (x_0, y_0) \in \partial S_{(x_0, y_0)} \quad \text{gilt.}$$

Bemerkung. Für ein beschränktes, konvexes Gebiet $\Omega \subset \mathbb{R}^2$ mit $\partial\Omega \in C^2$ gilt: Ω ist κ-konvex genau dann, wenn in jedem Punkte von $\partial\Omega$ die Kurvenkrümmung von $\partial\Omega$ größer oder gleich κ ist.

Wir legen nun die für das folgende wichtige Kurvenklasse fest:

Definition 5. Sei $\kappa \geqq 0$ vorgegeben. Weiter sei $\Omega \subset \mathbb{R}^2$ ein κ-konvexes Gebiet mit regulärer C^2-Randkurve $\partial\Omega$, $f: \partial\Omega \to \mathbb{R} \in C^2(\partial\Omega)$. Wir legen nun drei Punkte (x_ν, y_ν) $\nu = 0, 1, 2$ auf $\partial\Omega$ in positiver Orientierung in bezug auf die x, y-Ebene fest, und setzen

$$\pi_\nu := (x_\nu, y_\nu, f(x_\nu, y_\nu)) \qquad \nu = 0, 1, 2.$$

Die nun konstruierte Jordankurve

$$\Gamma := \{(x, y, f(x, y)) | \quad (x, y) \in \partial\Omega\}$$

mit der durch die Punkte $\{\pi_\nu\}_\nu$ vorgeschriebenen Orientierung nennen wir *zur Klasse $\mathfrak{C}_\kappa$ gehörige Jordankurve.* Ω nennen wir *Projektionsgebiet von Γ.*

Den über Ω errichteten Zylinder bezeichnen wir mit $\mathfrak{Z}_\Omega$, also

$$\mathfrak{Z}_\Omega := \{(x, y, z) \in \mathbb{R}^3 | (x, y) \in \Omega\}.$$

Hilfssatz 4. *Sei $h \geqq 0$, Γ Jordankurve der Klasse $\mathfrak{C}_{2h}$ mit Projektion Ω und $H = H(x, y, z) \in C^\alpha(\mathfrak{Z}_{2h})$ mit $\sup_{\mathfrak{Z}_{2h}} |H(x, y, z)| \leqq h$. Weiter sei $\mathfrak{x}(u, v): \bar{B} \to \mathfrak{Z}_{2h}$ $H(\ )$-Fläche zu Γ.*

Behauptung. Es gilt $\mathfrak{x}(B) \subset \mathfrak{Z}_\Omega$.

Beweis. (nur für $h > 0$).

Annahme. Es gibt ein $(\tilde{u}, \tilde{v}) \in B$ mit $\mathfrak{x}(\tilde{u}, \tilde{v}) \notin \mathfrak{Z}_\Omega$. Da Ω $(2h)$-konvex ist, gibt es einen Stützkreis S an $\partial\Omega$

$$S = \left\{(x, y) \in \mathbb{R}^2 | (x - x_0)^2 + (y - y_0)^2 < \frac{1}{4h^2}\right\},$$

so daß $\mathfrak{x}(\tilde{u},\tilde{v})\notin\mathfrak{Z}_S$. Betrachte nun die Schar von Zylindern

$$\mathfrak{Z}^{(\lambda)}:=\left\{(x,y,z)\,|\,(x-\lambda x_0)^2+(y-\lambda y_0)^2\leqq\frac{1}{4h^2}\right\},\qquad \lambda\in[0,1].$$

Es gilt:

$$\mathfrak{x}(\bar{B})\subset\mathfrak{Z}^{(0)}=\mathfrak{Z}_{2h}$$

und

$$\mathfrak{x}(B)\cap\partial\mathfrak{Z}^{(1)}=\mathfrak{x}(B)\cap\partial\mathfrak{Z}_S\neq\varnothing.$$

Man schließt aus Stetigkeitsgründen, daß es ein $\lambda_0\in[0,1]$ gibt mit

$$\mathfrak{x}(\bar{B})\subset\mathfrak{Z}^{(\lambda_0)},\ \mathfrak{x}(B)\cap\partial\mathfrak{Z}^{(\lambda_0)}\neq\varnothing.$$

Das Maximumsprinzip Hilfssatz 3 auf die Funktion

$$\Phi(u,v):=(x(u,v)-\lambda_0 x_0)^2+(y(u,v)-\lambda_0 y_0)^2$$

angewandt, liefert nun $\Phi\equiv\frac{1}{4h^2}$. Widerspruch!

Wir kommen nun zum Hauptresultat dieses Paragraphen, nämlich

Satz 2 (Transversalität an den Rand). *Sei $h\geqq 0$, $\Gamma\subset\mathfrak{Z}_{2h}$ sei Jordankurve der Klasse $\mathfrak{C}_{2h}$ mit Projektionsgebiet Ω. Weiter sei $H=H(x,y,z)\in C^\alpha(\mathfrak{Z}_{2h})$ $(\alpha\in(0,1))$ mit $\sup_{\mathfrak{Z}_{2h}}|H(x,y,z)|\leqq h$ vorgeschrieben.*

Schließlich sei

$$\mathfrak{x}(u,v)\colon\bar{B}\to\mathfrak{Z}_{2h}$$

Fläche vorgeschriebener mittlerer Krümmung zu Γ.

Behauptung. Für alle $\nu\in(0,1)$ gilt $\mathfrak{x}(u,v)\in C^{1+\nu}(\bar{B})$. Die Fläche $\mathfrak{x}(u,v)$ hat auf ∂B keine Verzweigungspunkte. Die Normale $\mathfrak{X}(u,v)$ an $\mathfrak{x}(u,v)$, die stetig in $\bar{B}$ ist, erfüllt die Bedingung:

$$\mathfrak{X}(u,v)\cdot\mathfrak{e}>0\qquad \textit{für alle } (u,v)\in\partial B.$$

Dabei ist $\mathfrak{e}=(0,0,1)$ gesetzt worden.

Bemerkung. Man siehe hierzu Gulliver, Spruck [8], §2.

Beweis (nur für $h>0$).

1. Da $\Gamma\in\mathfrak{C}_{2h}$ ist, gilt für eine sich einspannende $H(\ )$-Fläche nach Heinz [14]: $\mathfrak{x}(u,v)\in C^{1+\nu}(\bar{B})$ mit $\nu\in(0,1)$. Sei $w_0=(u_0,v_0)=e^{i\vartheta_0}$ ein beliebiger Punkt von ∂B. Da Ω $(2h)$-konvex ist, gibt es zu $(x(w_0),y(w_0))\in\partial\Omega$ einen Stützkreis S vom Radius $\frac{1}{2h}$, der o.E. beschrieben werde durch:

$$S:=\left\{(x,y)\,|\,x^2+y^2<\frac{1}{4h^2}\right\}.$$

Für die Funktion $\Phi(u,v):=(x(u,v))^2+(y(u,v))^2\in C^1(\bar{B})$ gilt nun wegen $\mathfrak{x}(B)\subset\mathfrak{Z}_\Omega \subset\mathfrak{Z}_S$:

$$\Phi(u,v)\leqq\frac{1}{4h^2} \quad \text{in } \bar{B}$$

$$\Phi(u,v)\not\equiv\frac{1}{4h^2} \quad \text{in } B$$

$$\Phi(w_0)=\frac{1}{4h^2}.$$

Setzen wir $w=re^{i\vartheta}$, so folgt nach Hilfssatz 3:

$$(xx_r+yy_r)|_{w_0}=\frac{1}{2}\left.\frac{\partial\Phi}{\partial r}\right|_{w_0}>0. \tag{1}$$

Da nun $\Phi(\vartheta):=\Phi(\cos\vartheta,\sin\vartheta)$ in ϑ_0 das Maximum annimmt, folgt:

$$(xx_\vartheta+yy_\vartheta)|_{w_0}=\frac{1}{2}\left.\frac{\partial\Phi}{\partial\vartheta}\right|_{\vartheta_0}=0. \tag{2}$$

2. In (1) lesen wir nun ab, daß w_0 kein Verzweigungspunkt von $\mathfrak{x}$ ist; genauer:

$$(\mathfrak{x}_\vartheta(w_0))^2=(\mathfrak{x}_r(w_0))^2>0. \tag{3}$$

Da nun $\Gamma\in\mathfrak{C}_{2h}$ ist, gibt es ein $K>0$, so daß gilt:

$$z_\vartheta^2\leqq K(x_\vartheta^2+y_\vartheta^2). \tag{4}$$

Aus (3) und (4) folgt nun:

$$0<(x_\vartheta^2+y_\vartheta^2+z_\vartheta^2)|_{w_0}\leqq(1+K)(x_\vartheta^2+y_\vartheta^2)|_{w_0}. \tag{5}$$

3. Benutzen wir nun, daß die Abbildung

$$(x(u,v),\,y(u,v)):\,\partial B\to\partial\Omega$$

positiv orientiert ist, so gibt es wegen (2) und (5) ein $\lambda>0$ mit:

$$x_\vartheta(w_0)=-\lambda y(w_0)$$
$$y_\vartheta(w_0)=\lambda x(w_0).$$

Mit (1) folgt nun:

$$(x_r y_\vartheta-x_\vartheta y_r)|_{w_0}=\lambda(xx_r+yy_r)|_{w_0}>0.$$

Also erhalten wir:

$$(x_u y_v-x_v y_u)|_{w_0}>0. \quad \text{w.z.b.w.}$$

3. Stabile Flächen vorgeschriebener mittlerer Krümmung mit eineindeutiger Projektion auf eine Ebene

Sei $\mathfrak{O}\subset\mathbb{R}^3$ offen, $H=H(x,y,z)\in C^{1+\alpha}(\mathfrak{O})$ mit $\alpha\in(0,1)$ vorgeschrieben, und $\mathfrak{e}=(0,0,1)$.

Definition 6. Eine $H(\ )$-Fläche $\mathfrak{x}(u,v)\colon \bar{B}\to\bar{\mathfrak{O}}$ nennen wir *stabil*, falls

$$\iint_B |\nabla\varphi|^2\,du\,dv\geqq 2\iint_B q(u,v)\,\varphi^2(u,v)\,du\,dv$$

für alle $\varphi\in C_0^\infty(B)$ gilt. Anderenfalls nennen wir sie *instabil*.

Dabei ist der $\mathfrak{x}$ assoziierte Ausdruck

$$q(u,v):=2E(H(\mathfrak{x}))^2-EK-E(\nabla H\cdot\mathfrak{X})\in C^\alpha(B)$$

gesetzt worden. Man siehe hierzu Satz 1. [Für reguläre Flächen ist im Falle $H=0$ $q\geqq 0$ und q verschwindet nur in den isolierten Nabelpunkten; im Falle $H=\text{const}\neq 0$ ist $q>0$; im Falle $H=$variabel könnte q in B das Vorzeichen wechseln!]

Bemerkung. Diese Definition ist sinnvoll auch für verzweigte Flächen. Über das Verhalten von $q(u,v)$ am Rande von B benötigen wir keine Information.

Zentrale Bedeutung hat nun

Satz 3. *Sei $\mathfrak{O}\subset\mathbb{R}^3$ offen; $H=H(x,y,z)\in C^{1+\alpha}(\mathfrak{O})$ mit $\dfrac{\partial}{\partial z}H(x,y,z)\geqq 0$ sei in $\mathfrak{O}$ vorgeschrieben. Weiter sei $\mathfrak{x}(u,v)\colon \bar{B}\to\bar{\mathfrak{O}}$ (mit $\mathfrak{x}(B)\subset\mathfrak{O}$) eine stabile $H(\ )$-Fläche, deren Normale $\mathfrak{X}(u,v)$ stetig in $\bar{B}$ ist, und $\mathfrak{X}(u,v)\cdot\mathfrak{e}>0$ für alle $(u,v)\in\partial B$ erfüllt.*

Behauptung. $\mathfrak{X}(u,v)\cdot\mathfrak{e}>0$ *für alle* $(u,v)\in\bar{B}$.

Kurz: Eine stabile $H(\ )$-Fläche, die am Rand projizierbar ist, ist auch im Innern projizierbar.

Beweis. Wir betrachten die mit der $H(\ )$-Fläche $\mathfrak{x}$ assoziierte Hilfsfunktion

$$F(u,v):=\mathfrak{X}(u,v)\cdot\mathfrak{e}\in C^{2+\alpha}(B)\cap C^0(\bar{B}).$$

Sie erfüllt $F|_{\partial B}>0$ als Randbedingung und genügt nach Satz 1 der Differentialgleichung:

$$\Delta F+2qF=-2E(\nabla H\cdot\mathfrak{e})\quad\text{in } B.$$

Beachten wir nun $\dfrac{\partial}{\partial z}H\geqq 0$, so folgt die Behauptung mit dem weiter unten bewiesenen Hilfssatz 6. w.z.b.w.

Zunächst formulieren wir noch:

Hilfssatz 5. *Sei $q(u,v)\in C^\alpha(B)$. Weiter sei $F(u,v)\in C^2(B)$ nicht negative Lösung der Differentialgleichung*

$$\Delta F+2qF\leqq 0\quad\text{in } B$$
$$F\geqq 0\quad\text{in } B.$$

Behauptung: Gibt es ein $(u_0, v_0) \in B$ *mit* $F(u_0, v_0) = 0$, *so folgt* $F(u, v) \equiv 0$ *für alle* $(u, v) \in B$.

Beweis. Siehe Heinz [11], Lemma 6.

Hilfssatz 6. *Sei* $q(u, v) \in C^\alpha(B)$ *mit der folgenden Eigenschaft:*

$$\iint\limits_B |\nabla \varphi|^2 \, du dv \geqq 2 \iint\limits_B q \varphi^2 \, du dv \quad \text{für alle} \quad \varphi \in C_0^\infty(B). \tag{*}$$

Nun sei $F(u, v) \in C^2(B) \cap C^0(\bar{B})$ *eine Lösung von*

$$\begin{aligned} \Delta F + 2qF &\leqq 0 \quad \text{in } B \\ F &> 0 \quad \text{auf } \partial B. \end{aligned}$$

Behauptung. $F(u, v) > 0$ *auf* $\bar{B}$.

Beweis. Aufgrund von Hilfssatz 5 reicht es zu zeigen, daß $F(u, v) \geqq 0$ in B gilt. Hierzu betrachten wir die Funktion

$$F^-(u, v) := \min(F(u, v), 0)$$

und weisen nach, daß $F^-(u, v) \equiv 0$ in $\bar{B}$ gilt.

1. Wegen $F|_{\partial B} > 0$ hat $F^- \in C^0(\bar{B})$ für hinreichend großes $\rho \in (0, 1)$ kompakten Träger in der Kreisscheibe

$$b := \{(u, v) \mid u^2 + v^2 < \rho^2\}.$$

Weiter ist $F^- \in H_2^1(b) \cap C_0^0(b)$, und es gilt:

$$\nabla F^- = \begin{cases} 0 & \text{falls } F \geqq 0 \\ \nabla F & \text{falls } F < 0. \end{cases}$$

Ein verallgemeinerter Gaußscher Satz liefert nun:

$$\iint\limits_b |\nabla F^-|^2 = \iint\limits_b \Delta F(-F^-) \leqq 2 \iint\limits_b qF \cdot F^- = 2 \iint\limits_b q |F^-|^2. \tag{1}$$

2. Mit $\psi(u, v) \in C_0^\infty(b)$ setzen wir nun in (*) die Testfunktion

$$\varphi(u, v) := F^-(u, v) + \varepsilon \psi(u, v) \in H_2^1(b)$$

ein, was offenbar zulässig ist ($\varepsilon \in \mathbb{R}$):

$$\begin{aligned} &\iint\limits_b |\nabla F^-|^2 + 2\varepsilon \iint\limits_b \nabla F^- \cdot \nabla \psi + \varepsilon^2 \iint\limits_b |\nabla \psi|^2 \\ &\quad \geqq 2 \iint\limits_b q |F^-|^2 + 4\varepsilon \iint\limits_b q F^- \psi + 2\varepsilon^2 \iint\limits_b q \psi^2. \end{aligned}$$

Wegen (1) folgt nun für alle $\varepsilon \in \mathbb{R}$:

$$2\varepsilon \iint\limits_b (\nabla F^- \cdot \nabla \psi - 2qF^- \psi) + \varepsilon^2 \iint\limits_b (|\nabla \psi|^2 - 2q\psi^2) \geqq 0.$$

Somit gilt für alle $\psi \in C_0^\infty(b)$:

$$\iint_b (\nabla F^- \cdot \nabla \psi - 2qF^- \psi) = 0. \tag{2}$$

3. Mit einem bekannten Regularitätssatz für lineare elliptische Differentialgleichungen (s. Hellwig [17], IV, 4.2) folgt nun

$$F^- \in C_0^2(b).$$

Die Arbeit von Hartman und Wintner [9] zeigt nun:

$$F^-(u, v) \equiv 0 \quad \text{in } \bar{B}. \quad \text{w.z.b.w.}$$

Bemerkungen zu Hilfssatz 6. Hierzu bin ich angeregt worden durch die Abhandlungen von Schwarz zur zweiten Variation des Flächeninhaltes [25], S. 151–167. In diesem Zusammenhang seien auch die Arbeiten von Barbosa, do Carmo [1] und Nitsche [22] genannt. Zum eindimensionalen Fall vergleiche man auch Bolza [3], S. 84–85. Die hier verwendeten Methoden stehen auch in gewissem Zusammenhang mit den Sätzen von Courant über die Knotenlinien der Eigenfunktionen (Courant; Hilbert [4], Band I, VI, §6).

Ebenso wie den Satz über die konforme Abbildung im Großen beweist man

Hilfssatz 7. *Sei $\Gamma \subset \mathbb{R}^3$ eineindeutig projizierbare Jordankurve mit Projektionsgebiet Ω. Sei weiter*

$$\mathfrak{x}(u, v): \bar{B} \to \mathbb{R}^3 \in C^0(\bar{B}) \cap C^k(B) \qquad (k \in \mathbb{R}, k \geqq 2)$$

$H(\ )$-Fläche zu Γ, deren Normale $\mathfrak{X}(u, v)$ die Bedingung erfüllt:

$$\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0 \quad \textit{für alle } (u, v) \in B.$$

Dann ist die Abbildung

$$(x(u, v), y(u, v)): \bar{B} \to \bar{\Omega} \in C^0(\bar{B}) \cap C^k(B)$$

ein C^k-Diffeomorphismus von $\bar{B}$ auf $\bar{\Omega}$.

Beweis. Im Zusammenhang mit Definition 2 erwähnten wir, daß es zu jedem $w_0 = (u_0, v_0) \in B$ einen Vektor $\mathfrak{a} \in \mathbb{C}^3 - \{0\}$ und ein $n \in \mathbb{N} \cup \{0\}$ gibt, so daß

$$\mathfrak{x}_w(w_0 + w) = \mathfrak{a} w^n + o(|w|^n) \quad |w| \to 0 \tag{1}$$

gilt.

Der Fall $n = 0$ entspricht einem regulären Punkt der Fläche, $n > 0$ einem Verzweigungspunkt. Setzen wir $\mathfrak{a} = \mathfrak{a}_1 - i\mathfrak{a}_2$ mit $\mathfrak{a}_1 = (a_1, b_1, c_1)$, $\mathfrak{a}_2 = (a_2, b_2, c_2) \in \mathbb{R}^3$, so gilt:

$$\left(\frac{\mathfrak{a}_1 \times \mathfrak{a}_2}{|\mathfrak{a}_1 \times \mathfrak{a}_2|}, \mathfrak{e} \right) = (\mathfrak{X}(w_0), \mathfrak{e}) > 0.$$

$$\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix} > 0. \tag{2}$$

Wir betrachten nun die Abbildung

$$f(u,v):=x(u,v)+iy(u,v):\bar{B}\to\mathbb{C}.$$

1. Die ersten beiden Komponenten von (1) integrieren wir und erhalten in Polarkoordinaten ($w=re^{i\varphi}$)

$$x(w_0+w)=x(w_0)+\frac{2}{n+1}r^{n+1}(a_1\cos(n+1)\varphi+a_2\sin(n+1)\varphi)+o(r^{n+1})$$

$$y(w_0+w)=y(w_0)+\frac{2}{n+1}r^{n+1}(b_1\cos(n+1)\varphi+b_2\sin(n+1)\varphi)+o(r^{n+1}). \qquad (3)$$

Mit (2) erkennt man nun, daß für hinreichend kleines $\varepsilon>0$ gilt:

$$f(w_0+w)\neq f(w_0) \quad \text{für } 0<|w|<\varepsilon,$$

und daß für den Index

$$i(f,w_0)=n+1\geqq 1$$

richtig ist.

2. Sei nun $z_0=x_0+iy_0\in\mathbb{C}-\partial\Omega$. Dann hat z_0 endliche viele Urbilder $w_\nu=u_\nu+iv_\nu\in B(\nu=1,\ldots,N)$; $f(w_\nu)=z_0$. Nun ist $f:\partial B\to\partial\Omega$ topologische, positiv orientierte Abbildung. Somit folgt:

$$\sum_{\nu=1}^{N} i(f,w_\nu)=\begin{cases}1 & z_0\in\Omega\\ 0 & z_0\in\mathbb{C}-\bar{\Omega}.\end{cases}$$

Da nach 1. $i(f,w_\nu)\in\mathbb{N}-\{0\}$ gilt, folgt, daß es zu $z_0\in\mathbb{C}-\bar{\Omega}$ kein Urbild gibt, also $f(\bar{B})\subset\bar{\Omega}$.

Zu $z_0\in\Omega$ gibt es genau ein Urbild w_0 mit $i(f,w_0)=1$. Somit ist in (3) $n=0$ richtig, und wegen (2) folgt:

$$\left.\frac{\partial(x,y)}{\partial(u,v)}\right|_{w_1}>0.$$

Da nun f eine offene Abbildung ist, folgt $f(B)\subset\Omega$. Wir erhalten also mit $f:\bar{B}\to\bar{\Omega}$ eine topologische Abbildung.

Der Fundamentalsatz über die inverse Abbildung liefert nun die restliche Behauptung. w.z.b.w.

Um mit unseren Überlegungen zu einer Lösung des Dirichletproblems der $H(x,y,z)$-Flächen-Gleichung zu gelangen, stützen wir uns auf die Existenz einer $H(\ )$-Fläche im Zylinder.

Hilfssatz 8 (Gulliver, Spruck: Hildebrandt). *Sei* $h\geqq 0$, $H=H(x,y,z)\in C^{1+\alpha}(\mathfrak{Z}_{2h})$ *mit* $\sup_{\mathfrak{Z}_{2h}}|H(x,y,z)|\leqq h$ *vorgeschrieben.* ($\alpha\in(0,1)$). $\Gamma\subset\mathfrak{Z}_{2h}$ *sei rektifizierbare Jordankurve. Dann gibt es eine* $H(\ )$*-Fläche zu* Γ,

$$\mathfrak{x}(u,v):\bar{B}\to\mathfrak{Z}_{2h}\in C^2(B)\cap C^0(\bar{B})\cap H_2^1(B),$$

die stabil ist im Sinne von Definition 6.

Beweis. In [18] hat Hildebrandt eine $H(\ )$-Fläche konstruiert, die das Energiefunktional $E(\mathfrak{x})$ minimiert. Unter $A(\mathfrak{x})$ verstehen wir das zugehörige parameterinvariante Funktional, das wir im Anhang betrachtet haben.

Indirekt beweisen wir nun, daß $\mathfrak{x}$ stabil ist im Sinne von Definition 6. Wäre $\mathfrak{x}$ instabil, so gibt es eine Funktion $\varphi \in C_0^\infty(B)$:

$$\iint_B |\nabla\varphi|^2 < 2\iint_B q\varphi^2$$

wobei $q(u,v) := -EK + 2EH^2 - E(\nabla H \cdot \mathfrak{X}) \in C^\alpha(B)$ ist. Wir betrachten nun die „normalvariierte" Fläche ($\varepsilon \in \mathbb{R}$)

$$\bar{\mathfrak{x}}(u,v) := \mathfrak{x}(u,v) + \varepsilon\varphi(u,v)\,\mathfrak{X}(u,v) \in C^2(B) \cap C^0(\bar{B}) \cap H_2^1(B).$$

1. Seien $\{w_\nu\}_{\nu=1,\ldots,N}$ die endlich vielen Verzweigungspunkte von $\mathfrak{x}$ in

$$\operatorname{tr}\varphi := \overline{\{w \in B \mid \varphi(w) \neq 0\}},\quad C_\delta := \bigcup_{\nu=1}^{N} \{|w - w_\nu| < \delta\},$$

$B_\delta := B - C_\delta (\delta > 0)$. Wir wählen nun $\delta > 0$ so klein, daß

$$I_\delta := \iint_{C_\delta} (|\nabla\varphi|^2 - 2EK\varphi^2 + 2E(\nabla H \cdot \mathfrak{X})\,\varphi^2)\,du\,dv + \iint_{B_\delta} (|\nabla\varphi|^2 - 2q\varphi^2)\,du\,dv < 0$$

gilt. Nach dem Anhang über die zweite Variation von A gibt es ein $\varepsilon_0 = \varepsilon_0(\delta) > 0$, so daß für $|\varepsilon| \leqq \varepsilon_0(\delta)$ gilt:

$$A[\bar{\mathfrak{x}}] - A[\mathfrak{x}] \leqq \varepsilon^2 \cdot I_\delta + (\varepsilon^2).$$

Also existiert $\varepsilon_1 > 0$, so daß $A[\bar{\mathfrak{x}}] < A[\mathfrak{x}]$ für alle $0 < \varepsilon < \varepsilon_1$ richtig ist.

Außerdem können wir $\bar{\mathfrak{x}}(\bar{B}) \subset \mathfrak{Z}_{2h}$ für alle $\varepsilon \in [0, \varepsilon_1)$ realisieren.

2. Zur nun notwendigen Einführung fast-konformer Parameter verweisen wir auf Morrey [21], Chap. I, Theorem 1.2.

Sei $\eta := \frac{1}{2}(A[\mathfrak{x}] - A[\bar{\mathfrak{x}}]) > 0$. Dann gibt es eine Abbildung

$$\begin{array}{ll} \tau\colon \bar{B} \to \bar{B} & \text{topologisch} \\ e^{\frac{2\pi i}{3}\nu} \mapsto e^{\frac{2\pi i}{3}\nu} & (\nu = 0, 1, 2), \end{array}$$

so daß $\mathfrak{y}(u,v) := \bar{\mathfrak{x}} \circ \tau(u,v) \in H_2^1(B) \cap C^0(\bar{B})$ ist, für τ die Transformationsformel gilt, und so daß gilt:

$$E(\mathfrak{y}) \leqq A(\mathfrak{y}) + \eta.$$

Also folgt:

$$E(\mathfrak{y}) \leqq A(\mathfrak{y}) + \eta = A(\bar{\mathfrak{x}}) + \eta = A(\mathfrak{x}) - 2\eta + \eta = E(\mathfrak{x}) - \eta.$$

Mit $\mathfrak{y} = \bar{\mathfrak{x}} \circ \tau$ haben wir also eine zulässige Fläche gefunden, die kleinere Energie als die Lösung des Variationsproblems hat. Widerspruch! w.z.b.w.

Bemerkung 1. Da wir im obigen Beweis *keinen* Gebrauch von dem sehr aufwendigen Satz gemacht haben, daß Flächen vom Minimumtyp verzweigungspunktfrei in B sind, ist eine genaue Kontrolle der zweiten Variation des Flächenfunktionals in den Verzweigungspunkten nötig. Diese ist im Anhang durchgeführt.

Bemerkung 2. Nach einem Vorschlag von Heinz kann man folgendermaßen im Rahmen der klassischen Räume fast-konforme Parameter einführen:

Sei

$$X(u,v)=(x^1(u,v),\ldots,x^q(u,v)):\bar{B}\to\mathbb{R}^q\in C^2(\bar{B}),\ q\geqq 2.$$
$$E=X_u\cdot X_u,\ F=X_u\cdot X_v,\ G=X_v\cdot X_v.$$

Wir betrachten dann eine assoziierte reguläre Fläche, nämlich zu $\varepsilon>0$:

$$\tilde{X}(u,v)=(x^1(u,v),\ldots,x^q(u,v);\varepsilon u,\varepsilon v):\bar{B}\to\mathbb{R}^{q+2}\in C^2(\bar{B}).$$
$$\tilde{E}=\tilde{X}_u\cdot\tilde{X}_u=E+\varepsilon^2,\ \tilde{F}=\tilde{X}_u\cdot\tilde{X}_v=F,\ \tilde{G}=\tilde{X}_v\cdot\tilde{X}_v=G+\varepsilon^2.$$

Somit ist:

$$\tilde{E}\tilde{G}-\tilde{F}^2=EG-F^2+\varepsilon^2(E+G)+\varepsilon^4>0.$$

Nach den klassischen Sätzen von Korn und Lichtenstein führen wir nun auf $\tilde{X}$ konforme Parameter ein mit der Abbildung:

$$\tau(\alpha,\beta)=u(\alpha,\beta)+iv(\alpha,\beta):\bar{B}\to\bar{B}\in C^1(\bar{B})\,(\text{topologisch}).$$

Setzen wir nun

$$\tilde{Y}(\alpha,\beta):=\tilde{X}\circ\tau(\alpha,\beta)=(X\circ\tau(\alpha,\beta);\varepsilon\cdot\tau(\alpha,\beta))$$

und

$$Y(\alpha,\beta):=X\circ\tau(\alpha,\beta),$$

so gilt also:

$$\tilde{Y}_\alpha\cdot\tilde{Y}_\beta=0=\tilde{Y}_\alpha^2-\tilde{Y}_\beta^2.$$

Bezeichnen wir mit D das Dirichletintegral, so erhalten wir mit der Transformationsformel:

$$D(Y)+\varepsilon^2D(\tau)=D(\tilde{Y})=2\iint_B(\tilde{E}\tilde{G}-\tilde{F}^2)^{1/2}dudv$$
$$\leqq 2\iint_B\sqrt{EG-F^2}\,dudv+2\varepsilon\iint_B\sqrt{E+G}\,dudv+2\varepsilon^2\pi.$$

Zu vorgegebenem $\eta>0$ können wir also ein $\varepsilon>0$ und eine zugehörige Transformation τ finden, so daß für $Y=X\circ\tau$ gilt:

$$D(Y)\leqq 2\cdot\iint_B\sqrt{EG-F^2}\,dudv+\eta.$$

Definition 7. Sei $\Omega \subset \mathbb{R}^2$ offen, $\zeta \in C^2(\Omega)$. Unter "$\mathscr{L}$" verstehen wir den *Minimalflächenausdruck:*

$$\mathscr{L}(\zeta) := (1+\zeta_y^2)\,\zeta_{xx} - 2\zeta_x\,\zeta_y\,\zeta_{xy} + (1+\zeta_x^2)\,\zeta_{yy}.$$

Wir kommen nun zum Hauptergebnis dieses Abschnitts, nämlich

Satz 4. *Sei* $h \geqq 0$, $\Gamma \subset \mathfrak{Z}_{2h}$ *Jordankurve der Klasse* $\mathfrak{C}_{2h}$ *mit Projektionsgebiet* Ω, *und* $H = H(x, y, z) \in C^{1+\alpha}(\mathfrak{Z}_{2h})$ *mit* $\dfrac{\partial}{\partial z} H \geqq 0$ *und* $\sup\limits_{\mathfrak{Z}_{2h}} |H(x, y, z)| \leqq h$ *vorgegeben* ($\alpha \in (0, 1)$).

Behauptung. Es gibt zu Γ *genau eine stabile* $H(\ )$*-Fläche* $\mathfrak{x}(u, v)$: $\bar{B} \to \mathfrak{Z}_{2h}$. *Diese Fläche ist verzweigungspunktfrei in* $\bar{B}$ *und projizierbar auf die* x, y*-Ebene. Sie besitzt also eine Darstellung* $\zeta = \zeta(x, y) \in C^2(\Omega) \cap C^{1+\nu}(\bar{\Omega})$ (*mit* $\nu \in (0, 1)$), *die das folgende Dirichletproblem löst:*

$$\begin{aligned} \mathscr{L}(\zeta) &= 2H(x, y, \zeta(x, y))\,(1+|\nabla\zeta|^2)^{3/2} && \text{in } \Omega \\ \zeta &= f && \text{auf } \partial\Omega \end{aligned} \qquad (*)$$

Beweis. Von Gulliver und Hildebrandt ist im Zylinder $\mathfrak{Z}_{2h}$ eine $H(\ )$-Fläche $\mathfrak{x}$ konstruiert worden, die in unserem Sinne stabil ist (s. hierzu Hilfssatz 8).

Da nun $\Gamma \in C^2$ gilt, folgt für die sich einspannende $H(\ )$-Fläche:

$$\mathfrak{x} \in C^{1+\nu}(\bar{B}) \qquad \nu \in (0, 1).$$

Die Normale $\mathfrak{X}$ an $\mathfrak{x}$ ist stetig in $\bar{B}$ und hat als Randbedingung:

$$\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0 \qquad \text{für alle } (u, v) \in \partial B$$

nach Satz 2.

Weiter ist $\mathfrak{x}_u \neq 0$ auf ∂B.

Da nun $\dfrac{\partial}{\partial z} H \geqq 0$ ist, folgt nun für stabile $H(\ )$-Flächen nach Satz 3:

$$\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0 \qquad \text{für alle } (u, v) \in \bar{B}.$$

Hilfssatz 7 liefert nun, daß die Abbildung

$$(x, y): \bar{B} \to \bar{\Omega} \in C^2(B) \cap C^{1+\nu}(\bar{B})$$

ein Diffeomorphismus ist. Insbesondere ist dann $\mathfrak{x}_u \neq 0$ in $\bar{B}$. Nennen wir $u = u(x, y)$, $v = v(x, y) \in C^2(\Omega) \cap C^{1+\nu}(\bar{\Omega})$ die Umkehrabbildung zu $x = x(u, v)$, $y = y(u, v)$, so erhalten wir in

$$\zeta(x, y) := z(u(x, y), v(x, y)) \in C^2(\Omega) \cap C^{1+\nu}(\bar{\Omega})$$

die Höhendarstellung der $H(\ )$-Fläche $\mathfrak{x}$. Diese Höhendarstellung ζ genügt der angegebenen Differentialgleichung mit den, die Kurve Γ beschreibenden, Randwerten:

$$\zeta|_{\partial\Omega} = f.$$

Da nun das Dirichletproblem (*) wegen $\frac{\partial}{\partial z}H \geqq 0$ höchstens eine Lösung hat (siehe Courant; Hilbert [5] S. 322, 323), so gibt es auch höchstens eine stabile $H(\ )$-Fläche zu Γ. w.z.b.w.

4. Lösbarkeit des Dirichletproblems der nichtparametrischen $H(x, y, z)$-Flächen-Gleichung für stetige Randwerte - Ein Beispiel

Wie in der Einleitung beschrieben, lösen wir nun das Dirichletproblem (*) in Satz 4 zu stetigen Randwerten durch Approximation. Hierzu benötigen wir zunächst

Hilfssatz 9. *Sei* $\kappa \geqq 0$, $\Omega \subset \mathbb{R}^2$ *sei ein* κ*-konvexes Gebiet, und* $f: \partial\Omega \to \mathbb{R}$ *sei stetig. Wir setzen:*

$$\Gamma := \{(x, y, z) \in \mathbb{R}^3 \mid (x, y) \in \partial\Omega,\ z = f(x, y)\} = \{\mathfrak{x}(\vartheta) \mid \vartheta \in [0, 2\pi]\}.$$

Wir können Γ *approximieren durch Kurven* $\Gamma^n = \{\mathfrak{x}^n(\vartheta) \mid \vartheta \in [0, 2\pi]\}$ *der Klasse* $\mathfrak{C}_\kappa$ *mit Projektionsgebieten* $\Omega^n \subset \Omega$, $n = 1, 2, \ldots$. *Es gilt:*

$$\mathfrak{x}^n(\vartheta) \xrightarrow[n \to \infty]{} \mathfrak{x}(\vartheta) \quad \text{gleichmäßig auf } [0, 2\pi]$$

und

$$|\partial\Omega^n| \leqq l \quad n = 1, 2, \ldots \quad \text{mit } l > 0.$$

Bezeichnen wir mit $\eta^n(\varepsilon)$ *die den Kurven* Γ^n *zugeordneten Oszillationsfunktionen* (s. Nitsche [23], §23), *so werden diese gleichmäßig majorisiert durch eine Funktion* $\eta(\varepsilon)$: $[0, \varepsilon_0] \to [0, \infty)$ *mit* $\lim\limits_{\varepsilon \to 0} \eta(\varepsilon) = 0$, *also gilt*

$$\eta^n(\varepsilon) \leqq \eta(\varepsilon) \quad \text{für alle } n \in \mathbb{N}, \varepsilon \in [0, \varepsilon_0].$$

Beweis. Sei $C^k_{2\pi}(\mathbb{R}) := \{g: \mathbb{R} \to \mathbb{R} \mid g \in C^k(\mathbb{R});\ g(\vartheta + 2\pi) = g(\vartheta)\}$.

Der Rand des Gebietes Ω sei o.E. beschrieben durch:

$$\partial\Omega: x(\vartheta) = r(\vartheta)\cos\vartheta,\ y(\vartheta) = r(\vartheta)\sin\vartheta,\ \vartheta \in [0, 2\pi]$$

mit $r(\vartheta) \in C^0_{2\pi}(\mathbb{R})$. Wir konstruieren nun eine Ω ausschöpfende Folge $\{\Omega^n\}_n$ κ-konvexer Gebiete mit $\Omega^n \subset \Omega$, $\partial\Omega^n \in C^2$ und $|\partial\Omega^n| \leqq l$ für $n = 1, 2, \ldots$.

Hierzu beschreiben wir in Ω konvexe Kurven ein, die aus Kreissegmenten der Krümmung κ bestehen. Die Ecken glätten wir dann entsprechend ab.

In der Darstellung von $\partial\Omega^n$:

$$\partial\Omega^n: x(\vartheta) = r^n(\vartheta) \cdot \cos\vartheta,\ y(\vartheta) = r^n(\vartheta) \cdot \sin\vartheta,\ \vartheta \in [0, 2\pi]$$

mit Funktionen $r^n(\vartheta) \in C^2_{2\pi}(\mathbb{R})$ gilt nun:

$$r^n(\vartheta) \to r(\vartheta)\ (n \to \infty) \text{ gleichmäßig in } [0, 2\pi].$$

Die Randfunktion $f(\vartheta)\in C^0_{2\pi}(\mathbb{R})$ approximieren wir gleichmäßig in $[0, 2\pi]$ mit Funktionen $f^n(\vartheta)\in C^2_{2\pi}(\mathbb{R})$. Die so konstruierten Jordankurven

$$\Gamma^n=\{\mathfrak{x}^n(\vartheta)\,|\,\mathfrak{x}^n(\vartheta)=(r^n(\vartheta)\cos\vartheta, r^n(\vartheta)\sin\vartheta, f^n(\vartheta)), \vartheta\in[0, 2\pi]\}$$

gehören zur Klasse $\mathfrak{C}_\kappa$, und es gilt:

$$\mathfrak{x}^n(\vartheta)\to\mathfrak{x}(\vartheta) \quad (n\to\infty) \quad \textit{gleichmäßig auf } [0, 2\pi]. \tag{1}$$

Dabei ist $\mathfrak{x}(\vartheta)=(\mathfrak{r}(\vartheta)\cos\vartheta, \mathfrak{r}(\vartheta)\sin\vartheta, \mathfrak{f}(\vartheta))$ eine Darstellung von Γ.

Wegen (1) ist $\{\mathfrak{x}^n\}_n$ gleichgradig stetig, und wir können somit leicht die Oszillation der Kurven Γ^n gleichmäßig abschätzen. w.z.b.w.

Für eine in unserem Sinne kleine $H(\ )$-Fläche, für die also $\mathfrak{x}(\bar B)\subset\mathfrak{Z}_{2h}$ gilt, kann man a priori die dritte Komponente abschätzen. Setzen wir $\Omega=\{4h^2(x^2+y^2)<1\}$, so ist diese Aussage enthalten in dem folgenden

Hilfssatz 10. *Sei* $h\geqq 0$; $\Omega\subset\mathbb{R}^2$ *ein beschränktes Gebiet mit*

$$\bar\Omega\subset\{(x, y)\,|\,h^2(x^2+y^2)<1\}.$$

Nun sei $H=H(x, y, z)\in C^0(\mathfrak{Z}_\Omega)$ *mit* $\sup_{\mathfrak{Z}_\Omega}|H|\leqq h$ *vorgegeben, wobei wie üblich* $\mathfrak{Z}_\Omega:=\{(x, y, z)\,|\,(x, y)\in\Omega\}$ *ist.*

Wir betrachten nun die „Dosen":

Fall $h=0$: $D_m:=\{(x, y, z)\,|\,(x, y)\in\bar\Omega, |z|\leqq m\}$.

Fall $h>0$: $D_m:=\left\{(x, y, z)\,|\,(x, y)\in\bar\Omega, |z|\leqq m+\left(\frac{1}{h^2}-(x^2+y^2)\right)^{1/2}\right\}$.

Weiter sei $\mathfrak{x}(u, v)\colon \bar B\to\mathfrak{Z}_\Omega$ *eine* $H(\ )$*-Fläche, die*

$\mathfrak{x}(\partial B)\subset D_m$ *und* $\mathfrak{x}(B)\subset\mathfrak{Z}_\Omega$ *erfüllt.*

Behauptung. $\mathfrak{x}(\bar B)\subset D_m$.

Beweis (nur für $h>0$). Wir betrachten die „Dose" D_μ, deren obere Begrenzung wir mit Δ_μ bezeichnen, also:

$$\Delta_\mu:=\left\{(x, y, z)\in\mathbb{R}^3\,|\,(x, y)\in\Omega, z=\mu+\left(\frac{1}{h^2}-(x^2+y^2)\right)^{1/2}\right\}.$$

Wir zeigen nun, daß $\mathfrak{x}(\bar B)\subset D_m$ gilt.

Wenn nicht, so gäbe es ein $\mu>m$, so daß o.E. gilt:

$$\mathfrak{x}(\bar B)\subset D_\mu \quad \text{und} \quad \mathfrak{x}(B)\cap\Delta_\mu\neq\varnothing.$$

Das Maximumprinzip für die Funktion

$$\Phi(u, v):=(x(u, v))^2+(y(u, v))^2+(z(u, v)-\mu)^2$$

zusammen mit einem Fortsetzungsargument liefert nun $\mathfrak{x}(B)\subset\Delta_\mu$ im Widerspruch zu $\mathfrak{x}(\partial B)\subset D_m$. w.z.b.w.

Zur Anwendung des Lemmas von Lebesgue und Courant ist die folgende a priori Abschätzung des Flächeninhaltes einer projizierbaren Fläche notwendig, die man mit der in [16], S. 155 angegebenen Anleitung beweisen kann:

Hilfssatz 11. *Sei* $h \geqq 0$, $H = H(x, y, z) \in C^{1+\alpha}(\mathfrak{Z}_{2h})$ *mit* $\sup_{\mathfrak{Z}_{2h}} |H(x, y, z)| \leqq h$. $\Omega \subset \mathbb{R}^2$ *sei ein beschränktes Gebiet, mit* C^2*-Jordankurve* $\partial\Omega$ *als Rand, für dessen Länge* $|\partial\Omega|$ *gilt:*

$$|\partial\Omega| \leqq l.$$

Die Funktion $\zeta = \zeta(x, y) \in C^2(\Omega) \cap C^1(\bar{\Omega})$ *genüge*

$$\mathscr{L}(\zeta) = 2H(x, y, \zeta(x, y))(1 + |\nabla\zeta|^2)^{3/2} \quad \text{in } \Omega \tag{1}$$

und

$$|\zeta(x, y)| \leqq M \quad \text{für alle } (x, y) \in \bar{\Omega}. \tag{2}$$

Behauptung. Es gibt eine Konstante $c(h, l, M) > 0$ *so daß*

$$\iint_{\Omega} \sqrt{1 + |\nabla\zeta|^2}\, dx\, dy \leqq c(h, l, M)$$

für alle Lösungen ζ *von* (1) *und* (2) *richtig ist.*

Entscheidende Bedeutung für unsere Approximation in Satz 5 hat die folgende Aussage über die „Abgeschlossenheit" der Klasse der projizierbaren $H(\ )$-Flächen. (Man vergleiche hierzu Heinz [16], Lemma 2 für den Fall $H = H(x, y)$.)

Hilfssatz 12. *Sei* $\mathfrak{O} \subset \mathbb{R}^3$ *offen;* $H = H(x, y, z) \in C^{1+\alpha}(\mathfrak{O})$ *mit* $\dfrac{\partial}{\partial z} H \geqq 0$ *sei vorgeschrieben. Sei* $\mathfrak{x}^n(u, v)$: $B \to \mathfrak{O}$ $(n = 1, 2, \ldots)$ *eine Folge von* $H(\ )$*-Flächen, für deren Normalen* $\mathfrak{X}^n(u, v)$ *gilt:*

$$\mathfrak{e} \cdot \mathfrak{X}^n(u, v) > 0 \quad \text{für alle } (u, v) \in B,\ n \in \mathbb{N}.$$

Weiter gebe es eine $H(\ )$*-Fläche* $\mathfrak{x}(u, v)$: $B \to \mathfrak{O}$, *so daß deren Normale* $\mathfrak{X}(u, v)$ *die Bedingung* $\mathfrak{X}(u, v) \cdot \mathfrak{e} \not\equiv 0$ *erfüllt. Weiter gelte für jede kompakte Teilmenge* $b \subset B$*:*

$$\begin{matrix} \mathfrak{x}_u^n \to \mathfrak{x}_u \\ \mathfrak{x}_v^n \to \mathfrak{x}_v \end{matrix} \quad (n \to \infty) \text{ gleichmäßig in } b. \tag{*}$$

Behauptung. $\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0$ *für alle* $(u, v) \in B$.

Beweis. 1. Es gilt $\mathfrak{X}(u, v) \cdot \mathfrak{e} \geqq 0$ in B; denn wenn es ein $(u_0, v_0) \in B$ gäbe mit $\mathfrak{X}(u_0, v_0) \cdot \mathfrak{e} < 0$, so gäbe es ein $(u_1, v_1) \in B$ mit $\mathfrak{X}(u_1, v_1) \cdot \mathfrak{e} < 0$ und $\mathfrak{x}_u(u_1, v_1) \neq 0$ (denn die Verzweigungspunkte von $\mathfrak{x}$ sind isoliert). Wegen (*) ist dann für $n \geqq n_0$ $\mathfrak{x}_u^n(u_1, v_1) \neq 0$, und es gilt:

$$\mathfrak{X}^n(u_1, v_1) \cdot \mathfrak{e} \xrightarrow[n \to \infty]{} \mathfrak{X}(u_1, v_1) \cdot \mathfrak{e} < 0.$$

Dies steht aber im Widerspruch zu: $\mathfrak{X}^n(u_1, v_1) \cdot \mathfrak{e} > 0$ für $n = 1, 2, \ldots$.

2. Es gilt sogar $\mathfrak{X}(u, v) \cdot \mathfrak{e} > 0$ für alle $(u, v) \in B$. Würde nämlich ein $(u_0, v_0) \in B$ existieren mit $\mathfrak{X}(u_0, v_0) \cdot \mathfrak{e} = 0$, so betrachten wir die Funktion

$$\Phi(u, v) := \mathfrak{X}(u, v) \cdot \mathfrak{e} \geqq 0 \text{ in } B.$$

Nach Satz 1 genügt Φ in B der Differentialgleichung ($q := E(2H^2 - K - \nabla H \cdot \mathfrak{x})$):

$$\Delta\Phi + 2q\Phi = -2E(\nabla H \cdot \mathfrak{e}) \leqq 0.$$

Hilfssatz 5 liefert nun $\Phi \equiv 0$ im Widerspruch zur Voraussetzung. w.z.b.w.

Satz 5. *Sei* $h \geqq 0$, $\Omega \subset \{(x, y) \in \mathbb{R}^2 \mid 4h^2(x^2 + y^2) \leqq 1\}$ *sei ein* $(2h)$*-konvexes Gebiet in der* x, y*-Ebene. Weiter sei*

$$H = H(x, y, z) \in C^{1+\alpha}(\mathfrak{Z}_{2h}) \text{ mit } \sup_{\mathfrak{Z}_{2h}} |H(x, y, z)| \leqq h$$

und $\dfrac{\partial}{\partial z} H \geqq 0$ *auf* $\mathfrak{Z}_{2h}$ *vorgeschrieben* ($\alpha \in (0, 1)$).

Behauptung. Zu jeder stetigen Funktion $f: \partial\Omega \to \mathbb{R}$ *gibt es genau eine Funktion* $\zeta = \zeta(x, y) \in C^2(\Omega) \cap C^0(\bar{\Omega})$, *die das Dirichletproblem*

$$\begin{aligned} \mathfrak{L}(\zeta) = 2H(x, y, \zeta(x, y))(1 + |\nabla\zeta|^2)^{3/2} \quad & \text{in } \Omega \\ \zeta = f \quad & \text{auf } \partial\Omega \end{aligned} \tag{*}$$

für die $H(x, y, z)$*-Flächengleichung löst.*

Beweis. Zunächst besitzt (*) wegen $\dfrac{\partial}{\partial z} H \geqq 0$ höchstens eine Lösung (siehe Courant, Hilbert [5], Band 2, S. 322, 323). Eine Lösung von (*) konstruieren wir auf der Grundlage von Satz 4 folgendermaßen:

1. Sei $\Gamma := \{(x, y, f(x, y)) \mid (x, y) \in \partial\Omega\}$, und $\{\pi_\nu\}_{\nu = 0, 1, 2}$ seien drei Punkte auf Γ in positiver Orientierung in bezug auf die (x, y)-Ebene.

Gemäß Hilfssatz 9 approximieren wir Γ durch Kurven Γ^n der Klasse $\mathfrak{C}_{2h}$ mit den Projektionsgebieten $\Omega^n (n = 1, 2, \ldots)$.

Zu jeder Kurve Γ^n konstruieren wir nun mit Satz 4 eine projizierbare $H(\ )$-Fläche $\mathfrak{x}^n(u, v)$: $\bar{B} \to \mathfrak{Z}_{2h}$ $(n = 1, 2, \ldots)$, für die also

$$\frac{\partial(x^n, y^n)}{\partial(u, v)} > 0, (u, v) \in \bar{B}$$

für alle $n = 1, 2, \ldots$ gilt. Durch Normierung können wir erreichen, daß

$$\mathfrak{x}^n(e^{\frac{2\pi i}{3}\nu}) \to \pi_\nu (n \to \infty); \ \nu = 0, 1, 2; \text{ richtig ist.}$$

2. Zunächst gilt:

$$\sup_{\partial B} |\mathfrak{x}^n(u, v)| = \sup_{\pi \in \Gamma^n} |\pi| \leqq \text{const } (n = 1, 2 \ldots),$$

also folgt mit Hilfssatz 10:

$$\sup_{B} |\mathfrak{x}^n(u, v)| \leqq \text{const } (n = 1, 2, \ldots).$$

Bezeichnen wir mit $\zeta^n(x, y)$: $\bar{\Omega}^n \to \mathbb{R}$ die Höhendarstellung die projizierbaren Fläche $\mathfrak{x}^n(u, v)$, so folgt wegen $|\partial\Omega^n| \leqq l$ $(n = 1, 2, \ldots)$ mit Hilfssatz 11:

$$D[\mathfrak{x}^n] = 2 \iint_{\Omega^n} \sqrt{1 + |\nabla\zeta^n|^2}\, dx\, dy \leqq \text{const} \quad (n = 1, 2, \ldots).$$

$\{\mathfrak{x}^n\}_n$ ist also eine beschränkte Funktionenfolge mit gleichmäßig beschränktem Dirichletintegral.

3. Zum Nachweis der gleichgradigen Stetigkeit der Folge $\{\mathfrak{x}^n\}_n$ mit dem Oszillationslemma von Lebesgue und Courant sind die folgenden Umgebungen von besonderer Bedeutung: Zu $\mathfrak{a}=(a,b,c)\in\mathbb{R}^3$ definieren wir:

Im *Fall* $h=0$ setzen wir:

$$D_\varepsilon(\mathfrak{a}):=\{(x,y,z)\in\mathbb{R}^3\,|\,(x-a)^2+(y-b)^2\leqq\varepsilon^2,\ |z-c|\leqq\varepsilon\}.$$

Im *Fall* $h>0$ setzen wir:

$$D_\varepsilon(\mathfrak{a}):=\{(x,y,z)\in\mathbb{R}^3|(x-a)^2+(y-b)^2\leqq\varepsilon^2, |z-c|\leqq\varepsilon + (1/h^2-(x-a)^2-(y-b)^2)^{1/2}-(1/h^2-\varepsilon^2)^{1/2}\}.$$

Bezeichne $V_\delta(w_0):=\{w\in\bar{B}||w-w_0|\leqq\delta\}$ eine Umgebung um ein $w_0\in\bar{B}$.

Da nun die Oszillation der Kurven Γ^n nach Hilfssatz 9 durch eine Funktion $\eta(\varepsilon)$ majorisiert wird, können wir mit dem Lemma von Lebesgue und Courant folgendes nachweisen:

Zu vorgegebenem $\varepsilon>0$ gibt es ein $\delta\in(0,1)$ mit der Eigenschaft: Es gibt ein $\delta^*=\delta^*(w_0,n)\in[\delta,\sqrt{\delta}]$, so daß gilt:

$$|\mathfrak{x}^n(w_1)-\mathfrak{x}^n(w_2)|\leqq\varepsilon \quad \text{für alle } w_1,w_2\in\partial V_{\delta^*}.$$

Somit gibt es ein $\mathfrak{a}\in\mathbb{R}^3$ so daß gilt:

$$\mathfrak{x}^n(\partial V_{\delta^*})\subset D_\varepsilon(\mathfrak{a}),$$

und mit Hilfssatz 10 folgt:

$$\mathfrak{x}^n(V_\delta)\subset\mathfrak{x}^n(V_{\delta^*})\subset D_\varepsilon(\mathfrak{a}).$$

Also ist $\{\mathfrak{x}^n\}_n$ gleichgradig stetige Funktionenfolge.

4. Mit $\mathfrak{x}^n(u,v)\colon\bar{B}\to\mathfrak{Z}_{2h}$ haben wir also eine gleichgradig stetige, beschränkte Funktionenfolge, welche der Differentialungleichung

$$|\Delta\mathfrak{x}|\leqq h(\mathfrak{x}_u^2+\mathfrak{x}_v^2) \quad \text{in } B$$

genügt. Nach Heinz [12], Hilfssatz 5 kann man nun a priori im Innern die $(1+\alpha)$-Norm abschätzen: $B_r:=\{(u,v)\in\mathbb{R}^2|u^2+v^2<r^2\}$. Zu $r,\alpha\in(0,1)$ ist:

$$\sup_{n\in\mathbb{N}}\|\mathfrak{x}^n\|^{B_r}_{1+\alpha}<+\infty.$$

Mit einem Schluß der Potentialtheorie folgt:

$$\sup_{n\in\mathbb{N}}\|\mathfrak{x}^n\|^{B_r}_{2+\alpha}<+\infty.$$

Mit dem Satz von Arzelà-Ascoli können wir nun eine Teilfolge auswählen, die wir wiederum mit $\{\mathfrak{x}^n\}_n$ bezeichnen, die auf $\bar{B}$ gleichmäßig gegen eine Funktion $\mathfrak{x}\colon\bar{B}\to\mathfrak{Z}_{2h}$ konvergiert, und deren erste und zweite Ableitungen im Innern von

B gegen eine Funktion

$$\mathfrak{x}\in C^2(B)\cap C^0(\bar{B})$$

konvergieren. Also genügt $\mathfrak{x}$ den Differentialgleichungen

$$\Delta\mathfrak{x}=2H(\mathfrak{x})\,\mathfrak{x}_u\times\mathfrak{x}_v$$
$$\mathfrak{x}_u\cdot\mathfrak{x}_v=0=\mathfrak{x}_u^2-\mathfrak{x}_v^2 \quad \text{in } B,$$

durchläuft die Kurve Γ zunächst schwach monoton, und dann mit dem üblichen Schluß auch stark monoton.

5. Da nun $\mathfrak{X}^n(u,v)\cdot\mathfrak{e}>0$ für $(u,v)\in B$, $n=1,2,\ldots$ gilt, folgt nach Hilfssatz 12 $\mathfrak{X}(u,v)\cdot\mathfrak{e}>0$ für $(u,v)\in B$, wobei $\mathfrak{X}$ die Normale an $\mathfrak{x}$ bezeichnet. Hilfssatz 7 zeigt dann, daß die Abbildung $(x(u,v), y(u,v)):\bar{B}\to\mathbb{R}^2$ einen Diffeomorphismus der Klasse C^2 darstellt.

Die Höhendarstellung $\zeta=\zeta(x,y)\in C^2(\Omega)\cap C^0(\bar{\Omega})$ löst dann das Dirichletproblem (*) der $H(\ )$-Flächengleichung für stetige Daten. w.z.b.w.

Bemerkung 1. Ist Ω ein $(2h)$-konvexes Gebiet in der Ebene, dessen Rand durch die stetige, schwach monotone Abbildung $(\varphi(\vartheta), \psi(\vartheta)):[0,2\pi]\to\partial\Omega$ parametrisiert sei, und bezeichne $\chi(\vartheta):[0,2\pi]\to\mathbb{R}$ eine beliebig vorgegebene, stetige, periodische Funktion auf $[0,2\pi]$, so können wir auch zu der so entstehenden Jordankurve

$$\Gamma:=\{x=\varphi(\vartheta), y=\psi(\vartheta), z=\chi(\vartheta)|\vartheta\in[0,2\pi]\}$$

eine projizierbare $H(\ )$-Fläche konstruieren. Diese Jordankurve Γ kann auch vertikale Kurvenstücke enthalten.

Bemerkung 2. Ein anderer Zugang zum Dirichletproblem (*), nämlich im Rahmen einer Theorie für quasilineare Differentialgleichungen in mehreren Variablen, wird durch die Arbeiten von Serrin [26, 27] gegeben.

Zum Schluß behandeln wir nun noch eine wichtige Gleichung der Kapillaritätstheorie, zu der man nun=bis zu stetigen Daten=einen rein parameterischen Zugang hat:

Beispiel (Gauß-Laplace-Gleichung). Sei $h>0$, $\Omega\subset\mathbb{R}^2$ sei ein $(2h)$-konvexes Gebiet, $c>0$. Weiter sei $f:\partial\Omega\to\mathbb{R}\in C^0(\partial\Omega)$ vorgegeben, mit $\sup_{\partial\Omega}|f|<\frac{2h}{c}$. Dann gibt es genau eine Lösung $\zeta(x,y)\in C^2(\Omega)\cap C^0(\bar{\Omega})$ des Dirichletproblems:

$$\begin{aligned}\mathcal{L}(\zeta)=c\cdot\zeta(1+|\nabla\zeta|^2)^{3/2} &\quad \text{in } \Omega\\ \zeta=f &\quad \text{auf } \partial\Omega.\end{aligned}\tag{*}$$

Beweis. 1. Wegen $c>0$ ist (*) eindeutig lösbar.

2. Setze $\varepsilon:=\frac{2h}{c}-\sup_{\partial\Omega}|f|$ und konstruiere $H(z)\in C^\infty(\mathbb{R})$, so daß $\frac{\partial}{\partial z}H(z)\geqq 0$ für alle $z\in\mathbb{R}$ gilt, und so daß

$$H(z)=\begin{cases}\frac{1}{2}cz & |z|\leqq\frac{2h}{c}-\varepsilon\\ -h & z\leqq-\frac{2h}{c}\\ h & z\geqq\frac{2h}{c}\end{cases}$$

richtig ist. Mit Satz 5 lösen wir nun das Dirichletproblem

$$\begin{aligned}\mathscr{L}(\zeta)=2H(\zeta)(1+|\nabla\zeta|^2)^{3/2} &\quad \text{in } \Omega\\ \zeta=f &\quad \text{auf } \partial\Omega.\end{aligned}$$

Mit dem Maximumprinzip stellt man nun fest, daß gilt:

$$\sup_{\Omega}|\zeta(x,y)|\leqq\sup_{\partial\Omega}|\zeta(x,y)|=\frac{2h}{c}-\varepsilon.$$

Somit löst ζ auch das Dirichletproblem (*).

Anhang. *Die zweite Variation des Flächenfunktionals ($H=H(\mathfrak{x})$ vorgeschrieben)*

Sei $\mathfrak{D}\subset\mathbb{R}^3$ offen, $H=H(x,y,z)\in C^{1+\alpha}(\mathfrak{D})$ vorgeschrieben. Sei weiter $\mathfrak{x}(u,v)$: $\bar{B}\to\bar{\mathfrak{D}}\in C^2(B)\cap C^0(\bar{B})\cap H_2^1(B)$ Fläche vorgeschriebener mittlerer Krümmung $H(x,y,z)$ mit endlichem Flächeninhalt. Nach § 1, Satz 1 ist die Normale $\mathfrak{X}(u,v)$ auch in den Verzweigungspunkten von $\mathfrak{x}$ differenzierbar, genauer gilt:

$$\mathfrak{X}(u,v)\in C^{2+\alpha}(B).$$

Wir betrachten nun eine Normalvariation der Fläche $\mathfrak{x}$, d.h. mit $\varphi\in C_0^\infty(B)$ die Fläche

$$\bar{\mathfrak{x}}(u,v)=\mathfrak{x}(u,v)+\varepsilon\,\varphi(u,v)\cdot\mathfrak{X}(u,v),\qquad (u,v)\in B,$$

wobei ε reeller Parameter ist.

Mit $H(x,y,z)$ ist folgendes parameterinvariante Funktional assoziiert:

$$A[\mathfrak{x}]:=2\mathbf{O}[\mathfrak{x}]+4\mathbf{V}[\mathfrak{x}]$$

mit

$$\mathbf{O}[\mathfrak{x}]:=\iint_B|\mathfrak{x}_u\times\mathfrak{x}_v|\,du\,dv \qquad \textit{Oberfläche}$$

$$\mathbf{V}[\mathfrak{x}]:=\iint_B(Q(\mathfrak{x}),\mathfrak{x}_u,\mathfrak{x}_v)\,du\,dv \qquad \textit{Volumen},$$

wobei $Q=Q(x,y,z)$: $\mathfrak{D}\to\mathbb{R}^3$ ein Vektorfeld mit $\operatorname{div}Q(x,y,z)=H(x,y,z)$ ist.

1. Seien $\{w_\nu\}_{\nu=1,\ldots,N}$ die endlich vielen Verzweigungspunkte von $\mathfrak{x}$ in $\operatorname{tr}\varphi:=\overline{\{w\in B|\varphi(w)\neq 0\}}$. Für $\delta>0$ hinreichend klein sei

$$C_\delta:=\bigcup_{\nu=1}^{N}\{w\in B||w-w_\nu|<\delta\}\text{ und }B_\delta:=B-C_\delta$$

erklärt. In B_δ gilt für $|\varepsilon| \leqq \varepsilon_0(\delta)$ (siehe Blaschke [2], § 116):

$$|\bar{\mathfrak{x}}_u \times \bar{\mathfrak{x}}_v| - |\mathfrak{x}_u \times \mathfrak{x}_v| = -2\varepsilon EH\varphi + \frac{\varepsilon^2}{2}(|\nabla\varphi|^2 + 2EK\varphi^2) + o(\varepsilon^2).$$

Für $|\varepsilon| \leqq \varepsilon_0(\delta)$ folgt:

$$2\iint_{B_\delta}(|\bar{\mathfrak{x}}_u \times \bar{\mathfrak{x}}_v| - |\mathfrak{x}_u \times \mathfrak{x}_v|) = -4\varepsilon\iint_{B_\delta} EH\varphi + \varepsilon^2\iint_{B_\delta}(|\nabla\varphi|^2 + 2EK\varphi^2) + o(\varepsilon^2).$$

Weiter gilt in C_δ:

$$2\iint_{C_\delta}(|\bar{\mathfrak{x}}_u \times \bar{\mathfrak{x}}_v| - |\mathfrak{x}_u \times \mathfrak{x}_v|) \leqq \iint_{C_\delta}(\bar{E} + \bar{G} - 2E) = -4\varepsilon\iint_{C_\delta} EH\varphi + \varepsilon^2\iint_{C_\delta}(|\nabla\varphi|^2 - 2EK\varphi^2 + 4EH^2\varphi^2).$$

Also gilt für $|\varepsilon| \leqq \varepsilon_0(\delta)$:

$$2(\mathbf{O}[\bar{\mathfrak{x}}] - \mathbf{O}[\mathfrak{x}]) \leqq -4\varepsilon\iint_B EH\varphi\, du\, dv + \varepsilon^2\Big(\iint_{C_\delta}(|\nabla\varphi|^2 - 2EK\varphi^2 + 4EH^2\varphi^2)\, du\, dv + \iint_{B_\delta}(|\nabla\varphi|^2 + 2EK\varphi^2)\, du\, dv\Big) + o(\varepsilon^2)$$

2. Wir betrachten nun die Variation des Volumens:
Sei $Q = Q(x, y, z) = (a(x, y, z),\ b(x, y, z),\ c(x, y, z)) \in C^1$ das Vektorfeld mit $\operatorname{div} Q(x, y, z) = H(x, y, z)$. Mit

$$\partial Q = \begin{pmatrix} a_x & a_y & a_z \\ b_x & b_y & b_z \\ c_x & c_y & c_z \end{pmatrix}$$

bezeichnen wir dessen Funktionalmatrix. Wir betrachten nun das Feld:

$$\mathfrak{y}(u, v; t) := \mathfrak{x}(u, v) + t\,\varphi(u, v)\,\mathfrak{X}(u, v), (u, v) \in \bar{B},\ |t| \leqq \varepsilon_0$$

und setzen $\mathfrak{q}(u, v; t) := Q(\mathfrak{y}(u, v; t))$.

Nun berechnen wir:

$$\begin{aligned}
&\frac{\partial}{\partial t}(\mathfrak{q}(u, v; t), \mathfrak{y}_u(u, v; t), \mathfrak{y}_v(u, v; t)) \\
&= (\mathfrak{q}_t, \mathfrak{y}_u, \mathfrak{y}_v) + (\mathfrak{q}, (\varphi\mathfrak{X})_u, \mathfrak{y}_v) + (\mathfrak{q}, \mathfrak{y}_u, (\varphi\mathfrak{X})_v) \\
&= (\partial Q \cdot \mathfrak{y}_u, \mathfrak{y}_v, \mathfrak{y}_t) + (\partial Q \cdot \mathfrak{y}_v, \mathfrak{y}_t, \mathfrak{y}_u) + (\partial Q \cdot \mathfrak{y}_t, \mathfrak{y}_u, \mathfrak{y}_v) \\
&\quad + ((\mathfrak{q}, \varphi\mathfrak{X}, \mathfrak{y}_v)_u + (\mathfrak{q}, \mathfrak{y}_u, \varphi\mathfrak{X})_v) \\
&= (\mathfrak{y}_u, \mathfrak{y}_v, \mathfrak{y}_t) \cdot H(\mathfrak{y}(u, v, t)) + ((\mathfrak{q}, \varphi\mathfrak{X}, \mathfrak{y}_v)_u + (\mathfrak{q}, \mathfrak{y}_u, \varphi\mathfrak{X})_v).
\end{aligned}$$

Somit folgt:

$$\frac{\partial}{\partial t}\mathbf{V}[\mathfrak{y}]=\iint\limits_B H(\mathfrak{y}(u,v,t))(\mathfrak{y}_u,\mathfrak{y}_v,\mathfrak{y}_t)\,du\,dv.$$

Mit den Ableitungsgleichungen berechnen wir nun:

$$(\mathfrak{y}_u,\mathfrak{y}_v,\mathfrak{y}_t)=\mathrm{E}\varphi-2\mathrm{EH}(\mathfrak{x})\cdot\varphi^2 t+EK\varphi^3 t^2.$$

Wir erhalten dann:

$$\frac{\partial}{\partial t}\mathbf{V}[\mathfrak{y}]\big|_{t=0}=\iint\limits_B EH(\mathfrak{x}(u,v))\cdot\varphi(u,v)\,du\,dv$$

und

$$\frac{\partial}{\partial t^2}\mathbf{V}[\mathfrak{y}]\big|_{t=0}=\iint\limits_B (E(\nabla H\cdot\mathfrak{X})\varphi^2-2EH^2\varphi^2)\,du\,dv.$$

Also folgt:

$$\begin{aligned}&4(\mathbf{V}[\bar{\mathfrak{x}}]-\mathbf{V}[\mathfrak{x}])\\&\quad=4\varepsilon\iint\limits_B EH\varphi\,du\,dv+\varepsilon^2\iint\limits_B(2E(\nabla H\cdot\mathfrak{X})\varphi^2-4EH^2\varphi^2)\,du\,dv+o(\varepsilon^2).\end{aligned}$$

3. Zusammenfassung: Zu jedem $\delta>0$ gibt es ein $\varepsilon_0=\varepsilon_0(\delta)>0$, so daß für $|\varepsilon|\leqq\varepsilon_0(\delta)$ die folgende Ungleichung richtig ist:

$$\begin{aligned}&(A[\bar{\mathfrak{x}}]-A[\mathfrak{x}])\\&\quad\leqq\varepsilon^2\Big(\iint\limits_{C_\delta}(|\nabla\varphi|^2-2EK\varphi^2+2E(\nabla H\cdot\mathfrak{X})\varphi^2)\,du\,dv\\&\qquad+\iint\limits_{B_\delta}(|\nabla\varphi|^2+2EK\varphi^2-4EH^2\varphi^2+2E(\nabla H\cdot\mathfrak{X})\varphi^2)\,du\,dv\Big)\\&\qquad+o(\varepsilon^2).\end{aligned}$$

Literatur

1. Barbosa, J.L., do Carmo, M.: On the size of a stable minimal surface in $\mathbb{R}^3$. Amer. J. Math. **98**, 515-528 (1976)
2. Blaschke, W.: Vorlesung über Differentialgeometrie. I. Elementare Differentialgeometrie, 4. Aufl. Berlin: Springer 1945
3. Bolza, O.: Vorlesungen über Variationsrechnung. Leipzig, Berlin: Teubner 1909
4. Courant, R., Hilbert, D.: Methoden der Mathematischen Physik I, 3. Aufl. Berlin-Heidelberg-New York: Springer 1968
5. Courant, R., Hilbert, D.: Methods of Mathematical Physics II. New York: Interscience 1962
6. Douglas, J.: Solution of the problem of Plateau. Trans. Amer. Math. Soc. **33**, 263-321 (1931)
7. Gulliver, R., Spruck, J.: The Plateau Problem for surfaces of prescribed mean curvature in a cylinder. Invent. Math. **13**, 169-178 (1971)
8. Gulliver, R., Spruck, J.: Surfaces of constant mean curvature which have a simple projection. Math. Z. **129**, 95-107 (1972)
9. Hartman, P., Wintner, A.: On the local behavior of solutions of nonparabolic partial differential equations. Amer. J. Math. **75**, 449-476 (1953)

10. Heinz, E.: Über die Existenz einer Fläche konstanter mittlerer Krümmung bei vorgegebener Berandung. Math. Ann **127**, 258–287 (1954)
11. Heinz, E.: On certain nonlinear elliptic differential equations and univalent mappings. J. Analyse Math. **5**, 197–272 (1956/57)
12. Heinz, E.: Neue a-priori-Abschätzungen für den Ortsvektor einer Fläche positiver Gauß'scher Krümmung durch ihr Linienelement. Math. Z. **74**, 129–157 (1960)
13. Heinz, E.: Existence theorems for one-to-one mappings associated with elliptic systems of second order I. J. Analyse Math. **15**, 325–352 (1965)
14. Heinz, E.: Über das Randverhalten quasilinearer elliptischer Systeme mit isothermen Parametern. Math. Z. **113**, 99–105 (1970)
15. Heinz, E., Hildebrandt, S.: The number of branch points of surfaces of bounded mean curvature. J. Differential Geometry **4**, 227–235 (1970)
16. Heinz, E.: Interior gradient estimates for surfaces $z=f(x,y)$ with prescribed mean curvature. J. Differential Geometry **5**, 149–157 (1971)
17. Hellwig, G.: Partial differential equations. New York: Blaisdell 1964
18. Hildebrandt, S.: Über einen neuen Existenzsatz für Flächen vorgeschriebener mittlerer Krümmung. Math. Z. **119**, 267–272 (1971)
19. Hopf, E.: A remark on linear elliptic differential equations of second order. Proc. Amer. Math. Soc. **3**, 791–793 (1952)
20. Kneser, H.: Lösung der Aufgabe 41. Jber. Deutsch. Math.-Verein. **35**, 123–124 (1926)
21. Morrey, C.B.: The problem of Plateau on a Riemannian manifold. Ann. of Math. (2) **49**, 807–851 (1948)
22. Nitsche, J.C.C.: A new uniqueness theorem for minimal surfaces. Arch. Rational Mech. Anal. **52**, 319–329 (1973)
23. Nitsche, J.C.C.: Vorlesungen über Minimalflächen. Berlin-Heidelberg-New York: Springer 1975
24. Radó, T.: The problem of the least area and the problem of Plateau. Math. Z. **32**, 763–796 (1930)
25. Schwarz, H.A.: Gesammelte Mathematische Abhandlungen, Erster Band. Berlin: Springer 1890
26. Serrin, J.: The problem of Dirichlet for quasi-linear elliptic differential equations with many independent variables. Phil. Trans. Roy. Soc. London Ser. A**264**, 413–496 (1969)
27. Serrin, J.: The Dirichlet problem for surfaces of constant mean curvature. Proc. London Math. Soc. **21**, 361–384 (1970)
28. Tomi, F.: Ein einfacher Beweis eines Regularitätssatzes für schwache Lösungen gewisser elliptischer Systeme. Math. Z. **112**, 214–218 (1969)

Eingegangen am 20. Oktober 1981

Math. Z. 180, 69–77 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

The Gap Phenomena of Yang-Mills Fields over the Complete Manifold

Chun-li Shen*

Mathematisches Institut der Universität,
Sonderforschungsbereich "Theoretische Mathematik" SFB40, Beringstraße 4,
D-5300 Bonn 1, Federal Republic of Germany
and Department of Mathematics, Fudan University,
Shanghai, People's Republic of China

1. Introduction

It is a well known problem whether a sourceless $SU(2)$ gauge field over S^4 is self-dual (or anti-self-dual). So it is very interesting to study the relation between the sourceless fields and the self-dual (or anti-self-dual) fields.

In [5], we proved the following

Theorem A. *Let M be a four-dimensional self-dual compact Riemannian manifold with positive scalar curvature R, and assume that there is a sourceless $SU(N)$ gauge field over M. If the norm of the field strength is less than $\frac{R}{12}$, then this field is a self-dual gauge field.*

Bourguignon, Lawson [2] and Min-Oo [4] also discussed the related problem and obtained a better theorem [2]:

Theorem B. *Let M be a four-dimensional self-dual compact Riemannian manifold with positive scalar curvature R, and there is a sourceless $SU(N)$ gauge field over M. If the norm of the anti-self-dual part of the field strength is less than $\frac{R}{12}$, then this field is a self-dual gauge field.*

In this paper we generalize the above theorems to the non-compact case. We have

Theorem. *Let M be a four-dimensional complete self-dual Riemannian manifold, R the scalar curvature of M, f the strength of a Yang-Mills field over M with the gauge group $SU(N)$, and $|f^-|$ the norm of the anti-self-dual part of f. If*

(1) $|f^-| < \frac{R}{12}$,

* This work was done under the program Sonderforschungsbereich Theoretische Mathematik (SFB40) at University of Bonn

0025-5874/82/0180/0069/$01.80

(2) *there is a point* $x_0 \in M$ *such that*

$$\int_{B_k} |f^{-}|^2 \, d\,\mathrm{vol} = o(k^2),$$

where B_k *is a geodesic ball with the center* x_0 *and radius k, (for example, if the action of this field is finite), then this Yang-Mills field must be self-dual.*

Corollary. *Let M be a four-dimensional conformally flat, non-compact but complete Riemannian manifold with positive scalar curvature R, and with a sourceless* $SU(N)$ *gauge field over M. If the norm of the field strength of M is less than* $\frac{R}{12}$, *and the action of this field is finite, then the field is a vacuum one.*

The author is gratefully indebted to Professors Klingenberg and Ruh for their encouragements and concern. He would like to thank Dr. Min-Oo for helpful discussions. He is also grateful to the Mathematics Institute of Bonn University for its hospitality.

2. Notations

In this paper, let M be a four-dimensional complete Riemannian manifold. In a local coordinate system, the metric of M is defined by

$$ds^2 = g_{\mu\nu}\, dx^\mu\, dx^\nu, \qquad (\mu, \nu = 0, 1, 2, 3), \tag{1}$$

and the *Riemannian curvature tensor* and the *conformal curvature tensor* are $R_{\nu\mu\alpha\beta}$ and

$$\begin{aligned} C_{\mu\nu\alpha\beta} = R_{\mu\nu\alpha\beta} &- \tfrac{1}{2}(g_{\mu\alpha}R_{\nu\beta} - g_{\mu\beta}R_{\nu\alpha} + g_{\nu\beta}R_{\mu\alpha} - g_{\mu\alpha}R_{\nu\beta}) \\ &- \frac{R}{6}(g_{\mu\beta}g_{\nu\alpha} - g_{\mu\alpha}g_{\nu\beta}). \end{aligned} \tag{2}$$

respectively. $R_{\mu\alpha} = g^{\nu\beta}R_{\mu\nu\alpha\beta}$ and R are the *Ricci curvature tensor* and the *scalar curvature* of M respectively.

To any 2nd order covariant tensor $T_{\mu\nu}$, we can apply the $*$-operator

$$T^*_{\mu\nu} = \tfrac{1}{2}\eta_{\mu\nu}{}^{\alpha\beta}\, T_{\alpha\beta} \tag{3}$$

where

$$\eta_{\mu\nu\alpha\beta} = \sqrt{\det |g_{\lambda\delta}|}\ \varepsilon_{\mu\nu\alpha\beta}$$

and $\varepsilon_{\mu\nu\alpha\beta}$ is the complete skew-symmetric tensor with $\varepsilon_{0123} = 1$. From this, we can define the following expressions for the pair $(\mu\nu)$ of indices as

$$T^+_{\mu\nu} = \tfrac{1}{2}(T_{\mu\nu} + T^*_{\mu\nu}), \qquad T^-_{\mu\nu} = \tfrac{1}{2}(T_{\mu\nu} - T^*_{\mu\nu}). \tag{4}$$

We call $T^+_{\mu\nu}$ or $T^-_{\mu\nu}$ the *self-dual* or *anti-self-dual* part of $T_{\mu\nu}$ respectively. Similarly, we may define these expressions for the pairs of indices $(\mu\nu)$ or $(\alpha\beta)$ of the conformal curvature tensor $C_{\mu\nu\alpha\beta}$ and obtain

$$C_{\mu\nu\alpha\beta} = C^{+\,+}_{\mu\nu\alpha\beta} + C^{+\,-}_{\mu\nu\alpha\beta} + C^{-\,+}_{\mu\nu\alpha\beta} + C^{-\,-}_{\mu\nu\alpha\beta}.$$

It is evident that $C^{+-}=C^{-+}=0$ (to be brief, we shall suppress the lower indices).

If we have $C^{++}=0$ or $C^{--}=0$, then the Riemannian manifold M is said to be *self-dual* or *anti-self-dual* respectively [1].

Now let us study a gauge field the base manifold of which is a self-dual with $R>0$ whose gauge group is $SU(N)$.

Let X_a be a basis of the Lie algebra $SU(N)'$ and C^a_{bc} the structure constants. We can define the Cartan-Killing inner product in the $SU(N)'$ by

$$\langle A, B\rangle = -\operatorname{Tr}(AB) \tag{5}$$

where $A, B\in SU(N)'$, and let

$$G_{ab}=\langle X_a, X_b\rangle. \tag{6}$$

Then we define the *norm* $\|f\|$, $\|f^+\|$, $\|f^-\|$ of the field strength $f=\frac{1}{2}f_{\mu\nu}\,dx^\mu\wedge dx^\nu$ and its self-dual parts and anti-self-dual parts by

$$\begin{aligned}\|f\|^2&=g^{\mu\alpha}g^{\nu\beta}\langle f_{\mu\nu},f_{\alpha\beta}\rangle=g^{\mu\alpha}g^{\nu\beta}f^a_{\mu\nu}f^b_{\alpha\beta}\,G_{ab},\\ \|f^+\|^2&=g^{\mu\alpha}g^{\nu\beta}\langle f^+_{\mu\nu},f^+_{\alpha\beta}\rangle,\\ v=\|f^-\|^2&=g^{\mu\alpha}g^{\nu\beta}\langle f^-_{\mu\nu},f^-_{\alpha\beta}\rangle\end{aligned} \tag{7}$$

respectively. The field is said to be *self-dual* if $f^-=0$, and *anti-self-dual* if $f^+=0$.

3. Proof of the Theorem

The outline of the proof is that we first compute the Laplacian of the function $\sqrt{v}$. With a complicated calculation we can prove $\Delta(\sqrt{v})\geqq 0$. We know a theorem of Yau [6]:

If M is a non-compact Riemannian manifold, u is a non-negative subharmonic function on M, and $\int_M u^2\,d\,\mathrm{vol}<\infty$, then u must be a constant function.

We would like to use this theorem to prove $v=$const. and then prove $v=0$, i.e., the field is self-dual. However it is not a priori evident that $\sqrt{v}$ is of class C^1 or C^2 while Yau's theorem requires the smoothness of this function. Therefore we must use the standard trick of approximating $\sqrt{v}$ with $\sqrt{v+\varepsilon}$, for some $\varepsilon>0$.

3.1. Now we compute the Laplacian of $\sqrt{v+\varepsilon}$. By the Bianchi identities, sourceless condition and Ricci identities, we obtain

$$\Delta\sqrt{v+\varepsilon}=g^{\alpha\beta}(\sqrt{v+\varepsilon})_{\|\alpha\beta}=\frac{1}{2\sqrt{v+\varepsilon}}g^{\alpha\beta}v_{\|\alpha\beta}$$

$$-\frac{1}{4}\frac{1}{(v+\varepsilon)^{3/2}}g^{\alpha\beta}v_{\|\alpha}\cdot v_{\|\beta}$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta}(g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha} f^{-b}_{\lambda\delta} G_{ab})_{\|\beta} \\
&\quad - \frac{1}{4} \frac{1}{(v+\varepsilon)^{3/2}} g^{\alpha\beta} v_{\|\alpha} \cdot v_{\|\beta} \\
&= \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha\beta} f^{-b}_{\lambda\delta} G_{ab} \\
&\quad + \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha} f^{-b}_{\lambda\delta\|\beta} G_{ab} \\
&\quad - \frac{1}{4} \frac{1}{(v+\varepsilon)^{3/2}} g^{\alpha\beta} v_{\|\alpha} \cdot v_{\|\beta} \\
&= -\frac{2}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\nu\alpha\|\mu\beta} f^{-b}_{\lambda\delta} G_{ab} \\
&\quad + \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha} f^{-b}_{\lambda\delta\|\beta} G_{ab} \\
&\quad - \frac{1}{4} \frac{1}{(v+\varepsilon)^{3/2}} g^{\alpha\beta} v_{\|\alpha} \cdot v_{\|\beta} \\
&= -\frac{2}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} (f^{-a}_{\nu\alpha\|\beta\mu} \\
&\quad + f^{-a}_{\rho\alpha} R^{\rho}_{\nu\mu\beta} + f^{-a}_{\nu\rho} R^{\rho}_{\alpha\mu\beta} + f^{-c}_{\nu\alpha} C^{a}_{cd} f^{d}_{\mu\beta}) f^{-b}_{\lambda\delta} G_{ab} \\
&\quad + \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha} f^{-b}_{\lambda\delta\|\beta} G_{ab} \\
&\quad - \frac{1}{4} \frac{1}{(v+\varepsilon)^{3/2}} g^{\alpha\beta} v_{\|\alpha} \cdot v_{\|\beta} \\
&= -\frac{2}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} R^{\rho}_{\nu\mu\beta} f^{-a}_{\rho\alpha} f^{-b}_{\lambda\delta} G_{ab} \\
&\quad - \frac{2}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} R^{\rho}_{\alpha\mu\beta} f^{-a}_{\nu\rho} f^{-b}_{\lambda\delta} G_{ab} \\
&\quad - \frac{2}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} C^{a}_{cd} f^{-c}_{\nu\alpha} f^{d}_{\mu\beta} f^{-b}_{\lambda\delta} G_{ab} \\
&\quad + \frac{1}{\sqrt{v+\varepsilon}} g^{\alpha\beta} g^{\mu\lambda} g^{\nu\delta} f^{-a}_{\mu\nu\|\alpha} f^{-b}_{\lambda\delta\|\beta} G_{ab} \\
&\quad - \frac{1}{4} \frac{1}{(v+\varepsilon)^{3/2}} g^{\alpha\beta} v_{\|\alpha} \cdot v_{\|\beta}. \qquad (8)
\end{aligned}
$$

We denote the sum of the first and second terms in the right hand side by I, the third term by II, and the remaining terms by III, then

$$\Delta(\sqrt{v+\varepsilon}) = \mathrm{I} + \mathrm{II} + \mathrm{III}. \qquad (9)$$

First, from the definition of the conformal curvature tensor, we have

$$\begin{aligned}
\mathrm{I} = & -\frac{2}{\sqrt{v+\varepsilon}}(C^{\rho\delta\lambda\alpha}\langle f_{\rho\alpha}^{-}, f_{\lambda\delta}^{-}\rangle + \tfrac{1}{2}(g^{\rho\lambda}R^{\delta\alpha} \\
& - g^{\rho\alpha}R^{\delta\lambda} + g^{\delta\alpha}R^{\rho\lambda} - g^{\delta\lambda}R^{\rho\alpha})\langle f_{\rho\alpha}^{-}, f_{\lambda\delta}^{-}\rangle \\
& + \frac{R}{6}(g^{\rho\alpha}g^{\delta\lambda} - g^{\rho\lambda}g^{\delta\alpha})\langle f_{\rho\alpha}^{-}, f_{\lambda\delta}^{-}\rangle) \\
& - \frac{2}{\sqrt{v+\varepsilon}} g^{\nu\delta}R^{\rho\lambda}\langle f_{\nu\rho}^{-}, f_{\lambda\delta}^{-}\rangle \\
= & -\frac{2}{\sqrt{v+\varepsilon}} C^{\rho\delta\lambda\alpha}\langle f_{\rho\alpha}^{-}, f_{\lambda\delta}^{-}\rangle + \tfrac{1}{3}R\frac{v}{\sqrt{v+\varepsilon}}.
\end{aligned} \tag{10}$$

If we adapt the normal coordinate system about a point x of M such that $g_{\mu\nu}|_x = \delta_{\mu\nu}$, then, by the self-dual condition of M, we have

$$C_{\rho\delta\lambda\alpha} = C^{+\ +}_{\rho\delta\lambda\alpha}$$

and

$$\mathrm{I} = \frac{2}{\sqrt{v+\varepsilon}} C^{+\ +}_{\rho\delta\alpha\lambda}\langle f_{\rho\alpha}^{-}, f_{\lambda\delta}^{-}\rangle + \tfrac{1}{3}R\cdot\frac{v}{\sqrt{v+\varepsilon}}. \tag{11}$$

Through a straightforward but complicated computation, we know that the first term on the right hand vanishes. So

$$\mathrm{I} = \frac{R}{3}\cdot\frac{v}{\sqrt{v+\varepsilon}}. \tag{12}$$

Secondly, in the normal coordinate system,

$$\mathrm{II} = \frac{2}{\sqrt{v+\varepsilon}}\langle [f_{\delta\alpha}^{-}, f_{\alpha\mu}], f_{\mu\delta}^{-}\rangle. \tag{13}$$

Since the Cartan-Killing inner product is invariant under the adjoint representation, we have

$$\begin{aligned}
\mathrm{II} = & \frac{2}{\sqrt{v+\varepsilon}}\langle [f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}], f_{\alpha\mu}\rangle \\
= & \frac{2}{\sqrt{v+\varepsilon}}(\langle [f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}], f_{\alpha\mu}^{+}\rangle + \langle [f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}], f_{\alpha\mu}^{-}\rangle).
\end{aligned}$$

It is easy to see that $[f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}]$ are anti-self-dual, so $\langle [f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}], f_{\alpha\mu}^{+}\rangle = 0$. Thus

$$\mathrm{II} = -\frac{2}{\sqrt{v+\varepsilon}}\mathrm{Tr}([f_{\mu\delta}^{-}, f_{\delta\alpha}^{-}]f_{\alpha\mu}^{-}) = -\frac{4}{\sqrt{v+\varepsilon}}\mathrm{Tr}(f_{\mu\delta}^{-}f_{\delta\alpha}^{-}f_{\alpha\mu}^{-}). \tag{14}$$

Let $(f_{\mu\delta}^{-ij})$, $(f_{\delta\alpha}^{-jk})$ and $(f_{\alpha\mu}^{-ki})$ be the elements of the matrices $f_{\mu\delta}^{-}$, $f_{\delta\alpha}^{-}$ and $f_{\alpha\mu}^{-}$ respectively. Then

$$\mathrm{II}=-\frac{4}{\sqrt{v+\varepsilon}} f_{\mu\delta}^{-ij} f_{\delta\alpha}^{-jk} f_{\alpha\mu}^{-ki} \tag{15}$$

If we denote the double indices $\binom{i}{\mu}$, $\binom{j}{\delta}$ and $\binom{k}{\alpha}$ by A, B and C respectively, then

$$\mathrm{II}=-\frac{4}{\sqrt{v+\varepsilon}} f_{AB}^{-} f_{BC}^{-} f_{CA}^{-}. \tag{16}$$

Consider the matrices

$$T=(f_{CA}^{-}), \qquad S=(f_{AB}^{-} f_{BC}^{-}).$$

It is clear that S, T are both Hermitian matrices and further S is semi-positive definite. Therefore, it is easy to prove that

$$\mathrm{Tr}(ST)\leqq\sqrt{\mathrm{Tr}(T^2)}\cdot\mathrm{Tr}(S).$$

Hence

$$\begin{aligned}
\mathrm{II}&\geqq -\frac{4}{\sqrt{v+\varepsilon}}\cdot\sqrt{f_{AB}^{-}\cdot f_{BA}^{-}}\cdot f_{CD}^{-} f_{DC}^{-}\\
&=-\frac{4}{\sqrt{v+\varepsilon}}\cdot\sqrt{f_{\mu\delta}^{-ij}\cdot f_{\delta\mu}^{-ji}}\cdot f_{\alpha\beta}^{-kl} f_{\beta\alpha}^{-lk}\\
&=-\frac{4}{\sqrt{v+\varepsilon}}\cdot\sqrt{-\mathrm{Tr}(f_{\mu\delta}^{-}\cdot f_{\mu\delta}^{-})}\cdot(-\mathrm{Tr}(f_{\alpha\beta}^{-} f_{\alpha\beta}^{-}))\\
&=-\frac{4}{\sqrt{v+\varepsilon}}\cdot v^{3/2}.
\end{aligned} \tag{17}$$

Finally, in the normal coordinate system,

$$\begin{aligned}
\mathrm{III}&=\frac{1}{\sqrt{v+\varepsilon}}\left(\langle f_{\mu\nu\|\alpha}^{-}, f_{\mu\nu\|\alpha}^{-}\rangle-\frac{1}{4(v+\varepsilon)} v_{\|\mu}\cdot v_{\|\mu}\right)\\
&=\frac{1}{\sqrt{v+\varepsilon}}\left(\langle f_{\mu\nu\|\alpha}^{-}, f_{\mu\nu\|\alpha}^{-}\rangle-\frac{1}{4(v+\varepsilon)}\sum_{\mu}(2\langle f_{\lambda\delta\|\mu}^{-}, f_{\lambda\delta}^{-}\rangle)^2\right).
\end{aligned} \tag{18}$$

Observe that

$$\begin{aligned}
\langle f_{\lambda\delta\|\mu}^{-}, f_{\lambda\delta}^{-}\rangle^2&=(-\mathrm{Tr}(f_{\lambda\delta\|\mu}^{-}\cdot f_{\lambda\delta}^{-}))^2\\
&=f_{\lambda\delta}^{-ij}\cdot f_{\lambda\delta}^{-ji}\cdot f_{\alpha\beta\|\mu}^{-kl}\cdot f_{\alpha\beta}^{-lk}\\
&=(f_{\lambda\delta\|\mu}^{-ij}\cdot\overline{f_{\alpha\beta\|\mu}^{-lk}})\cdot(f_{\alpha\beta}^{-lk}\cdot\overline{f_{\lambda\delta}^{-ij}}).
\end{aligned} \tag{19}$$

Introduce the notations $A=\begin{pmatrix} i & j\\ \lambda & \delta\end{pmatrix}$, $B=\begin{pmatrix} l & k\\ \alpha & \beta\end{pmatrix}$. Then

$$\langle f_{\lambda\delta\|\mu}^{-}, f_{\lambda\delta}^{-}\rangle^2=(f_{A\|\mu}^{-}\cdot\overline{f_{B\|\mu}^{-}})\cdot(f_{B}^{-}\cdot\overline{f_{A}^{-}}). \tag{20}$$

Let

$$P=(P_{AB}), \quad Q=(Q_{AB}),$$

where

$$P_{AB}=f^{-}_{A\|\mu}\cdot\overline{f^{-}_{B\|\mu}}$$
$$Q_{AB}=f^{-}_{A}\cdot\overline{f^{-}_{B}}.$$

It is clear that both P and Q are semi-positive definite Hermitian matrices. Hence it is easy to prove that

$$\mathrm{Tr}(PQ)\leqq\mathrm{Tr}\,P\cdot\mathrm{Tr}\,Q.$$

It follows from (19) that

$$\begin{aligned}\langle f^{-}_{\lambda\delta\|\mu},f^{-}_{\lambda\delta}\rangle^2=P_{AB}\cdot Q_{BA}&=\mathrm{Tr}(PQ)\\&\leqq\mathrm{Tr}\,P\cdot\mathrm{Tr}\,Q=(f^{-}_{A\|\mu}\overline{f^{-}_{A\|\mu}})\cdot(f^{-}_{B}\overline{f^{-}_{B}})\\&=(f^{-ij}_{\lambda\delta\|\mu}\cdot\overline{f^{-ij}_{\lambda\delta\|\mu}})\cdot(f^{-lk}_{\alpha\beta}\cdot\overline{f^{-lk}_{\alpha\beta}})\\&=\langle f^{-}_{\lambda\delta\|\mu},f^{-}_{\lambda\delta\|\mu}\rangle\cdot\langle f^{-}_{\alpha\beta},f^{-}_{\alpha\beta}\rangle.\end{aligned}$$

Hence from (18), we obtain

$$\mathrm{III}\geqq\frac{1}{\sqrt{v+\varepsilon}}\cdot\langle f^{-}_{\mu\nu\|\alpha},f^{-}_{\mu\nu\|\alpha}\rangle\cdot\left(1-\frac{v}{v+\varepsilon}\right)\geqq 0$$

$$\Delta(\sqrt{v+\varepsilon})\geqq 4\left(\frac{R}{12}-\sqrt{v}\right)\frac{v}{\sqrt{v+\varepsilon}}. \tag{21}$$

From the condition of the theorem, we have $\Delta(\sqrt{v+\varepsilon})\geqq 0$.

3.2. Now we want to prove $v=0$. Even though we have proved $\sqrt{v+\varepsilon}$ is subharmonic, we still cannot apply the Yau's theorem to $\sqrt{v+\varepsilon}$ directly. We do not know whether $\int_M (v+\varepsilon)\,d\,\mathrm{vol}<\infty$, because $\mathrm{vol}(M)$ may be infinite. Now we follow the proof of Yau's theorem [6] with some modifications as Hildebrandt did in [3].

As Yau [6] pointed out, for every $R>0$, we can find a Lipschitz function η on M such that $\eta(x)=1$ for $x\in B_R$, $\eta(x)=0$ on $M\backslash B_{2R}$, $0\leqq\eta\leqq 1$, and $|\nabla\eta|\leqq\frac{C}{R}$, where C is a constant which is independent of R, and B_R denotes a geodesic ball with the fixed center x_0 and radius R. Thus

$$\begin{aligned}0&\leqq\int_{B_{2R}}(\eta^2\sqrt{v+\varepsilon})\,\Delta(\sqrt{v+\varepsilon})\,d\,\mathrm{vol}\\&=-\int_{B_{2R}}g^{\mu\nu}\frac{\partial\sqrt{v+\varepsilon}}{\partial x^{\mu}}\frac{\partial(\eta^2\sqrt{v+\varepsilon})}{\partial x^{\nu}}\,d\,\mathrm{vol}\\&=-\int_{B_{2R}}g^{\mu\nu}\frac{\partial\sqrt{v+\varepsilon}}{\partial x^{\mu}}\frac{\partial\sqrt{v+\varepsilon}}{\partial x^{\nu}}\cdot\eta^2\,d\,\mathrm{vol}\\&\quad-\int_{B_{2R}}2g^{\mu\nu}\cdot\eta\cdot\frac{\partial\sqrt{v+\varepsilon}}{\partial x^{\mu}}\cdot\frac{\partial\eta}{\partial x^{\nu}}\cdot\sqrt{v+\varepsilon}\,d\,\mathrm{vol}.\end{aligned}$$

Therefore

$$\int_{B_R} |\nabla\sqrt{v+\varepsilon}|^2 \eta^2 \, d\,\mathrm{vol} + \int_{B_{2R}-B_R} |\nabla\sqrt{v+\varepsilon}|^2 \eta^2 \, d\,\mathrm{vol}$$
$$= \int_{B_{2R}} |\nabla\sqrt{v+\varepsilon}|^2 \eta^2 \, d\,\mathrm{vol}$$
$$\leqq 2\sqrt{\int_{B_{2R}-B_R} \eta^2 \cdot |\nabla\sqrt{v+\varepsilon}|^2 \, d\,\mathrm{vol}} \cdot \sqrt{\int_{B_{2R}-B_R} (v+\varepsilon)\,|\nabla\eta|^2 \, d\,\mathrm{vol}}$$

Let

$$\mathsf{X} = \sqrt{\int_{B_{2R}-B_R} \eta^2 \, |\nabla\sqrt{v+\varepsilon}|^2 \, d\,\mathrm{vol}},$$

then the above inequality becomes

$$\mathsf{X}^2 - 2\sqrt{\int_{B_{2R}-B_R} (v+\varepsilon)\,|\nabla\eta|^2 \, d\,\mathrm{vol}} \cdot \mathsf{X} + \int_{B_R} |\nabla\sqrt{v+\varepsilon}|^2 \, d\,\mathrm{vol} \leqq 0.$$

Since X is real, the discriminant of this quadratic must be non-negative. Thus

$$\int_{B_R} |\nabla\sqrt{v+\varepsilon}|^2 \, d\,\mathrm{vol} \leqq \int_{B_{2R}-B_R} (v+\varepsilon) \cdot |\nabla\eta|^2 \, d\,\mathrm{vol} \leqq \frac{C^2}{R^2} \int_{B_{2R}-B_R} (v+\varepsilon) \, d\,\mathrm{vol}.$$

Set $B'_R = B_R \setminus \{x \mid v=0\}$, and let $\varepsilon \to 0$. We have

$$\int_{B'_R} \frac{|\nabla v|^2}{4v} \, d\,\mathrm{vol} \leqq \frac{C^2}{R^2} \int_{B_{2R}-B_R} v \, d\,\mathrm{vol} \leqq \frac{C^2}{R^2} \int_{B_{2R}} v \, d\,\mathrm{vol}.$$

Let $R \to \infty$, using the assumption (2) in the theorem, we obtain

$$\int_{M \setminus \{x \mid v=0\}} \frac{|\nabla v|^2}{4v} \, d\,\mathrm{vol} = 0.$$

Thus $|\nabla v| = 0$ on $M \setminus \{x \mid v=0\}$, and therefore also on M. Hence $v = \text{const.}$ on M. Then, substituting this into the inequality (21), we have $v=0$, i.e.,

$$f^-_{\mu\nu} = 0. \quad \text{Q.E.D.}$$

Remark. If we replace the self-dual condition of M in the Theorem by the anti-self-dual condition and f^- by f^+, then this field would be an anti-self-dual field.

Finally, we point out that the corollary is a direct consequence of the theorem and the remark above.

References

1. Atiyah, M.F., Hitchin, N.J., Singer, I.M.: Self duality in four dimensional Riemannian Geometry. Proc. Roy. Soc. London Ser. A **362**, 425–461 (1978)
2. Bourguignon, J.P., Lawson, H.B.: Stability and isolation phenomena for Yang-Mills fields. Comm. Math. Phys. **79**, 189–230 (1981)

3. Hildebrandt, S.: Liouville theorems for harmonic mappings and an approach to Bernstein theorem. Annals of Mathematics Studies. Princeton: Princeton University Press (to appear)
4. Min-Oo: An $\mathscr{L}_2$-isolation theorem for Yang-Mills fields. Preprint
5. Shen, C.L.: On the sourceless $SU(N)$ gauge field over the four-dimensional self-dual compact Riemannian manifold with positive scalar curvature. In: Proceedings of the Beijing Symposium in Differential Geometry and Partial Differential Equations (Beijing 1980) (to appear)
6. Yau, S.T.: Some function-theoretic properties of complete Riemannian manifolds and their applications to geometry. India Univ. Math. J. **25**, 659–670 (1976)

Received July 17, 1981

Math. Z. 180, 79–80 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Zu einer Aussage über die kettengeometrischen Erzeugenden

Michael Werner

Mathematisches Institut B der Universität, Pfaffenwaldring 57, D-7000 Stuttgart 80, Bundesrepublik Deutschland

In [3, (7.7)] heißt es, daß sich die nach [3, S. 12] durch

$$E=\{\mathfrak{L}^\times(a_{11}k_1+a_{12}k_2, a_{21}k_1+a_{22}k_2)\in\mathbf{P}(\mathfrak{L}):\mathfrak{K}^\times(k_1,k_2)\in\mathbf{P}(\mathfrak{K})\}$$
$$a_{ik}\in\mathfrak{L},\quad |a_{ik}|\in\mathfrak{L}\setminus\mathfrak{L}^\times,\quad |E|\geqq 2$$

definierten Erzeugenden E einer Kettengeometrie $\Sigma(\mathfrak{K},\mathfrak{L})$ gemäß [1] im Fall einer kinematischen [4] Algebra $\mathfrak{L}$ endlicher Dimension im Hotje-Modell [2; 3]

$$Q\colon \{\mathfrak{K}^\times(y,y_{n+1},y_{n+2})\in(\mathfrak{L}\times\mathfrak{K}\times\mathfrak{K}\setminus\{0,0,0\})/\mathfrak{K}^\times : N(y)=y_{n+1}y_{n+2}\},$$
$$h\colon\begin{cases}\mathbf{P}(\mathfrak{L}) & \leftrightarrow Q\setminus R\\ \mathfrak{L}^\times(u,v)&\mapsto\mathfrak{K}^\times(u\bar{v},N(u),N(v)),\end{cases}$$
$$R=\{\mathfrak{K}^\times(z,0,0): z\in\mathfrak{L}\setminus(\mathfrak{L}^\times\cup\{0\})\wedge z\bar{z}'+\bar{z}z'=0 \text{ für alle } z'\in\mathfrak{L}\}$$

als Spuren der Erzeugenden der Quadrik Q in $Q\setminus R$ realisieren.

Die Minkowski-Geometrien zeigen, daß diese Aussage nicht uneingeschränkt gilt.

Der Minkowski-Geometrie über einem Körper $\mathfrak{K}$ liegt bis auf Isomorphie die Algebra $\mathfrak{L}=\mathfrak{K}\times\mathfrak{K}$ der Doppelzahlen über $\mathfrak{K}$ mit komponentenweiser Addition und Multiplikation zugrunde, die für $|\mathfrak{K}|>2$ kinematisch ist.

Da der Automorphismus $^-$ die Doppelzahl $y=(y_1,y_2)$ in (y_2,y_1) überführt, hat die Modellquadrik Q hier die Gleichung $y_1y_2=y_3y_4$. Das Radikal R von Q ist leer.

Die Menge

$$E=\{\mathfrak{L}^\times(ka,1): k\in\mathfrak{K}\}$$

mit der Doppelzahl $a=(0,1)$ ist eine kettengeometrische Erzeugende; man beachte, daß das Paar $(a,0)$ wegen $\langle a,0\rangle\neq\mathfrak{L}$ auf keinen Punkt der Kettengeometrie führt.

Die Modellmenge $h(E)$ von E liegt auf der Erzeugenden e

$$e\colon y_1=0, y_3=0$$

von Q.

0025-5874/82/0180/0079/$01.00

Es gilt aber

$$e \setminus h(E) = \{\mathfrak{R}^{\times}(0,1,0,0)\} = \{h(\mathfrak{L}^{\times}((0,1),(1,0)))\} \neq \varnothing.$$

Die sich hierdurch ergebende Situation wird in [5] weiter verfolgt.

Literatur

1. Benz, W.: Vorlesungen über Geometrie der Algebren. Berlin-Heidelberg-New York: Springer 1973
2. Hotje, H.: Einbettung gewisser Kettengeometrien in projektive Räume. J. Geometry **5**, 85–94 (1974)
3. Hotje, H.: Zur Einbettung von Kettengeometrien in projektive Räume. Math. Z. **151**, 5–17 (1976)
4. Karzel, H.: Kinematische Algebren und ihre geometrischen Ableitungen. Abh. Math. Sem. Univ. Hamb. **41**, 158–171 (1974)
5. Werner, M.: Zur Problematik des Einbettungsbegriffs bei Kettengeometrien. Geometriae Dedicata (erscheint demnächst)

Eingegangen am 5. Oktober 1981

Math. Z. 180, 81–89 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Quadratic Forms and Artin's Reciprocity Law

Manfred Kolster

Mathematisches Institut der Universität, Roxeler Straße 64, D-4400 Münster, Federal Republic of Germany

Introduction

A. Weil extended Hilbert's reciprocity law to a reciprocity law for arbitrary quadratic forms (cf. [12], Proposition 5, p. 179). Using standard methods from the theory of quadratic forms Knebusch and Scharlau [4] gave a quite elementary proof of Weil's reciprocity law. The aim of this paper is to show, that a similar treatment applies to Artin's reciprocity law for multiquadratic extensions of an algebraic number field F. In the special situation considered, Artin's reciprocity law may be expressed via Kummer theory as the non-degeneracy of a certain pairing defined by the 2-power residue symbol. We show, that this pairing has natural generalizations to reciprocity pairings with values in the group of 8^{th} roots of unity on the level of quadratic forms (Theorem 7), and that the non-degeneracy of these pairings can easily be proved using only the classification of forms over local and global fields.

The general reciprocity pairings are defined by means of certain characters on the Witt rings of the completions of F at finite and real primes. The study of these characters – taken up in the first section – also leads to a new simple proof of Weil's reciprocity law. In the second section we prove our main result (Theorem 7) and show, how it is related to Artin's reciprocity law.

1. S-reciprocity Laws

For any commutative ring A with 1 let $W(A)$ denote the Witt ring of symmetric bilinear forms over A, $I(A)$ the fundamental ideal of forms of even rank, $W'(A) := W(A)/I^3(A)$ and $WQ(A)$ the Witt group of quadratic forms over A.

Let F be an algebraic number field. The Clifford invariant (cf. Lam [5], Chap. 5, § 3), which is a homomorphism from the Witt ring $W(F)$ to the graded Brauer group $GB(F)$ of F, factors through $W'(F)$ and gives an injective homomorphism from $W'(F)$ to $GB(F)$ (cf. Lam [5], Chap. 6, Corollary 3.10). Thus the elements from $W'(F)$ are classified by the rank, the discriminant and the

0025-5874/82/0180/0081/$01.80

Witt-invariant. For any prime $\mathfrak{p}$ of F we denote by $F_\mathfrak{p}$ the completion of F at $\mathfrak{p}$. If $\mathfrak{p}$ is a finite prime, the third power $I^3(F_\mathfrak{p})$ of the fundamental ideal $I(F_\mathfrak{p})$ vanishes (cf. Lam [5], Chap. 6, Corollary 2.15), hence $W'(F_\mathfrak{p})=W(F_\mathfrak{p})$. If $\mathfrak{p}$ is a real infinite prime we get $W'(F_\mathfrak{p})=\mathbb{Z}/8\mathbb{Z}$. Thus $W'(F)$ is obtained from $W(F)$ by "reducing the signatures mod 8". Since the characters and pairings we are going to study take values in the group μ_8 of 8^{th} roots of unity, they all vanish on $I^3(F)$, hence we are lead to consider $W'(F)$ instead of $W(F)$ (cf. also Knebusch-Scharlau [4], § 6).

Let R denote the ring of integers of F and $R_\mathfrak{p}$ the completion of R at a finite prime $\mathfrak{p}$. As for any Dedekind ring we have injective maps $WQ(R_\mathfrak{p})\to W(F_\mathfrak{p})$ and $W(R_\mathfrak{p})\to W(F_\mathfrak{p})$ (cf. Knebusch-Scharlau [4], Lemma 1.4.). If $\mathfrak{p}$ is a nondyadic prime, the groups $W(R_\mathfrak{p})$ and $WQ(R_\mathfrak{p})$ are equal and their elements are classified by the rank and the discriminant, which has to be a unit in $R_\mathfrak{p}$. If $\mathfrak{p}$ is dyadic, the situation is different: The elements of $W(R_\mathfrak{p})$ are classified by the rank, the discriminant, which again has to be a unit in $R_\mathfrak{p}$, and the Witt-invariant. Thus the cokernel of the map $W(R_\mathfrak{p})\to W(F_\mathfrak{p})$ has order 2. The Witt group $WQ(R_\mathfrak{p})$ at a dyadic prime $\mathfrak{p}$ is isomorphic with $\mathbb{Z}/2\mathbb{Z}$, where the non-trivial element is represented by a form of even rank and discriminant $\Delta_\mathfrak{p}$, such that $F_\mathfrak{p}(\sqrt{\Delta_\mathfrak{p}})/F_\mathfrak{p}$ is the unique unramified extension of degree 2 over $F_\mathfrak{p}$. For a more detailed discussion of the local situation the reader is referred to Lannes [7], Sect. 2.

If η is an element of $W'(F)$ we denote by $\eta_\mathfrak{p}$ the image of η in $W'(F_\mathfrak{p})$. If $\mathfrak{p}$ is a nondyadic finite prime, the element $\eta_\mathfrak{p}$ is contained in $W(R_\mathfrak{p})$ if and only if the Witt-invariant of η is trivial at $\mathfrak{p}$ and the $\mathfrak{p}$-adic valuation $v_\mathfrak{p}(d(\eta))$ of the discriminant $d(\eta)$ of η is even. The latter condition is equivalent with the fact, that $F_\mathfrak{p}(\sqrt{d(\eta)})/F_\mathfrak{p}$ is unramified. If $\mathfrak{p}$ is a dyadic prime, the element $\eta_\mathfrak{p}$ is contained in $W(R_\mathfrak{p})$ if and only if $v_\mathfrak{p}(d(\eta))$ is even, and $\eta_\mathfrak{p}$ is contained in $WQ(R_\mathfrak{p})$ if and only if η has even rank, trivial Witt-invariant at $\mathfrak{p}$ and $F_\mathfrak{p}(\sqrt{d(\eta)})/F_\mathfrak{p}$ is unramified. These elementary facts will be frequently used in the following considerations.

Let W denote the set of real primes of F and let S_f be a finite set of finite primes of F. Let $S=W\cup S_f$. Motivated by a similar definition in [4] we define an *S-reciprocity law* γ to be a family of homomorphisms $\gamma_\mathfrak{p}\colon W'(F_\mathfrak{p})\to\mu_8$, defined at all finite and real primes $\mathfrak{p}$, satisfying the following conditions:

i) $\gamma_\mathfrak{p}(WQ(R_\mathfrak{p}))=1$ for all $\mathfrak{p}\notin S$

ii) $\prod_\mathfrak{p}\gamma_\mathfrak{p}(\eta_\mathfrak{p})=1$ for all $\eta\in W'(F)$.

γ is called a *generating S-reciprocity law*, if in addition

iii) $\gamma_\mathfrak{p}(\varphi_\mathfrak{p})=-1$ for all $\mathfrak{p}$,

where $\varphi_\mathfrak{p}$ denotes the unique anisotropic 4-dimensional form over $F_\mathfrak{p}$ (cf. Lam [5], Chap. 6, Theorem 2.2.).

Let $K(S)$ denote the kernel of the homomorphism

$$\partial_S\colon W'(F)\to\coprod_{\mathfrak{p}\notin S} W(F_\mathfrak{p})/W(R_\mathfrak{p})$$

and let $C(S)$ denote the cokernel of the homomorphism

$$\partial^S: W'(F) \to \coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W'(F_\mathfrak{p}),$$

where the summation on the right-hand side only runs over finite and real primes. The complex primes of F don't occur. The S-reciprocity laws are just the characters of $C(S)$. The subgroup $K(S)$ of $W'(F)$ is actually a subring, which acts on $W'(F)$ and on $\coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W'(F_\mathfrak{p})$. Note, that $WQ(R_\mathfrak{p})$ is a $W(R_\mathfrak{p})$ - module (cf. Milnor-Husemoller [8], Appendix 1). Since the map ∂^S is $K(S)$ - linear, the group $C(S)$ is a $K(S)$ - module. Thus any S-reciprocity law γ defines a pairing

$$(*) \qquad K(S) \times C(S) \to \mu_8$$

via $(k, c) \to (k \cdot c)$ for $k \in K(S)$ and $c \in C(S)$. In this section we reveal the group structure of $C(S)$ and show, that generating S-reciprocity laws exist. In the next section we study the pairing $(*)$ defined by a generating S-reciprocity law.

As a first step we look at the structure of $K(S)$. By definition an element $\eta \in W'(F)$ lies in $K(S)$ if and only if $\eta_\mathfrak{p}$ is contained in $W(R_\mathfrak{p})$ for all $\mathfrak{p} \notin S$. This is equivalent with the conditions, that the Witt-invariant of η is trivial at all nondyadic primes $\mathfrak{p} \notin S$ and that the $\mathfrak{p}$-adic valuation $v_\mathfrak{p}(d(\eta))$ is even at all $\mathfrak{p} \notin S$. Thus, if we denote by $\mathrm{Br}_2(S)$ the subgroup of the Brauer group of F of all quaternion algebras, which split at all nondyadic primes outside of S, and if we define $D(S) := \{\alpha \in F^* \mid v_\mathfrak{p}(\alpha) \equiv 0 \bmod 2 \text{ at all } \mathfrak{p} \notin S\}/F^{*2}$, then as a set $K(S)$ is equal to the product $\mathbb{Z}/2\mathbb{Z} \times D(S) \times \mathrm{Br}_2(S)$. The description of the group structure of $K(S)$ is obtained from that of the graded Brauer group (cf. Lam [5], Chap. 5; Scharlau [10], Sect. 3.5): First one defines a group $K_0(S)$ as a group extension

$$0 \to D(S) \to K_0(S) \to \mathbb{Z}/2\mathbb{Z} \to 0$$

given by the cocycle $f: \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z} \to D(S)$,

$$f(e, e') = (-1)^{ee'} \bmod F^{*2}$$

and then $K(S)$ is given by the group extension

$$0 \to \mathrm{Br}_2(S) \to K(S) \to K_0(S) \to 0$$

with cocycle $g: K_0(S) \times K_0(S) \to \mathrm{Br}_2(S)$ defined by

$$g((e, d), (e', d')) = \begin{cases} \left(\dfrac{d, d'}{F}\right) & \text{if } e = e' \\ \left(\dfrac{(-1)^e d, (-1)^{e'} d'}{F}\right) & \text{if } e \neq e' \end{cases}.$$

Hence we have

Proposition 1. *The group $K(S)$ is described by a two-fold group extension*

$$0 \to D(S) \to K_0(S) \to \mathbb{Z}/2\mathbb{Z} \to 0$$
$$0 \to \mathrm{Br}_2(S) \to K(S) \to K_0(S) \to 0.$$

To compute the order of the finite group $K(S)$ we have to know the size of $D(S)$. We fix the following notations:

$$C = \text{class group of } F$$
$$D = \text{set of dyadic primes not contained in } S$$
$$r = |W| = \#\text{ of real primes of } F$$
$$c = \#\text{ of complex primes of } F$$
$$S' = S \cup D.$$

Lemma 2. *Let $S_0 \subset S_f$ be maximal with the property, that the primes in S_0 are linearly independant in C/C^2. Then the group $D(S)$ has order $2^{|S|+c-|S_0|} \cdot |C/C^2|$.*

Proof. The group $D(W)$ has order $2^{r+c}|C/C^2|$ (cf. Knebusch-Scharlau [4], Lemma 6.3). Thus it suffices to show the following:

a) $D(S_0 \cup W) = D(W)$,
b) $|D(S)| = 2^{|S_f|-|S_0|} \cdot |D(S_0 \cup W)|$.

Part a) is an immediate consequence of the definition of S_0. To prove b) let $\mathfrak{p} \in S_f \setminus S_0$. Thus there are $\mathfrak{p}_1, \ldots, \mathfrak{p}_r \in S_0$, such that $\mathfrak{p} \cdot \prod_{i=1}^{r} \mathfrak{p}_i = (\alpha) \cdot \mathfrak{b}^2$. For this α the p-adic valuation is odd, but $v_{\mathfrak{q}}(\alpha)$ is even at all $\mathfrak{q} \neq \mathfrak{p}$, $\mathfrak{q} \in S_f \setminus S_0$. Clearly this implies b).

Corollary 3. *$K(S)$ has order $2^{2|S|+|D|+c-|S_0|} \cdot |C/C^2|$.*

Proof. Since $\mathrm{Br}_2(S)$ has order $2^{|S|+|D|-1}$, the result follows from Proposition 1 and Lemma 2.

In a next step we look at the structure of $C(S)$. Let $C_0(S)$ denote the cokernel of the map

$$\partial_0^S : W_0'(F) \to \coprod_{\mathfrak{p} \notin S'} W(F_\mathfrak{p})/W(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in D} W_0(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W_0'(F_\mathfrak{p}),$$

where the subscript "o" means restriction to forms of even rank. We get an extension

$$0 \to C_0(S) \to C(S) \to Rk(S) \to 0,$$

where $Rk(S)$ is an elementary abelian 2-group of order $2^{|S|+|D|-1}$. To study $C_0(S)$ we look at the map

$$\varphi : W_0'(F) \to \coprod_{\mathfrak{p} \in D} W_0(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W_0'(F_\mathfrak{p})$$

φ is surjective, since the discriminants at a finite number of primes can be approximated suitably. Now we compare the two sequences

$$\begin{array}{ccccccc}
W_0'(F) & \xrightarrow{\partial_0^S} & \coprod_{\mathfrak{p} \notin S'} W(F_\mathfrak{p})/W(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in D} W_0(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W_0'(F_\mathfrak{p}) & \longrightarrow & C_0(S) \\
\uparrow & & \uparrow & & \nearrow \\
\ker \varphi & \xrightarrow{\psi} & \coprod_{\mathfrak{p} \notin S'} W(F_\mathfrak{p})/W(R_\mathfrak{p}) & \longrightarrow & \operatorname{coker} \psi.
\end{array}$$

Here the map ψ denotes the restriction of ∂_0^S to $\ker\varphi$. Since the maps ∂_0^S and ψ have the same kernel and since φ is surjective, the snake lemma implies, that $\operatorname{coker}\psi \xrightarrow{\sim} C_0(S)$. Now let $I^{S'}$ denote the ideal group of all ideals relatively prime to S'. We have a surjective homomorphism

$$\delta\colon \coprod_{\mathfrak{p}\notin S'} W(F_\mathfrak{p})/W(R_\mathfrak{p}) \to I^{S'}/(I^{S'})^2$$

defined by $\delta(\{\eta_\mathfrak{p}\}) = \text{class of } \prod_{\mathfrak{p}\notin S'} \mathfrak{p}^{v_\mathfrak{p}(d(\eta_\mathfrak{p}))}$.

As is easily seen the discriminant maps $\ker\varphi$ onto $K_{1,S} := \{\alpha\in F^* \mid \alpha\in F_\mathfrak{p}^{*2}$ at $\mathfrak{p}\in S$, $F_\mathfrak{p}(\sqrt{\alpha})/F_\mathfrak{p}$ unramified at $\mathfrak{p}\in D\}/F^{*2}$. Thus, if we let $i(K_{1,S})$ denote the image of $K_{1,S}$ in $I^{S'}/(I^{S'})^2$ and $R(S) := I^{S'}/(I^{S'})^2 \cdot i(K_{1,S})$, we get a commutative diagram

$$\begin{array}{ccccc} \ker\varphi & \longrightarrow & \coprod_{\mathfrak{p}\notin S'} W(F_\mathfrak{p})/W(R_\mathfrak{p}) & \to C_0(S) & \longrightarrow 0 \\ \downarrow & & \downarrow{\scriptstyle\delta} & \downarrow & \\ K_{\cdot,S} & \longrightarrow & I^{S'}/(I^{S'})^2 & \longrightarrow R(S) & \longrightarrow 0. \end{array}$$

Now, elements from $\ker\delta$ are classified by the Witt invariant, hence due to Hilbert's reciprocity law, the kernel of $C_0(S)\to R(S)$ is $\{\pm 1\} = \mu_2$.

Proposition 4. *$C(S)$ is described by a two-fold group extension*

$$0\to\mu_2\to C_0(S)\to R(S)\to 0,$$
$$0\to C_0(S)\to C(S)\to Rk(S)\to 0.$$

Remark. We shall see below (Corollary 9), that these extensions are dual to those defining $K(S)$. For this reason we did not write down the cocycles belonging to the extensions. These can be described, but not in an easy way.

The proof of Proposition 4 showed, that the generating S-reciprocity laws correspond with those characters of $C(S)$ that do not vanish on the subgroup μ_2. Since any character of μ_2 is extendible to a character of $C(S)$ we get the following.

Corollary 5. *There exist generating S-reciprocity laws.*

Corollary 6. (Weil's reciprocity law (cf. [12], Proposition 5, p. 179)). *There exist homomorphisms $\gamma_\mathfrak{p}\colon W(F_\mathfrak{p})\to\mu_8$, defined for all finite and real primes $\mathfrak{p}$, such that*

i) *for each $\eta\in W(F)$: $\gamma_\mathfrak{p}(\eta_\mathfrak{p}) = 1$ for almost all $\mathfrak{p}$ and $\prod_\mathfrak{p}\gamma_\mathfrak{p}(\eta_\mathfrak{p}) = 1$*

ii) $\gamma_\mathfrak{p}(\varphi_\mathfrak{p}) = -1$ for all $\mathfrak{p}$.

2. Reciprocity Pairings

The groups $D(S)$ and $R(S)$, which occured in the description of $K(S)$ and $C(S)$, are related to each other in the following way: The 2-power residue symbol

$\left(\frac{\alpha}{\mathfrak{p}}\right)$ is defined by

$$\left(\frac{\alpha}{\mathfrak{p}}\right)=\begin{cases} 1 & \text{if } \alpha\in F_{\mathfrak{p}}^{*2} \\ -1 & \text{otherwise.}\end{cases}$$

The definition of $D(S)$ implies, that for $\alpha,\beta\in D(S)$, $\mathfrak{p}\notin S'$ the equation

$$\left(\frac{\alpha}{\mathfrak{p}}\right)\left(\frac{\beta}{\mathfrak{p}}\right)=\left(\frac{\alpha\beta}{\mathfrak{p}}\right)$$

holds. Thus, if we extend the symbol multiplicatively to the idealgroup $I^{S'}$, we get a bimultiplicative pairing

$$D(S)\times I^{S'}\to\mu_2.$$

Let $x\in K_{1,S}$ and $\alpha\in D(S)$. We have the relation $\left(\frac{\alpha}{\mathfrak{p}}\right)^{v_{\mathfrak{p}}(x)}=(\alpha,x)_{\mathfrak{p}}$, where $(,)_{\mathfrak{p}}$ denotes the Hilbert-symbol over $F_{\mathfrak{p}}$. Thus $\left(\frac{\alpha}{(x)}\right)=\prod_{\mathfrak{p}\notin S'}(\alpha,x)_{\mathfrak{p}}=\prod_{\mathfrak{p}\in S'}(\alpha,x)_{\mathfrak{p}}$ by Hilbert's reciprocity law. The definition of $D(S)$ and $K_{1,S}$ implies, that $(\alpha,x)_{\mathfrak{p}}=1$ for each $\mathfrak{p}\in S'$. Thus factoring out $(I^{S'})^2\cdot i(K_{1,S})$ we get a pairing

$$D(S)\times R(S)\to\mu_2.$$

Theorem 7. *For each generating S-reciprocity law γ the pairing $K(S)\times C(S)\to\mu_8$, defined by $(k,c)\to\gamma(k\cdot c)$, is a perfect duality, which extends the pairing $D(S)\times R(S)\to\mu_2$ given by the 2-power residue symbol.*

Remark. Before we give a proof of Theorem 7, we want to show, that it implies Artin's reciprocity law for multiquadratic extensions of F. The pairing $D(S)\times R(S)\to\mu_2$ given by the 2-power residue symbol is non-degenerate in the first argument by the global square theorem (cf. O'Meara [9], 65:15). Theorem 7 implies, that it is a perfect duality.

Let $L_S=F(\sqrt{D(S)})$ be the field extension of F, obtained by adjoining all square roots of elements from $D(S)$, and let $I_{L_S}^{S'}$ denote the ideal group of L_S of all ideals relatively prime to S'. Via Kummer theory the Artin map $\varphi_{L_S/F}$: $I^{S'}\to\mathrm{Gal}(L_S/F)$ induces the pairing $D(S)\times I^{S'}\to\mu_2$ given by the 2-power residue symbol. Thus our result so far implies, that $i(K_{1,S})\cdot(I^{S'})^2$ is precisely the kernel of the Artin map. Now clearly the group $N_{L_S/F}(I_{L_S}^{S'})$ of norms is contained in the kernel of $\varphi_{L_S/F}$ (cf. Janusz [2], Corollary 3.2.).

On the other hand the group $(I^{S'})^2$ is contained in $N_{L_S/F}(I_{L_S}^{S'})$, since the Galois group of L_S over F has exponent 2. Thus we see, that the kernel of the Artin map is precisely $N_{L_S/F}(I_{L_S}^{S'})\cdot i(K_{1,S})$ which proves Artin's reciprocity law for L_S over F.

Let $L=F(\sqrt{D})$ be an arbitrary multiquadratic extension of F. We find a set S, such that $D\subset D_{S'}$ hence $L\subset L_S$. By Hasse [1], Satz 135, Artin's reciprocity law is true for L/F as well.

Proof of Theorem 7. Let $k \in K(S)$ be non-trivial. By the theorem of Hasse-Minkowski there is a prime $\mathfrak{p}$, such that $k_\mathfrak{p}$ is non-trivial. As is easily seen (cf. Knebusch-Scharlau [4], Satz 6.1.) there is some $c \in W'(F_\mathfrak{p})$, such that $k_\mathfrak{p} \otimes c = \varphi_\mathfrak{p}$. This implies, that the pairing $K(S) \times C(S) \to \mu_8$ is non-degenerate in the first argument. To show, that it is non-degenerate in the second argument as well, it suffices to prove, that the groups $K(S)$ and $C(S)$ have the same order. We compare the two sequences

$$\begin{array}{ccccccc} \ker \partial^S & \longrightarrow & W'(F) & \xrightarrow{\partial_S} & \coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W'(F_\mathfrak{p}) & \longrightarrow & C(S) \\ \downarrow & & \downarrow \wr & & \downarrow & & \downarrow \\ K(S) & \longrightarrow & W'(F) & \xrightarrow{\partial_S} & \coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/W(R_\mathfrak{p}) & \longrightarrow & \operatorname{coker} \partial_S. \end{array}$$

The snake lemma implies, that

$$\frac{|K(S)|}{|\ker \partial^S|} \cdot \frac{|C(S)|}{|\operatorname{coker} \partial_S|} = \prod_{\mathfrak{p} \in S} |W'(F_\mathfrak{p})| \cdot \prod_{\mathfrak{p} \in D} \frac{|W(R_\mathfrak{p})|}{|WQ(R_\mathfrak{p})|}.$$

As is well-known (cf. Lam [5], Chap. 6, Theorem 2.2., Corollary 2.30) the order of $W'(F_\mathfrak{p})$ is computed as follows:

$$|W'(F_\mathfrak{p})| = \begin{cases} 2^3 & \text{if } \mathfrak{p} \text{ is real} \\ 2^4 & \text{if } \mathfrak{p} \text{ is finite, non-dyadic} \\ 2^{N_\mathfrak{p}+4} & \text{if } \mathfrak{p} \text{ is dyadic, } [F_\mathfrak{p} : \mathbb{Q}_2] = N_\mathfrak{p}. \end{cases}$$

Now we observe, that for a dyadic prime $\mathfrak{p}$ we have $\dfrac{|W(R_\mathfrak{p})|}{|WQ(R_\mathfrak{p})|} = \dfrac{|W(F_\mathfrak{p})|}{4} = 2^{N_\mathfrak{p}+2}$.

Since furthermore $\sum_{\mathfrak{p} \text{ dyadic}} N_\mathfrak{p} = r + 2c$, we get

$$(*) \qquad \frac{|K(S)|}{|\ker \partial^S|} \cdot \frac{|C(S)|}{|\operatorname{coker} \partial_S|} = 2^{2c+4|S|+2|D|}.$$

To go further we have to describe $\ker \partial^S$ and $\operatorname{coker} \partial_S$. Now, elements from $\ker \partial^S$ must have even rank and trivial Witt-invariant, hence the discriminant defines an isomorphism between $\ker \partial^S$ and the group $\Delta(S)$, which is defined by

$$\Delta(S) = \{\alpha \in F^* \mid F_\mathfrak{p}(\sqrt{\alpha})/F_\mathfrak{p} \text{ unramified for } \mathfrak{p} \notin S,\ \alpha \in F_\mathfrak{p}^{*2} \text{ for } \mathfrak{p} \in S\}/F^{*2}.$$

For the group $\operatorname{coker} \partial_S$ the following holds:

Lemma 8. *Let $I(S_f)$ denote the image of S_f in the class group C. Then* $\operatorname{coker} \partial_S \tilde{\to} C/C^2 \cdot I(S_f)$.

Proof. There is a surjective homomorphism

$$\coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/W(R_\mathfrak{p}) \to C/C^2 \cdot I(S_f),$$

which sends a family $\{\eta_\mathfrak{p}\}$ to the class of $\prod_{\mathfrak{p}\notin S}\mathfrak{p}^{v_\mathfrak{p}(d(\eta_\mathfrak{p}))}$. Clearly this factors through coker ∂_S. If the image of $\{\eta_\mathfrak{p}\}$ vanishes, we have an equation $\prod_{\mathfrak{p}\notin S}\mathfrak{p}^{v_\mathfrak{p}(d(\eta_\mathfrak{p}))}$ $=\mathfrak{b}^2\cdot\mathfrak{a}\cdot(\alpha)$ for some ideal $\mathfrak{b}$, some $\mathfrak{a}\in S_f$ and some $\alpha\in F^*$. We find a global form of discriminant α, which maps onto $\eta_\mathfrak{p}$ for all $\mathfrak{p}$. Thus $\{\eta_\mathfrak{p}\}$ is trivial in coker ∂_S.

It is easy to verify, that the 2-power residue symbol induces a bimultiplicative pairing

$$\Delta_S\times C/C^2\cdot I(S_f)\to\mu_2,$$

which is non-degenerate in the first argument. Hence $|\ker\partial^S|\leqq|\operatorname{coker}\partial_S|$ and thus $\dfrac{|K(S)|}{|\operatorname{coker}\partial_S|}\leqq\dfrac{|C(S)|}{|\ker\partial^S|}$. We are done, if we show, that $\dfrac{|K(S)|}{|\operatorname{coker}\partial_S|}=2^{c+2|S|+|D|}$, since then the equation (*) implies, that $\dfrac{|K(S)|}{|\operatorname{coker}\partial_S|}=\dfrac{|C(S)|}{|\ker\partial^S|}$, hence $|K(S)|$ $=|C(S)|$. Now by Corollary 3 we have $|K(S)|=2^{c+2|S|+|D|-|S_0|}\cdot|C/C^2|$ and coker $\partial_S\simeq C/C^2\cdot I(S_f)$, which has order $2^{-|S_0|}\cdot|C/C^2|$. Thus we get $\dfrac{|K(S)|}{|\operatorname{coker}\partial_S|}$ $=2^{c+2|S|+|D|}$, as claimed.

We still have to show, that the pairing $K(S)\times C(S)\to\mu_8$ extends the pairing $D(S)\times R(S)\to\mu_2$ given by the 2-power residue symbol. Let $\eta\in Br_2(S)$. Then for each $\mathfrak{p}\in S'$ $\eta_\mathfrak{p}$ is either trivial or represented by $\varphi_\mathfrak{p}$. Now, since $\varphi_\mathfrak{p}\cong\varphi_\mathfrak{p}\otimes\langle a\rangle$ (cf. Lam [5], Chap. 6: Corollary 2.11, Chap. 10: Corollary 1.7) for all $a\in F_\mathfrak{p}^*$, we have $\gamma(\eta\cdot c)=1$ for all $c\in C_0(S)$. Thus $Br_2(S)$ is orthogonal to $C_0(S)$ and hence we get a pairing

$$K_0(S)\times C_0(S)\to\mu_4.$$

The same argument as above shows, that the subgroup μ_2 of $C_0(S)$ is orthogonal to the subgroup $D(S)$ of $K_0(S)$ and hence we get a perfect duality $D(S)$ $\times R(S)\to\mu_2$. Let $\alpha\in D(S)$ and represent α by the form $\langle 1,-\alpha\rangle\in K(S)$. Let $\mathfrak{p}$ be any prime not contained in S' and choose a uniformizing parameter $\pi_\mathfrak{p}$ at $\mathfrak{p}$. The value of the pairing between α and the class of $\mathfrak{p}$ in $R(S)$ is now given by $\gamma_\mathfrak{p}(\langle 1,-\alpha_\mathfrak{p}\rangle\otimes\langle 1,-\pi_\mathfrak{p}\rangle)=\gamma_\mathfrak{p}(\langle 1,-\alpha_\mathfrak{p},-\pi_\mathfrak{p},\alpha_\mathfrak{p}\pi_\mathfrak{p}\rangle)$ which is trivial except if $\varphi_\mathfrak{p}$ $=\langle 1,-\alpha_\mathfrak{p},-\pi_\mathfrak{p},\alpha_\mathfrak{p}\pi_\mathfrak{p}\rangle$. Now, since $\alpha_\mathfrak{p}$ is a unit mod squares, the latter happens precisely, if $\left(\dfrac{\alpha}{\mathfrak{p}}\right)=-1$. Thus we have $\gamma_\mathfrak{p}(\langle 1,-\alpha_\mathfrak{p}\rangle\otimes\langle 1,-\pi_\mathfrak{p}\rangle)=\left(\dfrac{\alpha}{\mathfrak{p}}\right)$ as claimed.

Corollary 9. *The group extensions defining $C(S)$ are dual to those defining $K(S)$.*

Remarks. 1) The special case, that $S=W$ has been treated by Knebusch-Scharlau [4] and Lannes [7]. In this case we have $K(W)=W(R)/I^3(F)$ and the pairing $\ker\partial^W\times\operatorname{coker}\partial_W\to\mu_2$ is just the pairing $WQ(R)_t\times C/C^2\to\mu_2$ described in [7].

2) Let $\mathfrak{a}$ be a fractional ideal in F. In [4] Knebusch-Scharlau studied reciprocity laws for $\mathfrak{a}$. There is an obvious notion of an S-reciprocity law for $\mathfrak{a}$: replace condition i) by the condition

$$\text{i')}\quad \gamma_\mathfrak{p}(WQ(\mathfrak{a}_\mathfrak{p}))=1\quad\text{for all } \mathfrak{p}\notin S.$$

The existence of generating S-reciprocity laws for $\mathfrak{a}$ only depends on the class of $\mathfrak{a}$ in $C/C^2 \cdot I(S_f)$. Now imitating the proof of Lemma 5.7 in [4], one can easily show, that there exist generating S-reciprocity laws for $\mathfrak{a}$ if and only if $\mathfrak{a}$ vanishes in $C/C^2 \cdot I(S_f)$.

3) It is worth mentioning, that if the set S is sufficiently large, i.e. contains a basis of C/C^2, we get short exact sequences

$$0 \to K(S) \to W'(F) \to \coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/W(R_\mathfrak{p}) \to 0,$$

$$0 \to W'(F) \to \coprod_{\mathfrak{p} \notin S} W(F_\mathfrak{p})/WQ(R_\mathfrak{p}) \oplus \coprod_{\mathfrak{p} \in S} W'(F_\mathfrak{p}) \to C(S) \to 0.$$

References

1. Hasse, H.: Vorlesungen über Klassenkörpertheorie. Würzburg: Physica-Verlag 1967
2. Janusz, G.J.: Algebraic number fields. New York-London: Academic Press 1973
3. Knebusch, M.: Grothendieck- und Wittringe von nichtausgearteten symmetrischen Bilinearformen. S.-ber. Heidelberger Akad. Wiss. 1969/70, 3. Abh. Berlin-Göttingen-Heidelberg: Springer 1970
4. Knebusch, M., Scharlau, W.: Quadratische Formen und quadratische Reziprozitätsgesetze. Math. Z. **121**, 346-368 (1971)
5. Lam, T.Y.: The Algebraic Theory of Quadratic Forms. Reading: Benjamin 1973
6. Lang, S.: Algebraic Number Theory. Reading-Menlo Park-London-Don Mills: Addison-Wesley 1970
7. Lannes, J.: Formes quadratique d'enlacement sur l'anneau des entiers d'un corps de nombres. Ann. Sci. Ecole. Norm. Sup. **8**, 535-579 (1975)
8. Milnor, J., Husemoller, D.: Symmetric bilinear forms. Ergebnisse der Mathematik. Berlin-Heidelberg-New York: Springer 1973
9. O'Meara, O.T.: Introduction to quadratic forms. Berlin-Heidelberg-New York: Springer 1963
10. Scharlau, W.: Quadratic Forms. Queen's Papers on Pure and Applied Mathematics 22. Queen's University, Kingston, Ontario 1969
11. Scharlau, W.: Quadratic Reciprocity Laws. J. Number Theory **4**, 78-97 (1972)
12. Weil, A.: Sur certains groupes d'operateurs unitaires. Acta Math. **111**, 143-211 (1964)

Received July 10, 1981; received in final form October 14, 1981

Math. Z. 180, 91–105 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On Extreme Bloch Functions with Prescribed Critical Points

Stephan Ruscheweyh [1] and Karl-Joachim Wirths [2]

[1] Mathematisches Institut der Universität, Am Hubland, D-8700 Würzburg, Federal Republic of Germany

[2] Institut C für Mathematik der Technischen Universität, Pockelstraße 14, D-3300 Braunschweig, Federal Republic of Germany

Dedicated to Prof. Dr. Dr. h.c. E. Peschl on the occasion of his 75th birthday

I. Introduction

For F analytic in the unit disc $\Delta=\{z\,|\,|z|<1\}$ let

$$\mu_F(z)=(1-|z|^2)\,|F'(z)|.$$

The Bloch space $\mathscr{B}$ consists of those F for which

$$\mathrm{M}(F)=\sup_{z\in\Delta}\mu_F(z)<\infty.$$

In the present paper we are dealing with the functions $F\in\mathscr{B}$ for which the set

$$\Lambda(F)=\{z\in\Delta\,|\,\mu_F(z)=\mathrm{M}(F)\}$$

has an accumulation point in Δ. The study of these functions (characterization and construction) is relevant to distortion theorems for Bloch functions with prescribed critical points.

It is a consequence of the work of Cima and Wogen [2] that any extreme point of the unit ball $\mathscr{B}_0\subset\mathscr{B}$ (see [2]) has this property. This fact was one of the motivations for our paper.

A general description of the shape of such sets $\Lambda(F)$ is given in Theorem 1:

Theorem 1. *For $F\in\mathscr{B}$ let $\Lambda(F)$ have an accumulation point in Δ. Then exactly one of the following two statements holds:*

a) *$\Lambda(F)$ consists of one analytic Jordan curve $\gamma\subset\Delta$ and at most finitely many points, all located in the bounded component Γ of $\mathbb{C}\setminus\gamma$.*

b) *$\Lambda(F)$ is the union of isolated points and disjoint simple analytic arcs terminating on $\partial\Delta$.*

In the sequel we study case a) of Theorem 1. For convenience we denote by $\mathscr{E}$ the set of all $F\in\mathscr{B}$ such that $F(0)=0$, $\mathrm{M}(F)=1$ and condition a) of Theorem 1 holds. For these functions we give the following constructive description.

0025-5874/82/0180/0091/$03.00

Definition 1. Let $\mathscr{R}$ be the class of rational functions which map a domain $\Pi = \Pi_f \supset \bar{\Delta}$ conformally onto Δ. For $f \in \mathscr{R}$ let

$$Q_f(w) = \frac{2f'(w)\,\overline{f(1/\bar{w})}}{1 - f(w)\,\overline{f(1/\bar{w})}}, \quad w \in \bar{\Delta}, \tag{1}$$

$$G_f(w) = (1 - |f(1)|^2)^{-1} \exp\left(\int_1^w Q_f(t)\,dt\right). \tag{2}$$

f is said to be in $\mathscr{R}_0$ if and only if $G_f(w)$ has an analytic continuation into $\bar{\Delta}$.

For $f \in \mathscr{R}$ we write f^{-1} for the inverse of $f|_\Pi$ and

$$\mathsf{A}(f) = \{f(w) \mid w \in \bar{\Delta},\ f(w) = f(1/\bar{w})\}. \tag{3}$$

In Lemmas 4, 5 we prove that G_f is rational for $f \in \mathscr{R}_0$ and that

$$F_f(z) = \int_0^z (G_f \circ f^{-1})(t)\,dt \tag{4}$$

is analytic with bounded derivative in Δ.

Theorem 2. *$F \in \mathscr{E}$ if and only if $\mathsf{M}(F) = 1$ and $F = \varepsilon F_f$ for certain $f \in \mathscr{R}_0$ and $\varepsilon \in \mathbb{C}$, $|\varepsilon| = 1$. In this case*

$$f(\partial\Delta) \subset \Lambda(F) \subset \mathsf{A}(f). \tag{5}$$

The statements concerning F_f and Theorem 2 lead to the following

Corollary 1. *If $F = \varepsilon F_f \in \mathscr{E}$ then F' is algebraic of genus zero and has a rational uniformization $z = f(w)$, $F' = G_f(w)$. Furthermore,*

$$\lim_{z \to \partial\Delta} \mu_F(z) = 0. \tag{6}$$

The members of $\mathscr{E}$ are characterized by the location and orders of their critical points. We set

$$\Omega_F = \{z \mid F'(z) = 0\}. \tag{7}$$

Theorem 3. a) *Let $F \in \mathscr{E}$ and Γ as in Theorem* 1, a). *Then we have $\varnothing \neq \Omega_F \subset \Gamma$. Clearly $\#\Omega_F < \infty$.*

b) *If $F_1, F_2 \in \mathscr{E}$ and F_1'/F_2' analytic and nonvanishing in Δ then $F_1 = \varepsilon F_2$ for a certain ε, $|\varepsilon| = 1$.*

In the proof of Theorem 3 the following useful fact turns out as a by-product. If $f \in \mathscr{R}_0$, $F = F_f \in \mathscr{E}$, then $\#\Omega_F$ equals the number of poles of f counted according to their orders. This is important for the construction of $F \in \mathscr{E}$ with prescribed Ω_F and leads - via (5) - to

Theorem 4. *Let $F \in \mathscr{E}$ with $\#\Omega_F = n$. Then $\#(\Lambda(F) \setminus \gamma) \leqq (n-1)(2n-3)$.*

In Chap. 4 we construct all $F \in \mathscr{E}$ with $\#\Omega_F \leqq 2$. In particular, our results show that $\Lambda(F)$ is a circle if $\#\Omega_F = 1$. If $\#\Omega_F = 2$ $\Lambda(F)$ is not a circle and $\Lambda(F)$

$\setminus \gamma = \varnothing$. These results are related to the work of Cima and Wogen [2]: their characterization and our Corollary 1 prove that $\mathscr{E}$ is the set of extreme points of the unit ball $\tilde{\mathscr{B}}_0$ (see [2]). Their problem whether $\Lambda(F)$ is necessarily a circle has therefore found a negative solution. Another problem namely to show that $\Lambda(F)$ is necessarily connected (equivalent to $\Lambda(F) \setminus \gamma = \varnothing$) remains open although our results prove this for $F \in \mathscr{E}$ with $\# \Omega_F \leqq 2$. It may be of interest that our results imply that not every finite set in Δ can occur as Ω_F, $F \in \mathscr{E}$, if the orders of the critical points are prescribed. The following theorem proves the relevance of $\mathscr{E}$ for the whole Bloch space and the importance of its explicit construction.

Theorem 5. *Let $G \in \mathscr{B}$, $F \in \mathscr{E}$ with $\Gamma = \Gamma(F)$ as above. If G'/F' is analytic in Γ we have*

$$|G'(z)| \leqq \mathrm{M}(G)|F'(z)|, \qquad z \in \Gamma. \tag{8}$$

If equality holds in (8) *for a certain $z_0 \in \Gamma \setminus \Omega_F$ then $G' = \varepsilon \mathrm{M}(G) F'$ for a certain ε, $|\varepsilon| = 1$.*

We wish to mention that other sharp estimates for $|F'|$, $F \in \mathscr{B}$, depending on the minimal order of the critical points are due to E. Peschl [4, 5]. Some of our present results are contained in [8] where a different method has been applied.

II. Some Lemmas

Our method to describe $\Lambda(F)$, $F \in \mathscr{B}$, rests on a careful study of the sets

$$\Sigma(H) = \{z \in \Delta \mid H(z) = \bar{z}\} \tag{9}$$

where H is meromorphic in Δ. In fact, if $z_0 \in \Lambda(F)$ then $\mu_F(z)$ has a local maximum in z_0 and thus

$$\frac{\partial}{\partial z} \mu_F(z) = 0, \qquad z = z_0 \tag{10}$$

or equivalently

$$F''(z_0)(1 - |z_0|^2) = 2\bar{z}_0 F'(z_0). \tag{11}$$

Let

$$H(z) = \frac{F''(z)}{zF''(z) + 2F'(z)}. \tag{12}$$

Then $H(z)$ is meromorphic in Δ and from (11) we obtain $z_0 \in \Sigma(H)$. Thus the fundamental relation

$$\Lambda(F) \subset \Sigma(H) \tag{13}$$

holds for $F \in \mathscr{E}$ non constant and H given by (12).

Some of the following observations concerning $\Sigma(H)$ are known. For example, if an analytic arc γ is contained in $\Sigma(H)$ then H is called the Schwarz function of γ (compare Davis [3], Aharonov and Shapiro [1]). Since our aim is different we prefer to include a complete discussion of the relevant properties.

Lemma 1. *Let $z_0 \in \Delta$ be an accumulation point of $\Sigma(H)$ for H meromorphic in Δ. Then there is an analytic arc γ passing through z_0 and a neighborhood $N(z_0) \subset \Delta$ such that for $z \in N(z_0)$ we have $z \in \Sigma(H)$ if and only if $z \in \gamma$.*

Proof (compare [6], Chap. 3). There exists P meromorphic in Δ, conformal in z_0 such that $P(z) \in \mathbb{R}$ for $z \in \Sigma(H)$ (for instance, one of the functions $H(z)+z$, $i(H(z)-z)$). Choose $N(z_0) \subset \Delta$ such that both, $N(z_0)$ and $\{\overline{H(z)} \mid z \in N(z_0)\}$, are in the range of conformality of P. The functions $P(z)$ and $\overline{P(\overline{H(z)})}$ are analytic in $N(z_0)$ and coincide on $\Sigma(H) \cap N(z_0)$. Thus $P(z) = \overline{P(\overline{H(z)})}$ in $N(z_0)$ by the identity principle. This shows that in $N(z_0)$ the equation $z = \overline{H(z)}$ holds if and only if $P(z) \in \mathbb{R}$ which is the case if and only if $z \in \gamma = P^{-1}(P(N(z_0)) \cap \mathbb{R})$.

Remark 1. Let $F \in \mathcal{B}$ and H as in (12). If $\Sigma(H)$ contains an analytic arc $\gamma = \gamma(t)$, $0 \leqq t \leqq 1$, the function $\mu_F(\gamma(t))$ has vanishing derivative in t, since (10) holds for arbitrary $z \in \Sigma(H)$. Thus $\mu_F(z) = \text{const.}$ on γ.

Lemma 2. *Let H be meromorphic in Δ. For $\Sigma(H)$ to contain an analytic Jordan curve γ it is necessary and sufficient that*

$$H(z) = \overline{f(1/\overline{f^{-1}(z)})}, \quad z \in \Delta, \tag{14}$$

for a certain $f \in \mathcal{R}$. In this case

$$\gamma = f(\partial\Delta) \subset \Sigma(H) = A(f). \tag{15}$$

The set $\Sigma(H) \setminus \gamma$ is finite and lies in the bounded component Γ of $\mathbb{C} \setminus \gamma$.

Proof. If $f \in \mathcal{R}$ then H given by (14) is clearly meromorphic in Δ. Furthermore, if $z \in \gamma = f(\partial\Delta)$ we have $1/\overline{f^{-1}(z)} = f^{-1}(z)$ and thus $H(z) = \bar{z}$. This proves the sufficiency part. Now let $\gamma \subset \Sigma(H)$ be an analytic Jordan curve and f a conformal mapping of Δ onto the bounded component Γ of $\mathbb{C} \setminus \gamma$. Then f has a conformal extension into a domain $\Delta^* \supset \bar{\Delta}$ and $f(\partial\Delta) = \gamma$. The functions $H(f(w))$, $\overline{f(1/\bar{w})}$ are analytic in a ring domain N containing $\partial\Delta$ and on $\partial\Delta$ we have

$$H(f(w)) = \overline{f(1/\bar{w})}. \tag{16}$$

Thus (16) holds in N and (16) gives the meromorphic continuation of f into $\overline{\mathbb{C}}$ which implies that f is rational.

Now let f_R^{-1} be the algebraic inverse to f defined on the Riemann surface $f(\overline{\mathbb{C}})$. Clearly

$$H_R(p) = \overline{f(1/\overline{f_R^{-1}(p)})}, \quad p \in f(\overline{\mathbb{C}}), \tag{17}$$

is meromorphic in $f(\overline{\mathbb{C}})$ and is the meromorphic continuation of H by (16). Since H is meromorphic in Δ there is a certain sheet in $f(\overline{\mathbb{C}})$ containing a unit

disc, say Δ_R, on which $H_R(p)=H(z)$, trace $(p)=z$, and $f(\bar{\Delta})\subset\Delta_R$. Let $\Pi=f_R^{-1}(\Delta_R)$ such that trace $(f(\Pi))$ is the unit disc and $\Pi\supset\bar{\Delta}$. The next step is to prove univalence of f on Π. To this end assume there exist w_1, $w_2\in\Pi\setminus\bar{\Delta}$, $w_1\neq w_2$, with $f(w_1)=f(w_2)$. Since $f(w_1)$, $f(w_2)\in\Delta_R$ this implies

$$H(f(w_1))=H_R|_{\Delta_R}(f(w_1))=H_R|_{\Delta_R}(f(w_2))=H(f(w_2)).$$

On the other hand, f is univalent in Δ and thus $H(f(w_1))\neq H(f(w_2))$ by (16). It remains to consider the case $w_1\in\bar{\Delta}$, $w_2\in\Pi\setminus\{\infty\}$. Since Π is a domain containing $\bar{\Delta}$ we find an analytic Jordan curve σ such that its interior domain belongs to Π and contains $\bar{\Delta}\cup\{w_2\}$. An application of the argument principle to $f(\sigma)$ shows $f(w_1)\neq f(w_2)$. Thus f is univalent on $\Pi\setminus\{\infty\}$ and hence on Π. This completes the proof of $f\in\mathscr{R}$ and (14) follows from (17).

It is clear from our construction that for H given in (14) we have $\gamma=f(\partial\Delta)\subset\Sigma(H)$. We wish to show $\Sigma(H)\subset\bar{\Gamma}$. Assume $z_0\in\Sigma(H)\setminus\bar{\Gamma}$. Then $w_0=f^{-1}(z_0)\in\Pi\setminus\bar{\Delta}$ and (16) leads to the contradiction

$$z_0=\overline{H(z_0)}=\overline{H(f(w_0))}=f(1/\overline{w_0})\in\Gamma.$$

For $z_0\in\bar{\Gamma}$ we have $z_0\in\Sigma(H)$ if and only if $z_0=f(1/\overline{f^{-1}(z_0)})$, or

$$f(w_0)=f(1/\overline{w_0}),\qquad z_0=f(w_0),\qquad w_0\in\Delta.$$

This is equivalent to $\Sigma(H)=\mathsf{A}(f)$. It follows from Lemma 1 that $\Sigma(H)\setminus\gamma$ cannot accumulate on γ. Thus if $\Sigma(H)\setminus\gamma$ is infinite it has an accumulation point in Γ. This implies the existence of a second analytic Jordan curve $\gamma^*\subset\Gamma$, $\gamma^*\subset\Sigma(H)$. The above arguments, applied to γ^*, rule out such a possibility since $\gamma\subset\Sigma(H)$ is in the exterior domain of γ^*. The proof of Lemma 2 is complete.

The next Lemma states the linear invariance of $\mathscr{R}_0$.

Lemma 3. *Let* $f\in\mathscr{R}_0$, $T(w)=\varepsilon\dfrac{w-a}{1-\bar{a}w}$ *for certain* ε, $a\in\mathbb{C}$, $|\varepsilon|=1$, $|a|<1$. *Then* $f_T=f\circ T\in\mathscr{R}_0$ *and there exists* $c\in\mathbb{C}$ *such that* $G_{f_T}=c(G_f\circ T)$.

Proof. Clearly $f_T\in\mathscr{R}$. We note that $f\in\mathscr{R}$ is in $\mathscr{R}_0$ if and only if there is a solution G_f, analytic in $\bar{\Delta}$, of the equation

$$G_f'(w)=Q_f(w)G_f(w) \tag{18}$$

with

$$G_f(1)=(1-|f(1)|^2)^{-1}. \tag{19}$$

A straightforward calculation shows that $G=G_f\circ T$ is a solution of $G'=Q_{f_T}\cdot G$ in $\bar{\Delta}$, $G\not\equiv 0$. Since $Q_{f_T}\neq 0,\infty$ on $\partial\Delta$ there is a constant $c\in\mathbb{C}$ with $G_{f_T}=cG$.

Lemma 4. *Let* $f\in\mathscr{R}_0$. *Then* F_f (*see* (4)) *satisfies*

i) F_f *is analytic with bounded derivative in* Δ,

ii) $\Lambda(F_f)\subset\mathsf{A}(f)$,

iii) $\mu_{F_f}(z)=1$, $z\in f(\partial\Delta)$.

Proof. First assume $f(\infty)=\infty$ which implies the boundedness of Π. For $w\in\overline{\Pi}\setminus\Delta$ we have $|f(w)|\leqq 1$, $|f(1/\bar{w})|<1$. Thus Q_f is bounded and analytic on $\overline{\Pi}\setminus\Delta$. Since Π is a bounded simply connected domain we see that G_f has an analytic continuation into $\overline{\Pi}$. Therefore $F_f'=G_f\circ f^{-1}$ is analytic and bounded in Δ. This proves i) and $F_f\in\mathscr{B}$. Define $H_f(z)$ by (12) using $F=F_f$. Using the relations ($z=f(w)$)

$$F_f'(z)=G_f(w),\qquad F_f''(z)=G_f'(w)/f'(w),\qquad G_f'(w)=Q_f(w)G(w)$$

together with the explicit form of Q_f (see (1)) we obtain $H_f(z)=\overline{f(1/\overline{f^{-1}(z)})}$. ii) follows now from (13) and Lemma 2. Since $f(\partial\Delta)\subset\Sigma(H_f)$ by Lemma 2 it follows from Remark 1 that $\mu_{F_f}(z)=\text{const.}$ on $f(\partial\Delta)$. But $\mu_{F_f}(f(1))=1$ by construction and iii) is settled.

Let f be as above and $f_T=f\circ T$ where T is an automorphism of Δ. It follows from Lemma 3 that every function in $\mathscr{R}_0$ has a representation of this form. Furthermore, since

$$1=(1-|f_T(1)|^2)|G_{f_T}(1)|=(1-|f(T(1))|^2)|c|\,|G_f(T(1))|=|c|$$

by Lemma 3 and iii) above, we see that

$$F_{f_T}'=G_{f_T}\circ f_T^{-1}=c(G_f\circ T\circ T^{-1}\circ f^{-1})=cG_f\circ f^{-1}=cF_f'.$$

This completes the proof of Lemma 4.

Lemma 5. *Let $f\in\mathscr{R}_0$. Then G_f is a rational function.*

Proof. Let

$$B(w)=(1-f(w)\overline{f(1/\bar{w})})G_f(w).$$

B is meromorphic in $\bar{\Delta}$ and for $w\in\partial\Delta$ we have $|B(w)|=\mu_{F_f}(f(w))=1$. Thus B is the quotient of finite Blaschke products which implies G_f rational.

III. Proofs of Theorems 1–5

Proof of Theorem 1. From (13) we have $\Lambda(F)\subset\Sigma(H)$. Thus $\Sigma(H)$ has an accumulation point $z_0\in\Delta$. Lemma 1 then shows that $\Sigma(H)$ contains a simple analytic arc γ passing through z_0 and γ cannot terminate in Δ. Remark 1 shows that $\gamma\subset\Lambda(F)$. If, however, γ is a closed curve Lemma 2 applies and proves that $\Lambda(F)\setminus\gamma\subset\Sigma(H)\setminus\gamma$ is finite and located in Γ.

Proof of Theorem 2. Let $F\in\mathscr{E}$ and the analytic Jordan curve $\gamma\subset\Lambda(F)\subset\Sigma(H)$ where H is given by (12). Lemma 2 shows $H(f(w))=\overline{f(1/\bar{w})}$ for a certain $f\in\mathscr{R}_0$, which gives the differential equation

$$\frac{F''(f(w))}{f(w)F''(f(w))+2F'(f(w))}=\overline{f(1/\bar{w})}.$$

Using (18) we obtain $F'(f(w))=\varepsilon G_f(w)$ with a certain $\varepsilon\in\mathbb{C}$. The special case $w=1$ (where $f(w)\in\gamma$) shows that $|\varepsilon|=1$. The remaining conclusions follow from

Lemma 4. Corollary 1 is a consequence of Theorem 2 and Lemma 4; i), 5.

We turn to the proof of Theorem 3, part a). Note that $F=\varepsilon F_f$ for certain $f\in\mathscr{R}_0$, $|\varepsilon|=1$. Furthermore, in view of (12) we have

$$\frac{F''(z)}{F'(z)}=\frac{2H(z)}{1-zH(z)}=L(z), \qquad z\in\Delta. \tag{20}$$

Proposition 1. *The number of critical points of F in Γ agrees with the number of poles (counted according to their orders) of H in Γ.*

Proof. Since F is analytic in Δ L has just simple poles with positive residues in Δ, and every such pole corresponds to a critical point of F. A simple calculation shows that $z_0=0$ is such a critical point of F if and only if H has a simple pole in z_0. Thus in order to prove the proposition, it suffices to study the singularities of $zL(z)$. These occur exactly where $D(z)=1-zH(z)$ has its zeros (and all these zeros are simple in Δ). Since $D(z)>0$ on γ, the argument principle shows that D has as many zeros as poles (orders counted) in Γ. But $z=0$ is no pole of D and thus the poles of D agree with the poles of H (orders counted) in $\Gamma\setminus\{0\}$.

Proposition 2. *The poles of H in Δ are located in Γ. Their number (orders counted) agrees with the number of poles (orders counted) of f in $\overline{\mathbb{C}}$.*

Proof. Recall $\Pi\supset\bar{\Delta}$, $f(\Pi)=\Delta$. The relation $H(f(w))=\overline{f(1/\bar{w})}$, $w\in\Pi$, shows that $|H(f(w))|\leqq 1$ for $w\in\Pi\setminus\Delta$. Thus H has no pole in $\Delta\setminus\Gamma$. Since f is conformal in Δ the same relation shows that $z_0=f(w_0)\in\Gamma$ is a pole of order m of H if and only if f has a pole of the same order in $1/\overline{w_0}$.

Proof of Theorem 3. a) From Propositions 1, 2 we conclude $\varnothing\neq\Omega_F$ since $f\in\mathscr{R}_0$ has at least one pole. It remains to show that $\Omega_F\subset\Gamma$. Let $z_0\in\Omega_F$, $z_0\in\Delta\setminus\Gamma$. If $z_0=0$ it follows from the arguments in the proof of Proposition 1 that H has a pole in z_0. This contradicts Proposition 2. If $z_0\neq 0$ we obtain from $D(z_0)=0$ the relation $H(z_0)=1/z_0$ or $1/f(w_0)=H(f(w_0))=\overline{f(1/\bar{w}_0)}$ with $z_0=f(w_0)$, $w_0\in\Pi\setminus\Delta$. But $\overline{f(1/\bar{w}_0)}\in\bar{\Gamma}\subset\Delta$ and $1/f(w_0)\in\mathbb{C}\setminus\Delta$.

b) Let γ_i, $i=1,2$, be the analytic Jordan curves with $\mu_{F_i}(z)=1$, $z\in\gamma_i$ and Γ_i the corresponding interior domains of γ_i. For $z\in\gamma_1$ we have

$$\frac{\mu_{F_2}(z)}{\mu_{F_1}(z)}=\frac{|F_2'(z)|}{|F_1'(z)|}\leqq 1.$$

By the maximum principle we deduce $|F_2'(z)|\leqq|F_1'(z)|$, $z\in\Gamma_1$. By a similar argument we obtain the reverse inequality in Γ_2, and thus $|F_1'(z)|=|F_2'(z)|$ in $\Gamma_1\cap\Gamma_2$. But $\Gamma_1\cap\Gamma_2$ contains an open neighborhood of each $z_0\in\Omega_{F_1}=\Omega_{F_2}$ and the latter sets are nonempty by part a). The result follows from the identity principle.

Note that the remark following the statement of Theorem 3 is a consequence of Propositions 1, 2. Before we turn to the construction of functions $f\in\mathscr{R}_0$ belonging to $F_f\in\mathscr{E}$ in Chap. 4 we wish to introduce some normalizations which are also useful in the proof of Theorem 4. First of all we may assume that F has one of its critical points of lowest order at $z=0$. This is clear from

the linear invariance of $\mathscr{E}$: for $F\in\mathscr{E}$, $T(z)=\dfrac{z+a}{1+\bar{a}z}$ we have $F^*=(F\circ T)(z)-(F\circ T)(0)\in\mathscr{E}$ and

$$\Lambda(F^*)=\{T^{-1}(z)|z\in\Lambda(F)\},\qquad \Omega_{F^*}=\{T^{-1}(z)|z\in\Omega_F\}.$$

The second relation is obvious and the first one follows from the identity

$$|F'(T(z))|(1-|T(z)|^2)=|F^{*\prime}(z)|(1-|z|^2).$$

If $F=F_f$ has a critical point at $z=0$ we see from Theorem 3 that $0\in\Gamma$ and thus f has a zero in Δ. We may assume $f(0)=0$, $f'(0)>0$, since an automorphism of the unit disc in the w-plane does not change F_f (see last section of the proof of Lemma 4). The critical point of F in $z=0$ implies a simple pole of H at the same point (see proof of Proposition 1). Since $H(f(w))=\overline{f(1/\bar{w})}$ we conclude that f has a simple pole at ∞. We proved

Proposition 3. *If $F\in\mathscr{E}$, $\#\Omega_F=n$, $0\in\Omega_F$, then $F=\varepsilon F_f$ for a certain $\varepsilon\in\mathbb{C}$, $|\varepsilon|=1$, and*

$$f(w)=w\frac{P(w)}{Q(w)}\in\mathscr{R}_0, \tag{21}$$

where P, Q are polynomials without common zeros and

$$\deg P\leqq n-1,\qquad \deg Q=n-1,\qquad P(0)/Q(0)>0.$$

Proof of Theorem 4. In view of Theorem 2 and Propositions 3 we have to estimate the number of solutions of

$$f(w)=f(1/\bar{w}),\qquad w\in\Delta, \tag{22}$$

where f is given by (21). (22) is equivalent to an equation

$$\sum_{j,k=0}^{n} a_{jk}w^j\bar{w}^k=0,\qquad a_{jk}\in\mathbb{C},\ w\in\Delta \tag{23}$$

where we may assume $a_{n,n}\geqq 0$. Taking the real part and imaginary part of (23) we obtain two real algebraic equations of orders $\leqq 2n$ and $\leqq 2n-1$, respectively. In both of these equations we may factor out the only common factor $(x^2+y^2-1)^k$ for a certain $k\in\mathbb{N}$. According to Bezout's theorem the remaining algebraic curves of orders $\leqq 2n-2$ and $\leqq 2n-3$ will intersect in at most $2(n-1)(2n-3)$ points. Since (22) is symmetric to $|w|=1$ it is clear that (22) has at most $(n-1)(2n-3)$ solutions in Δ.

The proof of Theorem 5 is similar to the proof of Theorem 3, b).

IV. Construction of Functions in $\mathscr{E}$

In this chapter we completely determine all $F\in\mathscr{E}$ with $\#\Omega_F\leqq 2$.

Theorem 6. *Let $F\in\mathscr{E}$, $\Omega_F=\{z_0\}$. If the order of the zero of F' in z_0 is $k\in\mathbb{N}$, then*

$$F(z)=\varepsilon\frac{k+2}{2k+2}\left(\frac{k+2}{k}\right)^{k/2}\left\{\left(\frac{z-z_0}{1-\bar{z}_0z}\right)^{k+1}-(-z_0)^{k+1}\right\} \tag{24}$$

for a certain $\varepsilon\in\mathbb{C}$, $|\varepsilon|=1$. *In this case*

$$\Lambda(F)=\left\{z\,\middle|\,\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|^2=\frac{k}{k+2}\right\}. \tag{25}$$

Proof. We may assume $z_0=0$ and according to Proposition 3 we have to find functions $f(w)=\alpha w\in\mathscr{R}_0$. In this case

$$Q_f(w)=\frac{2|\alpha|^2}{1-|\alpha|^2}\,\frac{1}{w}$$

and G_f is analytic in Δ if and only if $|\alpha|=\left(\frac{n}{n+2}\right)^{1/2}$, $n\in\mathbb{N}$. In this case $G_f(w)=(n+2)w/n$, $F_f'(z)=\frac{n+2}{2}\left(\frac{n+2}{n}\right)^{n/2}z^n$. It follows from our theory that εF_f, $|\varepsilon|=1$, are the only candidates for functions in $\mathscr{E}$ with exactly one critical point, located at $z=0$. In this particularly simple case $\mathsf{M}(F_f)=1$ and (25) follow from the simple observation $\mathsf{A}(f)=f(\partial\Delta)$.

Remark 2. Theorem 6 shows that in these cases $\Lambda(F)$ is a circle. These examples for members in $\mathscr{E}$ have previously been studied, see [2], and were the only known functions in $\mathscr{E}$. There is no other function $F\in\mathscr{E}$ with $\Lambda(F)$ a circle. In fact, if $F=F_f\in\mathscr{E}$ with $\Lambda(F)$ a circle then $f\in\mathscr{R}_0$ maps Δ conformally onto a disc and is therefore a Moebius transform. Thus f has only one (simple) pole in $\overline{\mathbb{C}}$ and this shows that $\#\Omega_F=1$.

The following Corollary to Theorem 6 plays a crucial role in our further development.

Corollary 3. *For* $F\in\mathscr{E}$ *let* γ, Γ *be as above. If* F' *has a zero* z_0 *of order* k *then*

$$\mathsf{N}=\left\{z\,\middle|\,\left|\frac{z-z_0}{1-\bar{z}_0 z}\right|^2<\frac{k}{k+2}\right\}\subset\Gamma. \tag{26}$$

Proof. Let F_0 be a function as in (24) (z_0 and k as in Corollary 3). Then by Theorem 5

$$(1-|z|^2)\,|F'(z)|\leqq(1-|z|^2)\,|F_0'(z)|<1,\qquad z\in\mathsf{N}.$$

Since $z_0\in\mathsf{N}\cap\Gamma$ this proves (26).

Lemma 6. *Let* α, $\beta\in\mathbb{C}\setminus\{0\}$, $\mu>0$, $\mu|\alpha|<1$. *Then*

$$f(w)=w\frac{\alpha w+\beta}{w+\mu^2\bar{\alpha}\beta} \tag{27}$$

is univalent in the disc $\{w\,||w|<\mu|\beta|\}$ *and* f^{-1} *has two branch points on* $\partial\Delta_{|\beta|}$ *where* $\Delta_{|\beta|}=\{z\,||z|<|\beta|\}$. f *maps*

1) *the domain*

$$\Theta=\{w\,||w+\beta/\alpha|^2>|\beta/\alpha|^2(1-\mu^2|\alpha|^2)\}\cap\{w\,||w|<\mu|\beta|\}$$

conformally onto $\Delta_{|\beta|}$,

2) *the disc* $\{w\,||w|<\mu|\beta|\}$ *onto* $\overline{\mathbb{C}}\setminus S$ *where* S *is the arc on* $\partial\Delta_{|\beta|}$ *connecting the branch points and containing* β.

Proof. The branch points are

$$z_{1,2}=\beta(1-2\mu^2|\alpha|^2\pm 2\mu|\alpha|\sqrt{\mu^2|\alpha|^2-1}),\qquad |z_{1,2}|=|\beta|,$$

and the equation $|f(w)|=|\beta|$ is equivalent to

$$(|w|^2-\mu^2|\beta|^2)(|w+\beta/\alpha|^2-|\beta/\alpha|^2(1-\mu^2|\alpha|^2))=0.$$

This together with $f(0)=0$, $0\in\Theta$ gives 1).

2) follows by reflection at corresponding circular arcs on the boundaries of Θ and $\Delta_{|\beta|}$.

Lemma 7. *Let* k, $l\in\mathbb{N}$, $k\leqq l$, *and* $b\geqq 1$. *The function* $a=a(b)$ *is defined as the largest positive root of the equation*

$$16(l+1)^2a^2b^2k^2+p(k+l+2)\,p(k-l)=0 \tag{28}$$

where

$$p(q)=a^2k^2(q+2)-(k+2)^2a^2b^2(q-2)+k^2(b^2-1)q.$$

Then

$$f(w)=w\frac{aw+b}{w+\dfrac{k+2}{k}ab}\in\mathscr{R}_0. \tag{29}$$

Proof. 1) For $b\geqq 1$ fixed (28) is quadratic in a^2 and has two solutions $a^2\in\left[0,\dfrac{k}{k+2}\right)$. In fact, the transformation

$$a^2=x+\frac{k}{k+2},\qquad b^2=y+\frac{k}{k+2}$$

applied to (28) gives

$$\begin{aligned}S(x,y)&=x^2y^2(k+2)^4(l(l+2)-k(k-2))+4x^2yk(k+2)^2(l(l+2)+k^2)\\&\quad+4xy^2k(k+2)^2(l(l+2)+k)+4x^2k^2(l(l+2)-k(k+2))\\&\quad+4y^2k^2l(l+2)+8xyk^2(k+1)l(l+2)\\&=A(y)x^2+B(y)x+C(y)=0.\end{aligned}$$

An immediate calculation shows that the discriminant $B^2(y)-4A(y)C(y)$ is positive under our assumptions so that there are two real roots in x. Since $S(x,y)>0$ for $x\geqq 0$, $y>0$, and $S\left(-\dfrac{k}{k+2},y\right)\geqq 0$, $S_x\left(-\dfrac{k}{k+2},y\right)<0$ for $y\geqq\dfrac{2}{k+2}$ we deduce that the parabola $S(\cdot,y)$ has its zeros within $\left[-\dfrac{k}{k+2},0\right)$. This shows that Lemma 6 applies to the functions (29).

2) We wish to show

$$a>(b+1)\Big/\left(1+\frac{k+2}{k}b\right). \tag{30}$$

In fact, we have

$$S\left(\left(\frac{b+1}{1+\frac{k+2}{2}b}\right)^2-\frac{k}{k+2},y\right)=S\left(\frac{-2y}{k\left(1+\frac{k+2}{k}b\right)^2},y\right)<0, \qquad b\geqq 1,$$

which shows that $\dfrac{b+1}{1+\frac{k+2}{2}b}$ lies between the smaller and the larger non-negative solution of (28). Note that this implies

$$a^*<(b+1)\Big/\left(1+\frac{k+2}{k}b\right) \tag{31}$$

for the smaller nonnegative solution a^* of (28), a fact used in the sequel.

3) Next we prove $f\in\mathscr{R}$. We use the notations of Lemma 6:

$$\mu=((k+2)/k)^{1/2}, \qquad \alpha=a, \qquad \beta=b.$$

First we show $\Delta\subset\Theta$. Since $|\mu|\beta>1$ and thus $\Delta\subset\Delta_{|\mu|\beta}$ we only have to verify $b\left(\left(1-\frac{k+2}{k}a^2\right)^{1/2}-1\right)\Big/a<-1$ which is equivalent with $a>2b\Big/\left(1+\frac{k+2}{k}b^2\right)$. The latter inequality follows from (30). In particular, f has no pole in Δ. It remains to prove $f(\partial\Delta)\subset\Delta$. But $|f(e^{i\varphi})|=|g(e^{i\varphi})|$ where

$$g(w)=\frac{aw+b}{w+\frac{k+2}{k}ab}$$

is a Moebius transform with real coefficients and without pole in Δ.

Thus for $w\in\bar{\Delta}$

$$|f(w)|\leqq|g(w)|\leqq\max\{|g(1)|,|g(-1)|\}.$$

The result follows from

$$|g(1)|=\frac{a+b}{1+\frac{k+2}{k}ab}<\frac{a-b}{1-\frac{k+2}{k}ab}=|g(-1)|$$

and $(a-b)\Big/\left(1-\frac{k+2}{k}ab\right)<1$ by (30). These conclusions imply the existence of a domain $\Pi\supset\bar{\Delta}$ with $f(\Pi)=\Delta$, conformally.

4) To complete the proof we have to show that G_f is analytic in $\bar{\Delta}$. This is equivalent to the statement that Q_f has just simple poles in Δ with natural numbers as residues. A straightforward calculation gives

$$Q_f(w)=\frac{k}{w}+\frac{2}{w+\frac{k+2}{k}ab}+\frac{s-2w}{1+tw+w^2}$$

with

$$s=\frac{bk(k+2)}{2a}+\frac{ak(k+4)}{2b}-\frac{k(k+2)}{2ab}-\frac{(k+2)^2}{2}ab,$$

$$t=\frac{k}{2ab}+\frac{(k+2)^2}{2k}ab-\frac{ka}{2b}-\frac{kb}{2a}.$$

Since Q_f is bounded on $\partial\Delta(f(\bar{\Delta})\subset\Delta)$ we deduce that Q_f cannot have a double pole which would necessarily lie on $\partial\Delta$. Therefore

$$\frac{s-2w}{1+tw+w^2}=\frac{r}{w-w_1}-\frac{r+2}{w-w_2},$$

where

$$2w_1(r+1)=s-tr,\quad 2w_2(r+1)=-s-t(r+2),\quad w_1\cdot w_2=1. \tag{32}$$

Eliminating w_1, w_2 from the equations (32) we are left with the equation

$$16(r+1)^2a^2b^2k^2+p(k+r+2)\,p(k-r)=0 \tag{33}$$

with $p(q)$ as in the statement of Lemma 7. By (28) we see that $r=l$ is a solution of (33) and it remains to show: $|w_1|<1$. For $b=1$ the relevant solution a is given by

$$a^2=\frac{k^2(l+1)^2}{(k+1)^2(l+1)^2-1},\quad a>0. \tag{34}$$

For k, l fixed $w_1=w_1(b)$, $w_2=w_2(b)$ are continuous in $b\geqq 1$. Furthermore, a direct calculation using (34) shows

$$w_2(1)=-\left(\frac{(k+1)(l+1)+1}{(k+1)(l+1)-1}\right)^{1/2}<-1.$$

From $w_1(1)\cdot w_2(1)=1$ we get $-1<w_1(1)<0$. Using the continuity of w_1, w_2 and the facts $w_1(b)\neq w_2(b)$, $w_1(b)\,w_2(b)=1$ for $b\geqq 1$ we obtain $-1<w_1(b)<0$, $b\geqq 1$.

Remark 3. For $f\in\mathcal{R}_0$ given as in Lemma 7 we obtain explicit representations for G_f:

$$G_f(w)=w^k\frac{(1-w_1)^2(w-w_1)^l\left(w+\frac{k+2}{k}ab\right)^2}{\left[\left(1+\frac{k+2}{k}ab\right)^2-(a+b)^2\right](1-w_1w)^{l+2}}$$

where $w_1=p(k+l+2)/(4abk(l+1))$, p as in Lemma 7.

Lemma 8. *Let f be as in Lemma 7. Then $F_f \in \mathscr{E}$ and $\Lambda(F_f) = f(\partial\Delta)$.*

Proof. In view of Theorem 2 and Lemma 4, iii) it remains to show $\mu_{F_f}(f(w)) < 1$ for $f(w) \in A(f)$, $w \in \Delta$. Let $z_0 = f(w_0) \in A(f)$, $w_0 \in \Delta$ and $c = \mu_{F_f}(z_0) \geqq 1$ such that $F := F_f/c$ fulfills $\mathsf{M}(F) = 1$. Since F' has a zero of order k at the origin (compare the representation for G_f above) we obtain from Theorems 5, 6

$$(1-|z|^2)|F'(z)| < 1, \quad |z| < \sqrt{\frac{k}{k+2}}.$$

This shows $|z_0| = |f(w_0)| \geqq \sqrt{k/(k+2)}$. Since $|f(w)| < 1$, $w \in \bar{\Delta}$, $f(0) = 0$, we obtain from Schwarz' Lemma $|f(w_0)| < |w_0|$ and thus $|1/w_0| < \sqrt{(k+2)/k}$. But f is univalent in $|w| < b\sqrt{(k+2)/k}$ by Lemma 6 and this contradicts the relation $f(w_0) = f(1/\overline{w_0})$ or $f(w_0) \in A(f)$.

Theorem 7. *Let $F \in \mathscr{E}$ have a critical point of order k at $z = 0$ and a critical point of order l at z_0, $-1 < z_0 < 0$, $k \leqq l$, and $F' \neq 0$ elsewhere. Then there exists $f \in \mathscr{R}_0$ as in Lemma 7 such that $F = \varepsilon F_f$ for a certain $\varepsilon \in \mathbb{C}$, $|\varepsilon| = 1$.*

Proof. 1) With $F \in \mathscr{E}$ we have $F^* \in \mathscr{E}$ where $F^*(z) = \overline{F(\bar{z})}$, $z \in \Delta$, and since under our assumptions F^* has the same critical points as F we obtain $F = \delta F^*$, $|\delta| = 1$, by Theorem 3, b). For the analytic Jordan curve $\gamma \subset \Lambda(F)$ this implies: $z \in \gamma$ if and only if $\bar{z} \in \gamma$. From Theorem 2 and Proposition 3, we have the existence of $f \in \mathscr{R}_0$ of the form (21) with $n = 2$ which maps Δ onto the interior domain Γ of γ and $F = \varepsilon F_f$. Furthermore we may assume $f'(0) > 0$. From the symmetry of γ (and Γ) mentioned above we then deduce the relation $f(w) = \overline{f(\bar{w})}$ in Δ and thus

$$f(w) = w\frac{aw+b}{w+c}, \quad a, b, c \in \mathbb{R}, \quad c \neq 0. \tag{35}$$

2) It follows from our assumptions that F_f' and - because of (35) - G_f has a k-fold zero at the origin. Thus

$$\operatorname*{res}_{w=0} \frac{G_f'}{G_f} = \operatorname*{res}_{w=0} Q_f = k.$$

A direct calculation of Q_f using (1), (35) then gives $c = \dfrac{k+2}{k} ab$ and therefore $a \neq 0$, $b \neq 0$. Now we obtain

$$Q_f(w) = \frac{k}{w} + \frac{2}{w + \dfrac{k+2}{k} ab} + \frac{l}{w-w_1} - \frac{l+2}{w-w_2} \tag{36}$$

where $w_1 = f^{-1}(z_0)$, $|w_1| < 1$, and $w_2 = 1/w_1$. Here we used that F_f has a critical point of order l in z_0. From part 4) of the proof of Lemma 7 we see that Q_f has the form (36) for $f \in \mathscr{R}_0$ like (35) only if (28) is fulfilled.

3) It remains to show that f also necessarily fulfills the other assumptions of Lemma 7. From Corollary 3 we see that Γ properly contains the disc $|w|<\sqrt{k/(k+2)}$ which shows, by the principle of subordination, that

$$|f'(0)|=\frac{k}{k+2}\frac{1}{a}>\sqrt{\frac{k}{k+2}}.$$

Thus $\frac{k+2}{k}a^2<1$ and f is of the form studied in Lemma 6. From that lemma we obtain that f^{-1} has branch points on $|z|=b$. Since $f\in\mathscr{R}_0$ all branch points of f^{-1} are in $\mathbb{C}\setminus\Delta$: $b\geqq 1$. The condition

$$|f(-1)|=\frac{a-b}{1-\frac{k+2}{k}ab}<1$$

implies

$$a>\frac{b+1}{1+\frac{k+2}{k}b}$$

and from part 2) in the proof of Lemma 7 we see that this is true only for the bigger solution of (28). We have shown that under our assumptions f is one of the functions as described in Lemma 7. This completes the proof of Theorem 7.

Corollary 4. *Let $F\in\mathscr{E}$, $\Omega_F=\{z_0,z_1\}$, z_0 of order k, z_1 of order l, $k\leqq l$. Let T be an automorphism of Δ with $T(z_0)=0$, $T(z_1)<0$. Then there exists a function $f\in\mathscr{R}_0$ as in Lemma 7 such that $F=\varepsilon F_f(T^{-1}(z))$ for a certain $\varepsilon\in\mathbb{C}$, $|\varepsilon|=1$.*

Corollary 5. *Let $F\in\mathscr{E}$, $\#\Omega_F=2$. Then $\Lambda(F)=\gamma$.*

These are immediate consequences of the results proved before.

Remark 4. A more detailed study of the relation between the second critical point z_0 of a function F_f as in Theorem 7 and the parameter b in Lemma 7 yields that - for k, l fixed - z_0 cannot approach -1. This shows that not every configuration of critical points can occur in $\mathscr{E}$.

V. Concluding Remarks

Many questions remain open. As it is clear from Chap. 4 the construction of members in $\mathscr{E}$ is a difficult task. Are there members in $\mathscr{E}$ with $\#\Omega_F=n$ for arbitrary n? And are there functions $F\in\mathscr{E}$ with $\Lambda(F)\neq\gamma$? There is some evidence that the answer in both cases is affirmative.

The functions $F\in\mathscr{E}$ with $\#\Omega_F=n$ seem to play a similar role in the space $\{F'|F\in\mathscr{B}\}$ as the Blaschke products with n factors in the theory of bounded analytic functions in Δ. For instance, it is known that the functional $|F''(z)|$ in the unit ball of the Bloch space is maximized by functions $F\in\mathscr{E}$, $\#\Omega_F=1$ (see [7]). Is it perhaps true that functionals like $|(F^{(n+1)}(z)|$ will be maximized

within the functions $F \in \mathscr{E}$ with $\# \Omega_F = n$? The complete determination of all functions with $\# \Omega_F = 2$ given in the present paper may be useful in this context.

References

1. Aharonov, D., Shapiro, H.S.: Domains on which analytic functions satisfy quadrature identities. J. Analyse Math. **30**, 39–73 (1976)
2. Cima, J.A., Wogen, W.R.: Extreme points of the unit ball of the Bloch space B_0. Michigan Math. J. **25**, 213–222 (1978)
3. Davis, P.J.: The Schwarz functions and its applications. Washington: Mathematical Association of America 1974
4. Peschl, E.: Les invariants différentiels non holomorphes et leur rôle dans la théorie des fonctions. Rend. Sem. Mat. Messina **1**, 100–108 (1955)
5. Peschl, E.: Über die Weiterentwicklung einer Majorisierungsmethode der Funktionentheorie, die auf Differentialinvarianten aufbaut. In: C. Carathéodory, International Symposium (Athens 1973), pp. 461–482. Athens: The Greek Mathematical Society 1974
6. Ruscheweyh, St.: Gewisse Klassen verallgemeinerter analytischer Funktionen. Bonner Mathematische Schriften **39**. Bonn: Mathematisches Institut der Universität 1969
7. Wirths, K.-J.: Über holomorphe Funktionen, die einer Wachstumsbeschränkung unterliegen. Arch. Math. (Basel) **30**, 606–612 (1978)
8. Wirths, K.-J.: On holomorphic functions satisfying $|f(z)|(1-|z|^2) \leqq 1$ in the unit disc. Preprint

Received October 26, 1981

Math. Z. 180, 107–117 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Weak Containment and Tensor Products of Group Representations

Eberhard Kaniuth

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

Let G be a locally compact group and $1=1_G$ the one-dimensional identity representation of G. For any two unitary representations π and ρ of G let $\pi\otimes\rho$ denote their tensor product and $\bar{\rho}$ the conjugate representation of ρ. It is well known that if π and ρ are finite-dimensional, then $\pi\otimes\bar{\rho}$ contains 1 if and only if π and ρ have a common subrepresentation. Can this be generalized to the case where π and ρ are infinite-dimensional and containment is replaced by weak containment?

Fell [6] answered this question affirmatively for the case that, say, π is arbitrary but ρ is still finite-dimensional. Concerning the general case he has shown that for $G=SO(3)\ltimes\mathbb{R}^3$, the motion group of $\mathbb{R}^3$, there exist $\pi, \rho\in\hat{G}$, the set of equivalence classes of irreducible unitary representations of G, such that 1 is weakly contained in $\pi\otimes\bar{\rho}$ ($1\prec\pi\otimes\bar{\rho}$) but $\pi\neq\rho$ (recall that $(SO(3)\ltimes\mathbb{R}^3)^\wedge$ is a T_1 space)). The same occurs with many simply connected nilpotent Lie groups.

The main purpose of this paper is to prove the following two results. Suppose first that G is a locally nilpotent discrete group and π and ρ are unitary representations of G. Then $1\prec\pi\otimes\bar{\rho}$ if and only if $\operatorname{supp}\pi\cap\operatorname{supp}\rho\neq\varnothing$ (Theorem 1), where $\operatorname{supp}\pi=\{\tau\in\hat{G};\ \tau\prec\pi\}$. Secondly, let G be a locally compact group containing a central subgroup Z such that G/Z is a group with relatively compact conjugacy classes. If $\pi, \rho\in\hat{G}$, then $1\prec\pi\otimes\bar{\rho}$ if and only if π and ρ are weakly equivalent ($\pi\sim\rho$) (Theorem 3). The proof of the latter result uses the corresponding one for groups with relatively compact inner automorphism groups [10].

1. Preliminaries and Examples

The following notations will be used throughout the paper:

$C^*(G)$ = C^*-algebra of the locally compact group G

$P(G)$ = set of all continuous positive definite functions φ on G such that $\varphi(e)=1$

0025-5874/82/0180/0107/$02.20

π_φ = cyclic representation of G (resp. $C^*(G)$) defined by $\varphi \in P(G)$
I_φ = $\{a \in C^*(G);\ \langle a^*a, \varphi\rangle = 0\}$ for $\varphi \in P(G)$
$\ker \pi$ = $\{a \in C^*(G);\ \pi(a) = 0\}$ for a representation π
Prim $C^*(G)$ = set of primitive ideals of $C^*(G)$ endowed with the hull-kernel topology
$\hat{G}$ = set of equivalence classes of irreducible unitary representations of G with the inverse image of the hull-kernel topology under the mapping $\pi \to \ker \pi$, $\hat{G} \to \text{Prim } C^*(G)$
$U^\pi = {}_GU^\pi$ = the representation obtained by inducing the unitary representation of the subgroup H up to G.
$f|H$ = restriction of a mapping f to H
$\check{f}(x)$ = $f(x)$ for $x \in H$ and $= 0$ for $x \in G \setminus H$ for a function f on H
$f^x(y)$ = $f(xyx^{-1})$ for a mapping f of the normal subgroup N of G, $y \in N$ and $x \in G$
$Z(G)$ = center of G
$[A, B]$ = $\{xyx^{-1}y^{-1};\ x \in A,\ y \in B\}$ for $A, B \subseteq G$
$[G:H]$ = index of the subgroup H in G
$\text{supp}\,\mu$ = support of the measure μ

If S and T are sets of unitary representations of G, then S is *weakly contained* in $T (S \prec T)$ if $\bigcap_{\sigma \in S} \ker \sigma \supseteq \bigcap_{\tau \in T} \ker \tau$. S and T are weakly equivalent $(S \sim T)$ if $S \prec T$ and $T \prec S$. For any unitary representation π of G set $\text{supp}\,\pi = \{\tau \in \hat{G};\ \tau \prec \pi\}$. We are dealing with the following question: When is $1 \prec \pi \otimes \bar{\rho}$ equivalent to $\text{supp}\,\pi \cap \text{supp}\,\rho \neq \varnothing$? Somewhat restricting the problem we are led to formulate the following two tensor product properties which a locally compact group G may have or not:

(TP 1) $1_G \prec \pi \otimes \bar{\pi}$ for all $\pi \in \hat{G}$

(TP 2) $1_G \prec \pi \otimes \bar{\rho} \Rightarrow \rho \sim \pi$ for all $\pi, \rho \in \hat{G}$.

Since forming tensor products preserves weak containment [6], (TP 1) implies that $1 \prec \pi \otimes \bar{\rho}$ for all unitary representations π and ρ of G such that $\text{supp}\,\pi \cap \text{supp}\,\rho \neq \varnothing$. Of course, if G satisfies (TP 1) as well as (TP 2), then Prim $C^*(G)$ must be a T_1 space. Fell [6] observed that (TP 1) can hold for G only when G is amenable. But, conversely, no example seems to be known of an amenable group for which (TP 1) fails. We start with some simple fact concerning (TP 1).

Remarks. a) Let G be a locally compact group and $\mathfrak{H}$ the set of all open, compactly generated subgroups of G. It is well known that $I = \overline{\bigcup_{H \in \mathfrak{H}} I \cap C^*(H)}$ for every closed ideal I in $C^*(G)$ (see [20, Lemma I.2.1]).

Thus, if π is a unitary representation of G and $1_H \prec \pi|H$ for all $H \in \mathfrak{H}$, then $1_G \prec \pi$. Hence (TP1) holds for G provided that it holds for all $H \in \mathfrak{H}$. b) Suppose that G is σ-compact. It was shown in [8] that G is a projective limit of second countable groups (Corollary A.10) and that every cyclic representation of G is on a separable Hilbert space (Lemma 2.2) and hence lives on one of the approximating second countable groups (Corollary A.12). Thus, if all these approximating second countable groups satisfy (TP 1), then so does G.

Lemma 1. *Let G be a locally compact group which is either nilpotent or a compact extension of an abelian group. Then* (TP 1) *holds for G.*

Proof. Notice first that if H is a closed subgroup of the amenable group G and τ a unitary representation of H such that $1_H \prec \tau \otimes \bar{\tau}$, then $1_G \prec U^\tau \otimes \overline{U^\tau}$. Indeed, $1_G \prec U^{1_H}$ as G is amenable [7, Theorem 5.1], $\tau \prec U^\tau | H$ (see [17, Proposition 4.2]), and $\rho \otimes U^{\bar{\tau}} = U^{\rho | H \otimes \bar{\tau}}$ (see [7, proof of Theorem 5.1]), hence

$$1_G \prec U^{\tau \otimes \bar{\tau}} \prec U^{U^\tau | H \otimes \bar{\tau}} = U^\tau \otimes U^{\bar{\tau}} = U^\tau \otimes \overline{U^\tau}.$$

Suppose now that G contains an abelian normal subgroup N such that G/N compact, and let $\pi \in \hat{G}$. Then $\pi | N$ is weakly equivalent to the compact orbit $G(\lambda)$ for some $\lambda \in \hat{N}$. Let H be the stability subgroup of λ. Since G/N is compact, the Mackey theory works (see [17, Main Theorem]). Thus $\pi = U^\tau$ for some $\tau \in \hat{H}$ such that $\tau | N \sim \lambda$. Let $K = \{x \in N;\ \lambda(x) = 1\}$, then $\tau \in \widehat{H/K}$ and $N/K \subseteq Z(H/K)$. By [9, Theorem 2.1] H/K has only finite-dimensional irreducible representations. It follows that $1_H \prec \tau \otimes \bar{\tau}$ and $1_G \prec \pi \otimes \bar{\pi}$.

For the nilpotent case, by the preceding remark, we can assume that G is second countable, and, moreover, that $\pi \in \hat{G}$ is faithful, i.e. $\{x \in G;\ \pi(x) = 1\} = \{e\}$. If G nilpotent, then by [13, Proposition 5] there exist a closed subgroup H of G and a 1-dimensional representation λ of H such that $\pi \sim U^\lambda$. As above we obtain $1_G \prec \pi \otimes \bar{\pi}$.

Examples. As already mentioned in the introduction, Fell [6] has constructed two inequivalent irreducible representations π and ρ of the motion group of $\mathbb{R}^3$ such that $1 \prec \pi \otimes \bar{\rho}$.

Let now G be a simply connected nilpotent Lie group and $\mathfrak{g}$ its Lie algebra. By the Kirillov theory there is a one-to-one correspondence $\pi \to B_\pi$ between $\hat{G}$ and the set of orbits in the dual space $\mathfrak{g}^*$ of $\mathfrak{g}$ under the coadjoint representation of G. Unter this correspondence, $1 \to \{0\}$ and $\bar{\pi} \to -B_\pi$, and $\sigma \prec \pi \otimes \rho$ iff $B_\sigma \subseteq \overline{B_\pi + B_\rho}$ [5, Lemma 2.1]. Hence G satisfies (TP 2) iff $0 \in \overline{B(f) - B(g)}$, $f, g \in \mathfrak{g}^*$, implies $B(f) = B(g)$. In particular, (TP 2) holds for G if all coadjoint orbits in $\mathfrak{g}^*$ are affine subspaces.

Felix has pointed out that the six-dimensional simply connected nilpotent Lie group G considered in [5, (2.3)] does not satisfy (TP 2). Indeed, if

$$f = (1, 0, 0, 0, -1, 0), \qquad g = (1, 0, 0, 0, 0, 0) \in \mathbb{R}^6 = \mathfrak{g}^*,$$

then $B(f) \neq B(g)$, but

$$\left(1, n, 0, \frac{1}{n}, 0, 0\right) \in B(f), \qquad (1, n, 0, 0, 0, 0) \in B(g)$$

(see (5, 2.3)]). The same argument can be used to show that among the simply connected nilpotent Lie groups Γ of dimension $\leqq 5$ (TP 2) fails exactly for Γ_4, $\Gamma_{5,4}$, $\Gamma_{5,5}$ and $\Gamma_{5,6}$ - in the notation of [3]. Notice first that if $\Gamma \neq \Gamma_4$, $\Gamma_{5,4}$, $\Gamma_{5,5}$ and $\Gamma_{5,6}$, the coadjoint orbits are affine. On the other hand, Γ_4 is a factor group of $\Gamma_{5,4}$, $\Gamma_{5,5}$, and $\Gamma_{5,6}$. Thus we are reduced to Γ_4. The coadjoint orbit of

$(\alpha, \beta, \gamma, \delta) \in \mathbb{R}^4 = \mathfrak{g}_4^*$ consists of all

$$\left(a, \beta, r\alpha + s\beta + \gamma, \left(\frac{rs}{2} + t\right)\alpha + \frac{s^2}{2}\beta + s\gamma + \delta\right), \quad r, s, t \in \mathbb{R}.$$

It follows that if $f = (0, 1, 0, 0)$, $g = (0, 1, 0, -1) \in \mathfrak{g}_4^*$, then $B(f) \neq B(g)$, but

$$\left(0, 0, \frac{1}{n}, \frac{1}{2n^2}\right) = \left(0, 1, n, \frac{n^2}{2}\right) - \left(0, 1, n + \frac{1}{n}, \frac{1}{2}\left(n + \frac{1}{n}\right)^2 - 1\right) \in B(f) - B(g).$$

2. Locally Nilpotent Discrete Groups

Let G be a locally compact group and N a closed normal subgroup. We denote by $K(N, G)$ the convex set of all $\varphi \in P(N)$ such that $\varphi^x = \varphi$ for all $x \in G$, and by $E(N, G)$ the set of extreme points of $K(N, G)$. On $P(N) \subseteq L^\infty(N)$, hence on $K(N, G)$, the weak $*$-topology coincides with the topology of uniform convergence on compact subsets of N. We write $K(G)$, $E(G)$ instead of $K(G, G)$, $E(G, G)$ etc.

Suppose now that G is discrete. Then $K(N, G)$ is a simplex, and for $\alpha \in K(N, G)$ there exists a unique normalized boundary measure ν_α on $K(N, G)$ representing α [1, Theorem II. 3.6]. If G is countable, $E(N, G)$ is a G_δ-set in $K(N, G)$ and $\nu_\alpha(E(N, G)) = 1$ [1, § I.4]. In this case we denote by μ_α the measure on $E(N, G)$ induced by ν_α.

For any group G denote by $Z(G) = Z_1(G) \subseteq Z_2(G) \subseteq \ldots$ the ascending central series of G and by G_f the normal subgroup consisting of all elements with finite conjugacy classes. Let now G be a finitely generated nilpotent discrete group. Then $a \to \ker \pi_\alpha = I_\alpha$ is a bijection between $E(G)$ and Prim $C^*(G)$ [15, § 1]. Moreover, the mapping

$$\text{Prim } C^*(G) \to E(G_f, G), \quad I_\alpha \to a | G_f$$

is continuous (compare [11, p. 400]) (notice that Prim $C^*(G)$ carries the hull-kernel topology, $E(G_f, G)$ the topology of pointwise convergence). In view of later applications the following lemma will be proved in greater generality than it is needed here. Our proof is a refinement of arguments used in the proof of [12, Theorem 2].

Lemma 2. *Let G be a finitely generated discrete group containing a nilpotent normal subgroup N of finite index. Suppose that $\alpha_n \in E(G)$, $n \in \mathbb{N}$, such that $\alpha_n | N \in E(N)$ and $\alpha_n \to \alpha$ in $K(G)$. Then there exist normal subgroups K and L in G such that $K \subseteq L$, $L/K = (G/K)_f$, $\alpha_n | K \equiv 1$ for $n \geqq n_0$ and $\alpha = \widetilde{a | L}$.*

Proof. Consider the subgroups

$$K_n = \{x \in G;\ \alpha_n(x) = 1\}, \quad C_n = \{x \in G;\ [x, G] \subseteq K_n\},$$

$$K = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} K_j \quad \text{and} \quad C = \bigcup_{n=1}^{\infty} \bigcap_{j=n}^{\infty} C_j.$$

Since every subgroup of a finitely generated nilpotent group is again finitely generated, the same is true for G. It follows that $K=\bigcap_{j=n_0}^{\infty} K_j$ and $C=\bigcap_{j=n_0}^{\infty} C_j$ for some n_0. As $\alpha_n\in E(G/K)$ for $n\geqq n_0$, we can assume that $K=\bigcap_{j=1}^{\infty} K_j=\{e\}$ and $C=\bigcap_{j=1}^{\infty} C_j$, so that $C=Z(G)$. Let

$$A=\{x\in G;\ xC\in(G/C)_f\} \quad \text{and} \quad B=\{x\in N;\ xZ(N)\in(N/Z(N))_f\}.$$

We assert that $[A:Z_2(N)]<\infty$. Notice first that $[B:Z_2(N)]<\infty$, because in a finitely generated nilpotent group H the center of H has finite index in H_f. On the other hand, since $\alpha_n|N\in E(N)$,

$$\alpha_n([x,y])=\alpha_n(x)\,\alpha_n(yx^{-1}y^{-1})=1$$

for $x\in Z(N)$ and $y\in G$, hence $[x,y]\in\bigcap_{n=1}^{\infty} K_n=\{e\}$. This shows $[Z(G):Z(N)]=[Z(G):Z(G)\cap N]<\infty$, and it is easily seen that this implies $A\cap N=B$. Therefore

$$[A:Z_2(N)]=[A:A\cap N]\,[B:Z_2(N)]<\infty.$$

Now, every $\sigma\in E(N)$ is linear on some subgroup H, $Z_1(N)\subseteq H\subseteq Z_2(N)$, and vanishes on $Z_2(N)\setminus H$. Hence, if $x\in Z_2(N)$ and $|\alpha_n(x)|=1$ for all $n\geqq n_x$, then as above $\alpha_n([x,y])=1$ for all $y\in G$, $n\geqq n_x$, so that $x\in\bigcap_{n=n_x}^{\infty} C_n=Z_1(G)$. Thus $\alpha(x)=\lim_{n\to\infty}\alpha_n(x)=0$ for $x\in Z_2(N)\setminus Z_1(G)$, and [15, Lemma 1] implies that $\alpha=\widetilde{\alpha|A}$.

It remains to show that $\alpha(x)=0$ for $x\in A\setminus G_f$. Since $x\notin G_f$ and $xZ(G)\in(G/Z(G))_f$, there exist $a,y,z\in G$ such that yxy^{-1}, $zxz^{-1}\in aZ(G)$ and $yxy^{-1}\neq zxz^{-1}$. Then $[y^{-1}z,x]\in Z(G)$ and

$$\alpha_n(x)=\alpha_n([y^{-1}z,x]\,x)=\alpha_n([y^{-1}z,x])\,\alpha_n(x).$$

If now $[y^{-1}z,x]\in K_n$ for all $n\geqq m$, then $y^{-1}xy=z^{-1}xz$, a contradiction. Hence $[y^{-1}z,x]\notin K_{n_j}$ and $\alpha_{n_j}([y^{-1}z,x])\neq 1$ for some infinite sequence n_j, so that $\alpha_{n_j}(x)=0$ and $\alpha(x)=\lim_{j\to\infty}\alpha_{n_j}(x)=0$.

Lemma 3. *Let G be a finitely generated nilpotent discrete group, and let π and ρ be unitary representations of G. If $1\prec\pi\otimes\bar{\rho}$, then* $\operatorname{supp}\pi\cap\operatorname{supp}\rho\neq\varnothing$.

Proof. Since G is second countable, there exist

$$\pi_n\in\operatorname{supp}\pi,\ \rho_n\in\operatorname{supp}\rho \text{ and } \omega_n\in\operatorname{supp}\pi_n\otimes\overline{\rho_n},\qquad n\in\mathbb{N},$$

such that $\omega_n\to 1$. Let $\pi_n\sim\pi_{\alpha_n}$ and $\rho_n\sim\pi_{\beta_n}$, α_n, $\beta_n\in E(G)$. Of course, we can assume that $\alpha_n\to\alpha$ and $\beta_n\to\beta$ in $K(G)$. By Lemma 2 there are normal

subgroups K, L, A and B in G such that

$$L/K=(G/K)_f, \quad \alpha_n|K\equiv 1, \quad \alpha=\widetilde{\alpha|L}$$

and

$$B/A=(G/A)_f, \quad \beta_n|A\equiv 1, \quad \beta=\widetilde{\beta|B} \quad \text{for } n\geqq n_0.$$

Thus, we can moreover assume that $K\cap A=\{e\}$. We are going to prove $\alpha|L\cap B=\beta|L\cap B$. Let M and C be the subgroups of G defined by $K\subseteq M$, $M/K=Z(G/K)$ and $A\subseteq C$, $C/A=Z(G/A)$. Then $M\cap C\subseteq Z(G)$ and

$$[L\cap B: M\cap C]\leqq[L:M][B:C]<\infty.$$

Now, in a finitely generated nilpotent group H, any subgroup F with $[F:F\cap Z(H)]<\infty$ is contained in H_f. Hence $L\cap B\subseteq G_f$, and it suffices to show $\alpha|G_f=\beta|G_f$. To that end let $\omega_n\sim\pi_{\chi_n}$, $\chi_n\in E(G)$, and notice first that $\omega_n\to 1$ implies $\chi_n|G_f\to 1$ in $E(G_f,G)$, since the mapping $I_\omega\to\omega|G_f$, $\operatorname{Prim}C^*(G)\to E(G_f,G)$ is continuous. On the other hand, $\pi_\gamma\otimes\pi_\delta\sim\pi_{\gamma\delta}$ for $\gamma,\delta\in E(G_f,G)$. This can be established as follows. If K is a compact group, then the orthogonality relations for the characters of K show that $\pi_\sigma\otimes\pi_\tau\sim\pi_{\sigma\tau}$ for $\sigma,\tau\in E(K)$, and by applying a simple Bohr compactification argument this follows for arbitrary K provided that π_σ and π_τ are finite dimensional. Since, $[G_f:Z(G)]<\infty$, we thus have $\pi_\sigma\otimes\pi_\tau\sim\pi_{\sigma\tau}$ for all $\sigma,\tau\in E(G_f)$. Finally, every π_γ, $\gamma\in E(G_f,G)$, is weakly equivalent to the finite orbit $\{\pi_{\sigma^x};\ x\in G\}$ for some $\sigma\in E(G_f)$. Hence we conclude $\pi_\gamma\otimes\pi_\delta\sim\pi_{\gamma\delta}$ for $\gamma,\delta\in E(G_f,G)$. Recall next that $\pi_\eta|G_f\sim\pi_{\eta|G_f}$ for $\eta\in K(G)$ [11, p. 396]. Therefore we have

$$\pi_{\chi_n|G_f}\prec\pi_{\alpha_n|G_f}\otimes\pi_{\beta_n|G_f}\sim\pi_{\alpha_n|G_f\cdot\beta_n|G_f},$$

hence $\chi_n|G_f\in\operatorname{supp}\mu_{\alpha_n|G_f\cdot\beta_n|G_f}$. But $\alpha_n|G_f\to\alpha|G_f$, $\beta_n|G_f\to\beta|G_f$, and $\chi_n|G_f\to 1$, and the mapping $(\gamma,\delta)\to\operatorname{supp}\mu_{\gamma\delta}$ from $E(G_f,G)\times E(G_f,G)$ into the space Γ of all closed subsets of $E(G_f,G)$ is continuous [10, Lemma 3], where Γ carries the so-called "finite" topology. Thus $1\in\operatorname{supp}\mu_{\alpha\beta|G_f}$, and this implies $\alpha|G_f=\beta|G_f$. So far we have shown that $\alpha|L\cap B=\beta|L\cap B$.

Let $M\subseteq N$ be normal subgroups of G, and denote by $r_{N,M}$ the restriction map $E(N,G)\to E(M,G)$. For $\gamma\in K(N,G)$ $\mu_{\gamma|M}$ is the image of μ_γ under $r_{N,M}$, so that $\operatorname{supp}\mu_{\gamma|M}=r_{N,M}(\operatorname{supp}\mu_\gamma)$. Moreover, it follows from the nilpotence of G that if $\delta\in K(M,G)$, then its trivial extension $\tilde{\delta}$ to N satisfies $\operatorname{supp}\mu_{\tilde{\delta}}=r_{N,M}^{-1}(\operatorname{supp}\mu_\delta)$. To verify this it is evidently enough to consider the case $N/M\subseteq Z(G/M)$. But then $r_{N,M}^{-1}(\omega|M)=\{\omega\lambda;\ \lambda\in\widehat{N/M}\}$ for $\omega\in E(N,G)$, hence

$$\begin{aligned} r_{N,M}^{-1}(\operatorname{supp}\mu_\delta)&=\bigcup\{\lambda\cdot\operatorname{supp}\mu_{\tilde{\delta}};\ \lambda\in\widehat{N/M}\}\\ &=\bigcup\{\operatorname{supp}\mu_{\lambda\tilde{\delta}};\ \lambda\in\widehat{N/M}\}=\operatorname{supp}\mu_{\tilde{\delta}}.\end{aligned}$$

Now we can finish the proof of Lemma 3. Since $\alpha=\widetilde{\alpha|L}$, $\beta=\widetilde{\beta|B}$ and $\alpha|L\cap B=\beta|L\cap B$, we conclude that

$$\begin{aligned}\operatorname{supp}\mu_{\alpha|L}&\subseteq r_{L,L\cap B}^{-1}(\operatorname{supp}\mu_{\alpha|L\cap B})=r_{L,L\cap B}^{-1}(\operatorname{supp}\mu_{\beta|L\cap B})\\ &=\operatorname{supp}\mu_{B|L} \quad\text{and}\quad \operatorname{supp}\mu_\alpha=r_{G,L}^{-1}(\operatorname{supp}\mu_{\alpha|L}),\end{aligned}$$

and these relations imply $\operatorname{supp}\mu_\alpha \cap \operatorname{supp}\mu_\beta \neq \varnothing$. But $\omega \to \langle f^* * f, \omega\rangle$ is continuous on $E(G)$ and

$$\langle f^* * f, \gamma\rangle = \int_{E(G)} \langle f^* * f, \omega\rangle \, d\mu_\gamma(\omega) \quad \text{for } \gamma \in K(G),$$

so that $I_\gamma \subseteq I_\omega$ for all $\omega \in \operatorname{supp}\mu_\gamma$. It follows that $\operatorname{supp}\pi_\alpha \cap \operatorname{supp}\pi_\beta \neq \varnothing$. On the other hand $\pi_{\alpha_n} \in \operatorname{supp}\pi$ and $\alpha_n \to \alpha$ yield $I_\alpha \supseteq \bigcap_{n=1}^{\infty} I_{\alpha_n} \supseteq \ker\pi$, in the same way $I_\beta \supseteq \ker\rho$. Thus we have shown $\ker\pi \cap \ker\rho \neq \varnothing$.

Remark. Consider a discrete group G and unitary representations π and ρ of G. Suppose that $\operatorname{supp}\pi|H \cap \operatorname{supp}\rho|H \neq \varnothing$ for all $H \in \mathfrak{H}$, the set of all finitely generated subgroups of G. Then also $\operatorname{supp}\pi \cap \operatorname{supp}\rho \neq \varnothing$. For otherwise $\ker\pi + \ker\rho = C^*(G)$, and by [20, Lemma I.2.1]

$$\ker\pi = \overline{\bigcup_{H\in\mathfrak{H}} \ker\pi \cap C^*(H)} \quad \text{and} \quad \ker\rho = \overline{\bigcup_{H\in\mathfrak{H}} \ker\rho \cap C^*(H)}.$$

Hence there exist $u \in \ker\pi|H$ and $v \in \ker\rho|H$ such that $\|\delta_e - (u+v)\| < 1$, and this gives the contradiction $\ker\pi|H + \ker\rho|H = C^*(H)$.

Combining this remark with Lemma 3 and the fact that (TP 1) holds for locally nilpotent groups, we obtain

Theorem 1. *Let G be a locally nilpotent discrete group, and let π and ρ be unitary representations of G. Then $1 \prec \pi \otimes \bar{\rho}$ if and only if* $\operatorname{supp}\pi \cap \operatorname{supp}\rho \neq \varnothing$.

Somewhat deviating from the main subject of this paper we show now how Lemma 2 can be used to describe the hull-kernel topology on Prim $C^*(G)$ for finitely generated nilpotent discrete groups G. Recall that if K is a simplex and $\partial_e K$ its set of extreme points, then

$$\{F \cap \partial_e K;\ F \text{ a closed face in } K\}$$

is the collection of closed sets for a topology on $\partial_e K$ [1, § II.6], which is called the facial topology and has been introduced by Effros.

Theorem 2. *Let G be a finitely generated nilpotent discrete group. Then the bijection $\alpha \to I_\alpha = \ker\pi_\alpha$ between $E(G)$ and* Prim $C^*(G)$ *is a homeomorphism provided that $E(G) = \partial_e K(G)$ is endowed with the facial topology.*

Proof. Let φ denote the mapping $\alpha \to \ker\pi_\alpha$, and let $A \subseteq \operatorname{Prim} C^*(G)$. First we have to show that if A is closed, then $\varphi^{-1}(A) = F \cap E(G)$ for some closed face F in $K(G)$. Set $F = \overline{\operatorname{co}\varphi^{-1}(A)}$, the closed convex hull of $\varphi^{-1}(A)$. Then $F \cap E(G) = \overline{\varphi^{-1}(A)} \cap E(G) = E(G)$ [1, p. 50, (5.22)], since $\varphi^{-1}(A)$ is clearly closed in $E(G)$. By [4, Theorem 3.3] F is a face provided that $\varphi^{-1}(A)$ is dilated. To see this, let $\gamma = \lim_{n\to\infty} \alpha_n \in K(G)$, $\alpha_n \in \varphi^{-1}(A)$, and $\alpha \in \operatorname{supp} \nu_\gamma \cap E(G)$. Since $\beta \to \langle f^* * f, \beta\rangle$ is continuous on $K(G)$ and $\langle f^* * f, \gamma\rangle = \int_{K(G)} \langle f^* * f, \beta\rangle \, d\nu_\gamma(\beta)$, it follows that $\bigcap_{n=1}^{\infty} I_{\alpha_n} \subseteq I_\gamma \subseteq I_\alpha$, i.e. $\alpha \in \varphi^{-1}(\bar{A}) = \varphi^{-1}(A)$. Thus $\operatorname{supp}\nu_\gamma = \overline{\operatorname{supp}\nu_\gamma \cap E(G)} \subseteq \overline{\varphi^{-1}(A)}$.

Conversely, suppose that $\varphi^{-1}(A)=F\cap E(G)$ for some closed face F in $K(G)$. Let $\alpha_n\in\varphi^{-1}(A)$, $n\in\mathbb{N}$, and $\beta\in E(G)$ such that $I_{\alpha_n}\to I_\beta$. We can assume that $\alpha_n\to\alpha$ in $K(G)$. By Lemma 2 there are normal subgroups K and L in G such that $L/K=(G/K)_f$, $\alpha_n|K\equiv 1$ for $n\geqq n_0$ and $\alpha=\widetilde{\alpha|L}$. Hence we can moreover assume that $K=\{e\}$. Then $\alpha=\widetilde{\alpha|G_f}$, and since G is nilpotent, $\operatorname{supp}\mu_\alpha=\{\omega\in E(G);\ \omega|G_f=\alpha|G_f\}$ (compare the proof of Lemma 3). $I_{\alpha_n}\to I_\beta$ implies that $\alpha_n|G_f\to\beta|G_f$, and it follows that $\beta\in\operatorname{supp}\mu_\alpha$. On the other hand, $\alpha\in F$ and $\operatorname{supp}\mu_\alpha\subseteq F$, since F is a closed face in $K(G)$. Hence $\beta\in F\cap E(G)=\varphi^{-1}(A)$, i.e. $\ker\pi_\beta\in A$.

3. Groups with Compactness Conditions

Let G be an $[FIA]^-$ group, i.e. a locally compact group whose inner automorphism group has compact closure in the full automorphism group of G. Hartmann, Henrichs, and Lasser [10] have recently shown that $E(G)$ is a hypergroup, the crucial point being $1\prec\pi_{\alpha\bar{\beta}}$ iff $\pi_\beta\sim\pi_\alpha$ for $\alpha,\beta\in E(G)$. Now, $\alpha\to\ker\pi_\alpha=I_\alpha$ is a bijection (in fact, a homeomorphism) between $E(G)$ and $\operatorname{Prim}C^*(G)$. Moreover, $\pi_{\alpha\beta}$ is a subrepresentation of $\pi_\alpha\otimes\pi_\beta$, and $\bar{\pi}_\beta=\pi_{\bar{\beta}}$. Hence (TP 1) holds for G. On the other hand, $[FIA]^-$ groups have the Frobenius reciprocity property (FP) [11, § 3]. From (FP) and $1\prec\pi_{\alpha\bar{\alpha}}$ one can deduce that $1\prec\pi_\alpha\otimes\pi_\beta$ implies $\pi_\beta\sim\pi_\alpha$ [10, Remark p. 237]. Thus G also has property (TP2).

Suppose now that G is an $[FC]^-$ group, i.e. a group with relatively compact conjugacy classes. By [18] G contains a compact normal subgroup K such that $G/K=\mathbb{R}^n\times D$, where D is a discrete $[FC]^-$ group. In particular, G/K is $[FIA]^-$. Therefore the first attempt to prove (TP 1) and (TP 2) for $[FC]^-$ groups, could be to try to show that extensions of compact groups by groups with property (TP 1) resp. (TP 2) also satisfy (TP 1) resp. (TP 2). But at least for (TP 2) this is not true (compare the remark following Theorem 3).

To handle $[FC]^-$ groups we shall nevertheless use the result of [10]. We start with a rather technical lemma.

Lemma 4. *Let G be a locally compact group, and suppose that $K\subseteq N$ are normal subgroups of G such that K is compact and N/K is second countable and contained in $Z(G/K)$. For each $\pi\in\hat{G}$, $\pi|N$ is weakly equivalent to the G-orbit $G(\sigma)$ for some $\sigma\in\hat{N}$. If $\pi,\rho\in\hat{G}$ with $\pi|K\sim\rho|K$, then $\rho|N\sim\pi|N\otimes\lambda$ for some $\lambda\in\widehat{N/K}$.*

Proof. Choose $\Delta\subseteq\hat{N}$ such that $\hat{K}$ is the disjoint union of the sets $\operatorname{supp}\delta|K$, $\delta\in\Delta$. By [14, Theorem 2] $\operatorname{Prim}C^*(N)$ is the disjoint union of the open sets $\ker\delta\otimes\widehat{N/K}=\{\ker\delta\otimes\lambda;\ \lambda\in\widehat{N/K}\}$. Moreover, the mappings

$$\widehat{N/K}\to\ker\delta\otimes\widehat{N/K},\ \lambda\to\ker\delta\otimes\lambda$$

are continuous and open [14, Proposition], hence every $\ker\delta\otimes\widehat{N/K}$ is second countable.

Let E denote the projection valued measure on Prim $C^*(N)$ associated with $\pi|N$. Since E is regular, $E(S)>0$ for some $S=\ker\delta\otimes\widehat{N/K}$. If F is closed in Prim $C^*(N)$, then $E(F)=I$ iff F contains the hull $h(\ker\pi|N)$. Let $\mathfrak{U}$ be a countable basis of $S\cap h(\ker\pi|N)$ and $U^G=\bigcup_{x\in G}U^x$ for $U\in\mathfrak{U}$. Then $E(U)>0$, hence $E(U^G)=I$ for all $U\in\mathfrak{U}$, and $E(\bigcap_{U\in\mathfrak{U}}U^G)=I$. Now, for $P\in\bigcap_{U\in\mathfrak{U}}U^G$, the G-orbit $G(P)$ is dense in $h(\ker\pi|N)$ as is easily seen. This shows the existence of σ.

To prove the second assertion, notice that $\sigma|K\sim\tau|K$ for some $\tau\in\operatorname{supp}\rho|N$. Thus $\tau\sim\sigma\otimes\lambda$ for some $\lambda\in\widehat{N/K}$, hence $G(\tau)\sim G(\sigma)\otimes\lambda$ since $N/K\subseteq Z(G/K)$. Therefore $\pi|N\otimes\lambda\prec\rho|N$, and by the same argument, $\rho|N\otimes\mu\prec\pi|N$ for some $\mu\in\widehat{N/K}$. It follows that $\pi|N\otimes\lambda\sim\rho|N$.

Lemma 5. *Let G be an $[FC]^-$ group and $\pi,\rho\in\hat{G}$. Then $1\prec\pi\otimes\bar{\rho}$ iff $\rho\sim\pi$.*

Proof. There exists a compact normal subgroup C in G such that $G/C=\mathbb{R}^n\times D$, where D is a discrete $[FC]^-$ group. We show first that if $1\prec\pi\otimes\bar{\rho}$, then $\rho|N\sim\pi|N$ for all open, compactly generated normal subgroups N of G with $C\subseteq N$. Since every open, compactly generated subgroup of G is contained in some such N, it will follow that $\rho\sim\pi$.

We have $N/C=\mathbb{R}^n\times F$, where F is a finitely generated normal subgroup of D. It follows from [19, Theorem 4.32] that $[F,D]$ is finite, and hence $K=C[N,G]$ is a compact normal subgroup of G with N/K second countable and $N/K\subseteq Z(G/K)$. As $\rho|K\sim\pi|K$, by Lemma 4 there are $\sigma\in\hat{N}$ and $\lambda\in\widehat{N/K}$ such that $\pi|N\sim\{\sigma^x;\ x\in G\}$ and $\rho|N\sim\pi|N\otimes\lambda$. Fix $\omega\in\operatorname{supp}\sigma|K$, and let H be the open stability subgroup of ω in N. H/K is second countable and $H/K\subseteq Z(G/K)$. As K is compact, by Mackey's theory there is a $\tau\in\hat{H}$ such that $\sigma=U^\tau$ and $\tau|K\sim\omega$. We show that $\lambda|H\prec\tau\otimes\bar{\tau}$.

$$\pi|N\sim\{U^{\tau^x};\ x\in G\}\Rightarrow 1_N\prec\{U^{\tau^x}\otimes U^{\bar{\tau}^y}\otimes\bar{\lambda};\ x,y\in G\}$$
$$\Rightarrow 1_H\prec\{\tau^x\otimes\bar{\tau}^y\otimes\bar{\lambda}|H;\ x,y\in G\}\quad[17,\text{ Proposition 4.2}].$$

But $\operatorname{supp}\tau^x\otimes\bar{\tau}^y\cap\widehat{H/K}\neq\varnothing$ for some $x,y\in G\Rightarrow\tau^x|K\sim\tau^y|K$. By the definition of H, this implies $\tau^x=\tau^y$, and it follows that

$$1_H\prec(\tau\otimes\bar{\tau})^x\otimes\bar{\lambda}|H\quad\text{or}\quad 1_H\prec\tau\otimes\overline{\tau\otimes\lambda|H}.$$

We assert $\tau\otimes\lambda|H\sim\tau$. Of course, we can assume that τ is faithful on K. Then H is second countable, K has small H-invariant neighbourhoods of the identity, and $H/K=\mathbb{R}^n\times E$, where E is discrete abelian. Let now H_0 denote the abstract group of H endowed with the topology obtained by making K open, and for any unitary representation χ of H, χ viewed as a representation of H_0 will be denoted by χ_0. By a deep result of Liukkonen and Mosak [16, p. 284, (**)] the mapping

$$\ker\tau\otimes\widehat{H/K}\to\ker\tau_0\otimes\widehat{H_0/K},\qquad\ker\tau\otimes\mu\to\ker\tau_0\otimes\mu_0$$

is injective. Thus, if $\ker\tau\otimes\lambda|H\neq\ker\tau$, then $\ker\tau_0\otimes(\lambda|H)_0\neq\ker\tau_0$. But H_0 is an $[FIA]^-$ group, and $1_H\prec\tau\otimes\overline{\tau\otimes\lambda|H}$ implies $1_{H_0}\prec\tau_0\otimes\overline{(\tau\otimes\lambda|H)_0}$. This contradicts the fact that (TP 2) holds for $[FIA]^-$ groups.

Finally, $\tau\otimes\lambda|H\sim\tau$ yields

$$\rho|N\sim\pi|N\otimes\lambda\sim\{U^{\tau^x}\otimes\lambda;\ x\in G\}$$
$$=\{U^{\tau^x\otimes\lambda|H};\ x\in G\}=\{(U^{\tau\otimes\lambda|H})^x;\ x\in G\}\sim\{(U^\tau)^x;\ x\in G\}\sim\pi|N.$$

To prove the other direction of Lemma 5, i.e. $1\prec\pi\otimes\bar{\pi}$ for $\pi\in\hat{G}$, we can assume that G is compactly generated. Thus there exists a compact normal subgroup K in G such that G/K is abelian. Since $1_K\prec\pi\otimes\bar{\pi}|K$, there is a $\lambda\in\widehat{G/K}\cap\operatorname{supp}\pi\otimes\bar{\pi}$. But we have already shown that G satisfies (TP 2), hence $1\prec\pi\otimes\overline{\pi\otimes\lambda}$ implies $\pi\sim\pi\otimes\lambda$, so that $1\prec\pi\otimes\bar{\pi}$.

We say that the topological group G is a *central extension* (of N) by H if G contains a closed central subgroup Z (isomorphic to N) such that G/Z is isomorphic to H.

Lemma 6. *Let H be a locally compact group, and suppose that* (TP 1) *resp.* (TP 2) *is satisfied by all central extensions of the circle group $\mathbb{T}$ by H. Then* (TP 1) *resp.* (TP 2) *holds for all central extensions by H.*

Proof. Let G be a locally compact group containing a central subgroup Z such that $G/Z=H$, and let $\lambda\in\hat{Z}$. By the unitary representation theory of group extensions, there exist a multiplier ω on H and an irreducible $\bar{\omega}$-representation ρ of G with $\rho|Z=\lambda$. Moreover, the mapping $\chi\to\rho\otimes\chi$ is a bijection between the set $(\hat{H},\omega)$ of all irreducible ω-representations of H and the set of all irreducible representations of G which restrict to a multiple of λ (see [2, Theorem 2.3] for these facts).

Let K denote the central extension of $\mathbb{T}$ by H defined by ω, i.e. $K=H\times\mathbb{T}$ with the product topology and multiplication $(x,s)(y,t)=(xy,st\,\omega(x,y))$. Any unitary ω-representation χ of H defines a unitary representation χ^* of K by $\chi^*(x,t)=t\chi(x)$. $\chi\to\chi^*$ gives a bijection between $(\hat{H},\omega)$ and $\hat{K}$. If $\pi_i=\rho\otimes\chi_i$, $\chi_i\in(\hat{H},\omega)$, $i=1,2$, then

$$\pi_1\otimes\bar{\pi}_2=(\rho\otimes\bar{\rho})\otimes(\chi_1\otimes\bar{\chi}_2)=\chi_1\otimes\bar{\chi}_2,$$

since ρ is 1-dimensional. It follows that

$$1_G\prec\pi_1\otimes\bar{\pi}_2\quad\text{iff}\quad 1_H\prec\chi_1\otimes\bar{\chi}_2.$$

The compactness of $\mathbb{T}$, the definition of χ^* and the fact that $\overline{\chi^*}(x,t)=\bar{t}\bar{\chi}(x)$ immediately imply

$$1_H\prec\chi_1\otimes\overline{\chi_2}\quad\text{iff}\quad 1_K\prec\chi_1\otimes\overline{\chi_2}=\chi_1^*\otimes\overline{\chi_2^*}.$$

Suppose now that (TP 2) holds for K and that $1_G\prec\pi_1\otimes\bar{\pi}_2$, $\pi_1,\pi_2\in\hat{G}$. Then of course π_1 and π_2 restrict to multiples of λ for some $\lambda\in\hat{Z}$, hence $\pi_i=\rho\otimes\chi_i$, $\chi_i\in(\hat{H},\omega)$, $i=1,2$. It follows that $\chi_1^*\sim\chi_2^*$, and from that we easily deduce that $\pi_1\sim\pi_2$ by using the equation

$$\langle\pi_i(x)(1\otimes\xi),\ 1\otimes\eta\rangle=\rho(x)\langle\chi_i(xZ)\,\xi,\eta\rangle$$
$$=\langle\chi_i^*(xZ,\rho(x))\,\xi,\eta\rangle$$

for $x \in G$ and ξ, η in the Hilbert space of χ_i. In the same way it is shown that (TP 1) for K implies (TP 1) for G.

Since extensions of compact groups by $[FC]^-$ groups are again $[FC]^-$, we obtain from Lemma 5 and Lemma 6.

Theorem 3. *Let G be a central extension by a group with relatively compact conjugacy classes, and let $\pi, \rho \in \hat{G}$. Then $1 \prec \pi \otimes \bar{\rho}$ iff $\rho \sim \pi$.*

Remark. Lemma 6 and the examples in Part 1 show that (TP 2) does not even hold for central extensions of $\mathbb{T}$ by simply connected nilpotent Lie groups of class 2.

References

1. Alfsen, E.M.: Compact convex sets and boundary integrals. Berlin-Heidelberg-New York: Springer 1971
2. Baggett, L., Kleppner, A.: Multiplier representations of abelian groups. J. Functional Analysis **14**, 299–324 (1973)
3. Dixmier, J.: Sur les représentations unitaires des groupes de Lie nilpotents III. Canad. J. Math. **10**, 321–348 (1958)
4. Effros, E.G.: Structure in simplexes II. J. Functional Analysis **1**, 379–391 (1967)
5. Felix, R.: Über Integralzerlegungen von Darstellungen nilpotenter Liegruppen. Manuscripta Math. **27**, 279–290 (1979)
6. Fell, J.M.G.: Weak containment and Kronecker products of group representations. Pacific J. Math. **13**, 503–510 (1963)
7. Greenleaf, F.P.: Amenable actions of locally compact groups. J. Functional Analysis **4**, 295–315 (1969)
8. Greenleaf, F.P., Moskowitz, M.: Cyclic vectors for representations associated with positive definite measures: Non separable groups. Pacific J. Math. **45**, 165–186 (1973)
9. Grosser, S., Moskowitz, M.: Representation theory of central topological groups. Trans. Amer. Math. Soc. **129**, 361–390 (1967)
10. Hartmann, K., Henrichs, R.W., Lasser, R.: Duals of orbit spaces in groups with relatively compact inner automorphism groups are hypergroups. Monatsh. Math. **88**, 229–238 (1979)
11. Henrichs, R.W.: Weak Frobenius reciprocity and compactness conditions in topological groups. Pacific J. Math. **82**, 387–406 (1979)
12. Howe, R.: On representations of discrete, finitely generated, torsion-free nilpotent groups. Pacific J. Math. **73**, 281–305 (1977)
13. Howe, R.: The Fourier transform for nilpotent locally compact groups. Pacific J. Math. **73**, 307–327 (1977)
14. Kaniuth, E.: Primitive ideal spaces of groups with relatively compact conjugacy classes. Arch. Math. (Basel) **32**, 16–24 (1979)
15. Kaniuth, E.: Ideals in group algebras of finitely generated FC-nilpotent discrete groups. Math. Ann. **248**, 97–108 (1980)
16. Liukkonen, J., Mosak, R.: The primitive dual space of $[FC]^-$ groups. J. Functional Analysis **15**, 279–296 (1974)
17. Rieffel, M.A.: Unitary representations of group extensions; an algebraic approach to the theory of Mackey and Blattner. Advances in Math., Suppl. Series **4**, 43–82 (1979)
18. Robertson, L.: A note on the structure of Moore groups. Bull. Amer. Math. Soc. **75**, 594–599 (1969)
19. Robinson, D.: Finiteness conditions and generalized soluble groups I. Berlin-Heidelberg-New York: Springer 1972
20. Strătilă, S., Voiculescu, D.: Representations of AF-algebras and of the group $U(\infty)$. Lecture Notes in Mathematics **486**. Berlin-Heidelberg-New York: Springer 1975.

Received November 17, 1980; received in final form January 8, 1982

Math. Z. 180, 119–140 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Ergodic Properties of Invariant Measures for Piecewise Monotonic Transformations

Franz Hofbauer[1,*] and Gerhard Keller[2,**]

[1] Institut für Mathematik, Universität Wien, Strudlhofgasse 4, A-1090 Wien, Austria
[2] Institut für Mathematische Statistik der Universität, Lotzestraße 13, D-3400 Göttingen, Federal Republic of Germany

Introduction

During the last decade a lot of research has been done on onedimensional dynamics. Different kinds of invariant measures for certain classes of piecewise monotonic transformations have been considered. In most of these cases the Perron-Frobenius-operator plays an important role. In this paper we try to unify these different examples, which are discussed in detail below. First we give a discription of the results proved in this paper.

The general setting we propose is the following.

X is a totally ordered, order-complete set and $T: X \to X$ a transformation which is piecewise monotonic and order-continuous. P is a Perron-Frobenius-operator on the space of all bounded measurable functions on X given by

$$Pf(x) = \sum_{y \in T^{-1}x} g(y) f(y),$$

where g is a function of bounded variation on X with $0 < g(x) \leqq d < 1$. Furthermore we assume the existence of a Borel probability measure m on X, which is invariant for P, i.e. $m(Pf) = m(f)$ for all bounded measurable $f: X \to \mathbb{R}$.

Under these assumptions we give a kind of spectral representation for P (Theorem 1) which is fundamental for the subsequent results:

(1) There is a "greatest" T-invariant m-absolutely continuous probability μ (also denoted by $\mu \ll m$), $\mu = hm$, where h is of bounded variation. "Greatest" means that all other T-invariant, m-absolutely continuous probabilities are absolutely continuous with respect to μ (Theorem 1).

(2) The σ-algebra of T-invariant sets is μ-a.s. finite (Theorem 2). This gives the possibility to split up μ into its finitely many ergodic components. Furthermore, there is a k such that the finitely many ergodic components $W_i \subset X$ of

* Research for this paper was done, when the first author visited the Institut für mathematische Statistik, Göttingen

** The second author was supported by the Deutsche Forschungsgemeinschaft

0025-5874/82/0180/0119/$04.40

(X, μ, T^k) are already weak mixing. We refer to $(W_i, \mu|W_i, T^k|W_i)$ as a weak mixing component of μ (Remark 2).

(3) The asymptotic σ-algebra $\mathfrak{B}_\infty = \bigcap_{j=0}^{\infty} T^{-j}\mathfrak{B}$ of T, where $\mathfrak{B}$ denotes the Borel sets of X, is μ-a.s. finite. It is trivial for a weak mixing component of μ and hence such a component is exact (Theorem 3).

(4) Each finite partition into intervals is a weak Bernoulli partition for every weak mixing component W_i of μ. The corresponding mixing-coefficient decreases exponentially (Theorem 4 and its Corollary). For an $f: W_i \to \mathbb{R}$ of bounded p-variation the stationary stochastic process $\xi_n = f \circ T^{kn}$ $(n \geqq 0)$ satisfies a central limit theorem and an a.s. invariance principle with respect to μ (Theorem 5).

The examples of invariant measures, which fit in our situation, are the following:

(i) Lasota and Yorke [11] consider piecewise monotonic and C^2 transformations T on $[0, 1]$ such that $\inf|T'| > 1$. If one sets $g = |1/T'|$, then the operator P satisfies $\lambda(Pf) = \lambda(f)$ for the Lebesguemeasure λ. In [11] the existence of a function h with $Ph = h$ is proved, which implies the existence of an invariant $\mu \ll \lambda$. In [13] μ is split up into its finitely many ergodic components. Bowen [3] shows that μ is weak Bernoulli, if it is weak mixing. Wong [23] generalizes the results of [11] and [13] to the class of T's, which are piecewise monotonic and C^1, such that T' is of bounded variation and satisfies $\inf|T'| > 1$. Wagner [20] shows that the asymptotic σ-algebra $\mathfrak{B}_\infty$ is μ-a.s. finite. For the class of T considered in [11] the spectral decomposition of P can be found in [10] (cf. also [14]), from which a central limit theorem for stationary processes $f \circ T^n$ on $([0, 1], \mu)$ is deduced, if μ is weak mixing (a special case of this is shown in [24]).

As $g = |1/T'|$, where T is as in [23], satisfies our requirements, all these results are contained in this paper. Furthermore, we show for a weak mixing component of μ, that the mixing coefficients in the weak Bernoulli property decrease at an exponential rate and that the process $f \circ T$ satisfies for certain f a central limit theorem and an almost sure invariance principle.

(ii) Next we consider any piecewise monotonic transformation T on $[0, 1]$ with $h_{\text{top}}(T) > 0$. In [6] and [7] it is shown that every maximal measure can be split up into finitely many ergodic components, which are isomorphic to Markov measures on a finite type subshift with finite or countable alphabet.

If one sets $g(x) = \exp(-h_{\text{top}}(T)) < 1$ for all $x \in [0, 1]$, it is shown in §5 of Chap. III that there is an m with $m(Pf) = m(f)$ and that μ is then a measure with maximal entropy for T. Hence the results of this paper give a decomposition of the maximal measure into its finitely many weak mixing components, each of which is weak Bernoulli and the process $f \circ T^n$ on such a component satisfies for certain f a central limit theorem and an almost sure invariance principle.

(iii) Now let T be the transformation $x \to \beta x (\text{mod } 1)$ on $[0, 1]$. Using a limit theorem for the operator P, Walters [22] shows, that there is a unique equilibrium state μ for every Lipschitz continuous function $\varphi: [0, 1] \to \mathbb{R}$, and that μ is weak Bernoulli.

In §4 of Chap. III we consider $\varphi: [0, 1] \to \mathbb{R}$ of bounded variation, which satisfy $\sum_{i=1}^{\infty} \operatorname{var}_i \varphi < \infty$, where $\operatorname{var}_i \varphi = \sup \{|\varphi(x) - \varphi(y)|: x \text{ and } y \text{ are contained in the same inerval, on which } T^i \text{ is monotone}\}$. Using the same technique as Walters [22], we show the existence of a probability m, a $\lambda > 0$ and an $n \in \mathbb{N}$ such that $m(Pf) = m(f)$ and $\|g\|_\infty < 1$ is satisfied for $g(x) = \exp\left(\sum_{i=0}^{n-1} \varphi(T^i x)\right) \Big/ \lambda^n$. Furthermore we get that μ is then an equilibrium state for φ.

This means that we generalize the results of Walters to a larger class of functions φ and show besides the weak Bernoulli property also a central limit theorem and an almost sure invariance principle for certain processes $f \circ T^n$.

(iv) Also subshifts X of finite type with finite alphabet are piecewise monotonic, if one introduces the lexicographic ordering in X.

Using a limit theorem for the operator P, Bowen [2] has shown the weak Bernoulli property with exponentially decreasing coefficient and a central limit theorem for equilibrium states of Hölder continuous functions $\varphi: X \to \mathbb{R}$, i.e. $\operatorname{var}_i \varphi \leqq C q^i$ for some $C > 0$ and $0 < q < 1$.

In §3 of Chap. III we show for every $\varphi: X \to \mathbb{R}$ of bounded variation, which satisfies $\sum_{i=1}^{\infty} \operatorname{var}_i \varphi < \infty$, the existence of a probability measure m, a $\lambda > 0$ and an $n \in \mathbb{N}$ such that $m(Pf) = m(f)$ and $\|g\|_\infty < 1$ is satisfied for $g(x) = \exp\left(\sum_{i=0}^{n-1} \varphi(T^i x)\right) \Big/ \lambda^n$. Furthermore μ is then an equilibrium state for φ. Hence for subshifts of finite type we show Bowen's results for a slightly different class of φ's.

Finally we should like to thank M. Denker for encouraging discussions during the preparation of the paper.

I. Spectral Decomposition of the Operator P and the Measure μ

Let X be a totally ordered, complete set. Then the order topology, this is that one generated by open intervals, makes X into a compact space [see e.g. Birkhoff, Lattice Theory, A.M.S. Colloquium Publications 25, Chap. X, §7, Theorem 12]. We call a transformation T on X *piecewise monotonic*, if $X = \bigcup_{i=1}^{N} I_i$, the I_i are disjoint intervals and $T | I_i$ is continuous and preserves or reverses the ordering such that $T(I_i)$ is again an interval. We consider $g: X \to (0, d]$ with

(a) $$d < 1 \quad \text{and} \quad V_X g < \infty$$

where the variation $V_X g$ of g over X is defined by

$$\sup \left\{ \sum_{i=1}^{k-1} |g(a_i) - g(a_{i+1})|: k \geqq 2, a_1 < \ldots < a_k, a_i \in X \right\}.$$

The *Perron-Frobenius-operator* P on the set $\mathfrak{F}$ of all bounded complex-valued functions on X, which are measurable with respect to the Borel-σ-algebra $\mathfrak{B}$, is defined by

$$Pf(x)=\sum_{y\in T^{-1}x} g(y)f(y)=\sum_{i=1}^{N} g(T_i^{-1}x)f(T_i^{-1}x)\,1_{T(I_i)}(x)$$

where $T_i=T\mid I_i$. Finally we assume that there is a Borel-probability-measure m on X which satisfies

(b) $$m(Pf)=m(f) \quad \text{for all } f\in\mathfrak{F}.$$

§1. We begin with the investigation of the support of the measure m.

Lemma 1. *For $B\in\mathfrak{B}$, $m(B)=0$ implies that $m(TB)=0$.*

Proof. We can assume that $B\subset I_i$ for some i. Suppose that $m(TB)>0$. We set $A_n=\left\{x\in T(I_i): g(T_i^{-1}x)\geqq\frac{1}{n}\right\}$. As $g(x)>0$ for all $x\in X$, it follows that $\bigcup_{n=1}^{\infty} A_n = T(I_i)$. This implies that there is a k with $m(A_k\cap TB)>0$. Then we have

$$m(B)=m(P1_B)=m\left(\sum_{j=1}^{N} g(T_j^{-1}x)\,1_B(T_j^{-1}x)\,1_{TI_j}(x)\right)$$

$$=m(g(T_i^{-1}x)\,1_{TB}(x))\geqq\frac{1}{k}\,m(TB\cap A_k)>0.$$

This contradicts $m(B)=0$. Hence $m(TB)=0$.

Let $\mathfrak{P}=\{I_1,\ldots,I_N\}$ denote the partition of X into intervals of continuity and monotonicity of T. Then $\mathfrak{P}_0^k=\bigvee_{i=0}^{k-1} T^{-i}\mathfrak{P}$ is the finite partition of X into intervals, on which T^k is monotone and continuous.

Lemma 2. *$m(A)\leqq d^k$ for all $A\in\mathfrak{P}_0^k$. In particular, $m(\{x\})=0$ for all $x\in X$.*

Proof. This is proved by induction. Let $A\in\mathfrak{P}_0^n$ with $n\geqq 1$ be such that $A\subset I_i$. Then

$$m(A)=m(P1_A)=m(g(T_i^{-1}(x))\,1_{TA}(x))\leqq d\,m(TA).$$

One sees easily that TA is contained in an element of $\mathfrak{P}_0^{n-1}$. Hence $m(TA)\leqq d^{n-1}$ by induction hypothesis and $m(A)\leqq d^n$.

Lemma 3. *Let $Y=\bigcap\{F: F \text{ is a closed subset of } X \text{ with } m(F)=1\}$ be the support of m. T can be changed at a finite number of points in such a way that then $T(Y)\subseteq Y$ and $T^{-1}\{x\}\subseteq Y$ for all $x\in Y$ except finitely many ones.*

Proof. Let J be a subinterval of X, which is contained in I_i for some i. We show that $m(TJ)>0$ if and only if $m(J)>0$. It follows from Lemma 1 that $m(TJ)>0$ implies $m(J)>0$. The other direction follows from

$$m(T^{-1}B)=m(1_B\circ T)=m(P(1_B\circ T))=m(1_B\cdot P1)\leqq N d\,m(B)$$

because $P1(x)\leqq Nd$ for all $x\in X$ ($g(x)\leqq d$ and the set $T^{-1}\{x\}$ contains at most N elements).

If x is now in the interior of some I_i then one takes for J open subintervals of I_i containing x and it follows that $x\in Y$ if and only if $Tx\in Y$. In particular, $T^{-1}\{x\}\subseteq Y$ for all $x\in Y$, which are not endpoints of some TI_i. As we shall change T only at endpoints of I_i's, this gives the second assertion. By Lemma 2, Y contains no isolated points. Hence, for an endpoint x of some I_i, we can redefine $T(x)$ as a one-sided limit of $T(t)$ for $t\in Y$. Then $T(Y)\subseteq Y$, because Y is closed and $T(t)\in Y$ for all $t\in Y$, which are in the interior of some I_i. This proves the lemma.

If T is changed on a finite set, $Pf(x)$ changes only for finitely many x. By Lemma 2, we have then $m(Pf)=m(f)$ also for the modified T of Lemma 3. We consider now the modified T, restrict it to the closed set $Y\subseteq X$, which is T-invariant by Lemma 3, and denote it again by T. As the sets $T^{-1}\{x\}$ for the original and the restricted T differ only for finitely many $x\in Y$ by Lemma 3, it follows from Lemma 2 that $m(Pf)=m(f)$ holds also for the restricted T. Y is again a totally ordered set, which is compact and has the advantage, that $m(J)>0$ for every nontrivial interval $J\subseteq Y$ (important for Lemmas 5 and 7 below). T on Y is again piecewise monotonic. Hence we consider from now on only the dynamical system (Y, T).

For $f\colon Y\to\mathbb{C}$ we define the *variation over a subset J of Y* by

$$V_J f=\sup\left\{\sum_{i=1}^{k-1}|f(a_i)-f(a_{i+1})|\colon k\geqq 2,\ a_1<\ldots<a_k,\ a_i\in J\right\}.$$

We shall need a variation $v(f)$ for $f\in L^1_m$, the set of all equivalence classes of complex-valued, m-integrable functions on Y. We define

$$v(f)=\inf\{V_Y\tilde{f}\colon \tilde{f} \text{ is a version of } f\}.$$

Remark 1. If $\tilde{f}$ is defined on $Y\setminus N$, if $Y\setminus N$ is dense in Y – this is satisfied, if N has measure zero, because every interval has positive m-measure – and if $V_{Y\setminus N}\tilde{f}<\infty$, then one shows easily that for all $x\in Y$ at least one of the limits $\tilde{f}(x^+)=\lim_{z\downarrow x}\tilde{f}(z)$ and $\tilde{f}(x^-)=\lim_{z\uparrow x}\tilde{f}(z)$ exists, since each $x\in Y$ can be approximated from above or below (Y contains no isolated points). Extend $\tilde{f}$ to the whole of Y such that for each $x\in N$ $\tilde{f}(x)=\tilde{f}(x^+)$ or $\tilde{f}(x)=\tilde{f}(x^-)$. Then $V_Y\tilde{f}=V_{Y\setminus N}\tilde{f}$ as is easily seen.

Now set $BV=\{f\in L^1_m\colon v(f)<\infty\}$, which is a linear subspace of L^1_m, but it is not closed with respect to $\|\ \|_1$. We define for $f\in BV$

$$\|f\|_v=\int|f|\,dm+v(f)=\|f\|_1+v(f).$$

It is easy to see that $\|\ \|_v$ is a norm on BV. The following lemma is needed for Lemma 5.

Lemma 4. *Let $f\in L^1_m$ satisfy $v(f)\leqq c<\infty$. Then $\int|f-E_m(f\,|\,\mathfrak{A}_n)|\,dm\leqq cd^n$, where $E_m(f\,|\,\mathfrak{A}_n)$ denotes the conditional expectation of f with respect to the measure m and the σ-algebra $\mathfrak{A}_n$ generated by $\mathfrak{P}_0^n$.*

Proof.

$$\int |f - E_m(f \mid \mathfrak{A}_n)|\, dm \leqq \sum_{A \in \mathfrak{P}_0} m(A)(\operatorname{ess\,sup}_{x \in A} f(x) - \operatorname{ess\,inf}_{x \in A} f(x)) \leqq d^n \cdot v(f) \leqq c d^n.$$

For the second inequality we have used Lemma 2.

The following lemma and Lemmas 6 and 7 will be requirements for the theorem of Ionescu-Tulcea and Marinescu (Theorem 1 below).

Lemma 5. (i) *For every $c>0$ the set $E=\{f\in L^1_m\colon \|f\|_v \leqq c\}$ is compact in $(L^1_m, \|\ \|_1)$.*

(ii) *$(BV, \|\ \|_v)$ is a Banach space*

(iii) *BV is dense in $(L^1_m, \|\ \|_1)$.*

Proof. (i) Let f_n be in E ($n \geqq 1$). From Lemma 4 and Theorem IV.8.18 in [4] it follows that there is a subsequence (g_n) of (f_n) and a $f \in L^1_m$ with $\lim_{n\to\infty} \|g_n - f\|_1 = 0$. We must show that $f \in E$: Let $\varepsilon > 0$ be arbitrary. There are versions $\tilde{g}_n$ of g_n such that $V_Y(\tilde{g}_n) \leqq v(g_n) + \varepsilon$ for all n. As g_n converges to f in L^1_m, we can find a subsequence $(\tilde{h}_n)$ of $(\tilde{g}_n)$ and a set $N \subseteq Y$ of measure zero such that $\lim_{n\to\infty} \tilde{h}_n(x) = \tilde{f}(x)$ for all $x \in Y \setminus N$, where $\tilde{f}$ is a version of f defined on $Y \setminus N$. The corresponding subsequence of (g_n) will be called (h_n). We thus have

$$\begin{aligned} V_{Y\setminus N}(\tilde{f}) &\leqq \varepsilon + \sum_{i=1}^{k} |\tilde{f}(a_i) - \tilde{f}(a_{i-1})| \qquad \text{for suitable } k \in \mathbb{N} \text{ and } a_i \in Y \setminus N, \\ &= \varepsilon + \lim_{n\to\infty} \sum_{i=1}^{k} |\tilde{h}_n(a_i) - \tilde{h}_n(a_{i-1})| \\ &\leqq \varepsilon + \limsup_{n\to\infty} V_Y(\tilde{h}_n) \\ &\leqq 2\varepsilon + \limsup_{n\to\infty} v(h_n). \end{aligned}$$

Extending $\tilde{f}$ to the whole of Y as in Remark 1 we obtain

$$v(f) \leqq V_Y(\tilde{f}) = V_{Y\setminus N}(\tilde{f}) \leqq \limsup_{n\to\infty} v(h_n),$$

as $\varepsilon > 0$ was arbitrary and $h_n \in E$. Together with $\lim_{n\to\infty} \|h_n - f\|_1 = 0$ this implies $\|f\|_v \leqq \limsup_{n\to\infty} \|h_n\|_v \leqq c$, i.e. $f \in E$. This proves (i).

(ii) follows immediately from (i) and I.1.6 in [18].

(iii) follows from the fact that the functions constant on the elements of some $\mathfrak{P}_0^n$ are dense in $(L^1_m, \|.\|_1)$, which is a standard result of measure theory, as $\bigcup_{n=1}^{\infty} \mathfrak{P}_0^n$ generates the Borel-σ-algebra m-a.e.

§2. Now we turn to the investigation of the operator P. The following result is an easy consequence of (b) and the formula $\varphi \cdot Pf = P(f \cdot (\varphi \circ T))$, which is easiliy verified.

Lemma 6. (i) *P is an operator on L^1_m satisfying $m(Pf) = m(f)$ for $f \in L^1_m$.*

(ii) *$\|Pf\|_1 \leqq \|f\|_1$ for all $f \in L^1_m$.*

Proof. We show (ii) first for $f \in \mathfrak{F}$ using (b). We have $|Pf(x)| = Pf(x) \cdot \varphi(x)$, where $\varphi(x) = Pf(x)/|Pf(x)|$. Since $|\varphi(x)| = 1$ for all $x \in Y$, we have

$$\begin{aligned} \|Pf\|_1 &= \int (Pf)\,\varphi\,dm = \int P(f \cdot (\varphi \circ T))\,dm \\ &= \int f(\varphi \circ T)\,dm \leqq \int |f| \cdot |\varphi \circ T|\,dm = \|f\|_1. \end{aligned}$$

If one changes f on a set B of measure zero, then Pf is changed on TB, which has measure zero by Lemma 1. Hence P can be considered as an operator on the set of all bounded $f \in L^1_m$. Because of $\|Pf\|_1 \leqq \|f\|_1$ for $f \in \mathfrak{F}$ and as $\mathfrak{F}$ is dense in L^1_m, P can be extended continuously to all $f \in L^1_m$. This proves the lemma.

Now we choose an $M \in \mathbb{N}$ such that $\|g_M\|_\infty = \sup\{g_M(x): x \in Y\} < \frac{1}{3}$, where g_n for $n \geqq 1$ is given by $g_n(x) = g(x)\,g(Tx) \ldots g(T^{n-1}x)$. Such an M exists, because $\|g\|_\infty \leqq d < 1$ (cf. (a)). The following result is a modification of a result in [23] (cf. also [11]).

Lemma 7. *There is an $\alpha \in (0, 1)$ and a $\beta \in (0, \infty)$ with*

$$v(P^M f) \leqq \alpha v(f) + \beta \|f\|_1 \quad \text{for all } f \in L^1_m.$$

Proof. Writing Q for P^M and U for T^M, we have $Qf(x) = \sum_{y \in U^{-1}x} g_M(y) f(y)$. Let $(J_1, \ldots, J_{\tilde{N}})$ be a partition of Y into intervals such that U/J_i is monotone and continuous and that $V_{J_i} g_M \leqq \|g_M\|_\infty$. As $T(I_i)$ is an interval in Y, also $U(J_i)$ is an interval in Y. For $\varepsilon > 0$, let $\tilde{f}$ be a version of $f \in L^1_m$ such that $V_Y \tilde{f} \leqq v(f) + \varepsilon$. As $m(J_i) > 0$ for $1 \leqq i \leqq \tilde{N}$ and as $U(J_i)$ is an interval in Y, the same computation as in Theorem 1 of [23] works for Q and $\tilde{f}$. Hence we get $V_Y Q\tilde{f} \leqq \alpha V_Y \tilde{f} + \beta \|\tilde{f}\|_1$, where $\alpha = 3\|g_M\|_\infty < 1$ and $\beta = 3\|g_M\|_\infty / \min\{m(J_i): 1 \leqq i \leqq \tilde{N}\}$. Since $Q\tilde{f}$ is a version of Qf (cf. the proof of Lemma 6) we have

$$v(Qf) \leqq V_Y Q\tilde{f} \leqq \alpha V_Y \tilde{f} + \beta \|\tilde{f}\|_1 \leqq \alpha \cdot v(f) + \beta \|f\|_1 + \alpha\varepsilon.$$

Because ε was arbitrary, we get the desired result.

The properties of the space BV and of the operator P, which are listed in Lemma 5, 6 and 7 allow to apply an ergodic theorem of Ionescu-Tulcea and Marinescu [9], stating in an abstract Banach space setting, the following results (they are formulated here for the special situation under consideration).

Theorem 1. *Under the above assumptions on P, T, X and m we have*

(i) *$P: L^1_m \to L^1_m$ only has a finite number of eigenvalues $\lambda_1, \ldots, \lambda_r$ of modulus 1.*

(ii) *Set $E_i = \{f \in L^1_m: Pf = \lambda_i f\}$ for $1 \leqq i \leqq r$. Then $E_i \subset BV$ and $\dim(E_i) < \infty$.*

(iii) *The operator P can be represented as*

$$P = \sum_{i=1}^{r} \lambda_i \phi_i + \psi$$

where the ϕ_i are projections onto the eigenspaces E_i, $\|\phi_i\|_1 \leqq 1$, and ψ is a linear operator on L^1_m with $\sup_{n \geqq 1} \|\psi^n\|_1 < \infty$. Furthermore $\phi_i \phi_j = 0 (i \neq j)$ and $\phi_i \psi = 0$ (all i).

(iv) $\psi(BV) \subset BV$ *and, considered as a linear operator on* $(BV, \| \ \|_v)$, ψ *satisfies* $\|\psi^n\|_v \leqq H q^n$ $(n \geqq 1)$ *for some constants* H *and* q *with* $H > 0$ *and* $0 < q < 1$.

(v) 1 *is an eigenvalue of* P. *We assume* $\lambda_1 = 1$.

(vi) *Writing* $h := \phi_1(1)$ *and* $\mu = hm$ *one has* $h \geqq 0$, $\int h\,dm = 1$ *and* $Ph = h$. *Thus* μ *is a* T*-invariant probability on* Y *and hence on* X. *If* μ' *is* T*-invariant and* $\mu' \ll m$, *then* $\mu' \ll \mu$.

Proof. (i) to (iv) are immediate consequences of the theorem of Ionescu-Tulcea and Marinescu. It remains to prove (v) and (vi).

Set $h_n = \frac{1}{n} \sum_{k=0}^{n-1} P^k(1)$. Then $h_n \geqq 0$, because P is a positive operator and $\int h_n\,dm = 1$ because of $m(Pf) = m(f)$. Using (iii) and evaluating geometric series it follows that h_n converges in L^1_m. The limit function is invariant under P, nonnegative and has integral 1. Thus 1 is an eigenvalue of P, say $\lambda_1 = 1$, and it follows from (iii) as above that h_n converges to $\phi_1(1) = h$. That μ is T-invariant follows from the following equalities, because $Ph = h$:

$$\mu(f \circ T) = m(h \cdot (f \circ T)) = m(P(h \cdot (f \circ T))) = m(f \cdot Ph)$$
$$\mu(f) = m(fh).$$

If $\mu' = h'm$ is T-invariant, it follows from these equalities that $Ph' = h'$. Thus $h' \in BV$ by (ii) and h' is bounded by some constant c. This implies $h' = \frac{1}{n} \sum_{k=0}^{n-1} P^k(h') \leqq c h_n$, hence $h' \leqq ch$ and $\mu' \ll \mu$.

§3. Once having obtained the invariant measure μ for T, we focus attention to the dynamical system (T, μ). Sometimes we shall adopt the Hilbert-space-viewpoint, interpreting T as a linear operator on the complex Hilbert space L^2_μ and writing $\langle f, g \rangle$ for $\int f \bar{g}\,d\mu$, such that $\langle Tf, Tg \rangle = \langle f, g \rangle$.

The next lemma will provide a description of the adjoint T^* of T in terms of P.

Lemma 8. (i) $T^* f = \frac{P(f \cdot h)}{h}$ *for all* $f \in L^2_\mu$.

(ii) $T^* f = \lambda f \Leftrightarrow P(fh) = \lambda f h$ *for all* $f \in L^2_\mu$ *and* $\lambda \in \mathbb{C}$ *(all identities are* L^2_μ*-identities).*

Proof. (i) for $f, g \in L^2_\mu$ one has

$$\langle T^* f, g \rangle = \langle f, Tg \rangle = \int f \cdot (\bar{g} \circ T)\, h\,dm = \int \frac{P(fh)}{h}\, \bar{g}\,d\mu = \left\langle \frac{P(fh)}{h}, g \right\rangle.$$

(ii) $T^* f = \lambda f \Leftrightarrow P(fh) = \lambda f h$ because of (i).

Lemma 9. (i) $(T^*)^n \circ T^n = Id$ *for each* $n \in \mathbb{N}$.

(ii) $T^n \circ (T^*)^n = Pr_n$ *for each* n, *where* Pr_n *denotes the orthogonal projection onto the subspace* $T^n(L^2_\mu)$.

Proof. (i) It suffices to show that $T^* T = Id$. But for $f, g \in L^2_\mu$ one has $\langle T^* Tf, g \rangle = \langle Tf, Tg \rangle = \langle f, g \rangle$.

(ii) Since $(T^n \circ (T^*)^n)^* = T^n \circ (T^*)^n$ and $(T^n \circ (T^*)^n)^2 = T^n \circ (T^*)^n$ (because of (i)), $T^n \circ (T^*)^n$ is an orthogonal projection. On the other hand,

$$(T^n \circ (T^*)^n)(L^2_\mu) \subset T^n(L^2_\mu) \quad \text{and} \quad (T^n \circ (T^*)^n)(T^n f) = T^n f$$

for each $f \in L^2_\mu$. Thus $T^n \circ (T^*)^n = Pr_n$.

Now we can easily clarify the ergodic structure of (T, μ) up to the level of weak mixing (for weak mixing cf. [21]).

Theorem 2. *Let $G \subseteq \{z \in \mathbb{C}: |z| = 1\}$ be the set of eigenvalues of T. Then*

(i) $G = \{\lambda_1, \ldots, \lambda_r\}$

(ii) $Tf = \lambda f \Leftrightarrow T^* \bar{f} = \lambda \bar{f}$ *and* $|\lambda| = 1$ *for all* $f \in L^2_\mu$ *and* $\lambda \in \mathbb{C}$.

(iii) $\dim(E_{(T,\mu)}(\lambda)) < \infty$ *for each* $\lambda \in G$, *where* $E_{(T,\mu)}(\lambda) = \{f \in L^2_\mu: Tf = \lambda f\}$.

Proof. Let us first assume that $Tf = \lambda f$ for some $f \in L^2_\mu$, $f \neq 0$. Then $|\lambda| = 1$, as μ is T-invariant, and $T^* \bar{f} = \lambda T^*(T\bar{f}) = \lambda \bar{f}$. On the other hand, if $T^* f = \lambda f$ for some $f \in L^2_\mu$, $f \neq 0$ and $\lambda \in \mathbb{C}$, $|\lambda| = 1$, we have

$$\langle f, f \rangle = \langle T^* f, \lambda f \rangle = \langle f, \lambda T f \rangle \quad \text{and} \quad \langle \lambda T f, \lambda T f \rangle = |\lambda|^2 \langle Tf, Tf \rangle = \langle f, f \rangle.$$

This implies $\langle \lambda Tf - f, f \rangle = 0$ and $\langle \lambda Tf - f, \lambda Tf \rangle = 0$, hence $\lambda Tf = f$ or, equivalently, $T\bar{f} = \lambda \bar{f}$. This proves (ii).

If now $Tf = \lambda f$, we get from (ii) and Lemma 8 that $|\lambda| = 1$ and $P(\bar{f}h) = \lambda(\bar{f}h)$. Hence it follows from Theorem 1 that $G \subseteq \{\lambda_1, \ldots, \lambda_r\}$. If $Pg = \lambda_i g$ for some i, then $g \in BV$ by Theorem 1 and $|g| = |P^n g| \leqq P^n |g| \leqq \|g\|_\infty \cdot P^n(1)$. Hence

$$|g| = \frac{1}{n} \sum_{i=0}^{n-1} |P^i g| \leqq \|g\|_\infty \frac{1}{n} \sum_{i=0}^{n-1} P^i(1) \to \|g\|_\infty h \quad \text{for } n \to \infty.$$

This gives $\frac{|g|}{h} \leqq \|g\|_\infty$ such that $f := \frac{g}{h} \in L^2_\mu$. Furthermore it follows from Lemma 8 and (ii) that $T\bar{f} = \lambda_i \bar{f}$. Hence $G \supseteq \{\lambda_1, \ldots, \lambda_r\}$ proving (i). The same arguments show also that $\dim(E_{(T,\mu)}(\lambda_i)) = \dim(E_i) < \infty$. This is (iii).

The next theorem describes the structure of (T, μ) up to the level of exactness (for exactness see [17]).

Theorem 3. *Let $\mathfrak{B}_\infty(T) := \bigcap_{n=1}^{\infty} T^{-n}\mathfrak{B}$, where $\mathfrak{B}$ denotes the Borel-σ-algebra on Y. Then*

(i) $\{f \in L^2_\mu: f$ *is* $\mathfrak{B}_\infty(T)$*-measurable*$\} = \bigcap_{n=1}^{\infty} T^n(L^2_\mu) = \sum_{i=1}^{r} E_{(T,\mu)}(\lambda_i)$. *In particular,* $\mathfrak{B}_\infty(T)$ *is finite* μ-a.s.

(ii) *If (T, μ) is weakly mixing, then $\mathfrak{B}_\infty(T)$ is trivial, i.e. (T, μ) is exact.*

Proof. (i) The first of the above equalities is generally true. Hence we only have to show the second one. First observe that for $f = \sum_{i=1}^{r} f_i$, $f_i \in E_{(T,\mu)}(\lambda_i)$, we have $f = T^n\left(\sum_{i=1}^{r} \bar{\lambda}_i^n f_i\right)$ such that $\sum_{i=1}^{r} E_{(T,\mu)}(\lambda_i) \subseteq \bigcap_{n=1}^{\infty} T^n(L^2_\mu)$.

The other direction is a bit harder to prove. Let $g \in L^2_\mu \cap BV$. Then $gh \in L^1_m \cap BV$ and since $P(\phi_i(gh)) = \lambda_i \phi_i(gh)$, we have $T^*(\phi_i(gh)/h) = \lambda_i \phi_i(gh)/h$ by Lemma 8. Hence $T(\phi_i(gh)/h) = \bar{\lambda}_i \phi_i(gh)/h$ by Theorem 2.

As $G = \{\lambda_1, \ldots, \lambda_r\}$ generates a finite subgroup of the circle, there is a $p \in \mathbb{N}$ such that $\lambda_i^p = 1$ for $1 \leqq i \leqq r$. This gives for any $n \in \mathbb{N}$

$$\begin{aligned} Pr_{np}(g) &= T^{np}(T)^{*np}(g) = T^{np}\left(\sum_{i=1}^{r} \phi_i(gh)/h + \psi^{np}(gh)/h\right) \\ &= \sum_{i=1}^{r} \phi_i(gh)/h + T^{np}(\psi^{np}(gh)/h). \end{aligned}$$

We have used Lemma 9 and 8 and (iii) of Theorem 1. Now one has for any $f \in \bigcap_{n=1}^{\infty} T^n(L^2_\mu)$, any $g \in L^2_\mu \cap BV$ and any $n \in \mathbb{N}$ that

$$\begin{aligned} \int \left| f - \sum_{i=1}^{r} \phi_i(fh)/h \right| d\mu &\leqq \int |Pr_{np}(f) - Pr_{np}(g)|\, d\mu \\ &\quad + \sum_{i=1}^{r} \int |\phi_i(fh)/h - \phi_i(gh)/h|\, d\mu + \int |\psi^{np}(gh)/h| \circ T^{np}\, d\mu \\ &\leqq \|Pr_{np}(f-g)\|_{2,\mu} + r \int |f-g|\, h\, dm + \int |\psi^{np}(gh)|\, dm \\ &\leqq (r+1)\, \|f-g\|_{2,\mu} + H q^{np}\, \|gh\|_v. \end{aligned}$$

Since BV is dense in L^2_μ this shows that $f = \sum_{i=1}^{r} \phi_i(fh)/h$ μ-a.e., i.e. $f \in \sum_{i=1}^{r} E_{(T,\mu)}(\lambda_i)$.

(ii) If (T, μ) is weakly mixing, then 1 is the only eigenvalue, and the space of fix-elements of T only contains the constant functions. Hence $\mathfrak{B}_\infty(T)$ is trivial by (i).

Remark 2. Let $W \subset Y$ satisfy $T^{-1} W = W$, i.e. $1_W \circ T = 1_W$. Define

$$m'(A) = m(A \cap W)/m(W).$$

Then

$$\begin{aligned} m'(Pf) &= m(1_W\, Pf)/m(W) = m(P(f \cdot (1_W \circ T)))/m(W) \\ &= m(f \cdot (1_W \circ T))/m(W) = m(f \cdot 1_W)/m(W) = m'(f). \end{aligned}$$

Similarly one can show that $P(h \cdot 1_W) = h\, 1_W$. Hence every ergodic component of μ is again of the form $h \cdot m$ for some m and h satisfying $m(Pf) = m(f)$ and $Ph = h$.

If one considers T^k instead of T, where k is such that $\lambda_i^k = 1$ for all i, then an ergodic component of μ is already weak mixing for T^k (cf. Theorem 2). Hence every weak mixing component of μ is again of the form hm for some m and h with $m(Pf) = m(f)$ and $Ph = h$. Here one has to take T^k instead of T and g_k instead of g in the definition of P. Therefore we can assume that μ is weak mixing. We shall do this in the next chapter.

Remark 3. The spectral representation of Theorem 1 together with Theorem 2 allows an easy proof of a central limit theorem for stationary processes of the

kind $\xi_n = f \circ T^n$ with $f \in BV$. The proof is completely analogous to the one in [10], where it is shown how to reduce this problem to a central limit theorem of Gordin [5]. As in the next chapter we shall even derive almost sure invariance principles, an exact statement of this theorem is omitted.

II. Weak Bernoullicity and Invariance Principles

In this chapter we shall show that, if (T, μ) is weakly mixing (cf. Remark 2), then the natural partition $\mathfrak{P} = (I_1, \ldots, I_N)$ of Y into intervals of continuity and monotonicity of T (and even more each finite partition of Y into intervals) is a weak-Bernoulli-partition for T. This means that the associated "label-process" (ξ_n), where $\xi_n(x) = i$, if $T^n x \in I_i$, is absolutely regular. The corresponding mixing-coefficients will be shown to decrease at an exponential rate.

§1. We start with the investigation of the weak Bernoulli property. Technically we can follow very closely Ratner's article [16]. Remember that, for all $k \geqq 1$, $\mathfrak{P}_0^k = \bigvee_{i=0}^{k-1} T^{-i}\mathfrak{P}$ is a finite partition of Y into intervals of m-measure $\leqq d^k$ (cf. Lemma 2) such that m-measure plays the same role here as Lebesgue-measure (respectively "length") does in the article of Ratner [16]. In that paper the invariant measure also is called μ, but its density is called p and the function $\varphi(x) = 1/|f'(x)|$ plays the role of our $g(x)$.

The first lemma is proved in Bowen's paper [3] and is easily seen to be true also in this situation.

Lemma 10. *If $A \in \mathfrak{P}_0^k$, and if $\bar{A}$ contains no endpoint of an interval $I_i \in \mathfrak{P}$, then $TA \in \mathfrak{P}_0^{k-1}$.*

The next lemma is the same as Lemma 3.2 in [16].

Lemma 11. *There is a constant $L > 0$ such that, given $\varepsilon > 0$, we can find for all $n > L \cdot \log \frac{1}{\varepsilon}$ a collection $\alpha_n \subseteq \mathfrak{P}_0^n$ with $\mu(\bigcup \alpha_n) > 1 - \varepsilon$ such that for any $x, y \in A \in \alpha_n$*

$$\frac{h(x)}{h(y)} \in [e^{-\varepsilon}, e^{\varepsilon}] \quad \text{and} \quad \frac{g(x)}{g(y)} \in [e^{-\varepsilon}, e^{\varepsilon}].$$

Proof. The proof is quite analogous to the one in [16] except for the estimate of the measure of the set $S_l = \{x \in Y: g(x) < l\}$ $(l > 0)$, which must be replaced by

$$\begin{aligned} \mu(S_l) &\leqq \|h\|_\infty \cdot m(S_l) = \|h\|_\infty \int P(1_{S_l})\, dm \\ &\leqq \|h\|_\infty \sum_{i=1}^{N} \|(g \circ T_i^{-1}) \cdot 1_{T_i(S_l)}\|_\infty, \qquad \text{where } T_i = T/I_i \\ &\leqq \|h\|_\infty \cdot N \cdot l, \qquad \text{by definition of } S_l. \end{aligned}$$

The next lemma, which is basic for our further considerations, can be proved just as the corresponding one (3.3) in [16], too.

Lemma 12. *There is a constant $D>0$ such that for all sufficiently large $M\in\mathbb{N}$ and all $m\geqq 0$ one can find a collection of atoms $\beta=\beta_{m,M}\subseteq\mathfrak{P}_0^{m+M}$ with*

(i) $T^m B\in\mathfrak{P}_0^M$ *for* $B\in\beta$

(ii) $\mu(\bigcup\beta)>1-e^{-M/D}$

(iii) $\left|\frac{\mu(T^m\tilde{B})}{\mu(T^m B)}-\frac{\mu(\tilde{B})}{\mu(B)}\right|<e^{-M/D}\frac{\mu(\tilde{B})}{\mu(B)}$ *for any measurable* $\tilde{B}\subseteq B\in\beta$ *with* $\mu(B)>0$.

Proof. The proof is again as in [16] using Lemmas 10 and 11 instead of the corresponding ones in that paper. One only has to observe that for any measurable $\tilde{B}\subset B\in\mathfrak{P}_0^n$ (n arbitrary) the following holds. Put $q(x):=h(T^n x)/(h(x)\,g_n(x))$, where $g_n(x)=g(x)\,g(Tx)\ldots g(T^{n-1}x)$. Then

$$\begin{aligned}\int_{\tilde{B}} q\,d\mu&=\int P^n(1_{\tilde{B}}\cdot(h\circ T^n)/g_n)\,dm\\&=\int\sum_{y\in T^{-n}x} g_n(y)\,1_{\tilde{B}}(y)(h(T^n y)/g_n(y))\,dm(x)\\&=\int\Big(\sum_{y\in T^{-n}x}1_{\tilde{B}}(y)\Big)\,h(x)\,dm(x)\\&=\int 1_{T^n\tilde{B}}(x)\,h(x)\,dm(x)\qquad\text{since }\tilde{B}\subseteq B\in\mathfrak{P}_0^n\\&=\mu(T^n\tilde{B})\end{aligned}$$

and

$$\frac{q(y)}{q(z)}=\frac{h(T^n y)}{h(T^n z)}\cdot\frac{h(z)}{h(y)}\cdot\prod_{k=0}^{n-1}\frac{g(T^k z)}{g(T^k y)}.$$

The next lemma "translates" the information about P, which is contained in the spectral representation, into the "language" of partitions. We introduce some new notation before. Let $\mathfrak{R}$ and $\mathfrak{S}$ be two finite collections of pairwise disjoint measurable subsets of Y. Define

$$D(\mathfrak{R},\mathfrak{S}):=\sum_{R\in\mathfrak{R}}\sum_{S\in\mathfrak{S}}|\mu(R\cap S)-\mu(R)\,\mu(S)|.$$

Lemma 13. *Under the assumption that (T,μ) is weakly mixing there is a constant $C>0$ such that for all M, n, $l\geqq 0$*

$$D(\mathfrak{P}_0^M,\mathfrak{P}_{M+l}^{M+l+n})\leqq C\cdot N^M q^l$$

where q is the constant of (iv) *of Theorem* 1 *and N the number of elements of $\mathfrak{P}$.*

Proof. Set $\mathfrak{R}=\mathfrak{P}_0^M$ and $\mathfrak{S}=\mathfrak{P}_{M+l}^{M+l+n}=T^{-l}(\mathfrak{P}_M^{M+n})$. Then

$$D(\mathfrak{R},\mathfrak{S})=\sum_{R\in\mathfrak{R}}\{|\mu(R\cap S_R^+)-\mu(R)\,\mu(S_R^+)|+|\mu(R\cap S_R^-)-\mu(R)\,\mu(S_R^-)|\},$$

where

$$S_R^+=\cup\{S\in\mathfrak{S}:\mu(R\cap S)>\mu(R)\,\mu(S)\}$$

and

$$S_R^-=\cup\{S\in\mathfrak{S}:\mu(R\cap S)\leqq\mu(R)\,\mu(S)\}=Y\setminus S_R^+.$$

Because of $\mathfrak{S}=T^{-l}(\mathfrak{P}_M^{M+n})$, there are measurable sets $\tilde{S}_R^+$ and $\tilde{S}_R^-$ with $S_R^+=T^{-l}(\tilde{S}_R^+)$ and $S_R^-=T^{-l}(\tilde{S}_R^-)$. Hence

$$D(\mathfrak{R}, \mathfrak{S}) = \sum_{R\in\mathfrak{R}} \{|\int 1_{S_R^+}(1_R - \mu(R))\, h\, dm| + |\int 1_{S_R^-}(1_R - \mu(R))\, h\, dm|\}$$
$$= \sum_{R\in\mathfrak{R}} \{|\int 1_{\tilde{S}_R^+} P^l((1_R - \mu(R))\, h)\, dm| + |\int 1_{\tilde{S}_R^-} P^l((1_R - \mu(R))\, h)\, dm|\}$$
$$\leqq 2 \sum_{R\in\mathfrak{R}} \|P^l((1_R - \mu(R))\, h)\|_v.$$

Since (T, μ) is weakly mixing, it follows from Theorems 2 and 1 that $P^l = \phi_1 + \psi^l$. Furthermore this implies $\phi_1((1_R - \mu(R))\, h) = (\int (1_R - \mu(R))\, h\, dm) \cdot h \equiv 0$. Thus

$$D(\mathfrak{R}, \mathfrak{S}) \leqq 2 \sum_{R\in\mathfrak{R}} \|\psi^l((1_R - \mu(R))\, h)\|_v$$
$$\leqq 2 \sum_{R\in\mathfrak{R}} H q^l \|(1_R - \mu(R))\, h\|_v$$
$$\leqq 2\, |\mathfrak{R}|\, H q^l 3 \|h\|_v$$
$$\leqq 6H \|h\|_v N^M q^l = C N^M q^l.$$

The third inequality follows because the function $1_R - \mu(R)$ has the value $1 - \mu(R)$ on the interval R and the value $-\mu(R)$ on $Y \setminus R$. Hence $v((1_R - \mu(R))\, h) \leqq 3v(h)$. The last inequality follows from $|\mathfrak{R}| \leqq N^M$, which is true, because $\mathfrak{R} = \mathfrak{P}_0^M$.

Now we can derive an estimate of $D(\mathfrak{P}_0^n, \mathfrak{P}_{n+l}^{2n+l})$ independent of n.

Theorem 4. *If (T, μ) is weakly mixing, then there are constants $K > 0$ and p with $0 < p < 1$ such that*

$$D(\mathfrak{P}_0^n, \mathfrak{P}_{n+l}^{n+l+k}) \leqq K p^l$$

for arbitrary n and k.

Proof. Fix $\delta > 0$ such that $\tilde{q} = q N^\delta < 1$ (q as in Lemma 13). It suffices to show Theorem 4 for $n \geqq [\delta l] + 1$. The case $n \leqq [\delta l]$ reduces immediately to $n = [\delta l] + 1$ using the triangle inequality.

Set $M = [\delta l] + 1$ and $m = n - M \geqq 0$. Then

$$D(\mathfrak{P}_0^n, \mathfrak{P}_{n+l}^{n+l+k}) \leqq D(\beta_{m,M}, \mathfrak{P}_{n+l}^{n+l+k}) + 2e^{-M/D}$$

by Lemma 12. We set $\mathfrak{S} = \mathfrak{P}_{n+l}^{n+l+k}$ and $\hat{\mathfrak{S}} = \mathfrak{P}_{M+l}^{M+l+k}$. Because $\mu(T^m S) = \mu(S)$ for every $S \in \mathfrak{S}$, we get ($\beta = \beta_{m,M}$)

$$D(\beta, \mathfrak{S}) = \sum_{B\in\beta} \mu(B) \sum_{S\in\mathfrak{S}} \left\{ \left| \frac{\mu(B\cap S)}{\mu(B)} - \frac{\mu(T^m(B\cap S))}{\mu(T^m B)} \right| + \left| \frac{\mu(T^m(B\cap S))}{\mu(T^m B)} - \mu(T^m S) \right| \right\}$$
$$\leqq \sum_{B\in\beta} \sum_{S\in\mathfrak{S}} e^{-M/D} \mu(B\cap S) + \sum_{\hat{B}\in\mathfrak{P}_0^M} \left\{ \sum_{\substack{B\in\beta \\ T^m B = \hat{B}}} \mu(B) \sum_{\hat{S}\in\hat{\mathfrak{S}}} \left| \frac{\mu(\hat{B}\cap\hat{S})}{\mu(\hat{B})} - \mu(\hat{S}) \right| \right\}$$

by Lemma 12

$$\leqq e^{-M/D} + \sum_{\hat{B}\in\mathfrak{P}_0^M} \sum_{\hat{S}\in\hat{\mathfrak{S}}} |\mu(\hat{B}\cap\hat{S}) - \mu(\hat{B})\, \mu(\hat{S})|$$
$$\leqq e^{-M/D} + D(\mathfrak{P}_0^M, \mathfrak{P}_{M+l}^{M+l+k})$$
$$\leqq e^{-M/D} + C N^M q^l \quad \text{by Lemma 13.}$$

Putting together the two estimates and observing that $M=[\delta l]+1$ we get

$$D(\mathfrak{P}_0^n, \mathfrak{P}_{n+l}^{n+l+k}) \leqq 3e^{-l\delta/D} + CN\tilde{q}^l.$$

This gives the desired result if one sets $p=\max\{\tilde{q}, e^{-\delta/D}\}$ and $K=3+CN$.

Corollary. *If (T, μ) is weakly mixing and $\mathfrak{D}$ is an arbitrary finite partition of Y into intervals, then there are constants $\tilde{K}$ and $\tilde{p}$, $0<\tilde{p}<1$ such that $D(\mathfrak{D}_0^n, \mathfrak{D}_{n+l}^{n+l+k}) \leqq \tilde{K}\tilde{p}^l$ for any k and n.*

Proof. The refinement $\mathfrak{P} \vee \mathfrak{D}$ of $\mathfrak{P}$ and $\mathfrak{D}$ is again a finite partition of Y into intervals, on which T is monotone and continuous. Hence Theorem 4 is also true for $\mathfrak{P} \vee \mathfrak{D}$ instead of $\mathfrak{P}$ (with different constants $\tilde{K}$ and $\tilde{p}$). But then it is also true for each coarser partition, in particular for $\mathfrak{D}$.

§2. In order to show central limit theorems, a.s. invariance principles and other limit theorems for stationary processes governed by the dynamical system (T, μ) one only has to observe the following: For $\mathfrak{P}=\{I_1, \ldots, I_N\}$ let $(\xi_n)_{n \geqq 0}$ be the associated label-process, i.e. $\xi_n(x)=i$, if $T^n x \in I_i$. If $\mathfrak{F}_a^b$ denotes the σ-algebra generated by $\{\xi_n : a \leqq n < b\}$, then $\mathfrak{F}_a^b$ is the σ-algebra generated by $\mathfrak{P}_a^b$. Define

$$\beta(n) := \sup_{t \geqq 1} E_\mu[\sup_{A \in \mathfrak{F}_{t+n}^\infty} |\mu(A \mid \mathfrak{F}_0^t) - \mu(A)|].$$

If $\beta(n) \to 0$, (ξ_n) is called *absolutely regular*. One easily checks that (cf. [19])

$$2\beta(n) = \sup_{t \geqq 1} \sup_{k \geqq 1} D(\mathfrak{P}_0^t, \mathfrak{P}_{t+n}^{t+n+k}).$$

Hence $\beta(n) \leqq Kp^n$ by Theorem 4.

This enables us to apply some well-known limit theorems to the process (ξ_n) and to functionals of it. First define that $f: Y \to \mathbb{R}$ is of *bounded p-variation* on an interval $A \subseteq Y$ if

$$V_A^p f := \sup \sum_{i=1}^{n} |f(a_i) - f(a_{i-1})|^p < \infty,$$

where the supremum is taken over all finite subsets $a_0 < a_1 < \ldots < a_n$ of A.

Theorem 5. *Assume that (T, μ) is weakly mixing. Let $f: Y \to \mathbb{R}$ be a function of bounded p-variation, $p \geqq 1$, and $\int f d\mu = 0$. Define $S(t) := \sum_{i=0}^{t-1} f \circ T^i$, which is a stochastic process on (Y, μ). Then the series*

$$\sigma^2 := \int f^2 d\mu + 2 \sum_{k=1}^{\infty} \int f(f \circ T^k) d\mu$$

converges absolutely, $\int S(t)^2 d\mu = t\sigma^2 + O(1)$, and if $\sigma^2 \neq 0$ the following holds

(i) $$\sup_{z \in \mathbb{R}} \left| \mu\left\{\frac{1}{\sigma\sqrt{t}} S(t) \leqq z\right\} - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-x^2/2} dx \right| = O(t^{-\nu}) \text{ for some } \nu > 0.$$

(ii) *Without changing its distribution one can redefine the process $(S(t))_{t \geqq 0}$ on a richer probability space together with standard Brownian motion $(B(t))_{t \geqq 0}$ such that*

$$|\sigma^{-1} S(t) - B(t)| = O(t^{1/2-\lambda}) \quad \mu\text{-a.s.}$$

for some $\lambda > 0$.

Proof. We show that theorems in [15] and [1] are applicable here. First observe that the Borel-σ-algebra on Y is generated by the family $(\mathfrak{P}_0^k : k \in \mathbb{N})$, because each interval in Y has positive m-measure and $m(A) \leqq d^k$ for each $A \in \mathfrak{P}_0^k$ (see Lemma 2). Thus f can be considered to be μ-a.e. a functional of the label-process (ξ_n). Furthermore

(i) as a function of bounded p-variation, f is bounded.

(ii) For each $r \geqq p$

$$\begin{aligned}\int |f - E_\mu(f \mid \mathfrak{F}_0^k)|^r \, d\mu &\leqq \sum_{A \in \mathfrak{P}_0^k} \mu(A) \cdot \sup_{x,\, y \in A} |f(x) - f(y)|^r \\ &\leqq d^k \, \|h\|_\infty \sum_{A \in \mathfrak{P}_0^k} V_A^p f \cdot (2 \|f\|_\infty)^{r-p} \quad \text{by Lemma 2} \\ &\leqq d^k \, \|h\|_\infty \, V_Y^p f \cdot (2 \|f\|_\infty)^{r-p}.\end{aligned}$$

(iii) Set $R(k) := \int f(f \circ T^k) \, d\mu$. Then

$$\begin{aligned}|R(k)| &\leqq \int |f - E_\mu(f \mid \mathfrak{F}_0^{[k/2]})| \cdot |f \circ T^k| \, d\mu + \int |E_\mu(f \mid \mathfrak{F}_0^{[k/2]}) \cdot (f \circ T^k)| \, d\mu \\ &\leqq (d^{[k/2]} \, \|h\|_\infty \, V_Y^p f)^{1/p} \, \|f\|_\infty + 4 \, \|f\|_\infty^2 \cdot \beta([k/2])\end{aligned}$$

by Theorem 17.2.1 of [8]. Hence $|R(k)|$ decreases exponentially with $k \to \infty$ such that $\sigma^2 = R(0) + 2 \sum_{k=1}^{\infty} R(k)$ converges absolutely, and

$$\begin{aligned}\int S(t)^2 \, d\mu &= t \cdot R(0) + 2 \sum_{k=1}^{t-1} (t-k) \, R(k) \\ &= t\sigma^2 - 2t \sum_{k=t}^{\infty} R(k) - 2 \sum_{k=1}^{t-1} k \cdot R(k) \\ &= t\sigma^2 + O(1)\end{aligned}$$

since $|R(k)|$ decreases exponentially.

Thus Theorem 7.1 in [15] can be applied to obtain (ii), while (i) follows from [1].

Remark 4. (i) The order of approximation in (ii) of Theorem 5 is good enough to imply integral tests, log-log-laws, and weak invariance principles for the process $(f \circ T^i)_{i \geqq 0}$ (see e.g. [15]).

(ii) Each Hölder-continuous function on [0, 1] with exponent $\frac{1}{p}$ is of bounded p-variation.

III. Equilibrium States

In this chapter we show first that μ is an equilibrium state for the function $\log g$, i.e. $h(\mu) + \mu(\log g) = \sup \{h(\nu) + \nu(\log g) : \nu$ is a T-invariant probability on

$Y\}$, where $h(\mu)$ denotes the entropy of μ. Then we give a method, how one can show the existence of a measure m and a function g such that (a) and (b) of Chap. I are satisfied. This method is then applied in three different situations.

§1. In order to show that μ is an equilibrium state for $\log g$, we need the following lemmas. For a function $f: Y \to \mathbb{R}$ let f^+ be defined by $f^+(x) = \max\{f(x), 0\}$.

Lemma 14. *If $f: Y \to \mathbb{R}$ is measurable and ν is a T-invariant probability on Y, then $(f - f \circ T)^+ \in L^1_\nu$ implies $f - f \circ T \in L^1_\nu$ and $\int (f - f \circ T)\, d\nu = 0$.*

Proof. Setting $F := (f - f \circ T)^+$, we have $0 \leqq F + f \circ T - f$, $F \geqq 0$ and $F \in L^1_\nu$. Define $f_n: Y \to \mathbb{R}$ by $f_n(x) = \min\{f(x), n\}$. Then we have $0 \leqq F + f_n \circ T - f_n$ and

$$\lim_{n \to \infty} (f_n \circ T - f_n) = f \circ T - f.$$

Hence by Fatou's lemma

$$\begin{aligned} \int (F + f \circ T - f)\, d\nu &= \int \liminf (F + f_n \circ T - f_n)\, d\nu \\ &\leqq \liminf \int (F + f_n \circ T - f_n)\, d\nu = \int F\, d\nu < \infty. \end{aligned}$$

This says that $F + f \circ T - f \in L^1_\nu$ and hence $f \circ T - f \in L^1_\nu$. As $|f_n \circ T - f_n| \leqq |f \circ T - f|$, it follows from Lebesgue's theorem of dominated convergence that

$$\int (f \circ T - f)\, d\nu = \lim_{n \to \infty} \int (f_n \circ T - f_n)\, d\nu = 0.$$

Lemma 15. *Set $\bar{g} := g h / (h \circ T)$. If ν is a T-invariant probability on Y, then*

$$\int \log \bar{g}\, d\nu = \int \log g\, d\nu.$$

Proof. It follows from $g(x) \leqq d < 1$ (cf. (a)) that $\log g(x) < 0$ for all $x \in Y$ and from $\bar{g}(x) \geqq 0$ and $\sum_{y \in T^{-1}x} \bar{g}(y) = 1$, which is implied by $Ph = h$, that $\log \bar{g}(x) \leqq 0$ for all $x \in Y$.

If $\int \log g\, d\nu = \int \log \bar{g}\, d\nu = -\infty$, we have nothing to show. If $\int \log g\, d\nu > -\infty$, we have $\log g \in L^1_\nu$. Because of $\log h - \log h \circ T = \log \bar{g} - \log g$, it follows from $\log \bar{g} \leqq 0$ that $(\log h - \log h \circ T)^+ \leqq |\log g|$ and hence from Lemma 14 that $\int (\log h - \log h \circ T)\, d\nu = 0$. This implies $\log \bar{g} \in L^1_\nu$ and $\int \log g\, d\nu = \int \log \bar{g}\, d\nu$. If $\int \log \bar{g}\, d\nu > -\infty$, we apply the same arguments to $\log h^{-1} - \log h^{-1} \circ T = \log g - \log \bar{g}$ and get again $\int \log g\, d\nu = \int \log \bar{g}\, d\nu$.

Theorem 6. *μ is an equilibrium state for $\log g$ on Y.*

Proof. Define the operator $\bar{P}$ by $\bar{P}f(x) = \sum_{y \in T^{-1}x} \bar{g}(y) f(y)$. It follows from $m(Pf) = m(f)$ and $Ph = h$ that $\mu(\bar{P}f) = \mu(f)$ and $\bar{P}1 = 1$. Now we can use a result of Ledrappier (Theorem 1 of [12]), which says that μ is then an equilibrium state for the function $\log \bar{g}$, i.e. $h(\mu) + \mu(\log \bar{g}) = \sup_\nu \{h(\nu) + \nu(\log \bar{g})\}$. But by Lemma 15, we have $\nu(\log \bar{g}) = \nu(\log g)$ for every T-invariant probability ν on Y and hence $h(\mu) + \mu(\log g) = \sup_\nu \{h(\nu) + \nu(\log g)\}$, which says that μ is an equilibrium state for $\log g$.

§2. Let T be a piecewise monotonic transformation on a totally ordered set X, such that the order topology is compact and separable. Let $\varphi \in C(X)$ satisfy $\sum_{i=1}^{\infty} \operatorname{var}_i \varphi < \infty$, where $\operatorname{var}_i \varphi = \sup\{|\varphi(x)-\varphi(y)|: x, y \in A, A \in \mathfrak{P}_0^i\}$. A method, to show that an equilibrium state for φ fits in the situation described at the beginning of Chap. I, is the following one. It is adapted from [22].

We have to modify X a little bit. Set $\tilde{W} = \bigcup_{j=0}^{\infty} T^{-j}\{\lim_{t \uparrow x} T^k(t), \lim_{t \downarrow x} T^k(t): k \geqq 1$ and x is an endpoint of some $I_i \in \mathfrak{P}, 1 \leqq i \leqq N\}$.

Let W be the set of those elements of $\tilde{W}$, which are not an endpoint of an interval, which is open and closed. Now we consider $X' = X \cup W'$, where W' is a copy of W disjoint from X. We extend the order relation to X'. If $y < x < z$ in X, $x \in W$ and x' is the element of W' corresponding to x, then we define $y < x < x' < z$ or $y < x' < x < z$, such that we can extend T continuously to X'. T is then continuous on X' and if one defines $\tilde{W}$ for X', all elements of $\tilde{W}$ are endpoints of intervals, which are open and closed. The order topology of X' is again compact and separable. We extend also φ continuously to X' and define $\tilde{P}$ by $\tilde{P}f(x) = \sum_{y \in T^{-1}x} e^{\varphi(y)} f(y)$ for $x \in X'$.

Because the intervals $T(I_i)$ for $1 \leqq i \leqq N$ are open and closed in X', $\tilde{P}$ maps $C(X')$ into $C(X')$. We can apply the Schauder-Tychonoff-theorem (see Dunford and Schwartz, Linear operators I, p. 456) to the continuous map $m \mapsto \tilde{P}^* m / \tilde{P}^* m(1)$ on the set of all Borel-probability measures, which is a compact, convex subset of the dual of $C(X')$ with w^*-topology. $\tilde{P}^*$ denotes the dual operator of $\tilde{P}$. We get a fixed point m for this map, i.e. $m(\tilde{P}f) = \lambda m(f)$ for all $f \in C(X')$, where $\lambda = \tilde{P}^* m(1)$.

Now set $\tilde{g}(x) = e^{\varphi(x)}/\lambda$. The problem is to show that $\|\tilde{g}_n\|_\infty < 1$ for some n ($\tilde{g}_n(x) = \tilde{g}(x)\,\tilde{g}(Tx) \ldots \tilde{g}(T^{n-1}x)$) and that $m(W') = 0$. To this end we prove two lemmas.

Lemma 16. *Let A be a Borel subset of X' such that T^n/A is monotone for all n. Suppose that for every $j \geqq 1$ there is a B_j, which is a Borel subset of some element of $\mathfrak{P}_0^{k_j}$ and which satisfies $T^{k_j}(B_j) = A$ for some k_j. Furthermore suppose that $B_i \cap B_j = \varnothing$ for $i \neq j$ and that there is an $l \in \mathbb{N}$ such that B_j and A are contained in the same element of $\mathfrak{P}_0^{k_j - l}$ for all $j \geqq 1$. Then $m(A) = 0$.*

Proof. We have $m(B_j) = \lambda^{-k_j} m(\tilde{P}^{k_j} 1_{B_j}) \geqq \inf\{\tilde{g}_{k_j}(y): y \in B_j\} \cdot m(A)$ and similarly $m(A) \leqq \sup\{\tilde{g}_{k_j}(y): y \in A\} \cdot m(T^{k_j} A)$. This implies

$$\begin{aligned} m(B_j)\, m(T^{k_j} A) &\geqq (\inf\{\tilde{g}_{k_j}(y): y \in B_j\} / \sup\{\tilde{g}_{k_j}(y): y \in A\})\, m(A)^2 \\ &\geqq \exp\left(-\sum_{i=0}^{\infty} \operatorname{var}_i \varphi - 2l \|\varphi\|_\infty\right) \cdot m(A)^2 \end{aligned}$$

because A and B_j are contained in the same element of $\mathfrak{P}_0^{k_j - l}$. As the B_j are disjoint, we have $m(B_j) \to 0$ as $j \to \infty$. Of course, $m(T^{k_j} A) \leqq 1$, hence $m(A) = 0$.

Lemma 17. *Let J be a closed interval, which is a subset of some element of $\mathfrak{P}_0^n$. Then*

$$\sup_J \tilde{g}_n \leqq K m(J)/m(T^n J), \quad \text{where } K = \exp\left(\sum_{i=1}^{\infty} \operatorname{var}_i \varphi\right).$$

Proof.
$$m(J) = \lambda^{-n} m(\tilde{P}^n 1_J) \geqq (\inf_J \tilde{g}_n)\, m(T^n J)$$
$$\geqq (\sup_J \tilde{g}_n) \exp\left(-\sum_{i=1}^{n} \operatorname{var}_i \varphi\right) \cdot m(T^n J)$$
$$\geqq K^{-1} (\sup_J \tilde{g}_n)\, m(T^n J).$$

This implies the desired inequality.

We shall apply these lemmas in three special cases in §§ 3–5.

§3. Let $X = \{x_0 x_1 \ldots \in \{1, 2, \ldots, N\}^{\mathbb{N}} : M(x_i x_{i+1}) = 1\}$ be a finite type subshift, which is topologically transitive, i.e. the transition matrix M is irreducible. Furthermore we assume that $h_{\text{top}}(X) > 0$, i.e. X consists not only of a periodic orbit. If we introduce the lexicographic ordering, the shift transformation σ becomes a piecewise monotinic transformation on X. We have here $W = \varnothing$ and $X' = X$. We apply Lemma 16 to $A = \{x\}$ for $x \in X$. Because X is not only a periodic orbit, we can find an inverse image $y_0 y_1 \ldots y_{m-1} x_0 x_1 \ldots$ of $x = x_0 x_1 \ldots$, which is not periodic. As M is irreducible, there is a k_0 such that for any x_j, $x_0 \in \{1, \ldots, N\}$ there are $i_0 = x_j$, $i_1, \ldots, i_k = y_0$ with $k \leqq k_0$ and $M(i_r, i_{r+1}) = 1$ for $0 \leqq r \leqq k = 1$. Then among the points

$$x_0 x_1 \ldots x_j i_1 \ldots i_{k-1} y_0 y_1 \ldots y_{m-1} x_0 x_1 x_2 \ldots \in X \quad \text{for } j = 1, 2, 3, \ldots$$

there are infinitely many different ones. They can be used in Lemma 16 as B_j $(l = m + k_0)$ to get $m(\{x\}) = 0$.

Set $q_k = \max\{m(A) : A \in \mathfrak{P}_0^k\}$. In this case $\mathfrak{P}_0^k$ is the set of all cylinders $[x_0 \ldots x_{k-1}] = \{y_0 y_1 \ldots \in X : y_i = x_i \text{ for } 0 \leqq i \leqq k-1\}$. We have $q_k \to 0$ $(k \to \infty)$, because otherwise one finds an $x \in X$ and an $\varepsilon > 0$, such that each neighbourhood of x contains an $A \in \mathfrak{P}_0^k$ for some k with $m(A) \geqq \varepsilon$. This implies $m(\{x\}) \geqq \varepsilon$, a contradiction.

Next we show that the support Y of m is all of X. There is at least one i with $m([i]) > 0$, because otherwise $m(X) = 0$. Choose some cylinder $[j_0 \ldots j_{n-1}] \subset X$. As above one finds $i_0 = j_{n-1}, i_1, \ldots, i_k = i$ with $M(i_r, i_{r+1}) = 1$. Then

$$m([j_0 \ldots j_{n-1} i_1 \ldots i_k]) \geqq (\inf \tilde{g})^{k+n-1} m([i_k]) > 0$$

and hence $m([j_0 \ldots j_{n-1}]) > 0$. This gives $Y = X$. In particular $m([i]) > 0$ for $1 \leqq i \leqq N$.

Now set $J := [x_0 \ldots x_{n-1}]$. Then $T^n(J) \supset [i]$ for some i, hence

$$m(T^n J) \geqq \min\{m([i]) : 1 \leqq i \leqq N\} =: c > 0.$$

Lemma 17 implies then that $\sup_J \tilde{g}_n \leqq \frac{K}{c} q_n$. As J was an arbitrary element of $\mathfrak{P}_0^n$, this shows that there is an n with $\|\tilde{g}_n\|_\infty < 1$. Setting $g = \tilde{g}_n$, we have $P = \tilde{P}^n/\lambda^n$ and (a) and (b) are satisfied, if we assume additionally that φ and

hence also $\tilde{g}$ and g are of bounded variation. As $Y=X$ and as $h(\mu, T^n)=nh(\mu, T)$, it follows from Theorem 6 that μ is then an equilibrium state for φ. (cf. the results of Chap. 1 of [2]).

§4. We consider the transformation $T: x\mapsto\beta x \pmod 1$ on $[0, 1)$. Let $\chi: [0, 1)\to\{1, \ldots, N\}^{\mathbb{N}}$ (N such that $\beta\leqq N<\beta+1$) be the β-expansion $x\mapsto i_0 i_1, \ldots, i_j$ such that $i_j-1\leqq\beta T^j(x)<i_j$. Set $e_0 e_1 \ldots=\lim_{t\uparrow 1}\chi(t)$ and

$$X=\{x_0 x_1 \ldots\in\{1, \ldots, N\}^{\mathbb{N}}: x_k x_{k+1}\ldots\leqq e_0 e_1 \ldots \text{ for } k\geqq 0\},$$

where $\leqq$ denotes the lexicographic ordering. It follows from Theorem 1 of [6] that $\overline{\chi([0, 1))}=X$. Furthermore $\chi\circ T=\sigma\circ\chi$ and $X\setminus\chi([0, 1))$ is countable. Hence we can consider (X, σ) instead of $([0, 1), T)$, because m will have no atoms. It follows from the definitions, that $e_k e_{k+1}\ldots\neq 111\ldots$ for all $k\geqq 0$.

We proceed as in §3. Let $x\in X'$ and $y=y_0\ldots y_{m-1}x_0 x_1\ldots\in X'$ be an inverse image of x, which is not periodic. We choose a sequence (k_j) and integers z_j with $1\leqq z_j\leqq x_{k_j}$ as follows. If $\sigma^i(y)\neq e_0 e_1\ldots$ for all i then choose k_1 such that $x_i=e_i$ for $0\leqq i\leqq k_1-1$ and $x_{k_1}<e_{k_1}$. Having chosen $k_1<\ldots<k_{n-1}$ choose k_n such that $x_{k_{n-1}+i}=e_i$ for $0\leqq i\leqq k_n-k_{n-1}-1$ and $x_{k_n}<e_{k_n-k_{n-1}}$. Using the definition of X, one checks easily that $x_0\ldots x_{k_i}$ does not end with an initial segment of $e_0 e_1\ldots$. Set $z_j=x_{k_j}$. If $\sigma^i(y)=e_0 e_1\ldots$ for some i choose the $k_j\geqq\max\{i, m\}$ such that $e_{k_j-i}=x_{k_j}>1$. This is possible, because $e_j e_{j+1}\ldots\neq 11\ldots$ for all j. Set $z_j=x_{k_j}-1$. One has again that $x_0\ldots x_{k_j-1}z_j$ does not end with an initial segment of $e_0 e_1\ldots$. We set

$$B_j=\{x_0\ldots x_{k_j-1}z_j y_0\ldots y_{m-1}x_0 x_1\ldots\}\subset X'.$$

Then the requirements of Lemma 16 are satisfied for these B_j, $A=\{x\}$ and $l=m+1$. Hence $m(\{x\})=0$. In particular $m(W')=0$ and we can consider again X instead of X'.

As in §3 we get that $q_k=\sup\{m(A): A\in\mathfrak{P}_0^k\}$ tends to zero as $k\to\infty$. We set $J=[x_0\ldots x_{n-1}]$. If $x_0\ldots x_{n-1}$ does not end with an initial segment of $e_0 e_1\ldots$, then $\sigma^n(J)=X$. Hence by Lemma 17 we get

$$\sup_J\tilde{g}_n\leqq K q_n.$$

If $x_0\ldots x_{n-1}$ ends with an initial segment of $e_0 e_1\ldots$, choose k minimal such that $x_{n-k}\ldots x_{n-1}=e_0\ldots e_{k-1}$. Then $J'=[x_0\ldots x_{n-k-1}(x_{n-k}-1)x_{n-k+1}\ldots x_{n-1}]$ satisfies $\sigma^n(J')=X$. Hence

$$\sup_J\tilde{g}_n\leqq(\sup_{J'}\tilde{g}_n)\cdot\exp\left(\sum_{i=1}^{\infty}\operatorname{var}_i\varphi+2\|\varphi\|_\infty\right)\leqq K q_n\cdot\tilde{K}.$$

Therefore we find again an n such that $\|\tilde{g}_n\|_\infty<1$. We set $g=\tilde{g}_n$. As in §3, one can show that the support Y of m is all of X. Hence by Theorem 6, μ is an equilibrium state for φ.

This means that we have generalized the results of [22] from the class of all Lipschitz continuous φ to the class of φ satisfying $\sum\operatorname{var}_i\varphi<\infty$ and $V_X\varphi<\infty$.

§5. Now let X be the interval $[0, 1]$ and T any piecewise monotonic transformation on X with positive topological entropy. For φ we take the constant function zero. Then $\tilde{g}$ is the constant function λ^{-1}. We have to show that $\lambda > 1$. Then we can set $g = \tilde{g}$ and (a) and (b) are satisfied.

To this end we need some results of [6] (cf. also [7]). Let D be the following set of subintervals of $I_i \subset X'$ $(1 \leqq i \leqq N)$, on which T is monotone. An interval K is called a successor of an interval $J \subset I_i$, if K is one of the nonempty intervals among $T(J) \cap I_j$ for $1 \leqq j \leqq N$. We define D in the following way: D contains I_i for $1 \leqq i \leqq N$ and with every interval also its successors. Because all images of I_i's are open and closed subsets of X', all elements of D are open and closed. In [6] the elements of D are subsets of a shift space isomorphic to $([0, 1], T)$, but this makes no difference. Furthermore we define a $D \times D$-matrix M with entries 0 and 1 by $M(J, K) = 1$ if and only if K is a successor of J, for $J, K \in D$. It happens that elements of D are successors of more than one element of D.

It is shown in [6] that every irreducible submatrix L of M corresponds to a closed T-invariant subset Ω of $[0, 1]$. If $F \subset D$ is the index set of L, i.e. $L = M/F$, then $I = \cup \{J : J \in \tilde{F}\}$, where $\tilde{F} := \{J \in D$: there are $J_0, J_1, \ldots, J_n = J$ with $J_0 \in F$ and $M(J_i, J_{i+1}) = 1\}$, and $I' = \cup \{J : J \in \tilde{F} \setminus F\}$ are T-invariant sets, which are finite unions of intervals and $\Omega = \bigcap_{i=0}^{\infty} T^{-i}(\overline{I \setminus I'})$. We choose an Ω, such that $r(L) = r(M)$, where r denotes the spectral radius of the l^1-operators L and M. Because I is T-invariant, T/I is again a piecewise monotonic transformation. Hence we can take I as X and assume that m is concentrated on I.

Lemma 18. (i) *If* $M(J_i J_{i+1}) = 1$ *for* $0 \leqq i \leqq k-1$ *then* $T^k(J_0 \cap \ldots \cap T^{-k} J_k) = J_k$.

(ii) *For every* $J \in F$, *we have* $m(J) > 0$.

Proof. (i) is proved by induction. The induction step is as follows

$$\begin{aligned} T^k(J_0 \cap T^{-1} J_1 \cap \ldots \cap T^{-k} J_k) &= T^{k-1}(TJ_0 \cap J_1 \cap \ldots \cap T^{-k+1} J_k) \\ &= T^{k-1}(J_1 \cap \ldots \cap T^{-k+1} J_k) \end{aligned}$$

because $TJ_0 \supset J_1$, since J_1 is a successor of J_0.

In order to show (ii) note that there is a $K \in \tilde{F}$ with $m(K) > 0$, since $m(I) = 1$. We can find $J_0 = J, J_1, \ldots, J_k = K$ with $M(J_i, J_{i+1}) = 1$ and $J \in F$. As L is irreducible, we can do this for all $J \in F$. Set $Z = J_0 \cap T^{-1} J_1 \cap \ldots \cap T^{-k} J_k \subset J$. Then $m(Z) = \lambda^{-k} m(\tilde{P}^k 1_Z) = \lambda^{-k} m(T^k Z) = \lambda^{-k} m(K) > 0$, using (i). Hence $m(J) > 0$ proving (ii).

It follows from §3 of [6] that $r(M) = \exp h_{\text{top}}(X)$. Thus $r(L) = r(M) > 1$, and L consists not only of a cycle. Therefore we can find $J_0, J_1, \ldots, J_n = J_0 \in F$ and $K_0 = J_0$, $K_1, \ldots, K_k = J_r \in F$ for some $n, k \geqq 0$ and $0 \leqq r \leqq n-1$ such that $M(J_i, J_{i+1}) = M(K_i, K_{i+1}) = 1$. Now set

$$A_k = \bigcap_{i=0}^{k-1} T^{-ni} \left(\bigcap_{j=0}^{n-1} T^{-j} J_j \right), \qquad A = \bigcap_{k=0}^{\infty} A_k,$$

$$B_j = \bigcap_{i=0}^{jn} T^{-i} J_{i(\text{mod } n)} \cap \bigcap_{i=1}^{k} T^{-jn-i} K_i \cap \bigcap_{i=1}^{\infty} T^{-jn-k-i} J_{r+i(\text{mod } n)}.$$

The B_j are disjoint, because $T^{jn}B_j \subset K_1$, $T^{jn}B_{j+m} \subset J_1$ and K_1 and J_1 are disjoint as different successors of J_0. Furthermore $T^{jn+k+n-r}B_j = A$ and A and B_j are contained in the same element of $\mathfrak{P}_0^{jn}$. Hence it follows from Lemma 16 that $m(A)=0$. Now we apply Lemma 17 and get

$$\sup_{A_k} \tilde{g}_{kn} \leqq K m(A_k)/m(T^{kn}A_k).$$

Because of $T^{kn}A_k = T(J_{n-1}) \supset J_0$ ((i) of Lemma 18) and $m(J_0)>0$ ((ii) of Lemma 18), the right hand side of this inequality tends to zero. Hence there is a k with $\sup_{A_k} \tilde{g}_{kn} < 1$. Because of $\tilde{g}_{kn} \equiv \lambda^{-kn}$, we get $\lambda > 1$, the desired result. It follows now from Lemma 2 that $m(\{x\})=0$ for all $x \in X'$. Hence $m(W')=0$.

A similar proof as that of (ii) of Lemma 18 shows that the support Y of m contains Ω. As $h_{\text{top}}(\Omega) = \log r(L) = \log r(M) = h_{\text{top}}(X)$, it follows from Theorem 6, that μ is a measure with maximal entropy.

Acknowledgement. We thank the referee for some comments which simplified the proof of Lemma 5.

References

1. Babbel, B.: Diplomarbeit, Institut für Mathematische Statistik und Wirtschaftsmathematik der Universität Göttingen 1980
2. Bowen, R.: Equilibrium states and the ergodic theory of Anosov diffeomorphisms. Lecture Notes in Mathematics **470**. Berlin-Heidelberg-New York: Springer 1975
3. Bowen, R.: Bernoulli maps of an interval. Israel J. Math. **28**, 298–314 (1978)
4. Dunford, N., Schwartz, J.T.: Linear Operators, Part I. New York: Interscience 1957
5. Gordin, M.I.: The central limit theorem for stationary processes. Soviet Math. Dokl. **10**, 1174–1176 (1969)
6. Hofbauer, F.: On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. Israel J. Math. **34**, 213–237 (1979)
7. Hofbauer, F.: On intrinsic ergodicity of piecewise monotonic transformations with positive entropy II. Israel J. Math. (To appear)
8. Ibragimov, I.A., Linnik, Y.V.: Independent and stationary sequences of random variables. Groningen: Wolters-Noordhoff 1971
9. Ionescu-Tulcea, C., Marinescu, G.: Théorie ergodique pour des classes d'opérations non complètement continues. Ann. of Math. (2) **52**, 140–147 (1950)
10. Keller, G.: Un théorème de la limite centrale pour une classe de transformations monotones per morceaux. C.R. Acad. Sci. Paris Sér. A **291**, 155–158 (1980)
11. Lasota, A., Yorke, J.: On the existence of invariant measures for piecewise monotonic transformations. Trans. Amer. Math. Soc. **186**, 481–488 (1973)
12. Ledrappier, F.: Principe variationnel et systemes dynamiques symboliques. Z. Wahrscheinlichkeitstheorie verw. Gebiete **30**, 185–202 (1974)
13. Li, T., Yorke, J.: Ergodic transformations from an interval into itself. Trans. Amer. Math. Soc. **235**, 183–192 (1978)
14. Li, T., Yorke, J.: Iterating piecewise expanding maps: Asymptotic dynamics of probability densities. Preprint
15. Philipp, W., Stout, W.: Almost sure invariance principles for partial sums of weakly dependent random variables. Mem. Amer. Math. Soc. **161** (1975)
16. Ratner, M.: Bernoulli flows over maps of the interval. Israel J. Math. **31**, 298–314 (1978)
17. Rohlin, V.A.: Exact endomorphisms of Lebesgue spaces. Amer. Math. Soc. Transl. (2) **39**, 1–36 (1964)

18. Schäfer, H.H.: Topological Vector Spaces. New York: MacMillan 1966
19. Volkonski, V.A., Rozanov, Y.A.: Some limit theorems for random functions II. Theor. Probability Appl. **6**, 186–198 (1961)
20. Wagner, G.: The ergodic behavior of piecewise monotonic transformations. Z. Wahrscheinlichkeitstheorie und verw. Gebiete **46**, 317–324 (1979)
21. Walters, P.: Ergodic theory. Lecture Notes in Mathematics **458**. Berlin-Heidelberg-New York: Springer 1975
22. Walters, P.: Equilibrium states for β-transformations and related transformations. Math. Z. **159**, 65–88 (1978)
23. Wong, S.: Some metric properties of piecewise monotonic mappings of the unit interval. Trans. Amer. Math. Soc. **246**, 493–500 (1978)
24. Wong, S.: A central limit theorem for piecewise monotonic mappings of the unit interval. Ann. Probability **7**, 500–514 (1979)

Received March 25, 1981; received in final form December 7, 1981

Math. Z. 180, 141–149 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On the Degrees of Primitive Permutation Groups

Peter J. Cameron[1], Peter M. Neumann[2] and David N. Teague[3]

[1] Merton College, Oxford OX1 4JD, England
[2] The Queen's College, Oxford OX1 4AW, England
[3] Brasenose College, Oxford OX1 4AJ, England

1. Introduction

The recently announced classification of the finite simple groups has made possible some striking results about permutation groups which had previously defied all attempts at proof. Some of these are discussed in [1]. It is our purpose here to add another result: *for almost all integers n, the only primitive groups of degree n are the symmetric and alternating groups.*

Let E be the set of natural numbers n for which there exists a primitive group of degree n other than S_n and A_n. If $n \leqq 4$ then $n \notin E$ for trivial reasons. It was known to Émile Mathieu by 1861 that $n \in E$ for $5 \leqq n \leqq 33$. He had been searching quite systematically (see [11, p. 242]) for multiply transitive groups other than S_n, A_n, and he observed [11, p. 274, footnote] explicitly that for $n < 34$ (and, one must add, $n > 4$) there is at least one such group for each degree. (See also Jordan [9] for an announcement of a contemporary but slightly inaccurate enumeration of all primitive groups of degrees $n \leqq 17$, and Sims [16] for a more recent enumeration up to degree 20.) It is a non-trivial exercise in permutation group theory, an exercise which could probably not have been completed without inordinate labour even 100 years after Mathieu's paper appeared, to show that S_{34}, A_{34} are the only primitive groups of degree 34. The only numbers from 5 to 50 that are not in E are 34, 39, 46. Numbers of the form $4p$, where $p \geqq 23$, p and $\frac{1}{2}(p-1)$ are prime, and $4p-1$ is not a prime (the first few such are 92, 188, 236, ...) do not lie in E; similarly, if

$$n = 2p = 4q + 2 = r + 3 > 22,$$

where p, q, r are primes (that is, if $n = 46, 166, 214, \ldots$), then $n \notin E$. These facts, and several more of the same nature which are proved in [12] and [13], strongly suggest that $\mathbb{N} - E$ is infinite. But they do no more than that: up to now the set of numbers n for which it was known that the only primitive groups of degree n are S_n and A_n was of density zero. Our theorem reverses this situation and shows that E has density 0, hence $\mathbb{N} - E$ has density 1.

0025-5874/82/0180/0141/$01.80

Let $e(x) := \#\{n \in E \mid n \leqq x\}$ and let $\pi(x)$ denote, as usual, the number of prime numbers $p \leqq x$.

Theorem. $e(x) = 2\pi(x) + (1+\sqrt{2})\, x^{\frac{1}{2}} + O(x^{\frac{1}{2}}/\log x)$
$\sim 2x/\log x.$

There are primitive groups other than S_n, A_n of degree $n \geqq 5$
if n is prime (the cyclic group Z_n acting regularly);
if $n = p+1$ where p is prime (the group $PSL(2,p)$ in its natural action);
if $n = m^2$ (the wreath product $S_m \operatorname{wr} S_2$ with the product action);
if $n = \frac{1}{2}m(m-1)$ (the symmetric group S_m acting on unordered pairs).
Moreover, if p is a prime and $p > 5$ then neither p nor $p+1$ can be a square or a triangular number $\frac{1}{2}m(m-1)$; and a coincidence $m_1^2 = \frac{1}{2}m_2(m_2-1)$ gives the Pell equation $u^2 - 2v^2 = 1$ which has only $O(\log x)$ solutions below a given bound x. Therefore we certainly have

$$e(x) \geqq 2\pi(x) + (1+\sqrt{2})\, x^{\frac{1}{2}} - O(\log x).$$

The remainder of the paper is devoted to a proof that

$$e(x) \leqq 2\pi(x) + (1+\sqrt{2})\, x^{\frac{1}{2}} + O(x^{\frac{1}{2}}/\log x).$$

2. First Steps

In this section we write E as a union of four subsets, and we estimate the density of three of them. We postpone study of the fourth since it is the most interesting and most difficult. Let

$E_1 := \{p \mid p \text{ prime}, p \geqq 5\}$,

$E_2 := \{m^k \mid m \geqq 2,\ k \geqq 2,\ m^k > 4\}$,

$E_3 := \{n \mid \text{there is a non-abelian simple group of order } n\}$,

$E_4 := \{n \mid \text{there is a primitive group of degree } n \text{ whose socle is a non-abelian simple group other than } A_n\}$.

As we have already seen, $E_1 \subseteq E$; also $E_2 \subseteq E$ because if $m > 2$ and $k \geqq 2$ then $S_m \operatorname{wr} S_k$ with the product action is primitive of degree m^k, while the affine group $AGL(k,2)$ is primitive of degree 2^k; and $E_3 \subseteq E$ because if S is a non-abelian simple group then $S \times S$ acts primitively on S by the rule (g,h): $x \to g^{-1}xh$; finally, $E_4 \subseteq E$ by definition. The following lemma is an immediate consequence of a theorem of O'Nan and Scott [15], although what we need of that theorem is hardly more than what is to be found already in Jordan's thesis (see [8]). Since we need so little we shall, for completeness, sketch a proof.

Lemma. $E = E_1 \cup E_2 \cup E_3 \cup E_4$.

Proof. Let G be a primitive group of degree n that is not S_n or A_n, and let N be a minimal non-trivial normal subgroup of G. If the centraliser $C_G(N)$ is non-

trivial then $C_G(N)$, as well as N, would be transitive and both these groups would act regularly: thus $n=|N|$ and, as N is a direct power of some simple group, we have $n\in E_1\cup E_2\cup E_3$.

Suppose therefore that $C_G(N)=\{1\}$. Then N is non-abelian and is the only minimal non-trivial normal subgroup of G—so N is the socle of G. Now N is a direct product $X_1\times X_2\times\ldots\times X_k$ of subgroups all isomorphic to the same non-abelian simple group S, and the factors are permuted transitively under conjugation by elements of G. Let α be an element of the set Ω on which G acts, and let H be its stabiliser G_α. Let $M:=N_\alpha=N\cap H$, and let Y_i be the projection of M into X_i. Notice that N is transitive on Ω, so $H\cdot N=G$ and therefore H acts transitively on $\{X_1,\ldots,X_k\}$, and also on $\{Y_1,\ldots,Y_k\}$ by conjugation. If now $Y_i=X_i$ for some i then $Y_i=X_i$ for all i and M is a subdirect power of S. Then M is isomorphic to a direct power of S, so we have $n=|N:M|=|S|^{k'}$ for some k', and $n\in E_2\cup E_3$. Suppose therefore that $Y_i<X_i$ for all i. Since H normalises $Y_1\times\ldots\times Y_k$ the set $H\cdot(Y_1\times\ldots\times Y_k)$ is a subgroup of G, and from the maximality of H it follows that $Y_1\times\ldots\times Y_k\leqq H$. Hence $M=Y_1\times\ldots\times Y_k$. But now if $m:=|X_1:Y_1|$ then $m=|X_i:Y_i|$ for all i and $n=|N:M|=m^k$. Thus either $k>1$ and $n\in E_2$ or $k=1$ and $n\in E_4$. This shows that $E\subseteq E_1\cup E_2\cup E_3\cup E_4$ and, as we have already explained the reverse inclusion, the proof is complete.

Now let
$$e_i(x):=\#\{n\in E_i\mid n\leqq x\}.$$
Then certainly
$$e(x)\leqq e_1(x)+e_2(x)+e_3(x)+e_4(x).$$

Clearly, $e_1(x)=\pi(x)-2$. Also,
$$e_2(x)\leqq([x^{\frac{1}{2}}]-1)+([x^{\frac{1}{3}}]-1)+([x^{\frac{1}{4}}]-1)+\ldots$$
where $[y]$ denotes the integer part of y. The sum has fewer than $\log_2(x)$ non-zero terms so $e_2(x)\leqq x^{\frac{1}{2}}+x^{\frac{1}{3}}+x^{\frac{1}{4}}\log_2(x)$ and so $e_2(x)=x^{\frac{1}{2}}+O(x^{\frac{1}{3}})$.

To estimate $e_3(x)$, and later to deal with $e_4(x)$, we need to invoke the classification of finite simple groups. The Classification Theorem asserts that the non-abelian finite simple groups are of three kinds: alternating groups, groups of Lie type, and known sporadic groups. The groups of Lie type may be further subdivided into classical groups of which there are six families each parametrised by a rank r and a field order q, and exceptional groups, of which there are ten families, each parametrised by a field order q. There are just 26 known sporadic groups. Since finite collections of groups do not affect the asymptotic form of $e(x)$ we may ignore these.

If $|A_m|\leqq x$ and $m\geqq 5$ then certainly $2^m<x$, so $m<\log_2(x)$. For groups of Lie type one may inspect the order formulae (see, for example, [5, p. 491] or [2, pp. 144, 262]) to see that, with the exception of $PSL(2,q)$ whose order is $\frac{1}{2}q(q^2-1)$ or $q(q^2-1)$, a group of rank r defined over $GF(q)$ has order at least $2^{\frac{1}{2}r(r+1)}$ and at least $q^2(q^2+1)(q-1)$. Thus if G is such a group and $|G|\leqq x$ then certainly $r<\log_2(x)$ and $q<(2x)^{\frac{1}{4}}$ unless G is $PSL(2,q)$. Consequently
$$e_3(x)<\log_2(x)+(2x)^{\frac{1}{4}}+16\cdot\log_2(x)\cdot(2x)^{\frac{1}{4}}+26=O(x^{\frac{1}{4}}).$$
In fact, $e_3(x)\sim\pi((2x)^{\frac{1}{4}})\sim 3.2^{\frac{1}{4}}x^{\frac{1}{4}}/\log x$, but we do not need such precision.

We are left with the problem of estimating $e_4(x)$. To do this we subdivide E_4 into three sets E_{4A}, E_{4L}, and E_{4S} corresponding to primitive groups whose socles are alternating groups, groups of Lie type, and known sporadic simple groups. Notice that if G_0 is the socle of G then $G_0 \leqq G \leqq \mathrm{Aut}(G_0)$, so that each simple group can contribute only finitely many numbers to E_4. In particular, E_{4S} is finite and may be ignored.

3. Alternating Groups

For primitive groups in which the socle is an alternating group we have the estimate:

Lemma. *If* $e_{4A}(x) := \#\{n \in E_{4A} \mid n \leqq x\}$ *then* $e_{4A}(x) = \sqrt{2}x^{\frac{1}{2}} + O(x^{\frac{1}{3}})$.

Proof. Suppose that $A_m \trianglelefteq G \leqq \mathrm{Aut}(A_m)$ and that $n \leqq x$. We may suppose also that $m > 6$ so that G is A_m or S_m. Let H be a stabiliser in G, so that $n = |G:H|$.

Consider first the cases where H is intransitive as a permutation group on $\{1, 2, \ldots, m\}$. Then, since H is maximal in G it must be the stabiliser of an r-subset, where $1 < r < \frac{1}{2}m$, and $n = \binom{m}{r}$. The contribution to E_{4A} from groups of this kind consists therefore of binomial coefficients. Now if $\binom{m}{2} \leqq x$ then $m < (2x)^{\frac{1}{2}} + 1$; if $\binom{m}{3} \leqq x$ then $m < (6x)^{\frac{1}{3}} + 2$; and if $\binom{m}{r} \leqq x$ for $4 \leqq r < \frac{1}{2}m$ then $(m-3)^4/24 < \binom{m}{4} \leqq \binom{m}{r} \leqq x$ so that $m < (24x)^{\frac{1}{4}} + 3$, and also $2^r < \binom{2r}{r} < \binom{m}{r} \leqq x$, whence $r < \log_2(x)$. Thus these groups contribute at most $(2x)^{\frac{1}{2}} + (6x)^{\frac{1}{3}} + {} + O((24x)^{\frac{1}{4}} \cdot \log_2(x))$, that is, $2 \cdot x^{\frac{1}{2}} + O(x^{\frac{1}{3}})$ to $e_{4A}(x)$.

Suppose next that H is transitive but imprimitive on $\{1, 2, \ldots, m\}$. Then H must be the stabiliser of a system consisting of k blocks each of length l, where $kl = m$ and $1 < k \leqq \frac{1}{2}m$, and the degree n of G is $m!/(l!)^k \cdot k!$. If $k = 2$ we have

$$n = \frac{m}{l} \cdot \frac{m-1}{l-1} \cdot \ldots \frac{l+1}{1} \cdot \frac{1}{2} > 2^{\frac{1}{2}m} \quad \text{(since } l \geqq 3\text{)}.$$

If $k \geqq 3$ we have

$$n = \frac{m(m-1)\ldots(m-k+1)}{l^k} \cdot \frac{(m-k)\ldots(m-2k+1)}{(l-1)^k} \cdot \ldots \cdot \frac{(2k)\ldots(k+1)}{2^k}$$

$$> 2^{m-k} \geqq 2^{\frac{1}{2}m}.$$

Since $n \leqq x$ we have $m \leqq 2\log_2(x)$. For each value of m the number of possibilities for n is the same as the number of proper divisors k of m and this is certainly less than m. The total contribution to $e_{4A}(x)$ from groups of this kind is therefore less than $(2\log_2(x))^2$.

Suppose finally that H is primitive on $\{1, 2, \ldots, m\}$. Then $|H| < 4^m$ (Praeger and Saxl [14]), so $n > \frac{1}{2}m!/4^m$ and for any positive number c we have $n > c^m$ if m is sufficiently large. Taking $c := 2^8$ and ignoring a (large but) finite number of degrees we have $2^{8m} < x$, and so $m < \frac{1}{8}\log_2(x)$. For each value of m there are

fewer than 4^m possibilities for the order of $H \cap A_m$, hence at most 4^m possible degrees n. Consequently the contribution to $e_{4A}(x)$ from groups of this kind is at most

$$\tfrac{1}{8}\log_2(x)\cdot 4^{\frac{1}{2}\log_2(x)}+O(1), \quad \text{and this is} \quad O(x^{\frac{1}{2}}\log x).$$

4. Groups of Lie Type

Our estimate for primitive groups with simple socles of Lie type is contained in the

Lemma. *If* $e_{4L}(x) := \#\{n \in E_{4L} | n \leqq x\}$ *then* $e_{4L}(x) = \pi(x) + O(x^{\frac{1}{2}}/\log x)$.

For the proof we shall need the following facts:

(A) *If* G_0 *is a classical simple group of rank* r *defined over* $GF(q)$ *that has a proper subgroup of index* n *then, with finitely many exceptions,* $n \geqq q^r$. (See Cooperstein [3] and references quoted there. In fact, the only exception is $PSL(2, 9)$ acting as a group of degree 6.)

(B) *If* r *is sufficiently large, if* G_0 *is a classical special linear group (i.e. a linear group consisting of transformations of determinant* 1*) of rank* r *defined over* $GF(q)$*, and if* H_0 *is an irreducible proper subgroup of index* n *in* G_0*, then, except in the case mentioned below,* $|G_0| < n^5$.

(See Kantor [10]: $r > 100$ is certainly sufficient. In our statement the case where $G_0 = PSp(2r, 2^d)$ must be excluded. However, $PSp(2r, 2^d) \cong P\Omega(2r+1, 2^d)$, and Kantor's theorem applies if $G_0 = P\Omega(2r+1, 2^d)$ and H_0 fixes no proper subspace other than the radical.)

(C) *For any* $\delta > 0$ *the number of divisors of a positive integer* m *is* $O(m^\delta)$. (See, for example, Hardy and Wright [6, p. 260].)

Our notation will be this: G is a primitive permutation group of degree n (where $n \leqq x$) whose socle G_0 is a simple group of Lie type of rank r over $GF(q)$ where q is a power of the prime number p. Note that for groups of twisted type we take q to be the size of the fixed field of the relevant automorphism—thus for example we write $PSU(s+1, q)$ for the group ${}^2A_s(q^2)$, that is, for the unitary subgroup of $PSL(s+1, q^2)$ (for which $r = [\frac{1}{2}(s+1)]$). Let $H := G_\alpha$, the stabiliser of a point α of the set of size n on which G acts, and let $H_0 := H_\alpha = H \cap G_0$. Thus $n = |G:H| = |G_0:H_0|$.

We begin the proof with the observation that *if* y, ε *are fixed positive numbers then the contribution to* $e_{4L}(x)$ *from groups* G *with socle* G_0 *for which* $p \leqq y$ *and* $r \leqq y$ *is* $O(x^\varepsilon)$. For, any group of Lie type over $GF(q)$ has a section isomorphic to $PSL(2, q)$ and if $q > 9$ then this has no faithful permutation representation of degree less than q. Thus if $q > 9$ then $q \leqq n \leqq x$ and the number of possibilities for q is therefore at most $y \cdot \log_2(x)$. Given q and r we have $|G_0| < q^{4r^2}$ (see [5, p. 491] or [2, pp. 144, 262]) and so, given q, r and the type of G_0 the number of possibilities for n is $O(q^{4r^2\delta})$ for any $\delta > 0$ by Fact (C). Consequently if $r \leqq y$ and $p \leqq y$ then the number of possibilities for n is at most

$16 \cdot y^2 \log_2(x) \cdot O(x^{4y^2\delta})$. Taking δ to be $\varepsilon/(4y^2+1)$ shows that the contribution from these groups to $e_{4L}(x)$ is $O(x^\varepsilon)$ as we claimed.

The rest of the proof of our lemma is organised as an examination of various possibilities case by case, working from groups of small rank to those of higher rank. We begin with those which contribute significantly to $e(x)$:

Case 1: $G_0 = PSL(2,q)$. In this case $PSL(2,q) = G_0 \trianglelefteq G \leqq P\Gamma L(2,q)$ and it happens that if $q > 11$ then maximal subgroups of G intersect G_0 in maximal subgroups. The list of these is well known (see, for example, [4, Chap. XII]). If q is odd, as we may assume by the observation above, then they are

affine subgroups of index $q+1$;
dihedral subgroups of index $\frac{1}{2}q(q-1)$;
dihedral subgroups of index $\frac{1}{2}q(q+1)$;
tetrahedral, octahedral or icosahedral subgroups of index $q(q^2-1)/2d$ where d is 12, 24, or 60;
subgroups $PSL(2,q_0)$ or $PGL(2,q_0)$ of index $q(q^2-1)/dq_0(q_0^2-1)$, where $q = q_0^r$, r is prime, and d is 1 or 2.

The numbers $\frac{1}{2}q(q-1)$, $\frac{1}{2}q(q+1)$ have already been counted in $e(x)$ as binomial coefficients but in any case their contribution to $e_{4L}(x)$ is only $O(\pi(\sqrt{2x}))$; numbers of the form $q(q^2-1)/2d$ contribute at most $O(\pi((120 \cdot x)^{\frac{1}{3}}))$; and numbers of the form $q(q^2-1)/dq_0(q_0^2-1)$ contribute $O(\pi(x^{\frac{1}{3}}))$. Since the number of prime-powers less than x is $\pi(x)+\pi(\sqrt{x})+o(x^{\frac{1}{3}})$ and since, by the Prime Number Theorem (or much easier, but weaker versions of it (see, for example [6, pp. 9, 345])) $\pi(\sqrt{x}) = O(\sqrt{x}/\log x)$, the contribution from $PSL(2,q)$ to $e_{4L}(x)$ is $\pi(x)+O(\sqrt{x}/\log x)$. In fact, the most significant contribution comes from the groups $PSL(2,p)$ of degree $p+1$ that we mentioned in the introduction.

Case 2: *Other groups of small rank, say* $r \leqq 100$. Consider first permutation representations for which $n > q^3$. The number of possibilities for q is $O(\pi(x^{\frac{1}{3}}))$, which is $O(x^{\frac{1}{3}}/\log x)$. The number of possibilities for r is 100 and the number of types of group is at most 16, so the total number of possibilities for G_0 is $O(x^{\frac{1}{3}}/\log x)$. Now $|G_0| < q^{4r^2} < q^{40000}$ and, as n divides $|G_0|$, the number of possibilities for n when G_0 is given is $O(|G_0|^\delta)$ for any $\delta > 0$, which is $O(x^{40000\delta/3})$ since $q < n^{\frac{1}{3}} \leqq x^{\frac{1}{3}}$. If we take $\delta := 1/80000$ than we see that the total number of possibilities for n is $O(x^{\frac{1}{3}}/\log x) \cdot O(x^{\frac{1}{2}})$, and this is $O(x^{\frac{1}{2}}/\log x)$.

We are left with situations where $n \leqq q^3$. This only arises for groups of small rank. For the classical groups what we need is a slight refinement of the result of Cooperstein and others quoted above as Fact (A): if we ignore the groups $PSL(2,q)$ and finitely many others then the only groups that arise are $PSL(3,q)$ of degree q^2+q+1. But these contribute only $O(\pi(\sqrt{x}))$, that is, only $O(\sqrt{x}/\log x)$ to $e_{4L}(x)$. For the exceptional groups we need a result corresponding to Fact (A). The groups ${}^2B_2(q)$, ${}^2F_4(q)$, and ${}^2G_2(q)$ need not concern us because they are defined only over fields of characteristic 2, 2, 3 respectively. Now $PSU(3,q) < G_2(q) < {}^3D_4(q)$ and $PSU(3,q)$ has no faithful representations of degree less than q^3+1 if $q > 2$: therefore $G_2(q)$ and ${}^3D_4(q)$ do not have faithful representations of degree less than q^3+1 if $q > 2$. This same argument, based

on the inclusions $PSp(6,q)<F_4(q)$ and $PSL(6,q)<E_6(q)<E_7(q)<E_8(q)$ proves that the groups $F_4(q)$, $E_6(q)$, $E_7(q)$, $E_8(q)$ do not have faithful permutation representations of degree less than $q^5+q^4+q^3+q^2+q+1$; and the inclusion $PSU(6,q)<{}^2E_6(q)$ shows that there is no faithful representation of ${}^2E_6(q)$ of degree less than $q^9+q^7+q^5+q^4+q^2+1$. Thus at most finitely many of the exceptional groups remain, and these do not concern us.

Case 3: *Groups with* $r>100$. Such groups are of one of the six classical types and Facts (A) and (B) quoted above apply. From (A) we see that $r\leqq\log_2(x)$ and $q<x^{1/100}$. Therefore the number of possibilities for G_0 is at most $6x^{1/100}\cdot\log_2(x)$. Now $G_0\trianglelefteq G\leqq\operatorname{Aut}G_0$. For our purposes we may replace G_0 with the linear group of which it is the projective version, so that there is a vector space V over $GF(q)$ (or over $GF(q^2)$ in the unitary case) on which G_0 naturally acts. By a theorem of Steinberg (see [2, p. 211]) $\operatorname{Aut}G_0$ may be identified with the appropriate group of semi-linear transformations of V except when G_0 is $SL(r+1,q)$, when $\operatorname{Aut}G_0$ also contains dualities.

Suppose for the moment that H_0 (which, recall, was a point-stabiliser in G_0, and which is now to be thought of as lifted to a linear group) is reducible in its action on V. The maximality of H in G ensures that H_0 is either the stabiliser of a subspace U of V or the stabiliser of an ordered pair (U_1, U_2) of subspaces where

$$\dim U_1+\dim U_2=\dim V \quad\text{and}\quad \dim U_1\leqq\dim U_2.$$

This latter situation can only occur if G_0 is $SL(r+1,q)$ and $U_2=U_1^\sigma$ where σ is some element of G inducing a duality of the projective space of V. Then, since H_0 fixes $U_1\cap U_2$ and U_1+U_2, which are interchanged by σ, the maximality of H in G implies that either $U_1\cap U_2=\{0\}$ or $U_1\leqq U_2$. If G_0 is $SL(r+1,q)$ there are r different permutation representations arising where H_0 is the stabiliser of a subspace, and these contribute $[\frac{1}{2}(r+1)]$ different degrees since there are as many subspaces of dimension k as of dimension $r+1-k$. Also, there are $[\frac{1}{2}(r+1)]$ inequivalent choices for (U_1, U_2) in each of the cases $U_1\cap U_2=\{0\}$, $U_1\leqq U_2$. Thus when G_0 is $SL(r+1,q)$ there are in all at most $\frac{3}{2}(r+1)$ possible degrees.

In the cases where G_0 preserves a non-singular bilinear or an hermitian form on V, and G_0 is not an orthogonal group over a field of characteristic 2, when H_0 stabilises U it also stabilises $U^\perp$, hence $U\cap U^\perp$. So again we may suppose that $U\cap U^\perp=\{0\}$ (U is anisotropic) or $U\leqq U^\perp$ (U is totally isotropic). In each case the degree n is the number of subspaces of the appropriate type. The number of possible values for n is therefore at most the number of inequivalent choices for the subspace U. There is one choice of totally isotropic U for each dimension from 1 up to r; if U is anisotropic we may assume that $\dim U\leqq\dim U^\perp$ and the number of inequivalent subspaces of each dimension is at most 2. Consequently there are certainly no more than $3r$ possible degrees (in fact, the maximum number is $\frac{5}{2}r+2$ in the orthogonal case and $2r$ in other cases).

This reasoning covers all cases except those where G_0 is an orthogonal group and q is a power of 2. If $G_0=\Omega(2r+1,2^d)$ then the appropriate bilinear form is singular, its radical $V^\perp$ has dimension 1, and we may suppose that

$U \cap U^\perp$ is $\{0\}$ or $V^\perp$, or that $U \leqq U^\perp$. In all cases (i.e. if G_0 is $\Omega^+(2r, 2^d)$, $\Omega(2r+1, 2^d)$ or $\Omega^-(2r+2, 2^d)$), if $U \cap U^\perp$ is $\{0\}$ or $V^\perp$ then the restriction of the quadratic form to U is non-singular, so there are at most two inequivalent choices of subspace of any dimension up to $r+1$. If $U \leqq U^\perp$ then the restriction of the quadratic form to U is a semilinear form whose kernel is fixed by H_0, so we may suppose that either U is totally singular or $\dim U = 1$; but in this latter case U is non-singular. Thus the number of inequivalent permutation representations is again at most $\frac{5}{2}(r+1)$.

In all cases where H_0 is an irreducible subgroup of G_0 we have fewer than $3r$ possibilities for the degree n, and so the contribution to $e(x)$ from groups of this kind is at most $O(\log x)$.

Finally, we must consider the cases where H_0 is irreducible and G_0 is not $PSp(2r, 2^d)$, or where $G_0 = \Omega(2r+1, 2^d)$ (which, recall, is isomorphic to $PSp(2r, 2^d)$) and H_0 fixes no proper subspace other than the radical. In this case $|G_0| < n^5 \leqq x^5$ by Fact (B), and, as the number of possibilities for the degree n is less than the number of divisors of $|G_0|$, it is $O(x^{5\delta})$ for any $\delta > 0$ by Fact (C). In all, the contribution to $e_{4L}(x)$ from groups of rank $r > 100$ is therefore at most

$$6 \cdot x^{1/100} \cdot \log_2(x)\,(O(\log x) + O(x^{5\delta})).$$

Taking δ to be $1/11$ we see that this contribution is $O(x^{\frac{1}{2}}/\log x)$, which completes the proof of the lemma and of the theorem.

5. A Last Word

The analysis can be refined to give further terms in our description of $e(x)$: the next few are $2\pi(x^{\frac{1}{2}})$ from numbers of the form p^2+1 (coming from the groups $PSL(2, p^2)$) and p^2+p+1 (from $PSL(3, p)$), then $(1+6^{\frac{1}{3}})x^{\frac{1}{3}}$ from cubes and binomial coefficients $\binom{m}{3}$, so that

$$e(x) = 2\pi(x) + (1+\sqrt{2})x^{\frac{1}{2}} + 2\pi(x^{\frac{1}{2}}) + (1+6^{\frac{1}{3}})x^{\frac{1}{3}} + O(x^{\frac{1}{3}}/\log x).$$

The number of groups contributing to each term soon increases however, and questions of overlap become more involved. Besides, our estimate for $e(x)$ as $2\pi(x) + (1+\sqrt{2})x^{\frac{1}{2}} + O(x^{\frac{1}{2}}/\log x)$ is already a rather dubious hybrid. We have given it in this form because the origin of each term is identifiable. But in fact the difference between $\pi(x)$ and its smooth asymptotic approximation $x/\log x$ is of the order of magnitude $x/\log^2 x$, and even the difference $\pi(x) - Li(x)$ is, apparently, not known to be smaller than $x^{\frac{1}{2}}$. (See, e.g. Huxley [7].)

References

1. Cameron, P.J.: Finite permutation groups and finite simple groups. Bull. London Math. Soc. **13**, 1–22 (1981)
2. Carter, R.W.: Simple groups of Lie type. London-New York-Sydney-Toronto: Wiley 1972

3. Cooperstein, B.N.: Minimal degree for a permutation representation of a classical group. Israel J. Math. **30**, 213–235 (1978)
4. Dickson, L.E.: Linear groups. Reprint. New York: Dover 1972
5. Gorenstein, D.: Finite groups. New York: Harper and Row 1968
6. Hardy, G.H., Wright, E.M.: An introduction to the theory of numbers. Fifth edition. Oxford: Clarendon Press 1979
7. Huxley, M.N.: The distribution of prime numbers—large sieves and zero-density theorems. Oxford: Clarendon Press 1972
8. Jordan, C.: Mémoire sur le nombre des valeurs des fonctions. J. École Poly. **22**, 113–194 (1861) = Oeuvres, Tome 1, 1–82. Paris: Gauthier-Villars 1961
9. Jordan, C.: Sur l'énumération des groupes primitifs pour les dix-sept premiers degrés. C.R. Acad. Sci. Paris **75**, 1754–1757 (1872) = Oeuvres, Tome 1, 341–344. Paris: Gauthier-Villars 1961
10. Kantor, W.M.: Permutation representations of the finite classical groups of small degree or rank. J. Algebra **60**, 158–168 (1979)
11. Mathieu, E.: Mémoire sur l'étude des fonctions des plusiers quantités, sur le manière de les former et sur les substitutions qui les laissent invariables. J. Math. Pure Appl. (Liouville) (Second series) **6**, 241–323 (1861)
12. Miyamoto, I.: On primitive permutation groups of degree $2p=4q+2$, p and q being prime numbers. J. Fac. Sci. Univ. Tokyo Sect. IA Math. **22**, 17–23 (1975)
13. Neumann, P.M., Saxl, J.: The primitive permutation groups of some special degrees, II: small multiples of certain large primes. Math. Z. **169**, 205–222 (1979)
14. Praeger, C.E., Saxl, J.: On the orders of primitive permutation groups. Bull. London Math. Soc. **12**, 303–307 (1980)
15. Scott, L.L.: Representations in characteristic p. In: The Santa Cruz Conference on Finite Groups (Cooperstein, B., Mason, G. eds.) pp. 319–331. Proceedings of Symposia in Pure Mathematics **37**. Providence, R.I.: American Mathematical Society 1980
16. Sims, C.C.: Computational methods in the study of permutation groups. In: Computational problems in abstract algebra, (ed. Leech, J.). Proceedings of a Conference (Oxford 1967), pp. 169–183. Oxford: Pergamon 1970

Received October 31, 1981

Math. Z. 180, 151-177 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Projective Modules over Graded Lie Algebras. I

Alvany Rocha-Caridi and Nolan R. Wallach*

Department of Mathematics, Rutgers University, New Brunswick, New Jersey 08903, U.S.A.

Table of Contents

§1. Preliminaries 153
§2. The Category $\mathcal{C}$ 154
§3. The Category $\mathcal{C}^S$ 155
§4. Projectives in $\mathcal{C}^S$ 156
§5. The Irreducible Subquotients of an Object in $\mathcal{C}^S$ 160
§6. The Duality Theorem 161
§7. Extensions and Relative Cohomology 163
§8. Embeddings of Verma Modules over Kac-Moody Lie Algebras 165
§9. The Generalized Strong BGG Resolution for Kac-Moody Lie Algebras 171

Introduction

Perhaps the most surprising aspect of the work of Bernstein-Gelfand-Gelfand (BGG) on Verma modules is their work on projective modules in their category $\mathcal{O}$ (in particular their beautiful duality theorem). This theorem has been used in [16] and [17] to simplify and sharpen the so-called strong and generalized strong BGG resolution of a finite dimensional irreducible module ([2] and [12]).

In this paper we introduce a class of graded Lie algebras, that is more general than the symmetrizable generalized Cartan matrix (GCM) algebras, for which one can by studying certain "truncated categories" of modules develop a theory of projectives strong enough to prove an analog of the BGG duality theorem. We then show how the duality theorem combined with results of Kac and Kazhdan [11] can be used to prove, following the outline of [16] and [17] and using Garland-Lepowsky's exact sequence [8], the analog, for symmetrizable Kac-Moody Lie algebras, of the generalized strong BGG resolution.

An example of our algebras which is different from the symmetrizable GCM algebras is the Witt algebra. Certain formal similarities of the Witt algebras and the GCM algebras motivated the definition of the graded Lie

* Research partially supported by National Science Foundation Grant

0025-5874/82/0180/0151/$05.40

algebras in question. In fact the conditions that our algebras should satisfy are abstractions of properties of both types of algebras. Hence the results of this paper apply to the Witt algebras as well. In the second part of this paper we will show how one can study the analog of the BGG resolution for the Witt algebra using our theory of projectives.

We now describe in detail the contents of this article.

In [1], I.N. Bernstein, I.M. Gelfand and S.I. Gelfand introduced the subcategory $\mathcal{O}$ of modules over a complex semi-simple Lie algebra. The universal highest weight modules, or Verma modules, are natural objects of $\mathcal{O}$. An important step in BGG's study is the construction of projectives in $\mathcal{O}$. BGG obtained a duality theorem which relates the indecomposable projectives, the Verma modules and the irreducibles in $\mathcal{O}$. In [16] and in [17], the categories $\mathcal{O}_S$ were introduced which include the category $\mathcal{O}$ of Bernstein-Gelfand-Gelfand as a particular case. BGG's duality theorem was then generalized. Here, generalized Verma modules (GVM's) play the role originally played by Verma modules in BGG's duality theorem. Using the duality theorem, together with an Yoneda type interpretation of n-extensions given in [19], vanishing theorems on the Ext bifunctor were obtained.

In the present article we consider certain (possibly infinite dimensional) complex graded Lie algebras $\mathfrak{g}$ of which complex reductive Lie algebras are an example. We define suitable subcategories $\mathscr{C}^S$ of $\mathfrak{g}$-modules, which contain generalized Verma modules as objects. In $\mathscr{C}^S$ we construct projective objects and obtain a duality theorem (Theorem 6.4) which generalizes [1, Proposition 6.2] and [17, Theorem 6.1]. Our construction is different from the ones obtained in [1] and [17]. For instance, while both [1] and [17] use infinitesimal characters in their constructions, the center of the universal enveloping algebra of $\mathfrak{g}$ does not play any role in ours. Also, unlike those of $\mathcal{O}$ and $\mathcal{O}_S$, the objects of $\mathscr{C}^S$ may not satisfy chain conditions for submodules. As a consequence, various results, which in [1] and [17] follow from the general theory of modules over a Noetherian ring, require a separate proof. For example, the multiplicity of an irreducible subquotient of an object of $\mathscr{C}^S$ must be carefully redefined.

Following the general lines of [17], we couple Theorem 6.4 with the interpretation of n-extensions mentioned above to obtain Theorem 7.5 on the vanishing of the Ext bifunctor. Theorem 7.5 generalizes [17, Lemma 9.1]. As in [16, introduction] and in [17], Theorem 7.5 implies vanishing theorems on the Lie algebra cohomology of a certain subalgebra of $\mathfrak{g}$.

In §8 we obtain results on the structure theory of Verma modules over Kac-Moody Lie algebras with symmetrizable GCM. Specifically, we use the Bruhat ordering to describe the inclusions among Verma modules with highest weights in the Weyl group orbit of a fixed dominant integral weight. This is the content of Theorem 8.15 which is an extension of [5, Theorem 7.7.7] for Kac-Moody Lie algebras with symmetrizable GCM. Preliminary to Theorem 8.15 we obtained Lemma 8.14. Lemma 8.14 establishes that the space of $\mathfrak{g}$-module homomorphisms of a Verma module into another is at most one-dimensional, for Verma modules with integral highest weights. This is an analog, for Kac-Moody Lie algebras with symmetrizable GCM, of a theorem

of Verma's [5]. The proof of Lemma 8.14 is based on Enright's theory of completions [6] and therefore differs from the existing proofs of its analog in the semisimple case.

As an application of the theory developed in §§1–7, we use the structure Theorem 8.15 to obtain, in a way similar to that followed in [17], a generalized strong type BGG resolution for Kac-Moody Lie algebras with symmetrizable GCM. In Theorem 9.2 we formulate criteria for the splitting of certain induced modules into direct sums of GVM's, generalizing [17, Theorem 9.1]. Theorems 9.7 and 9.12 prove the existence of a strong BGG type resolution by Verma modules and by GVM's, respectively. This generalized strong BGG type resolution for Kac-Moody Lie algebras contains the generalized strong BGG resolution for a complex semi-simple Lie algebra as a particular case. In addition, Theorem 9.12 proves the equivalence of the generalized strong type BGG resolution constructed here to a generalized weak type BGG resolution for Kac-Moody Lie algebras obtained by Garland and Lepowsky [8].

We thank V.G. Kac for posing the question of the construction of a BGG type resolution for the Witt algebra. We are grateful to J. Lepowsky for helpful conversations.

§1. Preliminaries

Let $\mathfrak{g}$ be a complex Lie algebra. If $\mathfrak{a}\subset\mathfrak{g}$ is a subalgebra, M an $\mathfrak{a}$-module and $\mu\in\mathfrak{a}^*$, we define

$$M_\mu=\{v\in M \mid Xv=\mu(X)v \quad \text{for all } X\in\mathfrak{a}\}.$$

Let $\mathfrak{h}\subset\mathfrak{g}$ be a subalgebra and let Q be the free abelian group with basis $\{\alpha_1,\dots,\alpha_n\}\subset\mathfrak{h}^*$. We assume, throughout, that $\mathfrak{g}$ is a graded Lie algebra: $\mathfrak{g}=\bigoplus_{\alpha\in Q}\mathfrak{g}_\alpha$, with $\mathfrak{g}_0=\mathfrak{h}$, $\dim\mathfrak{g}_\alpha<\infty$, and $[\mathfrak{g}_\alpha,\mathfrak{g}_\beta]\subset\mathfrak{g}_{\alpha+\beta}$ for all $\alpha,\beta\in Q$. We make two additional assumptions. Our first assumption is:

(T1) Let $0\neq\alpha=\sum_{i=1}^{n} m_i\alpha_i\in Q$. $\mathfrak{g}_\alpha\neq(0)$ only if $m_i\in\mathbb{Z}_+$ for all i, or $-m_i\in\mathbb{Z}_+$ for all i.

Here $\mathbb{Z}_+$ denotes the set of all nonnegative integers.

Let Q^+ be the subsemigroup of Q generated by $\{\alpha_1,\dots,\alpha_n\}$. We set

$$\mathfrak{n}=\bigoplus_{\alpha\in Q^+\setminus\{0\}}\mathfrak{g}_\alpha,\qquad \mathfrak{n}^-=\bigoplus_{\alpha\in Q^+\setminus\{0\}}\mathfrak{g}_{-\alpha},\qquad \mathfrak{b}=\mathfrak{h}\oplus\mathfrak{n}.$$

Then $\mathfrak{g}=\mathfrak{n}^-\oplus\mathfrak{h}\oplus\mathfrak{n}$, a direct sum of subalgebras.

We now state our second additional assumption:

(T2) There is an $\mathfrak{h}$-invariant, nondegenerate pairing between the spaces $\mathfrak{n}$ and $\mathfrak{n}^-$.

Remark. Examples of such $\mathfrak{g}$ are:

1) a finite dimensional complex semi-simple Lie algebra; here we take $\mathfrak{h}$ to be a Cartan subalgebra and $\mathfrak{n}^-\oplus\mathfrak{h}\oplus\mathfrak{n}$ is a triangular decomposition of $\mathfrak{g}$. The nondegenerate $\mathfrak{h}$-invariant pairing of $\mathfrak{n}$ and $\mathfrak{n}^-$ is given by the Killing form.

2) the extended Kac-Moody Lie algebra $\mathfrak{g}=\mathfrak{d}\times\hat{\mathfrak{g}}$ obtained by adjoining degree derivations to a Kac-Moody Lie algebra $\hat{\mathfrak{g}}$. If the generalized Cartan matrix of the Kac-Moody Lie algebra $\hat{\mathfrak{g}}$ is symmetrizable, then all the assumptions made hold for the extended Kac-Moody Lie algebra (see §8).

Let $\mathfrak{a}\subset\mathfrak{g}$ be a subalgebra. We denote by $U(\mathfrak{a})$ the universal enveloping algebra of $\mathfrak{a}$ and identify $U(\mathfrak{a})$ with a subalgebra of $U(\mathfrak{g})$. We will use, interchangeably, the languages of Lie algebra modules and universal enveloping algebra modules.

§2. The Categories $\mathcal{C}(\lambda, q)$

If $k\in\mathbb{Z}_+$, we set ${}_kQ^+=\left\{\mu=\sum_{i=1}^n m_i\alpha_i\in Q^+ \,\middle|\, \sum_{i=1}^n m_i>k, m_i\in\mathbb{Z}_+\right\}$.

Let λ be an arbitrary, but fixed, linear functional on $\mathfrak{h}$. Let $q\in\mathbb{Z}_+$ be also arbitrary, but fixed. If $\mu=\sum_{i=1}^n m_i\alpha_i\in Q, m_i\in\mathbb{Z}$, set $\mu^+=\sum m_i\alpha_i$, $m_i\in\mathbb{Z}_+$, and $\mu^-=\mu-\mu^+$. We call μ^+ the Q^+-part of μ. We denote by $\Pi=\Pi(\lambda,q)$ the set of all $\mu\in(\mathfrak{h})^*$ such that $\lambda-\mu\in Q$ and the Q^+-part of $\mu-\lambda$ is in $Q^+\setminus_q Q^+$.

Definition 1. Let $\mathcal{C}=\mathcal{C}(\lambda,q)$ be the full subcategory of the category of $\mathfrak{g}$-modules M such that:

(i) $M=\bigoplus_{\mu\in\mathfrak{h}^*} M_\mu$, and $\dim M_\mu<\infty$,

(ii) $M_\mu=(0)$ for all $\mu\in\mathfrak{h}^*\setminus\Pi$.

Let $\mu\in\mathfrak{h}^*$ and M a $\mathfrak{g}$-module. We call μ a *weight* of M if $M_\mu\neq(0)$, and refer to the nonzero elements of M_μ as *weight vectors* of weight μ. The subspace M_μ of M is called the *weightspace* of M relative to the weight μ. M is said to be a *weight module* if it is the direct sum of its weightspaces.

Let $\mu\in\mathfrak{h}^*$, M a $\mathfrak{g}$-module and v a nonzero vector of M_μ such that $\mathfrak{n}v=0$. M is called a *highest weight module* if $M=U(\mathfrak{g})v$. In this case, μ is called the *highest weight* of M and M_μ the *highest weightspace* of M; M_μ is one-dimensional, and its nonzero elements are the *highest weight vectors* of M. A highest weight module is clearly a weight module with finite dimensional weightspaces. The weights of a highest weight module of highest weight μ lie in the set $\{\mu-\alpha\,|\,\alpha\in Q^+\}$. In particular, if $\mu\in\Pi$, a highest weight module of highest weight μ is an object of $\mathcal{C}$.

Let $\mu\in\mathfrak{h}^*$. We denote by $\mathbb{C}(\mu)$ the vector space $\mathbb{C}$ when looked upon as a $U(\mathfrak{b})$-module via the action:

$$(H+X)z=\mu(H)z, \quad \text{for all } H\in\mathfrak{h}, \quad X\in\mathfrak{n} \quad \text{and} \quad z\in\mathbb{C}.$$

The left $U(\mathfrak{g})$-module induced by $\mathbb{C}(\mu)$, $U(\mathfrak{g})\otimes_{U(\mathfrak{b})}\mathbb{C}(\mu)$ is called the *Verma module* associated with $\mathfrak{g},\mathfrak{h},\{\alpha_1,\ldots,\alpha_n\}$ and μ. We denote it by $M(\mu)$. $M(\mu)$ is the universal highest weight module with highest weight μ. Every highest weight module has a unique maximal proper submodule and, therefore, a unique irreducible quotient. We denote by $L(\mu)$ the unique irreducible quotient of $M(\mu)$.

Following [9] and [15] we define a generalization of Kostant's partition function (cf. [8]):

We let Ω be a set and f: $\Omega \to Q$ a surjection such that for all $\alpha \in Q$, $\#(f^{-1}(\alpha)) = \dim \mathfrak{g}_\alpha$. We define $\mathscr{P}$: $\mathfrak{h}^* \to \mathbb{Z}_+$ by: $\mathscr{P}(\mu)$ is the number of maps τ: $\Omega \to \mathbb{Z}_+$ such that $\mu = \sum_{a \in \Omega} \tau(a) f(a)$.

It is clear that $\dim(M(\mu))_\nu = \mathscr{P}(\mu - \nu)$.

The following result is straightforward:

Proposition 2. (1) *$\mathscr{C}$ is stable under the operations of taking submodules, quotients and finite direct sums.*

(2) *The modules $L(\mu)$ with $\mu \in \Pi$ exhaust the class of irreducible modules in $\mathscr{C}$.*

§3. The Categories $\mathscr{C}^S(\lambda, q)$

We fix an arbitrary subset S of $\{1, \ldots, n\}$. Set $Q_S = Q \cap \sum_{i \in S} \mathbb{Z}\alpha_i$ where $\mathbb{Z}$ is the ring of integers, and $Q_S^+ = Q^+ \cap Q_S$. Let $\mathfrak{n}_S = \bigoplus_{\alpha \in Q_S^+ \setminus \{0\}} \mathfrak{g}_\alpha$, $\mathfrak{n}_S^- = \bigoplus_{\alpha \in Q_S^+ \setminus \{0\}} \mathfrak{g}_{-\alpha}$, $\mathfrak{m}_S = \mathfrak{n}_S^- \oplus \mathfrak{h} \oplus \mathfrak{n}_S$, $\mathfrak{u}_S = \bigoplus_{\alpha \in Q^+ \setminus Q_S^+} \mathfrak{g}_\alpha$ and $\mathfrak{u}_S^- = \bigoplus_{\alpha \in Q^+ \setminus Q_S^+} \mathfrak{g}_{-\alpha}$. Set $\mathfrak{p}_S = \mathfrak{m}_S \oplus \mathfrak{u}_S$. Since $[\mathfrak{m}_S, \mathfrak{u}_S] \subset \mathfrak{u}_S$, $\mathfrak{p}_S$ is a subalgebra of $\mathfrak{g}$.

Assume that S satisfies the following conditions:

(S1) $\mathfrak{h} = \mathfrak{a}_S \oplus \mathfrak{h}_S$, a direct sum of subalgebras, such that $\{\alpha_i|_{\mathfrak{h}_S} | i \in S\}$ is linearly independent.

(S2) If $\bar{Q}_S$ is the free abelian group with basis $\{\alpha_i|_{\mathfrak{h}_S} | i \in S\}$, and $\mathfrak{g}_S = \mathfrak{n}_S^- \oplus \mathfrak{h}_S \oplus \mathfrak{n}_S$, then $\mathfrak{g}_S = \bigoplus_{\alpha \in \bar{Q}_S} (\mathfrak{g}_S)_\alpha$ and $(\mathfrak{g}_S)_0 = \mathfrak{h}_S$.

(S3) $\{\alpha \in Q_S | \mathfrak{g}_\alpha \neq 0\}$ is a finite set.

(S4) Let $\mathfrak{g}_S$ be as in (S2). Every $\mathfrak{g}_S$-invariant subspace of $\mathfrak{g}$ is a direct sum of finite-dimensional irreducible $\mathfrak{g}_S$-modules.

(S5) Every finite-dimensional $\mathfrak{g}_S$-module is completely reducible.

(S6) The $\mathfrak{h}$-invariant pairing of $\mathfrak{n}$ and $\mathfrak{n}^-$ restricts to an $\mathfrak{m}_S$-invariant pairing of $\mathfrak{u}_S$ and $\mathfrak{u}_S^-$.

Remarks. 1) The pairing in (S6) extends to an $\mathfrak{m}_S$-invariant pairing of the symmetric algebras of $\mathfrak{u}_S$ and $\mathfrak{u}_S^-$ and, therefore, to an $\mathfrak{m}_S$-invariant pairing of $U(\mathfrak{u}_S)$ and $U(\mathfrak{u}_S^-)$ by the Poincaré-Birkhoff-Witt theorem.

2) Examples of such S are: (i) $S = \varnothing$. Here $\mathfrak{m}_\varnothing = \mathfrak{h}$. If we set $\mathfrak{a}_\varnothing = (0)$, then (S1)–(S6) are satisfied; (ii) Let $\mathfrak{g} = \mathfrak{d} \rtimes \hat{\mathfrak{g}}$ be an extended Kac-Moody Lie algebra [8]. Any set S defining an F-parabolic subalgebra of $\hat{\mathfrak{g}}$ satisfies (S1)–(S3) for suitable choices of $\mathfrak{h}_S$ and $\mathfrak{a}_S$ [8]. (S4) follows from [8, Proposition 5.1] and (S5) by Weyl's theorem. If the generalized Cartan matrix defining $\hat{\mathfrak{g}}$ is symmetrizable, then (S6) holds.

For $k \in \mathbb{Z}_+$, set ${}_kQ_S^+ = {}_kQ^+ \cap Q_S^+$, and denote by $\bar{Q}_S^+$ and ${}_k\bar{Q}_S^+$ the set of restrictions to $\mathfrak{h}_S^*$ of the elements of Q_S^+ and ${}_kQ_S^+$, respectively. If $q' \in \mathbb{Z}_+$ and $\lambda' \in \mathfrak{h}_S^*$, we denote by $\Pi_{S'}(\lambda', q')$ the set of all $\mu' \in \mathfrak{h}_S^*$ such that $\lambda' - \mu' \in \bar{Q}_S$ and the

$\bar{Q}_S^+$-part of $\mu'-\lambda'$ is in $\bar{Q}_S^+ \setminus_{q'} \bar{Q}_S^+$. (The definition of the $\bar{Q}_S^+$-part of an element of $\bar{Q}_S$ is similar to the definition of the Q^+-part of an element of Q given in §2.) Let $\mathscr{C}_S(\lambda', q')$ be the full subcategory of the category of $\mathfrak{g}_S$-modules (see (S2) above for the definition of $\mathfrak{g}_S$) such that:

(i') $M = \bigoplus_{\mu' \in \mathfrak{h}_S^*} M_{\mu'}$, $\dim M_{\mu'} < \infty$, and (ii') $M_{\mu'} = (0)$ for all μ' in $\mathfrak{h}_S^* \setminus \Pi_S(\lambda', q')$.

We now apply the results of §2 to $\mathfrak{g}_S$ and $\mathscr{C}_S(\lambda', q')$. If $\mu' \in \mathfrak{h}_S^*$ we let $\mathbb{C}_S(\mu')$ be the one-dimensional $U(\mathfrak{n}_S \oplus \mathfrak{h}_S)$-module where $\mathfrak{n}_S$ acts trivially and $H \in \mathfrak{h}_S$ acts by the scalar $\mu'(H)$ for all $H \in \mathfrak{h}_S$. We denote by $M_S(\mu')$ the induced $U(\mathfrak{g}_S)$-module $U(\mathfrak{g}_S) \otimes_{U(\mathfrak{n}_S \oplus \mathfrak{h}_S)} \mathbb{C}_S(\mu')$ and by $L_S(\mu')$ the unique irreducible quotient of $M_S(\mu')$. By Proposition 2.2 the modules $L_S(\mu')$, with $\mu' \in \Pi_S(\lambda', q')$, exhaust the irreducible modules in $\mathscr{C}_S(\lambda', q')$. Let $\Pi_S^+(\lambda', q') = \{\mu \in \mathfrak{h}^* \mid \mu_S = \mu_{|\mathfrak{h}_S^*} \in \Pi_S(\lambda', q')$ and $L_S(\mu_S)$ is finite dimensional$\}$. There is a natural bijection: $\mu \mapsto [R_S(\mu)]$ between $\Pi_S^+(\lambda', q')$ and the set of isomorphism classes of finite dimensional irreducible $\mathfrak{m}_S$-modules $R_S(\mu)$ such that, as a $\mathfrak{g}_S$-module, $R_S(\mu) = L_S(\mu_S)$.

Let $\lambda \in \mathfrak{h}^*$, $q \in \mathbb{Z}_+$ be fixed. Let $\Pi^S(\lambda, q)$ be $\bigcup_{\lambda' \in \Pi(\lambda, q)} \Pi_S^+(\lambda', 0)$.

We now introduce the category $\mathscr{C}^S(\lambda, q)$.

Definition 1. Let $\mathscr{C}^S = \mathscr{C}^S(\lambda, q)$ be the full subcategory of the category $\mathscr{C}(\lambda, q)$ of $\mathfrak{g}$-modules M such that:

(i) M is a direct sum of $R_S(\mu)$'s, where $\mu \in \Pi^S(\lambda, q)$.

Let L be an irreducible $\mathfrak{m}_S$-module. Because $[\mathfrak{p}_S, \mathfrak{u}_S] \subset \mathfrak{u}_S$, we can extend the action of $\mathfrak{m}_S$ on L to an action of $\mathfrak{p}_S$ on L, by letting $\mathfrak{u}_S$ act trivially on L. We set $M(L) = U(\mathfrak{g}) \otimes_{U(\mathfrak{p}_S)} L$, and call $M(L)$ the *generalized Verma module* associated with $\mathfrak{g}$, $\mathfrak{h}$, Q, S and L. (S4) implies that (i) above is satisfied. Hence $M(R_S(\mu)) \in \mathscr{C}^S(\lambda, q)$ for all $\mu \in \Pi^S(\lambda, q)$. It is clear that $M(R_S(\mu))$ is a highest weight module. We denote by $L^S(\mu)$ its unique irreducible quotient.

Definition 2. Let $\mathscr{F}(\mathscr{C}^S)$ denote the full subcategory of the category $\mathscr{C}^S$ consisting of the finitely generated objects of $\mathscr{C}^S$.

We now state the analog or Proposition 2.2 for the category $\mathscr{C}^S$. Its proof is the same as that of Proposition 3.3 of [17], and will be omitted.

Proposition 3. (1) *$\mathscr{C}^S$ is stable under the operations of taking submodules, quotients and finite direct sums.*

(2) *For each $\mu \in \Pi^S$, $L^S(\mu) = L(\mu)$ and the modules $L(\mu)$, $\mu \in \Pi^S$ exhaust the class of irreducible modules in $\mathscr{C}^S$.*

Remark. If $S = \varnothing$ then $Q_\varnothing = Q_\varnothing^+ = \{0\}$ and $\mathfrak{m}_\varnothing = \mathfrak{h}$. We set $\mathfrak{h}_\varnothing = \mathfrak{h}$. Then $\mathscr{C}^\varnothing(\lambda, q) = \mathscr{C}(\lambda, q)$, and $M(R_\varnothing(\mu)) = M(\mu)$, $\mu \in \Pi^\varnothing(\lambda, q) = \Pi(\lambda, q)$.

§4. Projectives in $\mathscr{C}^S$

By the Poincaré-Birkhoff-Witt theorem, $U(\mathfrak{g}) = \bigoplus_{\alpha \in Q} U(\mathfrak{g})_\alpha$. For each $k \in \mathbb{Z}_+$, we let $U(\mathfrak{g})^{(k)} = \bigoplus_{\alpha \in kQ^+} U(\mathfrak{g})_\alpha$. Let $\lambda \in \mathfrak{h}^*$, $q \in \mathbb{Z}_+$ be fixed, and let $\mathscr{C}^S = \mathscr{C}^S(\lambda, q)$, $\Pi^S = \Pi^S(\lambda, q)$ be defined as in §3. If $\mathfrak{a} \subset \mathfrak{g}$ we set $U(\mathfrak{a})^{(k)} = U(\mathfrak{a}) \cap U(\mathfrak{g})^{(k)}$, $k \in \mathbb{Z}_+$.

Proposition 1. *Let $\mu\in\Pi^S$. There is a $\mathfrak{g}$-module $P=P(\mu)$ in $\mathscr{C}^S$ such that, for every $M\in\mathscr{C}^S$, $\mathrm{Hom}_{\mathfrak{g}}(P(\mu), M)$ is canonically isomorphic with $\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M)$.*

Proof. There is a nonnegative integer k, depending on λ, q and μ such that if v is a weight vector in $R_S(\mu)$ then the weights of $U(\mathfrak{u})^{(k)}v$ are in $\mathfrak{h}^*\setminus\Pi(\lambda, q)$. We set

$$P(\mu)=U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}((U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})\otimes R_S(\mu)).$$

By (S4) and (S5) $P(\mu)\in\mathscr{C}^S$. Let Φ: $\mathrm{Hom}_{\mathfrak{g}}(P(\mu), M)\to\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M)$ be the map defined by: $\Phi(A)(v)=A(1\otimes\bar{1}\otimes v)$, where $A\in\mathrm{Hom}_{\mathfrak{g}}(P(\mu), M)$, $v\in R_S(\mu)$ and $\bar{Y}$ denotes the canonical image of $Y\in U(\mathfrak{u}_S)$ in $U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)}$. If $f\in\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M)$, $X\in U(\mathfrak{g})$, $Y\in U(\mathfrak{u}_S)$, let $\Psi(f)(X\otimes\bar{Y}\otimes v)=XYf(v)$. By our choice of k, $U(\mathfrak{u}_S)^{(k)}f(v)=0$. Hence $\Psi(f)\in\mathrm{Hom}_{\mathfrak{g}}(P(\mu), M)$. Clearly $\Psi\circ\Phi=\Phi\circ\Psi=\mathrm{id}$. This concludes the proof of the proposition. Q.E.D.

As immediate consequences of Proposition 1 we obtain:

Corollary 2. *For each $\mu\in\Pi^S$, $P(\mu)$ is a projective object of $\mathscr{C}^S$.*

Corollary 3. *Every object of $\mathscr{F}(\mathscr{C}^S)$ is the image of a projective in $\mathscr{C}^S$.*

Definition 4. An object M in $\mathscr{C}^S$ is said to have a *generalized Verma composition series* (GVCS) if there is a filtration:

$$M=M^0\supset M^1\supset\ldots\supset M^l\supset M^{l+1}=(0)\quad\text{of submodules,}$$

such that $M^i/M^{i+1}\simeq M(R_S(\mu_i))$, $\mu_i\in\Pi^S$, $i=0, 1, \ldots, l$.

Proposition 5. *Let $\mu\in\Pi^S$. The object $P(\mu)$ in Proposition 1 has a GVCS.*

Proof. Let $u_1, \ldots, u_r$ be a collection of vectors of weights $\mu_1', \ldots, \mu_r'$, respectively, such that the projections $\bar{u}_1, \ldots, \bar{u}_r$ constitute a basis of $U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)}$, where the nonnegative integer k is defined in Proposition 1. We order the weights $\mu_1', \ldots, \mu_r'$ as follows: If the $(Q\setminus Q_S)$-part[1] of $\mu_i'-\mu_j'$ is in Q^+, then we set $i>j$. Let M_i' be the $\mathfrak{p}_S$-submodule of $(U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})\otimes R_S(\mu)$ generated by $\bar{u}_i\otimes v_\mu, \ldots, \bar{u}_r\otimes v_\mu$, where v_μ is a generator of $R_S(\mu)$, $i=0, 1, \ldots, r$. Then $(U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})\otimes R_S(\mu)=M_0'\supset M_1'\supset\ldots\supset M_r'\supset M_{r+1}'=(0)$. Since $\mathfrak{u}_S M_i'\subset M_{i+1}'$ and by (S5), we can find a filtration $(U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})\otimes R_S(\mu)=M_0\supset M_1\supset\ldots\supset M_l\supset M_{l+1}=(0)$, such that $M_i/M_{i+1}\simeq R_S(\mu_i)$, $\mu_i\in\Pi^S$ and $\mathfrak{u}_S R_S(\mu_i)=(0)$, $i=0, 1, \ldots, l$. Now, set $M^i=U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}M_i$ to get the filtration of $P(\mu)$ we sought. Q.E.D.

We define a partial order on Π by setting $\mu_1>\mu_2$ if and only if $\mu_1-\mu_2\in Q^+\setminus\{0\}$, for any $\mu_1, \mu_2\in\Pi^S$.

The same arguments used to prove Proposition 6.2 in [1] (cf. also [17], Lemmas 5.4 and 5.5) can be used to prove Lemma 6 and Proposition 7:

Lemma 6. *Let $M\in\mathscr{C}^S$ and assume that M has a GVCS. Let μ be an element of Π^S maximal among those weights of M such that $\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M)\neq 0$. Let f be*

[1] This is defined in the same manner as is done for parabolics of finite dimensional semi-simple lie algebras

a nonzero element of $\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M)$, *and set* $v = f(v_\mu)$, $M' = U(\mathfrak{g})v$, *where* v_μ *is a generator of* $R_S(\mu)$. *Then* $M' = M(R_S(\mu))$ *and* M/M' *admits a GVCS.*

Applying Lemma 6 we obtain

Proposition 7. *If* $M = M_1 \oplus M_2$ *in* $\mathscr{C}^S$ *admits a GVCS then both* M_1 *and* M_2 *admit a GVCS.*

Proposition 8. *Every object in* $\mathscr{F}(\mathscr{C}^S)$ *is a finite direct sum of indecomposable modules.*

Proof. If $M \in \mathscr{F}(\mathscr{C}^S)$ we say that M has rank $\leqq r$ if there exists $F \subset \mathfrak{h}^*$, F finite, so that $M = U(\mathfrak{g}) \sum_{\mu \in F} M_\mu$ and $\dim \sum_{\mu \in F} M_\mu \leqq r$.

We prove the result by induction on the rank.

If M has rank $\leqq 1$ then M is indecomposable. Indeed, $M = U(\mathfrak{g})M_\mu$, $\dim M_\mu \leqq 1$. If $M = M^{(1)} \oplus M^{(2)}$ and $M^{(i)}$ is a proper submodule, $i = 1, 2$, then $p_i(M_\mu) \neq 0$, $i = 1, 2$, where p_i: $M \to M^{(i)}$ is the projection. Thus $\dim M_\mu \geqq 2$, contrary to the assumption.

Suppose that we have shown that if M has rank $\leqq r$ then M is a finite direct sum of indecomposable submodules. We now assume $r k M \leqq r + 1$. If M is indecomposable then we are done. Otherwise $M = M^{(1)} \oplus M^{(2)}$, $M^{(i)}$ a proper $\mathfrak{g}$-submodule, $i = 1, 2$. Let $M = U(\mathfrak{g}) \sum_{\mu \in F} M_\mu$ with $\dim(\sum_{\mu \in F} M_\mu) \leqq r + 1$. Then

$$\sum_{\mu \in F} M_\mu = p_1(\sum_{\mu \in F} M_\mu) \oplus p_2(\sum_{\mu \in F} M_\mu)$$

and $M^{(i)} = U(\mathfrak{g})(p_i(\sum_{\mu \in F} M_\mu))$. Hence rank of $M^{(i)} \leqq r$ for $i = 1, 2$. Now apply the inductive hypothesis to $M^{(1)}$ and $M^{(2)}$. Q.E.D.

Corollary 9. *For each* $\mu \in \Pi^S$, $P(\mu)$ *is a finite direct sum of indecomposable projectives in* $\mathscr{C}^S$. *Furthermore, the set of all indecomposable direct summands of* $P(\mu)$ *for all* $\mu \in \Pi^S$ *exhaust the class of indecomposable projectives in* $\mathscr{C}^S$ *which are also objects of* $\mathscr{F}(\mathscr{C}^S)$.

Proof. By the proof of Proposition 1 it is clear that $P(\mu) \in \mathscr{F}(\mathscr{C}^S)$ for all $\mu \in \Pi^S$. Hence the first statement is true. Now, let I be an indecomposable projective in $\mathscr{C}^S$ such that $I \in \mathscr{F}(\mathscr{C}^S)$. Then I is a quotient of $P(\mu)$ for some $\mu \in \Pi^S$ by Corollary 3. Therefore I is a summand of $P(\mu)$ for some $\mu \in \Pi^S$. Q.E.D.

Corollary 10. *Every indecomposable projective of* $\mathscr{C}^S$ *which is an object of* $\mathscr{F}(\mathscr{C}^S)$ *has a GVCS.*

Remark. If M is a finitely generated module over an arbitrary ring with unity, then the set $\mathscr{M}$ of all proper submodules of M is inductive. By the maximum principle (see e.g. [4, 57.1]) $\mathscr{M}$ has a maximal element. Therefore M has a maximal proper submodule. Also, every proper submodule of M is contained in a maximal proper submodule of M.

The proof of the next lemma follows the general lines of [16, Lemma III.1.3].

Lemma 11. *Each indecomposable projective in $\mathscr{C}^S$ which is an object of $\mathscr{F}(\mathscr{C}^S)$ has a unique maximal proper submodule.*

Proof. Let I be an indecomposable projective in $\mathscr{C}^S$ which is an object of $\mathscr{F}(\mathscr{C}^S)$. By Corollary 9, there is a weight vector v of weight $\mu\in\Pi^S$ such that $I=U(\mathfrak{g})v$ since I is summand of $P(\mu)$. By the remark above, I has at least one maximal proper submodule. Suppose that M_1 and M_2 are two maximal proper submodules of I. Then there is a $\mathfrak{g}$-module exact sequence $M_1\oplus M_2\xrightarrow{\phi} I\to 0$. Since I is projective there is a $\mathfrak{g}$-module map $\eta\colon I\to M_1\oplus M_2$ such that $\phi\circ\eta=\mathrm{id}_I$. Set $f_i=\pi_i\circ\eta\colon I\to M_i$, where $\pi_i\colon M_1\oplus M_2\to M_i$, $i=1,2$ are the natural projections. Then $f_1+f_2=\mathrm{id}_I$ and $f_1\circ f_2=f_2\circ f_1$. We now consider the chain

$$f_1(I)\supset f_1^2(I)\supset\ldots\supset f_1^j(I)\supset\ldots$$

of submodules of $f_1(I)$. Clearly, $f_1^j(I)=U(\mathfrak{g})f_1^j(v)$ for all $j\in\mathbb{N}$ and $f_1^j(v)\in(f_1(I))_\mu$. Since $\dim f_1(I)_\mu<\infty$, there is $s\in\mathbb{N}$ such that

$$(*)\qquad f_1^s(I)=f_1^{s+1}(I)=\ldots.$$

Now, $f_1(I)\subsetneqq I$, since $M_1\subsetneqq I$. Hence we have the exact sequence:

$$0\to\mathrm{Ker}\,f_1^s\to I\to f_1^s(I)\to 0$$

with $\mathrm{Ker}\,f_1^s\neq(0)$. Let $\mu\in\mathrm{Ker}\,f_1^s\cap f_1^s(I)$. Then $u=f_1^s(w)$ and $f_1^{2s}(w)=0$. We assert that $f_1\colon f_1^j(I)\to f_1^{j+1}(I)$ is bijective for all $j\geqq s$. Indeed, let $j\geqq s$ and suppose that $\mathrm{Ker}\ f_{1|f_1^j(I)}\neq 0$. Then $(\mathrm{Ker}\,f_{1|f_1^j(I)})_\nu\neq(0)$ for some $\nu\in\Pi$. Hence $\dim f_1^{j+1}(I)_\nu<\dim f_1^j(I)_\nu$ which contradicts (*) above. This proves our assertion. Hence $f_1^s\colon f_1^s(I)\to f_1^{2s}(I)$ is bijective. Thus $f_1^s(u)=f_1^{2s}(w)=0$ implies $u=0$. Therefore $I=\mathrm{Ker}\,f_1^s\oplus f_1^s(I)$, with $\mathrm{Ker}\,f_1^s\neq(0)$. Since I is indecomposable, this implies that $f_1^s=0$. Similarly, we show that $f_2^t=0$ for some $t\in\mathbb{N}$. But this implies that the identity map on I is nilpotent since $(f_1\circ f_2=f_2\circ f_1)$, which is a contradiction. Q.E.D.

Lemma 12. *Let L be an irreducible in $\mathscr{C}^S$. There is a unique indecomposable projective in $\mathscr{C}^S$ which is an object of $\mathscr{F}(\mathscr{C}^S)$ and maps onto L.*

Proof. Suppose I_1 and I_2 are two indecomposable projective modules in $\mathscr{C}^S$, which are also objects of $\mathscr{F}(\mathscr{C}^S)$, such that $I_i\xrightarrow{\pi_i} L\to 0$, $i=1,2$, are exact sequences. By projectivity, there is $f\colon I_2\to I_1$ such that $\pi_1\circ f=\pi_2$. If $\mathrm{Im}\,f\neq I_1$ then $\mathrm{Im}\,f$ is contained in a maximal proper submodule of I, by the remark preceeding Lemma 11. Since $\mathrm{Ker}\,\pi_1$ is a maximal proper submodule of I_1, Lemma 11 implies that $\mathrm{Im}\,f\subset\mathrm{Ker}\,\pi_1$. Hence $\pi_2=0$, a contradiction. Therefore $f\colon I_2\to I_1$ is surjective. This implies that I_1 is a summand of I_2, hence $I_1=I_2$, since I_2 is indecomposable. Q.E.D.

Corollary 13. *There is a one-to-one correspondence between the irreducible objects in $\mathscr{C}^S$ and the indecomposable projective objects in $\mathscr{C}^S$ which are objects of $\mathscr{F}(\mathscr{C}^S)$. We denote by $I(\mu)$ the unique object of $\mathscr{F}(\mathscr{C}^S)$, indecomposable projective in $\mathscr{C}^S$, which is assigned to $L(\mu)$, $\mu\in\Pi^S$, by this correspondence.*

§5. The Irreducible Subquotients of an Object in $\mathscr{C}^S$

We remind the reader that we will be using the partial order on Π introduced before Lemma 4.6.

The proof of the next lemma follows the general lines of [8, Lemma 4.4]. We sketch the argument.

Lemma 1. *Let $M\in\mathscr{C}^S$. There are submodules $(0)=M_0\subset M_1\subset\dots$, such that $\bigcup_{i=0}^{\infty} M_i=M$, M_i/M_{i-1} is a highest weight module with highest weight $\mu_i\in\Pi$ such that $\mu_i<\mu_j$ implies $i>j$, and if $\mu_i=\mu_{i+1}$ then $\dim(M_i/M_{i-1})_{\mu_i}-1=\dim(M_{i+1}/M_i)_{\mu_i}$.*

Proof. Let $\mu_1\in\Pi^S$ be maximal such that $M_{\mu_1}\neq(0)$ and let v be a nonzero element of M_{μ_1}. Then $\mathfrak{n}\cdot v_1=0$. Set $M_1=U(\mathfrak{g})v$. Considering the quotient module M/M_1 we can find submodules $M_1\subset M_2\subset\dots\subset M_{d_1}$ of M, where $d_1=\dim M_{\mu_1}$ such that M_i/M_{i-1} is a highest weight module of highest weight μ_1, $\dim(M_i/M_{i-1})_{\mu_1}-1=\dim(M_{i+1}/M_i)_{\mu_1}$, $i=1,\dots,d_1-1$, and $(M/M_{d_1})_{\mu_1}=(0)$. Induction gives the filtration we seek. Q.E.D.

We will define next a generalization of the notion of number of occurrences of an irreducible module in a Jordan-Hölder series of a module with finite length, for an object in $\mathscr{C}^S$.

Let $M\in\mathscr{C}^S$, $\mu\in\Pi^S$. We define a number $(M:L(\mu))$ (which apparently depends on the choices of various filtrations) as follows:

Suppose first that M is a highest weight module of highest weight $\nu\in\Pi^S$. We set $(M:L(\mu))=0$, if $\nu\not\geqq\mu$, and $(M:L(\mu))=1$ if $\nu=\mu$. If $\nu>\mu$, we define $(M:L(\mu))$ inductively in the following way: Assume that $(V:L(\mu))$ has been defined for all highest weight modules V with highest weight $\nu'\in\Pi^S$ such that $\nu'-\mu<\nu-\mu$. Let M' be the unique maximal submodule of M. Let $(0)=M'_0\subset M'_1\subset\dots$ be a sequence of submodules of M' such that $M'=\bigcup_{i=0}^{\infty} M'_i$ and M'_i/M'_{i-1} is a highest weight module of highest weight ν'_i as in Lemma 1. Let s be the largest index such that $(M'_s/M'_{s-1})_\mu\neq(0)$. Then set

$$(M:L(\mu))=(M':L(\mu)=(M'_1:L(\mu))+(M'_2/M'_1:L(\mu))+\dots+(M'_s/M'_{s-1}:L(\mu)).$$

Clearly $\nu'_i-\mu<\nu-\mu$, for all i such that $\nu'_i\geqq\mu$ and, therefore, the numbers $(M'_i/M'_{i-1}:L(\mu))$ have been defined, by the induction hypothesis, for all $i=1,\dots,s$.

Let now M be an arbitrary object in $\mathscr{C}^S$. Let $(0)=M_0\subset M_1\subset\dots$ be a sequence of submodules of M such that $\bigcup_{i=0}^{\infty} M_i=M$ and M_i/M_{i-1} is a highest weight module of highest weight μ_i as in Lemma 1. Let r be the largest index such that $(M_r/M_{r-1})_\mu\neq(0)$. Then set $(M:L(\mu))=(M_1:L(\mu))+(M_2/M_1:L(\mu))+\dots+(M_r/M_{r-1}:L(\mu))$.

Remark. It is not hard to prove directly that $(M:L(\mu))$ is well-defined. However, the next proposition implies that $(M:L(\mu))$ is independent of the choices. Therefore the notation is justified.

Definition 2. Let $M \in \mathscr{C}^S$, $\mu \in \Pi^S$. The number $(M: L(\mu))$ is called the *multiplicity of* $L(\mu)$ *in* M.

Remark. It is clear from the definition of $(M: L(\mu))$ that $(M: L(\mu)) \neq 0$ if and only if $L(\mu)$ is a subquotient of M.

Proposition 3. *Let* $M \in \mathscr{C}^S$ *and* $\mu \in \Pi^S$. *Then* $(M: L(\mu)) = \dim \mathrm{Hom}(I(\mu), M)$.

Proof. Let $(0) = M_0 \subset M_1 \subset \ldots$ be a sequence of submodules of M such that M_i/M_{i-1} is a highest weight module of highest weight μ_i as in Lemma 1. Let r be the largest index such that $(M_r/M_{r-1})_\mu \neq (0)$. We assert that $\mathrm{Hom}(I(\mu), M) = \mathrm{Hom}(I(\mu), M_r)$. First we show that, if $i \geqq r$ then $\mathrm{Hom}(I(\mu), M_i) = \mathrm{Hom}(I(\mu), M_r)$. Indeed, if $i > r$, then $(M_i/M_{i-1})_\mu = (0)$, hence $\mathrm{Hom}(I(\mu), M_i/M_{i-1}) = 0$. Since $I(\mu)$ is projective $\mathrm{Hom}(I(\mu), M_i) = \mathrm{Hom}(I(\mu), M_{i-1})$. After applying this argument a finite number of times we obtain the desired result. Now, let $f \in \mathrm{Hom}(I(\mu), M)$ and let v be a generator of $I(\mu)$. $f(v) \in M_i$, for some i. If $i < r$, then $f \in \mathrm{Hom}(I(\mu), M_r)$. If $r \geqq i$, then $f \in \mathrm{Hom}(I(\mu), M_r)$, by the above. This proves our assertion. By projectivity of $I(\mu)$, $\mathrm{Hom}(I(\mu), M) = \mathrm{Hom}(I(\mu), M_1) \oplus \mathrm{Hom}(I(\mu), M_2/M_1) \oplus \ldots \oplus \mathrm{Hom}(I(\mu), M_r/M_{r-1})$. To prove the proposition we may now assume that M is a highest weight module with highest weight $\nu \in \Pi^S$. Clearly $\dim \mathrm{Hom}(I(\mu), M) = 0$, if $\nu \ngeqq \mu$. Let M' be the unique maximal submodule of M. If $\nu = \mu$, then $\mathrm{Hom}(I(\mu), M') = (0)$, by weight considerations. Hence $\dim \mathrm{Hom}(I(\mu), M) = \dim \mathrm{Hom}(I(\mu), L(\mu)) = 1$, by Lemma 4.11, and, therefore, $\dim \mathrm{Hom}(I(\mu), M) = (M: L(\mu))$. If $\nu > \mu$ we proceed by induction on $\nu - \mu$. So, we assume that the proposition has been proved for all highest weight modules with highest weight ν' such that $\nu' - \mu < \nu - \mu$. Now we can follow the recursive construction of the numbers $(M: L(\mu))$ and use the projectivity of $I(\mu)$ as above. Q.E.D.

§6. The Duality Theorem

Lemma 1. *Suppose* $M \in \mathscr{C}^S$ *has a GVCS*

$$M = M^1 \supset M^2 \supset \ldots \supset M^l \supset M^{l+1} = (0),$$

$M^i/M^{i+1} \simeq M(R_S(\mu_i))$, $\mu_i \in \Pi^S$, $i = 0, 1, \ldots, l$. *Then as an* $\mathfrak{m}_S$*-module* $M/\mathfrak{u}_S^- M \simeq \bigoplus_{i=1}^{l} R_S(\mu_i)$.

Proof. We prove by induction on l that as a $\mathfrak{u}_S^- \oplus \mathfrak{m}_S$-module $M \simeq U(\mathfrak{u}_S^-) \otimes \left(\bigoplus_{i=1}^{l} R_S(\mu_i) \right)$. If $l = 1$ this is clear. Suppose the result has been proved for $l - 1 \geqq 1$. Then we have the $\mathfrak{g}$-module exact sequence

$$0 \to M^l \to M \to M/M_l \to 0$$

which as a $\mathfrak{u}_S^- \oplus \mathfrak{m}_S$-module sequence gives

$$0 \to U(\mathfrak{u}_S^-) \otimes R_S(\mu_l) \xrightarrow{\alpha} M \xrightarrow{\beta} U(\mathfrak{u}_S^-) \otimes \bigoplus_{i=1}^{l-1} R_S(\mu_i) \to 0,$$

by the induction assumption. Let $V \subset M$ be such that β: $V \to \bigoplus_{i=1}^{l-1} R_S(\mu_i)$ is an $\mathfrak{m}_S$-isomorphism. Then β: $U(\mathfrak{u}_S^-)V \to U(\mathfrak{u}_S^-) \otimes \bigoplus_{i=1}^{l-1} R_S(\mu_i)$ is bijective. Thus $M \simeq U(\mathfrak{u}_S^-) \otimes R_S(\mu_l) \oplus U(\mathfrak{u}_S^-) \otimes \bigoplus_{i=1}^{l-1} R_S(\mu_i)$. This proves the assertion. The lemma now follows. Q.E.D.

Lemma 1 justifies the following definition.

Definition 2. Suppose $M \in \mathscr{C}^S$ has a GVCS

$$M = M^0 \supset M^1 \supset \ldots \supset M^l \supset M^{l+1} = (0),$$

$M^i/M^{i+1} \simeq M(R_S(\mu_i))$, $\mu_i \in \Pi^S$, $i = 0, 1, \ldots, l$. Let $\mu \in \Pi^S$. We define the number $(M: M(R_S(\mu)))$ to be the number of indices i in $\{0, 1, \ldots, l\}$ such that $\mu = \mu_i$.

By Proposition 4.5, $P(\mu)$ has a GVCS, for all $\mu \in \Pi^S$.

Lemma 2. *Let $\mu, \nu \in \Pi^S$. Then*

$$(P(\mu): M(R_S(\nu))) = \dim \mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), M(R_S(\nu))).$$

Proof. By the proof of Proposition 4.5,

$$(P(\mu): M(R_S(\nu))) = \dim \mathrm{Hom}_{\mathfrak{m}_S}(R_S(\nu), U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)} \otimes R_S(\mu)),$$

where k is a nonnegative integer which depends on μ and is defined in the proof of Proposition 4.1. Now,

$$\begin{aligned} &\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\nu), U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)} \otimes R_S(\mu)) \\ &\quad = \mathrm{Hom}_{\mathfrak{m}_S}((U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})^* \otimes R_S(\nu), R_S(\mu)), \end{aligned}$$

and 3.1(i) implies that

$$\begin{aligned} &\mathrm{Hom}_{\mathfrak{m}_S}((U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})^* \otimes R_S(\nu), R_S(\mu)) \\ &\quad = \mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), (U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})^* \otimes R_S(\nu)). \end{aligned}$$

By condition (S6) in §3 and weightspace considerations there is an $\mathfrak{m}_S$-invariant pairing of $(U(\mathfrak{u}_S)^{(k)})$ and $U(\mathfrak{u}_S^-)^{(k)}$, where $U(\mathfrak{u}_S^-)^{(k)} = \bigoplus_{\alpha \in kQ^+} U(\mathfrak{u}_S^-)_{-\alpha}$. This induces an $\mathfrak{m}_S$-invariant pairing of $(U(\mathfrak{u}_S)/U(\mathfrak{u}_S)^{(k)})$ and $U(\mathfrak{u}_S^-)/U(\mathfrak{u}_S^-)^{(k)}$. By the definition of k in the proof of Proposition 4.1, it is clear that

$$\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), U(\mathfrak{u}_S^-) \otimes R_S(\nu)) = \mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), U(\mathfrak{u}_S^-)/U(\mathfrak{u}_S^-)^{(k)} \otimes R_S(\nu)).$$

Now $U(\mathfrak{u}_S^-) \otimes R_S(\nu) \simeq M(R_S(\nu))$ as $\mathfrak{m}_S$-modules. This concludes the proof of the lemma.

As an immediate consequence of Lemma 2 above, we obtain, using Proposition 4.1, the following

Proposition 3. *Let* $\mu, \nu \in \Pi^S$. *Then*

$$(P(\mu): M(R_S(\nu))) = \dim \mathrm{Hom}(P(\mu), M(R_S(\nu))).$$

By Propositions 4.5 and 4.7 and by Corollary 4.13, $I(\mu)$ possesses a GVCS, for all $\mu \in \Pi^S$. The following is a generalization of Theorem 6.1 of [17] and Proposition 6.2.2 of [1].

Theorem 4. *Let* $\mu, \nu \in \Pi^S$. *Then*

$$(I(\mu): M(R_S(\nu))) = (M(R_S(\nu)): L(\mu)).$$

Proof. By Proposition 5.3 it suffices to show

$$(I(\mu): M(R_S(\nu))) = \dim \mathrm{Hom}(I(\mu), M(R_S(\nu))). \tag{1}$$

By the proof of Proposition 4.8 and by Corollary 4.13 $P(\mu) = \bigoplus_{\chi \in \Pi^S} m_\mu(\chi) I(\chi)$, and $m_\mu(\chi) < \infty$. Furthermore $m_\mu(\chi) = \dim \mathrm{Hom}(P(\mu), L(\chi)) = \mathrm{Hom}_{\mathfrak{m}_S}(R_S(\mu), L(\chi))$, by Corollary 4.13 and Proposition 4.1. Hence, $m_\mu(\mu) = 1$ and $m_\mu(\chi) \neq 0$ only if $\mu \leqq \chi$. Let $\mu_1, \ldots, \mu_s$ be the maximal weights of Π^S with respect to the partial ordering in $\mathfrak{h}^*$. Then $P(\mu_i) = I(\mu_i)$, $i = 1, \ldots, s$, and, hence, (1) holds for $\mu = \mu_i$, $i = 1, \ldots, s$, by Proposition 3. We proceed by finite induction, as follows: Suppose (1) has been proved for all $\chi \in \Pi^S$ such that $\chi > \mu$. By the above, $P(\mu) = I(\mu) \oplus \bigoplus_{\chi \in \Pi^S} m_\mu(\chi) I(\chi)$, where $m_\mu(\chi) \neq 0$ only if $\mu < \chi$. Therefore, (1) follows by the induction hypothesis and by Proposition 3. Q.E.D.

§7. Extensions and Relative Cohomology

If $M \in \mathscr{C}^S(\lambda, q)$ we denote by M^μ the isotypic component of type $R_S(\mu)$ in M, $\mu \in \Pi^S(\lambda, q)$. Then $M = \bigoplus_{\mu \in \Pi^S(\lambda, q)} M^\mu$. Allowing λ, q to vary, we let $\Pi^S(\lambda, \infty) = \bigcup_{q \in \mathbb{Z}_+} \Pi^S(\lambda, q)$ and $\Pi_{(\mathfrak{g}, \mathfrak{m}_S)} = \bigcup_\lambda \Pi^S(\lambda, \infty)$, where λ ranges through a set of representatives of $\mathfrak{h}^*/Q$. Let $\mathscr{C}_{(\mathfrak{g}, \mathfrak{m}_S)}$ denote the full subcategory of the category of $\mathfrak{g}$-modules M such that $M = \bigoplus_{\mu \in \Pi_{(\mathfrak{g}, \mathfrak{m}_S)}} M^\mu$. If $M \in \mathscr{C}_{(\mathfrak{g}, \mathfrak{m}_S)}$ let ${}_\lambda M = \bigoplus_{\mu \in \Pi^S(\lambda, \infty)} M^\mu$. Then $M = \bigoplus_\lambda {}_\lambda M$ is a direct sum of submodules, where λ ranges through a set of representatives of $\mathfrak{h}^*/Q$.

If M and N are objects in an abelian category $\mathscr{A}$, we call an exact sequence

$$E: \ 0 \to N \to E_{p-1} \to E_{p-2} \to \ldots \to E_0 \to M \to 0,$$

with $E_0, \ldots, E_{p-1}$ in $\mathscr{A}$, a *p-extension* of M by N in $\mathscr{A}$. We denote by $[E]$ the equivalence class of the p-extension E of M by N in $\mathscr{A}$, and by $\mathrm{Ext}^p_{\mathscr{A}}(M, N)$ the abelian group of all equivalence classes of p-extensions of M by N in $\mathscr{A}$ (cf. [14]). If $\mathscr{A} = \mathscr{C}_{(\mathfrak{g}, \mathfrak{m}_S)}$, we denote $\mathrm{Ext}^p_{\mathscr{A}}(M, N)$ by $\mathrm{Ext}^p_{(\mathfrak{g}, \mathfrak{m}_S)}(M, N)$.

Remark. $\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$ has enough projectives. Hence Yoneda's theorem implies that $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}$ is the usual relative Ext bifunctor.

Using the additivity of $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(\ ,\ )$ we observe that Lemma 8.1 of [17] can be extended to arbitrary abelian categories of modules. In particular, we obtain the following:

Lemma 1. *Let* $M=M_0\supset\ldots\supset M_s=M_{s+1}=(0)$ *be a filtration of* $M\in\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$ *by submodules, such that* $M_i/M_{i+1}\simeq V_i$, $i=0,1,\ldots,s$. *If* $\mathrm{Ext}^1_{(\mathfrak{g},\mathfrak{m}_S)}(V_i,V_j)$ *is trivial for all* $0\leqq i<j\leqq s$, *then* $M=\bigoplus_{i=0}^{s} V_i$.

Theorem 2 [19]. *Let* $V\in\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$, $\nu\in\Pi_{(\mathfrak{g},\mathfrak{m}_S)}$. $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(M(R_S(\nu)),V)$ *is naturally isomorphic with* $\mathrm{Hom}_{\mathfrak{m}_S}(R_S(\nu),H^p(\mathfrak{u}_S,V))$.

Proof. Let for $j=0,1,\ldots,$ and $\nu\in\Pi_{(\mathfrak{g},\mathfrak{m}_S)}$, $R_{j,\nu}=U(\mathfrak{u}_S)\otimes\Lambda^j(\mathfrak{u}_S)\otimes R_S(\nu)$ and β_j: $R_{j,\nu}\to R_{j-1,\nu}$, $j\geqq 1$, be defined as follows:

$$\begin{aligned}\beta_j(x\otimes x_1\wedge\ldots\wedge x_j\otimes v)&=\sum_{k=1}^{j}(-1)^{k+1}x x_k\otimes x_1\wedge\ldots\wedge\hat{x}_k\wedge\ldots\wedge x_j\otimes v\\&+\sum_{1\leqq k<l\leqq j}(-1)^{k+l}x\otimes[x_k,x_l]\wedge x_1\wedge\ldots\wedge\hat{x}_k\wedge\ldots\wedge\hat{x}_l\wedge\ldots\wedge x_j\otimes v.\end{aligned}$$

Let σ: $R_{0,\nu}\to R_S(\nu)$ be the trivial module action. It is well-known that

$$\ldots\to R_{j,\nu}\xrightarrow{\beta_j}\ldots\xrightarrow{\beta_1}R_{0,\nu}\xrightarrow{\sigma}R_S(\nu)\to 0$$

is an exact sequence of $U(\mathfrak{p}_S)$-modules.

Let $V\in\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$ and $f\in\mathrm{Hom}_{U(\mathfrak{p}_S)}(R_{p,\nu},V)$ be such that $f\circ\beta_{p+1}=0$. By a standard homological algebra argument we obtain the following diagram with exact rows:

$$\begin{array}{ccccccccccc}R_{p+1,\nu}&\xrightarrow{\beta_{p+1}}&R_{p,\nu}&\xrightarrow{\beta_p}&R_{p-1,\nu}&\to\ldots\xrightarrow{\beta_1}&R_{0,\nu}&\xrightarrow{\sigma}&R_S(\nu)&\to 0\\ \downarrow&&\downarrow f&&\downarrow\mu_{p-1}&&\downarrow\mu_j&&\|&\\ 0&\longrightarrow&V&\xrightarrow{\alpha_p}&E_{p-1,\nu}&\to\ldots\xrightarrow{\alpha_1}&E_{0,\nu}&\xrightarrow{\alpha_0}&R_S(\nu)&\to 0.\end{array}$$

Here (μ_{p-1},α_p) is the push-out of (β_p,f), $E_{j,\nu}=R_{j,\nu}$ and $\mu_j=\mathrm{id}$ for $j=0,\ldots,p-2$. Hence, we obtain the g-module exact sequence:

$$\begin{aligned}0\to U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}V&\xrightarrow{I\otimes\alpha_p}U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}E_{p-1,\nu}\to\ldots\\&\xrightarrow{I\otimes\alpha_1}U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}E_{0,\nu}\xrightarrow{I\otimes\alpha_0}M(R_S(\nu))\to 0.\end{aligned}$$

Let ϕ: $U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}V\to V$ be the surjective g-module homomorphism such that $\phi(1\otimes v)=v$, for all $v\in V$. Let $D=\mathrm{Ker}\,\phi$. Then, the sequence

$$\bar{E}\colon\ 0\to V\xrightarrow{\bar{\alpha}_p}\bar{E}_{p-1}\xrightarrow{\bar{\alpha}_{p-1}}\ldots\xrightarrow{\bar{\alpha}_1}\bar{E}_0\xrightarrow{\bar{\alpha}_0}M(R_S(\nu))\to 0\quad\text{in }\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$$

is exact, where $\bar{E}_{p-1}=U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}E_{p-1,\nu}/(I\otimes\alpha_p)(D)$, $\bar{E}_j=U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}E_{j,\nu}$, $\bar{\alpha}_j=I\otimes\alpha_j$, $j=0,\ldots,p-2$. $[\bar{E}]$ depends only on the cohomology class ω of f. We set $\hat{\omega}=[E]$. Conversely, let E: $0\to V\xrightarrow{\alpha_p}E_{p-1}\xrightarrow{\alpha_{p-1}}\ldots\to E_0\xrightarrow{\alpha_0}M(R_S(\nu))\to 0$ be a representative of an element of $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(M(R_S(\nu)),V)$. Let ϕ_j: $\mathrm{Im}\,\alpha_j\to E_j$, $j=0,\ldots,p$ be an $\mathfrak{m}_S$-module homomorphism such that $\alpha_j\circ\phi_j(z)=z$ for all $z\in\mathrm{Im}\,\alpha_j$. Set $u_0=\phi_0(1\otimes v_\nu)$. Then u_0 can be looked upon as an element of $\mathrm{Hom}_{U(\mathfrak{p}_S)}(R_{0,\nu},E_0)$. Let $X\in U(\mathfrak{u}_S)$, $Y\in\mathfrak{u}_S$, $(u_0\circ\beta_1)(X\otimes Y\otimes v_\nu)=u_0(XY\otimes v_\nu)=XYu_0$. Hence $(\alpha_0\circ u_0\circ\beta_1)(X\otimes Y\otimes v_\nu)=0$. Therefore $(u_0\circ\beta_1)(X\otimes Y\otimes v_\nu)\in\mathrm{Im}\,\alpha_1$. Now, set $u_1(X\otimes Y\otimes v_\nu)=(\phi_1\circ\mu_1\circ\beta_1)(X\otimes Y\otimes v_\nu)$, $X\in U(\mathfrak{u}_S)$, $Y\in\mathfrak{u}_S$. Then u_1 can be looked upon as an element of $\mathrm{Hom}_{U(\mathfrak{p}_S)}(R_{1,\nu},E_1)$. Inductively, we obtain an element f in $\mathrm{Hom}_{U(\mathfrak{p}_S)}(R_{p,\nu},V)$ such that $f\circ\beta_{p+1}=0$. The cohomology class ω of f depends only on $[E]$. We set $[E]^\vee=\omega$. Then $(\omega)^\vee=\omega$ and $([E]^\vee)^\wedge=[E]$. Q.E.D.

Lemma 3. *Let V be an object of $\mathscr{C}_{(\mathfrak{g},\mathfrak{m}_S)}$ such that $V={}_\mu V$, $\nu\in\Pi_{(\mathfrak{g},\mathfrak{m}_S)}$, and $p\in\mathbb{Z}_+$. If $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(M(R_S(\nu)),V)$ is nontrivial, then $\nu\equiv\mu \bmod Q$.*

Proof. The conclusion of the lemma is clear if $p=0$. Assume $p\geqq 1$ and let $\nu-\mu\in Q$. Let E be a non-split p-extension of $M(R_S(\nu))$ by V. By Theorem 2, we may assume that E is an exact sequence

$$0\to V\xrightarrow{\bar{\alpha}_p}\bar{E}_{p-1}\xrightarrow{\bar{\alpha}_{p-1}}\ldots\xrightarrow{\bar{\alpha}_1}\bar{E}_0\xrightarrow{\bar{\alpha}_0}M(R_S(\nu))\to 0$$

with the following properties: 1) $\bar{E}_{p-1}$ is a quotient of $U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}W$, where W is a quotient of $V\oplus R_{p-1,\nu}$; 2) if $p\geqq 2$, $\bar{E}_j=U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}R_{j,\nu}$ for $j\leqq p-2$. Hence $\bar{E}_{p-1}={}_\mu\bar{E}_{p-1}\oplus{}_\nu\bar{E}_{p-1}$, and if $p\geqq 2$, $\bar{E}_j={}_\nu\bar{E}_j$ for all $j\leqq p-2$. Therefore $\bar{\alpha}_p(V)\subset{}_\mu\bar{E}_{p-1}\subset\mathrm{Ker}\,\bar{\alpha}_{p-1}=\bar{\alpha}_p(V)$. Thus E splits, a contradiction. Q.E.D.

For each $l\in\mathbb{Z}_+$, let $(U(\mathfrak{u}_S)\otimes\Lambda^j(\mathfrak{u}_S))^{(l)}=\bigoplus_{\alpha\in lQ^+}(U(\mathfrak{u}_S)\otimes\Lambda^j(\mathfrak{u}_S))_\alpha$ and let $R^l_{j,\nu}=(U(\mathfrak{u}_S)\otimes\Lambda^j(\mathfrak{u}_S))^{(l)}\otimes R_S(\nu)$. Then $R_{j,\nu}=R^0_{j,\nu}\supset R^1_{j,\nu}\supset\ldots$ and $\bigcap_{l=0}^{\infty}R_{j,\nu}=(0)$. This implies the following

Lemma 4. *Let $\lambda\in\mathfrak{h}^*$, $q',q''\in\mathbb{Z}_+$ be fixed. If $\mu\in\Pi^S(\lambda,q')$, $\nu\in\Pi^S(\lambda,q'')$ and $p\in\mathbb{Z}_+$, then*

$$\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(M(R_S(\nu)),M(R_S(\mu)))=\mathrm{Ext}^p_{\mathscr{C}^S(\lambda,q)}(M(R_S(\nu)),M(R_S(\mu))),$$

where $q=\max\{q',q''\}$.

Following the same argument used in the proof of [15, Lemma 9.1], we apply Theorem 6.4 and [14, Theorem 5.1, Chap. VII], to obtain the following

Theorem 5. *Let $\mu,\nu\in\Pi_{(\mathfrak{g},\mathfrak{m}_S)}$. Then $\mathrm{Ext}^p_{(\mathfrak{g},\mathfrak{m}_S)}(M(R_S(\nu)),M(R_S(\mu)))$ is non-trivial only if there are distinct elements $\mu_1,\ldots,\mu_p\in\Pi_{(\mathfrak{g},\mathfrak{m}_S)}$, $\mu_1\neq\nu$, such that:*

$$(M(R_S(\mu)):L(\mu_p))(M(R_S(\mu_p)):L(\mu_{p-1}))\ldots(M(R_S(\mu_1)):L(\nu))\neq 0.$$

§8. Embeddings of Verma Modules over Kac-Moody Lie Algebras

We recall the definition and basic facts about Kac-Moody Lie algebras with symmetrizable generalized Cartan matrices. Following the point of view of [8]

we will consider the extended Kac-Moody Lie algebras. These Lie algebras have a grading of the type described in §1 and are, therefore, suitable for the treatment of highest weight modules.

We let $A=(A_{ij})$ be an $n\times n$ *generalized Cartan matrix* (GCM), i.e., $A_{ii}=2$ for all i, $-A_{ij}\in\mathbb{Z}_+$ whenever $i\neq j$, and $A_{ji}=0$ whenever $A_{ij}=0$. We let $\mathfrak{g}_1$ denote the complex Lie algebra defined by the $3n$ generators e_i, f_i, h_i, $i=1,\ldots,n$, and by the relations: $[e_i,f_j]=\delta_{ij}h_i$, $[h_i,e_j]=A_{ij}e_j$, $[h_i,f_j]=-A_{ij}f_j$, $[h_i,h_j]=0$ for all i,j, and $(\operatorname{ad}e_i)^{-A_{ij}+1}e_j=(\operatorname{ad}f_i)^{-A_{ij}+1}f_j=0$ for all $i\neq j$. It follows from the definition that $\mathfrak{g}_1$ has a natural $\mathbb{Z}^n$-gradation. We denote by $\mathfrak{g}_1(k_1,\ldots,k_n)$ the subspace spanned by the elements $[e_{i_1},[e_{i_2},\ldots,[e_{i_{r-1}},e_{i_r}]\ldots]]$ (resp. $[f_{i_1},[f_{i_2},\ldots,[f_{i_{r-1}},f_{i_r}]\ldots]]$), where e_j (resp. f_j) occurs $|k_j|$ times, for every n-tuple $(k_1,\ldots,k_n)$ of nonnegative (resp. nonpositive) integers not all zero. $\mathfrak{g}_1$ has a unique $\mathbb{Z}^n$-graded ideal $\mathfrak{r}_1$ maximal among those $\mathbb{Z}^n$-graded ideals of $\mathfrak{g}_1$ which intersect the span of $\{h_i|i=1,\ldots,n\}$ trivially. The *Kac-Moody Lie algebra* $\hat{\mathfrak{g}}$ associated with A is the $\mathbb{Z}^n$-graded Lie algebra $\mathfrak{g}_1/\mathfrak{r}_1$. We denote by e_i, f_i and h_i the canonical images in $\hat{\mathfrak{g}}$ of the elements of the same name in $\mathfrak{g}_1$. $\{e_i,f_i,h_i|i=1,\ldots,n\}$ is a linearly independent set in $\hat{\mathfrak{g}}$ and we denote by $\hat{\mathfrak{h}}$ the linear span of $\{h_i|i=1,\ldots,n\}$.

The n linear functionals on $\hat{\mathfrak{h}}$ arising from the adjoint action of $\hat{\mathfrak{h}}$ on the spaces $\hat{\mathfrak{g}}(0,\ldots,1,\ldots,0)$ are not always linearly independent. We denote by D_i the degree derivation of $\hat{\mathfrak{g}}$ that acts by the scalar k_i on $\hat{\mathfrak{g}}(k_1,\ldots,k_n)$. Let $\mathfrak{d}_0$ denote the span of $\{D_i|i=1,\ldots,n\}$. If $\mathfrak{d}$ is a subspace of $\mathfrak{d}_0$ we define the Lie algebra $\mathfrak{g}$ to be the semidirect product $\mathfrak{d}\rtimes\hat{\mathfrak{g}}$. Set $\mathfrak{h}=\mathfrak{d}\oplus\hat{\mathfrak{h}}$. Let $\alpha_i\in\mathfrak{h}^*$ be defined by $[H,e_i]=\alpha_i(H)e_i$ for all $H\in\mathfrak{h}$. We choose $\mathfrak{d}$ such that $\{\alpha_i|i=1,\ldots,n\}$ is linearly independent. If $\phi\in\mathfrak{h}^*$, then $\mathfrak{g}_\phi=\{X\in\mathfrak{g}|[H,X]=\phi(H)X \text{ for all } H\in\mathfrak{h}\}$. If $\alpha=\sum_{i=1}^n k_i\alpha_i$, $k_i\in\mathbb{Z}$, then $\mathfrak{g}_\alpha=\mathfrak{g}(k_1,\ldots,k_n)$. We define the set Δ of *roots* of $\mathfrak{g}$ to be the set of all nonzero elements α of $\mathfrak{h}^*$ such that $\mathfrak{g}_\alpha\neq 0$. Then $\mathfrak{g}=\bigoplus_{\alpha\in\Delta\cup\{0\}}\mathfrak{g}_\alpha$, every $\alpha\in\Delta\cup\{0\}$ is an element of the free abelian group with basis $\{\alpha_1,\ldots,\alpha_n\}$, $\mathfrak{h}$ is an abelian subalgebra of $\mathfrak{g}$ such that $\mathfrak{g}_0=\mathfrak{h}$, $\dim\mathfrak{g}_\alpha<\infty$ for all $\alpha\in\Delta\cup\{0\}$, and $[\mathfrak{g}_\alpha,\mathfrak{g}_\beta]\subset\mathfrak{g}_{\alpha+\beta}$ for all $\alpha,\beta\in\Delta\cup\{0\}$. If $\alpha=\sum_{i=1}^n k_i\alpha_i$ is an element of $\Delta\cup\{0\}$ then $k_i\in\mathbb{Z}^+$ for all i, or $-k_i\in\mathbb{Z}^+$ for all i. Accordingly, we define the set Δ_+ of *positive roots* to be the set of all roots of the form $\sum_{i=1}^n k_i\alpha_i$ with $k_i\in\mathbb{Z}^+$ for all i, and the set Δ_- of negative roots to be $-\Delta_+$. Let $\mathfrak{n}=\bigoplus_{\alpha\in\Delta_+}\mathfrak{g}_\alpha$, $\mathfrak{n}^-=\bigoplus_{\alpha\in\Delta_+}\mathfrak{g}_{-\alpha}$, $\mathfrak{b}=\mathfrak{h}\oplus\mathfrak{n}$. Then $\mathfrak{g}=\mathfrak{n}^-\oplus\mathfrak{h}\oplus\mathfrak{n}$.

We assume, throughout, that the GCM A is *symmetrizable*, i.e., that there are positive rational numbers $c_1,\ldots,c_n$ such that $\operatorname{diag}(c_1,\ldots,c_n)A$ is a symmetric matrix. Let R denote the linear span of $\{\alpha_1,\ldots,\alpha_n\}$ in $\mathfrak{h}^*$. We define a symmetric, bilinear form $(\ ,\ )$ on R by $(\alpha_i,\alpha_j)=c_iA_{ij}$, $i,j=1,\ldots,n$. Set $h_{\alpha_i}=c_ih_i=\dfrac{(\alpha_i,\alpha_i)}{2}h_i$, and for

$$\alpha=\sum_{i=1}^n k_i\alpha_i\in\Delta, \qquad \text{set } h_\alpha=\sum_{i=1}^n k_ih_{\alpha_i}.$$

We extend $(\ ,\ )$ to a symmetric, bilinear form $(\ ,\ )$ on $\mathfrak{h}^*$ such that $(\mu, \alpha)=\mu(h_\alpha)$ for all $\alpha\in\Delta$, $\mu\in\mathfrak{h}^*$. Define $B(h_\alpha, h_\beta)=(\alpha, \beta)$ for all $\alpha, \beta\in\Delta$. B extends to a symmetric, $\mathfrak{g}$-invariant, bilinear form on $\mathfrak{g}$, denoted by B, such that $[X, Y]=B(X, Y)h_\alpha$ for all $\alpha\in\Delta$, $X\in\mathfrak{g}_\alpha$, $Y\in\mathfrak{g}_{-\alpha}$. If $\alpha\in\Delta$ and $\beta\in\Delta\cup\{0\}$, $B(\mathfrak{g}_\alpha, \mathfrak{g}_\beta)=0$, *unless* $\beta=-\alpha$. In particular, $\mathfrak{g}$ satisfies the conditions (T1) and (T2) of §1.

For each i let $\sigma_i: \mathfrak{h}^*\to\mathfrak{h}^*$ be the linear transformation defined by $\sigma_i(\mu)=\mu-\mu(h_i)\alpha_i$, $\mu\in\mathfrak{h}^*$. The *Weyl group* W of $(\mathfrak{g}, \mathfrak{h})$ is the group of automorphisms of $\mathfrak{h}^*$ generated by the reflections σ_i, $i=1, \ldots, n$. W is a Coxeter group (cf. [13]).

We note that, for each i, the subalgebra $\mathfrak{a}_i$ spanned by $\{e_i, f_i, h_i\}$ is a copy of $\mathfrak{sl}(2, \mathbb{C})$. It follows from the defining relations of $\mathfrak{g}$ that every $\mathfrak{a}_i$-invariant submodule of $\mathfrak{g}$ is a direct sum of finite dimensional, irreducible $\mathfrak{a}_i$-modules. This implies that $W\Delta=\Delta$ and that $\dim\mathfrak{g}_\alpha=\dim\mathfrak{g}_{w\alpha}$ for all $\alpha\in\Delta$ and all $w\in W$. It is easy to show that, for each i, $\sigma_i(\Delta_+\setminus\{\alpha_i\})=\Delta_+\setminus\{\alpha_i\}$.

We define the set Δ_R of *real roots* to be the set $\{w\alpha_i | w\in W, i=1, \ldots, n\}$ and the set Δ_I of *imaginary roots* to be $\Delta\setminus\Delta_R$. It is obvious that $W\Delta_R=\Delta_R$, $W\Delta_I=\Delta_I$, and $W(\Delta_I\cap\Delta_+)=\Delta_I\cap\Delta_+$.

For each $w\in W$, define $\Phi_w=\Delta_+\cap w\Delta_-=\{\alpha\in\Delta_+ | w^{-1}\alpha\in\Delta_-\}$. Let $l(w)$ be the length of w, that is, the smallest nonnegative integer r such that $w=\sigma_{i_1}\sigma_{i_2}\ldots\sigma_{i_r}$. Then the number of elements of Φ_w is $l(w)$ (cf. [8]). Let $\varepsilon(w)=(-1)^{l(w)}$, for all $w\in W$. ε is a homomorphism of groups.

Let $\mu\in\mathfrak{h}^*$. μ is said to be *integral* if $\mu(h_i)\in\mathbb{Z}$, for all $i=1, \ldots, n$, and μ is said to be *dominant integral* if $\mu(h_i)\in\mathbb{Z}_+$ for all $i=1, \ldots, n$. Let P (respectively P^+) denote the set of integral (respectively dominant integral) elements of $\mathfrak{h}^*$.

Let $\rho\in\mathfrak{h}^*$ be any fixed element such that $\rho(h_i)=1$ for all $i=1, \ldots, n$.

If $\alpha\in\Delta_R$ we let $\sigma_\alpha\in\mathrm{End}(\mathfrak{h}^*)$ be defined by: $\sigma_\alpha(\mu)=\mu-\dfrac{2(\mu, \alpha)}{(\alpha, \alpha)}\alpha$. Let $i\in\{1, \ldots, n\}$ and $w\in W$ be such that $\alpha=w\alpha_i$. Then it is easy to check that $\sigma_\alpha=w\sigma_i w^{-1}$. In particular $\sigma_\alpha\in W$, if $\alpha\in\Delta_R$.

We recall a useful characteristization of imaginary roots.

Lemma 1 ([10, Lemma II, §1, 14). *The following conditions are equivalent:*

(a) $\alpha\in\Delta_I$,

(b) $(\alpha, \alpha)\leqq 0$.

Extending the argument of [5, Lemma 7.7.2] we obtain

Lemma 2. *Let $\mu\in P^+$, $w\in W$, and $\beta\in\Delta_+$. The following conditions are equivalent:*

(i) $2w(\mu+\rho)(h_\beta)=n(\beta, \beta)$ *with* $n\in\mathbb{N}$

(ii) $\beta\in\Delta_R$ *and* $l(\sigma_\beta w)>l(w)$.

Proof. Suppose (i) holds. If $(\beta, \beta)=0$ then $(\mu+\rho, w^{-1}\beta)=w(\mu+\rho)(h_\beta)=0$. Hence $w^{-1}\beta=0$, contradicting the assumption that $\beta\in\Delta$. If $(\beta, \beta)<0$, then we conclude that $(\mu+\rho, w^{-1}\beta)$ is a negative number. Therefore $w^{-1}\beta\in\Delta_-$. This implies that $\beta\in\Delta_R$ and, hence, that $(\beta, \beta)>0$, by Lemma 1. Thus, we obtain a contradiction. This shows that $(\beta, \beta)>0$ and hence $\beta\in\Delta_R$ by Lemma 1.

It is obvious that $\beta\notin\Phi_w$ and, hence, that $\beta\in\Phi_{\sigma_\beta w}$. Let $S=\{\psi\in\Phi_w | \sigma_\beta\psi\in\Delta_+\}$ and $S'=\{\psi\in\Phi_w | \sigma_\beta\psi\in\Delta_-\}$. If $\psi\in S$ then $(\sigma_\beta w)^{-1}\sigma_\beta\psi=w^{-1}\psi\in\Delta_-$. Hence

$\sigma_\beta S \subset \Phi_{\sigma_\beta w}$. If $\psi \in S'$ then $\sigma_\beta \psi = \psi - 2\dfrac{(\psi,\beta)}{(\beta,\beta)}\beta \in \Delta_-$. Hence $p = 2\dfrac{(\psi,\beta)}{(\beta,\beta)} < 0$. This implies that

$$(\sigma_\beta w)^{-1}\psi = (\sigma_\beta w)^{-1}\left(\sigma_\beta \psi + 2\frac{(\psi,\beta)}{(\beta,\beta)}\beta\right) = w^{-1}\psi - 2\frac{(\psi,\beta)}{(\beta,\beta)}w^{-1}\beta \in \Delta_- - p\Delta_-.$$

Hence $S' \subset \Phi_{\sigma_\beta w}$. Since $\sigma_\beta(\sigma_\beta S) \subset \Delta_+$ and $\sigma_\beta S' \subset \Delta_-$, $\sigma_\beta S$ and S' are disjoint and $\Phi_{\sigma_\beta w} \supset \sigma_\beta S$, $\Phi_{\sigma_\beta w} \supset S'$. Clearly $\Phi_w = S \cup S'$ (disjoint union). Also, $\beta \in \Phi_{\sigma_\beta w}$, $\beta \notin S'$ and $\beta \notin \sigma_\beta S$. Therefore $l(\sigma_\beta w) = \#\Phi_{\sigma_\beta w} > \#\Phi_w = l(w)$ and (i) $\Rightarrow$ (ii). Conversely, suppose (ii) holds. Then

$$\sigma_\beta w(\mu+\rho) = w(\mu+\rho) - 2\frac{w(\mu+\rho)(h_\beta)}{(\beta,\beta)}\beta.$$

Since $\mu \in P^+$, $2\dfrac{w(\mu+\rho)(h_\beta)}{(\beta,\beta)} = n$, $n \in \mathbb{Z}\setminus\{0\}$. If $n<0$ then $2\sigma_\beta w(\mu+\rho)(h_\beta) = -n(\beta,\beta)$ with $-n \in \mathbb{N}$. By the above, $l(w) > l(\sigma_\beta w)$, a contradiction. Hence $n \in \mathbb{N}$ and (ii) implies (i). Q.E.D.

The following is the usual Bruhat order of a Coxeter pair, here defined on W.

Definition 3. Let $w, w' \in W$ and $\beta \in \Delta_+ \cap \Delta_R$. We write $w \xleftarrow{\beta} w'$ if $w = \sigma_\beta w'$ and $l(w) = l(w') + 1$. We write $w \leftarrow w'$ if there is some $\gamma \in \Delta_+ \cap \Delta_R$ such that $w \xleftarrow{\gamma} w'$. We define the following partial ordering on W: $w \leqq w'$ if $w = w'$ or if there are $w_1, \ldots, w_r \in W$ such that

$$w = w_1 \leftarrow w_2 \leftarrow \ldots \leftarrow w_r = w'.$$

We can easily adapt the arguments in the proofs of Lemmas 7.7.4, 7.7.5 and 7.7.6 of [5], to obtain the following three lemmas:

Lemma 4. *Let $w_1, w_2 \in W$, $\beta \in \Delta_+ \cap \Delta_R$, $i \in \{1, \ldots, n\}$, $\alpha_i \neq \beta$. The following conditions are equivalent:*

(i) $\sigma_i w_1 \xleftarrow{\alpha} w_1$ *and* $\sigma_i w_1 \xleftarrow{\beta} w_2$,

(ii) $w_2 \xleftarrow{\alpha_i} \sigma_i w_2$ *and* $w_1 \xleftarrow{\sigma_i \beta} \sigma_i w_2$.

Lemma 5. *Let $w \in W$ and $\beta \in \Delta_+ \cap \Delta_R$ be such that $l(w) > l(\sigma_\beta w)$. Then $w < \sigma_\beta w$.*

Lemma 6. *Let $w_1, w_2 \in W$. The number of elements $w \in W$ such that $w_1 \leftarrow w \leftarrow w_2$ is 0 or 2.*

The following result gives a description of the irreducible subquotients of a Verma module with highest weight in $W(P^+ + \rho) - \rho$. If $w \in W$ and $\mu \in \mathfrak{h}^*$ we set $w \cdot \mu = w(\mu + \rho) - \rho$.

Lemma 7. *Let* $\mu\in P^+$, $w, w'\in W$, $w\neq w'$. *Then*

(i) $(M(w\cdot\mu)\colon L(w'\cdot\mu))\neq 0$ *if and only if* $w'<w$.

(ii) *If* $(M(w\cdot\mu)\colon L(w'\cdot\mu))\neq 0$ *then there are* $w_0, w_1, \ldots, w_r\in W$ *such that* $l(w_{i+1})=l(w_i)-1$, $i=0,1,\ldots,r-1$, $w_0=w'$, $w_r=w$ *and*

$$(M(w_r\cdot\mu)\colon L(w_{r-1}\cdot\mu))(M(w_{r-1}\cdot\mu)\colon L(w_{r-2}\cdot\mu))\ldots(M(w_1\cdot\mu)\colon L(w_0\cdot\mu))\neq 0.$$

Proof. By [11, Theorem 4.2], $(M(w\cdot\mu)\colon L(w'\cdot\mu))\neq 0$ if and only if there are a sequence of positive roots $\beta_1,\ldots,\beta_r$ and a sequence of positive integers $m_1,\ldots,m_r$ such that

$$2w(\mu+\rho)(h_{\beta_r})=m_r(\beta_r,\beta_r),$$

$$2(w(\mu+\rho)-m_r\beta_r-\ldots-m_{i+1}\beta_{i+1})(h_{\beta_i})=m_i(\beta_i,\beta_i)\quad\text{for } i=1,\ldots,r-1,$$

and $w'(\mu+\rho)=w(\mu+\rho)-\sum_{i=1}^{r} m_i\beta_i$. Suppose such sequences exist. By Lemma 2, $\beta_i\in\Delta_+\cap\Delta_R$, for all $i=1,\ldots,r$, $\sigma_{\beta_i}\sigma_{\beta_{i+1}}\ldots\sigma_{\beta_r}w(\mu+\rho)=w(\mu+\rho)-m_r\beta_r-\ldots-m_{i+1}\beta_{i+1}-m_i\beta_i$ for all $i=1,\ldots,r-1$, and $l(\sigma_{\beta_i}\ldots\sigma_{\beta_r}w)>l(\sigma_{\beta_{i+1}}\ldots\sigma_{\beta_r}w)$, $i=1,\ldots,r-1$. Now set $w_{i-1}=\sigma_{\beta_i}\ldots\sigma_{\beta_r}w$, $i=1,\ldots,r$, $w_r=w$. Then $w'=w_0<\ldots<w_r=w$, by Lemma 5. Conversely, suppose $w'<w$. Then there are $w_0, w_1, \ldots, w_r\in W$ such that

$$w'=w_0\xleftarrow{\beta_1}w_1\leftarrow\ldots\xleftarrow{\beta_r}w_r=w,\qquad \beta_i\in\Delta_+\cap\Delta_r,\qquad i=1,\ldots,r.$$

Then, by Lemma 2, there are positive integers $m_1,\ldots,m_r$ such that

$$2w(\mu+\rho)(h_{\beta_r})=m_r(\beta_r,\beta_r),$$

$$2(w(\mu+\rho)-m_r\beta_r-\ldots-m_{i+1}\beta_{i+1})(h_{\beta_i})=m_i(\beta_i,\beta_i),\qquad i=1,\ldots,r-1$$

and $w'(\mu+\rho)=w(\mu+\rho)-\sum_{i=1}^{r} m_i\beta_i$. (i) and (ii) now follow. Q.E.D.

We now recall the construction of completion of modules, due to Enright. Let $\mathfrak{a}=\mathfrak{sl}(2,\mathbb{C})$ and denote by $\{e,f,h\}$ its standard basis, satisfying: $[e,f]=h$, $[h,e]=2e$ and $[h,f]=-2f$. Then $\mathfrak{a}=\mathfrak{a}_{-\alpha}\oplus\mathbb{C}h\oplus\mathfrak{a}_\alpha$, where $\alpha\in(\mathbb{C}h)^*$ is defined by the condition: $\alpha(h)=2$. We identify $(\mathbb{C}h)^*$ with $\mathbb{C}$ so that α corresponds to 2 under this identification.

Let M be an $\mathfrak{a}$-module. We define $M_\mu=\{v\in M\mid hv=\mu v\}$, for each $\mu\in\mathbb{C}$, as in §1. For each $\mu\in\mathbb{C}$, let $M(\mu)$ denote the Verma module associated with $\mathfrak{a}$, $\mathbb{C}h$, α and μ (cf. §2).

Definition 8 ([6, Definition 3.2]). Let $\mathcal{I}(\mathfrak{a})$ denote the full subcategory of the category of $\mathfrak{a}$-modules consisting of all $\mathfrak{a}$-modules M such that (i) M is a weight module, (ii) f acts injectively on M and (iii) e acts locally nilpotently on M (i.e., for each $v\in M$ there is $m\in\mathbb{N}$ such that $(\operatorname{ad}e)^m v=0$).

Definition 9 ([6, Definition 3.1]). Let $M\in\mathcal{I}(\mathfrak{a})$. An object $C_{\mathfrak{a}}(M)$ of $\mathcal{I}(\mathfrak{a})$ together with an $\mathfrak{a}$-module injection $i_M\colon M\to C_{\mathfrak{a}}(M)$ is said to be a *completion* of

M if (i) given $m\in\mathbb{Z}_+$ and an $\mathfrak{a}$-module injection $M(-m-2)\to M$ there is a unique $\mathfrak{a}$-module injection $M(m)\to C_{\mathfrak{a}}(M)$ such that the following diagram commutes:

$$\begin{array}{ccc} M(-m-2) & \longrightarrow & M \\ \downarrow & & \downarrow{\scriptstyle i_M} \\ M(m) & \longrightarrow & C_{\mathfrak{a}}(M), \end{array} \quad \text{and}$$

(ii) $C_{\mathfrak{a}}(M)/M$ is $U(\mathfrak{a})$-finite.

An $\mathfrak{a}$-module M in $\mathscr{I}(\mathfrak{a})$ is *complete* if $M\simeq C_{\mathfrak{a}}(M)$.

We now apply the above construction to the subalgebras $\mathfrak{a}_i$ of $\mathfrak{g}$ spanned by $\{e_i,f_i,h_i\}$, $i=1,\dots,n$. If M is a $\mathfrak{g}$-module we denote by $M_{\mathfrak{a}_i}$ the underlying $\mathfrak{a}_i$-module and by $\mathscr{I}_{\mathfrak{g}}(\mathfrak{a}_i)$ the full subcategory of the category of $\mathfrak{g}$-modules M such that $M_{\mathfrak{a}_i}\in\mathscr{I}(\mathfrak{a}_i)$.

Lemma 10. *Let i be fixed. Let M be a $\mathfrak{g}$-module, and set $M'=\{v\in M\mid e_i^m v=0$ for some $m\in\mathbb{Z}_+\}$ (respectively $M''=\{v\in M\mid f_i^m v=0$ for some $m\in\mathbb{Z}_+$). Then M' (respectively M'') is a $\mathfrak{g}$-submodule of M.*

Proof. $\mathfrak{g}$ is a direct sum of finite dimensional irreducible $\mathfrak{a}_i$-modules under $\mathrm{ad}|_{\mathfrak{a}_i}$. Hence $\mathrm{ad}\,e_i$ acts locally nilpotently on $\mathfrak{g}$. Let $X\in\mathfrak{g}$, $s\in\mathbb{Z}_+$ and $t\in\mathbb{Z}_+$ be such that $(\mathrm{ad}\,e_i)^s X=0$ and $e_i^t v=0$. Let $l=s+t$. Then

$$e_i^l X v=\sum_{j=0}^{l}\binom{l}{j}((\mathrm{ad}\,e_i)^j X)\,e_i^{l-j}v=0.$$

This proves that M' is a $\mathfrak{g}$-submodule of M. The same argument applied to $\mathrm{ad}\,f_i$ proves the statement of the proposition for M''. Q.E.D.

Using Lemma 10 we can extend the arguments of [6, Propositions 3.6 and 3.15] to obtain the following two propositions:

Proposition 11. (i) *Let $M\in\mathscr{I}_{\mathfrak{g}}(\mathfrak{a}_i)$. Then the completion $C_{\mathfrak{a}_i}(M)$ of M exists and is unique up to isomorphism in $\mathscr{I}_{\mathfrak{g}}(\mathfrak{a}_i)$.*

(ii) *Let $M,N\in\mathscr{I}_{\mathfrak{g}}(\mathfrak{a}_i)$, $(C_{\mathfrak{a}_i}(M),i_M)$ and $(C_{\mathfrak{a}_i}(N),i_N)$ the completions of M and N, respectively, and T: $M\to N$ a $\mathfrak{g}$-module map. Then there is a $\mathfrak{g}$-module map $C_{\mathfrak{a}_i}(T)$: $C_{\mathfrak{a}_i}(M)\to C_{\mathfrak{a}_i}(N)$ such that the following diagram commutes:*

$$\begin{array}{ccc} M & \xrightarrow{T} & N \\ {\scriptstyle i_M}\downarrow & & \downarrow{\scriptstyle i_N} \\ C_{\mathfrak{a}_i}(M) & \xrightarrow{C_{\mathfrak{a}_i}(T)} & C_{\mathfrak{a}_i}(N). \end{array}$$

Proposition 12 [6]. *Let $M\in\mathscr{I}_{\mathfrak{g}}(\mathfrak{a}_i)$ and F a finite dimensional $\mathfrak{g}$-module. Then $C_{\mathfrak{a}_i}(F\otimes M)\cong F\otimes C_{\mathfrak{a}_i}(M)$.*

We use Proposition 12 as in the proof of [6, Proposition 3.16] to obtain

Proposition 13. *Let $\mu\in\mathfrak{h}^*$. Then $C_{\alpha_i}(M(\mu))\cong M(\sigma_i\cdot\mu)$ or $M(\mu)$ according to whether $\mu(h_i)$ belongs to $\{-2,-3,\ldots\}$ or not.*

Lemma 14. *Let $\mu\in P^+$, $w,w'\in W$. Then* $\dim \mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(w\cdot\mu))\leqq 1$.

Proof. The statement of the lemma is clear if $w=w'=1$. Assume that the statement is true if $w=1$ and $l(w')<k$. Suppose $l(w')=k$. Let $i\in\{1,\ldots,n\}$ be such that $l(\sigma_i w')<l(w')$. By Lemma 2 $w'\cdot\mu(h_i)\in\{-2,-3,\ldots\}$. By Proposition 13, $C_{\alpha_i}(M(w'\cdot\mu))=M(\sigma_i w'\cdot\mu)$. Let f be a nonzero element of $\mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(\mu))$. By Proposition 11(i), there is a $\mathfrak{g}$-module map $C_{\alpha_i}(f)\colon M(\sigma_i w'\cdot\mu)\to M(\mu)$ such that

$$\begin{array}{ccc} M(w'\cdot\mu) & \xrightarrow{\ f\ } & M(\mu) \\ {\scriptstyle i_{M(w'\cdot\mu)}}\downarrow & & \| \\ M(\sigma_i w'\cdot\mu) & \xrightarrow{\ C_{\alpha_i}(f)\ } & M(\mu) \end{array}$$

is a commutative diagram, where $i_{M(w'\cdot\mu)}(1\otimes 1)=f_i^{\sigma_i w'\cdot\mu(h_i)+1}(1\otimes 1)$. By the induction hypothesis $\dim \mathrm{Hom}_{\mathfrak{g}}(M(\sigma_i w'\cdot\mu), M(\mu))\leqq 1$. Hence $\dim \mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(\mu))\leqq 1$ for all $w'\in W$. Applying the argument of [5, Proposition 7.6.8] we see that $\dim \mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(\mu))=1$ for all $w'\in W$. From this it is clear that $\dim \mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(w\cdot\mu))\leqq 1$ for all $w,w'\in W$. Q.E.D.

By the above we now write $M(w'\cdot\mu)\subset M(w\cdot\mu)$, whenever

$$\mathrm{Hom}_{\mathfrak{g}}(M(w'\cdot\mu), M(w\cdot\mu))\neq 0, \qquad w,w'\in W,\ \mu\in P^+.$$

Using Lemmas 6 and 7 we obtain the following generalization of [5, Theorem 7.7.7]:

Theorem 15. *Let $\mu\in P^+$, $w,w'\in W$. Then:*

(i) $M(w'\cdot\mu)\subset M(w\cdot\mu) \Leftrightarrow w'\leqq w \Leftrightarrow (M(w\cdot\mu): L(w'\cdot\mu))\neq 0$.

(ii) *If $M(w'\cdot\mu)\subset M(w\cdot\mu)$ then there are $w_0,w_1,\ldots,w_r\in W$ such that $l(w_{i+1})=l(w_i)-1$, $i=0,1,\ldots,r-1$, $w_0=w'$, $w_r=w$, and $M(w_0\cdot\mu)\subset M(w_1\cdot\mu)\subset\ldots\subset M(w_r\cdot\mu)$.*

(iii) *If $l(w)=l(w')-2$, the number of $w''\in W$ such that $M(w'\cdot\mu)\subset M(w''\cdot\mu)\subset M(w\cdot\mu)$, $M(w'\cdot\mu)\neq M(w''\cdot\mu)\neq M(w\cdot\mu)$ is 0 or 2.*

Proof. It remains to prove only that whenever $w_1\leftarrow w_2$ then $M(w_1\cdot\mu)\subset M(w_2\cdot\mu)$. This follows using the same argument used in the proof of [5, Lemma 7.6.11] with the following modification: In order to prove [5, Lemma 7.6.9] for $\mathfrak{n}^-$ and f_i, which is needed in the argument, we replace the filtration used for the filtration obtained by choosing a basis $u_0=1, u_1, u_2,\ldots$ of $U(\mathfrak{n}^-)$ and then setting $\mathcal{U}_0=\mathbb{C}$, $\mathcal{U}_1=\mathbb{C}+\mathrm{ad}\,U(\mathbb{C}f_i)u_1,\ldots,\mathcal{U}_{j+1}=\mathcal{U}_j+\mathrm{ad}\,U(\mathbb{C}f_i)u_{j+1},\ldots$. Q.E.D.

§9. The Generalized Strong BGG Resolution for Kac-Moody Lie Algebras

Throughout this section, $\mathfrak{g}$ will denote an extended Kac-Moody Lie algebra with symmetrizable generalized Cartan matrix $A=(A_{ij})$; we will keep the notation of the previous section.

We recall the construction of the generalized weak BGG resolution following [8]. Let S be a subset of $\{1,\dots,n\}$ such that (A_{ij}), $i,j\in S$, is a classical Cartan matrix. The Kac-Moody Lie algebra associated with the submatrix (A_{ij}), $i,j\in S$, of A, is a finite dimensional, complex simple Lie subalgebra of $\hat{\mathfrak{g}}$. Let $\mathfrak{h}_S$ be the linear span of $\{h_i | i\in S\}$, $\Delta_S=\Delta\cap\sum_{i\in S}\mathbb{Z}\alpha_i$, $\Delta_S^+=\Delta^+\cap\Delta_S$, $\Delta_S^-=\Delta^-\cap\Delta_S$. Set $\mathfrak{n}_S=\bigoplus_{\alpha\in\Delta_S^+}\mathfrak{g}_\alpha$, $\mathfrak{n}_S^-=\bigoplus_{\alpha\in\Delta_S^+}\mathfrak{g}_{-\alpha}$, $\mathfrak{u}_S=\bigoplus_{\alpha\in\Delta^+\setminus\Delta_S^+}\mathfrak{g}_\alpha$ and $\mathfrak{u}_S^-=\bigoplus_{\alpha\in\Delta^+\setminus\Delta_S^+}\mathfrak{g}_{-\alpha}$. Then $\mathfrak{g}_S=\mathfrak{n}_S^-\oplus\mathfrak{h}_S\oplus\mathfrak{n}_S$. Set $\mathfrak{h}^S=\bigcap_{i\in S}\operatorname{Ker}\alpha_{i|\mathfrak{h}_S}$, $\mathfrak{a}_S=\mathfrak{d}\oplus\mathfrak{h}^S$, $\mathfrak{m}_S=\mathfrak{n}_S^-\oplus\mathfrak{a}_S\oplus\mathfrak{h}_S\oplus\mathfrak{n}_S$, $\mathfrak{p}_S=\mathfrak{m}_S\oplus\mathfrak{u}_S$. Then (S1)–(S6) of §3 are satisfied (cf. the remark following (S6) in §3). $\mathfrak{p}_S$ is called the *F-parabolic subalgebra* of $\mathfrak{g}$ defined by S.

Let $\mathscr{A}$ denote the full subcategory of the category of $\mathfrak{g}$-modules M such that $M=\bigoplus_{\nu\in\mathfrak{h}^*}M_\nu$, $\dim M_\nu<\infty$, and ν is in a finite union of sets of the form $\Pi(\mu,0)$, $\mu\in\mathfrak{h}^*$ (cf. §2 for the definition of $\Pi(\mu,0)$). Clearly, a highest weight module is in $\mathscr{A}$. We now recall the definition of the Casimir operator Γ_M of $M\in\mathscr{A}$ ([9], cf. [8]). If $\alpha\in\Delta$, let $\{X_1,\dots,X_d\}$ be a basis of $\mathfrak{g}_\alpha$, $\{Z_1,\dots,Z_d\}$ a dual basis relative to the bilinear form B. Set $\Gamma_\alpha=\sum_{i=1}^{d}Z_iX_i\in U(\mathfrak{g})$. Then Γ_α is independent of the basis of $\mathfrak{g}_\alpha$ used in its definition. If $M=\bigoplus_{\nu\in\mathfrak{h}^*}M_\nu\in\mathscr{A}$, we define the *Casimir operator* Γ_M *of* M by the formula:

$$\Gamma_{M|M_\nu}=(\nu+\rho,\nu+\rho)I+2\sum_{\alpha\in\Delta^+}\Gamma_\alpha$$

$\Gamma_M\in\operatorname{End}_{\mathfrak{g}}(M)$ for all $M\in\mathscr{A}$ ([9], cf. [8]). It follows immediately that the Casimir operator of a highest weight module with highest weight $\nu\in\mathfrak{h}^*$ acts by scalar multiplication by $(\nu+\rho,\nu+\rho)$.

If $M\in\mathscr{A}$, let $\Theta(M)=\{\theta\in\mathbb{C}\,|\,\Gamma_M v=\theta v$ for some nonzero $v\in M\}$. If $\theta\in\mathbb{C}$, let

$$M_{(\theta)}=\{v\in M\,|\,(\Gamma_M-\theta)^n v=0 \text{ for some } n\in\mathbb{N}\}.$$

Then $\Theta(M)=\{\theta\in\mathbb{C}\,|\,M_{(\theta)}\neq 0\}$ and $M=\bigoplus_{\theta\in\Theta(M)}M_{(\theta)}$ [8, Proposition 4.5].

A highest weight module with a highest weight vector v such that $f_i^m\cdot v=0$ for some $m\in\mathbb{N}$ and all $i=1,\dots,n$ is said to be a *standard* module. The highest weight of a standard module lies in P^+ ([9], cf. [8, Proposition 6.2]).

Let D_j denote the $\mathfrak{g}$-module $U(\mathfrak{g})\otimes_{U(\mathfrak{p}_S)}\Lambda^j(\mathfrak{g}/\mathfrak{p}_S)$, $j\in\mathbb{Z}_+$. Let $\partial_j\colon D_j\to D_{j-1}$ be defined by

$$\partial_j(X\otimes X_1\wedge\dots\wedge X_j)=\sum_{i=1}^{j}(-1)^{i+1}XY_i\otimes X_1\wedge\dots\wedge\hat{X}_i\wedge\dots\wedge X_j$$

$$+\sum_{1\le k<l\le j}(-1)^{k+l}X\otimes\overline{[Y_k,Y_l]}\wedge X_1\wedge\dots\wedge\hat{X}_k\wedge\dots\wedge\hat{X}_l\wedge\dots\wedge X_j,$$

where $\bar{Y}$ represents the canonical image of $Y\in\mathfrak{g}$ in $\mathfrak{g}/\mathfrak{p}_S$, and $\bar{Y}_i=X_i$, $i=1,\dots,j$, $j\in\mathbb{N}$. Let $\varepsilon\colon D_0\to\mathbb{C}$ be defined by $\varepsilon(X)=$the constant term of X, $X\in U(\mathfrak{g})$. The following is an exact complex of $\mathfrak{g}$-modules

$$\dots\to D_j\xrightarrow{\partial_j}D_{j-1}\dots\to D_1\xrightarrow{\partial_1}D_0\xrightarrow{\varepsilon}\mathbb{C}\to 0$$

[2, Theorem II.9.1]. We denote this exact complex by $V(\mathfrak{g}, \mathfrak{p}_S)$. Let X be a standard module. Set $D_j^X = U(\mathfrak{g}) \otimes_{U(\mathfrak{p}_S)} (\Lambda^j(\mathfrak{g}/\mathfrak{p}_S) \otimes X)$, $j \in \mathbb{Z}_+$ and let $V(\mathfrak{g}, \mathfrak{p}_S) \otimes X$ be the exact complex of tensor product $\mathfrak{g}$-modules:

$$\ldots \to D_j \otimes X \xrightarrow{\partial_j \otimes 1} D_{j-1} \otimes X \to \ldots \to D_1 \otimes X \xrightarrow{\partial_1 \otimes 1} D_0 \otimes X \xrightarrow{\varepsilon \otimes 1} X \to 0.$$

Now, D_j^X is naturally isomorphic with $D_j \otimes X$, $j \in \mathbb{Z}_+$ (cf. [8, Proposition 1.7]). Making this identification we abuse notation and write

$$\ldots \to D_j^X \xrightarrow{\partial_j} D_{j-1}^X \to \ldots \to D_1^X \xrightarrow{\partial_1} D_0^X \xrightarrow{\varepsilon} X \to 0$$

for the exact complex so obtained. We denote it by $V(\mathfrak{g}, \mathfrak{p}_S, X)$.

Let μ be the highest weight of X. Then $\Theta(X) = \{(\mu+\rho, \mu+\rho)\}$. It is easy to see that $D_j^X \in \mathscr{A}$, $j \in \mathbb{Z}_+$. Projecting D_j^X onto $(D_j^X)_{((\mu+\rho, \mu+\rho))}$ we obtain a new exact complex. We abuse notation and write

$$(1) \qquad \ldots \to (D_j^X)_{((\mu+\rho, \mu+\rho))} \xrightarrow{\partial_j} (D_{j-1}^X)_{((\mu+\rho, \mu+\rho))} \xrightarrow{\partial_{j-1}} \ldots$$

$$\ldots \to (D_1^X)_{((\mu+\rho, \mu+\rho))} \xrightarrow{\partial_1} (D_0^X)_{((\mu+\rho, \mu+\rho))} \xrightarrow{\varepsilon} X \to 0$$

for the resulting exact complex of $\mathfrak{g}$-modules.

Set $P_S^+ = \{\nu \in \mathfrak{h}^* \mid \nu(h_i) \in \mathbb{Z}_+ \text{ for all } i \in S\}$. Then $P_S^+ = \bigcup_{\mu', q'} \Pi_S^+(\mu', q')$ by [8, Proposition 3.1]. Let W_S denote the subgroup of W generated by $\{\sigma_i \mid i \in S\}$, and let $W^S = \{w \in W \mid w^{-1} \Delta_S^+ \subset \Delta^+\}$. Then every element of W can be uniquely expressed in the form ww', where $w \in W_S$ and $w' \in W^S$ ([8]). Let $W^{(j)} = \{w \in W \mid l(w) = j\}$, $j \in \mathbb{Z}_+$, and set $(W^S)^{(j)} = W^S \cap W^{(j)}$, $j \in \mathbb{Z}_+$. Let $\Psi_\mu^j = \{w \cdot \mu \mid w \in (W^S)^{(j)}\}$, $j \in \mathbb{Z}_+$. Clearly X, $D_j^X \in \mathscr{C}^S(\mu, q)$ for any $q \in \mathbb{Z}_+$ all $j \in \mathbb{Z}_+$. Also, $\Psi_\mu^j \subset \Pi^S(\mu, q)$ for every $q \in \mathbb{Z}_+$, (cf. [8, Proposition 6.1]). For simplicity we choose $q = 0$.

The following is contained in [8, Theorem 8.6].

Theorem 1 ([8, Theorem 8.6]). *For each $j \in \mathbb{Z}_+$, there are submodules of $(D_j^X)_{((\mu+\rho, \mu+\rho))}$:*

$$(D_j^X)_{((\mu+\rho, \mu+\rho))} = M^0 \supset M^1 \supset \ldots \supset M^r \supset M^{r+1} = (0)$$

such that

$$M^i / M^{i+1} \simeq M(R_S(\mu_i)), \quad \text{where } \{\mu_i\} = \Psi_\mu^j.$$

The exact sequence (1) is known as the generalized weak BGG resolution. The next theorem generalized Theorem 9.3 of [15]:

Theorem 2. *Let X be a standard module with highest weight μ. Then*

$$(D_j^X)_{((\mu+\rho, \mu+\rho))} = \oplus M(R_S(\mu_i)), \text{ where } \{\mu_i\} = \Psi_\mu^j.$$

Proof. By Theorem 1 and Lemma 7.1 it suffices to prove that $\mathrm{Ext}^1_{(\mathfrak{g}, \mathfrak{m}_S)}(M(R_S(\mu_i)), M(R_S(\mu_j))) = 0$ for all $0 \leqq i < j < r$. Suppose $\mathrm{Ext}^1_{(\mathfrak{g}, \mathfrak{m}_S)}(M(R_S(\mu_i)), M(R_S(\mu_j)))$ is not trivial. By Theorem 7.5,

$(M(R_S(\mu_j))\colon L(v_1))(M(R_S(v_1))\colon L(\mu_i)) \neq 0$. Now, $\mu_j = w\cdot\mu$ and $\mu_i = w'\cdot\mu$, for some $w, w' \in (W^S)^{(j)}$, $w \neq w'$. By Lemma 8.7 $w' < w'' \leqq w$, where $v_1 = w''\cdot\mu$, a contradiction, since $l(w') = l(w)$. Q.E.D.

We assume temporarily that $S = \varnothing$. In this case Theorem 2 says that

$$(D_j^X)_{((\mu+\rho,\mu+\rho))} = \bigoplus_{w\in W^{(j)}} M(w\cdot\mu).$$

Abusing notation we write

$$\ldots \to C_j \xrightarrow{\partial_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{\partial_1} C_0 \xrightarrow{\varepsilon} X \to 0,$$

where $C_j = \bigoplus_{w\in W^{(j)}} M(w\cdot\mu)$.

The same argument used in Lemma 10.1 of [17] can be used to prove the following

Lemma 3. *Let*

$$\ldots \to C_j \xrightarrow{\alpha_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{\alpha_1} C_0 \xrightarrow{\alpha_0} X \to 0$$

be an exact complex, $j\in\mathbb{Z}_+$. Then $\alpha_{j|_{M(w\cdot\mu)}}$ is injective for all $w\in W^{(j)}$, $j\in\mathbb{Z}_+$.

By Theorem 8.15 we can fix injections $i_{w,1}\colon M(w\cdot\mu)\to M(\mu)$ for all $w\in W$ and, hence, injections $i_{w',w}\colon M(w'\cdot\mu)\to M(w\cdot\mu)$ for all $w'\leqq w$. By Lemma 8.14 every g-homomorphism of C_j into C_{j-1} is of the form $\oplus c^j_{w_1,w_2} i_{w_1,w_2}$, with $c^j_{w_1,w_2}\in\mathbb{C}$ and the sum over all $w_1\in W^{(j)}$, $w_2\in W^{(j-1)}$, $j\in\mathbb{N}$.

Using Lemmas 8.4 and 8.6 and Theorem 8.15 as in the proof of [17, Lemma 10.2] we obtain the following:

Lemma 4. *Let $\ldots \to C_j \xrightarrow{\alpha_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{\alpha_1} C_0 \xrightarrow{\alpha_0} X \to 0$ be a complex, $\alpha_j = \oplus a^j_{w_1,w_2}$ $(w_1\in W^{(j)},\ w_2\in W^{(j-1)})$, $j\in\mathbb{N}$. If $\alpha_{j|_{M(w\cdot\mu)}} \neq 0$ for all $w\in W^{(j)}$ and all $j\in\mathbb{N}$, then $a^j_{w_1,w_2} \neq 0$ for all $w_1\in W^{(j)}$, $w_2\in W^{(j-1)}$ and all $j\in\mathbb{N}$.*

Definition 5 [2]. A *square* is a quadrupole (w_1, w_2, w_3, w_4) of elements of W such that $w_1 \leftarrow w_2 \leftarrow w_4$ and $w_1 \leftarrow w_3 \leftarrow w_4$, $w_2 \neq w_3$.

The lemma that follows is a generalization of [17, Lemma 10.4].

Lemma 6. *To every pair $(w_1, w_2)\in W\times W$ satisfying $w_1 \leftarrow w_2$ we can associate a number $c(w_1, w_2)\in\{+1, -1\}$, such that the product of all numbers associated to the four arrows of any square (w_1, w_2, w_3, w_4) is equal to -1.*

Proof. We note that all the statements made in this article hold if we replace $\mathbb{C}$ by any field of characteristic zero. Replacing $\mathbb{C}$ by $\mathbb{R}$ in Theorem 2 we obtain for every pair $w, w'\in W^{(j)}\times W^{(j-1)}$ a real number $s(w, w')$ such that $\partial_j = \oplus s(w, w') i_{w,w'}$ $(w\in W^{(j)},\ w'\in W^{(j-1)})$ for all $j\in\mathbb{N}$. The argument in the proof of [17, Lemma 10.2] in the real case implies that if (w_1, w_2, w_3, w_4) is a square then

$$s(w_1, w_2)\, s(w_2, w_4) = -s(w_1, w_3)\, s(w_3, w_4).$$

The exactness of (1) and Lemmas 3 and 4 in the real case imply that $s(w_1, w_2) \neq 0$ for all $w_1 \in W^{(j)}$, $w_2 \in W^{(j-1)}$, $j \in \mathbb{N}$. Now set $c(w_1, w_2) = \dfrac{s(w_1, w_2)}{|s(w_1, w_2)|}$. Q.E.D.

Let X, C_j, ∂_j and ε be as before. Let $d_j: C_j \to C_{j-1}$ be defined by $d_j = \oplus b^j_{w_1, w_2} i_{w_1, w_2} (w_1 \in W^{(j)}, w_2 \in W^{(j-1)})$, where $b^j_{w_1, w_2} = c(w_1, w_2)$ if $w_1 \leftarrow w_2$ and $b^j_{w_1, w_2} = 0$ otherwise, for all $j \in \mathbb{N}$. Let $\eta: C_0 \to X$ be any surjective $\mathfrak{g}$-homomorphism.

Using the argument in the proof of [17, Lemma 10.2] as in the proof of [17, Lemma 10.5] we obtain a strong BGG resolution for Kac-Moody Lie algebras with symmetrizable generalized Cartan matrix.

Theorem 7. *There exist* $\mathfrak{g}$*-isomorphisms* $\mu_j: C_j \to C_j$ *such that all the diagrams*

$$\begin{array}{ccccccccccc}
\ldots & C_j & \xrightarrow{d_j} & C_{j-1} & \to \ldots \to & C_1 & \xrightarrow{d_1} & C_0 & \xrightarrow{\eta} & X & \to 0 \\
 & \uparrow{\scriptstyle \mu_j} & & \uparrow{\scriptstyle \mu_{j-1}} & & \uparrow{\scriptstyle \mu_1} & & \uparrow{\scriptstyle \mu_0} & & \| & \\
\ldots & C_j & \xrightarrow{\partial_j} & C_{j-1} & \to \ldots \to & C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\varepsilon} & X & \to 0 \qquad (j \in \mathbb{N})
\end{array}$$

commute. In particular, the sequence

$$\ldots C_j \xrightarrow{d_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{d_1} C_0 \xrightarrow{\eta} X \to 0 \qquad (j \in \mathbb{N})$$

is exact.

Proof. The first part of the statement is now clear. For the second part, use the exactness of (1). Q.E.D.

Proceeding as in the proof of [17, Corollary 7] we obtain

Corollary 8. *Any complex*

$$\ldots C_j \xrightarrow{\alpha_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{\alpha_1} C_0 \xrightarrow{\alpha_0} X \to 0 \qquad (j \in \mathbb{N})$$

such that $\alpha_{j|_{M(w \cdot \mu)}} \neq 0$ *for all* $w \in W^{(j)}$, $j \in \mathbb{N}$, *is exact and equivalent to* (1).

We return to the general case where S is a fixed but arbitrary subset of $\{1, \ldots, n\}$ as in the beginning of this section. The next proposition is obtained exactly as in [12] using semi-simplicity with respect to $\mathfrak{m}_S$.

Proposition 9. *Let* $\mu \in \mathfrak{h}^*$, $v \in P_S^+$ *and* $f: M(\mu) \to M(v)$ *a non-zero map. Denote by* $\pi_v: M(v) \to M(R_S(v))$ *the natural projection. If* $\pi_v \circ f: M(\mu) \to M(R_S(v))$ *is nonzero, then* $\mu \in P_S^+$ *and* $(\pi_v \circ f)(K_\mu) = 0$, *where* $K_\mu = \operatorname{Ker} \pi_\mu$.

Remark. Let $\mu \in P_S^+$, $v \in P_S^+$ and f as above. If $(\pi_v \circ f)(K_\mu) = 0$ then f induces

$$\tilde{f}: M(R_S(\mu)) \to M(R_S(v)) \quad \text{defined by } \tilde{f}(v + K_\mu) = f(v) + K_v.$$

Definition 10 [12]. Let $\mu, \nu \in P_S^+$, $f: M(\mu) \to M(\nu)$ a nonzero map. The map $\tilde{f}$: $M(R_S(\mu)) \to M(R_S(\nu))$ defined above is called a *standard map* associated with f.

The following proposition is also obtained exactly as in [12] using Theorem 8.15.

Proposition 11. *Let $\mu \in P^+$ and $w, w' \in W^{(S)}$ such that $l(w) = l(w') + 1$. Then there exists a nonzero* g*-homomorphism $M(R_S(w \cdot \mu)) \to M(R_S(w' \cdot \mu))$ if and only if $w \leftarrow w'$. If $w \leftarrow w'$ then the standard map from $M(R_S(w \cdot \mu))$ to $M(R_S(w' \cdot \mu))$ (which is unique up to a scalar multiple) is nonzero.*

Using standard maps, (1) and Theorem 7 as in the proof of [17, Theorem 11.4] we obtain the generalized strong BGG resolution for Kac-Moody Lie algebras as a resolution equivalent to the weak form obtained in [8]. Here $\tilde{d}_j = \oplus b^j_{w_1, w_2} \tilde{i}_{w_1, w_2} (w_1 \in W^S)^{(j)}$, $w_2 \in (W^S)^{(j-1)}$), $j \in \mathbb{N}$.

Theorem 12. *There exist* g*-isomorphisms $\tilde{\mu}_j$ such that all the diagrams:*

$$\begin{array}{ccccccccccc}
\ldots \to C_j & \xrightarrow{\tilde{d}_j} & C_{j-1} \to \ldots \to C_1 & \xrightarrow{\tilde{d}_1} & C_0 & \xrightarrow{\tilde{\eta}} & X \to 0 \\
& \uparrow{\scriptstyle \tilde{\mu}_{j-1}} & & \uparrow{\scriptstyle \tilde{\mu}_1} & & \uparrow{\scriptstyle \tilde{\mu}_0} & \\
\ldots \to C_j & \xrightarrow{\partial_j} & C_{j-1} \to \ldots \to C_1 & \xrightarrow{\partial_1} & C_0 & \xrightarrow{\varepsilon} & X \to 0 \qquad (j \in \mathbb{N})
\end{array}$$

commute. In particular, the sequence

$$\ldots \to C_j \xrightarrow{\tilde{d}_j} C_{j-1} \to \ldots \to C_1 \xrightarrow{\tilde{d}_1} C_0 \xrightarrow{\tilde{\eta}} X \to 0 \qquad (j \in \mathbb{N})$$

is exact.

References

1. Bernstein, I.N., Gelfand, I.M., Gelfand, S.I.: Category of g-modules. Funkcional. Anal. i. Prilożen. **10**, Nr. 2, 1-8 (1976) [Russian]. Engl. Transl.: Functional. Anal. Appl. **10**, 87-92 (1976)
2. Bernstein, I.N., Gelfand, I.M., Gelfand, S.I.: Differential operators on the base affine space and a study of g-modules, Lie groups and their representations. Proceedings of the Summer School on Group Representations (I.M. Gelfand, ed.), Bolyai János Mathematical Society (Budapest 1971), pp. 39-69. London: Hilger 1975
3. Bernstein, I.N., Gelfand, I.M., Gelfand, S.I.: Structure of representations generated by vectors of highest weight. Funkcional. Anal. i Prilożen. **5**, Nr. 1, 1-9 (1970) [Russian]. Engl. Transl.: Functional. Anal. Appl. **5**, 1-8 (1971)
4. Curtis, C.W., Reiner, I.: Representation theory of finite groups and associative algebras. New York-London-Sydney: Wiley-Interscience 1966
5. Dixmier, J.: Enveloping Algebras. North-Holland Mathematical Library **14**. Amsterdam-New York-Oxford: North-Holland 1977
6. Enright, T.: On the fundamental series of a real semisimple Lie algebra: their irreducibility, resolutions and multiplicity formulae. Ann. of Math. (2) **110**, 1-82 (1979)
7. Enright, T., Wallach, N.R.: The fundamental series of representations of a real semisimple Lie algebra. Acta Math. **140**, 1-32 (1978)

8. Garland, H., Lepowsky, J.: Lie algebra homology and the Macdonald-Kac formulas. Invent. Math. **34**, 37–76 (1976)
9. Kac, V.G.: Infinite-dimensional Lie algebras and Dedekind's η-function. Funkcional. Anal. i Prilozen. **8**, Nr. **1**, 77–78 (1974) [Russian]. Engl. Transl.: Functional. Anal. Appl. **8**, 68–70 (1974)
10. Kac, V.G.: Simple irreducible graded Lie algebras of finite growth. Izo. Akad. Nauk SSSR Ser. Mat. **32**, 1323–1367 (1968) [Russian]. Engl. Transl.: Math. USSR-Izv. **2**, 1271–1311 (1968)
11. Kac, V.G., Kazhdan, D.A.: Structure of representations with highest weight of infinite dimensional Lie algebras. Advances in Math. **34**, 97–108 (1979)
12. Lepowsky, J.: A generalization of the Bernstein-Gelfand-Gelfand resolution J. Algebra **49**, 496–511 (1977)
13. Lepowsky, J.: Lectures on Kac-Moody Lie algebras Mimeographed Notes. Paris: Université de Paris VI 1978
14. Mitchell, B.: Theory of Categories. Pure and Applied Mathematics XVII. New York-San Francisco-London: Academic Press 1965
15. Moody, R.V.: Macdonald identities and Euclidean Lie algebras Proc. Amer. Math. Soc. **48**, 43–52 (1975)
16. Rocha-Caridi, A.: Splitting criteria for $\mathfrak{g}$-modules induced from a parabolic and the Bernstein-Gelfand-Gelfand resolution of a finite dimensional irreducible $\mathfrak{g}$-module. Ph.D. Thesis. New Brunswick: Rutgers University 1978
17. Rocha-Caridi, A.: Splitting Criteria for $\mathfrak{g}$-modules induced from a parabolic and the Bernstein-Gelfand-Gelfand resolution of a finite dimensional irreducible $\mathfrak{g}$-module. Trans. Amer. Math. Soc. **262**, 335–366 (1980)
18. Wallach, N.R.: On the Enright-Varadarajan modules, a construction of the discrete series. Ann. Sci. École Norm. Sup (4) **9**, 81–102 (1976)
19. Wallach, N.R.: Unpublished manuscript notes on the Borel-Weil theorem.

Received August 28, 1981

Added in Proof

Subsequent to our original submission of this paper, the authors received the manuscript "Structure of some categories of representations of infinite dimensional Lie algebras," by V. Deodhar, O. Gabber and V.G. Kac, which contains an alternate proof of Theorem 7.5 for $p=1$, $S=\varnothing$.

Math. Z. 180, 179–192 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Estimates for Evolutionary Surfaces of Prescribed Mean Curvature

Klaus Ecker

Institut für Angewandte Mathematik, Universität Heidelberg, Im Neuenheimer Feld 294, D-6900 Heidelberg 1, Federal Republic of Germany

0. Introduction

We consider solutions of the *parabolic* equation

$$\begin{aligned} \dot{u}+Au+H(x,u)&=0 \quad \text{in } \Omega\times(0,T)\\ u(0)&=u_0 \quad \text{in } \Omega \end{aligned} \tag{0.1}$$

where

$$Au=-D_i(a^i(Du)),\qquad a^i(p)=p^i\cdot(1+|p|^2)^{-1/2}$$

is the minimal surface operator.

Lichnewsky and Temam [6, 7] have first studied generalized solutions of the evolutionary minimal surface equation.

Gerhardt [5] proved under the condition

$$\frac{\partial H}{\partial\tau}\geqq 0,\qquad H=H(x,\tau) \tag{0.2}$$

estimates for $|u|$, $|\dot{u}|$ and an interior gradient bound. The estimate

$$|u(x_0)|\leqq \sup_{B_d(x_0)}|u_0|+d+T\left(\frac{n}{d}+\sup_{B_d(x_0)}|H(\cdot,0)|\right) \tag{0.3}$$

where $x_0\in\Omega$ and $d=\operatorname{dist}(x_0,\partial\Omega)$, is an analogue to the estimate for capillary surfaces given by Concus and Finn [2]. It is an interesting fact that solutions of (0.1) behave in a very similar manner as solutions of capillary type equations, since the term $\dot{u}$ takes the role of the condition

$$\frac{\partial H}{\partial\tau}\geqq\kappa>0. \tag{0.4}$$

Thus in the sequel it will be sufficient to assume the weaker condition

$$\frac{\partial H}{\partial\tau}\geqq-\bar{\kappa},\qquad \bar{\kappa}\geqq 0. \tag{0.5}$$

0025-5874/82/0180/0179/$02.80

In the stationary case we cannot expect to prove estimates if we drop the assumption concerning the monotonicity of $H(x,\cdot)$.

In Sect. 2 of this paper we show that the a-priori estimates for $|u|$ and $|\dot{u}|$ given in [5] remain valid under the condition (0.5).

In Sect. 3 we use a common approach to establish both a global Hölder continuity result for solutions of the Dirichlet problem corresponding to (0.1) and an interior gradient bound of the form

$$|Du(x_0,t)| \leqq c \cdot d^{-2}, \quad 0 \leqq t \leqq T$$

with

$$c = c(n, T, \bar{\kappa}, H_0 d, H_1 d^2, (1 + \sup_{B_d(x_0)} |Du_0|) \cdot d^2) \tag{0.6}$$

and

$$H_0 = \sup_{\Omega \times (0,T)} |H(\cdot, u(\cdot))|$$

$$H_1 = \sup_{\Omega \times (0,T)} \left(\frac{\partial H}{\partial x^k} \cdot D_k u \cdot (1 + |Du|^2)^{-1/2} \right)^+ .$$

A corresponding estimate for capillary surfaces is given in [3].

The author wishes to thank Professor Gerhardt for having acquainted him with this problem and for helpful advice.

1. Definitions and Some Preliminary Remarks

Ω is a bounded open set in $\mathbb{R}^n$, $n \geqq 2$, with Lipschitz-boundary $\partial\Omega$. In the case of the interior gradient bound and for the proof of the global Hölder continuity of solutions to the Dirichlet problem we do not have to make any assumptions concerning the boundary of Ω. The cylinder Q_T is defined by $Q_T = \Omega \times (0,T)$, $0 < T < \infty$.

Let $\mathscr{S} = \mathscr{S}(t) = \text{graph } u(t)$ over Ω, where $u(t) = u(\cdot, t)$. The components of the outward unit normal at a point $(x, u(x,t)) \in \mathscr{S}$ are then given by

$$\nu^i = -\frac{D_i v}{v}, \quad 1 \leqq i \leqq n$$

$$\nu^{n+1} = \frac{1}{v}, \quad \text{where } v = \sqrt{1 + |Du|^2}.$$

We define the tangential gradient on $\mathscr{S}$ by

$$\delta_i g = D_i g - \nu^i \sum_{j=1}^{n+1} \nu^j D_j g, \quad 1 \leqq i \leqq n+1, \ g \in C^1(\Omega \times \mathbb{R})$$

for which the relations

$$|\delta g| \leqq |Dg| \tag{1.1}$$

and

$$|Df| \leqq v |\delta f| \quad \text{for all } f \in C^1(\Omega) \tag{1.2}$$

are valid.

Applying the operator $v^{-1}D_k u D_k$ to both sides of (0.1) we get by noting (0.5) and the relation

$$D_i v = v^{-1} D_k u D_i D_k u$$

the crucial inequality

$$\dot{v} + v^{-1} a^{ij} D_i D_k u D_j D_k u - v^{-1} D_i (a^{ij} D_j v \cdot v) \leqq H_1 + \bar{\kappa} \frac{|Du|^2}{v} \tag{1.3}$$

where

$$a^{ij} := \frac{\partial a^i}{\partial p^j} = v^{-1}(\delta_{ij} - \nu_i \nu_j)$$

and

$$H_1 := \sup_{Q_T} \left(\frac{\partial H}{\partial x^k} \frac{D_k u}{v} \right)^+ . \tag{1.4}$$

Here and in the following we assume the convention that we sum over repeated indices from 1 to n. One can also derive the simpler inequality

$$\dot{v} - D_i (a^{ij} D_j v) \leqq H_1 + \bar{\kappa} \frac{|Du|^2}{v} \tag{1.5}$$

which shall be needed in the later sections.

Defining

$$\mathscr{H} = a^{ij} D_i D_j u \tag{1.6}$$

we derive as in [10]:

$$\begin{aligned} |\mathscr{H}| &= |a^{ij} D_i D_j u| = |a^{ij} \delta_{ki} D_k D_j u| \\ &\leqq (a^{kl} \delta_{ki} \delta_{lj})^{1/2} (a^{ij} D_k D_j u D_k D_i u)^{1/2} \\ &\leqq \left(\sum_{i=1}^{n} a^{ii} \right)^{1/2} (a^{ij} D_i D_k u D_j D_k u)^{1/2} \\ &\leqq n v^{-1/2} (a^{ij} D_i D_k u D_j D_k u)^{1/2}. \end{aligned} \tag{1.7}$$

Hence we get from (1.3)

$$\dot{v} + \frac{\mathscr{H}^2}{n^2} - v^{-1} D_i (a^{ij} D_j v v) \leqq H_1 + \bar{\kappa} \frac{|Du|^2}{v} \tag{1.8}$$

If $H(x, \cdot)$ satisfies (0.2), the inequalities (1.5) and (1.8) can be simplified to

$$\dot{v} - D_i (a^{ij} D_j v) \leqq H_1 \tag{1.9}$$

and

$$\dot{v} + \frac{\mathscr{H}^2}{n^2} - v^{-1} D_i (a^{ij} D_j v v) \leqq H_1 \tag{1.10}$$

respectively.

Finally we want to note the relations

$$a^{ij} D_i f D_j g \leqq v^{-1} |\delta f| |\delta g| \quad \text{for all } f, g \in C^1(\Omega) \tag{1.11}$$

$$a^{ij}D_i f D_j f = v^{-1}|\delta f|^2 \tag{1.12}$$

$$a^{ij}\xi_i\eta_j \leqq (a^{ij}\xi_i\xi_j)^{1/2}(a^{ij}\eta_i\eta_j)^{1/2} \tag{1.13}$$

$$a\cdot b \leqq \varepsilon a^p + \varepsilon^{-p/q} b^q \quad \text{for } a,b,\varepsilon>0, \frac{1}{p}+\frac{1}{q}=1 \tag{1.14}$$

which we frequently use in the sequel.

2. A-priori Estimates for $|u|$ and $|\dot{u}|$

2.1 Lemma. *Let u be a smooth solution of problem* (0.1) *satisfying the boundary condition*

$$u=\varphi \quad \text{on } \partial\Omega\times(0,T). \tag{2.1}$$

Then

$$\sup_{Q_T}|u| \leqq (1+e^{2\bar{\kappa}T})\max(\sup_{\Omega}|\dot{u}(0)|, \sup_{\partial\Omega\times(0,T)}|\dot{\varphi}|) \tag{2.2}$$

Proof. As in [5] we differentiate (0.1) and (2.1) with respect to t to get

$$\ddot{u} - D_i(a^{ij}D_j\dot{u}) + \frac{\partial H}{\partial \tau}\cdot\dot{u}=0 \tag{2.3}$$

and

$$\dot{u}=\dot{\varphi}. \tag{2.4}$$

Let $k=\max(\sup_{\Omega}|\dot{u}(0)|, \sup_{\partial\Omega\times(0,T)}|\dot{\varphi}|)$ and let $\eta=\operatorname{sgn}\dot{u}\cdot\max(|\dot{u}|-k,0)$. Multiplying (2.2) by η^{2p-1}, $p>1$, and integrating over Q_t, $0<t\leqq T$, we obtain

$$\begin{aligned}&\int_\Omega \eta^{2p}(t)dx + \int_0^t\int_\Omega a^{ij}D_i\eta D_j\eta\eta^{2p-2}dx\,ds\\ &\quad\leqq 2p\bar{\kappa}\int_0^t\int_\Omega |\dot{u}|\cdot\max(|\dot{u}|-k,0)^{2p-1}dx\,ds\end{aligned} \tag{2.5}$$

in view of (2.4) and the choice of η.

Since $|\dot{u}|\leqq\max(|\dot{u}|-k,0)+k$ we get after applying Young's inequality (1.14)

$$\int_\Omega \eta^{2p}(t)dx \leqq 2p\bar{\kappa}\int_0^t\int_\Omega k^{2p}dx\,ds + 4p\bar{\kappa}\int_0^t\int_\Omega \eta^{2p}dx\,ds. \tag{2.6}$$

By means of Gronwall's lemma it follows

$$\sup_{[0,T]}\int_\Omega \eta^{2p}(t)dx \leqq e^{4p\bar{\kappa}T}2p\bar{\kappa}\int_0^T\int_\Omega k^{2p}dx\,dt. \tag{2.7}$$

Exponating by $1/2p$ and letting p tend towards infinity yields

$$\sup_{Q_T}|\dot{u}| \leqq \sup_{Q_T}|\eta| + k \leqq k(1+e^{2\bar{\kappa}T}). \tag{2.8}$$

The condition $\sup_\Omega|\dot{u}(0)|<\infty$ is satisfied provided $u_0\in C^2(\bar{\Omega})$.

2.2 Remark. The boundedness of $|\dot{u}|$ immediately implies an a-priori bound for $|u|$. Nevertheless a more natural bound can be obtained by multiplying equation (0.1) by η^{2p-1}, where $\eta = \operatorname{sgn} u \cdot \max(|u|-k, 0)$, k suffiently large, and proceeding in the same way as in 2.1.

We then get

2.3 Lemma. *Under the same assumptions as in* 2.1 *we have*

$$\sup_{Q_T} |u| \leqq e^{(1+\bar{\kappa})T} \tilde{H}_0 + \max\Big(\sup_{\partial\Omega \times (0,T)} |\varphi|, \sup_{\Omega} |u_0|\Big) \tag{2.9}$$

where $\tilde{H}_0 := \sup_{\Omega} |H(\cdot, 0)|$.

2.4 Lemma. *Let B be the ball $B = B_R(x_0)$ in $\mathbb{R}^n$ and let u be a suffiently smooth solution of*

$$\begin{aligned} \dot{u} + Au + H(x,u) &= 0 \quad \text{in } B \\ u(0) &= u_0. \end{aligned} \tag{2.10}$$

Then the estimate

$$|u| \leqq (1+e^{2\bar{\kappa}T}) \left[R + \sup_B |u_0| + T\left(\frac{n}{R} + \sup_B |H(\cdot,0)|\right)\right] \tag{2.11}$$

is valid in $B \times (0, T)$.

Proof. As in [5] we shall prove the estimate only for u, the estimate for $-u$ can be proved similarly. Choose $0 < R_0 < R_1 < R$. We are going to show the validity of (2.11) in $B_{R_0} = B_{R_0}(x_0)$, where $u(t)$ is uniformly Lipschitz continuous. Since the estimate will hold for all $R_0 < R$, it will also be valid for R.

We assume $x_0 = 0$ and define in B_{R_0}

$$\delta(x,t) = -(R_1^2 - |x|^2)^{1/2} + M_0 + t \cdot M \tag{2.12}$$

where $M_0 = \sup_B |u_0| + R_1$, and

$$M = \frac{n}{R_1} + \sup_B |H(\cdot, 0)|.$$

Then we have

$$A\delta = -\frac{n}{R_1} \tag{2.13}$$

$$\dot{\delta} = M \tag{2.14}$$

and

$$a^i(D\delta)\, \nu_i = \frac{R_0}{R_1} \quad \text{on } \partial B_{R_0} \tag{2.15}$$

where ν is the outward unit normal.

We also know

$$\sup_{B_{R_0} \times (0,T)} |Du| \leqq L < \infty; \tag{2.16}$$

hence

$$|a^i(Du) \cdot \nu_i| \leqq \frac{L}{\sqrt{1+L^2}} < 1 \quad \text{on } \partial B_{R_0}. \tag{2.17}$$

Now we choose R_1 close enough to R_0 such that

$$a^i(D\delta)\cdot\nu_i \geqq \frac{L}{\sqrt{1+L^2}} \quad \text{on } \partial B_{R_0}. \tag{2.18}$$

For δ the inequality

$$\dot{\delta} + A\delta + H(x,\delta) + \bar{\kappa}\delta \geqq 0 \quad \text{on } B_{R_0}\times(0,T) \tag{2.19}$$

is valid.

Subtracting (2.19) from (2.10), multiplying the difference by η^{p-1}, $p>1$, where $\eta=\max(u-\delta,0)$ and integrating over $B_{R_0}\times(0,t)$ we obtain in view of (2.18), (0.5) and the monotonicity of A

$$\int_{B_{R_0}} \eta^p(t)\,dx \leqq \bar{\kappa}\int_0^t\int_{B_{R_0}} \delta\cdot\eta^{p-1}\,dx\,ds + \bar{\kappa}\int_0^t\int_{B_{R_0}} |u-\delta|\,\eta^{p-1}\,dx\,ds. \tag{2.20}$$

Since $|u-\delta|\,\eta^{p-1}\leqq\eta^p$ we get after applying Young's inequality to the first term of the right hand side of (2.20)

$$\int_{B_{R_0}} \eta^p(t)\,dx \leqq \bar{\kappa}p\int_0^t\int_{B_{R_0}} \delta^p\,dx\,ds + 2\bar{\kappa}p\int_0^t\int_{B_{R_0}} \eta^p\,dx\,ds. \tag{2.21}$$

Using now again Gronwall's lemma and proceeding as in Lemma 2.1 we obtain

$$u \leqq (1+e^{2\bar{\kappa}T}) \sup_{B_{R_0}\times(0,T)} \delta \tag{2.22}$$

q.e.d.

2.5 Remark. Under the conditions of 2.1 we have

$$\sup_{Q_T} |Du| \leqq c(\bar{\kappa},H_1,T)\left(\max\left(\sup_{\partial\Omega\times(0,T)} |Du|, \sup_{\Omega} |Du_0|\right)+1\right) \tag{2.23}$$

Proof. Let $k=\max\left(\sup_{\partial\Omega\times(0,T)} \sqrt{1+|Du|^2}, \sup_{\Omega}\sqrt{1+|Du_0|^2}\right)$ and $v_k=\max(v-k,0)$. Multiplying (1.5) by v_k^{p-1}, $p>1$, integrating over Q_t and taking the relations (1.11)–(1.14) into account we obtain after some calculation

$$\begin{aligned}\int_\Omega v_k^p(t)\,dx &\leqq \int_0^t\int_\Omega (H_1+\bar{\kappa}v)\,v_k^{p-1}\,dx\,ds \\ &\leqq \int_0^t\int_\Omega (H_1^p+\bar{\kappa}k^p)\,dx\,ds + 2p\bar{\kappa}\int_0^t\int_\Omega v_k^p\,dx\,ds\end{aligned} \tag{2.24}$$

where we used the relation $v\leqq v_k+k$.

As in 2.1, it follows after applying Gronwall's lemma

$$\sup_{Q_T} v_k \leqq e^{2\bar{\kappa}T}(H_1+\bar{\kappa}k) \tag{2.25}$$

which immediately yields the result.

3. The Main Results

3.1 Theorem. *Let u be a smooth solution of* (0.1). *Then for $x_0 \in \Omega$ and for $t \in [0, T]$ we have the estimate*

$$|Du(x_0, t)| \leqq c \cdot d^{-2} \tag{3.1}$$

where $d = \operatorname{dist}(x_0, \partial\Omega)$ *and* c *depends on* n, T, $\bar{\kappa}$, $H_0 d$, $H_1 d^2$ *and* $(1) + \sup_{B_d(x_0)} |Du_0|) d^2$.

The form of the estimate (3.1) one could expect looking at the estimate (0.3) for $|u|$.

Beside the interior gradient bound we prove an estimate of global Hölder continuity for solutions of the Dirichlet problem corresponding to (0.1) similar to the estimate for a large class of elliptic quasilinear equations given by Simon [10]. The Hölder exponent will be independent of all data, as it is valid for capillary surfaces [3].

Let us denote for $\rho > 0$

$$\Omega_{x_0, \rho} := \Omega \cap B_\rho(x_0), \; Q^T_{x_0, \rho} := \Omega_{x_0, \rho} \times (0, T)$$

where $\Omega_{x_0, \rho/2} \neq \varnothing$ is always assumed.

We consider smooth solutions of (0.1) with the Dirichlet boundary condition

$$u = \varphi \quad \text{on } \partial\Omega \times (0, T) \tag{3.2}$$

where the boundary values φ are assumed to be Lipschitz continuous in $\bar{Q}_T$.

3.2 Theorem. *Let u be continuous on $\bar{Q}^T_{x_0, \rho}$, φ Lipschitz on $\bar{Q}^T_{x_0, \rho}$ and $u = \varphi$ on $(\partial\Omega \cap B_\rho(x_0)) \times (0, T)$, then we have the estimate*

$$\sup_{Q^T_{x_0, \rho/2}} |u - \varphi|^3 \, |Du| \leqq c \tag{3.3}$$

where c depends on n, $\bar{\kappa}$, T, $\sup_{Q^T_{x_0, \rho}} |\varphi|$, $\sup_{Q^T_{x_0, \rho}} |\dot{\varphi}|$, $\sup_{Q^T_{x_0, \rho}} |D\varphi|$, ρ, $H_0\rho$, $H_1\rho^2$, $\sup_{\Omega_{x_0, \rho}} |u_0|$ and $\sup_{\Omega_{x_0, \rho}} |Du_0|$.

By the same arguments as in [10] we get the following

3.3 Theorem. *Under the conditions of Theorem* 3.2 *the inequality*

$$|u(x, t) - u(y, t)| \leqq (c^{3/4} + L|x - y|^{3/4}) \, |x - y|^{1/4} \tag{3.4}$$

for all $x, y \in \bar{\Omega}_{x_0, \rho/2}$, for all $t \in [0, T]$

is valid, where L is the Lipschitz constant for φ and c depends on the same quantities as in Theorem 3.2.

3.4 Remark. The global Hölder continuity under the corresponding assumptions follows immediately.

Consider now smooth solutions of the general boundary value problem

$$\begin{aligned} \dot{u}+Au+H(x,u)&=0\\ u(0)&=u_0\\ -a^i\cdot\nu_i&=\beta(x,u-\varphi) \quad \text{on } \partial\Omega\times(0,T) \end{aligned} \tag{3.5}$$

for C^3-boundary $\partial\Omega$ with normal $\nu=(\nu_1,\dots,\nu_n)$, where $\beta=\beta(x,\tau)$, $H=H(x,\tau)$ and φ are given Lipschitz functions satisfying

$$\frac{\partial\beta}{\partial\tau}\geqq 0, \quad \frac{\partial H}{\partial\tau}\geqq 0$$

and

$$|\beta|\leqq 1-a, \quad a>0. \tag{3.6}$$

In [5] a bound for the tangential derivative of u is given under the above conditions. As I heard from Professor Gerhardt his result remains valid even under the most general assumption

$$|\beta|\leqq 1 \tag{3.7}$$

if either $H=H(\tau)$ or $\frac{\partial H}{\partial\tau}\geqq\kappa>0$ in the case $H=H(x,\tau)$ and if $\beta=\beta(\tau)$. In view of Theorem 3.3 we can now state

3.5 Corollary. *Under the assumptions mentioned above every solution of* (3.5) *is global Hölder continuous.*

If $H(x,\cdot)$ is monotone one can use the Moser iteration technique [9] for the proof of Theorem 3.1 as was done for capillary type equations in [3].

Here we choose a slightly different approach yielding also the Hölder continuity result. First we derive two crucial lemmata.

3.6 Lemma. *Let* $k_0\geqq\sup_\Omega\sqrt{1+|Du_0|^2}$ *and* $v_{k_0}=\max(v-k_0,0)$.

Then for $0\leqq\eta\in H_0^{1,\infty}(\Omega)$, $p\geqq 1$, $q\geqq 1$ *and* $c(\eta):=\sup_\Omega(H_1\eta^2+|\delta\eta|^2)$ *the inequality*

$$\begin{aligned} &\int_0^T\int_\Omega v_{k_0}^p\eta^{2q}v\,dx\,dt\\ &\quad\leqq c(p,q,\bar\kappa,T)\left(c(\eta)\cdot\int_0^T\int_\Omega v_{k_0}^{p-1}\eta^{2q-2}v\,dx\,dt+k_0^p\int_0^T\int_\Omega v\eta^{2q}\,dx\,dt\right) \end{aligned} \tag{3.8}$$

is valid.

Proof. Multiplying (1.5) by $pv_{k_0}^{p-1}v\eta^{2q}$ and $v_{k_0}^p\eta^{2q}$, $p\geqq 1$, $q\geqq 1$ adding up and integrating over Q_t we obtain in view of (1.11)–(1.14) and of the relation $v_{k_0}^{p-1}v\leqq c(v_{k_0}^p+k_0^p)$

$$\begin{aligned} \int_\Omega(v_{k_0}^pv)(t)\eta^{2q}\,dx&\leqq c(p,q)\int_0^t\int_\Omega v_{k_0}^{p-1}v\eta^{2q-2}(H_1\eta^2+|\delta\eta|^2)\,dx\,ds\\ &\quad+c(\bar\kappa,p,q)\left(\int_0^t\int_\Omega k_0^pv\eta^{2q}\,dx\,ds+\int_0^t\int_\Omega v_{k_0}^pv\eta^{2q}\,dx\,ds\right). \end{aligned} \tag{3.9}$$

The result now follows by Gronwall's lemma.

3.7 Remark. If $p\in\mathbb{N}$, $q=p$ and $k_0>c(\eta)$ we derive by iterating (3.8)

$$\int_0^T\int_\Omega v_{k_0}^p\eta^{2p}v\,dx\,dt\leqq c(p,\bar\kappa,T,\sup_\Omega\eta)\,k_0^p\cdot\int_0^T\int_{\operatorname{supp}\eta}v\,dx\,dt \tag{3.10}$$

3.8 Lemma. *Let $\eta\in H_0^{1,\infty}(\Omega)$ be nonnegative. Then we have for*

$$k_0=\sup_{\operatorname{supp}\eta}\sqrt{1+|Du_0|^2}\cdot\max(1,\sup_\Omega\eta^3)+c(\eta),$$

$$\sup_{Q_T}(v\eta^3)\leqq k_0+c\cdot k_0^{1+\frac{n}{4(n+1)}}\left(\int_0^T\int_{\operatorname{supp}\eta}v\,dx\,dt\right)^{\frac{1}{2(n+1)}} \tag{3.11}$$

where $c(\eta):=\sup_\Omega(H_1\eta^2+|\delta\eta|^2)$ *and c depends on $n,\bar\kappa,T$ and* $\sup_\Omega\eta$.

Proof. For $k\geqq k_0$ we define $\zeta_k:=\max(v\eta^3-k,0)$.

Let $A(k):=\{(x,t)\in Q_T:\zeta_k(x,t)>0\}$ and $|A(k)|:=\int_{A(k)}v\,dx\,dt$. Note that on $A(k)$ we have

$$D_i\zeta_k=D_i(v\eta^3)\qquad 1\leqq i\leqq n$$

and

$$\dot\zeta_k=\dot v\eta^3. \tag{3.12}$$

Multiplying inequality (1.8) by $\zeta_k v^2\eta^3$ and by $v\zeta_k^2$, adding up and integrating over Q_T we obtain after some calculations using (1.11)–(1.14) and neglecting some nonnegative terms on the left hand side

$$\sup_{[0,T]}\int_\Omega\zeta_k^2v^2\,dx+\int_0^T\int_\Omega\frac{\mathscr{H}^2}{n^2}\zeta_k^2v\,dx\,dt+\int_{A(k)}|\delta v|^2\eta^6v\,dx\,dt$$
$$\leqq c(\bar\kappa)\Big(\int_{A(k)}v^3\eta^4(H_1\eta^2+|\delta\eta|^2)\,dx\,dt+\int_{A(k)}v^4\eta^6\,dx\,dt\Big). \tag{3.13}$$

Since the relation

$$|\delta\zeta_k|^2v\leqq c\cdot(|\delta v|^2\eta^6v+\eta^4|\delta\eta|^2v^3) \tag{3.14}$$

is valid on $A(k)$ we get from (3.13) integrating over $\mathscr{S}$ instead of Ω and denoting by $d\mathscr{H}_n=v\,dx$ the n-dimensional Hausdorff measure for the surface $\mathscr{S}$

$$\sup_{[0,T]}\int_{\mathscr{S}}\zeta_k^2v\,d\mathscr{H}_n+\int_0^T\int_{\mathscr{S}}|\delta\zeta_k|^2\,d\mathscr{H}_n\,dt+\int_0^T\int_{\mathscr{S}}\frac{\mathscr{H}^2}{n^2}\zeta_k^2\,d\mathscr{H}_n\,dt$$
$$\leqq c(\bar\kappa)\Big(\int_{A(k)}v^3\eta^4(H_1\eta^2+|\delta\eta|^2)\,dx\,dt+\int_{A(k)}v^4\eta^6\,dx\,dt\Big). \tag{3.15}$$

We now want to apply Sobolev's inequality on hypersurfaces in $\mathbb{R}^{n+1}$ [8]

$$\Big(\int_{\mathscr{S}}\varphi^{n/n-1}\,d\mathscr{H}_n\Big)^{\frac{n-1}{n}}\leqq c(n)\int_{\mathscr{S}}\left(|\delta\varphi|+\frac{|\mathscr{H}|}{n}|\varphi|\right)d\mathscr{H}_n \tag{3.16}$$

which is valid for every $\varphi\in H_0^{1,1}(\Omega)$ and proceed in a very similar way as in [10], Lemma 1.

Inserting ζ_k^2 in (3.16) we obtain using the relation $ab \leqq \varepsilon a^2 + (4\varepsilon)^{-1} b^2$

$$\begin{aligned}
&\left(\int_{\mathscr{S}} \zeta_k^{2n/n-1}\, d\mathscr{H}_n\right)^{\frac{n-1}{n}} \leqq c(n) \int_{\mathscr{S}} \left(|\delta\zeta_k|\,\zeta_k + \zeta_k^2 \frac{|\mathscr{H}|}{n}\right) d\mathscr{H}_n \\
&\leqq c(n)\left(\varepsilon \cdot \int_{\mathscr{S}} \left(|\delta\zeta_k|^2 + \frac{\mathscr{H}^2}{n^2}\,\zeta_k^2\right) d\mathscr{H}_n + \varepsilon^{-1} \cdot \int_{\mathscr{S}} \zeta_k^2\, d\mathscr{H}_n\right). \qquad (3.17)
\end{aligned}$$

In view of (3.15) and since $v > k \geqq k_0$ on $A(k)$ it follows

$$\begin{aligned}
&\varepsilon \cdot k_0 \sup_{[0,T]} \int_{\mathscr{S}} \zeta_k^2\, d\mathscr{H}_n + \int_0^T \left(\int_{\mathscr{S}} \zeta_k^{2n/n-1}\, d\mathscr{H}_n\right)^{\frac{n-1}{n}} dt \\
&\leqq c(n,\bar{\kappa}) \cdot \left(\varepsilon \int_{A(k)} v^3 \eta^4 (H_1 \eta^2 + |\delta\eta|^2)\, dx\, dt\right. \\
&\left. + \varepsilon^{-1} \int_{A(k)} v^3 \eta^6\, dx\, dt + \varepsilon \int_{A(k)} v^4 \eta^6\, dx\, dt\right). \qquad (3.18)
\end{aligned}$$

We are going to use an interpolation argument similar to that given in [5]: For $1 < \alpha < \frac{n}{n-1}$ with $\frac{1}{\alpha} = a \cdot \frac{n-1}{n} + 1 - a$, $0 < a < 1$, we have the interpolation inequality

$$\left(\int_{\mathscr{S}} \zeta_k^{2\alpha}\, d\mathscr{H}_n\right)^{1/\alpha} \leqq \left(\int_{\mathscr{S}} \zeta_k^{2n/n-1}\, d\mathscr{H}_n\right)^{\frac{n-1}{n} \cdot a} \left(\int_{\mathscr{S}} \zeta_k^2\, d\mathscr{H}_n\right)^{1-a}. \qquad (3.19)$$

Choosing α such that $\alpha \cdot a = 1$ yields $\alpha = \frac{n+1}{n}$ and

$$\begin{aligned}
&(\varepsilon \cdot k_0)^{\alpha-1} \int_0^T \int_{\mathscr{S}} \zeta_k^{2\alpha}\, d\mathscr{H}_n\, dt \leqq (\varepsilon \cdot k_0)^{\alpha-1} \left(\sup_{[0,T]} \int_{\mathscr{S}} \zeta_k^2\, d\mathscr{H}_n\right)^{\alpha-1} \\
&\cdot \int_0^T \left(\int_{\mathscr{S}} \zeta_k^{2n/n-1}\, d\mathscr{H}_n\right)^{\frac{n-1}{n}} dt \\
&\leqq \left(\varepsilon \cdot k_0 \sup_{[0,T]} \int_{\mathscr{S}} \zeta_k^2\, d\mathscr{H}_n + \int_0^T \left(\int_{\mathscr{S}} \zeta_k^{2n/n-1}\, d\mathscr{H}_n\right)^{\frac{n-1}{n}} dt\right)^{\alpha}. \qquad (3.20)
\end{aligned}$$

Combined with (3.18) this gives for $\varepsilon = k_0^{-1/2}$

$$\begin{aligned}
&\left(\int_{A(k)} \zeta_k^{2\alpha} v\, dx\, dt\right)^{1/\alpha} \\
&\leqq c(n,\bar{\kappa})\, k_0^{-\frac{\alpha-1}{2\alpha}} \left(k_0^{1/2} \cdot \sup_{\Omega} (1+\eta^2) \int_{A(k)} v^3 \eta^4\, dx\, dt\right. \\
&\left. + k_0^{-1/2} \cdot \int_{A(k)} v^4 \eta^6\, dx\, dt\right). \qquad (3.21)
\end{aligned}$$

Since on $A(h)$ we have $\zeta_k > (h-k)$ for $h > k \geqq k_0$, we obtain by applying Hölder's inequality for arbitrary γ, $0 < \gamma < 1$

$$|h-k|^2\,|A(h)| \leqq c\cdot k_0^{-\frac{\alpha-1}{2\alpha}}\cdot |A(k)|^{\gamma+\frac{\alpha-1}{\alpha}}$$
$$\cdot\left(k_0^{1/2}\sup_{\Omega}(1+\eta^2)\left(\int_{A(k)} v^{\frac{2}{1-\gamma}+1}\eta^{\frac{4}{1-\gamma}}\,dx\,dt\right)^{1-\gamma}\right.$$
$$\left.+k_0^{-1/2}\left(\int_{A(k)} v^{\frac{3}{1-\gamma}+1}\eta^{\frac{6}{1-\gamma}}\,dx\,dt\right)^{1-\gamma}\right) \tag{3.22}$$

where c depends on the above quantities.

Choosing γ such that $\gamma+\frac{\alpha-1}{\alpha}>1$ and $\frac{1}{1-\gamma}\in\mathbb{N}$ (with this choice γ depends on n) we are now in the situation to apply a lemma due to Stampacchia [12], Lemma 4.1, to derive

$$\sup_{Q_T}(v\eta^3) \leqq k_0 + c\cdot k_0^{-\frac{\alpha-1}{4\alpha}+\frac{1}{4}}\cdot|A(k_0)|^{\frac{\gamma}{2}+\frac{\alpha-1}{2\alpha}-\frac{1}{2}}$$
$$\cdot\left(\sup_{\Omega}(1+\eta^2)^{1/2}\left(\int_{A(k_0)} v^{\frac{2}{1-\gamma}+1}\eta^{\frac{4}{1-\gamma}}\,dx\,dt\right)^{\frac{1-\gamma}{2}}\right.$$
$$\left.+k_0^{-1/2}\left(\int_{A(k_0)} v^{\frac{3}{1-\gamma}+1}\eta^{\frac{6}{1-\gamma}}\,dx\,dt\right)^{\frac{1-\gamma}{2}}\right) \tag{3.23}$$

where c depends on the same quantities as in (3.22).

Since $v\leqq v_{k_0}+k_0$ we have

$$\int_{A(k_0)} v^{\frac{3}{1-\gamma}+1}\eta^{\frac{6}{1-\gamma}}\,dx\,dt$$
$$\leqq c(\gamma)\left(\int_0^T\int_{\Omega} v^{\frac{3}{1-\gamma}}\eta^{\frac{6}{1-\gamma}}v\,dx\,dt+\sup_{\Omega}\eta^{\frac{6}{1-\gamma}}\cdot k^{\frac{3}{1-\gamma}}\int_0^T\int_{\operatorname{supp}\eta} v\,dx\,dt\right). \tag{3.24}$$

Applying now inequality (3.10) with $p=\frac{3}{1-\gamma}$, and $p=\frac{2}{1-\gamma}$ respectively yields

$$\left(\int_{A(k_0)} v^{\frac{3}{1-\gamma}+1}\eta^{\frac{6}{1-\gamma}}\,dx\,dt\right)^{\frac{1-\gamma}{2}} \leqq c(\gamma,\bar{\kappa},T,\sup_{\Omega}\eta)\cdot k_0^{3/2}\cdot\left(\int_0^T\int_{\operatorname{supp}\eta} v\,dx\,dt\right)^{\frac{1-\gamma}{2}} \tag{3.25}$$

and

$$\left(\int_{A(k_0)} v^{\frac{2}{1-\gamma}+1}\eta^{\frac{4}{1-\gamma}}\,dx\,dt\right)^{\frac{1-\gamma}{2}} \leqq c(\gamma,\bar{\kappa},T,\sup_{\Omega}\eta)\cdot k_0\left(\int_0^T\int_{\operatorname{supp}\eta} v\,dx\,dt\right)^{\frac{1-\gamma}{2}}. \tag{3.26}$$

We obtain the result combining (3.23), (3.25) and (3.26) and noting the relations $\frac{\alpha-1}{\alpha}=\frac{1}{n+1}$ and

$$|A(k_0)| \leqq \int_0^T\int_{\operatorname{supp}\eta} v\,dx\,dt. \tag{3.27}$$

The Theorems 3.1 and 3.2 will be proved by choosing appropriate test-functions η in Lemma 3.8.

Proof of Theorem 3.1. Let $R>0$ such that $B=B_R(x_0)\subset\subset\Omega$ and define

$$\eta(x)=\vartheta(x)\cdot\xi\left(\sup_{[0,T]}(u(x)-u(x_0))\right),\ \vartheta\in C_c^{0,1}(\Omega),\ \xi\in C_c^{0,1}(\mathbb{R})$$

such that $0\leqq\vartheta\leqq 1$, $0\leqq\xi\leqq 1$

$$\begin{aligned}&\vartheta\equiv 1 \text{ in } B_{\frac{R}{4}}, \qquad \vartheta\equiv 0 \text{ in } B_R\setminus B_{\frac{R}{2}}, \qquad |D\vartheta|\leqq\frac{c}{R}\\ &\xi\equiv 1 \text{ in } \left(-\frac{R}{4},\frac{R}{4}\right), \qquad \xi\equiv 0 \text{ in } \mathbb{R}\setminus\left(-\frac{R}{2},\frac{R}{2}\right), \qquad |\xi'|\leqq\frac{c}{R}.\end{aligned} \tag{3.28}$$

Noting that $\sup_{[0,T]} u\in H_0^{1,\infty}(\Omega)$ and

$$|D_i\sup_{[0,T]}u|\leqq\sup_{[0,T]}|D_iu| \qquad \text{a.e. in } \Omega, \qquad 1\leqq i\leqq n$$

we have furthermore

$$|\delta\eta|\leqq|\xi|\,|\delta\vartheta|+|\vartheta|\,|\xi'|\sup_{[0,T]}|\delta u|\leqq\frac{c}{R} \tag{3.29}$$

where we used the relation $|\delta u(t)|\leqq 1, t\in[0,T]$.

Hence the following is valid for $0\leqq\eta\leqq 1$

$$\begin{aligned}&\eta\equiv 1 \text{ in } C_{\frac{R}{4}}:=\left\{x\in B_{\frac{R}{4}}:\ \sup_{[0,T]}|u(x)-u(x_0)|<\frac{R}{4}\right\}\\ &\eta\equiv 0 \text{ in } \Omega\setminus C_{\frac{R}{2}}\end{aligned}$$

Inserting now η in (3.11) yields

$$\sup_{C_{\frac{R}{4}}\times[0,T]} v\leqq c\cdot R^{-2}\left(1+R^{-\frac{n}{2(n+1)}}\cdot\left(\int_0^T\int_{C_{\frac{R}{2}}} v\,dx\,dt\right)^{\frac{1}{2(n+1)}}\right) \tag{3.30}$$

The constant depends on $n,\bar{\kappa},T,H_1R^2$ and $(1+\sup_{B_R}|Du_0|)R^2$.

It remains to estimate the integral in (3.30) appropriately. This we accomplish using methods given in [14, 15].

We note that $C_R\subset C_R(t):=\{x\in B_R:\ |u(x,t)-u(x_0,t)|<R\}$ and define

$$u_R(t)=\begin{cases}\frac{R}{2};\ u(x,t)-u(x_0,t)\geqq\frac{R}{2}\\ u(x,t)-u(x_0,t),\ -\frac{R}{2}\leqq u(x,t)-u(x_0,t)\leqq\frac{R}{2}\\ -\frac{R}{2},\ u(x,t)-u(x_0,t)\leqq-\frac{R}{2}.\end{cases} \tag{3.31}$$

Let χ be a C^1-cut-off function with

$$\chi \equiv 1 \quad \text{on} \quad B_{\frac{R}{2}}, \quad \chi \equiv 0 \quad \text{on} \quad \Omega \setminus B_{\frac{3}{4}R}.$$

Multiplying (0.1) by $u_R \cdot \chi$ and integrating over Q_T we obtain

$$\int_0^T \int_{C_{\frac{R}{2}}(t)} \frac{|Du|^2}{v}\, dx\, dt \leqq \int_0^T \int_{B_{\frac{3}{4}R}} (|D_x| + H_0 + |\dot{u}|)\, u_R\, dx\, dt. \tag{3.32}$$

Since $u_R \leqq R$ it follows

$$\int_0^T \int_{C_{\frac{R}{2}}(t)} v\, dx\, dt \leqq c(n) \left(T \cdot R^n (1 + H_0 R + R^{\frac{n}{2}+1}) \left(\int_0^T \int_{B_{\frac{3}{4}R}} |\dot{u}|^2\, dx\, dt \right)^{1/2} \right). \tag{3.33}$$

Multiplying (0.1) by $\dot{u} \cdot \psi^2$ where ψ is the cut-off-function for $B_{\frac{3}{4}R}$ we derive in view of the relation $\dot{v} = v^{-1} D_i u\, D_i \dot{u}$ and of Young's inequality $ab \leqq \varepsilon a^2 + (4\varepsilon)^{-1} b^2$

$$\int_0^T \int_\Omega |\dot{u}|^2 \psi^2\, dx\, dt \leqq c \cdot \int_0^T \int_\Omega (H_0^2 \psi^2 + |D\psi|^2)\, dx\, dt + c \int_\Omega v(0) \psi^2\, dx \tag{3.34}$$

from which we immediately obtain

$$\int_0^T \int_{B_{\frac{3}{4}R}} |\dot{u}|^2\, dx\, dt \leqq c(n, T) R^{n-2} (1 + (H_0^2 + \sup_{B_R} v(0)) R^2) \tag{3.35}$$

and thus from (3.33)

$$\int_0^T \int_{C_{\frac{R}{2}}} v\, dx\, dt \leqq c(n, T, H_0 R, \sup_{B_R} v(0) R^2)\, R^n. \tag{3.36}$$

Inserting (3.36) in (3.30) yields the result.

Proof of Theorem 3.2. Choosing $\eta = \sup_{[0,T]} \max(u - \varphi, 0) \max(1 - \frac{4}{3} r/\rho, 0)$, where $r = |x - x_0|$ we have

$$\begin{aligned} \sup_\Omega \eta^2 &\leqq \sup_{Q_{x_0,\rho}^T} |u - \varphi|^2 \\ \sup_\Omega |\delta \eta|^2 &\leqq c \cdot (1 + \sup_{Q_{x_0,\rho}^T} |D\varphi|^2 + \sup_{Q_{x_0,\rho}^T} |u - \varphi|^2 \rho^{-2}). \end{aligned} \tag{3.37}$$

Inserting η in (3.11), noting that $1 - \frac{4}{3} r/\rho > \frac{1}{3}$ on $\Omega_{x_0, \rho/2}$, yields

$$\sup_{Q_{x_0,\rho/2}^T} (v \cdot \max(u - \varphi, 0)^3) \leqq c \cdot \left(\int_0^T \int_{\Omega_{x_0, \frac{3}{4}\rho}} v\, dx\, dt \right)^{\frac{1}{2(n+1)}} \tag{3.38}$$

where c depends on the quantities mentioned in Theorem 3.2. A bound for $|u|$ is given in Sect. 2.

To estimate the integral in (3.38) we multiply (0.1) by

$$(u-\varphi)\cdot\max(1-\tfrac{4}{3}r/\rho,0)$$

and integrate to obtain

$$\int_0^T\int_{\Omega_{x_0},\frac{3}{4}\rho} v\,dx\,dt\leqq c(n)\cdot T\cdot\rho^n+c\cdot\int_0^T\int_{\Omega_{x_0},\rho}|D\varphi|\,dx\,dt$$
$$+c\cdot\rho^{-1}\int_0^T\int_{\Omega_{x_0},\rho}|u-\varphi|\,(1+(H_0+|\dot\varphi|)\,\rho)\,dx\,dt+c\int_{\Omega_{x_0},\rho}|u_0-\varphi(0)|^2\,dx \tag{3.39}$$

which completes the proof.

3.9 Remark. If Ω is a bounded set one can choose $\rho=\operatorname{diam}\Omega$ to obtain

$$\eta=\sup_{[0,T]}\max(u-\varphi,0).$$

References

1. Brakke, K.A.: The motion of surface by its mean curvature. Mathematical Notes 20. Princeton: Princeton University Press 1978
2. Concus, P., Finn, R.: On capillary surfaces in a gravitational field. Acta Math. **132**, 207–223 (1974)
3. Ecker, K.: An interior gradient bound for solutions of equations of capillary type. Preprint
4. Gerhardt, C.: Boundary value problems for surfaces of prescribed mean curvature. J. Math. Pures Appl. **58**, 75–109 (1979)
5. Gerhardt, C.: Evolutionary surfaces of prescribed mean curvature. J. Differential Equations **36**, 139–172 (1980)
6. Lichnewski, A., Temam, R.: Surfaces minimales d'évolution: le concept de pseudo-solution. C.R. Acad. Sci. Paris Sér. A **284**, 853–856 (1977)
7. Lichnewski, A., Temam, R.: Pseudo-solutions of the time dependent minimal surface problem. J. Differential Equations **30**, 340–364 (1978)
8. Michael, J.H., Simon, L.M.: Sobolev and mean value inequalities on generalized submanifolds of $\mathbb{R}^n$. Comm. Pure Appl. Math. **26**, 361–379 (1973)
9. Moser, J.: A new proof of De Giorgi's theorem concerning the regularity problem for elliptic differential equations. Comm. Pure Appl. Math. **13**, 457–468 (1960)
10. Simon, L.: Global estimates of Hölder continuity for a class of divergence-form elliptic equations. Arch. Rational Mech. Anal. **56**, 253–272 (1974)
11. Simon, L., Spruck, J.: Existence and regularity of a capillary surface with prescribed contact angle. Arch. Rational Mech. Anal. **61**, 19–34 (1976)
12. Stampacchia, G.: Equations elliptiques au second ordre à coéfficients discontinues. Séminaire de Mathématiques supérieures 16. Montreal: Les Presses de l'Université de Montreal 1966
13. Temam, R.: Applications de l'analyse convexe au calcul des variations. In: Nonlinear Operators and the Calculus of Variations (Gossez, Lami Dozo, Mawhin, Waelbroeck, eds.). Summer School (Bruxelles 1975), pp. 208–237. Lecture Notes in Mathematics **543**. Berlin-Heidelberg-New York: Springer 1976
14. Trudinger, N.S.: A new proof of the interior gradient bound for the minimal surface equation in n-dimensions. Proc. Nat. Acad. Sci. USA **69**, 821–823 (1972)
15. Trudinger, N.S.: Gradient estimates and mean curvature. Math. Z. **131**, 165–175 (1973)

Received September 14, 1981

Math. Z. 180, 193–201 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Positive Definite Kernels on the Complex Hilbert Sphere

Jens Peter Reus Christensen[1]* and Paul Ressel[2]

[1] Københavns Universitets Matematiske Institut, Universitetsparken 5, DK-2100 København, Danmark

[2] Mathematisch-Geograph. Fakultät der Katholischen Universität Eichstätt, Residenzplatz 12, D-8078 Eichstätt, Federal Republic of Germany

In 1942 Schoenberg published the following result: every continuous function $f\colon [-1,1]\to\mathbb{R}$ having the property that $f(\langle x,y\rangle)$ is a positive definite kernel on the unit sphere of a real (infinite dimensional) Hilbert space, has the power series representation $f(t)=\sum_{n\geq 0} a_n t^n$ where $a_n\geq 0$ and $\sum a_n<\infty$ (cf. [4]). He obtained this result by finite dimensional approximation. In [1] we gave a completely different new proof for Schoenberg's theorem. We defined K_1 to be the set of all functions $f\colon [-1,1]\to[-1,1]$ operating on real positive definite matrices with entries in $[-1,1]$, normalized by $f(1)=1$, and showed that K_1 is a Bauer simplex whose extreme points are the closure of the monomials $t\mapsto t^n, n\geq 0$, which of course means, that the two limit points $1_{\{-1,1\}}$ and $1_{\{1\}}-1_{\{-1\}}$ have to be added. A simple and short approximation argument then gave Schoenberg's Theorem, even for non-continuous functions.

In the present paper we want to prove the complex analogue of Schoenberg's Theorem which is far from being an obvious generalization. It turns out that a continuous function f such that $f(\langle x,y\rangle)$ is positive definite on the complex Hilbert sphere has a unique representation of the form

$$f(z)=\sum_{n,m=0}^{\infty} a_{n,m} z^n \bar{z}^m$$

where $a_{n,m}\geq 0$ and $\sum_{n,m=0}^{\infty} a_{n,m}<\infty$. The discontinuous functions with this property however are of a more complicated nature than in the real case.

Our Theorem 1 is motivated by results of Rider, Herz and Moran concerning functions operating on positive definite functions on abelian groups, a good reference to which can be found in the recent book of Graham and McGehee ([2], Chap. 9.6).

* The first named author gratefully acknowledges the generous support given by the Thyssen-Stiftung during his stay at the University of Münster

0025-5874/82/0180/0193/$01.80

Let us recall some definitions and simple facts. A complex $n \times n$-matrix (a_{jk}) is called *positive* (semi-)*definite* if and only if

$$\sum_{j,k=1}^{n} c_j \bar{c}_k a_{jk} \geqq 0$$

for all $(c_1, \ldots, c_n) \in \mathbb{C}^n$. The matrix then is automatically hermitian, i.e. $a_{kj} = \bar{a}_{jk}$, and $a_{jj} \geqq 0$ for all j. If X is a nonempty set, then $\varphi: X \times X \to \mathbb{C}$ is called a *positive definite kernel* if and only if for every finite subset $\{x_1, \ldots, x_n\} \subseteq X$ the $n \times n$-matrix $(\varphi(x_j, x_k))$ is positive definite. If this holds then $\bar{\varphi}$ and $\mathrm{Re}(\varphi) = \frac{1}{2}(\varphi + \bar{\varphi})$ are positive definite, too. We will make repeated use of the fundamental fact, known as Schur's Product Theorem, that the (closed, convex) cone of all positive definite kernels on some set X is also stable under multiplication.

Let $D := \{z \in \mathbb{C}: |z| \leqq 1\}$ be the closed unit disc of the complex plane and let $T := \{z \in \mathbb{C}: |z| = 1\}$ be the torus group. When we consider T as a discrete group, we shall write T_d instead of T. We introduce C to be the set of all functions $f: D \to \mathbb{C}$ operating on positive definite matrices in the sense that $(f(a_{jk}))$ is positive definite whenever (a_{jk}) is a positive definite matrix with entries in D. By the Schur Product Theorem C is a multiplicative closed convex cone in $\mathbb{C}^D$. It is immediately seen that each $f \in C$ is nonnegative on $[0, 1]$. For $z \in D$ the matrix $\begin{pmatrix} 1 & z \\ \bar{z} & 1 \end{pmatrix}$ is positive definite hence so is

$$\begin{pmatrix} f(1) & f(z) \\ f(\bar{z}) & f(1) \end{pmatrix}$$

if $f \in C$, and therefore $f(\bar{z}) = \overline{f(z)}$ as well as $|f(z)| \leqq f(1)$ for all $z \in D$, showing in particular, that

$$K := \{f \in C: f(1) = 1\}$$

is a compact convex base for C.

For the proof of Theorem 1 below, which determines the extreme points of K, we need several lemmas, which we prove first.

Lemma 1. *For $c \in \mathbb{C}$, $|c| \leqq 1$, define $M_c := \begin{pmatrix} 1 & c & c \\ \bar{c} & 1 & c \\ \bar{c} & \bar{c} & 1 \end{pmatrix}$. Then*

(i) *M_c is positive definite if and only if $[3 - 2\,\mathrm{Re}(c)]\,|c|^2 \leqq 1$.*

(ii) *M_t is positive definite for all $t \in [-\frac{1}{2}, 1]$, but not for $t < -\frac{1}{2}$.*

(iii) *If $t \in [0, 1[$ and $c = te^{i\theta}$, where $0 \leqq \theta < 2\pi$, then M_c is positive definite if and only if*

$$\theta \leqq \arccos(\tfrac{3}{2}t^{-1} - \tfrac{1}{2}t^{-3}).$$

Hence M_c is positive definite in a complex neighborhood of t.

Proof. (i) We have $\det \begin{pmatrix} 1 & c \\ \bar{c} & 1 \end{pmatrix} = 1 - |c|^2$ and

$$\det(M_c) = 1 - [3 - 2\,\mathrm{Re}(c)]\,|c|^2.$$

(ii) follows from $\det(M_t)=2t^3-3t^2+1=(t-1)^2(t+\frac{1}{2})$.

(iii) Solving the equation

$$(3-2t\cos\theta)\,t^2=1$$

for θ gives immediately the result. □

Lemma 2. *Let* $f(t):=\frac{3}{2}t^{-k}-\frac{1}{2}t^{-3k}$ *and*

$$g(t):=\cos[(k+2)\arccos(\tfrac{3}{2}t^{-1}-\tfrac{1}{2}t^{-3})]$$

where $t>0$ *and* $k\in\mathbb{N}$. *Then there is an open neighborhood* U *of* 1 *such that* $f(t)>g(t)$ *for all* $t\in U\setminus\{1\}$.

Proof. First note that $|\frac{3}{2}s-\frac{1}{2}s^3|\leqq 1$ for $|s|\leqq 2$ so that g is well defined in a neighborhood of 1. Further the expansion

$$\cos[(k+2)\arccos x]=x^{k+2}-\binom{k+2}{2}x^k(1-x^2)+\binom{k+2}{4}x^{k-2}(1-x^2)^2-+\ldots$$

shows that g is C^∞ for t near to 1.

We have

$$f'(t)=-\tfrac{3}{2}k\,t^{-k-1}+\tfrac{3}{2}k\,t^{-3k-1}$$

$$f''(t)=\tfrac{3}{2}k(k+1)\,t^{-k-2}-\tfrac{3}{2}k(3k+1)\,t^{-3k-2}$$

and so $f(1)=1, f'(1)=0, f''(1)=-3k^2$.

Furthermore

$$g'(t)=\sin[(k+2)\arccos(\tfrac{3}{2}t^{-1}-\tfrac{1}{2}t^{-3})]\,(k+2)\frac{-\frac{3}{2}t^{-2}+\frac{3}{2}t^{-4}}{\sqrt{1-(\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3})^2}}$$

and

$$\frac{1}{k+2}g''(t)=-\cos[\ ](k+2)\frac{(-\frac{3}{2}t^{-2}+\frac{3}{2}t^{-4})^2}{1-(\frac{3}{2}t^{-1}-\frac{3}{2}t^{-3})^2}$$

$$+\frac{\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3}}{\sqrt{\quad}^3}(-\tfrac{3}{2}t^{-2}+\tfrac{3}{2}t^{-4})^2\cdot\sin[\ ]+\sin[\ ]\frac{1}{\sqrt{\quad}}(3t^{-3}-6t^{-5}).$$

By l'Hospital's rule

$$\lim_{t\to 1}\frac{(\frac{3}{2}t^{-4}-\frac{3}{2}t^{-2})^2}{1-(\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3})^2}=\lim_{t\to 1}\frac{2(\frac{3}{2}t^{-4}-\frac{3}{2}t^{-2})(-6t^{-5}+3t^{-3})}{-2(\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3})(-\frac{3}{2}t^{-2}+\frac{3}{2}t^{-4})}=3$$

and

$$\lim_{t\to 1}\frac{\sin[\ ]}{\sqrt{\quad}}=\lim_{t\to 1}\frac{\cos[\ ]\cdot(k+2)\cdot\frac{-1}{\sqrt{\quad}}(-\frac{3}{2}t^{-2}+\frac{3}{2}t^{-4})}{\frac{1}{2\sqrt{\quad}}\cdot(-2)\cdot(\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3})(-\frac{3}{2}t^{-2}+\frac{3}{2}t^{-4})}=k+2$$

so that $g(1)=1$, $g'(1)=0$ and $g''(1)=-3(k+2)^2<f''(1)$. □

Lemma 3. *Let $\varphi(z)=z^n\bar{z}^m$, where $n, m\in\mathbb{Z}$ and $n+m\geqq 0$. Then, if φ operates on positive definite matrices, it follows $n\geqq 0$ and $m\geqq 0$.*

Proof. If n or m is strictly negative, then

$$|n-m|\geqq n+m+2.$$

Let $c=t\,e^{i\theta}$, $0\leqq t<1$, and put $k:=n+m\in\mathbb{N}_0$. Then $|\varphi(c)|=t^k$ and $\operatorname{Re}\varphi(c)=t^k\cdot\cos(|n-m|\,\theta)$ and by Lemma 1 the matrix $M_{\varphi(c)}$ is positive definite if and only if

$$|n-m|\,\theta\leqq\arccos(\tfrac{3}{2}t^{-k}-\tfrac{1}{2}t^{-3k})$$

whereas M_c is positive definite if and only if

$$\theta\leqq\arccos(\tfrac{3}{2}t^{-1}-\tfrac{1}{2}t^{-3}).$$

Therefore if $\theta\leqq\arccos(\frac{3}{2}t^{-1}-\frac{1}{2}t^{-3})$ but

$$\theta>\frac{1}{k+2}\arccos(\tfrac{3}{2}t^{-k}-\tfrac{1}{2}t^{-3k})\geqq\frac{1}{|n-m|}\arccos\ldots$$

then M_c is positive definite, but not so $M_{\varphi(c)}$. Such a choice of θ is possible for t close to 1 by Lemma 2, thus finishing the proof of Lemma 3. □

Lemma 4. *Let $f\in C$ and $z\in D$ and define the functions g_1, g_2, h_1, h_2: $D\to\mathbb{C}$ by $g_{1,2}(w)=f(w)\pm\frac{1}{2}[f(z\,w)+f(\bar{z}\,w)]$ and $h_{1,2}(w)=f(w)\pm\frac{1}{2i}[f(z\,w)-f(\bar{z}\,w)]$. Then g_1, g_2, h_1, h_2 all belong also to C.*

Proof. Let (a_{jk}) be a fixed positive definite $n\times n$-matrix with entries in D, and let $(c_1,\ldots,c_n)\in\mathbb{C}^n$ also be fixed. Define the function F: $D\to\mathbb{C}$ by

$$F(z):=\sum_{j,k=1}^{n}c_j\bar{c}_k f(z\,a_{jk}).$$

We claim that $F\in C$. In fact for any positive definite $m\times m$-matrix (b_{pq}) with entries in D and for all $(d_1,\ldots,d_m)\in\mathbb{C}^m$ we have

$$\sum_{p,q=1}^{m}d_p\bar{d}_q F(b_{pq})=\sum_{p,q=1}^{m}\sum_{j,k=1}^{n}d_p c_j\overline{d_q c_k}\,f(b_{pq}\,a_{jk})\geqq 0$$

because the tensor product $(b_{pq})\otimes(a_{jk})$ is positive definite, too. Therefore

$$|F(z)|\leqq F(1)=\sum_{j,k=1}^{n}c_j\bar{c}_k f(a_{jk}),$$

hence, denoting $G=\operatorname{Re}(F)$ and $H=\operatorname{Im}(F)$,

$$|G(z)|\leqq F(1)\quad\text{and}\quad|H(z)|\leqq F(1).$$

But $G(z)=\frac{1}{2}\sum c_j\bar{c}_k[f(z\,a_{jk})+f(\bar{z}\,a_{jk})]$ and $H(z)=\frac{1}{2i}\sum c_j\bar{c}_k[f(z\,a_{jk})-f(\bar{z}\,a_{jk})]$ and thus

$$\sum c_j\bar{c}_k\,g_\alpha(a_{jk})\geqq 0,\qquad \sum c_j\bar{c}_k\,h_\alpha(a_{jk})\geqq 0$$

for $\alpha=1,2$, i.e. $\{g_1,g_2,h_1,h_2\}\subseteq C$. □

Theorem 1. *The set K of normalized operating functions is a Bauer simplex whose extreme points consist of the closure of $\{z^n\bar{z}^m\colon n,m\in\mathbb{N}_0\}$ in the pointwise topology of $\mathbb{C}^D$. More precisely, denoting by ρ_0 the zero extension of any character ρ on the discrete torus T_d (i.e. $\rho_0=\rho$ on T and $\rho_0|\mathring{D}=0$), we have*

$$ex(K)=\{z^n\bar{z}^m\colon n,m\in\mathbb{N}_0\}\cup\{\rho_0\colon \rho\in\hat{T}_d\}.$$

Proof. Let $f\in ex(K)$ and $z\in D$. By Lemma 4 there exist $\alpha=\alpha(z)\geqq 0$ and $\beta=\beta(z)\geqq 0$ such that

$$f(w)+\tfrac{1}{2}[f(z\,w)+f(\bar{z}\,w)]=\alpha f(w)\qquad \text{for all } w\in D.$$

$$f(w)+\frac{1}{2i}[f(z\,w)-f(\bar{z}\,w)]=\beta f(w)\qquad \text{for all } w\in D.$$

This implies $\alpha=1+\frac{1}{2}[f(z)+f(\bar{z})]$ and $\beta=1+\frac{1}{2i}[f(z)-f(\bar{z})]$ and further

$$f(z\,w)+f(\bar{z}\,w)=f(z)\,f(w)+f(\bar{z})\,f(w)$$

as well as

$$f(z\,w)-f(\bar{z}\,w)=f(z)\,f(w)-f(\bar{z})\,f(w)$$

so that finally $f(z\,w)=f(z)\,f(w)$ for all $z,w\in D$, that is, f is multiplicative.

We conclude that $f|T$ is a character on the (discrete) torus and from [1], Theorem 1 we know that either $f]-1,1[\equiv 0$ in which case the multiplicativity of f would imply $f|\mathring{D}\equiv 0$, or there is a $k\in\mathbb{N}_0$ such that $f(t)=t^k$ for all $t\in[-1,1]$.

If $f|\mathring{D}$ is identically zero, then $f=\rho_0$ for some $\rho\in\hat{T}_d$. If $f(0)\neq 0$ then

$$f(0)=(f(0))^2=1=f(0\cdot z)=f(0)\,f(z)=f(z)$$

for all $z\in D$, i.e. $f(z)=z^0\bar{z}^0$.

We now assume that there is a $z\in\mathring{D}\setminus\{0\}$ such that $f(z)\neq 0$. Then $f(|z|)=|f(z)|>0$ and $f(\sqrt[n]{|z|})=\sqrt[n]{|f(z)|}\to 1$ for $n\to\infty$ implying that for some large enough $t\in]0,1[$ we have $f(t)>\frac{1}{2}$. Making use of Lemma 1 we find the matrix

$$M_c=\begin{pmatrix}1 & c & c\\ \bar{c} & 1 & c\\ \bar{c} & \bar{c} & 1\end{pmatrix}$$

to be positive definite for $c=t\,e^{i\theta}$, if $e^{i\theta}$ lies in some neighborhood U of 1 in T. If $f|T$ were discontinuous, i.e. $f|T\in\hat{T}_d\setminus\hat{T}$, then it is well known that $\overline{f(U)}=T$, so that we could find some $e^{i\theta_0}$ in U, such that $f(e^{i\theta_0})=e^{i\theta_1}$ is close to -1, say

$\cos\theta_1 < -\frac{3}{4}$. Put $c_0 := t\,e^{i\theta_0}$; then M_{c_0} is positive definite, but not so $M_{f(c_0)}$, because

$$[3-2\,\mathrm{Re}\,f(c_0)]\cdot|f(c_0)|^2 \geqq (3+\tfrac{3}{2})\cdot(\tfrac{1}{2})^2 = \tfrac{9}{8} > 1.$$

This contradiction shows that $f|T$ is continuous and hence for some $p\in\mathbb{Z}$

$$f(z) = z^p \quad \text{for all } z\in T.$$

On the other hand, as already mentioned above, there is a $k\in\mathbb{N}_0$ such that

$$f(t) = t^k \quad \text{for all } t\in[-1,1].$$

The equality $f(-1)=(-1)^k=(-1)^p$ shows k and p to be of equal parity; consequently $n := (k+p)/2$ and $m := (k-p)/2$ are integers and $n+m\geqq 0$. For each $z = t\,e^{i\theta}\in D$ we get

$$f(z) = t^k\,e^{ip\theta} = t^{n+m}\,e^{i\theta(n-m)} = z^n\,\bar{z}^m$$

and Lemma 3 implies finally $n\geqq 0$ and $m\geqq 0$.

We have now proved that an extreme point of K either is of the form $z^n\,\bar{z}^m$ for some $n,m\in\mathbb{N}_0$, or else is the zero extension of some character on the (discrete) torus.

To prove the reverse inclusion we remark, that Schur's Product Theorem together with the elementary fact, that $(\overline{a_{jk}})$ is positive definite if (a_{jk}) is so, imply $\{z^n\,\bar{z}^m\colon n,m\in\mathbb{N}_0\}\subseteq K$. Furthermore we see that

$$1_T(z) = \lim_{n\to\infty} z^n\,\bar{z}^n$$

belongs to K, so that the zero extension of every continuous character belongs to K. Now it is well known (and follows easily from Pontryagins Duality Theorem) that the continuous characters on T are dense in the group of all characters on T, i.e. in $\hat{T}_d$, and this fact shows that the zero extension of discontinuous characters on T, too, belong to K.

Now let $K_0 := \{f\colon D\to D\colon f$ is bounded, $f(1)=1$, $f(\bar{z})=\overline{f(z)}$ for all $z\in D$, $(f(a_j\bar{a}_k))$ is positive definite for every finite set $a_1,\ldots,a_n\in D\}$. In [3], Theorem 2.1 it was shown that K_0 is a Bauer simplex whose extreme points are precisely the so called $*$-semicharacters on D, i.e. the multiplicative functions $\rho\colon D\to D$ with the additional properties $\rho(\bar{z})=\overline{\rho(z)}$ and $\rho(1)=1$. Since $K\subseteq K_0$ obviously, we conclude

$$ex(K)\subseteq\overline{\{z^n\,\bar{z}^m\colon n,m\in\mathbb{N}_0\}}\subseteq K\cap ex(K_0)\subseteq ex(K).$$

The fact, that $ex(K)$ is closed under multiplication, together with the Stone-Weierstrass Theorem show that K is a Bauer simplex. □

In the proof of Theorem 2 below we will use the following approximation result.

Lemma 5. *Let (a_{jk}) be a hermitian $p\times p$-matrix, let $X\neq\varnothing$ be a set and let $\varphi\colon X\times X\to\mathbb{C}$ be positive definite and bounded. Suppose that for every $n\in\mathbb{N}$ there exist $\{x_{j,\alpha}^n\colon j=1,\ldots,p;\ \alpha=1,\ldots,n\}\subseteq X$ such that*

$$\varphi(x^n_{j,\alpha}, x^n_{k,\beta}) = \begin{cases} a_{jk}, & \text{if } j \neq k \\ a_{jj}, & \text{if } j = k \text{ but } \alpha \neq \beta. \end{cases}$$

Then the matrix (a_{jk}) is positive definite, too.

Proof. Let $c_1, \ldots, c_p \in \mathbb{C}$ be given. For $n \in \mathbb{N}$ choose $\{x^n_{j,\alpha}\}$ as indicated and put $d_{j,\alpha} := \frac{c_j}{n}$, $1 \leqq j \leqq p$, $1 \leqq \alpha \leqq n$. Then

$$\begin{aligned} 0 &\leqq \sum_{j,k=1}^{p} \sum_{\alpha,\beta=1}^{n} d_{j\alpha}\, \overline{d_{k\beta}}\, \varphi(x^n_{j,\alpha}, x^n_{k,\beta}) \\ &= \sum_{\substack{j,k=1 \\ j \neq k}}^{p} c_j \bar{c}_k a_{jk} + \frac{n^2 - n}{n^2} \sum_{j=1}^{p} |c_j|^2 a_{jj} + \frac{1}{n^2} \sum_{j=1}^{p} \sum_{\alpha=1}^{n} |c_j|^2 \varphi(x^n_{j,\alpha}, x^n_{j,\alpha}) \\ &= \sum_{j,k=1}^{p} c_j \bar{c}_k a_{jk} + R_n \end{aligned}$$

where

$$R_n = -\frac{1}{n} \sum_{j=1}^{p} |c_j|^2 a_{jj} + \frac{1}{n^2} \sum_{j=1}^{p} \sum_{\alpha=1}^{n} |c_j|^2 \varphi(x^n_{j,\alpha}, x^n_{j,\alpha}).$$

The kernel φ being bounded, R_n tends to zero for $n \to \infty$. □

Let H denote a fixed infinite dimensional complex Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and denote by $S := \{x \in H: \|x\| = 1\}$ the unit sphere of H. Restricted to S the scalar product is a mapping into the closed unit disc D and plainly for every $f \in K$ the composed map $f(\langle x, y \rangle)$ is a positive definite kernel. The next theorem shows that the converse also holds.

Theorem 2. *Let $f: D \to \mathbb{C}$ be a function such that*

(1) $f(1) = 1$

(2) *$f(\langle x, y \rangle)$ is a positive definite kernel on $S \times S$.*

Then f belongs to K.

Proof. Let (a_{jk}) be a positive definite $p \times p$-matrix with entries in D. We have to show that $(f(a_{jk}))$ is also positive definite. Without loss of generality assume $H = l^2_{\mathbb{C}}(\mathbb{N})$ and let in the terminology of Lemma 5 be $X = S$ and $\varphi(x, y) = f(\langle x, y \rangle)$. We first choose p vectors $z_1, \ldots, z_p \in \mathbb{C}^p$ such that $a_{jk} = \langle z_j, z_k \rangle$ for all j, k. For a given $n \in \mathbb{N}$ we define $x^n_{j,\alpha} \in H$ by their coordinates

$$x^n_{j,\alpha}(q) = \begin{cases} z_j(q) & \text{if } 1 \leqq q \leqq p \\ \sqrt{1 - \|z_j\|^2} & \text{if } q = p + (j-1)\, n + \alpha \\ 0 & \text{else.} \end{cases}$$

Then $\langle x^n_{j,\alpha}, x^n_{k,\beta} \rangle$ equals $\langle z_j, z_k \rangle = a_{jk}$ for $j \neq k$, and also for $j = k$ but $\alpha \neq \beta$, whereas $\|x^n_{j,\alpha}\|^2 = 1$ for all j, α and n. Lemma 5 now shows the positive definiteness of $(f(a_{jk}))$. □

Corollary 1. *A continuous complex valued function f on D such that $f(\langle x, y\rangle)$ is positive definite on the unit sphere of some infinite dimensional complex Hilbert space, has the unique series representation*

$$f(z)=\sum_{n,m=0}^{\infty} a_{n,m} z^n \bar{z}^m \quad \text{for all } z\in D,$$

where all coefficients $a_{n,m}$ are nonnegative and $\sum_{n,m=0}^{\infty} a_{n,m}<\infty$.

Proof. By Theorem 1 there is a uniquely determined nonnegative Radon measure μ on $ex(K)$ such that

$$f(z)=\int_{ex(K)} \rho(z)\, d\mu(\rho) \quad \text{for all } z\in D.$$

Let $\tilde{K}$ be the countable set of functions $z\mapsto z^n \bar{z}^m$, $n,m\in\mathbb{N}_0$. Of course $\tilde{K}$ is measurable and with $a_{n,m}:=\mu(\{z^n \bar{z}^m\})$ we have

$$f(z)-\sum_{n,m=0}^{\infty} a_{n,m} z^n \bar{z}^m=\int_{ex(K)\setminus\tilde{K}} \rho(z)\, d\mu(\rho)$$

showing that the function on the righthandside is also continuous on all of D. The description of $ex(K)$ given in Theorem 1 however shows that this function is identically zero in $\mathring{D}$; therefore $\mu(ex(K)\setminus\tilde{K})=0$ □

Corollary 2. *Let $f_1, f_2, \ldots$ be a sequence of continuous functions of the type described in Corollary 1 and let them have the series representations*

$$f_k(z)=\sum_{n,m=0}^{\infty} a_{n,m}^{(k)} z^n \bar{z}^m, \quad z\in D,\ k=1,2,\ldots.$$

If the pointwise limit $f(z)=\lim_{k\to\infty} f_k(z)$ exists for all $z\in D$ and if $f|[0,1]$ is continuous in 1, then f is continuous everywhere and for the coefficients $a_{n,m}$ in

$$f(z)=\sum_{n,m=0}^{\infty} a_{n,m} z^n \bar{z}^m$$

we have $a_{n,m}=\lim_{k\to\infty} a_{n,m}^{(k)}$ for all $n,m\in\mathbb{N}_0$.

Proof. The cone C being closed we have of course $f\in C$. Let f have the integral representation

$$f(z)=\int_{ex(K)} \rho(z)\, d\mu(\rho).$$

Then

$$0=\int_{ex(K)\setminus\tilde{K}} \rho\left(1-\frac{1}{j}\right) d\mu(\rho)\to\int_{ex(K)\setminus\tilde{K}} \rho(1)\, d\mu(\rho)=\mu(ex(K)\setminus\tilde{K}),$$

so that μ is in fact concentrated on $\tilde{K}$. The affine isometrie between the two Bauer simplices K and $M_+^1(ex(K))$, the Radon probabilities on $ex(K)$, and the fact that the functions $z^n \bar{z}^m$ are isolated points in $ex(K)$ (which is easily seen), imply the convergence of the corresponding coefficients. □

References

1. Christensen, J.P.R., Ressel, P.: Functions operating on positive definite matrices and a theorem of Schoenberg. Trans. Amer. Math. Soc. **243**, 89–95 (1978)
2. Graham, C.C., McGehee, O.C.: Essays in commutative harmonic analysis. Berlin-Heidelberg-New York: Springer 1979
3. Lindahl, R.J., Maserick, P.H.: Positive-definite functions on involution semigroups. Duke Math. J. **38**, 771–782 (1971)
4. Schoenberg, I.J.: Positive definite functions on spheres. Duke Math. J. **9**, 96–108 (1942)

Received September 4, 1981; received in final form December 16, 1981

Math. Z. 180, 203–215 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Räume normierter Bilinearformen und Cliffordstrukturen

Thomas Bier und Ulrich Schwardmann

Mathematisches Institut der Universität, Bunsenstraße 3–5, D-3400 Göttingen, Bundesrepublik Deutschland

Sei $N_{k+1,n}(\mathbb{R}^n)=\{f\colon \mathbb{R}^{k+1}\times\mathbb{R}^n\to\mathbb{R}^n \mid \|f(x,y)\|=\|x\|\cdot\|y\|\}$ der Raum aller $\mathbb{R}$-bilinearen normierten Abbildungen. In der vorliegenden Arbeit soll dieser Raum untersucht werden. Er ist im wesentlichen der Quotient von zwei wohlbekannten Liegruppen. In dieser Darstellung können wir die Homotopiegruppen (in einem stabilen Sinn) und die Kohomologiegruppen mit $\mathbb{Z}/2$-Koeffizienten berechnen.

Im ersten Abschnitt studieren wir die Zerlegung des Raumes $N_{k+1,n}(\mathbb{R}^n)$ in ein Produkt $O(n)\times N'_{k+1,n}(\mathbb{R}^n)$ und zeigen, daß der Raum $N'_{k+1,n}(\mathbb{R}^n)$ ein homogener Raum der Form $O(n)/G^k(m)$ ist. Im zweiten Abschnitt berechnen wir die Gruppen $G^k(m)$. Sie sind selbst wieder Gruppen der Form $O(m)$, $U(m)$, $Sp(m)$ oder Produkte zweier solcher Gruppen. Im dritten Abschnitt führen wir eine stabile Version dieser Räume ein (N'_{k+1}) und zeigen, daß $\pi_i(N'_{k+1})\cong KO^{-i}(\mathbb{R}P(k+1))$ ist. Damit kennt man die Homotopiegruppen von N'_{k+1} in einem stabilen Sinn. Im vierten und fünften Abschnitt berechnen wir die additive Kohomologie mit $\mathbb{Z}/2$-Koeffizienten von $N'_{k+1,n}(\mathbb{R}^n)$. Dazu werden im vierten Abschnitt die Abbildungen $Bi_k\colon BG^k(m)\to BO(n)$ in der Kohomologie berechnet. Die Ergebnisse dieser Rechnung werden auch in [6] benutzt. Im fünften Abschnitt werden dann die additiven Kohomologiegruppen angegeben. Ein in diesem Zusammenhang auftretender Modul über der mod 2 Steenrod-Algebra wird in einem kurzen Anhang definiert.

Wir danken Professor L. Smith für seine Anregungen und seine Unterstützung.

1. Darstellung von $N_{k+1,n}(\mathbb{R}^n)$ als homogener Raum

Sei $\mathbb{R}^{k+1}$ mit der kanonischen Basis $b_0, b_1, \ldots, b_k$ und mit der Norm $\|x\|=x_0^2+\ldots+x_k^2$ und $\mathbb{R}^n$ mit der kanonischen Basis $c_1,\ldots,c_n$ und der Norm $\|y\|=y_1^2+\ldots+y_n^2$ versehen.

Definition. $N_{k+1,n}(\mathbb{R}^n)=\{f\colon \mathbb{R}^{k+1}\times\mathbb{R}^n\to\mathbb{R}^n \mid f\ \mathbb{R}\text{-bilinear},\ \|f(x,y)\|=\|x\|\cdot\|y\|\}$.

0025-5874/82/0180/0203/$02.60

$$N'_{k+1,n}(\mathbb{R}^n) = \{f \in N_{k+1,n}(\mathbb{R}^n) \mid f(b_0, y) = y \text{ für alle } y \in \mathbb{R}^n\}.$$

$N_{k+1,n}(\mathbb{R}^n)$ heißt der *Raum der normierten Bilinearformen auf* $\mathbb{R}^n$ und $N'_{k+1,n}(\mathbb{R}^n)$ der *Raum der (k-ten) Cliffordstrukturen auf* $\mathbb{R}^n$. Die Räume $N_{k+1,n}(\mathbb{R}^n)$ und $N'_{k+1,n}(\mathbb{R}^n)$ sind als Teilräume des $(k+1)$-fachen Produkts von $\mathbb{R}^{n\times n}$ topologische Räume, die kompakt sind.

Lemma 1. $N_{k+1,n}(\mathbb{R}^n) \cong O(n) \times N'_{k+1,n}(\mathbb{R}^n)$.

Beweis. Es gibt eine freie $O(n)$-Operation auf $N_{k+1,n}(\mathbb{R}^n)$, die durch $(A \circ f)(x, y) = A \cdot f(x, y)$ für $A \in O(n)$ gegeben ist. Dazu erhält man das assoziierte Prinzipalbündel

$$O(n) \to N_{k+1,n}(\mathbb{R}^n) \to B := N_{k+1,n}(\mathbb{R}^n)/O(n). \tag{1}$$

Sei $F: \mathbb{R}^n \to \mathbb{R}^n$ durch $F(y) := f(b_0, y)$ definiert. Dann ist die Abbildung

$$B \to N'_{k+1,n}(\mathbb{R}^n),$$
$$[f] \mapsto F^{-1} \circ f$$

eine bijektive stetige Abbildung mit der Umkehrabbildung $g \to [g]$, also ein Homöomorphismus.

Die Inklusion $N'_{k+1,n}(\mathbb{R}^n) \subset N_{k+1,n}(\mathbb{R}^n)$ liefert einen Schnitt des Prinzipalbündels (1) und deshalb zerlegt sich $N_{k+1,n}(\mathbb{R}^n)$ wie angegeben.

Man kann ein $f \in N'_{k+1,n}(\mathbb{R}^n)$ als k lineare schiefsymmetrische Abbildungen $f_1, \ldots, f_k: \mathbb{R}^n \to \mathbb{R}^n$ auffassen, die den Bedingungen

$$f_i \in O(n), \quad f_i^2 = -\mathrm{id}, \quad f_i f_j = -f^j f^i \quad i,j = 1, \ldots, k; \; i \neq j \tag{2}$$

genügen. Umgekehrt definieren k solche Abbildungen mit den Bedingungen (2) ein Element aus $N'_{k+1,n}(\mathbb{R}^n)$ [3].

Sei G_k die Gruppe, die durch Erzeugende und Relationen wie folgt gegeben ist:

Erzeugende $\quad \varepsilon, e_1, \ldots, e_k$

Relationen $\quad \varepsilon e_i = e_i \varepsilon, \quad e_i^2 = \varepsilon, \quad e_i e_j = \varepsilon e_j e_i, \quad \varepsilon^2 = 1.$

G_k ist eine extraspezielle Gruppe der Ordnung 2^{k+1}.

Die Elemente aus $N'_{k+1,n}(\mathbb{R}^n)$ und die Darstellungen ρ von G_k über $\mathbb{R}$ mit $\rho(\varepsilon) = -\mathrm{id}$ entsprechen sich eineindeutig wegen der Relationen (2) [3].

Also gilt:

Lemma 2. *Der Raum* $N'_{k+1,n}(\mathbb{R}^n)$ *und der Raum aller Darstellungen von* G_k, $D_k := \{\rho: G_k \to O(n) \mid \rho(\varepsilon) = -\mathrm{id}\}$ *sind homöomorph.*

Beweis. Der Raum D_k ist kompakt und Hausdorffsch, und die oben angegebene Abbildung ist stetig und bijektiv.

Bemerkung. Es ist wohlbekannt, daß man die Elemente aus D_k auch als Darstellungen einer Clifford-Algebra auffassen kann [2], daher der Name „Raum der Clifford-Strukturen".

Nun ist auf dem Raum $N'_{k+1,n}(\mathbb{R}^n)=D_k$ eine weitere $O(n)$-Operation durch Konjugation gegeben:

$$\rho \to A\rho A^{-1} \quad \text{für } A\in O(n).$$

Man kann die Darstellungstheorie von G_k aus [3] dazu verwenden, um Aussagen über diese $O(n)$-Operation zu machen.

Lemma 3. *Die $O(n)$-Operation durch Konjugation ist transitiv auf Paaren von Zusammenhangskomponenten.*

Beweis. Äquivalente Darstellungen liegen in demselben Paar von Zusammenhangskomponenten. Denn falls ρ, ρ' äquivalent sind, so gibt es ein $A\in O(n)$ mit $A\rho A^{-1}=\rho'$. Man braucht dann nur A mit einem Weg $A(t)$ mit id bzw. id$_-$ $=\mathrm{Diag}(-1,1,\ldots,1)$ zu verbinden, und erhält einen Weg zwischen ρ und ρ' oder zwischen ρ und $\rho'_-:=\mathrm{id}_-\circ\rho'\circ\mathrm{id}_-$.

Andererseits liefern nichtäquivalente Darstellungen in der diskreten Menge der Äquivalenzklassen von Darstellungen verschiedene Elemente, können also nicht äquivalent sein. Insbesondere entsprechen die Paare von Zusammenhangskomponenten genau den Äquivalenzklassen von Darstellungen.

Man wählt nun eine feste Darstellung ρ von G_k mit $\rho(\varepsilon)=-\mathrm{id}$. Sei $a_{k+1}=2^{\varphi(k)}$ mit $\varphi(k)=\#\{i\,|\,1\leqq i\leqq k,\ i\equiv 0,1,2,4(8)\}$.

Definition. Die (*k-te*) *Cliffordgruppe* ist die Matrizengruppe

$$G^k(m)=\{A\in O(n)\,|\,\rho(e_i)A\rho(e_i)^{-1}=A \text{ für } i=1,\ldots,k\}\subset O(n).$$

Dabei ist $m=m(n,k):=n/a_{k+1}$.

Bemerkung. Wie im folgenden deutlich wird, hängt die Gruppe $G^k(m)$ in den Fällen $k\equiv 3(4)$ von der Wahl der Darstellung ρ ab.

Für die Gruppen $G^k(m)$ gilt

$$G^k(m)\subset G^{k-1}(m')\subset G^{k-2}(m'')\subset\ldots\subset G^0(n):=SO(n)$$

falls man die Darstellungen $\rho', \rho''\ldots$ als Einschränkung von ρ wählt.

Falls nun $k\not\equiv 3(4)$ ist, so gibt es für G_k nur eine irreduzible Darstellungsklasse $[\rho]$, und der Raum ist

$$N'_{k+1,n}(\mathbb{R}^n)=O(n)/G^k(m).$$

Falls $k\equiv 3(4)$ ist, gibt es für G_k zwei verschiedene irreduzible Darstellungsklassen $[\rho_1]$ und $[\rho_2]$. Falls dann $\rho=m_1\rho_1\oplus m_2\rho_2$ ist, schreibt man für $G^k(m)$ genauer $G^k(m_1,m_2)$. In diesem Fall hat der Raum $N'_{k+1,n}(\mathbb{R}^n)$ genau $(m+1)$ verschiedene Zusammenhangskomponenten, die im allgemeinen nicht zueinander homöomorph sind.

Wir haben also bisher gesehen, daß man die Zusammenhangskomponenten der Räume $N'_{k+1,n}(\mathbb{R}^n)$ als $SO(n)/G^k(m)$ schreiben kann. Im nächsten Abschnitt werden die Gruppen $G^k(m)$ ausgerechnet.

2. Die Berechnung der Gruppen $G^k(m)$

Im folgenden Satz geben wir die Liste der Gruppen $G^k(m)$ zusammen mit den Inklusionsabbildungen $i\colon G^k(m)\to O(n)$ an. Dazu benutzt man die Darstellungsmatrizen, wie sie in [1] angegeben sind.

Satz 1. *Die Gruppen $G^k(m)$ sind mit einer Periode 8 in k wie folgt gegeben (sei $k=8r+s$):*

$$
\begin{array}{llll}
s & a_{k+1} & i_s\colon G^k(m) & \longrightarrow O(a_{k+1}m)\\
0 & 16^r & i_0\colon O(m) & \longrightarrow O(16^r m)\\
1 & 16^r 2 & i_1\colon U(m) \longrightarrow O(2m) & \xrightarrow{i_0} O(16^r 2m)\\
2 & 16^r 4 & i_2\colon Sp(m) \longrightarrow U(2m) & \xrightarrow{i_1} O(16^r 4m)\\
3 & 16^r 4 & i_3\colon Sp(m_1)\times Sp(m_2) \longrightarrow Sp(m) & \xrightarrow{i_2} O(16^r 4m)\\
4 & 16^r 8 & i_4\colon Sp(m) \longrightarrow Sp(2m) & \xrightarrow{i_2} O(16^r 8m)\\
5 & 16^r 8 & i_5\colon U(m) \longrightarrow Sp(m) & \xrightarrow{i_4} O(16^r 8m)\\
6 & 16^r 8 & i_6\colon O(m) \longrightarrow U(m) & \xrightarrow{i_5} O(16^r 8m)\\
7 & 16^r 8 & i_7\colon O(m_1)\times O(m_2) \longrightarrow O(m) & \xrightarrow{i_6} O(16^r 8m).
\end{array}
$$

Dabei ist für $s=0$ die Inklusionsabbildung die diagonale Einbettung

$$
A\to\begin{bmatrix} A & & \\ & A & 0 \\ 0 & & \ddots \\ & & & A \end{bmatrix}.
$$

Im Falle $s=3$ und 7, also $k\equiv 3 \bmod 4$ ist eine Äquivalenzklasse von Darstellungen ausgewählt. Die Räume $N'_{k+1,n}(\mathbb{R}^n)$ bestehen dort aus $m+1$ verschiedenen Komponenten von $O(a_{k+1}m)/G^k(m_1,m_2)$. Die Abbildungen links im Falle $s=1,2,4,5,6$ entstehen aus den üblichen Inklusionen dieser Liegruppen.

Beweis. Man wählt möglichst einfache Darstellungsmatrizen $\rho(e_i)$ und berechnet die Cliffordgruppen $G^k(m)$. Der Beweis wird beispielhaft für den Fall $k=4$ durchgeführt.

In diesem Fall kann man die Darstellungsmatrizen für eine irreduzible Darstellung $\bar\rho$ als

$$
\bar\rho(e_1)=\begin{bmatrix} i & 0 \\ 0 & i \end{bmatrix}\quad \bar\rho(e_2)=\begin{bmatrix} j & 0 \\ 0 & j \end{bmatrix}\quad \bar\rho(e_3)=\begin{bmatrix} k & 0 \\ 0 & -k \end{bmatrix}\quad \bar\rho(e_4)=\begin{bmatrix} 0 & k \\ k & 0 \end{bmatrix}
$$

wählen, wobei i,j,k die üblichen reellen 4×4-Matrizen sind, die man durch die Identifikation $\mathbb{H}\cong\mathbb{R}^4$ erhält.

Nun ist $G^4(m)=\{A\in O(8m)\mid \rho(e_i)\,A\,\rho(e_i)^{-1}=A\}$, wobei ρ die m-fache Summe der irreduziblen Darstellung $\bar\rho$ ist.

Man schreibt A blockweise als $A=\begin{bmatrix} B & C \\ D & E \end{bmatrix}$ mit reellen $4m \times 4m$-Matrizen B, C, D, E. Wegen $\rho(e_i) A \rho(e_i)^{-1} = A$ für $i=1,2$ sind die B, C, D, E symplektische Matrizen. Weiterhin ist $\rho(e_3) A \rho(e_3)^{-1} = A$, also

$$Bk=kB \quad kC=-Ck \quad kD=-kD \quad kE=Ek, \quad \text{also } C=D=0.$$

Schließlich ist $\rho(e_4) A \rho(e_4)^{-1} = A$, also

$$Ek=kB, \quad \text{also } E=B.$$

Damit hat A die Form $A=\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix}$ mit $B \in Sp(m)$.

Also ist $G^4(m)=\left\{\begin{bmatrix} B & 0 \\ 0 & B \end{bmatrix} \in Sp(2m) \subset O(8m) \mid B \in Sp(m)\right\} \cong Sp(m)$. Damit ist der Satz für den Fall $k=4$ gezeigt. Die Beweise für die anderen Fälle $k \not\equiv 3(4)$ gehen analog. Man benutzt die Periodizität von 8 in k, um von k auf $k+8$ zu schließen.

In den Fällen $k \equiv 3(4)$ hat man als letzte Darstellungsmatrix der Darstellung $\rho = m_1 \rho_1 \oplus m_2 \rho_2$ eine Diagonalmatrix mit m_1 positiven und m_2 negativen Eingängen:

$$\rho(e_k)=\begin{bmatrix} E_k & & & & \\ & \ddots & & & \\ & & E_k & & \\ & & & -E_k & \\ & & & & \ddots \\ & & & & & -E_k \end{bmatrix}.$$

Daraus ergibt sich wegen $\rho(e_k) A \rho(e_k)^{-1} = A$ für A die Gestalt

$$A=\begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$$

mit $A_1 \in G^{k-1}(m_1)$ und $A_2 \in G^{k-1}(m_2)$.

Damit ist die Berechnung der Gruppen durchgeführt. Die Inklusionsabbildungen ergeben sich (wie im Fall $k=4$ angedeutet) daraus, daß man für die Darstellungen von G_{k-1} die Matrizen $\rho(e_i)$, $i=1, \ldots, k-1$ einer Darstellung von G_k wählen kann.

3. Die stabilen Räume und die Beziehungen mit den KO-Gruppen der reellen projektiven Räume

Man erhält eine stabile Version der Räume durch Stabilisierungsabbildungen, die durch die folgenden Diagramme erklärt sind:

1) $k \not\equiv 3(4)$, $a := a_{k+1}$

$$\begin{array}{ccccc} A & & G^k(m) & \longrightarrow & O(am) & & B \\ \downarrow & & \downarrow & & \downarrow & & \downarrow \\ \begin{bmatrix} 1 & 0 \\ 0 & A \end{bmatrix} & & G^k(m+1) & \longrightarrow & O(a(m+1)) & & \hat{B} \end{array}$$

wobei

$$\hat{B} = \begin{bmatrix} 1 & 0 & 0 & 0 & & & \\ 0 & B_{11} & 0 & B_{12} & \cdots & & \\ 0 & 0 & 1 & 0 & & & \\ 0 & B_{21} & 0 & B_{22} & & & \\ & & & & & 1 & 0 \\ & & & & & 0 & B_{aa} \end{bmatrix}$$

und B_{ij} der $m \times m$ Block der Matrix B an der Stelle (i, j) ist. Dann erhält man die Inklusionen

$$\frac{O(am)}{G^k(m)} \to \frac{O(a(m+1))}{G^k(m+1)}$$

und kommutative Diagramme

$$\begin{array}{ccc} N'_{k+1,n}(\mathbb{R}^n) & \longrightarrow & N'_{k+1,n+a}(\mathbb{R}^{n+a}) \\ \| \wr & & \| \wr \\ \dfrac{O(am)}{G^k(m)} & \longrightarrow & \dfrac{O(a(m+1))}{G^k(m+1)}. \end{array}$$

2) $k \equiv 3(4)$, $a := a_{k+1}$, $m = m_1 + m_2$,

$$\begin{array}{ccc} G^k(m_1, m_2) & \longrightarrow & O(am) \\ \downarrow & & \downarrow \\ G^k(m_1+1, m_2+1) & \longrightarrow & O(a(m+2)) \end{array}$$

mit den Abbildungen ähnlich wie 1) erhält man dann die Inklusionen

$$O(am)/G^k(m) \to O(a(m+2))/G^k(m+2)$$

und entsprechende kommutative Diagramme.

Definition. Sei $a = a_{k+1}$ für $k \not\equiv 3(4)$ und $a = 2a_{k+1}$ für $k \equiv 3(4)$. Der *stabile Raum der (k-ten) Cliffordstrukturen* ist

$$N'_{k+1} := \lim_{\substack{n \to \infty \\ n \equiv 0(a)}} N'_{k+1,n}(\mathbb{R}^n)$$

mit den oben angegebenen Stabilisierungsabbildungen.

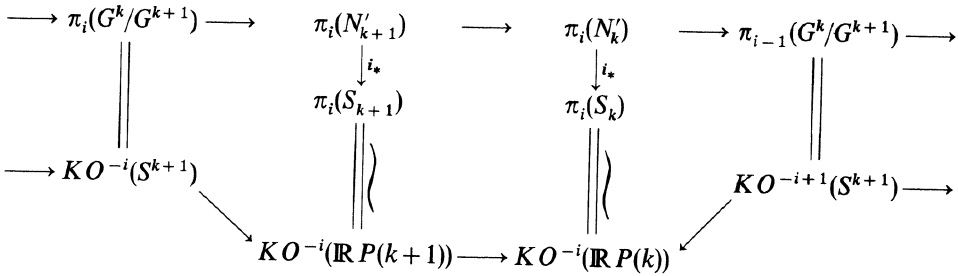
Insbesondere kann man für eine Untergruppe $G^k \subset O$ schreiben

$$N'_{k+1} \cong O/G^k.$$

Für diese stabilen Räume N'_k gilt

Satz 2. $\pi_i(N'_k) \cong KO^{-i}(\mathbb{R}P(k))$.

Beweisskizze. Betrachte $S_{k+1} = \{f\colon S^k \to 0 \mid f(-x) = -f(x)\}$ mit der Topologie des Abbildungsraumes und einer festen Stabilisierung von $O = \bigcup_{n\to\infty} O(n)$. Mit einer Konstruktion des äquivarianten Zusammenklebens von Übergangsfunktionen sieht man, daß $\pi_i(S_{k+1}) \cong KO^{-i}(\mathbb{R}P(k+1))$. Andererseits definiert die Einschränkung von Abbildungen eine Inklusion $N'_{k+1} \to S_{k+1}$. Daraus ergibt sich ein kommutatives Diagramm mit den exakten Zeilen



Dabei ist G^k/G^{k+1} mit der üblichen Stabilisierung ein Bott-Raum. Dann folgt der Satz aus dem 5-er Lemma mit Induktion.

Folgerung. Die Inklusion

$$N'_{k+1} \to S_{k+1} = \{f\colon S^k \to O \mid f(-x) = -f(x)\}$$

ist eine schwache Homotopieäquivalenz.

Diese geometrische Aussage erscheint durchsichtiger als ein entsprechendes Ergebnis in [2].

4. Die Abbildung Bi_k in der $\mathbb{Z}/2$-Kohomologie

Durch die Abbildungen

$$Bi_k\colon BG^k(m) \to BO(a_{k+1}m)$$

werden auf Kohomologieebene Abbildungen der Art:

$$Bi_k^*\colon H^*(BO(a_{k+1}m)) \to H^*(BG^k(m))$$

induziert. Ziel dieses Abschnitts ist es, diese Abbildungen für die $\mathbb{Z}/2$-Kohomologie zu berechnen. Es ist:

$$H^*(BG^k(m);\ \mathbb{Z}/2) = \begin{cases} H^*(BO(m) = \mathbb{Z}/2[w_1, \ldots, w_m] & \text{für } k \equiv 0,6 \bmod 8 \\ H^*(BO(m_1) \times BO(m_2) = \mathbb{Z}/2[w'_1, \ldots, w'_{m_1}, w''_1, \ldots, w''_{m_2}] & \\ & \text{für } k \equiv 7 \bmod 8 \\ H^*(BU(m) = \mathbb{Z}/2[c_1, \ldots, c_m] & \text{für } k \equiv 1,5 \bmod 8 \\ H^*(BSp(m_1) \times BSp(m_2) = \mathbb{Z}/2[p'_1, \ldots, p'_{m_1}, p''_1, \ldots, p''_{m_2}] & \\ & \text{für } k \equiv 3 \bmod 8 \\ H^*(BSp(m) = \mathbb{Z}/2[p_1, \ldots, p_m] & \text{für } k \equiv 2,4 \bmod 8. \end{cases}$$

Die Inklusion der Liegruppen, wie sie in Satz 1 aufgeführt sind, setzen sich aus Inklusionen zusammen, die die folgenden bekannten Ringhomomorphismen in der $\mathbb{Z}/2$-Kohomologie induzieren:

(a) $H^*(BO(2m)) \to H^*(BU(m))$ $\quad w_i \to \begin{cases} c_{i/2} & \text{falls } i \equiv 0(2) \\ 0 & \text{sonst} \end{cases}$

(b) $H^*(BU(2m)) \to H^*(BSp(m))$ $\quad c_i \to \begin{cases} p_{i/2} & \text{falls } i \equiv 0(2) \\ 0 & \text{sonst} \end{cases}$

(c) $H^*(BO(2m)) \to H^*(BO(m))$ $\quad w_i \to \begin{cases} (w_{i/2})^2 & \text{falls } i \equiv 0(2) \\ 0 & \text{sonst} \end{cases}$

(d) $H^*(BSp(2m)) \to H^*(BSp(m))$ $\quad p_i \to \begin{cases} (p_{i/2})^2 & \text{falls } i \equiv 0(2) \\ 0 & \text{sonst} \end{cases}$

(e) $H^*(BSp(m)) \to H^*(BU(m))$ $\quad p_i \to c_i^2$

(f) $H^*(BU(m)) \to H^*(BO(m))$ $\quad c_i \to w_i^2$.

Ist G eine der verwendeten Liegruppen O, U oder Sp, so erhält man für die Abbildung

$$q\colon\ BG(m_1) \times BG(m_2) \to BG(m_1 + m_2)$$

auf Kohomologieebene den induzierten Homomorphismus:

$$q\colon\ H^*(BG(m_1 + m_2)) \to H^*(BG(m_1) \times BG(m_2))$$
$$q(v_i) = \sum_{i=k+l} v'_k v''_l.$$

Dabei sollen $k, l \in \mathbb{N}$ sein und kleiner oder gleich m_1 bzw. m_2. All diese Abbildungen lassen sich elementar aus dem Whitney-Produkt-Satz für siese charakteristischen Klassen in der $\mathbb{Z}/2$-Kohomologie ableiten, oder aus der Beschreibung der verwendeten charakteristischen Klassen als elementarsymmetrische Polynome errechnen (s. dazu z.B. Milnor und Stasheff [5]).

So, wie sich die in Satz 1 beschriebenen Inklusionen induktiv aus den bekannten Inklusionen für Liegruppen zusammensetzen, ergeben sich die induzierten Homomorphismen in der $\mathbb{Z}/2$-Kohomologie der klassifizierenden Räume induktiv aus den bekannten Abbildungen oben. Dies ist im folgenden Satz zusammengefaßt.

Satz 3. *Die Ringhomomorphismen der $\mathbb{Z}/2$-Kohomologie:*

$$Bi_k^*\colon\ H^*(BO(a_{k+1}m)) \to H^*(BG^k(m))$$

lassen sich mit einer Periode 8 wie folgt beschrieben (sei $k=8r+s$ *und* $a=a_{k+1}$*):*

s	a_{k+1}	$BG^k(m)$	$H^*(BO(am))\to H^*(BG^k(m))$	
0	16^r	$BO(m)$	$w_i \to \begin{cases}(w_{i/16^r})^{16^r}\\ 0\end{cases}$	*falls* $i\equiv 0 \bmod a$ *sonst*
1	$16^r\cdot 2$	$BU(m)$	$w_i \to \begin{cases}(c_{i/16^r\cdot 2})^{16^r}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
2	$16^r\cdot 4$	$BSp(m)$	$w_i \to \begin{cases}(p_{i/16^r\cdot 4})^{16^r}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
3	$16^r\cdot 4$	$BSp(m_1)\times BSp(m_2)$	$w_i \to \begin{cases}(q(p_{i/16^r\cdot 4}))^{16^r}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
4	$16^r\cdot 8$	$BSp(m)$	$w_i \to \begin{cases}(p_{i/16^r\cdot 8}))^{16^r\cdot 2}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
5	$16^r\cdot 8$	$BU(m)$	$w_i \to \begin{cases}(c_{i/16^r\cdot 8})^{16^r\cdot 4}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
6	$16^r\cdot 8$	$BO(m)$	$w_i \to \begin{cases}(w_{i/16^r\cdot 8})^{16^r\cdot 8}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst*
7	$16^r\cdot 8$	$BO(m_1)\times BO(m_2)$	$w_i \to \begin{cases}(q(w_{i/16^r\cdot 8}))^{16^r\cdot 8}\\ 0\end{cases}$	*falls* $i\equiv 0(a)$ *sonst.*

Es ist also

$$Bi_k^*(w_i)=\begin{cases}0 & \text{falls } i\not\equiv 0 \bmod a\\ (w_{i/a})^a & \text{falls } i\equiv 0 \bmod a \quad \text{und} \quad k\equiv 0,6 \bmod 8\\ (q(w_{i/a}))^a & \text{falls } i\equiv 0 \bmod a \quad \text{und} \quad k\equiv 7 \bmod 8\\ (c_{i/a})^{a/2} & \text{falls } i\equiv 0 \bmod a \quad \text{und} \quad k\equiv 1,5 \bmod 8\\ (p_{i/a})^{a/4} & \text{falls } i\equiv 0 \bmod a \quad \text{und} \quad k\equiv 2,4 \bmod 8\\ (q(p_{i/a}))^{a/4} & \text{falls } i\equiv 0 \bmod a \quad \text{und} \quad k\equiv 3 \bmod 8.\end{cases}$$

Beweis. Der Beweis wird beispielhaft für den Fall $k=5$ durchgeführt. Das Argument gilt entsprechend für jeden anderen Fall, indem man die dort gegebenen Abbildungen genauso zerlegt und schrittweise behandelt. Die Abbildungen und damit auch ihre Periodizität werden durch die Inklusionen der Liegruppen induziert. Für den Fall $k=5$ ist Bi_k die Verknüpfung der Abbildungen:

$$Bi_5: BU(m)\xrightarrow{j_1} BSp(m)\xrightarrow{j_2} BSp(2m)\xrightarrow{j_3} BO(8m)$$

damit haben die induzierten Ringhomomorphismen der $\mathbb{Z}/2$-Kohomologie, wie unter (a), (b), (d), (e) angegeben, die Gestalt:

$$j_1^*(p_i) = c_i^2$$

$$j_2^*(p_i) = \begin{cases} (p_{i/2})^2 & \text{für } i \equiv 0 \bmod 2 \\ 0 & \text{sonst} \end{cases}$$

$$j_3^*(w_i) = \begin{cases} p_{i/4} & \text{für } i \equiv 0 \bmod 4 \\ 0 & \text{sonst.} \end{cases}$$

Insgesamt ist also:

$$Bi_5^*(w_i) = j_1^* j_2^* j_3^*(w_i) = \begin{cases} (c_{i/8})^4 & \text{für } i \equiv 0 \bmod 8 \\ 0 & \text{sonst} \end{cases}$$

was der Behauptung für diesen Fall entspricht.

Definition. Ist $J \subset R$ ein Ideal, so heißt J *Borelideal*, wenn es von einer Sequenz $(l_1, \ldots, l_i, \ldots)$ erzeugt wird, in der kein l_j Nullteiler von $R/(l_1, \ldots, l_{j-1})$ ist.

Lemma 4. *Die nichtverschwindenden Bilder der Klassen w_i unter Bi_k^* erzeugen ein Borelideal.*

Beweis. Ist $k \not\equiv 3 \bmod 4$, so sind die nichtverschwindenden Bilder von w_i Potenzen der charakteristischen Klassen, die den jeweiligen $\mathbb{Z}/2$-Kohomologiering erzeugen. Eine Sequenz von solchen Elementen erzeugt sicherlich ein Borelideal.

Im Falle $k \equiv 3 \bmod 4$ sind die nichtverschwindenden Bilder unter Bi_k^* Potenzen von Bildern der charakteristischen Klassen unter q. Die Abbildung q ist nun verträglich mit den Homomorphismen

$$H^*(BSp(m)) \to H^*(\mathbb{H}P(\infty) \times \ldots \times \mathbb{H}P(\infty)) \qquad \text{für } k \equiv 3(8)$$

und

$$H^*(BO(m)) \to H^*(\mathbb{R}P(\infty) \times \ldots \times \mathbb{R}P(\infty)) \qquad \text{für } k \equiv 7(8)$$

unter denen die charakteristischen Klassen auf die elementarsymmetrischen Polynome im entsprechenden Grad abgebildet werden. Die elementarsymmetrischen Polynome besitzen ebenfalls die geforderte Eigenschaft: eine Sequenz über Polynome, die jeweils eine fest vorgegebene Potenz der elementarsymmetrischen Polynome sind, erzeugen ein Borelideal. Deshalb müssen auch die Potenzen von Bildern der charakteristischen Klassen unter q^* ein Borelideal erzeugen.

5. Die Kohomologie des Raumes $N'_{k+1,n}(\mathbb{R}^n)$

In diesem Abschnitt sollen Berechnungen zur $\mathbb{Z}/2$-Kohomologie des Raumes der k-ten Cliffordstrukturen gemacht werden. Außerdem wird die Steenrodoperation auf den Erzeugern ausgerechnet. Als Hilfsmittel wird die Eilenberg-Moore-Spektralsequenz benutzt, mittels der Smith [7] ein Verfahren zur Berechnung der Kohomologie von Räumen angibt.

Für $k>0$ läßt sich $i_k: G^k(m)\to O(a_{k+1}m)$ als Abbildung auf eine Komponente $SO(a_{k+1}m)\subset O(a_{k+1}m)$ auffassen. Die induzierte Faserung

$$SO(a_{k+1}m)/G^k(m)\to BG^k(m)\to BSO(a_{k+1}m)$$

ist dann Teil folgenden Diagramms:

$$\begin{array}{ccc} SO(a_{k+1}m) & = & SO(a_{k+1}m) \\ \downarrow & & \downarrow \\ \dfrac{SO(a_{k+1}m)}{G^k(m)} & \longrightarrow & ESO(a_{k+1}m) \\ \downarrow & & \downarrow \\ BG^k(m) & \xrightarrow{Bi_k} & BSO(a_{k+1}m). \end{array}$$

In dieser Betrachtungsweise ist $N'_{k+1,n}(\mathbb{R}^n)$ eine disjunkte Vereinigung von Paaren von Zusammenhangskomponenten der Form:

$$SO(a_{k+1}m)/G^k(m)\cup SO(a_{k+1}m)/G^k(m).$$

Die Kohomologie dieser Räume ist demzufolge die direkte Summe über die Kohomologie der Komponenten. Es reicht daher, jene zu berechnen.

Dies ist ein Diagramm, wie es von Smith zur Anwendung des großen Kollapssatzes ([7], S. 84ff.) vorausgesetzt wird. Zu der in der Eilenberg-Moore-Spektralsequenz gewählten Filtrierung läßt sich dann mit dem großen Kollapssatz die Struktur des assoziiert-graduierten Moduls

$$E^0H^*(SO(a_{k+1}m)/G^k(m))$$

des gesuchten Kohomologierings und damit dessen additive Struktur angeben. Die Voraussetzungen dieses Satzes sind insbesondere wegen Lemma 4 erfüllt.

Satz 4.

$$E^0H(SO(a_{k+1}m)/G^k(m))=\frac{H^*(BG^k(m))}{(\text{im } Bi_k)}\otimes E(u_1,\ldots)$$

wobei mit $a:=a_{k+1}$

$$(\text{im } Bi_k^*)=\begin{cases} (w_1^a,\ldots,w_m^a) & \text{falls } k\equiv 0,6 \bmod 8 \\ (q^*(w_1)^a,\ldots,q^*(w_m)^a) & \text{falls } k\equiv 7 \bmod 8 \\ (c_1^{a/2},\ldots,c_m^{a/2}) & \text{falls } k\equiv 1,5 \bmod 8 \\ (p_1^{a/4},\ldots,p_m^{a/4}) & \text{falls } k\equiv 2,4 \bmod 8 \\ (q^*(p_1)^{a/4},\ldots,q^*(p_m)^{a/4}) & \text{falls } k\equiv 3 \bmod 8 \end{cases}$$

ist.

Die Erzeuger der äußeren Algebra werden in ([7], S. 75ff.) als Bilder einer additiven Relation beschrieben. Da die in dieser Relation auftretende Kohomologie des Faserraumes $ESO(n)$ verschwindet, läßt sich die gegebene Relation

in unserem Fall als die bekannte Kohomologiesuspension:

$$S\colon H^*(BSO(n)) \longrightarrow H^*(BG^k(m), SO(n)/G^k(m)) \xleftarrow{\delta} H^*(SO(n)/G^k(m))$$

auffassen. Sie wird auf alle Elemente angewandt, die im Kern der Abbildung Bi_k^* liegen.

Damit liegen die Erzeuger der äußeren Algebra im homologischen Grad -1 und ihre inneren Grade sind $i \not\equiv -1 \bmod a_{k+1}$ und $0 < i < n$.

Sie sind mittels Kohomologiesuspension bis auf eine Unbestimmtheit festgelegt, die gegeben ist durch das Bild von $H^*(BG^k(m))$.

Weil die Steenrodoperation natürlich ist gegenüber der Kohomologiesuspension, lassen sich so die Bilder der Erzeuger der äußeren Algebra unter dieser Operation berechnen.

Lemma 5.

$$Sq^j(u_i) = \binom{i}{j} u_{i+j}.$$

Beweis. Die Kohomologiesuspension, angewandt auf ein Produkt $u \cdot v$ im Kern von Bi_k, zerfällt in ein Produkt, dessen einer Faktor das Bild von u oder v unter der Verknüpfung:

$$H^*(BSO(n)) \to H^*(BG^k(m)) \to H^*(SO(n)/G^k(m))$$

ist. Weil diese Verknüpfung die Nullabbildung ist, verschwindet auch die Kohomologiesuspension von Produkten.

Mit $u_i = S(w_{i+1})$ und der *Wu*-Formel

$$Sq^k(w_m) = \binom{m-k-1}{0} w_m w_k + \binom{m-k}{1} w_{m+1} w_{k-1} + \ldots + \binom{m-1}{k} w_{m+k}$$

erhält man

$$\begin{aligned} Sq^j(u_i) &= Sq^j(S(w_{i+1})) = S(Sq^j(w_{i+1})) \\ &= S\left(\binom{i}{j} w_{i+j+1}\right) = \binom{i}{j} S(w_{i+j+1}) = \binom{i}{j} u_{i+j}. \end{aligned}$$

Insbesondere sind die Erzeuger u_i unter der A-Modulstruktur nicht unabhängig, A die mod 2 Steenrod-Algebra.

Anhang: Ein Modul über der Steenrod-Algebra

Wir möchten hier einen Modul über der Steenrod-Algebra A beschreiben, der als zweiter „Faktor" in dem Kohomologiering $H^*(N_k'; \mathbb{Z}/2)$ über A im Satz auftreten könnte.

Sei $P = \mathbb{Z}/2[x] \cong H^*(\mathbb{R}P(\infty); \mathbb{Z}/2)$ und setze $x_i := x^i$. Dann operiert A auf P durch $Sq^k(x_i) = \binom{i}{k} x_{i+k}$. Sei $P(2^s)$ der über $\mathbb{Z}/2$ von allen Klassen x_i mit $i \not\equiv$

$-1(2^s)$ erzeugte graduierte $\mathbb{Z}/2$-Untermodul. Man sieht leicht, daß $P(2^s)$ ein unstabiler A-Untermodul von P ist.

Man sieht, daß $P(2^s)$ als A-Modul von den Klassen x_{2^i-1} für $0<i<s$ und $x_{2^i-2^{s-1}-1}$ für $i>s$ erzeugt ist.

Bezeichne mit $U(2^s)$ die von $P(2^s)$ erzeugte freie A-Algebra nach [8, 5.4 Definition, S. 29]. Als graduierter $\mathbb{Z}/2$-Modul ist $U(2^s)$ isomorph zu einer äußeren Algebra $E(u_i | i \not\equiv -1(2^s))$ auf Erzeugenden u_i vom Grad i mit $i \not\equiv -1(2^s)$ [4, Theorem 3.12].

Wir vermuten, daß der Satz 4 auch als Aussage über A-Moduln richtig ist, wenn man den Modul $E(u_1, \ldots)$ durch $U(2^s)$ ersetzt.

Literatur

1. Atiyah, M.F., Bott, R., Shapiro, A.: Clifford modules. Topology **3**, Suppl. 1, 3–33 (1964)
2. Atiyah, M.F., Dupont, J.L.: Vector fields with finite singularities. Acta Math. **128**, 1–40 (1972)
3. Eckmann, B.: Gruppentheoretischer Beweis des Satzes von Hurwitz-Radon über die Komposition quadratischer Formen. Comment. Math. Helv. **15**, 358–367 (1942)
4. Massey, W.S., Peterson, F.P.: Cohomology of certain fibre spaces I. Topology **4**, 47–65 (1965)
5. Milnor, J.W., Stasheff, J.D.: Characteristic classes. Princeton: Princeton University Press 1974
6. Schwardmann, U.: Obstructions for clifford structures on vector bundles. Math. Z. **179**, 69–72 (1982)
7. Smith, L.: Homological algebra and the Eilenberg-Moore spectral sequence. Trans. Amer. Math. Soc. **129**, 58–93 (1967)
8. Steenrod, N.E., Epstein, D.B.A.: Cohomology Operations. Princeton: Princeton University Press 1962

Eingegangen am 14. August 1981

Math. Z. 180, 217-224 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Galois Coverings of the Non-Archimedean Projective Line

Guido Van Steen

Department of Mathematics, Rijksuniversitair Centrum Antwerpen, Groenenborgerlaan 171, B-2020 Antwerpen, Belgium

Introduction

Let K be a complete non-archimedean valued field. In this paper we study curves over K which admit a Galois covering of $\mathbb{P}^1(K)$ of prime degree p. In particular we are concerned with curves which can be parameterised by Schottky groups over K, i.e. by free discontinuous subgroups of $\mathrm{PGL}(2, K)$.

The special case $p=2$ – i.e. the case of so-called hyperelliptic Mumford curves – has been studied in the papers [3, 4] and [5]. The results presented here generalise certain of the results of these papers.

In § 1 we define p-cyclic coverings of $\mathbb{P}^1(K)$; i.e. Galois coverings of $\mathbb{P}^1(K)$ of prime degree p. We give a canonical form for the equation of such a curve and we calculate a genus formula.

In § 2 we confine to the case of Mumford curves; i.e. curves which can be parametrised by Schottky groups over K. A necessary and sufficient condition for a Mumford curve to admit a p-cyclic covering is given. It is remarkable that the Schottky group of such a curve is the free product of some of its subgroups of rank $p-1$. Each of these subgroups parameterises itself a Mumford curve which admits a p-cyclic covering of $\mathbb{P}^1(K)$.

In the last paragraph we calculate the exponents which occur in the equation of curves as studied in § 2 in the case $p \neq \mathrm{char}(K)$.

§ 1. p-Cyclic Coverings of $\mathbb{P}^1(K)$

Let K be a complete non-Archimedean valued field. We study curves over K which admit a Galois covering of $\mathbb{P}^1(K)$ of prime degree. To avoid rationality problems we assume that K is algebraically closed. Let p be a prime number, fixed throughout this paper.

Definition 1.1. A *p-cyclic covering* of $\mathbb{P}^1(K)$ is a Galois covering $\rho: C \to \mathbb{P}^1(K)$ of degree p, where C is a nonsingular projective curve over K with genus $g>0$.

0025-5874/82/0180/0217/$01.60

Let F be the algebraic function field of C. The function field of $\mathbb{P}^1(K)$ will be called $K(T)$, with T a variable. Let $X=\rho^*(T)\in F$. Then F is a Galois extension of the field $K(X)$ of degree p. Let σ^* be a generator for the Galois group of F over $K(X)$. The corresponding automorphism of C is denoted by σ.

Since F is separable over $K(X)$, there exist elements $f\in K(X)$ and $Y\in F$ such that $F=K(X, Y)$ and such that Y satisfies the following equations

a) $\mathrm{char}(K)\neq p$: $Y^p=f$.

b) $\mathrm{char}(K)=p$: $Y^p-Y=f$.

Using standard techniques, one can transform the generator Y such that these equations have the following form:

a) $\mathrm{char}(K)\neq p$: $Y^p=\prod_{i=1}^{r}(X-a_i)^{n_i}$

with $0<n_i<p$ and $a_i\in K$ for all $i=1, \ldots, r$.

b) $\mathrm{char}(K)=p$: $Y^p-Y=\sum_{i=1}^{r}\left(\sum_{\substack{j=1\\ j\not\equiv 0 \bmod (p)}}^{n_i}\frac{\alpha_{ij}}{(X-a_i)^j}\right)$

with $\alpha_{ij}\in K$ for all $j=1, \ldots, n_i$ and $a_i\in K$ and $n_i\not\equiv 0 \bmod (p)$ for all $i=1, \ldots, r$.

Moreover, if $\mathrm{char}(K)\neq p$, we can transform the parameter function T of $\mathbb{P}^1(K)$ in such a way that $\sum_{i=1}^{r} n_i\equiv 0 \bmod (p)$.

Proposition 1.2 ($\mathrm{char}(K)\neq p$ case). *Assume that $F=K(X, Y)$ where Y satisfies the equation $Y^p=\prod_{i=1}^{r}(X-a_i)^{n_i}$ with $0<n_i<p$ and $a_i\in K$ for all $i=1, \ldots, r$ and let $\sum_{i=1}^{r} n_i\equiv 0 \bmod (p)$. The p-cyclic covering $\rho: C\to\mathbb{P}^1(K)$ is then ramified in the points of C lying above $a_1, \ldots, a_r$. Moreover the genus $g=(r-2)\frac{(p-1)}{2}$.*

Proof. We may assume that $a_1=0$. Let P be a point of C lying above a_1. We prove for a suitable local parameter t at P, that $dX/dt=p-1$.

Let $U_\varepsilon=\{x\in\mathbb{P}^1(K)\,|\,|x|\leqq\varepsilon\}$ and let $V_\varepsilon=\rho^{-1}(U_\varepsilon)$. The ring of analytic functions $\mathcal{O}(U_\varepsilon)$ of U_ε is $\mathcal{O}(U_\varepsilon)=\{f\in K[\![X]\!]\,|\,f$ is convergent on $U_\varepsilon\}$. For the ring of analytic functions $\mathcal{O}(V_\varepsilon)$ of V_ε we find that $\mathcal{O}(V_\varepsilon)$ is the normalisation of

$$A=\mathcal{O}(U_\varepsilon)[Y]/\left(Y^p-\prod_{i=1}^{r}(X-a_i)^{n_i}\right).$$

For ε small enough the factor $f_0=\prod_{i=2}^{r}(X-a_i)^{n_i}$ can be written in the form $f_0=\lambda(1+f)$ with $\lambda\in K$ and f such that $|f(x)|<1$ for all $x\in U_\varepsilon$. Therefore, $f_0=\lambda h^p$ for some $h\in\mathcal{O}(U_\varepsilon)$. Replacing Y by $\lambda^{1/p}hY$ we find that $A\cong\mathcal{O}(U_\varepsilon)[Y]/(Y^p-X^{n_1})$. If t is a local parameter at P, then $X=t^pf_1$ and $Y=t^{n_1}f_2$ for some $f_i\in\mathcal{O}(U_\varepsilon)$ such that f_i has order 0 at P; $(i=1, 2)$. It follows that $f_1^{n_1}=f_2^p$. Since $n_1\not\equiv 0 \bmod (p)$ there exists $k, l\in\mathbb{Z}$ such that $kp+ln_1=1$. Hence $f_1=f_1^{kp+ln_1}=(f_1^kf_2^l)^p$ and $f_2=f_2^{kp+ln_1}=(f_1^kf_2^l)^{n_1}$. Replacing t by the local parameter $f_1^kf_2^lt$ we find that $X=t^p$ and $Y=t^{n_1}$. Consequently $dX/dt=pt^{p-1}$ and dX/dt has order $(p-1)$ at P. It follows that ρ is ramified at P. Analogous

arguments show that the points of C lying above $a_2, \ldots, a_r$ are also ramification points of ρ. It is easy to see that these points are the *only* ramification points. From the Hurwitz formula it follows that $g=(r-2)\dfrac{(p-1)}{2}$. □

Proposition 1.3 (char $(K)=p$ case). *Assume that $F=K(X, Y)$ where Y satisfies the equation*

$$Y^p-Y=\sum_{i=1}^{r}\left(\sum_{\substack{j=1\\ j\not\equiv 0 \bmod (p)}}^{n_i}\frac{\alpha_{ij}}{(X-a_i)^j}\right)$$

with $n_i\not\equiv 0 \bmod (p)$.

Then ρ is ramified in the points of C lying above $a_1, \ldots, a_r$ and

$$g=\left(\sum_{i=1}^{r}(n_i+1)-2\right)\frac{(p-1)}{2}$$

Proof. We proceed in the same way as for char $(K)\neq p$. We may assume that $a_1=0$. Let P be a point of C lying above a_1. Let $U_\varepsilon=\{x\in\mathbb{P}^1(K)\,|\,|x|\leqq\varepsilon\}$ and let $V_\varepsilon=\rho^{-1}(U_\varepsilon)$. The rings of analytic functions on U_ε and V_ε are $\mathcal{O}(U_\varepsilon)=\{f\in K[\![X]\!]\,|\,f \text{ convergent on } U_\varepsilon\}$ and $V_\varepsilon=A_{normalised}$ with

$$A=\mathcal{O}(U_\varepsilon)[Y]\Big/\left(Y^p-Y-\sum_{i=1}^{r}\left(\sum_{j=1}^{n_i}\left(\frac{\alpha_{ij}}{(X-a_i)^j}\right)\right)\right).$$

Each of the terms $\dfrac{\alpha_{ij}}{(X-a_i)^j}$ can be written in the form $\dfrac{\alpha_{ij}}{(X-a_i)^j}=\lambda(1+f)$ with $\lambda\in K$ and $f=\sum_{k=1}^{j}\binom{j}{k}\dfrac{(-X)^k}{(X-a_i)^k}$; $(i=2, \ldots, r)$. If ε is small enough, then $|\lambda f(x)|<1$ for all $x\in U_\varepsilon$. Hence there exists a $g\in\mathcal{O}(U_\varepsilon)$ such that $\lambda f=g^p-g$. Let $\mu\in K$ such that $\lambda=\mu^p-\mu$. Replacing Y by $Y+\mu+g$ we find a new equation for Y in which the $(X-a_i)^j$-term does not occur. Repeating this argument for all $j=1, \ldots, n_i$ and $i=2, \ldots, r$ we find that $A\cong\mathcal{O}(U_\varepsilon)[Y]\Big/\left(Y^p-Y-\sum_{j=1}^{n_1}\dfrac{\alpha_{1j}}{X^j}\right)$. If X has order a and if Y has order b at P we find that $bp+an_1=0$ and hence $a=p$ and $b=-n_1$. Let t be a local parameter at P. We have $X=\sum_{k=0}^{\infty}\alpha_k t^{k+p}$ and $Y=\sum_{l=0}^{\infty}\beta_l t^{l-n_1}$ with $\alpha_k, \beta_l\in K$ and $\alpha_0, \beta_0\in K^*$. It follows that

$$\frac{dY}{dt}=\sum_{l=0}^{\infty}(l-n_1)t^{l-n_1-1}=\sum_{\substack{n=1\\ n\not\equiv 0 \bmod (p)}}^{n_1}\frac{n\alpha_{1n}}{X^{(n+1)}}\frac{dX}{dt},$$

and consequently dX/dt has order $(n_1+1)(p-1)$ at P. So ρ is ramified at P.

Analogous arguments show that ρ is also ramified in the points of C lying above $a_2, \ldots, a_r$. It follows from the Hurwitz formula that

$$g=\left(\sum_{i=1}^{r}(n_i+1)-2\right)\frac{(p-1)}{2}.$$ □

§ 2. Schottky Groups and p-Cyclic Coverings of $\mathbb{P}^1(K)$

Let Γ be a Schottky group of rank g over K; i.e. Γ is a free discontinuous subgroup of $\mathrm{PGL}(2, K)$ of rank g. Let $\Omega = \{x \in \mathbb{P}^1(K) \mid x \text{ is ordinary point for } \Gamma\}$ and let $\mathscr{L} = \mathbb{P}^1(K) - \Omega = \{x \in \mathbb{P}^1(K) \mid x \text{ is a limitpoint for } \Gamma\}$. We assume that $\infty \notin \mathscr{L}$. The group Γ acts freely on Ω and the analytic space $X_\Gamma = \Omega/\Gamma$ is analytically isomorphic to a nonsingular projective curve of genus g. We identify X_Γ with this curve.

Let $\pi: \Omega \to \Omega/\Gamma$ be the quotient map. For each automorphism σ of $X_\Gamma = \Omega/\Gamma$ there exists an element s in the normaliser $N(\Gamma)$ of Γ in $\mathrm{PGL}(2, K)$ such that $\pi \circ s(x) = \sigma \circ \pi(x)$ for all $x \in \Omega$, in such manner that σ is the identity map on X_Γ if and only if $s \in \Gamma$. Hence $\mathrm{Aut}(X_\Gamma) \cong N(\Gamma)/\Gamma$; (see [2]).

Proposition 2.1. *The curve X_Γ admits a p-cyclic covering of $\mathbb{P}^1(K)$ if and only if there exists a $s_0 \in N(\Gamma)$ such that $s_0^p = id$ and such that $\prod_{k=0}^{p-1} s_0^k \gamma s_0^{-k} \equiv 1 \bmod [\Gamma, \Gamma]$ for all $\gamma \in \Gamma$.*

Proof. a) Assume that the element $s_0 \in N(\Gamma)$ exists. Let Γ_0 be the group generated by Γ and s_0. Since $s_0 \in N(\Gamma)$, Γ_0 is a discontinuous group and Γ_0 has the same set of limit points as Γ. It follows that the quotient space $X_{\Gamma_0} = \Omega/\Gamma_0$ is analytically isomorphic to a nonsingular projective curve. The genus of this curve is equal to the free rank of the group $\Gamma_0/[\Gamma_0, \Gamma_0]$; (see [2]). Let σ be the automorphism of X_Γ induced by s_0 and let $\rho: \Omega/\Gamma \to \Omega/\Gamma_0$ be the canonical morphism defined by $\rho(x \bmod (\Gamma)) = x \bmod (\Gamma_0)$.

It is clear that ρ is a Galois covering of degree p of X_{Γ_0}. The Galois automorphism σ^* of the function field F of X_Γ is induced by σ. We prove that X_{Γ_0} is isomorphic to $\mathbb{P}^1(K)$.

For each $\gamma \in \Gamma_0$ there exists a $w \in \Gamma$ and $n \in \mathbb{N}$ such that $\gamma = w s_0^n$. So $\prod_{k=0}^{p-1} s_0^k \gamma s_0^{-k} \equiv \prod_{k=0}^{p-1} s_0^k w s_0^{-k} s_0^{pn} \bmod [\Gamma_0, \Gamma_0]$. Since $\prod_{k=0}^{p-1} s_0^k w s_0^{-k} \in [\Gamma, \Gamma] \subset [\Gamma_0, \Gamma_0]$ and since $s_0^0 = id$ we have $\gamma^p \equiv \prod_{s=0}^{p-1} s_0^k \gamma s_0^{-k} \equiv 1 \bmod [\Gamma_0, \Gamma_0]$. It follows that each element of $\Gamma_0/[\Gamma_0, \Gamma_0]$ has order p and consequently the rank of $\Gamma_0/[\Gamma_0, \Gamma_0]$ is 0. It follows that the genus of X_{Γ_0} is 0 and $X_{\Gamma_0} \cong \mathbb{P}^1(K)$.

b) Assume that X_Γ admits a p-cyclic covering of $\mathbb{P}^1(K)$. There exists an automorphism σ of X_Γ of order p such that $X_\Gamma/\langle\sigma\rangle \cong \mathbb{P}^1(K)$. Let $s \in N(\Gamma)$ such that $\pi \circ s = \sigma \circ \pi$. From 1.2 and 1.3 it follows that the p-cyclic covering $\rho: X_\Gamma \to X_\Gamma/\langle\sigma\rangle$ is ramified in at least one point P of X_Γ. Hence $\sigma(P) = P$ and so there exists a $\gamma \in \Gamma$ and a point $x \in \Omega$ such that $\pi(x) = P$ and such that $s(x) = \gamma(x)$. Hence $(\gamma^{-1} s)(x) = (\gamma^{-1} s)^2(x) = \ldots = (\gamma^{-1} s)^p(x)$. We have

$$(\gamma^{-1} s)^p = \gamma^{-1}(s\gamma^{-1} s^{-1})(s^2 \gamma^{-1} s^{-2}) \ldots (s^{p-1} \gamma^{-1} s^{-(p-1)}) \in \Gamma$$

and $\lim_{n \to \infty} (\gamma^{-1} s)^{np}(x) = x$. Since $x \in \Omega$ we find that $(\gamma^{-1} s)^p = id$. Let $s_0 = \gamma^{-1} s$. So $s_0^p = id$ and s_0 also induces the automorphism σ on X_Γ. Let Γ_0 be the group generated by Γ and s_0. This group is discontinuous and has the same set of

limitpoints as Γ. The quotient space $X_{\Gamma_0}=\Omega/\Gamma_0$ is isomorphic to $X_\Gamma/\langle\sigma\rangle\cong\mathbb{P}^1(K)$. Hence the rank of $\Gamma_0/[\Gamma_0,\Gamma_0]$ is 0.

Each $\gamma\in\Gamma_0$ can be written in the form $\gamma=ws_0^m$ with $w\in\Gamma$ and $0\leqq m<p$. Therefore each commutator $\delta\in[\Gamma_0,\Gamma_0]$ has a decomposition of the form $\delta=as_0^nbs_0^ms_0^{-n}a^{-1}s_0^{-m}b^{-1}$ with $a,b\in\Gamma$ and $0\leqq n,\ m<p$. It follows that $\delta=a(s_0^nbs_0^{-n})(s_0^ma^{-1}s_0^{-m})b^{-1}\in\Gamma$ and consequently $[\Gamma_0,\Gamma_0]\subset\Gamma$. It is not difficult to see that

$$\prod_{k=0}^{p-1}s_0^k\delta s_0^{-k}=\prod_{k=0}^{p-1}s_0^kas_0^{-k}\cdot s_0^k(s_0^nbs_0^{-n})s_0^{-k}s_0^k(s_0^ma^{-1}s_0^{-m})s_0^{-k}s_0^kb^{-1}s_0^{-k}$$
$$\equiv 1 \bmod [\Gamma,\Gamma].$$

Because the rank of $\Gamma_0/[\Gamma_0,\Gamma_0]$ is 0, for each $\gamma\in\Gamma$ there exists a $n\in\mathbb{N}$ such that $\gamma^n\in[\Gamma_0,\Gamma_0]$ and hence $\left(\prod_{k=0}^{p-1}s_0^k\gamma s_0^{-k}\right)^n\equiv 1 \bmod [\Gamma,\Gamma]$. Since $\Gamma/[\Gamma,\Gamma]$ is a free abelian group we have $\prod_{k=0}^{p-1}s_0^k\gamma s_0^{-k}\equiv 1 \bmod [\Gamma,\Gamma]$.

From now on we assume that X_Γ admits a p-cyclic covering of $\mathbb{P}^1(K)$. We fix $s_0\in N(\Gamma)$ such that s_0 satisfies the conditions of the previous proposition. The automorphism induced by s_0 is denoted by σ and the curve X_{Γ_0} is identified with $\mathbb{P}^1(K)$, so the p-cyclic covering $\rho\colon X_\Gamma\to\mathbb{P}^1(K)$ is just the canonical morphism $X_\Gamma\to X_{\Gamma_0}$.

It follows from a theorem on automorphisms of free groups of prime degree by Dyer and Scott that Γ has a free basis B of the following form:

$$B=\{x_{1,1},\ldots,x_{1,p-1},x_{2,1},\ldots,x_{2,p-1},\ldots,x_{r,1},\ldots,x_{r,p-1}\}$$

with $s_0x_{i,j}s_0^{-1}=x_{i,j+1}$ for each $j=1,\ldots,p-2$ and $i=1,\ldots,r$ and with $s_0x_{i,p-1}s_0^{-1}=(x_{i,1},\ldots,x_{i,p-1})^{-1}$ for each $i=1,\ldots,r$; (see [1]). These generators satisfy the condition $\prod_{k=0}^{p-1}s_0^kx_{i,j}s_0^{-k}\equiv 1 \bmod [\Gamma,\Gamma]$. For each $j=1,\ldots,r$ the element $s_j=x_{j,1}s_0$ has order p and Γ_0 is generated by $s_0,\ldots,s_r$. It is clear that $s_0^p=\ldots=s_r^p=id$ are the only defining relations among these generators. Furthermore, the elements $z_{jk}=s_j^ks_0^{-k}$ also form a free basis for Γ. We can summarize as follows.

Proposition 2.2. *There exist $s_0,\ldots,s_r\in\Gamma_0$ such that $s_i^p=id$ for each $i=0,\ldots,r$ and such that Γ_0 is the free product of the groups $S_0,\ldots,S_r$ where each S_i is generated by s_i. The set $\{s_j^ks_0^{-k}\mid j=1,\ldots,r$ and $k=1,\ldots,p-1\}$ is a set of free generators for Γ.*

Let $u_{j,k}$ be the residue of $s_j^ks_0^{-k}$ in $\Gamma/[\Gamma,\Gamma]$. The automorphism of $\Gamma/[\Gamma,\Gamma]$, induced by s_0, has a matrix G of the following type with respect to the set generators $\{u_{j,k}\mid j=1,\ldots,r,k=1,\ldots,p-1\}$:

$$G=\begin{pmatrix}\boxed{G_p} & & \\ & \boxed{G_p} & 0 \\ & \ddots & \\ 0 & & \boxed{G_p}\end{pmatrix}\in\mathrm{GL}(g,\mathbb{Z})$$

with

$$G_p = \begin{pmatrix} -1 & -1 & \cdots & -1 \\ 1 & 0 & & 0 \\ & \ddots & \ddots & \\ 0 & & 1 & 0 \end{pmatrix} \in \mathrm{GL}(p-1, \mathbb{Z}).$$

The matrix G_p is known as the companion matrix of the polynomial $X^{p-1}+X^{p-2}+\ldots+1$.

Remark. It follows from 1.2. and 1.3. that $(p-1)|2g$ for each p-cyclic covering of $\mathbb{P}^1(K)$ of genus g. However, for curves which can be parameterised by Schottky groups we see that $(p-1)|g$. This is not true for general coverings; e.g. if $\mathrm{char}(K)\neq p$, then curves with equations $Y^p=(X-a_1)^{n_1}(X-a_2)^{n_2}(X-a_3)^{n_3}$ with $0<n_1, n_2, n_3<p$ and $n_1+n_2+n_3=p$ admit p-cyclic coverings of $\mathbb{P}^1(K)$, but their genus is $\left(\frac{p-1}{2}\right)$. Hence, these curves cannot be parameterised by Schottky groups.

§ 3. The Equation of a p-Cyclic Covering of $\mathbb{P}^1(K)$

We keep the notations of the previous paragraph. So Γ_0 is generated by $s_0, \ldots, s_r$ such that $s_0^p=\ldots=s_r^p=id$ are the only defining relations among these generators. In [4] we proved that the fixpoints of s_i are in Ω for each $i=0, \ldots, r$. If z is a fixpoint of s_i, then $\sigma(\pi(z))=\pi(z)$. It follows that $\pi(z)$ is a fixpoint of σ and hence a ramification point for the p-cyclic covering ρ. Conversely, if $P\in X_\Gamma$ is a fixpoint of Γ, then there exists some $\gamma\in\Gamma$ and $z\in\Omega$ such that $\pi(z)=P$ and $s_0(z)=\gamma(z)$. It follows that $\gamma^{-1}s_0(z)=z$. Exactly as in the proof of 2.1. we find that $(\gamma^{-1}s_0)^p=id$. It follows that $\gamma^{-1}s_0$ is conjugate to some s_i^k by an element $w\in\Gamma$; $(0<k<p)$. So $w(z)$ is a fixpoint of s_i and $\pi(w(z))=P$. It is easy to see that the fixpoints of $s_0, \ldots, s_r$ determine different points on X_Γ. If $\mathrm{char}(K)\neq p$, then each s_i is elliptic and has two fixpoints. So σ has $2r+2$ fixpoints. In fact, we knew this already from 1.2.

If $\mathrm{char}(K)=p$, then each s_i is parabolic and has one fixpoint. So σ has $r+1$ fixpoints $P_0, \ldots, P_r$. If we assume $\rho(P_i)\neq\infty$ for each $i=0, \ldots, r$ then X_Γ has an equation of the form

$$Y^p-Y=\sum_{i=0}^{r}\left(\sum_{\substack{j=1 \\ j\not\equiv 0 \bmod(p)}}^{n_i}\frac{\alpha_{ij}}{(X-\rho(P_i))^j}\right)$$

with $n_i\not\equiv 0 \bmod(p)$. From 1.4. we have

$$g=r(p-1)=\left(\sum_{i=0}^{r}(n_i+1)-2\right)\left(\frac{p-1}{2}\right)$$

and consequently $n_i=1$ for all $i=0, \ldots, r$.

We summarise as follows.

Proposition 3.1. *Let $s_0, \ldots, s_r$ be generators for Γ_0 such that $s_0^p = \ldots = s_r^p = id$ are the only defining relations. Assume that for each fixpoint x of s_i, $\rho \circ \pi(x) \neq \infty$; $(i = 0, \ldots, r)$*

a) *If $\mathrm{char}(K) \neq p$ then each s_i has two fixpoints x_i and y_i and X_Γ has an equation of the form $Y^p = \prod_{i=0}^{r} (X - \rho \circ \pi(x_i))^{n_i} \cdot (X - \rho \circ \pi(y_i))^{m_i}$ with $0 < n_i,\ m_i < p$ for each $i = 0, \ldots, r$ and such that $\sum_{i=0}^{r} (n_i + m_i) \equiv 0 \bmod (p)$.*

b) *If $\mathrm{char}(K) = p$ then each s_i has one fixpoint x_i and X_Γ has an equation of the form*

$$Y^p - Y = \sum_{i=0}^{r} \frac{\alpha_i}{(X - \rho \circ \pi(x_i))}$$

with $\alpha_i \in K^*$ for each $i = 0, \ldots, r$.

If $\mathrm{char}(K) \neq p$ then it is possible to calculate the exponents $n_0, m_0, \ldots, n_r, m_r$ in the equation of X_Γ.

We identify the function field of $\mathbb{P}^1(K)$ with $K(T)$ such that $\rho^*(T) = X$. Let F be the function field of X_Γ and let $Y \in F$ such that Y satisfies the equation of 3.1. Since $\sigma^*(X) = X$ we have $\sigma^*(Y^p) = Y^p$ and hence $\sigma^*(Y) = \zeta Y$ for some primitive p-th root of unity ζ.

Since s_i has order p there exists a $w_i \in \mathrm{PGL}(2, K)$ such that $s_i = w_i t_i w_i^{-1}$ where $t_i \in \mathrm{PGL}(2, K)$ is represented by a matrix $\begin{pmatrix} \zeta^{v_i} & 0 \\ 0 & 1 \end{pmatrix}$ with $0 < v_i < p$. The fixpoints of t_i are 0 and ∞ and we may assume that $w_i(x_i) = 0$ and $w_i(y_i) = \infty$.

Proposition 3.2. *Y satisfies the equation $Y^p = \prod_{i=0}^{r} (X - \rho \circ \pi(x_i))^{n_i} (X - \rho \circ \pi(y_i))^{m_i}$ with $0 < n_i,\ m_i < p$ for each $i = 0, \ldots, r$ and such that $v_i \cdot n_i \equiv 1 \bmod (p)$ and $(p - v_i) \cdot m_i \equiv 1 \bmod (p)$.*

Proof. We prove the proposition in the case that $i = 0$. After a suitable transformation of T, X and Y we may assume that $\rho^*(T) = X$, $F = K(X, Y)$ and $Y^p = X^{n_0} \cdot \prod_{i=1}^{r} (X - a_i)^{n_i} (X - b_i)^{m_i}$ with $a_i, b_i \in K^*$, (same n_i, m_i as in 3.1.).

For this $Y \in F$ we still have $\sigma^*(Y) = \zeta Y$.

We assume that the fixpoints x_i and y_i of s_i are 0 and ∞ respectively; i.e. s_i is represented by the matrix $S_i = \begin{pmatrix} \zeta^{v_i} & 0 \\ 0 & 1 \end{pmatrix}$.

For each $z \in \Omega$ we put

$$\Theta(z) = \prod_{\gamma \in \Gamma} \frac{z - \gamma(0)}{z - \gamma(\infty)} \quad \text{and} \quad \Theta_0(z) = \prod_{\gamma \in \Gamma_0} \frac{z - \gamma(0)}{z - \gamma(\infty)},$$

where $\dfrac{z - \gamma(0)}{z - \gamma(\infty)} = z$ if $\gamma = id$.

It was proved in [3] that Θ and Θ_0 define meromorphic functions on Ω. Furthermore, Θ_0 and $X \circ \pi$ have the same sets of zeroes and poles. Hence

$\Theta_0(z)=\lambda X\circ\pi(z)$ for each $z\in\Omega$ with $\lambda\in K^*$. We may assume $\lambda=1$. Since $s_0\in N(\Gamma)$ we have:

$$\Theta_0(z)=\prod_{\gamma\in\Gamma_0}\frac{z-\gamma(0)}{z-\gamma(\infty)}=\prod_{k=0}^{p-1}\left(\prod_{\gamma\in\Gamma}\frac{z-\gamma(s_0^k(0))}{z-\gamma(s_0^k(\infty))}\right)$$
$$=\prod_{k=0}^{p-1}\left(\prod_{\gamma\in\Gamma}\frac{z-\gamma(0)}{z-\gamma(\infty)}\right)=\Theta^p(z).$$

It follows that $(Y\circ\pi)^p=\Theta^{pn_0}\cdot\prod_{i=1}^{r}(\Theta^p-a_i)^{n_i}(\Theta^p-b_i)^{m_i}$. And hence $Y\circ\pi=\Theta^{n_0}\cdot F$, where F is a meromorphic function on Ω such that

$$F^p=\prod_{i=1}^{r}(\Theta^p-a_i)^{n_i}(\Theta^p-b_i)^{m_i}.$$

Since $F^p(s_0(z))=F^p(z)$ for each $z\in\Omega$ we have $F(s_0(z))=\zeta^kF(z)$ for some $k\in\mathbb{Z}$. But since $\zeta^kF(0)=F(s_0(0))=F(0)$ and, since $F(0)\neq 0$, we have $\zeta^k=1$. Furthermore, for each $z\in\Omega$:

$$\Theta(s_0(z))=(\zeta^{v_0}z)\cdot\prod_{\substack{\gamma\in\Gamma\\ \gamma\neq id}}\left(\frac{\zeta^{v_0}z-\gamma(0)}{\zeta^{v_0}z-\gamma(\infty)}\right)=\zeta^{v_0}z\cdot\prod_{\substack{\gamma\in\Gamma\\ \gamma\neq id}}\left[\frac{z-\zeta^{-v_0}\gamma(0)}{z-\zeta^{-v_0}\gamma(\infty)}\right]$$
$$=\zeta^{v_0}\left[z\cdot\prod_{\substack{\gamma\in\Gamma\\ \gamma\neq id}}\frac{z-s_0^{-1}\gamma s_0(0)}{z-s_0^{-1}\gamma s_0(\infty)}\right]=\zeta^{v_0}\Theta(z).$$

It follows that for each $z\in\Omega$:

$$\zeta Y\circ\pi(z)=Y\circ\pi(s_0(z))=\Theta^{n_0}(s_0(z))\cdot F(s_0(z))$$
$$=\zeta^{v_0n_0}\cdot\Theta(z)\cdot F(z)=\zeta^{v_0n_0}Y\circ\pi(z).$$

Consequently $\zeta=\zeta^{v_0n_0}$ and hence $v_0n_0\equiv 1 \bmod(p)$.

In an analogous way we find $(p-v_0)m_0\equiv 1 \bmod(p)$. □

Bibliography

1. Dyer, J.L., Scott, P.: Periodic automorphisms of free groups. Comm. Algebra 3, 195–201 (1975)
2. Gerritzen, L.: On automorphism group of p-adic Schottky curves. In: Journées d'analyse ultramétrique (Y. Amice, D. Barsky, P. Robba, eds.) 1976/1977. Paris: Secrétariat Mathématique 1977
3. Gerritzen, L., van der Put, M.: Schottky groups and Mumford curves. Lecture Notes in Mathematics 817. Berlin-Heidelberg-New York: Springer 1981
4. Van Steen, G.: Hyperelliptic curves defined by Schottky groups over a non-Archimedean valued field. Thesis, University of Antwerpen 1981
5. Van Steen, G.: Non-archimedean Schottky groups and hyperelliptic curves (to appear)

Received December 16, 1981

Math. Z. 180, 225-234 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Positivity Improving Operators and Hypercontractivity

Christer Borell

Department Mathematics, Chalmers University of Technology and the University of Göteborg, S-412 96 Göteborg, Sweden

1. Introduction

The purpose of this paper is to broaden the meaning of so-called hypercontractivity (Sect. 2) and to relate the definition to certain spectral properties of self-adjoint operators.

To begin with, consider the number operator $N=-d^2/dx^2+xd/dx$ in $L_2([\exp(-x^2/2)]\,dx/(2\pi)^{1/2})$. Then, for each $t>0$,

$$\|e^{-tN}f\|_{1+e^{2t}(p-1)} \leqq \|f\|_p, \quad f\in L_2,\ 1<p<+\infty, \tag{1.1}$$

and

$$\|e^{-tN}f\|_{1+e^{2t}(p-1)} \geqq \|f\|_p, \quad f\in L_2^+,\ -\infty<p<1. \tag{1.2}$$

Here (1.1) is the famous Nelson hypercontractivity inequality [14], which has proved to be extremely useful in Euclidean quantum field theory (see e.g. [16]). Like (1.1) the estimate (1.2) follows very easily from the Itô calculus of Brownian motion [3]. Here, in order to make the paper self-contained and to introduce the reader to some general principles of interest, we submit a new proof of (1.2) by mimicing a standard proof of (1.1). We should like to add that (1.2) has at least a formal resemblance to the familiar correlation inequalities from statistical mechanics (see e.g. [11, 12, 16]). Using (1.1) and (1.2) it is simple to obtain similar estimates for the biquantization of a linear contraction on Hilbert space (Sect. 4).

Now suppose A is a self-adjoint operator *on* L_2 of a probability space and set $\Delta(A)=\max\operatorname{spec}(A)-\sup(\operatorname{spec}(A)\setminus\{\max\operatorname{spec}(A)\})$, where $\operatorname{spec}(A)$ denotes the spectrum of A. Below we will state conditions that guarantee that the gap $\Delta(A)$ is proper. In fact, a simplified version of our main result (Theorem 5.3) gives a positive numerical lower bound of $\Delta(A)$ if

(i) A is positivity preserving,

(ii) $\sup\{\|Af\|_q;\ \|f\|_2=1\}=K<+\infty$ some $q>2$, and

(iii) $\inf\{\|Af\|_{p/(p-1)};\ \|f\|_p=1, f\geqq 0\}=k>0$ some $0<p<1$.

Here (i)–(iii) imply that A has a unique ground state. It shall be empha-

0025-5874/82/0180/0225/$02.00

sized that the presence of a proper gap $\Delta(A)>0$, in general, does not force hypercontractivity. If A is positivity preserving and $\Delta(A)>0$, then any of the statements (ii) $\wedge$ ($\neg$(iii)), ($\neg$(ii)) $\wedge$ (iii), and ($\neg$(ii)) $\wedge$ ($\neg$(iii)) may, in fact, be true.

Finally, in Sect. 6, we show that the results obtained so far apply to the spatially cut-off Hamiltonian in the so-called $P(\phi)_2$ model.

2. Hypercontractivity

Throughout, (S, Σ, μ) denotes a fixed probability space and L_2 stands for the real space $L_2(S, \Sigma, \mu)$ if there is no risk of confusion. We recall that a linear operator A on L_2 is said to be positivity improving (preserving) if $Af>0$ ($Af \geqq 0$) for each $f \in L_2^+ \setminus \{0\}$.

Assume $A \neq 0$ is a positivity preserving, self-adjoint, linear operator *on* L_2 and let $p, q \in \mathbb{R}$, $K>0$, be fixed. Then we shall write $A \in \mathcal{O}(p, q, K)$ (or $A \in \mathcal{O}(p, q, K; \mu)$) if either

$$1<p<q \quad \textit{and} \quad \|Af\|_q \leqq K\|f\|_p, \quad f \in L_2^+,$$

or

$$q<p<1 \quad \textit{and} \quad \|Af\|_q \geqq K\|f\|_p, \quad f \in L_2^+.$$

Suppose $A \in \mathcal{O}(p, q, K)$. Then, if $p>1$, it is consistent with the literature to call A hypercontractive. Therefore, if $p<1$, A is also said to be hypercontractive since there can be no confusion.

To begin with, we prove some very simple properties of hypercontractive operators.

Theorem 2.1. *If* $A \in \mathcal{O}(p, q, K)$ *and* $0<p<1$, $q \leqq 0$, *then* A *is positivity improving.*

Proof. Suppose $g \in L_2^+$ is such that $g=0$ on a set of positive measure. Then $\|g\|_q = 0$, $q \leqq 0$. Theorem 2.1 follows at once from this remark.

Theorem 2.2. *If* $A \in \mathcal{O}(p, q, K)$ *then* $A \in \mathcal{O}(q', p', K)$.

Proof. Set $q' = q/(q-1)$, $q \neq 1$, and let $g \in L_2^+$. Then by Hölder's inequality

$$\|g\|_q = \sup\{(g, h);\ \|h\|_{q'} = 1, h \in L_2^+\}, \quad q>1,$$

and

$$\|g\|_q = \inf\{(g, h);\ \|h\|_{q'} = 1, h \in L_2^+\}, \quad q<1. \tag{2.1}$$

Clearly, Theorem 2.2 is an immediate consequence of these representations.

The next important message is due to Bonami [2] and Segal [15].

Theorem 2.3. *If* $A \in \mathcal{O}(p, q, K_A; \mu_A)$ *and* $B \in \mathcal{O}(p, q, K_B; \mu_B)$, *then* $A \otimes B \in \mathcal{O}(p, q, K_A K_B; \mu_A \otimes \mu_B)$.

Actually, Bonami and Segal only treat the case $p>1$. However, from (2.1), we have the well-known Minkowski inequality

$$\|f+g\|_q \geqq \|f\|_q + \|g\|_q, \quad f, g \in L_2^+, \ q<1,$$

which is sufficient to prove the remaining part of Theorem 2.3. We omit the simple details here.

Theorem 2.3, which we will refer to as the Bonami-Segal lemma, is a powerful tool for constructing interesting hypercontractive operators.

3. The Bonami Hypercontractivity Theorem and its Complement

Let m be Lebesgue measure on $[0,1]$ and let $(r_n)_{n\in\mathbb{N}}$ be the Rademacher functions

$$r_n(x)=\operatorname{sgn}\sin(2^{n+1}\pi x), \quad 0\leqq x\leqq 1.$$

We define a self-adjoint operator S in $L_2(m)$ by linearity and the rules

$$S1=0,$$

$$S(r_{n_1}\cdot\ldots\cdot r_{n_k})=k\,r_{n_1}\cdot\ldots\cdot r_{n_k}, \quad n_1<\ldots<n_k.$$

Theorem 3.1. *For any* $t>0$,

$$e^{-tS}\in\mathcal{C}(p, 1+e^{2t}(p-1), 1; m), \quad p\neq 1.$$

The Rademacher functions are known to be stochastically independent based on the Lebesgue probability space $([0,1], \mathcal{B}, m)$. In view of the Bonami-Segal lemma, Theorem 3.1 may thus be restated as follows.

Theorem 3.2. *Let* $\beta=(\delta_{-1}+\delta_{+1})/2$ *and let* Q *denote the orthogonal projection of* $L_2(\beta)$ *onto* $\operatorname{span}\{x\curvearrowright x\}$. *If* $t>0$, *then*

$$e^{-tQ}\in\mathcal{C}(p, 1+e^{2t}(p-1), 1; \beta), \quad p\neq 1. \tag{3.1}$$

Theorems 3.1 and 3.2, $p>1$, are due to Bonami [2].

Proof of Theorem 3.2, $p<1$. The following line of reasoning is very close to that of Beckner [1].

Set $q=1+e^{2t}(p-1)$. Then $p'=1+e^{2t}(q'-1)$. Therefore, if (3.1) is known for $0\leqq q<p<1$, Theorem 2.2 implies (3.1) for $p\leqq 0$. On the other hand, if both these cases are known, we may apply the semi-group property to conclude that (3.1) is true for all $p<1$. Therefore, if suffices to prove the case $0<q<p<1$. To this end, let $f(x)=1+ax$, $x\in\mathbb{R}$, where $-1<a<1$. Then

$$\|e^{-tQ}f\|_q^q=1+\sum_{n=1}^{\infty}\binom{q}{2n}e^{-2nt}a^{2n}$$

and, by convexity,

$$\|e^{-tQ}\|_q^p\geqq 1+\sum_{n=1}^{\infty}\frac{p}{q}\binom{q}{2n}e^{-2nt}a^{2n}.$$

Thus, it only remains to establish the estimate

$$\frac{p}{q}\binom{q}{2n}e^{-2nt}\geqq\binom{p}{2n}$$

which follows from the simple inequality

$$(i-q)(1-p)^{1/2} \leqq (i-p)(1-q)^{1/2}, \quad i \geqq 2.$$

This concludes the proof of Theorem 3.2, $p<1$.

Corollary 3.1. *For any $t>0$, e^{-tS} is positivity improving.*

4. Hypercontractivity in Q-space

Let $(Q, \mathscr{B}, \gamma)$ be a centred Gaussian Radon space with reproducing kernel Hilbert space $(\mathscr{K}, \| \ \|_\gamma)$ (see e.g. [4]). Here the Q-space is assumed to be a real, locally convex Hausdorff space and, hence, $\mathscr{K}$ is a real Hilbert space. Two elements $f, g \in L_2$ are identified if $f=g$ a.s. $[\gamma]$. For any $a \in \mathscr{K}$, $\phi(a)$ denotes the unique γ-measurable linear extension of $(\cdot, a)_\gamma$ to Q.

Suppose $f \in L_2$ is fixed. Then there exists a sequence $(a_n)_{n\in\mathbb{N}}$ in $\mathscr{K}$ and a Borel measurable function $f_0: \mathbb{R}^{\mathbb{N}} \to \mathbb{R}$ such that

$$f = f_0(\{\phi(a_n)\}_{n\in\mathbb{N}}). \tag{4.1}$$

If $U: \mathscr{K} \to \mathscr{K}$ is a linear contraction and $f \in L_2$ is given by (4.1), we define

$$[\Gamma(U)f](x) = \int f_0(\{[\phi(Ua_n)](x) + [\phi((1-U^*U)^{1/2}a_n)](y)\}_{n\in\mathbb{N}})\, d\gamma(y). \tag{4.2}$$

In this way $\Gamma(U)$ becomes a well-defined positivity preserving contraction on L_2. It is readily seen that the above definition of the biquantization $\Gamma(U)$ of U is equivalent to any of the usual ones [16, §1.4].

Theorem 4.1. *Suppose $0<\|U\|<1$. Then*

$$\Gamma(U) \in \mathscr{C}(p, 1+\|U\|^{-2}(p-1), 1; \gamma), \quad p \neq 1.$$

Theorem 4.1, $(-1)^i p < (-1)^i$, is equivalent to (1.i), $i=1, 2$.

Proof of Theorem 4.1, $p<1$. If $q<1$, then

$$\|\Gamma(U)f\|_q \geqq \|f\|_q, \quad f \in L_2^+,$$

by the Jensen inequality. Since $\Gamma(U) = \Gamma(U/\|U\|)\Gamma(\|U\|)$ it thus suffices to consider the special case $U=\|U\|$. Set $\|U\| = e^{-t}$. Then either (4.2) or the Bonami-Segal lemma reduces the dimension of Q to one. Consequently, we may assume that $\Gamma(U) = \exp(-tN)$, where N is as in Sect. 1. By applying Theorem 3.1 and the central limit theorem, (1.2) follows at once. We omit the details, which do not depend on p ([1], [9, p. 1072]).

Below we need the following

Corollary 4.1. *Let F, G be linear subspaces of $\mathscr{K}$ and define*

$$\theta = \sup\{(a, b)_\gamma;\ \|a\|_\gamma \leqq 1, a \in F,\ \|b\|_\gamma \leqq 1, b \in G\}.$$

Suppose $0<p, q<1$ and $(1-p)(1-q) \geqq \theta^2$. Then

$$(f, g) \geqq \|f\|_p \|g\|_q, \quad f \in L_2^+(\gamma_{|\Sigma(F)}), \quad g \in L_2^+(\gamma_{|\Sigma(G)}),$$

where

$$\sigma(H)=\sigma\{\phi(c);c\in H\},\qquad H=F,G.$$

A similar correlation inequality, but in the opposite direction, is known to be a corollary to Theorem 4.1, $p>1$ [16, p. 103]. The point is that the conditional expectation $\mathbb{E}[\cdot|\Sigma(F)]$ equals $\Gamma(\mathrm{proj}_F)$. Corollary 4.1 follows at once from this remark.

For more general correlation inequalities in Q-space, the reader may consult [3].

By Simon [16, Theorem 1.16], $\Gamma(U)$ is positivity improving if $\|U\|<1$. This also follows from Theorems 2.1 and 4.1, $p<1$. Here we shall prove

Theorem 4.2. *Let $U\colon \mathscr{K}\to\mathscr{K}$ be a linear contraction. Then $\Gamma(U)$ is positivity improving if and only if*

$$G:=(U^*)^{-1}(\mathscr{R}((1-U^*U)^{1/2}))$$

is a dense linear subspace of $\mathscr{K}$.

Proof. Suppose first that G is not dense and choose $a\in G^{\perp}\setminus\{0\}$. Then $U^*a\neq 0$ and $U(1-U^*U)^{1/2}b\in G$, $b\in\mathscr{K}$. In particular, $(1-U^*U)^{1/2}U^*a=0$ and, hence, $UU^*a\neq 0$. However, it then follows that for any $f_0\in L_2(\gamma_{\phi(U^*a)})$,

$$\Gamma(U)f_0(\phi(U^*a))=f_0(\phi(UU^*a))$$

and, consequently, $\Gamma(U)$ cannot be positivity improving.

Conversely, assume G is dense and let $f\in L_2^+\setminus\{0\}$. We first choose $f_0\geqq 0$ such that (4.1) holds and set

$$M=\{x\in Q;[\Gamma(U)f](x)>0\}.$$

Actually, here $\Gamma(U)f$ is regarded as a point function defined by (4.2). Clearly, $\gamma(M)>0$ since $\|\Gamma(U)f\|_1=\|f\|_1$.

Suppose now that $x\in M$ and $a\in G$. By assumption there exists a $b\in\mathscr{K}$ such that

$$(1-U^*U)^{1/2}b=U^*a.$$

By applying the translation formula for Gauss measures it follows that

$$[\Gamma(U)f](x+a)$$
$$=\int f_0(\{[\phi(Ua_n)](x)+[\phi((1-U^*U)^{1/2}a_n)](y)\}_{n\in\mathbb{N}})\,e^{[\phi(b)](y)-\|b\|_{\mathscr{K}}^2/2}\,d\gamma(y)$$

and, hence, $x+a\in M$. Thus, $M+G\subseteq M$, and the Kallianpur lemma (see e.g. [3]) gives $\gamma(M)=1$.

This concludes the proof of Theorem 4.2.

5. On a Gap in the Spectrum below the Operator Norm

Suppose $A\in\mathscr{O}(2,q,K)$. Then, by a beautiful theorem of Gross [10, Theorem 1], $\|A\|$ is an eigenvalue of finite multiplicity not exceeding $(K/\|A\|)^{2q/(q-2)}$. Hence, if, in addition, A is positivity improving, then $\|A\|$ is a simple eigenvalue and

the corresponding normalized eigenfunction e_A may be chosen strictly positive [7].

Theorem 5.1. *Assume A is positivity improving and let $A\in\mathcal{O}(p,q,\|A\|)$, where $1<p\leqq 2$, $q>2$. Then*

$$\operatorname{spec}(A)\cap]({p'}^{-1}+q^{-1})\,\|A\|,\|A\|[=\phi.$$

Here $\operatorname{spec}(A)$ denotes the spectrum of A.

Theorem 5.1 is an abstraction from Guerra et al. [12, Theorem III.14] and Simon [17]. In spite of its elegance, Theorem 5.1 is probably of limited value in practice. In fact, the condition $A\in\mathcal{O}(p,q,K)$, with $K=\|A\|$, is, generally, extremely hard to verify without imposing some further restrictions. We do not prove Theorem 5.1 here since it follows as in [12].

The next theorem is an interesting complement to Theorem 5.1.

Theorem 5.2. *Suppose $A\in\mathcal{O}(2,q,K;\mu)$ is positivity improving and*

$$\|e_A\|_{(2-p)q/(q-p)}>0 \tag{5.1}$$

for some fixed $p\in]2,q[$. Let $0<\theta<1$ be fixed and assume

$$\operatorname{spec}(|A|)\cap]\theta\,\|A\|,\|A\|[=\phi.$$

Choose $n\in\mathbb{N}_+$ so large that

$$2(p-1)^{1/2}\,\theta^{n-1}K\,\|e_A\|_{\frac{(2-p)/p}{(2-p)q/(q-p)}}^{(2-p)/p}\leqq\|A\|.$$

Set $v=e_A^2\mu$ and $Bf=[A(e_A f)]/e_A$, $f\in L_2(v)$. Then $B^n\in\mathcal{O}(2,p,\|B^n\|;v)$.

The phenomenon we meet in Theorem 5.2 goes back to Glimm [6] in his study of the Hermite semi-group [13, p. 148]. Theorem 5.2 (with $e_A=1$) is (essentially) proved in [5, Lemma 3.3] and the proof, which is based on the Bonami hypercontractivity theorem and the Jensen inequality for conditional expectation, will not be repeated here. Note that (5.1) is satisfied for some $p\in]2,q[$ if $A\in\mathcal{O}(r,s,k;\mu)$ for some $s<0<r$.

With this background we are now ready to formulate the main contribution of this paper.

Theorem 5.3. *Suppose $A\in\mathcal{O}(2,q,K)$ and let k: $[0,+\infty]^2\to[0,+\infty[$ be a function, which decreases in each variable separately, and such that $k(s,t)>0$, $0\leqq s,t<+\infty$. Let $0<p<1$ be fixed and assume*

$$(Af,g)\geqq k\left(\frac{\|f\|_q}{\|f\|_p}\,\frac{\|g\|_q}{\|g\|_p}\right)\|f\|_p\|g\|_p,\qquad f,g\in L_2^+\setminus\{0\}. \tag{5.2}$$

Then A is positivity improving. If $0<\lambda\in(\operatorname{spec}(A)\setminus\{\|A\|\})$, then

$$\|A\|-\lambda\geqq 2a(\lambda)\,b^3(\lambda)\,c(\lambda)/(1+a^2(\lambda)\,b^2(\lambda))$$

where

$$a(\lambda)=k((K/\lambda)^{\alpha+1},(K/\lambda)^{\alpha+1}),\qquad b(\lambda)=(\lambda/K)^{\alpha},$$

and

$$c(\lambda)=\|A\|^{-1}(\|A\|/K)^{\alpha}\,k((K/\|A\|)^{\alpha+1},(K/\lambda)^{\alpha+1}),$$

$$\alpha=(q(2-p))/(p(q-2)).$$

Note that (5.2) is satisfied if $A \in \mathcal{O}(p, p', k)$, where k is a positive constant. The condition (5.2) is motivated by various forms of the Feynman-Kac formula. We discuss one example in Sect. 6.

The proof of Theorem 5.3 depends on the following lemma, which is based on the same assumptions as in Theorem 5.3.

Lemma 5.1. a) *A is positivity improving.*

b) *Set $e = e_A$. Then*

$$\|e\|_p \geqq (\|A\|/K)^{\alpha}$$

and

$$\|e\|_p \geqq (\|A\|/K)^{\alpha+1} \|e\|_q.$$

c) *If $0 < \lambda_0 < \lambda$, then there exists a unit vector $h \perp e$ such that*

$$(Ah, h) \geqq \lambda_0, \tag{5.3}$$

$$\|h^{(\pm)}\|_p \geqq b(\lambda_0) \|h^{(\pm)}\|_2 \tag{5.4}$$

and

$$\|h^{(\pm)}\|_p \geqq (\lambda_0/K)^{\alpha+1} \|h^{(\pm)}\|_q.$$

Proof. a) By (5.2), $Af > 0$ if $f \in L_\infty^+ \setminus \{0\}$. Since A is positivity preserving it follows at once that A is, in fact, positivity improving.

b) Suppose $g \in L_2$ and $v > 0$ are such that

$$v \|g\|_q \leqq K \|g\|_2.$$

We claim that

$$\|g\|_p \geqq (v/K)^{\alpha} \|g\|_2 \tag{5.5}$$

and

$$\|g\|_p \geqq (v/K)^{\alpha+1} \|g\|_q.$$

The first inequality obviously implies the second one. Since

$$\|g\|_2^2 \leqq \|g\|_p^{\theta p} \|g\|_q^{(1-\theta)q} \tag{5.6}$$

where $\theta = (q-2)/(q-p)$, (5.5) also follows in a straight forward manner.

Finally, by choosing $(g, v) = (e, \|A\|)$, we are done.

c) Let $f_n \perp e$, $\|f_n\|_2 = 1$, $n \in \mathbb{N}$, be such that

$$\lim_{n \to +\infty} \|Af_n - \lambda f_n\|_2 = 0. \tag{5.7}$$

We claim that

$$\varliminf_{n \to +\infty} \|(Af_n)^{(\pm)}\|_2 > 0. \tag{5.8}$$

To see this, we use (5.6) and have

$$\|Af_n\|_2 \leqq K^{(1-\theta)q} \|Af_n\|_p^{\theta p}.$$

Therefore, remembering (5.7),

$$\varliminf_{n \to +\infty} \|Af_n\|_p > 0.$$

By applying (5.2) and Part b) it now follows that

$$\underline{\lim}_{n\to+\infty}(e,|Af_n|)>0$$

and, moreover, as $(e,Af_n)=0$, we have

$$\underline{\lim}_{n\to+\infty}(e,(Af_n)^{(\pm)})>0$$

which proves (5.8).

Now set $g_n=\lambda f_n-Af_n$, $n\in\mathbb{N}$. Since A is positivity preserving, necessarily

$$\lambda(Af_n)^{(\pm)}\leqq A[(Af_n)^{(\pm)}]+A[|g_n|]$$

and hence

$$\lambda\,\|(Af_n)^{(\pm)}\|_q\leqq K(\|(Af_n)^{(\pm)}\|_2+\|g_n\|_2).$$

Suppose now that n_0 is so large that

$$(A^2f_{n_0},Af_{n_0})>\lambda_0\,\|Af_{n_0}\|_2^2$$

and

$$\lambda_0(\|(Af_{n_0})^{(\pm)}\|_2+\|g_{n_0}\|_2)\leqq\lambda\,\|(Af_{n_0})^{(\pm)}\|_2.$$

The existence of n_0 is obvious from (5.7) and (5.8). By setting $h=(Af_{n_0})/\|Af_{n_0}\|_2$, we have (5.3) and, also, $h\perp e$. Moreover, $\lambda_0\,\|h^{(\pm)}\|_q\leqq K\,\|h^{(\pm)}\|_2$ and the remaining part of Part c) follows from the proof of Part b).

Proof of Theorem 5.3. Throughout the proof, λ_0, e, and h have the same meaning as in Lemma 5.1. Then, since,

$$\|h^-\|_2^2+\|h^+\|_2^2=1 \tag{5.9}$$

the inequality (5.3) implies that

$$\|A\|-\lambda_0\geqq 2(Ah^+,h^-).$$

By applying (5.2) and Lemma 5.1, c), we thus have

$$\|A\|-\lambda_0\geqq 2a(\lambda_0)\,\|h^+\|_p\,\|h^-\|_p. \tag{5.10}$$

Moreover, as $h\perp e$, it follows that

$$\|h^{(\pm)}\|_2\geqq(e,h^{(\mp)}).$$

By applying (5.2) and Lemma 5.3 again, we obtain

$$\|h^{(\pm)}\|_2\geqq c(\lambda_0)\,\|h^{(\mp)}\|_p. \tag{5.11}$$

However, if $0<b,\ c\leqq 1$, then the optimization problem

$$\begin{aligned}&\min xy\\&x\geqq bu,\quad y\geqq bv\\&u\geqq cy,\quad v\geqq cx\\&u^2+v^2=1,\quad u,v,x,y\geqq 0\end{aligned}$$

has the optimal value $b^3c/(1+b^2c^2)$. In view of (5.4), (5.11), and (5.9), we thus have

$$\|h^+\|_p\|h^-\|_p \geqq b^3(\lambda_0)\,c(\lambda_0)/(1+b^2(\lambda_0)\,c^2(\lambda_0)).$$

From this and (5.10), Theorem 5.3 follows at once by letting $\lambda_0 \nearrow \lambda$.

6. The Spatially Cut-Off Hamiltonian in $P(\phi)_2$

The aim of this section is to show how the condition (5.2) enters in a concrete case. The following example is not the simplest one but it seems to be the most natural in this connection.

Let γ be a centred Gaussian Radon measure on real $\mathscr{D}'(\mathbb{R}^2)$ with

$$\int \langle u,a\rangle^2\, d\gamma(u) = 2\int \frac{|\hat{a}(\xi)|^2}{\xi^2+m^2}\,d\xi, \qquad a\in\mathscr{D}(\mathbb{R}^2),$$

where $m>0$ is a fixed constant. The reproducing kernel Hilbert space of γ is denoted by $\mathscr{K}=(\mathscr{K},\|\ \|_\gamma)$. We have the following situation:

$$\mathscr{D}(\mathbb{R}^2)\subseteq\mathscr{K}\subseteq\mathscr{D}'(\mathbb{R}^2).$$

Set $\mathscr{K}_{t_0}=\{a\in\mathscr{K};\operatorname{supp} a\subseteq\{t=t_0\}\}$. Then, by [16, Theorem III.12],

$$|(a,b)_\gamma| \leqq e^{-m|t|} \quad \text{if}$$

$$\|a\|_\gamma\leqq 1,\ a\in\mathscr{K}_0, \qquad \|b\|_\gamma\leqq 1,\ b\in\mathscr{K}_t. \tag{6.1}$$

Suppose the linear contraction $j_t\colon \mathscr{K}\to\mathscr{K}$ is given by $j_t(h\otimes\delta_0)=h\otimes\delta_t$ and $\mathscr{N}(j_t)=\mathscr{K}_0^\perp$. Set $\mathscr{F}_t=\Sigma(\mathscr{K}_t)$ and note that $I_t=\Gamma(j_t)$ maps $\mathscr{F}_0$-measurable L_2-functions into $\mathscr{F}_t$-measurable L_2-functions.

The spatially cut-off Hamiltonian H in the $P(\phi)_2$ model is of the form

$$e^{-tH}f = \mathbb{E}[(I_tf)\,e^{-V(t)}|\mathscr{F}_0], \qquad f\in L_2(\gamma_{|\mathscr{F}_0}),$$

where $V(t)$, $\exp(-V(t))\in L_r$, $1\leqq r<+\infty$.

Below we let $t>0$ and $2<q<1+\exp(2mt)$ be fixed. It is well known (and follows from Theorem 4.1, $p>1$) that

$$e^{-tH}\in\mathscr{O}(2,q,K;\gamma_{|\mathscr{F}_0})$$

for an appropriate K [16, Theorem V.10]. Now set $p=1-\exp(-mt)$ and suppose $f,\ g\in L_2^+(\gamma_{|\mathscr{F}_0})\setminus\{0\}$. We shall prove the estimate

$$(e^{-tH}f,g)\geqq\left\{\exp\left[-\|V(t)\|_{q/(q-2)}\frac{\|f\|_q}{\|f\|_p}\cdot\frac{\|g\|_q}{\|g\|_p}\right]\right\}\|f\|_p\|g\|_p. \tag{6.2}$$

To see this, first note that

$$(e^{-tH}f,g)\geqq\{\exp-[\textstyle\int V(t)(I_tf)\,g\,d\gamma/(I_tf,g)]\}(I_tf,g)$$

by the Jensen inequality. Moreover,

$$\int V(t)(I_t f)\, g\, d\gamma \leqq \|V(t)\|_{q/(q-2)} \|f\|_q \|g\|_q.$$

By applying (6.1) and Corollary 4.1 we also have

$$(I_t f, g) \geqq \|f\|_p \|g\|_p$$

and (6.2) follows at once.

Of course, it is well-known that H possesses a mass gap (see e.g. [13, Theorem 4.5]).

Note, finally, that a non-cut-off Dirichlet ϕ_2^4-theory may have a non-unique vacuum [8].

References

1. Beckner, W.: Inequalities in Fourier analysis. Ann. of Math. (2) **102**, 159–182 (1975)
2. Bonami, A.: Étude des coefficients de Fourier des fonctions de $L^p(G)$. Ann. Inst. Fourier (Grenoble) **20**; 2, 335–402 (1970)
3. Borell, C.: Convexity in Gauss space. In: Aspects statistiques et aspects physiques des processus gaussiens. (Saint-Flour 1980), 27–37. Paris: CNRS No. 307, 1981
4. Borell, C.: Gaussian Radon measures on locally convex spaces. Math. Scand. **38**, 265–284 (1976)
5. Borell, C.: Polynomial chaos and integrability. (To appear)
6. Glimm, J.: Boson fields with nonlinear self-interaction in two dimensions. Comm. Math. Phys. **8**, 12–25 (1968)
7. Glimm, J., Jaffe, A.: The $\lambda(\phi^4)_2$ quantum field theory without cutoffs, II. The field operators and the approximate vacuum. Ann. of Math. (2) **91**, 362–401 (1970)
8. Glimm, J., Jaffe, A.: Phase transitions for ϕ_2^4 quantum fields. Comm. Math. Phys. **45**, 203–216 (1975)
9. Gross, L.: Logarithmic Sobolev inequalities. Amer. J. Math. **97**, 1061–1083 (1975)
10. Gross, L.: Existence and uniqueness of physical ground states. J. Functional Analysis **10**, 52–109 (1972)
11. Guerra, F., Rosen, L., Simon, B.: Correlation inequalities and the mass gap in $P(\phi)_2$, III. Mass gap for a class of strongly coupled theories with nonzero external field. Comm. Math. Phys. **41**, 19–32 (1975)
12. Guerra, F., Rosen, L., Simon, B.: The $P(\phi)_2$ Euclidean quantum field theory as classical statistical mechanics. Ann. of Math. (2) **101**, 111–259 (1975)
13. Hoegh-Krohn, R., Simon, B.: Hypercontractive semigroups and two dimensional self-coupled Bose fields. J. Functional Analysis **9**, 121–180 (1972)
14. Nelson, E.: The free Markoff field. J. Functional Analysis **12**, 211–227 (1973)
15. Segal, I.: Construction of non-linear local quantum processes I. Ann. of Math. (2) **92**, 462–481 (1970)
16. Simon, B.: The $P(\phi)_2$ Euclidean (Quantum) Field Theory. Princeton Series in Physics. Princeton, New Jersey: Princeton University Press 1974
17. Simon, B.: A remark on Nelson's best hypercontractive estimates. Proc. Amer. Math. Soc. **55**, 376–378 (1976)

Received March 23, 1981; received in final form December 21, 1981

Math. Z. 180, 235–247 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Differentiability of Busemann Functions and Total Excess

Nobuhiro Innami

Institute of Mathematics, University of Tsukuba, Sakura-mura Ibaraki, 305 Japan

1. Introduction

A G-space is defined by Busemann [1] as a metric space in which any two points can be joined by a segment and in which any segment may be prolonged uniquely to a geodesic. All complete Riemannian manifolds are, of course, G-spaces.

If A is a ray with origin $q \in R$ in a noncompact G-space R, then we can define a function on R by $f_A(p) := \lim \{qz - pz\}$ where $qz \to \infty$ with $z \in A$, because $qz - pz \leqq qz' - pz' \leqq pq$ for z, $z' \in A$ with $qz \leqq qz'$. The function f_A is called the Busemann function of A.

It is recently known that Busemann functions play important roles in the study of complete manifolds. For example, Cheeger-Gromoll [3], Gromoll-Meyer [7] and Wu [17] have obtained the topological and differentiable structures of complete manifolds of positive or nonnegative sectional curvature by using convexity of Busemann functions on such spaces. Eberlein-O'Neill [5] used them as a useful implement in determining their axial and parabolic manifolds. Shiohama [15, 16] discovered that the existence of exhaustion or nonexhaustion Busemann functions was controlled by the total curvature in noncompact complete 2-manifolds. Nevertheless we have not yet known what points spoil the differentiability of a Busemann function in non-straight spaces. In straight Riemann G-spaces, i.e., complete simply connected Riemannian manifolds without conjugate points, all Busemann functions are differentiable. Eschenburg [6] has proved this.

The purposes of the present paper are to exhibit points which spoil the differentiability of a Busemann function and to investigate the set of all such points.

We are, of course, interested in Riemann and Finsler G-spaces. However in order to clarify the properties which ensure our conclusions we make the arguments start from general G-spaces and put the requisite assumptions on the spaces case by case.

0025-5874/82/0180/0235/$02.60

In order to discuss differentiability of functions on a G-space we need the following notions. A G-space R is in general not a differentiable manifold, so the differentiability of a function on R at a point $p \in R$ is defined along a geodesic curve through p. Let f be a function on an open set U of R. f is by definition *differentiable* at $p \in U$ if f restricted to any geodesic curve through p is differentiable at p for a parameter of a representation of the geodesic curve. Also f is *differentiable* on U if for each point p of U f is differentiable at p. The distance on R is by definition *differentiable around* p if there exists an open neighborhood V_p of p such that the distance function from p is differentiable on $V_p - \{p\}$. If for each point p of R the distance on R is differentiable around p and if there exists a positive ρ_p such that for each $r \in S(p, \rho_p) := \{r;\ pr < \rho_p\}$ we can choose an above open neighborhood V_r of r with $V_r \ni p$, then we say that the distance of R is "*locally differentiable*" on R.

In a G-space R we need to assume another important property concerning the distance function which holds in differentiable manifolds. For any point $p \in R$ the distance function from p restricted to a segment from p is differentiable except p for a parameter of a representation of the segment and its derivative is equal to 1 or -1. We say that the distance from $p \in R$ is "*regular* at $q \in R$" if the differential coefficient of the distance function from p restricted to any geodesic curve through q exists at q and it is equal neither to 1 nor to -1 whenever the geodesic curve through q is not contained in any segment from p. And also we say that the distance of R is *regular* if for each point $p \in R$ the distance from p is regular at every point $q \in R$ at which it is differentiable.

We denote by $B(A)$ the set of all points at each of which the Busemann function of a ray A is not differentiable. Our first aim is to explain $B(A)$ in very geometrical words. To do this, we need the notion of co-points to a ray A which are defined as follows. Let A be a ray with origin $q \in R$ in a noncompact G-space R. A *co-ray* B from p to A is by definition the limit of a converging sequence of segments $T(p_n, z_n)$ where $p_n \to p$ and $z_n \in A$ with $qz_n \to \infty$. The union of all co-rays which contain a co-ray B to A is either a straight line or a ray. In the first case, we call the line, with the orientation for which B is a positive subray, an *asymptote* to A; in the second a *maximal co-ray* to A and its origin a *co-point* to A. We denote by $C(A)$ the set of all co-points to A and by $C_2(A)$ the subset of $C(A)$ whose points are origins of two or more co-rays to A. $C(A)$ and $C_2(A)$ are in general not closed in R (see [12]). Lewis [9] and Nasu [12–14] developed the theory of $C(A)$ and $C_2(A)$ (also see Busemann [2]).

Under these notation we prove

Theorem 2. *Let R be a noncompact G-space with locally differentiable and regular distance and let A be a ray in R. Then $C_2(A) \subset B(A) \subset C(A)$. Further, if all spheres $K(p, \rho) := \{r;\ pr = \rho\}$ with $0 < \rho < \rho(p)$ are not contractible, then $B(A)$ is dense in $C(A)$* (see [1], p. 33 for the definition of $\rho(p)$).

From this fact we understand the study of $C(A)$ to be very interesting. We have in the case where $C(A)$ is bounded

Corollary 10. *Let R be a noncompact G-space with domain invariance. If there exists a ray A in R such that $C(A)$ is nonempty and bounded, then R has exactly*

one end, no asymptote to A exists, and therefore f_A is an exhaustion function (see §2 for the definitions of domain invariance, ends and exhaustion functions).

These are main contents of §2.

In §3 we clarify the relation between the boundedness of $C(A)$ and the total excess of a G-surface (for the implication of a G-surface see [1], p. 52). According Corollary 10, this is a detail of our investigation. To do this, we need the notion of the angular measure which is defined by Busemann [1] as follows. Let p be a point of a G-surface R. A *direction* from, or with origin p, is an oriented segment $T^+(p, q)$ of length $\sigma(p) := \min\{\rho(p)/4, 1\}$. Then for each $q \in K := K(p, \sigma(p))$ there exists a unique direction from p to q. We conceive of an *angle* A with vertex p as a set of directions with origin p passing through the points of a sub-arc of K. A *measure* for the angles at p is a nonnegative function $|A|$ which is defined for all angles A with vertex p and has the following properties: 1) $|A| = \pi$ if and only if A contains exactly one segment through p and connecting points of K, which is called a *straight angle*, 2) If $A_1 \cap A_2$ consists of exactly one direction, then $|A_1 \cup A_2| = |A_1| + |A_2|$. We speak of an *angular measure* on a G-surface R if a definite angular measure has been defined at each of its points.

Using the angular measure above we can define the excesses of triangles and the total excesses of certain regions on a G-surface (see §3) which is, of course, in the case of Riemannian G-surfaces with the Riemannian angular measure, identified with the total curvatures of triangles and the total curvatures of regions by the Gauss-Bonnet Theorem.

The main result of §3 is

Theorem 14. *Let R be a noncompact G-surface with a continuous angular measure and let the total excess $\varepsilon(R)$ of R exist. If there is a ray A in R such that $C(A)$ is nonempty and bounded, then $\varepsilon(R) \geqq 2\pi\chi(R)$ where $\chi(R)$ is the Euler characteristic of R.*

If the angular measure has the property which insures that, in a uniform way, an angle cannot be nearly straight without having a measure close to π, then in the inequality above $\varepsilon(R) = 2\pi\chi(R)$ holds. Such an angular measure is said to be "uniform at π" on R (see §3). Of course, the Riemannian angular measures on Riemannian G-surfaces are uniform at π.

In §4 we investigate the union of the angles composed of the directions of co-rays from each point to a ray. In a certain G-plane R the totality of the measures of these angles at each point is at most the total excess (see Theorem 21). As an application, we show, in combination with Maeda [10] and [11], the existence of infinitely many rays from each point $p \in R$ which are not co-rays from p to a ray (see Corollary 24).

Lastly the author wishes the readers to accept the synthetic approach to differential geometry.

2. Differentiability of Busemann Functions and Co-Points

Let R be a noncompact G-space and let A be a ray in R. If B is a co-ray to A, then B is also a co-ray to any ray contained in or containing A as a sub-ray. Furthermore, the limit of a converging sequence of co-rays to a ray A is

likewise a co-ray to A. The existence of co-rays to A from a point $p \in R$ is in general not unique. However the co-ray to A from any point of a co-ray B other than the origin p of B is unique and a sub-ray of B (see [1], p. 136). For co-rays B_1 and B_2 to A unless one is a sub-ray of the other, then B_1 and B_2 do not intersect each other except at the origins.

Our first observation is

Proposition 1. *Let R be a noncompact G-space with locally differentiable distance and let A be a ray from q. If $p \in R$ is not a co-point to A, then the Busemann function f_A is differentiable at p.*

From this we have

Theorem 2. *Let R be a noncompact G-space with locally differentiable and regular distance. Let A be a ray. Then $C_2(A) \subset B(A) \subset C(A)$. Further, if all spheres $K(p, \rho) := \{r;\ pr = \rho\}$ with $0 < \rho < \rho(p)$ are not contractible, then $B(A)$ is dense in $C(A)$.*

It is known that if a G-space R is finite-dimensional, then all spheres $K(p, \rho)$ with $0 < \rho < \rho(p)$ are not contractible (see [2], p. 16).

From Theorem 2 any noncompact G-space in which no straight lines exist does not admit any differentiable Busemann function.

We say that a G-space R is *straight* if all geodesics are straight lines. In a straight G-space $C(A)$ is empty for every ray A (see [1], p. 138).

Corollary 3. *Let R be as in Proposition 1. If R is straight, then all Busemann functions are differentiable on R.*

This is the version of the result of Eschenburg [6].

Corollary 4. *Let R and $K(p, \rho)$ be as in Theorem 2. If $B(A)$ is empty for a ray A in R, then there is a homeomorphism h of R onto a product $R_1 \times \mathbb{R}$ such that R_1 is homeomorphic to $f_A^{-1}(0)$ by h^{-1} and $\{p\} \times \mathbb{R}$ is the image of an asymptote to A by h.*

For the homeomorphism h is given as follows. If for each $q \in R$ $h_1(q)$ is the point at which the asymptote through q to A intersects $f_A^{-1}(0)$, then we have only to put $h(q) := (h_1(q), f_A(q))$, since generally $f_A(y(s)) = s + f_A(y(0))$ holds where $y(s)$ is a representation of a co-ray to A (see [1], p. 134) and since the asymptote through $p \in R$ to A continuously depends on p.

An *end* τ is by definition an assignment to each compact set K in R a component $\tau(K)$ of $R - K$ in such a way that $\tau(K_1) \supset \tau(K_2)$ if $K_1 \subset K_2$.

From the definition of ends a product $R_1 \times \mathbb{R}$ has one end if R_1 is noncompact and otherwise two ends. Hence we have

Corollary 5. *Let R and $K(p, \rho)$ be as in Theorem 2. If R has at least three ends, then there are no differentiable Busemann functions on R.*

Nasu [14] stated that in a noncompact G-space if for a ray A $C(A)$ is nonempty and compact, then no asymptote to A exists (also see [2], p. 89). Since $f_A(p) \geqq p\,C(A) + \inf f_A(C(A))$ for every $p \in R$, we have

Corollary 6. *Let R and $K(p, \rho)$ be as above. If $B(A)$ is nonempty and compact, then f_A is exhaustive, i.e., for every $a \in \mathbb{R}$ $f_A^{-1}((-\infty, a])$ is compact in R.*

In combination with Nasu's structure theorem [12] of a G-space having an isolated co-point to a ray, we have

Corollary 7. *Let R and $K(p, \rho)$ be as in Theorem 2. If $B(A)$ contains an isolated point p, then $B(A)=\{p\}$ and the space is the union of all co-rays to A with origin p and R is contractible to p. The levels of f_A are the ordinary spheres $K(p, \rho)$ with center p.*

We return to the proofs of Proposition 1 and Theorem 2.

Proof of Proposition 1. Let B be a co-ray through p to A. By the definitions of co-rays to A and of co-points to A there is a sequence of points w_n and $z_n \in A$ such that $w_n \to w (\in B)$, p follows w on B, $pz_n \to \infty$, and segments $T(w_n, z_n) \to B$. Further, from local differentiability of the distance, choose points u_n and v_n on $T(w_n, z_n)$ such that w_n, u_n, p_n and v_n are in this order on $T(w_n, z_n)$ and $u_n \to u \in B$, $p_n \to p \in B$, $v_n \to v \in B$ with $p \in V_u \cap V_v$ where V_u and V_v are the domains as in the definition of local differentiability of the distance. Let $x(s)$, $-\varepsilon \leqq s \leqq \varepsilon$, be a representation of any geodesic curve with $x(0)=p$ in $V_u \cap V_v$. We must prove that $f_A(x(s))$ is differentiable at $s=0$.

By the triangle inequalities, $x(s)z_n \leqq x(s)v_n + p_n z_n - p_n v_n$ for every n and for every s, and hence $p_n v_n - x(s)v_n \leqq p_n z_n - x(s)z_n$. Also $u_n p_n + p_n z_n \leqq x(s)u_n + x(s)z_n$ for every n and for every s, and therefore $p_n z_n - x(s)z_n \leqq x(s)u_n - p_n u_n$. Hence, because $p_n z_n - x(s)z_n = \{qz_n - x(s)z_n\} - \{qz_n - p_n z_n\}$ for every s, we have, as $n \to \infty$, $-(x(s)v - pv) \leqq f_A(x(s)) - f_A(x(0)) \leqq x(s)u - pu$ for every s. Thus we have only to show that $\lim_{s \to 0}(x(s)u - pu)/s = -\lim_{s \to 0}(x(s)v - pv)/s$. This is proved as follows.

Since the function $x(s)u + x(s)v$ for s has a minimum at $s=0$ and since $x(s)u$ and $x(s)v$ are differentiable at $s=0$, $0 = \lim_{s \to 0}(x(s)u + x(s)v - uv)/s = \lim_{s \to 0}(x(s)u - pu)/s + \lim_{s \to 0}(x(s)v - pv)/s$. The proof is complete.

It should be noted that, under the notations of the proof, even if p is a co-point to A, the inequality $-(x(s)v - pv) \leqq f_A(x(s)) - f_A(p)$ holds for every s and the equality sign holds at $s=0$, so that if f_A is differentiable at p, then the differential coefficient of f_A at p along $x(s)$ is equal to the one of the negative distance function from v at p along $x(s)$.

Proof of Theorem 2. From Proposition 1 it follows that $B(A) \subset C(A)$. Suppose there exists a point $p \in C_2(A) - B(A)$. Let A_1 and A_2 be two distinct co-rays from p to A. Let $y(s)$, $0 \leqq s < \infty$, be a representation of a ray A_2 with $y(0)=p$ and let v be a point on A_1 such that it follows p and $p \in V_v$. From the fact that $f_A(y(s)) = s + f_A(p)$ it follows that the differential coefficient of $f_A(y(s))$ at $s=0$ is equal to 1. However, by regularity of the distance, the differential coefficient of $y(s)v$ is not equal to -1, a contradiction to the above remark. Thus $C_2(A) \subset B(A)$.

The density of $B(A)$ in $C(A)$ follows from [2], p. 89. The proof is complete.

We notice that if R is a Riemannian G-space, then the gradient of f_A is continuous on $R-B(A)$.

In the following we will discuss and partially answer the question; in what spaces there exists a ray A such that $C(A)$ is bounded.

We say that a Hausdorff space R has the property of *domain invariance* if a subset of R which is homeomorphic to an open set of R is open in R. According to Busemann [2], p. 16, domain invariance holds in finite-dimensional G-spaces.

Theorem 8. *Let R be a noncompact G-space with domain invariance and let A be a ray in R. If $C(A)$ is nonempty, then $f_A(C(A))$ is dense in the interval* $[\inf f_A(R), \sup f_A(C(A))]$.

To give a corollary of Theorem 8 we need a short remark. If a noncompact G-space R has at least two ends, then for every ray A there exists an asymptote to A, and f_A, in particular, takes no minimums. In fact, from the assumption, there exists a compact set $K \subset R$ such that $R-K$ contains at least two unbounded components and a sub-ray of A is contained in one component W of $R-K$. Take an unbounded sequence p_j of points in $R-K$ each of which is contained in distinct components from W. Since for each j a co-ray A_j from p_j to A intersects K, there exists a subsequence A_k of the sequence A_j converging to an asymptote to A.

Corollary 9. *Let R be a noncompact G-space with domain invariance and with at least two ends. If for a ray A in R $C(A)$ is nonempty, then $C(A)$ is unbounded. In particular, if R has at least three ends, then for every ray A $C(A)$ is nonempty and unbounded.*

Corollary 10. *Let R be a noncompact G-space with domain invariance. If there exists a ray A in R such that $C(A)$ is nonempty and bounded, then R always has exactly one end, no asymptote to A exists, and hence f_A is exhaustive.*

This corollary should be compared with Corollary 6.

To the proof of Theorem 8 we need the following lemma.

Lemma 11. *Let R be a noncompact G-space and let A be a ray in R. Then, for every $a \in f_A(R)$ $f_A^{-1}([a, \infty))$ is pathwise connected.*

Proof. We first prove that we connect a point p with $f_A(p) > a$ and a point q with $f_A(q) > a$ by a curve. To show this we will obtain two curves in $f_A^{-1}([a, \infty))$ from p to some point $z \in A$ and from q to the same point $z \in A$. If for each $z \in A$ a segment $T(p, z)$ from p to z (or $T(q, z)$) intersects $f_A^{-1}(a)$, then we denote by $p(z)$ (or $q(z)$) a point in the intersection. Thus $rz - pz = rz - p(z)z - pp(z)$ or $rz - qz = rz - q(z)z - qq(z)$ for every $z \in A$ respectively, where r is the origin of A. In general if z' follows z on A, then $rz - sz \leqq rz' - sz'$ for every point $s \in R$. Hence $rz - pz \leqq a - pp(z)$ or $rz - qz \leqq a - qq(z)$ for every $z \in A$. From this, for every $z \in A$ such that rz is sufficiently large, $pp(z)$ and $qq(z)$ is negative, a contradiction. Therefore using such a point $z \in A$ we can obtain a curve $T(p, z) \cup T(q, z)$ from p to q.

For point p with $f_A(p)=a$ and for a point q with $f_A(q)>a$, for example, we pick out a point $p_1 \neq p$ on a co-ray from p to A and a curve T from p_1 to q in $f_A^{-1}([a, \infty))$ as above. Then we get a curve $T(p, p_1) \cup T$ from p to q in $f_A^{-1}([a, \infty))$.

Proof of Theorem 8. Suppose that there exist values a and b such that for every t, $a<t<b \leqq \sup f_A(C(A))$, $f_A^{-1}(t)$ contains no co-points to A. Fix t_0, $a<t_0<b$. From Lemma 11 $f_A^{-1}([t_0, \infty))$ is connected. Now let X be the set of points each of which is contained in the co-ray to A from some point $p \in f_A^{-1}(t_0)$. Clearly $X \subset f_A^{-1}([t_0, \infty))$. We show that X is closed and open in $f_A^{-1}([t_0, \infty))$.

Assume that $p_j \in X$ converges to p_0 in $f_A^{-1}([t_0, \infty))$. By the construction of X, for each p_j there exists a unique point $q_j \in f_A^{-1}(t_0)$ the co-ray to A from which contains p_j. For each j the distance between p_j and q_j is at most $\sup_j \{f_A(p_j)\} - t_0$ and hence q_j is bounded. Since $f_A^{-1}(t_0)$ is closed, there exists a subsequence q_k of q_j converging to $q_0 \in f_A^{-1}(t_0)$. Because $f_A^{-1}(t_0)$ contains no co-points to A and therefore the co-ray from a point q in $f_A^{-1}(t_0)$ to A continuously depends on q, p_k converges to p_0, which is contained in a unique co-ray from q_0 to A, namely $p_0 \in X$. Hence X is closed in $f_A^{-1}([t_0, \infty))$.

To prove that X is open in $f_A^{-1}([t_0, \infty))$ we need the property of domain invariance of R. For any point q' in X there is a unique point q in $f_A^{-1}(t_0)$ the co-ray to A from which contains q'. If $f_A(q')=t_0$, then we have nothing to prove, because $f_A([t_0, b))$ is a neighborhood of q' in X. So we may assume that $f_A(q')>t_0$. Since $f_A^{-1}((t_0, b))$ is open in R, for given point $p \in T(q, q') \cap f_A^{-1}((t_0, b))$ there exists an open neighborhood V of p in R which is contained in $f_A^{-1}((t_0, b))$. Define a map of V into X by sending $r \in V$ to a point r' such that r' is contained in the co-ray from r to A and the distance between r and r' is pq' ($q'=p'$ under our new notation). Obviously this map h is a homeomorphism of V onto its image containing p. From domain invariance of R the image $h(V) \subset X$ is open in $f_A^{-1}([t_0, \infty))$. Hence X is open in $f_A^{-1}([t_0, \infty))$.

Therefore $X=f_A^{-1}([t_0, \infty))$. However this is impossible, because X contains no co-points to A from the construction of X and because $f_A^{-1}([t_0, \infty))$ contains co-points to A. The proof is complete.

In the proof above we have just shown that there are no values a and $b \leqq \sup f_A(C(A))$ such that no components of $f_A^{-1}((a, b))$ exist such that it does not contain any co-points to A. Hence we have

Proposition 12. *Let R be a noncompact G-space with domain invariance and with at least two ends. Let A be a ray in R. If $C(A)$ is nonempty, then for each compact set K there are co-points to A in all unbounded components of $R-K$, except only one which contains a sub-ray of A.*

For the investigations in §3 it is convenient to notice here the case of G-surfaces in more details.

Proposition 13. *Let R be a noncompact G-surface and let A be a ray in R. If $C(A)$ is nonempty and bounded, then no asymptote to A exists, f_A is exhaustive, and moreover there is a t_0 such that $f_A^{-1}(t)$ is a one-dimensional embedded sub-*

manifold of R which is homeomorphic to a circle S^1 for every $t \geqq t_0$ and $f_A^{-1}([t_0, \infty))$ is homeomorphic to a product $f_A^{-1}(t_0) \times [0, \infty)$.

Proof. The first two assertions are contained in Corollary 10. Since $C(A)$ is bounded, there is a t_0 such that $f_A^{-1}([t_0-1, \infty)) \cap C(A)$ is empty. By the same reasoning as in Corollary 4 $f_A^{-1}([t_0, \infty))$ is homeomorphic to a product $f_A^{-1}(t_0) \times [t_0, \infty)$. The proof of the remainder is as follows.

For given $p \in f_A^{-1}(t)$, $t \geqq t_0$, there is a point q such that $t_0 > f_A(q) > t_0 - 1$ and the co-ray B from q to A contains p. Let T be a segment through q such that T is contained in $f_A^{-1}((t_0-1, t_0))$ and the intersection $T \cap B$ consists of only one point q. Note that for every co-ray B' from $r \in T$ to A the intersection $T \cap B'$ consists only one point, since if $T \cap B'$ contains two points, then q must be a co-point to A. Let D be the set of all points each of which is contained in the co-ray from some point $r \in T$ to A. Then, D is homeomorphic to a half-strip $[0, 1] \times [0, \infty)$, since the co-ray from $r \in T$ to A continuously depends on r. In particular, D is a neighborhood of p. Obviously the intersection $D \cap f_A^{-1}(t)$ is homeomorphic to T, so that $f_A^{-1}(t)$ is a one-dimensional embedded submanifold of R without boundary. Since f_A is exhaustive, $f_A^{-1}(t)$ is compact, and hence $f_A^{-1}(t)$ is topologically a circle S^1. The proof is complete.

3. Co-points and Total Excesses

In this section we see to what extent the total excess of a G-surface influences the set of all co-points to a ray. We start to confirm the notions which are used in this section. They are found in Busemann [1], pp. 273–305.

We have already defined the angles and the angular measures in the introduction. We say that the angles A_n with vertex p_n *tends* to the angle A with vertex p if $p_n \to p$ and the point set carring the directions in A_n tends to the point set carring the directions in A. An angular measure is by definition *continuous* if $A_n \to A$ implies $|A_n| \to |A|$. The excess of a triangle abc in $S(p, \rho(p)/8)$ is the value $\varepsilon(abc) = |abc| + |bca| + |cab| - \pi$, where $|abc|$, for example, implies the measure of the angle composed of the directions with origin b and through points of the segment $T(a, c)$. Let D be a polygonal compact domain, i.e., its boundary, if exists, consists of finitely many simple closed geodesic polygons. We put $\varepsilon(D) = \sum_{i=1}^{n} \varepsilon(a_i b_i c_i)$, where $a_i b_i c_i$ is the decomposition of D into triangles each of which is contained in $S(p, \rho(p)/8)$ for some p. $\varepsilon(D)$ is independent of the choice of the decomposition and is called the *total excess* of D. Busemann has stated in [1], p. 283 that $\varepsilon(D) = 2\pi\chi(D) - \sum_{j=1}^{m} (\pi - \beta_j)$, where β_j is the measure of the inner angle of D with vertex p_j on the boundary of D and $\chi(D)$ is the Euler characteristic of D. An angular measure is said to be "*uniform at* π" in the subset G of R if a nondecreasing positive function $\delta(\varepsilon) < 1$ and a positive function $\omega(p, \varepsilon) \leqq \rho(p)/4$ defined for $0 < \varepsilon < \pi$ and $p \in G$ exist, such that the relations $0 < a_1 p = a_2 p < \omega(p, \varepsilon)$ and $a_1 a_2/(a_1 p + p a_2) > 1 - \delta(\varepsilon)$ imply $|a_1 p a_2| > \pi - \varepsilon$. This property insures that, in a uniform way, an angle cannot

be nearly straight without having a measure close to π. Of course, the Riemannian angular measure on a Riemannian G-surface R is uniform at π on R (see [1], p. 293). Finally we need the notion of the total excess of a noncompact region G. Let G be a subset of R which is the union of an increasing sequence D_n of polygonal domains $G=\bigcup D_n$ such that $p_i \in D_{n_i+1}-D_{n_i}$ implies that p_i has no accumulation point for any sequence $n_i \to \infty$. If $\varepsilon(G) := \lim_{n\to\infty} \varepsilon(D_n)$ exists for each such sequence D_n, $\pm\infty$ admitted, then we call it the total excess of G. Busemann [1], p. 300 has proved that if the G-surface R of finite connectivity possesses the total excess with respect to the angular measure which is uniform at π on R, then $\varepsilon(R) \leqq 2\pi\chi(R)$. This is a generalization of Cohn-Vossen [4].

Under the notions above we establish

Theorem 14. *Let R be a noncompact G-surface with a continuous angular measure and let the total excess $\varepsilon(R)$ of R exist. If there exists a ray A in R such that $C(A)$ is nonempty and bounded, then $\varepsilon(R) \geqq 2\pi\chi(R)$.*

According to the remark preceding Theorem 14 we have

Corollary 15. *Let R and $C(A)$ be as above. Further, if the angular measure is uniform at π on R, then $\varepsilon(R)=2\pi\chi(R)$. In particular, if R is a Riemannian G-surface with the Riemannian angular measure, then total curvature of R, $\varepsilon(R)$, is equal to $2\pi\chi(R)$.*

If for any point p outside a bounded set in a G-surface there is only one ray starting from p, then the assumption concerning $C(A)$ is automatically satisfied. Hence we obtain

Corollary 16. *Let R be as in Theorem 14. If there is a bounded set K of R such that for every $p \notin K$ the only one ray from p exists, then $\varepsilon(R) \geqq 2\pi\chi(R)$.*

Before starting to prove Theorem 14 we provide the notations and three lemmas under the assumption of Theorem 14.

From Proposition 13 there is a $t_0 > 0$ such that $C(A)$ is contained in $f_A^{-1}((-\infty, t_0-2])$ and $f_A^{-1}([t_0-1, \infty))$ is topologically a cylinder $S^1 \times [t_0, \infty)$.

Lemma 17. *For every $t \geqq t_0$ there is a point $z(t) \in A$ such that for every $z \in A$ which follows $z(t)$ on A and for every $q \in f_A^{-1}(t)$ all segments from q to z does not intersect $f_A^{-1}((-\infty, t-1])$.*

Proof. Suppose this is false. Then there exists a sequence $q_j \in f_A^{-1}(t)$ and a sequence $z_j \in A$ such that $q_j z_j \to \infty$ and a segment from q_j to z_j, for each j, intersects $f_A^{-1}((-\infty, t-1])$. The sequence of these segment contains a subsequence converging to a co-ray B from a point in $f_A^{-1}(t)$ to A. f_A is monotone increasing on B, so B cannot intersect $f_A^{-1}((-\infty, t-1])$, a contradiction.

Lemma 18. *Let $t \geqq t_0$. Then, for every $z \in A$ which follows $z(t) \in A$ as above there exists a point $q(z) \in f_A^{-1}(t)$ such that at least two segments from $q(z)$ to z exist and the union of two of these segments, if chosen, surrounds a compact polygonal region $D(z)$ which contains $f_A^{-1}((-\infty, t-1])$ in its interior.*

Proof. Let p be a point on A such that $f_A(p)=t$ and let $c(s)$, $0\leqq s\leqq 1$, be a representation of $f_A^{-1}(t)$ with $c(0)=c(1)=p$. Fix a $z\in A$ which follows $z(t)$ on A. For each s, $0\leqq s\leqq 1$, we denote by $X(s)$ (or $Y(s)$) the set of all curves from $c(s)$ to z in $R':=R-f_A^{-1}((-\infty, t-1))$ each of which is homotopic to the curve $c([0, s])\cup T(p, z)$ (or $c([s, 1])\cup T(p, z)$) in R'. We note that for every $\tilde{x}\in X(s)$ and for every $\tilde{y}\in Y(s)$ the union $\tilde{x}\cup\tilde{y}$ is homotopic to the union $T(z, p)\cup f_A^{-1}(t)\cup T(p, z)$ in R', so that $\tilde{x}\cup\tilde{y}$ surrounds $f_A^{-1}((-\infty, t-1])$. For each s, $0\leqq s\leqq 1$, let $\tilde{x}(s)$ be an element of $X(s)$ such that for every element $\tilde{x}$ of $X(s)$ the length of $\tilde{x}(s)$ is not greater than the one of $\tilde{x}$. Similarly let $\tilde{y}(s)$ be an element of $Y(s)$ with minimum length in $Y(s)$. We denote by $\lambda(s)$ and $\mu(s)$ the lengths of $\tilde{x}(s)$ and $\tilde{y}(s)$ respectively. Obviously $\tilde{x}(0)=T(p, z)$ and $\tilde{y}(1)=T(p, z)$, so $\lambda(0)<\mu(0)$ and $\lambda(1)>\mu(1)$, because of the uniqueness of the existence of a segment $T(p, z)$. By semicontinuity of the lengths of curves (see [1], p. 20) $\lambda(s)$ and $\mu(s)$, $0\leqq s\leqq 1$, are continuous, so that there is an s_0 with $0<s_0<1$ such that $\lambda(s_0)=\mu(s_0)$. Then both $\tilde{x}(s_0)$ and $\tilde{y}(s_0)$ are segments from $c(s_0)$ to z in R. In fact, for every s, $0\leqq s\leqq 1$, Lemma 17 implies that $\tilde{x}(s)$, $\tilde{y}(s)$ or both are segments in R, because R' is topologically a half cylinder. Thus we have only to put $q(z)=c(s_0)$, and then $\tilde{x}(s_0)\cup\tilde{y}(s_0)$ surrounds a desired compact domain $D(z)$. The proof is complete.

Note that $\chi(R)=\chi(f_A^{-1}((-\infty, t-1]))=\chi(D(z))$ for every $t\geqq t_0$.

Lemma 19. *For any $\varepsilon>0$ and for every $t\geqq t_0$ there is a $w(t)\in A$ such that $w(t)$ follows $z(t)$ on A and $\varepsilon(D(w(t)))>2\pi\chi(R)-\varepsilon$ where $D(w(t))$ is as in Lemma* 18.

Proof. Since $f_A^{-1}(t)$ is compact, there exists a sequence z_i on A such that $pz_i\to\infty$ and $q(z_i)$ converges to some point in $f_A^{-1}(t)$. If $\alpha(z_i)$ is the measure of the inner angle of $D(z_i)$ at $q(z_i)$ and so is $\beta(z_i)$ at z_i, then $\alpha(z_i)\to 2\pi$ as $pz_i\to\infty$, since the angular measure is continuous and since for every point in $f_A^{-1}(t)$ the existence of a co-ray from it to A is unique. Hence there exists a $w(t):=z_{i_0}\in A$ which follows $z(t)$ on A such that $\varepsilon(D(w(t)))=2\pi\chi(R)-(\pi-\alpha(w(t)))-(\pi-\beta(w(t)))>2\pi\chi(R)-\varepsilon$.

Now we prove Theorem 14.

Proof of Theorem 14. We want to prove that $\varepsilon(R)>2\pi\chi(R)-\varepsilon$ for any $\varepsilon>0$. Put $D_0:=D(w(t_0))$ as above. We assume by induction hypothesis that there exists a sequence $D_0, D_1, \ldots, D_n$ of compact polygonal domains such that $D_i\subset D_{i+1}$ for $i=0, 1, \ldots, n-1$ and $\varepsilon(D_i)>2\pi\chi(R)-\varepsilon$ for $i=0, 1, \ldots, n$. If $a:=\max f_A(D_n)$, then we can put $D_{n+1}:=D(w(a+2))$ from Lemma 19. Obviously the sequence D_i satisfies the conditions which are needed for the definition of the total excess of R. Thus $\varepsilon(R)\geqq 2\pi\chi(R)$. The proof is complete.

4. Remarks on the Existence of Co-Rays

In this section we treat the case that a Busemann function f_A on a G-plane, i.e., topologically a plane, is convex. It is known that every Busemann function on a Riemannian manifold of nonnegative sectional curvature is convex (see [3, 17]). In a G-plane if a convex function takes a minimum, then the mini-

mum set is either a point, a segment, a ray, or a straight line (see [8]). However the last does not occur in the case of convex Busemann functions as in the following.

Proposition 20. *Let R be a noncompact G-space. The minimum set of a Busemann function does not separate the space R.*

For we have already proved this in the proof of Lemma 11.

It is proved in [8] that unless a convex function on a G-surface R has nonempty minimum set which is a one-dimensional manifold without boundary, then all levels are connected and all levels except the minimum set are one-dimensional embedded submanifolds without boundary, and hence they are topologically circles or real lines. Moreover they are homeomorphic to each other.

Let R be a G-plane and let a Busemann function f_A be convex and nonexhaustive. Then from the fact above every level except the minimum set is homeomorphic to a real line. If A_1 and A_2 are distinct co-rays from $p \in C_2(A)$ to A, then they intersect $f_A^{-1}(t)$ at distinct points for all $t > f_A(p)$, and therefore they cut off the sub-arc $\tilde{c}(t)$ of $f_A^{-1}(t)$ for all $t > f_A(p)$.

Theorem 21. *Let R be a G-plane which possesses an continuous angular measure with nonnegative excess and let a Busemann function f_A be convex and non-exhaustive. Then for each point $p \in C_2(A)$ the measure of the inner angle at p of the domain $D := \bigcup_{t > f_A(p)} \tilde{c}(t)$ as above is not greater than the total excess $\varepsilon(R)$ of R.*

We notice that since the excess of every triangle is nonnegative, $\varepsilon(R)$ always exists (see [1], p. 299).

Proof. We have only to prove the statement for two co-rays A_1 and A_2 from p to A such that any co-ray from p to A passes through the sub-arcs cut off from levels by A_1 and A_2. Let p_1 and p_2 be points in A_1 and A_2 respectively with $f_A(p_1) = f_A(p_2) > f_A(p)$. First we show that there exists a $z_0 \in A$ such that if $z \in A$ follows z_0 on A, then the union $T(p, p_1) \cup T(p_1, z) \cup T(z, p_2) \cup T(p_2, p)$ of four segments is a simple 4-gon, and such that the compact domain $D(z)$ which is surrounded by the 4-gon tends to the domain D in the statement. If q is the origin of A, then $qz - p_i z \leqq f_A(p_i) - p_i r_i(z)$, $i = 1, 2$, as in the proof of Lemma 11, where $z \in A$ and $r_i(z)$ is, if it exists, a point of the intersection of a segment $T(p_i, z)$ and $f_A^{-1}(f_A(p_i))$. Hence $p_i r_i(z)$, $i = 1, 2$, tends to zero. Thus there is a $z_0 \in A$ such that $p_i r_i(z) < p_i T(p_j, p)$, $i \neq j$, for every z following z_0 on A. Then our 4-gon is simple, because, for example, a segment $T(p_1, z)$ intersects the union $T(p, p_1) \cup T(p, p_2) \cup T(p_2, z)$ of segments only at two points p_1 and z. Since each of components of $R - f_A^{-1}(f_A(p_1))$ is homeomorphic to a plane and since segments $T(p_1, z)$ and $T(p_2, z)$ tends to the sub-rays of A_1 and A_2 respectively as $pz \to \infty$ with $z \in A$, the domain $D(z)$ surrounded by the 4-gon tends to the domain D in the statement.

Because the excesses are nonnegative,

$$\varepsilon(D(z)) = 2\pi - (\pi - \alpha(z)) - (\pi - \beta(z)) - (\pi - \gamma(z)) - (\pi - \delta(z)) \leqq \varepsilon(R)$$

where $\alpha(z)$, $\beta(z)$, $\gamma(z)$ and $\delta(z)$ are the measures of the inner angles of $D(z)$ at p, p_1, p_2 and z respectively. Hence $\alpha(z)+\beta(z)+\gamma(z)-2\pi \leqq \varepsilon(R)$. Since the angular measure is continuous, $\beta(z)$ and $\gamma(z)$ tends to π as $pz \to \infty$ with $z \in A$. Therefore $\alpha(z) \leqq \varepsilon(R)$, which is our goal.

Let D be the region which is composed of all co-rays from p to a ray A. In general $D-\{p\}$ is not connected and each component may contain many co-rays from p to A. For each component D_λ of $D-\{p\}$, if we apply Theorem 21 to the boundary of $D_\lambda \cup \{p\}$, then the measure of the inner angle of $D_\lambda \cup \{p\}$ at p is at most $\varepsilon(R)$. Nevertheless the proof of Theorem 21 yields a stronger result than this.

Corollary 22. *Let R be a G-plane which possesses a continuous angular measure with nonnegative excess and let a Busemann function f_A be convex and non-exhaustive. Then for each point p the totality of the measures of the inner angles of $D_\lambda \cup \{p\}$'s at p is at most the total excess $\varepsilon(R)$ of R.*

In the Riemannian G-plane M^2 with the Riemannian angular measure Shiohama [15] has proved that if M^2 is of nonnegative Gaussian curvature and if the total curvature is not greater than π, then all Busemann functions on M^2 are nonexhaustive. Hence we have

Corollary 23. *Let M be a Riemannian G-plane with the Riemannian angular measure and of nonnegative Gaussian curvature. If the total curvature $\varepsilon(M)$ is not greater than π, then for each ray A and for every point p the totality of the measures of the inner angles of $D_\lambda \cup \{p\}$'s at p is at most $\varepsilon(M)$.*

As another application we prove the existence of rays from each point p which are not co-rays from p to a given ray A. To do this, we carefully observe Maeda's result in [10] and [11], i.e., if M is a Riemannian G-plane with the Riemannian angular measure and of nonnegative Gaussian curvature, then the totality of the measures of the inner angles at each point p of the domains which are composed of all rays from p is at least $2\pi-\varepsilon(M)$ where $\varepsilon(M)$ is the total curvature of M. Using this we have

Corollary 24. *Let M be a Riemannian G-plane with the Riemannian angular measure and of nonnegative Gaussian curvature. If the total curvature $\varepsilon(M)$ is smaller than π, then for each ray A and for every point $p \in R$ there are infinitely many rays from p which are not co-rays from p to A, more precisely, the totality of the measures of the angles at p of domains which are composed of rays from p that are not co-rays to A is at least $2\pi-2\varepsilon(M)$.*

Our estimate is, of course, valid only at points of $C_2(A)$.

References

1. Busemann, H.: The geometry of geodesics. New York: Academic Press 1955
2. Busemann, H.: Recent synthetic differential geometry. Berlin-Heidelberg-New York: Springer 1970
3. Cheeger, J., Gromoll, D.: On the structure of complete manifolds of nonnegative curvature. Ann. of Math. (2) **96**, 415-443 (1972)

4. Cohn-Vossen, S.: Kürzeste Wege und Totalkrümmung auf Flächen. Compositio Math. **2**, 63–133 (1935)
5. Eberlein, P., O'Neill, B.: Visibility manifolds. Pacific J. Math. **46**, 45–109 (1973)
6. Eschenburg, J.-H.: Horospheres and the stable part of the geodesic flow. Math. Z. **153**, 237–251 (1977)
7. Gromoll, D., Meyer, W.: On complete open manifolds of positive curvature. Ann. of Math. (2) **90**, 75–90 (1969)
8. Innami, N.: A classification of Busemann G-surfaces which possess convex functions. Acta Math. (to appear)
9. Lewis, G.M.: Cut loci of points at infinity. Pacific J. Math. **43**, 675–690 (1972)
10. Maeda, M.: On the existence of rays. Sci. Rep. Yokohama Nat. Univ. Sect. I **26**, 1–4 (1979)
11. Maeda, M.: Remarks on the distribution of rays. Preprint
12. Nasu, Y.: On asymptotic conjugate points. Tôhoku Math. J. Ser. 2, **7**, 157–165 (1955)
13. Nasu, Y.: On asymptotes in metric spaces with non-positive curvature. Tôhoku Math. J. Ser. 2, **9**, 68–95 (1957)
14. Nasu, Y.: On asymptotes in two-dimensional metric spaces. Tensor New Ser. **7**, 173–184 (1957)
15. Shiohama, K.: Busemann function and total curvature. Invent. Math. **53**, 281–297 (1979)
16. Shiohama, K.: A role of total curvature on complete non-compact Riemannian 2-manifolds. Preprint
17. Wu, H.: An elementary method in the study of nonnegative curvature. Acta Math. **142**, 57–78 (1979)

Received August 26, 1981; received in final form December 28, 1981

Math. Z. 180, 249–255 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Compact Integral Operators on Banach Function Spaces

Peter G. Dodds[1] and Anton R. Schep[2]

[1] Department of Mathematics, Flinders University, Bedford Park, SA 5042, Australia
[2] Department of Mathematics, California Institute of Technology, Pasadena, California 91125, U.S.A.

1. Introduction

In the monograph [8] of Halmos and Sunder, it is shown that if (X, μ), (Y, ν) are finite measure spaces, then each integral operator $T: L^2(Y, \nu) \to L^2(X, \mu)$ is compact if considered as a map from $L^2(Y, \nu)$ to $L^1(X, \mu)$. In the present note we shall present two much more general results on compactness of integral operators, each one of them containing the above so-called $\langle 2.1 \rangle$-compactness as a special case. To be specific, our first result is that each integral operator from the Banach function L_{ρ_1} to the Banach function space L_{ρ_2} is compact provided $\sigma(L_{\rho_2}) < \min(s(L_{\rho_1}), 2)$, where σ (respectively s) denotes the upper (respectively lower) indices introduced for Banach function spaces by Grobler in [7]. Our second result is that if L_{ρ_1} and L_{ρ_2} are as before and $L_{\rho_2} \subseteq L_{\rho_3}$, then each integral operator from L_{ρ_1} into L_{ρ_2} is compact as a mapping from L_{ρ_1} into L_{ρ_3} provided ρ_1' is order continuous and $\sigma(L_{\rho_3}) < s(L_{\rho_2})$. The approach of the present paper is entirely different to that of [8] and is partly based on a compactness criterion given in [4] which derives from a characterization due to Luxemburg and Zaanen [13] for compactness of order-bounded kernel operators on Banach function spaces. The methods of [4] not only yield the $\langle 2.1 \rangle$-compactness result as a special case, but place it in perspective with reference to related work of Andô [1], who treated order-bounded kernel operators on Orlicz spaces, and Rosenthal [15], who considered continuous linear mappings from (subspaces of) L^p-spaces. It is perhaps appropriate to point out that these latter papers have a common antecedent in the work of Pitt [14] for sequence spaces.

We refer the reader to [19] for the basic facts and terminology of Banach function spaces and we shall use the monograph [16] as general reference on Riesz spaces and Banach lattices. We shall use $L^0(X, \mu)$ to denote the space of equivalence classes of realvalued μ-measurable functions. If L_ρ is a Banach function space, then L_ρ' denotes the first (Köthe) associate space and $(L_\rho')^+$ shall denote the set of non-negative elements of L_ρ'. The result contained in Theorem 2.7 was announced by the first author at the Oberwolfach meeting on

0025-5874/82/0180/0249/$01.40

Riesz spaces and order bounded operators in June, 1979. The results contained in Theorem 2.8 and Theorem 2.11 were announced by the second author on the Conference on Measure Theory and its applications in October, 1980 (see [18]).

2. Compact Integral Operators

Proposition 2.1. *Let $L_{\rho_1} \subseteq L^0(Y, \nu)$, $L_{\rho_2} \subseteq L^0(X, \mu)$ be Banach function spaces and let $T: L_{\rho_1} \to L_{\rho_2}$ be an integral operator. If T' is the restriction to L'_{ρ_2} of the Banach space adjoint T^*, then T' maps L'_{ρ_2} into L'_{ρ_1}.*

Proof. Let $0 \leqq \phi \in L'_{\rho_2}$ be given. To show that $T'\phi \in L'_{\rho_1}$, it suffices to show that $\langle f_n, T^*\phi \rangle \to 0$ as $n \to \infty$ for each sequence $\{f_n\} \subseteq L_{\rho_1}$ with $f_n \downarrow_n 0$, by Theorem 69.3 of [20]. Suppose then that $0 \leqq f_n \in L_{\rho_1}$, $n = 1, 2, \ldots$ and that $f_n \downarrow_n 0$ holds. Since T is an integral operator, it follows that $Tf_n \to 0$ μ-a.e. (see for example Theorem 3.3 of [17]). By a well known theorem of Grothendieck [6, 2], it follows from the continuity of T that T maps order intervals of L_{ρ_1} to relatively weakly compact sets in $L^1(\phi d\mu)$ and consequently the sequence $\{Tf_n\}_{n=1}^{\infty}$ is relatively weakly compact in $L^1(\phi d\mu)$. It now follows from Theorem IV, 8.12 of [5] that $\int |Tf_n| \phi d\mu \to 0$ as $n \to \infty$ and this clearly suffices to complete the proof of the Proposition.

Corollary 2.2. *If L_{ρ_2}, L_{ρ_1} are as above and if $T: L_{\rho_1} \to L_{\rho_2}$ is an integral operator, then T is $\sigma(L_{\rho_1}, L'_{\rho_1})$ to $\sigma(L_{\rho_2}, L'_{\rho_2})$ continuous.*

Proposition 2.3. *Let $L_{\rho_1} \subseteq L^0(Y, \nu)$, $L_{\rho_2} \subseteq L^0(X, \mu)$ be Banach function spaces and let $T: L_{\rho_1} \to L_{\rho_2}$ be an integral operator. Assume that L'_{ρ_1} has order continuous norm. Then*

(i) *T is measure compact, i.e. if $\{f_n\}_{n=1}^{\infty} \subseteq L_{\rho_1}$ is norm bounded, then $\{f_n\}_{n=1}^{\infty}$ contains a subsequence $\{g_n\}_{n=1}^{\infty}$ such that $\{Tg_n\}$ is convergent in μ-measure on every measurable subset of X of finite μ-measure*

(ii) *T maps the unit ball of L_{ρ_1} to a set which is relatively compact in $L^1(\phi d\mu)$ for each $0 \leqq \phi \in L'_{\rho_2}$.*

Proof. Statement (i) is well-known [9]; see also [10], Theorem 1.10. To see (ii), observe that the unit ball of L_{ρ_1} is relatively $\sigma(L_{\rho_1}, L'_{\rho_1})$ sequentially compact since L'_{ρ_1} has order continuous norm, by [13], Theorem 5.4. From Corollary 2.2 it follows that T maps the unit ball of L_{ρ_1} to a set which is relatively weakly compact in $L_1(\phi d\mu)$ for each $0 \leqq \phi \in L'_{\rho_2}$, since $L_1(\phi d\mu)^*$ may be identified with the order ideal generated in L'_{ρ_2} by the element ϕ. By the measure compactness of T, the assertion of (ii) now follows from Theorem IV, 8.12 of [5] and the proof is complete.

Statement (ii) of the above proposition says that every integral operator from L_{ρ}, with ρ' order continuous, into $L_1(X, \mu)$ is compact. One might conjecture therefore a dual version, which says that every integral operator from L_{∞} into L_{ρ}, with ρ order continuous, is compact. This is however false as Example 2.10 will show. Statement (ii) above for order bounded integral

operators is contained in [13], Lemma 7.2. Further, in the terminology of [4], it asserts that the integral operator T is PL-compact. We now present a criterion for compactness of integral operators.

Theorem 2.4. *Let T be a continuous linear map from the Banach function space $L_{\rho_1} \subseteq L^0(Y, \nu)$ into the Banach function space $L_{\rho_2} \subseteq L^0(X, \mu)$. Assume that L'_{ρ_1} and L_{ρ_2} have order continuous norms.*

(i) *If T is an integral operator, then T is compact if and only if $\rho'_1(T'\phi_n) \to 0$ as $n \to \infty$, whenever the sequence $\{\phi_n\}_{n=1}^{\infty} \subseteq L'^{+}_{\rho_2}$ is norm bounded and disjointly supported.*

(ii) *Assume that T' maps L'_{ρ_2} into L'_{ρ_1}. If T' is an integral operator, then T is compact if and only if $\rho_2(Tf_n) \to 0$ as $n \to \infty$ whenever $\{f_n\} \subseteq L^{+}_{\rho_1}$ is norm bounded and disjointly supported.*

(iii) *If both T and T' are integral operators, then T is compact if and only if $\int \phi_n Tf_n \, d\mu \to 0$ as $n \to \infty$, whenever $\{\phi_n\}_{n=1}^{\infty}$ and $\{f_n\}_{n=1}^{\infty}$ are norm bounded and disjointly supported elements of $L'^{+}_{\rho_2}$, $L^{+}_{\rho_1}$ respectively.*

The above Theorem 2.4 is essentially a consequence of [4], Theorem 5.5 although direct application of this Theorem requires, at least superficially, that $L^*_{\rho_1}$ should have order continuous norm. However, via Proposition 2.1 above, it should be noted that in each of the statement (i), (ii), (iii) of the preceding Theorem 2.4, the operator T' has range in L'_{ρ_1}, which is assumed to have order continuous norm and it is not difficult to see that the proof of Theorem 5.5 of [4] remains valid in this situation.

The above Theorem 2.4 may be viewed as an extension to general integral operators of the criterion given by Luxemburg and Zaanen [13] for order bounded kernels. To indicate some applications, we need some further terminology which is conveniently stated in the setting of Banach lattices.

Let E be a Banach lattice and let $1 \leqq p \leqq \infty$. E is said to have the *l^p-composition property* if and only if, whenever $\{f_n\}_{n=1}^{\infty} \subseteq E^+$ is a disjoint sequence for which the sequence $\{\|f_n\|\}_{n=1}^{\infty} \in l^p$, it follows that $\sup_n \|f_1 + \ldots + f_n\| < \infty$. The lower index $s(E)$ is then defined by

$$s(E) = \sup\{p \geqq 1 : E \text{ has the } l^p\text{-composition property}\}.$$

E is said to have the *l^p-decomposition property* if and only if $\{\|f_n\|\}_{n=1}^{\infty} \in l^p$ whenever $\{f_n\} \subseteq E^+$ is a disjoint order bounded sequence. The upper index $\sigma(E)$ is then defined by

$$\sigma(E) = \inf\{p \geqq 1 : E \text{ has the } l^p\text{-decomposition property}\}.$$

We require the following properties of indices, for which the reader is referred to [3] and [4]. For any Banach lattice E, $1 \leqq s(E) \leqq \sigma(E) \leqq \infty$. If $\sigma(E) < \infty$ then E has order continuous norm; if $s(E) > 1$, then E^* has order continuous norm. Further, $s(E)^{-1} + \sigma(E^*)^{-1} = 1 = \sigma(E)^{-1} + s(E^*)^{-1}$. If moreover L_ρ is a Banach function space, then also the equalities $s(L_\rho)^{-1} + \sigma(L'_\rho)^{-1} = 1 = \sigma(L_\rho)^{-1} + s(L'_\rho)^{-1}$ are valid.

For a discussion of these indices in the context of Banach function spaces, the reader is referred to the paper of Grobler [7] where it is shown that the

indices introduced above coincide in the case of Orlicz spaces, with the indices of Orlicz-Matuszewska. The relation with the Shimogaki indices has been discussed in [3]. We state now, for easy reference, some results from [4].

Lemma 2.5. *Let E, F be Banach lattices and let $T: E \to F$ be a continuous linear map.*

(i) *If $s(E)>2$, then $\phi(|Tf_n|)\to 0$ as $n\to\infty$ whenever $\{f_n\}_{n=1}^{\infty}\subseteq E^+$ is a norm bounded disjoint sequence and $0\leqq\phi\in F^*$.*

(ii) *If $s(E)>\sigma(F)$, then $\phi_n(Tf_n)\to 0$ as $n\to\infty$ whenever $\{f_n\}_{n=1}^{\infty}\subseteq E^+$ and $\{\phi_n\}_{n=1}^{\infty}\subseteq F^{*+}$ are norm bounded disjoint sequences.*

This is Corollary 7.5 and Theorem 6.2 of [4].

We shall use also the following characterization of norm convergence in a Banach lattice.

Lemma 2.6. *Let E be a Banach lattice.*

(a) *A sequence $\{f_n\}_{n=1}^{\infty}\subseteq E$ is norm convergent to 0 if and only if*

(i) $\lim_{n\to\infty}\phi(|f_n|)=0$ *for each* $\phi\in E^*$

(ii) $\lim_{n\to\infty}\phi_n(f_n)=0$ *for each norm-bounded disjoint sequence* $\{f_n\}_{n=1}^{\infty}$ *in* E^{*+}.

(b) *A sequence $\{\phi_n\}_{n=1}^{\infty}\subseteq E^*$ is norm convergent to 0 if and only if*

(i) $\lim_{n\to\infty}|\phi_n|(f)=0$ *for each* $f\in E$

(ii) $\lim_{n\to\infty}\phi_n(f_n)=0$ *for each norm-bounded disjoint sequence* $\{f_n\}_{n=1}^{\infty}$ *in* E^+.

See, for example, [4] Corollary 2.6, 2.7.

Theorem 2.7. *Let T be a continuous linear map from the Banach function space $L_{\rho_1}\subseteq L^0(Y, \nu)$ into the Banach function space $L_{\rho_2}\subseteq L^0(X, \mu)$. Assume that $\sigma(L_{\rho_2})<s(L_{\rho_1})$.*

(i) *If T is an integral operator and if $\sigma(L_{\rho_2})<2$, then T is compact.*

Assume that T' maps L'_{ρ_2} into the associate space L'_{ρ_1}.

(ii) *If T' is an integral operator and $s(L_{\rho_1})>2$, then T is compact.*

(iii) *If both T and T' are integral operators, then T is compact.*

Proof. (i) The assumption $\sigma(L_{\rho_2})<s(L_{\rho_1})$ implies $\sigma(L_{\rho_2})<\infty$ and $s(L_{\rho_1})>1$ and so L_{ρ_2} and $L^*_{\rho_1}$ (and hence L'_{ρ_1}) have order continuous norms. Consider now assertion (i) and let $\{\phi_n\}_{n=1}^{\infty}$ be a norm bounded disjointly supported sequence of positive elements of L'_{ρ_2} and let $0\leqq f\in L_{\rho_1}$. From $\sigma(L_{\rho_2})<2$, it follows $s(L'_{\rho_2})>2$ and it follows from Lemma 2.5(i) that $|T'\phi_n|(f)\to 0$ as $n\to\infty$. From Lemma 2.5(ii) and Lemma 2.6(b) it follows that $\rho'_1(T'\phi_n)\to 0$ as $n\to\infty$ and consequently T is compact by Theorem 2.4(i).

The proof (ii) is similar while the proof of statement (iii) follows directly from Lemma 2.5(ii) and part (iii) of Theorem 2.4.

In order to make some comments on the preceding Theorem 2.7, we first make the simple observation that if the Banach function space $L_{\rho_1}\subseteq L^0(Y, \nu)$ has order continuous norm and if $1\leqq r<\infty$, then each continuous linear map $T: L_{\rho_1}\to l^r$ is an integral operator. In fact, in this case the Banach dual $L^*_{\rho_1}$ coincides with L'_{ρ_1} and so if $\{e'_n\}_{n=1}^{\infty}$ denotes the usual vector basis of $l^{r'}$, where

$1/r+1/r'=1$, then it is easily checked that T is given by the kernel $T(n, y)=(T'e_n')(y)$ for $(n, y)\in N\times Y$. From this remark, it may be readily seen that Theorem 2.7 contains as a special case some compactness results of Rosenthal [15] for linear operators from $L^p(\mu)$ to $L^r(\nu)$, although it should be mentioned that Rosenthal's results remain valid for linear operators defined on subspaces of $L^p(\mu)$ and our methods do not apply to this setting. We point out also that part (iii) of the preceding Theorem 2.7 is a simultaneous extension to general integral operators of a Theorem of Andô [1] for order bounded integral operators on Orlicz spaces and a result of Pitt [14] for continuous maps on sequence spaces.

It is to be observed that part (i) of the preceding Theorem 2.7 contains as a special case the $\langle 2.1\rangle$-compactness result of Halmos-Sunder [8]. We now give a somewhat different extension of this $\langle 2.1\rangle$-compactness theorem. Our method is essentially that of Korotkov [10], which carries over without difficulty to the Banach function space setting; an order bounded version for L_p-spaces may be found in the monograph [12].

Theorem 2.8. *Let $L_{\rho_1}\subseteq L^0(Y, \nu)$ and L_{ρ_2}, $L_{\rho_3}\subseteq L^0(Y, \mu)$ be Banach function spaces. Assume that L'_{ρ_1} has order continuous norm, that $L_{\rho_2}\subseteq L_{\rho_3}$ and that $s(L_{\rho_2})>\sigma(L_{\rho_3})$. Then every integral operator from L_{ρ_1} into L_{ρ_2} is compact if considered as a map from L_{ρ_1} into L_{ρ_3}.*

Proof. Let $T: L_{\rho_1}\to L_{\rho_2}$ be an integral operator. From Theorem 4.3 of [7], it follows from the assumption that $s(L_{\rho_2})>\sigma(L_{\rho_3})$ that T maps bounded subsets of L_{ρ_1} to sets in L_{ρ_3} which are of uniformly absolutely continuous norm. Since $\sigma(L_{\rho_3})<\infty$, it follows that L_{ρ_3} has order continuous norm. Moreover, by Proposition 2.3(i) T is measure compact. Now it is well known that, since L_{ρ_3} has order continuous norm, a subset of L_{ρ_3} is relatively compact if and only if it is measure compact and of uniformly absolutely continuous norm. The compactness of T now follows.

We now present some examples. The first example shows that it is not possible to omit the assumption that L'_{ρ_1} has order continuous norm in Theorem 2.8.

Example 2.9. Let $X=Y=[0, 1]$ with Lebesgue measure and let

$$T(x, y)=\sum_{n=1}^{\infty}\sin n\pi x\,\chi_{[2^{-n},\,2^{-n+1}]}(y).$$

If $f\in L^1[0, 1]$, then

$$\int|T(x, y)f(y)|\,dy\leqq\sum_{n=1}^{\infty}\int_{2^{-n}}^{2^{-n+1}}|f(y)|\,dy=\int_0^1|f(y)|\,dy=\|f\|_1.$$

Hence $T(x, y)$ defines an (order-bounded) integral operator T from $L^1[0, 1]$ into $L^\infty[0, 1]$. Now let $f_n=2^n\chi_{[2^{-n},\,2^{-n+1}]}$. Then $\|f_n\|_1=1$ and $Tf_n(x)=\sin(n\pi x)$ for $n=1, 2, \ldots$. Now it is easy to see that the sequence $\{\sin(n\pi x): n=1, 2, \ldots\}$ is not compact in measure. It follows that T is not compact in measure and so T is *not* compact as a mapping from $L^1[0, 1]$ into $L^p[0, 1]$ for p with $1\leqq p\leqq\infty$.

The next example shows that the condition $\sigma(L_{\rho_2}) < \min(s(L_{\rho_1}), 2)$ cannot be omitted in Theorem 2.7(i). In this context, however, the reader's attention is drawn to Theorem 2.11 below.

Example 2.10. Let $X = Y = [0, 1]$ with Lebesgue measure. Let r_n denote the n^{th} Rademacher function and define g_n via $g_n(x) = 2^{-(n+1)/s} \chi_{[2^{-n}, 2^{-n+1}]}(x)$ for $x \in [0, 1]$ and $n = 1, 2, \ldots$. For $2 \leqq s < p$ and $f \in L^p[0, 1]$ define

$$Tf(x) = \sum_{n=1}^{\infty} \left(\int_0^1 r_n(y) f(y)\, dy \right) g_n(x).$$

Then T is an integral operator from $L^p[0, 1]$ into $L^s[0, 1]$ which is not compact. Observe that T' is not an integral operator.

It has already been observed that if $T: L_{\rho_1}(Y, \nu) \to L_{\rho_2}(X, \mu)$ is an integral operator, then T' maps the associate space L'_{ρ_2} into the associate space L'_{ρ_1}. That T' need not be an integral operator is shown by Example 2.10. We remark that it is not too difficult to prove that T' is "nearly" an integral operator in the sense that if (X, μ) is a finite measure space, then for each $\varepsilon > 0$ there exists $X_\varepsilon \subseteq X$ with $\mu(X_\varepsilon^c) \leqq \varepsilon$ such that $T' \circ P_{X_\varepsilon}$ is an integral operator from L'_{ρ_2} into L'_{ρ_1}. Here $P_{X_\varepsilon} \cdot f$ denotes $f \cdot \chi_{X_\varepsilon}$. Together with part (iii) of Theorem 2.7, this observation implies the final result of the paper, which in view of parts (i) and (ii) of Theorem 2.7 and the examples is not entirely without interest.

Theorem 2.11. *Let (X, μ) be a finite measure space and let $L_{\rho_1} \subseteq L^0(Y, \nu)$, $L_2 \subseteq L^0(X, \mu)$ be Banach function spaces. Assume that $s(L_{\rho_1}) > \sigma(L_{\rho_2})$ and let $T: L_{\rho_1} \to L_{\rho_2}$ be an integral operator. Then for every $\varepsilon > 0$ there exists $X_\varepsilon \subseteq X$ with $\mu(X_\varepsilon^c) \leqq \varepsilon$ such that $P_{X_\varepsilon} \circ T$ is a compact operator from L_{ρ_1} into L_{ρ_2}.*

References

1. Andô, T.: On compactness of integral operators. Indag. Math. **24**, 235–239 (1962)
2. Dodds, P.G.: 0-weakly compact mappings of Riesz spaces. Trans. Amer. Math. Soc. **214**, 389–402 (1975)
3. Dodds, P.G.: Indices for Banach lattices. Indag. Math. **39**, 73–86 (1977)
4. Dodds, P.G., Fremlin, D.H.: Compact operators in Banach lattices. Israel J. Math. **34**, 287–320 (1979)
5. Dunford, N., Schwartz, J.: Linear Operators, Part I. New-York: Interscience 1964
6. Grothendieck, A.: Sur les applications linéaires faiblement compactes d'espaces du type $C(K)$. Canad. J. Math. **5**, 129–173 (1953)
7. Grobler, J.J.: Indices for Banach function spaces. Math. Z. **145**, 99–109 (1975)
8. Halmos, P.R., Sunder, V.S.: Bounded integral operators on L^2-spaces. Berlin-Heidelberg-New York: Springer 1978
9. Krein, S.G.: Functional Analysis, 2nd Ed. Moscow: Nauka 1972 [Russian]
10. Korotkov, V.B.: Strong Integral Operators. Mat. Zametki **16**, 907–912 (1974) [Russian]. Engl. Transl.: Math. Notes **16**, 1137–1140 (1974)
11. Korotkov, V.B.: Integral and partially integral operators. Sibirsk. Mat. Ž. **19**, 70–90 (1978) [Russian]. Engl. Transl.: Siberian Math. J. **19**, 48–63 (1978)
12. Krasnoselskii, M.A. et al.: Integral operators in spaces of summable functions. Leyden: Noordhoff 1976
13. Luxemburg, W.A.J., Zaanen, A.C.: Compactness of integral operators in Banach function spaces. Math. Ann. **149**, 150–180 (1963)

14. Pitt, H.R.: A note on bilinear forms. J. London Math. Soc. **11**, 174–180 (1936)
15. Rosenthal, H.P.: On quasi-complemented subspaces of Banach spaces, with an appendix on compactness of operators from $L^p(\mu)$ to $L^r(\nu)$. J. Functional Analysis **2**, 176–214 (1969)
16. Schaefer, H.H.: Banach lattices and Positive operators. Berlin-Heidelberg-New York: Springer 1974
17. Schep, A.R.: Kernel operators. Nederl. Akad. Wetensch. Proc. Ser. A **82**, 39–53 (1979)
18. Schep, A.R.: On compactness of integral operators. In: Measure Theory and its Applications Proceedings. (Dekalb 1980). Dekalb: Northern Illinois University 1981
19. Zaanen, A.C.: Integration. Amsterdam: North-Holland 1967

Received July 2, 1981; received in final form January 4, 1982

Math. Z. 180, 257–263 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Ricci Curvature of Left Invariant Metrics on Solvable Unimodular Lie Groups

Isabel Dotti Miatello

Departamento de Matemática, Universidade Federal de Pernambuco, 50000 Recife, Brasil

Introduction

In studying curvature properties of left invariant metrics on a Lie group, many results have been obtained in the past years. Most of them are contained in a survey article by Milnor [4].

It remains open the classification of Lie groups admitting left invariant metrics with all Ricci curvatures <0 or $\leqq 0$. In this paper we prove

Theorem 3.1. *A solvable unimodular Lie group G with a left invariant metric has all Ricci curvatures $\leqq 0$ if and only if its derived Lie algebra $[\mathfrak{g}, \mathfrak{g}]$ is commutative, $[\mathfrak{g}, \mathfrak{g}]^{\perp}$ is a commutative subalgebra and the linear transformation ad_b is normal for every $b \in [\mathfrak{g}, \mathfrak{g}]^{\perp}$.*

We show that the above condition on ad_b, $b \in [\mathfrak{g}, \mathfrak{g}]^{\perp}$, is equivalent to the metric being Ricci flat on $[\mathfrak{g}, \mathfrak{g}]$. Hence we obtain

Corollary 3.3. *No solvable unimodular Lie group admits a left invariant metric of strictly negative Ricci curvature.*

It follows from Azencott and Wilson [1] that a left invariant metric on a unimodular solvable Lie group is either flat or has both positive and negative sectional curvatures. An analogous result does not hold for Ricci curvature as the example of $E(1, 1)$ shows (see [4], Sect. 4). However, related to this problem, as a consequence of Theorem 3.1, we obtain

Corollary 3.4. *A unimodular solvable Lie group such that its derived Lie algebra is not commutative has for every left invariant metric directions of strictly positive Ricci curvature and directions of strictly negative Ricci curvature.*

As a final application we find all possible signatures of the self adjoint Ricci transformation on a two step solvable unimodular Lie group and provide examples where those signatures occur.

1. Notation

Let G be an n-dimensional Lie group with Lie algebra $\mathfrak{g}$ consisting of all smooth vector fields invariant under left translations.

0025-5874/82/0180/0257/$0.1.40

If $\langle\ ,\ \rangle$ denotes a left invariant Riemannian metric on G, we recall that for x, y, z in $\mathfrak{g}$, the Riemannian connection and Riemann curvature tensor associated with $\langle\ ,\ \rangle$ are given by

$$\langle \nabla_x y, z\rangle = \tfrac{1}{2}\{\langle [x, y], z\rangle - \langle [y, z], x\rangle + \langle [z, x], y\rangle\},$$

$$R_{xy}(z) = \nabla_{[x,y]} z - \nabla_x \nabla_y z + \nabla_y \nabla_x z.$$

The self-adjoint Ricci transformation $\hat{r}$ of $\mathfrak{g}$ is defined by

$$\hat{r}(x) = \sum R_{e_i x}(e_i),$$

where $e_1, \ldots, e_n$ can be any orthonormal basis of $\mathfrak{g}$, and if u is a unit vector in $\mathfrak{g}$, then

$$r(u) = \langle \hat{r}(u), u\rangle$$

is the Ricci curvature in the direction of u.

2. Curvature Formula

For computational purposes it will be more convenient to work with the following expression of the Ricci curvature:

$$r(u) = -\tfrac{1}{2}\operatorname{tr} \operatorname{ad}_u^2 - \tfrac{1}{2}\operatorname{tr} \operatorname{ad}_u^t \operatorname{ad}_u + \tfrac{1}{2}\sum_{i<j} \langle [e_i, e_j], u\rangle^2 - \operatorname{tr} \operatorname{ad}_{\nabla_u u}, \tag{2.1}$$

where $e_1, \ldots, e_n$ denotes an orthonormal basis of $\mathfrak{g}$ and $\operatorname{ad}_u(x)$ is the Lie bracket $[u, x]$.

To obtain the above formula we recall that

$$\nabla_u - \nabla u = \operatorname{ad}_u,$$

$$\langle \nabla_u x, y\rangle + \langle x, \nabla_u y\rangle = 0$$

for all u, x, y in $\mathfrak{g}$. Using the above properties, an easy computation shows that

$$r(u) = -\operatorname{tr}(\nabla u)^2 - \operatorname{tr} \operatorname{ad}_{\nabla_u u}.$$

From the connexion expression we have

$$\langle \nabla_{e_i} u, e_j\rangle = \tfrac{1}{2}\{-\langle \operatorname{ad}_u e_i, e_j\rangle - \langle \operatorname{ad}_u e_j, e_i\rangle + \langle [e_j, e_i], u\rangle\}$$

therefore

$$\begin{aligned}
\operatorname{tr}(\nabla u)^2 &= \sum_{i,j} \langle \nabla_{e_i} u, e_j\rangle \langle \nabla_{e_j} u, e_i\rangle \\
&= \tfrac{1}{4}\sum_{i,j} (\langle \operatorname{ad}_u e_i, e_j\rangle + \langle \operatorname{ad}_u e_j, e_i\rangle)^2 - \tfrac{1}{4}\sum_{i,j} \langle [e_i, e_j], u\rangle^2 \\
&= \tfrac{1}{2}\operatorname{tr} \operatorname{ad}_u^2 + \tfrac{1}{2}\operatorname{tr} \operatorname{ad}_u^t \operatorname{ad}_u - \tfrac{1}{2}\sum_{i<j} \langle [e_i, e_j], u\rangle^2,
\end{aligned}$$

thus obtaining expression 2.1.

3. Ricci Curvature of Solvable, Unimodular Lie Groups

We recall that a connected Lie group G is unimodular if the linear transformation ad_x has trace zero for every x in the associated Lie algebra $\mathfrak{g}$.

Let $\mathfrak{g}$ denote the Lie algebra of a connected, solvable unimodular Lie group G. For each u in $[\mathfrak{g},\mathfrak{g}]$ and b in $[\mathfrak{g},\mathfrak{g}]^\perp$ Cartan's criteria for solvability (see [3]) implies that $\operatorname{tr}\mathrm{ad}_u\mathrm{ad}_b=\operatorname{tr}\mathrm{ad}_u^2=0$. Therefore given a metric on $\mathfrak{g}$ with orthonormal basis $v_1,\ldots,v_m$, from 2.1 we have

$$\begin{aligned} r(u)&=-\tfrac{1}{2}\operatorname{tr}\mathrm{ad}_u^t\mathrm{ad}_u+\tfrac{1}{2}\sum_{i<j}\langle[v_i,v_j],u\rangle^2,\\ r(u+b)&=r(u)+r(b)-\operatorname{tr}\mathrm{ad}_u^t\mathrm{ad}_b,\\ r(b)&=-\tfrac{1}{4}\operatorname{tr}(\mathrm{ad}_b+\mathrm{ad}_b^t)^2. \end{aligned} \tag{3.1}$$

In what follows we classify the solvable unimodular Lie groups admitting left invariant metrics with non positive Ricci curvature.

Theorem 3.1. *A solvable unimodular Lie group G with a left invariant metric has all Ricci curvatures $\leqq 0$ if and only if its derived Lie algebra $[\mathfrak{g},\mathfrak{g}]$ is commutative, $[\mathfrak{g},\mathfrak{g}]^\perp$ is a commutative subalgebra and the linear transformation ad_b is normal for every $b\in[\mathfrak{g},\mathfrak{g}]^\perp$.*

Proof. Assume that G has a left invariant metric of non positive Ricci curvature. Decompose its Lie algebra $\mathfrak{g}$ as an orthogonal direct sum of subspaces

$$\mathfrak{g}=[\mathfrak{g},\mathfrak{g}]\oplus[\mathfrak{g},\mathfrak{g}]^\perp=\mathfrak{z}\oplus\mathfrak{h}\oplus[\mathfrak{g},\mathfrak{g}]^\perp,$$

where $\mathfrak{z}$ denotes the center of $[\mathfrak{g},\mathfrak{g}]$.

We may assume $\mathfrak{g}$ is not abelian. Consequently being $[\mathfrak{g},\mathfrak{g}]$ nilpotent its center is non trivial.

Denote by $x_1,\ldots,x_n$; $e_1,\ldots,e_m$; $f_1,\ldots,f_p$ an orthonormal basis of $\mathfrak{z}$, $\mathfrak{h}$ and $[\mathfrak{g},\mathfrak{g}]^\perp$ respectively.

Choosing a unit vector $u\in\mathfrak{z}$ the Ricci curvature becomes

$$r(u)=-\tfrac{1}{2}\sum_k\|[u,f_k]\|^2+\tfrac{1}{2}\sum_{i,k}\langle[x_i,f_k],u\rangle^2+\beta(u)$$

where

$$\beta(u)=\tfrac{1}{2}\sum_{j<s}\langle[e_j,e_s],u\rangle^2+\tfrac{1}{2}\sum_{j,k}\langle[e_j,f_k],u\rangle^2+\tfrac{1}{2}\sum_{k<s}\langle[f_k,f_s],u\rangle^2.$$

Since ad_{f_k} preserves $\mathfrak{z}$,

$$r(u)=-\tfrac{1}{2}\sum_{i,k}\langle[u,f_k],x_i\rangle^2+\tfrac{1}{2}\sum_{i,k}\langle[x_i,f_k],u\rangle^2+\beta(u)$$

and

$$\sum_t r(x_t)=\sum\beta(x_t)\geqq\sum_t\sum_{j<s}\langle[e_j,e_s],x_t\rangle^2.$$

The right hand side equal to zero in the above inequality implies that the center is perpendicular to $[[\mathfrak{g},\mathfrak{g}],[\mathfrak{g},\mathfrak{g}]]$. Hence $[\mathfrak{g},\mathfrak{g}]$ is abelian.

Since

$$0 \geqq \sum_t r(x_t) = \tfrac{1}{2} \sum_{k<s} \langle [f_k, f_s], x_t \rangle^2,$$

the orthogonal complement of $[\mathfrak{g}, \mathfrak{g}]$ is a commutative subalgebra.

In particular, we have also obtained above that $r(u)=0$ for u in $[\mathfrak{g}, \mathfrak{g}]$. To finish the proof we need the following technical lemma (to be proved later).

Lemma 3.2. *If the unimodular Lie algebra* $\mathfrak{g}$ *splits as an orthogonal direct sum*

$$\mathfrak{g} = \mathfrak{u} \oplus \mathfrak{b}$$

where $\mathfrak{u}$ *is an abelian ideal and* $\mathfrak{b}$ *is an abelian subalgebra, then*

(i) *Given an orthonormal basis* $b_1, \ldots, b_n$ *of* $\mathfrak{b}$ *and* $u \in \mathfrak{u}$

$$\hat{r}(u) = \tfrac{1}{2} \sum [L_i, L_i^*](u),$$

where L_i *denotes the transformation* ad_{b_i} *restricted to* $\mathfrak{u}$ *and* L_i^* *its adjoint.*

(ii) *The self adjoint Ricci transformation* $\hat{r} \equiv 0$ *on* $\mathfrak{u}$ *if and only if* ad_b *restricted to* $\mathfrak{u}$ *is a normal transformation for every* $b \in \mathfrak{b}$.

It follows from Lemma 3.2 (i) that the self adjoint Ricci transformation $\hat{r}$ preserves $[\mathfrak{g}, \mathfrak{g}]$. Hence $r(u)=0$, $u \in [\mathfrak{g}, \mathfrak{g}]$ is only possible if $\hat{r} \equiv 0$ on $[\mathfrak{g}, \mathfrak{g}]$. Therefore (ii) of same lemma and the fact that $[\mathfrak{g}, \mathfrak{g}]^\perp$ is commutative imply that ad_b is a normal transformation for every $b \in [\mathfrak{g}, \mathfrak{g}]^\perp$.

The converse follows easily from Lemma 3.2 together with (3.1). In fact, in those Lie groups G satisfying the conditions of the theorem, the Ricci curvature becomes

$$r(u+b) = r(b) = -\tfrac{1}{4} \mathrm{tr}(\mathrm{ad}_b + \mathrm{ad}_b^t)^2 \tag{3.2}$$

where $u \in [\mathfrak{g}, \mathfrak{g}]$ and $b \in [\mathfrak{g}, \mathfrak{g}]^\perp$.

Remarks. If $\mathrm{Ricci} \equiv 0$ on a solvable Lie groups, then all of the ad_b, $b \in [\mathfrak{g}, \mathfrak{g}]^\perp$ are skew adjoint ([4], Lemma 2.3). Since such a group is unimodular ([4], Lemma 6.4) Theorem 3.1 gives a flat decomposition.

It also follows from Theorem 3.1 that any left invariant metric on a nilpotent, non abelian Lie group has directions of strictly positive Ricci curvature (see [4], Theorem 2.4), since the only normal and nilpotent transformation is the trivial one.

Corollary 3.3. *No solvable unimodular Lie group* G *admits a metric of strictly negative Ricci curvature.*

We should remark that the unimodular hypothesis is definitely needed. According to Heintze [2], there exist solvable non unimodular Lie groups with left invariant metrics of negative sectional curvature.

Proof. As noted earlier, in (3.2), a left invariant metric of non positive Ricci curvature, on a solvable unimodular Lie group, must be Ricci flat on $[\mathfrak{g}, \mathfrak{g}]$. Hence, the corollary follows.

Corollary 3.4. *A unimodular solvable Lie group such that its derived Lie algebra is not commutative has for every left invariant metric directions of strictly positive Ricci curvature and directions of strictly negative Ricci curvature* (compare [4], corollary 3.2).

Proof. It follows from Theorem 3.1 that there exist directions of strictly positive Ricci curvature.

It is known (see [4], Theorem 3.1) that any left invariant metric on a solvable Lie group is either flat or else has strictly negative scalar curvature. Since a Lie group as in corollary does not admit a flat metric, there exist directions of strictly negative Ricci curvature.

Corollary 3.5. *A solvable unimodular Lie group G admits a left invariant metric with all Ricci curvatures $\leqq 0$ if and only if its Lie algebra $\mathfrak{g} \simeq \mathfrak{u} \oplus \mathfrak{b}$ where $\mathfrak{u}$ is a commutative ideal, $\mathfrak{b}$ is a commutative subalgebra, and $\mathfrak{u} \otimes C$ admits a basis consisting of common eigenvectors for all of the ad_b, $b \in \mathfrak{b}$.*

Proof. This follows immediately from Theorem 3.1.

Proof of Lemma 3.2. Let $e_1, \ldots, e_n$; $b_1, \ldots, b_m$ be an orthonormal basis of $\mathfrak{u}$ and $\mathfrak{b}$ respectively.

Given $u \in \mathfrak{u}$,

$$\begin{aligned} \hat{r}(u) &= \sum_i R_{e_i u}(e_i) + \sum_j R_{b_j u}(b_j) \\ &= \sum_i \nabla_u \nabla_{e_i} e_i - \nabla_{e_i} \nabla_u e_i + \sum_j \nabla_{[b_j, u]} b_j - \nabla_{b_j} \nabla_u b_j, \end{aligned}$$

since $\mathfrak{u}$ is abelian and $\nabla_x y = 0$, for $x, y \in \mathfrak{b}$.

Let D_j and S_j denote the symmetric and skew-symmetric parts of ad_{b_j} restricted to $\mathfrak{u}$.

After some computations one finds that

$$\sum_i \nabla_u \nabla_{e_i} e_i = -\sum_j (\operatorname{tr} D_j) D_j u = 0 \qquad (\mathfrak{g} \text{ unimodular}),$$

$$\sum_i \nabla_{e_i} \nabla_u e_i = -\sum_{i,j} \langle D_j u, e_i \rangle D_j e_i = -\sum_j D_j^2 u,$$

$$\sum_j \nabla_{[b_j, u]} b_j - \nabla_{b_j} \nabla_u b_j = \sum_j -D_j^2 u + [S_j, D_j] u.$$

Hence

$$\hat{r}(u) = \sum_j [S_j, D_j](u) = \tfrac{1}{2} \sum_j [L_j, L_j^*](u),$$

which completes the proof of (i).

The proof of (ii) is an immediate consequence of the following assertion.

"Let L_i, $i = 1, \ldots, k$ be a commuting family of linear transformations of R^n. Denote by L_i^*, $i = 1, \ldots, k$ the corresponding adjoint transformations with respect to an inner product $\langle\ ,\ \rangle$.

If $\sum_{i=1}^{k} [L_i, L_i^*] = 0$ then each L_i, $i = 1, \ldots, k$, is a normal transformation."

We prove the above result by induction on n. Since the L_i commute, the induced transformations $\tilde{L}_i$ on C^n have a common eigenvector x with eigenvalue λ_i, $i = 1, \ldots, k$.

The new family $\tilde{L}_i - \lambda_i I$, $i = 1, \ldots, k$ satisfy

$$\sum_{i=1}^{k} [\tilde{L}_i - \lambda_i I, (\tilde{L}_i - \lambda_i I)^*] = 0.$$

(Here we are considering C^n with the Hermitian inner product induced by $\langle\ ,\ \rangle$.)

Then

$$0=\left\langle \sum_{i=1}^{k} [\tilde{L}_i-\lambda_i I,(\tilde{L}_i-\lambda_i I)^*]\, x, x \right\rangle$$

$$=\sum_{i=1}^{k} \|(\tilde{L}_i-\lambda_i I)^* x\|^2$$

which shows that x is common eigenvector of L_i^*, $i=1,\ldots,k$. Therefore, we may restrict L_i to the orthogonal complement of x and by inductive hypothesis (since it is true for $n=1$) the assertion follows.

4. Signature of the Self Adjoint Ricci Transformation

Given a left invariant metric on a Lie group G, let us denote by n, p and z the number of negative, positive and zero eigenvalues, respectively, of the self adjoint Ricci transformation $\hat{r}$.

If G is solvable and the metric is not flat, then $n>0$ ([4], Theorem 3.1). It follows from Corollary 3.3 that on solvable unimodular Lie groups $p+z>0$.

Proposition 4.1. *Given any signature with $p>0$ and $n>1$ or $p=0$ and $z>1$ there exists a two step solvable unimodular Lie group and a left invariant metric such that the signature of $\hat{r}$ is the given one. Conversely, these are the only possible signatures on a two step solvable unimodular Lie group of dimension >1.*

Proof. If $p>0$, we may consider in R^{z+p+1} (with usual metric) a self adjoint transformation U of trace zero and signature

$$(\lambda_n, \mu_1, \ldots, \mu_p, \delta_1, \ldots, \delta_z), \qquad \lambda_n<0, \qquad \mu_j>0, \qquad \delta_r=0.$$

It is not hard to verify that there exists a trace zero transformation T in R^{z+p+1} such that $[T, T^*]=U$.

Build now the Lie algebra

$$\mathfrak{g}=R^{z+p+1}\oplus R^{n-1}$$

where bracket is defined by being zero on R^{z+p+1} and R^{n-1} and by $[e_i, x]=Tx$, $\{e_i\}$ $i=1,\ldots,n-1$ canonical basis of R^{n-1}.

Endow $\mathfrak{g}$ with the product metric, euclidean on each factor. It follows from Lemma 3.2 that

$$\hat{r}(u)=(z+p+1)\,[T, T^*]\,(u)=(z+p+1)\,U(u), \qquad u\in R^{z+p+1}$$

and from [4], Lemma 2.3 that $r(x)<0$, $x\in R^{n-1}$, thus obtaining the desired signature.

If $p=0$, a similar argument allows us to obtain any signature with $z>1$ and $n>1$. It suffices to consider any non trivial, trace zero, self adjoint transformation T in R^z.

We prove the converse by induction. It is true if dimension of G is 2. We recall that for $\dim G=3$, the possible signatures are $(0, 0, 0)$, $(-, 0, 0)$, $(-, -, +)$ (see [4], Sect. 4).

Assume the converse is true for dimension k. Let G be a two step solvable unimodular Lie group of dimension $k+1$ with left invariant metric. If $\mathfrak{g}$ denotes its Lie algebra and $\langle b, [\mathfrak{g}, \mathfrak{g}]\rangle=0$ we denote by $\mathfrak{u}$ the orthogonal complement of b, by L the transformation $\mathrm{ad}_{b|u}$ and by L^* its adjoint.

Following computations made in [4], Lemma 5.5, it is not hard to obtain

$$\hat{r}(u)=\hat{r}_{\mathfrak{u}}(u)-\tfrac{1}{2}\mathrm{tr}[(L+L^*)\,\mathrm{ad}_u]\,b+\tfrac{1}{2}[L, L^*](u),$$

where $u\in\mathfrak{u}$ and $\hat{r}_{\mathfrak{u}}$ denote the self adjoint Ricci transformation of $\mathfrak{u}$ with the restricted metric.

If there exists b perpendicular to $[\mathfrak{g}, \mathfrak{g}]$ with $r(b)=0$, it follows from Lemma 2.3 [4] that L is skew adjoint. Thus, $\hat{r}$ preserves $\mathfrak{u}$ and by inductive hypothesis we conclude the proof in this case.

Assume now that $r(b)<0$ for every non zero b perpendicular to $[\mathfrak{g}, \mathfrak{g}]$.

If dimension $[\mathfrak{g}, \mathfrak{g}]^{\perp}\geqq 2$ there exist at least two negative eigenvalues of $\hat{r}$. According to Corollary 3.3, $p+z>0$. Hence, if $p>0$ we obtain one of the desired signatures. If $p=0$, $r(x)\leqq 0$ for every $x\in\mathfrak{g}$. Applying Theorem 3.1 and Lemma 3.2 we obtain zero eigenvalues on $[\mathfrak{g}, \mathfrak{g}]$. Since $\dim[\mathfrak{g}, \mathfrak{g}]>1$ (otherwise $\mathfrak{g}$ would be commutative) it follows that $z>1$.

If dimension $[\mathfrak{g}, \mathfrak{g}]^{\perp}=1$, $\hat{r}$ preserves $\mathfrak{u}$ since $\mathfrak{u}=[\mathfrak{g}, \mathfrak{g}]$ is commutative. Therefore, for $u\in\mathfrak{u}$

$$\hat{r}(u)=\tfrac{1}{2}[L, L^*](u),$$

which completes the proof. We should remark that only here we have used that G is two step solvable. I don't know whether this proposition is true for solvable unimodular Lie groups.

References

1. Azencott, R., Wilson, E.: Homogeneous manifolds with negative curvature, Part I. Trans. Amer. Math. Soc. **215**, 323-362 (1976)
2. Heintze, E.: On homogeneous manifolds of negative curvature. Math. Ann. **211**, 23-34 (1974)
3. Humphreys, J.: Introduction to Lie Algebras and Representation Theory. New York-Heidelberg-Berlin: Springer 1972
4. Milnor, J.: Curvatures of left invariant metrics on Lie groups. Advances in Math. **21**, 293-329 (1976)

Received August 24, 1981; received in final form January 8, 1982

Math. Z. 180, 265–273 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Superconvexity of the Spectral Radius, and Convexity of the Spectral Bound and the Type

Tosio Kato

Department of Mathematics, University of California, Berkeley, California 94720, U.S.A.

1. Introduction

One of the objects of this paper is to generalize a theorem of Cohen [2] (see also Friedland [5]) to the effect that the *principal eigenvalue* of an essentially nonnegative matrix (a real $n \times n$ matrix with all off-diagonal elements nonnegative) is a *convex* function of its diagonal elements.

The simplest proof of Cohen's theorem seems to be the one based on a theorem of Kingman [12] to the effect that the *spectral radius* of a matrix whose elements are *superconvex* functions of a parameter h is likewise superconvex in h. (We say a real-valued function $\phi(h)$ is superconvex in h if either $\phi(h)=0$ identically or $\phi(h)>0$ and $\log \phi(h)$ is convex in h, deviating slightly from the usage in [12]). Therefore, we want to generalize Kingman's theorem too.

The intended generalizations are to linear operators in an ordered Banach space X. Naturally, the spectral radius, denoted by $\operatorname{spr} T$ in the sequel, is defined only for $T \in B(X)$ (the set of all bounded linear operators on X to X). But analogs of the principal eigenvalue may be defined for an unbounded operator A in X, which is assumed to be closed for technical convenience. A candidate for such an analog is the *spectral bound*, $\operatorname{spb} A$ in symbol, given by

$$\operatorname{spb} A = \sup \operatorname{Re} \sigma(A) = \sup \{\operatorname{Re} \lambda;\ \lambda \in \sigma(A)\}, \tag{1.1}$$

where $\sigma(A)$ is the spectrum of A. If $\sigma(A)$ is empty, we set $\operatorname{spb} A = -\infty$.

Another candidate is the *type* of the semigroup $\{e^{tA};\ t>0\}$, denoted by $\operatorname{type} A$, in case A is the infinitesimal generator of a strongly continuous semigroup. Thus

$$\operatorname{type} A = \lim_{t \to \infty} t^{-1} \log \|e^{tA}\| = \log \operatorname{spr} e^{A}. \tag{1.2}$$

Note that $\operatorname{type} A$ is defined only for a generator A, while $\operatorname{spb} A$ is defined for any closed linear operator A.

0025-5874/82/0180/0265/$01.80

It is well known that

$$-\infty \leqq \operatorname{spb} A \leqq \operatorname{type} A < +\infty \tag{1.3}$$

whenever A is a generator in a rather wide class (see Hille-Phillips [8; p. 352]). It is not so well known, however, under what conditions the equality $\operatorname{spb} A = \operatorname{type} A$ holds, although it is shown in [8; p. 460] that this is the case if e^{tA} is norm-continuous for $t \geqq t_0 > 0$. Cf. also the discussions in Greiner-Voigt-Wolff [6] and Kato [11].

The link between the spectral radius and the type is given in (1.2). A similar relationship between the spectral radius and the spectral bound is given by

$$\operatorname{spb} A = \lim_{\xi \to +\infty} [\xi^2 \operatorname{spr}(\xi - A)^{-1} - \xi] \quad \text{if } \operatorname{spb} A < +\infty. \tag{1.4}$$

The proof of (1.4) is straightforward and may be omitted.

We note that the facts summarized above are valid in any Banach space X, independently of any possible order relations in X.

We now consider the convexity of $\operatorname{spb} A(h)$ or $\operatorname{type} A(h)$, when $A(h)$ is a linear operator in X depending on a parameter h. It is easy to see from (1.2) and (1.4) that the convexity in question is a consequence of the superconvexity of $\operatorname{spr}(\xi - A(h))^{-1}$ or $\operatorname{spr} e^{A(h)}$. Thus it is clear that a generalization of Kingman's theorem should be useful for the purpose.

At this point we shall have to introduce an order relation into X, since Kingman's theorem is essentially tied to positivity.

In the following sections we prove general theorems related to the convexity and superconvexity for the families considered. Applications to concrete problems, including some differential operators, will be given in the last section.

Acknowledgment. The writer thanks W. Arendt and S. Friedland for useful discussions and suggestions, and Mathematics Research Center, University of Wisconsin, where part of this work was done. This work was partially supported by National Science Foundation Grant MCS-79-02578.

2. Superconvexity

In what follows X is either an ordered real Banach space or its complexification, with positive cone X_+ (which is as usual assumed to be proper and closed). Since there is no essential difference between the two cases, we shall consider only the complex space X.

We denote by X^* the dual (Banach adjoint) space of X, with the associated positive cone X_+^*. B, B_+, B^*, and B_+^* denote, respectively, the unit balls in X, X_+, X^*, and X_+^*. An operator $T \in B(X)$ is said to be positive if $TX_+ \subset X_+$. T is positive if and only if the adjoint operator $T^* \in B(X^*)$ is positive.

Throughout this paper we assume that X_+ is generating and normal. This means that $X_+ - X_+$ is identical with the underlying real space of X and that $\|x+y\| \geqq c > 0$ for $x, y \in X_+$ with $\|x\| = \|y\| = 1$. The following lemma is a direct consequence of this assumption (see Schaefer [14, p. 222]).

Lemma 2.1. *For any positive operator T on X, one has*

$$\|T\| \leqq c \sup_{x \in B_+,\, x^* \in B_+^*} \langle Tx, x^* \rangle, \tag{2.1}$$

where $c>0$ is a constant, and

$$\operatorname{spr} T = \lim_{n \to \infty} [\sup_{x \in B_+,\, x^* \in B_+^*} \langle T^n x, x^* \rangle]^{1/n}. \tag{2.2}$$

We now define vector-valued and operator-valued superconvex functions in a parameter h. In what follows h is assumed to vary on a closed finite interval of real numbers. But the results are immediately generalized to a parameter h varying on any convex set H, since convexity and superconvexity of functions on H can be characterized by their behavior on each segment in H.

We say a vector-valued function $x(h) \in X$ is *superconvex* in h if for each $\varepsilon>0$ and each triplet $h_1 \leqq h_0 \leqq h_2$, there are a finite number of vectors $x_j \in X_+$ and real-valued superconvex functions (see Introduction) $\phi_j(h)$, $j=1, 2, \ldots, m$, such that

$$\left\| x(h_k) - \sum_{j=1}^{m} \phi_j(h_k) x_j \right\| \leqq \varepsilon \quad \text{for } k=0, 1, 2. \tag{2.3}$$

We say an operator-valued function $T(h) \in B(X)$ is *superconvex* if $T(h)x$ is superconvex for each fixed $x \in X_+$.

Lemma 2.2. (a) *If $x(h) \in X$ is superconvex, then $x(h) \in X_+$ and $\langle x(h), x^* \rangle$ is a real-valued superconvex function for each $x^* \in X_+^*$.*

(b) *If $x(h)$ and $y(h)$ are superconvex, so is $ax(h)+by(h)$ for $a, b \in \mathbb{R}_+$.*

(c) *If $x_n(h)$, $n=1, 2, \ldots$, are superconvex and $x(h) = \lim_{n \to \infty} x_n(h)$ exists for each h pointwise, then $x(h)$ is superconvex.*

Proof. (a) $x(h) \in X_+$ follows from (2.3), which shows that $x(h)$ is approximated by elements of X_+. To prove that $\langle x(h), x^* \rangle$ is superconvex in h, fix $h_1 \leqq h_2$ and set $h_0 = \theta_1 h_1 + \theta_2 h_2$, where $\theta_1, \theta_2 \geqq 0$ with $\theta_1 + \theta_2 = 1$. We have to show that

$$0 \leqq \langle x(h_0), x^* \rangle \leqq \langle x(h_1), x^* \rangle^{\theta_1} \langle x(h_2), x^* \rangle^{\theta_2}. \tag{2.4}$$

With h_1, h_2, h_0 and $\varepsilon>0$ thus fixed, let the x_j and ϕ_j be as in (2.3). Then

$$\left| \langle x(h_k), x^* \rangle - \sum_{j=1}^{m} \phi_j(h_k) \langle x_j, x^* \rangle \right| \leqq \varepsilon \|x^*\| \tag{2.5}$$

for $k=0, 1, 2$. Since $\sum \phi_j(h) \langle x_j, x^* \rangle$ is superconvex in h as a positive combination of superconvex functions (see [12]), we see that (2.4) is true up to an error of order ε. On letting $\varepsilon \to 0$, we thus obtain (2.4).

(b) is obvious.

(c) Fix $\varepsilon>0$ and h_k, $k=0, 1, 2$. Choose an n such that $x_n(h_k)$ approximates $x(h_k)$ within ε for $k=0, 1, 2$. Then choose the x_j and ϕ_j satisfying (2.3) for x replaced with x_n. Then (2.3) follows with ε replaced with 2ε. Hence $x(h)$ is superconvex in h.

Lemma 2.3. *If $x(h)\in X$ and $T(h)\in B(X)$ are superconvex in h, so is $T(h)x(h)\in X$.*

Proof. Let $\varepsilon>0$ and h_k, $k=0,1,2$, be given. Choose the x_j and ϕ_j satisfying (2.3) with ε replaced by $\varepsilon/2M$, where $M=\max_k \|T(h_k)\|$. Then

$$\left\|T(h_k)x(h_k)-\sum_{j=1}^{m}\phi_j(h_k)\,T(h_k)\,x_j\right\|\leqq\varepsilon/2 \qquad (k=0,1,2).$$

Since $T(h)x_j\in X$ is superconvex in h for each j, there are $x_{ji}\in X_+$ and real-valued superconvex functions ϕ_{ji}, $i=1,\dots,m_j$, such that

$$\left\|T(h_k)x_j-\sum_{i=1}^{m_j}\phi_{ji}(h_k)\,x_{ji}\right\|\leqq\varepsilon/2mN \qquad (k=0,1,2),$$

where $N=\max_{j,k}\phi_j(h_k)$. Then

$$\|T(h_k)x(h_k)-\sum_j\sum_i\phi_j(h_k)\,\phi_{ji}(h_k)\,x_{ji}\|\leqq\varepsilon$$

for $k=0,1,2$. Since the $\phi_j\phi_{ji}$ are superconvex (see [12]), we have proved the lemma.

Corollary 2.4. (a) *If $T(h)\in B(X)$ is superconvex, then $T(h)$ is positive and $\langle T(h)x, x^*\rangle$ is a real-valued superconvex function for each $x\in X_+$ and $x^*\in X_+^*$.*

(b) *If $T(h)$ and $S(h)$ are superconvex, so are $aT(h)+bS(h)$ and $T(h)S(h)$, where $a, b\in\mathbb{R}_+$.*

(c) *If $T_n(h)$, $n=1,2,\dots$, are superconvex and $T(h)=s-\lim_{n\to\infty}T_n(h)$ exists for each h pointwise, then $T(h)$ is superconvex.*

Theorem 2.5 (generalized Kingman's theorem). *If $T(h)\in B(X)$ is superconvex in h, so is* spr $T(h)$.

Proof. According to Corollary 2.4, (b), $T(h)^n$ is superconvex in h for $n=1,2,\dots$. By (a) of the same Corollary, $\langle T(h)^n x, x^*\rangle$ is superconvex for each $x\in X_+$ and $x^*\in X_+^*$. If we take the supremum of this function for $x\in B_+$ and $x^*\in B_+^*$ and then take its n-th root and go to the limit $n\to\infty$, superconvexity is preserved (see [12]). In view of (2.2), we conclude that spr $T(h)$ is superconvex.

Remark 2.6. We may say a vector-valued function $x(h)$ is *weakly superconvex* in h if $\langle x(h), x^*\rangle$ is superconvex for each $x^*\in X_+^*$, and accordingly define weak superconvexity for an operator-valued function $T(h)$. Lemma 2.2, (a) and Corollary 2.4, (a), show that superconvexity implies weak superconvexity in either case. But in general weak superconvexity of $T(h)$ and $S(h)$ is not likely to imply that of $T(h)S(h)$, a crucial property needed in the proof of Kingman's theorem.

On the other hand, weak superconvexity does imply superconvexity if X_+ has a certain positive approximation property (cf. Walsh [15]). Since this would impose further restrictions on X_+, however, we prefer to use the "strong" definition of superconvexity given above (which has been suggested to the author by S. Friedland).

3. Resolvent-Positive Operators and Resolvent Superconvex Families

Let A be a closed linear operator in X. We say A is *resolvent-positive* if the resolvent $(\xi-A)^{-1}\in B(X)$ of A exists and is positive for all sufficiently large real number ξ. This implies that $\operatorname{spb} A<+\infty$ (for the definition of the spectral bound $\operatorname{spb} A$ see Introduction) if A is resolvent-positive. Moreover, in view of the resolvent equation

$$(\xi-A)^{-1}-(\xi'-A)^{-1}=(\xi'-\xi)(\xi'-A)^{-1}(\xi-A)^{-1} \tag{3.1}$$

and the assumptions on X_+, one concludes that

$$\operatorname{spb} A\in\sigma(A) \quad \text{if } \sigma(A) \text{ is nonempty,} \tag{3.2}$$

(cf. [6]) and that $(\xi-A)^{-1}$ is positive and decreasing in ξ for $\operatorname{spb} A<\xi<+\infty$.

We say a family $A(h)$ of closed operators is *resolvent-superconvex* in h if $A(h)$ is resolvent-positive with $\operatorname{spb} A(h)$ bounded from above and $(\xi-A(h))^{-1}$ is superconvex in h for sufficiently large ξ.

Theorem 3.1. *If $A(h)$ is resolvent-superconvex in h, then* $\operatorname{spb} A(h)$ *is convex in h. In particular if* $\operatorname{spb} A(h)=-\infty$ *for some h, it is identically equal to* $-\infty$.

Proof. We have only to apply (1.4) to $A=A(h)$. Since $(\xi-A(h))^{-1}$ is superconvex in h by hypothesis, its spectral radius is superconvex in h by Theorem 2.5. Hence it is convex a fortiori, and $\xi^2\operatorname{spr}(\xi-A(h))^{-1}-\xi$ is also convex in h. The same is true of its limit $\operatorname{spb} A(h)$ as $\xi\to+\infty$.

Remark 3.2. If $A(h)$ is a family of resolvent-positive operators with $\operatorname{spb} A(h)$ bounded from above, $\operatorname{spb} A(h)$ is convex if and only if $\operatorname{spr}(\xi-A(h))^{-1}$ is superconvex. Indeed, we have $\operatorname{spr}(\xi-A(h))^{-1}=(\xi-\operatorname{spb} A(h))^{-1}$ by (3.2). But it is easy to see that a real-valued function $\phi(h)$ bounded from above is convex if and only if $(\xi-\phi(h))^{-1}$ is superconvex for sufficiently large ξ.

Theorem 3.3 (perturbation theorem). *Let $A(h)$ and $B(h)$ be resolvent-superconvex in h. Assume, moreover, that $B(h)\in B(X)$ with $\|B(h)\|$ bounded. Then $A(h)+B(h)$ is also resolvent-superconvex in h. (Hence* $\operatorname{spb}[A(h)+B(h)]$ *is convex.)*

The proof of Theorem 3.3 will be given after a similar perturbation theorem for generators has been proved. The following section contains preliminary material for that purpose.

4. Yosida Approximations to a Resolvent-Positive Operator

Let A be a resolvent-positive operator in X, with $\operatorname{spb} A=s<+\infty$. Then $(\xi-A)^{-1}$ is positive for $\xi>s$ (see the previous section). In particular

$$J_n=(1-n^{-1}A)^{-1}=n(n-A)^{-1}\in B(X) \tag{4.1}$$

is defined and is positive for $n>s_0=\max\{s,0\}$. The Yosida approximations

$$A_n=AJ_n=n(J_n-1)\in B(X), \quad (n-[s_0]=1,2,\ldots) \tag{4.2}$$

are essentially positive (positive modulo a scalar operator).

Lemma 4.1. *The A_n are resolvent-positive, uniformly in $n>s_0$, and*

$$\|(\xi-A_n)^{-1}-(\xi-A)^{-1}\|\to 0 \quad (n\to\infty) \quad \text{for each } \xi>s_0. \tag{4.3}$$

Proof. A simple algebra gives

$$(\xi-A_n)^{-1}=(n+\xi)^{-1}+n^2(n+\xi)^{-2}[n\xi(n+\xi)^{-1}-A]^{-1}, \tag{4.4}$$

which exists and is positive if

$$n\xi(n+\xi)^{-1}>s_0 \quad \text{i.e.} \quad \xi>ns_0(n-s_0)^{-1}. \tag{4.5}$$

Thus A_n is resolvent-positive, uniformly in n since $ns_0(n-s_0)^{-1}\to s_0$ as $n\to\infty$. If $\xi>s_0$ is fixed, (4.4) is true for $n>\xi s_0(\xi-s_0)^{-1}$ and (4.3) follows easily from (4.4).

Lemma 4.2. *If $A(h)$ is resolvent-superconvex in h, the associated Yosida approximations $A_n(h)$ are also resolvent-superconvex.*

Proof. Obvious from (4.4).

5. Semigroup-Positive Operators and Semigroup-Superconvex Families

In what follows we mean by a *generator* the infinitesimal generator A of a semigroup $\{e^{tA}; t\geqq 0\}$ of class C_0 in the sense of Hille-Phillips [8] (which we call simply a semigroup). (Some of the following results are valid for more general semigroups of class $(0, A)$.)

If A is a generator, type A is defined by (1.2) and satisfies (1.3). The following relations between e^{tA} and the resolvent $(\xi-A)^{-1}$ are well known.

$$(\xi-A)^{-1}=\int_0^\infty e^{-\xi t}e^{tA}\,dt \quad (\xi>\text{type } A) \tag{5.1}$$

$$e^{tA}=s-\lim_{n\to\infty}(n/t)^n(n/t-A)^{-n} \quad (t>0). \tag{5.2}$$

We say a generator A is *semigroup-positive* if e^{tA} is positive for each $t>0$. A family $A(h)$ of generators is *semigroup-superconvex* if $e^{tA(h)}$ is superconvex in h for each $t>0$ with type $A(h)$ bounded from above.

Lemma 5.1. *A generator A is semigroup-positive if and only if it is resolvent-positive.*

Proof. This is well known (see [8, p. 353]), and is a direct consequence of (5.1) and (5.2).

Theorem 5.2. *If $A(h)$ is a family of generators, the following conditions are equivalent.*

(a) *$A(h)$ is resolvent-superconvex in h.*
(b) *$A(h)$ is semigroup-superconvex in h.*

When these conditions are met, both spb $A(h)$ *and* type $A(h)$ *are convex in h.*

Proof. The equivalence of (a) and (b) is again a consequence of (5.1) and (5.2), as is easily seen by Corollary 2.4. Indeed, (5.1) shows that $(\xi-A(h))^{-1}$ is the

limit of certain positive combinations of $e^{tA(h)}$ for different t's, and (5.2) shows that $e^{tA(h)}$ is the limit of positive number times powers of $(\xi - A(h))^{-1}$.

The convexity of spb $A(h)$ was already proved (Theorem 3.1). Also, type $A(h) = \log \operatorname{spr} e^{A(h)}$ (see (1.2)) is convex in h because spr $e^{A(h)}$ is superconvex by Theorem 2.5.

Theorem 5.3 (perturbation theorem). *Let $A(h)$ and $B(h)$ be semigroup-superconvex families of generators. Assume, moreover, that $B(h) \in B(X)$ with $\|B(h)\|$ bounded. Then $A(h)+B(h)$ is also a semigroup-superconvex family of generators. (Hence* spb $[A(h)+B(h)]$ *and* type $[A(h)+B(h)]$ *are convex in h.)*

Proof. Under the assumptions, it is known that $A(h)+B(h)$ is also a generator and the product formula

$$e^{t[A(h)+B(h)]} = s - \lim_{n \to \infty} [e^{tA(h)/n} e^{tB(h)/n}]^n \tag{5.3}$$

holds (see e.g. Chernoff [1]). Here $e^{tA(h)/n}$ and $e^{tB(h)/n}$ are both superconvex in h by hypothesis, hence their product and its n-th power are superconvex by Corollary 2.4, (b), and the same is true of the limit (same Corollary, (c)). Thus $A(h)+B(h)$ is semigroup-superconvex.

We can now prove Theorem 3.3 using Theorem 5.3. Let $A_n(h)$ be the Yosida approximations for $A(h)$ given by (4.2), where $n > \max\{0, \text{spb } A(h)\}$ (recall that spb $A(h)$ is by definition bounded from above). Then $A_n(h)$ and $B(h)$ are both generators and resolvent-superconvex (see Lemma 4.2). Hence they are semigroup-superconvex by Theorem 5.2, and $A_n(h)+B(h)$ is semigroup-superconvex by Theorem 5.3.

On the other hand, $(\xi - A_n(h))^{-1} \to (\xi - A(h))^{-1}$ for sufficiently large ξ (see Lemma 4.1). It follows easily that $(\xi - A_n(h) - B(h))^{-1} \to (\xi - A(h) - B(h))^{-1}$. Since the former is superconvex by Theorem 5.2, we conclude that the latter is also superconvex. This completes the proof of Theorem 3.3.

Remark 5.4. (a) We could not prove Theorem 3.3 earlier since we have no analog of the product formula (5.3) for the resolvents.

(b) We assumed the boundedness of $B(h)$ in Theorems 3.3 and 5.3 only to facilitate application of the product formula (5.3). There are many cases in which (5.3) is true for unbounded $B(h)$ (see [1]). Theorems 3.3 and 5.3 can easily be extended to such cases.

6. Applications

Theorems 3.3 and 5.3 may be used to construct a nontrivial resolvent-superconvex or semigroup-superconvex family $A(h)+B(h)$ from two such families $A(h)$ and $B(h)$, where $B(h)$ is bounded. We shall consider a rather simple case in which $A(h) = A$ is a constant operator A which is resolvent-positive (so that it forms a trivial resolvent-superconvex family).

It is in general difficult to find a nontrivial semigroup-superconvex family $B(h)$ even among bounded families. Thus we shall have to specialize the space X to concrete function spaces such as $X = C(S)$ (continuous functions on a compact Hausdorff space S) or $X = L^p(S)$, $1 \leqq p < \infty$, on a measure space S.

More generally, X may be, for example, the intersection of two L^p-spaces with different p's and different weight functions (see [6]).

Let B be an operator of multiplication on X with a real-valued function b, where $b \in C(S)$ for $X = C(S)$ or $b \in L^\infty(S)$ for the other cases. Then the family $e^{hB} \in B(X)$ is superconvex in h on any interval, as is easily verified by using appropriate partitions of unity on S in the case $X = C(S)$ or by approximating b in $L^\infty(S)$ with simple functions in the other cases. Thus $hB \in B(X)$ is semigroup-superconvex in h, and we obtain

Theorem 6.1 (generalized Cohen's theorem). *Let X, A, and B be as above. Then* $\operatorname{spb}(A+B)$ *is a convex function of B. If in particular A is a generator, then both* $\operatorname{spb}(A+B)$ *and* $\operatorname{type}(A+B)$ *are convex in B.*

Proof. Theorems 3.3 and 5.3 show that $\operatorname{spb}(A+hB)$ and/or $\operatorname{type}(A+hB)$ are convex in h. These are equivalent to the statements of the theorem in which B is supposed to vary.

Example 6.2. In [6], examples of $X = L^p \cap L^q$ on $(0, \infty)$ (with certain weight functions) are given in which $A = d/dx$ is semigroup-positive with $-\infty \leqq \operatorname{spb} A < \operatorname{type} A$. If B is the operator of multiplication with a real $b \in L^\infty$, it follows from Theorem 6.1 that $\operatorname{spb}(A+B)$ and $\operatorname{type}(A+B)$ are convex in B. These are two distinct quantities. If in particular $\operatorname{spb} A = -\infty$ (which is possible as shown in [6]), then $\operatorname{spb}(A+B) = -\infty$ while $\operatorname{type}(A+B) > -\infty$ (because $\operatorname{type} A > -\infty$).

Example 6.3. Another example of Theorem 6.1 is given by a second-order elliptic differential operator

$$A = \sum_{j,k=1}^{n} a_{jk}(x)\, \partial^2/\partial x_j \partial x_k + \sum_{j=1}^{n} a_j(x)\, \partial/\partial x_j + a(x) \tag{6.1}$$

with real coefficients, with the matrix $(a_{jk}(x))$ symmetric and positive-definite. The space variable x may vary on a domain $S \subset \mathbb{R}^n$ or an n-dimensional manifold S. A becomes a semigroup-positive operator in $X = L^p(S)$, possibly under a certain boundary condition (such as Dirichlet or Neumann), depending on the smoothness of the boundary ∂S. $X = C_0(S)$ (the subspace of $C(S)$ with Dirichlet boundary condition) is another choice for X.

If B is an operator of multiplication considered above, it follows from Theorem 6.1 that $\operatorname{spb}(A+B)$ and $\operatorname{type}(A+B)$ are convex functions of B. In most cases, however, these two quantities coincide, since $A+B$ will generate an analytic semigroup (see remarks after (1.3)). If in particular S is bounded, $A+B$ usually has compact resolvent and $\operatorname{spb}(A+B)$ is equal to the principal eigenvalue $\lambda(A+B)$ (cf. Protter-Weinberger [14]). It will be noted that under more restrictive assumptions, the convexity of $\lambda(A+B)$ in B is implied by the variational formulas given by Donsker-Varadhan [3, 4] and Holland [9]. Since such variational formulas depend heavily on the boundary conditions, it would be worthwhile to have an independent proof of the convexity.

Remark 6.4. In Example 6.3, consider a family $A - hB$ with a $B = b(x)\cdot$ fixed and with a real parameter h. Then $\lambda(A - hB)$ is convex in h. If $\lambda(A) < 0$ (which will be the case if $a(x) \leqq 0$ and the Dirichlet boundary condition is used), it

follows that $\lambda(A-hB)=0$ occurs for at most two values of h, and we may expect that there are in fact exactly two such values $h=h_{\pm}\gtrless 0$ unless $b(x)$ is of fixed sign. Since $\lambda(A-hB)$ is a simple eigenvalue with a positive eigenfunction (Perron-Frobenius theorem), we expect that the eigenvalue problem

$$Au=hBu \tag{6.2}$$

has two distinguished eigenvalues $h_{\pm}$ with the following properties. h_+ [h_-] is the smallest positive [largest negative] eigenvalue for (6.2). Both $h_{\pm}$ have geometric multiplicity one and have positive eigenfunctions. We may proceed to show that these eigenvalues have algebraic multiplicity one, in a sense to be specified. We do not go into details here, but note that the problem was studied in Hess-Kato [7] by a different method.

Remark 6.5. Convexity of spb A in A (or part of A) is by no means a general phenomenon for generators A. There are examples of the opposite situation. For example, let $X=L^p(0,1)$, $A=d^2/dx^2$ with the boundary condition $u(0)=u(1)=0$, and $B=-d/dx$ with boundary condition $u(0)=0$. Both A and B are semigroup-positive. But $\mathrm{spb}(kA+hB)=-\pi^2 k-h^2/4k$ for $k,h>0$ (see [10, Example VIII-1.19]), which is *concave* in k and h jointly rather than convex.

References

1. Chernoff, P.R.: Product formulas, nonlinear semigroups, and addition of unbounded operators. Mem. Amer. Math. Soc. **140** (1974)
2. Cohen, J.E.: Convexity of the dominant eigenvalue of an essentially nonnegative matrix. Proc. Amer. Math. Soc. **81**, 657-658 (1981)
3. Donsker, M.D., Varadhan, S.R.S.: On a variational formula for the principal eigenvalue for operators with maximum principle. Proc. Nat. Acad. Sci. U.S.A **72**, 780-783 (1975)
4. Donsker, M.D., Varadhan, S.R.S.: On the principal eigenvalue of second-order elliptic differential operators. Comm. Pure Appl. Math. **29**, 595-621 (1976)
5. Friedland, S.: Convex spectral functions. Linear and Multilinear Algebra **9**, 299-316 (1981)
6. Greiner, G., Voigt, J., Wolff, M.: On the spectral bound of the generator of semigroups of positive operators. Preprint 1981
7. Hess, P., Kato, T.: On some linear and nonlinear eigenvalue problems with an indefinite weight function. Comm. Partial Differential Equations **5**, 999-1030 (1980)
8. Hille, E., Phillips, R.S.: Functional analysis and semi-groups. Amer. Math. Soc. Colloq. Publ. Vol. 31, Rev. Ed. Providence, Rhode Island: American Mathematical Society 1957
9. Holland, C.J.: A minimum principle for the principal eigenvalue for second-order linear elliptic equations with natural boundary conditions. Comm. Pure Appl. Math. **31**, 509-519 (1978)
10. Kato, T.: Perturbation theory for linear operators, 2nd Ed. Berlin-Heidelberg-New York: Springer 1980
11. Kato, T.: A spectral mapping theorem for the exponential function, and some counterexamples. Mathematical Research Center Technical Report #2316, Madison: University of Wisconsin 1982
12. Kingman, J.F.C.: A convexity property of positive matrices. Quart. J. Math. Oxford (2) **12**, 283-284 (1961)
13. Protter, M.H., Weinberger, H.F.: On the spectrum of general second order operators. Bull. Amer. Math. Soc. **72**, 251-255 (1966)
14. Schaefer, H.H.: Topological vector spaces. New York-Heidelberg-Berlin: Springer 1971
15. Walsh, B.: Positive approximation identities and lattice-ordered dual spaces. Manuscripta Math. **14**, 57-63 (1974)

Received January 7, 1982

Math. Z. 180, 275–286 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Long-Time Asymptotic Properties of Dynamical Semigroups on W^*-algebras

Alberto Frigerio[1] and Maurizio Verri[2]

[1] Istituto di Scienze Fisiche, Università di Milano, Milano, Italy
[2] Istituto di Matematica, Informatica e Sistemistica, Università di Udine, Udine, Italy

Introduction

The theory of quantum dynamical semigroups provides a convenient mathematical description of the irreversible dynamics of an open quantum system. This motivates an interest in the study of conditions for a dynamical semigroup to induce approach to a stationary state [7, 8, 13, 14] and of related subjects such as irreducibility [2, 4], ergodic theorems [8, 10, 15] and Perron-Frobenius type results [1, 5, 16]. Most of the results have been shown so far in the finite-dimensional case [5, 13, 14] or when there exists a faithful (family of) stationary state(s) [1, 7, 8, 10, 15]. Here we prove more general theorems of ergodic type and on approach to equilibrium, by suitably modifying some of the techniques of the above quoted papers and using a result of [12]. We also give some applications to dynamical semigroups of Lindblad type [6, 11] and asymptotically finite-dimensional.

1. Preliminaries

Let $T=\{T_t\colon t\in\mathbb{R}^+\}$ be a *dynamical semigroup* on a W^*-algebra $\mathcal{M}$, i.e. a weakly* continuous one-parameter semigroup of completely positive identity preserving normal maps of $\mathcal{M}$ into itself, with T_0 the identity map of $\mathcal{M}$. Denote by T_{t*} the preadjoint map of T_t on the predual space $\mathcal{M}_*$ of normal linear functionals on $\mathcal{M}$: then $T_*=\{T_{t*}\colon t\in\mathbb{R}^+\}$ is a strongly continuous one-parameter semigroup of positive contractions on $\mathcal{M}_*$. Let also $\mathcal{F}(T)$, resp. $\mathcal{F}(T_*)$, be the fixed point set of T in $\mathcal{M}$, respectively of T_* in $\mathcal{M}_*$. If there exists a faithful family of T-invariant normal states on $\mathcal{M}$, then $\mathcal{F}(T)$ is a W^*-subalgebra of $\mathcal{M}$ [7, 14]. If P is a projection in $\mathcal{M}$ and φ is in $\mathcal{M}_*$, we denote by $P\varphi P$ the element of $\mathcal{M}_*$ defined as $P\varphi P(A)=\varphi(PAP)$ for all A in $\mathcal{M}$, and by $P\mathcal{M}_*P$ the set of such elements as φ spans $\mathcal{M}_*$. Then the hereditary W^*-subalgebra $P\mathcal{M}P$ of $\mathcal{M}$ is (canonically isomorphic to) the dual space of $P\mathcal{M}_*P$. A non-zero projection P in $\mathcal{M}$ is said to *reduce* T_* [2] if $P\mathcal{M}_*P$ is

0025-5874/82/0180/0275/$02.40

globally invariant under T_* or, equivalently, if

$$P T_t(A) P = P T_t(PAP) P \quad \text{for all } A \text{ in } \mathcal{M},\ t \text{ in } \mathbb{R}^+. \tag{1.1}$$

Then

$$T_t^P(A) = P T_t(A) P \quad \text{for all } A \text{ in } P\mathcal{M}P \tag{1.2}$$

defines a dynamical semigroup $T^P = \{T_t^P\colon t \in \mathbb{R}^+\}$ on $P\mathcal{M}P$. The support projection $S(\omega)$ of any state ω in $\mathcal{F}(T_*)$ reduces T_* since $S(\omega)\,\mathcal{M}_*\,S(\omega)$ is the norm closure of the set of linear combinations of the normal states on $\mathcal{M}$ which are majorized by a scalar multiple of ω, and the same is true for the *recurrent subspace projection* (cf. [5])

$$R = \sup\{S(\omega)\colon \omega \text{ is a state in } \mathcal{F}(T_*)\} \tag{1.3}$$

since $R\mathcal{M}_*R$ is the norm closure of $\bigcup\{S(\omega)\,\mathcal{M}_*\,S(\omega)\colon \omega$ is a state in $\mathcal{F}(T_*)\}$. We have $R=0$ when the semigroup T has no normal stationary state (see examples in [2, 4]), and $R=\mathbb{1}$ when T has a faithful family of normal stationary states.

The application of the mean ergodic theorems of [8, 10, 15] to T^R leads to the following.

Theorem 1.1. *For any dynamical semigroup $T=\{T_t\colon t\in\mathbb{R}^+\}$ on a W*-algebra $\mathcal{M}$,*

$$E(A) = w^*\text{-}\lim_{t\to\infty} \frac{1}{t}\int_0^t R\,T_s(A)\,R\,ds \tag{1.4}$$

exists for all A in $\mathcal{M}$, and defines a T-invariant normal conditional expectation E onto the W-subalgebra $\mathcal{F}(T^R)$ of $R\mathcal{M}R$; and a normal state ω on $\mathcal{M}$ is T-invariant if and only if $\omega\circ E=\omega$.*

As a consequence, the study of normal T-invariant states on $\mathcal{M}$ is reduced to the study of normal states on $\mathcal{F}(T^R)$ (cf. [8]).

It should be remarked that the above theorem and all results of Sect. 2 and 3 are also valid if complete positivity is replaced by the requirement that the maps T_t satisfy the Kadison-Schwarz inequality $T_t(A^*A)\geqq T_t(A)^*T_t(A)$ for all A in $\mathcal{M}$. As is well known, this property is stronger than positivity, but weaker than 2-positivity [6].

2. The Ergodic Theorem

In this Section we extend to the general case the mean ergodic theorems which were proved in [8, 10, 15] under the assumption $R=\mathbb{1}$.

Theorem 2.1. *For a dynamical semigroup $T=\{T_t\colon t\in\mathbb{R}^+\}$ on a W*-algebra $\mathcal{M}$ the following are equivalent:*

(i) *there exists a normal T-invariant norm one projection F of $\mathcal{M}$ onto $\mathcal{F}(T)$;*

(ii) $w\text{-}\lim_{t\to\infty} \frac{1}{t}\int_0^t T_{s*}(\varphi)\,ds$ *exists in $\mathcal{M}_*$ for all φ in $\mathcal{M}_*$;*

(iii) $w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(R)\,ds=\mathbb{1}$;

(iv) $\mathscr{F}(T_*)$ *separates* $\mathscr{F}(T)$.

If the above conditions are satisfied, then

$$F(A)=w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(A)\,ds \quad \textit{for all } A \textit{ in } \mathscr{M}. \tag{2.1}$$

Proof. (i)⇒(ii): For each A in $\mathscr{M}$, the net $\left\{\frac{1}{t}\int_0^t T_s(A)\,ds\right\}_{t>0}$ is compact in the weak* topology of $\mathscr{M}$. Any of its limit points, say A_∞, is in $\mathscr{F}(T)$, hence, if (i) holds, $A_\infty=F(A_\infty)$; on the other hand, $F\circ T_s=F$ for all s and F is normal, hence $F(A_\infty)=F(A)$. So, $\frac{1}{t}\int_0^t T_s(A)\,ds$ converges in the weak* topology to $F(A)$ as $t\to\infty$. This proves (2.1). As a consequence, for each φ in $\mathscr{M}_*$, the net $\left\{\frac{1}{t}\int_0^t T_{s*}(\varphi)\,ds\right\}_{t>0}$ converges to $\varphi\circ F$ in the $\sigma(\mathscr{M}^*,\mathscr{M})$ topology of $\mathscr{M}^*$. Both the net and its limit point are in $\mathscr{M}_*$, since F is assumed to be normal, and the $\sigma(\mathscr{M}^*,\mathscr{M})$ topology of $\mathscr{M}^*$, restricted to $\mathscr{M}_*$, is the weak topology of $\mathscr{M}_*$.

(ii)⇒(iii): For any state φ in $\mathscr{M}_*$, the support projection of $w\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_{s*}(\varphi)\,ds=\varphi_\infty$ satisfies $S(\varphi_\infty)\leqq R$, hence $\lim_{t\to\infty}\frac{1}{t}\int_0^t\varphi(T_s(R))\,ds=\varphi(\mathbb{1})$ for all φ in $\mathscr{M}_*$, which is (iii).

(iii)⇒(iv): If A is in $\mathscr{F}(T)$, then $RAR=E(A)$ is in $\mathscr{F}(T^R)$, where E and T^R have been defined in Section 1. We know from Theorem 1.1 that $\mathscr{F}(T_*)$ is isomorphic to the predual space of $\mathscr{F}(T^R)$, hence it separates $\mathscr{F}(T)$ if and only if $RAR=0$, $A\in\mathscr{F}(T)$, implies $A=0$. Now, from (iii) it follows that $w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s((\mathbb{1}-R)A)\,ds=w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(A(\mathbb{1}-R))\,ds=0$ for all A in $\mathscr{M}$, by application of the Kadison-Schwarz inequality to the maps $\frac{1}{t}\int_0^t T_s(.)\,ds$. Hence, if A is in $\mathscr{F}(T)$,

$$A=w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(RAR)\,ds$$

which proves that $A=0$ if $RAR=0$.

(iv)⇒(i): This is a special case of a theorem due to Nagel [12].

Remark 2.2. It should be noticed that for *any* dynamical semigroup T on a W^*-algebra $\mathscr{M}$ there exists a unique T-invariant norm one projection F, which satisfies (2.1); however, F need not be normal. The proof of this fact may be obtained by using Theorem 5.1 of [3], where the weak topology should be replaced by the weak* topology. It follows also from [3] that the Cesaro limit or the Abel limit may be used equivalently in (2.1).

Remark 2.3. When $\mathcal{M}$ is finite-dimensional, the map F must be normal, so that (iii) always holds. Actually, more is true in this case: Evans and Høegh-Krohn have shown that $\lim_{t\to\infty} T_t(R)=\mathbb{1}$ [5].

In the general, infinite-dimensional case, the problem of finding R and of ascertaining whether $w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(R)\,ds=\mathbb{1}$ seems hard and a detailed analysis of concrete cases would be necessary. There is, however, a class of dynamical semigroups for which the above problem can be reduced to the analogous problem for a Markov semigroup on a discrete state space. Suppose there exist a totally atomic abelian W^*-subalgebra $\mathcal{Z}$ of $\mathcal{M}$ and a normal conditional expectation N of $\mathcal{M}$ onto $\mathcal{Z}$, commuting with the dynamical semigroup T. (For instance, this is the case when $\mathcal{M}$ is the algebra $\mathcal{B}(\mathcal{H})$ of all bounded linear operators on a separable Hilbert space $\mathcal{H}$, $\mathcal{Z}=\{P_i\colon i\in I\}''$ where $\{P_i\colon i\in I\}$ is a family of mutually orthogonal one-dimensional projections with $\sum_{i\in I} P_i=\mathbb{1}$, and T commutes with a group $\{\alpha_t\colon t\in\mathbb{R}\}$ of *-automorphisms of $\mathcal{M}$ determined by a Hamiltonian $H=\sum_{i\in I}\varepsilon_i P_i$ where the ε_i are distinct real numbers). Then, for all t in $\mathbb{R}^+$, T_t may be restricted to a map $\tilde{T}_t$ on $\mathcal{Z}$, and $\tilde{T}=\{\tilde{T}_t\colon t\in\mathbb{R}^+\}$ is a Markov semigroup on a discrete state space, the "states" corresponding to the atoms of $\mathcal{Z}$. Now, we claim that $R=\tilde{R}$, where R and $\tilde{R}$ are the recurrent subspace projections of T_* and $\tilde{T}_*$, respectively. Indeed, each state φ in $\mathcal{F}(\tilde{T}_*)$ can be extended to a state $\varphi\circ N$ in $\mathcal{F}(T_*)$ and $\tilde{R}$ is just $\sup\{S(\omega)\colon \omega$ is an N-invariant state in $\mathcal{F}(T_*)\}$. This proves $R\geqq\tilde{R}$. On the other hand, we have $\tilde{R}=N(\tilde{R})$ and, for any state ω in $\mathcal{F}(T_*)$, $\omega\circ N\restriction\mathcal{Z}$ is a state in $\mathcal{F}(\tilde{T}_*)$. Then, $\omega(\tilde{R})=(\omega\circ N)(\tilde{R})=\mathbb{1}$, so that $\tilde{R}\geqq R$.

Finally, since $R=\tilde{R}=N(\tilde{R})$ and N is normal, for each φ in $\mathcal{M}_*$ we have

$$\varphi\left(\frac{1}{t}\int_0^t T_s(R)\,ds\right)=(\varphi\circ N)\left(\frac{1}{t}\int_0^t T_s(\tilde{R})\,ds\right),$$

hence the existence of $\sigma(\mathcal{M},\mathcal{M}_*)\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(R)\,ds$ is equivalent to the existence of $\sigma(\mathcal{Z},\mathcal{Z}_*)\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(\tilde{R})\,ds$, and the former is $\mathbb{1}$ if and only if the latter is.

3. Approach to Equilibrium

In this Section we study conditions under which any normal state on $\mathcal{M}$ approaches a normal state under the action of T_{t*} as $t\to\infty$. We reduce this problem to the corresponding problem for T^R, which has a faithful family of normal stationary states, plus the condition that $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$. Then, by restricting considerations to T^R, we may assume $R=\mathbb{1}$; for this case we give some extension of the results of [8].

The following Lemma is needed repeatedly.

Lemma 3.1. *Let $T=\{T_t: t\in\mathbb{R}^+\}$ be a dynamical semigroup on a W*-algebra $\mathcal{M}$, and P a projection in $\mathcal{M}$ reducing T_*. Then the following are equivalent:*

(i) $w^*\text{-}\lim_{t\to\infty} T_t(P)=\mathbb{1}$;

(ii) $\lim_{t\to\infty}\|T_{t*}\varphi - P(T_{t*}\varphi)P\|=0$ *for all* φ *in* $\mathcal{M}_*$;

(iii) *for all* φ *in* $\mathcal{M}_*$ *and* $\varepsilon>0$, *there exists* ψ *in* $P\mathcal{M}_*P$ *and* $t>0$ *such that* $\|T_{t*}\varphi-\psi\|<\varepsilon$.

Proof. (i)⇒(ii): We can assume $\mathcal{M}$ to be concretely represented as a von Neumann algebra of bounded linear operators on a Hilbert space $\mathcal{H}$. For all A in $\mathcal{M}$, ξ in $\mathcal{H}$, and $t\geqq 0$, we have, using the Kadison-Schwarz inequality for T_t,

$$\begin{aligned}\|T_t(A(\mathbb{1}-P))\,\xi\|^2 &\leqq (\xi, T_t((\mathbb{1}-P)\,A^*A(\mathbb{1}-P))\,\xi)\\ &\leqq \|A\|^2(\xi, T_t(\mathbb{1}-P)\,\xi)\end{aligned}$$

so that $T_t(A(\mathbb{1}-P))$ tends to zero as $t\to\infty$ in the strong operator topology, uniformly in A. Thus, $T_t(A-PAP)$ tends to zero as $t\to\infty$ in the weak (hence ultraweak, being a bounded net) operator topology, uniformly in A. This proves (ii).

(ii)⇒(i): For any state φ in $\mathcal{M}_*$, (ii) gives

$$0\leqq\varphi(T_t(\mathbb{1}-P))=(T_{t*}\varphi-P(T_{t*}\varphi)P)(\mathbb{1})\to 0$$

as $t\to\infty$, hence (i) follows.

(ii)⇒(iii): Given φ in $\mathcal{M}_*$ and $\varepsilon>0$, choose t large enough and $\psi=P(T_{t*}\varphi)P$.

(iii)⇒(ii): Let φ, ε, ψ, t be given such that (iii) holds. For all s in $\mathbb{R}^+$ we have

$$\begin{aligned}\|P(T_{t+s*}\varphi)P-P(T_{s*}\psi)P\| &\leqq \|T_{t+s*}\varphi-T_{s*}\psi\|\\ &\leqq \|T_{t*}\varphi-\psi\|<\varepsilon.\end{aligned}$$

But ψ is in $P\mathcal{M}_*P$ and P reduces T_*, hence $P(T_{s*}\psi)P=T_{s*}\psi$. Then, $\|T_{t+s*}\varphi - P(T_{t+s*}\varphi)P\|<2\varepsilon$ for all s in $\mathbb{R}^+$ and a suitable t, which is (ii).

Theorem 3.2. *Let $T=\{T_t: t\in\mathbb{R}^+\}$ be a dynamical semigroup on a W*-algebra $\mathcal{M}$, and let T^R and E be defined as in Section* 1. *Then the following conditions are equivalent:*

(i) $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$ *and* $w\text{-}\lim_{t\to\infty}\psi\circ T_t^R=\psi\circ E$ *for all* ψ *in* $R\mathcal{M}_*R$;

(ii) $w\text{-}\lim_{t\to\infty} T_{t*}\varphi$ *exists in* $\mathcal{M}_*$ *for all* φ *in* $\mathcal{M}_*$.

Similarly, the following conditions are equivalent:

(i′) $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$ *and* $\lim_{t\to\infty}\|\psi\circ T_t^R-\psi\circ E\|=0$ *for all* ψ *in* $R\mathcal{M}_*R$;

(ii′) $n\text{-}\lim_{t\to\infty} T_{t*}\varphi$ *exists in* $\mathcal{M}_*$ *for all* φ *in* $\mathcal{M}_*$.

Furthermore, conditions (i′), (ii′) *imply* (i), (ii), *and* (i) *implies the equivalent conditions of Theorem* 2.1.

Proof. (ii)$\Rightarrow$(i) and (ii')$\Rightarrow$(i'): If the limit in (ii) (or in (ii')) exists, it coincides with $\varphi \circ F$, where F is defined by (2.1) and is normal since the limit is assumed to exist in $\mathcal{M}_*$. For any state φ in $\mathcal{M}_*$, the support projection of $\varphi \circ F$ is not greater than R, so that $\lim_{t\to\infty} \varphi(T_t(R))=\varphi(\mathbb{1})$ for all φ in $\mathcal{M}_*$, i.e. $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$. Moreover, it is clearly $RF(.)R=E$, hence the conditions $w\text{-}\lim_{t\to\infty} \psi \circ T_t^R=\psi \circ E$ and $\lim_{t\to\infty} \|\psi \circ T_t^R-\psi \circ E\|=0$ for all ψ in $R\mathcal{M}_*R$ are the specializations of (ii) and (ii'), respectively, to ψ in $R\mathcal{M}_*R$.

(i)$\Rightarrow$(ii) and (i')$\Rightarrow$(ii'): Notice, first of all, that (i')$\Rightarrow$(i), which in turn implies (iii) of Theorem 2.1, hence F exists, given by (2.1), and is normal. For all A in $\mathcal{M}$ with $\|A\| \leqq 1$, φ in $\mathcal{M}_*$ and t, s in $\mathbb{R}^+$, we have

$$
\begin{aligned}
&|\varphi(T_{t+s}(A))-\varphi(F(A))| \\
&\quad \leqq |\varphi \cdot T_t(T_s(A)-R\,T_s(A)\,R)|+|\varphi \circ T_t(F(A)-RF(A)\,R)|+|\varphi \circ T_t(T_s^R(A)-E(A))| \\
&\quad \leqq \|T_{t*}\varphi-R(T_{t*}\varphi)\,R\|\,[\|T_s(A)\|+\|F(A)\|]+|T_{t*}\varphi(T_s^R(A)-E(A))| \\
&\quad \leqq 2\|T_{t*}\varphi-R(T_{t*}\varphi)R\|+|T_{t*}\varphi(T_s^R(A)-E(A))|,
\end{aligned}
$$

where $T_t \circ F=F$ and $RF(.)R=E$ have been used.

Take $\varepsilon>0$. Because of Lemma 3.1, there exists $t(\varepsilon, \varphi)$ such that

$$\|T_{t*}\varphi-R(T_{t*}\varphi)R\|<\varepsilon/3 \quad \text{for all } t\geqq t(\varepsilon, \varphi).$$

Under condition (i), there exists $s(t, \varepsilon, A, \varphi)$ such that

$$|T_{t*}\varphi(T_s^R(A)-E(A))|<\varepsilon/3 \quad \text{for all } s\geqq s(t, \varepsilon, A, \varphi)$$

where we have taken into account that $T_{t*}\varphi(T_s^R(A)-E(A))=(RT_{t*}\varphi R)\,(T_t^R(A)-E(A))$. If condition (i') holds, $s(t, \varepsilon, A, \varphi)$ may be chosen independent of A for $\|A\| \leqq 1$. So, if $u\geqq u(\varepsilon, A, \varphi)=t(\varepsilon, \varphi)+s(t(\varepsilon, \varphi), \varepsilon, A, \varphi)$, we may decompose u as $t(\varepsilon, \varphi)+s$ with $s\geqq s(t(\varepsilon, \varphi), \varepsilon, A, \varphi)$ and we have

$$|\varphi(T_u(A))-\varphi(F(A))|<\varepsilon \quad \text{for all } u\geqq u(\varepsilon, A, \varphi).$$

This proves (ii) from (i). If (i') holds, $u(\varepsilon, A, \varphi)$ may be chosen independent of A for $\|A\| \leqq 1$ and (ii') follows.

In the rest of the Section, we assume that $R=\mathbb{1}$ and we given some extension of the results of [8] concerning the weak approach to equilibrium. These considerations could, in principle, be applied to T^R. As regards the condition $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$, we shall discuss it in some special case in the next Section.

Let $\mathcal{M}_\Gamma(T)$ be the weak closure of the linear span of the eigenvectors of T_t corresponding to eigenvalues of modulus one, and define

$$
\begin{aligned}
\mathcal{N}(T)=\{&A\in\mathcal{M}:\ T_t(A^*A)=T_t(A^*)\,T_t(A), \\
&T_t(AA^*)=T_t(A)\,T_t(A^*) \text{ for all } t \text{ in } \mathbb{R}^+\}.
\end{aligned}
$$

It has been shown by Evans [4, Theorem 3.1] that $\mathcal{N}(T)$ is a W^*-subalgebra of $\mathcal{M}$. When there exists a faithful family of normal T-invariant states, $\mathcal{M}_\Gamma(T)$ is a

W^*-subalgebra of $\mathcal{M}$ and is contained in $\mathcal{N}(T)$, as shown by Albeverio and Høegh-Krohn in [1]; furthermore, when $R=\mathbb{1}$, $\mathcal{N}(T)$ can be equivalently characterized as

$$\mathcal{N}(T)=\{A\in\mathcal{M}\colon \omega(A^*A)=\omega(T_t(A^*)\,T_t(A)),\ \omega(AA^*) =\omega(T_t(A)\,T_t(A^*)) \text{ for all states } \omega \text{ in } \mathcal{F}(T_*) \text{ and } t \text{ in } \mathbb{R}^+\},$$

which shows that $\mathcal{N}(T)$ is globally invariant under T when $R=\mathbb{1}$.

Theorem 3.3. *Let $T=\{T_t\colon t\in\mathbb{R}^+\}$ be a dynamical semigroup on a W^*-algebra $\mathcal{M}$. If there is a faithful family of normal T-invariant states, then $\mathcal{N}(T)=\mathcal{F}(T)$ implies that*

$$w^*\text{-}\lim_{t\to\infty} T_t(A)=E(A) \qquad \text{for all } A \text{ in } \mathcal{M}, \tag{3.1}$$

which in turn implies that $\mathcal{M}_\Gamma(T)=\mathcal{F}(T)$.

Proof. Under the above assumptions, we have

$$E(A)=F(A)=w^*\text{-}\lim_{t\to\infty}\frac{1}{t}\int_0^t T_s(A)\,ds \qquad \text{for all } A \text{ in } \mathcal{M}.$$

Then the first implication is a straightforward generalization of Theorem 3.1 of [8] (we warn the reader that the set $\mathcal{N}(T)$ defined here is $\mathcal{N}(T)\cap\mathcal{N}(T)^*$ in the notation of [8]), and the second one is obvious.

Theorem 3.4. *Under the assumptions of Theorem* 3.3, *suppose in addition that either*

(a) *$\mathcal{M}$ is finite-dimensional,*

or

(b) *$\mathcal{M}$ is the algebra $\mathcal{B}(\mathcal{H})$ of all bounded linear operators on a separable Hilbert space $\mathcal{H}$, and T is norm continuous and satisfies detailed balance, in the sense of* [9], *with respect to a faithful normal state ω.*

Then $\mathcal{N}(T)=\mathcal{F}(T)$ is a necessary and sufficient condition for the validity of (3.1).

Proof. We show that in both cases (a) and (b) $\mathcal{M}_\Gamma(T)=\mathcal{N}(T)$ and then we apply Theorem 3.3. Notice that, under the present assumptions, $\mathcal{N}(T)$ is the largest T-invariant W^*-subalgebra of $\mathcal{M}$ such that the restriction of T to it is a semigroup of *-homomorphisms.

When $\mathcal{M}$, hence $\mathcal{N}(T)$, is finite-dimensional, a semigroup of *-homomorphisms can always be extended to a group of *-automorphisms with pure point spectrum. This proves $\mathcal{N}(T)=\mathcal{M}_\Gamma(T)$ in case (a).

In case (b), the detailed balance condition of [9] means that the infinitesimal generator L of T can be decomposed as $L=L_h+L_s$, where $L_h(A)=[iH, A]$ for all A in $\mathcal{M}$, H being a self-adjoint operator on $\mathcal{H}$ commuting with the density matrix ρ determining ω, and where L_s satisfies $\omega(AL_s(B))=\omega(L_s(A)B)$ for all A, B in $\mathcal{M}$. Recalling Theorem 3.1 of [4], we see that, if A is in $\mathcal{N}(T)$, we have $\omega(AL(B))=-\omega(L(A)B)$ for all B in $\mathcal{M}$, so that $L(A)=L_h(A)$ for all A in $\mathcal{N}(T)$. Then, $T_t(A)=e^{iHt}Ae^{-iHt}$ for all A in $\mathcal{N}(T)$, t in $\mathbb{R}^+$, and H has pure point spectrum since it commutes with the strictly positive density matrix ρ. This proves $\mathcal{N}(T)=\mathcal{M}_\Gamma(T)$ in case (b).

4. Applications

Let $\mathcal{M}$ be the algebra $\mathcal{B}(\mathcal{H})$ of all bounded linear operators on a separable Hilbert space $\mathcal{H}$ and let $T=\{T_t: t\in\mathbb{R}^+\}$ be a norm continuous dynamical semigroup on $\mathcal{M}$ with infinitesimal generator L. Then L is of the form (Lindblad [11])

$$L(A)=K^*A+AK+W(A) \quad \text{for all } A \text{ in } \mathcal{B}(\mathcal{H}) \tag{4.1}$$

where K is in $\mathcal{B}(\mathcal{H})$ and W is a completely positive normal map of $\mathcal{B}(\mathcal{H})$ into itself, satisfying $K^*+K+W(\mathbb{1})=0$. By Stinespring's theorem, W can be written as

$$W(A)=\sum_{i=1}^{\infty} V_i^* A V_i \quad \text{for all } A \text{ in } \mathcal{B}(\mathcal{H}) \tag{4.2}$$

where the V_i are in $\mathcal{B}(\mathcal{H})$ and the convergence is ultraweak. We can allow for an unbounded K in (4.1) in the following way: if $C_t=\exp(Kt)$ is a strongly continuous contraction semigroup on $\mathcal{H}$ such that $(K\xi,\ \eta)+(\xi,\ K\eta)+(\xi,\ W(\mathbb{1})\eta)=0$ for all ξ, η in $\mathrm{dom}(K)$, and $S_t(A)=C_t^* A C_t$, $A\in\mathcal{B}(\mathcal{H})$, $t\in\mathbb{R}^+$, then there exists a dynamical semigroup $T=\{T_t: t\in\mathbb{R}^+\}$ of $\mathcal{B}(\mathcal{H})$ satisfying

$$T_t(A)=S_t(A)+\int_0^t S_{t-s} W T_s(A)\,\mathrm{d}s \quad \text{for all } A \text{ in } \mathcal{B}(\mathcal{H}). \tag{4.3}$$

Such dynamical semigroups will be called *of Lindblad type* [6].

A projection P reduces the preadjoint of a dynamical semigroup T of Lindblad type if and only if [2]

$$(\mathbb{1}-P)\,V_i P=0 \text{ for all } i \text{ and } (\mathbb{1}-P)\,C_t P=0 \text{ for all } t \text{ in } \mathbb{R}^+, \tag{4.4}$$

and if T is norm continuous with generator of the form (4.1), T^P is norm continuous with a generator of the same form, in which K is replaced by PKP and V_i by PV_iP. Moreover, $\mathcal{N}(T)\subseteq\{V_i^*, V_i\}'$, and if T has a faithful family of normal stationary states, then $\mathcal{M}(T)=\{V_i^*, V_i, K, K^*\}'$ [4, 7]. If $\mathrm{lin}\{V_i\}$ is self-adjoint, a projection reducing T is also a fixed point of T, hence in this case the equivalent conditions of Theorem 2.1 give $R=\mathbb{1}$.

For some semigroup of Lindblad type, it is possible to check concretely the equivalent conditions of Lemma 3.1.

Proposition 4.1. *Let $T=\{T_t: t\in\mathbb{R}^+\}$ be a dynamical semigroup of Lindblad type on $\mathcal{M}=\mathcal{B}(\mathcal{H})$ and let $\{P_n: n=0,1,2,\ldots\}$ be an increasing (finite or infinite) sequence of projections in $\mathcal{B}(\mathcal{H})$ satisfying*

(i) $P_0=P_1=P$, $\sup_n P_n=\mathbb{1}$;

(ii) $(\mathbb{1}-P_n)\,V_i P_{n+1}=0$ *for all n and for all i;*

(iii) $(\mathbb{1}-P_n)\,C_t P_n=0$ *for all n and for all t in $\mathbb{R}^+$;*

(iv) $\lim_{t\to\infty} C_t(\mathbb{1}-P)\,\xi=0$ *for all ξ in $\mathcal{H}$.*

Then P satisfies the equivalent conditions of Lemma 3.1.

Proof. Note, first of all, that all P_n reduce T_* by (ii), (iii) and (4.4). Then it follows from (ii), (iii), (4.3) and (4.2) that $(T_{t*}-S_{t*})(\varphi)$ is in $\mathcal{T}(P_{n-1}\mathcal{H})$

$=P_{n-1}\mathcal{M}_*P_{n-1}$ if φ is in $\mathcal{T}(P_n\mathcal{H})=P_n\mathcal{M}_*P_n$ (here we denote by $\mathcal{T}(\mathcal{K})$ the space of all trace class operators on a Hilbert space $\mathcal{K}$). By (iv), $\lim_{t\to\infty}\|S_{t*}(\varphi-P\varphi P)\|=0$ for all φ in $\mathcal{M}_*$. Thus, for all $n=2,3,\ldots$, for all φ_n in $\mathcal{T}(P_n\mathcal{H})$ and all $\varepsilon>0$, it is possible to find φ_{n-1} in $\mathcal{T}(P_{n-1}\mathcal{H})$ and $t=t(\varphi_n,\varepsilon)>0$ such that $\|T_{t*}\varphi_n-\varphi_{n-1}\|<\varepsilon$; take, e.g., $\varphi_{n-1}=S_{t*}(P\varphi_nP)+\int_0^t T_{s*}W_*S_{t-s*}(\varphi_n)\,ds$ with t large enough: then, $\|T_{t*}\varphi_n-\varphi_{n-1}\|=\|S_{t*}(\varphi_n-P\varphi_nP)\|$. By iterating this procedure, given φ_n in $\mathcal{T}(P_n\mathcal{H})$ and $\varepsilon>0$, it is possible to find ψ in $\mathcal{T}(P\mathcal{H})$ and $s=s(n,\varphi_n,\varepsilon)>0$ such that

$$\|T_{s*}\varphi_n-\psi\|<\varepsilon;$$

here, $s(n,\varphi_n,\varepsilon)=\sum_{j=2}^{n} t(\varphi_j,\varepsilon/(n-1))$ and $\psi=\varphi_1$. Then, by Lemma 3.1, we have for all n

$$\lim_{t\to\infty}\|T_{t*}\varphi_n-P(T_{t*}\varphi_n)P\|=0 \quad \text{for all } \varphi_n \text{ in } \mathcal{T}(P_n\mathcal{H}). \tag{4.5}$$

On the other hand, by (i) we have that

$$\lim_{n\to\infty}\|\varphi-P_n\varphi P_n\|=0 \quad \text{for all } \varphi \text{ in } \mathcal{T}(\mathcal{H}). \tag{4.6}$$

Combining (4.5) and (4.6), with $\varphi_n=P_n\varphi P_n$, we conclude by an $\varepsilon/3$ argument that

$$\lim_{t\to\infty}\|T_{t*}\varphi-P(T_{t*}\varphi)P\|=0 \quad \text{for all } \varphi \text{ in } \mathcal{T}(\mathcal{H}),$$

which is (ii) of Lemma 3.1.

Remark 4.2. Conditions (ii) to (iv) of Proposition (4.1) are most easily checked when there is a family $\{Q_k: k\in I\}$ of mutually orthogonal projections with $\sum_{k\in I}Q_k=\mathbb{1}$ such that $\{Q_k: k\in I\}''$ contains all P_n and C_t, and is mapped into itself by W. Explicitly, let

$$P_n=\sum_{k\in I_n}Q_k \quad \text{with } I_n\subseteq I_{n+1}(n=0,1,\ldots), \quad \sup_n I_n=I;$$

$$C_t=\sum_{k\in I}\exp[-(i\omega_k+\gamma_k)t]\,Q_k \quad \text{with } \omega_k\in\mathbb{R},\ \gamma_k\in\mathbb{R}^+;$$

$$W(Q_k)=\sum_{j\in I}W_{kj}Q_j \quad \text{with } W_{kj}\in\mathbb{R}^+.$$

Then (iii) is satisfied and (iv) becomes $\gamma_k>0$ for all $k\notin I_1$; it is easily shown that (ii) is equivalent to $W(\mathcal{B}(P_{n+1}\mathcal{H}))\subseteq\mathcal{B}(P_n\mathcal{H})$ for all n, which in turn is equivalent to $W_{kj}=0$ whenever $k\notin I_n$, $j\in I_{n+1}$.

Note that, when $\operatorname{tr}Q_k<\infty$ for all $k\in I$, the W_{kj} are just the transition rates of the Pauli master equation

$$\dot{n}_k=\sum_j W_{kj}n_j-W_{jk}n_k$$

governing the evolution under T_* of the diagonal density matrices of the form $\sum_{k\in I} n_k Q_k$, $\sum_{k\in I} n_k \operatorname{tr} Q_k < \infty$.

The following Proposition is the generalization of Theorem 3.2 of [8].

Proposition 4.3. *Under the assumptions of Proposition 4.1, suppose in addition that*

(a) *there exists a T-invariant normal state ω, and*

(b) $\operatorname{lin}\{PV_iP\}$ *is self-adjoint and its commutant in $\mathcal{B}(P\mathcal{H})$ is reduced to the multiples of P.*

Then $S(\omega)=P$, $R=P$ and $w\text{-}\lim_{t\to\infty} \varphi\circ T_t=\varphi(\mathbb{1})\,\omega$ for all φ in $\mathcal{M}_*$.

Proof. Apply Theorem 3.2 of [8] to T^P to prove that $S(\omega)=P$ and that $w\text{-}\lim_{t\to\infty} \varphi\circ T_t^P=\varphi(\mathbb{1})\,\omega$ for all φ in $P\mathcal{M}_*P$. Then $R\geqq P$. On the other hand, by Proposition 4.1, $w^*\text{-}\lim_{t\to\infty} T_t(P)=\mathbb{1}$, hence $\varphi(P)=\mathbb{1}$ for each state φ in $\mathcal{F}(T_*)$ and $P\geqq R$. It follows that $R=P$ and, by Theorem 3.2, $\varphi\circ T_t$ tends weakly as $t\to\infty$ to $\varphi(\mathbb{1})\,\omega$ for all φ in $\mathcal{M}_*$.

Let $\mathcal{M}$ be any W^*-algebra. We say that the preadjoint semigroup T_* of a dynamical semigroup T on $\mathcal{M}$ is *asymptotically finite-dimensional* if there exists a projection P in $\mathcal{M}$ reducing T_* and satisfying the conditions of Lemma 3.1 such that $P\mathcal{M}_*P$ is finite-dimensional.

Proposition 4.4. *Let T_* be asymptotically finite-dimensional Then $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$.*

Proof. Let R^P be the recurrent subspace projection of T_*^P. Clearly, $R^P\leqq R$. It is also easy to show that R^P reduces T_* since it reduces T_*^P and P reduces T_*. Next, we show that $w^*\text{-}\lim_{t\to\infty} T_t(R^P)=\mathbb{1}$. Since, by assumption, P satisfies the conditions of Lemma 3.1, for all φ in $\mathcal{M}_*$ and $\varepsilon>0$ there exist ψ in $P\mathcal{M}_*P$ and $t>0$ such that

$$\|T_{t*}\varphi-\psi\|<\varepsilon/2.$$

Similarly, by Lemma 3.1 and [5], for all ψ in $P\mathcal{M}_*P$ and $\varepsilon>0$ there exist ω in $R^P\mathcal{M}_*R^P$ and $s>0$ such that

$$\|T_{s*}^P\psi-\omega\|<\varepsilon/2.$$

But P reduces T, hence $T_{s*}^P\psi=T_{s*}\psi$ for all ψ in $P\mathcal{M}_*P$ and $s>0$, so that

$$\|T_{t+s*}\varphi-\omega\|<\varepsilon.$$

Therefore, since R^P reduces T_*, $w^*\text{-}\lim_{t\to\infty} T_t(R^P)=\mathbb{1}$ by Lemma 3.1. Finally, using the same argument as in Proposition 4.3, we find that $R^P\geqq R$, so $R^P=R$ and $w^*\text{-}\lim_{t\to\infty} T_t(R)=\mathbb{1}$.

Theorem 4.5. *Let $T=\{T_t\colon t\in\mathbb{R}^+\}$ be a dynamical semigroup on a W^*-algebra $\mathcal{M}$ such that T_* is asymptotically finite-dimensional. Then the following are equivalent:*

(i) *there is a state* ω *in* $\mathcal{M}_*$ *such that*

$$\lim_{t\to\infty} \|T_{t*}\varphi - \varphi(\mathbb{1})\,\omega\| = 0 \qquad \text{for all } \varphi \text{ in } \mathcal{M}_*;$$

(ii) *there is exactly one* T*-invariant state* ω *in* $\mathcal{M}_*$;

(iii) *the set* $\mathcal{P}(T_*)$ *of all projections in* $\mathcal{M}$ *reducing* T_* *has a non-zero smallest element.*

If these conditions hold, then $\inf \mathcal{P}(T_*) = S(\omega) = R$.

Proof. (i)$\Rightarrow$(ii): Obvious.

(i)$\Rightarrow$(iii): The support $S(\omega)$ of ω reduces T_* and, if Q is in $\mathcal{P}(T_*)$, it follows from (i) that ω is in $Q\mathcal{M}_*Q$, so that $S(\omega) \leqq Q$ for all Q in $\mathcal{P}(T_*)$.

(iii)$\Rightarrow$(ii): Let ψ be an extremal T-invariant normal state on $\mathcal{M}$, and consider the dynamical semigroup $T^{S(\psi)}$ on $S(\psi)\mathcal{M}S(\psi)$. For all φ in $(\mathcal{M}_*)_+$ with $S(\varphi) \leqq S(\psi)$, we have $T_{t*}\varphi = T^{S(\psi)}_{t*}\varphi$ for all t in $\mathbb{R}^+$. Hence, ψ is also extremal $T^{S(\psi)}$-invariant, and since it is faithful on $S(\psi)\mathcal{M}S(\psi)$, it follows from [7] that no projection in $S(\psi)\mathcal{M}S(\psi)$ reduces $T^{S(\psi)}$. Then, no projection Q in $\mathcal{M}$ with $Q \leqq S(\psi)$ can reduce T. As a consequence, the support of any extremal T-invariant normal state is a minimal element of $\mathcal{P}(T_*)$. By (iii), $\mathcal{P}(T_*)$ has a unique minimal element, hence there exists just one extremal T-invariant normal state. Since T_* is asymptotically finite-dimensional, $\mathcal{F}(T_*)$ is non-zero and finite-dimensional, hence any state in $\mathcal{F}(T_*)$ can be decomposed (not uniquely, in general) into extremal T-invariant normal states. We can conclude that there exists exactly one T-invariant state ω in $\mathcal{M}_*$, with support $S(\omega) = \inf \mathcal{P}(T_*)$.

(ii)$\Rightarrow$(i): If ω is the unique T-invariant state in $\mathcal{M}_*$ we have $R = S(\omega)$. Since T_* is asymptotically finite-dimensional, $w^*\text{-}\lim_{t\to\infty} T_t(R) = \mathbb{1}$ by Proposition 4.4. Hence, by Theorem 3.2, it suffices to show that

$$\lim_{t\to\infty} \|\varphi \circ T_t - \varphi(\mathbb{1})\,\omega\| = 0 \qquad \text{for all } \varphi \text{ in } S(\omega)\mathcal{M}_*S(\omega). \tag{4.7}$$

Let P be the projection satisfying $w^*\text{-}\lim_{t\to\infty} T_t(P) = \mathbb{1}$ such that $P\mathcal{M}_*P$ is finite-dimensional, whose existence has been assumed. Then $\omega(P) = 1$ and $S(\omega) \leqq P$. It follows that $S(\omega)\mathcal{M}_*S(\omega)$ is finite-dimensional, and we get (4.7) by the uniqueness of the invariant state, as shown by Evans and Høegh-Krohn [5].

References

1. Albeverio, S., Høegh-Krohn, R.: Frobenius theory for positive maps of von Neumann algebras. Comm. Math. Phys. **64**, 83–94 (1978)
2. Davies, E.B.: Quantum stochastic processes II. Comm. Math. Phys. **19**, 83–105 (1970). See also Davies, E.B.: Quantum theory of open systems (Chap. 6). New York-San Francisco-London: Academic Press 1976
3. Davies, E.B.: One-parameter semigroups. New York-San Francisco-London: Academic Press 1980

4. Evans, D.E.: Irreducible quantum dynamical semigroups. Comm. Math. Phys. **54**, 293–297 (1977)
5. Evans, D.E., Høegh-Krohn, R.: Spectral properties of positive maps on C^*-algebras. J. London Math. Soc. (2) **17**, 345–355 (1978)
6. Evans, D.E., Lewis, J.T.: Dilations of irreversible evolutions in algebraic quantum theory. Comm. Dublin Inst. Adv. Studies. Ser. A **24**, 1977
7. Frigerio, A.: Quantum dynamical semigroups and approach to equilibrium. Lett. Math. Phys. **2**, 79–87 (1977)
8. Frigerio, A.: Stationary states of quantum dynamical semigroups. Comm. Math. Phys. **63**, 269–276 (1978)
9. Kossakowski, A., Frigerio, A., Gorini, V., Verri, M.: Quantum detailed balance and KMS condition. Comm. Math. Phys. **57**, 97–110 (1977)
10. Kümmerer, B., Nagel, R.J.: Mean ergodic semigroups on W^*-algebras. Acta Sci. Math. (Szeged) **41**, 151–159 (1979). See also: Kümmerer, B.: Mean Ergodic Semigroups of contractions in W^*-algebras. In: Group Theoretical Methods in Physics. Proceedings (Tübingen 1977), pp. 475–481. Lecture Notes in Physics **79**. Berlin-Heidelberg-New York: Springer 1978
11. Lindblad, G.: On the generators of quantum dynamical semigroups. Comm. Math. Phys. **48**, 119–130 (1976)
12. Nagel, R.J.: Mittelergodische Halbgruppen linearer Operatoren. Ann. Inst. Fourier (Grenoble) **23**, No. 4, 75–87 (1973). See also: Nagel, R.J.: Mean ergodic semigroups and invariant ideals in ordered Banach spaces. In: Foundations of Quantum Mechanics and Ordered Linear Spaces. Advanced Study Institute (Marburg 1973), pp. 309–315. Lecture Notes in Physics **29**. Berlin-Heidelberg-New York: Springer 1974
13. Spohn, H.: Approach to equilibrium for completely positive dynamical semigroups of N-level systems. Rep. Mathematical Phys. **10**, 189–194 (1976)
14. Spohn, H.: An algebraic condition for the approach to equilibrium of an open N-level system. Lett. Math. Phys. **2**, 33–38 (1977)
15. Watanabe, S.: Ergodic theorems for W^*-dynamical semigroups. Hokkaido Math. J. **8**, 176–190 (1979)
16. Watatani, Y., Enomoto, M.: A Perron-Frobenius type theorem for positive linear maps on C^*-algebras. Math. Japan **24**, 53–63 (1979)

Received March 31, 1981; received in final form January 15, 1982

Math. Z. 180, 287–290 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

A Note on Metrizable Linear Topologies Strictly Finer than a Given Metrizable Complete Linear Topology

Jerzy Kąkol

Institute of Mathematics, A. Mickiewicz University, ul. Matejki 48/49, PL-60-769 Poznań, Poland

Introduction

In [3] De Wilde and Tsirulnikov proved the following result: If (E, t) is a barrelled locally convex Hausdorff space and E has finite codimension in the B-complete completion of (E, t), then E admits a minimal barrelled Hausdorff topology weaker than t. This combined with Corollary 1.3 of [3] gives the result of Saxon and Wilansky ([8], 3.3) that every finite codimensional linear subspace of a Banach space has a complete norm weaker than the original one.

An extension of these results to non locally convex case was given by Drewnowski [5]. Among other he proved that finite codimensional linear subspace of a metrizable complete topological linear space has a metrizable complete linear topology weaker than the original one.

In this paper we apply the above results to constructions of metrizable linear topologies which are strictly finer than a given metrizable complete linear topology. We shall show that every metrizable complete topological linear space (E, t) has a strictly increasing sequence of metrizable ultrabarrelled linear topologies strictly finer than t. Moreover, if (E, t) is dual-less (i.e. (E, t) has no non-trivial continuous linear functionals [7]) then we can find a family $(t_f)_{f \in T}$ of non comparable and non dual-less metrizable ultrabarrelled topologies t_f strictly finer than t, where $\operatorname{card} T = \dim E^*$. For the locally convex case we obtain that every Banach space (E, t) has a strictly increasing sequence of normed non-barrelled (hence also non-ultrabarrelled) topologies finer than t.

Throughout, we consider only infinite dimensional Hausdorff topological linear spaces (t.l.s.) over the field $\boldsymbol{K}$ of either real or complex numbers. F being a subspace of a t.l.s. (E, t), we denote by $(F, t|F)$ the space F endowed with topology induced by t. We denote by $(E, t)'$ and E^* the topological and algebraic duals of (E, t), respectively. If t and u are two linear topologies on E, we denote by $\sup(t, u)$ [$\inf(t, u)$] the coarsest [finest] linear topology on E

0025-5874/82/0180/0287/$01.00

finer [coarser] than t and u. We recall ([10], p. 10) that a t.l.s. (E, t) is *ultrabarrelled* if we have $u \leqq t$ for every linear topology u on E which is t-polar, i.e. has a base of neighbourhoods of zero consisting of t-closed sets.

Results

The following fact will be useful in the sequel (see also [3] Proposition 2.2 (ii)).

Lemma 1. *Let t and u be two linear topologies on E. If the topology* $\sup(t, u)$ *is complete, then* $\inf(t, u)$ *is Hausdorff. The converse is true if t and u are complete.*

Proof. Suppose $\sup(t, u)$ is complete. It is enough to prove that the graph of the identity map from (E, t) onto (E, u) is closed. Let $(x_a)_{a \in A}$ be an arbitrary net in E such that $x_a \to 0$ for t and $x_a \to y$ for u. We have to prove $y=0$. Clearly $(x_a)_{a \in A}$ is a Cauchy net in $(E, \sup(t, u))$. There exists $z \in E$ such that $x_a \to z$ for $\sup(t, u)$. Hence $z=y=0$. Conversely, suppose $\inf(t, u)$ is Hausdorff. By $\mathscr{B}(t)$ and $\mathscr{B}(u)$ we denote the bases of t and u-neighbourhoods of zero, respectively. Let $(x_a)_{a \in A}$ be a Cauchy net in $(E, \sup(t, u))$. Hence $(x_a)_{a \in A}$ is a Cauchy net in (E, t) and (E, u). By assumption there are x and y such that $x_a \to x$ for t and $x_a \to y$ for u. Choose $U \in \mathscr{B}(t)$ and $V \in \mathscr{B}(u)$. Then there exists $a_0 \in A$ such that for all $a \geqq a_0$

$$x - y = x - x_a + x_a - y \in U + V.$$

Since

$$x - y \in \bigcap (U + V\colon U \in \mathscr{B}(t), V \in \mathscr{B}(u)) = \{0\},$$

it follows that $x=y$. Hence $(x_a)_{a \in A}$ converges to x in $(E, \sup(t, u))$, which completes the proof.

Proposition 2. *Let (E, t) be a metrizable and complete dual-less t.l.s. Then there exists a family $(t_f)_{f \in T}$ of metrizable complete non comparable linear topologies, where* $\operatorname{card} T = \dim E^*$, *satisfying the following properties.*

(i) $\sup(t, t_f)$ *is ultrabarrelled for every $f \in T$.*
(ii) $\dim(E, t_f)' = \dim(E, \sup(t, t_f))' = 1$ *for every $f \in T$.*
(iii) *For no element f of T the topology* $\inf(t, t_f)$ *is Hausdorff.*
(iv) *If (E, t) is separable, then every (E, t_f) is also separable.*

Proof. (i) Let T be a Hamel basis of E. For every $f \in T$ let $S_f = (x \in E\colon f(x)=0)$. Then there exists a metrizable and complete linear topology u_f on S_f weaker than $t|S_f$. Then the topological direct sum $(E, t_f) = (S_f, u_f) \oplus \boldsymbol{K}$ is a metrizable and complete t.l.s. and the topologies t and t_f are non comparable. Since by a result of Adasch, Ernst, Keim ([1], p. 90) every countable codimensional linear subspace of an ultrabarrelled space is ultrabarrelled, the space $(S_f, t|S_f)$ is ultrabarrelled. On the other hand S_f being $\sup(t, t_f)$-closed is topologically complemented and thus the space $(E, \sup(t, t_f))$ is ultrabarrelled. This completes the proof of (i).

(ii) Let t_f be fixed. Clearly f belongs to $(E, \sup(t, t_f))'$ and $\sup(t, t_f)|S_f = t|S_f$. We shall prove that $(x \in E\colon f(x)=0)$ is contained in $(x \in E\colon g(x)=0)$ for

every linear functional g of $(E, \sup(t, t_f))'$. Suppose that is false. Then there exists $x_0 \in S_f$ such that $g(x_0) \neq 0$. Since $g|S_f$ is $t|S_f$-continuous, it follows ([6], Proposition 5, p. 129) that there exists a t-continuous extension $\tilde{g}$ of g on E such that $\tilde{g}(x_0) \neq 0$, contrarily to $(E, t)' = \{0\}$; thus $(x \in E: f(x) = 0) \subset (x \in E: g(x) = 0)$. To finish the proof of (ii) is suffices to apply Theorem 1 of [9], p. 7.

Now we prove that the topologies t_f, $f \in T$ are non comparable. Suppose that for some f and g of T the topologies t_f and t_g are comparable (and hence $t_f = t_g$ as well). From (ii) f and g are linearly dependent, which contradicts the definition of T.

(iii) If $\inf(t, t_f)$ were Hausdorff, then $(E, \sup(t, t_f))$ would be a complete space by Lemma 1. It would then follows that t and t_f are comparable, a contradiction.

The case (iv) is obvious.

Corollary 3. *Let (E, t) be a metrizable and complete dual-less t.l.s. Then there exists a family $(u_f)_{f \in T}$ of metrizable ultrabarrelled non dual-less and non comparable linear topologies strictly finer than t, where* card $T = \dim E^*$.

Proof. For the proof it is enough to put $u_f = \sup(t, t_f)$, where t_f are obtained as before, and apply Proposition 2.

Corollary 4. *Let (E, t) be a metrizable and complete dual-less t.l.s. Then for every linear functional f over E there exists a metrizable and ultrabarrelled topology u_f on E strictly finer than t and such that $f \in (E, u_f)'$ and* $\dim(E, u_f)' = 1$.

If (E, t) is a dual-less t.l.s., then by [7], Theorem 3.4 there exists a linear topology v on E such that (E, v) and (E, t) are linearly homeomorphic and $\dim(E, \sup(t, v))' > 1$. The next corollary is also an easy consequence of the proof of Proposition 2.

Corollary 5. *Let (E, t) be a metrizable and complete dual-less t.l.s. Let v be another linear topology on E such that $(E, \sup(t, v))$ is non dual-less. Let B be a Hamel basis of $(E, \sup(t, v))'$. Then there exists a family $(t_f)_{f \in B}$ of metrizable ultrabarrelled non dual-less and non comparable linear topologies strictly finer than t and weaker than* $\sup(t, v)$.

In [8], Theorem 3.2 it is proved that every Banach space has a larger non-barrelled norm. The next result is more general.

Proposition 6. (a) *Every Banach space (E, t) has a strictly increasing sequence of normed non-barrelled (and hence non-ultrabarrelled) linear topologies t_n finer than t.*

(b) *Every metrizable and complete (not necessarily locally convex) t.l.s. (E, t) has a strictly increasing sequence of metrizable ultrabarrelled topologies strictly finer than t.*

Proof. First we prove (a). Let (E, t) be a Banach space. We choose the sequence $(t_n)_{n \in N}$ inductively. There exists a finer norm topology $t_1 > t$ such that (E, t_1) is non-barrelled. Suppose that $t_1, t_2, \ldots, t_n$ have been found such that the desired conditions are satisfied. Now we consider the topologies t and t_n. Let S be a maximal t_n-dense linear subspace of E. There exists a normed complete

topology u on S weaker than $t|S$ ([3], 3.3). The topological direct sum $(E, v) = (S, u) \oplus K$ is a normed complete space and v and t_n are non comparable. Since $v|S < t|S < t_n|S$, it follows that $\sup(v, t_n)|S = t_n|S$. From this and [2] we conclude that the space $(S, \sup(t_n, v)|S)$ is non-barrelled, which implies that $(E, \sup(t_n, v))$ is also non-barrelled. To conclude the proof it suffices to put $t_{n+1} = \sup(t_n, v)$.

(b) Applying the proof of Proposition 2 (i) it is possible to find a metrizable and complete linear topology v non comparable with t and such that $t_1 = \sup(t, v)$ is ultrabarrelled. Now it is enough to apply the construction in the proof of the case (a).

References

1. Adasch, N., Ernst, B., Keim, D.: Topological vector spaces. Lecture Notes in Mathematics **639**, Berlin-Heidelberg-New York: Springer 1978
2. De Wilde, M., Houet, C.: On increasing sequences of absolutely convex sets in locally convex spaces. Math. Ann. **192**, 257–261 (1971)
3. De Wilde, M., Tsirulnikov, B.: Barrelledness and the supremum of two locally convex topologies. Math. Ann. **246**, 241–248 (1980)
4. De Wilde, M., Tsirulnikov, B.: Barrelled spaces with a B-complete completion. Manuscripta Math. **33**, 411–427 (1981)
5. Drewnowski, L.: Weaker ultrabarrelled topologies on topological vector spaces. Preprint
6. Horvath, J.: Topological vector spaces and distributions. Reading: Addison, Wesley 1966
7. Peck, N.T., Porta, H.: Linear topologies which are suprema of dual-less topologies. Studia Math. **47**, 63–73 (1973)
8. Saxon, S.A., Wilansky, A.: The equivalence of some Banach space problems. Colloq. Math. **37**, 217–226 (1977)
9. Wilansky, A.: Modern methods in topological vector spaces. New York: McGraw-Hill 1978
10. Waelbroeck, L.: Topological vector spaces and algebras. Lecture Notes in Mathematics **230**. Berlin-Heidelberg-New York: Springer 1971

Received October 5, 1981

Math. Z. 180, 291-305 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Einige einfach-zusammenhängende Hilbertschemata

Gerd Gotzmann

Steinburgring 21, D-4440 Rheine, Bundesrepublik Deutschland

1. Einleitung

Sei $Q(T)$ ein Polynom mit rationalen Koeffizienten vom Grad $r \geqq 1$. Nach einem Satz von F.S. Macaulay gibt es in dem graduierten Polynomring $S = k[X_0, \ldots, X_r]$ über dem Körper k ein graduiertes Ideal I mit $\dim I_n = Q(n)$, $n \gg 0$, genau dann, wenn Q von der Form

$$Q(T) = \binom{T-a_0+r}{r} + \cdots + \binom{T-a_s+r-s}{r-s} \tag{1}$$

ist, wobei $0 < a_0 \leqq \ldots \leqq a_s$, $0 \leqq s \leqq r-1$ eindeutig bestimmte ganze Zahlen oder $s = 0$ und $a_0 = 0$ ist. Die Polynome der Form (1) mit $a_0 > 0$ sind also genau die, für die das Hilbertschema $H_Q = \mathrm{Hilb}_X^p$, $X = \mathbb{P}_k^r$, $P(T) = \binom{T+r}{r} - Q(T)$, nicht leer ist.

Mit G bezeichnen wir die algebraische k-Gruppe der strikt oberen Dreiecksmatrizen aus $\mathbf{GL}(r+1, k)$, die auf H_Q über die Automorphismen

$$X_i \rightsquigarrow X_i + \sum_{j=0}^{i-1} a_{ij} X_j, \qquad a_{ij} \in A, \ 0 \leqq i \leqq r,$$

der A-Algebra $S \otimes_k A$ operiert, wobei A eine noethersche k-Algebra ist. Im folgenden wird gezeigt:

Satz 1. *Wenn in dem Polynom* (1) *die Zahl* $s \leqq r-2$ *ist, dann hat das Fixpunktschema* H_Q^G *einen einzigen Punkt.*

Im Fall $s = r-1$ gilt eine entsprechende Aussage nicht, wie das von Horrocks angegebene Beispiel der unter G invarianten Ideale (X_0^2, X_0X_1, X_1^2) und (X_0, X_1^3) in $k[X_0, X_1, X_2]$, die das Hilbertpolynom $\binom{T-1+2}{2} + \binom{T-3+1}{1}$ haben, zeigt.

Bei dem induktiven Beweis ist es nötig, einen entsprechenden Satz für eine anscheinend größere Klasse von Idealen von $\mathcal{O}_X$ zu beweisen, die numerisch

0025-5874/82/0180/0291/$03.00

maximal genannt werden. Der Zusammenhang mit dem Hilbertschema wird in Abschnitt 4 hergestellt. Nach Sätzen von Horrocks folgt für bestimmte Formen der Hilbertpolynome der einfache Zusammenhang von H_Q. Es wäre deshalb interessant, ein Polynom Q zu finden, für das H_Q nicht einfach zusammenhängend ist.

2. Ergebnisse und Hilfssätze über monomiale Ideale und Hilbertpolynome

(2.1) Die Zahl a_s in (1) wird mit $n(Q)$ bezeichnet; sie ist eine Regularitätsschranke für Ideale $\mathscr{I} \subset \mathscr{O}_X$ mit Hilbertpolynom Q, d.h. es ist $H^i(X, \mathscr{I}(n-i)) = (0)$ für alle $n \geqq n(Q)$, $i>0$, folglich $h^0(\mathscr{I}(n)) = \dim H^0(\mathscr{I}(n)) = Q(n)$ und $S_1 H^0(\mathscr{I}(n)) = H^0(\mathscr{I}(n+1))$ (vgl. [9] und [2], Lemma 2.9).

(2.2) Für irgendeine Menge $M \subset S_d$ von Monomen sei $S_\nu M = \{n \cdot m \mid n \in S_\nu$ Monom, $m \in M\}$. Zu jeder Zahl $0 < a \leqq \binom{d+r}{r}$ gibt es ein eindeutig bestimmtes Polynom Q wie in (1) mit $a = Q(d)$. Sei $M(d,a)$ die Menge der a ersten Monome vom Grad d in der lexikographischen Ordnung, N das von $M(d,a)$ erzeugte Ideal, $\mathscr{N}$ die zu N assoziierte Idealgarbe. Dann ist $N_n = H^0(X, \mathscr{N}(n))$ und $\dim N_n = Q(n)$, $n \geqq d$. Wenn $V \subset S_d$ irgendein Vektorraum mit $\dim V \geqq a = Q(d)$ und $S_1 V = \langle lv \mid l \in S_1, v \in V \rangle$ ist, dann besagt ein Satz von Macaulay, daß $\dim S_1 V \geqq Q(d+1)$ ist (vgl. [2], [10], ...).

(2.3) Für $d \geqq n(Q)$ ist

(i) $M(d, Q(d)) = X_0^{a_0} S_{d-a_0}$ im Fall $s=0$;

(ii) $M(d, Q(d)) = X_0^{a_0} S_{d-a_0} \bigcup_{1 \leqq \nu \leqq s} X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{\nu-1}^{a_{\nu-1}-a_{\nu-2}} X_\nu^{a_\nu - a_{\nu-1}+1} S_{d-a_\nu}$ im Fall $s>0$.

Beweis durch Induktion nach s. Zunächst ist $S_{d-a_s} M(a_s, Q(a_s)) = M(d, Q(d))$ ([10], p. 151) und daher (i) richtig im Fall $s=0$, da dann $M(a_0, Q(a_0)) = X_0^{a_0}$ ist. Sei $h(T) = Q(T) - \binom{T-a_s+r-s}{r-s}$. Nach Induktionsvoraussetzung ist

$$M(a_s, h(a_s)) = X_0^{a_0} \cdot S_{a_s - a_0} \cup \ldots \cup X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{s-1}^{a_{s-1}-a_{s-2}+1} S_{a_s - a_{s-1}}.$$

Da $h(a_s)+1 = Q(a_s)$, ist

$$M(a_s, Q(a_s)) = M(a_s, h(a_s)) \cup \{X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{s-1}^{a_{s-1}-a_{s-2}} X_s^{a_s - a_{s-1}+1}\},$$

denn das rechts hinzugefügte Monom ist das auf das letzte Monom von $M(a_s, h(a_s))$ folgende Monom in der lexikographischen Ordnung. Multiplikation mit S_{d-a_s} gibt die Formel (ii).

(2.4) Sei $d \geqq n(Q)$ und $\mathscr{N}$ die von $M(d, Q(d))$ erzeugte Idealgarbe, $\mathscr{I} \subset \mathscr{O}_X$ irgendeine Idealgarbe mit Hilbertpolynom $\chi_{\mathscr{I}}(T) = Q(T)$. Dann ist für alle

$0 \leqq t \leqq r-1$ und alle $a_t \leqq n < a_{t+1}$ die Dimension $h^0(\mathscr{I}(n))$ von $H^0(\mathscr{I}(n)) \leqq \binom{n-a_0+r}{r} + \cdots + \binom{n-a_t+r-t}{r-t}$.

Beweis. Man kann $t<s$ annehmen, da für $n \geqq a_s$ immer $h^0(\mathscr{I}(n))=Q(n)$ ist (2.1). Aus

$$\dim H^0(\mathscr{I}(n)) \geqq \binom{n-a_0+r}{r} + \cdots + \binom{n-a_t+r-t}{r-t} + \binom{n-n+r-t-1}{r-t-1}$$

folgt mit (2.2)

$$\dim H^0(\mathscr{I}(d)) \geqq \dim S_{d-n} H^0(\mathscr{I}(n)) \geqq \binom{d-a_0+r}{r} + \cdots + \binom{d-n+r-t-1}{r-t-1} > Q(d),$$

wenn man $n < a_{t+1} \leqq \ldots \leqq a_s$ und die Formel $\sum_{\rho=0}^{\rho=b} \binom{a+\rho}{\rho} = \binom{a+b+1}{b}$ beachtet.

(2.5) Für alle t und alle $a_t \leqq n < a_{t+1}$ ist

$$H^0(\mathscr{N}(n)) = X_0^{a_0} S_{n-a_0} + X_0^{a_0-1} X_1^{a_1-a_0+1} S_{n-a_1} + \cdots + X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{t-1}^{a_{t-1}-a_{t-2}} X_t^{a_t-a_{t-1}+1} S_{n-a_t},$$

folglich

$$h^0(\mathscr{N}(n)) = \binom{n-a_0+r}{r} + \cdots + \binom{n-a_t+r-t}{r-t}.$$

Beweis. Da $H^0(\mathscr{N}(a_s)) = \langle M(a_s, Q(a_s)) \rangle$ und X_r Nicht-Nullteiler auf S/N ist, wird $H^0(\mathscr{N}(n))$ von den Monomen

$$\{m \in S_n \text{ Monom} \mid X_r^{a_s-n} m \in M(a_s, Q(a_s))\}$$

erzeugt. Mit (2.3) folgt leicht die erste Aussage und hieraus die zweite.

(2.6) Das Polynom $Q'(T)=Q(T)-Q(T-1)$ ist wieder von der Form (1), und wenn N' das von $M(d, Q'(d))$ in $S'=S/X_rS(-1)=k[X_0, \ldots, X_{r-1}]$ erzeugte Ideal und $\mathscr{N}'=\tilde{N}'$ die assoziierte Idealgarbe in $\mathscr{O}_{X'}$, $X'=\mathbb{P}_k^{r-1}$ ist, dann ist $N'=N+X_rS(-1)/X_rS(-1)$ und

$$0 \longrightarrow (\mathscr{O}_X/\mathscr{N})(-1) \xrightarrow{X_r} \mathscr{O}_X/\mathscr{N} \longrightarrow \mathscr{O}_{X'}/\mathscr{N}' \longrightarrow 0$$

exakt.

Beweis. $N+X_rS(-1)/X_rS(-1)$ wird von $M(d, Q(d))-X_rS$ in S' erzeugt. Mit der Formel in (2.3) sieht man ohne Schwierigkeiten ein, daß diese Menge gleich $M(d, Q(d))-X_rM(d-1, Q(d-1))$ ist, daher aus den $Q'(d)=Q(d)-Q(d-1)$ ersten Monomen in $X_0, \ldots, X_{r-1}$ in der lexikographischen Ordnung besteht.

Da

$$0 \to S_n/N_n \to S_{n+1}/N_{n+1} \to S'_{n+1}/N'_{n+1} \to 0$$

exakt ist, $n \geqq d$, folgt die angegebene exakte Sequenz durch Übergang zu den Garben.

(2.7) Wenn $Q(T) = \binom{T-a_0+r}{r} + \cdots + \binom{T-a_{r-1}+1}{1}$ und $a_{r-3} < a_{r-2}$ ist, dann ist

(i) $H^0(\mathcal{N}(n)) \to H^0(\mathcal{N}'(n))$ surjektiv für $n < a_{r-2} - 1$ und

(ii) $\dim H^0(\mathcal{N}'(n))/\mathrm{Im}(H^0(\mathcal{N}(n)) = 1$ für $n = a_{r-2} - 1$.

Beweis. Man bekommt

$$Q'(T) = \binom{T-a_0+r-1}{r-1} + \cdots + \binom{T-a_{r-3}+2}{2} + \binom{T-(a_{r-2}-1)+1}{1}.$$

I. $a_{r-3} = a_{r-2} - 1$

1. Fall. Sei $a_{l-1} < a_l = \cdots = a_{r-3}$ mit $0 \leqq l \leqq r-3$ und $a_{l-1} \leqq n < a_l$. Dann ist nach (2.5)

$$H^0(\mathcal{N}(n)) = X_0^{a_0} S_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{l-1}^{a_{l-1}-a_{l-2}+1} S_{n-a_{l-1}}$$

und

$$H^0(\mathcal{N}'(n)) = X_0^{a_0} S'_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{l-1}^{a_{l-1}-a_{l-2}+1} S'_{n-a_{l-1}}$$

2. Fall. $n = a_{r-3}$. Dann ist

$$H^0(\mathcal{N}(n)) = X_0^{a_0} S_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{r-3}^{a_{r-3}-a_{r-4}+1} S_0$$

und

$$H^0(\mathcal{N}'(n)) = X_0^{a_0} S'_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{r-3}^{a_{r-3}-a_{r-4}+1} S'_0 + X_0^{a_0-1} \ldots X_{r-3}^{a_{r-3}-a_{r-4}} X_{r-2}^{(a_{r-2}-1)-a_{r-3}+1} S'_0.$$

II. $a_{r-3} < a_{r-2} - 1$.

Mit (2.5) folgt wieder die Surjektivität wie in I. Wenn $n = a_{r-2} - 1$, bekommt man

$$H^0(\mathcal{N}(n)) = X_0^{a_0} S_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{r-3}^{a_{r-3}-a_{r-4}+1} S_{n-a_{r-3}}$$

und

$$H^0(\mathcal{N}'(n)) = X_0^{a_0} S'_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{r-3}^{a_{r-3}-a_{r-4}+1} S'_{n-a_{r-3}} + X_0^{a_0-1} \ldots X_{r-2}^{(a_{r-2}-1)-a_{r-3}+1} S'_0.$$

Daraus folgen die Aussagen (i) und (ii).

(2.8) Man kann sich bei der Untersuchung von Idealen $I \subset S$ mit Hilbertpolynom Q meistens auf den Fall $a_0 = 1$ beschränken. Wenn man nämlich definiert

$$e(I) = \max\{v \in \mathbb{N} \mid \text{es gibt ein } u \in S_v \text{ mit } I \subset uS\},$$

dann hat bei algebraisch abgeschlossenem k das Hilbertpolynom $\chi_{S/I}$ von S/I die Form

$$e(T)\cdot\frac{T^{r-1}}{(r-1)!}+\text{Terme mit niedrigerem Grad.}$$

Außerdem ist

$$\binom{T+r}{r}-Q(T)=\sum_{\rho=0}^{\rho=r}\binom{T-1+\rho}{\rho}-Q(T)=\binom{T-1+r}{r}-\binom{T-a_0+r}{r}$$
$$+(\text{Terme vom Grad } <r-1)=(a_0-1)\frac{T^{r-1}}{(r-1)!}$$
$$+(\text{Terme vom Grad } <r-1).$$

3. Relativ maximale Ideale

Im folgenden soll k algebraisch abgeschlossen sein.

(3.1) **Definition.** Ein Ideal $\mathscr{J}\subset\mathscr{O}_X$ mit Hilbertpolynom Q heißt

(i) *linear maximal*, wenn es kein Ideal $\mathscr{J}\subset\mathscr{O}_X$ mit $\mathscr{J}\subsetneqq\mathscr{J}$ und $\chi_{\mathscr{J}/\mathscr{J}}(T)$ =konstant gibt;

(ii) *numerisch maximal*, wenn $\dim H^0(\mathscr{J}(n))=\dim H^0(\mathscr{N}(n))$ für alle $n\in\mathbb{N}$ ist (vgl. (2.4) und (2.5)).

Bemerkung 1. Wenn im Hilbertpolynom Q von $\mathscr{J}$ $s<r-1$ ist, dann ist $\mathscr{J}$ linear maximal. Wenn $s=r-1$, dann gilt für jedes linear maximale Ideal $\mathscr{L}\supset\mathscr{J}$ die Ungleichung $\chi_{\mathscr{L}}(T)-Q(T)\leqq a_{r-1}-a_{r-2}+1$. Dies folgt sofort aus $a_0\leqq\ldots\leqq a_s$ in (1).

Im ganzen Abschnitt soll das Ideal $\mathscr{J}\subset\mathscr{O}_X$ mit Hilbertpolynom Q invariant unter der kanonischen Operation von $G(k)$ auf $X=\mathbb{P}_k^r$ sein. Dann ist auch $I=\bigoplus_{v\geqq 0}H^0(\mathscr{J}(v))$ invariant unter $G(k)$.

Bemerkung 2. Die Linearform $l=X_r$ ist kein Nullteiler in S/I.

Wenn l in einem zu I assoziierten graduierten Primideal $\mathfrak{p}$ liegen würde, dann wegen der Invarianz von $\mathfrak{p}$ unter G auch $X_0,\ldots,X_r$, so daß $\mathfrak{p}=(X_0,\ldots,X_r)$ wäre.

Bemerkung 3. Das Ideal $I'=I+lS(-1)/lS(-1)\subset S/lS(-1)=S'$ hat das Hilbertpolynom Q'. Das Ideal $\mathscr{J}'=\mathscr{J}+l\mathscr{O}_X(-1)/l\mathscr{O}_X(-1)\subset\mathscr{O}_{X'}$ ist die zu I' assoziierte Garbe, daher $H^0(\mathscr{J}'(n))\subset S'_n\subset S_n$.

Dies folgt aus der exakten Sequenz

$$0\to(S/I)(-1)\xrightarrow{l}S/I\to S'/I'\to 0.$$

(3.2) **Hilfssatz.** *Für alle $n\in\mathbb{N}$ ist*

$$H^0(\mathscr{J}(n))\subset H^0(\mathscr{J}'(n))\oplus H^0(\mathscr{J}'(n-1))X_r\oplus\cdots\oplus H^0(\mathscr{J}'(0))X_r^n.$$

Beweis. Sei c die größte ganze Zahl, für die es ein Ideal $\mathscr{L}\supset\mathscr{J}$ gibt mit $\chi_{\mathscr{L}/\mathscr{J}}=c$. Dann ist $\mathscr{L}$ linear maximal. Der Funktor $\mathrm{Quot}^c(\mathscr{F}/X\times_k-/-)$, $\mathscr{F}=\mathscr{O}_X/\mathscr{J}$,

auf der Kategorie der lokal noetherschen k-Schemata, der die Quotientenmoduln von $\mathscr{F}$ mit Hilbertpolynom c parametrisiert, wird durch ein projektives k-Schema Z dargestellt. Da k algebraisch abgeschlossen ist, folgt $\mathscr{I}\otimes\mathcal{O}_Y$ invariant unter $G(Y)$ für alle k-Schemata Y; daher operiert G auf Z, und weil $\mathscr{L}/\mathscr{I}\in Z(k)$ ist, kann man nach dem Borelschen Fixpunktsatz o.E. annehmen, daß $\mathscr{L}$ invariant unter G ist. Da $0\to(\mathcal{O}_X/\mathscr{L})(-1)\xrightarrow{l}\mathcal{O}_X/\mathscr{L}$, $l=X_r$, exakt ist, hat $\mathscr{L}'=\mathscr{L}+l\mathcal{O}_X(-1)/l\mathcal{O}_X(-1)$ das Hilbertpolynom Q', und wegen $\mathscr{L}'\supset\mathscr{I}'$ ist $\mathscr{L}'=\mathscr{I}'$.

Sei $L=\bigoplus_{\nu\geqq 0}H^0(\mathscr{L}(\nu))$ und J das von L und $\bigoplus_{\nu\geqq 0}H^0(\mathscr{I}'(\nu))$ erzeugte Ideal in S. Da J invariant unter G ist, ist

$$0\to(\mathcal{O}_X/\tilde{J})(-1)\xrightarrow{l}\mathcal{O}_X/\tilde{J}\to\mathcal{O}_{X'}/\tilde{J}'\to 0$$

exakt, $\tilde{J}'=\tilde{J}+l\mathcal{O}_X(-1)/l\mathcal{O}_X(-1)=\mathscr{L}'=\mathscr{I}'$. Wegen der Maximalität von $\mathscr{L}$ muß $\mathscr{L}=\tilde{J}$ und daher $H^0(\mathscr{L}(n))\supset H^0(\mathscr{I}'(n))=H^0(\mathscr{L}'(n))$ sein.

Schreibt man $x\in H^0(\mathscr{L}(n))$ in der Form $x=u+lv$ mit $u\in S'_n$, $v\in S_{n-1}$, dann ist $u\in H^0(\mathscr{L}'(n))\subset H^0(\mathscr{L}(n))$, daher $lv\in H^0(\mathscr{L}(n))$ und folglich $v\in H^0(\mathscr{L}(n-1))$. So bekommt man nach und nach

$$H^0(\mathscr{I}(n))\subset H^0(\mathscr{L}(n))\subset H^0(\mathscr{I}'(n))\oplus\cdots\oplus H^0(\mathscr{I}'(0))l^n.$$

(3.3) **Hilfssatz.** *Wenn $\mathscr{I}$ linear maximal ist, dann ist*

$$H^0(\mathscr{I}(n))=H^0(\mathscr{I}'(n))\oplus\cdots\oplus H^0(\mathscr{I}'(0))X_r^n.$$

Beweis. Durch $J_n=H^0(\mathscr{I}'(n))\oplus\cdots\oplus H^0(\mathscr{I}'(0))X_r^n$ wird ein Ideal $J\subset S$ bzw. $\mathscr{J}\subset\mathcal{O}_X$ definiert, das I bzw. $\mathscr{I}$ enthält und invariant unter G ist. Wegen $\mathscr{J}'=\mathscr{I}'$ muß $\mathscr{J}/\mathscr{I}$ ein konstantes Hilbertpolynom haben und deshalb $\mathscr{J}=\mathscr{I}$ sein. So folgt $J_n\subset H^0(\mathscr{I}(n))$.

(3.4) **Hilfssatz.** *Wenn in dem Hilbertpolynom Q die Zahl $s\leqq r-2$ ist, dann sind $\mathscr{I}'$ und $\mathscr{I}$ numerisch maximal.*

Beweis. Mit Bemerkung 1 und den Hilfssätzen (3.3), (2.4) und (2.5) folgt

$$\begin{aligned}h^0(\mathscr{I}(n))&=h^0(\mathscr{I}'(n))+\cdots+h^0(\mathscr{I}'(0))\\&\leqq h^0(\mathscr{N}'(n))+\cdots+h^0(\mathscr{N}'(0))=h^0(\mathscr{N}(n)).\end{aligned}$$

Wäre hier $h^0(\mathscr{I}'(m))<h^0(\mathscr{N}'(m))$ für ein $m\in\mathbb{N}$, dann würde folgen $Q(n)=h^0(\mathscr{I}(n))<h^0(\mathscr{N}(n))=Q(n)$, $n\gg 0$. Deshalb gilt in den angegebenen Ungleichungen das Gleichheitszeichen.

(3.5) **Hilfssatz.** *Wenn $\mathscr{I}$ numerisch maximal ist, dann ist auch $\mathscr{I}'$ numerisch maximal.*

Beweis. Wegen (3.4) kann man o.E. annehmen:

$$Q(T)=\binom{T-a_0+r}{r}+\cdots+\binom{T-a_{r-1}+1}{1}.$$

Wenn hier $a_{l-1}<a_l=\cdots=a_{r-2}$ und $0\leqq l\leqq r-2$ sind, dann ist

$$Q'(T)=\binom{T-a_0+r-1}{r-1}+\cdots+\binom{T-(a_l-1)+r-l-1}{r-l-1}.$$

Im Fall $r-l-1\geqq 2$ gibt wieder (3.4) die Behauptung, so daß nur der Fall $l=r-2$, d.h.

$$a_{r-3}<a_{r-2}$$

übrigbleibt. Aus der exakten Sequenz

$$0\to\mathscr{I}(n-1)\to\mathscr{I}(n)\to\mathscr{I}'(n)\to 0,$$

in der der zweite Pfeil die Multiplikation mit X_r ist, erhält man für $n\geqq m=a_{r-2}-1$ exakte Sequenzen

$$\begin{aligned}0\to H^0(\mathscr{I}(n-1))&\to H^0(\mathscr{I}(n))\to H^0(\mathscr{I}'(n))\\&\to H^1(\mathscr{I}(n-1))\to H^1(\mathscr{I}(n))\to 0\end{aligned}\tag{2}$$

und

$$H^i(\mathscr{I}(n-i))=0\quad\text{für } i\geqq 2,\ n\geqq m,\tag{3}$$

da $\mathscr{I}'$ m-regulär ist ([2], 2.9 und [9], p. 99). Außerdem hat man für $\mathscr{N}$ entsprechende exakte Sequenzen und verschwindende Kohomologiegruppen. Deshalb ist

$$h^0(\mathscr{I}'(n))\geqq h^0(\mathscr{I}(n))-h^0(\mathscr{I}(n-1))=h^0(\mathscr{N}(n))-h^0(\mathscr{N}(n-1)).$$

1. Fall. $n<m$. Hilfssatz (2.7) gibt dann $h^0(\mathscr{I}'(n))\geqq h^0(\mathscr{N}'(n))$.

2. Fall. $n=m$. Zuerst zeigen wir mit einem Schluß wie in ([9], p. 102), daß $H^0(\mathscr{I}(m))\to H^0(\mathscr{I}'(m))$ nicht surjektiv ist. Nimmt man nämlich dies an, dann folgt aus der Betrachtung des kommutativen Diagramms

$$\begin{array}{ccc}H^0(\mathscr{I}(m))\otimes S_\nu & \longrightarrow & H^0(\mathscr{I}'(m))\otimes S_\nu\\ \downarrow & & \downarrow\\ H^0(\mathscr{I}(m+\nu)) & \longrightarrow & H^0(\mathscr{I}'(m+\nu))\end{array}$$

mit surjektivem zweiten senkrechten Pfeil die Surjektivität von $H^0(\mathscr{I}(n))\to H^0(\mathscr{I}'(n))$ für alle $n\geqq m$. Dann folgt mit (2) für alle $n\geqq m$, daß $H^1(\mathscr{I}(n-1)\simeq H^1(\mathscr{I}(n))$ und deshalb $H^1(\mathscr{I}(m-1))=0$ ist. Hieraus folgt mit (3)

$$\begin{aligned}h^0(\mathscr{I}(m-1))&=Q(m-1)+h^1(\mathscr{I}(m-1))\\&=Q(m-1)\\&<Q(m-1)+h^1(\mathscr{N}(m-1))\\&=h^0(\mathscr{N}(m-1)).\end{aligned}$$

Hierbei ist $h^1(\mathscr{N}(m-1))\neq 0$, da $H^0(\mathscr{N}(m))\to H^0(\mathscr{N}'(m))$ nicht surjektiv ist (2.7). Insgesamt erhält man einen Widerspruch zur vorausgesetzten numerischen Maximalität von $\mathscr{I}$.

Aus (2) folgt mit dem gerade Bewiesenen und (2.7)

$$\begin{aligned} h^0(\mathscr{I}'(m)) &\geqq h^0(\mathscr{I}(m)) - h^0(\mathscr{I}(m-1)) + 1 \\ &= h^0(\mathscr{N}(m)) - h^0(\mathscr{N}(m-1)) + 1 \\ &= h^0(\mathscr{N}'(m)). \end{aligned}$$

Wenn schließlich $n > m$ ist, hat man wegen der m-Regularität von $\mathscr{I}'$ und $\mathscr{N}'$

$$h^0(\mathscr{I}'(n)) = Q'(n) = h^0(\mathscr{N}'(n)).$$

(3.6) **Satz.** *Wenn k algebraisch abgeschlossen ist, $\mathscr{I}$ das Hilbertpolynom Q hat, numerisch maximal und invariant unter $G(k)$ ist, dann ist $\mathscr{I} = \mathscr{N}$ das von den Monomen $M(d, Q(d))$ mit $d \geqq n(Q)$ erzeugte Ideal.*

Zunächst kann man im Fall $a_0 > 1$ das Ideal $\mathscr{I}$ in der Form $\mathscr{I} = u\mathscr{J}$ mit einem $u \in S_{a_0-1}$ und einem Ideal $\mathscr{J} \subset \mathscr{O}_X$ schreiben (2.8). Da u invariant unter $G(k)$ sein muß, folgt $u = X_0^{a_0-1}$, so daß es reichen würde, die Behauptung für $\mathscr{J}$ nachzuweisen. Im Fall $r = 1$ ist $n(Q) = a_0$ und $Q(a_0) = 1$, so daß $\mathscr{I}$ von einer Form vom Grad a_0 erzeugt wird, die invariant unter $G(k)$ ist, daher gleich $X_0^{a_0}$ sein muß (2.1). Man kann für den folgenden Induktionsschluß $r \geqq 2$ und die Richtigkeit des Satzes für $r-1$ voraussetzen. Wenn $s \leqq r-2$ ist, dann ergibt sich aus der linearen Maximalität (Bemerkung 1), daß

$$H^0(\mathscr{I}(n)) = H^0(\mathscr{I}'(n)) \oplus \cdots \oplus H^0(\mathscr{I}'(0)) X_r^n$$

ist (3.3). Da $\mathscr{I}'$ numerisch maximal ist (3.5), ist nach Induktionsvoraussetzung $\mathscr{N}' = \mathscr{I}'$, und mit

$$H^0(\mathscr{N}(n)) = H^0(\mathscr{N}'(n)) \oplus \cdots \oplus H^0(\mathscr{N}'(0)) X_r^n$$

(3.3) folgt $\mathscr{I} = \mathscr{N}$.

Es soll im folgenden sein:

$$Q(T) = \binom{T-a_0+r}{r} + \cdots + \binom{T-a_{r-1}+1}{1},$$

$a_0 = 1$, $r \geqq 2$ und $a_{l-1} < a_l = \cdots = a_{r-2}$ mit $0 \leqq l \leqq r-2$. Dann ist

$$Q'(T) = \binom{T-a_0+r-1}{r-1} + \cdots + \binom{T-a_{l-1}+r-l}{r-l} + \binom{T-(a_l-1)+r-l-1}{r-l-1},$$

und mit

$$H(T) = \binom{T-a_0+r}{r} + \cdots + \binom{T-a_{l-1}+r-l+1}{r-l+1} + \binom{T-(a_l-1)+r-l}{r-l}$$

ist auch $H'(T) = Q'(T)$. Sei M das von $M(d, H(d))$, $d \geqq a_l$, in S erzeugte Ideal, $\mathscr{M}$ die assoziierte Garbe. Wegen $l \leqq r-2$ ist $\mathscr{M}$ linear maximal; aus $\mathscr{M}' = \mathscr{N}'$ und $\mathscr{N}' = \mathscr{I}'$ (Induktionsvoraussetzung und (3.5)) folgt mit (3.3) und (3.2)

(#) $H^0(\mathscr{I}(n)) \subset H^0(\mathscr{M}(n)) = H^0(\mathscr{N}'(n)) \oplus \cdots \oplus H^0(\mathscr{N}'(0)) X_r^n$ für alle $n \in \mathbb{N}$.

I. $l=0$. Dann ist $1=a_0=\cdots=a_{r-2}$.

1. Fall. $a_{r-1}=1$. Da $\mathscr{I}$ 1-regulär ist, folgt mit der Invarianz von $\mathscr{I}$ unter G, daß $H^0(\mathscr{I}(1))=\langle X_0,\ldots,X_{r-1}\rangle$, daher $\mathscr{I}=\mathscr{N}$ ist.

2. Fall. $1<a_{r-1}$. Nach Voraussetzung ist $h^0(\mathscr{I}(1))=h^0(\mathscr{N}(1))=r-1$, daher wieder $H^0(\mathscr{I}(1))=\langle X_0,\ldots,X_{r-2}\rangle=H^0(\mathscr{N}(1))$ (2.5).

Wir wollen zeigen, daß auch für $l>0$ immer gilt

$$(*)\qquad H^0(\mathscr{I}(a_l))=H^0(\mathscr{N}(a_l)).$$

II. Beweis von (*) für $l>0$.

1° $a_{l-1}=a_l-1$, und sei etwa $a_{t-1}<a_t=\cdots=a_{l-1}$ mit $0\leqq t\leqq l-1$.

1. Fall. $t=0$. Dann ist $1=a_0=\cdots=a_{l-1}$, daher $H^0(\mathscr{I}(a_{l-1}))=\langle X_0,\ldots,X_{l-1}\rangle$. Weil $H^0(\mathscr{M}(a_l))=X_0S_1+\cdots+X_lS_1$ ist, hat man folgende Inklusionen von k-Vektorräumen:

$$U=H^0(\mathscr{I}(a_l))/S_1H^0(\mathscr{I}(a_{l-1}))$$
$$\hookrightarrow H^0(\mathscr{M}(a_l))/S_1H^0(\mathscr{I}(a_{l-1}))\xrightarrow{\sim}X_l^2\cdot k\oplus\cdots\oplus X_lX_r\cdot k=V.$$

Sei G' die Untergruppe von G der Automorphismen von S

$$X_i\mapsto X_i+\sum_{j=l}^{j=i-1}a_{ij}X_j,\qquad l+1\leqq i\leqq r,$$
$$X_i\mapsto X_i,\qquad 0\leqq i\leqq l.$$

Da U invariant unter der natürlichen Operation von G' auf V ist, folgt

$$U=\langle X_l^2,\ldots,X_lX_{r-1}\rangle\quad\text{im Fall } a_{r-2}=a_{r-1},$$

bzw.

$$U=\langle X_l^2,\ldots,X_lX_{r-2}\rangle\quad\text{im Fall } a_{r-2}<a_{r-1}.$$

Beidesmal ist (*) erfüllt.

2. Fall. $t>0$. Aus $h^0(\mathscr{I}(n))=h^0(\mathscr{N}(n))=h^0(\mathscr{M}(n))$ für $n\leqq a_{t-1}$ (nach Voraussetzung und (2.5)) folgt mit (#)

$$H^0(\mathscr{I}(a_{t-1}))=H^0(\mathscr{M}(a_{t-1}))$$
$$=X_0^{a_0}S_{a_{t-1}-a_0}+\cdots+X_0^{a_0-1}X_1^{a_1-a_0}\ldots X_{t-1}^{a_{t-1}-a_{t-2}+1}S_0,$$

daher

$$V=H^0(\mathscr{M}(a_t))/S_{a_t-a_{t-1}}H^0(\mathscr{M}(a_{t-1}))$$
$$\xrightarrow{\sim}\mu X_tk\oplus\cdots\oplus\mu X_lk,\qquad \mu=X_0^{a_0-1}\ldots X_t^{a_t-a_{t-1}}.$$

Wenn man jetzt die Untergruppe G' der Automorphismen

$$X_i\mapsto X_i+\sum_{j=t}^{j=i-1}a_{ij}X_j,\qquad t+1\leqq i\leqq l,$$
$$X_i\mapsto X_i\quad\text{für die anderen } i,$$

auf V operieren läßt, dann sieht man, daß der $(l-t)$-dimensionale Untervektorraum $H^0(\mathscr{I}(a_t))/S_{a_t-a_{t-1}}H^0(\mathscr{M}(a_{t-1}))$ gleich $\mu X_t k \oplus \ldots \oplus \mu X_{l-1} k$ sein muß. Deshalb ist $H^0(\mathscr{I}(a_t)) = H^0(\mathscr{I}(a_{l-1})) = H^0(\mathscr{N}(a_{l-1}))$. Weiter hat man

$$h^0(\mathscr{I}(a_l)) = \binom{a_l - a_0 + r}{r} + \cdots + \binom{a_l - a_{l-1} + r - l + 1}{r - l + 1} + (r-l)$$

für $a_{r-2} = a_{r-1}$, bzw.

$$h^0(\mathscr{I}(a_l)) = \binom{a_l - a_0 + r}{r} + \cdots + \binom{a_l - a_{l-1} + r - l + 1}{r - l + 1} + (r-l-1)$$

für $a_{r-2} < a_{r-1}$.

Außerdem ist $H^0(\mathscr{M}(a_l)) = S_1 H^0(\mathscr{N}(a_{l-1})) + \mu X_l S_1$, $\mu = X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{l-1}^{a_{l-1}-a_{l-2}}$. Wegen der Invarianz des Untervektorraumes

$$U = H^0(\mathscr{I}(a_l))/S_1 H^0(\mathscr{N}(a_{l-1})) \hookrightarrow H^0(\mathscr{M}(a_l))/S_1 H^0(\mathscr{N}(a_{l-1}))$$
$$\xrightarrow{\sim} \mu X_l^2 k \oplus \cdots \oplus \mu X_l X_r k$$

folgt ähnlich wie oben, daß

$$U = \mu X_l^2 k \oplus \cdots \oplus \mu X_l X_{r-1} k \qquad \text{im Fall } a_{r-2} = a_{r-1}$$

bzw.

$$U = \mu X_l^2 k \oplus \cdots \oplus \mu X_l X_{r-2} k \qquad \text{im Fall } a_{r-2} < a_{r-1}$$

ist. Beide Male folgt (*).

2°. $a_{l-1} \leqq a_l - 2$, und es sei

$$h(T) = \binom{T - a_0 + r}{r} + \cdots + \binom{T - a_{l-1} + r - l + 1}{r - l + 1}.$$

Aus $h^0(\mathscr{I}(n)) = h^0(\mathscr{N}(n)) = h^0(\mathscr{M}(n))$ für $n = a_{l-1}$ und (#) folgt $H^0(\mathscr{I}(n)) = H^0(\mathscr{N}(n))$ für $n = a_{l-1}$; wegen $h^0(\mathscr{I}(n)) = h^0(\mathscr{N}(n))$ und $S_{a_l-1-a_{l-1}} \cdot H^0(\mathscr{N}(a_{l-1})) = H^0(\mathscr{N}(a_l - 1))$ (vgl. (2.5)) gilt dies auch für $n = a_l - 1$, d.h.

$$H^0(\mathscr{I}(n)) = X_0^{a_0} S_{n-a_0} + \cdots + X_0^{a_0-1} \ldots X_{l-1}^{a_{l-1}-a_{l-2}+1} S_{n-a_{l-1}}.$$

Wegen $H^0(\mathscr{M}(a_l)) = S_1 H^0(\mathscr{N}(a_l - 1)) + \mu \cdot S_1$ mit

$$\mu = X_0^{a_0-1} X_1^{a_1-a_0} \ldots X_{l-1}^{a_{l-1}-a_{l-2}} X_l^{a_l-1-a_{l-1}+1}$$

und

$$H^0(\mathscr{I}(a_l))/S_1 H^0(\mathscr{N}(a_l - 1)) \hookrightarrow H^0(\mathscr{M}(a_l))/S_1 H^0(\mathscr{N}(a_l - 1))$$
$$\xrightarrow{\sim} \mu X_l k \oplus \cdots \oplus \mu X_r k$$

folgt mit $h^0(\mathscr{I}(a_l)) = h(a_l) + (r-l)$ im Falle $a_{r-2} = a_{r-1}$ bzw. $h^0(\mathscr{I}(a_l)) = h(a_l) + (r-l-1)$ im Falle $a_{r-2} < a_{r-1}$ analog wie oben, daß auch in diesem Fall (*) gilt.

III. Wenn $a_{r-2}=a_{r-1}$, dann sind wir fertig, da $H^0(\mathscr{I}(n))=S_{n-a_{r-1}}H^0(\mathscr{I}(a_{r-1}))$ (2.1). Sei deshalb $a_{r-2}<a_{r-1}=d$ und $h(T)=Q(T)-\binom{T-d+1}{1}$. Wegen (*) und $Q(d)=h(d)+1$ wird $H^0(\mathscr{I}(d))$ von $L=\langle M(d,h(d))\rangle$ und einem weiteren Element $x\in S_d$ erzeugt. Wir schreiben $x=y+m$ mit $m\in S_d$ ein Monom und y Linearkombination von Monomen aus S_d, die kleiner als m und größer als das letzte Monom von $M(d,h(d))$ sind. Man hat $\dim I_{d+1}=Q(d+1)=h(d+1)+2$, $\dim S_1L=h(d+1)$ und $I_{d+1}=S_1L+S_1x$. Sei a die Anzahl der Monome X_im, die nicht in S_1L liegen. Wenn $a>2$ wäre, würde $\dim I_{d+1}>h(d+1)+2$ folgen. Wenn $a<2$ wäre, dann würden mehr als $r-1$ der Monome in S_1L liegen, also auch $X_{r-1}m$ oder X_rm. Da $S_1M(d,h(d))=M(d+1,\ h(d+1))$, folgt $X_im\in S_1L$, $0\leqq i\leqq r-1$. Weil alle in y vorkommenden Monome $<m$ sind, folgt $X_ix\in S_1L$, $0\leqq i\leqq r-1$, und daher $\dim I_{d+1}\leqq h(d+1)+1<Q(d+1)$. Also muß $X_im\in S_1L$, $0\leqq i\leqq r-2$, und $X_im\notin S_1L$, $i=r-1$ und $i=r$ sein. Im Fall $r=2$ ist $L=X_0^{a_0}S_{d-a_0}$, daher $m=X_0^{a_0-1}m'$ mit einem Monom m' in X_1 und X_2. Im Fall $r>2$ ist

$$S_1M(d,h(d))=X_0^{a_0}S_{d-a_0+1}+\cdots+X_0^{a_0-1}\ldots X_{r-2}^{a_{r-2}-a_{r-3}+1}S_{d-a_{r-2}+1}.$$

Aus $X_{r-2}m\in X_0^{a_0-1}X_1^{a_1-a_0}\ldots X_t^{a_t-a_{t-1}+1}S_{d-a_t+1}$ und $t<r-2$ folgt $m\in L$, daher muß $X_{r-2}m\in X_0^{a_0-1}\ldots X_{r-2}^{a_{r-2}-a_{r-3}+1}S_{d-a_{r-2}+1}$ sein. Man kann dann schreiben: $m=c\cdot m'$ mit $c=X_0^{a_0-1}\ldots X_{r-2}^{a_{r-2}-a_{r-3}}$ und m' Monom in X_{r-1} und X_r. Das letzte Element l von $M(d,h(d))$ ist $X_0^{a_0}X_r^{d-a_0}=cX_0X_r^{d-a_0}$ im Fall $r=2$, bzw. $X_0^{a_0-1}X_1^{a_1-a_0}\ \ldots\ X_{r-2}^{a_{r-2}-a_{r-3}+1}X_r^{d-a_{r-2}}=c\cdot X_{r-2}X_r^{d-a_{r-2}}$ im Fall $r>2$. Jedes Monom $l<n<m$ ist daher von der Form $c\cdot m'$ mit einem Monom m' vom Grad $d-a_{r-2}+1$ in X_{r-1} und X_r. Sei G' die Untergruppe von G der Automorphismen $X_i\mapsto X_i$, $0\leqq i\leqq r-1$, $X_r\mapsto X_r+aX_{r-1}$. Dann ist $x=c\cdot x'$, $x'\in S_{d-a_{r-2}+1}$ Polynom in X_{r-1} und X_r allein, folglich $\varphi(x)=c\varphi(x')$, $\varphi\in G'(k)$. Hier ist $\varphi(x')$ ebenfalls Polynom in X_{r-1} und X_r allein. Da $c(\varphi(x')-x')=\varphi(x)-x\in L$ ist und $\varphi(x')-x'$ die Variablen $X_0,\ldots,X_{r-2}$ nicht enthält, muß $\varphi(x')-x'=0$ sein, $\varphi\in G'$. Hieraus folgt $x'=X_{r-1}^{d-a_{r-2}+1}$, und daher ist x das $Q(d)$-te Monom in der lexikographischen Ordnung.

(3.7) Aus (3.6) erhält man sofort den in der Einleitung angegebenen Satz 1 (k beliebiger Körper), da jedes $\mathscr{I}\in H_Q^G(K)$, K algebraisch abgeschlossene Erweiterung von k, numerisch maximal ist (3.4) und daher gleich $\mathscr{N}$ sein muß.

4. Einige einfach-zusammenhängende Hilbertschemata

Nach einem Theorem von Horrocks ist

$$H_Q^G\hookrightarrow H_Q$$

ein 1-zusammenhängender Morphismus, d.h. $\pi_0(H_Q^G)\to\pi_0(H_Q)$ eine bijektive Abbildung der Zusammenhangskomponenten und $\pi_1(H_Q^G)\to\pi_1(H_Q)$ ein Isomorphismus der Fundamentalgruppen ([5], Theorem 6.2, p. 240). Nach Satz 1 ist $(H_Q^G)_{\mathrm{red}}\simeq\mathrm{Spec}(k)$, und wegen $\pi_1((H_Q^G)_{\mathrm{red}})\simeq\pi_1(H_Q^G)$ erhält man

Satz 2. *Wenn Q ein Hilbertpolynom wie in* (1) *mit $s \leqq r-2$ ist, dann ist über einem algebraisch abgeschlossenen Körper k das Hilbertschema H_Q zusammenhängend und einfach zusammenhängend.*

Mit Sätzen von Horrocks sollen diese Ergebnisse auf weitere Klassen von Hilbertpolynomen ausgedehnt und gleichzeitig einige étale Kohomologiegruppen von H_Q ausgerechnet werden.

Satz 3. *Wenn $r \geqq 3$ und im Hilbertpolynom*

$$Q(T) = \binom{T-a_0+r}{r} + \cdots + \binom{T-a_{r-1}+1}{1}$$

die Koeffizienten a_{r-2} und a_{r-3} gleich sind, dann gilt über einem algebraisch abgeschlossenen Körper k:

(i) *Das Hilbertschema H_Q ist zusammenhängend und einfach zusammenhängend (für $k=\mathbb{C}$ auch im topologischen Sinne).*

(ii) $H^i_{\text{et}}(H_Q, \mathbb{Z}/n) = \begin{cases} k & \text{für } i=0 \\ 0 & \text{für } i=1 \end{cases}$ wenn $(n, \operatorname{char} k)=1$.

(iii) $H^i_{\text{et}}(H_Q, \mathbb{Z}/p)=0$ für alle $i>0$, wenn $\operatorname{char} k=p$.

Beweis. 1° Für $c=a_{r-1}-a_{r-2}+1$ ist $H(T)=Q(T)+c$ ein Hilbertpolynom, so daß $H'(T)$ die Voraussetzung von Satz 1 erfüllt. Wenn $d=a_{r-1}$ ist, soll M das von $M(d, H(d))$ in S erzeugte Ideal und $\mathcal{M} \subset \mathcal{O}_X$ die assoziierte Idealgarbe mit Hilbertpolynomen H sein. Dann ist $Z=\operatorname{Quot}^c(\mathcal{M}/X/k)$ ein abgeschlossenes Unterschema von H_Q. Sei nun K/k eine Körpererweiterung und $\mathcal{I} \in H_Q^G(\operatorname{Spec} K)$. Um zu zeigen, daß $\mathcal{I} \subset \mathcal{M}$ ist, kann man zu $\bar{K}$ übergehen und dann sofort $K=k$ annehmen. Nach Satz 1 ist $\mathcal{I}'=\mathcal{M}'$, daher $\mathcal{I} \subset \mathcal{M}$ wegen der linearen Maximalität von $\mathcal{M}$. Es folgt $(Z^G)_{\text{red}}=(H_Q^G)_{\text{red}}$. Man kann jetzt die Aussagen (i)–(iii) mit Hilfe von Sätzen von Horrocks daraus ableiten, daß Z das Fixpunktschema eines Grassmann-Schemas unter einer unipotenten Gruppe ist. Da in ([7], Prop. 4.2.1, p. 116) bezüglich dieser Tatsache auf den Beweis von ([1], Prop. 2.2) hingewiesen wird, sollen hier die nötigen Schritte angegeben werden:

2° Wenn $I \subset M$ ein graduiertes Ideal und das Hilbertpolynom von M/I gleich c ist, dann ist $\mathfrak{p}^c M \subset I$, $\mathfrak{p}=(X_0, \ldots, X_{r-1})$.

3° Die graduierten S-Moduln $\mathfrak{p}^c M \subset L \subset M$ modulo der Äquivalenzrelation $L \sim F$, falls $L_n=F_n$ für alle $n \gg 0$, entsprechen bijektiv den $S_{(X_r)}$-Moduln $\mathfrak{p}^c M_{(X_r)} \subset L \subset M_{(X_r)}$ bezüglich der Zuordnung $L \mapsto L_{(X_r)}$.

4° Sei $R=S_{(X_r)}$, $\mathfrak{m}=\mathfrak{p}R$, $E=M_{(X_r)}$. Dann entsprechen die R-Moduln $\mathfrak{m}^c E \subset L \subset E$ bijektiv den $R_{\mathfrak{m}}$-Moduln $\mathfrak{m}^c E_{\mathfrak{m}} \subset L \subset E_{\mathfrak{m}}$ bezüglich der Zuordnung $L \mapsto L_{\mathfrak{m}}$.

5° Setzt man nun $A=R_{\mathfrak{m}}/\mathfrak{m}^c R_{\mathfrak{m}}$, $V=E_{\mathfrak{m}}/\mathfrak{m}^c E_{\mathfrak{m}}$, dann ergibt sich aus 2°–4°, daß $Z(\operatorname{Spec} K)$ isomorph ist zu

$$\operatorname{Quot}^c(V \otimes_k K/\operatorname{Spec}(A \otimes_k K)/K) \quad \text{für jede Körpererweiterung } K/k,$$

d.h. es ist $Z_{\mathrm{red}} \simeq \mathrm{Quot}^e(V/\mathrm{Spec}\, A/k)_{\mathrm{red}}$. Mit dem Schluß von ([1], p. 516) zeigt man nun, daß dieses Schema das Fixpunktschema von $W=\mathrm{Grass}_e(V)$ unter der Operation einer unipotenten Gruppe ist. Wegen $A=k\oplus\mathfrak{M}$, $\mathfrak{M}=\mathfrak{m}R_{\mathfrak{m}}/\mathfrak{m}^e R_{\mathfrak{m}}$, ist ein k-Untervektorraum $L\subset V$ genau dann ein A-Modul, wenn $(1+x)L\subset L$, d.h. $(1+x)L=L$ ist, für alle $x\in\mathfrak{M}$. Denn die Multiplikation mit $1+x$ ist eine unipotente lineare Abbildung des endlich-dimensionalen k-Vektorraumes V. Also ist $\mathrm{Quot}^e(V/\mathrm{Spec}\, A/k)_{\mathrm{red}}$ das reduzierte Fixpunktschema von W unter der unipotenten Gruppe $1+\mathfrak{M}$.

6° Wenn $e=\dim_k V$ und U die Gruppe der unitriangulären $e\times e$-Matrizen ist, dann hat das Fixpunktschema W^U nur einen Punkt. Wendet man nun Theorem 6.1 in [5] oder Korollar 2.2.1 bzw. Theorem 3.2 in [6] insgesamt viermal an, dann erhält man die Behauptungen des Satzes.

Korollar. *Wenn Q die Voraussetzungen von Satz 3 erfüllt, dann hat* $\mathrm{Pic}(H_Q)$ *keine Torsion prim zu* $\mathrm{char}(k)$.

Um die Ergebnisse auf eine weitere Klasse von Hilbertpolynomen zu übertragen, benötigt man zunächst ein weiteres

Lemma. *Wenn das Ideal $\mathscr{I}\subset\mathscr{O}_X$ invariant unter G ist und in dem Hilbertpolynom $Q(T)$ von $\mathscr{I}$ die Ungleichung $a_{r-1}\leqq 2a_{r-2}-4-a_{r-3}$ gilt, dann ist $\mathscr{I}$ nicht linear maximal.*

Beweis. Wenn $a_{l-1}<a_l=\cdots=a_{r-2}$ für ein $0\leqq l\leqq r-2$ und $c=a_{r-1}-a_{r-2}+1$ ist, dann ist

$$H(T)=Q(T)+c$$
$$=\binom{T-a_0+r}{r}+\cdots+\binom{T-a_{l-1}+r-l+1}{r-l+1}+\binom{T-(a_{r-2}-1)+r-l}{r-l}.$$

Wegen

$$Q'(T)=\binom{T-a_0+r-1}{r-1}+\cdots+\binom{T-(a_{r-2}-1)+r-1-l}{r-l-1}=H'(T)$$

ist $a_{r-2}-1$ eine Regularitätsschranke von $\mathscr{I}'$ (2.1), daher $h^0(\mathscr{I}'(n))=Q'(n)=h^0(\mathscr{N}'(n))$, $n\geqq a_{r-2}-2$, wobei $\mathscr{N}'$ die von $M(a_{r-2}-1,\ Q'(a_{r-2}-1))\subset S'$ erzeugte Idealgarbe mit Hilbertpolynom Q' ist (2.6). Wir nehmen an, daß $\mathscr{I}$ linear maximal ist. Dann ist nach (3.3) $h^0(\mathscr{I}(n))=h^0(\mathscr{I}'(n))+\cdots+h^0(\mathscr{I}'(0))$ für alle $n\in\mathbb{N}$ und $H(n)=h^0(\mathscr{N}'(n))+\cdots+h^0(\mathscr{N}'(0))$ für alle $n\geqq a_{r-2}-1$, da auch das von $M(a_{r-2}-1,\ H(a_{r-2}-1))$ erzeugte Ideal $\mathscr{M}$ mit Hilbertpolynom H linear maximal und $\mathscr{M}'=\mathscr{N}'$ ist (3. Bemerkung 1 und 2.6). Man erhält

$$\sum_{n=0}^{a_{r-2}-3} h^0(\mathscr{N}'(n))=\sum_{n=0}^{a_{r-2}-3} h^0(\mathscr{I}'(n))+c. \tag{4}$$

Wir betrachten nun die $c+1$ Zahlen $h^0(\mathscr{I}'(a_{r-2}-3))$, $h^0(\mathscr{I}'(a_{r-2}-4))$, ..., $h^0(\mathscr{I}'(a_{r-2}-3-c))$. Für jedes $n\in\mathbb{N}$ ist die Differenz $h^0(\mathscr{N}'(n))-h^0(\mathscr{I}'(n))\geqq 0$ (2.4 und 2.5). Wenn mehr als c dieser Zahlen größer als 0 wären, könnte (4) nicht richtig sein. Also muß für einen Index m mit $a_{r-2}-3-c\leqq m\leqq a_{r-2}-3$

gelten $h^0(\mathscr{I}'(m))=h^0(\mathscr{N}'(m))$. Da aus der Voraussetzung $a_{r-3}\leqq a_{r-2}-3-c$ folgt, hat man für den Index $n=m$ und den Index $n=a_{r-2}-2$ die Gleichung $h^0(\mathscr{I}'(n))=h^0(\mathscr{N}'(n))=h(n)$, wobei

$$h(T)=\binom{T-a_0+r-1}{r-1}+\cdots+\binom{T-a_{l-1}+r-l}{r-l}$$

ist (2.5). Das von $H^0(\mathscr{I}'(m))$ in S' erzeugte Ideal J hat dann das Hilbertpolynom h ([2], 2.6 und 3.9). Da J invariant unter $G'(k)$ ist, muß die zu J assoziierte Idealgarbe gleich dem von $M(a_{r-3}, h(a_{r-3}))$ erzeugten Ideal $\mathscr{J}$ sein (Satz 1). Da $h^0(\mathscr{N}'(n))=h^0(\mathscr{J}(n))$ für $n\leqq a_{r-2}-2$ ist (2.5), folgt mit $\mathscr{J}\subset\mathscr{I}'$, daß $h^0(\mathscr{I}'(n))=h^0(\mathscr{N}'(n))$, $n\in\mathbb{N}$, ist, Widerspruch zu (4).

Satz 4. *Wenn $r\geqq 3$ und im Hilbertpolynom Q die letzten drei Koeffizienten $a_{r-1}\leqq 2a_{r-2}-a_{r-3}-4$ erfüllen, dann gelten für das Hilbertschema H_Q dieselben Aussagen wie in Satz 3.*

Beweis. Wie im Teil 1° des Beweises von Satz 3 hat man für $\mathscr{I}\in H_Q^G(k)$ nachzuweisen $\mathscr{I}\subset\mathscr{M}$. Wenn $\mathscr{L}$ ein linear maximales Ideal ist, das $\mathscr{I}$ enthält, dann kann man annehmen, daß $\mathscr{L}$ invariant unter $G(k)$ ist; nach dem Lemma hat $\mathscr{L}$ das Hilbertpolynom H und nach Satz 1 ist $\mathscr{M}=\mathscr{L}$.

Bemerkung. Durch Satz 4 werden die Hilbertpolynome mit ziemlich kleinem Koeffizienten a_{r-1} erfaßt. Im gegenteiligen Fall, daß a_{r-1} ziemlich groß ist, kann man versuchen, die in [7] entwickelte Theorie anzuwenden. Zunächst wird in [7] ein Morphismus $f: H_Q^G\to H_{Q'}^{G'}$ konstruiert, der sich durch die Zuordnung $\mathscr{I}\mapsto\mathscr{I}'$ beschreiben läßt (Bezeichnungen wie im 3. Abschnitt). Weiter wird gezeigt $\pi_1(H_Q^G)\simeq\pi_1(H_{Q'}^{G'})$ für surjektives f. Schließlich beweist Horrocks die Existenz einer Zahl $c(Q')$ derart, daß $f: H_{Q-c}^G\to H_{Q'}^{G'}$ surjektiv ist für $c\geqq c(Q')$.

Explizit erhält man $c(Q')$ wie folgt. Wegen Satz 3 kann man o.E. $a_{r-3}<a_{r-2}$ annehmen und

$$H(T)=\binom{T-a_0+r}{r}+\cdots+\binom{T-a_{r-3}+3}{3}+\binom{T-(a_{r-2}-1)+2}{2}$$

bilden. Startet man mit einem $\mathscr{J}\in H_{Q'}^{G'}(k)$ und setzt $L_n=\sum_{\nu=0}^{n} X_r^\nu H^0(\mathscr{J}(n-\nu))$, dann hat $\tilde{L}$ das Hilbertpolynom $H(T)-c$. Eine einfache Rechnung (man muß benutzen, daß $a_{r-2}-1$ Regularitätsschranke von $\mathscr{J}$ ist) zeigt, daß $c\leqq\mu=\min\{H(n)\,|\,a_{r-3}-4\leqq n\leqq a_{r-2}-4\}$. Es folgt, für $a_{r-1}\geqq a_{r-2}+\mu-1$ ist $f: H_Q^G\to H_{Q'}^{G'}$ surjektiv und damit $\pi_1(H_Q)\simeq\pi_1(H_{Q'})$.

Literatur

1. Fogarty, J.: Algebraic families on an algebraic surface. Amer. J. Math. **90**, 511–521 (1968)
2. Gotzmann, G.: Eine Bedingung für die Flachheit und das Hilbertpolynom eines graduierten Ringes. Math. Z. **158**, 61–70 (1978)

3. Grothendieck, A.: Techniques de construction et théorèmes d'existence en géométrie algébrique IV: Les schémas de Hilbert. In: Seminaire Bourbaki, 13e année 1960/61, Exposé 221. Paris: Sécrétariat Mathématique 1961
4. Hartshorne, R.: Connectedness of the Hilbertschema. Inst. Hautes Études Sci. Publ. Math. **29**, 261–309 (1966)
5. Horrocks, G.: Fixed point schemes of additive group actions. Topology **8**, 233–242 (1969)
6. Horrocks, G.: Properties of schemes inherited from orbits. J. Algebra **40**, 406–417 (1976)
7. Horrocks, G.: Sheaves on projective space invariant under the unitriangular group. Invent. Math. **10**, 108–118 (1970)
8. Katz, N.: Une formule de congruence pour la fonction ζ. In: Groupes de Monodromie en Géométrie Algébrique. SGA7II. Lecture Notes in Mathematics **340**. Berlin-Heidelberg-New York: Springer 1973
9. Mumford, D.: Lectures on curves on an algebraic surface. Princeton: Princeton University Press 1966
10. Sperner, E.: Über einen kombinatorischen Satz von Macaulay und seine Anwendungen auf die Theorie der Polynomideale. Abh. Math. Sem. Univ. Hamburg **7**, 149–163 (1930)

Eingegangen am 27. Juni 1981

Nachtrag bei der Korrektur

Wenn im Hilbertpolynom (1) $s=r-1$ ist, hat das Fixpunktschema H_Q^G genau dann nur einen einzigen Punkt, wenn $a_0+1=\cdots=a_{t-1}+1=\alpha_t=\cdots=a_{r-1}$ für eine ganze Zahl $0\leqq t\leqq r-1$ ist.

Math. Z. 180, 307-329 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Das Modulproblem für holomorphe Einbettungen mit konkaver Umgebungsstruktur

Siegmund Kosarew

Fakultät für Mathematik der Universität, Universitätsstraße 31, D-8400 Regensburg, Bundesrepublik Deutschland

Einleitung

Das Ziel dieser Arbeit ist es, für abgeschlossene holomorphe Einbettungen $X \hookrightarrow Y$ mit konkaver Umgebungsstruktur einen lokalen Modulraum zu konstruieren. In unserem Fall ist X ein kompakter komplexer Raum und Y eine q-konkave (im Sinne von Andreotti-Grauert) n-dimensionale komplexe Mannigfaltigkeit mit $q < n-1$, insbesondere muß dann Y mindestens dreidimensional sein. Unter diesen Voraussetzungen existiert eine semiuniverselle Deformation der Einbettung $X \hookrightarrow Y$. Grob gesprochen ist eine Deformation einer holomorphen Einbettung eine Deformation von Y, bei der X festbleibt. Dieser Begriff wird im ersten Kapitel präzisiert, wo sich auch einige allgemein bekannte Tatsachen über Deformationen von Mannigfaltigkeiten finden. Ein wesentlicher Punkt ist dann die Reduktion auf den Fall $X = \varnothing$, was letztlich mit Hilfe des relativen Hilbertschen Modulraumes bewiesen wird. Es genügt daher, eine semiuniverselle Deformation der konkaven Mannigfaltigkeit Y zu finden. Wir gehen dabei ähnlich wie in [7] vor und konstruieren zunächst eine konvergente, formal semiuniverselle Deformation von Y. Der wesentliche Punkt hierbei ist der Satz II.2.3 zusammen mit II.2.5. Der Satz II.2.3 besagt, daß man über gewissen offenen Teilmengen des $\mathbb{C}^n$ einen Deformationscozyklus zerfallen lassen kann, und zwar in einem gewissen Sinne auf differenzierbare Weise. In [7] wird die analoge Aussage mit Hilfe des Cousinschen Induktionsprinzips bewiesen. Da man bei der Konstruktion von Deformationen über relativ kompakten offenen q-konkaven Teilmengen von Y allgemein die Deformationen in der Nähe des Randes untersuchen muß, kann man sich nicht ohne weiteres auf offene Quader im $\mathbb{C}^n$ beschränken. Bei dem vorliegenden Beweis wird stattdessen die Lösung des $\bar{\partial}$-Problems mit Abschätzung in der Nähe eines q-konkaven Randpunktes herangezogen, wie sie in Hörmander [10] bei Verwendung von L^2-Normen dargestellt ist. Hieraus folgt dann auf die übliche Weise der Glättungssatz für Deformationscozyklen. Die weitere Konstruktion einer konvergenten, formal semiuniversellen Deformation von Y kann dann genauso wie in [7] vorgenommen werden. Es bleibt somit nur noch die

0025-5874/82/0180/0307/$04.60

Vollständigkeit der so konstruierten Deformation nachzurechnen. Dies geschieht in Kapitel III. Hierbei wird öfters benutzt, wie auch schon in Kapitel II, § 4, daß sich holomorphe Funktionen in der Nähe eines q-konkaven Randpunktes mit $q<n$ über diesen hinweg fortsetzen lassen. Eine solche Vorgehensweise läßt sich daher nicht auf den Fall von Deformationen 1-konvexer Mannigfaltigkeiten übertragen.

Da die in Kapitel II konstruierte Deformation nur auf einem relativ kompakten Teil von Y gegeben ist, muß man diese noch auf ganz Y - bis auf Äquivalenz - ausdehnen. Um das durchzuführen, werden einige Aussagen über die Cohomologie nicht abelscher Garben benutzt, welche im Anhang zusammengestellt worden sind. Dabei ist es zweckmäßig, abgeschlossene Überdeckungen anstatt offener zu betrachten.

Es sei noch angemerkt, daß im Falle $H^1(Y,\mathcal{I}\theta_Y)=0$, wobei $\mathcal{I}$ das definierende Ideal von X in Y sei, die Einbettung $X\hookrightarrow Y$ starr ist, da sich die infinitesimalen Deformationen von $X\hookrightarrow Y$ gerade mit dieser Cohomologiegruppe identifizieren.

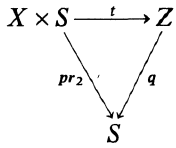
Meinen besonderen Dank für Ratschläge und Hinweise verschiedenster Art möchte ich hier noch Herrn Prof. Dr. J. Bingener sowie den Herren Dr. D. Leistner und Dipl.-math. H. Stieber aussprechen.

Alle im folgenden auftretenden komplexen Räume werden als separiert mit abzählbarer Topologie vorausgesetzt.

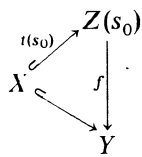
Kapitel I. Deformationen holomorpher Einbettungen

§ 1. Allgemeine Bemerkungen

1.1. Definition. Sei $\tau\colon X \hookrightarrow Y$ eine abgeschlossene Einbettung komplexer Räume, S ein Raumkeim mit ausgezeichnetem Punkt s_0. Unter einer *Deformation von τ über S* versteht man ein kommutatives Diagramm komplexer Räume über S



wobei q eine (flache) Deformation von Y über S ist und $t(s_0)$ eine Einbettung, die zu τ isomorph ist, d.h. es gibt ein kommutatives Diagramm



mit einem Isomorphismus f.

1.2. Satz. *Sei $t\colon X \times S \to Z$ eine Deformation von τ über dem komplexen Raumkeim S. Dann gibt es einen Repräsentanten $\tilde{t}\colon W \to \tilde{Z}$ von t, der eine abgeschlossene Einbettung ist.*

Bemerkung. Insbesondere sind dann auch die Morphismen $\tilde{t}(s)\colon X \to \tilde{Z}(s)$, s nahe bei s_0, abgeschlossene Einbettungen.

Der Beweis des obigen Satzes ergibt sich leicht aus folgender rein topologischen Aussage (vgl. dazu [3], (12.7)):

Lemma. *Seien X, Y parakompakte topologische Räume, $X' \subset X$, $Y' \subset Y$ abgeschlossene Unterräume und $f\colon X \to Y$ stetig. Ist f lokal eine abgeschlossene Einbettung (bzw. ein lokaler Homöomorphismus) und $f|_{X'}\colon X' \to Y'$ eine abgeschlossene Einbettung (bzw. ein Homöomorphismus), dann existieren offene Umgebungen $V \subset X$ von X' und $W \subset Y$ von $f(X')$, so daß $f|_V\colon V \to W$ eine abgeschlossene Einbettung (bzw. ein Homöomorphismus) ist.*

1.3. Morphismen von Deformationen. Seien $t_i\colon X \times S \to Z_i$, $i = 1, 2$ zwei Deformationen von τ über S. Ein Morphismus von t_1 und t_2 ist ein S-Morphismus $f\colon Z_1 \to Z_2$ mit $f \circ t_1 = t_2$.

1.4. Bemerkung. Jeder Morphismus von Deformationen von τ über einem Raumkeim S ist ein Isomorphismus.

Dies folgt mit Hilfe des letzten Lemmas.

1.5. Basiswechsel. Sei $\alpha\colon S' \to S$ ein Morphismus analytischer Raumkeime, dann ist $\alpha^* t$ eine Deformation von τ über S' für jede Deformation t von τ über S.

1.6. Bemerkung. Sei $X' \subset X$ ein abgeschlossener Unterraum von X. Jede Deformation von τ über S induziert auf natürliche Weise eine Deformation von $\tau|_{X'}$ über S.

Man beachte jedoch, daß die Isomorphieklassen von Deformationen von τ über S bzw. $\tau|_{X'}$ über S i.allg. verschieden sind.

Man hat somit einen kontravarianten Deformationsfunktor

$$F\colon (\mathrm{Gan}) \to (\mathrm{Ens})$$

von der Kategorie der analytischen Raumkeime in die Kategorie der Mengen, der jedem Keim S die Menge der Isomorphieklassen von Deformationen von τ über S zuordnet. F erfüllt die Schlessingerbedingungen (H1) und (H2) von [15], Theorem 2.11.

§ 2. Reduktion auf den Fall $X = \varnothing$

Wir betrachten nun den Spezialfall, daß X kompakt ist und der komplexe Raum Y eine verselle Deformation besitzt. Unter diesen Voraussetzungen kann man den folgenden Satz beweisen:

2.1. Satz. *Der Funktor F besitzt ein semiuniverselles Element.*

Beweis. Wir können wie in [4], Theorem 8.5 vorgehen. Sei $Z \to S$ eine verselle Deformation von Y. Nach [13] ist der Funktor $T \mapsto \underline{\mathrm{Mor}}_S(X \times S, Z)(T)$ durch einen (separierten) komplexen Raum $\mathrm{Mor}_S(X \times S, Z)$ über S darstellbar. Für die unterliegende Menge gilt:

$$|\mathrm{Mor}_S(X \times S, Z)| = \dot{\bigcup_{s \in S}} |\mathrm{Mor}(X, Z(s))|.$$

Insbesondere ist $\tau \in \mathrm{Mor}_S(X \times S, Z)$. Sei

$$M := \mathrm{Mor}_S(X \times S, Z) \times_{\mathrm{Mor}(X,Y)} \{\tau\}.$$

M ist ein abgeschlossener Unterraum von $\mathrm{Mor}_S(X \times S, Z)$. Wir setzen $Z_M = Z \times_S M$. Wie man feststellt, ist der universelle M-Morphismus $X \times M \to Z_M$ eine verselle Deformation von τ, wobei $\tau \in M$ der ausgezeichnete Punkt ist. Insbesondere existiert dann ein semiuniverselles Objekt für F (s. [4], Satz 8.1).

2.2. Ab jetzt sei Y stets eine zusammenhängende komplexe Mannigfaltigkeit der Dimension n. Wir wenden den Satz 2.1 auf die folgende Situation an:

$\tau\colon X \hookrightarrow Y$ ist eine abgeschlossene Einbettung des kompakten komplexen Raumes in die n-dimensionale q-konkave Mannigfaltigkeit, wobei $q < n-1$. Insbesondere muß also dann $n \geqq 3$ sein. Den Begriff „q-konkav" verwenden wir hier im Sinne von Andreotti-Grauert ([1]). Gemäß Satz 2.1 genügt es also, für Y eine verselle Deformation zu finden.

§ 3. Beschreibung der Isomorphieklassen von Deformationen von Mannigfaltigkeiten

3.1. Bezeichnungen. i) Sei S ein komplexer Raumkeim. Mit $\mathfrak{G}_S$ bezeichnen wir die Garbe der vertikalen Automorphismen auf Y, die definiert ist durch

$$\Gamma(U, \mathfrak{G}_S) := \{g \in \mathrm{Aut}_S(U \times S) : g(s_0) = \mathrm{id}_U\}$$

für $U \subset Y$ offen. s_0 sei dabei der ausgezeichnete Punkt von S.

ii) Mit B sei in Zukunft stets der Keim des $\mathbb{C}^m$ im Nullpunkt bezeichnet. Wir können die Elemente $g \in \Gamma(U, \mathfrak{G}_B)$ auf eindeutige Weise in eine konvergente Potenzreihe entwickeln:

$$g(z,t) = (z + \sum_{\substack{\nu \in \mathbb{N}^m \\ |\nu| \geqq 1}} g_\nu(z)\, t^\nu, t).$$

iii) $\mathrm{Def}(Y, S)$ sei die Menge der Isomorphieklassen von Deformationen von Y über dem Raumkeim S.

Es gilt dann die Aussage:

3.2. Satz. *Es existiert ein Isomorphismus von Mengen mit ausgezeichnetem Punkt*

$$\mathrm{Def}(Y, S) \to H^1(Y, \mathfrak{G}_S).$$

Eine kurze Beweisskizze für kompaktes Y befindet sich in [6], 5.4. Die Übertragung auf den allgemeinen Fall, wo Y abzählbare Topologie besitzt, erfolgt durch ein bekanntes Ausschöpfungsargument.

Die Zuordnung $S \mapsto \mathfrak{G}_S$ definiert offensichtlich einen kontravarianten Funktor von der Kategorie der analytischen Raumkeime in die Kategorie der Garben von Gruppen auf Y.

3.3. Satz. *Sei $Y \subset \mathbb{C}^n$ offen, $S \subset S'$ ein Unterkeim. Dann ist die Sequenz*

$$\Gamma(Y, \mathfrak{G}_{S'}) \to \Gamma(Y, \mathfrak{G}_S) \to 1$$

exakt.

Beweis. Sei $g \in \Gamma(Y, \mathfrak{G}_S)$. Offensichtlich ist dann g durch n Schnitte in $\Gamma(Y, \mathfrak{m}_S \mathcal{O}_{Y \times S})$ gegeben. Es genügt daher die Surjektivität der kanonischen Abbildung

$$\Gamma(Y, \mathfrak{m}_{S'} \mathcal{O}_{Y \times S'}) \to \Gamma(Y, \mathfrak{m}_S \mathcal{O}_{Y \times S})$$

einzusehen. Dies folgt aber mit [11], (0.3) aus der Nuklearität von $\Gamma(Y, \mathcal{O}_Y)$.

3.4. Corollar. *Für jeden Unterkeim $S \subset S'$ ist die kanonische Abbildung*

$$\mathfrak{G}_{S'} \to \mathfrak{G}_S$$

ein Garbenepimorphismus.

Es sei hier an folgende Sprechweise erinnert: Eine Erweiterung $S \subset S'$ analytischer Raumkeime heißt *klein*, falls für $K := \ker(\mathcal{O}_{S'} \to \mathcal{O}_S)$ gilt $\mathfrak{m}_{S'} K = 0$. Insbesondere ist dann K ein endlich dimensionaler $\mathbb{C}$-Vektorraum.

3.5. Satz. *Sei $S \subset S'$ eine kleine Erweiterung analytischer Raumkeime, $K := \ker(\mathcal{O}_{S'} \to \mathcal{O}_S)$, $\mathcal{K} := \mathcal{K}er(\mathfrak{G}_{S'} \to \mathfrak{G}_S)$.*

Behauptung: i) $\mathscr{K}$ *ist im Zentrum von* $\mathfrak{G}_S$, *enthalten.*
ii) *Es gibt einen kanonischen Isomorphismus*

$$\vartheta\colon \mathscr{K} \to \theta_Y \otimes_{\mathbb{C}} K.$$

Der Beweis dieser beiden Aussagen findet sich - in einer etwas spezielleren Situation - in [6], 5.7, 5.9.

3.6. Sei $U \subset Y$ eine offene Teilmenge.
Folgendes Diagramm ist dann kommutativ:

$$\begin{array}{ccc} \mathrm{Def}(Y,S) & \xrightarrow{\sim} & H^1(Y,\mathfrak{G}_S) \\ \downarrow & & \downarrow \\ \mathrm{Def}(U,S) & \xrightarrow{\sim} & H^1(U,\mathfrak{G}_S). \end{array}$$

3.7. Sei $Q \subset Y$ eine abgeschlossene Teilmenge.
Wir setzen $\mathrm{Def}(Q,S) := \varinjlim_{U \supset Q} \mathrm{Def}(U,S)$. Die kanonische Abbildung

$$\mathrm{Def}(Q,S) \to H^1(Q,\mathfrak{G}_S)$$

ist dann ein Isomorphismus (vgl. Anhang 2.3).

3.8. Die infinitesimalen Deformationen von $U \subset Y$ bzw. $Q \subset Y$ identifizieren sich dann vermöge Satz 3.2 und 3.5 mit $H^1(U,\theta_Y)$ bzw. $H^1(Q,\theta_Y)$. Sind diese $\mathbb{C}$-Vektorräume jeweils endlich dimensional, so erfüllen die Funktoren

$$\begin{aligned} F_U\colon (\mathrm{Gan}) &\to (\mathrm{Ens}) \\ S &\mapsto \mathrm{Def}(U,S), \\ F_Q\colon (\mathrm{Gan}) &\to (\mathrm{Ens}) \\ S &\mapsto \mathrm{Def}(Q,S) \end{aligned}$$

sämtliche Schlessingerbedingungen und es existieren somit formale formal semiuniverselle Deformationen $a \in \hat{F}_Q(S)$ bzw. $b \in \hat{F}_U(T)$, wobei S und T komplette analytische Raumkeime sind und $\hat{F}$ die jeweilige Fortsetzung von F auf die Kategorie der kompletten analytischen Raumkeime bezeichne, welche durch $\hat{F}(S) := \varprojlim F(S_e)$ gegeben ist, $S_e := (\{s_0\}, \mathcal{O}_S/\mathfrak{m}_S^{e+1})$.

Wir schreiben $\mathcal{O}_S = \mathbb{C}[\![t_1,\ldots,t_m]\!]/\mathfrak{a}$, $\mathfrak{a}$ werde von $f_1,\ldots,f_r$ erzeugt. Auf Grund der Effektivität von a gilt $f_i \equiv 0 \bmod \mathfrak{m}^2$, falls $m = \dim_{\mathbb{C}} H^1(Q,\theta_Y)$.

3.9. Ist nun Y q-konkav, $q < n-1$ und $\varphi\colon Y \to \mathbb{R}$ eine außerhalb eines Kompaktums K streng q-konkave Ausschöpfungsfunktion, d.h. die Leviform $\mathrm{Lev}(\varphi)$ hat in jedem Punkt von $Y \setminus K$ mindestens $(n-q+1)$ negative Eigenwerte, so gilt nach [1]

$$\dim_{\mathbb{C}} H^1(B_c,\theta_Y) < \infty,$$

$$\dim_{\mathbb{C}} H^1(\bar{B}_c,\theta_Y) < \infty$$

wobei $B_c := \{y \in Y: \varphi(y) < c\}$ und c so gewählt ist, daß

$$\sup_K \varphi < c \leqq \sup_Y \varphi.$$

Man beachte, daß für obige c gilt $\bar{B}_c = \{y \in Y: \varphi(y) \leqq c\}$.

Kapitel II. Ein Glättungssatz und die Konstruktion einer konvergenten formal semiuniversellen Deformation

§ 1. Definitionen und Bezeichnungen

1.0. Es sei wieder $B = (\mathbb{C}^m, 0)$, $t = (t_1, \ldots, t_m)$.

$$M(t) := \prod_{j=1}^{m} \left(\tfrac{1}{2} - \tfrac{1}{4}\sqrt{1-t_j}\right) = \sum_{\nu \in \mathbb{N}^m} \gamma_\nu t^\nu$$

bezeichne die *Standardmajorante.* Für einen Banachraum $(E, \|.\|)$ wird dann $E\{t\}$ mit den Pseudonormen $\|f\|_\rho := \sup_{\nu \in \mathbb{N}^m} \frac{\|a_\nu\|}{\gamma_\nu} \rho^\nu$, $f = \sum a_\nu t^\nu$, $\rho \in \mathbb{R}^m_{++}$, ein induktiv normierter $\mathbb{C}\{t\}$-Modul ([7], § 8), $E\{t\}^0 := \mathfrak{m} E\{t\}$, $\mathfrak{m} \subset \mathbb{C}\{t\}$ sei das maximale Ideal.

1.1. Sei $S \subset B$ ein Unterkeim, $\mathcal{J} \subset \mathbb{C}\{t\}$ sein Ideal. $\Delta := \mathbb{N}^m \setminus \mathrm{ord}(\mathcal{J})$ nennt man die *S-reduzierten Indizes* ([5], Kap. I). Die *S-reduzierten Potenzreihen* sind Reihen der Gestalt $f = \sum_{\nu \in \Delta} a_\nu t^\nu$.

Für einen Banachraum E bezeichne $\mathfrak{R}_S \subset E\{t\}$ den Untervektorraum der S-reduzierten Potenzreihen, $\mathfrak{N}_S := \mathcal{J} E\{t\}$. Nach dem Grauertschen Divisionssatz (vgl. [5], Satz 2.9) hat man dann die direkte Summenzerlegung $E\{t\} = \mathfrak{R}_S \oplus \mathfrak{N}_S$.

1.2. Sei $U \subset \mathbb{C}^n$ offen. Wir setzen

$$\mathfrak{H}(U) := \Gamma(U, \mathcal{O}^n_{\mathbb{C}^n \times B}).$$

$\mathfrak{H}$ definiert dann auf natürliche Weise eine Garbe auf dem $\mathbb{C}^n$. Für $f \in \Gamma(U, \mathcal{O}^n_{\mathbb{C}^n})$ sei

$$|f|_U := \sup_{z \in U} |f(z)|, \qquad \|f\|_U := \left(\int_U |f(z)|^2 \, dz\right)^{1/2},$$

$$\Gamma_b(U, \mathcal{O}^n_{\mathbb{C}^n}) := \{f \in \Gamma(U, \mathcal{O}^n_{\mathbb{C}^n}): |f|_U < \infty\},$$

$$\Gamma_h(U, \mathcal{O}^n_{\mathbb{C}^n}) := \{f \in \Gamma(U, \mathcal{O}^n_{\mathbb{C}^n}): \|f\|_U < \infty\}.$$

Abkürzend schreiben wir

$$\Gamma_b(U, \mathfrak{H}) := \Gamma_b(U, \mathcal{O}^n_{\mathbb{C}^n})\{t\},$$

$$\Gamma_{\mathcal{J}}(U, \mathfrak{H}) := \Gamma_h(U, \mathcal{O}^n_{\mathbb{C}^n})\{t\}.$$

Bemerkung. $\Gamma_{\mathscr{O}}(U,\mathfrak{H})$ und $\Gamma_{\mathscr{L}}(U,\mathfrak{H})$ sind auf natürliche Weise Unterräume von $\Gamma(U,\mathfrak{H})$. Ist $U'\subset\subset U$ relativ kompakt, so gilt

$$\mathrm{Im}\,(\Gamma(U,\mathfrak{H})\to\Gamma(U',\mathfrak{H}))\subset\Gamma_{\mathscr{O}}(U',\mathfrak{H})\subset\Gamma_{\mathscr{L}}(U',\mathfrak{H}).$$

1.3. Sei $U\subset\mathbb{C}^n$ offen, wir schreiben $\mathfrak{G}:=\mathfrak{G}_B$. Es gilt $\mathfrak{G}(U)\cong\{g\in\mathfrak{H}(U)\colon g(z,0)=z\}$. Sei

$$\Gamma_{\mathscr{O}}(U,\mathfrak{G}):=\{g\in\Gamma(U,\mathfrak{G})\colon g-\mathrm{id}\in\Gamma_{\mathscr{O}}(U,\mathfrak{H})\}$$

und entsprechend $\Gamma_{\mathscr{L}}(U,\mathfrak{G})$.

Bemerkung. $\Gamma_{\mathscr{O}}(U,\mathfrak{G})$ ist ein affiner Raum über dem induktiv normierten $\mathbb{C}\{t\}$-Modul $\Gamma_{\mathscr{O}}(U,\mathfrak{H})^0$.

Die Operationen Komposition, Inversenbildung, Konjugation mit biholomorphen Abbildungen von vertikalen Automorphismen sind nach Beschränkung auf relativ kompakte offene Teilmengen differenzierbare Abbildungen zwischen induktiv normierten $\mathbb{C}\{t\}$-Moduln ([7], § 10).

Für $g\in\Gamma(U,\mathfrak{G})$ setzen wir $\hat{g}:=g-\mathrm{id}$. Wir führen folgende vier Lemmata an.

1.4. Lemma. *Seien $V\subset\subset U\subset\subset\mathbb{C}^n$ jeweils offen, $f\in\Gamma_{\mathscr{O}}(U,\mathfrak{G})$, $g\in\Gamma_{\mathscr{O}}(V,\mathfrak{G})$. Sei $f\circ g=g+\hat{f}+R$.*

Behauptung. Es gibt von f und g unabhängige Konstanten $K>0$, $\varepsilon>0$, so daß für jedes $\rho\in\mathbb{R}^m_{++}$ mit $\|g\|_{V,\rho}\leqq\varepsilon$ gilt

$$\|R\|_{V,\rho}\leqq K\,\|f\|_{U,\rho}\,\|g\|_{V,\rho}.$$

Bemerkung. Die entsprechende Aussage bleibt richtig (mit Konstanten K', $\varepsilon'>0$) bei Verwendung der anfangs eingeführten L^2-Normen.

Der Beweis der Behauptung kann wie in [6], Lemma 2.15 geführt werden.

1.5. Lemma. *Sei $U\subset\mathbb{C}^n$ offen, $I\subset\mathbb{C}\{t\}$ ein Ideal und $f,g\in\Gamma(U,\mathfrak{G})$.*

Behauptung:

i) *gilt $f\equiv\mathrm{id}\bmod\mathfrak{m}^e$, $e\in\mathbb{N}\setminus\{0\}$, so folgt $f^{-1}\equiv\mathrm{id}-\hat{f}\bmod\mathfrak{m}^{2e}$*

ii) *gilt $f\circ g^{-1}\equiv\mathrm{id}\bmod I$, so folgt $f\circ g^{-1}\equiv\mathrm{id}+f-g\bmod\mathfrak{m}I$*

Beweis. Beide Aussagen ergeben sich durch Betrachtung der jeweiligen Taylorentwicklungen. Wir verifizieren ii und überlassen i dem Leser. Sei $f\circ g^{-1}=\mathrm{id}+R$. Nach Voraussetzung ist $R\equiv 0\bmod I$. Aus der Gleichung $f=g+R\circ(\mathrm{id}+\hat{g})$ folgt, da $R\circ(\mathrm{id}+\hat{g})\equiv R\bmod\mathfrak{m}I$, dann die Behauptung.

Seien $t=(t_1,\ldots,t_m)$, $s=(s_1,\ldots,s_l)$ Unbestimmte, $\alpha\in\mathbb{C}^m\{s\}^0$, E ein Banachraum. α induziert einen Morphismus

$$\begin{aligned}\alpha^*\colon\ & E\{t\}\to E\{s\}\\ & f\mapsto f(\alpha).\end{aligned}$$

1.6. Lemma. *Für $\alpha_1,\alpha_2\in\mathbb{C}^m\{s\}^0$, $\alpha_2-\alpha_1=R$, R homogen vom Grad e, folgt für jedes $f\in E\{t\}$*

$$\alpha_2^*f\equiv\alpha_1^*f+Df(0)\,R\bmod\mathfrak{m}^{e+1}.$$

Für $\rho\in\mathbb{R}^k_{++}$ sei stets $\gamma(\rho):=\min\limits_{1\leqq i\leqq k}\rho_i$, $\rho=(\rho_1,\ldots,\rho_k)$.

1.7. Lemma. *Sei* $f \in E\{t\}^0$, $\alpha \in \mathbb{C}^m\{s\}^0$, $\rho \in \mathbb{R}^l_{++}$, $\tilde{\rho} \in \mathbb{R}^m_{++}$.

Behauptung. Gilt $\|\alpha\|_\rho \leqq \frac{\gamma(\tilde{\rho})}{2}$, *so folgt*

$$\|\alpha^* f\|_\rho \leqq \frac{1}{\gamma(\tilde{\rho})} \|\alpha\|_\rho \, \|f\|_{\tilde{\rho}}.$$

Beweis. Sei $f = \sum_{\nu \in \mathbb{N}^m \setminus \{0\}} c_\nu t^\nu$. Dann ist $|c_\nu| \leqq \|f\|_{\tilde{\rho}} \frac{\gamma_\nu}{\tilde{\rho}^\nu}$ für alle $\nu \in \mathbb{N}^m \setminus \{0\}$.

$$(\alpha^* f)(s) = \sum_{\nu \in \mathbb{N}^m \setminus \{0\}} c_\nu (\alpha(s))^\nu \ll \|f\|_{\tilde{\rho}} \sum_{\nu \in \mathbb{N}^m \setminus \{0\}} \gamma_\nu$$

$$\frac{\|\alpha\|_\rho^{|\nu|}}{\tilde{\rho}^\nu} M\left(\frac{s}{\rho}\right) \ll \|f\|_{\tilde{\rho}} \|\alpha\|_\rho \left(\sum_{\nu \in \mathbb{N}^m \setminus \{0\}} \gamma_\nu \frac{(\gamma(\tilde{\rho})/2)^{|\nu|-1}}{\tilde{\rho}^\nu} \right) M\left(\frac{s}{\rho}\right).$$

Da nun $0 < \gamma_\nu \leqq 4^{-m}$ für alle $\nu \in \mathbb{N}^m$ und $\sum_{\nu \in \mathbb{N}^m \setminus \{0\}} (\frac{1}{2})^{|\nu|-1} = 2^{m+1} - 2$ folgt die gewünschte Abschätzung.

§ 2. Ein Theorem B

Wir beweisen in diesem Abschnitt, daß Deformationen über gewissen offenen Mengen trivial sind, d.h., daß man unter gewissen Bedingungen Cozyklen zerfallen lassen kann. Dieses Zerfallen muß für die spätere Konstruktion differenzierbar von dem Ausgangscozyklus abhängen. Dazu dienen die folgenden Hilfsaussagen.

2.1. Lemma. *Seien E und F (komplexe) Banachräume,* $(F_i)_{i \in \mathbb{N}}$ *eine aufsteigende Kette direkter Unterräume von F. Es gebe verträgliche Familien von Projektionen* $p_i\colon F \to F_i$, $p_i^{i+1}\colon F_{i+1} \to F_i$ $i \in \mathbb{N}$, *mit* $\|p_i\|$, $\|p_i^{i+1}\| \leqq 1$. *Der kanonische Morphismus* $\alpha\colon F \to \varprojlim_b F_i$ *sei ein Isomorphismus topologischer Vektorräume, wobei*

$$\varprojlim_b F_i := \{(f_i) \in \varprojlim F_i\colon \sup_{i \in \mathbb{N}} |f_i| < \infty\}.$$

Behauptung. Sei $U \subset E$ *offen,* $f\colon U \to F$ *eine lokal beschränkte Abbildung. Dann sind folgende Aussagen äquivalent:*

i) *f ist differenzierbar,*

ii) *für jedes* $i \in \mathbb{N}$ *ist* $f_i := p_i \circ f$ *differenzierbar.*

Die Richtung i ⇒ ii ist trivial. Der Schluß ii ⇒ i folgt so: Sei $x \in U$. Die homogenen Anteile vom Grad ν $\left(\frac{1}{\nu!} D^\nu f_i(x)\right)_{i \in \mathbb{N}}$ bilden ein projektives System von ν-multilinearen Abbildungen von E nach F, welche aufgrund der Cauchyabschätzung und der lokalen Beschränktheit von f gleichmäßig beschränkt sind. Da α ein Isomorphismus ist, kann man $\varprojlim_i \frac{1}{\nu!} D^\nu f_i(x) =: f^{(\nu)}$ als (stetige) ν-multilineare Abbildung von E nach F auffassen. Es ist dann klar, daß $\sum_{\nu \in \mathbb{N}} f^{(\nu)}$ die Taylorentwicklung von f um x ist.

2.2. Lemma (Zyklenprojektion). *Sei* $\mathfrak{U}=(U_i)_{i\in I}$ *eine endliche Familie offener, beschränkter Teilmengen des* $\mathbb{C}^n$.

Behauptung: Es gibt eine $\mathbb{C}\{t\}$*-lineare Projektion*

$$\lambda\colon C^1_{\mathscr{L}}(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n\times B}) \to Z^1_{\mathscr{L}}(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n\times B})$$

mit $\|\lambda(c)\|_{\mathfrak{U},\rho} \leqq \|c\|_{\mathfrak{U},\rho}$ *für alle* $\rho\in\mathbb{R}^m_{++}$.

Beweis. Es genügt λ koeffizientenweise anzugeben. Da $Z^1_h(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n}) = Z^1(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n}) \cap C^1_h(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n})$ ein abgeschlossener Unterraum des Hilbertraumes $C^1_h(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n})$ ist, gibt es eine orthogonale Projektion auf diesen.

Daraus folgt sofort die Behauptung.

Wir formulieren die Bedingung

(+) Sei $\mathfrak{U}$ wie in 2.2.

Es gebe einen $\mathbb{C}\{t\}$-linearen Morphismus

$$\kappa\colon Z^1_{\mathscr{L}}(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n\times B}) \to C^0_{\mathscr{L}}(\mathfrak{U}, \mathcal{O}_{\mathbb{C}^n\times B})$$

mit $\delta\circ\kappa = \mathrm{id}$, wobei δ den Čech-Corand bezeichne.

2.3. Satz (Theorem B für Deformationen). *Seien* $\mathfrak{V}\ll\mathfrak{U}$ *endliche Familien offener, beschränkter Teilmengen des* $\mathbb{C}^n$. *Es gebe eine Zwischenüberdeckung* $\mathfrak{U}'$, $\mathfrak{V}\ll\mathfrak{U}'\ll\mathfrak{U}$, *so daß* (+) *auf* $\mathfrak{U}'$ *erfüllt ist.*

Dann gibt es eine differenzierbare Abbildung induktiv normierter $\mathbb{C}\{t\}$*-Moduln*

$$\vartheta\colon C^1_{\delta}(\mathfrak{U}, \mathfrak{G}) \to C^0_{\delta}(\mathfrak{V}, \mathfrak{G})$$

mit folgender Eigenschaft: Ist $S\subset B$ *ein Unterkeim und* $g\in C^1_{\delta}(\mathfrak{U}, \mathfrak{G})$ *ein Cozyklus über S, so zerfällt g bzgl.* $\mathfrak{V}$ *über S in die Cokette* $\vartheta(g)$.

2.4. Bemerkung. Aufgrund der Definition der Topologie von $C^1_{\delta}(\mathfrak{U}, \mathfrak{G})$ (s. [7], 8.5) genügt es, ϑ auf einer offenen Umgebung der Identität anzugeben.

Beweis von 2.3. Es reicht offenbar aus, ein $\vartheta_1\colon C^1_{\mathscr{L}}(\mathfrak{U}, \mathfrak{G}) \to C^0_{\mathscr{L}}(\mathfrak{U}', \mathfrak{G})$ mit den entsprechenden Eigenschaften zu finden. Für ϑ wähle man dann die Komposition

$$C^1_{\delta}(\mathfrak{U}, \mathfrak{G}) \xleftarrow{\text{Incl.}} C^1_{\mathscr{L}}(\mathfrak{U}, \mathfrak{G}) \xrightarrow{\vartheta_1} C^0_{\mathscr{L}}(\mathfrak{U}', \mathfrak{G}) \xrightarrow{\text{Beschr.}} C^0_{\delta}(\mathfrak{V}, \mathfrak{G}).$$

Sei $g\in C^1_{\mathscr{L}}(\mathfrak{U}, \mathfrak{G})$ vorgegeben. Wir konstruieren $h=\vartheta_1(g)$ induktiv nach der Ordnung: $h^0 := \mathrm{id}$.

Sei h^e bereits konstruiert. Wir setzen

$$\xi_{ij} := (g_{ij}\circ h^e_j - h^e_i)\,[e+1]$$

(die eckige Klammer bezeichne dabei den homogenen Anteil der Ordnung $e+1$)

$$\xi := (\xi_{ij})_{i,j\in I} \in C^1_{\mathscr{L}}(\mathfrak{U}', \mathcal{O}^n_{\mathbb{C}^n\times B}).$$

Sei $\Delta := (\kappa\circ\lambda)(\xi)$ (komponentenweise)

$$h^{e+1} := h^e + \Delta.$$

Somit ist eine Abbildung $\vartheta_1\colon C^1_{\mathscr{L}}(\mathfrak{U},\mathfrak{G})\to C^0(\mathfrak{U}',\mathfrak{G})$ definiert. Wir zeigen nun: es gibt eine offene Umgebung W der Identität in $C^1_{\mathscr{L}}(\mathfrak{U},\mathfrak{G})$, so daß $\vartheta_1\colon W\to C^0_{\mathscr{L}}(\mathfrak{U}',\mathfrak{G})$ differenzierbar ist.

Nach dem Lemma 1.4 gibt es Konstanten $K\geqq 1$, $1\geqq\varepsilon_0>0$, so daß stets

$$\|\xi\|_{\mathfrak{U}',\rho}\leqq\|g\|_{\mathfrak{U},\rho}+K\|g\|_{\mathfrak{U},\rho}\,\|h^e\|_{\mathfrak{U}',\rho}$$

für jedes $\rho\in\mathbb{R}^m_{++}$ mit $\|h^e\|_{\mathfrak{U}',\rho}\leqq\varepsilon_0$. Sei

$$W:=\bigcup_{\rho\in\mathbb{R}^m_{++}}\left\{g\in C^1_{\mathscr{L}}(\mathfrak{U},\mathfrak{G})\colon\|g\|_{\mathfrak{U},\rho}<\frac{\varepsilon_0}{2KL}\right\};$$

hierbei bezeichne L eine Stetigkeitskonstante von κ.

Sei $g\in W$, also gilt insbesondere $\|g\|_{\mathfrak{U},\rho}\leqq\frac{\varepsilon_0}{2KL}$ für ein $\rho\in\mathbb{R}^m_{++}$, $\sigma:=\|g\|_{\mathfrak{U},\rho}$, $\sigma_1:=2KL\sigma$. Induktiv zeigt man nun $\|h^e\|_{\mathfrak{U}',\rho}\leqq\sigma_1$. Denn es ist

$$\|\xi\|_{\mathfrak{U}',\rho}\leqq\sigma+K\sigma\sigma_1,$$
$$\|\Delta\|_{\mathfrak{U}',\rho}\leqq L(\sigma+K\sigma\sigma_1)\leqq\frac{L\sigma_1}{2KL}+\frac{LK\sigma_1}{2KL}\sigma_1\leqq\sigma_1.$$

Somit ist ϑ_1 lokal beschränkt im Sinne von [7], 8.6. Für $\rho\in\mathbb{R}^m_{++}$ setzen wir $F:=(C^0_{\mathscr{L}}(\mathfrak{U}',\mathfrak{G}))_\rho=\{g\in C^0_{\mathscr{L}}(\mathfrak{U}',\mathfrak{G})\colon\|g-\mathrm{id}\|_{\mathfrak{U}',\rho}<\infty\}$. $F_i\subset F$ bezeichne die Menge aller Polynome vom Grad $\leqq i$ in t. Offensichtlich folgt die Differenzierbarkeit von ϑ_1 aus dem Lemma 2.1.

Sei nun g ein Zykel auf S, $h=\vartheta_1(g)$. Für jedes $e\in\mathbb{N}$ genügt es zu zeigen, daß $g_{ij}\circ h^e_j=h^e_i$ auf $S\cap B_e$ gilt, wobei $B_e=(0,\mathbb{C}\{t\}/\mathfrak{m}^{e+1})$. Dies folgt induktiv, denn es gilt $\mathrm{id}+\xi_{ij}=(h^e_i)^{-1}\circ g_{ij}\circ h^e_j$ über $S_{e+1}=S\cap B_{e+1}$ und somit $\xi\in Z^1_{\mathscr{L}}(\mathfrak{U}',\mathcal{O}^n_{\mathbb{C}^n\times S_{e+1}})$, d.h. es ist $\lambda(\xi)=\xi$ bzw. $\xi=\delta\Delta$, jeweils über S_{e+1}. Hieraus erhält man dann $\mathrm{id}=(h^{e+1}_i)^{-1}\circ g_{ij}\circ h^{e+1}_j$ über S_{e+1}. Der Satz ist damit bewiesen.

Mit Hilfe der von Hörmander [10] entwickelten Technik zur Lösung des $\bar\partial$-Problems mit L^2-Abschätzungen kann man nun die folgende Aussage beweisen:

2.5. Satz. *Sei $\Omega\subset\mathbb{C}^n$ offen, $\varphi\colon\Omega\to\mathbb{R}$ differenzierbar, $c\in\mathbb{R}$ ein regulärer Wert von φ und $B_c:=\{z\in\Omega\colon\varphi(z)<c\}$. Die Leviform von φ habe in jedem Punkt von Ω mindestens drei negative oder n-positive Eigenwerte.*

Behauptung: Sei $V\subset\subset\Omega$ offen und Steinsch. Für jede endliche Familie $(W_i)_{i\in I}$ offener Teilmengen des $\mathbb{C}^n$ mit $B_c\cap V\subset\subset\bigcup_{i\in I}W_i$ ist für die Überdeckung $\mathfrak{U}:=(B_c\cap V\cap W_i)_{i\in I}$ von $B_c\cap V$ die Bedingung $(+)$ erfüllt.

Beweis. Wir setzen abkürzend $U:=B_c\cap V$, $U_i:=B_c\cap V\cap W_i$, $\mathfrak{U}:=(U_i)_{i\in I}$. Es gibt differenzierbare Funktionen $\sigma_i\colon W_i\to\mathbb{R}_+$ mit $\operatorname{supp}\sigma_i\subset\subset W_i$ und $\sum_{i\in I}\sigma_i=1$ auf U.

Zum Nachweis der Behauptung genügt es zu zeigen, daß der Čech-Corand-operator

$$\delta\colon C^0_h(\mathfrak{U},\mathcal{O})\to Z^1_h(\mathfrak{U},\mathcal{O})$$

surjektiv ist. Sei also $\xi=(\xi_{ij})\in Z_h^1(\mathfrak{U},\mathcal{O})$. Wir setzen $\eta_i:=\sum_{\nu\in I}\sigma_\nu\xi_{\nu i}$, dann ist $\eta:=(\eta_i)\in C_h^0(\mathfrak{U},\mathcal{E})$ und $\xi=\delta\eta$. Es gilt für $\omega:=(\bar\partial\eta_i)$, $\omega\in\Gamma_h(U,\mathcal{E}^{0,1})$, $\bar\partial\omega=0$.

Aufgrund von [10], Abschnitt 3.3 und [10], Theorem 1.1.4 existiert ein $f\in\Gamma_h(U,\mathcal{E})$ mit $\omega=\bar\partial f$. Sei $f_i:=f|_{U_i}$, $\varphi_i:=\eta_i-f_i$.

Es folgt dann: $\bar\partial\varphi_i=0$,

$$\xi_{ij}=\eta_j-\eta_i=\eta_j-f_j-(\eta_i-f_i)=\varphi_j-\varphi_i.$$

Also $\xi=\delta\varphi$, $\varphi=(\varphi_i)\in C_h^0(\mathfrak{U},\mathcal{O})$.

Der Satz 2.3 liefert unter Verwendung von Anhang 2.3:

2.6. Corollar. *Mit den Voraussetzungen und Bezeichnungen von 2.5 gilt:*

Besitzt V ein Fundamentalsystem offener Steinscher Umgebungen, so gilt für jeden komplexen Raumkeim S die Identität $H^1(\overline{B_c\cap V},\mathfrak{G}_S)=1$.

§ 3. Der Glättungssatz

3.0. Sei Y eine q-konkave oder 1-konvexe Mannigfaltigkeit, $q<n-1$, $\varphi\colon Y\to\mathbb{R}$ eine q-konkave (bzw. 1-konvexe) Ausschöpfungsfunktion mit Ausnahmemenge K, ferner sei $c\in\mathbb{R}$ regulärer Wert von φ, $\sup_K\varphi<c<\sup_Y\varphi$, $B_c:=\{y\in Y\colon\varphi<c\}$. $\mathfrak{U}^0$ bezeichne eine lokal endliche Kartenüberdeckung von Y.

Unter Verwendung der von Grauert eingeführten Ausbeulungstechnik folgt mit den Ergebnissen des letzten Paragraphen der folgende Satz:

3.1. Satz (Glättungssatz). *Es gibt ein $c_1\in\mathbb{R}$, $\sup_K\varphi<c_1<c$, beliebig feine Meßfamilien (bzgl. der Karten von $\mathfrak{U}^0$) $\mathfrak{W}\ll\mathfrak{V}\ll\mathfrak{U}$ mit $\bar B_c\subset|\mathfrak{V}|$, und eine differenzierbare Abbildung*

$$\Omega\colon C_\delta^1(\mathfrak{V},\mathfrak{G})\to C_\delta^1(\mathfrak{U},\mathfrak{G})\times C_\delta^0(\mathfrak{W},\mathfrak{G})$$

mit folgenden Eigenschaften:

Ist $S\subset B$ ein Unterkeim und $g\in C_\delta^1(\mathfrak{V},\mathfrak{G})$ ein Cozyklus über S, $(G,h):=\Omega(g)$, so ist G ein Cozyklus über S und bzgl. $\mathfrak{W}_{c_1}$ gilt die Äquivalenz $h\colon g\to G$ von Cozyklen über S. Dabei sei $\mathfrak{W}_{c_1}:=\{W\in\mathfrak{W}\colon W\subset B_{c_1}\}$.

Der Beweis der entsprechenden Aussage für kohärente analytische Garben findet sich in [12], §§ 3, 4. Verwendet man nun die Aussagen 2.3 und 2.5 sowie eine 0-Zyklenprojektion, so folgt mit diesem Beweis die Behauptung von 3.1. Man vergleiche auch hierzu die detaillierte Darstellung in [7], § 14.

§ 4. Konstruktion einer konvergenten formal semiuniversellen Deformation

4.1. Sei Y eine q-konkave komplexe Mannigfaltigkeit mit $q<n-1$, $n=\dim Y$. B_c sei, wie in 3.0 vorgegeben, $Q:=\bar B_c$. Unser Ziel ist es, für den Funktor F_Q (vgl. I.3.8) ein konvergentes formal semiuniverselles Element x zu finden. Wir gehen dabei wie in [7], Kapitel III vor, wobei sich in unserem Fall erhebliche

Vereinfachungen ergeben. Nach Schlessinger [15] gibt es ein formales formal semiuniverselles Element

$$\mathfrak{x} \in \hat{F}_Q(T), \qquad \mathcal{O}_T = \mathbb{C}[\![t_1, \ldots, t_m]\!]/\mathfrak{a}, \qquad m = \dim_{\mathbb{C}} H^1(Y, \theta_Y).$$

Λ sei das reduzierende System von $\mathfrak{a}$, $a := \sup_{\lambda \in \Lambda} |\lambda|$. Das formale Element $\mathfrak{x}$ wird von der Ordnung a ab sukzessive *konvergent* abgeändert, so daß die formale Semiuniversalität erhalten bleibt.

Mit den entsprechenden Bezeichnungen kann der Konvergenzbeweis *genauso* wie in [7], Kapitel III geführt werden, wobei hier noch folgende zwei Punkte zu beachten sind:

4.2. Nach [14], Théorème 2 ist die Cohomologiegruppe $H^2(B_d, \theta_Y)$ für $d \in \mathbb{R}$, $\sup_K \varphi < d < \sup_Y \varphi$, separiert. Insbesondere folgt dann aus der exakten Sequenz der Terme niedrigen Grades der Spektralsequenz für die Čech-Cohomologie auf einer Überdeckung $\mathfrak{U} = (U_i)_{i \in I}$ von B_d mit $H^1(U_i, \theta_Y) = 0$, für alle $i \in I$, auch die Separiertheit von $H^2(\mathfrak{U}, \theta_Y)$.

Aus diesem Grund bleibt der Fortsetzungssatz um eine Ordnung [7], 17.8 richtig.

4.3. Der Glättungssatz zweiter Art (Glättung in H^0) [7], 19.3 ist gültig, da das dabei verwendete Lemma 7.8 von [7] im q-konkaven Fall für $q < n$ richtig bleibt.

Die somit konstruierte konvergente *formal* semiuniverselle Deformation für F_Q wird im folgenden Kapitel zu einer Deformation von ganz Y ausgedehnt und die *konvergente* Semiuniversalität nachgewiesen.

Kapitel III. Der Vollständigkeitssatz

§ 1. Der Vollständigkeitssatz für den Funktor F_Q

Das Ziel dieses Paragraphen ist es, den folgenden Satz zu beweisen:

1.1. Satz. *Mit den Bezeichnungen von* II.4 *gilt:*

Ist $a \in F_Q(S)$ ein konvergentes, formal semiuniverselles Element, so ist a bereits semiuniversell.

Bemerkung. Es sei darauf hingewiesen, daß diese Aussage nicht ohne weiteres aus dem allgemeinen Vollständigkeitssatz von Bingener [2], (3.2) folgt.

1.2. Festlegung der Daten. Wir können annehmen, daß es – bezogen auf eine Kartenüberdeckung von Y – eine Meßkette $\mathfrak{W} \ll \mathfrak{V} \ll \mathfrak{U}$ gibt, die folgenden Bedingungen genügt:

i) $Q_0 = \bar{B}_{c_0} \subset |\mathfrak{W}|$, $c_0 > c$, die Beschränkungsabbildung $\Gamma(V_i, \theta_Y) \to \Gamma(V_i \cap B_{c_0}, \theta_Y)$ ist bijektiv, $H^1(V_i \cap Q_0, \theta_Y) = 0$, wobei $\mathfrak{V} = (V_i)_{i \in I}$

ii) $a \in F_Q(S)$, $\mathcal{O}_S = \mathbb{C}\{t_1, \ldots, t_m\}/\mathfrak{a}$, $\mathfrak{a} \subset \mathfrak{m}^2$, $m = \dim_{\mathbb{C}} H^1(Q, \theta_Y) = \dim_{\mathbb{C}} H^1(Y, \theta_Y)$, a werde durch ein $x \in C^1_\delta(\mathfrak{U}, \mathfrak{G}_B)$ gegeben

iii) die Testdeformation $b \in F_Q(S')$, $\mathcal{O}_{S'} = \mathbb{C}\{s_1, \ldots, s_l\}/\mathcal{J}$ sei durch ein $\mathcal{J}$-reduziertes Element $y \in C^1_\delta(\mathfrak{V}, \mathfrak{G}_{B'})$ repräsentiert, wobei $B' = (\mathbb{C}^l, 0)$.

Sei $\mathfrak{V}_0 := \mathfrak{V} \cap B_{c_0}$. Zum Beweis des Satzes genügt es, einen Morphismus α: $\mathbb{C}\{t\} \to \mathbb{C}\{s\}$, $\alpha(\mathfrak{a}) \subset \mathcal{J}$, und eine Cokette $h \in C^0_\delta(\mathfrak{V}_0, \mathfrak{G}_{B'})$ zu finden, die eine Äquivalenz h: $\alpha^* x \to y$ über S' liefert.

Δ bezeichne im folgenden stets die $\mathcal{J}$-reduzierten Indizes im $\mathbb{N}^l$, $S'_e := B'_e \cap S'$ für $e \in \mathbb{N}$.

1.3. *Beschränkte Cohomologie im konkaven Fall*

1.3.1. Lemma. *Sei $y \in \partial B_{c_0}$ und $q < n$. Dann ist die Beschränkungsabbildung $\Gamma_b(V_i, \mathcal{O}_Y) \to \Gamma_b(V_i \cap B_{c_0}, \mathcal{O}_Y)$ eine Isometrie für alle $i \in I$.*

Beweis. Offensichtlich genügt es, folgende Aussage zu beweisen: Seien $V' \subset V \subset\subset \mathbb{C}^n$ offen, V zusammenhängend. Ist die Beschränkung $\Gamma(V, \mathcal{O}) \to \Gamma(V', \mathcal{O})$ bijektiv, so ist $\Gamma_b(V, \mathcal{O}) \to \Gamma_b(V', \mathcal{O})$ eine Isometrie. Wir nehmen an, es gebe ein $f \in \Gamma(V, \mathcal{O})$ mit $|f|_{V'} < \infty$ aber $|f|_V > |f|_{V'}$. Es existiert dann ein $x_0 \in V \setminus V'$ mit $|f(x_0)| > |f|_{V'}$. Sei $\varphi := \dfrac{1}{f - f(x_0)}$, $\varphi \in \Gamma_b(V', \mathcal{O})$. Es gibt nun ein $\Phi \in \Gamma(V, \mathcal{O})$ mit $\Phi|_{V'} = \varphi$, also $\Phi \cdot (f - f(x_0)) \equiv 1$ auf V' und nach dem Identitätssatz auch auf V. Dies ist aber ein Widerspruch.

1.3.2. Satz. *Sei $q < n - 1$. Die kanonischen Morphismen*

$$H^1_b(\mathfrak{V}_0, \theta_Y) \to H^1(\mathfrak{V}_0, \theta_Y),$$
$$H^1(\mathfrak{V}_0, \theta_Y) \to H^1(Q, \theta_Y)$$

sind bijektiv.

Beweis. Es gilt $H^1(\mathfrak{V}_0, \theta_Y) \subset H^1(B_{c_0}, \theta_Y) = H^1(Q, \theta_Y)$. Daß die erste Abbildung injektiv ist, folgt aus der Fortsetzbarkeit der 0-Coketten und Betrachtung der Cozyklenrelation. Es ist dann nur noch die Surjektivität von $H^1_b(\mathfrak{V}_0, \theta_Y) \to H^1(Q, \theta_Y)$ einzusehen. Dies ist aber klar, da $H^1(Y, \theta_Y) \to H^1(Q, \theta_Y)$ surjektiv ist.

1.3.3. Wir schreiben $x^1 = \mathrm{id} + \sum_{|\nu|=1} \xi_\nu t^\nu$. Aufgrund der Effektivität von a bilden die Elemente $(\xi_\nu)_{|\nu|=1}$ eine Basis von $H^1(Q, \theta_Y)$ und damit auch von $H^1_b(\mathfrak{V}_0, \theta_Y)$. Wir definieren

$$T\colon \mathbb{C}^m \times C^0(\mathfrak{V}_0, \theta_Y) \to Z^1(\mathfrak{V}_0, \theta_Y)$$
$$((c_\nu)_{|\nu|=1}, H) \to \sum_{|\nu|=1} c_\nu \xi_\nu + \delta H.$$

Entsprechend sei $T_b\colon \mathbb{C}^m \times C^0_b(\mathfrak{V}_0, \theta_Y) \to Z^1_b(\mathfrak{V}_0, \theta_Y)$ definiert. T und T_b sind surjektiv. Insbesondere existiert dann nach dem Satz von Banach eine Rückzugskonstante L zu T_b.

1.4. Formale Fortsetzung um eine Ordnung. Wir nehmen an, wir hätten für ein $e \in \mathbb{N}$ schon $\alpha^e \in \mathbb{C}^m\{s\}^0$, $\deg \alpha^e \leqq e$, $h^e \in C^0(\mathfrak{V}_0, \mathfrak{G}_{B'})$, $\deg h^e \leqq e$, gefunden, so daß h^e: $(\alpha^e)^* x^e \to y^e$ eine Äquivalenz über S'_e ist. O.B.d.A. seien alle diese

Objekte $\mathscr{J}_e$-reduziert, $\mathscr{J}_e = \mathscr{J} + \mathfrak{m}_{B'}^{e+1}$. Sei $w := ((\alpha^e)^* x - h^e_* y)[e+1]$, so ist $w \in C^1(\mathfrak{B}_0, \mathfrak{H})$ homogen vom Grad $e+1$. $h^e_* y$ bedeute dabei die Konjugation von y mit h^e. $v := \mathrm{Red}_{\mathscr{J}_{e+1}} w$.

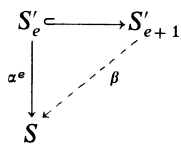
1.4.1. Lemma. $v \in Z^1(\mathfrak{B}_0, \theta_Y) \otimes_{\mathbb{C}} (\mathfrak{m}_{S'}^{e+1}/\mathfrak{m}_{S'}^{e+2})$

Beweis. Wir wollen die Aussage Anhang 2.1 auf die exakte Sequenz

$$0 \to \theta_Y \otimes_{\mathbb{C}} (\mathfrak{m}_{S'}^{e+1}/\mathfrak{m}_{S'}^{e+2}) \to \mathfrak{G}_{S'_{e+1}} \to \mathfrak{G}_{S'_e} \to 1$$

anwenden. Aus Lemma II.1.5 folgt zunächst $((\alpha^e)^* x) \circ (h^e_* y)^{-1} \equiv \mathrm{id} + (\alpha^e)^* x - (h^e_* y) \bmod \mathscr{J}_{e+1}$. Somit genügt es zu zeigen, daß das Bild von $(\alpha^e)^* x$ in $C^1(\mathfrak{U}, \mathfrak{G}_{S'_{e+1}})$ ein Zykel ist. Dies folgt so:

Es reicht zu zeigen, daß α^e sogar einen Morphismus $S'_{e+1} \to S_{e+1}$ definiert, d.h. $f_k(\alpha^e) \equiv 0 \bmod \mathscr{J}_{e+1}$ gilt, für ein Erzeugendensystem $f_1, \ldots, f_r$ von $\mathfrak{a}$. Sei abkürzend $\mathrm{Red} := \mathrm{Red}_{\mathscr{J}_{e+1}}$. Wegen der formalen Semiuniversalität von a existiert insbesondere ein kommutatives Diagramm



Es gibt also zu jedem $\mu \in \Delta_{e+1}$ ein $c_\mu \in \mathbb{C}^m$, so daß für $1 \leqq k \leqq r$ gilt:

$$\mathrm{Red}\,(f_k(\alpha^e + \sum_{\mu \in \Delta_{e+1}} c_\mu s^\mu)) = 0.$$

Taylorentwickeln von f_k um α_e ergibt:

$$f_k(\alpha^e + \textstyle\sum c_\mu s^\mu) \equiv f_k(\alpha^e) + Df_k(0) \sum c_\mu s^\mu \bmod \mathfrak{m}_{B'}^{e+2}.$$

Da nun $\mathfrak{a} \subset \mathfrak{m}_{B'}^2$ folgt $0 = \mathrm{Red}\,(f_k(\alpha^e + \sum c_\mu s^\mu)) = \mathrm{Red}\,(f_k(\alpha^e))$, womit das Lemma bewiesen ist.

1.4.2. Satz (Fortsetzung um eine Ordnung). *Es existiert ein Polynom* $\alpha^{e+1} \in \mathbb{C}^m[s]^0$ *und eine Cokette* $h^{e+1} \in C^0(\mathfrak{B}_0, \mathfrak{G}_{B'})$, *jeweils* $\mathscr{J}_{e+1}$*-reduziert, mit:*

i) $\alpha^{e+1} \equiv \alpha^e \bmod \mathfrak{m}_{B'}^{e+1}$,

ii) $h^{e+1} \equiv h^e \bmod \mathfrak{m}_{B'}^{e+1}$,

iii) α^{e+1} *induziert einen Morphismus* $S'_{e+1} \to S_{e+1}$,

iv) $h^{e+1}\colon (\alpha^{e+1})^* x^{e+1} \to y^{e+1}$ *ist eine Äquivalenz über* S'_{e+1}.

Beweis. Sei $v = \sum_{\mu \in \Delta_{e+1}} v_\mu s^\mu$, $v_\mu \in Z^1(\mathfrak{B}_0, \theta_Y)$ und

$$T((c_\nu^{(\mu)}), H^{(\mu)}) = -v_\mu, \beta := \sum_{\mu \in \Delta_{e+1}} c^{(\mu)} s^\mu, \; H := \sum_{\mu \in \Delta_{e+1}} H^{(\mu)} s^\mu,$$

wobei $c^{(\mu)} := (c_\nu^{(\mu)})_{|\nu|=1}$.

Wir setzen $\alpha^{e+1} := \alpha^e + \beta$, $h^{e+1} := h^e + H$.

Die Eigenschaften i und ii sind klar, iii folgt aus dem Beweis von 1.4.1. Sei $\Phi_{ij}=(u_{ij}, \mathrm{id}_{B'})$ der Kartenwechsel. Mit Hilfe von Lemma II.1.5 folgt:

$$\begin{aligned}&((\alpha^{e+1})^* x - h^{e+1}\dot{.} y)_{ij}\\ &\quad=(\alpha^{e+1})^* x_{ij} - h_i^{e+1}\circ y_{ij}\circ\Phi_{ij}\circ(h_j^{e+1})^{-1}\circ\Phi_{ji}\\ &\quad\equiv((\alpha^e)^* x - h_{\cdot}^e y)_{ij}+\Big(\sum_{|\nu|=1}\xi_\nu\sum_{\mu\in\Lambda_{e+1}}c_\nu^{(\mu)}s^\mu\Big)_{ij}\\ &\qquad-(H_i - D\Phi_{ij}(\Phi_{ji})H_j)\equiv 0 \bmod \mathfrak{m}_{B'}\mathscr{J}_e.\end{aligned}$$

Somit ist

$$(\alpha^{e+1})^* x - h^{e+1}\dot{.} y\equiv(\alpha^e)^* x - h_{\cdot}^e y - v \bmod \mathscr{J}_{e+1}.$$

Damit ist Satz 1.3.2 bewiesen.

1.5. Konvergente Fortsetzung um eine Ordnung. Die im vorigen Abschnitt durchgeführte Konstruktion wird nun mit Abschätzung vollzogen.

Für die Weierstraßfamilie $\omega_\lambda = s^\lambda + \alpha_\lambda$, $\lambda\in\Lambda$, des Ideals $\mathscr{J}$ sei $R_{\mathscr{J}}:=\left\{\rho\in\mathbb{R}^s_{++}:\rho\leqq(1,\ldots,1),\ \|\alpha_\lambda\|_\rho\leqq\frac{\rho^\lambda}{2|\Lambda|}\right\}$. Für $\rho\in R_{\mathscr{J}}$ und jeden Banachraum E gilt $\|\mathrm{Red}_{\mathscr{J}}(f)\|_\rho\leqq 2\|f\|_\rho$, $f\in E\{s\}$. Zu jedem $\varepsilon>0$ gibt es stets ein $\rho\in R_{\mathscr{J}}$ mit $\rho_i\leqq\varepsilon$, $1\leqq i\leqq l$.

1.5.1. Überschußlemma. *Mit den vorherigen Bezeichnungen gilt: Es gibt Konstanten $\varepsilon>0$, $K_0\geqq 1$ mit folgenden Eigenschaften:*

Sei $\rho\in R_{\mathscr{J}}$, $0<\sigma\leqq\tau\leqq\varepsilon$, es gelte

i) $\|y\|_{\mathfrak{B},\rho}\leqq\sigma$,

ii) $\|(\alpha^e)^* x\|_{\mathfrak{U},\rho}\leqq\tau$,
$\|((\alpha^e)^* x)[e+1]\|_{\mathfrak{U},\rho}\leqq\sigma$,

iii) $\|h^e\|_{\mathfrak{B}_0,\rho}\leqq\tau$.

Behauptung. Es gilt die Abschätzung

$$\|v\|_{\mathfrak{B}_0,\rho}\leqq K_0(\sigma+\tau^2).$$

Beweis. Es ist $w=((\alpha^e)^* x - h_{\cdot}^e y)[e+1]$ und daher

$$w_{ij}\equiv(\alpha^e)^* x_{ij}\circ\Phi_{ij}\circ h_j^e\circ\Phi_{ji} - h_i^e\circ y_{ij}\bmod\mathscr{J}_{e+1}.$$

Wegen 1.3.1 können wir annehmen, daß $h^e\in C^0_\varepsilon(\mathfrak{B},\mathfrak{G}_{B'})$, $\|h^e\|_{\mathfrak{B},\rho}\leqq\tau$. Wir kürzen ab $f_{ij}:=((\alpha^e)^* x_{ij})[\leqq e+1]$ und schreiben

$$f_{ij}\circ\Phi_{ij}\circ h_j^e\circ\Phi_{ji}=(f_{ij}[\leqq e]+f_{ij}[e+1])\circ\Phi_{ij}\circ h_j^e\circ\Phi_{ji}.$$

Für hinreichend kleines $\varepsilon>0$ gilt mit II.1.4:

$$\|(f_{ij}\circ\Phi_{ij}\circ h_j^e\circ\Phi_{ji})[\leqq e+1]\|_{V_i\cap V_j,\rho}\leqq K_1(\sigma+\tau^2).$$

Wir betrachten $h_i^e\circ y_{ij}$:

$$\begin{aligned}h_i^e\circ y_{ij}&=y_{ij}+\hat{h}_i^e\circ(\mathrm{id}+\hat{y}_{ij})\\ \hat{h}_i^e\circ(\mathrm{id}+\hat{y}_{ij})&=\hat{h}_j^e+R_{ij}\\ \|R_{ij}\|_{W_i\cap V_j,\rho}&\leqq K_2\tau\sigma.\end{aligned}$$

Insgesamt folgt also:

$$\|w_{ij}\|_{W_i \cap V_j, \rho} \leqq K_3(\sigma+\tau^2), \quad \text{mit } K_3 := \max\{1+K_1, K_1+K_2\}$$

und daher

$$\|v_{ij}\|_{W_i \cap V_j, \rho} \leqq 2K_3(\sigma+\tau^2).$$

Weil $v \in Z^1(\mathfrak{V}_0, \theta_Y) \otimes_{\mathbb{C}} (\mathfrak{m}_{S'}^{e+1}/\mathfrak{m}_{S'}^{e+1})$ und $B_{c_0} \subset |\mathfrak{V}|$ ergibt sich mit Hilfe der Cozyklenrelation $\|v\|_{\mathfrak{V}_0, \rho} \leqq 4K_3(\sigma+\tau^2)$. Wir setzen $K_0 := 4K_3$ und man erhält die Behauptung.

Wir wählen jetzt $\tilde{\rho} \in \mathbb{R}_{++}^m$ so klein, daß

$$\|x\|_{\mathfrak{U}, \tilde{\rho}} \leqq 1, \qquad \|\Gamma\|_{\mathfrak{U}, \tilde{\rho}} \leqq \frac{\gamma(\tilde{\rho})}{2LK_0}$$

wobei $\Gamma := x - x^1$ gesetzt wurde und $\gamma(\tilde{\rho}) = \min_{1 \leq i \leq m} \tilde{\rho}_i$. Sei $\tau \leqq \min\left\{\varepsilon, \frac{\gamma(\tilde{\rho})}{2LK_0}\right\}$, $\sigma := \frac{\gamma(\tilde{\rho})\tau}{2LK_0}$ und ferner $\rho \in R_{\mathscr{J}}$ so klein, daß $\|y\|_{\mathfrak{V}, \rho} \leqq \sigma$.

1.5.2. Konvergente Fortsetzung. Es gibt $\alpha^e \in \mathbb{C}^m[s]^0$, $h^e \in C_{\mathscr{J}}^0(\mathfrak{V}_0, \mathfrak{G}_{B'})$, jeweils $\mathscr{J}_e$-reduziert, mit folgenden Eigenschaften

i) $\alpha^{e+1} \equiv \alpha^e \bmod \mathfrak{m}_{B'}^{e+1}$

ii) $h^{e+1} \equiv h^e \bmod \mathfrak{m}_{B'}^{e+1}$

iii) h^e: $(\alpha^e)^* x \to y$ ist eine Äquivalenz über S'_e

iv) $\|\alpha^e\|_\rho \leqq \gamma(\tilde{\rho})\tau$, $\|h^e\|_{\mathfrak{V}_0, \rho} \leqq \tau$

Setzt man $h := \varprojlim h^e$, $\alpha := \varprojlim \alpha^e$, so ist h: $(\alpha)^* x \to y$ eine Äquivalenz über S'.

Beweis von 1.5.2. $\alpha_0 := 0$, $h^0 := \mathrm{id}$ definiert den Induktionsanfang. Die Behauptung sei also schon für $e \in \mathbb{N}$ gezeigt. Nach Lemma II.1.7 gilt

$$\|(\alpha^e)^* x\|_{\mathfrak{U}, \rho} \leqq \frac{1}{\gamma(\tilde{\rho})} \|\alpha^e\|_\rho \|x\|_{\mathfrak{U}, \tilde{\rho}},$$

$$\|((\alpha^e)^* x)[e+1]\|_{\mathfrak{U}, \rho} = \|((\alpha^e)^* \Gamma)[e+1]\|_{\mathfrak{U}, \rho} \leqq \tau \|\Gamma\|_{\mathfrak{U}, \tilde{\rho}} \leqq \sigma.$$

Aus dem Überschußlemma folgt:

$$\|v\|_{\mathfrak{V}_0, \rho} \leqq K_0(\sigma+\tau^2).$$

Es gibt dann β und H mit $\alpha^{e+1} = \alpha^e + \beta$, $h^{e+1} = h^e + H$ und $\|H\|_{\mathfrak{V}_0, \rho} \leqq LK_0(\sigma + \tau^2) \leqq \gamma(\tilde{\rho})\tau$, $\|\beta\|_\rho \leqq \gamma(\tilde{\rho})\tau$, dies aufgrund der Surjektivität von T_b. Damit ist 1.5.2 und auch der Satz 1.1 bewiesen.

§ 2. Beweis des Vollständigkeitssatzes für F_Y

Wir legen die Situation von II.4 zugrunde.

2.1. Satz. *Es gibt ein $\delta > 0$, so daß für $c - \delta \leqq c_1 < c_2 \leqq c + \delta$ und $Q_i := \bar{B}_{c_i}$, $i = 1, 2$ gilt:*

Für jedes $z \in H^1(Q_2, \mathfrak{G}_S)$ sind die Beschränkungsabbildungen bijektiv

$$H^1(Q_2, \mathfrak{G}_S) \to H^1(Q_1, \mathfrak{G}_S),$$
$$\Gamma(Q_2, {}^z\mathfrak{G}_S) \to \Gamma(Q_1, {}^z\mathfrak{G}_S).$$

Dabei sei S ein komplexer Raumkeim.

Wir notieren die Folgerung:

2.2. Corollar. *Es gibt eine streng monoton wachsende Folge reeller Zahlen $(c_i)_{i\in\mathbb{N}}$ mit $\sup_K \varphi < c_i < \sup_Y \varphi$, $Q_i := \bar{B}_{c_i}$, und $Y = \bigcup_{i\in\mathbb{N}} Q_i$, so daß gilt:*
Für jedes $i \in \mathbb{N}$ und $z \in H^1(Y, \mathfrak{G}_S)$ sind die Beschränkungen bijektiv

$$H^1(Q_{i+1}, \mathfrak{G}_S) \to H^1(Q_i, \mathfrak{G}_S),$$
$$\Gamma(Q_{i+1}, {}^z\mathfrak{G}_S) \to \Gamma(Q_i, {}^z\mathfrak{G}_S).$$

Der Satz Anhang 3.1 liefert mit II.2.6 das

2.3. Corollar. *Die Beschränkung $H^1(Y, \mathfrak{G}_S) \to H^1(Q_0, \mathfrak{G}_S)$ ist bijektiv für jeden komplexen Raumkeim S.*

Somit erhält man aus Satz 1.1 die Aussage:

2.4. Vollständigkeitssatz. *Jedes formal semiuniverselle Element aus $F_Y(S) = H^1(Y, \mathfrak{G}_S)$ ist bereits semiuniversell.*

Es ist somit nur noch der Satz 2.1 zu beweisen. Dies geschieht mit Hilfe der Grauertschen Ausbeulungstechnik [1].

Wir setzen abkürzend $\mathfrak{G} := \mathfrak{G}_S$. $\mathfrak{U} = (U_i^0)_{1\leqq i\leqq t}$ sei eine Kartenüberdeckung von ∂B_c.

Ferner sei $U_i \subset C_i \subset\subset U_i^0$, $1 \leqq i \leqq t$, wobei C_i ein Steinsches Kompaktum ist und $\partial B_c \subset \bigcup_{i=1}^{t} U_i$ gilt. Abkürzend schreiben wir $U := \bigcup_i U_i$, $U^0 := \bigcup_i U_i^0$. Seien $\sigma_i : Y \to \mathbb{R}_+$ differenzierbare Funktionen mit $\operatorname{supp} \sigma_i \subset \mathring{C}_i$ und $\sum_{i=1}^{t} \sigma_i = 1$ auf U. Es gibt ein $\varepsilon_0 > 0$, so daß für alle ε mit $0 \leqq \varepsilon \leqq \varepsilon_0$ die Funktionen $\tilde{\varphi}_r := \varphi - \varepsilon \sum_{i=1}^{r} \sigma_i$ streng q-konkav auf U^0 sind für $0 \leqq r \leqq t$. Außerdem sei $\delta_0 > 0$ so klein, daß $\partial B_{c+\delta} \subset U$ für alle $\delta \in \mathbb{R}$ mit $|\delta| \leqq \delta_0$. Wir setzen $\delta := \min\left\{\delta_0, \frac{\varepsilon_0}{2}\right\}$. Seien $c_1, c_2 \in \mathbb{R}$ mit $c - \delta \leqq c_1 < c_2 \leqq c + \delta$,

$$\varphi_r := \varphi - (c_2 - c_1) \sum_{i=1}^{r} \sigma_i,$$
$$Q^r := B_{c_1} \cup \{y \in U : \varphi_r(y) \leqq c_1\}$$

$0 \leqq r \leqq t$. Insbesondere gilt dann

$$Q^0 = \bar{B}_{c_1}, \quad Q^t = \bar{B}_{c_2},$$
$$Q^r \supset Q^{r-1}, \quad Q^r \setminus Q^{r-1} \subset \mathring{C}_r, \quad 1 \leqq r \leqq t.$$

Nach eventueller Verkleinerung von ε_0 kann man nach II.2.5, II.2.6 annehmen, daß die obige Konstruktion so eingerichtet ist, daß folgende Bedingungen erfüllt sind:

Die Beschränkungsabbildung $\Gamma(C_i, \mathfrak{G}) \to \Gamma(Q^r \cap C_i, \mathfrak{G})$ ist bijektiv, und es ist $H^1(Q^r \cap C_i, \mathfrak{G}) = 1$ für $1 \leqq i \leqq t$, $0 \leqq r \leqq t$.

Die Mayer-Vietoris-Sequenz (Anhang 1.4) liefert für $Q^r = Q^{r-1} \cup (Q^r \cap C_r)$ die Bijektivität der Beschränkung $H^1(Q^r, \mathfrak{G}) \to H^1(Q^{r-1}, \mathfrak{G})$ und somit auch die von $H^1(Q_2, \mathfrak{G}) \to H^1(Q_1, \mathfrak{G})$.

Sei $z \in H^1(Q_2, \mathfrak{G})$ vorgegeben. Wir betrachten die exakte Sequenz

$$1 \to \Gamma(Q^r, {}^z\mathfrak{G}) \to \Gamma(Q^{r-1}, {}^z\mathfrak{G}) \times \Gamma(Q^r \cap C_r, {}^z\mathfrak{G}) \rightrightarrows \Gamma(Q^{r-1} \cap C_r, {}^z\mathfrak{G}).$$

Über C_r gilt nun ${}^z\mathfrak{G} \cong \mathfrak{G}$. Da aber $\Gamma(C_r, \mathfrak{G}) \to \Gamma(Q^{r-1} \cap C_r, \mathfrak{G})$ ein Isomorphismus ist, folgt somit auch die Bijektivität der Beschränkungsabbildung $\Gamma(Q^r, {}^z\mathfrak{G}) \to \Gamma(Q^{r-1}, {}^z\mathfrak{G})$.

Damit erhält man den zweiten Teil der Behauptung von Satz 2.1.

Anhang. Hilfsmittel aus der Cohomologietheorie nicht abelscher Garben

§ 1. Eine Mayer-Vietoris Sequenz

Sei X ein topologischer Raum, $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$ Garben von Gruppen auf X. Gegeben sei die Sequenz

$$(*) \qquad 1 \to \mathscr{A} \xrightarrow{\alpha} \mathscr{B} \underset{\gamma}{\overset{\beta}{\rightrightarrows}} \mathscr{C} \to 1$$

wobei α, β, γ Gruppenhomomorphismen sind und

$$1 \to \mathscr{A} \xrightarrow{\alpha} \mathscr{B} \xrightarrow{\beta^{-1} \cdot \gamma} \mathscr{C} \to 1$$

exakt ist. D.h., insbesondere $\mathscr{C} \cong \mathscr{B}/\mathscr{A}$.

1.1. Satz. *Unter den obigen Voraussetzungen ist die folgende Sequenz exakt*

$$1 \to \Gamma(X, \mathscr{A}) \xrightarrow{\alpha} \Gamma(X, \mathscr{B}) \underset{\gamma}{\overset{\beta}{\rightrightarrows}} \Gamma(X, \mathscr{C}) \xrightarrow{\delta} H^1(X, \mathscr{A}) \to H^1(X, \mathscr{B}) \underset{\gamma}{\overset{\beta}{\rightrightarrows}} H^1(X, \mathscr{C}).$$

Die Exaktheit bis zur Stelle $H^1(X, \mathscr{B})$ folgt aus [8], III Proposition 3.2.2. Eine einfache Rechnung liefert auch die Exaktheit bei $H^1(X, \mathscr{B})$.

1.2. Sei nun $z \in H^1(X, \mathscr{A})$, $u := \alpha(z)$, $v := (\beta \circ \alpha)(z) = (\gamma \circ \alpha)(z)$. Dann hat man bekanntlich die exakte Sequenz

$$1 \to {}^z\mathscr{A} \xrightarrow{\tilde{\alpha}} {}^u\mathscr{B} \underset{\tilde{\gamma}}{\overset{\tilde{\beta}}{\rightrightarrows}} {}^v\mathscr{C} \to 1$$

wobei $\tilde{\alpha}$, $\tilde{\beta}$, $\tilde{\gamma}$ durch α, β, γ induziert werden.

Sei $z' \in H^1(X, \mathscr{A})$ mit $\alpha(z) = \alpha(z')$.

Es existiert ein Element $\xi' \in \Gamma(X, {}^v\mathscr{C})$ mit $\tilde{\delta}\xi' = z'$, wobei hier z' mit dem Bild unter der bijektiven Abbildung $\theta_z \colon H^1(X, \mathscr{A}) \to H^1(X, {}^z\mathscr{A})$ identifiziert wurde. Ist insbesondere $\tilde{\beta}^{-1} \cdot \tilde{\gamma} \colon \Gamma(X, {}^u\mathscr{B}) \to \Gamma(X, {}^v\mathscr{C})$ surjektiv, so folgt $z = z'$. Gilt diese Surjektivität für jedes $z \in H^1(X, \mathscr{A})$, so ist $H^1(X, \mathscr{A}) \xrightarrow{\alpha} H^1(X, \mathscr{B})$ injektiv.

1.3. Sei X ein topologischer Raum, $X=A_1\cup A_2$, $A_i\subset X$ abgeschlossen, $A_3:=A_1\cap A_2$, $\mathfrak{G}$ eine Garbe von Gruppen auf X, i_k: $A_k\to X$ bezeichne die jeweiligen Inklusionen für $i=1,2,3$. Wir setzen $\mathfrak{G}_k:=(i_k)_*(i_k)^*\mathfrak{G}$. Es gibt dann die kanonische Sequenz

$$1\to\mathfrak{G}\xrightarrow{j}\mathfrak{G}_1\times\mathfrak{G}_2\underset{q}{\overset{p}{\rightrightarrows}}\mathfrak{G}_3\to 1$$

j,p,q sind die entsprechenden Adjunktionsmorphismen. Man zeigt leicht, z.B. durch halmweises Nachrechnen, daß diese Sequenz exakt ist.

Der Satz 1.1 liefert dann die exakte Cohomologiesequenz

$$\begin{aligned}1\to\Gamma(X,\mathfrak{G})\to\Gamma(A_1,\mathfrak{G})\times\Gamma(A_2,\mathfrak{G})\underset{q}{\overset{p}{\rightrightarrows}}\Gamma(A_3\mathfrak{G})\xrightarrow{\delta}H^1(X,\mathfrak{G})\\ \xrightarrow{j}H^1(A_1,\mathfrak{G})\times H^1(A_2,\mathfrak{G})\underset{q}{\overset{p}{\rightrightarrows}}H^1(A_3,\mathfrak{G}).\end{aligned}\tag{1.4}$$

1.5. Ist z.B. $H^1(A_2,\mathfrak{G})$ und $H^1(A_3,\mathfrak{G})$ trivial, so folgt die Surjektivität von $H^1(X,\mathfrak{G})\to H^1(A_1,\mathfrak{G})$.

1.6. Sei $H^1(A_2,\mathfrak{G})$ trivial, $z\in H^1(X,\mathfrak{G})$. Es ist dann $j(z)=(z_1,1)$, $1=(p\circ j)(z)=(q\circ j)(z)$. Man erhält somit

$$\begin{aligned}{}^{j(z)}(\mathfrak{G}_1\times\mathfrak{G}_2)&={}^{z_1}\mathfrak{G}_1\times\mathfrak{G}_2,\\ {}^{(p\circ j)}\mathfrak{G}_3&=\mathfrak{G}_3.\end{aligned}$$

Mit 1.2 ist daher die Abbildung $H^1(X,\mathfrak{G})\to H^1(A_1,\mathfrak{G})$ injektiv, falls $\Gamma(A_1,{}^{z_1}\mathfrak{G}_1)\times\Gamma(A_2,\mathfrak{G})\xrightarrow{p^{-1}\cdot q}\Gamma(A_3,\mathfrak{G})$ surjektiv ist für alle $z\in H^1(X,\mathfrak{G})$. Dies folgt insbesondere dann, wenn $\Gamma(A_2,\mathfrak{G})\xrightarrow{q}\Gamma(A_3,\mathfrak{G})$ schon surjektiv ist.

§ 2. Cohomologie auf Überdeckungen

Die folgenden drei elementaren Aussagen geben wir ohne Beweis an.

2.1. Satz. *Sei* $1\to\mathscr{F}\to\mathscr{G}\to\mathscr{H}\to 1$ *eine exakte Sequenz von Garben von Gruppen auf dem topologischen Raum* X, $\mathfrak{U}=(U_i)_{i\in I}$ *eine offene Überdeckung von* X. $\mathscr{G}(U_i)\to\mathscr{H}(U_i)$ *sei surjektiv für alle* $i\in I$.

Behauptung. Liegt $\mathscr{F}$ *im Zentrum von* $\mathscr{G}$, *so operiert* $H^r(\mathfrak{U},\mathscr{F})$ *transitiv auf den Fasern von* $H^r(\mathfrak{U},\mathscr{G})\to H^r(\mathfrak{U},\mathscr{H})$ *für* $r=0,1$.

2.2. Satz. *Sei* X *ein topologischer Raum,* $\mathfrak{U}=(U_i)_{i\in I}$ *eine offene oder abgeschlossene Überdeckung von* X; $I=\bigcup_{\nu\in\mathbb{N}}I_\nu$, *wobei* $I_0\subset I_1\subset I_2\subset\ldots$, $\mathfrak{G}$ *eine Garbe von Gruppen auf* X. *Wir setzen* $\mathfrak{U}_\nu:=(U_i)_{i\in I_\nu}$.

Behauptung. i) *die kanonische Abbildung*

$$H^1(\mathfrak{U},\mathfrak{G})\to\varprojlim H^1(\mathfrak{U}_\nu,\mathfrak{G})$$

ist surjektiv

ii) *erfüllt das projektive System* $(H^0(\mathfrak{U}_\nu, \mathfrak{G}))_{\nu\in\mathbb{N}}$ *die Mittag-Leffler Bedingung, so ist die Sequenz exakt*

$$1 \to H^1(\mathfrak{U}, \mathfrak{G}) \to \varprojlim H^1(\mathfrak{U}_\nu, \mathfrak{G}) \to 1.$$

2.3. Satz. *Sei X ein parakompakter topologischer Raum, $A \subset X$ eine abgeschlossene Teilmenge, $\mathfrak{G}$ eine Garbe von Gruppen auf X.*

Behauptung. Für $r=0,1$ ist die kanonische Abbildung $\varinjlim_{U \supset A} H^r(U, \mathfrak{G}) \to H^r(A, \mathfrak{G})$ *bijektiv.*

Dies ist die entsprechende Aussage von [9], Théorème 4.11.1 für eine Garbe von Gruppen.

Sei X ein topologischer Raum. Wir formulieren die Bedingung

(N) für je zwei abgeschlossene Teilmengen $A, B \subset X$ gelte für die vollen Umgebungskeime

$$(A_*) \cap (B_*) = (A \cap B)_*.$$

Die nicht triviale Inklusion bedeutet hierbei:

Zu jeder offenen Umgebung U von $A \cap B$ gibt es offene Umgebungen U_1 von A und U_2 von B mit $U_1 \cap U_2 \subset U$. Es gilt die

Hilfsaussage. *Sei jede offene Teilmenge von X ein T_4-Raum, so ist* (N) *stets erfüllt.*

Insbesondere besitzen metrische Räume stets die Eigenschaft (N).

2.4. Lemma. *Sei X ein metrisierbarer topologischer Raum, $\mathfrak{A} = (A_i)_{i\in I}$ eine abgeschlossene Überdeckung von X. Es gebe eine offene Überdeckung $\mathfrak{W} = (W_i)_{i\in I}$ mit $A_i \subset W_i$ und $W_i \cap W_j \neq \varnothing$ nur für endlich viele $j \in I$ zu vorgegebenem $i \in I$.*

Behauptung. Sei $\mathscr{F}$ eine Garbe auf X und $r \in \mathbb{N}$. Dann ist die kanonische Abbildung

$$\varinjlim_{\mathfrak{U} > \mathfrak{A}} C^r(\mathfrak{U}, \mathscr{F}) \to C^r(\mathfrak{A}, \mathscr{F})$$

bijektiv.

Dabei bedeutet $\mathfrak{U} > \mathfrak{A}$, $\mathfrak{U} = (U_i)_{i\in I}$, daß $U_i \supset A_i$ für alle $i \in I$.

Beweis. Zur Surjektivität: Sei $(f_{i_0 \ldots i_r}) \in C^r(\mathfrak{A}, \mathscr{F})$, $j = (i_0, \ldots, i_r) \in \mathrm{Nerf}_r\, \mathfrak{A}$ ein festes Tupel. Nach Satz 2.3 existiert ein offenes U_j und ein $g_j \in \Gamma(U_j, \mathscr{F})$ mit $A_{i_0} \cap \ldots \cap A_{i_r} \subset U_j$ und $g_j|_{A_{i_0} \cap \ldots \cap A_{i_r}} = f_j$. Da (N) erfüllt ist, gibt es offene $U_{i_0}^{(j)}, \ldots, U_{i_r}^{(j)}$ mit $A_{i_k} \subset U_{i_k}^{(j)} \subset W_{i_k}$ und $U_{i_0}^{(j)} \cap \ldots \cap U_{i_r}^{(j)} \subset U_j$. Für $i \in I$ setzen wir

$$U_i := \bigcap_{\substack{j \in \mathrm{Nerf}_r\, \mathfrak{A} \\ i \in |j|}} U_i^{(j)}.$$

Dieser Durchschnitt ist endlich und es ist $A_i \subset U_i$, $U_{i_0} \cap \ldots \cap U_{i_r} \subset U_{(i_0, \ldots, i_r)}$ für $(i_0, \ldots, i_r) \in \mathrm{Nerf}_r\, \mathfrak{A}$. Sei $\mathfrak{U} := (U_i)_{i\in I}$. Die Surjektivität ist damit klar. Auf dieselbe Weise zeigt man auch die Injektivität.

2.5. Satz. *Die Voraussetzungen seien wie in* 2.4.

Behauptung. Für jede Garbe von Gruppen $\mathfrak{G}$ auf X ist die kanonische Abbildung

$$\varinjlim_{\mathfrak{U}>\mathfrak{A}} H^r(\mathfrak{U},\mathfrak{G})\to H^r(\mathfrak{A},\mathfrak{G})$$

bijektiv für $r=0,1$.

Beweis. Da $\varinjlim_{\mathfrak{U}>\mathfrak{A}} H^r(\mathfrak{U},\mathfrak{G})\to H^r(\varinjlim_{\mathfrak{U}>\mathfrak{A}} C^\cdot(\mathfrak{U},\mathfrak{G}))$ für $r=0,1$ bijektiv ist, erhält man aus 2.4 die Behauptung.

2.6. Satz. *Die Voraussetzungen seien wieder wie in* 2.4.

Behauptung. Für jede Garbe $\mathfrak{G}$ von Gruppen auf X ist die kanonische Sequenz exakt

$$1\to H^1(\mathfrak{A},\mathfrak{G})\to H^1(X,\mathfrak{G})\to\prod_{i\in I} H^1(A_i,\mathfrak{G}).$$

Beweis. Diese Aussage ist allgemein bekannt für jede offene Überdeckung $\mathfrak{U}>\mathfrak{A}$. Ferner ist die kanonische Abbildung

$$\varinjlim_{\mathfrak{U}>\mathfrak{A}}\prod_{i\in I} H^1(U_i,\mathfrak{G})\to\prod_{i\in I} H^1(A_i,\mathfrak{G})$$

nach 2.3 bijektiv. Durch Übergang zum Limes folgt die Behauptung auch für $\mathfrak{A}$ unter Benutzung von 2.5.

§ 3. Berechnung gewisser projektiver Limites

3.0. Sei X ein Hausdorffraum und $\mathfrak{G}$ eine Garbe von Gruppen auf X. Es gebe eine kompakte Ausschöpfung $Q_0\subset Q_1\subset Q_2\subset\ldots\subset X$ von X mit $Q_\nu\subset\subset\mathring{Q}_{\nu+1}$ für alle $\nu\in\mathbb{N}$. Jedes $x\in X$ besitze ein Fundamentalsystem $\mathscr{B}_x$ abgeschlossener Umgebungen A mit $H^1(A\cap Q_\nu,\mathfrak{G})=1$ für alle $A\in\mathscr{B}_x$ und alle $\nu\in\mathbb{N}$.

3.1. Satz. *Unter den obigen Voraussetzungen gilt für die kanonische Abbildung*

$$\rho\colon H^1(X,\mathfrak{G})\to\varprojlim H^1(Q_\nu,\mathfrak{G})$$

i) *ρ ist stets surjektiv,*

ii) *erfüllt für jedes $z\in H^1(X,\mathfrak{G})$ das projektive System $(\Gamma(Q_\nu,{}^z\mathfrak{G}))_{\nu\in\mathbb{N}}$ die ML-Bedingung, so ist ρ injektiv.*

Beweis. Man stellt zunächst fest, daß X lokal kompakt und parakompakt ist. Es gibt dann beliebig feine abgeschlossene Überdeckungen $\mathfrak{A}=(A_i)_{i\in I}$ von X mit folgenden Eigenschaften:

$\mathfrak{A}$ besitzt eine offene Verdickung $\mathfrak{W}=(W_i)_{i\in I}$ mit $W_i\cap W_j\neq\varnothing$ nur für endlich viele $j\in I$ zu vorgegebenem $i\in I$

$$H^1(A_i\cap Q_\nu,\mathfrak{G})=1,\quad\text{für alle } i\in I \text{ und für alle } \nu\in\mathbb{N}.$$

Wir definieren induktiv die Überdeckungsfolge

$$\mathfrak{R}_0:=\mathfrak{A}\cap Q_0,\qquad \mathfrak{R}_{\nu+1}:=\mathfrak{R}_\nu\cup(\mathfrak{A}\cap Q_{\nu+1}),$$

$\mathfrak{R}_\infty := \bigcup_{\nu\in\mathbb{N}} \mathfrak{R}_\nu$, $Q_\infty := X$. Insbesondere ist $|\mathfrak{R}_\nu| = Q_\nu$. Für die kanonischen Abbildungen gilt:

$H^1(\mathfrak{R}_\nu, \mathfrak{G}) \to H^1(Q_\nu, \mathfrak{G})$ ist bijektiv für $\nu \in \mathbb{N} \cup \{\infty\}$, (vgl. 2.6)
$H^1(\mathfrak{R}, \mathfrak{G}) \to \varprojlim H^1(\mathfrak{R}_\nu, \mathfrak{G})$ ist surjektiv (vgl. 2.2).

Daraus folgt i.

Seien nun $z, z' \in H^1(X, \mathfrak{G})$ vorgegeben. Wir können annehmen, daß z und z' über einer offenen Überdeckung $\mathfrak{U}$ von X gegeben sind, wobei $\mathfrak{U}$ eine abgeschlossene Verfeinerung $\mathfrak{A}$ mit den obigen Eigenschaften besitzt. Da über U_i ${}^z\mathfrak{G} = \mathfrak{G}$ gilt, folgt $H^1(A_i \cap Q_\nu, {}^z\mathfrak{G}) = 1$ für alle $i \in I$ und alle $\nu \in \mathbb{N}$. Die Sequenz $1 \to H^1(\mathfrak{R}, {}^z\mathfrak{G}) \to \varprojlim H^1(\mathfrak{R}_\nu, {}^z\mathfrak{G})$ ist nach 2.2 exakt, falls das projektive System $(H^0(\mathfrak{R}_\nu, {}^z\mathfrak{G}))_{\nu\in\mathbb{N}}$ die ML-Bedingung erfüllt. Mit Hilfe von Satz 2.5 folgt $H^0(\mathfrak{R}_\nu, {}^z\mathfrak{G}) = \Gamma(Q_\nu, {}^z\mathfrak{G})$ und damit auch ii.

Literatur

1. Andreotti, A., Grauert, H.: Théorèmes de finitude pour la cohomologie des espaces complexes. Bull. Soc. Math. France **90**, 193–259 (1962)
2. Bingener, J.: Offenheit der Versalität in der analytischen Geometrie. Math. Z. **173**, 241–281 (1980)
3. Bröcker, T., Jänich, K.: Einführung in die Differentialtopologie. Heidelberger Taschenbuch **143**. Berlin-Heidelberg-New York: Springer 1973
4. Flenner, H.: Über Deformationen holomorpher Abbildungen. Habilitationsschrift, Osnabrück 1978
5. Forster, O., Knorr, K.: Über die Deformationen von Vektorraumbündeln auf kompakten komplexen Räumen. Math. Ann. **209**, 291–346 (1974)
6. Forster, O., Knorr, K.: Ein neuer Beweis des Satzes von Kodaira-Nirenberg-Spencer. Math. Z. **139**, 257–291 (1974)
7. Forster, O., Knorr, K.: Konstruktion verseller Deformationen kompakter komplexer Räume. Lecture Notes in Mathematics **705**. Berlin-Heidelberg-New York: Springer 1979
8. Giraud, J.: Cohomologie non abélienne. Grundlehren der mathematischen Wissenschaften **179**. Berlin-Heidelberg-New York: Springer 1971
9. Godement, R.: Théorie des faisceaux. Paris: Hermann 1964
10. Hörmander, L.: L^2-estimates and existence theorems for the $\bar{\partial}$-Operator. Acta. Math. **113**, 89–152 (1965)
11. Kaup, L.: Eine Künnethformel für Fréchetgarben. Math. Z. **97**, 158–168 (1967)
12. Leistner, D.: Cohomologie mit Schranken für pseudokonvexe und pseudokonkave komplexe Räume. J. Reine Angew. Math. **311**, 315–329 (1979)
13. Pourcin, G.: Théorème de Douady au-dessus de S. Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mat. Ser. 3 **23**, 451–459 (1969)
14. Ramis, J.P.: Théorèmes de separation et de finitude pour l'homologie et la cohomologie des espaces (p, q) - convexes - concaves. Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mat. Ser. 3 **27**, 933–997 (1973)
15. Schlessinger, M.: Functors of Artin rings. Trans. Amer. Math. Soc. **130**, 208–222 (1968)
16. Schuster, H.W.: Zur Theorie der Deformationen kompakter komplexer Räume. Invent. Math. **9**, 284–294 (1970)

Eingegangen am 16. September 1981

Math. Z. 180, 331–334 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Some Properties of a Regular Sequence in Hilbert Space

Andrés Reyes

Facultad de Ciencias, Cátedra de Geometría Analítica, Zaragoza, Spain

1. Introduction

Let us consider the separable complex Hilbert space H. If $\int=(a_n)_{n\in N}$ is a complete, minimal sequence in H, for $S\subseteq N$ denote

$$W_S=\overline{\text{span}}(a_k)_{k\in S}$$

and

$$W_S^*=\bigcap_{k\notin S}\text{Ker}\, a_k^*,$$

being $(a_n^*)_{n\in N}$ the conjugate sequence of $(a_n)_{n\in N}$ (i.e. $\langle a_j, a_k^*\rangle=\delta_{k,j}$). It is clear that $W_S\subseteq W_S^*$ and $\bigcap_\alpha W_{S_\alpha}^*=W^*_{\bigcap_\alpha S_\alpha}$.

In the sequel a sequence in H will mean a complete minimal sequence.

A sequence is *weak* if its kernel, W_Φ^*, is the null subspace [1]. The following fact is proved in [2]: $\int=(a_n)_{n\in N}$ is a weak sequence if and only if $W_S\cap W_T=\{0\}$, for every disjoint $S, T\subseteq N$.

Observe that for a weak sequence one may have $S, T\subseteq N$ such that

$$W_S\cap W_T\neq W_{S\cap T}\quad [4].$$

1.1. **Definition.** The sequence $\int=(a_n)_{n\in N}$ is *regular* if

$$W_S\cap W_T=W_{S\cap T},$$

for any $S, T\subseteq N$.

Every regular sequence is a weak sequence and every Schauder sequence is regular. Let $(e_n)_{n\in N}$ be an orthonormal basis in H and let

$$b_n=e_1+e_2+\ldots+e_n\quad (n\in N).$$

Then, $(b_n)_{n\in N}$ is regular, but is not Schauder not even for any rearrangement.

We have proved the following

Theorem. *If $\int=(a_n)_{n\in N}$ is a regular sequence in H, then*

$$H=[a_i;\, i\in S]\overset{\perp}{\oplus}[a_k^*;\, k\notin S],$$

for any $S\subseteq N$.

0025-5874/82/0180/0331/$01.00

2. Some Previous Results

In a Hilbert space H, the sets

$$L(H)=\{M;\ M \text{ is a subspace of } H\}$$

and

$$G(H)=\{M;\ M \text{ is a closed subspace of } H\}$$

are complete lattices with respect to the order relation given by inclusion of sets. The lattice $L(H)$ is modular or Dedekind, but $G(H)$ is not modular in general. In fact, $G(H)$ is modular if and only if H has a finite dimension [3]. Denote by $G(\int)=\{W_S;\ S\subseteq N\}$. Then, $G(\int)$ is an upper semi-lattice of $G(H)$ verifying in addition

$$\sup_{G(H)}\{W_{S_i};\ i\in I\}=W_{\bigcup_{i\in I} S_i}.$$

Thus a sequence $\int$ will be *regular if and only if* $G(\int)$ *is a sublattice of* $G(H)$.

2.1. *Claim.* Let $\int=(a_n)_{n\in N}$ be a sequence in H. Then

(i) $W_S^*=W_S$, if S is cofinite,

(ii) $W_S^*=W_S\oplus W_\Phi^*$, if S is finite.

Proof. (i) Let $S^c=\{q_1,q_2,\ldots,q_r\}$ be the complement of S with respect to N. Then

$$W_S^*=\bigcap_{j=1}^{r} W_{N-\{q_j\}}=W_{\bigcap_{j=1}^{r} N-\{q_j\}}=W_S.$$

(ii) If S is finite, it is clear that

$$H=W_S\oplus W_{S^c},$$

so that

$$W_S^*=(W_S+W_{S^c})\cap W_S^*$$

or

$$W_S^*=W_S+(W_{S^c}\cap W_S^*),$$

since $L(H)$ is modular and $W_S\subseteq W_S^*$.

But

$$W_{S^c}\cap W_S^*\underset{(i)}{=}W_{S^c}^*\cap W_S^*=W_\Phi^*$$

Hence

$$W_S^*=W_S+W_\Phi^*.$$

The result follows since $W_S\cap W_{S^c}=\{0\}$ and $W_\Phi^*\subseteq W_{S^c}$. □

2.2. **Corollary.** *If* $\int=(a_n)_{n\in N}$ *is a weak sequence in* H, *then*

$$W_S^*=W_S,$$

for any finite or cofinite $S\subseteq N$. □

2.3. *Claim.* The sequence $\int=(a_n)_{n\in N}$ is regular if and only if

$$W_S^*=W_S,$$

for any $S\subseteq N$.

Proof. The condition is easily seen to be sufficient. In order to prove the necessity, let us suppose on the contrary that there exists a subset U of N such that $W_U^* \neq W_U$. By the former corollary, U must be infinite and with infinite complement. Let $U=\{p_n; n\in N\}$, $U^c=\{q_n; n\in N\}$ and let $x\in W_U^* - W_U$. There must exist a finite R_1 subset of N such that

$$\|x - \sum_{k\in R_1} \lambda_k^{(1)} a_k\| < 1.$$

Suppose $R_1=\{p_{r_1},\ldots,p_{r_{l_1}}; q_{s_1},\ldots,q_{s_{m_1}}\}$. Since $x\in W_U^*$, it follows $x\in W_{N-\{q_{s_1},\ldots,q_{s_{m_1}}\}}$. Therefore we can find a finite R_2 in $N-\{q_{s_1},\ldots,q_{s_{m_1}}\}$ such that

$$\|x - \sum_{k\in R_2} \lambda_k^{(2)} a_k\| < 1/2.$$

Let $R_2=\{p_{r_2},\ldots,p_{r_{l_2}}; q_{s_2},\ldots,q_{s_{m_2}}\}$. Since $x\in W_U^*$, $x\in W_{N-\{q_{s_1},\cdots,q_{s_{m_1}},q_{s_2},\cdots,q_{s_{m_2}}\}}$ and thus, there exists a finite R_3 in $N-\{q_{s_1},\ldots,q_{s_{m_1}},q_{s_2},\ldots,q_{s_{m_2}}\}$ such that

$$\|x - \sum_{k\in R_3} \lambda_k^{(3)} a_k\| < 1/3.$$

By iteration, we obtain a sequence of finite subsets of N, $\{R_n; n\in N\}$ such that

$$x=\lim_n \sum_{k\in R_n} \lambda_k^{(n)} a_k.$$

Take $S=\bigcup_{n\in N} R_{2n-1}$ and $T=\bigcup_{n\in N} R_{2n}$. It follows

$$x\in W_S\cap W_T$$

and

$$x\notin W_{S\cap T},$$

since $S\cap T\subseteq U$. □

2.4. **Corollary.** *For a given sequence* $\int$, *the following statements are equivalent:*

(i) $\int$ *is regular,*

(ii) $G(\int)$ *is a sublattice of* $G(H)$,

(iii) $G(\int)$ *is a complete sublattice of* $G(H)$. □

3. Proof of the Theorem

For any sequence $\int=(a_n)_{n\in N}$ in H, it is easy to prove

$$(W_S^*)^\perp=[a_k^*; k\notin S].$$

By 2.3 and taking into account that a Hilbert space can be represented as a direct sum of a closed subspace and its orthogonal complement, we obtain the Theorem. □

Bibliography

1. Dixmier, J.: Étude sur les variétés et les opérateurs de Julia avec quelques applications. Bull Soc. Math. France **77**, 11-101 (1949)
2. Plans, A.: A generalization of heterogonal systems. Arch. Math. (Basel) **26**, 398-401 (1975)
3. Pasenkov, V.V.: The lattice of subspaces of Hilbert space. Soviet. Math. Dokl. **2**, 775-778 (1961)
4. Reyes, A.: Aspectos reticulares y geométricos de sistemas de vectores en espacios de Banach y de Hilbert. Problema de la intersección. Tesis doctoral, Universidad de Zaragoza 1980

Received August 25, 1981; received in final form October 5, 1981

Math. Z. 180, 335-348 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Primes in Short Intervals

Glyn Harman*

Department of Mathematics, Royal Holloway College, Egham, Surrey TW20 OEX, United Kingdom

1. Introduction

It was conjectured by Cramér [1] that every interval of the form $[n, n+f(n)\log^2 n]$ contains a prime for some $f(n)\to 1$ as $n\to\infty$. Assuming the Riemann Hypothesis, Selberg [12] has shown that almost all intervals of the above form contain a prime providing $f(n)\to\infty$ with n. "Almost all" in this context indicates that the number of $n\leqq X$ for which the statement is false is $o(X)$. Selberg's proof essentially gave a relationship between the density of zeros of $\zeta(s)$ and the length of the interval. This was used by Montgomery (Chap. 14 of [11]) to show that, for almost all n, $[n, n^{1/5+\varepsilon}]$ contains a prime. The exponent 1/5 may be improved to 1/6 using the zero density estimate of Huxley [6] which he obtained to show that $p_{n+1}-p_n \ll p_n^{7/12+\varepsilon}$ where p_n is the nth prime. This result on the difference between consecutive primes has been improved by Iwaniec and Jutila [8] to $p_n^{5/9+\varepsilon}$, and by Heath-Brown and Iwaniec [5] to $p_n^{11/20+\varepsilon}$. These last two results were obtained by a sieve method. We shall use similar arguments to prove the following result:

Theorem. *For almost all n, the interval*

$$[n, n+n^{(1/10)+\varepsilon}) \tag{1}$$

contains a prime number.

It is the hypothesis of Lemma 5 below which sets $(1/10)+\varepsilon$ as the limit of the present method. We shall in fact show that for almost all n the interval (1) contains $\gg n^{(1/10)+\varepsilon}(\log n)^{-1}$ primes. This exhibits a common feature of sieve results: we obtain a lower bound which is a fraction of the "expected" number of primes under consideration.

2. Outline of Method

In Sect. 2-4 we use the following standard notation:

$$P(z)=\prod_{p<z} p, \qquad V(z)=\prod_{p<z}(1-1/p),$$

* Written while the author held a London University Postgraduate Studentship

0025-5874/82/0180/0335/$02.80

and note the well known asymptotic formula:

$$V(z)=\frac{e^{-\gamma}}{\log z}+O((\log z)^{-2}), \tag{2}$$

where γ is Euler's constant. For a finite set of integers $\mathscr{A}$ write

$$\mathscr{A}_d=\{n\in\mathscr{A}; d\mid n\},$$
$$S(\mathscr{A}, z)=|\{n\in\mathscr{A}, (n, P(z))=1\}|.$$

We shall consider the set

$$\mathscr{A}=\{n; x-y<n\leqq x\}.$$

Here x is a real number satisfying $X<x\leqq 2X$, where X will be assumed "large", and $y=xX^{-(9/10)-\varepsilon}/2$. The fundamental Buchstab identity states that

$$S(\mathscr{A}, z_1)=S(\mathscr{A}, z_2)-\sum_{z_1\leqq p<z_2} S(\mathscr{A}_p, p).$$

Using this we find that, for $z_1<z_2<x^{\frac{1}{2}}$, we have

$$\begin{aligned}\pi(x)-\pi(x-y)&=S(\mathscr{A}, x^{\frac{1}{2}})\\&=S(\mathscr{A}, z_1)-\sum_{z_1\leqq p<x^{\frac{1}{2}}} S(\mathscr{A}_p, p)\\&=S(\mathscr{A}, z_1)-\sum_{z_2\leqq p<x^{\frac{1}{2}}} S(\mathscr{A}_p, z_3(p))-\sum_{z_1\leqq p<z_2} S(\mathscr{A}_p, z_4(p))\\&\quad+\sum_{\substack{z_3(p)\leqq q<p<x^{\frac{1}{2}}\\ p\geqq z_2}} S(\mathscr{A}_{pq}, q)+\sum_{\substack{z_4(p)\leqq q<p<z_2\\ p\geqq z_1}} S(\mathscr{A}_{pq}, q)\\&=\Sigma_1-\Sigma_2-\Sigma_3+\Sigma_4+\Sigma_5 \text{ say.}\end{aligned}$$

In the above it is also necessary to have $z_3(p)\leqq p$ for $p>z_2$, $z_4(p)\leqq p$ for $z_1\leqq p<z_2$. Since we will only give a lower bound for $\pi(x)-\pi(x-y)$ we shall consider Σ_2 with $x^{\frac{1}{2}}$ replaced with $(2X)^{\frac{1}{2}}$, and Σ_4 with $x^{\frac{1}{2}}$ replaced by $X^{\frac{1}{2}}$. Then

$$\pi(x)-\pi(x-y)\geqq\Sigma_1-\Sigma_2-\Sigma_3+\Sigma_4+\Sigma_5. \tag{3}$$

We shall give a lower bound for Σ_1 and an upper bound for Σ_2 and Σ_3 by means of the linear sieve (Lemma 1). It should be noted that one can make use of certain subsums over almost-primes which arise in the sieve results in [7] to improve the lower bound for the left hand side of (3). The inclusion of these sums has significance in other problems, but is unnecessary here. We deal with the remainder terms in Lemma 1 by bounding the integral of their square over $[X, 2X]$, using Dirichlet polynomials (see Lemmas 2–6). The motivating principle here is to be able to choose the D which occurs in Lemma 1 as large as possible. To this end we shall pick z_1, z_2, $z_3(p)$, $z_4(p)$ so that the hypotheses for our estimates of mean values of Dirichlet polynomials match the form of the remainder term given by Iwaniec. Eventually we obtain a lower bound of the form $C'y/\log X$ for $\Sigma_1-\Sigma_2-\Sigma_3$, valid for all x except on a set of measure

$o(X)$. Here C' is negative, but quite small. On the other hand, we are able to give an asymptotic formula for a subsum of Σ_4, with an error which is also considered by its integral over $[X, 2X]$, again using the mean-value estimates for Dirichlet polynomials given in Sect. 4. The remainder of the sum, together with the whole of Σ_5, being non-negative, is discarded. Although the proportion of Σ_4 for which the asymptotic formula can be shown to hold is small, and would diminish further if y were decreased, we do obtain a lower bound of $C'' y/\log X$, where $C''+C'>0$.

In proving the theorem we may, of course, assume that the ε in (1) is "sufficiently small". We use ϕ for a positive function of ε, and henceforth reserve the letter C for an absolute constant. We write λ for $(\log X)^B$ where B is bounded by a function of ε. The entities ϕ, C, λ need not be the same at each occurrence. We may thus write, for example,

$$(X^{\phi})^{\varepsilon} \gg X^{\phi}, \quad \text{or} \quad \lambda^3 \ll \lambda.$$

The constants implied by Vinogradov's $\ll$ notation depend here, and elsewhere in this paper, on at most ε. We write

$$T=X^{9(1-\varepsilon)/10}, \quad Y=X^{(1/10)+\varepsilon}, \quad z_1=X^{26(1-3\varepsilon)/105},$$
$$z_2=X^{9/35}, \quad z_3(p)=(X^{1-3\varepsilon}/p)^{1/3}, \quad z_4(p)=(X^{(26/35)-2\varepsilon}/p)^{\frac{1}{2}}.$$

The reason for this particular choice of z_1, z_2, z_3, z_4 will become apparent in Sect. 5 when the remainder term of the sieve is estimated.

3. The Fundamental Sieve Result

We use the linear sieve result of Iwaniec [7] in the form stated in [5]. When we come to apply Lemma 1, the $\mathscr{A}$ occurring in its statement will not always be that specified in Sect. 2. For the properties of the standard functions $f(s)$ and $F(s)$ see [2] (Chap. 8). We write

$$r(\mathscr{A}, d)=|\mathscr{A}_d|-W/d$$

for the remainder term, where W is independent of d.

Lemma 1. *Let $z \geqq 2$, $D \geqq z^2$ and $\varepsilon>0$. Then*

$$S(\mathscr{A}, z) \leqq WV(z)\{F(s)+E\}+R^+, \tag{5}$$

$$S(\mathscr{A}, z) \geqq WV(z)\{f(s)-E\}-R^- \tag{6}$$

where $s=(\log D)/(\log z)$ and $E=C\varepsilon+O((\log D)^{-1/3})$. The remainder terms $R^{\pm}$ are of the form

$$R^{\pm}=\sum_{(D)} R^{\pm}_{(D)}=\sum_{(D)} \sum_{v<D^{\varepsilon}} C^{\pm}_{(D)}(v, \varepsilon) \sum_{D_i \leqq p_i \leqq D_i^{1+\varepsilon^7}}{}' r(\mathscr{A}, v p_1 \dots p_r) \tag{7}$$

where (D) runs over all subsequences $D_1 \geqq D_2 \dots \geqq D_r$, including the empty subsequence, of the sequence

$$D^{\varepsilon^2(1+\varepsilon^7)^n}, \quad n \geqq 0,$$

for which

$$D_1 D_2 \dots D_{2k} D_{2k+1}^3 \leqq D \qquad (0 \leqq k \leqq (r-1)/2)$$

in the case of R^+, *and*

$$D_1 D_2 \dots D_{2k-1} D_{2k}^3 \leqq D \qquad (0 \leqq k \leqq r/2)$$

in the case of R^-. *Moreover,* $\sum'$ *indicates that* ν *and* p_i, $(1 \leqq i \leqq r)$, *are restricted by the conditions*

$$\nu \mid P(D^{\varepsilon^2}), \qquad p_i \mid P(z).$$

Finally, the coefficients $C_{(D)}^{\pm}(\nu, \varepsilon)$ *depend at most on* (D), ν, ε *and the* $\pm$ *signs and satisfy*

$$|C_{(D)}^{\pm}(\nu, \varepsilon)| \leqq 1.$$

4. Some Preliminary Results

In this section we give the relation between the remainder terms and integrals of Dirichlet polynomials, and various estimates for such integrals. For any upper case latin letter B other than Q, P, and H we write

$$B(s) = \sum_{B < b \leqq 2B} \alpha_b b^{-s}$$

where α_n is real and $\sum_{B < b \leqq 2B} \alpha_b^k \ll B\lambda$, for any integer $k \leqq C_1(\varepsilon)$.

For the letters L, K we stipulate other conditions. For both K and L, $\alpha_n = 1$ for some set of consecutive integers and $\alpha_n = 0$ otherwise. For L the condition $L < l \leqq 2L$ is to be replaced by $C_2(\varepsilon) \leqq l/L \leqq C_3(\varepsilon)$.

For $1 \leqq H \leqq D^{\varepsilon}$ we define $H^{\pm}(s)$ by

$$H^{\pm}(s) = \sum_{H \leqq \nu < 2H} C_{(D)}^{\pm}(\nu, \varepsilon)\, \nu^{-s}.$$

For numbers $P_j (j = 1, \dots, r)$ with $D_j \leqq P_j \leqq D_j^{1+\varepsilon^7}$ we write

$$P_j(s) = \sum_{P_j \leqq p_j < 2P_j} P_j^{-s} \alpha_{p_j}^{(j)}$$

where $\alpha_n^{(j)} = 1$ for some set of consecutive integers n, and is zero otherwise.

We write $R_{(D)}^{\pm}(H, P_1, \dots, P_r)$ for that subsum of $R_{(D)}^{\pm}$ corresponding to $P_j \leqq p_j < 2P_j$. To consider $R_{(D)}^{+}$ for Σ_2 we put

$$P(s) = \sum_{\substack{P \leqq n < 2P \\ n < (2X)^{\frac{1}{2}}}} n^{-s} \Lambda(n)/(\log P)$$

for any number P with $z_2 \leqq P < (2X)^{\frac{1}{2}}$. We replace $\mathscr{A}$ by $\mathscr{A}_n$ in (5) and write $R_{(D)}^{+}(P, H, P_1, \dots, P_r)$ for the sum of the remainder terms $R_{(D)}^{+}(H, P_1, \dots, P_r)$ weighted with the factor $\Lambda(n)/(\log P)$ for $P \leqq n < \min(2P, (2X)^{\frac{1}{2}})$. We use the

same notation for the remainder terms from Σ_3, replacing $(2X)^{\frac{1}{2}}$ by z_2 and z_2 by z_1.

Also we write

$$H(s)=\min\left(\frac{X^{9/5}}{|s|^2+1},1\right),$$

$$c=1+(\log X)^{-1}.$$

Lemma 2. *For any continuous function $g(s)$, and $0<T_0<T<X$, we have*

$$\frac{1}{X}\int_X^{2X}\left|\int_{c+iT_0}^{c+iT} g(s)\,x^s\,ds\right|^2 dx \ll X^2(\log X)\int_{c+iT_0}^{c+iT}|g(s)|^2\,|ds|.$$

Proof. This is easily established by squaring out the inner integral on the left, integrating with respect to X first and making use of the inequality:

$$|g(c+it_1)\,g(c+it_2)|\leqq|g(c+it_1)|^2+|g(c+it_2)|^2.$$

Lemma 3. *Suppose $YX^{-\varepsilon}\leqq A\leqq X^{1-\varepsilon}$, then*

$$\sum_{A<n\leqq 2A} a_n\left(\left[\frac{x}{n}\right]-\left[\frac{x-y}{n}\right]\right)=y\sum_{A<n\leqq 2A}\frac{a_n}{n}+O(yX^{-\phi})+I$$

where

$$I=\frac{1}{2\pi i}\left(\int_{c-iT}^{c-iT_0}+\int_{c+iT_0}^{c+iT}\right)L(s)\,A(s)\,\frac{x^s-(x-y)^s}{s}\,ds.$$

Here $L=X/A$, $T_0=L^{\frac{1}{2}}$, $C_2(\varepsilon)=\frac{1}{3}$, $C_3(\varepsilon)=3$.

Proof. This is demonstrated in all essentials in [5], pp. 53–54.

Corollary (1). *Write $L=X(HP_1\dots P_r)^{-1}$. Let $\mathscr{A}$ be as in Sect. 2 and let $W=y$. Then we have, for $D=X^{1-3\varepsilon}$,*

$$\frac{1}{X}\int_X^{2X}|R^-_{(D)}(H,P_1,\dots,P_r)|^2\,dx$$

$$\ll Y^2\lambda\int_{c+iT_0}^{c+iT} H(s)\,|H^-(s)\,L(s)\,P_1(s)\dots P_r(s)|^2\,|ds|+Y^2X^{-\phi}. \tag{8}$$

For $T_0=L^{\frac{1}{2}}$.

Corollary (2). *Write $L=X(HPP_1\dots P_r)^{-1}$. Replace $\mathscr{A}$ by $\mathscr{A}_n$, put $D=X^{1-3\varepsilon}/P$, $z=z_3(2P)$ for Σ_2, $z=z_4(2P)$ for Σ_3, and let $W=y/n$. Then we have*

$$\frac{1}{X}\int_X^{2X}|R^+_{(D)}(P,H,P_1,\dots P_r)|^2\,dx$$

$$\ll Y^2\lambda\int_{c+iT_0}^{c+iT} H(s)\,|H^+(s)\,L(s)\,P(s)\,P_1(s)\dots P_r(s)|^2\,|ds|+Y^2X^{-\phi}. \tag{9}$$

The above corollaries follow by combining Lemmas 2 and 3 and using the definition of the remainder terms.

We write $T_1 = X^{9/10-\varepsilon}$ and note that the presence of the $H(s)$ factors in (8) and (9) indicate that it suffices to consider integrals from $c+iT_2$ to $c+iT_3$ where $T_3 - T_2 \leqq T_1$ and $T_2 \geqq T_0$. We write $\mathscr{J}$ for such a line of integration.

We note that the classical mean value theorem for Dirichlet Polynomials (Theorem 6.1 of [11]) gives, for a positive integer h,

$$\int_{\mathscr{J}} |M(s)|^{2h} |ds| \ll \lambda(1 + T/M^h). \tag{10}$$

This result shall be used repeatedly in the following. The next lemma gives a bound for integrals of the above type when h is not an integer, and Lemma 5 continues in a similar vein.

Lemma 4. *Let h be an integer $\geqq 1$ and β a real number with $6h > \beta > 2h$. Then*

$$\int_{\mathscr{J}} |M(s)|^{\beta} |ds| \ll \lambda \left(\mu + \left(\frac{T_1}{M^{2h}\mu} \right)^{(\beta-2h)/(6h-\beta)} \left(1 + \frac{T_1}{M^h} \right) \right) \tag{11}$$

where μ is any number $\geqq \max_{s \in \mathscr{J}} |M(s)|^{(\beta-2h)}$.

Proof. Put $Q(s) = M(s)^h$, $Q = M^h$. Let s_r be any points in $\mathscr{J}$ with $|s_j - s_k| \geqq 1$ for $j \neq k$, and $V < |Q(s_r)| \leqq 2V$ $(r = 1, \ldots, R)$. By the form of the Halász lemma due to Huxley [6] we have

$$R \ll \lambda(V^{-2} + T_1 Q^{-2} V^{-6}).$$

Thus

$$\begin{aligned} RV^{2+(\beta-2h)/h} &\ll \lambda(V^{(\beta-2h)/h} + T_1 Q^{-2} V^{-4+(\beta-2h)/h}) \\ &\ll \lambda\mu \end{aligned}$$

providing

$$V > \left(\frac{T_1}{M^{2h}\mu} \right)^{h/(6h-\beta)}. \tag{12}$$

If (12) holds it follows that the integral over that part of $\mathscr{J}$ for which $|Q(s)| > V$ is $\ll \lambda\mu$. For the remainder of the integral it follows from (10) that we get the bound

$$\ll \left(\frac{T_1}{M^{2h}\mu} \right)^{(\beta-2h)/(6h-\beta)} \left(1 + \frac{T_1}{M^h} \right).$$

Which completes the proof of (11). It should be noted that this result is considerably weaker than what might be expected (cf. Conjecture 9.2 of [11]).

By Van der Corput's bounds for exponential sums (see Chap. 5 of [13] for example), for $X \geqq K \geqq X^{\varepsilon}$, we have

$$\max_{s \in \mathscr{J}} |K(s)| < X^{-\phi}. \tag{13}$$

This, together with (11) for $h = 3$ gives the following:

Corollary. *If $K > T_1^{2/7} X^{\varepsilon/10}$ then*

$$\int_{\mathscr{J}} |K(s)|^8 |ds| \ll X^{-\phi}.$$

Lemma 5. *Suppose* $KMN \geqq T_1^{10/9} X^{\varepsilon/10}$, $K \geqq X^{\varepsilon}$, $N \gg X^{\varepsilon^2/2}$, *and* $\max(KM, MN) \geqq T_1^{6/7} X^{\varepsilon}$. *Then*

$$\int_{\mathscr{J}} |K(s) M(s) N(s)|^2 |ds| \ll X^{-\phi}.$$

Proof. We assume $KM \geqq MN$, the other case follows similarly. We write $J(\alpha)$ for the largest even integer $\leqq \alpha$, and define β to be the smallest number not less than 8 for which

$$N^{\beta + J(\beta)} \geqq T_1^4. \tag{14}$$

We put $\tau = 2\beta(\beta-2)^{-1}$, then, by Hölder's inequality

$$\int_{\mathscr{J}} |K(s) M(s) N(s)|^2 |ds| \leqq \left(\int_{\mathscr{J}} |N(s)|^{\beta} |ds|\right)^{2/\beta} \left(\int_{\mathscr{J}} |K(s) M(s)|^{\tau}\right)^{2/\tau}. \tag{15}$$

From (14) and (11) with $h = J(\beta)/2$, together with $|N(s)| \ll \lambda$, it follows that

$$\int_{\mathscr{J}} |N(s)|^{\beta} |ds| \ll \lambda. \tag{16}$$

We now apply Lemma 3 to $K(s) M(s)$ with τ in place of β and $h = 1$. Using (13) and $|M(s)| \ll \lambda$ we obtain the inequality

$$\int_{\mathscr{J}} |K(s) M(s)|^{\tau} |ds| \ll X^{-\phi}, \tag{17}$$

if

$$(MK)^{2+\tau} > T_1^4 X^{\varepsilon}. \tag{18}$$

Now, for $\beta = 8$, (18) follows from the hypothesis $KM \geqq T_1^{6/7} X^{\varepsilon}$. Otherwise we use the definition of β to obtain

$$(MK)^{2+\tau} > X^{\varepsilon} T_1^{\theta(\beta)}$$

where

$$\theta(\beta) = 4\left(\frac{\beta-1}{\beta-2}\right)\left(\frac{10}{9} - \frac{4}{8+\beta}\right) \quad \text{for } 8 < \beta < 10$$

$$= \frac{8}{9}\left(5 - \frac{4}{\beta-2}\right) \quad \text{for } \beta \geqq 10.$$

It may easily be verified that $\theta(\beta) \geqq 4$ for all $\beta > 8$. It follows that (18) holds, and so the result of the lemma follows from (15), (16), (17).

We require one more lemma on mean values of Dirichlet polynomials.

Lemma 6. *For* $K > T^{1/2} X^{\varepsilon}$, $KM > X^{1-\varepsilon}$, $T_0 \gg X^{\varepsilon/10}$, *we have*

$$\int_{\mathscr{J}} |M(s) K(s)|^2 |ds| \ll X^{-\phi}. \tag{19}$$

Proof. By the well known Perron formula (Lemma 3.12 of [13]), we have

$$K(s) = \frac{1}{2\pi i} \int_{c-iT-\sigma}^{c+iT-\sigma} \zeta(s+w) \frac{K^w(2^w-1)}{w} dw + O\left(K^{-1} + \frac{\log X}{T}\right)$$

$$= \frac{1}{2\pi i} \int_{\frac{1}{2}-iT-\sigma}^{\frac{1}{2}+iT-\sigma} \zeta(s+w) \frac{K^w(2^w-1)}{w} dw + O(K^{-1} + T^{-5/6} + (|s|+1)^{-1}), \tag{20}$$

after shifting the integral to the $\frac{1}{2}-\sigma$ line and using $|\zeta(\sigma+it)| \ll t^{1/6}$ for $\sigma \geqq \frac{1}{2}$. We write $\mathscr{J}(s)$ for the integral in (20).

We have

$$\int_{\mathscr{J}} |M(s)|^2 (K^{-1}+(1+|s|)^{-1}+T^{-5/6})^2 \, |ds| \ll X^{-\phi}$$

by an easy calculation using (10). Also, using the fourth power moment of $\zeta(s)$ (see Chap. 7 of [13], for example) we have

$$\int_{\mathscr{J}} |\mathscr{J}(s) M(s)|^2 \, |ds| \ll \lambda \left(\frac{T}{K^2}\right)^{\frac{1}{2}} \left(1+\frac{T}{M^2}\right)^{\frac{1}{2}} \ll X^{-\phi}.$$

This completes the proof of (19).

5. Estimation of Σ_1, Σ_2, Σ_3

(I) *The main terms*

We first note the forms which the functions $f(u)$ and $F(u)$ take for the present choice of parameters. We have, for $4 \leqq u \leqq 6$

$$f(u)=\frac{2e^{\gamma}}{u}\left(\log(u-1)+\int_2^{u-2} \frac{\log(t-1)}{t} \log\left(\frac{u-1}{t+1}\right) dt\right).$$

In the case of Σ_1 $u=105/26$, so

$$f(u)>\frac{2e^{\gamma}}{105/26} \log\left(\frac{79}{26}\right).$$

Also, for $3 \leqq u \leqq 5$,

$$F(u)=\frac{2e^{\gamma}}{u}\left(1+\int_2^{u-1} \frac{\log(t-1)}{t} dt\right).$$

In the case of Σ_2 $u<3$, and here $F(u)=2e^{\gamma}/u$. For $\Sigma_3 u$ varies, but it is never more than 52/17. Thus in this case we have

$$\begin{aligned} F(u) &\leqq \frac{2e^{\gamma}}{u}\left(1+\int_2^{35/17} \frac{\log(t-1)}{t} dt\right) \\ &\leqq \frac{2e^{\gamma}}{u}\left(1+\int_2^{35/17} (1-2/t)\, dt\right) = 2e^{\gamma} u^{-1}(1+(1/17-2\log(35/34)). \end{aligned}$$

In the following we write $\mathscr{E}^{\pm} = \pm C\varepsilon + O((\log X)^{-1/3})$, keeping to our convention that the Cs need not be the same at each occurrence. By (6) the main term from Σ_1 is thus

$$\geqq \frac{2y}{\log X} \log(79/26)(1+\mathscr{E}^{-}).$$

To estimate the main terms for Σ_2 and Σ_3 we use the following trivial inequality in order to simplify the estimation of the remainder terms:

$$\sum_{P\leqq p<2P} S(\mathscr{A}_p, z_j(p)) \leqq \frac{1}{\log P} \sum_{P\leqq n<2P} \Lambda(n)\, S(\mathscr{A}_n, z_j(P)) \qquad (j=3,4).$$

Thus the main term from Σ_2 using (2), (5) and the formula for $F(u)$ is

$$\begin{aligned} &\leqq \sum_{z_2\leqq n<(2X)^{\frac{1}{2}}} \frac{2y\Lambda(n)(1+\mathscr{E}^+)}{(\log n)\, n(\log(X/n))} \\ &= 2\int_{9/35}^{1/2} \frac{dt}{t(1-t)}(1+\mathscr{E}^+)\frac{y}{\log X} \\ &= \frac{y}{\log X}\, 2\log(26/9)(1+\mathscr{E}^+). \end{aligned}$$

Similarly, the main term from Σ_3 is

$$\leqq \frac{y}{\log X}\, 2\log\left(\frac{79\times 27}{78\times 26}\right)(1+(1/17-2\log(35/34))(1+\mathscr{E}^+).$$

Hence the main term from $\Sigma_1-\Sigma_2-\Sigma_3$ is

$$\begin{aligned} &\geqq \frac{-2y}{\log X}\log\left(\frac{79\times 27}{78\times 26}\right)(1/17-2\log(35/34))(1+\mathscr{E}^+) \\ &> -10^{-4}\, y/(\log X), \end{aligned}$$

assuming ε is sufficiently small and X sufficiently large.

(II) *The remainder terms*

We wish to show that

$$\frac{1}{X}\int_X^{2X} R_{(D)}^{\pm 2}\, dx \ll Y^2 X^{-\phi}.$$

After a division of $R_{(D)}^{\pm}$ into $\ll\lambda$ subsums (note that the number of sequences (D) is $\ll 1$), this reduces to obtaining the same estimate for

$$\frac{1}{X}\int_X^{2X} |R_{(D)}^{-}(H, P_1, \ldots, P_r)^2\, dx \quad \text{and} \quad \frac{1}{X}\int_X^{2X} |R_{(D)}^{+}(P, H, P_1, \ldots, P_r)|^2\, dx.$$

We now appeal to Lemma 3, Corollaries (1) and (2), to reduce our task to obtaining the upper bound $X^{-\phi}$ for the integrals which occur on the right in (8) and (9).

We first consider Σ_1. If (D) is the empty sequence or has only 1 or 2 members then Lemma 6 may be applied since $P_1 P_2 \leqq (X^{26/105})^{2+3\varepsilon} < X^{\frac{1}{2}}$. This lemma gives the desired result whenever $HP_1 \ldots P_r < X^{11/20}$, so we henceforth

suppose this condition is violated. If $r \geqq 4$ then $P_r^6 < X$, so Lemma 5 may be applied $\ll 1$ times with $K(s)$ as a subsum of $L(s)$, $M(s)$ as a subsum of $H^-(s)P_1(s)\dots P_{r-1}(s)$ and $N(s)$ as $P_r(s)$. For a sequence consisting of three members Lemma 5 may be applied if $P_3 < X^{8/35}$ or $H > X^{\varepsilon^2}$. Otherwise we note that P_1, P_2, P_3 are all greater than $T_1^{1/4}$, while $L > X^{1-2\varepsilon^2}/z_1^3 > T_1^{2/7} X^{\varepsilon/10}$ (z_1 was chosen so that this last condition would hold). An appeal to Holder's inequality, Lemma 4 Corollary and (10) completes the proof in this case.

For Σ_2 the empty subsequence may be dealt with simply by appealing to Lemma 6 since $P \ll X^{\frac{1}{2}}$. If a sequence has one member then at least one of $PD_1 < X^{11/20}$, $D_1 < X^{8/35}$ holds, for otherwise we would have $PD_1^3 > X$, which is impossible. For a sequence with r members where $r \geqq 3$ we note that $D_r^6 < X$ so Lemma 5 is applicable. The remaining case of $r=2$ is a little more troublesome. We may assume

$$P_2 > X^{8/35}, \qquad X/(PP_1P_2) < X^{9/35},$$

for otherwise the proof may be completed as for Σ_1 with $r=3$. We note that we must have $P < X^{16/35}$ since $P_2 > X^{8/35}$. In this final case we must split $Q(s)$ by Vaughan's identity (see [14]) into $\ll \lambda$ double sums of the form

$$\sum_{M < m \leqq 2M} a_m \sum_{P \leqq mn < Q} b_n (mn)^{-s}, \tag{23}$$

where either:

1) $PX^{-\varepsilon^2} \geqq M \geqq X^{\varepsilon^2}$, or
2) $M < X^{\varepsilon^2}$ and $b_n \equiv 1$ or $\equiv \log n$.

In both cases

$$\sum_{M < m \leqq 2M} a_m^k \ll \lambda M, \qquad \sum_{N < n \leqq N} b_n^k \ll \lambda N$$

and $a_m, b_n \ll X^{\eta}$ for any $\eta > 0$.

We now write

$$Q_1(s) = \sum_{M < m \leqq 2M} a_m m^{-s}, \qquad Q_2(s) = \sum_{P/4M < n < 2P/M} b_n n^{-s}. \tag{24}$$

Then the sum (23) may be expressed in the form

$$\frac{1}{2\pi i} \int_{-iT}^{iT} \left(\frac{Q^w - P^w}{w} \right) Q_1(s+w) Q_2(s+w)\, dw + (X^{\varepsilon} T^{-1}), \tag{25}$$

using the Perron formula and assuming, without loss of generality, that $\|P\| = \|Q\| = \frac{1}{4}$.

Now, if we denote by $Q_3(s)$ a double sum of the form (23), we have, by (25)

$$\int_{\mathcal{J}} |H^+(s) P_1(s) Q_3(s) P_2(s) L(s)|^2 \, |ds|$$

$$\ll \lambda \int_{-T}^{T} \frac{dt}{1+|t|} \int_{\mathcal{J}} |H^+(s) L(s) Q_1(s+it) Q_2(s+it) P_1(s) P_2(s)|^2 \, |ds| + O(X^{-\phi}). \tag{26}$$

It is easily demonstrated, using (13), that, for any value of t, the portion of the inner integral on the right of (26) corresponding to $|s+ti| \leqq T^{\frac{1}{4}}$ is $\ll X^{-\phi}$. It thus suffices to prove that

$$\int_{s \in \mathscr{S},\ |s+it| > T^{\frac{1}{4}}} |H(s)\, Q_1(s+it)\, Q_2(s+it)\, P(s)\, P_2(s)|^2\, |ds| \ll X^{-\phi} \tag{27}$$

for any t with $|t| \leqq T$. If $Q_3(s)$ is a type 1) sum then

$$X^{\varepsilon^2} \leqq \min(M, P/M) \leqq P^{\frac{1}{4}} \leqq X^{8/35}.$$

In this case Lemma 5 may be applied. For a sum of type 2) we note that $P/M \geqq X^{9/35-\varepsilon^2} > T_1^{2/7} X^{\varepsilon/10}$. The proof may then be completed as in other cases by Hölder's inequality and Lemma 4 Corollary (with slight modifications to allow for $K(s)$ having the different form $Q_2(s+it)$). The motivation in the choice for z_2 is now clear.

We shall be brief in our discussion of Σ_3. The only sequence to give any trouble is, as with Σ_2, $r=2$, $H > X^{\varepsilon^2}$, $P_2 > X^{8/35}$. In this case, however, $X/(PP_1P_2) > X^{1-2\varepsilon^2}/(P(z_4(P))^2) \geqq X^{9/35-2\varepsilon^2} > T^{2/7} X^{\varepsilon/10}$. The proof may then be completed as for Σ_1 and Σ_2. The reason for the choice of $z_4(p)$ is now obvious.

6. Estimating P_3s in the Interval

We write

$$\psi(a) = \sum_{n \leqq a} \Lambda(n).$$

The following lemma gives the asymptotic formula for estimating subsums of Σ_4. The author first proved the result of this paper using the method of [3] in order to establish a lemma of the following type. Lemma 7 here is stronger and the proof is simpler, though for very short interval lengths the zero density method of [3] appears to be stronger.

Lemma 7. *Suppose* $X^{1-\varepsilon} > PR \geqq X^{27/35}$. *Write*

$$S = \sum_{P \leqq n < 2P} \sum_{R \leqq r < 2R} \Lambda(n)\, \Lambda(r) \left(\psi\left(\frac{x}{nr}\right) - \psi\left(\frac{x-y}{nr}\right) \right).$$

Then we have

$$S = \sum_{P \leqq n < 2P} \sum_{R \leqq r < 2R} \left(\frac{\Lambda(n)\, \Lambda(r)}{nr} \right) y + E_1 + E_2 \tag{28}$$

where

$$\frac{1}{X} \int_X^{2X} |E_j|^j\, dx \ll Y^j (\log X)^{-10} \qquad (j = 1, 2).$$

Proof. Our starting point will be the familiar formula, which follows from Lemma 3.12 of [13]:

$$\sum_{A < n \leqq 9A} \Lambda(n)\, n^{-s} = -\frac{1}{2\pi i} \int_{c-\sigma-iT}^{c-\sigma+iT} \frac{\zeta'}{\zeta}(s+w)\, \frac{A^w (9^w - 1)\, dw}{w} + O\left(\frac{A^{1-\sigma} (\log A)^2}{T} + \frac{\log A}{A^\sigma} \right). \tag{29}$$

Now we suppose $A \gg T^{\varepsilon}$, then, using the zero free region of $\zeta(s)$ (see Corollary 11.4 of [11], for example) we may shift the integral in (29) to

$$\operatorname{Re} w = 1 - (\log A)^{-\frac{3}{4}} - \sigma$$

and only encounter the pole at $s+w=1$, while on the new line of integration

$$\frac{\zeta'}{\zeta}(s) \ll (\log A)^3.$$

The "horizontal" parts of the contour are

$$\ll A^{c-\sigma}(\log A)^3 T^{-1}.$$

We thus find that

$$\sum_{A \leqq n < 9A} \Lambda(n) n^{-s} = \frac{A^{1-s}(9^{1-s}-1)}{1-s} + O(A^{1-\sigma} \exp(-(\log A)^{\frac{1}{4}})(\log A)^4).$$

We write

$$P(s) = \sum_{P \leqq n < 2P} \Lambda(n) n^{-s}$$

and define $R(s)$ similarly. We also put $Q=(X-Y)/(4PR)$,

$$Q(s) = \sum_{Q \leqq n \leqq 9Q} n^{-s} \Lambda(n).$$

Then

$$S = \frac{1}{2\pi i} \int_{c-iT}^{c+iT} P(s) Q(s) R(s) \frac{(x+y)^s - x^s}{s} ds + O(YX^{-\phi}) + O(E_1)$$

from another application of Perron's formula. We write $B=\exp((\log P)^{1/5})$ and then

$$\frac{1}{2\pi i} \int_{c-iB}^{c+iB} P(s) Q(s) R(s) \frac{(x+y)^s - x^s}{s} ds = \frac{1}{2\pi i} \int_{c-iB}^{c+iB} \frac{Q^{1-s}(9^{1-s}-1)}{1-s} P(s) R(s) \frac{(x+y)^s - x^s}{s} ds + O(Y(\log X)^{-10}). \tag{30}$$

Now, in (30),

$$\frac{(x+y)^s - x^s}{s} = y x^{s-1} + O(B y^2 x^{c-2}).$$

The expression (30) thus becomes

$$\frac{y}{2\pi i} \int_{c-iB}^{c+iB} P(s) R(s) \frac{(x/Q)^{s-1}(1-9^{1-s})}{s-1} ds + O(Y(\log X)^{-10}) = y \sum_{P \leqq n < 2P} \sum_{R \leqq r < 2R} \frac{\Lambda(n)\Lambda(r)}{nr} + O(Y(\log X)^{-10})$$

using Perron's formula again. To complete the proof it thus suffices to show that

$$\int_{c+iB}^{c+iT} |P(s)\,Q(s)\,R(s)|^2\,|ds| \ll X(\log X)^{-10}. \tag{31}$$

The inequality (31) may be proved in the same way as Lemma 5 was demonstrated, using the present $Q(s)$ in the rôle played in Lemma 5 by $K(s)$ since

$$Q(s) \ll B^{-1}$$

on the line of integration.

We are now in a position to give a lower bound for Σ_4. This sum counts all integers in $(x-y, x]$ of the form pqr where p, q, r are primes with

$$(X^{1-3\varepsilon}/p)^{1/3} \leqq q < p < X^{\frac{1}{2}} \quad \text{and} \quad r \geqq q,\ p > X^{9/35}. \tag{32}$$

We divide the sum into subsums $S(P, R)$ corresponding to $P \leqq p < 2P$, $R \leqq r < 2R$. Then

$$S(P,R) = \frac{(1+O((\log X)^{-1}))}{\log(X/PR)} \sum_{\substack{P \leqq n < 2P \\ R \leqq m < 2R}} \frac{\Lambda(n)\,\Lambda(m)}{\log P \log R} \left(\psi\left(\frac{x}{mn}\right) - \psi\left(\frac{x-y}{mn}\right)\right) + E_3$$

where

$$\frac{1}{X}\int_X^{2X} |E_3|\,dx \ll Y(P^{-1/3} + R^{-1/3} + (X/PR)^{-1/3}).$$

We may estimate $S(P, R)$ by Lemma 7 if $X^{1-\varepsilon} > PR \geqq X^{27/35}$. If

$$64P^2R^3X^{1-3\varepsilon} \leqq X^3, \quad X^{9/35} \leqq P \leqq X^{\frac{1}{2}}/2, \quad \text{and} \quad PR \geqq X^{27/35}$$

then (32) holds ($r \geqq q$ since $PR^2 = (PR)^2 P^{-1} > X^{54/35 - \frac{1}{2}} > X$) and $PR < X^{1-\varepsilon}$ so

$$\Sigma_4 \geqq \frac{y}{\log X}(C_4 + O((\log X)^{-1})) + (\log X)^2(E_1 + E_2 + E_3).$$

Here

$$C_4 = \iint_{\substack{\frac{9}{35} \leqq v \leqq \frac{1}{2} \\ u+v \geqq \frac{27}{35},\ 2v+3u \leqq 2}} \frac{du\,dv}{uv(1-u-v)}.$$

The integral is estimated crudely as 27 times the area of integration and we find that

$$C_4 > 0.14.$$

7. Conclusion

Assembling all our information, we have shown, assuming ε is sufficiently small and X sufficiently large, that

$$\pi(x) - \pi(x-y) > y/(8\log x)$$

for all $x \in [X, 2X]$, except on a set of measure $< CX(\log X)^{-1}$. We note that $[x]$ takes on $\geqq X - CX(\log X)^{-1}$ values, that

$$\pi([x]) - \pi([x] - y) > y(10 \log x)^{-1},$$

and $[x]^{(1/10)+\varepsilon} \leqq 2y \leqq 2[x]^{(1/10)+\varepsilon}$. This completes the proof of the theorem. Possibly by using a more complicated decomposition instead of (3) one could estimate the remainder term in a satisfactory manner with an exponent smaller than 1/10.

By a similar method to that used in this paper, it will be shown elsewhere that there are infinitely many solutions of the inequality

$$\{\sqrt{p} - \beta\} < p^{-\frac{1}{4}+\varepsilon}$$

for any real β. Hence $\{x\}$ denotes the fractional part of x. The exponent $\frac{1}{4}$ improves on an exponent slightly smaller than 1/6 given by Kaufman [10].

Acknowledgements. I would like to thank R.C. Baker for his careful checking of my manuscript and his helpful suggestions on matters of exposition. I would also like to thank R. Heath-Brown for an interesting discussion on the subject, and H. Iwaniec for his encouragement.

References

1. Cramér, H.: On the order of magnitudes of the difference between consecutive prime numbers. Acta Arith. **2**, 23–46 (1937)
2. Halberstam, H., Richert, H.E.: Sieve Methods. New York-London: Academic Press 1974
3. Harman, G.: Almost-primes in short intervals. Math. Ann. **258**, 107–112 (1981)
4. Heath-Brown, D.R.: Zero density estimates for the Riemann Zeta-function and Dirichlet *L*-functions. J. London Math. Soc. (2), **19**, 221–232 (1979)
5. Heath-Brown, D.R., Iwaniec, H.: On the difference between consecutive primes. Invent. Math. **55**, 49–69 (1979)
6. Huxley, M.N.: On the difference between consecutive primes. Invent. Math. **15**, 164–170 (1972)
7. Iwaniec, H.: A new form of the error term in the linear sieve. Acta Arith. **37**, 307–320 (1980)
8. Iwaniec, H., Jutila, M.: Primes in short intervals. Ark. Mat. **17**, 167–176 (1979)
9. Jutila, M.: Zero density estimates for *L*-functions. Acta Arith. **32**, 55–62 (1977)
10. Kaufman, R.M.: The distribution of $\{\sqrt{p}\}$. Mat. Zametki **26**, 497–504 (1979) [Russian]. Eng. Transl.: Math. Notes 26, (1979)
11. Montgomery, H.L.: Topics in multiplicative number theory. Berlin-Heidelberg-New York: Springer 1971
12. Selberg, A.: On the normal density of primes in short intervals, and the difference between consecutive primes. Arch. Math. Naturvid. **47**, 87–105 (1943)
13. Titchmarsh, E.C.: The Theory of the Riemann Zeta-function. Oxford at the Clarendon Press 1951
14. Vaughan, R.C.: Somes trigonometriques sur les nombres premiers. Comptes Rendus Acad. Sci. Paris, Serie A. **285**, 981–983 (1977)

Received November 11, 1981

Math. Z. 180, 349–359 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Das Nielsensche Realisierungsproblem für hinreichend große 3-Mannigfaltigkeiten

Bruno Zimmermann

Institut für Mathematik der Ruhr-Universität, Postfach 102148, D-4630 Bochum, Bundesrepublik Deutschland

Es sei M eine geschlossene orientierbare irreduzible hinreichend große 3-Mannigfaltigkeit. Die Gruppe Abb M der Abbildungsklassen von M ist die Gruppe aller Homöomorphismen von M modulo Isotopie; nach [14] ist sie isomorph zur äußeren Automorphismengruppe Out $\pi_1 M$ von $\pi_1 M$. Wir untersuchen in dieser Arbeit das Problem, ob sich endliche Untergruppen A von Abb M realisieren lassen durch isomorphe Gruppen von Homöomorphismen von M, die Repräsentanten der Abbildungsklassen in A sind. Dieses Problem wurde zuerst von Nielsen im Falle von Flächen untersucht [9].

Für Seifertsche Faserräume M wurde das Problem bereits in [19] und [20] behandelt; eine endliche Untergruppe $A \subset$ Abb M läßt sich genau dann realisieren, wenn eine Erweiterung zum abstrakten Kern $(A, \pi_1 M, i\colon A \hookrightarrow \text{Out } \pi_1 M)$ existiert. Wir werden im folgenden voraussetzen, daß M kein Seifertscher Faserraum ist; dann gibt es stets eine Erweiterung zum Kern $(A, \pi_1 M, i)$, d.h. das algebraische Hindernis gegen eine Realisierung von A ist trivial.

Besitzt M eine hyperbolische Struktur, so ist die Lösung des Realisierungsproblems eine Folgerung aus Mostow's "rigidity theorem". Nach Resultaten von Jaco-Shalen und Johannson läßt sich jede hinreichend große 3-Mannigfaltigkeit zerlegen in Faktoren, die entweder Seifertsche Faserräume sind oder eine hyperbolische Struktur besitzen; dabei ergibt sich die Existenz der hyperbolischen Struktur aus tiefliegenden Resultaten von Thurston. Wir verwenden diese Sätze, um das Realisierungsproblem für beliebige hinreichend große 3-Mannigfaltigkeiten zu lösen.

0.1. Satz. *Es sei M eine geschlossene orientierbare irreduzible hinreichend große 3-Mannigfaltigkeit und $A \subset$ Abb $M \cong$ Out $\pi_1 M$ eine endliche Gruppe von Abbildungsklassen von M. Es sei M kein Seifertscher Faserraum. Dann läßt sich A realisieren durch eine isomorphe Gruppe von Homöomorphismen von M.*

Im Falle $A \cong \mathbb{Z}_2$ wurde Satz 0.1 bereits von Heil und Tollefson bewiesen [1]; ihre Methoden lassen sich nach einer Mitteilung von Heil auch für Gruppen $A \cong \mathbb{Z}_n$ durchführen. Eine Folgerung aus Satz 0.1 sowie, im Falle Seifertscher Faserräume, der Lösung des Nielsenschen Realisierungsproblems für Flächen [5] ist folgender Satz:

0025-5874/82/0180/0349/$02.20

0.2. Satz. *Es sei M eine geschlossene orientierbare irreduzible hinreichend große 3-Mannigfaltigkeit und E eine endliche effektive torsionsfreie Erweiterung von $\pi_1 M$. Dann ist E isomorph zur Fundamentalgruppe $\pi_1 M^*$ einer geschlossenen irreduziblen 3-Mannigfaltigkeit M^*. Ist M^* nicht hinreichend groß, so ist M ein Seifertscher Faserraum oder besitzt eine vollständige hyperbolische Struktur.*

1. Grundlegende Resultate über 3-Mannigfaltigkeiten

Wir stellen in diesem Abschnitt einige Resultate über 3-Mannigfaltigkeiten zusammen, die wir beim Beweis von Satz 0.1 benötigen. Zur Definition der verschiedenen Begriffe im Zusammenhang mit 3-Mannigfaltigkeiten vergleiche man das Buch von Hempel [2].

1.1. Satz [3, 4]. *Es sei M eine orientierbare geschlossene irreduzible hinreichend große 3-Mannigfaltigkeit. Dann gibt es ein endliches System V (eindeutig bis auf Isotopie) von inkompressiblen paarweise nichtparallelen Tori in M mit folgenden Eigenschaften:*

(i) *V ist charakteristisch, d.h. ist $f\colon M \to M$ ein Homöomorphismus, so ist $f(V)$ ambient isotop zu V.*

(ii) *Der Abschluß N jeder Komponente von $M-V$ ist entweder ein Seifertscher Faserraum oder jede Untergruppe der Form $\mathbb{Z}\oplus\mathbb{Z}$ von $\pi_1 N$ ist periphär, d.h. konjugiert in $\pi_1 N$ zur Fundamentalgruppe einer Randkomponente von N.*

1.2. Satz [13]. *Es sei M eine orientierbare irreduzible randirreduzible hinreichend große 3-Mannigfaltigkeit, deren Randkomponenten alle Tori sind. M sei nicht homöomorph zu einem I-Bündel über einer Fläche, $I=[0,1]$. Genau dann besitzt das Innere $\mathring{M}=M-\partial M$ von M eine vollständige hyperbolische Struktur mit endlichem Volumen, wenn jede Untergruppe der Form $\mathbb{Z}\oplus\mathbb{Z}$ von $\pi_1 M$ periphär ist.*

1.3. Satz ([8], vgl. auch [7, 5.1]). *Es sei M eine orientierbare 3-Mannigfaltigkeit, deren Inneres $\mathring{M}=M-\partial M$ eine vollständige hyperbolische Struktur mit endlichem Volumen besitzt (insbesondere sind alle Randkomponenten von M Tori). Dann ist $\mathrm{Out}\,\pi_1 M$ endlich und die Operation von $\mathrm{Out}\,\pi_1 M$ auf $\pi_1 M$ läßt sich realisieren durch die Operation einer isomorphen Gruppe von Isometrien von M. Die (bis auf Äquivalenz eindeutige) Erweiterung $1\to\pi_1 M\hookrightarrow E\to\mathrm{Out}\,\pi_1 M\to 1$ zur Operation von $\mathrm{Out}\,\pi_1 M$ auf $\pi_1 M$ läßt sich realisieren durch eine zu E isomorphe Gruppe von Isometrien der universellen Überlagerung von M mit der Decktransformationsgruppe $\pi_1 M$ als Untergruppe.*

2. Seifertsche Faserräume

Die Komponenten von $M-V$ in Satz 1.1 sind entweder Seifertsche Faserräume mit nichtleerem Rand oder besitzen nach Satz 1.2 eine vollständige hyperbolische Struktur mit endlichem Volumen. Für diese hyperbolischen Mannigfaltigkeiten folgt der Realisierungssatz 0.1 aus Satz 1.3. In diesem Abschnitt be-

trachten wir das Realisierungsproblem für Seifertsche Faserräume mit nichtleerem Rand.

Es sei S ein kompakter Seifertscher Faserraum mit nichtleerem Rand und $H \subset \pi_1 S$ die Homotopieklasse der regulären Fasern. Dann ist $\pi_1 S/H$ isomorph zu einer Gruppe F von Bewegungen der Sphäre, der euklidischen oder der hyperbolischen Ebene. Wir nennen S hinreichend kompliziert, falls F isomorph ist zu einer (nichtelementaren) Gruppe von Bewegungen der hyperbolischen Ebene. In diesem Fall gilt $H \cong \mathbb{Z}$ (vgl. [10]).

In [18, 5.7] (vgl. auch [21, 2.2, insbesondere die Fußnote S. 366]) ist folgender Satz bewiesen.

2.1. Satz. *Es sei S ein kompakter hinreichend komplizierter Seifertscher Faserraum mit nichtleerem Rand und A eine endliche Gruppe von Abbildungsklassen von M. Existiert eine Erweiterung zum abstrakten Kern $(A, \pi_1 S, i\colon A \hookrightarrow \mathrm{Out}\,\pi_1 M)$, so läßt sich A realisieren durch eine Gruppe von (fasertreuen) Homöomorphismen von S. Darüber hinaus läßt sich jede Erweiterung $1 \to \pi_1 S \hookrightarrow E \to A \to 1$ zum Kern $(A, \pi_1 S, i)$ realisieren durch eine zu E isomorphe Gruppe von Homöomorphismen der universellen Überlagerung von S mit der Decktransformationsgruppe $\pi_1 S$ als Untergruppe.*

Eine Erweiterung $1 \to \pi_1 S \hookrightarrow E \to A \to 1$ zum abstrakten Kern $(A, \pi_1 S, i)$ ist effektiv (d.h. kein Element aus $E - \pi_1 S$ operiert durch Konjugation trivial auf $\pi_1 S$), da A als Gruppe von Abbildungsklassen effektiv auf $\pi_1 S$ operiert.

Ist eine beliebige (d.h. nicht notwendigerweise effektive) endliche Erweiterung $1 \to \pi_1 S \hookrightarrow E' \to A' \to 1$ von $\pi_1 S$ gegeben, so läßt sich E' unter Umständen immer noch als isomorphe Gruppe von Homöomorphismen der universellen Überlagerung von S realisieren. Dies ist in [21, Abschnitt 2.A] untersucht. Aus dem Beweis von [21, Satz 2.2] ergibt sich folgendes

2.2. Lemma. *Genau dann läßt sich E' realisieren, falls keine Untergruppe der Form $\pi_1 S \oplus \mathbb{Z}_p$ von E' existiert* (man vergleiche hierzu auch die folgenden ähnlichen Schlüsse).

Die Komponenten von $M - V$ in Satz 1.1, die (orientierbare) Seifertsche Faserräume sind, sind nicht notwendigerweise hinreichend kompliziert. Wie man aus der Klassifikation der Seifertschen Faserräume (mit nichtleerem Rand) jedoch leicht sieht (vgl. [2, 10]), ist die einzige weitere Möglichkeit der Seifertsche Faserraum S^*, der als Zerlegungsfläche (Menge der Fasern) die Kreisscheibe besitzt mit genau zwei ausgezeichneten Punkten, über denen Ausnahmefasern der Ordnung zwei liegen. (S^* besitzt eine weitere Seifertsche Faserung ohne Ausnahmefasern mit dem Möbiusband als Zerlegungsfläche.) Da dieser Fall auch in anderer Hinsicht eine Ausnahmestellung einnimmt, wollen wir ihn hier direkt behandeln. Die Gruppe F ist in diesem Falle isomorph zu $\mathbb{Z}_2 * \mathbb{Z}_2$, das auf der euklidischen Ebene operiert bzw. auf $\mathbb{R} \times [0, 1]$, falls man die Randkomponenten berücksichtigt. Es gilt

$$\pi_1 S^* \cong \langle h, x, y \mid [x, h] = [y, h] = 1, x^2 = h, y^2 = h \rangle,$$

und es gibt eine exakte Sequenz

$$1 \to H = \langle h \rangle \cong \mathbb{Z} \hookrightarrow \pi_1 S^* \to \mathbb{Z}_2 * \mathbb{Z}_2 \to 1.$$

Die universelle Überlagerung von S^* ist $U := \mathbb{R} \times (\mathbb{R} \times [0,1])$ und hat genau 2 Randkomponenten, auf denen dieselbe Untergruppe der Decktransformationsgruppe operiert.

Es sei $1 \to \pi_1 S^* \hookrightarrow E' \to A' \to 1$ eine endliche Erweiterung. Wir wollen untersuchen, wann sich E' durch eine Gruppe von Homöomorphismen von U realisieren läßt. Es sei zusätzlich ein Homomorphismus $\varphi: E' \to \mathbb{Z}_2 = \{+1, -1\}$ gegeben, der festlegt, ob die beiden Randkomponenten von U durch einen Homöomorphismus e einer Realisierung von E' festbleiben (falls $\varphi(e) = 1$) oder vertauscht werden sollen (falls $\varphi(e) = -1$) (im Gegensatz zu den „hinreichend komplizierten" Fällen ist dies hier nicht schon rein algebraisch durch die Erweiterung festgelegt).

Es sei $H' := \{x \in E' : f^{-1} x^{-1} f x \in H\}$ für alle $\{f \in \pi_1 S^*\}$. Es gilt $H' \cap \pi_1 S^* = H$, da das Zentrum von $\mathbb{Z}_2 * \mathbb{Z}_2$ trivial ist. Es folgt, daß H' eine endliche Erweiterung von $H \cong \mathbb{Z}$ ist. Nach [18, 3.5] gibt es eine endliche Untergruppe T von H' mit $H'/T \cong \mathbb{Z}$ oder $\mathbb{Z}_2 * \mathbb{Z}_2$, wobei die Elemente aus T durch Konjugation trivial auf H operieren; T ist Normalteiler von E'. Es sei $t \in T$, $t^n = 1$. Dann gilt für $f \in \pi_1 S^*$ eine Gleichung $t^{-1} f t = f \cdot z$, $z \in H$. Es folgt $f = t^{-n} f t^n = f \cdot z^n$ und daher $z = 1$. Daher operiert t trivial auf $\pi_1 S^*$ und es folgt $\pi_1 S^* \oplus T \subset E'$.

Nach Definition von H' ist die Gruppe E'/H' eine endliche effektive Erweiterung von $\pi_1 S^*/H \cong \mathbb{Z}_2 * \mathbb{Z}_2$ und daher ebenfalls isomorph zu $\mathbb{Z}_2 * \mathbb{Z}_2$, wobei $\pi_1 S^*/H$ vom Index zwei in $\mathbb{Z}_2 * \mathbb{Z}_2 = E'/H'$ liegt. Wir haben ein Diagramm

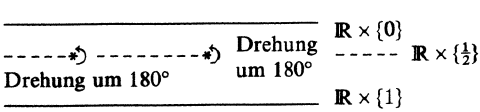
$$\begin{array}{ccccccc} 1 \to & H' & \hookrightarrow & E' & \longrightarrow & \mathbb{Z}_2 * \mathbb{Z}_2 & \to 1 \\ & \cup & & \cup & & \cup & \\ 1 \to & H & \hookrightarrow & \pi_1 S^* & \longrightarrow & \pi_1 S^*/H & \to 1 \end{array}$$

Wir unterscheiden drei Fälle.

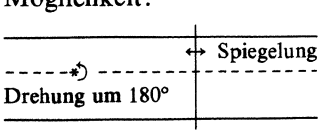
1. Fall. $T = 1$. Dann läßt sich eine Operation von E' auf U direkt konstruieren, wobei sich die Operation zusammensetzt aus kanonischen Operationen von $H' \cong \mathbb{Z}$ oder $\mathbb{Z}_2 * \mathbb{Z}_2$ auf $\mathbb{R}$ bzw. $\mathbb{Z}_2 * \mathbb{Z}_2 = E'/H'$ auf $(\mathbb{R} \times [0,1])$; dabei operiert ein $\mathbb{Z}_2 \subset \mathbb{Z}_2 * \mathbb{Z}_2 = E'/H'$ als Drehung oder Spiegelung auf $(\mathbb{R} \times [0,1])$, je nach dem Wert von φ:

Operation von $\mathbb{Z}_2 * \mathbb{Z}_2 = E'/H'$ auf $(\mathbb{R} \times [0,1])$:

1) Möglichkeit:



2) Möglichkeit:



2. Fall. $T \cong \mathbb{Z}_2 = \langle t \rangle$. Wir unterscheiden zwei Teilfälle:

(i) $\varphi(t) = 1$. Dann läßt sich E' nicht (treu!) realisieren, da $\mathbb{Z} \oplus \mathbb{Z} \oplus \mathbb{Z}_2$ sonst auf beiden Randkomponenten von U operieren würde, was unmöglich ist ($\mathbb{Z}^2 + \mathbb{Z}_2$ ist keine kristallographische Gruppe des $\mathbb{R}^2$, vgl. [16]).

(ii) $\varphi(t)=-1$. Es sei $E''\subset E'$ der Kern von φ. Nach dem 1. Fall läßt sich E'' realisieren; außerdem gilt $E'=E''\oplus\mathbb{Z}_2$, da $E''\cap\mathbb{Z}_2=1$. Dann läßt sich t auf $\mathbb{R}$ als Identität und auf $(\mathbb{R}\times[0,1])$ als Spiegelung an $\mathbb{R}\times\{1/2\}$ definieren.

3. Fall. $|T|>2$. Dann läßt sich E' nach dem analogen Schluß wie im 2. Fall (i) nicht realisieren.

Wir fassen unsere Betrachtungen zusammen:

2.3. Lemma. *Genau dann läßt sich E' realisieren, falls keine Untergruppe der Form $\pi_1 S^*\oplus\mathbb{Z}_p$ von E' existiert mit $p>2$ oder $p=2$ und $\varphi(\mathbb{Z}_p=\mathbb{Z}_2)=1$.*

3. Beweis von Satz 0.1

Da M kein Seifertscher Faserraum ist, ist das Zentrum von $\pi_1 M$ trivial [15] und es existiert eine (effektive) Erweiterung

$$1\to\pi_1 M\hookrightarrow E\longrightarrow A\longrightarrow 1$$

zum abstrakten Kern $(A,\pi_1 M, i\colon A\hookrightarrow \mathrm{Out}\,\pi_1 M)$ (vgl. [6, Theorem 9.1]).

Wir zerlegen den Beweis in mehrere Schritte.

A. Konstruktion der universellen Überlagerung von M und die algebraische Struktur von $\pi_1 M$

Es sei $V=\{T_1,\dots,T_n\}$ ein System von inkompressiblen disjunkten Tori in M wie in Satz 1.1. Es seien $M_1,\dots,M_m$ die Abschlüsse der Komponenten von $M-\bigcup_{i=1}^{n} T_i$ (wie erhalten die M_j z.B., indem wir das Innere hinreichend kleiner disjunkter regulärer Umgebungen der T_i aus M entfernen). Wegen Satz 1.1 und Satz 1.2 ist jedes M_j entweder ein Seifertscher Faserraum oder das Innere von M_j besitzt eine vollständige hyperbolische Struktur mit endlichem Volumen.

Es sei Γ der Graph, dessen Eckpunkte die Mannigfaltigkeiten M_j sind; zwei Eckpunkte M_{j_1} und M_{j_2} seien genau dann durch eine Kante T_i verbunden, wenn M_{j_1} und M_{j_2} in T_i aneinander grenzen.

Es sei $p\colon \tilde{M}\to M$ die universelle Überlagerung von M, die man folgendermaßen konstruieren kann.

Es sei $p_j\colon \tilde{M}_j\to M_j$ die universelle Überlagerung von M_j. Dann ist jede Komponente von $p_j^{-1}(T_i)$ homöomorph zu $\mathbb{R}^2$ und universelle Überlagerung $\tilde{T}_i$ von T_i, da $\pi_1 T_i\to\pi_1 M_j$ injektiv ist (T_i liege im Rand von M_j). Wir beginnen mit einem $\tilde{M}_{j_0}=:N_0$ und kleben entlang jeder Randkomponente $\mathbb{R}^2$ von N_0 die entsprechende Kopie eines $\tilde{M}_j$ an, wobei die Nachbarschaftsverhältnisse in M bzw. Γ zugrunde gelegt werden. Der entstehende Raum heiße N_1. Wir iterieren das Verfahren und erhalten Räume $N_2,N_3,\dots$; dann ist $\tilde{M}:=\bigcup_{i=1}^{\infty} N_i$ nach Konstruktion eine Überlagerung von M.

Völlig analog zur Konstruktion von Γ aus M kann man $\tilde{M}$ einen Graphen $\tilde{\Gamma}$ zuordnen, dessen Ecken die verschiedenen Kopien der $\tilde{M}_j$ sind. Nach Konstruktion von $\tilde{M}$ ist $\tilde{\Gamma}$ ein Baum, und wir haben eine kanonische Projektion

$p'\colon \tilde{\Gamma}\to\Gamma$. Außerdem läßt sich $\tilde{\Gamma}$ kanonisch in $\tilde{M}$ einbetten (ebenso wie Γ in M), und jeder geschlossene Weg in $\tilde{M}$ ist homotop zu einem Weg in $\tilde{\Gamma}$. Daher ist $\tilde{M}$ einfach zusammenhängend und universelle Überlagerung von M. Die Operation von $\pi_1 M$ auf $\tilde{M}$ als Decktransformationsgruppe induziert eine Operation von $\pi_1 M$ auf $\tilde{\Gamma}$ mit $\tilde{\Gamma}/\pi_1 M=\Gamma$, d.h. $\pi_1 M$ operiert auf einem Baum mit kompaktem Faktorgraphen.

Die algebraische Struktur von $\pi_1 M$ läßt sich direkt aus dem Satz von Seifert-Van Kampen bestimmen oder mit den Methoden von [12, § 5] unter Verwendung der Operation von $\pi_1 M$ auf $\tilde{\Gamma}$. Man erhält $\pi_1 M$ jeweils als HNN-Erweiterung mit einem Baumprodukt als Basisgruppe. Der zugrundeliegende Baum ist dabei ein maximaler Baum B^* in Γ bzw. dessen Liftung nach $\tilde{\Gamma}$, die Ecken- bzw. Kantengruppen sind die Fundamentalgruppen der entsprechenden Mannigfaltigkeiten bzw. die Stabilisatoren bzgl. der Operation von $\pi_1 M$ auf $\tilde{\Gamma}$. Die „stabilen Buchstaben" $s_1,\ldots,s_\nu$ der HNN-Erweiterung entsprechen bijektiv den Kanten in $\Gamma-B^*$ (eine analoge Konstruktion wird im nächsten Abschnitt für die Gruppe E beschrieben).

B. Konstruktion einer Operation der Erweiterung E von $\pi_1 M$ auf $\tilde{\Gamma}$

Nach Satz 1.1 können wir in jeder Abbildungsklasse $[a]$ in A als Repräsentanten einen Homöomorphismus a von M wählen mit $a\left(\bigcup_{i=1}^{n} T_i\right)=\bigcup_{i=1}^{n} T_i$. Es sei A' die Menge dieser Repräsentanten mit $\mathrm{id}_M\in A'$. Ist a^ν isotop zur Identität von M für $a\in A'$, $\nu\geqq 1$, so induziert a^ν einen inneren Automorphismus von $\pi_1 M$. Wegen der algebraischen Struktur von $\pi_1 M$ muß a^ν daher jedes M_j bzw. T_i auf sich abbilden. Es folgt, daß A' eine Operation der Gruppe A auf dem Graphen Γ induziert. Wir können ohne Einschränkung annehmen, daß A ohne Inversion auf Γ operiert, d.h. keine Kante auf sich abbildet unter Vertauschung der Ecken der Kante. Dies läßt sich durch entsprechendes Unterteilen der Kanten erreichen, wobei jedem neuen Eckpunkt ein neues M_j der Form $T_i\times[0,1]$ entspricht. Es sei E' die Menge aller Liftungen der Elemente aus A' nach $\tilde{M}$, mit $\pi_1 M\subset E'$. Gehen wir von $\tilde{M}$ nach $\tilde{\Gamma}$ über, so definiert jedes Element in E' einen Automorphismus von $\tilde{\Gamma}$. Insgesamt erhalten wir hierdurch eine Operation der Gruppe E auf dem Baum $\tilde{\Gamma}$; dabei ergibt die Projektion von E nach $\Gamma=\tilde{\Gamma}/\pi_1 M$ gerade die Operation von $A\cong E/\pi_1 M$ auf Γ. Die algebraische Struktur von E läßt sich wiederum nach [12, § 5] bestimmen. Es sei $\tilde{\Gamma}/E$ der Faktorgraph, in dem wir einen maximalen Baum B' wählen. Es sei B eine isomorphe Liftung von B' nach $\tilde{\Gamma}$.

Die Ecken bzw. Kanten von B bzw. $\tilde{\Gamma}$ seien durch die entsprechenden Stabilisatoren in E gekennzeichnet. Es seien $E_1,\ldots,E_\alpha$ bzw. $K_1,\ldots,K_\beta$ die (Stabilisatoren der) Ecken bzw. Kanten von B. Darüber hinaus liften wir jede Kante in $(\tilde{\Gamma}/E-B')$ nach $\tilde{\Gamma}$ mit Anfangspunkt in B; die so erhaltenen Kanten (bzw. Stabilisatoren) seien mit $F_1,\ldots,F_\gamma$ bezeichnet. Für jedes F_i gibt es ein Element t_i in E, daß den Endpunkt von F_i nach B abbildet. In E gilt $t_i^{-1}F_i t_i=G_i\subset E_{\delta(i)}$, wobei G_i der Stabilisator der Kante $t_i(F_i)$ ist.

Es sei $F:=B\cup F_1\cup\ldots\cup F_\gamma$, wobei die Endpunkte der F_i nicht zu F gehören sollen. Dann ist F ein Fundamentalbereich für die Operation von E auf $\tilde{\Gamma}$, d.h.

aus jeder Äquivalenzklasse von Ecken bzw. Kanten bzgl. der Operation von E auf $\tilde{\Gamma}$ liegt genau ein Repräsentant in F (Fig. 1).

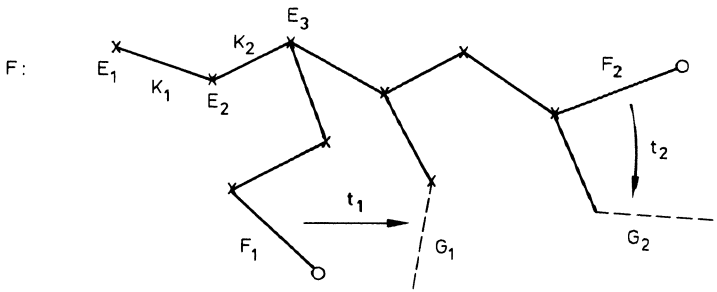


Fig. 1

Nach [12, § 5] gilt

$$E = \langle \text{Erzeugende von } \bar{B},\ t_1, \ldots, t_\gamma \mid \text{Relationen von } \bar{B},\ t_1^{-1} F_1 t_1 = G_1, \ldots, t_\gamma^{-1} F_\gamma t_\gamma = G_\gamma \rangle;$$

dabei ist $\bar{B}$ das Baumprodukt über dem Baum von Gruppen B, d.h. ein iteriertes freies Produkt mit Amalgam der Eckengruppen von B, wobei über den Kantengruppen von B amalgamiert wird.

C. Realisierung der Eckengruppen von B als Gruppen von Homöomorphismen

Jeder Ecke e bzw. Kante k von $\tilde{\Gamma}$ entspricht eine Mannigfaltigkeit $\tilde{M}_e \subset \tilde{M}$ bzw. $\tilde{T}_k \subset \tilde{M}$, wobei $\tilde{M}_e$ universelle Überlagerung eines $M_e \in \{M_1, \ldots, M_m\}$ und $\tilde{T}_k$ universelle Überlagerung eines $T_k \in \{T_1, \ldots, T_n\}$ ist. Wir betrachten insbesondere eine Ecke E_i bzw. Kante K_j von B, $1 \leqq i \leqq \alpha$, $1 \leqq j \leqq \beta$. Zu E_i bzw. K_j erhalten wir exakte Sequenzen

$$\begin{array}{ccccccc} 1 \to \pi_1 M_{E_i} = \pi_1 M \cap E_i & \hookrightarrow & E_i & \to & A_i := E_i / \pi_1 M_{E_i} & \to & 1 \\ \cap & & \cap & & \cap & & \\ 1 \to \pi_1 M & \hookrightarrow & E & \to & A & \longrightarrow & 1 \end{array}$$

bzw.

$$\begin{array}{ccccccc} 1 \to \pi_1 T_{K_j} = \pi_1 M \cap K_j & \hookrightarrow & K_j & \to & B_j := K_j / \pi_1 T_{K_j} & \to & 1 \\ \cap & & \cap & & \cap & & \\ 1 \to \pi_1 M & \hookrightarrow & E & \to & A & \longrightarrow & 1. \end{array}$$

Dabei ist $\pi_1 M_{E_i}$ bzw. $\pi_1 T_{K_j}$ Decktransformationsgruppe der universellen Überlagerung $\tilde{M}_{E_i}$ bzw. $\tilde{T}_{K_j}$ von M_{E_i} bzw. T_{K_j}; es besteht $\pi_1 M_{E_i}$ aus genau den Elementen aus $\pi_1 M$, die $\tilde{M}_{E_i} \subset \tilde{M}$ auf sich abbilden, analog für $\pi_1 T_{K_j}$.

Wir wollen zeigen, daß sich jedes E_i realisieren läßt als Gruppe von Homöomorphismen von $\tilde{M}_{E_i}$ mit der Decktransformationsgruppe $\pi_1 M_{E_i}$ als Untergruppe. Ist die Erweiterung $1 \to \pi_1 M_{E_i} \hookrightarrow E_i \to A_i \to 1$ effektiv, so folgt die Behauptung aus Satz 2.1 bzw. Satz 1.3.

Wir nehmen an, für ein $i\in\{1,\ldots,\alpha\}$ läßt sich E_i nicht realisieren. Dann gibt es eine Untergruppe $\mathbb{Z}_p\subset A_i\subset A$ mit $\pi_1 M_{E_i}\oplus\mathbb{Z}_p\subset E_i$: ist M_{E_i} ein Seifertscher Faserraum, so folgt dies aus Lemma 2.2 bzw. 2.3, ist M_{E_i} eine hyperbolische Mannigfaltigkeit, so folgt es aus der Tatsache, daß die Erweiterung $1\to\pi_1 M_{E_i}\hookrightarrow E_i\to A_i\to 1$ nicht effektiv und das Zentrum von $\pi_1 M_{E_i}$ trivial ist. Für den folgenden Widerspruchsbeweis sei ohne Einschränkung $A=\mathbb{Z}_p$ und p eine Primzahl. Die Untergruppe $\mathbb{Z}_p\subset E_i\subset E$ operiert durch Konjugation trivial auf $\pi_1 M_{E_i}$ und fixiert alle in E_i beginnenden Kanten von $\tilde\Gamma$ (vgl. Lemma 2.3). Ist K_j eine solche Kante mit Endpunkt $E_{i'}$, so läßt sich $K_j\cong\mathbb{Z}\oplus\mathbb{Z}\oplus\mathbb{Z}_p$ nicht als isomorphe Gruppe von Homöomorphismen von $\tilde T_{K_j}\cong\mathbb{R}^2$ realisieren, da $\mathbb{Z}^2\oplus\mathbb{Z}_p$ keine kristallographische Gruppe des $\mathbb{R}^2$ ist. Dann läßt sich aber auch $E_{i'}$ nicht als isomorphe Gruppe von Homöomorphismen von $\tilde M_{E_{i'}}$ realisieren; wegen 2.2 und 2.3 gilt wieder $\pi_1 M_{E_{i'}}\oplus\mathbb{Z}_p\subset E_{i'}$. Durch Induktion folgt, daß $\mathbb{Z}_p\subset E$ alle Ecken und Kanten von $\tilde\Gamma$ fixiert und trivial auf den entsprechenden Stabilisatoren operiert. Ist $s\in\{s_1,\ldots,s_\nu\}$ ein „stabiler Buchstabe" der HNN-Erweiterung $\pi_1 M$ (vgl. Teil A), so gilt $x^{-1}sx=s\cdot y$, wobei $y\in\pi_1 M$ im Stabilisator einer Ecke von $\tilde\Gamma$ liegt und x erzeugendes Element von $\mathbb{Z}_p$ ist. Es folgt $s=x^{-p}sx^p=s\cdot y^p$ und daher $y=1$, da $\pi_1 M$ keine Elemente endlicher Ordnung enthält. Daher gilt $\pi_1 M\oplus\mathbb{Z}_p=E$, was der Effektivität der Erweiterung $1\to\pi_1 M\hookrightarrow E\to A\to 1$ widerspricht. Damit erhalten wir, daß sich alle Stabilisatoren E_i der Ecken von B durch isomorphe Gruppen von Homöomorphismen realisieren lassen. Sind zwei Ecken E_{i_1}, E_{i_2} in B durch eine Kante K_j verbunden, so können wir annehmen, daß die Realisierungen von $K_j\subset E_{i_1}$ und $K_j\subset E_{i_2}$ auf $\tilde T_{K_j}=\tilde M_{E_{i_1}}\cap\tilde M_{E_{i_2}}$ übereinstimmen, da beide äquivalent zur Operation einer kristallographischen Gruppe des $\mathbb{R}^2$ sind.

Abschließend realisieren wir noch die Elemente $t_j\in E$ (die stabilen Buchstaben der HNN-Erweiterung E, vgl. Teil B) jeweils durch Homöomorphismen von $\tilde T_{F_j}$ nach $\tilde T_{G_j}$, die wir wieder mit t_j bezeichnen; dabei soll $t_j^{-1}F_jt_j=G_j$ gelten (die F_j bzw. G_j sind ja als Untergruppen der E_i auf $\tilde T_{F_j}$ bzw. $\tilde T_{G_j}$ bereits realisiert und äquivalent zu kristallographischen Gruppen des $\mathbb{R}^2$; jeder Isomorphismus zwischen zwei solchen Gruppen läßt sich durch einen Homöomorphismus induzieren).

Bemerkung. Ist M_{E_i} ein Seifertscher Faserraum, so braucht die Erweiterung $1\to\pi_1 M_{E_i}\hookrightarrow E_i\to A_i\to 1$ nicht effektiv zu sein. Es ist leicht, eine 3-Mannigfaltigkeit M wie in 0.1 zu konstruieren zusammen mit einer Involution von M, die auf der Fundamentalgruppe einer der Komponenten M_j der durch 1.1 gegebenen Zerlegung trivial operiert (notwendigerweise ist diese Komponente dann ein Seifertscher Faserraum).

D. Konstruktion einer Operation von E auf $\tilde M$

Wir wollen die im vorangehenden Teil definierten Operationen der Stabilisatoren E_i bzw. der stabilen Buchstaben t_j auf den $\tilde M_{E_i}$ bzw. den $\tilde T_{K_j}$ erweitern zu einer Operation von E auf $\tilde M$.

Für jede Ecke e von $\tilde\Gamma$ wählen wir im folgenden induktiv über den Abstand von B ein Element x_e in E mit $x_e(e)\in B$ und definieren dann zu x_e einen Homöomorphismus von $\tilde M_e$ auf $\tilde M_{x_e(e)}$, den wir ebenfalls mit x_e bezeichnen.

Für Ecken in B wählen wir jeweils die Identität. Es sei e eine Ecke im Abstand n von B, für die x_e bereits definiert sei. Es sei e' eine Ecke im Abstand $n+1$ von B, die mit e durch eine Kante k verbunden sei. Wir unterscheiden 3 Fälle:

(i) $x_e(k) \in B$. Dann sei $x_{e'} := x_e$. Wir setzen den Homöomorphismus $x_e | \tilde{T}_k : \tilde{T}_k \to \tilde{T}_{x_e(k)}$ fort zu einem Homöomorphismus $x_e : \tilde{M}_{e'} \to \tilde{M}_{x_e(e')}$. Als Homöomorphismus können wir z.B. den zu x_e korrespondierenden Homöomorphismus in E' nehmen (vgl. Teil B), den wir in einer Umgebung der Randkomponente $\tilde{T}_k$ von $\tilde{M}_{e'}$ geeignet deformieren.

(ii) $x_e(k) \in F$, $x_e(k) \notin B$. Dann gibt es einen stabilen Buchstaben $t_j \in E$ mit $t_j(x_e(e')) \in B$ und es sei $x_{e'} := t_j \circ x_e$. Es ist $x_{e'}$ auf $\tilde{T}_k$ bereits als Homöomorphismus definiert, und wir setzen $x_{e'} | \tilde{T}_k$ wie unter (i) fort zu einem Homöomorphismus $\tilde{M}_{e'} \to \tilde{M}_{x_{e'}(e')}$.

(iii) $x_e(k) \notin F$. Dann gibt es ein Element y aus dem Stabilisator von $x_e(e)$ in E oder einen stabilen Buchstaben $y \in \{t_1, \ldots, t_y\}$ (oder ein Produkt aus beiden) mit $y^{-1}(x_e(e')) \in B$ und es sei $x_{e'} := y^{-1} \circ x_e$. Es ist $x_{e'}$ auf $\tilde{T}_k$ bereits als Homöomorphismus definiert, und wir setzen wie oben $x_{e'} | \tilde{T}_k$ fort zu einem Homöomorphismus $\tilde{M}_{e'} \to \tilde{M}_{x_{e'}(e')}$.

Damit ist x_e für alle Ecken von $\tilde{\Gamma}$ als Homöomorphismus $\tilde{M}_e \to \tilde{M}_{x_e(e)}$ definiert. Nun sei e eine beliebige Ecke in $\tilde{\Gamma}$ und $y \in E$. Dann gilt $y = x_{y(e)}^{-1} \circ s_y \circ x_e$ für ein Element s_y aus dem Stabilisator von $x_e(e)$. Wir definieren den Homöomorphismus $y | \tilde{M}_e$ als $x_{y(e)}^{-1} \circ s_y \circ x_e$ von $\tilde{M}_e$ nach $\tilde{M}_{y(e)}$. Dann ist für jede Kante k von Γ der Homöomorphismus $y | \tilde{T}_k$ durch die beiden Eckpunkte zweimal definiert; aufgrund unserer Wahl (i)–(iii) der x_e stimmen diese beiden Definitionen jedoch überein.

Damit ist jedem Element y in E ein gleichbezeichneter Homöomorphismus von $\tilde{M}$ zugeordnet. Durch diese Zuordnung wird die Gruppe E als Gruppe von Homöomorphismen von $\tilde{M}$ definiert.

E. Beweis von Satz 0.1

Wir haben in Teil D eine Operation von E auf $\tilde{M}$ definiert. Dabei stimmt die Operation von $\pi_1 M \subset E$ nicht notwendigerweise mit der ursprünglichen Operation als Decktransformationsgruppe überein. Die Faktorräume beider Operationen sind jedoch hinreichend große 3-Mannigfaltigkeiten, die nach [14] homöomorph sind. Daher können wir ohne Einschränkung annehmen, daß die Operationen übereinstimmen. Dann operiert $E/\pi_1 M \cong A$ auf $\tilde{M}/\pi_1 M \cong M$ und realisiert A. Damit ist der Satz bewiesen.

F. Beweis von Satz 0.2

Ist M ein Seifertscher Faserraum, so folgt die erste Behauptung aus der Lösung des Nielsenschen Realisierungsproblems für Flächen [5], vgl. [11] und [17]. Ist M kein Seifertscher Faserraum, so ergibt sich die Behauptung aus dem Beweis von Satz 0.1: da E torsionsfrei ist, operiert E fixpunktfrei auf $\tilde{M}$ und $\tilde{M}/E$ ist eine geschlossene irreduzible 3-Mannigfaltigkeit. Ist $\tilde{M}/E$ nicht hinreichend groß, so ist $E = \pi_1(\tilde{M}/E)$ nach [14, 15] kein freies Produkt mit Amalgam und keine HNN-Erweiterung; wegen [12] besteht $\tilde{\Gamma}$ daher nur aus einer einzigen Ecke, d.h. M ist ein Seifertscher Faserraum oder eine hyperbolische Mannigfaltigkeit.

Addendum

Satz 0.1 obiger Arbeit ist ebenfalls richtig für kompakte irreduzible randirreduzible 3-Mannigfaltigkeiten $M(\partial M \neq \emptyset)$; dabei hat man Seifertsche Faserräume und I-Bündel über Flächen wieder gesondert zu betrachten, vgl. hierzu [18]. Sind alle Randkomponenten von M Tori, so kann man genau den Beweis von Satz 0.1 anwenden. Andernfalls sei $M' := M \cup_{\partial M} M$ die Verdopplung von M längs ∂M und i die kanonische Involution von M', die die beiden Kopien von M vertauscht. Aus A erhält man durch Verdopplung eine Gruppe $A' \cong A$ von Abbildungsklassen von M'. Dann definieren A' und die von i erzeugte Gruppe $\mathbb{Z}_2$ von Abbildungsklassen von M' eine Gruppe A'' von Abbildungsklassen von M' mit $A'' = A' \oplus \mathbb{Z}_2$. Nach Satz 0.1 läßt sich A'' realisieren durch eine isomorphe Gruppe $\tilde{A}$ von Homöomorphismen von M'. Dabei können wir wegen Tollefson, Involutions on sufficiently large 3-manifolds (Topology **20**, 323–352 (1981)) annehmen, daß die Gruppe $\mathbb{Z}_2$ gerade durch die Involution i realisiert wird (nach Tollefson sind je zwei Realisierungen äquivalent). Dann operiert $\tilde{A}/\langle i \rangle \cong A$ auf $M'/\langle i \rangle \cong M$ und realisiert A.

Außerdem bleibt Satz 0.1 richtig für nichtorientierbare $\mathbb{P}^2$-irreduzible hinreichend große 3-Mannigfaltigkeiten. Zum Beweis gehe man zur orientierbaren Überlagerung über.

Literatur

1. Heil, W., Tollefson, J.L.: Deforming homotopy involutions of 3-manifolds to involutions II. Topology (erscheint demnächst)
2. Hempel, J.: 3-manifolds. Annals of Mathematics Studies **86**. Princeton: Princeton University Press 1976
3. Jaco, W., Shalen, P.: Seifert fibered spaces in 3-manifolds. Mem. Amer. Math. Soc. **220** (1979)
4. Johannson, K.: Homotopy equivalences of 3-manifolds with boundaries. Lecture Notes in Mathematics **761**. Berlin-Heidelberg-New York: Springer 1979
5. Kerckhoff, S.P.: The Nielsen realization problem. Bull. Amer. Math. Soc. (2) **2**, 452–454 (1980)
6. MacLane, S.: Homology. Berlin-Heidelberg-New York: Springer 1963
7. Marden, A.: Geometrically finite Kleinian groups and their Deformation Spaces. In: Discrete groups and automorphic functions (W. Harvey, ed.). New York-London: Academic Press 1977
8. Mostow, G.D.: Quasi-conformal mappings in n-space and the rigidity of hyperbolic space forms. Inst. Hautes Etudes Sci. Publ. Math. **34**, 55–104 (1968)
9. Nielsen, J.: Abbildungsklassen endlicher Ordnung. Acta Math. **75**, 24–115 (1942)
10. Orlik, P., Vogt, E., Zieschang, H.: Zur Topologie dreidimensionaler gefaserter Mannigfaltigkeiten. Topology **6**, 49–64 (1967)
11. Peczynski, N., Zieschang, H.: On coverings of Seifert 3-manifolds. Arch. Math. (Basel) **31**, 382–386 (1978)
12. Serre, J.P.: Arbres, amalgames et SL_2. Astérisque **46**. Paris: Société Mathématique de France 1977
13. Thurston, W.: The geometry and topology of 3-manifolds. Lecture Notes, Princeton University
14. Waldhausen, F.: On irreducible 3-manifolds which are sufficiently large. Ann. of Math. (2) **87**, 56–88 (1968)
15. Waldhausen, F.: Gruppen mit Zentrum und 3-dimensionale Mannigfaltigkeiten. Topology **6**, 505–517 (1967)
16. Wolf, J.A.: Spaces of constant curvature. New York: McGraw-Hill (1967)

17. Zieschang, H.: On extensions of fundamental groups of surfaces and related groups. Bull. Amer. Math. Soc. **77**, 1116–1119 (1971)
18. Zieschang, H., Zimmermann, B.: Endliche Gruppen von Abbildungsklassen gefaserter 3-Mannigfaltigkeiten. Math. Ann. **240**, 41–62 (1979)
19. Zieschang, H., Zimmermann, B.: Über Erweiterungen von $\mathbb{Z}$ und $\mathbb{Z}_2 * \mathbb{Z}_2$ durch nichteuklidische kristallographische Gruppen. Math. Ann. (erscheint demnächst)
20. Zimmermann, B.: Periodische Homöomorphismen Seifertscher Faserräume. Math. Z. **166**, 289–297 (1979)
21. Zimmermann, B.: Über Gruppen von Homöomorphismen Seifertscher Faserräume und flacher Mannigfaltigkeiten. Manuscripta Math. **30**, 361–373 (1980)

Eingegangen am 27. März 1981; in endgültiger Form am 5. Februar 1982

Math. Z. 180, 361–371 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On Topological Stability of Divergent Diagrams of Folds

Marco Antonio Teixeira*

Instituto de Matemática, Estatística e Ciência da Computação, Universidade Estadual de Campinas, IMECC-UNICAMP C.P. 1170, 13100 Campinas, Sao Paulo, Brasil

The objective of this paper is to study a class of the diagram of C^∞ mappings via the structural stability of diffeomorphisms.

The study of divergent diagrams of mappings of type $P \xleftarrow{f} M \xrightarrow{g} Q$ is in fact the study of singularities of mappings of type $M \to P \times Q$ with the product structure of the target preserved. Diagrams of type $\mathbb{R} \leftarrow \mathbb{R} \to \mathbb{R}$, $\mathbb{R} \leftarrow \mathbb{R}^n \to \mathbb{R}^n$, $\mathbb{R} \leftarrow \mathbb{R}^2 \to \mathbb{R}^2$ have been discussed by several authors including Arnold [1] and Dufour [2]. In addition [7] contains a study on local and simultaneous structural stability of a pair of certain diffeomorphisms, in such a way that there is an application to the local theory of divergent diagrams of type $\mathscr{D}$: $\mathbb{R}^2 \xleftarrow{f} \mathbb{R}^2 \xrightarrow{g} \mathbb{R}^2$. In fact it says that the topological stability of $\mathscr{D}$ is never generic, under C^2 perturbations of f and g. Roughly speaking, if 0 is a fold point of both mappings f and g, then there are associated to them diffeomorphisms φ_f and φ_g (the symmetrics of f and g respectively) satisfying $f \circ \varphi_f = f$ and $g \circ \varphi_g = g$. One can show that if $\mathscr{D}$ is topologically stable then the pair (φ_f, φ_g) is simultaneously stable in the following sense: if $\tilde{f}$ and $\tilde{g}$ are close to f and g respectively then there is a homeomorphism $h: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ such that $h \circ \varphi_f = \varphi_{\tilde{f}} \circ h$ and $h \circ \varphi_g = \varphi_{\tilde{g}} \circ h$.

Our goal in this paper is to study the Local Topological Stability of diagrams of type

$$\mathscr{D}^*: \mathbb{R}^2, 0 \xleftarrow{f} \mathbb{R}^2, 0 \xrightarrow{g} \mathbb{R}^2, 0$$

where 0 is a fold point of C^∞ mappings f and g with the following additional hypothesis: the singular sets of both, f and g, coincide. We mention that there is a relationship between the simultaneous dynamic behavior of the respective diffeomorphisms associated to those mappings and the topological stability of $\mathscr{D}^*$. One also can observe that the results obtained here can have applications to the study of structural stability of ordinary differential equations on $\mathbb{R}^3$ when we allow non-uniqueness of solutions. More specifically,

* This research was supported by CNPq Brasil under Contract 40.0652/80

0025-5874/82/0180/0361/$02.20

consider a class of vector fields Z on $\mathbb{R}^3$, given by

$$Z(x, y, z) = \begin{cases} X(x, y, z) & \text{if } z > 0 \\ Y(x, y, z) & \text{if } z < 0 \end{cases}$$

such that $X^3(x, y, 0) = Y^3(x, y, 0)$ where $X = (X^1, X^2, X^3)$ and $Y = (Y^1, Y^2, Y^3)$ are C^∞ vector fields on $\mathbb{R}^3$. Consider now the following generic situation: the trajectories of X and Y passing through $0 = (0, 0, 0)$ have a 2^{nd} order contact with the plane $z = 0$. So the field Z has a fold of contact points with $z = 0$. Now recall the construction done by Sotomayor in [5, pp. 172 and 173] where we can see that both fields X and Y induce two C^∞ mappings f_X, f_Y: $\mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ respectively. Finally, we can say that the stability of the diagram

$$\mathbb{R}^2, 0 \xrightarrow[f_X]{} \mathbb{R}^2, 0 \xrightarrow[f_Y]{} \mathbb{R}^2, 0$$

is closely related with the structural stability of Z in 0.

The main result of the paper is

Theorem. a) *The topological stability of the diagrams $\mathscr{D}^*$ is generic.*

b) *We classify the topologically stable diagrams $\mathscr{D}^*$ by exhibiting their normal forms.*

In Sect. 1 we give definitions and recall standard facts.

In Sect. 2 we define the G-singularity of a fold and a G-diagram. Moreover, we analyse the associated diffeomorphism to that fold.

Section 3 contains the study of Simultaneous Structural Stability of the diffeomorphisms associated to $\mathscr{D}^*$. We have used a very interesting result obtained by Dumortier-Rodrigues-Roussarie (see [3]) concerning C^∞ embeddings of diffeomorphisms in vector fields.

Finally, Sect. 4 contains the main result of the paper. We observe that its proof is a corollary of the result contained in Sect. 3.

It is a pleasure to acknowledge useful conservations regarding this work with J. Palis and I. Malta. I am especially grateful to F. Dumortier who gave me a private talk on the subject matter of [3]. Finally, I want to thank the Mathematics Institute of the University of Warwick for the hospitality they offered to me during the time I prepared this paper.

1. Preliminaries

We will use the germ notation of the functions treated here. For simplicity, we can make no distinction between a germ of a function (at 0) and anyone of its representatives. We shall also use the concept and the usual notation of a k-jet of a function.

1.1. **Definition.** An *involution* is a germ of a C^∞ diffeomorphism φ: $\mathbb{R}^2$, $0 \to \mathbb{R}^2$, 0 satisfying $\varphi(0) = 0$, $\varphi \circ \varphi = \mathrm{Id}$ and $\det[\varphi'(0)] = -1$.

Let $\mathrm{Fix}\, \varphi = \{x \in \mathbb{R}^2;\ \varphi(x) = x\}$.

Denote by G^r the space of germs of involutions with the C^r topology. Let B^r be the subspace of $G^r \times G^r$ consisting of the pairs (φ_0, φ_1) satisfying $\mathrm{Fix}(\varphi_0) = \mathrm{Fix}(\varphi_1)$.

1.2. **Definition.** Let $\varphi=(\varphi_0,\varphi_1)$, $\tilde{\varphi}=(\tilde{\varphi}_0,\tilde{\varphi}_1)\in B^r$. Then φ and $\tilde{\varphi}$ are *topologically equivalent* if there is a homeomorphism $h\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ which satisfies $h\circ\varphi_0=\tilde{\varphi}_0\circ h$ and $h\circ\varphi_1=\tilde{\varphi}_1\circ h$. So the (local) structural stability in B^r is defined in a natural way.

If $f\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ is a fold we know that there is an involution φ_f associated to f which satisfies $f\circ\varphi_f=f$. Moreover, if there are a fold $g\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ and homeomorphisms $h,k\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ such that $g\circ h=k\circ f$ then h obviously satisfies $\varphi_g\circ h=h\circ\varphi_f$. We can still say that if φ is an involution then there is a fold $f\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ such that $\varphi=\varphi_f$ (see [7, p. 5]).

Let $\mathscr{D}$ be the space of the local divergent diagrams of C^∞ mappings as follows:

$$D\colon \mathbb{R}^2,0 \xleftarrow[f]{} \mathbb{R}^2,0 \xrightarrow[g]{} \mathbb{R}^2,0$$

with $S(f)=S(g)$, where $S(f)$ means the singular set of f. We are assuming that $0\in S(f)$.

We are considering $\mathscr{D}$ endowed with the C^r-topology and $r>1$.

The concept of topological equivalence in $\mathscr{D}$ is reached by the following definition:

1.3. **Definition.** We say that

$$D\colon \mathbb{R}^2,0 \xleftarrow[f]{} \mathbb{R}^2,0 \xrightarrow[g]{} \mathbb{R}^2,0$$

is C^0-*equivalent* (or *topologically equivalent*) *to*

$$\tilde{D}\colon \mathbb{R}^2,0 \xleftarrow[\tilde{f}]{} \mathbb{R}^2,0 \xrightarrow[\tilde{g}]{} \mathbb{R}^2,0$$

if there exist homeomorphisms k_1, k_2, $h\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ such that $\tilde{f}\circ h=k_1\circ f$ and $\tilde{g}\circ h=k_2\circ f$.

Let J be the set of homeomorphisms $h\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ satisfying $h(x,y)=(h^1(x,y), h^2(x,y))$ and $h(x,-y)=(h^1(x,y),-h^2(x,y))$.

Call by $\mathscr{D}^*$ the subset of $\mathscr{D}$ consisting of those diagrams

$$D\colon \mathbb{R}^2,0 \xleftarrow[f]{} \mathbb{R}^2,0 \xrightarrow[g]{} \mathbb{R}^2,0$$

such that 0 is a fold point of both mappings f and g.

2. The G-Diagram

Let $g\colon \mathbb{R}^2,0\to\mathbb{R}^2,0$ be a C^∞ mapping such that 0 is a fold point of it and its singular set is $S(g)=\{(x,y);\, y=0\}$. If $g=(\alpha,\beta)$ we are going to distinguish the following cases:

2i) Regular case: $\frac{\partial \alpha}{\partial y}(0,0) \neq 0$.

2ii) Singular case: $\frac{\partial \alpha}{\partial y}(0,0)=0$, $\frac{\partial^2 \alpha}{\partial x \partial y}(0,0) \neq 0$ and $\frac{\partial^3 \beta}{\partial y^3}(0,0) \neq 0$.

We remark that, the regular case means that Kern$(g'(0))$ is transverse to the y-axis at 0; in other words, if $\theta=\theta(x)$ is the angle between Kern$(g'(x,0)$ and the y-axis (with some fixed orientation) then the above case 2i is roughly expressed by $\theta(x)=\theta_0 \neq 0$. The singular case is a quasi-transversality condition, expressed as above by $\theta(x)=\theta_0 x$ with $\theta_0 \neq 0$ plus a generic condition.

2.1. **Definition.** $0 \in \mathbb{R}^2$ is a *G-singularity of* g if g satisfies either condition 2i) or condition 2ii) given above.

2.2. **Definition.** We say that $D \in \mathscr{D}^*$ is a *G-diagram* if D is C^0-equivalent to

$$D_0\colon \mathbb{R}^2,0 \xleftarrow{f} \mathbb{R}^2,0 \xrightarrow{g} \mathbb{R}^2,0$$

where $f(x,y)=(x,y^2)$ and 0 is a G-singularity of g.

2.3. *Study of the Regular Case*

2.3a. **Lemma.** *If g satisfies the above condition* 2i) *then there exists a pair of diffeomorphisms $h,k\colon \mathbb{R}^2,0 \to \mathbb{R}^2,0$ such that $h \in J$ and $g_0 \circ h = k \circ g$ where $g_0(x,y)=(x+y, y^2+y^3\lambda(x,y))$.*

Proof. At first, consider α written in the following form

$$\alpha(x,y)=x\,\eta(x,y^2)+y\,\mu(x,y^2)$$

with $a=\eta(0,0) \neq 0$ and $b=\mu(0,0) \neq 0$.

By the diffeomorphism $h_1 \in J$ given by

$$h_1(x,y)=(x\,\eta(x,y^2), y\,\mu(x,y^2))$$

we can take α as being $\alpha(x,y)=x+y$. So

$$g'(x,0)=\begin{pmatrix} 1 & 1 \\ \frac{\partial \beta}{\partial x}(x,0) & \frac{\partial \beta}{\partial y}(x,0) \end{pmatrix}.$$

Since rank$[g'(0,0)]=1$ we deduce that $g(\Sigma(g))$ is a one dimensional submanifold of $\mathbb{R}^2,0$. By the Implicit Funtion Theorem we can get a diffeomorphism $k\colon \mathbb{R}^2,0 \to \mathbb{R}^2,0$ (in the target of g) such that

$$(k \circ g)|_{\Sigma(g)}=\mathrm{Id}.$$

Hence $\beta(x,0)=\frac{\partial \beta}{\partial x}(x,0)=0$, $\beta(x,y)=y\,A(x,y)$, $\frac{\partial \beta}{\partial y}(x,0)=A(x,0)$ and $A(x,0)=\det(g'(x,0))=0$.

So we have $g(x,y)=x+y,\ y^2\theta(x,y)$.

Of course, we can write

$$\theta(x, y) = \theta_1(x, y) + y\,\theta_2(x + y, y).$$

The assumption $\dfrac{\partial^2 \beta}{\partial y^2}(0,0) \neq 0$ implies $\theta_1(0) \neq 0$.

By a trivial change of coordinates in the target of g we get

$$g(x, y) = (x + y, y^2 + y^3\,\lambda(x, y))$$

and consequently Lemma 2.3a is true. □

A direct calculation permits us to prove that:

2.3b. **Lemma.** i) *The involution associated to* $g_0(x, y) = (x + y, y^2 + y^3\,\lambda(x, y))$ *has the form* $\varphi_0(x, y) = (x + 2y - y^2\,a(x, y), -y + y^2\,a(x, y))$.

ii) *If* φ *is the involution associated to* g *(see Lemma* 2.3a*) then there is a diffeomorphism* $h \in J$ *which is an equivalence between* φ_0 *and* φ.

2.4. *Study of the Singular Case*

2.4a. **Lemma.** *If* g *satisfies the above conditions* 2ii) *then there exists a pair of diffeomorphisms* $h, k: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ *such that* $h \in J$, $g_1 \circ h = k \circ g$ *where* $g_1(x, y) = (x + \alpha_0\,x\,y + y^2,\ y^2 + y^3\,\lambda(x, y))$ *with* $\lambda_0 = \lambda(0,0) \neq 0$ *and* $\alpha_0 \neq 0$.

The proof of this lemma is an immediate consequence of Proposition 1.8 of [2, p. 98].

By a straightforward calculation one shows that:

2.4b. **Lemma.** i) *The involution* φ_1 *associated to*

$$g_1 = (x + \alpha_0\,x\,y + y^2,\ y^2 + y^3\,\lambda(x, y))$$

has the form

$$\varphi_1(x, y) = ([x + \alpha_0\,x\,y - 2y^3\,\lambda(x, y) - y^4(\lambda(x, y)^2]/(1 - y - y^2\,\lambda(x, y)),\ -y - y^2\,\lambda(x, y))$$

with $\lambda_0 = \lambda(0,0) \neq 0$ *and* $\alpha_0 \neq 0$.

ii) *There is a diffeomorphism* $h \in J$ *which is an equivalence between* φ_1 *and the involution associated to* g.

3. Simultaneous Stability of Involutions

3.1. *The Regular Case*

Consider the involution $\varphi(x, y) = (x + 2y - y^2\,a(x, y), -y + y^2\,a(x, y))$ given in Lemma 2.3b and the diffeomorphism $\psi = \theta \circ \varphi$ where $\theta(x, y) = (x, -y)$.

So $\psi(x, y) = (x + 2y - y^2\,a(x, y), y - y^2\,a(x, y))$.

Observe that $\psi^n(x, y) = (x, y) + 2n(y, 0) + H(x, y)$ with $H(x, y) = 0|(x, y)|^2$.

Consider the following functions:

$$p(x, y)=x+2y-y^2 a(x, y) \quad \text{and} \quad q(x, y)=y-y^2 a(x, y).$$

3.1a. **Lemma.** *There exists a neighborhood U of 0 in $\mathbb{R}$ such that for each $t\in U$ there is a germ of C^∞ function $\rho_t: \mathbb{R}, 0\to\mathbb{R}, t$ satisfying $\rho_t(0)=t$ and $q(x, \rho_t(x))=\rho_t(p(x, \rho_t(x)))$.*

Proof. At first, we are going to show that the diffeomorphism $\psi=(p, q)$ can be C^∞ imbedded in a flow X_t (X_t denotes the flow of some C^∞ vector field X; in this case $j_1(X)(0)=y\dfrac{\partial}{\partial x}$. After this, the present proof is immediate.

Due to the symmetry of φ with respect to the x-axis, we may restrict our analysis for $y\geqq 0$.

If we blow up ψ, by means of polar coordinates, we see that it has on $0\times\{s^1\}$, only two singularities which are $p_1=(0, 0)$ and $p_2=(0, \pi)$. The 2-jet of φ at these points takes the form

$$(r, \theta)\to(r+2r\theta, \theta\pm 2\theta^2).$$

So in neighborhoods of p_1 and p_2, ψ can be C^∞ imbedded in a flow (see for instance, the 11th singularity listed in [3], pp. 18, 19). Moreover, following the techniques developped by Dumortier-Rodrigues-Roussarie in [3], we see that there are no obstructions to getting the required imbedding in a neighborhood of $0\times\{s^1\}$. This finishes the proof. □

Let $E_t=\mathrm{Graf}(\rho_t)$ be the graffic of ρ_t.

Take now a very small positive number t_0. Let R be the region bounded by the following curves: $J_0=E_t$, $I_0=[0, t_0]$ and $I_1=\psi(I_0)$. (See Fig. 1.)

Observe that the region $R^-=\theta(R)$ is bounded by $J_0^-=E_{-t_0}$, $I_0^-=[0, -t_0]$ and $I_1^-=\psi^{-1}(I_0^-)$.

Furthermore $\psi(R^-)$ is in the region $x\leqq 0$, $y\leqq 0$, $\theta(\psi(R^-))$ is in the region $x\leqq 0$, $y\geqq 0$ and finally $\psi(\theta(\psi(R^-)))=R$.

If $p=(x, 0), (x>0)$ is close to the origin, we claim that, given $\varepsilon>0$ there exist $\delta=\delta(\varepsilon)$, $n=n(\varepsilon)$ and $p_n=(x_n, y_n)\in R$ such that $|y_n|<\delta$ and $|\psi^n(p_n)-p|<\varepsilon$. To see this it is enough to observe that

$$\frac{|q(x, y)-y|}{p(x, y)-x|}<y \quad \text{if } y\geqq 0.$$

So associated to each $(x, 0)$ there exists a sequence in R, $p_n=(x_n, y_n)\to(0, 0)$ in such a way that $|y_{n+1}|<|y_n|$, $|\psi^{n+1}(p_{n+1})<|\psi^n(p_n)|$ and $\psi^n(p_n)\to(x, 0)$.

In other words we are saying that R is a fundamental domain of ψ. So it is not difficult to find a homeomorphism $h: \mathbb{R}^2, 0\to\mathbb{R}^2, 0$ satisfying $h\circ\theta=\theta\circ h$ and $h\circ\psi=\psi_0\circ h$ where $\psi_0(x, y)=(x+2y, y)$. We add that the homeomorphism defined in the fundamental domain must be taken as being close to identity. Finally we have the following result.

3.1b. **Lemma.** *There is a homeomorphism $h\in J$ such that $h\circ\varphi=\varphi_0\circ h$ where $\varphi(x, y)=(x+2y-y^2 a(x, y)-y+y^2 a(x, y))$ and $\varphi_0(x, y)=x+2y, y)$.*

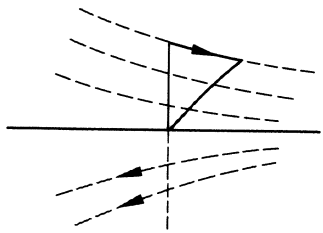


Fig. 1. The fundamental domain of ψ. The Regular Case

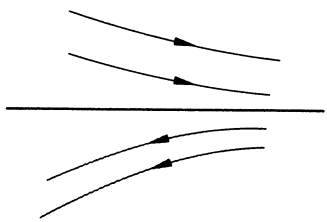


Fig. 2. The dynamic of ψ. The Regular Case

3.2. *The Singular Case*

Consider the involution φ_1 given in Lemma 2.3b and the diffeomorphism $\psi = \theta \circ \varphi_1$ where $\theta(x, y) = (x, -y)$.

We may assume, without loss of generality that $\lambda_0 < 0$. So $\psi = (p, q)$, where

$$p(x, y) = x + \alpha_0 x y - 2y^3 \lambda(x, y) - y^4(\lambda(x, y))^2/(1 - y - y^2 \lambda(x, y))$$

and

$$q(x, y) = y + y^2 \lambda(x, y).$$

The proof of the next lemma is an immediate consequence of the Implicit Function Theorem:

3.2.1. **Lemma.** *Consider $\psi = (p, q)$. Then there is a differentiable function $x = v(y)$ with $v(0) = 0$ and $p(x, y) = 0$ if and only if $x = v(y)$. Moreover $v'(0) = 0$.*

3.2.2. *Remark.* We can check by a simple inspection that for each $y > 0$ (resp. $y < 0$) the function $u_y = p(x, y) - x$ increases (resp. decreases) for $\alpha_0 > 0$ and decreases (resp. increases) for $\alpha_0 < 0$. It is not difficult to recognize that if $\alpha_0 < 0$ then $\psi^n(x, y) \to (0, 0)$ when $y > 0$ and $\psi^{-n}(x, y) \to (0, 0)$ when $y < 0$ (see Fig. 5).

We have to analyse the case $\alpha_0 > 0$ because we cannot say, a priori, what happens in the following region:

$$A = \{(x, y), v(y) < x < 0\}.$$

In this case we are going to show that φ has a unique invariant manifold (see Figs. 3 and 4). We apply here the method of blowing up singularity and use

a result of Dumortier-Rodrigues-Roussarie [3]. More specifically, consider the following steps:

1. We see that $j_1(\psi)(0)=\mathrm{Id}$. Because of a formal normal form theorem of Takens [6, pp. 1-59] we know the existence of a unique ∞-jet of a vector field X such that $\tilde{\psi}=\tilde{X}_1$ where $\tilde{\psi}=j_\infty(\psi)(0)$, X_1 is the time-one mapping of X and $\tilde{X}_1=j_\infty(X_1)(0)$.

2. We blow up X as well as ψ by means of the following mapping $\rho\colon(r,\theta)\to(r\cos\theta, r\sin\theta)$. We have $\rho_x(\tilde{X})\simeq X\circ\rho$ and $\rho\circ\tilde{t}=\psi\circ\rho$.

Along $\{r=0\}=S^1\times\{0\}$ we have, because of the 1st step that $j_\infty(\tilde{\psi})(u)=j_\infty(\tilde{X}_1)(u)$ for every $u\in S_1\times\{0\}$.

3. If we write $\bar{X}=\frac{1}{r}\tilde{X}$ then we see that $\bar{X}$ has on $S^1\times\{0\}$ only singularities in $\theta=0,\ \pi$ and $\theta=\pi/2, 3\pi/2$.

We see that $\bar{X}$ has in $\pi/2$ (resp. $3\pi/2$) a hyperbolic singularity with an unstable manifold (resp. stable manifold) inside $S^1\times\{0\}$.

Now because of the fundamental theorem of Dumortier, Rodrigues and Roussarie (Theorem 3.1 of [3]), we know the existence of a C^∞ diffeomorphism $\hat{f}$ defined in the neighborhood of $S^1\times\{0\}$ with $j_\infty(\hat{f})(u)=0$ for every $u\in S_1\times\{0\}$ such that f is a C^∞ conjugacy in between $\hat{\varphi}$ and $\hat{X}_1$ restricted to the following region

$$R_1=\{(r,\theta);\ r\geqq 0 \text{ and } \theta_0\leqq\theta\lessapprox\pi\sim\theta_0 \text{ or } \pi+\theta_0\leqq\theta\leqq 2\pi-\theta_0\}$$

where we may choose $\theta_0>0$ as small as we want. (See Fig. 3.)

If we define $f\colon\mathbb{R}^2,0\to\mathbb{R}^2,0$ to be such that $\rho\circ\hat{f}=f\circ\rho$ we find a C^∞ diffeomorphism on some neighborhood of 0 such that f is a C^∞ conjugacy in between ψ and X_1 on $\psi(R_1)$. This means that ψ on $\rho(R_1)$ C^∞ imbeds in a flow. (May be seen, up to the coordinates change f, as the time-one mapping of X on $\rho(R_1)$.)

The dynamic of the vector field of X can be seen in the following figure

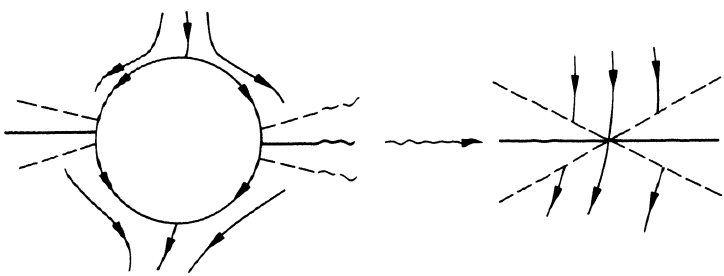


Fig. 3. Blowing up ψ. The singular case and $\alpha_0<0$

But now, we observe that we know the behavior of ψ off $\rho(R_1)$. So the dynamic of ψ can be seen in the following figure:

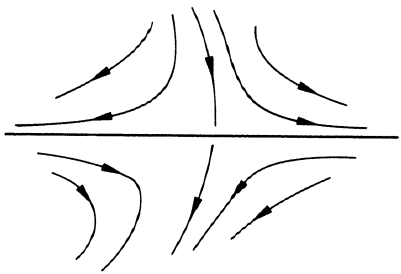


Fig. 4. The dynamic of ψ. The singular case and $\alpha_0 > 0$

3.2.3. *Remark.* The technique above described can also be used to the case $\lambda_0 > 0$ to obtain the behavior of ψ.

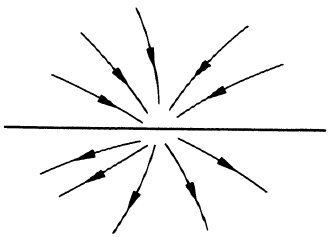


Fig. 5. The dynamic of ψ. Singular case and $\alpha_0 < 0$

To complete our analysis we would have to find a fundamental domain of ψ. For instance, we give the following rough idea:

a) for $\alpha_0 > 0$, a fundamental domain of ψ is the region bounded by L and $\psi(L)$ where L is a horizontal line intersecting the invariant submanifold of ψ at a point (x, y) with $y > 0$;

b) for $\alpha_0 < 0$, a fundamental domain of ψ can be the region bounded by L and $\psi(L)$ where L is a semi-circle in $y \geqq 0$ joining two points $(-\varepsilon, 0)$ and $(\varepsilon, 0)$ for ε positive and small enough.

The later, analysis allows us to find a homeomorphism $h: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ such that $h \circ \theta = \theta \circ h$ and $h \circ \psi = \tilde{\psi} \circ h$ where $\theta(x, y) = (x, -y)$, $\tilde{\psi} = \theta \circ \tilde{\varphi}$ and (θ, φ) is close to $(\theta, \tilde{\varphi})$ in B^r.

Given a diagram $D: \mathbb{R}^2, 0 \xleftarrow[f]{} \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ in $\mathscr{D}^*$ we may, of course, assume that $f(x, y) = (x, y^2)$. *If* φ is the involution associated to g we are able to state the following result:

3.3. **Proposition.** *(θ, φ) is structurally stable in B^r if and only if 0 is a G-singularity of g. Moreover, its normal forms are:*

$$\text{Type} G_1 - \theta(x, y) = (x, -y) \quad \textit{and} \quad \varphi(x, y) = (x + 2y, -y).$$

$$\text{Type} G_2 - \theta(x, y) = (x, -y) \quad \textit{and}$$

$$\varphi(x, y) = ((x + \alpha_0 x y - 2\lambda_0 y^3 - \lambda_0^2 y^4)/(1 - y - \lambda_0 y^2), \ -y - \lambda_0 y^2)$$

with $\lambda_0 = \pm 1$, *and* $\alpha_0 = \pm 1$.

4. Stability of Diagrams

Consider now, as before, folds $f, g: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ and let $\varphi, \tilde{\varphi}$ be their respective involutions. Suppose there is a homeomorphism $h: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$ such that $\varphi \circ h = h \circ \tilde{\varphi}$.

Define:

a) the following equivalence relations on $\mathbb{R}^2, 0$ by

i) $x S y$ if $\varphi(x) = y$.

ii) $x \tilde{S} y$ if $\tilde{\varphi}(x) = y$.

b) The following spaces by

$$E = \mathbb{R}^2/S, \quad \tilde{E} = \mathbb{R}^2/\tilde{S}, \quad I = \operatorname{Im} f \quad \text{and} \quad \tilde{I} = \operatorname{Im} g,$$

c) and the following homeomorphisms by

$$\hat{h} = h/E, \quad \hat{f} = f/E, \quad \hat{g} = g/\tilde{E} \quad \text{and} \quad \hat{k} = \hat{g} \circ \hat{h} \circ \hat{f}^{-1}.$$

Now, extend $\hat{k}$ some how to a homeomorphism $k: \mathbb{R}^2, 0 \to \mathbb{R}^2, 0$.

Of course, k satisfies $g \circ h = k \circ f$; this means that f and g are topologically equivalent.

The proof of the following lemma is trivial.

4.1. **Lemma.** *Let* $D: \mathbb{R}^2 \xleftarrow{f} \mathbb{R}^2 \xrightarrow{g} \mathbb{R}^2$ *and* $\tilde{\mathcal{D}}: \mathbb{R}^2 \xleftarrow{\tilde{f}} \mathbb{R}^2 \xrightarrow{\tilde{g}} \mathbb{R}^2$ *be two diagrams in* $\tilde{D}$ *and* $\theta, \varphi, \tilde{\theta}, \tilde{\varphi}$ *be the involution associated to* $p, g, \tilde{f}, \tilde{g}$ *respectively. If* (θ, φ) *is equivalent to* $(\tilde{\theta}, \tilde{\varphi})$ *in* B^r *then* D *is* C^0*-equivalent to* $\tilde{D}$ *in* $\tilde{\mathcal{D}}$.

The proof of the following result follows from Proposition 3.3 and Lemma 4.1

4.2. **Theorem.** 1. *A diagram* $D: \mathbb{R}^2 \xleftarrow{f} \mathbb{R}^2 \xrightarrow{g} \mathbb{R}^2$ *is* C^0*-stable in* $\mathcal{D}^*$ *if and only if* D *is a G-diagram.*

2. *The normal forms of a G-diagram are:*

$$G_1 - f(x, y) = (x, y^2) \quad \text{and} \quad g(x, y) = (x + y, y^2);$$

$$G_2 - f(x, y) = (x, y^2) \quad \text{and} \quad g(x, y) = (x + \alpha_0 x y + y^2, y^2 + \lambda_0 y^3)$$

with $\lambda_0 = \pm 1$ *and* $\alpha_0 = \pm 1$.

References

1. Arnold, V.F.: Evolution of Wave Fronts and Equivariant Morse Lemma. Comm. Pure Appl. Math. **29**, 557–582 (1976)
2. Dufour, J.P.: Diagrammes d'Applications Différentiables. Thesis, University of Montpellier, France (1979)
3. Dumortier, F., Rodrigues, P., Roussarie, R.: Germs of Diffeomorphisms in the Plane. Lecture Notes in Mathematics **902**. Berlin-Heidelberg-New York: Springer 1981

4. De Mello, W., Palis, J.: Introdução aos Sistemas Dinâmicos. Rio de Janeiro: Instituto de Matemática Pura e Applicada (1978)
5. Sotomayor, J.: Structural Stability in Manifolds with Boundary. In: Global Analysis and its Applications, Vol. III. Lectures presented at an International Seminar Course (Trieste 1972), pp. 167–176. Wien: International Atomic Energy Agency 1974
6. Takens, F.: Forced Oscillations and Bifurcations. In: Applications of Global Analysis I. Proceedings of a Symposium (Utrecht 1973), pp. 1–59. Comm. Math. Inst. Rijksuniv. Utrecht **3**. Utrecht: Mathematical Institute of the Rijksuniversiteit 1974
7. Teixeira, M.: Local and Simultaneous Structural Stability of Certain Diffeomorphisms. In: Dynamical Systems, Stability and Turbulence, Warwick 1980. Proceedings of a Symposium (Warwick 1979/1980), pp. 382–390. Lecture Notes in Mathematics **898**. Berlin-Heidelberg-New York: Springer 1981

Received March 16, 1981; received in final form January 11, 1982

Math. Z. 180, 373–378 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

The Dual of a Space of Vector Measures

Michael Cambern[1] and Peter Greim[2]

[1] Department of Mathematics, University of California, Santa Barbara, California 93106, U.S.A.
[2] Mathematisches Institut der Freien Universität, Arnimallee 2–6, D-1000 Berlin 33, Federal Republic of Germany

1. Introduction

Throughout this article, the letters X and Z will denote compact Hausdorff spaces, and E a Banach space. Given E, K will denote the one-dimensional Banach space consisting of the corresponding field of scalars. The interaction between elements of a Banach space and those of its dual is denoted by $\langle \cdot, \cdot \rangle$. We will write $E_1 \simeq E_2$ to indicate that the Banach spaces E_1 and E_2 are isometric.

Given X, $\mathscr{B}$ will denote the σ-algebra of Borel subsets of X, and $\mathscr{B}_0$ the σ-algebra of Baire sets. $C(X, E)$ denotes the space of continuous functions on X to E provided with the supremum norm. And $C(X, (E^*, \sigma^*))$ stands for the Banach space of continuous functions G on X to E^* when this latter space is provided with its weak * topology, again normed by $\|G\|_\infty = \sup_{x \in X} \|G(x)\|$.

If (Y, Σ) is a measurable space, then $m(Y, \Sigma, E)$ will denote the space of all E-valued measures m on Σ with finite variation $|m|$. The goal of this article is to provide a characterization of the dual of $m(Y, \Sigma, E)$ and describe the interaction between the elements of $m(Y, \Sigma, E)$ and those of the dual via an integration. This problem has been studied by Mauldin [11], who showed that, assuming that E has the Radon-Nikodym property, that the continuum hypothesis holds, and that the cardinality of $m(Y, \Sigma, K)$ is $\leqq 2^{\aleph_0}$, then the dual of $m(Y, \Sigma, E)$ can be isometrically represented as a space of bounded functions on Σ to E^* in such a way that the interaction between elements of $m(Y, \Sigma, E)$ and those of $m(Y, \Sigma, E)^*$ is indeed given by an integral.

What we seek here is a rather different characterization of the dual. We wish to represent $m(Y, \Sigma, E)^*$ as a space of continuous functions defined on a compact Hausdorff space, in analogy with the result for $m(Y, \Sigma, K)^*$ given by Gordon [8], and previously obtained, for the case in which the space of measures arises as the dual of $C(X, K)$, by Kakutani [9], Arens [1], and Kaplan [10] among others. The characterization we give represents $m(Y, \Sigma, E)^*$ as a space $C(X, (E^*, \sigma^*))$ whenever E has the Radon-Nikodym property, and

0025-5874/82/0180/0373/$01.20

does away with assumptions concerning the continuum hypothesis and the cardinality of $m(Y, \Sigma, K)$. When the space of vector measures is of the form $m(X, \mathscr{B}_0, E^*)$ and is hence the dual of a space $C(X, E)$, [14], or [5, p. 354 and p. 397], the fact that $m(X, \mathscr{B}_0, E^*)^* \simeq C(Z, (E^{**}, \sigma^*))$ for some compact Hausdorff space Z was proved by the authors in [3]. However, in that article, no attempt was made to describe the interaction between measures and elements of the dual space - in fact, such interaction is somewhat obscured by the sequence of isometries used to obtain the principal result of [3]. Here we make this interaction explicit.

Much of what we do in this article is dependent upon the notion of a perfect measure. If X is an extremally disconnected compact Hausdorff space then, following [2], we will call a nonnegative, extended real-valued Borel measure μ on X *perfect* if

(i) every nonempty clopen set has positive measure,

(ii) every nowhere dense Borel set has measure zero, and

(iii) every nonempty clopen set contains another nonempty clopen set with finite measure.

Property (iii), together with an application of Zorn's lemma, can be used to establish that such a μ is *strictly localizable*, by which we mean that X is the disjoint union of measurable subsets, $X = \bigcup_{i \in I} X_i$, with $\mu(X_i) < \infty$ for all i, and for all $B \in \mathscr{B}$, $\mu(B) = \sum_{i \in I} \mu(B \cap X_i)$. A fact which we shall need is that if μ is such a measure, then $L^1(X, \mathscr{B}, \mu, K)^* \simeq L^\infty(X, \mathscr{B}, \mu, K)$, [6, p. 290].

If μ is a perfect measure on the Borel sets of an extremally disconnected space X, then for any Banach space E we denote by $M(\mu, E)$ the subspace of $m(X, \mathscr{B}, E)$ consisting of all $m \in m(X, \mathscr{B}, E)$ which are absolutely continuous with respect to μ. And we remark, for future reference, that if E has the Radon-Nikodym property, then $M(\mu, E) \simeq L^1(X, \mathscr{B}, \mu, E)$. For although the Radon-Nikodym property is usually formulated with respect to a finite measure space, [4, p. 61], the fact that μ is strictly localizable gives the desired result. For clearly the indefinite integral defines an isometric embedding of $L^1(X, \mathscr{B}, \mu, E)$ into $M(\mu, E)$. And if X is the disjoint union $\bigcup_{i \in I} X_i$ of measurable sets X_i with $\mu(X_i) < \infty$ for all i, then for $m \in M(\mu, E)$, since $|m|$ is finite, we can have $m(X_i) \neq 0$ at most for i belonging to a countable subset $I' \subseteq I$. For $i \in I'$, the Radon-Nikodym derivative of m with respect to $\mu|_{X_i}$ can be defined, and the resulting functions can then be pasted together in the obvious way to obtain $\frac{dm}{d\mu}$.

Throughout the article scalar measures are denoted by μ, ν, and λ, while vector measures are denoted by m and n. Facts about vector measures used in this paper can be found in [4] and [5]. The authors are indebted to J.J. Uhl, Jr., for having called their attention to the work of Mauldin.

2. The Dual of an L^1-space

Theorem 1. *Let E be a Banach space, X an extremally disconnected compact Hausdorff space, and μ a perfect measure on $\mathscr{B}$. Then $L^1(X, \mathscr{B}, \mu, E)^*$*

$\simeq C(X,(E^*,\sigma^*))$. *Specifically, each* $G\in C(X,(E^*,\sigma^*))$ *gives rise to an element* $\Phi_G\in L^1(X,\mathscr{B},\mu,E)^*$ *determined by* $\langle F,\Phi_G\rangle=\int\langle F,G\rangle\,d\mu$, *for* $F\in L^1(X,\mathscr{B},\mu,E)$, *with* $\|\Phi_G\|=\|G\|_\infty$, *and every element of* $L^1(X,\mathscr{B},\mu,E)^*$ *is of this form.*

Proof. We have $L^1(X,\mathscr{B},\mu,E)\simeq L^1(X,\mathscr{B},\mu,K)\otimes_\pi E$, [12, p. 96] or [4, p. 228]. The dual of the latter space is $\mathscr{L}(E;L^\infty(X,\mathscr{B},\mu,K))$, the space of bounded linear operators on E to $L^\infty(X,\mathscr{B},\mu,K)$, [4, p. 230]. And, since μ is perfect, $L^\infty(X,\mathscr{B},\mu,K)\simeq C(X,K)$, [2, p. 31], or [13, p. 123].

It is straightforward (and is shown explicitly in [3]) that $\mathscr{L}(E;C(X,K))$ is isometric to $C(X,(E^*,\sigma^*))$ under the mapping which sends an element $\Phi'\in\mathscr{L}(E;C(X,K))$ to the element $G\in C(X,(E^*,\sigma^*))$ if and only if $(\Phi' e)(x)=\langle e,G(x)\rangle$ for all $e\in E$ and $x\in X$. Thus (see [4, pp. 228–230]), an element Φ of $L^1(X,\mathscr{B},\mu,E)^*$ is identified with the element $\Phi'\in\mathscr{L}(E;C(X,K))$ if and only if for all measurable sets A with $\mu(A)<\infty$ and all $e\in E$, $\langle\chi_A e,\Phi\rangle=\langle\chi_A,\Phi' e\rangle=\int\chi_A(x)\,(\Phi' e)\,(x)\,d\mu$. And since this latter quantity is equal to $\int\chi_A(x)\,\langle e,G(x)\rangle\,d\mu=\int\langle\chi_A(x)e,\,G(x)\rangle\,d\mu$, where Φ' and G are related as in the sentence before last, the correspondence between elements of $L^1(X,\mathscr{B},\mu,E)^*$ and those of $C(X,(E^*,\sigma^*))$ is as stated in the theorem.

Remark 1. If (Y,Σ,ν) is any σ-finite measure space, it is, of course, standard procedure to represent $L^1(Y,\Sigma,\nu,K)^*$ as $L^\infty(Y,\Sigma,\nu,K)$, [6, p. 289]. And, using Banach algebra techniques, one can obtain a second characterization of $L^1(Y,\Sigma,\nu,K)^*$ as a space of continuous functions on a certain extremally disconnected compact Hausdorff space, [7, p. 17].

In the case of vector-valued functions the characterization of $L^1(Y,\Sigma,\nu,E)^*$ as $L^\infty(Y,\Sigma,\nu,E^*)$ has been established only in the presence of additional assumptions about E^* – e.g. that E^* is separable [5, p. 282], or, more generally, that E^* has the Radon-Nikodym property with respect to ν, [4, p. 98]. However, the preceding theorem shows that $L^1(Y,\Sigma,\nu,E)^*$ is *always* isometric to a space of continuous functions defined on an extremally disconnected compact Hausdorff space, regardless of the nature of E^*. For if X is the Stonean space of the measure algebra Σ/ν, and μ is the perfect Borel measure on X uniquely determined by $\mu(\hat{A})=\nu(A)$, where the clopen subset $\hat{A}$ of X represents the equivalence class of $A\in\Sigma$, [5, p. 351], then the Stone representation can be extended, via simple functions, to an isometry of $L^1(Y,\Sigma,\nu,E)$ onto $L^1(X,\mathscr{B},\mu,E)$.

3. Representation of a Space of Measures

Theorem 2. *Given a space of vector measures* $m(Y,\Sigma,E)$, *there exists an extremally disconnected compact Hausdorff space* X, *and a perfect measure* μ *on* X *such that* $m(Y,\Sigma,E)$ *is isometric to the subspace* $M(\mu,E)$ *of* $m(X,\mathscr{B},E)$.

Proof. We let $\{\lambda_\gamma:\gamma\in\Gamma\}$ be a maximal subset of $m(Y,\Sigma,K)$ consisting of mutually singular positive measures. Next fixing a $\gamma\in\Gamma$, we let X_γ be the Stonean space of the Boolean algebra Σ/λ_γ. Then let Σ_γ denote the algebra of clopen subsets of X_γ. Given $A\in\Sigma$, we let $\hat{A}$ be the element of Σ_γ which represents the

equivalence class of A. Then given $m \in m(Y, \Sigma, E)$ such that $m \ll \lambda_\gamma$, we put

$$\hat{m}(\hat{A}) = m(A).$$

Since m vanishes on every λ_γ-null subset of Σ, it is clear that $\hat{m}$ is well defined on Σ_γ and is additive. Moreover, it is also clear that if $|\hat{m}|$ denotes the variation of $\hat{m}$ on Σ_γ, then $|\hat{m}|(\hat{A}) = |m|(A)$ for $\hat{A} \in \Sigma_\gamma$.

Since the algebra Σ_γ consists of compact sets, $\hat{m}$ and $|\hat{m}|$ are regular on the left; and since Σ_γ consists of open sets, $\hat{m}$ and $|\hat{m}|$ are regular on the right on Σ_γ, [5, p. 303]. Thus $\hat{m}$ and $|\hat{m}|$ are regular on Σ_γ. And since Σ_γ is dense in the power set of X_γ for the interval topology, $\hat{m}$ and $|\hat{m}|$ can be uniquely extended to regular Borel measures on X_γ, [5, p. 302 and 351]. We denote these extensions by the same characters $\hat{m}$ and $|\hat{m}|$, and note that $|\hat{m}|$ is still the variation of $\hat{m}$.

Let X be the Stone-Čech compactification of the disjoint union $\bigcup_{\gamma \in \Gamma} X_\gamma$. If $\hat{m}$ is as above, we next extend $\hat{m}$ to the σ-algebra $\mathscr{B}$ of Borel sets of X by defining $\hat{m}(B)$ to be equal to $\hat{m}(B \cap X_\gamma)$, for $B \in \mathscr{B}$. Clearly this definition extends $\hat{m}$ to a regular Borel measure on X, with $\|\hat{m}\| = \|\hat{m}|_{X_\gamma}\| = \|m\|$.

Now let m be an arbitrary element of $m(Y, \Sigma, E)$. There exists a sequence $\{\lambda_{\gamma_j} : j = 1, 2, \ldots\} \subseteq \{\lambda_\gamma : \gamma \in \Gamma\}$ and a sequence of elements $\{m_j : j = 1, 2, \ldots\} \subseteq m(Y, \Sigma, E)$ such that

(1) $m_j \ll \lambda_{\gamma_j}$ for all j,

(2) $\left\| m - \sum_{j=1}^{N} m_j \right\| \to 0$, and

(3) $\|m\| = \sum_{j=1}^{\infty} \|m_j\|$,

[11, p. 33]. We define, for each $j = 1, 2, \ldots$, $\hat{m}_j$ on $\mathscr{B}$ as in the previous paragraph. Then clearly $\left\{ \sum_{j=1}^{N} \hat{m}_j : N = 1, 2, \ldots \right\}$ is a Cauchy sequence of regular Borel vector measures and is hence convergent to a regular Borel measure on X. We now define $\hat{m}$, (the measure on X corresponding to m), by letting $\hat{m}$ be equal to the limit of the above Cauchy sequence. We have

$$\|\hat{m}\| = \lim_N \left\| \sum_{j=1}^{N} \hat{m}_j \right\| = \lim_N \sum_{j=1}^{N} \|\hat{m}_j\| = \lim_N \sum_{j=1}^{N} \|m_j\| = \|m\|,$$

so that the mapping $m \to \hat{m}$ is norm-preserving. The map is also clearly linear, and thus takes $m(Y, \Sigma, E)$ isometrically onto a subspace M of $m(X, \mathscr{B}, E)$. We would like to show M consists precisely of those elements of $m(X, \mathscr{B}, E)$ which are absolutely continuous with respect to a fixed perfect measure μ on X.

To this end, for each $\gamma \in \Gamma$, we let μ_γ be the perfect Borel measure on X_γ determined by $\mu_\gamma(\hat{A}) = \lambda_\gamma(A)$ where $\hat{A} \in \Sigma_\gamma$ corresponds to $A \in \Sigma$ as in the first paragraph of this proof. We then extend μ_γ to a regular Borel measure on X by setting $\mu_\gamma(B) = \mu_\gamma(B \cap X_\gamma)$, for $B \in \mathscr{B}$. Next define μ on $\mathscr{B}$ by setting $\mu(B) = \sum_{\gamma \in \Gamma} \mu_\gamma(B)$, for $B \in \mathscr{B}$. We thus have

$$\mu(B) = \sum_{\gamma \in \Gamma} \mu_\gamma(B \cap X_\gamma) = \sum_{\gamma \in \Gamma} \mu(B \cap X_\gamma), \qquad (*)$$

and one readily verifies from (*), and the fact that the μ_γ are perfect, that μ is a perfect Borel measure on X. We wish to show that $M=M(\mu,E)$.

We let $M(\lambda_\gamma,E)=\{m\in m(Y,\Sigma,E)\colon m\ll\lambda_\gamma\}$ and $M(\mu_\gamma,E)=\{n\in m(X,\mathscr{B},E)\colon n\ll\mu_\gamma\}$. The fact that the λ_γ are mutually singular and properties (1)–(3) above insure that $m(Y,\Sigma,E)$ is the ℓ^1-direct sum of the subspaces $M(\lambda_\gamma,E)$, $\gamma\in\Gamma$. On the other hand, since $|n|(X-\bigcup_{\gamma\in\Gamma}X_\gamma)=0$ for each μ-continuous n, and the X_γ are pairwise disjoint, $M(\mu,E)$ is the ℓ^1-direct sum of the subspaces $M(\mu_\gamma,E)$. Hence it suffices to show that $M(\lambda_\gamma,E)$ is mapped onto $M(\mu_\gamma,E)$ under the restriction to $M(\lambda_\gamma,E)$ of the map $m\to\hat{m}$.

Thus suppose that $m\in M(\lambda_\gamma,E)$ and let $B\in\mathscr{B}$ be given with $\mu_\gamma(B)=0$. We have to show $|\hat{m}|(B)=0$. Since $|\hat{m}|(B)=|\hat{m}|(B\cap X_\gamma)$, we may assume $B\subseteq X_\gamma$. Then, as μ_γ is regular and perfect on X_γ,

$$\begin{aligned}0&=\inf\{\mu_\gamma(C)\colon C\text{ is clopen and }B\subseteq C\subseteq X_\gamma\}\\&=\inf\{\lambda_\gamma(A)\colon A\in\Sigma\text{ and }B\subseteq\hat{A}\subseteq X_\gamma\}.\end{aligned}$$

Since $|m|$ is λ_γ-continuous, we thus have

$$\begin{aligned}0&=\inf\{|m|(A)\colon A\in\Sigma\text{ and }B\subseteq\hat{A}\subseteq X_\gamma\}\\&=\inf\{|\hat{m}|(C)\colon C\text{ is clopen and }B\subseteq C\subseteq X_\gamma\}.\end{aligned}$$

Thus $|\hat{m}|(B)=0$ and hence $M(\lambda_\gamma,E)$ is mapped into $M(\mu_\gamma,E)$.

Now suppose $n\in M(\mu_\gamma,E)$. Define m on Σ by the equation $m(A)=n(\hat{A})$ for $A\in\Sigma$. The set function m is countably additive, since $n\left(\left(\bigcup_{k=1}^{\infty}A_k\right)^{\wedge}\right)=n\left(\left(\bigcup_{k=1}^{\infty}\hat{A}_k\right)^{-}\right)=n\left(\bigcup_{k=1}^{\infty}\hat{A}_k\right)$ by the μ_γ-continuity of n. From its definition, m is λ_γ-continuous, and $\hat{m}=n$ on Σ_γ. Since the extension of $\hat{m}|_{\Sigma_\gamma}$ to a regular Borel measure on X_γ is unique, $\hat{m}=n$. And hence our mapping $m\to\hat{m}$, when restricted to $M(\lambda_\gamma,E)$, carries $M(\lambda_\gamma,E)$ onto $M(\mu_\gamma,E)$. This completes the proof of the theorem.

4. The Dual of a Space of Measures

We will henceforth assume that E has the Radon-Nikodym property. In view of Theorem 2, the problem of characterizing the dual of a space of vector measures may be reduced to that of finding the dual of the subspace $M(\mu,E)$ of $m(X,\mathscr{B},E)$, where μ is a perfect measure on the extremally disconnected space X. As we have already noted, $M(\mu,E)\simeq L^1(X,\mathscr{B},\mu,E)$, under the map $m\to\frac{dm}{d\mu}$. Thus, via Theorem 1, we have established the following:

Theorem 3. *$M(\mu,E)^*\simeq C(X,(E^*,\sigma^*))$. Specifically, each $G\in C(X,(E^*,\sigma^*))$ gives rise to an element $\Psi_G\in M(\mu,E)^*$ determined by $\langle m,\Psi_G\rangle=\int\left\langle\frac{dm}{d\mu},G\right\rangle d\mu$, for $m\in M(\mu,E)$, with $\|\Psi_G\|=\|G\|_\infty$, and every element of $M(\mu,E)^*$ is of this form.*

Remark 2. If E^* is separable, then each $G \in C(X,(E^*,\sigma^*))$ is measurable as well as bounded, [4, p. 42], and the interaction between elements of $M(\mu,E)$ and those of $C(X,(E^*,\sigma^*))$ in the statement of Theorem 3 becomes $\langle m, \Psi_G \rangle = \int \left\langle \frac{dm}{d\mu}, G \right\rangle d\mu = \int G\, dm$, [5, p. 169].

References

1. Arens, R.: Operations induced in function classes. Monatsh. Math. **55**, 1-19 (1951)
2. Behrends, E. et al.: L^p-structure in real Banach spaces. Lecture Notes in Mathematics **613**. Berlin-Heidelberg-New York: Springer 1977
3. Cambern, M., Greim, P.: The bidual of $C(X,E)$. Proc. Amer. Math. Soc. (to appear)
4. Diestel, J., Uhl, J.J., Jr.: Vector measures. Mathematical Surveys **15**. Providence, Rhode Island: American Mathematical Society 1977
5. Dinculeanu, N.: Vector measures. New York: Pergamon Press 1967
6. Dunford, N., Schwartz, J.T.: Linear operators, Part I. New York-London: Interscience 1958
7. Gamelin, T.: Uniform algebras. Englewood Cliffs, N.J.: Prentice-Hall 1969
8. Gordon, H.: The maximal ideal space of a ring of measurable functions. Amer. J. Math. **88**, 827-843 (1966)
9. Kakutani, S.: Concrete representation of abstract (M)-spaces. Ann. of Math. (2) **42**, 994-1024 (1941)
10. Kaplan, S.: On the second dual of the space of continuous functions. Trans Amer. Math. Soc. **86**, 70-90 (1957)
11. Mauldin, R.D.: The continuum hypothesis, integration and duals of measure spaces. Illinois J. Math. **19**, 33-40 (1975)
12. Schaefer, H.H.: Topological vector spaces. Graduate Texts in Mathematics **3**. New York-Heidelberg-Berlin: Springer 1971
13. Schaefer, H.H.: Banach lattices and positive operators. Berlin-Heidelberg-New York: Springer 1974
14. Singer, I.: Linear functionals on the space of continuous mappings of a compact space into a Banach space. (Russian), Rev. Roumaine Math. Pures Appl. **2**, 301-315 (1957)

Received September 2, 1981; received in final form February 27, 1982

Math. Z. 180, 379–393 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Unabhängige Topologien, Zerlegung von Ringtopologien

Hans Weber

Fachbereich Mathematik der Gesamthochschule, Universitätsstraße 2,
D-4300 Essen, Bundesrepublik Deutschland

0. Einleitung

In [12] (und anschließend in [6] mit modelltheoretischen Methoden) wurde eine Frage von Kowalsky [5], ob nämlich eine lokalbeschränkte Ringtopologie auf einem Körper, die gröber als das Supremum einer (endlichen) Familie von V-Topologien ist, selbst Supremum von V-Topologien ist, positiv beantwortet. Dieses Ergebnis wird durch den Satz (2.2) dieser Arbeit wesentlich verallgemeinert: Eine Ringtopologie $\mathfrak{O}'$ auf einem Ring mit Einzelelement, die gröber ist als das Supremum einer unabhängigen Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ von Ringtopologien, besitzt eine eindeutige Zerlegung der Form $\mathfrak{O}'=\sup_{\alpha\in A}\mathfrak{O}'_\alpha$, $\mathfrak{O}'_\alpha\subset\mathfrak{O}_\alpha(\alpha\in A)$. (Zu einer Charakterisierung der Unabhängigkeit von Topologien siehe Proposition (1.1).)

Mit Hilfe dieses Satzes wird in Abschnitt 4 ein allgemeines Verfahren angegeben, mit dem sich gewisse Ringtopologien auf einem Integritätsbereich R zerlegen lassen in der Form $\mathfrak{O}=\sup\{\mathfrak{O}_1,\hat{\mathfrak{O}}_2|R\}$ mit einer Indealtopologie $\mathfrak{O}_1$ und einer gewissen Ringtopologie $\hat{\mathfrak{O}}_2$ auf dem Quotientenkörper K von R (s. (4.2)). Insbesondere lassen sich auf diese Weise in Verbindung mit [13, Satz (3.3)], in dem sämtliche lokalbeschränkten Ringtopologien auf globalen Körpern bestimmt werden, auch sämtliche lokalbeschränkten Ringtopologien auf beliebigen Teilringen globaler Körper angeben. Mit einer anderen Methode hat Cohen [1] – ebenfalls unter Benutzung von [13, (3.3)] – alle lokalbeschränkten Ringtopologien auf dem Ring der ganzen Zahlen eines globalen Körpers bestimmt. Die Beweise der vorliegenden Arbeit – insbesondere in Abschnitt 4 – sind größtenteils rein topologisch; daher ist die Zerlegung $\mathfrak{O}=\sup\{\mathfrak{O}_1,\hat{\mathfrak{O}}_2|R\}$ von Proposition (4.2) in verschiedenen algebraischen Situationen anwendbar. Konkrete Darstellungen von Ringtopologien auf R haben wir in verschiedenen Fällen (s. (4.4), (4.9), (4.11), (4.12)) angegeben, in denen entsprechende Ringtopologien auf K in [12–14] bestimmt werden.

Grundlegend für diese Arbeit ist der Begriff der Unabhängigkeit von Topologien, der in [12] eingeführt wurde und der nicht nur in der topologischen Algebra von Interesse ist; vgl. dazu die Schlußbemerkungen von Abschnitt 1.

0025-5874/82/0180/0379/$03.00

1. Unabhängige Topologien und Uniformitäten

Im folgenden sei R eine nicht leere Menge.

Eine Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ von Topologien auf R heißt *unabhängig*, wenn für jede endliche Teilmenge A_0 von A und nicht leere, offene Mengen O_α in $(R,\mathfrak{O}_\alpha)$ stets $\bigcap_{\alpha\in A_0} O_\alpha \neq \varnothing$ ist. Eine Familie von Uniformitäten auf R heißt unabhängig, wenn die Familie der induzierten Topologien unabhängig ist.

(1.1) **Proposition** (s. [12, (1.4)]). *$(\mathfrak{O}_\alpha)_{\alpha\in A}$ sei eine Familie von Topologien auf R und $d\colon R\to R^A$ die „Diagonaleinbettung", $d(x)=(x_\alpha)_{\alpha\in A}$ mit $x_\alpha=x$ für alle $\alpha\in A$.*

Dann ist die Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ genau dann unabhängig, wenn $d(R)$ dicht in $(R^A,\prod_{\alpha\in A}\mathfrak{O}_\alpha)$ $(:=\prod_{\alpha\in A}(R,\mathfrak{O}_\alpha))$ liegt.

(1.2) **Lemma.** *$\mathfrak{O}_1,\mathfrak{O}_2$ seien unabhängige Topologien auf R. Ist $\mathfrak{O}$ eine reguläre (nicht notwendig hausdorffsche) Topologie auf R mit $\mathfrak{O}\subset\mathfrak{O}_1$ und $\mathfrak{O}\subset\mathfrak{O}_2$, dann ist $\mathfrak{O}$ trivial.*

Beweis. Mit $\mathfrak{O}_1':=\mathfrak{O}_2':=\mathfrak{O}$ sind dann auch $\mathfrak{O}_1',\mathfrak{O}_2'$ unabhängig. Ist U eine abgeschlossene $\mathfrak{O}$-Umgebung eines Punktes $x\in R$, dann ist $\mathring{U}\neq\varnothing$, $\mathring{U}\cap\tilde{U}=\varnothing$, also wegen der Unabhängigkeit von $\mathfrak{O}_1',\mathfrak{O}_2'$ $\tilde{U}=\varnothing$, d.h. $U=R$.

$\mathfrak{O}_1:=\mathfrak{O}_2:=\{O\subset\mathbb{N}\colon \mathbb{N}\setminus O$ ist endlich$\}\cup\{\varnothing\}$ ist ein Beispiel einer (nicht regulären) nicht trivialen Topologie, so daß $\mathfrak{O}_1,\mathfrak{O}_2$ unabhängig sind.

(1.3) **Lemma.** (a) *Topologien $\mathfrak{O}_1,\mathfrak{O}_2$ auf R sind genau dann unabhängig, wenn jede $\mathfrak{O}_1$-offene, nicht leere Menge in $(R,\mathfrak{O}_2)$ dicht liegt.*

(b) *Sind $\mathfrak{O}_1,\mathfrak{O}_2$ unabhängige Topologien auf R und ist S eine $\mathfrak{O}_1$-offene, nicht leere Teilmenge von R, dann sind auch die Relativtopologien $\mathfrak{O}_1|S$, $\mathfrak{O}_2|S$ unabhängig.*

Der Beweis von (a) ist trivial, (b) folgt aus (a).

Die Relativtopologien von unabhängigen Topologien brauchen i.a. nicht unabhängig zu sein: Ist $\mathfrak{O}_1$ für eine Primzahl p die p-adische Bewertungstopologie auf dem rationalen Zahlkörper $\mathbb{Q}$ und $\mathfrak{O}_2$ die übliche archimedische Topologie auf $\mathbb{Q}$, dann sind $\mathfrak{O}_1,\mathfrak{O}_2$ unabhängig, aber die Relativtopologien $\mathfrak{O}_1|\mathbb{Z}$, $\mathfrak{O}_2|\mathbb{Z}$ auf dem Ring $\mathbb{Z}$ der ganzen Zahlen nicht unabhängig.

Andererseits gilt:

(1.4) **Lemma.** *$\mathfrak{O}_\alpha$, $\alpha\in A$, seien Ringtopologien auf einem Ring R mit Einselement 1. S sei ein Teilring von R mit $1\in S$. Ist die Familie $(\mathfrak{O}_\alpha|S)_{\alpha\in A}$ unabhängig, dann ist auch die Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ unabhängig.*

Beweis. Man kann annehmen, daß $A=\{1,\dots,n\}$ endlich ist. Die Unabhängigkeit von $(\mathfrak{O}_i|S)_{i=1}^n$ ist gleichbedeutend mit der Existenz von gerichteten Systemen $(x_{i,\beta})_{\beta\in B}$ in S mit $\mathfrak{O}_i\text{-}\lim_\beta x_{i,\beta}=1$ und $\mathfrak{O}_j\text{-}\lim_\beta x_{i,\beta}=0$ $(i\neq j)$ (vgl. [12, (1.6) (b)]). Da die Systeme $(x_{i,\beta})_{\beta\in B}$ auch in R liegen, sind $\mathfrak{O}_1,\dots,\mathfrak{O}_n$ unabhängig.

Die Menge aller Uniformitäten bzw. aller Topologien auf R bildet in natürlicher Weise einen vollständigen Verband. Für eine Familie $(\mathfrak{U}_\alpha)_{\alpha\in A}$ von Uniformitäten (bzw. $(\mathfrak{O}_\alpha)_{\alpha\in A}$ von Topologien) auf R bezeichne $\sup_{\alpha\in A}\mathfrak{U}$ (bzw.

$\sup_{\alpha\in A}\mathfrak{O}_\alpha$) die gröbste Uniformität (bzw. Topologie) auf R, die für alle $\alpha\in A$ feiner ist als $\mathfrak{U}_\alpha$ (bzw. $\mathfrak{O}_\alpha$); ist dabei für alle $\alpha\in A$ $\mathfrak{O}_\alpha$ die von $\mathfrak{U}_\alpha$ induzierte Topologie, dann ist $\sup\mathfrak{O}_\alpha$ die von $\sup\mathfrak{U}_\alpha$ induzierte Topologie. Ist für alle $\alpha\in A$ $\mathfrak{O}_\alpha$ eine Gruppentopologie (Ringtopologie, Vektorraumtopologie), dann ist auch $\sup\mathfrak{O}_\alpha$ eine Gruppentopologie (Ringtopologie, Vektorraumtopologie).

(1.5) **Satz.** *Sei $\mathfrak{U}$ eine Uniformität auf R und S eine dichte Teilmenge von $(R,\mathfrak{U})$. Dann wird durch $\mathfrak{V}\mapsto\mathfrak{V}|S$ ein Verbandsisomorphismus definiert vom Verband der Uniformitäten auf R, die gröber als $\mathfrak{U}$ sind, auf den Verband der Uniformitäten auf S, die gröber als $\mathfrak{U}|S$ sind.*

Dieser Satz ist in dieser Arbeit und auch in [15, 16] ein wesentliches Hilfsmittel. In [15] wurde er etwas spezieller formuliert für *monotone*[1] Ringtopologien (=*beschränkte*[2] Ringtopologien auf einem Booleschen Ring). Wir skizzieren den

Beweis von (1.5). Im folgenden werden die Begriffe gleichmäßige Stetigkeit, Abschluß etc. stets bez. $\mathfrak{U}$ verwendet.

(1) Sei $\mathfrak{V}_0$ eine Uniformität auf S mit $\mathfrak{V}_0\subset\mathfrak{U}|S$ und $\mathfrak{M}_0$ eine Menge $[0,\infty[$-wertiger Semimetriken, welche $\mathfrak{V}_0$ erzeugt. Bezeichnet $\bar\rho$ die gleichmäßig stetige Fortsetzung von ρ auf $R\times R$ ($\rho\in\mathfrak{M}_0$), dann erzeugt $\{\bar\rho\colon \rho\in\mathfrak{M}_0\}$ eine Uniformität $\mathfrak{V}$ auf R mit $\mathfrak{V}\subset\mathfrak{U}$ und $\mathfrak{V}|S=\mathfrak{V}_0$.

(2) Sei $\mathfrak{V}$ eine Uniformität auf R mit $\mathfrak{V}\subset\mathfrak{U}$ und $\mathfrak{M}$ eine nach oben gerichtete Menge $[0,\infty[$-wertiger Semimetriken, welche $\mathfrak{V}$ erzeugt. Für $\rho\in\mathfrak{M}$ und $\varepsilon>0$ ist wegen der Stetigkeit von ρ

$$\{(x,y)\in R^2\colon \rho(x,y)<\varepsilon\}\subset\overline{\{(x,y)\in S^2\colon \rho(x,y)<\varepsilon\}}$$
$$\subset\{(x,y)\in R^2\colon \rho(x,y)\leqq\varepsilon\}$$

und daher $\overline{N\cap(S\times S)}$, $N\in\mathfrak{V}$, eine Basis von $\mathfrak{V}$.

Aus (1), (2) ist die Behauptung von (1.5) ablesbar.

Bei der Bezeichnungsweise von (2) ist für $x_0\in S$ wegen

$$\{y\in R\colon \rho(x_0,y)<\varepsilon\}\subset\overline{\{y\in S\colon \rho(x_0,y)<\varepsilon\}}$$
$$\subset\{y\in R\colon \rho(x_0,y)\leqq\varepsilon\}\qquad(\rho\in\mathfrak{M})$$

$\{\bar U\colon U$ ist Umgebung von x_0 in $(S,\mathfrak{V}|S)\}$ eine Umgebungsbasis von x_0 in $(R,\mathfrak{V})$. Benutzt man die Charakterisierung von Ringtopologien durch Nullumgebungen, so sieht man damit sofort, daß $\mathfrak{V}$ genau dann von einer Ringtopologie erzeugt wird, wenn $\mathfrak{V}|S$ von einer Ringtopologie erzeugt wird, vorausgesetzt, R ist ein Ring, $\mathfrak{U}$ die Uniformität einer Ringtopologie auf R und S ein Teilring von R. Das beweist mit (1.5):

(1.6) **Folgerung.** *Ist S ein dichter Teilring eines topologischen Rings $(R,\mathfrak{O}_0)$, dann ist durch $\mathfrak{O}\mapsto\mathfrak{O}|S$ ein Verbandsisomorphismus definiert vom Verband der*

[1] s. [15, Abschnitt 1]

[2] zur Definition s. [13, S. 193]

Ringtopologien auf R, die gröber als $\mathfrak{O}_0$ sind, auf den Verband der Ringtopologien auf S, die gröber als $\mathfrak{O}_0|S$ sind.

Ein Analogon zu (1.5) gilt natürlich auch für Gruppentopologien und Vektorraumtopologien.

(1.7) **Lemma.** *Ist $(\mathfrak{U}_\alpha)_{\alpha\in A}$ eine Schar von Uniformitäten auf R und d wie in* (1.1) *definiert, dann ist $d^{-1}(\prod_{\alpha\in A}\mathfrak{U}_\alpha)=\sup_{\alpha\in A}\mathfrak{U}_\alpha$.*

Dabei bezeichnet $d^{-1}(\prod_{\alpha\in A}\mathfrak{U}_\alpha)$ das inverse Bild der Uniformität des Produkts $\prod_{\alpha\in A}(R,\mathfrak{U}_\alpha)$ $(=:(R^A,\prod_{\alpha\in A}\mathfrak{U}_\alpha))$.

Mit (1.5), (1.7) erhält man die Eindeutigkeit gewisser Zerlegungen von Uniformitäten.

(1.8) **Satz.** *$(\mathfrak{U}_\alpha)_{\alpha\in A}$ sei eine unabhängige Familie von Uniformitäten auf R. Sind $\mathfrak{V}_\alpha$ und $\mathfrak{W}_\alpha$ Uniformitäten auf R mit $\mathfrak{V}_\alpha\subset\mathfrak{U}_\alpha$, $\mathfrak{W}_\alpha\subset\mathfrak{U}_\alpha$ $(\alpha\in A)$ und $\sup_{\alpha\in A}\mathfrak{V}_\alpha=\sup_{\alpha\in A}\mathfrak{W}_\alpha$, dann ist für alle $\alpha\in A$ $\mathfrak{V}_\alpha=\mathfrak{W}_\alpha$.*

Beweis. Definiert man d wie in (1.1), dann ist nach (1.7) $d^{-1}(\prod\mathfrak{V}_\alpha)=\sup\mathfrak{V}_\alpha=\sup\mathfrak{W}_\alpha=d^{-1}(\prod\mathfrak{W}_\alpha)$, also $\prod\mathfrak{V}_\alpha|d(R)=\prod\mathfrak{W}_\alpha|d(R)$. Nach Voraussetzung ist $\prod\mathfrak{V}_\alpha\subset\prod\mathfrak{U}_\alpha$ und $\prod\mathfrak{W}_\alpha\subset\prod\mathfrak{U}_\alpha$. Da nach (1.1) $d(R)$ dicht in $(R^A,\prod\mathfrak{U}_\alpha)$ liegt, erhält man mit (1.5) $\prod\mathfrak{V}_\alpha=\prod\mathfrak{W}_\alpha$ und damit $\mathfrak{V}_\alpha=\mathfrak{W}_\alpha$ für alle $\alpha\in A$.

(1.9) **Proposition.** *$\mathfrak{O}_\alpha(\alpha\in A)$ seien Ringtopologien auf einem Ring R, $\mathfrak{O}=\sup_{\alpha\in A}\mathfrak{O}_\alpha$, N bzw. N_α der topologische Abschluß von $\{0\}$ in $(R,\mathfrak{O})$ bzw. $(R,\mathfrak{O}_\alpha)$ und C bzw. C_α die vollständige Hülle von $(R,\mathfrak{O})/N$ bzw. $(R,\mathfrak{O}_\alpha)/N_\alpha$ $(\alpha\in A)$.*

Dann ist die Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ genau dann unabhängig, wenn die stetige Fortsetzung der Abbildung $x+N\mapsto(x+N_\alpha)_{\alpha\in A}$ $(x\in R)$ auf C $\prod_{\alpha\in A}C_\alpha$ als Bild hat (d.h. ein topologischer Isomorphismus von C auf $\prod_{\alpha\in A}C_\alpha$ ist). (Vgl. (1.1), [12, (1.5)], [2, (2.2) (i)]).

(1.10) **Folgerung.** *Ist $(\mathfrak{O}_\alpha)_{\alpha\in A}$ eine unabhängige Familie von Ringtopologien auf einem Ring R und $\mathfrak{O}=\sup_{\alpha\in A}\mathfrak{O}_\alpha$, dann gilt:*

(a) *Eine Teilmenge von R ist genau dann in $(R,\mathfrak{O})$ beschränkt, wenn sie für jedes $\alpha\in A$ in $(R,\mathfrak{O}_\alpha)$ beschränkt ist.*

(b) *$\mathfrak{O}$ ist genau dann lokalbeschränkt, wenn die Topologien $\mathfrak{O}_\alpha$ alle lokalbeschränkt und mit höchstens endlich vielen Ausnahmen sogar beschränkt sind.* (Vgl. Beweis von [13, (2.1)] und [2, (3.1)]).

Der Begriff der Unabhängigkeit von Topologien wurde in [12] bzw. in einem Preprint (Januar 1977), auf dem [12] und [13] basieren, eingeführt als Verallgemeinerung der Unabhängigkeit von Bewertungen und später (für Ringtopologien) von Stenzel [8] und Wieslaw [18] übernommen. Er spielt bei der Behandlung gewisser Fragen der topologischen Algebra eine wichtige Rolle, vgl. z.B. in dieser Arbeit (2.2), (4.2). Unabhängige Topologien haben auch in der Maßtheorie und Funktionalanalysis eine Bedeutung. Von den verschiedenen Singularitätsbegriffen für gruppenwertige Inhalte, die den klassischen Begriff der Singularität zweier reellwertiger Maße auf einer σ-Algebra verallge-

meinern, ist meines Erachtens der folgende am zweckmäßigsten: Sind μ_1, μ_2 zwei Inhalte auf einem Booleschen Ring und ist $\mathfrak{O}_i$ die gröbste monotone Ringtopologie, bez. der μ_i stetig ist $(i=1,2)$, dann heißen μ_1, μ_2 *singulär*, wenn die triviale Topologie die einzige monotone Ringtopologie ist, die gröber ist als $\mathfrak{O}_1, \mathfrak{O}_2$ (vgl. [16]); (der entsprechende Begriff in [10] ist „*extremely noncontinuous*"); äquivalent dazu ist, daß $\mathfrak{O}_1, \mathfrak{O}_2$ unabhängig sind (s. [16, Abschnitt 1] und (2.3)). Entsprechend nennen De Wilde und Tsirulnikov [2] zwei Vektorraumtopologien $\mathfrak{O}_1, \mathfrak{O}_2$ auf einem reellen linearen Raum „*extraneous*", wenn die triviale Topologie die einzige Vektorraumtopologie ist, die gröber als $\mathfrak{O}_1, \mathfrak{O}_2$ ist; da $\{U_1+U_2 : U_i$ ist eine $\mathfrak{O}_i$-Nullumgebung $(i=1,2)\}$ eine Nullumgebungsbasis der feinsten Vektorraumtopologie ist, die gröber als $\mathfrak{O}_1$ und $\mathfrak{O}_2$ ist, bedeutet das gerade, daß $\mathfrak{O}_1, \mathfrak{O}_2$ unabhängig sind (s. [12, (1.6) (a)]).

2. Zerlegung von Ringtopologien

Ab jetzt sei R ein Ring; für Ringtopologien $\mathfrak{O}_1, \mathfrak{O}_2$ auf R bezeichne $\inf\{\mathfrak{O}_1, \mathfrak{O}_2\}$ die feinste Ringtopologie, die gröber ist als $\mathfrak{O}_1$ und $\mathfrak{O}_2$.

Wir untersuchen zuerst, wann eine Ringtopologie auf einem Produkt von Ringen eine Produkttopologie ist.

(2.1) **Hilfssatz.** *Für $\alpha \in A$ sei $\mathfrak{O}_\alpha$ eine Ringtopologie auf einem Ring R_α. $\hat{\mathfrak{O}}$ sei eine Ringtopologie auf $\prod_{\alpha\in A} R_\alpha$, die gröber ist als die Produkttopologie $\prod_{\alpha\in A} \mathfrak{O}_\alpha$. Für $\gamma \in A$ sei $\pi_\gamma : \prod_{\alpha\in A} R_\alpha \to \prod_{\alpha\in A} R_\alpha$ durch $\pi_\gamma((x_\alpha)_{\alpha\in A}) = (y_\alpha)_{\alpha\in A}$, $y_\gamma = x_\gamma$, $y_\alpha = 0$ $(\alpha \neq \gamma)$ definiert. Dann sind äquivalent:*

(1) *Für jedes $\gamma \in A$ ist $\pi_\gamma : (\prod_{\alpha\in A} R_\alpha, \hat{\mathfrak{O}}) \to (\prod_{\alpha\in A} R_\alpha, \hat{\mathfrak{O}})$ stetig.*

(2) *Für jedes $\alpha \in A$ gibt es eine Ringtopologie $\hat{\mathfrak{O}}_\alpha$ auf R_α, die gröber als $\mathfrak{O}_\alpha$ ist, so daß $\hat{\mathfrak{O}} = \prod_{\alpha\in A} \hat{\mathfrak{O}}_\alpha$.*

Beweis. (2)$\curvearrowright$(1) ist trivial.

(1)$\curvearrowright$(2): Für $\gamma \in A$ bezeichne $j_\gamma : R_\gamma \to \prod_{\alpha\in A} R_\alpha$ die Einbettung $j_\gamma(x) = (y_\alpha)_{\alpha\in A}$, $y_\gamma = x$, $y_\alpha = 0$ $(\alpha \neq \gamma)$ und $\hat{\mathfrak{O}}_\gamma := j_\gamma^{-1}(\hat{\mathfrak{O}})$; dann ist $\hat{\mathfrak{O}}_\gamma \subset j_\gamma^{-1}(\prod_{\alpha\in A} \mathfrak{O}_\alpha) = \mathfrak{O}_\gamma$. Für ein gerichtetes System $(x^{(\iota)})_{\iota\in I} = ((x_\alpha^{(\iota)})_{\alpha\in A})_{\iota\in I}$ in $\prod R_\alpha$ sind äquivalent:

① $x^{(\iota)} \to 0$ $(\prod \hat{\mathfrak{O}}_\alpha)$.
② Für jedes $\alpha \in A$ gilt $x_\alpha^{(\iota)} \to 0$ $(\hat{\mathfrak{O}}_\alpha)$.
③ Für jedes $\alpha \in A$ gilt $j_\alpha(x_\alpha^{(\iota)}) \to 0$ $(\hat{\mathfrak{O}})$.
④ Für jedes $\alpha \in A$ gilt $\pi_\alpha(x^{(\iota)}) \to 0$ $(\hat{\mathfrak{O}})$.
⑤ $x^{(\iota)} \to 0$ $(\hat{\mathfrak{O}})$.

Offenbar gilt ①$\curvearrowright$②$\curvearrowright$③$\curvearrowright$④ und wegen (1) ⑤$\curvearrowright$④. Zum Nachweis von ④$\curvearrowright$⑤ seien U, V $\hat{\mathfrak{O}}$-Nullumgebungen mit $V+V \subset U$. Wegen $\hat{\mathfrak{O}} \subset \prod \mathfrak{O}_\alpha$ gibt es eine endliche Teilmenge A_0 von A und $\mathfrak{O}_\alpha$-Nullumgebungen W_α $(\alpha \in A_0)$ mit $W := \{(x_\alpha)_{\alpha\in A} \in \prod R_\alpha : x_\alpha \in W_\alpha$ für $\alpha \in A_0\} \subset V$. Da nach ④ $\sum_{\alpha\in A_0} \pi_\alpha(x^{(\iota)}) \to 0$ $(\hat{\mathfrak{O}})$, gibt es ein $\iota_0 \in I$ mit $\sum_{\alpha\in A_0} \pi_\alpha(x^{(\iota)}) \in V$ $(\iota \geqq \iota_0)$. Für $\iota \geqq \iota_0$ ist dann $x^{(\iota)} \in V+W \subset U$.

Wegen der Äquivalenz ①↷⑤ ist $\hat{\mathfrak{O}} = \prod \hat{\mathfrak{O}}_\alpha$.

Besitzen die Ringe R_α ($\alpha \in A$) sämtlich ein Einselement, dann ist (1) von (2.1) erfüllt; ist nämlich für $\gamma \in A$ $e_\gamma = (\delta_{\gamma\alpha})_{\alpha\in A} \in \prod R_\alpha$, $\delta_{\gamma\gamma} = 1$, $\delta_{\gamma\alpha} = 0$ ($\alpha \neq \gamma$), dann ist $\pi_\gamma(x) = e_\gamma \cdot x$ ($x \in \prod R_\alpha$) und daher wegen der Stetigkeit der Multiplikation π_γ stetig. Die Bedingung (1) von (2.1) ist auch erfüllt, wenn für jedes $\alpha \in A$ R_α ein Boolescher Ring und $\hat{\mathfrak{O}}$ eine monotone Ringtopologie ist.

(2.2) **Satz.** *$\hat{\mathfrak{O}}$ sei eine Ringtopologie auf einem Ring R mit Einselement oder eine monotone Ringtopologie auf einem Booleschen Ring R. $(\mathfrak{O}_\alpha)_{\alpha\in A}$ sei eine unabhängige Familie von Ringtopologien auf R und $\hat{\mathfrak{O}} \subset \sup_{\alpha\in A} \mathfrak{O}_\alpha$.*

Dann gibt es eine eindeutig bestimmte Familie $(\hat{\mathfrak{O}}_\alpha)_{\alpha\in A}$ von Ringtopologien auf R mit $\hat{\mathfrak{O}}_\alpha \subset \mathfrak{O}_\alpha$ ($\alpha \in A$) und $\hat{\mathfrak{O}} = \sup_{\alpha\in A} \hat{\mathfrak{O}}_\alpha$. Dabei ist $\hat{\mathfrak{O}}_\alpha = \inf\{\hat{\mathfrak{O}}, \mathfrak{O}_\alpha\}$.

Im Falle, daß $\mathfrak{O}_\alpha$ ($\alpha \in A$) und $\hat{\mathfrak{O}}$ monotone Ringtopologien sind, sind auch die $\hat{\mathfrak{O}}_\alpha$ ($\alpha \in A$) monotone Ringtopologien.

Beweis. Sei d wie in (1.1) definiert. Bezeichnet für eine Topologie $\mathfrak{O}$ auf R $d(\mathfrak{O}) = (d^{-1})^{-1}(\mathfrak{O})$ das inverse Bild von $\mathfrak{O}$ bez. d^{-1}, dann ist nach (1.7) $\prod \mathfrak{O}_\alpha | d(R) = d(\sup \mathfrak{O}_\alpha) \supset d(\hat{\mathfrak{O}})$. Da nach (1.1) $d(R)$ dicht in $(R^A, \prod \mathfrak{O}_\alpha)$ liegt, gibt es nach (1.6) eine Ringtopologie $\tilde{\mathfrak{O}}$ auf R^A mit $\tilde{\mathfrak{O}} | d(R) = d(\hat{\mathfrak{O}})$ und $\tilde{\mathfrak{O}} \subset \prod \mathfrak{O}_\alpha$. Nach (2.1) gibt es Ringtopologien $\hat{\mathfrak{O}}_\alpha$ auf R ($\alpha \in A$) mit $\hat{\mathfrak{O}}_\alpha \subset \mathfrak{O}_\alpha$ ($\alpha \in A$) und $\tilde{\mathfrak{O}} = \prod \hat{\mathfrak{O}}_\alpha$. Nun folgt $\hat{\mathfrak{O}} = d^{-1}(\tilde{\mathfrak{O}}) = d^{-1}(\prod \hat{\mathfrak{O}}_\alpha) = \sup \hat{\mathfrak{O}}_\alpha$ (s. (1.7)).

Die Eindeutigkeitsaussage gilt nach (1.8). Da $\hat{\mathfrak{O}}_\alpha \subset \inf\{\hat{\mathfrak{O}}, \mathfrak{O}_\alpha\} \subset \hat{\mathfrak{O}}$ ($\alpha \in A$), ist auch $\hat{\mathfrak{O}} = \sup_\alpha (\inf\{\hat{\mathfrak{O}}, \hat{\mathfrak{O}}_\alpha\})$; die Eindeutigkeitsaussage liefert $\hat{\mathfrak{O}}_\alpha = \inf\{\hat{\mathfrak{O}}, \mathfrak{O}_\alpha\}$. Allgemeiner als die letzte Aussage von (2.2) folgt aus der Gleichung $\hat{\mathfrak{O}}_\alpha = \inf\{\hat{\mathfrak{O}}, \hat{\mathfrak{O}}_\alpha\}$ und (b) des nächsten Lemmas, daß $\hat{\mathfrak{O}}_\alpha$ beschränkt ist, falls $\hat{\mathfrak{O}}$ und $\hat{\mathfrak{O}}_\alpha$ beschränkt sind.

(2.3) **Lemma.** *Für zwei beschränkte Ringtopologien $\mathfrak{O}_1, \mathfrak{O}_2$ auf einem Ring R gilt:*

(a) *$\{U_1 + U_2 : U_i$ ist $\mathfrak{O}_i$-Nullumgebung $(i = 1, 2)\}$ ist eine Nullumgebungsbasis von $\inf\{\mathfrak{O}_1, \mathfrak{O}_2\}$.*

(b) *$\inf\{\mathfrak{O}_1, \mathfrak{O}_2\}$ ist eine beschränkte Ringtopologie.*

(c) *$\mathfrak{O}_1, \mathfrak{O}_2$ sind unabhängig genau dann, wenn $\inf\{\mathfrak{O}_1, \mathfrak{O}_2\}$ die triviale Topologie ist.*

Beweis. Offenbar ist $\mathfrak{U}_0 := \{U_1 + U_2 : U_i$ ist $\mathfrak{O}_i$-Nullumgebung $(i = 1, 2)\}$ eine Filterbasis; zu jedem $U \in \mathfrak{U}_0$ gibt es ein $V \in \mathfrak{U}_0$ mit $V - V \subset U$, $RV \subset U$ und $VR \subset U$. Also ist $\mathfrak{U}_0$ Nullumgebungsbasis einer beschränkten Ringtopologie auf R. Daraus folgen sofort (a) und (b). (c) erhält man aus (a) und [12, (1.6) (a)].

Es ist eine interessante Frage, für welche Ringtopologien die Äquivalenz aus (2.3) (c) gilt. Man bestätigt leicht, daß (2.3) (a) und damit (2.3) (c) noch gültig bleiben, wenn nur eine der Ringtopologien $\mathfrak{O}_1, \mathfrak{O}_2$ beschränkt ist oder $\mathfrak{O}_1, \mathfrak{O}_2$ beide linksbeschränkt[2] oder $\mathfrak{O}_1, \mathfrak{O}_2$ beide rechtsbeschränkt[2] sind. Im Fall, daß R ein Körper, $\mathfrak{O}_1$ eine V-Topologie und $\mathfrak{O}_2$ eine Körpertopologie auf R ist, wurde die Äquivalenz aus (2.3) (c) unabhängig von Shanks [18, Theorem 3] und ein etwas allgemeineres Resultat von mir [12, (2.3)] bewiesen. Einfache

Beispiele zeigen, daß die Äquivalenz aus (2.3) (c) für *lokalbeschränkte*[2] Ringtopologien nicht zu gelten braucht (s. [12, S. 133]).

Ist $\mathfrak{O}$ eine nicht triviale Rintopologie und die triviale Topologie die einzige von $\mathfrak{O}$ verschiedene Ringtopologie, die gröber ist als $\mathfrak{O}$, dann heißt $\mathfrak{O}$ eine *gröbste* Ringtopologie.

(2.4) **Folgerung.** *Ist R ein Ring mit Einselement und $(\mathfrak{O}_\alpha)_{\alpha\in A}$ eine Familie von unabhängigen Ringtopologien auf R, dann sind äquivalent:*

(1) *$\mathfrak{O}_\alpha$, $\alpha\in A$, sind gröbste Ringtopologien.*

(2) *$B\mapsto \sup_{\alpha\in B}\mathfrak{O}_\alpha$ ist ein Ordnungsisomorphismus von der Potenzmenge $\mathbb{P}(A)$ von A auf den Verband aller Ringtopologien auf R, die gröber sind als $\sup_{\alpha\in A}\mathfrak{O}_\alpha$.*

Beweis. (2)$\curvearrowright$(1) ist trivial, (1)$\curvearrowright$(2) folgt aus (2.2).

Ein Satz der Art (2.4) (1)$\curvearrowright$(2) wurde erstmals in [12; (1.9), (2.5)] bewiesen – angeregt durch eine Fragestellung von Kowalsky [5, S. 185] (vgl. Einleitung von [12]), anschließend von Stenzel [8, (6.6)] und mit modelltheoretischen Methoden von Prestel und Ziegler [6, Theorem (4.4)]. (In [12, (2.5)] ist $(\mathfrak{O}_\alpha)_{\alpha\in A}$ eine Familie verschiedener V-Topologien auf einem Schiefkörper – $(\mathfrak{O}_\alpha)_{\alpha\in A}$ ist dann eine unabhängige Familie von gröbsten Ringtopologien –, in [6, (4.4)] ist $(\mathfrak{O}_\alpha)_{\alpha\in A}$ eine endliche Familie verschiedener V-Topologien auf einem Körper und in [8, (6.6)] eine unabhängige Familie von gröbsten Ringtopologien auf einem Körper.) Durch (2.2), (2.4) ist diese Fragestellung in gewisser Weise abschließend beantwortet. Wohl ist in diesem Zusammenhang noch von Interesse, ob sich in (2.2), (2.4) die Unabhängigkeitsvoraussetzung abschwächen läßt, und, wann Ringtopologien $\mathfrak{O}_1,\ldots,\mathfrak{O}_n$ auf R unabhängig sind, d.h. für sie der Approximationssatz gilt. Stone [9, S. 292] hat einen entsprechenden Approximationssatz erstmals (mit Non-Standard-Methoden) für V-Topologien bewiesen. (Von Prestel und Ziegler [6, Theorem (4.1)] wurde für diesen Satz von Stone ein modelltheoretischer Beweis und von mir [12, (2.3)] (im nicht-kommutativen Fall) ein elementarer Beweis angegeben.) Vgl. auch (2.8).

(2.5) **Lemma.** *R sei ein Ring; zu jedem $x\in R$ gebe es $y, z\in R$ mit $x=y\cdot z$ (das ist erfüllt, wenn R ein Einselement besitzt $(x=1\cdot x)$ oder R ein Boolescher Ring ist $(x=x\cdot x)$). $\mathfrak{O},\mathfrak{O}_\alpha$ $(\alpha\in A)$ seien beschränkte Ringtopologien auf R. Dann sind äquivalent:*

(1) *$\mathfrak{O}$, $\sup_{\alpha\in A}\mathfrak{O}_\alpha$ sind unabhängig.*

(2) *Für jedes $\alpha\in A$ sind $\mathfrak{O},\mathfrak{O}_\alpha$ unabhängig.*

Beweis. (1)$\curvearrowright$(2) ist trivial (vgl. [12, (1.1) (b)]).

(2)$\curvearrowright$(1): Wir betrachten zunächst den Fall, daß $A=\{1,2\}$ zweielementig ist. Sei U eine $\mathfrak{O}$-Nullumgebung, V eine $\sup\{\mathfrak{O}_1,\mathfrak{O}_2\}$-Nullumgebung und $x\in R$; wir zeigen, daß $x\in U+V$ (s. [12, (1.6) (a)]). Dazu wählt man eine $\mathfrak{O}$-Nullumgebung U_0, $\mathfrak{O}_i$-Nullumgebungen V_i, W_i $(i=1,2)$ und Elemente $y, z\in R$ mit den folgenden Eigenschaften: $U_0R+RU_0\subset U$, $V_1\cap V_2\subset V$, $W_1R\subset V_1$, $RW_2\subset V_2$, $x=yz$. Wegen der Unabhängigkeit von $\mathfrak{O},\mathfrak{O}_i$ gibt es nach [12, (1.6) (a)] Elemente $u_i\in U_0$ und $w_i\in W_i$ $(i=1,2)$ mit $y=u_1+w_1$, $z=u_2+w_2$. Folglich ist $x=yz=u_1z+w_1u_2+w_1w_2\in U_0R+RU_0+(W_1R\cap RW_2)\subset U+V$. Für zweielementiges A ist damit die Behauptung bewiesen, daraus erhält man sie induktiv für endliches A und daraus schließlich für beliebiges A.

(2.6) **Folgerung.** *R sei ein Ring; zu jedem $x\in R$ gebe es Elemente y, $z\in R$ mit $x=y\cdot z$. Dann sind für eine Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ beschränkter Ringtopologien auf R äquivalent:*

(1) *Die Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ ist unabhängig.*

(2) *Für je zwei verschiedene Indizes $\alpha,\beta\in A$ sind $\mathfrak{O}_\alpha,\mathfrak{O}_\beta$ unabhängig.*

(3) *Für je zwei verschiedene Indizes $\alpha,\beta\in A$ ist $\inf\{\mathfrak{O}_\alpha,\mathfrak{O}_\beta\}$ die triviale Topologie.*

Beweis. (3)↷(2) gilt nach (2.3) (c); (1)↷(2) ist trivial.

(2)↷(1): Wir beweisen die Behauptung induktiv für eine n-elementige Indexmenge A ($n\in\mathbb{N}$). Für $n=1,2$ ist nichts zu beweisen. Gilt die Behauptung für eine n-elementige Indexmenge und ist $A=\{1,\dots,n+1\}$, dann sind nach Induktionsannahme $\mathfrak{O}_1,\dots,\mathfrak{O}_n$ unabhängig und nach (2.5) $\mathfrak{O}_{n+1}$, $\sup_{i=1}^{n}\mathfrak{O}_i$ unabhängig; also sind $\mathfrak{O}_1,\dots,\mathfrak{O}_{n+1}$ unabhängig (s. [12, (1.2)]).

Ein Vergleich z.B. von (2.2) und [16, (2.8)] wird durch folgende Äquivalenz ermöglicht.

(2.7) **Folgerung.** *Eine Familie $(\mathfrak{O}_\alpha)_{\alpha\in A}$ monotoner Ringtopologien auf einem Booleschen Ring ist genau dann unabhängig, wenn $\mathfrak{O}_\alpha$, $\alpha\in A$, paarweise singulär sind.*

(2.8) **Folgerung.** *Eine Familie von verschiedenen, beschränkten, gröbsten Ringtopologien auf einem Ring mit Einselement ist unabhängig.*

Der Satz (2.2) ist ein wichtiges Hilfsmittel im Beweis von (4.2). Als weitere Anwendung von (2.2) (bzw. (2.4) oder [12, (2.5)]) notieren wir einen Satz über die Fortsetzung von gewissen Ringtopologien. Wir benutzen dabei das bekannte

(2.9) **Lemma.** *$\mathfrak{O}$ sei eine Ringtopologie auf einem Körper K. L sei eine endlich-algebraische Körpererweiterung von K und $c_1,\dots,c_n$ eine K-Basis von L. $\mathfrak{O}_0$ bezeichne die Topologie des Produkts $(K,\mathfrak{O})^n$ und $f\colon L\to K^n$ die durch $f\left(\sum_{i=1}^{n}a_ic_i\right)=(a_1,\dots,a_n)$ definierte Abbildung. Dann ist $f^{-1}(\mathfrak{O}_0)$ die feinste Ringtopologie auf L, welche $\mathfrak{O}$ fortsetzt (d.h. auf K $\mathfrak{O}$ induziert).*

Beweis. Offenbar ist $f^{-1}(\mathfrak{O}_0)$ eine Ringtopologie auf L mit $f^{-1}(\mathfrak{O}_0)|K=\mathfrak{O}$. Ist $\mathfrak{O}'$ eine Ringtopologie auf L mit $\mathfrak{O}'|K=\mathfrak{O}$, dann ist $f^{-1}\colon (K^n,\mathfrak{O}_0)\to(L,\mathfrak{O}')$ stetig, also auch $\mathrm{id}=f^{-1}\circ f\colon (L,f^{-1}(\mathfrak{O}_0))\to(L,\mathfrak{O}')$ stetig, d.h. $\mathfrak{O}'\subset f^{-1}(\mathfrak{O}_0)$.
$f^{-1}(\mathfrak{O}_0)$ nennen wir die Produkttopologie auf L bez. $\mathfrak{O}$; sie ist unabhängig von der Wahl der Basis $c_1,\dots,c_n$.

(2.10) **Satz.** *$\mathfrak{O}_\alpha$, $\alpha\in A$, seien verschiedene V-Topologien auf einem Körper K. L sei eine endlich-algebraische, separable Körpererweiterung von K. $\mathbb{T}_\alpha$ bezeichne die (nicht leere) Menge aller V-Topologien auf L, welche $\mathfrak{O}_\alpha$ fortsetzen ($\alpha\in A$).*

Dann gibt es zu jeder Ringtopologie $\mathfrak{O}$ auf L, welche $\sup_{\alpha\in A}\mathfrak{O}_\alpha$ fortsetzt, eine eindeutig bestimmte Familie $(\mathbb{T}'_\alpha)_{\alpha\in A}$ mit $\varnothing\neq\mathbb{T}'_\alpha\subset\mathbb{T}_\alpha$ und $\mathfrak{O}=\sup_{\alpha\in A}\bigcup\mathbb{T}'$.

Ist $L|K$ galoissch, G die zugehörige Galoisgruppe und $\mathfrak{O}'_\alpha\in\mathbb{T}_\alpha$, dann ist $\mathbb{T}_\alpha=\{\sigma^{-1}(\mathfrak{O}'_\alpha)\colon\sigma\in G\}$ ($\alpha\in A$).

Beweis. Wir können annehmen, daß $L|K$ galoissch ist; sonst betrachtet man einen endlich-algebraischen Erweiterungskörper M von L, so daß $M|K$ galoissch ist, und beachtet, daß sich nach (2.9) jede auf L gegebene Ringtopologie auf M fortsetzen läßt. Sei nun $L|K$ galoissch mit Galoisgruppe G, $\mathfrak{O}'_\alpha$ eine V-Topologie, die $\mathfrak{O}_\alpha$ fortsetzt, $\mathfrak{U}_\alpha := \{\sigma^{-1}(\mathfrak{O}'_\alpha): \sigma \in G\}$ $(\alpha \in A)$, $\mathfrak{O}_0 := \sup \bigcup_{\alpha \in A} \mathfrak{U}_\alpha$ und $\mathfrak{O}$ eine Ringtopologie auf L mit $\mathfrak{O}|K = \sup_{\alpha \in A} \mathfrak{O}_\alpha$.

Nach [17, Lemma 2] (– dort ist die Vollständigkeitsvoraussetzung überflüssig –) ist $\mathfrak{O}_0$ die Produkttopologie bez. $\sup_{\alpha \in A} \mathfrak{O}_\alpha$ und daher nach (2.9) $\mathfrak{O} \subset \mathfrak{O}_0$. Nach (2.4) bzw. [12, (2.5)] gibt es eindeutig bestimmte Teilmengen $\mathfrak{U}'_\alpha$ von $\mathfrak{U}_\alpha$ $(\alpha \in A)$ mit $\mathfrak{O} = \sup \bigcup_{\alpha \in A} \mathfrak{U}'_\alpha$; wegen $\mathfrak{O}|K = \sup_{\alpha \in A} \mathfrak{O}_\alpha$ ist $\mathfrak{U}'_\alpha \neq \varnothing$ $(\alpha \in A)$.

Für eine endliche Indexmenge A wurde der Satz (2.10) von Rigo und Warner [7, Theorem 5] bewiesen.

Genauso wie (2.10) läßt sich ein entsprechender Satz beweisen, wenn $(\mathfrak{O}_\alpha)_{\alpha \in A}$ eine unabhängige Familie von gröbsten Ringtopologien ist, für die sich jedes $\mathfrak{O}_\alpha$ zu einer gröbsten Ringtopologie auf die galoissche Hülle von L bez. K fortsetzen läßt. (Im Beweis hat man noch (1.4) zu benutzen.) Allerdings ist zur Zeit nicht bekannt, ob es überhaupt gröbste Ringtopologien auf einem Körper gibt, die keine V-Topologien sind.

3. Eine Charakterisierung globaler Körper und algebraischer Funktionenkörper in einer Variablen

Die in [14] angegebene topologische Charakterisierung globaler Körper und algebraischer Funktionenkörper in einer Variablen enthält als Spezialfall eine algebraische Charakterisierung.

Ist U eine Teilmenge eines Körpers K und $(aU)_{a \in K^*}$ eine Nullumgebungsbasis einer Ringtopologie $\mathfrak{O}$ auf K, so sagen wir: U *erzeugt* $\mathfrak{O}$; dabei sei für $M \subset K$ $M^* := M \setminus \{0\}$. Für eine Menge $\$$ von Bewertungen auf einem Körper K sei $O(\$) := \{a \in K$: für alle $\varphi \in \$$ ist $\varphi(a) \leqq 1\}$; erzeugt $O(\$)$ eine Ringtopologie auf K, so bezeichnen wir diese mit $\mathfrak{O}(\$)$.

Wir notieren zunächst in (3.1) algebraische Eigenschaften globaler Körper, die sich sehr einfach mit den topologischen Ergebnissen aus [13] beweisen lassen und auch in Abschnitt 4 benötigt werden.

(3.1) **Proposition.** *Sei K ein globaler Körper und R ein Teilring von K mit $R \cdot (R^*)^{-1} = K$.*

(a) *Ist $\$_0$ das normalisierte volle Bewertungssystem von K, $\$:= \{\varphi \in \$_0 : R \subset O(\{\varphi\})\}$, dann gibt es ein $a \in R^*$ mit $aO(\$) \subset R \subset O(\$)$.*

(b) *Für jedes von* (0) *verschiedene Ideal A in R ist der Faktorring R/A endlich.* (s. [4, Satz 5]).

Beweis. Sei A ein Ideal in R, $A \neq (0)$. A und R erzeugen dieselbe lokalbeschränkte Ringtopologie auf K; nach [13, (3.3)] ist diese Topologie für eine echte Teilmenge $\$'$ von $\$_0$ gleich $\mathfrak{O}(\$')$. Also ist für ein $a \in R^*$

$aO(\$') \subset A \subset R \subset O(\$) \subset O(\$')$ (s. [13, (2.2)]). Hieraus folgt (a) und $(R: A) \leqq (O(\$'): aO(\$')) < \infty$. Es folgt noch $\$' = \$$.

(3.2) **Satz.** *Für einen Körper K sind äquivalent:*

(1) *K ist ein globaler Körper, oder K ist algebraisch über einem endlichen Körper.*

(2) *Zu jedem Teilring R von K mit $R \cdot (R^*)^{-1} = K$ gibt es ein $a \in R^*$, so daß $R \cdot a^{-1}$ den ganzen Abschluß von R in K enthält.*

Beweis. (1)$\curvearrowright$(2) ist im Fall, daß K über einem endlichen Körper algebriasch ist, trivial. Ist K ein globaler Körper, folgt das aus (3.1) (a). Bei gleicher Bezeichnung wie in (3.1) (a) ist nämlich der ganze Abschluß von R ein Teilring von $O(\$)$, da $O(\$)$ ganz abgeschlossen und $R \subset O(\$)$ ist. ($O(\$)$ ist sogar der ganze Abschluß von R.)

(2)$\curvearrowright$(1): Gilt (1) nicht, dann gibt es auf K eine hausdorffsche, nicht diskrete, lokalbeschränkte Ringtopologie $\mathfrak{O}$, die nicht vom Typ $\mathfrak{O}(\$)$ ist; diese wird von einem Teilring R von K erzeugt (s. Beweis von [14, (3.3)]). Dabei ist $R \cdot (R^*)^{-1} = K$ (vgl. [13, (1.1) (a)]). Der ganze Abschluß von R ist für eine Menge nicht archimedischer Bewertungen auf K gleich $O(\$)$. Für kein $a \in R$ gilt $O(\$) \subset Ra^{-1}$, da sonst zusammen mit $R \subset O(\$)$ folgen würde, daß $\mathfrak{O} = \mathfrak{O}(\$)$ ist.

Genauso wie (3.1), (3.2) beweist man (3.3), (3.4), indem man [14, (4.8)] statt [13, (3.3)], [14, (3.3)] benutzt und beachtet, daß bei der Bezeichnung von (3.3) $O(\$)/aO(\$)$ endlich dimensional über k ist.

(3.3) **Proposition.** *K sei ein algebraischer Funktionenkörper in einer Variablen über k und R ein Teilring von K mit $R \cdot (R^*)^{-1} = K$ und $k \subset R$.*

(a) *Ist $\$_0$ das normalisierte volle Bewertungssystem der Bewertungen auf K, die auf k trivial sind, und $\$:= \{\varphi \in \$_0 : R \subset O(\{\varphi\})\}$, dann gibt es ein $a \in R^*$ mit $aO(\$) \subset R \subset O(\$)$.*

(b) *Für jedes von (0) verschiedene Ideal A in R ist R/A endlich dimensional über k.*

(3.4) **Satz.** *Für eine Körpererweiterung $K|k$ sind äquivalent:*

(1) *K ist ein algebraischer Funktionenkörper in einer Variablen über k oder algebraisch über k.*

(2) *Zu jedem Teilring R von K mit $k \subset R$ und $R(R^*)^{-1} = K$ gibt es ein $a \in R^*$, so daß Ra^{-1} den ganzen Abschluß von R in K enthält.*

4. Lokalbeschränkte Ringtopologien auf Teilringen globaler Körper und algebraischer Funktionenkörper in einer Variablen

In diesem Abschnitt sei R ein Integritätsbereich und K sein Quotientenkörper.

In [13, (3.3)] wurden sämtliche lokalbeschränkten Ringtopologien auf einem globalen Körper bestimmt. Ist K ein globaler Körper, dann sind damit auch die lokalbeschränkten Ringtopologien auf R bekannt, die sich zu einer lokalbeschränkten Ringtopologie auf K fortsetzen lassen; aber nicht jede hausdorffsche, lokalbeschränkte Ringtopologie $\mathfrak{O}$ auf R besitzt eine solche Fortset-

zung (Beispiel: $K=\mathbb{Q}$, $R=\mathbb{Z}$, $\mathfrak{O}$ sei die Ringtopologie mit $\{p\mathbb{Z}: p$ ist Primzahl$\}$ als Nullumgebungsbasis). Statt dessen gilt noch (vgl. [11, (3.2)], [1, Lemma 2]):

(4.1) **Lemma.** *$\mathfrak{O}$ sei eine nicht diskrete Ringtopologie auf R. Dann gilt:*

(a) *$\{aU: a\in R^*$, U ist eine $\mathfrak{O}$-Nullumgebung$\}$ ist eine Nullumgebungsbasis der gröbsten Ringtopologie $\hat{\mathfrak{O}}$ auf K, so daß $\mathfrak{O}\subset\hat{\mathfrak{O}}|R$ und R offen in $(K,\hat{\mathfrak{O}})$ ist. $\mathfrak{O}$ ist nicht diskret.*

(b) *Jede beschränkte Teilmenge von $(R,\mathfrak{O})$ ist auch beschränkt in $(K,\hat{\mathfrak{O}})$.*

(c) *Ist U eine beschränkte $\mathfrak{O}$-Nullumgebung, dann wird $\hat{\mathfrak{O}}$ von U erzeugt. Insbesondere ist $\hat{\mathfrak{O}}$ lokalbeschränkt, wenn $\mathfrak{O}$ lokalbeschränkt ist.*

Beweis. Zu (a) vgl. [11, (3.2)].

(b) Sei B beschränkt in $(R,\mathfrak{O})$, U eine $\mathfrak{O}$-Nullumgebung und $a\in R^*$. Dann gibt es eine $\mathfrak{O}$-Nullumgebung V mit $VB\subset U$. Dann ist $(aV)B\subset aU$, also nach (a) B $\hat{\mathfrak{O}}$-beschränkt.

(c) Eine beschränkte Nullumgebung U in $(R,\mathfrak{O})$ ist nach (a), (b) auch eine beschränkte Nullumgebung in $(K,\hat{\mathfrak{O}})$ und erzeugt daher $\hat{\mathfrak{O}}$, da $\hat{\mathfrak{O}}$ nicht diskret ist.

Eine Ringtopologie, die eine Nullumgebungsbasis aus Idealen besitzt, heißt *Idealtopologie*. $\{aR: a\in R^*\}$ ist eine Nullumgebungsbasis der feinsten nicht diskreten Idealtopologie auf R.

(4.2) **Proposition.** *Voraussetzung:* (1) *Sei $B\subset R$ und $\mathbb{T}$ die Menge aller nicht diskreten (lokalbeschränkten) Ringtopologien auf K, bez. denen B beschränkt ist. Zu je zwei Topologien $\hat{\mathfrak{O}}_1,\hat{\mathfrak{O}}_2\in\mathbb{T}$ mit $\hat{\mathfrak{O}}_2\subset\hat{\mathfrak{O}}_1$ gebe es ein $\hat{\mathfrak{O}}_3\in\mathbb{T}$, so daß $\hat{\mathfrak{O}}_1=\sup\{\hat{\mathfrak{O}}_2,\hat{\mathfrak{O}}_3\}$ und $\hat{\mathfrak{O}}_2,\hat{\mathfrak{O}}_3$ unabhängig sind.*

(2) *$\hat{\mathfrak{O}}_0$ bezeichne die von R erzeugte Ringtopologie auf K und $\mathfrak{O}_0$ die feinste nicht diskrete Idealtopologie auf R.*

(3) *$\mathfrak{O}$ sei eine nicht diskrete (lokalbeschränkte) Ringtoplogie auf R, bezüglich der für ein $a\in R^*$ aB beschränkt ist.*

Behauptung: (a) *Dann gibt es eindeutig bestimmte Ringtopologien $\mathfrak{O}_1,\mathfrak{O}_2$ auf R, so daß $\mathfrak{O}=\sup\{\mathfrak{O}_1,\mathfrak{O}_2\}$, $\mathfrak{O}_1\subset\mathfrak{O}_0$ und $\mathfrak{O}_2,\mathfrak{O}_0$ unabhängig sind.*

(b) *$\mathfrak{O}_1,\mathfrak{O}_2$ sind die einzigen Ringtopologien auf R, für die $\mathfrak{O}=\sup\{\mathfrak{O}_1,\mathfrak{O}_2\}$, $\mathfrak{O}_1\subset\mathfrak{O}_0$ und $\mathfrak{O}_2$ sich zu einer Ringtopologie $\hat{\mathfrak{O}}_2$ auf K fortsetzen läßt, so daß R dicht in $(K,\hat{\mathfrak{O}}_2)$ liegt. Dabei ist $\hat{\mathfrak{O}}_2\in\mathbb{T}$ und sind $\hat{\mathfrak{O}}_0,\hat{\mathfrak{O}}_2$ unabhängig.*

Beweis. Wählt man $\hat{\mathfrak{O}}$ wie in (4.1), dann ist $\hat{\mathfrak{O}}\in\mathbb{T}$, $\mathfrak{O}\subset\hat{\mathfrak{O}}|R$, $\hat{\mathfrak{O}}_0\subset\hat{\mathfrak{O}}$, $\mathfrak{O}_0=\hat{\mathfrak{O}}_0|R$ und $\hat{\mathfrak{O}}_0\in\mathbb{T}$. Wählt man gemäß der Voraussetzung (1) $\hat{\mathfrak{O}}_3\in\mathbb{T}$, so daß $\hat{\mathfrak{O}}=\sup\{\hat{\mathfrak{O}}_0,\hat{\mathfrak{O}}_3\}$ und $\hat{\mathfrak{O}}_0,\hat{\mathfrak{O}}_3$ unabhängig sind, dann sind nach (1.3) $\hat{\mathfrak{O}}_0|R$, $\hat{\mathfrak{O}}_3|R$ unabhängig, und liegt R dicht in $(K,\hat{\mathfrak{O}}_3)$, da R $\hat{\mathfrak{O}}_0$-offen ist. Nach (2.2) gibt es daher wegen $\mathfrak{O}\subset\sup\{\hat{\mathfrak{O}}_0|R,\ \hat{\mathfrak{O}}_3|R\}$ Ringtopologien $\mathfrak{O}_1,\mathfrak{O}_2$ auf R mit $\mathfrak{O}=\sup\{\mathfrak{O}_1,\mathfrak{O}_2\}$, $\mathfrak{O}_1\subset\hat{\mathfrak{O}}_0|R=\mathfrak{O}_0$ und $\mathfrak{O}_2\subset\hat{\mathfrak{O}}_3|R$ und nach (1.6) eine Ringtopologie $\hat{\mathfrak{O}}_2$ auf K mit $\mathfrak{O}_2=\hat{\mathfrak{O}}_2|R$ und $\hat{\mathfrak{O}}_2\subset\hat{\mathfrak{O}}_3$. Da $\mathfrak{O}_2\subset\hat{\mathfrak{O}}_3|R$ und $\mathfrak{O}_0$, $\hat{\mathfrak{O}}_3|R$ unabhängig sind, sind auch $\mathfrak{O}_0$, $\mathfrak{O}_2$ unabhängig. Da B beschränkt und R dicht in $(K,\hat{\mathfrak{O}}_3)$ ist und $\hat{\mathfrak{O}}_0,\hat{\mathfrak{O}}_3$ unabhängig sind, ist wegen $\hat{\mathfrak{O}}_2\subset\hat{\mathfrak{O}}_3$ auch B beschränkt und R dicht in $(K,\hat{\mathfrak{O}}_2)$ und sind $\hat{\mathfrak{O}}_0,\hat{\mathfrak{O}}_2$ unabhängig. Ist $\mathfrak{O}$ lokalbeschränkt, dann ist nach (1.10) auch $\mathfrak{O}_2$ lokalbeschränkt. Da R dicht in $(K,\hat{\mathfrak{O}}_2)$ und $\mathfrak{O}_2=\hat{\mathfrak{O}}_2|R$ ist, ist auch $\hat{\mathfrak{O}}_2$ lokalbeschränkt und damit $\hat{\mathfrak{O}}_2\in\mathbb{T}$.

Zur Eindeutigkeitsaussage in (a): Seien $\mathfrak{O}_1', \mathfrak{O}_2'$ Ringtopologien auf R mit $\mathfrak{O} = \sup\{\mathfrak{O}_1', \mathfrak{O}_2'\}$, $\mathfrak{O}_1' \subset \mathfrak{O}_0$, so daß $\mathfrak{O}_2', \mathfrak{O}_0$ unabhängig sind. Wegen $\mathfrak{O}_2' \subset \sup\{\mathfrak{O}_0, \hat{\mathfrak{O}}_3 | R\}$ gibt es nach (2.2) Ringtopologien $\mathfrak{O}_1'', \mathfrak{O}_2''$ auf R mit $\mathfrak{O}_2' = \sup\{\mathfrak{O}_1'', \mathfrak{O}_2''\}$, $\mathfrak{O}_1'' \subset \mathfrak{O}_0$ und $\mathfrak{O}_2'' \subset \hat{\mathfrak{O}}_3 | R$. Da $\mathfrak{O}_1'' \subset \mathfrak{O}_0$, $\mathfrak{O}_1'' \subset \mathfrak{O}_2'$ und $\mathfrak{O}_0, \mathfrak{O}_2'$ unabhängig sind, ist nach (1.2) $\mathfrak{O}_1''$ trivial. Also ist $\mathfrak{O}_2' = \mathfrak{O}_2'' \subset \hat{\mathfrak{O}}_3 | R$, $\mathfrak{O}_2 \subset \hat{\mathfrak{O}}_3 | R$, $\mathfrak{O}_1' \subset \mathfrak{O}_0$, $\mathfrak{O}_1 \subset \mathfrak{O}_0$, $\sup\{\mathfrak{O}_1', \mathfrak{O}_2'\} = \sup\{\mathfrak{O}_1, \mathfrak{O}_2\}$; und die Eindeutigkeitsaussage in (2.2) liefert $\mathfrak{O}_1 = \mathfrak{O}_1'$, $\mathfrak{O}_2 = \mathfrak{O}_2'$.

Zur Eindeutigkeitsaussage in (b): Seien $\mathfrak{O}_1', \mathfrak{O}_2'$ Ringtopologien auf R mit $\mathfrak{O} = \sup\{\mathfrak{O}_1', \mathfrak{O}_2'\}$ und $\mathfrak{O}_1' \subset \mathfrak{O}_0$ und $\hat{\mathfrak{O}}_2'$ eine Ringtopologie auf K, so daß $\mathfrak{O}_2 = \hat{\mathfrak{O}}_2' | R$ und R dicht in $(K, \hat{\mathfrak{O}}_2')$ liegt. Dann liegt für $a \in R^*$, $b \in K$ auch $b + aR$ dicht in $(K, \hat{\mathfrak{O}}_2')$. Da $\{b + aR : a \in R^*, b \in K\}$ eine Basis der $\hat{\mathfrak{O}}_0$-offenen Mengen bilden, sind also nach (1.3) (a) $\hat{\mathfrak{O}}_0, \hat{\mathfrak{O}}_2'$ unabhängig und damit nach (1.3) (b) auch $\mathfrak{O}_0, \mathfrak{O}_2'$ unabhängig. Mit (a) folgt nun $\mathfrak{O}_1' = \mathfrak{O}_1$, $\mathfrak{O}_2' = \mathfrak{O}_2$.

Wir werden mit Hilfe von (4.2) gewisse Ringtopologien $\mathfrak{O}$ auf R zerlegen in der Form $\mathfrak{O} = \sup\{\mathfrak{O}_1, \hat{\mathfrak{O}}_2 | R\}$ mit einer Ringtopologie $\hat{\mathfrak{O}}_2$ auf K und einer Idealtopologie $\mathfrak{O}_1$ auf R. Bedingungen, unter denen bei den Gegebenheiten von (4.2) $\mathfrak{O}_1$ eine Idealtopologie ist, sind in (4.3), (4.8) angegeben.

(4.3) **Proposition.** *Hat jedes von* (0) *verschiedene Ideal in R endlichen Index, dann sind für eine Ringtopologie $\mathfrak{O}$ auf R äquivalent:*

(1) *$\mathfrak{O}$ ist eine nicht diskrete Idealtopologie.*

(2) *$\mathfrak{O}$ ist nicht diskret und beschränkt.*

(3) *Es gibt eine nicht diskrete Idealtopologie, die feiner als $\mathfrak{O}$ ist.*

Beweis. (1)$\curvearrowright$(2) ist trivial. (2)$\curvearrowright$(3): Offenbar ist $\{aR : a \in R^*\}$ eine Nullumgebungsbasis einer solchen Idealtopologie. (3)$\curvearrowright$(1) beweist man wie [11, (1.5) (5)$\curvearrowright$(1)] bzw. (4.8); vgl. [19, Lemma 1].

(4.4) **Satz.** *Voraussetzung: K sei ein globaler Körper, $\mathfrak{S}_0$ das normalisierte volle Bewertungssystem von K, $\mathfrak{S} := \{\varphi \in \mathfrak{S}_0 : R \subset O(\{\varphi\})\}$ und $\mathfrak{O}$ eine nicht diskrete, lokalbeschränkte Ringtopologie auf R.*

Behauptung: (a) *Dann gibt es eine Idealtopologie $\mathfrak{O}_1$ auf R und eine Menge $\mathfrak{S}' \subsetneq \mathfrak{S}_0 \setminus \mathfrak{S}$, so daß $\mathfrak{O} = \sup\{\mathfrak{O}_1, \mathfrak{O}(\mathfrak{S}') | R\}$ ist; $\mathfrak{O}_1$ und $\mathfrak{S}'$ sind dabei eindeutig bestimmt. R liegt dicht in $(K, \mathfrak{O}(\mathfrak{S}'))$.*

(b) *$\mathfrak{O}_1$, $\mathfrak{O}_2 := \mathfrak{O}(\mathfrak{S}') | R$ sind die einzigen Ringtopologien auf R, so daß $\mathfrak{O} = \sup\{\mathfrak{O}_1, \mathfrak{O}_2\}$, $\mathfrak{O}_1$ eine Idealtopologie und $\mathfrak{O}_2$, $\mathfrak{O}'$ für jede Idealtopologie $\mathfrak{O}'$ auf R unabhängig sind.*

$\mathfrak{O}_1, \mathfrak{O}_2$ sind die einzigen Ringtopologien auf R, so daß $\mathfrak{O} = \sup\{\mathfrak{O}_1, \mathfrak{O}_2\}$, $\mathfrak{O}_1$ eine Idealtopologie und $\mathfrak{O}_2$ sich zu einer Ringtopologie $\hat{\mathfrak{O}}_2$ auf K fortsetzen läßt, so daß R dicht in $(K, \hat{\mathfrak{O}}_2)$ liegt. Dabei ist auch $\hat{\mathfrak{O}}_2$ eindeutig bestimmt.

Beweis. Nach [13, (3.4)] sind für $B = \varnothing$ die Voraussetzungen von (4.2) im lokalbeschränkten Fall erfüllt; man beachte, daß nach dem strengen Approximationssatz für Mengen $\mathfrak{S}_1, \mathfrak{S}_2$ mit $\mathfrak{S}_1 \cup \mathfrak{S}_2 \subsetneq \mathfrak{S}_0$ $\mathfrak{O}(\mathfrak{S}_1)$, $\mathfrak{O}(\mathfrak{S}_2)$ genau dann unabhängig sind, wenn $\mathfrak{S}_1$ und $\mathfrak{S}_2$ disjunkt sind (s. [13, (1.4)], (1.2)).

Bei der Bezeichnung von (4.2) ist nach (3.1) (a) $\hat{\mathfrak{O}}_0 = \mathfrak{O}(\mathfrak{S})$. Wählt man $\mathfrak{O}_1, \mathfrak{O}_2, \hat{\mathfrak{O}}_2$ wie in (4.2), dann ist nach (3.1) (b), (4.3) $\mathfrak{O}_1$ eine Idealtopologie und wegen der Unabhängigkeit von $\hat{\mathfrak{O}}_0, \hat{\mathfrak{O}}_2$ für eine Menge $\mathfrak{S}' \subsetneq \mathfrak{S}_0 \setminus \mathfrak{S}$ $\hat{\mathfrak{O}}_2 = \mathfrak{O}(\mathfrak{S}')$ (s. [13, (3.4)]), also $\mathfrak{O}_2 = \hat{\mathfrak{O}}_2 | R = \mathfrak{O}(\mathfrak{S}') | R$. Für jede Menge $\mathfrak{S}'' \subsetneq \mathfrak{S}_0 \setminus \mathfrak{S}$ ist R eine

dichte Teilmenge von $(K, \mathfrak{O}(\$''))$ (s. (1.3) (b)). Die Eindeutigkeitsaussagen für $\mathfrak{O}_1, \mathfrak{O}_2$ folgen damit sofort aus (4.2). Die Eindeutigkeit der Topologie $\hat{\mathfrak{O}}_2$ in (b) folgt aus der Eindeutigkeitsaussage in (a). Wegen der Unabhängigkeit von $\hat{\mathfrak{O}}_0, \hat{\mathfrak{O}}_2$ (s. (4.2) (b)) ist nämlich für ein $\$'' \subsetneq \$_0 \setminus \$_0$ $\hat{\mathfrak{O}}_2 = \mathfrak{O}(\$'')$, $\mathfrak{O} = \sup\{\mathfrak{O}_1, \mathfrak{O}(\$'') | R\}$.

(4.5) **Folgerung.** *K sei ein globaler Körper und $\$_0, \$$ seien wie in* (4.4) *gewählt. Dann sind äquivalent:*

(1) *Jede lokalbeschränkte Ringtopologie auf R ist eine Idealtopologie.*

(2) *$\$_0 \setminus \$$ ist einelementig.*

(4.5) (2)$\curvearrowright$(1) folgt auch aus den Überlegungen von [13, Abschnitt 4], mit Hilfe von [13, (3.2)].

Hat K von 0 verschiedene Charakteristik, d.h. enthält $\$_0$ nur nicht archimedische Bewertungen, dann ist (2) äquivalent dazu, daß der ganze Abschluß von R in K ein minimaler ganz abgeschlossener Teilring T mit $T \cdot (T^*)^{-1} = K$ ist. Hat K die Charakteristik 0, dann sind bekanntlich die Aussagen (2) aus (4.5) und (4.6) äquivalent; es gilt also:

(4.6) **Folgerung.** (s. [3, Theorem 8]). *Für einen Teilring R eines algebraischen Zahlkörpers sind äquivalent:*

(1) *Jede lokalbeschränkte Ringtopologie auf R ist eine Idealtopologie.*

(2) *Für ein $d \in \mathbb{N}$ ist R ein Teilring des Rings der ganzen Zahlen in $\mathbb{Q}(\sqrt{-d})$.*

(4.7) **Lemma.** *Sei S ein Ring, M ein S-Modul, der die absteigende Kettenbedingung* (DCC) *für Teilmoduln erfüllt, und $\mathfrak{O}$ eine hausdorffsche Gruppentopologie auf M.*

(a) *Dann gibt es eine $\mathfrak{O}$-Nullumgebung, die keinen von* (0) *verschiedenen Teilmodul enthält.*

(b) *Gibt es zu jeder $\mathfrak{O}$-Nullumgebung U eine $\mathfrak{O}$-Nullumgebung V mit $SV \subset U$ und ist $Sv \neq (0)$ für alle $v \in M \setminus \{0\}$, dann ist $\mathfrak{O}$ diskret.*

Beweis. (a) Wir nehmen an, daß jede Nullumgebung einen von (0) verschiedenen Teilmodul enthält. Da $\mathfrak{O}$ hausdorffsch, gibt es dann eine Folge von Nullumgebungen $(U_n)_{n \in \mathbb{N}}$ und von Teilmoduln $(M_n)_{n \in \mathbb{N}}$ mit $(0) \neq M_n \subset U_n$, $M_n \not\subset U_{n+1}$, $U_{n+1} + U_{n+1} \subset U_n$ $(n \in \mathbb{N})$. Dann ist $\sum_{i=n+1}^{n+k} U_i \subset U_n$, also $N_n := \bigcup_{k=1}^{\infty} \sum_{i=n+1}^{n+k} M_i \subset \bigcup_{k=1}^{\infty} \sum_{i=n+1}^{n+k} U_i \subset U_n$. Wegen DCC ist die absteigende Folge (N_n) von Teilmoduln schließlich konstant; es gibt also ein $r \in \mathbb{N}$ mit $N_r = N_{r+2}$. Daraus folgt $M_{r+1} \subset N_r = N_{r+2} \subset U_{r+2}$, ein Widerspruch zu $M_{r+1} \not\subset U_{r+2}$.

(b) Wählt man eine Nullumgebung U gemäß (a), dazu eine Nullumgebung V mit $SV \subset U$, dann ist $V = (0)$.

(4.8) **Folgerung.** *Sei S ein Teilring von R, $1 \in S$; für jedes von* (0) *verschiedene Ideal A in R erfülle R/A die absteigende Kettenbedingung für S-Teilmoduln. Dann sind für eine Ringtopologie $\mathfrak{O}$ auf R äquivalent:*

(1) $\mathfrak{O}$ *ist eine nicht diskrete Idealtopologie.*

(2) $\mathfrak{O}$ *ist nicht diskret und beschränkt.*

(3) *Es gibt eine nicht diskrete Idealtopologie, die feiner als* $\mathfrak{O}$ *ist; und* S *ist beschränkt in* $(R, \mathfrak{O})$.

Beweis. (1)$\curvearrowright$(2)$\curvearrowright$(3) ist trivial (vgl. (4.3)).

(3)$\curvearrowright$(1): Sei U eine abgeschlossene $\mathfrak{O}$-Nullumgebung und A ein Ideal in R mit $(0) \neq A \subset U$. Dann ist auch der topologische Abschluß $\bar{A}$ ein Ideal in R, $\bar{A} \subset U$ und nach (4.7) (b) $(R, \mathfrak{O})/A$ diskret, also $\bar{A}$ offen.

(4.9) **Satz.** *K sei ein algebraischer Funktionenkörper in einer Variablen über dem Teilkörper* k, $\mathfrak{S}_0$ *ein volles Repräsentantensystem der Bewertungen auf* K, *die auf* k *trivial sind,* $k \subset R$, $\mathfrak{S} := \{\varphi \in S_0 : R \subset O(\{\varphi\})\}$ *und* $\mathfrak{O}$ *eine nicht diskrete, lokalbeschränkte Ringtopologie, bez. der* k *beschränkt ist.*

Dann gilt dieselbe Behauptung wie in (4.4).

Zum *Beweis* wendet man (4.2) mit $B = k$ an und schließt dabei wie im Beweis von (4.4), indem man [14, (4.4)], (3.3), (4.8) statt [13, (3.4)], (3.1), (4.3) anwendet und (1.10) beachtet.

(4.10) **Folgerung.** *Wählt man* K, k, R, $\mathfrak{S}_0$, $\mathfrak{S}$ *wie in* (4.9), *dann sind äquivalent:*

(1) *Jede lokalbeschränkte Ringtopologie auf* R, *bezüglich der* k *beschränkt ist, ist eine Idealtopologie.*

(2) $\mathfrak{S}_0 \setminus \mathfrak{S}$ *ist einelementig.*

(4.11) **Satz.** *Voraussetzung:* D *sei ein Dedekindring in* K, $D \cdot (D^*)^{-1} = K$, $D \subset R$. *Für jedes von* (0) *verschiedene Ideal* A *in* R *erfülle* R/A *die absteigende Kettenbedingung für* D*-Teilmoduln.*

$\mathfrak{S}_1$ *bezeichne die Menge der von den maximalen Idealen in* D *auf* K *induzierten Bewertungen und* $\mathfrak{S} := \{\varphi \in \mathfrak{S}_1 : R \subset O(\{\varphi\})\}$.

$\mathfrak{O}$ *sei eine nicht diskrete, lokalbeschränkte Ringtopologie auf* R, *bez. der* D *beschränkt ist.*

Behauptung: Dann gibt es eine Idealtopologie $\mathfrak{O}_1$ *auf* R *und eine Teilmenge* $\mathfrak{S}'$ *von* $\mathfrak{S}_1 \setminus \mathfrak{S}$, *so daß* $\mathfrak{O} = \sup\{\mathfrak{O}_1, \mathfrak{O}(S') | R\}$ *ist;* $\mathfrak{O}_1$ *und* $\mathfrak{S}'$ *sind dabei eindeutig bestimmt.* R *liegt dicht in* $(K, \mathfrak{O}(\mathfrak{S}'))$. *Ferner gilt die Aussage* (b) *von* (4.4).

Der *Beweis* verläuft wie der von (4.4), hier mit $B = D$; man benutzt dabei [13, (2.4)], (4.8) statt [13, (3.4)], (4.3) und beachtet, daß $R = O(\mathfrak{S})$ ist, oder nur, daß $\mathfrak{O}(\mathfrak{S})$ die von R erzeugte Topologie ist.

(4.12) **Satz.** $\mathfrak{S}_1$ *sei eine endliche Menge nicht archimedischer Bewertungen auf* K *und* $O(\mathfrak{S}_1) \subset R$. *Für jedes von* (0) *verschiedene Ideal* A *in* R *erfülle* R/A *die absteigende Kettenbedingung für* $O(\mathfrak{S}_1)$*-Teilmoduln.*

$\mathfrak{O}$ *sei eine nicht diskrete Ringtopologie auf* R, *bezüglich der* $O(\mathfrak{S}_1)$ *beschränkt ist.*

Dann gibt es genau eine Teilmenge $\mathfrak{S}$ *von* $\mathfrak{S}_1$, *so daß* $\mathfrak{O}(\mathfrak{S})$ *die von* R *erzeugte Ringtopologie ist. Mit diesem* $\mathfrak{S}$ *gilt die Behauptung von* (4.11). *Insbesondere ist* $\mathfrak{O}$ *lokalbeschränkt.*

Zum *Beweis* benutzt man (2.4) bzw. [12, (2.5)] und (4.2) mit $B = O(\mathfrak{S}_1)$.

Ist in (4.11) bzw. (4.12) für jedes von (0) verschiedene Ideal A der Faktorring R/A sogar endlich, dann genügt es statt der Beschränktheit von D bzw. $O(\mathfrak{S}_1)$ vorauszusetzen, daß $\mathfrak{O}$ gröber ist als $\mathfrak{O}(\mathfrak{S}_1)|R$. (Mit (4.11) bzw. (4.12) folgt

dann, daß sogar D beschränkt ist.) Ferner darf dann in (4.12) $\$_1$ auch archimedische Bewertungen enthalten.

Die generelle Voraussetzung dieses Abschnitts, daß R die Eins enthält, ist z.B. in (4.4) nicht wesentlich. Ist nämlich S ein Ring mit Einselement 1 und T ein Teilring von S, so daß $1 \notin T$ und S von $T \cup \{1\}$ erzeugt wird, dann ist das Nullumgebungssystem einer (lokalbeschränkten) Ringtopologie $\mathfrak{O}$ auf T eine Nullumgebungsbasis einer (lokalbeschränkten) Ringtopologie $\mathfrak{O}'$ auf S mit $\mathfrak{O} = \mathfrak{O}' | T$. Andererseits läßt sich mit dieser Überlegung z.B. nicht (2.4) für Ringe ohne 1 verallgemeinern. Für Ringtopologien $\mathfrak{O}_1, \mathfrak{O}_2$ auf T sind nämlich bei der Bezeichnung von eben $\mathfrak{O}_1', \mathfrak{O}_2'$ nicht unabhängig, da T offen-abgeschlossen bezüglich $\mathfrak{O}_1'$ und bezüglich $\mathfrak{O}_2'$ ist (s. (1.3) (a)).

Literatur

1. Cohen, J.A.: Topologies on the ring of integers of a global field. Pacific J. Math. **93**, 269–276 (1981)
2. De Wilde, M., Tsirulnikov, B.: Barrelledness and the supremum of two locally convex topologies. Math. Ann. **246**, 241–248 (1980)
3. Dikranjan, D., Wieslaw, W.: Rings with only ideal topologies. Comment. Math. Univ. St. Paul. **29**, 157–167 (1980)
4. Dress, A.: Eine Bemerkung über Teilringe globaler Körper. Abh. Math. Sem. Univ. Hamburg **28**, 133–138 (1965)
5. Kowalsky, H.J.: Beiträge zur topologischen Algebra. Math. Nachr. **11**, 143–185 (1954)
6. Prestel, A., Ziegler, M.: Model theoretic methods in the theory of topological fields. J. Reine Angew. Math. **299/300**, 318–341 (1978)
7. Rigo, T., Warner, S.: Topologies extending valuations. Canad. J. Math. **30**, 164–169 (1978)
8. Stenzel, M.: Inversenbildung und Komplettierung bei topologischen Ringen und Körpern. Dissertation, Bielefeld 1978
9. Stone, A.L.: Nonstandard Analysis in Topological Algebra. In: Applications of Model Theory to Algebra, Analysis and Probability (W.A.J. Luxemburg, ed.). Proceedings of an International Symposium (Pasadena 1967), pp. 285–299. New York-Toronto-London-Sidney: Holt, Rinehart and Winston 1969
10. Traynor, T.: The Lebesgue decomposition for group-valued set functions. Trans. Amer. Math. Soc. **220**, 307–319 (1976)
11. Weber, H.: Ringtopologien auf $\mathbb{Z}$ und $\mathbb{Q}$. Math. Z. **155**, 287–298 (1977)
12. Weber, H.: Zu einem Problem von H.J. Kowalsky. Abh. Braunschweig. Wiss. Gesellsch. **24**, 127–134 (1978)
13. Weber, H.: Charakterisierung der lokalbeschränkten Ringtopologien auf globalen Körpern. Math. Ann. **239**, 193–205 (1979)
14. Weber, H.: Topologische Charakterisierung globaler Körper und algebraischer Funktionenkörper in einer Variablen. Math. Z. **169**, 167–177 (1979)
15. Weber, H.: Die atomare Struktur topologischer Boolescher Ringe und s-beschränkter Inhalte. Studia Math. **74**, 1
16. Weber, H.: Topologische Boolesche Ringe. Zerlegung von Inhalten. Preprint
17. Wieslaw, W.: On some characterizations of the complex number field. Colloq. Math. **24**, 139–145 (1972)
18. Wieslaw, W.: Independent topologies on fields. Comment. Math. Univ. St. Paul. **26**, 201–208 (1977)
19. Wieslaw, W.: Locally bounded topologies an Euclidean rings. Arch. Math. (Basel) **31**, 33–37 (1978)
20. Wieslaw, W.: Locally bounded topologies on some Dedekind rings. Arch. Math. (Basel) **33**, 41–44 (1979)

Eingegangen am 9. September 1981

Math. Z. 180, 395–411 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On Groupoids Having a Linear Congruence Class Geometry

Thomas Ihringer

Fachbereich Mathematik, AG1, Technische Hochschule, Schloßgartenstraße 7, D-6100 Darmstadt, Federal Republic of Germany

1. Introduction

1.1. Let (A, F) be a (universal) algebra[1] and define, for all $x \in A$, $X \subseteq A$, $\pi(x|X)$ to be the congruence class $[x]\Theta(X)$ containing x of the smallest congruence relation $\Theta(X)$ of the algebra (A, F) identifying all elements of X. In accordance with Werner [12] and Wille [13] the pair (A, π) then is dented the *congruence class geometry* $\Gamma(A, F)$ of the algebra (A, F).

In order to encourage geometric intuition it will be useful to replace some algebraic by geometric notations. The elements of A will be considered as *points*, the empty set $\varnothing$ and the congruence classes of congruence relations of (A, F) as *subspaces* of $\Gamma(A, F)$. For $X \subseteq A$ let $[X]$ be the smallest subspace containing X. Clearly, for $x \in X$, $[X] = \pi(x|X)$. If $X \subseteq A$ is a two-element set, $[X]$ is called a *line* and, for all $x \in A$, $X \subseteq A$, $\pi(x|X)$ the subspace through x *parallel* to $[X]$. A map $f: A \to A$ which, for all $x, y \in A$, satisfies $f(x) \in \pi(f(y)|x, y)$ is denoted a *dilatation*. The dilatations are exactly the unary admissible operations of (A, F), therefore they completely determine the congruence relations of (A, F). Thus, two algebras (A, F_1) and (A, F_2) have the same congruence relations exactly if they have the same congruence class geometry. Let, for an algebra (A, F), F^0 denote the set of all unary operations derived from F by replacing in every nonnullary $f \in F$ all variables except one by constant elements of A. Then the following lemma is an immediate consequence of Proposition 6.1 on p. 87 of Cohn [3].

Lemma 1.1.1. *$\Gamma(A, F) = \Gamma(A, F^0)$ for every algebra (A, F).* □

1.2. In [13] Wille examines linear congruence class geometries: In accordance with Libois [9], a congruence class geometry is denoted *linear* if every line is specified uniquely by any two of its points or, more formally, if $y \in [x_1, x_2]$ and $x_1 \neq x_2$ always implies $x_1 \in [y, x_2]$.

[1] A pair (A, F) is called a (universal) algebra if F is a (possibly infinite) set of finitary operations on the set A. Cf. Cohn [3], Grätzer [4] and Wille [13] for the notations of Universal Algebra.

0025-5874/82/0180/0395/$03.40

The following proposition is a part of Satz 4.1 of Wille [13]. How strong the assumption of linearity on an algebra is will be illustrated by Theorem 2.1.4 which generalizes

Proposition 1.2.1. *A nonsimple group $(G,\cdot)$ has a linear congruence class geomtry if and only if $(G,\cdot)$ is an elementary abelian group.* □

Moreover, if $(G,\cdot)$ is a group, one may consider the set $L(G):=\{L_g|g\in G\}$ of left multiplication maps $L_g\colon G\to G$ defined by $x\mapsto g\cdot x$. The proof of Satz 4.11(2) of Wille [13] then implies

Proposition 1.2.2. *For a group $(G,\cdot)$ the congruence class geometry $\Gamma(G,L(G))$ is linear if and only if every non-identity element of G is of prime order.* □

1.3. In this section some basic tools on linear congruence class geometries are provided. Much of this information is contained in Chap. 4 of Wille [13].

A congruence class geometry will be called *nondegenerate* if it contains at least two different lines. It is a trivial observation that a congruence class geometry $\Gamma(A,F)$ is nondegenerate if and only if (A, F) is not simple. Every dilatation in a nondegenerate linear congruence class geometry is determined uniquely by the image of any two points. As a consequence, every nonconstant non-identity dilatation is injective and leaves at most one point fixed. Fixed point free dilatations, and the identity, are called *translations*, and nonconstant dilatations with exactly one fixed point *proper dilatations*. A congruence class geometry is called *transitive* if its nonconstant dilatations operate transitively on the set of points of the geometry, and *intransitive* otherwise. An algebra with transitive linear congruence class geometry is regular, i.e. each of its congruence classes determines uniquely the congruence relation to which it belongs. In this case $\pi(x|y,z)$ is a line for all $x,y,z\in A$, $y\neq z$.

If a nondegenerate linear congruence class geometry has only finitely many points, then its nonconstant dilatations form a permutation group on the points of the geometry, which will henceforth be called *dilatation group*.[2] Let now the dilation group of a nondegenerate finite linear congruence class geometry operate transitively on the set of points. Then, by a well-known theorem of Frobenius (cf. Theorem 16.8.8 in Hall, Jr. [5]), the translations of the geometry form a subgroup operating transitively on the set of points.

The following two propositions give a classification of the finite linear congruence class geometries basing on an examination of their dilatation groups.

Proposition 1.3.1. *There are exactly three types of nondegenerate finite linear congruence class geometries with transitive dilatation group.*

Type I: *consists of all finite affine desarguesian geometries. It is well-known that every such geometry is equal to the congruence class geometry $(V,+,K)$ of a finite vector space V over the field K with vector addition $+$. Let $n\in K$ be a generating element of the multiplicative group $K\setminus\{0\}$ of K, and define Φ to be the automorphism of $(V,+)$ with $x\mapsto nx$. Then $\Gamma(V,+,K)=\Gamma(V,L(V),\Phi)$.*

Type II: *In* [6] *Herzer proves that every finite nonabelian group $(G,\cdot)$, equipped with a nontrivial partition ρ and a ρ-admissible fixed point free automorphism Φ,*

[2] Cf. Hall, Jr. [5] and Huppert [7] for the group theoretic notations.

is of nilpotence class 2 and of prime exponent. Then the congruence class geometry of the algebra $(G, L(G), \Phi)$ is linear, and Φ can be chosen so that it generates the group of dilatations of $\Gamma(G, L(G), \Phi)$ which fix the identity element e of G.

Type III: *Let $(G, \cdot)$ be a finite nonabelian group in which every non-identity element is of prime order. Then the congruence class geometry $\Gamma(G, L(G))$ is linear, has no proper dilatation, and $\Gamma(G, L(G)) = \Gamma(G, L(G), \Phi)$ for $\Phi := \mathrm{id}_G$.*

Summarizing, for every nondegenerate finite linear congruence class geometry with transitive dilatation group there is a group $(G, \cdot)$ and an automorphism Φ of $(G, \cdot)$, such that the geometry is equal to $\Gamma(G, L(G), \Phi)$ and the following conditions are satisfied:

(i) *Φ generates the group of those nonconstant dilatations of $\Gamma(G, L(G), \Phi)$ which fix the identity e of G, and thus, for every integer i, Φ^i is either a fixed point free automorphism of $(G, \cdot)$, or $\Phi^i = \mathrm{id}_G$.*

(ii) *Every nonconstant dilatation of $\Gamma(G, L(G), \Phi)$ is of the form $x \mapsto g \cdot \Phi^i(x)$ for some $g \in G$ and a nonnegative integer $i < \mathrm{ord}\, \Phi$. It is a translation if and only if $i = 0$.*

$L(G)$ consists of all translations of $\Gamma(G, L(G), \Phi)$, and is, as a group, isomorphic to $(G, \cdot)$. In particular, the geometry is of type I if and only if $(G, \cdot)$ is abelian (cf. Proposition A of Pasini [10]).

(iii) *$\Gamma(G, L(G), \Phi^i)$ is linear for every integer i.* □

In such a representation of a finite linear congruence class geometry $(G, \cdot)$ is denoted a *group* and Φ an *automorphism associated with the geometry.*

If $(G, \cdot)$ is a group and I a set, then $L_I(G)$ is defined to be the set of all maps $f_g: G \times I \to G \times I$ with $(x, i) \mapsto (g \cdot x, i)$ for some $g \in G$. Two congruence class geometries (A, π_A) and (B, π_B) are called *isomorphic* if there is a bijective map ψ: $A \to B$ such that, for all $x \in A$, $X \subseteq A$, $\psi(\pi_A(x|X)) = \pi_B(\psi(x)|\psi(X))$.

Satz 4.11 (2) of Wille [13] yields

Proposition 1.3.2. *If $(G, \cdot)$ is a group in which every non-identity element is of prime order, and I is a set, then $(G \times I, L_I(G))$ has a linear congruence class geometry, which is intransitive if $|I| \geqq 2$ ($|I| \geqq 3$ if $(G, \cdot)$ is trivial).*

On the other hand, for every intransitive linear congruence class geometry there exists a group $(G, \cdot)$ and a set I such that the geometry is isomorphic to $\Gamma(G \times I, L_I(G))$.

The nonconstant dilatations of $\Gamma(G \times I, L_I(G))$ are exactly the maps in $L_I(G)$. All these maps are translations and form a group isomorphic to $(G, \cdot)$. □

In such a representation of an intransitive linear congruence class geometry $(G, \cdot)$ is denoted a *group* and I a *set associated with the geometry.*

2. On Groupoids $(G, \cdot)$ with Linear Congruence Class Geometry $\Gamma(G, \cdot)$

2.1. If $\cdot$ is an arbitrary binary operation on a set G, then $(G, \cdot)$ is denoted a *groupoid.*[3] $e \in G$ is called a *neutral element* (sometimes *identity*) of the groupoid

[3] Cf. Bruck [2] and Cohn [3] for information concerning groupoids, quasigroups, semigroups etc.

$(G,\cdot)$ if $e\cdot g=g\cdot e=g$ for all $g\in G$. $L(G,\cdot):=\{L_g|g\in G\}$ is defined to be the set of all *left multiplication maps* $L_g\colon G\to G$ with $x\mapsto g\cdot x$. Analogously, $R(G,\cdot):=\{R_g|g\in G\}$ is the set of all *right multiplication maps* $R_g\colon G\to G$ with $x\mapsto x\cdot g$. Sometimes $L(G)$ and $R(G)$ are used instead of $L(G,\cdot)$ and $R(G,\cdot)$ if the groupoid multiplication is clear from the context.

Lemma 2.1.1. *If $(G,\cdot)$ is a groupoid, then $\Gamma(G,\cdot)=\Gamma(G,L(G)\cup R(G))$.*

Proof. Immediate from Lemma 1.1.1. □

Lemma 2.1.2. *Let $(G,\cdot)$ be a nonsimple groupoid with linear congruence class geometry $\Gamma(G,\cdot)$, and $g,a,b\in G$, $a\neq b$. Then $g\cdot a=g\cdot b$ or $a\cdot g=b\cdot g$ implies L_g or R_g to be constant respectively.*

Proof. Trivial, because every L_g and every R_g is a dilatation and therefore has to be constant if it is not injective. □

A congruence class geometry (A,π) is called *discrete* if $[X]=X$ for all $X\subseteq A$. (A,π) then is linear. Obviously, for $|A|\neq 2$, (A,π) is discrete if and only if it has no nonconstant dilatation except the identity.

Lemma 2.1.3. *Let the algebra (A,F) have a linear congruence class geometry and a congruence relation with congruence classes $\{a\}$ and $A\setminus\{a\}$ for some $a\in A$. Then $\Gamma(A,F)$ is discrete.*

Proof. Let $|A|\geqq 3$, because $\Gamma(A,F)$ obviously is discrete otherwise. Then, by Satz 4.10 of Wille [13], the nonconstant dilatations of $\Gamma(A,F)$ do not operate transitively on A because (A,F) is not regular, as $\{a\}$ is a congruence class which does not determine uniquely the congruence relation to which it belongs. Therefore, by Proposition 1.3.2, the nonconstant dilatations of $\Gamma(A,F)$ are translations and form a group. Assume f to be a noncontant dilatation of $\Gamma(A,F)$ with $f(a)\neq a$. Then $f^{-1}(f(a))=a$ implies $f^{-1}(b)\neq a$ for every $b\in A\setminus\{a,f(a)\}$, because f^{-1} would be constant otherwise. But this contradicts the existence of a congruence relation with congruence classes $\{a\}$ and $A\setminus\{a\}$. Therefore every noncontant dilatation of $\Gamma(A,F)$ satisfies $f(a)=a$, and is a translation, thus is the identity on A. □

Theorem 2.1.4. *If $(G,\cdot)$ is a nonsimple groupoid with neutral element e and linear congruence class geometry $\Gamma(G,\cdot)$, then $(G,\cdot)$ is an elementary abelian group.*

The *proof* of Theorem 2.1.4 is divided into several single steps.

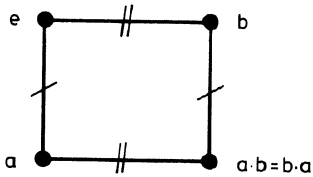
Claim 1. All maps in $L(G)$ and $R(G)$ are injective.

Proof of Claim 1. Assume L_g not to be injective for some $g\in G$. Then $g\neq e$, and there are two different elements $a,b\in G$ which satisfy $g\cdot a=g\cdot b$. By Lemma 2.1.2 therefore $g\cdot x=g\cdot g=g\cdot e=g$ for all $x\in G$, thus $e\cdot g=g=g\cdot g$, hence Lemma 2.1.2 yields $x\cdot g=g$ for all $x\in G$. $G\setminus\{g\}$ is then a subgroupoid of $(G,\cdot)$: If there would be elements $c,d\in G\setminus\{g\}$ with $c\cdot d=g$, then $c\cdot g=c\cdot d=g$, and, again by Lemma 2.1.2, $c=c\cdot e=g$, contradicting the choice of c. Accordingly, there is a congruence relation on $(G,\cdot)$ with the congruence classes $\{g\}$ and $G\setminus\{g\}$. Thus, by Lemma 2.1.3, every element of $L(G)\cup R(G)$ is a constant map

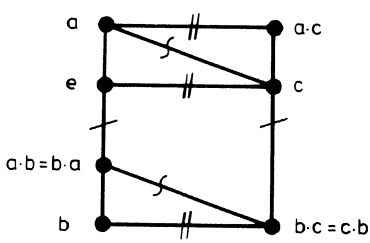
on G or the identity id_G. But, for all $c \neq e, g$, L_c cannot be constant, because otherwise $c = c \cdot e = c \cdot g = g$, and $L_c \neq \mathrm{id}_G$, because $L_c = \mathrm{id}_G$ would imply $c = c \cdot e = e$. Therefore $G = \{e, g\}$, and $(G, \cdot)$ thus is simple. This proves the injectivity of all L_g, $g \in G$. By symmetry the injectivity of all maps R_g can be proved. □

Claim 2. $(G, \cdot)$ is an abelian groupoid.

Proof of Claim 2. Let $a, b \in G \setminus \{e\}$, and $b \notin [e, a]$. Then $a \cdot b$, $b \cdot a \in (\pi(a|e, b) \cap \pi(b|e, a)$, i.e. $a \cdot b$ and $b \cdot a$ are contained in the intersection of two different lines. Linearity then yields $a \cdot b = b \cdot a$.



Now let $b \in [e, a]$, and $c \in G$ with $c \notin [e, a]$ (such an elememt exists, because $(G, \cdot)$ is nonsimple). As $b \cdot c = c \cdot b$, $a \cdot b$, $b \cdot a \in [e, a] \cap \pi(b \cdot c|a, c)$, which is the intersection of two different lines, whence $a \cdot b = b \cdot a$. □



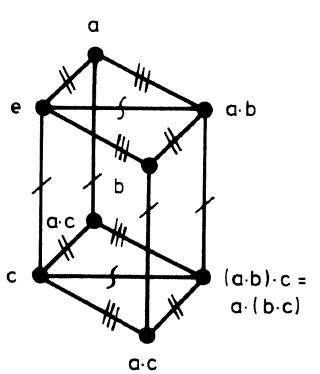
Claim 3. $(G, \cdot)$ is an associative groupoid.

Proof of Claim 3. Every congruence class geometry (A, π) satisfies, for all $x, y \in A$, $X, Y \subseteq A$, the rules

(C1) $\pi(x|\pi(y|X)) \subseteq \pi(x|X)$,

(C2) $X \subseteq Y \Rightarrow \pi(x|X) \subseteq \pi(x|Y)$,

(C3) $y \in \pi(x|X) \Rightarrow \pi(x|X) = \pi(y|X)$.



Let $a, b, c \in G \setminus \{e\}$. Then

$$a \cdot (b \cdot c) \in \pi(a \cdot b \mid b \cdot c, b) \subseteq \pi(a \cdot b \mid \pi(b \mid c, e)) \subseteq \pi(a \cdot b \mid c, e),$$

because of the rules (C1) and (C2). Obviously, $(a \cdot b) \cdot c \in \pi(a \cdot b \mid c, e)$. By rule (C3) $\pi(a \cdot b \mid c, e) = \pi((a \cdot b) \cdot c \mid c, e)$, i.e. $a \cdot (b \cdot c) \in \pi((a \cdot b) \cdot c \mid c, e)$. By interchanging a and c one obtains $c \cdot (b \cdot a) \in \pi((c \cdot b) \cdot a \mid a, e)$, hence $(c \cdot b) \cdot a \in \pi(c \cdot (b \cdot a) \mid a, e)$ by rule (C3). Claim 2 yields $(c \cdot b) \cdot a = a \cdot (b \cdot c)$ and $c \cdot (b \cdot a) = (a \cdot b) \cdot c$. Therefore $a \cdot (b \cdot c), (a \cdot b) \cdot c \in \pi((a \cdot b) \cdot c \mid c, e) \cap \pi((a \cdot b) \cdot c \mid a, e)$ which is the intersection of two different lines if $c \notin [e, a]$, whence in this case $a \cdot (b \cdot c) = (a \cdot b) \cdot c$ by linearity. Now let $c \in [e, a]$, but $b \notin [e, a]$. As proved above, $a \cdot (b \cdot c) \in \pi((a \cdot b) \cdot c \mid a, e)$. Interchanging of a and b yields $b \cdot (a \cdot c) \in \pi((b \cdot a) \cdot c \mid b, e)$. By claim 2 $b \cdot (a \cdot c) = b \cdot (c \cdot a)$ and $(b \cdot a) \cdot c = (a \cdot b) \cdot c$. $b \notin [e, a]$ implies $a \notin [e, b]$, thus, by claim 2 and the part of claim 3 which has already been proved, $b \cdot (a \cdot c) = b \cdot (c \cdot a) = (b \cdot c) \cdot a = a \cdot (b \cdot c)$, and $a \cdot (b \cdot c) \in \pi((a \cdot b) \cdot c \mid b, e)$. Therefore

$$a \cdot (b \cdot c), (a \cdot b) \cdot c \in \pi((a \cdot b) \cdot c \mid a, e) \cap \pi((a \cdot b) \cdot c \mid b, e)$$

which is the intersection of two different lines, whence $a \cdot (b \cdot c) = (a \cdot b) \cdot c$. The only case left is when $b, c \in [e, a]$. Then, for some $d \in G$ with $d \notin [e, a]$, one obtains

$$((a \cdot b) \cdot c) \cdot d = (a \cdot b) \cdot (c \cdot d) = a \cdot (b \cdot (c \cdot d)) = a \cdot ((b \cdot c) \cdot d) = (a \cdot (b \cdot c)) \cdot d,$$

as the associativity of $\cdot$ for any three points not lying alltogether on one line with e has already been proved. Claim 1 yields $(a \cdot b) \cdot c = a \cdot (b \cdot c)$. □

Claim 4. In $(G, \cdot)$ every element of G has an inverse.

Proof of Claim 4. By claim 1 $g \in G$ can have at most one inverse g^{-1}. For $g \in G$ and every positive integer i let $g^0 := e$, $g^i := g \cdot (g^{i-1})$ and, if g has an inverse g^{-1}, $g^{-i} := (g^{-1})^i$. Define, for $g \in G$,

$$K_g := \begin{cases} \{g^i \mid i \text{ nonnegative integer}\} & \text{if there exists no inverse of } g, \\ \{g^i \mid i \text{ integer}\} & \text{otherwise.} \end{cases}$$

Every K_g is a congruence class of $(G, \cdot)$: One has to show that $f(K_g) \cap K_g \neq \varnothing$ implies $f(K_g) \subseteq K_g$ for every $g \in G$ and every map f contained in the semigroup $S(L(G) \cup R(G))$ generated by $L(G) \cup R(G)$ (cf. Satz 1.6 in Wille [13]). As $L(G) = R(G)$ by claim 2, and as $L(G)$ is closed under composition ($L_g \circ L_h = L_{g \cdot h}$ for all $g, h \in G$ by claim 3), one obtains $S(L(G) \cup R(G)) = L(G)$. Assume g first to have no inverse, and let i, j be nonnegative integers such that $L_h(g^i) = g^j$ for some $h \in G$. It has to be shown that $L_h(g^k) \in K_g$ for all nonnegative integers k.

Case α) $j \geqq i$. Then, by claim 1, $h = g^{j-i}$ and $h \cdot g^k = g^{j-i+k} \in K_g$.

Case β) $j < i$. Then $h \cdot g^{i-j} \cdot g^j = h \cdot g^i = g^j = e \cdot g^j$, hence by claim 1 $h \cdot g^{i-j} = e$, thus $h \cdot g^{i-j-1} \cdot g = e$ and $h \cdot g^{i-j-1}$ is an inverse of g, contradicting the assumption on g.

If g has an inverse g^{-1}, and i, j are integers such that $L_h(g^i) = g^j$, then it is trivial to show that $L_h(g^k) \in K_g$ for all integers k, since $h \cdot g^i = g^j$ implies $h \cdot g^k = g^{k-i+j} \in K_g$

Now it can easily be proved that all elements of G have an inverse: As for every $g \in G$ $g^2 \in [e, g]$, either $g^2 = e$ and g thus is inverse to itself, or linearity of $\Gamma(G, \cdot)$ yields $g \in [e, g^2]$. As $[e, g^2] \subseteq K_2$, there is an integer $i \neq 0$ with $(g^2)^i = g$.

If $i > 0$, then $(g^2)^{i-1} \cdot g = e$, and g has an inverse $g^{-1} = (g^2)^{i-1}$.

If $i \leqq 0$, then $g \cdot (g^2)^{-i} = (g^2)^i \cdot (g^2)^{-i} = e$, and g has an inverse $g^{-1} = (g^2)^{-i}$. □

Proposition 1.2.1 completes the proof of Theorem 2.1.4, since claim 3 and claim 4 imply that $(G, \cdot)$ is a group. □

2.2. A groupoid $(Q, \cdot)$ is called a *quasigroup* if, for all $a \in Q$, L_a and R_a are bijective maps on Q. Therefore one can define, for all $a, b \in Q$, $a/b := R_b^{-1}(a)$ and $a \backslash b := L_a^{-1}(b)$. Very often the algebra $(Q, \cdot, /, \backslash)$ is considered instead of $(Q, \cdot)$.

Lemma 2.2.1. *The congruence class geometries $\Gamma(Q, \cdot)$ and $\Gamma(Q, L(Q, \cdot))$ of a quasigroup $(Q, \cdot)$ are transitive.*

Proof. By definition of a quasigroup for all $b, c \in Q$ there is an $a \in Q$ such that $L_a(b) = c$. □

Proposition 2.2.2. *If $(Q, \cdot)$ is a quasigroup with linear congruence class geometry $\Gamma(Q, \cdot)$, then $\Gamma(Q, \cdot) = \Gamma(Q, \cdot, /, \backslash)$, or, equivalently, $(Q, \cdot)$ and $(Q, \cdot, /, \backslash)$ have the same congruence relations.*

Proof. For a quasigroup $(Q, \cdot)$ let $L^{-1}(Q) := \{L_g^{-1} \mid g \in Q\}$, $R_g^{-1}(Q) := \{R_g^{-1} \mid g \in Q\}$. Then, as a consequence of Lemma 1.3(iii) in Smith [11],

$$\Gamma(Q, \cdot, /, \backslash) = \Gamma(Q, L(Q) \cup R(Q) \cup L^{-1}(Q) \cup R^{-1}(Q)),$$

and, by Lemma 2.1.1, $\Gamma(Q, \cdot) = \Gamma(Q, L(Q) \cup R(Q))$. Therefore $\Gamma(Q, \cdot) = \Gamma(Q, \cdot, /, \backslash)$ exactly if, for every $g \in Q$, L_g^{-1} and R_g^{-1} are admissible operations of $(Q, \cdot)$. If $(Q, \pi) := \Gamma(Q, \cdot)$, then, for all $a, b, g \in Q$, $a \in \pi(b \mid L_g^{-1}(a), L_g^{-1}(b))$, because $a = L_g(L_g^{-1}(a))$, $b = L_g(L_g^{-1}(b))$. Let $\Gamma(Q, \cdot)$ now be linear. By regularity then $L_g^{-1}(a) \in \pi(L_g^{-1}(b) \mid a, b)$, implying that L_g^{-1} is an admissible operation of $(Q, \cdot)$. Analogously every R_g^{-1}, $g \in Q$, can be shown to be an admissible operation of $(Q, \cdot)$. □

Lemma 2.2.3. *Let $(Q, \cdot)$ be a nonsimple quasigroup with linear congruence class geometry $\Gamma(Q, \cdot)$ such that, for all $g \in Q$, L_g and R_g are translations. Then $(Q, \cdot)$ is an elementary abelian group.*

Proof. Let $h \in Q$. Then there exists an element $g \in Q$ satisfying $g \cdot h = h$, because R_h is surjective. As L_g is a translation having h as fixed point, it must be the identity id_Q on Q, i.e. $g \cdot x = x$ for all $x \in Q$, in particular $g \cdot g = g$ which implies $R_g = \mathrm{id}_Q$, i.e. $x \cdot g = x$ for all $x \in Q$. Therefore g is a neutral element of $(Q, \cdot)$, and Theorem 2.1.4 yields the claim of the lemma. □

In Theorem 2.1.4 the assumption that the groupoid has a neutral element cannot be omitted. This is obvious from the following theorem, in which the finite quasigroups with linear congruence class geometry are characterized.

Theorem 2.2.4. (i) *Let V be a vector space over the finite field K with vector addition $+$, and define, for some $a \in V$, $m_1, m_2 \in K \backslash \{0\}$, the binary operation $*$ on*

V by $x*y:=a+m_1x+m_2y$. Then $(V,*)$ is a quasigroup with linear congruence class geometry $\Gamma(V,*)$. If K' is the subfield of K generated by m_1 and m_2, then $\Gamma(V,*)=\Gamma(V,+,K')$.

(ii) *The only nonsimple finite quasigroups $(Q,*)$ with linear congruence class geometry $\Gamma(Q,*)$ are those defined in* (i). *Therefore the congruence class geometry of every such quasigroup is affine and desarguesian.*

Proof. (i) It is an easy calculation that $x*y=a+m_1x+m_2y$ defines a quasigroup structure on V. By Lemma 2.1.1 $\Gamma(V,*)=\Gamma(V,L(V,*)\cup R(V,*))$. For every $v\in V$ $L_v\in L(V,*)$ maps x to $a+m_1v+m_2x$ and $R_v\in R(V,*)$ maps x to $a+m_2v+m_1x$, thus all maps in $L(V,*)\cup R(V,*)$ are dilatations of $\Gamma(V,+,K')$. Let $b\in V$, $u:=\frac{1}{m_1}(b-a)$, $v:=-\frac{1}{m_1}a$, $w:=-\frac{1}{m_2}a$, and k a positive integer such that $m_2^k=1$. Then $L_u, L_v, R_w\in L(V,*)$ satisfy $L_u\circ L_v^{k-1}(x)=b+x$, $R_w(x)=m_1x$, $L_v(x)=m_2x$ for all $x\in V$. Therefore all translations of $\Gamma(V,+,K')$ and the dilatations of $\Gamma(V,+,K')$ with $x\mapsto m_1x$ and $x\mapsto m_2x$ are dilatations of $\Gamma(V,*)$.

Let $M:=\{m\in K|f\colon V\to V$ with $x\mapsto mx$ is a dilatation of $\Gamma(V,*)\}$. It is easy to show that M is closed under the addition and the multiplication of the field K. This proves $M=K'$, as $m_1, m_2\in M$. Hence $\Gamma(V,*)$ and $\Gamma(V,+,K')$ have the same dilatations, thus $\Gamma(V,*)=\Gamma(V,+,K')$.

(ii) Let the finite nonsimple quasigroup $(Q,*)$ have a linear congruence class geometry $\Gamma(Q,*)$. Lemma 2.2.1 and Proposition 1.3.1 imply the existence of a group $(Q,\cdot)$ on Q and of an automorphism Φ of $(Q,\cdot)$ which are associated to $\Gamma(Q,*)$, i.e. $\Gamma(Q,*)=\Gamma(Q,L(Q,\cdot),\Phi)$ and Φ generates the group of those dilatations of $\Gamma(Q,*)$ which fix the identity e of $(Q,\cdot)$. Thus, by Proposition 1.3.1(ii), for every $g\in Q$ exist $a(g)$, $b(g)\in Q$ and $i(g)$, $j(g)\in\{0,1,2,\ldots,\operatorname{ord}\Phi-1\}$ such that $L_g\in L(Q,*)$, $R_g\in R(Q,*)$ satisfy $R_g(x)=a(g)\cdot\Phi^{i(g)}(x)$ and $L_g(x)=b(g)\cdot\Phi^{j(g)}(x)$. Let now $g_1,g_2\in Q$, $g_1\neq g_2$, $b_1:=b(g_1)$, $b_2:=b(g_2)$, $j_1:=j(g_1)$, $j_2:=j(g_2)$. Then, for all $x\in Q$, $g_1*x\neq g_2*x$, thus $b_1\cdot\Phi^{j_1}(x)\neq b_2\cdot\Phi^{j_2}(x)$, hence $\Phi^{-j_1}(b_2^{-1}\cdot b_1)\neq\Phi^{j_2-j_1}(x)\cdot x^{-1}$. This implies $j_1=j_2$, because otherwise, by Proposition 1.3.1(i), $\Phi^{j_2-j_1}$ would be a fixed point free automorphism of $(Q,\cdot)$, and therefore $x\mapsto\Phi^{j_2-j_1}(x)\cdot x^{-1}$ a surjective map on Q (cf. Satz 8.9b on p. 500 of Huppert [7]). Accordingly, there is an integer j and, by an analogous argument, an integer i such that $j=j(g)$ and $i=i(g)$ for all $g\in Q$. Let, for the identity e of $(Q,\cdot)$, $a:=a(e)$. Then, for every $x\in Q$, $b(x)=b(x)\cdot\Phi^j(e)=x*e=a(e)\cdot\Phi^i(x)=a\cdot\Phi^i(x)$, thus $x*y=b(x)\cdot\Phi^j(y)=a\cdot\Phi^i(x)\cdot\Phi^j(y)$ for all $x,y\in Q$. The next step will be to prove that the group $(Q,\cdot)$ associated with $\Gamma(Q,*)$ is abelian. Let $y\in Q\setminus\{e\}$, $z:=\Phi^{i-j}(y)$, and f_1,f_2,f_3 the dilatations of $\Gamma(Q,*)$ defined by $f_1(x):=x*z$, $f_2(x):=a^{-1}\cdot x$, $f_3(x):=\Phi^{-i}(x)$. Then the map with $x\mapsto x\cdot y$ is a dilatation of $\Gamma(Q,*)$, because $x\cdot y=f_3\circ f_2\circ f_1(x)$ for all $x\in Q$. Hence it is a translation, since it leaves no point of Q fixed. As the left multiplication by y in $(Q,\cdot)$ also is a translation of $\Gamma(Q,*)$, and every translation in this geometry is uniquely determined by the image of one point, $y\cdot e=e\cdot y$ implies $y\cdot x=x\cdot y$ for all $x\in Q$. This proves that $(Q,\cdot)$ is an abelian group. Thus, by Proposition 1.3.1(ii), $\Gamma(Q,*)$ is of type I and therefore an affine and desarguesian geometry. Hence $(Q,\cdot)$ can be regarded as the additive group $(Q,+)$ of a vector space Q over some field K, and there is an $n\in K$ such that $\Phi(x)=nx$ for all $x\in Q$. Thus

the quasigroup operation can be written as $x * y = a + m_1 x + m_2 y$ if one defines $m_1 := n^i$, $m_2 := n^j$. □

Remark 2.2.5. One of the quasigroups $(V, *)$ in Theorem 2.2.4 is associative if and only if $m_1 = m_2 = 1$. In this case the map f: $(V, +) \to (V, *)$ with $x \mapsto x - a$ is a quasigroup isomorphism, whence $(V, *)$ is a group which is elementary abelian by Proposition 1.2.1. □

2.3. An associative groupoid $(S, \cdot)$ is denoted a *semigroup*. A semigroup $(S, \cdot)$ is called *constant*, if $\cdot$ is a constant operation on S. The semigroup $(S, \cdot)$ which is defined by $x \cdot y = x$ for all $x, y \in S$ is called the *left trivial semigroup* on S or, if $x \cdot y = y$ for all $x, y \in S$, the *right trivial semigroup* on S. The congruence class geometry of every constant, left trivial or right trivial semigroup is discrete and therefore linear. A semigroup $(S, \cdot)$ is said to satisfy the *cancellation law* if all maps in $L(S) \cup R(S)$ are injective.

Theorem 2.3.1. *A nonsimple semigroup $(S, \cdot)$ with linear congruence class geometry $\Gamma(S, \cdot)$ either satisfies the cancellation law, or is left trivial, or right trivial, or constant.*

Proof. It is sufficient to show that a noninjective map in $L(S) \cup R(S)$ implies $(S, \cdot)$ to be left trivial, or right trivial, or constant. Let therefore $L_a \in L(S)$ be not injective for some $a \in S$, and assume $(S, \cdot)$ not to be left trivial, right trivial, or constant. Then, by Lemma 2.1.2, L_a is constant and there is an element $s \in S$ with $L_a(x) = s$ for all $x \in S$, thus $s = a \cdot (s \cdot x) = (a \cdot s) \cdot x = s \cdot x$ for all $x \in S$, hence $L_s = L_a$. Let $s \neq a$. Then, for all $g \in S$, $R_g(s) = s \cdot g = L_s(g) = L_a(g) = a \cdot g = R_g(a)$, whence R_g is constant and, for all $x, y \in S$, $x \cdot y = R_y(x) = R_y(s) = s \cdot y = L_s(y) = s$, thus $(S, \cdot)$ turns out to be constant, contradicting the assumption. Therefore $s = a$. If there are $b, c \in S \setminus \{s\}$ such that $b \cdot c = s$, then $R_c(b) = b \cdot c = s = L_s(c) = s \cdot c = R_c(s)$ which, by similar arguments as used above, implies R_c to be constant and $R_s = R_c$, thus again $\cdot$ to be constant. Therefore $b, c \neq s$ yields $b \cdot c \neq s$, whence $\{s\}$ and $S \setminus \{s\}$ are the congruence classes of a congruence relation of $(S, \cdot)$. By Lemma 2.1.3 therefore all maps in $L(S) \cup R(S)$ are constant or the identity id_S on S. But $\mathrm{id}_S \notin L(S) \cup R(S)$, because $L_g(s) = s$ or $R_g(s) = s$ only if $g = s$. Hence, for all $x, y \in S$, $x \cdot y = L_x(y) = L_x(s) = x \cdot s = R_s(x) = s$, contradicting the assumption. Therefore the only case left is, when $b \cdot c \neq s$ for all $b, c \in S$, $b \neq s$. But then again $\{s\}$ and $S \setminus \{s\}$ are the congruence classes of a congruence relation of $(S, \cdot)$ and, by Lemma 2.1.3, all maps in $L(S) \cup R(S)$ are constant or the identity id_S. For every $g \in S$ $R_g = \mathrm{id}_S$, because otherwise, for some $b \in G \setminus \{s\}$, $b \cdot g = R_g(b) = R_g(s) = s \cdot g = L_s(g) = s$. Hence $x \cdot y = R_y(x) = x$ for all $x, y \in S$. Thus $(S, \cdot)$ is left trivial, contradicting the assumption. Analogously one can show that the existence of a noninjective map in $R(S)$ implies the semigroup to be left trivial, right trivial, or constant. □

As every finite semigroup satisfying the cancellation law is a quasigroup, Theorem 2.3.1, Theorem 2.2.4 and Remark 2.2.5 yield

Theorem 2.3.2. *A finite nonsimple semigroup $(S, \cdot)$ with linear congruence class geometry $\Gamma(S, \cdot)$ either is left trivial, or right trivial, or constant, or an elementary abelian group.* □

2.4. In Theorem 2.4.4 it will be proved that a nonsimple finite groupoid with linear congruence class geometry either is constant, or a quasigroup, or one of the groupoids of

Definition 2.4.1. Let $(G,\cdot)$ be a group with identity e, I a set and $g\in G$. Then define, for all $x,y\in G$, $i,j\in I$,

$$(x,i)\oplus(y,j):=(g\cdot y,j),$$
$$(x,i)\otimes(y,j):=(g\cdot x,i).$$

Lemma 2.4.2. *Let $(G,\cdot)$ be a group of prime order, or of order* 1. *Then the groupoids $(G\times I,\oplus)$ and $(G\times I,\otimes)$ have linear congruence class geometries: For $g=e$ $\Gamma(G\times I,\oplus)$ and $\Gamma(G\times I,\otimes)$ are discrete, for $g\neq e$ $\Gamma(G\times I,\oplus)=\Gamma(G\times I,\otimes)=\Gamma(G\times I, L_I(G))$.* □

Proposition 2.4.3. *Every left trivial and every right trivial semigroup is isomorphic to one of the groupoids of Lemma* 2.4.2.

Proof. Let $(S,*)$ be the right trivial semigroup on S, i.e. $x*y=y$ for all $x,y\in S$. Define $(G,\cdot)$ to be the one-element group, $I:=S$, and $g:=e$. Then $(e,i)\oplus(e,j)=(e,j)$ for all $i,j\in I$. Therefore the map $\Psi\colon S\to G\times I$ with $i\mapsto(e,i)$ obviously is a groupoid isomorphism. Analogously it can be shown that the left trivial semigroup $(S,*)$ on S is isomorphic to one of the groupoids of Lemma 2.4.2 with multiplication $\otimes$. □

Theorem 2.4.4. *A nonsimple finite groupoid $(G,*)$ with linear congruence class geometry $\Gamma(G,*)$ either is constant, or one of the quasigroups of Theorem* 2.2.4, *or isomorphic to one of the groupoids of Lemma* 2.4.2.

Proof. In order to prove the theorem it is sufficient to show that every nonconstant groupoid, which is not a quasigroup and satisfies the assumptions of the theorem, is isomorphic to one of the groupoids of Lemma 2.4.2. Let therefore $(G,*)$ be such a groupoid. The defining properties of a quasigroup then imply that there is a nonbijective map in $L(G,*)\cup R(G,*)$: W.l.o.g. assume $L_a\in L(G,*)$ to be nonbijective for some $a\in G$. As G is finite, L_a is noninjective, hence a constant map by linearity of $(G,*)$. Thus there is a $c\in G$ such that $L_a(x)=a*x=c$ for all $x\in G$. Then two different cases can be distinguished.

Case 1. Let there be an element $b\in G\setminus\{a\}$ such that $L_b\in L(G,*)$ is a constant map. Then there is a $d\in G$ with $L_b(x)=b*x=d$ for all $x\in G$. Hence, for all $y\in G$, $R_y(a)=a*y=L_a(y)=c$ and $R_y(b)=b*y=L_b(y)=d$, thus all maps in $R(G,*)$ are equal. This implies all maps in $L(G,*)$ to be constant: For all $x,y\in G$ $L_x(y)=x*y=R_y(x)$, i.e. $L_x(y)$ does not depend on the choice of y. Let $f\colon G\to G$ with $x\mapsto f(x)$ be the unique map in $R(G,*)$. If f is constant, then $x*y=R_y(x)=R_y(a)=a*y=L_a(y)=c$ for all $x,y\in G$, and therefore $*$ is a constant operation on G. If f is the identity on G, then $x*y=R_y(x)=f(x)=x$ for all $x,y\in G$, hence $(G,*)$ is the left trivial semigroup on G and Proposition 2.4.3 implies $(G,*)$ to be one of the groupoids of Lemma 2.4.2. Now let f be a nonconstant non-identity map on G and $z\in G$ such that $f(z)\neq z$. Then $\Theta:=\{(x,y)\in G\times G\mid x=y$ or $x,y\in[z,f(z)]\}$ is a congruence relation of $(G,*)$, as

$f[z,f(z)]\subseteq[z,f(z)]$ and $\Gamma(G,*)=\Gamma(G,f)$. Θ proves $(G,*)$ not to be regular, hence $\Gamma(G,*)$ to be intransitive. Then, by Proposition 1.3.2, there exists a group $(H,\cdot)$ and a finite set I such that $\Gamma(G,*)$ and $\Gamma(H\times I,\ L_I(G))$ are isomorphic congruence class geometries. This isomorphism induces a binary operation $*$ on $H\times I$ such that $(G,*)$ and $(H\times I,*)$ are isomorphic groupoids, and $\Gamma(H\times I,*)=\Gamma(H\times I,L_I(H))$. The map $f\in R(G,*)$ then corresponds, for some non-identity element $h\in H$, to the map $f_h\in L_I(H)$. For all $x,y\in H$, $i,j\in I$ and $R_{(y,j)}\in R(H\times I,*)$ therefore $(x,i)*(y,j)=R_{(y,j)}(x,i)=f_g(x,i)=(g\cdot x,i)$, thus $(H\times I,*)$ is one of the groupoids of Definition 2.4.1. As the line $[z,f(z)]$ is a set of transitivity of $\Gamma(G,*)$, it can easily be shown that $(H,\cdot)$ is a group of prime order, whence $(H\times I,*)$ is one of the groupoids of Lemma 2.4.2.

Case 2. For every $x\in G\setminus\{a\}$ let $L_x\in L(G,*)$ be a nonconstant, i.e. a bijective map. Choose some $d\in G\setminus\{a\}$. Then there is a $b\in G$ such that $L_d(b)=c$, hence $R_b(d)=d*b=L_d(b)=c=L_a(b)=a*b=R_b(a)$ which implies $R_b(x)=c$ for all $x\in G$. For every $y\in G\setminus\{b\}$ R_y then is a nonconstant map, because otherwise, for any $x\in G\setminus\{a\}$, $L_x(y)=x*y=R_y(x)=R_y(a)=a*y=L_a(y)=c$, contradicting the injectivity of L_x. Now it can be proved that $\Gamma(G,*)$ is transitive. Let $u,v\in G$. For $u\neq a$ or $u\neq b$ there are nonconstant maps in $L(G,*)$ or $R(G,*)$ respectively mapping u to v. Thus $\Gamma(G,*)$ obviously is transitive if $a\neq b$. For $a=b$ $G\setminus\{a\}$ is contained in a set of transitivity of $\Gamma(G,*)$. If $G\setminus\{a\}$ itself would be a set of transitivity, then $\{a\}$ would be one, implying $G\setminus\{a\}$ to contain exactly one point, thus G to be a two-element set, which contradicts the assumption that $(G,*)$ is non-simple. As $\Gamma(G,*)$ therefore is transitive, Proposition 1.3.1 yields the existence of a group $(G,\cdot)$ on G and of an automorphism Φ of $(G,\cdot)$ such that $\Gamma(G,*)=\Gamma(G,L(G,\cdot),\Phi)$. For every $g\in G\setminus\{a\}$ then there are $r(g)\in G$ and $i(g)\in\{0,1,2,\ldots,\operatorname{ord}\Phi-1\}$ such that $L_g\in L(G,*)$ maps x to $r(g)\cdot\Phi^{i(g)}(x)$. For the identity e of $(G,\cdot)$ and some $g\in G\setminus\{e\}$ $\{e\}\cup\{\Phi^i(g)\mid i=0,1,2,\ldots,\operatorname{ord}\Phi-1\}$ is an $(\operatorname{ord}\Phi+1)$-element set which is contained in the line $[e,g]$. This implies $\operatorname{ord}\Phi<|G|-1$, as $\Gamma(G,*)$ is nondegenerate. Thus there are two different elements $g_1,g_2\in G\setminus\{a\}$ such that $i(g_1)=i(g_2)$, hence $r(g_1)\cdot\Phi^{i(g_1)}(b)=L_{g_1}(b)=c=L_{g_2}(b)=r(g_2)\cdot\Phi^{i(g_2)}(b)$ which implies $r(g_1)=r(g_2)$, and thus $L_{g_1}=L_{g_2}$. But then all maps in $R(G,*)$ are constant, because one obtains $R_y(g_1)=g_1*y=L_{g_1}(y)=L_{g_2}(y)=g_2*y=R_y(g_2)$ for all $y\in G$. This shows that the assumptions of case 2) lead to a contradiction, since R_y is nonconstant for $y\neq b$. This completes the proof of the theorem. □

3. On Groupoids $(G,\cdot)$ with Linear Congruence Class Geometry $\Gamma(G,L(G))$

3.1. In this chapter the following analogue of Lemma 2.1.2 will be used.

Lemma 3.1.1. *Let $(G,\cdot)$ be a groupoid with nondegenerate linear congruence class geometry $\Gamma(G,L(G))$, and $g,a,b\in G$, $a\neq b$. Then $g\cdot a=g\cdot b$ implies L_g to be constant.*

Proof. Trivial, because every L_g is a dilatation of $\Gamma(G,L(G))$ and has therefore to be contant if it is not injective. □

For a congruence class geometry (G,π) there exists a groupoid $(G,\cdot)$ with $(G,\pi)=\Gamma(G,L(G))$ if and only if there is a subset L of all dilatations of (G,π) such that $|L|\subseteq|G|$ and $(G,\pi)=\Gamma(G,L)$: The groupoid operation $\cdot$ then can be so defined that $L\subseteq L(G,\cdot)$. It is therefore an easy observation from Propositions 1.3.1 and 1.3.2 that every nondegenerate finite linear congruence class geometry can in many different ways be represented as the congruence class geometry $\Gamma(G,L(G))$ of a groupoid $(G,\cdot)$.

3.2. The following theorem, which is analogous to Theorem 2.2.4, uses the representation of transitive finite linear congruence class geometries given in Proposition 1.3.1.

Theorem 3.2.1. (i) *Let $(G,\cdot)$ be a group and Φ an automorphism of $(G,\cdot)$ which are associated with a transitive finite linear congruence class geometry, f: $G\to G$ a bijective map with $x\mapsto f(x)$, i an integer, and define the binary operation $*$ of G by $x*y:=f(x)\cdot\Phi^i(y)$. Then $(G,*)$ is a quasigroup with linear congruence class geometry $\Gamma(G,L(G,*))=\Gamma(G,L(G,\cdot),\Phi^i)$.*

(ii) *The only finite quasigroups $(Q,*)$ with nondegenerate linear congruence class geometry $\Gamma(Q,L(Q,*))$ are those defined in* (i).

Proof. (i) It is an easy calculation that $x*y=f(x)\cdot\Phi^i(y)$ defines a quasigroup structure on G. For every $g\in G$ $L_g\in L(G,*)$ maps x to $f(g)\cdot\Phi^i(x)$, thus all maps in $L(G,*)$ are dilatations of $\Gamma(G,L(G,\cdot),\Phi^i)$. Let $a\in G$, e the identity of $(G,\cdot)$, $u:=f^{-1}(a)$, $v:=f^{-1}(e)$, and $k:=\operatorname{ord}\Phi$. Then $L_u, L_v\in L(G,*)$ satisfy $L_u\circ L_v^{k-1}(x)=a\cdot x$, $L_v(x)=\Phi^i(x)$ for all $x\in G$. As therefore all translations of $\Gamma(G,L(G,\cdot),\Phi^i)$ and Φ^i are dilatations of $\Gamma(G,*)$, $\Gamma(G,*)$ and $\Gamma(G,L(G,\cdot),\Phi^i)$ have the same dilatations. This proves $\Gamma(G,*)=\Gamma(G,L(G,\cdot),\Phi^i)$.

(ii) Let the finite quasigroup $(Q,*)$ have a nondegenerate linear congruence class geometry $\Gamma(Q,L(Q,*))$. Lemma 2.2.1 and Proposition 1.3.1 imply the existence of a group $(Q,\cdot)$ on Q and of an automorphism Φ of $(Q,\cdot)$ which are associated with $\Gamma(Q,*)$, i.e. $\Gamma(Q,*)=\Gamma(Q,L(Q,\cdot),\Phi)$ and Φ generates the group of those dilatations of $\Gamma(Q,*)$ which fix the identity e of $(Q,\cdot)$. For every $g\in Q$ thus exist, by Proposition 1.3.1(ii), $f(g)\in Q$ and $i(g)\in\{0,1,2,\ldots,\operatorname{ord}\Phi-1\}$ such that $L_g\in L(Q,*)$ satisfies $L_g(x)=f(g)\cdot\Phi^{i(g)}(x)$. Let now $g_1,g_2\in Q$, $g_1\neq g_2$, $f_1:=f(g_1)$, $f_2:=f(g_2)$, $i_1:=i(g_1)$, $i_2:=i(g_2)$. Then, for all $x\in Q$, $g_1*x\neq g_2*x$, thus $f_1\cdot\Phi^{i_1}(x)\neq f_2\cdot\Phi^{i_2}(x)$, hence $\Phi^{-i_1}(f_2^{-1}\cdot f_1)\neq\Phi^{i_2-i_1}(x)\cdot x^{-1}$. This implies $i_1=i_2$, because otherwise, by Proposition 1.3.1 (i), $\Phi^{i_2-i_1}$ would be a fixed point free automorphism of $(Q,\cdot)$, and therefore $x\mapsto\Phi^{i_2-i_1}(x)\cdot x^{-1}$ a surjective map on Q (cf. Satz 8.9b on p. 500 of Huppert [7]). Accordingly, there is an integer i such that $i=i(g)$ for all $g\in Q$, thus $x*y=f(x)\cdot\Phi^i(y)$ for all $x,y\in Q$. The map f: $Q\to Q$ with $x\mapsto f(x)$ is injective, because $f(x_1)=f(x_2)$ for two different elements $x_1,x_2\in Q$ would imply $x_1*e=x_2*e$ which contradicts the defining properties of a quasigroup. □

Corollary 3.2.2. *If $(Q,*)$ is a nonsimple finite quasigroup with linear congruence class geometry $\Gamma(Q,*)$, then $\Gamma(Q,L(Q,*))$ also is linear.*

Proof. If $(Q,\cdot)$ is a group and Φ an automorphism of $(Q,\cdot)$ which are associated with $\Gamma(Q,*)$, then there are $a\in Q$, $i,j\in\{0,1,2,\ldots,\operatorname{ord}\Phi-1\}$ such that $x*y$

$= a \cdot \Phi^i(x) \cdot \Phi^j(y)$ for all $x, y \in Q$, as was shown in the proof of Theorem 2.2.4. Therefore $(Q, *)$ is one of the quasigroups of Theorem 3.2.1, because $f\colon Q \to Q$ with $x \mapsto a \cdot \Phi^i(x)$ is a bijective map. □

A quasigroup with neutral element is called a *loop*.

Theorem 3.2.3. *For a loop $(Q, *)$ with nondegenerate linear congruence class geometry $\Gamma(Q, L(Q, *))$ all left multiplications are translations.*

Proof. Assume, for some $a, b \in Q$, L_a to have the fixed point b. Then $a * b = L_a(b) = b = e * b$, hence $a = e$, $L_a = L_e = \mathrm{id}_Q$, and L_a therefore is a translation. □

Theorem 3.2.4. *A finite loop $(Q, *)$ with nondegenerate linear congruence class geometry $\Gamma(Q, L(Q, *))$ is a group in which every non-identity element is of prime order.*

Proof. Let $x, y, z \in Q$. As the translations of a nondegenerate finite linear congruence class geometry form a group and every translation is determined uniquely by the image of one point, one obtains $L_x \circ L_y = L_{x*y}$, because $L_x \circ L_y$ and L_{x*y} are both translations mapping e to $x * y$. Thus $L_x \circ L_y(z) = L_{x*y}(z)$, i.e. $x * (y * z) = (x * y) * z$, proving the associativity of $(Q, *)$. But as every associative loop is a group, Lemma 1.2.2. can be applied to show that in $(Q, *)$ every non-identity element is of prime order. □

3.3. As next step a class of semigroups is defined which will turn out to have some importance in this section.

Definition 3.3.1. Let $(G, \cdot)$ be a group, I, I_0, I_1, I_2 sets with $I = I_0 \cup I_1$, $I_2 \subseteq I_1$, and $a\colon I_0 \to G$, $b\colon I_1 \to G$, $k\colon I_1 \to I_2$ maps with $i \mapsto a_i$, $i \mapsto b_i$, $i \mapsto k_i$ respectively.

Let, for all $i \in I_2$, the following conditions be satisfied:

(i) $b_i = e$ (the identity of G),

(ii) $k_i = i$.

Then define, for all $x, y \in G$, $i, j \in I$,

$$(x, i) * (y, j) := \begin{cases} (x \cdot a_i \cdot y, j) & \text{if } i \in I_0, \\ (x \cdot b_i, k_i) & \text{if } i \in I_1. \end{cases}$$

Remark 3.3.2. Let $x \in G$. Then $L_{(x,i)} \in L(G \times I, *)$ is a constant map if $i \in I_1$. In particular, if $i \in I_2$, then $L_{(x,i)}(y, j) = (x, i)$ for all $y \in G$, $i \in I_2$. □

Lemma 3.3.3. *Every groupoid $(G \times I, *)$ of Definition* 3.3.1 *is a semigroup.*

Proof. Let $(x, i), (y, j), (z, k) \in G \times I$.

Case 1) $i, j \in I_0$.

Then $((x, i) * (y, j)) * (z, k) = (x \cdot a_i \cdot y, j) * (z, k) = (x \cdot a_i \cdot y \cdot a_j \cdot z, k)$,

$(x, i) * ((y, j) * (z, k)) = (x, i) * (y \cdot a_j \cdot z, k) = (x \cdot a_i \cdot y \cdot a_j \cdot z, k)$.

Case 2) $i \in I_1$.

Then $((x, i) * (y, j)) * (z, k) = (x \cdot b_i, k_i) * (z, k) \overset{k_i \in I_2}{=} (x \cdot b_i, k_i)$,

$(x, i) * ((y, i) * (z, k)) = (x \cdot b_i, k_i)$.

Case 3) $i \in I_0, j \in I_1$.
Then $((x,i)*(y,j))*(z,k) = (x \cdot a_i \cdot y, j)*(z,k) = (x \cdot a_i \cdot y \cdot b_j, k_j)$,
$(x,i)*((y,j)*(z,k)) = (x,i)*(y \cdot b_j, k_j) = (x \cdot a_i \cdot y \cdot b_j, k_j)$. □

Lemma 3.3.4. *In the case that $G \times I$ has at least two elements, one of the groupoids $(G \times I, *)$ of Definition* 3.3.1 *is a group if and only if $I_1 \equiv \varnothing$ and I_0 has exactly one element. $(G \times I, *)$ and $(G, \cdot)$ then are isomorphic groups.*

Proof. Let $(G \times I, *)$ be a group. Then $I_1 = \varnothing$, because otherwise, for some $i \in I_1$, $x \in G$, $L_{(x,i)} \in L(G \times I, *)$ is a constant map, which in a group with at least two elements never can happen. I_0 has exactly one element, because otherwise for two different elements $i, j \in I_0$ and some $x \in G$ $(x,i) \neq (x \cdot a_i \cdot a_j^{-1}, j)$, but $(x,i)*(x,i) = (x \cdot a_i \cdot a_j^{-1}, j)*(x,i)$, a contradiction.

Let now $I_1 = \varnothing$ and $I_0 = \{i\}$. It is then an easy calculation that the map f: $G \to G \times I$ with $x \mapsto (x \cdot a_i^{-1}, i)$ is a semigroup isomorphism, whence $(G \times I, *)$ is a group and f a group isomorphism. □

Lemma 3.3.5. *Let $(G, \cdot)$ be a group, I a set, $L_I(G)$ as defined in Sect.* 1.3, *and $(G \times I, *)$ one of the semigroups of Definition* 3.3.1. *Then $\Gamma(G \times I, L(G \times I, *)) = \Gamma(G \times I, L_I(G))$, if $I_0 \neq \varnothing$.*

For $I_0 = \varnothing$ $\Gamma(G \times I, *)$ is the discrete geometry on $G \times I$. In particular, $\Gamma(G \times I, L(G \times I, *))$ is linear if and only if $I_0 = \varnothing$ or in $(G, \cdot)$ every non-identity element is of prime order.

Proof. By Remark 3.3.2 $L(G \times I, *)$ only consists of constant maps if $I_0 = \varnothing$, whence $\Gamma(G \times I, *)$ is discrete and thus linear. Let now $I_0 \neq \varnothing$. To prove $\Gamma(G \times I, L(G \times I, *)) = \Gamma(G \times I, L_I(G))$ it is sufficient to show $L(G \times I, *) = L_I(G)$. For $(g,i) \in G \times I$ the map $L_{(g,i)} \in L(G \times I, *)$ is constant if $i \in I_1$, and $L_{(g,i)} = f_{g \cdot a_i} \in L_I(G)$ if $i \in I_0$. If $g \in G$ and $i \in I_0$, then $f_g \in L_I(G)$ satisfies $f_g = L_{(g \cdot a_i^{-1}, i)} \in L(G \times I, *)$. $I_0 \neq \varnothing$ and Proposition 1.3.2 yield that $\Gamma(G \times I, *)$ is linear if and only if in $(G, \cdot)$ every non-identity element is of prime order. □

In order to show that a finite semigroup $(S, *)$ has a nondegenerate linear congruence class geometry $\Gamma(S, L(S, *))$ only if it is isomorphic to one of the semigroups of Definition 3.3.1, some preliminaries are needed.

Lemma 3.3.6. *Let $(S, *)$ be a finite semigroup with nondegenerate linear congruence class geometry $\Gamma(S, L(S, *))$. Then the nonconstant maps of $L(S, *)$ form a subgroup of the full permutation group on S.*

Proof. As in a finite nondegenerate linear congruence class geometry the nonconstant dilatations form a group, it is sufficient to show for all $a, b \in S$ that $L_a \circ L_b$, $L_a^{-1} \in L(S, *)$ if $L_a, L_b \in L(S, *)$ are nonconstant maps. This is trivial for $L_a \circ L_b$, because associativity of $(S, *)$ yields $L_a \circ L_b = L_{a*b} \in L(S, *)$. As S is finite, there is a positive integer n such that $L_a^n = \mathrm{id}_S$. Associativity of $(S, *)$ then implies $L_{a^{n-1}} \circ L_a = L_a \circ L_{a^{n-1}} = L_{a^n} = L_a^n = \mathrm{id}_S$, hence $L_a^{-1} = L_{a^{n-1}} \in L(S, *)$. □

Lemma 3.3.7. *Under the assumptions of Lemma* 3.3.6 *the nonconstant maps of $L(S, *)$ form the full translation group of $\Gamma(S, L(S, *))$.*

Proof. Denote by C_0 the set of nonconstant maps of $L(S,*)$. By Lemma 3.3.6 there is an $e \in S$ with $L_e = \mathrm{id}_S$. For $A := \{s * e \in S \mid s \in S, L_s \in C_0\}$ a congruence relation of $(S, L(S,*))$ is defined by $\{(x,y) \in S \times S \mid x = y \text{ or } x, y \in A\}$, because for all $x, y \in A$, $s \in S$ either $L_s \notin C_0$ and thus $L_s(x) = L_s(y)$, or $L_s \in C_0$ and there are $s_1, s_2 \in S$ with $x = s_1 * e$, $y = s_2 * e$, whence $L_s(x) = s * x = s * (s_1 * e) = (s * s_1) * e \in A$ and $L_s(y) = s * y = s * (s_2 * e) = (s * s_2) * e \in A$.

As A obviously is a set of transitivity of $\Gamma(S, L(S,*))$, the Propositions 1.3.1 and 1.3.2 yield that there are exactly $|A|$ translations in $\Gamma(S, L(S,*))$. It is therefore sufficient to show that all maps in C_0 are translations, because then $|C_0| = |A|$. Let $a \in S$ such that $L_a \in C_0 \setminus \{\mathrm{id}_S\}$ and assume $L_a(b) = b$ for some $b \in S$. Then $a * b = b$, thus $a * (b * c) = (a * b) * c = b * c$ for all $c \in S$, hence $b * c = b$ for all $c \in S$, and L_b is a constant map which implies C_0 to have less elements than S, and therefoe $A \subsetneqq S$. The congruence relation defined above then shows that $\Gamma(S, L(S,*))$ is intransitive, because $(S, L(S,*))$ is not regular. Thus, by Proposition 1.3.2, all nonconstant dilatations of $\Gamma(S, L(S,*))$ are translations. □

Lemma 3.3.8. *Under the assumptions of Lemma* 3.3.6 *there is a group* $(G,\cdot)$ *and a set* I *such that* $\Gamma(S, L(S,*))$ *and* $\Gamma(G \times I, L_I(G))$ *are isomorphic congruence class geometries and, moreover,* $(G,\cdot)$ *is isomorphic to the group of translations of* $\Gamma(S, L(S,*))$. *Every nonidentity element of such a group* $(G,\cdot)$ *is of prime order.*

Proof. The claim of the lemma is implied by Proposition 1.3.2 if $\Gamma(S, L(S,*))$ is intransitive. If the geometry is transitive, Proposition 1.3.1 yields the existence of a group $(G,\cdot)$ and an automorphism Φ of $(G,\cdot)$ such that $\Gamma(S, L(S,*)) = \Gamma(G, L(G,\cdot), \Phi)$. But by Lemma 3.3.7 $L(S,*) = L(G,\cdot)$ which implies $\Gamma(S, L(S,*)) = \Gamma(G, L(G,\cdot))$. Trivially, $\Gamma(G, L(G,\cdot)) \cong \Gamma(G \times I, L_I(G))$ for every one-element set I. □

Theorem 3.3.9. *Let* $(S,*)$ *be a finite semigroup with nondegenerate linear congruence class geometry* $\Gamma(S, L(S,*))$. *Then* $(S,*)$ *is isomorphic to one of the semigroups* $(G \times I, *)$ *of Definition* 3.3.1. *On the other hand, such a semigroup has a linear congruence class geometry* $\Gamma(G \times I, *)$ *if and only if* $I_0 = \varnothing$ *or every nonidentity element of* G *is of prime order.*

Proof. The second part of this theorem was already proved in Lemma 3.3.5. In order to prove the first part let $(S,*)$ be a finite quasigroup with nondegenerate linear congruence class geometry $\Gamma(S, L(S,*))$. Lemma 3.3.8 then yields the existence of a group $(G,\cdot)$ and a set I such that $\Gamma(S, L(S,*))$ and $\Gamma(G \times I, L_I(G))$ are isomorphic congruence class geometries. Therefore henceforth in this proof instead of $(S,*)$ an isomorphic semigroup $(G \times I, *)$ will be considered which has the congruence class geometry $\Gamma(G \times I, L(G \times I, *)) = \Gamma(G \times I, L_I(G))$. Let C_0 consist of all nonconstant maps in $L(G \times I, *)$ and $M_0 := \{(x,i) \in G \times I \mid L_{(x,i)} \in C_0\}$. At first consider $C_0 = \varnothing$. In this case $G = \{e\}$. Define $I_0 := \varnothing$, $I_1 := I$, $k_i \in I$ such that $L_{(e,i)}(e,j) = (e,k_i)$ is satisfied for all $j \in I$ and $b_i := e$ for all $i \in I$, and $I_2 := \{k_i \in I \mid i \in I\}$. Then all properties given in Definition 3.3.1 trivially are satisfied except condition (ii). To prove this condition, let $i \in I$ and $j := k_i$. Then

$$((e,i) * (e,i)) * (e,i) = (e,k_i) * (e,i) = (e,j) * (e,i) = (e,k_j)$$

and

$$(e,i)*((e,i)*(e,i))=(e,k_i)=(e,j)$$

implies $j=k_j$, thus proving condition (ii) of Definition 3.3.1. Let now $C_0\neq\varnothing$. Then $C_0=L_I(G)$ by Lemma 3.3.8. For each $(x,i)\in M_0$ there is an element $a(x,i)\in G$ such that $(x,i)*(y,j)=(a(x,i)\cdot y,j)$ for all $(y,j)\in G\times I$. If $(x,i),(y,j)\in M_0$, then $(y,j)*(x,i)=(a(y,j)\cdot x,i)\in M_0$, whence $G\times\{i\}\subseteq M_0$, because $C_0=L_I(G)$ implies $G=\{a(y,j)|(y,j)\in M_0\}$. Thus $M_0=G\times I_0$ for $I_0:=\{i\in I|(e,i)\in M_0\}$. Let e be the identity of G and $(x,i),(y,j)\in G\times I_0$ such that $a(y,j)=x^{-1}$. Then, for all $(z,k)\in G\times I$,

$$\begin{aligned}((y,j)*(x,i))*(z,k)&=(a(y,j)\cdot x,i)*(z,k)\\&=(a(a(y,j)\cdot x,i)\cdot z,k)=(a(e,i)\cdot z,k),\\(y,j)*((x,i)*(z,k))&=(y,j)*(a(x,i)\cdot z,k)\\&=(a(y,j)\cdot a(x,i)\cdot z,k)=(x^{-1}\cdot a(x,i)\cdot z,k),\end{aligned}$$

hence by associativity $a(x,i)=x\cdot a(e,i)$. By defining $a_i:=a(e,i)$ for each $i\in I_0$ one obtains, for all $x,y\in G$, $i\in I_0$, $j\in I$, $(x,i)*(y,j)=(x\cdot a_i\cdot y,j)$. Let $I_1:=I\setminus I_0$. For every $(x,i)\in G\times I_1$ there are elements $b(x,i)\in G$, $k(x,i)\in I$ which satisfy $(x,i)*(y,j)=(b(x,i),k(x,i))$ for all $(y,j)\in G\times I$. If $(x,i)\in G\times I_1$, $(y,j)\in G\times I_0$, $(z,k)\in G\times I$ such that $x^{-1}=y\cdot a_j$, then

$$\begin{aligned}((y,j)*(x,i))*(z,k)&=(y\cdot a_j\cdot x,i)*(z,k)=(e,i)*(z,k)\\&=(b(e,i),k(e,i)),\\(y,j)*((x,i)*(z,k))&=(y,j)*b(x,i),k(x,i))\\&=(y\cdot a_j\cdot b(x,i),k(x,i))=(x^{-1}\cdot b(x,i),k(x,i)),\end{aligned}$$

hence by associativity $b(x,i)=x\cdot b(e,i)$ and $k(x,i)=k(e,i)$. By defining $b_i:=b(e,i)$, $k_i:=k(e,i)$ for each $i\in I_1$ one obtains, for all $x,y\in G$, $i\in I_1$, $j\in I$, $(x,i)*(y,j)=(x\cdot b_i,k_i)$, and $k_i\in I_1$, as $L_{(x,i)*(y,j)}$ obviously is constant. Let $I_2:=\{k_i\in I|i\in I_1\}$. All properties given in Definition 3.3.1 habe been proved now except the conditions (i) and (ii). Let therefore $i\in I_1$ and $j:=k_i$. Then

$$\begin{aligned}((e,i)*(e,i))*(e,i)&=(b_i,k_i)*(e,i)=(b_i,j)*(e,i)=(b_i\cdot b_j,k_j),\\(e,i)*((e,i)*(e,i))&=(e,i)*(b_i,k_i)=(b_i,k_i)=(b_i,j),\end{aligned}$$

hence by associativity $b_j=e$ and $k_j=j$, thus the conditions (i) and (ii) of Definition 3.3.1 are satisfied. □

References

1. Biliotti, M.: Strutture di André ed S-Spazi con traslationi. Geometriae Dedicata **10**, 113–128 (1981)
2. Bruck, R.H.: A Survey of Binary Systems. Berlin-Heidelberg-New York: Springer 1966
3. Cohn, P.M.: Universal Algebra. New York-Evanstone-London: Harper and Row 1965
4. Grätzer, G.: Universal Algebra. Princeton, N.J.: van Nostrand 1968

5. Hall, M., Jr.: The Theory of Groups, tenth printing. New York: Macmillan 1968
6. Herzer, A.: Endliche nichtkommutative Gruppen mit Partition π und fixpunktfreiem π-Automorphismus. Arch. Math. (Basel) **34**, 385–392 (1980)
7. Huppert, B.: Endliche Gruppen I. Berlin-Heidelberg-New York: Springer 1967
8. Ihringer, T.: On Certain Linear Congruence Class Geometries. Proc. Int. Conf. Combinatorial Geometries and their Applications (Rome 1981). Annales of Discrete Mathematics (to appear)
9. Libois, P.: Quelques espaces linéaires. Bull. Soc. Math. Belg. **16**, 13–22 (1964)
10. Pasini, A.: On the Finite Transitive Incidence Algebras. Boll. Un. Mat. Ital. (5) **17-B**, 373–389 (1980)
11. Smith, J.D.H.: Multiplication Groups of Quasigroups. Preprint. Darmstadt, 1981
12. Werner, H.: Produkte von Kongruenzklassengeometrien universeller Algebren. Math. Z. **121**, 111–140 (1971)
13. Wille, R.: Kongruenzklassengeometrien. Lecture Notes in Mathematics **113**. Berlin-Heidelberg-New York: Springer 1970

Received December 22, 1981

Math. Z. 180, 413–417 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Nicht-ausgeartete reguläre Überlagerungen

Erich Platte

Fachbereich 3, Naturwissenschaften/Mathematik der Universität Osnabrück, Abt. Vechta, Driverstraße 22, D-2848 Vechta, Bundesrepublik Deutschland

Einleitung

In den Bezeichnungen lehnen wir uns generell an [7] an; zur Theorie der Differentialmoduln verweisen wir ebenfalls auf [7].

In der vorliegenden Arbeit betrachten wir solche normalen analytischen Algebren A über einem bewerteten Körper k der Charakteristik Null, die von einem konvergenten Potenzreihenring $B=k\langle X_1,\ldots,X_s\rangle$ überlagert werden, derart, daß Potenzreihen $F_1,\ldots,F_m$ in $\mathfrak{m}_B$ existieren, so daß der Homomorphismus

$$A=k\langle F_1,\ldots,F_m\rangle \xrightarrow{\varphi} k\langle X_1,\ldots,X_s\rangle=B$$

nicht-ausgeartet ist, d.h. die Dimensionsgleichheit

$$\dim A+\dim B/\mathfrak{m}_A B=\dim B$$

erfüllt ist. (Liegt der Spezialfall vor, daß die Potenzreihen $F_1,\ldots,F_m$ ein $\mathfrak{m}_B$-primäres Ideal in B erzeugen, so ist φ ein endlicher Homomorphismus.) Da A normal und ausgezeichnet ist, ist φ rein, s. z.B. § 1 in [5], und wie im Fall, daß φ endlich ist, beweist man dann, daß A ein Macaulay-Ring ist. Der bereits genannte Spezialfall, daß φ endlich ist, schließt die Situation mit ein, daß auf B eine endliche Gruppe G von k-Algebra-Automorphismen mit $B^G=A$ operiert, und in dieser Hinsicht sind die in dieser Arbeit erzielten Ergebnisse teilweise Verallgemeinerungen aus [4, § 2], allerdings ausschließlich für den Fall der Charakteristik Null.

Im einzelnen sei zum Inhalt der vorliegenden Arbeit folgendes gesagt: In (1.1) wird gezeigt, daß der Modul der k-Derivationen auf A sowie der Modul der Zariski-Differentiale auf A Macaulay-Moduln sind. Hieraus folgt dann zusammen mit [5, § 2], daß der k-Derivationenmodul von A nur dann endliche homologische Dimensionen über A besitzt, wenn A bereits regulär ist. In (2.1) wird u.a. gezeigt, daß die Kotangentenmoduln $\mathrm{Ext}^i_A(\mathrm{D}_k(A),A)$ von A für $i=1$ bis zur Höhe $\mathrm{d}-2$ verschwinden, falls der singuläre Ort von A von der

0025-5874/82/0180/0413/$01.00

Kodimension $d \geqq 3$ ist. Insbesondere sind solche Singularitäten A starr, die in der Kodimension 2 regulär sind, s. (2.3). Hieraus folgt dann insbesondere, daß der singuläre Ort eines nicht-regulären vollständigen Durchschnittes A genau rein-2-kodimensional ist, (2.4). Eine ähnliche Aussage gilt für fastvollständige Durchschnitte A, s. (2.5).

1. Der Hauptsatz

Die Voraussetzungen und Bezeichnungen seien wie in der Einleitung.

(1.1) **Satz.** *Der Modul der k-Derivationen auf A*

(a) $(D_k(A))^*$

sowie der Modul der Zariski-Differentiale auf A

(b) $(D_k(A))^{**}$

sind Macaulay-Moduln maximaler Dimension über A.

Beweis. Wie in [5], § 1 können wir annehmen, daß A und B komplett sind. Es gibt dann eine endliche, reine Faktorisierung

$$\tilde{A} := A[\![T_1, \ldots, T_d]\!] \to B$$

mit $d := \dim B/\mathfrak{m}_A B$ und formalen analytischen Unbestimmten $T_1, \ldots, T_d$ über A. Es ist $D_A(\tilde{A})$ ein freier $\tilde{A}$-Modul des Ranges d. Aus der exakten Sequenz

$$D_k(A) \otimes_A \tilde{A} \to D_k(\tilde{A}) \to D_A(\tilde{A}) \to 0$$

folgen die Isomorphien, wobei mit $'$ das Dualisieren über $\tilde{A}$ bezeichnet ist:

(a) $(D_k(\tilde{A}))' = (D_k(A) \otimes_A \tilde{A})' \oplus (D_A(\tilde{A}))'$

(b) $(D_k(\tilde{A}))'' = (D_k(A) \otimes_A \tilde{A})'' \oplus (D_A(\tilde{A}))''$.

Weil die Erweiterung $A \to \tilde{A}$ treuflach ist, gilt:

$$(D_k(A))^* \otimes_A \tilde{A} = (D_k(A) \otimes_A \tilde{A})'$$
$$(D_k(A))^{**} \otimes_A \tilde{A} = (D_k(A) \otimes_A \tilde{A})''.$$

Wir können also A durch $\tilde{A}$ ersetzen und annehmen, daß die Erweiterung $A \to B$ endlich ist. Wie wir in Satz 1 aus [6] etwas allgemeiner als wir es hier benötigen gezeigt haben, besitzt der kanonische A-Homomorphismus

$$D\varphi\colon D_k(A) \to D_k(B)$$

eine eindeutig bestimmte A-lineare Fortsetzung

$$\widetilde{D\varphi}\colon (D_k(A))^{**} \to \mathrm{Hom}_B(\mathrm{Hom}_B(D_k(B), B), B) = D_k(B),$$

bezüglich der $(D_k(A))^{**}$ ein direkter A-Summand von $D_k(B)$ ist. Dann folgt die Behauptung (b) daraus, daß $D_k(B)$ Macaulay-Modul (auch) über A ist.

Zum Beweis der Behauptung (a) muß man etwas weitergehender argumentieren, denn Dualmoduln von Macaulay-Moduln sind i.a. keine Macaulay-Moduln. Da - wie bereits im Beweis von (b) verwendet - $(D_k(A))^{**} \subseteq D_k(B)$ ein direkter A-Summand von $D_k(B)$ ist, ist ebenfalls

$$(D_k(A))^* = ((D_k(A))^{**})^* \subseteq \mathrm{Hom}_A(D_k(B), A)$$

ein direkter A-Summand von $\mathrm{Hom}_A(D_k(B), A)$. Deshalb genügt zu zeigen, daß $\mathrm{Hom}_A(D_k(B), A)$ Macaulay-Modul über A ist. Weil $D_k(B)$ freier B-Modul ist, genügt zu wissen, daß der B-Modul

$$E := \mathrm{Hom}_A(B, A)$$

ein Macaulay-Modul ist. Es ist aber E sicherlich reflexiv über A und deshalb auch über B, weil B normal ist. Ferner hat E den Rang 1 über B. Weil B insbesondere faktoriell ist, folgt dann, daß E sogar frei über B ist. Hieraus folgt, daß $E = \mathrm{Hom}_A(B, A)$ und deshalb auch $\mathrm{Hom}_A(D_k(B), A)$ ein Macaulay-Modul über A ist.

In Verallgemeinerung von [5, § 2] folgt aus Satz (1.1) unmittelbar:

(1.2) **Korollar.** *Hat einer der Moduln*

(a) $(D_k(A))^*$

(b) $(D_k(A))^{**}$

endliche homologische Dimension über A, so ist A regulär.

2. Weitere Folgerungen

Die Voraussetzungen und Bezeichnungen seien wieder dieselben wie in der Einleitung. Mit ω_A sei der kanonische Modul von A im Sinne von [2] bezeichnet. Es ist ω_A ein endlicher reflexiver Macaulay-Modul über A.

(2.1) **Satz.** *Hat der singuläre Ort von A eine Kodimension* $d \geqq 3$, *so gilt:*

(a) $\mathrm{Ext}^i_A(D_k(A), A) = 0$ *für* $i = 1, \ldots, d-2$

(b) $\mathrm{Ext}^i_A(D_k(A), \omega_A) = 0$ *für* $i = 1, \ldots, d-2$.

Beweis. Nach Satz (1.1), Teil (a) ist

$$\mathrm{Hom}_A(D_k(A), A)$$

ein Macaulay-Modul über A, und nach Teil (b) ebenfalls $(D_k(A))^{**}$. Hieraus folgt, daß auch $\mathrm{Hom}_A((D_k(A))^{**}, \omega_A)$ ein Macaulay-Modul ist, siehe Satz (6.1) in [2]. Es ist also ebenfalls

$$\mathrm{Hom}_A(D_k(A), \omega_A)$$

ein Macaulay-Modul über A. Jetzt folgt die Behauptung von (2.1) aus dem folgenden interessanten Lemma von Jähner [3], das durch vollständige Induktion über n bewiesen wird.

(2.2) **Lemma** [3, S. 25]. *Seien M und N endliche Moduln über dem lokalen noetherschen Ring R. Das Ideal $\mathfrak{a} \subseteq R$ beschreibe die Varietät des nicht-freien Ortes von M, und es existiere in $\mathfrak{a}$ eine Primfolge in N der Länge $n \geqq 3$. Dann sind folgende Aussagen äquivalent:*

(1) *$\mathfrak{a}$ enthält eine Primfolge der Länge n in* $\mathrm{Hom}_R(M, N)$

(2) $\mathrm{Ext}^i_R(M, N) = 0$ *für* $i = 1, \ldots, n-2$.

Aus (2.1), Teil (a) folgt:

(2.3) **Korollar.** *Ist A in der Kodimension 2 regulär, so ist A starr.*

(2.4) **Korollar.** *Ist A ein in der Kodimension 2 regulärer vollständiger Durchschnitt, so ist A regulär.*

Der *Beweis* von (2.4) folgt aus (2.3), kann der Einfachheit halber aber auch direkt aus Satz (1.1) hergeleitet werden: Es hat $D_k(A)$ die homologische Dimension $\leqq 1$; ferner ist $D_k(A)$ in der Kodimension $\leqq 2$ frei. Daher ist $D_k(A)$ reflexiv. Nach (1.1), Teil (b) ist dann $D_k(A) = (D_k(A))^{**}$ sogar ein Macaulay-Modul. Hieraus folgt die Behauptung.

A heißt ein *fastvollständiger Durchschnitt*, wenn das die Singularität von A definierende Ideal I von $\operatorname{codim} I + 1$ Elementen erzeugt werden kann.

(2.5) **Korollar.** *Ist A ein in der Kodimension $\leqq 3$ regulärer fastvollständiger Durchschnitt, so ist A regulär.*

Beweis. Sei $\mathfrak{p} \subseteq A$ ein Primideal der Höhe $\geqq 4$, so daß $A_\mathfrak{p}$ höchstens in $\mathfrak{p}A_\mathfrak{p}$ singulär ist. Aus den Arbeiten von Aoyama, Matsuoka und Kunz, die in [1, § 2] zusammengestellt und verallgemeinert wurden, ergibt sich:

$$\operatorname{codh} D_k(A)_\mathfrak{p} \geqq \dim A_\mathfrak{p} - 2 \geqq 4 - 2 = 2.$$

Es ist also $D_k(A)_\mathfrak{p}$ reflexiv und wegen (1.1) darüber hinaus ein Macaulay-Modul. Deshalb ist auch der Konormalenmodul von $A_\mathfrak{p}$ ein Macaulay-Modul. Dann ist aber $A_\mathfrak{p}$ ein vollständiger Durchschnitt, wie in (2.13) aus [1] gezeigt wird. Weil $D_k(A)_\mathfrak{p}$ die homologische Dimension $\leqq 1$ hat, folgt, daß $A_\mathfrak{p}$ regulär ist.

Literatur

1. Herzog, J.: Ein Cohen-Macaulay-Kriterium mit Anwendungen auf den Konormalenmodul und den Differentialmodul. Math. Z. **163**, 149–162 (1978)
2. Herzog, J., Kunz, E. (Hrsg.): Der kanonische Modul eines Cohen-Macaulay-Rings. Lecture Notes in Mathematics **238**. Berlin-Heidelberg-New York: Springer 1971
3. Jähner, U.: Beispiele starrer analytischer Algebren. Dissertation Universität Clausthal 1974
4. Platte, E.: Differentielle Eigenschaften der Invarianten regulärer Algebren. J. Algebra **62**, 1–12 (1980)

5. Platte, E.: Pure descent for the module of Zariski-differentials. Proc. Amer. Math. Soc. **82**, 7–12 (1981)
6. Platte, E.: Jacobi-Algebra und Blätterzahl von analytisch-verzweigten Überlagerungen isolierter Hyperflächensingularitäten. Arch. Math. (Basel) (erscheint demnächst)
7. Scheja, G., Storch, U.: Differentielle Eigenschaften der Lokalisierungen analytischer Algebren. Math. Ann. **197**, 137–170 (1972)

Eingegangen am 7. Januar 1982

Math. Z. 180, 419–422 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On the Isomorphism Problem for Integral Group-Rings of Circle-Groups

Frank Röhl

Mathematisches Seminar der Universität, Bundesstraße 55, D-2000 Hamburg 13, Federal Republic of Germany

1. Introduction

In this paper we generalize Sandling's result that finite circle-groups G are characterized by their integral group-ring $\mathbb{Z}G$ (see [4]).

It is well-known that central units of finite order in $\mathbb{Z}G$ are trivial. Now, if the inclusion $G \hookrightarrow U_1(\mathbb{Z}G)$ splits, where $U_1(\mathbb{Z}G)$ denotes the group of units of augmentation one, it would be easy to prove G to be characterized by $\mathbb{Z}G$ by induction, if only there would be a *suitable* notion of torsion-length for G. But this will be given by the correspondence between torsion-elements of a nilpotent ring and torsion-elements of its circle-group, and we obtain: Circle-groups of nilpotent rings are characterized by their integral group-ring. Since any abelian group is the circle-group of its zero-ring (all products zero), this seems to be a natural generalization of the same, well-known statement for abelian groups.

The author wishes to thank Prof. Dr. K.W. Roggenkamp for his useful comments on an earlier version of this paper.

2. Exposition

Let R be a commutative ring with a unit and A an R-algebra. Then $(A, \circ)$, i.e. the set of elements of A under the circle-composition $x \circ y := x + y - xy$, forms a monoid, which is still a group, if and only if A is a Jacobson-radical-algebra.

If A contains a unit and is augmented, i.e. there exists an R-algebra-homomorphism $\varepsilon: A \to R$, then the group $U_1(A)$ of units $a \in A$ with $\varepsilon(a) = 1$ can be identified with the group of all invertible elements of $\ker \varepsilon$ under "$\circ$" via $a \mapsto 1 - a$. Universality of the group-ring RG, where G is a group, with respect to unit-groups thus translates into universality of the augmentation-ideal $\Delta_R G$ of RG with respect to circle-groups. This gives

0025-5874/82/0180/0419/$01.00

2.1. Proposition. *Let G be the circle-group of an R-algebra A. Then there exists a normal subgroup N of $U_1(RG)$ such that $U_1(RG)$ is the semi-direct product of N with G.*

In particular, $U_1(RG)$ splits, if G is any R-module, since G may be viewed as the circle-group of its zero-algebra.

Since central units in $\mathbb{Z}G$ of finite order are trivial (see [5], Corollary II.1.7 on p. 46), 2.1 implies that finite circle-groups, which are necessarily nilpotent, are characterized by their integral group-ring (see [4], Theorem 3.1 on p. 200). Clearly, this argumentation will work in any case, where G is finite nilpotent and there exists a splitting homomorphism $U_1(\mathbb{Z}G) \to G$. But in the infinite case a little more is needed.

For the following, let A denote a nilpotent ring with upper-annihilator-series $0 = A_0 \subset \operatorname{Ann} A = A_1 \subset \ldots \subset A = A_n$ and T its torsion-ideal. Then one has

2.2. Lemma. *The $N_i(A) := T \cap A_i$, $i = 0, \ldots, n$, form an ascending chain of ideals with the following properties*

a) $N_n(A) = T$; *especially, $A/N_n(A)$ is torsion-free*

b) $N_{i+j}(A)/N_i(A) = N_j(A/N_i(A))$.

2.2a) is obvious from $A_n = A$; and an easy induction on j, keeping i fixed and using only the definition of the A_i (and $N_i(A)$), gives 2.2b).

For brevity, let us define the torsion-length $t(A)$ of A to be the number of different members ($\neq 0$) of $N_1(A) \subset \ldots \subset N_n(A)$. Since every ideal in a nilpotent ring meets the annihilator non-trivial, $t(A) = 0$ says that A is torsion-free (or 0). And from 2.2b) follows that if $t(A) > 0$, then $t(A/N_1(A)) < t(A)$.

We have to cite two more technical observations, which can be found in [3] (Lemma 1 on p. 301, Corollary 1.5 on p. 302):

2.3. Lemma. *If A is a nilpotent ring with $(A, \circ)$ as circle-group, then*

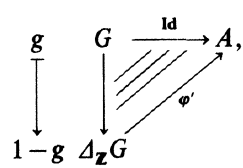
a) $(A/I, \circ) = (A, \circ)/(I, \circ)$ *for every ideal I of A*

b) *If p is any prime, then x is p-torsion in A, if and only if x is p-torsion in $(A, \circ)$.*

Now we can prove

2.4. Theorem. *Every circle-group of a nilpotent ring is characterized by its integral group-ring.*

Proof. Let A denote a nilpotent ring and G its circle-group. Since every isomorphism of group-rings yields an augmented isomorphism, we will show by induction on the torsion-length $t(A)$ of A that if H is a subgroup of $U_1(\mathbb{Z}G)$ with $\mathbb{Z}H = \mathbb{Z}G$, then the splitting homomorphism $\varphi\colon U_1(\mathbb{Z}G) \to G$ obtained by universality of $\Delta_{\mathbb{Z}}G$



induces an isomorphism $H \to G$.

Remember that circle-groups of nilpotent rings are nilpotent. Hence, if $t(A)=0$, then G is a torsion-free nilpotent group by 2.3b) (or 1, in which case the assertion is trivial), and the units of $\mathbb{Z}G$ are trivial by Bovdi's theorem (see for example [2], Theorem 1.11 on p. 592). In view of $\mathbb{Z}G=\mathbb{Z}H$ we thus obtain: $G=U_1(\mathbb{Z}G)=U_1(\mathbb{Z}H)=H$.

Let $t(A)>0$. Then $N:=N_1(A)$ is a non-trivial subgroup of the center of G, which elements are of finite order by 2.3b). But central units of finite order are trivial and hence are contained in H. Since N is normal in $U_1(\mathbb{Z}G)$, it is already a normal subgroup of H.

Let I be the ideal generated by all $1-n$, n in N. Now, N normal in H and $\mathbb{Z}G=\mathbb{Z}H$ implies $(1+I)\cap H=N$. In view of the natural isomorphism $\mathbb{Z}G/I\xrightarrow{\sim}\mathbb{Z}(G/N)$, we can identify $\mathbb{Z}(G/N)$ and $\mathbb{Z}(H/N)$. But N is an ideal of A with $\varphi' I=N$. By 2.3a) and b) we thus obtain

$$\begin{array}{ccc} G/N & \xrightarrow{\mathrm{Id}} & A/N \\ \downarrow & \nearrow_{\varphi''} & \\ \Delta_{\mathbb{Z}}(G/N), & & \end{array}$$

where φ'' is induced by φ'. (If there would only be a splitting homomorphism $U_1(\mathbb{Z}G)\to G$, one would not know, whether $(1+I)\cap U_1(\mathbb{Z}G)$ would be mapped into N; hence, the splitting of $U_1(\mathbb{Z}G)$ may not induce a splitting of a suitable subgroup of $U_1(\mathbb{Z}(G/N))$!) The inductive hypothesis yields an isomorphism $\bar{\varphi}$: $H/N\to G/N$, which is induced via φ'' by φ.

This implies

$$\begin{array}{ccccccccc} 1 & \longrightarrow & N & \longrightarrow & H & \longrightarrow & H/N & \longrightarrow & 1 \\ & & \downarrow{\scriptstyle \varphi|_N} & & \downarrow{\scriptstyle \varphi|_H} & & \downarrow{\scriptstyle \bar{\varphi}} & & \\ 1 & \longrightarrow & N & \longrightarrow & G & \longrightarrow & G/N & \longrightarrow & 1, \end{array}$$

where the (natural) rows are exact and $\bar{\varphi}$ as well as $\varphi|_N$ are isomorphisms. Qed.

There are several *remarks* in order:

1. The reader may wonder, why we cited Bovdi's result, which holds for all fields of characteristic zero. But the foregoing proof (and 2.2 and 2.3) would work over all integral-domains R of characteristic zero, if only one knows that central units of finite order in RG are trivial, which is a standing conjecture (see [5]).

2. Although it is easy to force $\bar{\varphi}$ to be an automorphism of H/N, it is not at all obvious from the above proof that the analogue holds for $\varphi|_H$. Under a corresponding inductive hypothesis it only would follow that $h\varphi(h^{-1})\in 1+(I\cap\ker\varphi')$.

3. It should be mentioned that the following groups G of nilpotence-class two are circle-groups of a nilpotent ring (see [1], Sect. 4, Corollary on p. 263):

a) The factor-commutator-group $\mathrm{Ab}(G)$ of G is a direct sum of cyclic groups

b) Ab(G) is divisible

c) Ab(G) is torsion

d) Ab(G) is torsion-free and completely decomposable (i.e. it is a direct sum of rank-one-groups).

References

1. Hales, A.W., Passi, I.B.S.: The second augmentation quotient of an integral group ring. Arch. Math. (Basel) **31**, 259–265 (1978)
2. Passman, D.S.: The algebraic structure of group rings. New York: Wiley 1977
3. Röhl, F., Ullrich, J.: Über Komplettierungen und Lokalisierungen von Zirkelgruppen nilpotenter Ringe. Arch. Math. (Basel) **37**, 300–305 (1981)
4. Sandling, R.: Group rings of circle and unit groups. Math. Z. **140**, 195–202 (1974)
5. Sehgal, S.K.: Topics in group rings. New York: Dekker 1978

Received February 5, 1982

Math. Z. 180, 423-427 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Absolut stetiges Spektrum bei Sturm-Liouville-Operatoren und Dirac-Systemen

Joachim Weidmann

Fachbereich Mathematik der Universität, Postfach 111932, D-6000 Frankfurt am Main, Bundesrepublik Deutschland

1. Einleitung

In [2], Abschnitt 5 und [3], Abschnitt 8 wurde vom Autor unter recht allgemeinen Voraussetzungen die absolute Stetigkeit des Spektrums von Sturm-Liouville-Operatoren und Dirac-Systemen in $L_2(a, \infty)$ bzw. $L_2(a, \infty)^2$ bewiesen. Dabei wurden Differentialausdrücke der folgenden Gestalt zugrunde gelegt:

(1) *Sturm-Liouville-Ausdruck*

$$(\ell u)(x) = -u''(x) + q(x)u(x) \quad \text{für } x \in (a, \infty),$$

wobei q reellwertig und in (a, ∞) lokal integrierbar ist.

(2) *Dirac-System*

$$(\ell u)(x) = Ju'(x) + P(x)u(x) \quad \text{für } x \in (a, \infty)$$

mit $u: (a, \infty) \to \mathbb{C}^2$ und

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad P(x) = \begin{pmatrix} p_1(x) & p_{12}(x) \\ p_{12}(x) & p_2(x) \end{pmatrix},$$

wobei p_1, p_2, p_{12} reellwertig und in (a, ∞) lokal integrierbar sind.

Neben den weiter unten formulierten Bedingungen an q bzw. P wurde in [2] bzw. [3] noch (implizit) eine Bedingung an das Verhalten von q bzw. P beim Randpunkt a benötigt: Um die absolute Stetigkeit des Spektrums jeder selbstadjungierten Realisierung von ℓ in einem Intervall $(\underline{\lambda}, \bar{\lambda})$ zu beweisen, wurde dort gefordert, daß jede selbstadjungierte Realisierung von ℓ in $L_2(a, c)$ bzw. $L_2(a, c)^2$ mit $a < c < \infty$ in $(\underline{\lambda}, \bar{\lambda})$ diskretes Spektrum hat. Diese Voraussetzung ist z.B. immer dann erfüllt, wenn für ℓ bei a der Grenzkreisfall (GKF) vorliegt.

Im Oktober 1981 teilte mir Herr Heinz (Göttingen) mit, daß er für Dirac-Systeme mit Hilfe der Resolventenmethode das gleiche Resultat ohne diese zusätzliche Bedingung bei a beweisen kann, vgl. [1]. Daraufhin habe ich meine

0025-5874/82/0180/0423/$01.00

früheren Beweise überprüft und festgestellt, daß eine relativ einfache Modifikation meiner Beweise ebenfalls dieses allgemeinere Resultat liefert, auch für Sturm-Liouville-Operatoren. Für Dirac-Systeme wird der Beweis in Abschnitt 2 durchgeführt. In Abschnitt 3 wird das entsprechende Resultat für Sturm-Liouville-Operatoren angegeben; der Beweis ergibt sich durch eine völlig analoge Modifikation der Beweise aus [2].

Herrn Heinz danke ich für die frühzeitige Mitteilung seines Resultats und die daran anschließende Korrespondenz.

2. Dirac-Systeme

Satz 2.1. *Sei ℓ wie in* (2) *erklärt. Für ein $c\in(a,\infty)$ gelte*

$P(x)=P_1(x)+P_2(x)$,

$|P_1(.)|\in \mathsf{L}_1(c,\infty)$,

$P_2(.)$ *ist in* $[c,\infty)$ *von beschränkter Variation mit*

$P_2(x)\to \begin{pmatrix}\mu_- & 0\\ 0 & \mu_+\end{pmatrix}$ *für* $x\to\infty$, $\mu_-<\mu_+$.

Dann hat jede selbstadjungierte Realisierung von ℓ in $\mathsf{L}_2(a,\infty)^2$ rein absolut stetiges Spektrum in $\complement[\mu_-,\mu_+]$.

Zum Beweis benötigen wir die folgenden Lemmata. Das erste ist mit Korollar 8.2 aus [3] identisch.

Lemma 2.2. *Sei ℓ wie in Satz* 2.1. *Für jedes Kompaktum $\mathsf{G}\subset(a,\infty)$ und jedes Kompaktum $\mathsf{K}\subset\complement[\mu_-,\mu_+]$ gibt es ein $C>0$ und ein $T_0>0$ so, daß für jede Lösung u von $(\ell-\lambda)u=0$ mit $\lambda\in\mathsf{K}$ gilt*

$$|u(x)|\left(\int_c^T |u(t)|^2\,dt\right)^{-1/2}\leqq C(T-c)^{-1/2} \quad \text{für } x\in\mathsf{G},\ T\geqq T_0.$$

Lemma 2.3. *Sei ℓ wie in Satz* 2.1. *Für b, d mit $a<b<d<\infty$ sei der Operator $A_{b,d}$ in $\mathsf{L}_2(b,d)^2$ erklärt durch*

$$\mathsf{D}(A_{b,d})=\{f\in\mathsf{L}_2(b,d)^2: f \text{ absolut stetig in } [b,d],\ \ell f\in\mathsf{L}_2(b,d)^2, f_1(b)=f_1(d)=0\},$$
$$A_{b,d}f(x)=(\ell f)(x) \quad \text{für } f\in\mathsf{D}(A_{b,d}).$$

Sei $[\underline{\lambda},\bar{\lambda}]\subset\complement[\mu_-,\mu_+]$ fest gewählt. Dann läßt sich für $[\lambda_1,\lambda_2]\subset[\underline{\lambda},\bar{\lambda}]$ die Zahl der Eigenwerte von $A_{b,d}$, die in $(\lambda_1,\lambda_2]$ liegen, abschätzen durch

$$N_{b,d}(\lambda_1,\lambda_2)\leqq C|\lambda_2-\lambda_1|\,d+C_b+o(d) \quad \text{für } d\geqq\max(b,1),$$

wobei C von b und d, C_b von d und der Term $o(d)$ von b unabhängig sind (natürlich hängen diese Abschätzungen von $[\underline{\lambda},\bar{\lambda}]$ ab).

Der *Beweis* dieses Lemmas folgt unmittelbar aus [3], Satz 3.1 und Lemma 6.11. Natürlich kann für $b \to a+$ die Konstante C_b gegen Unendlich streben; bei Beachtung der Reihenfolge der Grenzübergänge stört dies im folgenden Beweis nicht.

Beweis von Satz 2.1. Liegt bei a der Grenzkreisfall (GKF) vor, so ist die zusätzliche Bedingung von [3], Satz 8.3 erfüllt (vgl. Abschnitt 1): damit ist in diesem Fall nichts mehr zu beweisen. Ohne Einschränkung können wir also annehmen, daß bei a der Grenzpunktfall (GPF) vorliegt. Nach [3], Satz 5.1 liegt bei ∞ ebenfalls der GPF vor. Deshalb besitzt ℓ in $\mathsf{L}_2(a, \infty)^2$ nur eine selbstadjungierte Realisierung A und

$$\mathsf{D}_0 := \{f \in \mathsf{L}_2(a, \infty)^2 : f \text{ absolut stetig, } f \text{ hat kompakten Träger in } (a, \infty), \ell f \in \mathsf{L}_2(a, \infty)^2\},$$

ist ein determinierender Bereich (core) von A.

Für jedes $b \in (a, \infty)$ sei A'_b in $\mathsf{L}_2(a, \infty)^2$ definiert durch

$$\mathsf{D}(A'_b) = \{f \in \mathsf{L}_2(a, \infty)^2 : f \text{ absolut stetig in } [b, \infty), \ell f \in \mathsf{L}_2(b, \infty)^2, f_1(b+) = 0\},$$

$$A'_b f(x) = \begin{cases} (\ell f)(x) & \text{für } x \in (b, \infty), \\ 0 & \text{sonst.} \end{cases}$$

A'_b ist selbstadjungiert und es gilt $A'_b \to A$ für $b \to a+$ im Sinne der starken Resolventenkonvergenz (vgl. [4], Satz 9.16(i): für jedes $f \in \mathsf{D}_0$ ist $f \in \mathsf{D}(A'_b)$ für b hinreichend nahe bei a und $A'_b f = Af$). Da A in $\complement[\mu_-, \mu_+]$ keine Eigenwerte hat (vgl. [3], Satz 7.4) und da die Projektion auf $\mathsf{L}_2(a, b)^2$ für $b \to a+$ gegen Null konvergiert, folgt

$$E_b(\lambda) \xrightarrow{s} E(\lambda) \quad \text{für } \lambda \in \complement[\mu_-, \mu_+] \text{ und } b \to a+;$$

hier ist E_b die Spektralschar der Einschränkung A_b von A'_b auf $\mathsf{L}_2(b, \infty)^2$ und E die Spektralschar von A.

Für $d \in (b, \infty)$ sei der Operator $A'_{b,d}$ in $\mathsf{L}_2(a, \infty)^2$ erklärt durch

$$\mathsf{D}(A'_{b,d}) = \{f \in \mathsf{L}_2(a, \infty)^2 : f \text{ absolut stetig in } [b, d], \ell f \in \mathsf{L}_2(b, d)^2, f_1(b+) = f_1(d-) = 0\},$$

$$A'_{b,d} f(x) = \begin{cases} (\ell f)(x) & \text{für } x \in (b, d), \\ 0 & \text{sonst} \end{cases}$$

Dann gilt $A'_{b,d} \to A'_b$ für $b \to \infty$ im Sinne der starken Resolventenkonvergenz und

$$E_{b,d}(\lambda) \xrightarrow{s} E_b(\lambda) \quad \text{für } \lambda \in \complement[\mu_-, \mu_+] \text{ und } d \to \infty,$$

wobei $E_{b,d}$ die Spektralschar der Einschränkung $A_{b,d}$ von $A'_{b,d}$ auf $\mathsf{L}_2(b, d)^2$ ist (vgl. $A_{b,d}$ in Lemma 2.3).

Sei nun $[\underline{\lambda}, \bar{\lambda}] \subset \complement [\mu_-, \mu_+]$. Wir zeigen die absolute Stetigkeit des Spektrums von A in $[\underline{\lambda}, \bar{\lambda}]$. Dazu sei $f \in \mathsf{L}_2(a, \infty)^2$ mit kompaktem Träger in (a, ∞) (d.h. es gibt eine kompakte Teilmenge G von (a, ∞), so, daß f außerhalb von G fast überall verschwindet). Nach Lemma 2.2 gilt für die normierten Eigenfunktionen $\varphi_{b,d,m}$ von $A_{b,d}$, die zu den Eigenwerten $\lambda_{b,d,m} \in [\underline{\lambda}, \bar{\lambda}]$ gehören,

$$|\varphi_{b,d,m}(x)| \leqq C_1 d^{-1/2} \quad \text{für } x \in \operatorname{supp} f,\ d \geqq d_0.$$

Zusammen mit Lemma 2.3 folgt für $(\lambda_1, \lambda_2] \subset [\underline{\lambda}, \bar{\lambda}]$

$$\begin{aligned}\|(E_{b,d}(\lambda_2) - E_{b,d}(\lambda_1)) f\|^2 &= \| \sum_{\lambda_1 < \lambda_{b,d,m} \leqq \lambda_2} \langle \varphi_{b,d,m}, f \rangle \varphi_{b,d,m} \|^2 \\ &= \sum_{\lambda_1 < \lambda_{b,d,m} \leqq \lambda_2} |\langle \varphi_{b,d,m}, f \rangle|^2 \\ &\leqq \{ C|\lambda_2 - \lambda_1| d + C_b + o(d) \}\, C_1^2 d^{-1} \|f\|^2 \operatorname{meas}(\operatorname{supp} f) \\ &\leqq C(f)(|\lambda_2 - \lambda_1| + C_b d^{-1})\end{aligned}$$

für hinreichend große d, wobei die Konstanten nicht von d abhängen. Für $d \to \infty$ folgt hieraus

$$\|(E_b(\lambda_2) - E_b(\lambda_1)) f\|^2 \leqq C(f)|\lambda_2 - \lambda_1|,$$

und hieraus für $b \to a+$

$$\|E(\lambda_2) f\|^2 - \|E(\lambda_1) f\|^2 \leqq C(f)|\lambda_2 - \lambda_1|,$$

d.h. $\|E(\lambda) f\|^2$ ist absolut stetig in $[\underline{\lambda}, \bar{\lambda}]$ für alle f mit kompaktem Träger.

Da der absolut stetige Teilraum abgeschlossen ist, folgt daraus die absolute Stetigkeit von $\|E(\lambda) f\|^2$ in $[\underline{\lambda}, \bar{\lambda}]$ für alle $f \in \mathsf{L}_2(a, \infty)^2$; das ist die Behauptung des Satzes.

3. Sturm-Liouville-Operatoren

Satz 3.1. *Sei ℓ wie in* (1) *erklärt. Für ein $c \in (a, \infty)$ gelte*

$$q(x) = q_1(x) + q_2(x),$$

$q_1(.) \in \mathsf{L}_1(c, \infty)$,

$q_2(.)$ *von beschränkter Variation mit* $q_2(x) \to 0$ *für* $x \to \infty$.

Dann hat jede selbstadjungierte Realisierung A von ℓ in $\mathsf{L}_2(a, \infty)$ rein absolut stetiges Spektrum in $(0, \infty)$.

Der *Beweis* lehnt sich an den Beweis von [2], Satz 5.1 und an den Beweis des obigen Satzes 2.1 an. Wir können es uns sparen, die Einzelheiten durchzuführen. Für die, analog zu Abschnitt 2 definierten, Operatoren A_b und $A_{b,d}$ (wobei Bedingungen der Form $f_1(t) = 0$ durch $f(t) = 0$ zu ersetzen sind) ergibt sich:

Die Zahl der Eigenwerte von $A_{b,d}$ *in* $(\lambda_1, \lambda_2] \subset [\underline{\lambda}, \bar{\lambda}] \subset (0, \infty)$ *läßt sich abschätzen durch*

$$N_{b,d}(\lambda_1, \lambda_2) \leqq C|\lambda_2 - \lambda_1| + C_b + o(d)$$

([2], Lemma 5.3).

Für die normierten Eigenfunktionen zu Eigenwerten aus $[\underline{\lambda}, \bar{\lambda}]$ *und jedes Kompaktum* $\mathsf{G} \subset (a, \infty)$ *gilt*

$$|\varphi_{b,d,m}(x)| \leqq C_1 d^{-1/2} \quad \text{für } x \in \mathsf{G},\ d \geqq d_0$$

([2], S. 296–297).

Mit diesen beiden Abschätzungen kann wie im Beweis von Satz 2.1 geschlossen werden.

4. Schlußbemerkung

Die Tatsache, daß die absolute Stetigkeit des Spektrums in den oben betrachteten Fällen nicht durch das Verhalten am linken Rand gestört werden kann, legt die folgende allgemeine *Vermutung* nahe:

Sei ℓ *ein formal selbstadjungierter Differentialausdruck auf* (a, b) *und* A *eine selbstadjungierte Realisierung. Für ein* $c \in (a, b)$ *seien* A_a *und* A_b *selbstadjungierte Realisierungen von* ℓ *in* (a, c) *bzw.* (c, d). *Hat* A_a *oder* A_b *in* $[\underline{\lambda}, \bar{\lambda}]$ *absolut stetiges Spektrum, so gilt dies auch für* A.

Literatur

1. Heinz, E.: Über das absolut stetige Spektrum singulärer Differentialgleichungssysteme. Nachr. Akad. Wiss. Göttingen II. Math.-Phys. Kl. (erscheint demnächst)
2. Weidmann, J.: Zur Spektraltheorie von Sturm-Liouville-Operatoren. Math. Z. **98**, 268–302 (1967).
3. Weidmann, J.: Oszillationsmethoden für Systeme gewöhnlicher Differentialgleichungen. Math. Z. **119**, 349–373 (1971)
4. Weidmann, J.: Lineare Operatoren in Hilberträumen. Stuttgart: Teubner 1976

Eingegangen am 11. Februar 1982

Math. Z. 180, 429–444 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Ball-Homogeneous and Disk-Homogeneous Riemannian Manifolds

Oldřich Kowalski[1] and Lieven Vanhecke[2]

[1] Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 18600 Praha, Czechoslovakia

[2] Department of Mathematics, Katholieke Universiteit Leuven, Celestijnenlaan 200B, B-3030 Leuven, Belgium

1. Introduction

Let (M, g) be an analytic connected Riemannian manifold, $\dim M = n$, and let $i(m)$ denote the injectivity radius at a point $m \in M$. We shall denote by $B_m(r)$ the *geodesic ball* with center m and radius $r < i(m)$ in (M, g). For any unit vector $x \in T_m M$, we shall denote by $D_m^x(r)$ the *geodesic disk* with center m and radius $r < i(m)$ which is perpendicular to x. More specifically, we put

$$B_m(r) = \{m' \in M \mid d(m', m) \leqq r\},$$
$$D_m^x(r) = \{m' \in M \mid d(m', m) \leqq r\} \cap \exp_m(\{x\}^\perp)$$

where $\exp_m: T_m M \to M$ is the exponential map at m. Further, we shall denote by $V_m(r)$ the n-dimensional volume of $B_m(r)$ and by $V_m^x(r)$ the $(n-1)$-dimensional volume of $D_m^x(r)$.

The Riemannian manifold (M, g) will be said to be *ball-homogeneous* if $V_m(r)$ does not depend on the point $m \in M$ (whenever it is well-defined). (M, g) will be said to be *disk-homogeneous* if, at each point $m \in M$, $V_m^x(r)$ does not depend on the unit vector $x \in T_m M$. Finally (M, g) will be said to be *strongly disk-homogeneous* if it is disk-homogeneous at each point $m \in M$ and if, in addition, $V_m^x(r)$ does not depend on the point $m \in M$.

One can see easily that any homogeneous Riemannian space is ball-homogeneous. The Euclidean space and any rank one symmetric space are both ball-homogeneous and strongly disk-homogeneous. The converse for the ball-homogeneous manifolds (for $n > 2$) seems to be a difficult problem. As we shall see below, the converse for the manifolds which are simultaneously ball-homogeneous and disk-homogeneous or which are strongly disk-homogeneous is an interesting analogy of the famous problem on harmonic spaces (see [1, 7]).

We state the following *conjectures*:

0025-5874/82/0180/0429/$03.20

(BD) *Any analytic Riemannian manifold which is simultaneously ball-homogeneous and disk-homogeneous is either locally Euclidean, or it is locally isometric to a rank one symmetric space.*

(SD) *Any analytic Riemannian manifold of dimension* >2 *which is strongly disk-homogeneous is either locally Euclidean, or it is locally isometric to a rank one symmetric space.*

The proof of the conjecture (BD) is trivial for dimension 2 (in this case, "disk-homogeneous" is a trivial property and "ball-homogeneous" simply means "of constant Gaussian curvature"). As we shall see later, the proof of both conjectures is almost trivial for dimension 3.

The main purpose of this article is to provide a proof for dimension 4. Here we shall follow the main line of the proof given in [1] for the analogous theorem by Lichnerowicz and Walker, concerning infinitesimally harmonic manifolds of dimension 4. What is essentially new in our proof is the "source of information". Whereas the theorem by Lichnerowicz and Walker follows (by a non-trivial procedure) from the Ledger conditions of the second, fourth and sixth order, we are using the power series expansions up to the sixth order for the volume of geodesic balls and geodesic disks.

We shall also modify a bit the proof given in [1] to make it more elementary, and at the same time, we are correcting some minor errors involved in the latter.

2. General Formulas

First of all, we shall recall the power series expansion for the volume of geodesic balls (see [3]).

Theorem A. *For any point* $m\in M$ *and any small radius* $r>0$ we have

$$V_m(r)=V_0^n(r)\,(1+Ar^2+Br^4+Cr^6+O(r^8))_m$$

where $V_0^n(r)$ *denotes the volume of a Euclidean ball of dimension* n *and radius* r, *and the coefficients* A, B, C *are given by the following formulas:*

$$A(m)=-\frac{\tau(m)}{6(n+2)},$$

$$B(m)=\frac{(-3\|R\|^2+8\|\rho\|^2+5\tau^2-18\Delta\tau)_m}{360(n+2)(n+4)},$$

$$\begin{aligned}C(m)=\frac{1}{720(n+2)(n+4)(n+6)}\Big\{&-\tfrac{5}{9}\tau^3-\tfrac{8}{3}\tau\|\rho\|^2+\tau\|R\|^2+\tfrac{64}{63}\check{\rho}\\&-\tfrac{64}{21}\langle\rho\otimes\rho,\overline{R}\rangle+\tfrac{32}{7}\langle\rho,\dot{R}\rangle-\tfrac{110}{63}\check{R}-\tfrac{200}{63}\overline{\check{R}}+\tfrac{45}{7}\|\nabla\tau\|^2\\&+\tfrac{45}{14}\|\nabla\rho\|^2+\tfrac{45}{7}\alpha(\rho)-\tfrac{45}{14}\|\nabla R\|^2+6\tau\Delta\tau+\tfrac{48}{7}\langle\Delta\rho,\rho\rangle\\&+\tfrac{54}{7}\langle\nabla^2\tau,\rho\rangle-\tfrac{30}{7}\langle\Delta R,R\rangle-\tfrac{45}{7}\Delta^2\tau\Big\}_m.\end{aligned}$$

Here R, ρ and τ denote the *Riemannian curvature tensor* (of type (0, 4)), *the Ricci tensor*, and *the scalar curvature*, respectively. Δ denotes the *Laplacian* in (M, g), and the other symbols are given, with respect to an orthonormal frame $\{e_1, \ldots, e_n\}$ of $T_m M$ by the following formulas (in which we omit, for the sake of brevity, the suffix "m" on the left-hand sides and we make an obvious summation convention on the right-hand sides):

$$\|R\|^2 = \sum R_{ijkl}^2, \qquad \|\nabla\tau\|^2 = \sum (\nabla_i \tau)^2,$$

$$\|\rho\|^2 = \sum \rho_{ij}^2, \qquad \|\nabla\rho\|^2 = \sum (\nabla_i \rho_{jk})^2,$$

$$\check{\rho} = \sum \rho_{ij}\rho_{jk}\rho_{ki}, \qquad \alpha(\rho) = \sum (\nabla_i \rho_{jk})(\nabla_k \rho_{ij}),$$

$$\langle \rho, \dot{R} \rangle = \sum \rho_{ij} R_{ipqr} R_{jpqr}, \qquad \|\nabla R\|^2 = \sum (\nabla_i R_{jklq})^2,$$

$$\langle \rho \otimes \rho, \overline{R} \rangle = \sum \rho_{ij}\rho_{kl} R_{ikjl}, \qquad \langle \Delta\rho, \rho \rangle = \sum \rho_{ij} \nabla_{kk}^2 \rho_{ij},$$

$$\check{R} = \sum R_{ijkl} R_{klpq} R_{pqij}, \qquad \langle \nabla^2 \tau, \rho \rangle = \sum (\nabla_{ij}^2 \tau)\rho_{ij},$$

$$\check{\overline{R}} = \sum R_{ikjl} R_{kplq} R_{piqj}, \qquad \langle \Delta R, R \rangle = \sum R_{ijkl} \nabla_{pp}^2 R_{ijkl}.$$

Now, we shall provide the analogous power series expansion for geodesic disks in the form of two theorems.

Theorem 2.1. *For any point $m \in M$, any unit vector $x \in T_m M$ and any small radius $r > 0$ we have*

$$V_m^x(r) = V_0^{n-1}(r)\{1 + A(x) r^2 + B(x) r^4 + O(r^6)\}$$

where the coefficients $A(x)$, $B(x)$ are given, with respect to an adapted orthonormal basis $\{e_1 = x, e_2, \ldots, e_n\}$ by the following formulas:

$$A(x) = -\frac{1}{6(n+1)}(\tau - 2\rho_{11})_m,$$

$$\begin{aligned} B(x) = \frac{1}{360(n+1)(n+3)}\{&-3\|R\|^2 + 8\|\rho\|^2 + 5\tau^2 - 18\Delta\tau \\ &+ 24\sum R_{1ijk}^2 - 8\sum \rho_{1i}^2 + 12\rho_{11}^2 - 16\sum \rho_{ij} R_{1i1j} \\ &+ 20\sum R_{1i1j}^2 - 20\tau\rho_{11} + 36\sum \nabla_{ii}^2 \rho_{11} + 18\nabla_{11}^2 \tau\}_m. \end{aligned}$$

In the second formula, all the summations are ranging over $i, j, k = 2, 3, \ldots, n$.

Theorem 2.2. *Let, in addition, (M, g) be an Einstein manifold. Then the formula*

$$V_m^x(r) = V_0^{n-1}(r)\{1 + A(x) r^2 + B(x) r^4 + C(x) r^6 + O(r^8)\}$$

holds, where

$$\begin{aligned} C(x) = &\frac{(n+2)(n+4)(n+6)}{(n+1)(n+3)(n+5)} C(m) + \lambda\|R\|_m^2 + \mu\|R^x\|^2 + \nu\|R^{xx}\|^2 + \varepsilon \\ &+ \sum_{\alpha=1}^{7} a_\alpha C_\alpha(x) + \frac{1}{14(720)(n+1)(n+3)(n+5)}\{90D_1(x) + 10D_2(x) + 320D_3(x) \\ &+ 225D_4(x) + 35D_5(x) + 60D_6(x) + \tfrac{556}{3} D_7(x) + 322D_8(x) + 240D_9(x) \\ &+ 336D_{10}(x)\}. \end{aligned}$$

Here $a_1, \ldots, a_7, \lambda, \mu, \nu, \varepsilon$ *are constants (independent of* $m \in M$*), and with respect to an adapted orthonormal basis* $\{e_1 = x, e_2, \ldots, e_n\}$ *we have*

$$\|R^x\|^2 = \sum R_{1ijk}^2, \qquad D_1(x) = \sum \nabla_i R_{1j1k} \nabla_k R_{1i1j},$$

$$\|R^{xx}\|^2 = \sum R_{1i1j}^2, \qquad D_2(x) = \sum (\nabla_1 R_{ijkl})^2,$$

$$C_1(x) = \sum R_{1i1j} R_{1j1k} R_{1k1i}, \qquad D_3(x) = \sum (\nabla_i R_{1jkl})^2,$$

$$C_2(x) = \sum R_{1i1j} R_{1k1l} R_{ikjl}, \qquad D_4(x) = \sum (\nabla_k R_{1i1j})^2,$$

$$C_3(x) = \sum R_{1i1j} R_{ipqr} R_{jpqr}, \qquad D_5(x) = \sum (\nabla_1 R_{1i1j})^2,$$

$$C_4(x) = \sum R_{1jkl} R_{klpq} R_{1jpq}, \qquad D_6(x) = \sum R_{ijkl} \nabla_{11}^2 R_{ijkl},$$

$$C_5(x) = \sum R_{1j1l} R_{1lpq} R_{pq1j}, \qquad D_7(x) = \sum R_{1ijk} \nabla_{11}^2 R_{1ijk},$$

$$C_6(x) = \sum R_{1kjl} R_{kplq} R_{p1qj}, \qquad D_8(x) = \sum R_{1jkl} \nabla_{ii}^2 R_{1jkl},$$

$$C_7(x) = \sum R_{1kj1} R_{kp1q} R_{p1qj}, \qquad D_9(x) = \sum R_{1i1j} \nabla_{11}^2 R_{1i1j},$$

$$D_{10}(x) = \sum R_{1i1j} \nabla_{kk}^2 R_{1i1j},$$

all the summations being extended over $i, j, k, l, p, q = 2, 3, \ldots, n$.

Proof. Let $m \in M$ be a fixed point and $r > 0$ sufficiently small. According to [3], we have

$$V_m(r) = \int_0^r s^{n-1} \int_{S^{n-1}(1)} \theta(\exp_m su)\, du\, ds$$

where $S^{n-1}(1)$ is the unit sphere in $T_m M$ and θ denotes the "normal volume function" in the neighborhood of m. For $p = \exp_m(su)$, $u \in S^{n-1}(1)$, $s \in [0, r]$, we have

$$\begin{aligned}
\theta(p) = 1 &- \frac{s^2}{6} \rho_{uu}(m) - \frac{s^3}{12} (\nabla_u \rho_{uu})(m) \\
&+ \frac{s^4}{24} \left(-\tfrac{3}{5} \nabla_{uu}^2 \rho_{uu} + \tfrac{1}{3} \rho_{uu}^2 - \tfrac{2}{15} \sum_{a,b=1}^{n} R_{uaub}^2 \right)(m) \\
&+ \frac{s^5}{120} \left(-\tfrac{2}{3} \nabla_{uuu}^3 \rho_{uu} + \tfrac{5}{3} \rho_{uu} \nabla_u \rho_{uu} - \tfrac{2}{3} \sum_{a,b=1}^{n} R_{uaub} \nabla_u R_{uaub} \right)(m) \\
&+ \frac{s^6}{720} \Big\{ -\tfrac{5}{7} \nabla_{uuuu}^4 \rho_{uu} + 3 \rho_{uu} \nabla_{uu}^2 \rho_{uu} + \tfrac{5}{2} (\nabla_u \rho_{uu})^2 \\
&\qquad - \tfrac{5}{9} \rho_{uu}^3 - \tfrac{8}{7} \sum_{a,b=1}^{n} R_{uaub} \nabla_{uu}^2 R_{uaub} - \tfrac{15}{14} \sum_{a,b=1}^{n} (\nabla_u R_{uaub})^2 \\
&\qquad - \tfrac{16}{63} \sum_{a,b,c=1}^{n} R_{uaub} R_{ubuc} R_{ucua} + \tfrac{2}{3} \rho_{uu} \sum_{a,b=1}^{n} R_{uaub}^2 \Big\}(m) \\
&+ O(s^7).
\end{aligned}$$

Here only the *even* coefficients are essential because the integral of any odd coefficient over $S^{n-1}(1)$ is zero.

Next we consider the geodesic disk as a hypersurface $\tilde{M}$ of M and we shall mark by $\sim$ the corresponding quantities on $\tilde{M}$. Then we have

$$V_m^x(r)=\int_0^r s^{n-2}\int_{S^{n-2}(1)}\tilde{\theta}(\exp_m s\,u)\,du\,ds$$

where u remains perpendicular to x and the function $\tilde{\theta}(p)$ admits an analogous power series expansion as $\theta(p)$ with the tensors $R, \rho, \nabla R, \nabla\rho, \ldots$ replaced by the corresponding tensors $\tilde{R}, \tilde{\rho}, \tilde{\nabla}\tilde{R}, \tilde{\nabla}\tilde{\rho}, \ldots$ on the hypersurface $\tilde{M}$. Whenever it is necessary the quantities in the new expansion will be expressed with respect to an orthonormal basis $\{e_2, \ldots, e_n\}$ of $T_m\tilde{M}=\{x\}^{\perp}$. In accordance with that, all the summations will be taken with respect to the index set $\{2, 3, \ldots, n\}$.

To obtain the required formula, we first have to express the quantities which appear in the coefficients of s^2, s^4 and s^6 in $\tilde{\theta}(p)$ as functions of the curvature components $R_{ijkl}, \ldots, \nabla^4_{abcd}R_{ijkl}$. We shall use some well-known facts about hypersurfaces, here.

Let X, Y, Z, W be tangent vector fields of $\tilde{M}$ and let N be a unit normal vector field of $\tilde{M}$ in M. Then the shape operator S is defined by

$$\nabla_Y N=SY$$

and we have

$$\tilde{\nabla}_X Y=\nabla_X Y+g(SX, Y)N.$$

Further we will use the Gauss equation

$$\tilde{R}_{XYZW}=R_{XYZW}+S(X,Z)S(Y,W)-S(X,W)S(Y,Z)$$

where $S(X, Y)=g(SX, Y)$, and the Codazzi equation

$$R_{XYZN}=(\tilde{\nabla}_X S)(Y,Z)-(\tilde{\nabla}_Y S)(X,Z).$$

Along any geodesic γ of the form $s\mapsto\exp_m(s\,u)$, $u\in S^{n-2}(1)\subset T_m\tilde{M}$, we have $\tilde{\nabla}_{\gamma'}\gamma'=\nabla_{\gamma'}\gamma'=0$ and hence $S(\gamma', \gamma')=0$. Further we get from the Gauss and Codazzi equations

$$\tilde{R}_{\gamma' Y\gamma' Z}=R_{\gamma' Y\gamma' Z}-S(\gamma', Y)S(\gamma', Z), \tag{1}$$

$$\tilde{R}_{\gamma' Y\gamma' 1}=(\tilde{\nabla}_{\gamma'}S)(\gamma', Y)-(\tilde{\nabla}_Y S)(\gamma', \gamma') \tag{2}$$

along γ. Further since S is symmetric and $S(u, u)=0$ on $S^{n-2}(1)\subset T_m\tilde{M}$, we get at the center m

$$S(m)=0. \tag{3}$$

Putting $y=Y(m), z=Z(m)$ we obtain from the Gauss equation and (3):

$$\tilde{R}_{uyuz}=R_{uyuz}. \tag{4}$$

Next we take the covariant derivatives of (1) and we evaluate at the center m.

This gives

$$\begin{aligned}
\tilde{\nabla}_u \tilde{R}_{uyuz} &= \nabla_u R_{uyuz},\\
\tilde{\nabla}^2_{uu} \tilde{R}_{uyuz} &= \nabla^2_{uu} R_{uyuz} - R_{uyu1}\tilde{\nabla}_u S_{uz} - R_{uzu1}\tilde{\nabla}_u S_{uy} - 2\tilde{\nabla}_u S_{uy}\tilde{\nabla}_u S_{uz},\\
\tilde{\nabla}^4_{uuuu}\tilde{R}_{uyuy} &= \nabla^4_{uuuu} R_{uyuy} - 8\nabla^2_{uu} R_{uyu1}\tilde{\nabla}_u S_{uy} + 6R_{u1u1}(\tilde{\nabla}_u S_{uy})^2\\
&\quad - 6\sum_{i=2}^{n} R_{uyui}\tilde{\nabla}_u S_{uy}\tilde{\nabla}_u S_{ui} - 4\nabla_u R_{uyu1}\tilde{\nabla}^2_{uu} S_{uy} - 6(\tilde{\nabla}^2_{uu} S_{uy})^2\\
&\quad - 2R_{uyu1}\tilde{\nabla}^3_{uuu} S_{uy} - 8\tilde{\nabla}_u S_{uy}\tilde{\nabla}^3_{uuu} S_{uy}. \qquad (5)
\end{aligned}$$

Hence we need $\tilde{\nabla}_u S_{uy}$, $\tilde{\nabla}^2_{uu} S_{uy}$ and $\tilde{\nabla}^3_{uuu} S_{uy}$. First we note that

$$(\tilde{\nabla}_{\gamma'} S)(\gamma', \gamma') = 0$$

along γ and this implies at the center m, using the usual linearization procedure:

$$\tilde{\nabla}_y S_{uu} = 2\tilde{\nabla}_u S_{uy}.$$

So (2) gives after evaluation at m:

$$\tilde{\nabla}_u S_{uy} = \tfrac{1}{3} R_{uyu1}. \qquad (6)$$

Next we differentiate (2) covariantly along γ' and evaluate at m:

$$\nabla_u R_{uyu1} = \tilde{\nabla}^2_{uu} S_{uy} - \tilde{\nabla}^2_{uy} S_{uu}.$$

To compute the second term on the right-hand side we note that

$$(\tilde{\nabla}^2_{\gamma'\gamma'} S)(\gamma', \gamma') = 0.$$

Using the linearization at m and the Ricci identity we get

$$\tilde{\nabla}^2_{uy} S_{uu} + \tilde{\nabla}^2_{uu} S_{uy} = 0$$

and hence

$$\tilde{\nabla}^2_{uu} S_{uy} = \tfrac{1}{2}\nabla_u R_{uyu1}. \qquad (7)$$

Finally we differentiate (2) once again to obtain

$$\nabla^2_{uu} R_{uyu1} + \sum_{i=2}^{n} R_{uyui}\tilde{\nabla}_u S_{ui} - R_{u1u1}\tilde{\nabla}_u S_{uy} = \tilde{\nabla}^3_{uuu} S_{uy} - \tilde{\nabla}^3_{uuy} S_{uu}.$$

Using $(\tilde{\nabla}^3_{\gamma'\gamma'\gamma'} S)(\gamma', \gamma') = 0$, the linearization and the Ricci identity, we obtain

$$\tilde{\nabla}^3_{uuy} S_{uu} = -\tfrac{2}{3}\tilde{\nabla}^3_{uuu} S_{uy} + \tfrac{4}{3}\sum_{i=2}^{n} R_{uyui}\tilde{\nabla}_u S_{ui}$$

and hence

$$\tilde{\nabla}^3_{uuu} S_{uy} = \tfrac{3}{5}\tilde{\nabla}^2_{uu} R_{uyu1} - \tfrac{1}{5} R_{uyu1} R_{u1u1} + \tfrac{7}{15}\sum_{i=2}^{n} R_{uyui} R_{u1ui}. \qquad (8)$$

Now we substitute (6), (7) and (8) in the formulas (5). So we get

$$\begin{aligned}
\tilde{\nabla}_u \tilde{R}_{uyuz} &= \nabla_u R_{uyuz},\\
\tilde{\nabla}^2_{uu} \tilde{R}_{uyuz} &= \nabla^2_{uu} R_{uyuz} - \tfrac{8}{9} R_{uyu1} R_{uzu1},\\
\tilde{\nabla}^4_{uuuu} \tilde{R}_{uyuy} &= \nabla^4_{uuuu} R_{uyuy} - \tfrac{82}{15} R_{uyu1} \nabla^2_{uu} R_{uyu1} - \tfrac{7}{2} (\nabla_u R_{uyu1})^2\\
&\quad + \tfrac{8}{5} R_{uyu1} R_{uyu1} R_{u1u1} - \tfrac{128}{15} R_{uyu1} \sum_{i=2}^{n} R_{uyui} R_{u1ui}.
\end{aligned} \tag{9}$$

Finally, we substitute (4) and (9) into the coefficients of s^2, s^4, s^6 in $\tilde{\theta}(p)$ and we integrate over $S^{n-2}(1)$. Here we use the same technique as in [3] to obtain the required result. We omit the very long computations which do not involve new concepts or techniques.

We shall now draw some consequences from Theorem A and Theorem 2.1. We shall also use the denotations $\|R^x\|^2 = \sum (R_{1ijk})^2$, $\|R^{xx}\|^2 = \sum (R_{1i1j})^2$ from Theorem 2.2.

Proposition 2.3. *Let (M, g) be an analytic Riemannian manifold which satisfies the assumptions of the conjecture* (BD) *or the conjecture* (SD). *Then*

(a) *(M, g) is Einsteinian;*

(b) *$\|R\|^2$ is a constant function on M;*

(c) *$6\|R^x\|^2 + 5\|R^{xx}\|^2$ is a constant function on the unit sphere bundle $SM \subset TM$.*

Proof. For $\dim M = 2$ the theorem follows at once from the constancy of A on M. For $\dim M > 2$, both assumptions (BD) and (SD) imply that $A(x)$ and $B(x)$ are constant functions on each unit sphere $S_m M \subset T_m M$. Hence it follows that (M, g) is Einsteinian and τ is constant on M; consequently

$$-3\|R\|^2_m + 24\|R^x\|^2 + 20\|R^{xx}\|^2 \tag{10}$$

is constant on each $S_m M$ and since (see [2])

$$\begin{aligned}
\int_{S_m M} \|R^x\|^2 \, dx &= \frac{n-1}{n(n+2)} \left(\|R\|^2 - \frac{2}{n-1} \|\rho\|^2 \right)_m \operatorname{vol} S_m M,\\
\int_{S_m M} \|R^{xx}\|^2 \, dx &= \frac{1}{n(n+2)} \left(\tfrac{3}{2} \|R\|^2 + \|\rho\|^2 \right)_m \operatorname{vol} S_m M
\end{aligned} \tag{11}$$

we obtain

$$6\|R^x\|^2 + 5\|R^{xx}\|^2 = \frac{1}{n(n+2)} \left\{ \frac{12n+3}{2} \|R\|^2 - 7\|\rho\|^2 \right\}_m. \tag{12}$$

In the case of the (BD)-conjecture A, B and C are constant functions on M. Hence $\|R\|^2$ is constant on M. This proves (b) and (c).

For the (SD)-conjecture the assumptions imply that the expression (10) is constant on SM and since

$$-3\|R\|^2_m + 24\|R^x\|^2 + 20\|R^{xx}\|^2 = -\frac{1}{n(n+2)} \{3(n^2 - 6n - 2)\|R\|^2 + 28\|\rho\|^2\}_m$$

it follows that $\|R\|^2$ is constant on SM. Again, this proves the required result.

Corollary. *The conjecture* (BD) *is true for the dimensions* $n=2$ *and* $n=3$ *and the conjecture* (SD) *is true for dimension* 3.

We shall also need the following

Lemma 1. *On any Einstein manifold we have*

$$\Delta\|R\|^2=\frac{4\tau}{n}\|R\|^2-2\check{R}-8\check{\check{R}}+2\|\nabla R\|^2.$$

Proof. This is Lichnerowicz's formula (see [6], p. 9–10). See also [3, (2.18)].

3. Proof of the Conjecture (BD) for Dimension 4

We know already that (M,g) must be a four-dimensional Einstein manifold. Following [1], we connect with any fixed point $m\in M$ a *Singer-Thorpe* basis (shortly: ST-basis) $\{e_1,e_2,e_3,e_4\}$ of T_mM. This is an orthonormal basis with the following properties: let $a=\max\sigma$, $c=\min\sigma$ be the maximum and the minimum value of the sectional curvature σ at m, respectively, and put $b=\frac{\tau}{4}-a-c$. Then for suitable constants α,β,γ (depending on $m\in M$) such that $\alpha+\beta+\gamma=0$, all the components of the curvature tensor R are given by the following formulas:

$$R_{1212}=R_{3434}=a,\quad R_{1313}=R_{2424}=b,\quad R_{1414}=R_{2323}=c,$$
$$R_{1234}=\alpha,\quad R_{1342}=\beta,\quad R_{1423}=\gamma,$$

$R_{ijkl}=0$ whenever just 3 of the indices i,j,k,l are distinct. Equivalently, with respect to the corresponding basis $\{e_1\wedge e_2,e_1\wedge e_3,e_1\wedge e_4,e_3\wedge e_4,e_4\wedge e_2,e_2\wedge e_3\}$ of Λ^2T_mM, the curvature operator $\mathscr{R}$: $\Lambda^2T_mM\to\Lambda^2T_mM$ has the matrix of the form

$$\begin{bmatrix}A & B\\ B & A\end{bmatrix},$$

where

$$A=\begin{bmatrix}a & 0 & 0\\ 0 & b & 0\\ 0 & 0 & c\end{bmatrix},\qquad B=\begin{bmatrix}\alpha & 0 & 0\\ 0 & \beta & 0\\ 0 & 0 & \gamma\end{bmatrix}$$

(see also [8]).

Let us consider the symmetric bilinear form $\dot{R}$ on T_mM given by

$$\dot{R}_{uv}=\sum_{i,j,k=1}^{4}R_{uijk}R_{vijk}$$

for any orthonormal basis $\{f_1, f_2, f_3, f_4\}$. We shall prove that $\dot{R} = \frac{1}{4}\|R\|^2 g$ at m. Indeed, with respect to our ST-basis we get easily

$$\dot{R}_{ii} = 2(a^2+b^2+c^2+\alpha^2+\beta^2+\gamma^2), \quad i=1,\ldots,4;$$
$$\dot{R}_{ij} = 0, \quad i,j=1,\ldots,4,\ i\neq j,$$
$$\|R\|^2 = 8(a^2+b^2+c^2+\alpha^2+\beta^2+\gamma^2)$$

which proves our equality. Due to part (b) of Proposition 2.3, the quadratic form $\dot{R}(x,x) = \frac{1}{4}\|R\|^2 g(x,x)$ is constant on the whole unit sphere bundle $SM \subset TM$. On the other hand, for any $m \in M$ and any $x \in S_m M$ we get (using an adapted orthonormal frame $\{f_1 = x, f_2, f_3, f_4\}$

$$\dot{R}(x,x) = \|R^x\|^2 + 2\|R^{xx}\|^2.$$

Using part (c) of Proposition 2.3 we see that $\|R^x\|^2$ and $\|R^{xx}\|^2$ are constant functions on SM. For a special choice of x as the first vector e_1 of an ST-basis $\{e_1, e_2, e_3, e_4\}$ we can calculate explicitly:

$$\|R^{xx}\|^2 = a^2+b^2+c^2, \quad \|R^x\|^2 = 2(\alpha^2+\beta^2+\gamma^2).$$

Next we take a new orthonormal basis

$$\{f_1, f_2, f_3, f_4\} = \left\{\frac{1}{\sqrt{2}}(e_1+e_2), \frac{1}{\sqrt{2}}(e_1-e_2), \frac{1}{\sqrt{2}}(e_3+e_4), \frac{1}{\sqrt{2}}(e_3-e_4)\right\}$$

and we calculate $\|R^{f_1 f_1}\|^2$. We obtain easily

$$\|R^{f_1 f_1}\|^2 = a^2 + \tfrac{1}{2}(b+c)^2 + \tfrac{1}{2}(\gamma-\beta)^2$$

which must be equal to $a^2+b^2+c^2$. Hence we obtain $(b-c)^2 = (\beta-\gamma)^2$, and similarly, we get $(a-b)^2 = (\alpha-\beta)^2$, $(c-a)^2 = (\gamma-\alpha)^2$. Now, there is a unique orientation of the tangent space $T_m M$ such that for each *positive* ST-basis

$$a-b = \alpha-\beta, \quad b-c = \beta-\gamma, \quad c-a = \gamma-\alpha,$$

and hence

$$\alpha = \tfrac{1}{3}(2a-b-c), \quad \beta = \tfrac{1}{3}(2b-a-c), \quad \gamma = \tfrac{1}{3}(2c-a-b). \tag{13}$$

We know already that a, b, c are subjected on M to the relation

$$a+b+c = \rho_{11} = \frac{\tau}{4} = K \tag{14}$$

where K is a constant on M. We can rewrite (13) in the form

$$a-\alpha = b-\beta = c-\gamma = \tfrac{1}{3}K \tag{15}$$

or, more specifically

$$R_{1212} - R_{3412} = R_{3434} - R_{1234} = R_{1313} - R_{4213} = R_{4242} - R_{1342}$$
$$= R_{1414} - R_{2314} = R_{2323} - R_{1423} = \tfrac{1}{3}K \tag{16}$$

for a positive ST-basis $\{e_1, e_2, e_3, e_4\}$.

Let us recall that, for our fixed orientation of $V=T_mM$, we have the (uniquely determined) *Hodge operator* $*: \Lambda^2 V \to \Lambda^2 V$. For *any* positive orthonormal basis $\{f_1, f_2, f_3, f_4\}$ we get

$$*(f_1 \wedge f_2)=f_3 \wedge f_4, \quad *(f_1 \wedge f_3)=f_4 \wedge f_2, \quad *(f_1 \wedge f_4)=f_2 \wedge f_3.$$

Also, we have $*\circ *=\mathrm{Id}$, and $\langle *\lambda, *\lambda\rangle_2=\langle \lambda, \lambda\rangle_2$ for each $\lambda \in \Lambda^2 V$. The Einstein property of (M, g) means that the Hodge operator commutes with the curvature operator $\mathscr{R}: \Lambda^2 V \to \Lambda^2 V$ (see [8]).

Now, (16) can be easily rewritten in the following invariant form:

$$\left(\mathscr{R}-\frac{K}{3}I\right)\circ(I-*)=0 \quad \text{on } \Lambda^2 V.$$

But this means that (16) is true for *any* positive orthonormal basis $\{f_1, f_2, f_3, f_4\}$ (see [1]).

Proposition 3.1. *Along the manifold M, the extremal sectional curvatures $a=\max\sigma$, and $c=\min\sigma$ are constant.*

Proof. We put $b=K-a-c$, as usual. With respect to a ST-basis we have

$$a^2+b^2+c^2+\alpha^2+\beta^2+\gamma^2=\tfrac{1}{8}\|R\|^2$$

which is constant on M. Using (14) and (15) we see easily that the functions $a^2+b^2+c^2$, $ab+bc+ac$ are also constant on M; in particular we can write

$$ab+bc+ac=L. \tag{17}$$

Using the constancy of the function $C(m)$ from Theorem A, and Proposition 2.3, we obtain

$$\begin{aligned}\tfrac{64}{63}\check{\rho}-\tfrac{64}{21}\langle \rho\otimes\rho, \overline{R}\rangle+\tfrac{32}{7}\langle \rho, \dot{R}\rangle-\tfrac{110}{63}\check{R}-\tfrac{200}{63}\check{\overline{R}}-\tfrac{45}{14}\|\nabla R\|^2 \\ -\tfrac{30}{7}\langle \Delta R, R\rangle=\text{constant}.\end{aligned} \tag{18}$$

Now, with respect to a ST-basis we get easily by direct calculation

$$\begin{aligned}&\check{\rho}=4K^3, \quad \langle \rho\otimes\rho, \overline{R}\rangle=4K^3, \quad \langle \rho, \dot{R}\rangle=K\|R\|^2,\\ &\check{R}=16(a^3+b^3+c^3+3a\alpha^2+3b\beta^2+3c\gamma^2),\\ &\check{\overline{R}}=24(abc+a\beta\gamma+b\alpha\gamma+c\alpha\beta).\end{aligned} \tag{19}$$

Using (13), (14) and (17) we obtain on M:

$$\begin{aligned}\check{R}&=192abc+\text{constant},\\ \check{\overline{R}}&=96abc+\text{constant}.\end{aligned}$$

Now, $\langle \Delta R, R\rangle=\frac{1}{2}\Delta\|R\|^2-\|\nabla R\|^2=-\|\nabla R\|^2$, and taking (19) into account, we can reduce (18) to the form

$$-\tfrac{110}{63}\check{R}-\tfrac{200}{63}\check{\overline{R}}+\tfrac{15}{14}\|\nabla R\|^2=\text{constant}. \tag{21}$$

From lemma 1 we get

$$\check{R}+4\check{\bar{R}}-\|\nabla R\|^2=\text{constant} \tag{22}$$

and combining (21) and (22) we derive

$$28\check{\bar{R}}-17\check{R}=\text{constant}.$$

Hence and from (20) we have $abc=$constant on M, and this completes the proof.

Let us consider a quaternionic structure (I,J,K) on T_mM adapted to the given positive ST-basis $\{e_1,e_2,e_3,e_4\}$. This means, for $x=a^1e_1+a^2e_2+a^3e_3+a^4e_4$ we put

$$\begin{aligned} Ix&=-a^2e_1+a^1e_2-a^4e_3+a^3e_4,\\ Jx&=-a^3e_1+a^4e_2+a^1e_3-a^2e_4,\\ Kx&=-a^4e_1-a^3e_2+a^2e_3+a^1e_4. \end{aligned} \tag{23}$$

We shall derive a series of lemmas which will be needed in the crucial part of the proof.

Lemma 2. *For any unit vector $x\in S_mM$, $\{x,Ix,Jx,Kx\}$ is a positive ST-basis.*

Proof. For the curvature operator $\mathscr{R}\colon \Lambda^2T_mM\to\Lambda^2T_mM$ we have

$$\begin{aligned} \mathscr{R}(x\wedge Ix)&=a(x\wedge Ix)+\alpha(Jx\wedge Kx),\\ \mathscr{R}(x\wedge Jx)&=b(x\wedge Jx)+\beta(Kx\wedge Ix),\\ \mathscr{R}(x\wedge Kx)&=c(x\wedge Kx)+\gamma(Ix\wedge Jx),\\ \mathscr{R}(Jx\wedge Kx)&=a(Jx\wedge Kx)+\alpha(x\wedge Ix),\\ \mathscr{R}(Kx\wedge Ix)&=b(Kx\wedge Ix)+\beta(x\wedge Jx),\\ \mathscr{R}(Ix\wedge Jx)&=c(Ix\wedge Jx)+\gamma(x\wedge Kx). \end{aligned} \tag{24}$$

Indeed, by the definition of a positive ST-basis we obtain all the expressions $\mathscr{R}(e_i\wedge e_j)$, $i,j=1,\dots,4$, and using the linear coordinates of x with respect to $\{e_1,e_2,e_3,e_4\}$ we can prove easily our formulas. (In fact, it suffices to prove only the first half and then use the Hodge operator $*$.) Our lemma now follows from (24).

Lemma 3. *Let $\varphi(x)$ be a polynomial function defined on the unit sphere bundle $SM\subset TM$ in the following way: for any unit vector $x\in SM$ and any adapted orthonormal basis $\{f_1,f_2,f_3,f_4\}$ we have*

$$\varphi(x)=\sum R_{\alpha_1\alpha_2\alpha_3\alpha_4}R_{\alpha_5\alpha_6\alpha_7\alpha_8}\dots R_{\alpha_{4k-3}\alpha_{4k-2}\alpha_{4k-1}\alpha_{4k}}$$

where an even number of the indices $\alpha_1,\alpha_2,\dots,\alpha_{4k}$ is equal to 1, and the remaining α_i are distributed into pairs of equal summation indices, with the summation ranging over the set $\{1,2,3,4\}$, or over the set $\{2,3,4\}$ respectively. Then $\varphi(x)$ is a constant function on SM.

Proof. For each $x \in SM$ we can calculate $\varphi(x)$ by means of a ST-basis of the form $\{x, Ix, Jx, Kx\}$. Then $\varphi(x)$ becomes a well-determined function of the (constant!) parameters $a, b, c, \alpha, \beta, \gamma$ and hence it is constant.
As an immediate consequence we obtain

Lemma 4. *The functions $C_1(x), \ldots, C_7(x)$ defined in Theorem 2.2 are constant on the unit sphere bundle SM.*

Lemma 5. *The function* $\psi(x) = \sum_{\alpha,\beta,\gamma=1}^{4} (\nabla_\alpha R_{1\beta 1\gamma})^2$ *is constant on SM.*

Proof. By the same sort of calculation needed to prove Lemma 1 we obtain the general formula

$$\sum_1^n \nabla^2_{\gamma\gamma} R^2_{1\alpha 1\beta} = \frac{4\tau}{n} \sum_1^n R^2_{1\alpha 1\beta} - 2\sum_1^n R_{1\alpha\gamma\delta} R_{1\beta\gamma\delta} R_{1\alpha 1\beta} + 4\sum_1^n R_{1\delta\alpha\gamma} R_{1\gamma\beta\delta} R_{1\alpha 1\beta}$$
$$- 4\sum_1^n R_{\alpha\gamma\beta\delta} R_{1\alpha 1\beta} R_{1\gamma 1\delta} + 2\sum_1^n (\nabla_\gamma R_{1\alpha 1\beta})^2.$$

Now, our assertion follows due to Lemma 3 (for $n=4$).

Lemma 6. *For the functions $D_1(x), \ldots, D_{10}(x)$ from Theorem 2.2 the following identities hold:*

$$D_6(x) = -D_2(x),$$
$$D_7(x) = -\sum_2^4 (\nabla_1 R_{1ijk})^2,$$
$$D_8(x) = -D_3(x),$$
$$D_9(x) = -D_5(x),$$
$$D_{10}(x) = -D_4(x),$$
$$D_1(x) = D_4(x) - \tfrac{1}{2}\sum (\nabla_1 R_{1ijk})^2.$$

Proof. To obtain the last formula, we use the Bianchi identities. The other formulas follow from the fact that the functions

$$\sum_2^4 R^2_{1ijk} = \|R^x\|^2,$$
$$\sum_2^4 R^2_{1i1j} = \|R^{xx}\|^2,$$
$$\sum_2^4 R^2_{ijkl} = \|R\|^2 - 4\|R^x\|^2 - 4\|R^{xx}\|^2$$

are constant on the unit sphere bundle SM, by taking the second covariant derivatives of these functions.

Proposition 3.2. *The space (M, g) is locally symmetric.*

Proof. Choose a fixed point $m \in M$ and a positive ST-basis $\{e_1, e_2, e_3, e_4\}$ in $T_m M$. Let $\{E_1, E_2, E_3, E_4\}$ be the extension of our basis to a local frame field, using the parallel transport along the geodesics emanating from m. (The frames are no more ST-bases, in general.) Now, each covariant derivative $(\nabla_p R_{ijkl})_m$ is nothing but the usual derivative $e_p[R(E_i, E_j, E_k, E_l)](m)$. According to Proposition 3.1, the function $R(E_1, E_2, E_1, E_2)$ acquires its local maximum, a, at the point m. Hence $d\{R(E_1, E_2, E_1, E_2)\} = 0$ at m, i.e. $(\nabla_p R_{1212}) = 0$ for $p = 1, 2, 3, 4$. Analogously, we get $\nabla_p R_{3434} = 0$ and $\nabla_p R_{1414} = \nabla_p R_{2323} = 0$ *at* m. Because $R_{1212} + R_{1313} + R_{1414} = \rho_{11} = K$ in the whole neighborhood of m, we obtain $(\nabla_p R_{1313})_m = 0$, and analogously, we have $(\nabla_p R_{2424})_m = 0$. According to (16), which is valid for *any* positive orthonormal basis, we get $(\nabla_p R_{ijkl})_m = 0$ whenever i, j, k, l are all distinct.

We are left with the terms of the form $(\nabla_p R_{ijik})_m$, where i, j, k are all distinct. Using the identities

$$\sum_{p=1}^{4} \nabla_p R_{pijk} = 0, \qquad \sum_{p=1}^{4} \nabla_k R_{pipj} = 0, \qquad i, j, k = 1, \ldots, 4,$$

and the Bianchi identities we see that each term of the above type is equal (up to a sign) either to a term of the form $(\nabla_i R_{ijik})_m$, or to a term of the form $(\nabla_i R_{ijjk})_m$, i, j, k all distinct.

We can also see that

$$(\nabla_i R_{ijik})_m = (\nabla_j R_{jijk})_m = (\nabla_k R_{kikj})_m$$

for any 3 distinct indices i, j, k. The common value of that will be briefly denoted by the symbol $\textcircled{l}$, where l is the "missing" index.

Further, we see that, for any fixed k, the term $(\nabla_i R_{ijjk})^2$ is independent of the choice of i and j, assuming only that $i, j \neq k$, $i \neq j$. Therefore, we shall denote the common absolute value of $(\nabla_i R_{ijjk})_m$ by the symbol $\boxed{k}$. (We must keep in mind that the symbols $\textcircled{l}$ and $\boxed{k}$ are always connected with a concrete ST-basis of $T_m M$.)

So far, we followed closely the procedure from [1]. Now, we shall use the constancy of the function $C(x)$ on $S_m M$. According to Theorem 2.2, Lemma 4 and Lemma 6, we see easily that this condition implies

$$50 D_2(x) + 2 D_3(x) + 21 D_4(x) + 205 D_5(x) + \tfrac{691}{3} \sum_{2}^{4} (\nabla_1 R_{1ijk})^2 = \text{constant (on } S_m M). \tag{25}$$

Further, Lemma 5 implies that

$$D_4(x) + D_5(x) = \text{constant (on } SM). \tag{26}$$

If $\{e_1' = x, e_2', e_3', e_4'\}$ is any ST-basis at m, for instance $\{x, Ix, Jx, Kx\}$, we can check easily the following formulas

$$D_2(x)=8[②^2+③^2+④^2],$$
$$D_3(x)=8[②^2+③^2+④^2]+12\boxed{1}^2,$$
$$D_4(x)=4(\boxed{2}^2+\boxed{3}^2+\boxed{4}^2)+6①^2,$$
$$D_5(x)=2[②^2+③^2+④^2],$$
$$\sum(V_1R_{1ijk})^2=4(\boxed{2}^2+\boxed{3}^2+\boxed{4}^2).$$

Then (25) and (26) are reduced to the conditions

$$413(\boxed{2}^2+\boxed{3}^2+\boxed{4}^2)+12\boxed{1}^2+\tfrac{1508}{3}[②^2+③^2+④^2]+63①^2=\text{constant}, \tag{27}$$

$$[②^2+③^2+④^2]+2(\boxed{2}^2+\boxed{3}^2+\boxed{4}^2)+3①^2=\text{constant} \tag{28}$$

on the set of all ST-bases at m.

For a ST-basis $\Lambda=\{e_1,e_2,e_3,e_4\}$ put

$$A(\Lambda)=①^2+②^2+③^2+④^2,$$
$$B(\Lambda)=\boxed{1}^2+\boxed{2}^2+\boxed{3}^2+\boxed{4}^2,$$
$$\Lambda_i=\{e_i,Ie_i,Je_i,Ke_i\},\qquad i=1,2,3,4.$$

Then for each $i=1,\ldots,4$

$$A(\Lambda_i)=A(\Lambda),\qquad B(\Lambda_i)=B(\Lambda)$$

and $①^2,\boxed{1}^2$ are transformed into $ⓘ^2,\boxed{i}^2$ respectively.

(27) and (28) can be rewritten in the form

$$413A(\Lambda)+\tfrac{1508}{3}B(\Lambda)-350①^2-\tfrac{1472}{3}\boxed{1}^2=\text{constant},$$
$$A(\Lambda)+2B(\Lambda)+2①^2-2\boxed{1}^2=\text{constant}$$

on the set of all ST-bases at m. We conclude that

$$350①^2+\tfrac{1472}{3}\boxed{1}^2=\ldots=350④^2+\tfrac{1472}{3}\boxed{4}^2,$$
$$①^2-\boxed{1}^2=\ldots=④^2-\boxed{4}^2$$

and hence

$$①^2=②^2=③^2=④^2=\tfrac{1}{4}A(\Lambda),$$
$$\boxed{1}^2=\boxed{2}^2=\boxed{3}^2=\boxed{4}^2=\tfrac{1}{4}B(\Lambda).$$

Substituting again into (27) and (28) we get

$$615A(\Lambda)+760B(\Lambda)=\text{constant},$$
$$A(\Lambda)+B(\Lambda)=\text{constant}.$$

Hence the expressions $A(\Lambda)$, $B(\Lambda)$, $ⓘ^2$, $\boxed{i}^2$ are independent of the choice of an ST-basis Λ.

Let us denote by M^2, N^2 the constant values of $\textcircled{i}^2$, $\boxed{i}^2$ respectively, and consider two ST-bases:

$$\Lambda = \{e_1, e_2, e_3, e_4\},$$

$$\Lambda' = \left\{\frac{1}{\sqrt{2}}(e_1+e_2), \frac{1}{\sqrt{2}}(e_2-e_1), \frac{1}{\sqrt{2}}(e_3-e_4), \frac{1}{\sqrt{2}}(e_3+e_4)\right\}.$$

By a careful calculation (using various equalities between the components $\nabla_i R_{jklp}$) we obtain

$$M^2 = (\nabla_{1'} R_{1'2'1'3'})^2 = \tfrac{1}{2}(\boxed{3} \pm \boxed{4})^2 = N^2 \pm \boxed{3}\boxed{4},$$

(where either $(+)$ signs or $(-)$ signs take place)

$$N^2 = (\nabla_{1'} R_{1'2'2'3'})^2 = \tfrac{1}{2}[\textcircled{3} - \textcircled{4}]^2 = M^2 - \textcircled{3}\textcircled{4}.$$

Because $\textcircled{3}\textcircled{4} = \pm M^2$, $\boxed{3}\boxed{4} = \pm N^2$, we get immediately $M = N = 0$, and hence $\nabla R = 0$, q.e.d.

Proposition 3.3. *For each unit vector $x \in T_m M$ the endomorphism R_x: $u \mapsto R(x,u)x$ has the eigenvalues $0, a, b, c$.*

Proof. From Lemma 2 we see that $R_x(x) = 0 \cdot x$, $R_x(Ix) = aIx$, $R_x(Jx) = bJx$, $R_x(Kx) = cKx$.

Proposition 3.4. *If $abc = 0$, then $R = 0$ on M.*

Proof. Because (M, g) is locally symmetric, we have $R(u,v) \cdot R = 0$ for any $u, v \in T_m M, m \in M$. Here $R(u,v) = \nabla^2_{uv} - \nabla^2_{vu}$ acts as a derivation on the tensor algebra at m. In particular, let $\{e_1, e_2, e_3, e_4\}$ be a ST-basis at m. Then $(R(e_1, e_2) \cdot R)(e_1, e_3, e_1, e_4) = 0$ implies $(a+\alpha)(b-c) = 0$, and similarly we get $(b+\beta)(a-c) = 0$, $(c+\gamma)(a-b) = 0$. In other words, $(5a-b-c)(b-c) = 0$, $(5b-a-c)(a-c) = 0$, $(5c-a-b)(a-b) = 0$. Suppose that $a = 0$, for instance. Then we get $b^2 - c^2 = 0$, $(5b-c)c = 0$, which implies $b = c = 0$. We also obtain $\alpha = \beta = \gamma = 0$, and hence $R = 0$.

Theorem 3.5. *(M, g) is either locally isometric to a rank one symmetric space, or it is locally Euclidean.*

Proof. Let $(\tilde{M}, \tilde{g})$ be a globally symmetric space which is locally isometric to (M, g), and suppose that $(\tilde{M}, \tilde{g})$ has rank greater than 1. This means that $\tilde{M}$ contains a totally geodesic and flat submanifold S of dimension $\geqq 2$ (see [4]). At any point $m \in S$, choose a pair $\{f_1, f_2\}$ of orthogonal unit vectors which are tangent to S. Then $\mathscr{R}(f_1 \wedge f_2) = 0$ (see [5], Chap. VIII). This means that 0 is a multiple eigenvalue of R_{f_1}, and according to Proposition 3.3, one of the numbers a, b, c is zero. According to Proposition 3.4 we have $R = 0$ and (M, g) is locally Euclidean. q.e.d.

4. Proof of the Conjecture (SD) for Dimension 4

Instead of giving the long and tedious calculations for this case we give the essential steps to prove the conjecture (SD) for dimension 4.

The main difference with the case of the conjecture (BD) is that we do not have $C=$constant but now we have that $C(x)$ is a constant function on the unit sphere bundle $SM \subset TM$. So we have to replace the formula (18) by another relation which gives also, as in (21), a relation between the cubic invariants $\tilde{R}$, $\breve{R}$, $\|\nabla R\|^2$.

This relation can be obtained in a similar way as we obtained the result (c) in Proposition 2.3. Indeed, we have

$$C(x)=D$$

where D is a constant on SM. Hence

$$\int_{S^{n-1}(1)} C(x)=D \operatorname{vol} S^{n-1}(1).$$

Now using an integration technique as in Proposition 2.3 (see for example [2, 3]) the left-hand member can be written as a linear combination of the cubic invariants which appear in $C(m)$. Hence we obtain an explicit expression for D as a function of the cubic invariants.

Using this expression the rest of the proof is similar to that of the conjecture (BD).

References

1. Besse, A.L.: Manifolds all of whose geodesics are closed. Ergebnisse der Mathematik **93**, Berlin-Heidelberg-New York: Springer 1978
2. Chen, B.Y., Vanhecke, L.: Differential geometry of geodesic spheres. J. Reine Angew. Math. **325**, 28–67 (1981)
3. Gray, A., Vanhecke, L.: Riemannian geometry as determined by the volumes of small geodesic balls. Acta Math. **142**, 157–198 (1979)
4. Helgason, S.: Differential geometry and symmetric spaces. New York: Academic Press 1962
5. Kobayashi, S., Nomizu, K.: Foundations of differential geometry. Vol. II. New York-London: Interscience 1969
6. Lichnerowicz, A.: Géométrie des groupes de transformation. Paris: Dunod 1958
7. Ruse, H.S., Walker, A.G., Willmore, T.J.: Harmonic spaces. Roma: Cremonese 1961
8. Singer, I.M., Thorpe, J.A.: The curvature of 4-dimensional Einstein spaces. In: Global Analysis (Papers in honor of K. Kodaira), pp. 355–366. Princeton, New Jersey: Princeton University Press 1969

Received January 20, 1982

Math. Z. 180, 445-461 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On Locally Indicable Groups

James Howie*

Department of Mathematics, University of Edinburgh, Edinburgh EH9 3JZ, United Kingdom

1. Introduction

A group is said to be *locally indicable* if each of its non-trivial finitely generated subgroups has the infinite cyclic group as a homomorphic image. Such groups were studied by Higman [9] in connection with the zero-divisor and unit problems for group rings. More recently, they have arisen [2, 12] in the study of equations over groups.

In [12], I proved a Freiheitsatz for locally indicable groups. This has been proved independently by Brodskiĭ [2] and Short [22]. The present paper arises from an investigation of a question put to me by S.J. Pride - whether torsion-free 1-relator groups are locally indicable. The question was raised originally by Baumslag ([1], Problem 19) and an affirmative solution has recently been announced by Brodskiĭ [2].

As a consequence, the group algebra RG of a torsion-free 1-relator group G over an integral domain R has no non-trivial zero-divisors, and no non-trivial units (using Higman's results [9]). The first fact was also proved by Lewin and Lewin [16], by embedding RG in a division ring. The second appears to be new, as was pointed out to me by K.A. Brown.

A second consequence is that no 1-relator group has a non-trivial finitely generated perfect subgroup, which answers [1], Problem 7 and [10], Question 1.

In fact, using the Freiheitsatz, and the tower method described in [12], it is possible to prove the following general version of Brodskiĭ's theorem.

Theorem 4.2. *Let A and B be locally indicable groups, and let G be the quotient of $A * B$ by the normal closure of a cyclically reduced word r of length at least 2. Then the following are equivalent:*

(i) *G is locally indicable;*
(ii) *G is torsion-free;*
(iii) *r is not a proper power in $A * B$.*

* Research partly supported by a William Gordon Seggie Brown Fellowship

0025-5874/82/0180/0445/$03.40

The group G in Theorem 4.2 will be called a 1-relator product of A and B. Thus the class of locally indicable groups is closed under the formation of 1-relator products in which the relator is not a proper power.

The methods used in the proof of Theorem 4.2 can also be used to study asphericity in 2-complexes with locally indicable fundamental groups.

Theorem 5.2. *Let L be a connected 2-complex with $H_2(L)=0$ and $\pi_1(L)$ locally indicable. Then L is aspherical.*

Theorem 5.3. *Let L be a 2-complex formed by attaching a 1-cell and a 2-cell to a 2-complex K, each component of which is aspherical with locally indicable fundamental group. Suppose the attaching map for the 2-cell of $L\backslash K$ is not homotopic in $K\cup L^{(1)}$ to a map into K, and does not represent a proper power in $\pi_1(K\cup L^{(1)})$. Then L is aspherical.*

These results may be applied to a class of group presentations which I call "reducible", and which includes, for example, all 1-relator and staggered presentations.

Corollary 4.5. *A group given by a reducible presentation with no proper powers is locally indicable.*

Corollary 5.4. *A reducible presentation with no proper powers is aspherical.*

These terms will be made precise in § 2, but an "aspherical" presentation is one whose geometric realisation is an aspherical 2-complex. There exist in the literature a variety of inter-related notions of asphericity for presentations, which have been extensively investigated by Chiswell, Collins and Huebschmann [4] (see also [3, 7, 17, 21]). Any staggered presentation (with or without proper powers) is Cohen-Lyndon aspherical in the sense of [4]. In contrast, the proper power condition in Corollary 5.4 is crucial: a group given by a reducible presentation need not have finite quasi-projective dimension in the sense of Howie and Schneebeli [13], so the presentation need not satisfy any of the asphericity conditions discussed in [4].

A large class of locally indicable groups arise as fundamental groups of suitable 3-manifolds. In particular all knot-groups and link-groups are locally indicable, which answers Question 3 of [10]. This fact seems to be understood by 3-dimensional topologists, but does not seem to be in the literature. I have included a proof in § 6, based on Scott's compact submanifold theorem. I am grateful to G.A. Swarup for the ideas behind this proof.

Short ([22, 23]) applies the Freiheitsatz for knot-groups to a problem of Lickorish ([15], Problem 1.1): a band sum of two knots can be the unknot only if each of the original knots is the unknot.

I am grateful to the referee and to S.D. Brodskiĭ for a number of improvements to my original version of this paper.

2. Notation and Definitions

A group G is *locally indicable* if every finitely generated subgroup A of G (other than the identity) admits an epimorphism onto the infinite cyclic group.

Equivalent ways of saying this are that $H^1(A)\neq 0$ or that $H_1(A)$ $(=A^{ab}=A/[A,A])$ is finite. (All homology and cohomology is with integer coefficients unless stated otherwise.)

The main object of study of this paper is 1-relator products of locally indicable groups, that is groups of the form $G=(A*B)/N$, where A and B are locally indicable, and N is the normal closure of a single element r, not conjugate to an element of A or of B. The relator r may be represented (up to conjugacy) as a cyclically reduced word of length at least 2. We must therefore study cyclically reduced words in free products, or more generally in graphs of groups with trivial edge groups.

We use the notation of Dicks [5] for graphs of groups. Thus a graph of groups $(\mathcal{G}, X)$ consists of a connected, oriented graph X, with vertex set $V=V(X)$ and edge set $E=E(X)$, together with groups $\mathcal{G}(v)$ $(v\in V)$ and $\mathcal{G}(e)$ $(e\in E)$, and monomorphisms ι_e: $\mathcal{G}(e)\to\mathcal{G}(\iota e)$, τ_e: $\mathcal{G}(e)\to\mathcal{G}(\tau e)$ $(e\in E)$, where ιe, τe denote the initial and terminal vertices respectively of e.

Following Higgins [8], we define the *fundamental groupoid* $\mathcal{F}(\mathcal{G}, X)$ of $(\mathcal{G}, X)$ to be the groupoid generated by the groups $\mathcal{G}(v)$ $(v\in V)$ and the edges $e\in E$, subject to the relations $e\cdot\tau_e(g)=\iota_e(g)\cdot e$ $(e\in E, g\in\mathcal{G}(e))$.

The *fundamental group* $\pi(\mathcal{G}, X)$ is then just a vertex group of $\mathcal{F}(\mathcal{G}, X)$, or alternatively the quotient of $\mathcal{F}(\mathcal{G}, X)$ by a maximal tree subgroupoid.

An element of $\mathcal{F}(\mathcal{G}, X)$ can be expressed as a *word*

$$w=g_0\cdot e_1^{\varepsilon(1)}\cdot g_1\cdot\ldots\cdot e_n^{\varepsilon(n)}\cdot g_n$$

where $g_j\in\mathcal{G}(v_j)$ $(v_j\in V)$, $\varepsilon(j)=\pm 1$, $e_j\in E$ and $\iota(e_j)=v_{j-1}$, $\tau(e_j)=v_j$ (if $\varepsilon(j)=1$), or $\iota(e_j)=v_j$, $\tau(e_j)=v_{j-1}$ (if $\varepsilon(j)=-1$).

The integer n is the *length* of the word (w). A word (w) is *reduced* if it does not contain a sequence of the form $e\cdot\tau_e(h)\cdot e^{-1}$ or $e^{-1}\cdot\iota_e(h)\cdot e$ $(h\in\mathcal{G}(e))$. It is *closed* if $v_0=v_n$. A closed word (w) is *cyclically reduced* if it is reduced and if $e_n^{\varepsilon(n)}\cdot(g_n g_0)\cdot e_1^{\varepsilon(1)}$ is reduced.

Elements of $\mathcal{F}(\mathcal{G}, X)$ have a normal form [8], which in general depends on choices of transversals for the images of the maps ι_e, τ_e. In the case where the edge groups $\mathcal{G}(e)$ are trivial, there is a unique choice of transversals, so a unique normal form, and the normal forms are precisely the reduced words. Hence every element of $\mathcal{F}(\mathcal{G}, X)$ is uniquely defined by a reduced word. Also, every element in $\pi(\mathcal{G}, X)$ is conjugate in $\mathcal{F}(\mathcal{G}, X)$ to (an element represented by) a cyclically reduced closed word.

Let

$$w=g_0\cdot e_1^{\varepsilon(1)}\cdot g_1\cdot\ldots\cdot e_n^{\varepsilon(n)}\cdot g_n,$$

$$w'=h_0\cdot f_1^{\eta(1)}\cdot h_1\cdot\ldots\cdot f_m^{\eta(m)}\cdot h_m$$

be reduced words, such that $g_n, h_0\in\mathcal{G}(v)$ for the same vertex $v\in V$. Then one can form the product word

$$w\cdot w'=g_0\cdot e_1^{\varepsilon(1)}\cdot\ldots\cdot e_n^{\varepsilon(n)}\cdot(g_n h_0)\cdot f_1^{\eta(1)}\cdot\ldots\cdot f_m^{\eta(m)}\cdot h_m.$$

This is reduced if and only if $e_n^{\varepsilon(n)}\cdot(g_n h_0)\cdot f_1^{\eta(1)}$ is reduced.

Let w be a cyclically reduced closed word. If w_1, w_2 are reduced words such that $w_1 \cdot w_2$ is defined, and is a cyclically reduced conjugate of w, then we will call w_1 and w_2 *subwords* of w. If both w_1 and w_2 have positive length, then they are *proper subwords* of w.

A 2-*complex* is a CW-complex K of dimension at most 2. Its 1-skeleton $K^{(1)}$ may be thought of as a graph, and the attaching maps for 2-cells may be thought of as closed paths in $K^{(1)}$, which we will always assume to be cyclically reduced.

An *elementary reduction* consists of a pair (L, K) of 2-complexes, with $K \subseteq L$, such that $L \setminus K$ has no 0-cells, exactly one 1-cell, and at most one 2-cell; and such that the attaching map for the 2-cell (if there is one) *strictly involves* the 1-cell, in the sense that it is not homotopic in $K \cup L^{(1)}$ to a map into K.

We will be concerned with the case when L is connected, so that K has either 1 or 2 components. We associate a graph of groups $(\mathscr{G}, X)$ to the elementary reduction (L, K) as follows. The graph X is obtained from $K \cup L^{(1)}$ by shrinking each component of K to a point. Thus X has a single edge e and one vertex v for each component K_v of K. Define $\mathscr{G}(e) = 1$, and $\mathscr{G}(v) = \pi_1(K_v)$.

Clearly $\pi(\mathscr{G}, X) \cong \pi_1(K \cup L^{(1)})$. In particular, if there is no 2-cell in $L \setminus K$, then $\pi(\mathscr{G}, X) \cong \pi_1(L)$. If there is a 2-cell in $L \setminus K$, its attaching map "spells out" a closed word w in $\mathscr{F}(\mathscr{G}, X)$, and $\pi_1(L) \cong \pi(\mathscr{G}, X)/N$, where N is the normal closure of the element represented by w.

Note that, if K is connected, then w is not well-defined in general, but depends on a choice of maximal tree in $K^{(1)}$. In general also w is not cyclically reduced (although this is independent of a choice of tree), but is nevertheless conjugate to a cyclically reduced word of positive length. Replacing w by a conjugate does not alter the (simple) homotopy type of the pair (L, K), so there is no loss of generality in considering only the elementary reductions (L, K) for which the corresponding word w is cyclically reduced.

An important special case of an elementary reduction is an elementary collapse $L \searrow K$. An elementary reduction (L, K) is an elementary collapse precisely when (L, K) has a 2-cell, K is connected and w has length 1.

Recall [10] that a *combinatorial map* $f: L' \to L$ between CW-complexes is a cellular map which maps each cell of L' homeomorphically onto a cell of L (necessarily of the same dimension). Suppose (L', K'), (L, K) are elementary reductions, and $f: L' \to L$ is a combinatorial map. We will say f *carries* (L', K') *to* (L, K) if f maps each cell of $L' \setminus K'$ to a cell of $L \setminus K$. In particular if $L' \setminus K'$ has a 2-cell then so has $L \setminus K$. Note that f is *not* required to map K' into K.

By a *tower* (of height h) we mean a map between connected 2-complexes which can be expressed as a composite $i_0 \circ p_1 \circ i_1 \circ \ldots \circ p_h \circ i_h$, where each p_r is an infinite cyclic covering, and each i_r is the inclusion of a finite, connected subcomplex. (This is an $(\mathfrak{F}, \mathfrak{C}_\infty)$-tower in the notation of [12], where more general towers are considered.) Clearly towers are combinatorial maps.

Lemma 2.1. *Let (L, K) be an elementary reduction, and $g: L' \to L$ a tower such that $g(L') \not\subseteq K$. Then there exists an elementary reduction (L', K') which is carried to (L, K) by g.*

Proof. Since $g(L') \not\subseteq K$, it is clear that the 1-cell of $L \backslash K$ lies in $g(L')$, and it follows that $(g(L'), K \cap g(L'))$ is an elementary reduction. Hence there is no loss of generality in assuming g to be surjective. Also, by an inductive argument on the height of g, it is sufficient to consider the height 1 case. Thus we may take g to be $p \circ i$, where $p: \tilde{L} \to L$ is an infinite cyclic covering, and $i: L' \hookrightarrow \tilde{L}$ is an inclusion.

Let c denote the generator of the group of covering transformations of p. Since L' is finite and $p \circ i$ is surjective, we may choose a 1-cell y of L' such that $p(y) \notin K$ and $c^t(y) \notin L'$ for any $t \geqq 1$. If y is involved in the attaching map for at most 1 2-cell (α, say) of L', we may define K' to be $L' \backslash \{y\}$ or $L' \backslash \{y, \alpha\}$, and the result holds.

Suppose then that y is involved in the attaching maps for 2 distinct 2-cells α_1 and α_2 of L'. Then $p(y)$, the 1-cell of $L \backslash K$ is involved in the attaching maps for $p(\alpha_1)$ and $p(\alpha_2)$, which must therefore be 2-cells in $L \backslash K$, and so $p(\alpha_1) = p(\alpha_2)$. Thus there is a covering transformation c^s $(s \neq 0)$ such that $c^s(\alpha_1) = \alpha_2$. Since y is involved in the attaching map for α_1, it follows that $c^s(y)$ is involved in that for α_2, so $c^s(y) \in L'$. Similarly $c^{-s}(y) \in L'$, which contradicts the choice of y.

Define a 2-complex L to be *reducible* if for every finite subcomplex L' either $L' \subset L^{(0)}$ or there exists an elementary reduction (L', K'). Recall that to any group presentation

$$\mathscr{P}: \langle x_1, x_2, \ldots | r_1, r_2, \ldots \rangle$$

is associated a connected 2-complex $K(\mathscr{P})$, with a single 0-cell, a 1-cell for each generator, and a 2-cell for each defining relator - attached along a path which "spells" the relator. Define a presentation $\mathscr{P}$ to be *reducible* if $K(\mathscr{P})$ is reducible. Equivalently, $\mathscr{P}$ is reducible if every finite subpresentation with at least one defining relator has the form

$$\langle x_1, \ldots, x_m, y | r_1, \ldots, r_n, s \rangle,$$

such that the r_i are words in the x_j, and s is not conjugate in $G * \langle y \rangle$ to an element of $G = \langle x_1, \ldots, x_m | r_1, \ldots, r_n \rangle$.

Suppose $\mathscr{P}'$ is the presentation obtained by omitting the defining relator r from the presentation $\mathscr{P}$. Say that r is a *proper power* of $\mathscr{P}$ if it is a proper power in the group G presented by $\mathscr{P}'$. That is there exist a word q in the generators and an integer $m \geqq 2$ such that r and q^m represent the same element of G. We will say that a presentation has *no proper powers* if none of its defining relators is a proper power.

Clearly all 1-relator presentations, and indeed all staggered presentations, are reducible. In fact reducible presentations may be thought of as being "staggered on one side", and so are a natural generalisation of staggered presentations. It turns out that reducible presentations with no proper powers share some of the "nice" properties of staggered presentations, while those with proper powers do not, in general.

A topological space X is *aspherical* if $\pi_n(X)=0$ for each $n\geqq 2$. It is well known that a 2-complex L is aspherical if and only if $\pi_2(L)=0$. A presentation $\mathscr{P}$ is *aspherical* if the corresponding 2-complex $K(\mathscr{P})$ is aspherical. Now any presentation which is aspherical in this sense has no proper powers. There are, however, various slightly different asphericity conditions, for presentations, which allow proper powers (see [4] for a full discussion). Staggered presentations (even those with proper powers) are aspherical under the strongest of these conditions, that of Cohen-Lyndon asphericity. We will show in § 5 that reducible presentations with no proper powers are aspherical in the topological sense, but those with proper powers need not be aspherical in any sense.

Indeed, any group G with a presentation satisfying any of the asphericity conditions of [4] must have quasi-projective dimension (qpd) at most 2 in the sense of [13]. That is, there exists an exact sequence

$$0\to Q\oplus P_2\to P_1\to P_0\to \mathbb{Z}\to 0$$

of $\mathbb{Z}G$-modules, with Q a permutation module, and each P_i projective. Thus the invariant qpd may be used to show that certain presentations are not aspherical in any sense. We give examples of groups of qpd more than 2, with reducible presentations.

3. Preliminary Results

The simplest case of Theorem 4.2 is when r has length 2, that is when G is the free product of A and B with an infinite cyclic subgroup amalgamated. This case is a result of Karrass and Solitar [14]. In particular, it follows that the class of locally indicable groups is closed under the process of "adjoining an n^{th} root" to a non-trivial element. We will make use of this fact in the proof of Proposition 3.3 below, so it is convenient to state the special case of Theorem 4.2 separately.

Proposition 3.1 ([14], Theorem 9). *Let G be an amalgamated free product $G=A*_C B$, where A and B are locally indicable, and C is infinite cyclic. Then G is locally indicable.*

Corollary 3.2. *Let a be a non-trivial element of the locally indicable group A, and let n be a positive integer. Then the group $G=\langle A, t\,|\,t^n=a\rangle$ is locally indicable.*

Proof. Put $B=\langle t\rangle$ and $C=\langle t^n\rangle=\langle a\rangle$ in Proposition 3.1.

The following technical result generalises a result of Weinbaum [24], that any proper subword of the defining relator represents a non-trivial element in a one-relator group.

Proposition 3.3. *Let $(\mathscr{G}, X)$ be a graph of groups with trivial edge groups and locally indicable vertex groups. Let $w_1, \ldots, w_n$ $(n\geqq 2)$ be reduced closed words in $(\mathscr{G}, X)$, not all of length zero, such that the product $w=w_1\ldots w_n$ is defined and is cyclically reduced. Let N denote the normal closure of w in $\pi(\mathscr{G}, X)$. If $w_1N=\ldots=w_nN$, then either*

(i) $w_1 = \ldots = w_n \notin N$; *or*
(ii) *all but one of the* w_i *are the empty word.*

Proof. The proof is by induction on the length λ of w. If $\lambda = 1$, then w has the form $ge^{\pm 1}$ or $e^{\pm 1}g$, where $g \in \mathscr{G}(v)$, $v = \iota(e) = \tau(e)$. Let Y be the subgraph $X \setminus \{e\}$ of X, and denote the restriction of $(\mathscr{G}, X)$ to Y by $(\mathscr{G}, Y)$. Then the composite $\pi(\mathscr{G}, Y) \to \pi(\mathscr{G}, X) \to \pi(\mathscr{G}, X)/N$ is an isomorphism. Now $\pi(\mathscr{G}, Y)$ is a free product of the vertex groups $\mathscr{G}(v)$ $(v \in V(x))$ and the free group $\pi_1(X)$, all of which are locally indicable. It is an immediate consequence of the Kuroš subgroup theorem that a free product of locally indicable groups is locally indicable. Thus $\pi(\mathscr{G}, X)/N$ is locally indicable, so torsion-free. Since $(w_i N)^n = wN = N$, it follows that $w_i \in N$ for all i. But all but one of the w_i are words in $(\mathscr{G}, Y)$, and are thus equal to the empty word. Hence (ii) holds.

Suppose then that $\lambda > 1$ and that the result holds for words of length less than λ. Next suppose that w involves two distinct edges. Let f be one of those edges, let X_0 be the smallest subgraph containing f, and let X' be the graph obtained from X by shrinking X_0 to a vertex 0. Define a graph of groups $(\mathscr{G}', X')$ by $\mathscr{G}'(0) = \pi(\mathscr{G}, X_0)$, $\mathscr{G}'(v) = \mathscr{G}(v)$ for $v \in V(X) \setminus V(X_0)$ and $\mathscr{G}'(e) = 1$ for $e \in E(X')$. Then $\pi(\mathscr{G}', X') \cong \pi(\mathscr{G}, x)$, and the words w_i determine reduced closed words w_i' in $(\mathscr{G}', X')$, not all of length 0. Furthermore the product $w' = w_1' \cdot \ldots \cdot w_n'$ is defined, and is cyclically reduced of length $\lambda' < \lambda$. The result holds for w' by inductive hypothesis, and thus holds also for w.

Hence we may assume that precisely one edge, e say, occurs in w. Let X_1 denote the smallest subgraph containing e, let $(\mathscr{G}, X_1)$ denote the restriction of $(\mathscr{G}, X)$ to X_1, and let N_1 be the normal closure of w in $\pi(\mathscr{G}, X_1)$. Then $\pi(\mathscr{G}, X_1)/N_1$ is a free factor of $\pi(\mathscr{G}, X)/N$, so it is sufficient to prove the proposition for w as a word in $(\mathscr{G}, X_1)$. We may therefore assume that $X = X_1$, that is X consists only of the edge e and its initial and terminal vertices (which may coincide). By cyclically permuting the w_i, conjugating all the w_i by some fixed element of some vertex group, and possibly changing the orientation of e, we may assume that w_1 (and hence also w) begins $1 \cdot e \cdot \ldots$. If e has non-zero exponent sum in w, then X has a single vertex, v say, and w has a subword of the form *e.g.* e or $e^{-1} \cdot g^{-1} \cdot e^{-1}$ for some $g \in \mathscr{G}(v)$. There is an automorphism θ of $\pi(\mathscr{G}, X)$ defined by $\theta = \mathrm{id}$ on $\mathscr{G}(v)$ and $\theta(e) = eg^{-1}$. Then $\theta(w) = \theta(w_1) \cdot \ldots \cdot \theta(w_n)$ is a cyclically reduced word of length λ, beginning $1 \cdot e \cdot \ldots$, and containing a subword of the form $e \cdot 1 \cdot e$ or $e^{-1} \cdot 1 \cdot e^{-1}$. Clearly the result for w will follow from that for $\theta(w)$, so we may also assume that either e has exponent sum zero in w, or w has a subword of the form $e \cdot 1 \cdot e$ or $e^{-1} \cdot 1 \cdot e^{-1}$. Finally, we may assume that the vertex groups of $(\mathscr{G}, X)$ are generated by those elements which occur in some w_i.

Next suppose that some vertex group of $(\mathscr{G}, X)$, say $\mathscr{G}(v)$, vanishes. Then v is the only vertex of X (otherwise no closed reduced word in $(\mathscr{G}, X)$ has positive length). Thus $\pi(\mathscr{G}, X)$ is infinite cyclic, generated by e. Since the w_i and w are all reduced, we have $w_i = e^{\lambda(i)}$ and $w = e^{\lambda}$, where $0 \leqq \lambda(i)$ for all i, $1 \leqq \lambda(1)$, and $\lambda(1) + \ldots + \lambda(n) = \lambda$. Also $\pi(\mathscr{G}, X)/N \cong \mathbb{Z}/\lambda\mathbb{Z}$, so $\lambda(1) \equiv \ldots \equiv \lambda(n) \pmod{\lambda}$. Hence either $\lambda(1) = \ldots = \lambda(n) = \lambda/n \not\equiv 0 \pmod{\lambda}$, or $\lambda(1) = \lambda$ and $\lambda(2) = \ldots = \lambda(n) = 0$, corresponding to the possibilities (i) and (ii) in the statement.

We may therefore assume that no vertex group vanishes. Since the vertex groups are finitely generated and locally indicable, it follows that $H^1(\pi(\mathscr{G}, X))$ has rank at least 2, so we may choose an epimorphism $\phi: \pi(\mathscr{G}, X) \twoheadrightarrow \mathbb{Z}$ with $\phi(w)=0$. In particular, if $V(X)=\{v\}$ and e has exponent sum 0 in w, then ϕ may be defined by $\phi(\mathscr{G}(v))=0$ and $\phi(e)=1$. We make the convention that this choice of ϕ is used whenever possible. Clearly $\phi(w_i)=0$ for all i, since $\phi(w_i)^n = \phi(w)=0$.

For each vertex v of X, the image of $\mathscr{G}(v)$ under ϕ is a subgroup of $\mathbb{Z}$, so has the form $k\mathbb{Z}$ for some $k \geqq 0$. If $k \geqq 2$, choose $g \in \mathscr{G}(v)$ such that $\phi(g)=k$ and let $\bar{\mathscr{G}}(v)$ denote the group obtained by adjoining a k^{th} root to g. Then $\bar{\mathscr{G}}(v)$ is locally indicable, by Corollary 3.2, and the map ϕ on $\mathscr{G}(v)$ extends uniquely to an epimorphism $\phi: \bar{\mathscr{G}}(v) \twoheadrightarrow \mathbb{Z}$. Replacing each $\mathscr{G}(v)$ by $\bar{\mathscr{G}}(v)$ whenever $k \geqq 2$ gives a new graph of groups $(\bar{\mathscr{G}}, X)$, where $\mathscr{G}(v)=\mathscr{G}(v)$ if $k=0$ or $k=1$. Let $\bar{N}$ denote the normal closure of w in $\pi(\bar{\mathscr{G}}, X)$. It follows from the Freiheitsatz [12] that

$$\pi(\bar{\mathscr{G}}, X)/\bar{N} \cong \bar{\mathscr{G}}(u) *_{\mathscr{G}(u)} (\pi(\mathscr{G}, X)/N) *_{\mathscr{G}(v)} \bar{\mathscr{G}}(v) \quad \text{if } V(X)=\{u, v\},$$

or

$$\pi(\bar{\mathscr{G}}, X)/\bar{N} \cong \bar{\mathscr{G}}(v) *_{\mathscr{G}(v)} \pi(\mathscr{G}, X)/N \quad \text{if } V(x)=\{v\}.$$

Hence it is sufficient to prove the result for w as a word in $(\bar{\mathscr{G}}, X)$. Note that the vertex groups of $(\bar{\mathscr{G}}, X)$ are not in general generated by those elements which appear in the w_i. However, unless $\phi(\bar{\mathscr{G}}(v))=0$ for all $v \in V(X)$, some w_i contains an element g of some vertex group such that $\phi(g) \neq 0$. That is, either w_i begins or ends with such a g, or has a subword $e^{\mp 1} \cdot g \cdot e^{\mp 1}$. It follows that w also contains such a g. For otherwise $0=\phi(w)=\sigma \cdot \phi(e)$, where σ is the exponent sum of e in w. By our choice of ϕ, we must have $\sigma \neq 0$, so $\phi(e)=0$. Since the only elements with non-zero ϕ-image which can occur in w_j are the first and last, we have $w_j = g_j \cdot e^{\mp 1} \cdot \ldots \cdot e^{\mp 1} \cdot h_j$ with $\phi(g_j) = -\phi(h_j)$. In particular $\phi(g_i) \neq 0$. Since $h_{j-1} g_j$ is a vertex group element appearing in w, we also have $\phi(h_{j-1} g_j)=0$, and so $\phi(g_j) \neq 0$ for all j, contradicting the assumption that w_1 begins $1 \cdot e \ldots$. Thus either $\phi(\bar{\mathscr{G}}(v))=0$ for all v, or w ends $\ldots \cdot e^{\mp 1} \cdot g$ or has a subword $e^{\mp 1} \cdot g \cdot e^{\mp 1}$, with $\phi(g) \neq 0$.

Now let H denote the kernel of $\phi: \pi(\bar{\mathscr{G}}, X) \to \mathbb{Z}$. By Bass-Serre Theory [5, 20], $G=\pi(\bar{\mathscr{G}}, X)$ acts on a tree T, regularly on the edges of T, with quotient X. Also $H=\pi(\mathscr{H}, \tilde{X})$, where $\tilde{X}$ is the quotient of T by the action of H, and $(\mathscr{H}, \tilde{X})$ is a graph of groups with trivial edge groups, and vertex groups which are isomorphic to subgroups of those of $(\bar{\mathscr{G}}, X)$. The regular action of G on $E(T)$ induces a regular action of $\mathbb{Z} \cong G/H$ on $E(\tilde{X})$, which may be used to label the edges of $\tilde{X}$: $\ldots, e_{-1}, e_0, e_1, \ldots$. There are 4 possibilities for the isomorphism type of $(\mathscr{H}, \tilde{X})$ depending on X and ϕ:

A) $V(X)=\{u, v\}$, $\phi(\bar{\mathscr{G}}(u))=\mathbb{Z}=\phi(\bar{\mathscr{G}}(v))$. Then $V(\tilde{X})=\{u, v\}$, $\mathscr{H}(u)=H \cap \bar{\mathscr{G}}(u)$, $\mathscr{H}(v)=H \cap \bar{\mathscr{G}}(v)$, and each e_i connects u to v.

B) $V(X)=\{u, v\}$, $\phi(\bar{\mathscr{G}}(u))=\mathbb{Z}$, $\phi(\bar{\mathscr{G}}(v))=0$. Then $V(\tilde{X})=\{u; \ldots, v_{-1}, v_0, v_1, \ldots\}$, $\mathscr{H}(u)=H \cap \bar{\mathscr{G}}(u)$, $\mathscr{H}(v_i) \cong \bar{\mathscr{G}}(v)$, and each e_i connects u to v_i.

C) $V(X)=\{v\}$, $\phi(\bar{\mathscr{G}}(v))=\mathbb{Z}$. Then $V(\tilde{X})=\{v\}$ and $\mathscr{H}(v)=H \cap \bar{\mathscr{G}}(v)$.

D) $V(X)=\{v\}$, $\phi(\bar{\mathscr{G}}(v))=0$. Then $V(\tilde{X})=\{\ldots, v_{-1}, v_0, v_1, \ldots\}$, $\mathscr{H}(v_i) \cong \bar{\mathscr{G}}(v)$, and each e_i connects v_i to v_{i+1}.

The elements of G represented by the w_i all belong to H, and so can be represented by reduced closed words $\tilde{w}_i$ in $(\mathscr{H}, \tilde{X})$, such that $\tilde{w}_1$ begins $1 \cdot \tilde{e}_0 \cdot \ldots$, and such that the product $\tilde{w} = \tilde{w}_1 \cdot \ldots \cdot \tilde{w}_n$ is defined. Since w is cyclically reduced, so is $\tilde{w}$. Also $\tilde{w}$ has length λ.

Let s denote the least index such that e_s occurs in $\tilde{w}$, and let t denote the greatest such index. Let X_0 denote the smallest subgraph of $\tilde{X}$ containing the edges e_i $(s \leqq i \leqq t)$. Let X_1 be obtained from X_0 by omitting e_t (together with v_t in case B or v_{t+1} in case D), and let X_2 be obtained from X_0 by omitting e_s (together with v_s in cases B and D). Define $G_i = \pi(\mathscr{H}, X_i)$ $(i = 0, 1, 2)$, where $(\mathscr{H}, X_i)$ is the restriction of $(\mathscr{H}, \tilde{X})$, and let N_0 be the normal closure of $\tilde{w}$ in G_0. It is clear that X_1 and X_2 are connected, except possibly if $s = t$ in case A, when $X_1 = X_2$ consists of 2 vertices and no edges. However, we will see below that this possibility cannot arise: $\tilde{w}$ involves more than one edge in case A.

Now $\bar{G} = \pi(\bar{\mathscr{G}}, X)$ can be expressed as an HNN-extension of G_0, with associated subgroups G_1 and G_2, via the canonical embedding $\pi(\mathscr{H}, X_0) \to \pi(\mathscr{H}, \tilde{X}) \to \pi(\bar{\mathscr{G}}, X)$. In cases A and B, the stable letter may be taken to be an element p of $\bar{\mathscr{G}}(u)$ such that $\phi(p)$ is a generator of $\mathbb{Z}$. In case C it may be taken to be a similar element of $\bar{\mathscr{G}}(v)$, and in case D it may be taken to be the element e.

Also, the maps $G_i \to G_0 \to G_0/N_0$ $(i = 1, 2)$ are injective, by the Freiheitsatz [12], and hence $\bar{G}/\bar{N}$ can be expressed as an HNN-extension of G_0/N_0, also with associated subgroups G_1 and G_2. It follows that it is enough to prove the result for $\tilde{w}$, regarded as a word in $(\mathscr{H}, X_0)$.

If $s < t$, that is if $\tilde{w}$ involves more than one edge, then the result holds for $\tilde{w}$, and so for w. In particular this is true in cases A and B, for then w has a (cyclic) subword of the form $e \cdot g \cdot e^{-1}$ or $e^{-1} \cdot g \cdot e$ for some vertex group element g with $\phi(g) = k \neq 0$, and the corresponding cyclic subword of $\tilde{w}$ has the form $e_i \cdot h \cdot e_{i+k}^{-1}$ or $e_i^{-1} \cdot h \cdot e_{i+k}$ for some integer i and some element h of some vertex group of $(\mathscr{H}, X_0)$.

We also have $s < t$ in case C, for then $\phi(\bar{\mathscr{G}}(v)) \neq 0$, so e has non-zero exponent sum in w, by choice of ϕ. Then w has subwords

(a) $e \cdot 1 \cdot e$ or $e^{-1} \cdot 1 \cdot e^{-1}$.

(b) $e^{\pm 1} \cdot g \cdot e^{\pm 1}$ with $\phi(g) = k \neq 0$.

If $\phi(e) = l \neq 0$, then corresponding to (a) is a subword of $\tilde{w}$ of the form $e_i \cdot 1 \cdot e_{i+l}$ or $e_i^{-1} \cdot 1 \cdot e_{i-l}^{-1}$ for some i. If $\phi(e) = 0$, then corresponding to (b) is a subword of $\tilde{w}$ of the form $e_i^{\pm 1} \cdot h \cdot e_{i+k}^{\pm 1}$ for some i and some h.

Thus the only case in which $\tilde{w}$ can involve only one edge is case D, in which case X_0 is a graph with 1 edge and 2 vertices. Repeating the whole argument with w replaced by $\tilde{w}$ and $(\mathscr{G}, X)$ replaced by $(\mathscr{H}, X_0)$, the result holds for $\tilde{w}$ because we arrive in case A or B.

This completes the proof of Proposition 3.3.

Corollary 3.4. *Let $(\mathscr{G}, X)$ be a graph of groups as in Proposition* 3.3. *Let N be the normal closure in $\pi(\mathscr{G}, X)$ of a cyclically reduced closed word w in $(\mathscr{G}, X)$. Then no proper closed subword of w represents an element of N.*

Proof. Suppose $\bar{w} = w_1 \cdot w_2$ is a cyclically reduced conjugate of w, and w_1 represents an element of N. Then $w_2 N = (w_1 N)^{-1}(\bar{w} N) = N$, so $w_2 N = w_1 N$. By

Proposition 3.3, one of w_1, w_2 is the empty word, so w_1 is not a proper closed subword of w.

Corollary 3.5. *Let $(\mathcal{G}, X)$ be a graph of groups as in Proposition 3.3, and let N be the normal closure in $\pi(\mathcal{G}, X)$ of $w=u^m$, where u is a cyclically reduced closed word of positive length. Then uN is an element of order precisely m in $\pi(\mathcal{G}, X)/N$.*

Proof. Clearly $(uN)^m = wN = N$, so uN has order at most m. But for $k<m$, u^k is a proper closed subword of w, so $u^k \notin N$ by Corollary 3.4.

Corollary 3.6. *Let (L, K) be an elementary reduction, such that each component of K has locally indicable fundamental group. Suppose that $L\backslash K$ has a 2-cell, attached along a cyclically reduced closed path P in $L^{(1)}$, and let w denote the closed word defined by P in the graph of groups corresponding to (L, K) (see § 2). Suppose also that w is cyclically reduced and not a proper power. Let (L', K') be an elementary reduction with a 2-cell in $L'\backslash K'$, and let $g: L' \to L$ be a tower carrying (L', K') to (L, K), such that $g_*(\pi_1(K'')) = 1$ in $\pi_1(L)$ for some component K'' of K'. Then $K' = K''$ and (L', K') is an elementary collapse.*

Proof. Let $\tilde{e}$ denote the 1-cell of $L'\backslash K'$, and let e denote both the 1-cell $g(\tilde{e})$ of $L\backslash K$ and the unique edge of the graph X. The orientation of e as an edge of X induces an orientation of $\tilde{e}$. Since g is a tower, the 2-cell of $L'\backslash K'$ is attached along a cyclically reduced closed path $\tilde{P}$ in $L'^{(1)}$ such that $g(\tilde{P}) = P$.

Suppose both $\tilde{e}$ and $\tilde{e}^{-1}$ occur in the path $\tilde{P}$. Then there exist adjacent occurrences of $\tilde{e}$ and $\tilde{e}^{-1}$ separated by a reduced closed path $\tilde{Q}$ in $K''^{(1)}$. In particular this holds if K' is disconnected. Thus $\tilde{P}$ has a subpath $\tilde{e} \cdot \tilde{Q} \cdot \tilde{e}^{-1}$ or $\tilde{e}^{-1} \cdot \tilde{Q} \cdot \tilde{e}$, so P has a subpath $e \cdot Q \cdot e^{-1}$ or $e^{-1} \cdot Q \cdot e$, where $Q = g(\tilde{Q})$. If e occurs in Q, then Q represents a proper subword w_1 of w, which lies in the normal closure of w in $\pi(\mathcal{G}, X) = \pi_1(K \cup L^{(1)})$, since Q is nullhomotopic in L. This contradicts Corollary 3.4. If e does not occur in Q, then Q is a path in $K^{(1)}$ which is nullhomotopic in L. By the Freiheitsatz [12], Q is nullhomotopic in K, so w has a subword $e \cdot 1 \cdot e^{-1}$ or $e^{-1} \cdot 1 \cdot e$. This contradicts the assumption that w is cyclically reduced.

Hence $\tilde{e}$ and $\tilde{e}^{-1}$ cannot both occur in $\tilde{P}$. Without loss of generality we may assume only $\tilde{e}$ occurs. In particular, $K' = K''$ is connected, and $\tilde{P} = \tilde{e} \cdot \tilde{Q}_1 \cdot \tilde{e} \cdot \ldots \cdot \tilde{e} \cdot \tilde{Q}_n$ up to cyclic permutation, where $\tilde{Q}_1, \ldots, \tilde{Q}_n$ are reduced paths in K''. Let $Q_i = g(\tilde{Q}_i)$ $(i = 1, \ldots, n)$ and let w_i be the reduced closed word in $(\mathcal{G}, X)$ spelt out by the path $e \cdot Q_i$. Since $Q_i^{-1} \cdot Q_j = g(\tilde{Q}_i^{-1} \cdot \tilde{Q}_j)$ is the image in L of a closed path in K'', it is nullhomotopic in L, so $w_1 N = \ldots = w_n N$, where N is the normal closure of w in $\pi(\mathcal{G}, X)$. If $n \geqq 2$, we can apply Proposition 3.3 to get $w_1 = \ldots = w_n$, contradicting the hypothesis that w is not a proper power. Hence $n = 1$, and (L', K') is an elementary collapse, as claimed.

4. One-Relator Products

In this section we prove the main result, Theorem 4.2. First it is necessary to prove the following Lemma. The proof uses arguments similar to those of [11].

Lemma 4.1. *Let L be a connected 2-complex such that $\pi_1(L)$ is not locally indicable. Then there exists a tower $g: L' \to L$ such that*

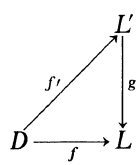
(a) $H^1(L')=0$; *and*

(b) $g_*(\pi_1(L')) \neq 1$ *in* $\pi_1(L)$.

Proof. Since $\pi_1(L)$ is not locally indicable, there is a finitely generated subgroup $G \neq 1$ of $\pi_1(L)$ such that G^{ab} is finite, of order m say. Since the inclusion-induced map $\eta: \pi_1(L^{(1)}) \to \pi_1(L)$ is surjective, there exists a finite subset $\{b_1, \ldots, b_n\}$ of $\pi_1(L^{(1)})$ such that $\{\eta(b_1), \ldots, \eta(b_n)\}$ generates G. Let H denote the subgroup of $\pi_1(L^{(1)})$ generated by $\{b_1, \ldots, b_n\}$.

For each i, we have $\eta(b_i^{-m}) \in [G, G] = \eta([H, H])$, so there exists $w_i \in [H, H]$ such that $\eta(b_i^{-m}) = \eta(w_i)$, that is $\eta(b_i^m w_i) = 1$. Now there exists a planar, 1-connected 2-complex D_i, and a combinatorial map $f_i: D_i \to L$, such that f_i maps the boundary path ∂D_i of D_i in the plane to a path in $L^{(1)}$ representing $b_i^m w_i$. (Indeed D_i may be taken as a wedge of complexes, each consisting of a disc with subdivided boundary, connected to the basepoint by a subdivided arc.) In particular ∂D_i is a composite $P_{i1} \cdot \ldots \cdot P_{im} \cdot Q_i$ of paths such that each $f_i(P_{ij})$ is a closed path representing b_i, and $f_i(Q_i)$ is a closed path representing w_i.

Let D denote the wedge $\bigvee_i D_i$ of the various D_i, and let $f = \bigvee_i f_i: D \to L$ be the induced map. Let



be a maximal tower-lifting of f [12, Lemma 3.1]. Now D is simply connected, so $H^1(D)=0$ and hence $H^1(L')=0$ (otherwise f' lifts over some infinite cyclic cover of L', contradicting maximality). Hence (a) holds.

To prove (b), suppose $\mathrm{Im}\, f \subseteq K \subseteq L$, and $\tilde{f}: D \to \tilde{L}$ is a lift of f over an infinite cyclic cover $p: \tilde{L} \to K$. Then $\tilde{f}(P_{ij})$ is a closed path in $\tilde{L}^{(1)}$, representing an element $\tilde{b}_i$ of $\pi_1(\tilde{L}^{(1)})$ (which depends only on i, not on j). Also each $\tilde{f}(Q_i)$ is a closed path representing some product of commutators of the $\tilde{b}_i$. Repeating this argument, we see that $f'(P_{ij})$ is a closed path in $L'^{(1)}$, representing an element b_i' (depending only on i). Clearly $g_*(b_i') = b_i$, so $1 \neq G \subseteq g_*(\pi_1(L')) \subseteq \pi_1(L)$.

Theorem 4.2. *Let A and B be locally indicable groups, and let G be the quotient of $A * B$ by the normal closure of a cyclically reduced word r of length at least 2. Then the following are equivalent:*

(i) *G is locally indicable;*

(ii) *G is torsion free;*

(iii) *r is not a proper power in $A * B$.*

Proof. That (i)⇒(ii) is immediate from the definition of locally indicable, and that (ii)⇒(iii) is Corollary 3.5, so it remains to prove (iii)⇒(i). Suppose then that r is not a proper power, and G is not locally indicable.

Choose presentations $\mathscr{P}_A, \mathscr{P}_B$ for A, B respectively, and let K be the disjoint union $K(\mathscr{P}_A) \cup K(\mathscr{P}_B)$. Attach a 1-cell e to K, connecting the two 0-cells of K. Then $\pi_1(K \cup \{e\}) \cong A * B \cong \pi(\mathscr{G}, X)$, where X is the tree with one edge and $(\mathscr{G}, X)$ has vertex groups A, B, and trivial edge group. The cyclically reduced word r of $A * B$ corresponds to a cyclically reduced closed word w in $(\mathscr{G}, X)$, and w can be represented by a cyclically reduced closed path P in $K^{(1)} \cup \{e\}$. Form a 2-complex L by attaching a 2-cell to $K \cup \{e\}$ along P. Then (L, K) is an elementary reduction, and $\pi_1(L) \cong G$.

Since G is not locally indicable, Lemma 4.1 applies, and there is a tower $g: L' \to L$ such that

(a) $H^1(L') = 0$; and

(b) $g_*(\pi_1(L')) \neq 1$ in $\pi_1(L)$.

Since L' is finite, there is no loss of generality in assuming it to be minimal with respect to (a) and (b), that is no proper connected subcomplex also satisfies these properties. Since each component of K has locally indicable fundamental group, it follows from (a) and (b) that $g(L') \not\subseteq K$. Hence, by Lemma 2.1, there is an elementary reduction (L', K') which is carried to (L, K) by g.

Suppose first that $L' \setminus K'$ has no 2-cell. Then by (a) it follows that K' is disconnected, say $K' = K'_1 \cup K'_2$, and so $H^1(K'_i) = 0$ for each $i = 1, 2$. By minimality of L', we must have $g_*(\pi_1(K_i)) = 1$ in $\pi_1(L)$, otherwise K'_i also satisfies (a) and (b). But $\pi_1(L') = \pi_1(K'_1) * \pi_1(K'_2)$, so $g_*(\pi_1(L')) = 1$, contradicting (b).

Now suppose $L' \setminus K'$ has a 2-cell. Then at least one component K'' of K' has $H^1(K'') = 0$. By minimality of L' we must have $g_*(\pi_1(K'')) = 1$ in $\pi_1(L)$. Hence by Corollary 3.6 (since w is not a proper power) it follows that $K' = K''$ is connected, and (L', K') is an elementary collapse. In particular, $\pi_1(K') \to \pi_1(L')$ is an isomorphism, so $g_*(\pi_1(L')) = g_*(\pi_1(K')) = 1$, again contradicting (b).

In either case, we have obtained a contradiction. Hence (iii)⇒(i), and the Theorem is proved.

Corollary 4.3 (Brodskiĭ [2]). *All torsion-free one-relator groups are locally indicable.*

Corollary 4.4 (Brodskiĭ [2]). *One-relator groups have no non-trivial finitely generated perfect subgroups.*

Proof. If G is a one-relator group, then there is an exact sequence

$$1 \to N \to G \to H \to 1,$$

where N is the normal closure in G of the root of the defining relator, and H is a torsion free one-relator group. Now N is a free product of finite cyclic groups [6], and so has no non-trivial finitely generated perfect subgroups. But neither has H, by Corollary 4.3, and hence neither has G.

Corollary 4.5. *Suppose the group G has a reducible presentation with no proper powers. Then G is locally indicable.*

Proof. Let $\mathscr{P}$ be a such a presentation. Since G is the direct limit of the groups presented by finite subpresentations of $\mathscr{P}$, it is sufficient to consider the case when $\mathscr{P}$ is finite. The proof is by induction on the number of defining relators

in $\mathscr{P}$. In the initial case there are no defining relators, so G is free. The inductive step is provided by Theorem 4.2 with B free.

The following consequence of Theorem 4.2 was pointed out to me by S.D. Brodskiĭ. Let us call a group G *equationally closed* if every nontrivial equation over G has a solution in G. That is, if F is a free group and N is the normal closure in $G*F$ of a word w not belonging to any conjugate of G, then the canonical map $G\to(G*F)/N$ is a split injection.

Corollary 4.6. *Every locally indicable group can be embedded in a locally indicable, equationally closed group.*

Proof. Let G be locally indicable, and let F be a fixed free group of countably infinite rank. Any element $w\in G*F$ not in any conjugate of G can be uniquely expressed $w=u^m$ where $m\geqq 1$ is an integer, and u is not a proper power. By Theorem 4.2 and the Freiheitsatz [12], G embeds in a group $G(u)$ in which the equation $u=1$ (and so also $w=1$) has a solution. If G' is the free product of all such $G(u)$, amalgamated over G, then G' is locally indicable, by an inductive argument using Theorem 4.2 and the Freiheitsatz. Also every nontrivial equation over G has a solution in G'. Define $G_0=G$, and inductively $G_i=G'_{i-1}$ for $i\geqq 1$. Then G embeds in the locally indicable, equationally closed group $G''=\bigcup_i G_i$.

5. Asphericity

We begin this section with a result similar to Lemma 4.1.

Lemma 5.1. *Let L be a connected, non-aspherical 2-complex. Then there exists a tower $g: L'\to L$ such that*

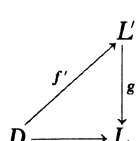
(a) $H^1(L')=0$; *and*

(b) $g_*(\pi_2(L'))\neq 0$ *in* $\pi_2(L)$.

Proof. Since L is non-aspherical, $\pi_2(L)\neq 0$.

Choose $\gamma\neq 0$ in $\pi_2(L)\subset\pi_2(L, L^{(1)})$. Then there exists a 1-connected planar 2-complex D, an element $\alpha\in\pi_2(D, D^{(1)})$, and a combinatorial map $f: D\to K$ such that $f_*(\alpha)=\gamma$. (As in Lemma 4.1, take D to be a wedge of complexes, each consisting of a disc with subdivided boundary, connected to the base point by a subdivided arc.)

Let



be a maximal tower lifting of f ([12], Lemma 3.1), and define $\beta=f'_*(\alpha)\in\pi_2(L', L'^{(1)})$.

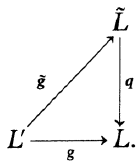
Let $\partial: \pi_2(L', L'^{(1)}) \to \pi_1(L'^{(1)})$ denote the boundary map. Then $g_*(\partial\beta) = 1$ in $\pi_1(L^{(1)})$. But $g_*: \pi_1(L'^{(1)}) \to \pi_1(L^{(1)})$ is injective, because g is a tower. Hence $\beta \in \operatorname{Ker} \partial = \pi_2(L')$, and so $0 \neq \gamma = g_*(\beta) \in g_*(\pi_2(L'))$ in $\pi_2(L)$, showing (b).

That (a) holds is a consequence of the maximality of f', since $H^1(D) = 0$.

Theorem 5.2. *Let L be a connected 2-complex with $H_2(L) = 0$ and $\pi_1(L)$ locally indicable. Then L is aspherical.*

Proof. Suppose L is not aspherical. Then by Lemma 5.1 there is a tower $g: L' \to L$ such that $H^1(L') = 0$ and $g_*(\pi_2(L')) \neq 0$ in $\pi_2(L)$.

Since $\pi_1(L)$ is locally indicable, $g_*(\pi_1(L')) = 1$ in $\pi_1(L)$, so g lifts over the universal cover $q: \tilde{L} \to L$,



Since $H_2(L) = 0$, it follows from [12], Lemma 3.2 that $H_2(L') = 0$.

From the commutative diagram

$$\begin{array}{ccc} \pi_2(L') & \xrightarrow{h} & H_2(L') = 0 \\ \downarrow{\scriptstyle \tilde{g}_*} & & \downarrow{\scriptstyle \tilde{g}_*} \\ \pi_2(\tilde{L}) & \xrightarrow[\cong]{h} & H_2(\tilde{L}) \end{array}$$

it follows that $\tilde{g}_*(\pi_2(L')) = 0$ in $\pi_2(\tilde{L})$, so $g_*(\pi_2(L')) = 0$ in $\pi_2(L)$, a contradiction.

Hence L is aspherical.

Theorem 5.3. *Suppose (L, K) is a simple reduction with a 2-cell whose attaching map does not represent a proper power in $\pi_1(K \cup L^{(1)})$. Suppose also that each component of K is aspherical with locally indicable fundamental group. Then L is aspherical.*

Proof. Let $(\mathscr{G}, X)$ denote the graph of groups corresponding to the reduction (L, K) (see § 2). Up to homotopy, we may assume the 2-cell of $L \setminus K$ is attached along a cyclically reduced path in $L^{(1)}$ which spells out a cyclically reduced closed word w in $(\mathscr{G}, X)$. Clearly w is not a proper power.

Suppose L is not aspherical. Then by Lemma 5.1 there is a tower $g: L' \to L$ such that $H^1(L') = 0$ and $g_*(\pi_2(L')) \neq 0$ in $\pi_2(L)$. Since L' is finite, we may assume without loss of generality that L' is minimal with respect to these properties, that is there is no proper connected subcomplex K' of L' such that $H^1(K') = 0$ and $g_*(\pi_2(K')) \neq 0$ in $\pi_2(L)$. From this assumption we derive a contradiction.

Since K is aspherical and $g_*(\pi_2(L')) \neq 0$, it follows that $g(L') \not\subseteq K$. Hence, by Lemma, by Lemma 2.1 there is an elementary reduction (L', K') which is carried to (L, K) by g.

First suppose that $L'\backslash K'$ has a 2-cell. Since $H^1(L')=0$, it follows that $H^1(K'')=0$ for some component K'' of K'. Since $\pi_1(L)$ is locally indicable, by Theorem 4.2, it follows that $g_*(\pi_1(K''))=1$ in $\pi_1(L)$. Hence, by Corollary 3.6, $K''=K'$ and (L', K') is an elementary collapse. Thus $H^1(K')=0$ and $g_*(\pi_2(K'))=g_*(\pi_2(L'))\neq 0$, contradicting the minimality of L'.

Now suppose that $L'\backslash K'$ has no 2-cell. Since $H^1(L')=0$, it follows that K' has 2 components, say K'_1 and K'_2, and that $H^1(K'_1)=0=H^1(K'_2)$. Also $\pi_2(L')$ is generated as a $\pi_1(L')$-module by the images of $\pi_2(K'_1)$ and $\pi_2(K'_2)$. Since $g_*(\pi_2(L'))\neq 0$, at least one of $g_*(\pi_2(K'_1))$, $g_*(\pi_2(K'_2))$ is non-zero, contradicting the minimality of L'.

In either case we have obtained a contradiction, so L is aspherical.

Corollary 5.4. *A reducible presentation with no proper powers is aspherical.*

Proof. It is enough to prove the result for finite presentations, and this is done by induction on the number of defining relators. The inductive step is provided by Theorem 5.3 and Corollary 4.5.

The following examples show that the condition on proper powers in Corollary 5.4 is essential for any form of asphericity to hold.

Example 1. $G=SL_2(\mathbb{Z})$ has a reducible presentation $\langle x, y | x^4, x^{-2}y^3\rangle$, but $\operatorname{qpd} G \doteq \infty$.

Example 2. $G=\mathbb{Z}\times\mathbb{Z}/m$ has a reducible presentation $\langle x, y | x^m, [x, y]\rangle$, but $\operatorname{qpd} G=\infty$.

Example 3. $G=S_3$, the symmetric group on 3 letters, has a reducible presentation $\langle x, y | x^2, xyxy^{-2}\rangle$, but $\operatorname{qpd} G=4$.

Example 4. $G=\mathbb{Z}$ has a presentation $\langle x | -\rangle$ which is aspherical in any sense, but also a reducible presentation $\mathscr{P}$: $\langle x, y, z | x^2, xyxy^{-2}, xzyz^{-1}\rangle$ which is not aspherical in any sense: $K(\mathscr{P})$ has the homotopy type of $S^1\vee S^2$.

Note also that Chiswell [4] and Sieradski [21] have constructed examples of aspherical presentations of the trivial group which are not *diagrammatically aspherical* (aspherical in the sense of Lyndon and Schupp [17]). Chiswell's example and one of Sieradski's are reducible presentations (with no proper powers).

6. 3-manifold Groups

The following shows that a large class of 3-manifolds have locally indicable fundamental groups.

Theorem 6.1. *Let M be a connected, orientable 3-manifold with $\pi_2(M)=0$. Let G be a non-trivial finitely generated subgroup of $\pi_1(M)$ with $H^1(G)=0$. Then M is a rational homology 3-sphere (in particular M is closed), and G has finite index in $\pi_1(M)$.*

Proof. Let $\bar{M}$ denote the covering of M corresponding to the subgroup G of $\pi_1(M)$. Since $G=\pi_1(\bar{M})$ is finitely generated, it follows [19] that $\bar{M}$ contains a

compact submanifold N such that the inclusion-induced map $\pi_1(N)\to\pi_1(\bar{M})$ is an isomorphism. Since $H^1(N)=H^1(G)=0$, it follows from Poincaré duality that $H^1(\partial N)=0$, so ∂N consists of a finite number $S_1,\ldots,S_n$ of 2-spheres.

Since $\pi_2(\bar{M})=\pi_2(M)=0$, each S_i separates $\bar{M}$ into 2 components, one of which is a (possibly fake) 3-cell D_i. Now $N\nsubseteq D_i$, for otherwise the isomorphism $\pi_1(N)\to\pi_1(\bar{M})=G$ would factor through $\pi_1(D_i)=1$, contradicting the hypothesis $G\neq 1$. Hence $\bar{M}=N\cup D_1\cup\ldots\cup D_n$.

In particular $\bar{M}$ is closed, so M is closed, and G has finite index in $\pi_1(M)$. Since $H^1(G)=0$, it follows that $H^1(M)=0$, so $H^1(M;\mathbb{Q})=0$, and M is a rational homology 3-sphere, as claimed.

Corollary 6.2. *Let M be a compact orientable 3-manifold. Then either $\pi_1(M)$ is locally indicable, or M has a connected summand which is a rational homology 3-sphere but not a homotopy 3-sphere.*

Proof. Let $M=M_1\#\ldots\#M_n$ be a prime factorisation of M [18], so that $\pi_1(M)=\pi_1(M_1)*\ldots*\pi_1(M_n)$. If $\pi_1(M)$ is not locally indicable, then some $\pi_1(M_i)$ is not locally indicable. In particular, $\pi_1(M_i)\ncong\mathbb{Z}$, so $\pi_2(M_i)=0$. Hence M_i is a rational homology 3-sphere, by the Theorem.

The following is an immediate consequence of Corollary 6.2.

Corollary 6.3. *Let ℓ be a tame link in S^3. Then $\pi_1(S^3\setminus\ell)$ is locally indicable.*

References

1. Baumslag, G.: Some problems on one-relator groups. Proceedings of the Second International Conference on the Theory of Groups (Canberra 1973), pp. 75–81. Lecture Notes in Mathematics **372**. Berlin-Heidelberg-New York: Springer 1974
2. Brodskii, S.D.: Equations over groups and groups with a single defining relator. Uspehi Mat. Nauk **35**, 4, 183 (1980) [Russian]
3. Brown, R., Huebschmann, J.: Identities among relations. In: Low Dimensional Topology vol. 1 (R. Brown and T.L. Thickstun, eds.). London Mathematical Society Lecture Note Series **48**, pp 153–202, Cambridge University Press 1982
4. Chiswell, I.M., Collins, D.J., Huebschmann, J.: Aspherical group presentations. Math. Z. **178**, 1–36 (1981)
5. Dicks, W.: Groups, trees and projective modules. Lecture Notes in Mathematics **790**. Berlin-Heidelberg-New York: Springer 1980
6. Fischer, J., Karrass, A., Solitar, D.: On one-relator groups having elements of finite order. Proc. Amer. Math. Soc. **33**, 297–301 (1972)
7. Gutierrez, M.A., Ratcliffe, J.G.: On the second homotopy group. Quart. J. Math. Oxford (2) **32**, 45–55 (1981)
8. Higgins, P.J.: The fundamental groupoid of a graph of groups. J. London Math. Soc. (2) **13**, 145–149 (1976)
9. Higman, G.: The units of groups rings. Proc. London Math. Soc. (2) **46**, 231–248 (1940)
10. Howie, J.: Aspherical and acyclic 2-complexes. J. London Math. Soc. (2) **20**, 549–558 (1979)
11. Howie, J.: On the fundamental group of an almost-acyclic 2-complex. Proc. Edinburgh Math. Soc. **24**, 119–122 (1981)
12. Howie, J.: On pairs of 2-complexes and systems of equations over groups. J. Reine Angew. Math. **324**, 165–174 (1981)
13. Howie, J., Schneebeli, H.R.: Groups of finite quasi-projective dimension. Comment. Math. Helv. **54**, 615–628 (1979)

14. Karrass, A., Solitar, D.: The subgroups of a free product of two groups with an amalgamated subgroup. Trans. Amer. Math. Soc. **149**, 227–255 (1970)
15. Kirby, R.: Problems in low-dimensional manifold theory. In: Algebraic and Geometric Topology (R.J. Milgram, ed.). Proceedings (Stanford 1976), pp. 273–312. Proceedings of Symposia in Pure Mathematics **32**, part 2. Providence, Rhode Island: American Mathematical Society 1978
16. Lewin, J., Lewin, T.: An embedding of the group algebra of a torsion-free 1-relator group in a field. J. Algebra **52**, 39–74 (1978)
17. Lyndon, R.C., Schupp, P.E.: Combinatorial Group Theory. Ergebnisse der Mathematik und ihrer Grenzgebiete **89**. Berlin-Heidelberg-New York: Springer 1977
18. Milnor, J.: A unique factorisation theorem for 3-manifolds. Amer. J. Math. **84**, 1–7 (1962)
19. Scott, G.P.: Compact submanifolds of 3-manifolds. J. London Math. Soc. (2) **7**, 246–250 (1973)
20. Serre, J.-P.: Arbres, amalgames, SL_2. Astérisque **46**. Paris: Société Mathematique de France 1977
21. Sieradski, A.J.: Framed links for Peiffer identities. Math. Z. **175**, 125–137 (1980)
22. Short, H.: Topological methods in group theory: the adjunction problem. Ph. D. Thesis, University of Warwick 1981
23. Short, H.: Knot groups are locally indicable. Preprint
24. Weinbaum, C.M.: On relators and diagrams for groups with a single defining relator. Illinois J. Math. **16**, 308–322 (1972)

Received September 4, 1981

Math. Z. 180, 463–468 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

The Banach Algebra of Conservative Triangular Matrices

James DeFranza[1] and Daniel J. Fleming[2]

[1] Mathematics Department, Youngstown State University, Youngstown, Ohio 44555, U.S.A.
[2] Mathematics Department, St. Lawrence University, Canton, New York 13617, U.S.A.

1. Introduction

If $A=(a_{nk})$ is an infinite matrix, A defines a sequence to sequence transformation, mapping the sequence x to Ax where

$$(Ax)_n=\sum_k a_{nk}x_k, \qquad n=1,2,\ldots.$$

Let $c_A=\{x|Ax\in c\}$ where c denotes the space of convergent sequences. If $c\subseteq c_A$, A is *conservative*. Well known necessary and sufficient conditions for A to be conservative are:

(i) $\|A\|\equiv\sup_n\sum_k|a_{nk}|<\infty$,

(ii) $t\equiv\lim_n\sum_k a_{nk}$ exists, and

(iii) $a_k\equiv\lim_n a_{nk}$ exists for $k=1,2,\ldots.$

If $\lim_n(Ax)_n=\lim_n x_n$ for all $x\in c$, then A is called *regular*. A conservative matrix is regular if and only if $t=1$ and $a_k=0$ for all k. For each conservative matrix A let

$$\chi(A)=t-\sum_k a_k.$$

The matrix A is *coregular* provided $\chi(A)\neq 0$ and *conull* if $\chi(A)=0$. A matrix A is of type M provided $b\in l_1\equiv\{x|\sum_k|x_k|<\infty\}$ and $bA=0$ implies $b=0$, where $(bA)_k=\sum_n b_n a_{nk}$.

Let $B(c)$ denote the Banach algebra of bounded linear operators on c, Γ the subalgebra of conservative matrices and Δ the subalgebra of triangular (i.e. $a_{nk}=0$ for $k>n$) conservative matrices. If A is triangular with $a_{nn}\neq 0$ for all n, A is called a *triangle*. With the norm given in (i) above Δ and Γ are closed Banach subalgebras of $B(c)$.

If X is any Banach algebra we denote the maximal group of invertible elements by $\mathscr{M}$, $\mathscr{B}$ the boundary of the maximal group, and $\mathscr{N}=X\setminus\overline{\mathscr{M}}$. Since

0025-5874/82/0180/0463/$01.20

$\mathcal{M}$ is open, $\mathcal{B}=\overline{\mathcal{M}}\setminus\mathcal{M}$ and $\mathcal{M}\cup\mathcal{B}\cup\mathcal{N}$ is a disjoint decomposition of X. An element $x\in X$ is called a left divisor of zero (l.d.z.) provided there exists $y\in X$, $y\neq 0$ such that $xy=0$. Right divisor of zero (r.d.z.) and two sided divisor of zero are defined similarily. An element $x\in X$ is called a *left topological divisor of zero* (l.t.d.z.) provided given $\varepsilon>0$ there exists a $y\in X$, $\|y\|=1$ such that $\|xy\|<\varepsilon$. Similar definitions are given for right topological divisor of zero (r.t.d.z.) and two sided topological divisor of zero.

2. Bi-diagonal Methods in Δ

In [6], some properties of $\mathcal{M}_\Delta$ and $\mathcal{B}_\Delta$ are given and the positioning of certain Nörlund and Hausdorff methods is presented. We now consider several questions raised in [6]. We begin by considering a class of triangles in Δ.

2.1. Definition. Let a, b be bounded sequences with $a_k, b_k\neq 0$ for all k and such that $\lim[b_{n-1}+a_n]$ exists. The triangle $A=(a_{nk})$ defined by $a_{nk}=a_k$ if $n=k$, $a_{nk}=b_k$ if $n-1=k$ and $a_{nk}=0$ otherwise is called a *bi-diagonal method.*

If A is the bi-diagonal method associated with the sequences a and b, then define

$$c_k=\frac{b_k}{a_k} \quad \text{and} \quad d_k=\prod_{j=1}^{k} c_j.$$

2.2. Theorem. *Let A be the bi-diagonal method associated with a and b. The matrix A is of type M if and only if* $\sum_k \frac{1}{|d_k|}=+\infty$.

Proof. ($\Rightarrow$) Assume $\sum_k \frac{1}{|d_k|}<\infty$. Let $t_1=1$, and for $n\geqq 2$

$$t_n=\frac{(-1)^{n-1}}{d_{n-1}}=(-1)^{n-1}\prod_{j=1}^{n-1}\left(\frac{a_j}{b_j}\right).$$

Now $tA=(t_n a_n+t_{n+1}b_n)_{n=1}^{\infty}$. Thus $t_1a_1+t_2b_1=a_1-(a_1/b_1)b_1=0$ and for $n\geqq 2$

$$\begin{aligned} t_n a_n+t_{n+1}b_n&=(-1)^{n-1}a_n\prod_{j=1}^{n-1}\left(\frac{a_j}{b_j}\right)+(-1)^n b_n\prod_{j=1}^{n}\left(\frac{a_j}{b_j}\right)\\ &=[(-1)^{n-1}+(-1)^n]\,a_n\prod_{j=1}^{n-1}\left(\frac{a_j}{b_j}\right)\\ &=0. \end{aligned}$$

Thus $t\in l_1$, $t\neq 0$ and $tA=0$ so A is not of type M.

($\Leftarrow$) Let $t\in l_1$, $t\neq 0$ with $tA=0$. Any such t must satisfy $t_k\neq 0$ for all k. Indeed, if $t_n=0$ for some n, then $(tA)_{n-1}=t_{n-1}a_{n-1}+t_nb_{n-1}=0$ and $(tA)_n=t_na_n+t_{n+1}b_n=0$. Since $a_j, b_j\neq 0$ for all j it follows that $t_{n-1}=0$ and $t_{n+1}=0$. Now $t_1a_1+t_2b_1=0\Rightarrow t_1=-t_2(b_1/a_1)=-t_2d_1$. Similarly $t_2a_2+t_3b_2=0\Rightarrow t_2=-t_3(b_2/a_2)$ so $t_1=t_3d_1(b_2/a_2)=t_3d_2$. Continuing we obtain

$$t_1=(-1)^n t_{n+1}d_n \quad \text{for all } n.$$

Thus

$$\sum_n (1/|d_n|) = (1/|t_1|) \sum_n |t_{n+1}| < \infty.$$

2.3. Lemma. *Let A be the bi-diagonal method associated with a and b. If*

$$L = \lim_n (|a_n| - |b_{n-1}|) \neq 0,$$

then A sums no bounded divergent sequence.

Proof. If $L>0$ the result follows from Theorem 2 of [1]. If $L<0$ let A' be the matrix obtained by deleting the first row of A. Then $c_A = c_{A'}$ and moreover

$$\liminf_n [|a'_{nn}| - \sum_{k \neq n} |a'_{nk}|] = -L > 0.$$

So again $c_A \cap m = c$ by Theorem 2 of [1].

2.4. Lemma. *Let A and B be triangles in Δ with $c_A = c_B$. Then*

(i) *$A \in \mathscr{B}_\Delta$ if and only if $B \in \mathscr{B}_\Delta$, and*

(ii) *$A \in \mathscr{N}_\Delta$ if and only if $B \in \mathscr{N}_\Delta$.*

Proof. This follows immediately from [6, Theorem 1] and the fact that a triangle is in $\mathscr{M}_\Delta$ if and only if it is equivalent to convergence.

The next result positions bi-diagonal methods associated with convergent sequences. Recall the Nörlund polynomial method $N = (a_{nk})$ associated with the polynomial $\sum_{k=0}^{n} p_k z^k$ is defined by $a_{nk} = p_{n-k}$ if $0 \leqq n-k \leqq n$ and 0 otherwise.

2.5. Theorem. *Let A be the bi-diagonal method associated with the convergent sequences a, b. Let $L_1 = \lim_n a_n$, $L_2 = \lim_n b_n$ and if $L_2 \neq 0$ let $L = |L_1/L_2|$.*

(i) *If $L_1 = L_2 = 0$ or if $L = 1$, then $A \in \mathscr{B}_\Delta$.*

(ii) *If $L < 1$, then $A \in \mathscr{N}_\Delta$,*

(iii) *If $L_1 \neq 0$, $L_2 = 0$ or if $L > 1$, then $A \in \mathscr{M}_\Delta$.*

Proof. (i) If $L_1 = L_2 = 0$, then A is coercive and hence in $\mathscr{B}_\Delta$. Suppose $L = 1$. Let a^j be the sequence with $a^j_i = a_i$ for $i \leqq j$ and $a^j_i = L_1$ for $i > j$ and b^j the sequence with $b^j_i = b_i$ for $i \leqq j-1$ and $b^j_i = L_2$ for $i \geqq j$. Let A^j be the bi-diagonal method associated with a^j and b^j. Consider the Nörlund method N defined by the polynomial $p(z) = L_1 + L_2 z$. By [6, Theorem 5(ii)] $N \in \mathscr{B}_\Delta$ and since $c_{A^j} = c_N$, $A^j \in \mathscr{B}_\Delta$ by 2.4.

Now

$$\|A^j - A\| = \sup_{n>j} \{|L_2 - b_n| + |L_1 - a_{n+1}|\}$$

tends to zero as $j \to +\infty$. Thus $A \in \overline{\mathscr{B}}_\Delta = \mathscr{B}_\Delta$.

(ii) Let A^j and N be as in part (i). Then by [6, Theorem 5, (iii)] and 2.4, N and hence A^j are in $\mathscr{N}_\Delta$. But by 2.3, $c_A \cap m = c$ and hence by [4, Theorem 11], $A \notin \mathscr{B}_\Delta$ and hence $A \in \mathscr{N}_\Delta$.

(iii) Arguing as above we conclude $A \in \overline{\mathscr{M}}_\Delta$. Since by 2.3 $c_A \cap m = c$ we must have $A \in \mathscr{M}_\Delta$ [4, Theorem 11].

In [6; 4(i), 3(ii)], Rhoades poses the following questions: Does there exist a triangle not of type M on $\mathscr{B}_\Delta$? If A is a coregular triangle not of type M is $A \in \mathscr{N}_\Delta$?

2.6. Example. Let $a \in c_0$ (the space of null sequences) and $b = ka$, $k > 1$. By 2.2 A is not of type M and by 2.5(i) $A \in \mathscr{B}_\Delta$. Note A is conull and answers question 4(i) of [6] in the affirmative.

Question 4(i) of [6] has also been answered (using conull methods) by Sharma in [7] and Hogan and Kelly in [5]. We now answer question 3(ii) of [6] in the negative by exhibiting a coregular triangle not of type M on $\mathscr{B}_\Delta$.

2.7. Example. Let $a_n = n^2/(n+1)^2$ and let $b_1 = 1/\ln(2)$, $b_n = \ln(n)/\ln(n+1)$ for $n \geqq 2$. Then $L = |\lim_n a_n / \lim_n b_n| = 1$, so by 2.5(i) the bi-diagonal method associated with $a = (a_n)$ and $b = (b_n)$ is in $\mathscr{B}_\Delta$. Moreover $\chi(A) = \lim_n [b_{n-1} + a_n] = 2$ so A is coregular.

Finally

$$1/(|d_n|) = \prod_{j=1}^{n} (a_j/b_j)$$

$$= \frac{\ln(n+1)}{(n+1)^2}$$

and hence $\sum_n (1/|d_n|) < \infty$. Therefore by 2.2 A is not of type M.

3. Topological Divisors of Zero in Δ

Still to be determined is: Does there exist a type M triangle in $\mathscr{N}_\Delta$? Must a type M coregular triangle be in $\overline{\mathscr{M}}_\Delta$? [6, 4(ii), 3(i)].

In [4, Theorem 11] Copping proved that a triangle in $\mathscr{B}_\Delta$ must sum a bounded divergent sequence and in turn exhibited a coregular triangle in $\mathscr{N}_\Delta$ with $c_A \cap m \neq c$, thus settling a conjecture of Wilansky and Zeller [10]. In light of Copping's result one method for positioning a triangle (not equivalent to convergence) in $\mathscr{N}_\Delta$ is to show $c_A \cap m = c$. However for type M triangles this latter condition is not possible.

3.1. Theorem. *If A is a type M triangle not in $\mathscr{M}_\Delta$, then $c_A \cap m \neq c$.*

Proof. Since any conull method satisfies $c_A \cap m \neq c$ we assume A is coregular. Now A is $1-1$ on c so by [9, Theorem 4] $\overline{A(c)} = c$. By [8] $c_A \cap m = c$ if and only if $A(c)$ is closed in c. Since $A(c)$ is dense in c this is equivalent to $A \in \mathscr{M}_\Delta$.

Another approach to positioning a method off the boundary is to construct one which fails to be a two sided topological divisor of zero. That this technique can not be used to obtain a type M coregular triangle in $\mathscr{N}_\Delta$ will follow from Theorem 3.5.

The key to Wilansky's [9] adaptation of Yood's [11] results to the algebra Γ is the fact that matrix methods which are divisors of zero or topological divisors of zero in $B(c)$ will, with appropriate hypothesis in the "right" case, be divisors of the same kind in Γ. The situation between Δ and Γ is identical with that between $B(c)$ and Γ in the "left" case.

3.2. Lemma. *Let $A \in \Delta$. If A is a l.d.z. (respectively l.t.d.z.) in Γ, then A is a l.d.z. (respectively l.t.d.z.) in Δ.*

Proof. If $A \in \Delta$ is a l.d.z. in Γ, then there exists $x \in c$, $x \neq 0$ such that $Ax = 0$. Thus $AB = 0$ where $B \in \Delta$ is the method with first column x and all other entries zero. Suppose now $A \in \Delta$ and A is a l.t.d.z. in Γ. Then clearly A is a l.t.d.z. in $B(c)$ so [9, Lemma 8] given any $\varepsilon > 0$ there exists $x \in c$, $\|x\| = 1$ such that $\|Ax\| < \varepsilon$. Let B be as above. Then $B \in \Delta$, $\|B\| = 1$ and

$$\begin{aligned} \|AB\| &= \sup_n |(Ax)_n| \\ &= \|Ax\| < \varepsilon. \end{aligned}$$

In [9, Corollary 2] Wilansky shows that if $A \in \Gamma$ is coregular, then A a r.d.z. in $B(c) \Rightarrow A$ is a r.d.z. in Γ. The analogous result for Δ versus Γ does not hold. Specifically let A be the bi-diagonal method with 1's on the main diagonal and 2's on the lower diagonal. Then A is a coregular r.d.z. in Γ (A is not of type M) but, as one easily checks, no bi-diagonal method is a r.d.z. in Δ. In fact the methods in Δ which are not right divisors of zero in Δ admit a characterization similar to the classical type M condition with l_1 replaced by Φ, the space of finitely non-zero sequences.

3.3. Lemma. *Let $A \in \Delta$. Then A is a r.d.z. in Δ iff and only if there exists $t \in \Phi$, $t \neq 0$ such that $tA = 0$.*

Proof. ($\Rightarrow$) Assume $tA = 0$, $t \in \Phi$ implies $t = 0$. Let $B \in \Delta$ with $BA = 0$. The n^{th} row of BA is $b^{(n)}A$ where $b^{(n)}$ is the n^{th} row of B. Since $b^{(n)} \in \Phi$ it follows that $b^{(n)} = 0$ for all n so $B = 0$ and hence A is not a r.d.z. in Δ.

($\Leftarrow$) Let $t \in \Phi$, $t \neq 0$ with $tA = 0$. Say $t = (t_1, \ldots, t_j, 0, 0, \ldots)$. Let B be the matrix with first $j-1$ rows zero and all other rows equal to t. Then $B \in \Delta$, $B \neq 0$ but $BA = 0$.

A t.d.z. which is not a divisor of zero of the same kind is called proper. In [9, Lemma 12] it is shown that if $A \in \Gamma$ is a proper r.t.d.z. in $B(c)$ then A is a proper r.t.d.z. in Γ. In going from Γ to Δ the assumption of proper is not necessary. In the following $\| \ \|_1$ denotes the norm in l_1.

3.4. Theorem. *Let $A \in \Delta$. Then A is a r.t.d.z. in Δ if and only if A is a r.t.d.z. in Γ.*

Proof. ($\Rightarrow$) obvious.

($\Leftarrow$) Let $\varepsilon > 0$ be given and choose $B \in \Gamma$ with $\|B\| > 1$ $\|BA\| < \varepsilon/2$. Let $b^{(n)}$ be the n^{th} row of B. Then $\|BA\| = \sup_n \|b^{(n)}A\|_1$ and $\|B\| = \sup_n \|b^{(n)}\|_1$. Choose $n_1 \ni \|b^{(n_1)}\|_1 > 1$ and $\|b^{(n_1)}A\|_1 < \varepsilon/2$. Let $t = b^{(n_1)}$ and let $t^n = (t_1, t_2, \ldots, t_n, 0, \ldots)$ be the n^{th} section of t. Now $t^n \to t$ in l_1 and A^T is continuous on l_1.

Thus choose n_2 such that $\|t^{n_2}\|_1 > 1$ and $\|A^T t^{n_2} - A^T t\|_1 < \varepsilon/2$. Then

$$\begin{aligned}\|t^{n_2}A\|_1 &= \|A^T t^{n_2}\|_1 \\ &= \|A^T t^{n_2} - A^T t + A^T t\|_1 \\ &\leqq \|A^T t^{n_2} - A^T t\|_1 + \|A^T t\|_1 \\ &< \varepsilon/2 + \varepsilon/2 = \varepsilon.\end{aligned}$$

Now $t^{n_2} \in \Phi$ so there exists j such that $t_i^{n_2} = 0$ for $i \geqq j$. Let C be the matrix with first $j-1$ rows zero and all others equal to t^{n_2}. Then $C \in \Delta$, $\|C\| > 1$ and $\|CA\| = \|t^{n_2}A\|_1 < \varepsilon$.

3.5. Theorem. *Let A be a type M triangle with $A \notin \mathcal{M}_\Delta$.*

(i) *Then A is a proper l.t.d.z. in Δ.*

(ii) *If, in addition, A is coregular, then A is a proper r.t.d.z. in Δ.*

Proof. (i) By 3.1 $c_A \cap m \neq c$ so by [2, Theorem 1] A is a l.t.d.z. in Γ and hence by 3.2 a l.t.d.z. in Δ. It is proper as A is a triangle.

(ii) By [9, Theorem 4] $\overline{A(c)} = c$. Since $A \notin \mathcal{M}_\Delta$ it cannot be onto as a map on c and so by [9, Theorem 5] A is a proper r.t.d.z. in Γ and so by 3.4 a proper r.t.d.z. in Δ.

3.6. Question. Does there exist a type M conull triangle which is not a r.t.d.z. in Δ?

3.7. Remark. From Lemma 3.2 it follows that the first two parts of Wilansky's adaptation of Yood's results to Γ [9, Theorem 5] carry over to Δ. The latter two do not since Example 2.7 is a coregular proper r.t.d.z. in Δ which is not range dense. [Note Theorem 5 of [9] has a typographical error. The implication ($\Rightarrow$) of part (iii) requires the hypothesis of coregularity.]

References

1. Agnew, R.P.: Equivalence of Methods for Evaluation of Sequences. Proc. Amer. Math. Soc. **3**, 550–556 (1952)
2. Berg, I.D.: A Banach Algebra Criterion for Tauberian Theorems. Proc. Amer. Math. Soc. **15**, 648–652 (1964)
3. Berg, I.D.: Open Sets of Conservative Matrices. Proc. Amer. Math. Soc. **16**, 719–724 (1965)
4. Copping, J.: Mercerian Theorems and Inverse Transformations. Studia Math. **21**, 177–194 (1962)
5. Hogan, D.A., Kelly, E.P., Jr.: Bounded, Conservative, Linear Operators, and the Maximal Group. Proc. Amer. Math. Soc. **32**, 195–200 (1972)
6. Rhoades, B.E.: Triangular Summability Methods and the Boundary of the Maximal Group. Math. Z. **105**, 284–290 (1968)
7. Sharma, N.K.: Spectra of Conservative Matrices. Proc. Amer. Math. Soc. **35**, 515–518 (1972)
8. Whitley, R.: Conull and Other Matrices which Sum a Bounded Divergent Sequence. Math. Monthly **74**, 798–801 (1967)
9. Wilansky, A.: Topological Divisors of Zero and Tauberian Theorems. Trans. Amer. Math. Soc. **113**, 240–251 (1964)
10. Wilansky, A., Zeller, K.: Banach Algebra and Summability. Illinois J. Math. **2**, 378–385 (1958)
11. Yood, B.: Transformations Between Banach Spaces in the Uniform Topology. Ann. of Math. (2) **50**, 486–503 (1949)

Received February 2, 1982

Math. Z. 180, 469–481 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On Convergence and Quasiregularity of Interpolating Complex Planar Splines

Gerhard Opfer[1] and Glenn Schober[2]

[1] Institut für Angewandte Mathematik der Universität, Bundesstraße 55, D-2000 Hamburg 13, Federal Republic of Germany
[2] Department of Mathematics, Indiana University, Bloomington, Indiana 47405, USA

1. Introduction

Complex planar splines and their elementary properties were first studied by Opfer-Puri [5]. Here we are concerned with the development of further properties of interpolating splines. Of particular interest are those splines which interpolate quasiregular or quasiconformal mappings. The question of *good* approximations by splines will be also discussed. We shall give more precise meanings later. Some of the results may be regarded as constructive extensions of results previously given by Jaenisch [2]. See [2] also for further references.

A *complex planar spline* is a continuous complex-valued function defined on a two-dimensional grid with finitely many meshes, where the spline is defined piecewise in the grid. The pieces will be called *elements*. As usual, we assume throughout that no vertex of a mesh is interior to an edge of another mesh.

In this investigation we shall admit only (arbitrary) triangular grids with elements defined by

$$p(z)=a+bz+c\bar{z}; \quad a, b, c\in\mathbb{C} \tag{1.1}$$

and rectangular grids whose meshes have sides parallel to the coordinate axes with elements defined by

$$p(z)=a+bz+c\bar{z}+d(z^2-\bar{z}^2); \quad a, b, c, d\in\mathbb{C}. \tag{1.2}$$

If in the sequel we refer to a rectangular grid, we always mean a grid whose sides are parallel to the coordinate axes. The explicit interpolation formulae are recalled in Sect. 2. More details, including extensions to parallelogram grids and grids on circular sectors can be found in Opfer-Puri [5].

Let $R\subset\mathbb{C}$ be a region under investigation. We assume throughout the paper that we fix a compact set $S\subset R$ (usually close to R) which can be partitioned into either a triangular or a rectangular grid. For such a grid we shall make the convention that its *meshsize* is the length of the longest edge of

0025-5874/82/0180/0469/$02.60

any mesh. If the partitioning of S results in a grid of meshsize h, we will designate that grid by S_h.

We adopt here the analytic definition of quasiconformality from Lehto-Virtanen [3, pp. 176 & 191]. The definition of quasiregularity will be found, for instance, in Martio-Rickman-Väisälä [4]; it is identical to the definition of quasiconformal function found in [3, Ch. VI].

Definition 1.1. Let $R \subset \mathbb{C}$ be a region, and let $f: R \to \mathbb{C}$ be a nonconstant continuous function. If f has the following two properties, it will be called *quasiregular.*

1. On each closed rectangle $T \subset R$ with sides parallel to the coordinate axes, f is absolutely continuous on a.e. horizontal and vertical line. In addition,

$$\iint_T (|f_z|^2 + |f_{\bar{z}}|^2) dx\, dy \quad \text{is finite.}$$

2. f satisfies a Beltrami equation $f_{\bar{z}} = \mu f_z$ a.e. in R for some measurable function μ with $\|\mu\|_\infty = \operatorname{ess\,sup}_R |\mu| < 1$.

It is a (nontrivial) consequence of the definition that $f_z(z) \neq 0$ for almost every $z \in R$ so that the function

$$\mu = f_{\bar{z}}/f_z \tag{1.3}$$

is associated with f. It is called the *(complex) dilatation* of f.

Definition 1.2. A homeomorphic quasiregular mapping will be called *quasiconformal.*

It is a consequence of the fact that the Jacobian

$$J_f = |f_z|^2 - |f_{\bar{z}}|^2 = (1 - |\mu|^2)\,|f_z|^2 \tag{1.4}$$

is positive a.e. that quasiconformal mappings are orientation-preserving.

Quasiregular mappings f are characterized by the representation $f = A \circ g$ where A is holomorphic and g is quasiconformal. Since homeomorphisms and holomorphic mappings carry open sets onto open sets, it follows that quasiregular mappings f are open and that $|f|$ cannot have an interior maximum.

The notion of a good approximation is also adopted from Lehto-Virtanen [3, p. 194].

Definition 1.3. Let $f, f_n: R \to \mathbb{C}$ be quasiregular mappings of a region R with dilatations μ, μ_n, respectively, $n = 1, 2, \ldots$. The sequence $\{f_n\}$ is called a *good approximation* of f if

$$\lim_{n \to \infty} f_n = f \quad \text{uniformly on each compact subset of } R, \tag{1.5}$$

$$\sup_n \|\mu_n\|_\infty < 1, \tag{1.6}$$

and

$$\lim_{n \to \infty} |\mu_n(z) - \mu(z)| = 0 \quad \text{for almost all } z \in R. \tag{1.7}$$

There is an example (see [3, pp. 195–196]) that shows that, in general, property (1.7) is not implied by properties (1.5) and (1.6).

The dilatation of a complex planar spline is piecewise defined in each mesh and has the form

$$\mu(z)=\frac{c}{b} \quad \text{in each triangle} \tag{1.8}$$

or

$$\mu(z)=\frac{c-2d\bar{z}}{b+2dz} \quad \text{in each rectangle.} \tag{1.9}$$

This is an easy consequence of the definition of the dilatation (1.3) and the elements (1.1), (1.2), respectively.

As a rule the dilatation of a complex planar spline is not well defined on the whole grid since it has jump discontinuities on all edges which are common to two meshes. However, in this case it is evident that the essential supremum of μ is given by

$$\|\mu\|_\infty = \max_{j=1,2,\ldots,M} \sup_{\mathring{m}_j} |\mu(z)| \tag{1.10}$$

where $m_1, m_2, \ldots, m_M$ are the meshes of the grid and $\mathring{m}_j$ is the set of interior points of m_j, $j=1, 2, \ldots, M$.

The following theorem is quite simple, nevertheless very useful for computing $\|\mu\|_\infty$ of (1.10).

Theorem 1.1. *Let μ be the complex dilatation of a complex planar spline on a grid S_h such that $\mu(z)$ of* (1.8) *or* (1.9) *is defined and finite in each mesh. Then there is a mesh m in S_h and a z_0 on an edge of m such that $|\mu(z_0)|=\|\mu\|_\infty$.*

Proof. If S_h is a triangular grid, then $\mu(z)$ is piecewise constant (i.e., constant in each mesh) and the assertion is trivial. In the rectangular case, the proof follows from the fact that

$$|\mu(z)|=\left|\frac{\bar{c}-2\bar{d}z}{b+2dz}\right| \tag{1.11}$$

is the absolute value of a holomorphic mapping in each rectangle. □

We shall see later that this theorem can be considerably improved in some situations by showing that $\|\mu\|_\infty$ is eventually taken on at a vertex of the underlying grid.

2. Interpolating Properties of Complex Planar Splines

In this section let $f: R\to\mathbb{C}$ be a *smooth* function on a region $R\subset\mathbb{C}$. By smooth we mean here that the second partial derivatives of f with respect to z and its complex conjugate $\bar{z}$ exist and are Lipschitz continuous.

Let S_h be a grid in R. Since we want to study interpolation problems

$$v(z_j)=f(z_j), \quad j=1, 2, \ldots, N \tag{2.1}$$

where z_j, $j=1, 2, \ldots, N$, are the grid points of S_h and v is a complex planar spline, we observe that (2.1) is equivalent to

$$\operatorname{Re} v(z_j)=\operatorname{Re} f(z_j); \quad \operatorname{Im} v(z_j)=\operatorname{Im} f(z_j) \tag{2.2}$$

where Re and Im stand for real part and imaginary part, respectively. That means the complex problem (2.1) can be regarded as a twice as large real interpolation problem.

For later reference we repeat the interpolation formulae for triangles and rectangles from Opfer-Puri [5].

Let P_1, P_2, P_3 be the vertices of a triangle, and p as defined in (1.1). Then $p(P_j)=\zeta_j$, $j=1, 2, 3$, for any three complex numbers $\zeta_1, \zeta_2, \zeta_3$ has the solution

$$a=(1/\delta)(\zeta_1(P_2\overline{P_3}-\overline{P_2}P_3)+\zeta_2(\overline{P_1}P_3-P_1\overline{P_3})+\zeta_3(P_1\overline{P_2}-\overline{P_1}P_2)), \tag{2.3}$$

$$b=(1/\delta)(\zeta_1(\overline{P_2}-\overline{P_3})+\zeta_2(\overline{P_3}-\overline{P_1})+\zeta_3(\overline{P_1}-\overline{P_2})), \tag{2.4}$$

$$c=(1/\delta)(\zeta_1(P_3-P_2)+\zeta_2(P_1-P_3)+\zeta_3(P_2-P_1)), \tag{2.5}$$

where

$$\delta=2i \operatorname{Im}(P_1\overline{P_2}+P_2\overline{P_3}+P_3\overline{P_1}). \tag{2.6}$$

Let P_1, P_2, P_3, P_4 be the vertices of a rectangle whose sides are parallel to the axes such that P_{j+1} follows P_j in positive orientation, $j=1, 2, 3, 4$ ($P_5=P_1$). Then $p(P_j)=\zeta_j$, $j=1, 2, 3, 4$, has the following solution where $\zeta_1, \zeta_2, \zeta_3, \zeta_4$ are any four complex numbers and p is defined in (1.2).

$$a=(1/\delta)((P_3^2-\overline{P_3}^2)\zeta_1-(P_4^2-\overline{P_4}^2)\zeta_2+(P_1^2-\overline{P_1}^2)\zeta_3-(P_2^2-\overline{P_2}^2)\zeta_4), \tag{2.7}$$

$$b=(2/\delta)(-P_3\zeta_1+P_4\zeta_2-P_1\zeta_3+P_2\zeta_4), \tag{2.8}$$

$$c=(2/\delta)(\overline{P_3}\zeta_1-\overline{P_4}\zeta_2+\overline{P_1}\zeta_3-\overline{P_2}\zeta_4), \tag{2.9}$$

$$d=(1/\delta)(\zeta_1-\zeta_2+\zeta_3-\zeta_4), \tag{2.10}$$

where

$$\delta=4(P_2-P_1)(P_3-P_2). \tag{2.11}$$

The error analysis for real interpolation problems is known within the framework of the theory of finite elements (see, e.g., Schultz [6]). Before summarizing these results we shall introduce some terminology.

Let S_h be a rectangular grid and σ_j the midpoints of the meshes m_j of S_h, $j=1, 2, \ldots, M$. Let g be any complex-valued function defined (at least) on all midpoints $\sigma_1, \sigma_2, \ldots, \sigma_M$. The number

$$\|\!|g|\!\|_\infty = \max_{j=1, 2, \ldots, M} |g(\sigma_j)| \tag{2.12}$$

will be referred to as the *uniform midpoint norm* of g.

As in Ciarlet-Raviart [1, p. 179], a family of grids S_h will be called *regular* if the quotient of the radii of circumscribed and inscribed circles for any mesh is

uniformly bounded independent of h. The least of all uniform upper bounds we denote by q.

The error analysis used can be summarized in

Theorem 2.1. *Let f be a smooth function in a region R, and let S_h be regular grids in R. If $v^{(h)}$ is the complex planar spline interpolating f on the gridpoints of S_h, then*

$$\|f-v^{(h)}\|_\infty \leqq c_1 h^2, \tag{2.13}$$

$$\|f_z-v_z^{(h)}\|_\infty \leqq c_2 h, \tag{2.14}$$

$$\|f_{\bar z}-v_{\bar z}^{(h)}\|_\infty \leqq c_2 h, \tag{2.15}$$

$$\hat{\|}f_z-v_z^{(h)}\hat{\|}_\infty \leqq c_3 h^2, \tag{2.16}$$

$$\hat{\|}f_{\bar z}-v_{\bar z}^{(h)}\hat{\|}_\infty \leqq c_3 h^2, \tag{2.17}$$

where the norms are relative to S_h and the constants c_1, c_2, c_3 depend on f and q.

Proof. Since (2.1) is equivalent to (2.2), we have for any norm $\|\ \|$ over $\mathbb{C}$:

$$\|f-v^{(h)}\| \leqq \|\mathrm{Re}(f-v^{(h)})\| + \|\mathrm{Im}(f-v^{(h)})\|.$$

Furthermore, both $\|f_z-v_z^{(h)}\|$ and $\|f_{\bar z}-v_{\bar z}^{(h)}\|$ are bounded by

$$\tfrac{1}{2}\{\|\mathrm{Re}(f_x-v_x^{(h)})\| + \|\mathrm{Im}(f_x-v_x^{(h)})\| + \|\mathrm{Re}(f_y-v_y^{(h)})\| + \|\mathrm{Im}(f_y-v_y^{(h)})\|\}.$$

Thus the estimations (2.13) to (2.15) can be directly obtained by applying the real theory from Ciarlet-Raviart [1, Theorem 2]. For these results to hold, it is sufficient to assume that f has continuous second derivatives with respect to z and $\bar z$.

For the derivation of (2.16) and (2.17) (which apply only to the rectangular case) we use

$$D(\tilde v^{(h)}-\tilde f)(z) = \tfrac{1}{2}\sum_{j=1}^{4}\{D^2\tilde f(\eta_j(z))\cdot(P_j-z)^2\}\,D\tilde v_j^h(z), \tag{2.18}$$

which is a special form of a formula given by Ciarlet-Raviart [1, formula (2.3)]. In the formula (2.18), D and D^2 mean the first and second Fréchet derivatives. The tilde indicates that we deal with either the real or the imaginary part of the corresponding function and $\eta_j(z)$ is a point in the segment $[z, P_j]$, where P_j are the vertices of the rectangle under investigation, $j=1, 2, 3, 4$. The function v_j stands for the polynomial defined by $v_j(P_k)=\delta_{jk}$; $j, k=1, 2, 3, 4$. If we insert $z=\frac{1}{4}(P_1+P_2+P_3+P_4)$ into the above formula and assume that $P_2-P_1=h_1$ and $P_4-P_1=ih_2$ where $0<h_1,\ h_2\leqq h$, then (2.18) will consist of terms of the form $h_1^2\tilde f_{xx}$, $h_1h_2\tilde f_{xx}$, $h_2^2\tilde f_{xx}$ and corresponding terms with $\tilde f_{xx}$ replaced by $\tilde f_{yy}$. In addition, there are terms of the form $h_2(f_{xy}(\eta_3)-\tilde f_{xy}(\eta_1))$ and similar terms where h_2 is replaced by h_1 and η_3 by η_4 and η_1 by η_2. That means if f_{xy} is Lipschitz continuous on the diagonals of the rectangle, these terms can be bounded by h_1^2, h_2^2 or h_1h_2, too. □

It should be remarked that the estimates (2.16) and (2.17) do not carry over to triangular meshes since the derivatives of $v^{(h)}$ are constant in each triangular

mesh. Also (2.16) and (2.17) are not valid if, in the definition (2.12) of $\hat{\|}\ \hat{\|}$, the midpoints σ_j are replaced by other points.

In order that the error analysis of Theorem 2.1 is in effect, we assume from now on that all grid families are regular.

3. Quasiregularity and Quasiconformality of Complex Planar Splines

We proceed to the question whether a complex planar spline is quasiregular.

Lemma 3.1. *Let $v^{(h)}$ be a complex planar spline on a grid S_h. Then $v^{(h)}$ is absolutely continuous on all straight lines parallel to the coordinate axes. In addition, $\iint_{S_h}(|v_x^{(h)}|^2+|v_{\bar{z}}^{(h)}|^2)\,dx\,dy$ is finite.*

Proof. The restriction of $v^{(h)}$ to a straight line parallel to one of the coordinate axes is a continuous piecewise linear function and, as such, is absolutely continuous. In addition, S_h consists of finitely many meshes on which the derivatives of $v^{(h)}$ are bounded, so that the given integral is finite. □

This lemma in connection with Definition 1.1 implies that we have to pay attention only to the dilatations $\mu^{(h)}$ (defined in (1.3)) of the complex planar splines when investigating quasiregularity.

Theorem 3.1. *Let $v^{(h)}$ interpolate a smooth function f at the gridpoints of a grid S_h, and denote the dilatations of f and $v^{(h)}$ by μ and $\mu^{(h)}$, respectively. Furthermore, assume that $f_z(z)\neq 0$ for all $z\in R$. Then for all h sufficiently small*

$$v_z^{(h)}(z)\neq 0, \qquad z\in S_h, \tag{3.1}$$

$$|\,\|\mu\|_\infty-\|\mu^{(h)}\|_\infty|\leqq\|\mu-\mu^{(h)}\|_\infty\leqq c_4 h, \tag{3.2}$$

$$|\,\hat{\|}\mu\hat{\|}_\infty-\hat{\|}\mu^{(h)}\hat{\|}_\infty|\leqq\hat{\|}\mu-\mu^{(h)}\hat{\|}_\infty\leqq c_5 h^2. \tag{3.3}$$

Proof. Since S_h is compact, $|f_z|$ has a positive lower bound on S_h. Therefore inequalities (3.1) to (3.3) follow directly from inequalities (2.14) through (2.17) and the triangle inequality. □

That we cannot omit the condition $f_z(z)\neq 0$ for $z\in S_h$ in the previous theorem is shown by the following example.

Example 3.1. Let $f(z)=z^2$ for $z\in\mathbb{C}$ and take a triangular grid which contains the mesh with the vertices $0,\ h/\sqrt{2},\ ih/\sqrt{2}$ or take a rectangular grid which contains the mesh with vertices $0, h, h(1+i), hi$. If $\mu^{(h)}$ is the dilatation of the complex planar spline interpolating f at the grid points, then (after some calculations) we obtain $\mu^{(h)}(0)=-i$. In this case $\|\mu^{(h)}\|_\infty=1$, whereas the dilatation of f is constantly zero.

Theorem 3.2. *Let f be a smooth quasiregular mapping with $f_z(z)\neq 0$ for $z\in R$. Then the interpolating complex planar spline $v^{(h)}$ on a grid $S_h\subset R$ is quasiregular, provided h is sufficiently small.*

Proof. Lemma 3.1 states that we only need to estimate the dilatation $\mu^{(h)}$ of $v^{(h)}$. For this purpose inequality (3.2) of Theorem 3.1 implies that

$$\|\mu^{(h)}\|_\infty \leqq \|\mu\|_\infty + c_4 h,$$

which is less than 1 if h is sufficiently small. □

Remark. Since quasiregular mappings are open, the maximum of $|v^{(h)}|$ is taken only at the boundary of S_h provided h is sufficiently small.

Theorems 3.1 and 3.2 imply that complex planar splines provide a good approximation to a given quasiregular function f. This is not entirely correct in the sense of Definition 1.2 since the splines are defined only on a compact subset. However, if we restrict f to a compact set from which a grid can be constructed, then we have not only the pointwise convergence of the dilatations in the sense of (1.7) but even the uniform approximation as stated in (3.2). This means that complex planar splines provide not only good approximations but even better approximations in the sense of (3.2).

Next we shall show that if f is quasiconformal, then so is the interpolating spline $v^{(h)}$ if h is sufficiently small. We shall prove this more generally in the following lemma.

Lemma 3.2. *Let f be a quasiconformal mapping of a region R, and assume that a sequence $\{f_n\}$ of quasiregular mappings is a good approximation of f on R. If K is a compact subset of R, then f_n is one-to-one, hence quasiconformal, on K for all n sufficiently large.*

Proof. If the assertion were not true, then there would exist points z_{n_j}, $\hat{z}_{n_j} \in K$, $z_{n_j} \neq \hat{z}_{n_j}$, such that $f_{n_j}(z_{n_j}) = f_{n_j}(\hat{z}_{n_j})$. Since K is compact, we may assume that $z_{n_j} \to p$ and $\hat{z}_{n_j} \to \hat{p}$ as $j \to \infty$. By the uniform convergence on K,

$$0 = f_{n_j}(z_{n_j}) - f_{n_j}(\hat{z}_{n_j}) \to f(p) - f(\hat{p}).$$

Consequently, $p = \hat{p}$ since f is one-to-one.

Since $\mu_{n_j} = (f_{n_j})_{\bar{z}}/(f_{n_j})_z \to \mu = f_{\bar{z}}/f_z$ and $\sup_j \|\mu_{n_j}\| < 1$, there exist in a neighborhood G of K (see [3, p. 218]) quasiconformal mappings g_k with the dilatations $\mu_{n_{j_k}}$ and with $g_k \to f$ locally uniformly in G. It follows that $f_{n_{j_k}} = A_k \circ g_k$ for some analytic functions A_k on $g_k(G)$. Furthermore, since $f_{n_{j_k}} \to f$ and $g_k \to f$ as $k \to \infty$, it follows that $\{A_k\}$ converges to the identity uniformly on compact subsets of $f(G)$.

Let $w_k = g(z_{n_{j_k}})$ and $\hat{w}_k = g(\hat{z}_{n_{j_k}})$. Then $\{w_k\}$ and $\{\hat{w}_k\}$ converge to the same limit and so

$$0 = \frac{f_{n_{j_k}}(z_{n_{j_k}}) - f_{n_{j_k}}(\hat{z}_{n_{j_k}})}{w_k - \hat{w}_k} = \frac{A_k(w_k) - A_k(\hat{w}_k)}{w_k - \hat{w}_k} \to 1$$

as $k \to \infty$. This is a contradiction. □

Of course, Lemma 3.2 remains valid if the discrete index n is replaced by a continuous index h.

Theorem 3.3. *Let f be a smooth quasiconformal mapping with $f_z(z) \neq 0$ for $z \in R$. Then the interpolating complex planar spline $v^{(h)}$ on a grid $S_h \subset R$ is quasiconformal, open, and the maximum of $|v^{(h)}|$ is taken on only at the boundary of S_h, provided h is sufficiently small.*

Proof. In view of Theorem 3.2 we only need to show that $v^{(h)}$ is one-to-one. In order to apply Lemma 3.2, observe that, if h is sufficiently small, we can find a slightly larger grid $T_h \subset R$ that contains S_h in its interior and contains the vertices of S_h among its vertices. Then the restrictions to S_h of interpolating splines on T_h agree with the interpolating splines on S_h. It now follows by applying Lemma 3.2 on the interior of T_h that $v^{(h)}$ is quasiconformal on the compact subset S_h, if h is sufficiently small. □

We would like to finish this section with some special properties of complex planar splines $v^{(h)}$ and their dilatations $\mu^{(h)}$ which will facilitate the numerical computation of these quantities.

In Theorem 1.1 we already stated that $\|\mu^{(h)}\|_\infty$ is achieved on an edge of S_h. But we can show more.

Theorem 3.4. *Let $v^{(h)}$ interpolate a smooth quasiregular mapping f on a rectangular grid S_h. For $z \in R$ assume that $f_z(z) \neq 0$, that $D(z) \equiv f_{zz}(z) - f_{z\bar{z}}(z) \neq 0$, and that one of the following conditions holds:*

$$\|\mu\|_\infty < \tfrac{1}{3}, \tag{3.4}$$

$$2|v(z)\,D(z)| < |D(z) + v(z)\,\overline{D(z)}|, \tag{3.5}$$

$$2|\mathrm{Re}\{\mu(z)[|D(z)|^2 + \overline{v(z)}\,D(z)^2]\}| < |D(z) + v(z)\,\overline{D(z)}|^2, \tag{3.6}$$

where $\mu = f_{\bar{z}}/f_z$ and $v = f_{\bar{z}}/\bar{f}_z$. Then in each mesh the modulus of the dilatation $\mu^{(h)}$ of $v^{(h)}$ assumes its maximum at a vertex if h is sufficiently small. Consequently, if h is small enough, there is a vertex z_1 in S_h such that $|\mu^{(h)}(z_1)| = \|\mu^{(h)}\|_\infty$.

Proof. We already know that $|\mu^{(h)}|$ cannot have a maximum at an interior meshpoint. We show that for any small mesh the restriction of $|\mu^{(h)}|$ to that mesh is a maximum only at a vertex of that mesh.

Consider the mesh with vertices z_0, $z_0 + h_1$, $z_0 + h_1 + ih_2$, $z_0 + ih_2$ where $0 < h_1, h_2 \leqq h$. For simplicity let us assume that $z_0 = 0$. Then we may parametrize the edges of the mesh by

$$\begin{aligned} z_1(t) &= h_1 t, & z_2(t) &= h_1 + ih_2 t, \\ z_3(t) &= h_1(1-t) + ih_2, & z_4(t) &= ih_2(1-t) \end{aligned} \tag{3.7}$$

where $0 \leqq t \leqq 1$. For this mesh we see from (2.7)–(2.11) that the coefficients of the element $p(z) = a + bz + c\bar{z} + d(z^2 - \bar{z}^2)$ satisfy

$$a = f(0) + o(h), \quad b = f_z(0) + o(h), \quad c = f_{\bar{z}}(0) + o(h), \quad d = \tfrac{1}{4}D(0) + o(h) \tag{3.8}$$

as $h \to 0$. The $o(h)$ terms are uniform in z.

For $j=1, 2, 3, 4$ let

$$\phi_j(t)=|\mu^{(h)}(z_j(t))|^2=\left|\frac{c-2d\overline{z_j(t)}}{b+2dz_j(t)}\right|^2.$$

If the dot denotes differentiation with respect to t, then

$$\dot{\phi}_j=\frac{-4}{|b+2dz_j|^2}\operatorname{Re}\left\{\frac{(b\bar{d}+\bar{c}d)(c-2dz_j)\dot{z}_j}{(b+2dz_j)}\right\}$$

and

$$\ddot{\phi}_j=\frac{8h_j^2}{|b+2dz_j|^2}\operatorname{Re}\left\{\frac{(b\bar{d}+\bar{c}d)}{(b+2dz_j)}\left[d+\frac{(c-2d\bar{z}_j)\bar{d}}{(\bar{b}+2\bar{d}\bar{z}_j)}-\frac{2(-1)^j(c-2d\bar{z}_j)d}{(b+2dz_j)}\right]\right\}$$

where $h_3=h_1$ and $h_4=h_2$. Using (3.7) and (3.8) we have

$$\begin{aligned}\ddot{\phi}_j(t)&=\frac{8h_j^2}{|b|^2}\operatorname{Re}\left\{\left(\bar{d}+\frac{\bar{c}}{b}d\right)\left(d+\frac{c}{\bar{b}}\bar{d}-2(-1)^j\frac{c}{b}d\right)\right\}+o(h^2)\\&=\frac{8h_j^2}{|b|^2}\left[\left|d+\frac{c}{\bar{b}}\bar{d}\right|^2-2(-1)^j\operatorname{Re}\left\{\frac{c}{b}\left(|d|^2+\frac{\bar{c}}{b}d^2\right)\right\}\right]+o(h^2)\\&=\frac{h_j^2}{2|f_z(0)|^2}\left[|D(0)+v(0)\overline{D(0)}|^2-2(-1)^j\right.\\&\quad\left.\cdot\operatorname{Re}\{\mu(0)[|D(0)|^2+\overline{v(0)}D(0)^2]\}\right]+o(h^2)\end{aligned}$$

where the $o(h^2)$ terms are uniform in t.

If condition (3.6) holds, then the functions

$$|f_z|^{-2}[|D+v\bar{D}|^2\pm 2\operatorname{Re}\{\mu[|D|^2+\bar{v}D^2]\}]$$

have a positive lower bound on the grids. Therefore $\ddot{\phi}_j(t)>0$ for all sufficiently small h depending only on f. In this case each ϕ_j is strictly convex, and the conclusion follows.

To complete the proof, observe that

$$|\operatorname{Re}\{\mu[|D|^2+\bar{v}D^2]\}|\leqq|\mu|\,|D|\,|\bar{D}+\bar{v}D|=|vD|\,|D+v\bar{D}|$$

so that condition (3.5) implies condition (3.6). Finally, if $\|v\|_\infty=\|\mu\|_\infty<\frac{1}{3}$, then

$$2|vD|<2|vD|+(1-3|v|)|D|=|D|-|v\bar{D}|\leqq|D+v\bar{D}|$$

so that condition (3.4) implies condition (3.5). □

Since the first three examples in Sect. 4 are analytic, condition (3.4) applies. In addition, example (4.4) satisfies condition (3.5). It follows then that the conclusion of Theorem 3.4 is valid for the first four examples in Sect. 4.

For certain functions the dilatation of the corresponding interpolating spline has the same values on each mesh. In this situation it is sufficient to compute the dilatation on a single mesh. In particular, we have

Theorem 3.5. *Let*

$$f(z)=e^{\alpha z+\beta\bar{z}} \quad \text{where } \alpha,\beta\in\mathbb{C}. \tag{3.9}$$

If S is a triangle or a rectangle and if v is the spline which interpolates f at the vertices of S and v_τ is the spline which interpolates f at the vertices of the mesh $S+\tau$, $\tau\in\mathbb{C}$, then for the corresponding dilatations μ and μ_τ we have

$$\mu(z)=\mu_\tau(z+\tau), \quad z\in S. \tag{3.10}$$

Proof. Assume that S is a triangle. In this case the formulae (1.8), (2.4), and (2.5) yield

$$\mu(z)=\frac{c}{b}=\frac{\zeta_1(P_3-P_2)+\zeta_2(P_1-P_3)+\zeta_3(P_2-P_1)}{\zeta_1(\bar{P}_2-\bar{P}_3)+\zeta_2(\bar{P}_3-\bar{P}_1)+\zeta_3(\bar{P}_1-\bar{P}_2)}$$

in S, where $\zeta_j=f(P_j)$. In the triangle $S+\tau$ one uses $\zeta_j=f(P_j+\tau)=f(P_j)f(\tau)$ and it is evident that the factor $f(\tau)$ cancels from the numerator and denominator. Finally, since only differences P_i-P_j occur, it follows that μ_τ in $S+\tau$ is independent of τ. If S is a rectangle, then formulae (1.9) and (2.8)–(2.10) imply that

$$\mu(z)=\frac{c-2d\bar{z}}{b+2dz}=\frac{\bar{P}_3\zeta_1-\bar{P}_4\zeta_2+\bar{P}_1\zeta_3-\bar{P}_2\zeta_4-(\zeta_1-\zeta_2+\zeta_3-\zeta_4)\bar{z}}{-P_3\zeta_1+P_4\zeta_2-P_1\zeta_3+P_2\zeta_4+(\zeta_1-\zeta_2+\zeta_3-\zeta_4)z}. \tag{3.11}$$

Again the multiplicative factor $f(\tau)$ cancels from the ζ_j's. From this, (3.10) follows since a translation of P_1, P_2, P_3, P_4 and the same translation in z cancel in (3.11). □

Theorem 3.5 applies to the examples f_1 and f_4 in Sect. 4. In these cases it is sufficient to compute $\mu^{(h)}$ only for one rectangle (or triangle).

In general, the computation of $\|\mu-\mu^{(h)}\|_\infty$ where μ is the dilatation of f and $\mu^{(h)}$ the dilatation of the interpolating spline $v^{(h)}$ seems to be a little more difficult, since we do not know whether a boundary maximum principle holds for $|\mu-\mu^{(h)}|$. But we have

Theorem 3.6. *Assume that $f(z)=g(z+k\bar{z})$ where g is holomorphic, $g'(z)\neq 0$, and $k\in\mathbb{C}$ with $|k|<1$. Let $v^{(h)}$ interpolate f on a grid S_h. If μ and $\mu^{(h)}$ are the dilatations of f and $v^{(h)}$, respectively, then there is a z_0 on an edge of S_h such that*

$$|\mu(z_0)-\mu^{(h)}(z_0)|=\|\mu-\mu^{(h)}\|_\infty, \tag{3.12}$$

provided h is sufficiently small.

Proof. In this case the dilatation μ is identically k. If the grid is triangular, then $\mu^{(h)}$ is constant in each triangle and the theorem is obvious. Assume therefore that S_h is rectangular. Let $\lambda^{(h)}=\mu-\mu^{(h)}$. In each rectangle, $\lambda^{(h)}$ has the form

$$\lambda^{(h)}(z)=k-\frac{c-2d\bar{z}}{b+2dz}.$$

We may restrict our attention to rectangles in which $d \neq 0$ since, otherwise, $\lambda^{(h)}$ is constant. The Jacobian J of $\lambda^{(h)}$ is

$$J(z)=|\lambda_z^{(h)}|^2-|\lambda_{\bar{z}}^{(h)}|^2=\frac{4|d|^2}{|b+2dz|^2}[|\mu^{(h)}(z)|^2-1].$$

By formula (3.2) of Theorem 3.1 this Jacobian is negative for small h. Thus the implicit function theorem implies that $\lambda^{(h)}$ is locally one-to-one, hence open, and satisfies a maximum modulus principle in such rectangles. Consequently, $|\lambda^{(h)}|$ assumes its maximum on an edge of S_h. □

In the following section the first four examples are of the form considered in Theorem 3.6.

4. Numerical Examples

For the numerical examples we use a subdivision of the standard square

$$Q=\{z: 0 \leqq \operatorname{Re} z, \operatorname{Im} z \leqq 1\}$$

as grid. S_h stands here for Q subdivided in little squares of equal sidelengths h. For h we choose

$$h=2^{-k}, \quad k=0, 1, \ldots, 6$$

and as functions to be interpolated on S_h we select

$$f_1(z)=e^z, \tag{4.1}$$

$$f_2(z)=1/(1+(z+\tfrac{1}{8})^2), \tag{4.2}$$

$$f_3(z)=\exp\{(z+\tfrac{1}{8})^2\}, \tag{4.3}$$

$$f_4(z)=\exp\{z+\tfrac{1}{2}\bar{z}\}, \tag{4.4}$$

$$f_5(z)=\exp\{(z+\tfrac{1}{8})^2+\tfrac{1}{2}(\bar{z}+\tfrac{1}{8})^2\}. \tag{4.5}$$

If μ_j is the dilatation of f_j, $1 \leqq j \leqq 5$, then

$$\mu_j=0, \quad \text{for } j=1, 2, 3$$

since the corresponding f_j is holomorphic and

$$\mu_4(z)=\tfrac{1}{2}, \quad \mu_5(z)=\tfrac{1}{2}\frac{\bar{z}+1/8}{z+1/8} \quad \text{for } z \in Q.$$

Note that $|\mu_5|=\frac{1}{2}$. In the following Tables 1–5 we list

$$\varepsilon_1=\|f-v^{(h)}\|_\infty, \quad \varepsilon_2=\|\mu-\mu^{(h)}\|_\infty \quad \text{and} \quad \varepsilon_3=\hat{\|}\mu-\mu^{(h)}\hat{\|}_\infty$$

where $\hat{\|}\ \hat{\|}$ is the uniform midpoint norm (2.12).

In each case we also computed the numerical order of convergence c_j of ε_j, $j=1, 2, 3$.

Table 1. Interpolating $f(z)=e^z$ by complex planar splines $v^{(h)}$

h	$\varepsilon_1=\|f-v^{(h)}\|_\infty$	c_1	$\varepsilon_2=\|\mu-\mu^{(h)}\|_\infty$	c_2	$\varepsilon_3=\|\|\mu-\mu^{(h)}\|\|_\infty$	c_3
1	$3.328\cdot10^{-1}$		$3.807\cdot10^{-1}$		$8.348\cdot10^{-2}$	
1/2	$8.450\cdot10^{-2}$	1.98	$1.841\cdot10^{-1}$	1.05	$2.084\cdot10^{-2}$	2.00
1/4	$2.121\cdot10^{-2}$	1.99	$9.024\cdot10^{-2}$	1.03	$5.208\cdot10^{-3}$	2.00
1/8	$5.307\cdot10^{-3}$	2.00	$4.466\cdot10^{-2}$	1.01	$1.302\cdot10^{-3}$	2.00
1/16	$1.327\cdot10^{-3}$	2.00	$2.221\cdot10^{-2}$	1.01	$3.255\cdot10^{-4}$	2.00
1/32	$3.318\cdot10^{-4}$	2.00	$1.108\cdot10^{-2}$	1.00	$8.138\cdot10^{-5}$	2.00
1/64	$8.296\cdot10^{-5}$	2.00	$5.531\cdot10^{-3}$	1.00	$2.035\cdot10^{-5}$	2.00

Table 2. Interpolating $f(z)=1/(1+(z+1/8)^2)$ by complex planar splines $v^{(h)}$

h	$\varepsilon_1=\|f-v^{(h)}\|_\infty$	c_1	$\varepsilon_2=\|\mu-\mu^{(h)}\|_\infty$	c_2	$\varepsilon_3=\|\|\mu-\mu^{(h)}\|\|_\infty$	c_3
1	2.643		$8.694\cdot10^{-1}$		$6.763\cdot10^{-1}$	
1/2	2.054	0.36	$7.636\cdot10^{-1}$	0.19	$5.878\cdot10^{-1}$	0.20
1/4	1.264	0.70	$6.253\cdot10^{-1}$	0.29	$3.926\cdot10^{-1}$	0.58
1/8	$6.322\cdot10^{-1}$	1.00	$4.441\cdot10^{-1}$	0.49	$1.981\cdot10^{-1}$	0.99
1/16	$2.169\cdot10^{-1}$	1.54	$2.759\cdot10^{-1}$	0.69	$7.643\cdot10^{-2}$	1.37
1/32	$6.016\cdot10^{-2}$	1.85	$1.555\cdot10^{-1}$	0.83	$2.427\cdot10^{-2}$	1.66
1/64	$1.547\cdot10^{-2}$	1.96	$8.270\cdot10^{-2}$	0.91	$6.866\cdot10^{-3}$	1.82

Table 3. Interpolating $f(z)=\exp\{(z+1/8)^2\}$ by complex planar splines $v^{(h)}$

h	$\varepsilon_1=\|f-v^{(h)}\|_\infty$	c_1	$\varepsilon_2=\|\mu-\mu^{(h)}\|_\infty$	c_2	$\varepsilon_3=\|\|\mu-\mu^{(h)}\|\|_\infty$	c_3
1	1.991		0.9336		0.6383	
1/2	0.6953	1.52	0.7101	0.40	0.1892	1.75
1/4	0.1899	1.87	0.4817	0.56	$5.206\cdot10^{-2}$	1.86
1/8	$4.854\cdot10^{-2}$	1.97	0.2925	0.72	$1.381\cdot10^{-2}$	1.91
1/16	$1.220\cdot10^{-2}$	1.99	0.1629	0.84	$3.585\cdot10^{-3}$	1.95
1/32	$3.055\cdot10^{-3}$	2.00	$8.613\cdot10^{-2}$	0.92	$9.134\cdot10^{-4}$	1.97
1/64	$7.640\cdot10^{-4}$	2.00	$4.430\cdot10^{-2}$	0.96	$2.305\cdot10^{-4}$	1.99

Table 4. Interpolating $f(z)=\exp\{z+0.5\bar{z}\}$ by complex planar splines $v^{(h)}$

h	$\varepsilon_1=\|f-v^{(h)}\|_\infty$	c_1	$\varepsilon_2=\|\mu-\mu^{(h)}\|_\infty$	c_2	$\varepsilon_3=\|\|\mu-\mu^{(h)}\|\|_\infty$	c_3
1	0.6238		0.2997		$7.349\cdot10^{-2}$	
1/2	0.2191	1.51	0.1508	0.99	$1.922\cdot10^{-2}$	1.93
1/4	$6.550\cdot10^{-2}$	1.74	$7.501\cdot10^{-2}$	1.01	$4.863\cdot10^{-3}$	1.98
1/8	$1.795\cdot10^{-2}$	1.87	$3.731\cdot10^{-2}$	1.01	$1.219\cdot10^{-3}$	2.00
1/16	$4.699\cdot10^{-3}$	1.93	$1.860\cdot10^{-2}$	1.00	$3.051\cdot10^{-4}$	2.00
1/32	$1.202\cdot10^{-3}$	1.97	$9.282\cdot10^{-3}$	1.00	$7.629\cdot10^{-5}$	2.00
1/64	$3.041\cdot10^{-4}$	1.98	$4.637\cdot10^{-3}$	1.00	$1.907\cdot10^{-5}$	2.00

Table 5. Interpolating $f(z)=\exp\{(z+1/8)^2+0.5(\bar{z}+1/8)^2\}$ by complex planar splines $v^{(h)}$

h	$\varepsilon_1=\|f-v^{(h)}\|_\infty$	c_1	$\varepsilon_2=\|\mu-\mu^{(h)}\|_\infty$	c_2	$\varepsilon_3=\hat{\|}\mu-\mu^{(h)}\hat{\|}_\infty$	c_3
1	2.109		1.238		$2.521\cdot10^{-1}$	
1/2	1.083	0.96	$8.023\cdot10^{-1}$	0.63	$1.229\cdot10^{-1}$	1.04
1/4	$4.327\cdot10^{-1}$	1.32	$7.211\cdot10^{-1}$	0.15	$3.807\cdot10^{-2}$	1.69
1/8	$1.409\cdot10^{-1}$	1.62	$5.374\cdot10^{-1}$	0.42	$1.043\cdot10^{-2}$	1.87
1/16	$4.052\cdot10^{-2}$	1.80	$3.143\cdot10^{-1}$	0.77	$2.774\cdot10^{-3}$	1.91
1/32	$1.089\cdot10^{-2}$	1.90	$1.651\cdot10^{-1}$	0.93	$7.137\cdot10^{-4}$	1.96
1/64	$2.825\cdot10^{-3}$	1.95	$8.463\cdot10^{-2}$	0.96	$1.809\cdot10^{-4}$	1.98

We computed corresponding norms also for other functions and also for grids not consisting of squares. The results were analogous. Interestingly, the norms were taken on always on meshes of the boundary of the whole grid.

Acknowledgement. The authors are grateful for several helpful comments from Professor Gaier at Giessen.

References

1. Ciarlet, P.G., Raviart, P.A.: General Lagrange and Hermite interpolation in $\mathbb{R}^n$ with applications to finite element methods. Arch. Rational Mech. Anal. **46**, 177–199 (1972)
2. Jaenisch, S.: Stückweise-linear approximierbare quasikonforme Homöomorphismen n-dimensionaler Gebiete. Mitt. Math. Sem. Giessen **99** (1973)
3. Lehto, O., Virtanen, K.I.: Quasikonforme Abbildungen. Berlin-Heidelberg-New York: Springer 1965
4. Martio, O., Rickman, S., Väisälä, J.: Definitions for quasiregular mappings. Ann. Acad. Sci. Fenn. Ser. AI **448** (1969)
5. Opfer, G., Puri, M.L.: Complex planar splines. J. Approximation Theory **31**, 383–402 (1981)
6. Schultz, M.H.: Spline Analysis. Englewood Cliffs: Prentice Hall 1973

Received September 28, 1981

Math. Z. 180, 483–487 (1982)

Mathematische Zeitschrift

© Springer-Verlag 1982

Variation du nombre de Lefschetz dans une famille d'automorphismes analytiques

Marcos Sebastiani

21, rue des Frères Vaillant, F-59800 Lille, France

1. On suppose données deux variétés analytiques complexes M, S (S connexe) avec $\dim M = n \geqq 1$ et $\dim S = 1$, et une application analytique *propre* $f: M \to S$. On dit que $p \in M$ est un *point critique* si $df(p) = 0$ et on dit que $q \in S$ est une *valeur critique* si $q = f(p)$ où p est un point critique.

On supposera toujours que les points critiques sont isolés.

Il en résulte que f est surjective et ouverte.

Pour chaque $q \in S$ on notera N_q la fibre de q: $N_q = f^{-1}(q)$, et on notera Σ_q l'ensemble des points critiques contenus dans N_q.

Si X est un espace topologique du type d'homotopie d'un complexe fini et si $T: X \to X$ est une application continue, on notera $L(T, X)$ le *nombre de Lefschetz* de T:

$$L(T, X) = \sum_j (-1)^j \operatorname{Trace}(T_* | H_j(X, Q)).$$

(C'est un nombre entier.) La fonction ζ de T est:

$$\zeta(s) = \exp \sum_m \frac{L(T^m, X)}{m} s^m.$$

Soit $F: M \to M$ un automorphisme analytique tel que $f \circ F = f$. Si l'ensemble des points critiques est vide, f est une fibration topologique, par le théorème d'Ehresmann. Dans ce cas, $L(F, N_q)$, $q \in S$, ne dépend pas de q. Mais si cet ensemble n'est pas vide, $L(F, N_q)$ n'est plus constant: il change quand q passe par une valeur critique. Cette variation a été étudiée dans [3]. Dans le présent article, nous reprenons l'étude des invariants introduits dans [3] et nous montrons que la variation ne dépend que de la différentielle de F dans chaque point critique fixe par F. Nous donnons aussi une méthode pour le calcul explicite de la variation à partir de ces différentielles.

2. Soit Ω^k le faisceau de germes de k-formes analytiques sur M. Soit Ω_S^1 l'image réciproque par f du faisceau de germes de 1-formes analytiques sur S. On peut

0025-5874/82/0180/0483/$01.00

supposer $\Omega^1_S \subset \Omega^1$. Soient $p \in M$ et $q = f(p) \in S$. Posons:

$$V_p = \Omega^n_p / (\Omega^1_{S,p} \wedge \Omega^{n-1}_p).$$

Soient $z_1, \ldots, z_n$ des coordonnées locales de M en p (nulles en p) et soit ζ une coordonnée locale de S en q. Soit $g = \zeta \circ f$. Alors, on a un isomorphisme $V_p \cong \mathcal{O}_0(\mathbb{C}^n)/\mathcal{I}$, où $\mathcal{I}$ est l'idéal engendré par $\partial g/\partial z_1, \ldots, \partial g/\partial z_n$, qui envoie la classe de $h dz_1 \wedge \ldots \wedge dz_n$ dans la classe de h. En particulier, $\dim_{\mathbb{C}} V_p = \mu_p$, nombre de Milnor de f au point p.

Soit $\mathrm{Aut}_f(M)$ le groupe des automorphismes analytiques F de M qui laissent invariante f (c'est à dire, $f \circ F = f$). Fixons $F \in \mathrm{Aut}_f(M)$. Pour chaque $q \in S$ on a $F(\Sigma_q) = \Sigma_q$. Soit $p \in N_q$. L'application $\Omega^n_{F(p)} \to \Omega^n_p$ qui envoie ω dans $F^*(\omega)$ passe au quotient et définit una application $\mathbb{C}$-linéaire $F^*_p : V_{F(p)} \to V_p$. Soit $W_q = \sum_p V_p$ et soit:

$$\bigwedge_q(F) = \sum_p F^*_p : W_q \to W_q$$

(obsérvons que $V_p = 0$ si $p \notin \Sigma_q$). Alors $\bigwedge_q : \mathrm{Aut}_f(M) \to GL(W_q)$ est un antihomomorphisme. Pour chaque $F \in \mathrm{Aut}_f(M)$ on définit:

$$\sigma_q(F) = (-1)^{n-1} \mathrm{Trace}(\bigwedge_q(F)).$$

Pour le calcul effectif de $\sigma_q(F)$ on observe, d'abord, que $\sigma_q(F) = \sum_p \nu_p(F)$, où la somme est prise pour $p \in \Sigma_q(F)$ (ensemble des points fixes de $F | \Sigma_q$) et où $\nu_p(F) = (-1)^{n-1} \mathrm{Trace}(F^*_p)$. Ensuite, on voit que, modulo l'identification $V_p = \mathcal{O}_0(\mathbb{C}^n)/\mathcal{I}$, F^*_p est l'application induite par $h \to J(h \circ F_p)$, $h \in \mathcal{O}_0(\mathbb{C}^n)$, où $F_p : (\mathbb{C}^n, 0) \to (\mathbb{C}^n, 0)$ est l'application locale définie par F dans les coordonnées choisies et J est son jacobien (on peut aussi faire le calcul avec les séries formelles).

Théorème 1. *Soit $F \in \mathrm{Aut}_f(M)$. Alors,*

a) *$\sigma_q(F)$ est un entier pour tout $q \in S$;*
b) *$\sigma_q(F) = 0$ si $q \in S$ est une valeur régulière;*
c) *l'entier $\lambda_q(F) = \sigma_q(F) + L(F, N_q)$, $q \in S$, ne dépend pas de q.*

Démonstration. Ce théorème découle de ce qui précède et des résultats de [3] (voir aussi [4]).

Exemples. 1) Si on prend $F =$ identité, pour la caractéristique d'Euler-Poincaré (notée χ) on a:

$$\chi(N_q) = \chi(N_r) + (-1)^{n-1} \sum_{f(p)=r} \mu_p$$

si $q, r \in S$ et q est une valeur régulière.

2) Soit Π un groupe cyclique d'ordre premier π qui opère analytiquement sur M en laissant f invariante. Soit X l'ensemble des points fixes de cette action et soit $X_q = X \cap N_q$, $q \in S$. Supposons que $1 + \mu_p < \pi$ pour tout point critique p. Alors,

$$\chi(X_q) = \chi(X_r) + (-1)^{n-1} \sum \{\mu_p : p \in \Sigma_r \cap X\},$$

si $q, r \in S$ et q est une valeur régulière.

En effet, soit F un générateur de Π. Alors $\chi(X_q)=L(F,N_q)$ pour tout $q\in S$. En effet, N_q/Π est un espace analytique et $X_q\subset N_q/\Pi$ est un sous-ensemble analytique [1]. Par le théorème de triangulation [2], il existe un système fondamental de voisinages fermés V de X_q dans N_q/Π tels que X_q est rétract par déformation de V. Comme la projection canonique $P: N_q-X\to(N_q/\Pi)-X_q$ est un revêtement non-ramifié (puisque π est prime), les $U=P^{-1}(V)$ forment un système fondamentale de voisinages fermés de X_q dans N_q qui se rétractent par déformation sur X_q et qui sont invariants par l'action de Π. La suite exacte d'homologie donne $L(F,N_q)=L(F,U)+L(F;N_q,U)$ (où le nombre Lefschetz relatif est défini de façon analogue à l'absolu). Comme l'inclusion $X_q\to U$ est équivariante et induit isomorphisme en homlogie, on a: $L(F,U)=L(F,X_q)=\chi(X_q)$.

Soit U' un voisinage fermé de X_q dans N_q, invariante par Π, qui se rétracte par déformation sur X_q et contenu dans l'intérieur de U. Alors, $H_*(N_q,U)=H_*(N_q-U_0',U-U_0')$, où U_0' est l'intérieur de U', et cet isomorphisme est équivariant. Alors,

$$L(F;N_q,U)=L(F;N_q-U_0',U-U_0')=L(F,N_q-U_0')-L(F,U-U_0')=0$$

par excision, par la suite exacte d'homologie et parce que F n'a pas de point fixe dans N_q-X_q. Ceci complète la preuve de $\chi(X_q)=L(F,N_q)$ pour tout $q\in S$.

On en déduit que:

$$\chi(X_q)=\chi(X_r)+\sum\{\nu_p(F): p\in\Sigma_r\cap X\}$$

pour tout $q,r\in S$, q étant une valeur régulière, d'après le théorème 1 et ce qu'on a vu plus haut, étant donné que $\Sigma_r\cap X=\Sigma_r(F)$. Pour chaque $p\in\Sigma_r\cap X$, F_p^* est un automorphisme linéaire cyclique de V_p d'ordre $\pi>1+\mu_p=1+\dim_{\mathbb{C}}V_p$. Donc, F_p^* est l'identité. Alors, $\nu_p(F)=(-1)^{n-1}\operatorname{Trace}(F_p^*)=(-1)^{n-1}\mu_p$.

3. Nous allons maintenant montrer comment peut-on calculer $\nu_p(F)$ à partir de $dF(p)$.

Soit $p\in M$. Soit T_p l'espace tangent à M en p. Soit $GL_f(T_p)$ le groupe des automorphismes linéaires L de T_p qui sont de la forme $L=dH(p)$, où $H:(M,p)\to(M,p)$ est une application analytique locale qui laisse f invariante. Soit $(\mathcal{O}_p,\mathfrak{m}_p)$ l'anneau local de M en p. Par définition, V_p est un $\mathcal{O}_p$-module. Soit E_p le module gradué associé à la filtration $\mathfrak{m}_p$-adique de V_p.

Soit $L\in GL_f(T_p)$ et soit $H:(M,p)\to(M,p)$ une application analytique locale, qui laisse f invariante, avec $L=dH(p)$. Alors, l'application induite $H^*:\Omega_p^n\to\Omega_p^n$ passe au quotient dans V_p en laissant la filtration invariante et, donc, induit une application $\mathbb{C}$-linéaire $\tau(L):E_p\to E_p$. On verra qu'elle ne dépend pas du choix de H.

Considérons une application analytique locale *quelconque* $\tilde{H}:(M,p)\to(M,p)$ telle que $d\tilde{H}(p)=L$ ($\tilde{H}$ ne laisse pas forcément f invariante). Alors, pour toute $\omega\in\mathfrak{m}_p^k\Omega_p^n$ on *a*:

$$H^*(\omega)-\tilde{H}^*(\omega)\in\mathfrak{m}_p^{k+1}\Omega_p^n.$$

De là on déduit que $\tilde{H}^*$ induit une application $\mathbb{C}$-linéaire $E_p\to E_p$ et que cette application coïncide avec $\tau(L)$. En effet,

$$E_p=E_p^0\oplus E_p^1\oplus\ldots\quad\text{où}\quad E_p^k=\mathfrak{m}_p^kV_p/\mathfrak{m}_p^{k+1}V_p.$$

Pour définir l'application induite par $\tilde{H}^*$ il suffit de la définir dans chaque E_p^k. Tout élément de E_p^k se représente par une $\omega \in \mathfrak{m}_p^k \Omega_p^n$. À cet élément on associe la classe de $\tilde{H}^*(\omega)$ dans E_p^k. Mais celle-ci coïncide avec la classe de $H^*(\omega)$ dans E_p^k, d'après ce qu'on vient de voir. Cela montre que la classe dans E_p^k de $\tilde{H}^*(\omega)$ indépend du choix de ω et que l'application $E_p \to E_p$ ainsi définie coïncide avec $\tau(L)$.

En particulier on a que $\tau(L)$ est bien définie: elle ne dépend pas du choix de H.

On va utiliser ce qui précède pour le calcul explicite de $\tau(L)$ (à l'équivalence linéaire près). Soit $z_1, \ldots, z_n$ un système de coordonnées locales de M en p et soit ζ une coordonnée de S en $q = f(p)$. L'application linéaire L s'exprime, dans ces coordonnées, par une application linéaire $A: \mathbb{C}^n \to \mathbb{C}^n$. Soit $\varphi: \mathcal{O}_0(\mathbb{C}^n) \to \mathcal{O}_0(\mathbb{C}^n)$ l'application $\varphi(h) = (h^\circ A)\det(A)$. Soit E le gradué associé à la filtration $\mathfrak{m}_0(\mathbb{C}^n)$-adique du module $V = \mathcal{O}_0(\mathbb{C}^n)/I$ où I est l'idéal engendré par les dérivées partielles de $\zeta^\circ f$ dans les coordonnées choisies. Alors, φ induit une application $\mathbb{C}$-linéaire $E \to E$ qui s'identifie à $\tau(L)$ modulo l'isomorphisme évi dent $E_p \cong E$. Pour le voir, considérons l'application analytique locale $\tilde{H}: (M, p) \to (M, p)$ défine par $\tilde{H}(u) = \alpha^{-1}(A(\alpha(u)))$ où $\alpha(u) = (z_1(u), \ldots, z_n(u))$. Alors $d\tilde{H}(p) = L$ et on peut faire le calcul de $\tau(L)$ en utilisant $\tilde{H}$, comme on a montré plus haut. (On peut, aussi, remplacer les séries convergentes par les séries formelles.)

Théorème 2. *$\tau: GL_f(T_p) \to GL(E_p)$ est un anti-homomorphisme et*

$$\nu_p(F) = (-1)^{n-1} \operatorname{Trace}(\tau(dF(p)))$$

pour tout $F \in \operatorname{Aut}_f(M)$ tel que $F(p) = p$.

Démonstration. En effet, si $L = dF(p)$, on peut prendre $H = F$ pour calculer $\tau(L)$. Il suffit, alors, d'observer que la trace d'une application linéaire qui conserve une filtration, est égale à la trace de l'application induite dans le gradué associé.

Corollaire 1. *La différence des nombres de Lefschetz des restrictions de $F \in \operatorname{Aut}_f(M)$ à une fibre singulière et à une fibre régulière ne dépend que des points critiques fixes par F dans la fibre singulière et des 1-jets de F dans ces points.*

Corollaire 2. *Soient $F_1, F_2 \in \operatorname{Aut}_f(M)$. Supposons qu'il existe $r \in M$ tel que:*

a) *$F_1 | N_r$ et $F_2 | N_r$ ont la même fonction ζ;*
b) *$F_1 | \Sigma_r = F_2 | \Sigma_r$;*
c) *pour tout $p \in \Sigma_r$, $dF_1(p) = dF_2(p)$.*

Alors, $F_1 | N_q$ et $F_2 | N_q$ ont la même fonction ζ pour toute valeur régulière $q \in S$.

Démonstration. Par (b), les points critiques de la fibre N_r fixes par F_1^k sont les mêmes que ceux fixes par F_2^k ($k = 1, 2, \ldots$). En plus, par (c), les différentielles de F_1^k et F_2^k coïncident dans chacun de ces points. Le corollaire 2 résulte, alors, du corollaire 1.

Corollaire 3. *Soient* $F_1, F_2 \in \mathrm{Aut}_f(M)$. *Supposons qu'il existe* $r \in S$ *tel que* $F_1|N_r = F_2|N_r$. *Alors,* $F_1|N_q$ *et* $F_2|N_q$ *ont la même fonction* ζ *pour toute valeur régulière* $q \in S$.

Démonstration. Les conditions (a) et (b) du corollaire 2 sont évidemment vérifiées. La condition (c) découle du lemma suivant:

Lemme. *Soient* T_1, T_2: $(\mathbb{C}^n, 0) \to (\mathbb{C}^n, 0)$ *des applications analytiques locales telles que l'ensemble* $T_1 - T_2 = 0$ *contient une hypersurface ayant une singularité en* 0. *Alors,* $dT_1(0) = dT_2(0)$.

Démonstration. En effet, si $g = 0$ est une équation minimale de cette hypersurface, alors $T_1 - T_2 = hg$ avec h analytique en 0. Donc,

$$dT_1(0) - dT_2(0) = h(0)\,dg(0) + g(0)\,dh(0) = 0.$$

Exemple. Nous allons donner un exemple dans lequel il existe $r \in S$ tel que $F_1|N_r$ et $F_2|N_r$ sont homologiquement équivalentes et, pourtant, $L(F_1, N_q) \neq L(F_2, N_q)$ pour toute valeur régulière $q \in S$.

Soit X la sous-variété algébrique de $\mathbb{P}_1(\mathbb{C}) \times \mathbb{P}_2(\mathbb{C})$ définie par:

$$s(x^3 - y^2 z) - t z^3 = 0,$$

où (t, s) sont les coordonnées de $\mathbb{P}_1(\mathbb{C})$ et (x, y, z), celles de $\mathbb{P}_2(\mathbb{C})$. Le seul point singulier de X est (1, 0; 0, 1, 0).

Soit $\pi: \mathbb{P}_1 \times \mathbb{P}_2 \to \mathbb{P}_1$ la projection canonique, soit $S = \mathbb{P}_1 - \{(1, 0)\}$, soit $M = X \cap \pi^{-1}(S)$ et soit $f = \pi|M$. Alors, $f: M \to S$ est analytique et propre, M et S sont des variétés analytiques complexes, $\dim S = 1$, $\dim M = 2$ et S est connexe. Le seul point de M où df s'annulle est $p = (0, 1; 0, 0, 1)$. L'unique fibre singulière est homéomorphe à S^2 et les fibres non-singulières sont des courbes elliptiques.

Soient $F_1 =$ identité, $F_2(t, s; x, y, z) = (t, s; x, -y, z)$. Alors $F_1, F_2 \in \mathrm{Aut}_f(M)$. Si $r = (0, 1)$, $F_2|N_r$ induit l'identité en homologie, puisque N_r est la fibre singulière, homéomorphe à S^2. Donc, $F_1|N_r$ et $F_2|N_r$ sont homologiquement équivalentes.

Si $q \in S$ et $q \neq (0, 1)$, alors $L(F_1, N_q) = \chi(N_q) = 0$, mais $L(F_2, N_q) = 4$, puisque $F_2|N_q$ est la transformation qui envoie chaque élément de N_q dans son inverse pour une structure naturelle de groupe sur la courbe elliptique N_q.

Bibliographie

1. Cartan, H.: Quotient d'un espace analytique par un groupe d'automorphismes. In: Algebraic Geometry and Topology. A Symposium in Honor of S. Lefschetz (Princeton 1954), pp. 90–102. Princeton University Press 1957
2. Lojasiewicz, S.: Triangulation of semi-analytic sets. Ann. Scuola Norm. Sup. Pisa Sci. Fis. Mat. Ser. 3 **18**, 449–474 (1964)
3. Sebastiani, M.: Sur les points fixes des automorphismes des fibrés avec singularités. Math. Ann. **248**, 267–273 (1980)
4. Wall, C.T.C.: Notes on symmetry of singularities. Bull. London Math. Soc. **12**, 169–175 (1980)

Received November 24, 1981

Math. Z. 180, 489-501 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Locally Truncated Buildings and M_{24}

Mark A. Ronan

Mathematics Department, University of Illinois at Chicago Circle, Chicago, Illinois 60680, USA

The purpose of this paper is to classify the finite geometries belonging to the diagram ○=○—○—□, showing they all arise from buildings with but two exceptions, one associated with the group M_{24} as in [5], and one thin example arising from $L_3(2) \cdot 2$, both of which are described below. As is natural we assume an intersection condition ((Int) in Section 1), although this is shown to be unnecessary if a group is assumed. In order to understand the diagram above, take a building [7] of type B_4 and remove the stars of all vertices belonging to the square node. This we refer to as a *truncated building* (with the above diagram), and any geometry locally isomorphic to it we call a *locally truncated building* belonging to this diagram.

It is a pleasure to acknowledge helpful comments on this work by A.E. Brouwer, A.M. Cohen, W.M. Kantor, S.D. Smith and the referee.

1. Main Results, Definitions and Examples

By a *geometry* we understand an incidence system as in [3], two such being *locally isomorphic* if they have the same indexing set with residues of the same type being isomorphic. Under the term *building* we also include apartments [7] (i.e. thin buildings over the "field" of one element).

Throughout this paper Δ will denote a connected geometry (i.e. connected under the incidence relation in the graph theoretical sense) belonging to the diagram above, the nodes corresponding to sets of *points*, *lines*, and *quads*. The diagram is equivalent to the following conditions:

i) Each quad is a generalized quadrangle and contains all points on each of its lines.

ii) The residue of each point p, written Res(p) (i.e. the lines and quads containing the point), forms the points and lines of projective 3-space (we allow the case of two points per line - i.e. the complete graph on 4 points).

If p and q are collinear points they evidently lie on a unique common line, denoted pq, since any two concurrent lines lie in a quad by (ii) above.

0025-5874/82/0180/0489/$02.60

Moreover assuming our intersection condition (Int), the projective spaces Res(p) and Res(q) give the same residual plane structure to Res(pq) (see (2.1)), and hence are defined over the same field. By connectivity therefore Res(p) is independent of the point p, and we let k denote the common field, tacitly allowing the "field" of one element.

Examples. 1) As explained in the introduction, examples arise from buildings of type B_4.

2) Take the octads, trios and sextets of the Steiner system $S(24,5,8)$ as points, lines, and quads, as described in [5]. Here k is the field of two elements, each quad has parameters $(2,2)$ and the automorphism group is M_{24}.

3) Take 14 points corresponding to the points and lines of the projective plane of order 2, take 28 lines corresponding to the antiflags, and let the 21 quads be all 4-gons. Here k is the "field" of one element and the automorphism group is $SL_3(2)\cdot 2$.

4) Take the complete bipartite graph on the sets $\{a_1,\ldots,a_4\}$, $\{b_1,\ldots,b_4\}$, and let the quads be 4-gons $(a_i b_k a_j b_l)$ where $\{i,j\}$ and $\{k,l\}$ are either equal or disjoint. The automorphism group 2^3S_4 is the quotient of the Weyl group $W(B_4)$ by its central involution.

For any point p of Δ, Res(p) may be regarded as a polar space (of type D_3) whose points are the quads containing p. If a set of quads forms a (projective) subspace of such a Res(p), then we shall call it a *quad subspace.* In other words a quad subspace is the set of quads of some Res(p) incident with a line L, a plane $\pi_p \in$ Res(p) or an incident (L,π_p) pair (a plane here is understood in the usual sense as a plane of the $PG_3(k)$ of Res(p)).

We may now state our intersection condition (cf. Buekenhout [3] and Tits [8; Sect. 6]).

(Int) Suppose X and Y are quad subspaces of, or all quads of, Res(x) and Res(y) respectively, where $x \neq y$. Then $X \cap Y$ is either the null set or is a quad subspace of both Res(x) and Res(y).

The reader may observe that examples 2 and 3 satisfy (Int) whilst example 4 does not. However example 4 has as a 2-fold cover the (truncated) tesseract (i.e. B_4 Weyl complex). The reader may also care to check that Example 3 is not a subgeometry of Example 2. We now state our main results.

Theorem 1. *A finite geometry belonging to the diagram* ○═○—○—□ *and satisfying* (Int) *is either a truncated building or arises from* M_{24} *or* $L_3(2)\cdot 2$ *as in examples* 2 *and* 3 *above.*

Before stating our corollary let us say that the automorphism group of a finite geometry Δ belonging to the above diagram is *locally of Lie type* if the group induced on the residue of each point (resp. line, resp. quad) contains $A_3(q)$ (resp. $A_1(q^e)\times A_2(q)$, resp. one of $B_2(q)$, $C_2(q)$, ${}^2A_3(q)$, ${}^2A_4(q)$, $e=\frac{1}{2},\frac{2}{3},1,\frac{3}{2},2$ as appropriate) as a normal subgroup; we include the case $q=1$, by which we understand the appropriate Weyl group.

Corollary. *Any finite geometry belonging to the diagram above whose automorphism group is locally of Lie type arises from one of* $B_4(q)$, $C_4(q)$, ${}^2A_7(q)$, ${}^2A_8(q)$, M_{24}, $L_3(2)\cdot 2$ *or* $B_4(1)/2$ (*Example* 4).

A Remark on M_{24}. The M_{24} geometry mentioned above arises from the maximal 2-local subgroups as explained in [5]. If however one starts with a Sylow 2-subgroup B (they are self-normalizing) and takes the three 2-local subgroups minimal with respect to containing B, then one obtains a slightly different geometry $\tilde{\Delta}$, one with diagram $\circ\tilde{=}\circ-\circ$, where $\circ\tilde{=}\circ$ denotes the triple cover of the $Sp_4(2)$ quadrangle, with automorphism group $3S_6$. This geometry $\tilde{\Delta}$ enters our work in section 3; see also section 7.

To prove the main theorem we proceed as follows. In section 2 we obtain a substructure in Res(x) for any point distant 3 from x, and we show this is always either a plane or the $Sp_4(k)$ quadrangle. In section 3 we use the covering $\tilde{\Delta}$ to show that the planar case leads buildings (no restriction on finiteness so far), and in section 4 we show that the finite non-planar case leads to quad parameters (1, 1) or (2, 2). The latter case gives rise to the Steiner system $S(24, 5, 8)$ in Sect. 5 by reconstructing the 24 symbols as sets of points in Δ. Section 6 shows that a group implies (Int) is satisfied or else one has example 4 above.

Standing Hypothesis. With the exception of Sect. 6, Δ is assumed to satisfy (Int).

2. Some Geometrical Properties

For any two points x and y we shall let $d(x, y)$ denote the distance between x and y in the collinearity graph. If $d(x, y)=3$ we shall show that the substructure in Res(x) induced by y (see (2.8)) is a plane or the $Sp_4(k)$ quadrangle, and this is independent of the pair (x, y).

(2.1) **Lemma.** *If p and q are collinear points and π_p is a plane of* Res(p) *containing pq, then the set of quads of π_p which contain pq equals the set of quads containing pq in a plane π_q of* Res(q). *In particular the quads containing pq form the points of a projective plane whose lines are the subsets of such quads in the planes π_p above.*

Proof. Let X be the set of quads in π_p containing pq. Clearly $X = X \cap \mathrm{Res}(q)$, and since X is neither a singleton nor the set of all quads on pq, it follows by (Int) that X is the set of quads on pq in a (unique) plane π_q of Res(q).

(2.2) **Lemma.** *If p and q are collinear points lying in a quad Q, then $pq \in Q$.*

Proof. By (Int) the quads of $\mathrm{Res}(p) \cap \mathrm{Res}(q)$ form a subspace of Res(p), containing (hence equal to) the maximal subspace of all quads containing pq. The result follows.

(2.3) **Lemma.** *There are no triangles in Δ.*

Proof. If p, q, r are the points of a triangle and P is the quad containing pq and pr, then by (2.2) $qr \in P$, a contradiction since P can contain no triangles.

(2.4) **Lemma.** *If p and q are points with $d(p, q)=2$, then they lie in a unique quad $Q(p, q)$.*

Proof. If p and q lie in two quads Q, Q' then by (Int) these quads lie in a subspace of Res(p), hence share a line L on p. Let q be collinear with points r, r' of L in Q, Q' respectively. By (2.3) $r=r'$, so $Q=Q'$ in Res(r).

(2.5) **Lemma.** *If p is collinear with two points x and y of a quad Q, then $p \in Q$.*

Proof. If x and y are collinear, then by (2.3) and (2.2) $p \in xy \in Q$. If x and y are not collinear, then the quad of Res(p) containing px and py must be Q by (2.4).

(2.6) **Lemma.** *Given collinear points p, q contained in quads P, Q respectively which do not contain the line $L=pq$, then there exists a quad R containing L and meeting each of P and Q in a line.*

Proof. In Res(p) (resp. Res(q)) let π_p (resp. π_q) denote the plane spanned by L and P (resp. Q). Then in Res(L), π_p and π_q meet in a quad R by (2.1). Clearly R meets P in π_p, so R and P share a line of Δ; similarly for R and Q.

(2.7) **Lemma.** *Every 5-gon lies in a quad. In particular a point distant 2 from two points of a line is collinear with a unique point of that line.*

Proof. Let (x, a, p, q, b) be a 5-gon. Set $P=Q(p, x)$, $Q=Q(q, x)$, and let R be the quad given by (2.6). Let y (resp. z) be the point collinear with x on a line through p (resp. q) in $R \cap P$ (resp. $R \cap Q$). Clearly $y \neq z$ by (2.3), so $x \in R$ by (2.5). Therefore the 5-gon lies in R.

Given points x, y with $d(x, y)=3$ we let x^y denote the substructure of Res(x) comprising quads, respectively lines, containing a point collinear with, respectively distant 2 from, y.

(2.8) **Lemma.** *With the above notation x^y is either a plane or the $Sp_4(k)$ generalized quadrangle in which case each line on x lies in x^y, moreover x^y is a plane if and only if y^x is.*

Proof. If L is a line of x^y, let p be the point of L distant 2 from y; then each line M of $Q(p, y)$ through p determines with L a quad of x^y, and all quads of x^y containing L are obtained in this way, by (2.7). Now considering Res(p) we see by (2.1) that these quads are precisely the quads containing L in the plane of Res(p) spanned by L and $Q(p, y)$; and since $x \in L$ we therefore see that these quads lie in a plane of Res(x). This means that x^y, regarded as a set of points and lines in the $PG_3(k)$ of Res(x), has the property that the lines through a point L comprise all lines on L in a plane on L. Moreover if L is a line on x in a quad of x^y then L is also in x^y, so all points on a line of x^y in $PG_3(k)$ are in x^y. It follows that x^y is a plane or a generalized quadrangle $\mathfrak{Q}$ of the $PG_3(k)$ on x. In the latter case, since the lines of $\mathfrak{Q}$ on a point X of $\mathfrak{Q}$ comprise all lines of a plane of $PG_3(k)$ on X, it easily follows that all points of $PG_3(k)$ lie in $\mathfrak{Q}$; in other words all lines on x lie in x^y, a fact we shall use repeatedly. In this case one immediately has a polarity of $PG_3(k)$, and $\mathfrak{Q}$ is the $Sp_4(k)$ quadrangle (see for example [7; 8.6 (I), 8.3.2 and 8.1.4]). For the second statement of the theorem it is easy to see using (2.7) that x^y contains a triangle of $PG_3(k)$ if and only if y^x does.

(2.9) **Lemma.** *Suppose y and y' are collinear points distant 3 from x. Then x^y is a plane if and only if $x^{y'}$ is.*

Proof. It clearly suffices to prove one implication, so suppose x^y is not a plane. Then y^x is not a plane either so x is distant 2 from a point p of yy'. Let L be any line through x not in $Q(x,p)$ and let $q \in L$ be distant 2 from y, which exists because x^y is not a plane. Now setting $P = Q(x,p)$, $Q = Q(y,q)$ we obtain a quad R containing L as in (2.6), and we let r,s be the points of $R \cap P$, $R \cap Q$ respectively collinear with p,y. Clearly r and s are collinear, otherwise $x, y \in Q(r,s)$ by (2.7), and so y' is distant 2, via a point of rs, with a point of L. Thus $x^{y'}$ is not a plane, as required.

(2.10) **Lemma.** *Let (x,p,s,y,q,r) form a hexagon, with $d(x,y) = d(p,q) = d(r,s) = 3$. Then x^y is a plane if and only if p^q is.*

Proof. We need only prove one implication, so suppose x^y is a plane, and thus $Q(p,r)$ contains a point collinear with y. This point must be collinear with p and r, otherwise we would obtain a 5-gon, showing that x and y lie in a quad by (2.7). Similarly since y^x is a plane we obtain a point collinear with x,s and q. We have thus constructed the graph of a cube, and it is clear that p^q contains a triangle, and must therefore be a plane by (2.8).

(2.11) **Corollary.** *The x^y subgeometries of $PG_3(k)$ are either all planes or all $Sp_4(k)$ quads.*

Proof. It suffices to show that if $d(x,y) = d(x,y') = 3$, then x^y and $x^{y'}$ are of the same type, so take a path from y to y'. If x^y is not a plane then every line through y contains a point at distance 2 from x, hence by repeated application of (2.9) and (2.10) we see that $x^{y'}$ is not a plane.

In the case that all x^y are (resp. are not) planes we shall say that Δ is of *planar type* (resp. *non-planar type*), and we shall show in (3.6) that the former are precisely those arising from buildings.

(2.12) **Proposition.** *If Δ is of non-planar type, then it is a near hexagon (defined in* [6]*).*

Proof. Clearly the diameter is 3 so we need only show that for a point p and a line L there is a unique point of L nearest p, but this follows from (2.3), (2.7) and the fact that non-planarity implies every line containing a point distant 3 from p contains a point distant 2 from p.

3. Planes and Coverings

Let p and q be collinear points, and let π_p be a plane of Res(p) containing pq. As in (2.1) we obtain a plane π_q of Res(q), which we call tha plane at q *induced* from π_p

(3.1) **Lemma.** *Let p and r be points with $d(p,r) = 2$ and suppose π_p is a plane in* Res(p) *containing $Q = Q(p,r)$. Let q, q' be points collinear with both p and r, let π_q, $\pi_{q'}$ be the planes induced from π_p, and let π_r, π'_r be the planes induced at r from π_q, $\pi_{q'}$ respectively. Then $\pi_r = \pi'_r$ if and only if Δ is of planar type.*

Proof. Take quads P, P' containing pq, pq' respectively and forming a triangle with Q in π_p. Let $x \neq p$ lie on the line $P \cap P'$, and notice that $d(r,x) = 3$ by (2.7).

Now take quads $R, R' \neq Q$ of $\pi_q, \pi_{q'}$ respectively containing qr, $q'r$, and notice that $R, R' \in r^x$ because R, R' meet P, P' respectively in lines which of course each contain a point collinear with x; moreover $R \in \pi_r$ and $R' \in \pi'_r$.

Now if $\pi_r = \pi'_r$, then R meets R' in a line of r, so r^x contains the triangle (Q, R, R') and hence is a plane. Conversely if r^x is a plane, then since Q and R (resp. R') generate π_r (resp. π'_r) we have $\pi_r = r^x = \pi'_r$.

(3.2) *The Covering.* Let $\tilde{\Delta}$ denote the geometry whose points and lines are pairs (p, π_p) and (L, π_L) respectively with π_p, π_L planes of Res(p), Res(L), and where incidence is given by $p \in L$ and $\pi_p = \pi_L$ in Res(L). We let $\varphi: \tilde{\Delta} \to \Delta$ denote the projection map sending (x, π_x) to x. If Q is a quad we let $\tilde{Q}$ denote the subgeometry of $\tilde{\Delta}$ comprising all (x, π_x) such that $x \in Q$ and $Q \in \pi_x$, and we let $\varphi: \tilde{Q} \to Q$ denote the restriction of φ to $\tilde{Q}$.

(3.3) **Lemma.** *The map $\varphi: \tilde{Q} \to Q$ is a covering, each fibre being parametrised by $PG_1(k)$. Moreover if Δ is of planar type, then $\tilde{Q}$ is a disjoint union of copies of Q, and if Δ is of non-planar type, then $\tilde{Q}$ contains no 4-gons.*

Proof. The first statement is obvious from the above construction since for any $x \in Q$ the planes π_x of Res(x) which contain Q are the planes of a $PG_3(k)$ containing a line. Now if Δ is of planar type, then it follows from (3.1) that any circuit in Q lifts to a circuit in $\tilde{Q}$ since all circuits in Q are composed of 4-gons; this proves the second statement. Conversely if Δ is of non-planar type then (3.1) shows that no 4-gon of Q lifts to a circuit in $\tilde{Q}$, so $\tilde{Q}$ contains no 4-gons.

(3.4) **Lemma.** *Any circuit of points and lines in Δ may be split into 4-gons.*

Proof. Let $(x_0, \ldots, x_n)$ be the points of a circuit, where we may assume of course that $d(x_i, x_{i+2}) = 2$. If $n = 4$, then the result follows from (2.7), so assume $n \geqq 5$. Now by (2.6) the quads $Q(x_0, x_2)$ and $Q(x_3, x_5)$ meet some quad R on $x_2 x_3$ in lines L and M. Let y, z be the points of L, M respectively collinear with x_0, x_5. Since $d(y, z) \leqq 2$, we have a chain of length 3 or 4 from x_0 to x_5 via y, z, so the result follows by induction.

(3.5) **Corollary.** *Regarded as a simplicial complex whose vertices are its points, lines and quads, Δ is simply-connected.*

Proof. Any 4-gon may be contracted to the vertex in whose residue it lies, and any circuit is homotopic to one of points and lines so the result follows from (3.4).

(3.6) **Proposition.** *If Δ is of planar type, then it is a truncated building.*

Proof. Let C be a connected component of $\tilde{\Delta}$, and consider $\varphi(C) \subset \Delta$. By (3.4) any circuit in $\varphi(C)$ based at a point p can be split into 4-gons, and consequently by (3.3) we see that this circuit lifts to a circuit in $\tilde{\Delta}$, and hence in C. In other words to each point p of $\varphi(C)$ there is associated a unique plane π_p with $(p, \pi_p) \in C$, and we extend Δ to a geometry $B\Delta$ by adding a vertex v (belonging to the square node) for each set $\varphi(C)$, taking incidence in the obvious manner.

1) For any vertex x of $B\Delta$, Res(x) is a building.

If x is a point, line, or quad this is obvious. If $x=v$ is a new vertex, then Res(v) is clearly a B_3 geometry and we need only check (LL) and (O) of [8]; (LL) is satisfied since it easily follows from (2.2) that two distinct lines can lie in at most one quad, and (O) is satisfied since by (2.2) and (2.4) the set of quads containing two distinct points p and q cannot form a plane of Res(p) or Res(q).

2) $B\Delta$ is simply-connected by the same proof as (3.5).

Now by 1) it follows that the universal cover of $B\Delta$ (in the sense of [8]) is unramified, and by 2) that it is $B\Delta$ itself. Moreover the B_3 residues are buildings, so it follows by [8] that $B\Delta$ is a building, as required.

4. The Finite Non-planar Case

In this section Δ is finite and of non-planar type; our field k contains q elements.

(4.1) **Lemma.** *Each quad has parameters* (q, q).

Proof. Considering the $PG_3(q)$ in the residue of a point, it is clear that the number of lines per point in a quad is $q+1$. Now consider a point p and a quad Q no point of which is collinear with p. By (2.12) each line of Q contains a point distant 2 from p, so the set of points of Q distant 2 from p forms an oval in Q. If $q=1$ this immediately implies a constant number of points per line, so in any case we have parameters (s, q), and each oval contains $1+sq$ points. Moreover the points of the oval give quads in Res(p) forming a set of mutually skew lines in $PG_3(q)$; since there can be at most $1+q^2$ such lines we see that $s \leqq q$.

Now fixing p we count the number of quads in the above relationship to p; it is

$$s^2q^5(q^2+1)(q^2+q+1)/(1+sq).$$

This being an integer implies

$$(1+sq) \mid (q^2+1)(q^2+q+1). \tag{1}$$

Moreover for any generalized quadrangle one knows [4] that $(s+q) \mid sq(1+sq)$. Now assume $s<q$; since q is a prime power this implies

$$(s+q) \mid s(1+sq). \tag{2}$$

Moreover

$$-q \equiv s \bmod (s+q). \tag{3}$$

Hence

$$s \equiv s^3 \bmod (s+q). \tag{4}$$

Using (1) and (2) we obtain

$$(s+q) \mid s(q^2+1)(q^2+q+1)$$

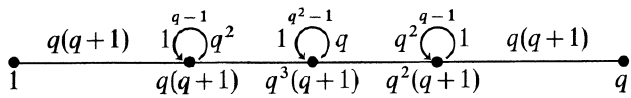
which implies, using (3) and (4) that

$$s+q \mid 6s.$$

With $s<q$ this implies $s=p^n$, $q=p^{n+1}$ where $p=2$ or 5, but this then contradicts (1), and so $s=q$ as required.

(4.2) **Proposition.** *Assuming Δ to be of finite non-planar type, the quads have parameters* (1, 1) *or* (2, 2).

Proof. We consider $\tilde{Q}$ as in (3.2). Given points $\tilde{p}=(p,\pi_p)$, $\tilde{q}=(q,\pi_q)$ of $\tilde{Q}$ with $d(p,q)=2$ in Q, it easily follows from (3.1) that $d(\tilde{p},\tilde{q})=2$. It follows that the point-graph of $\tilde{Q}$ is metrically regular of diameter 4 with parameters



or in other words intersection matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ q(q+1) & q-1 & 1 & 0 & 0 \\ 0 & q^2 & q^2-1 & q^2 & 0 \\ 0 & 0 & q & q-1 & (q+1)q \\ 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

One checks that a root α of $x^2-(q-1)x-(q+1)q$ is an eigenvalue with eigenvector $(1, \alpha/q(q+1), 0, -\alpha/q^2(q+1), -1/q)$. In the collinearity matrix for the point-graph, α therefore has multiplicity (see [1; p. 143])

$$\begin{aligned}&(q+1)^2(q^2+1)/(1+q(q+1)[\alpha/q(q+1)]^2+q^2(q+1)[\alpha/q^2(q+1)]^2+q(1/q)^2)\\ &\quad=q^2(q+1)^2(q^2+1)/(q^2+q+\alpha^2)\\ &\quad=q^2(q+1)^2(q^2+1)/(2q(q+1)+(q-1)\alpha).\end{aligned}$$

Therefore if $q\neq 1$, α must be rational, and since $\alpha^2-(q-1)\alpha-(q+1)q=0$ this implies $5q^2+2q+1$ is a perfect square. Since q is a prime power this implies $q=2$. Hence $q=1$ or 2 as required.

Notice that the proof above shows the non-existence of certain covers of (q,q) quadrangles. Proposition (4.2) may also be proved by considering the intersection matrix for the point-graph of Δ as shown to me by A.M. Cohen, to whom I am grateful for an explanation of eigenvalue techniques.

5. The M_{24} Case

This section deals with the non-planar case in which the quads have parameters (2, 2). By directly reconstructing the 24 symbols of the Steiner system we shall show that this case must arise from M_{24} as in Example 2 of section 1. An alternative proof of this fact has been communicated to me by Brouwer [2].

We now proceed to construct 24 sets S such that each line meets S in a unique point. First notice that any quad Q having parameters (2, 2) is isomor-

phic to the $Sp_4(2)$ quadrangle, and since $S \cap Q$ is clearly an oval $\mathcal{O}$ of 5 points one easily checks the following.

(5.1) Any two opposite points of Q lie in a unique oval.

(5.2) Any three points not in $\mathcal{O}$ collinear with a common point not in $\mathcal{O}$ form a hyperbolic line.

(5.3) Given two perpendicular hyperbolic lines, an oval meets one in 0 points and the other in 2 points. In particular a hyperbolic line meets an oval in 0 or 2 points.

Notation. For points p, q with $d(p,q)=2$, let $p*q$ denote the set of points collinear with both; it is a hyperbolic line of $Q(p,q)$. For a point x, let $\hat{x}$ denote the set of points collinear with x together with the lines and hyperbolic lines contained therein.

(5.4) **Lemma.** *For any point x, $\hat{x}$ forms the points and lines of $PG_4(2)$.*

Proof. We need only check Pasch's axiom, so take hyperbolic lines $H=x*a$, $K=x*b$ of $\hat{x}$ meeting in p and not lying in a quad. Clearly $d(a,b)=2$ and we let $p \neq y \in a*b$, so $d(x,y)=3$. Now suppose I and J are hyperbolic lines of $\hat{x}$ meeting H and K in four distinct points, and let $Q(I)$, $Q(J)$ denote the unique quads containing I, J respectively. Everything may be embedded in a plane of $\mathrm{Res}(x)$, so the third points of I and of J lie on a line L in $\mathrm{Res}(x)$, and we let q denote the point of L distant 2 from y. Since $Q(x,a)$, $Q(x,b) \in x^y$, and x^y contains no triangles, it follows that $Q(I)$, $Q(J) \notin x^y$, and therefore the points of $Q(I)$ and $Q(J)$ distant 2 from y form ovals. Since q lies in these ovals we conclude from (5.3) that $q \notin I, J$, and hence I meets J in the third point of L different from x and q. For lines containing x, or if H and K lie in a quad, then Pasch's axiom follows immediately since everything embeds in a quad.

We now fix a point x and construct $S=S_1 \cup S_2 \cup S_3$ where $S_i \subset \{\text{points distant } i \text{ from } x\}$.

(5.5) In the $PG_4(2)$ of $\hat{x}$ pick a hyperplane H not containing x, and set

$$S_1 = \{s \mid d(s,x)=1, s \notin H\}.$$

By (5.1) and (5.2), for any quad Q on x, $S_1 \cap Q$ is consistent with a unique oval $\mathcal{O}(Q)$ in Q. Set

$$S_2 = \{s \mid d(s,x)=2, s \in \mathcal{O}(Q(x,s))\}.$$

(5.6) Before constructing S_3 we check that no three points of S_2 lie on a

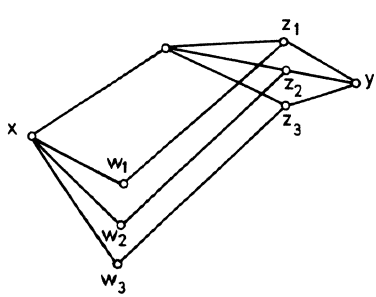


Fig. 1

hyperbolic line of a quad Q. Were this so we would have the situation illustrated in Fig. 1 where each $z_i \in S$, and consequently each $w_i \notin S$.
Since the $Q(x, z_i)$ lie in a plane of Res(x), we may pick the lines xw_i to lie in a quad R. Since x^y contains no triangle, $R \notin x^y$, so the w_i, being distant 2 from y, lie on an oval. However since each $w_i \notin S$, the w_i lie in the hyperplane H and hence form a hyperbolic line, contradicting (5.3).

(5.7) To construct S_3, consider the quads Q containing a unique point w collinear with x, $w \notin S_1$. By (5.1), (5.3) and (5.6) the three points of $S_2 \cap Q$ are consistent with a unique oval $\mathcal{O}(Q)$ in Q. If $d(x, y) = 3$, the points of x^y distant 2 from y cannot all lie in S_1, so y lies in one of the above quads. Now set

$$S_3 = \{y \mid d(x, y) = 3, y \in \mathcal{O}(Q), Q \text{ as above}\}$$

(5.8) Any two of the above quads containing y agree as to whether or not $y \in S_3$, otherwise we have the situation shown in Fig. 2, where points of S are circled.

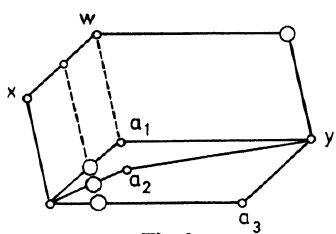


Fig. 2

Using (2.4) w may be chosen collinear to a_i for some i, and since clearly $w \notin S$, we see that x lies in the oval $Q(x, a_i) \cap S$, a contradiction.

Notice now that any line lies in one of the quads of (5.5) or (5.8) and these meet S in an oval, so the following is immediate.

(5.9) Every line of Δ contains a unique point of S.

Each set S will now be called a *symbol.*

The S we constructed is uniquely determined by the choice of hyperplane H in (5.5), so there are 16 symbols not containing x, and 8 symbols containing a given point. Thus there are 24 symbols in all. Moreover a quad Q on x gives a hyperbolic line in H, so there are 4 symbols agreeing on Q. These symbols clearly agree on no other point collinear with x, and hence with no other point collinear with a point of Q, so one easily checks the following.

(5.10) Four symbols agreeing on a quad have no further points in common.

Fixing a point we count $35 + 840 + 896 = 1771$ quads (see (2.5) and the proof of (4.1)), and since each quad contains 6 ovals and $6.1771 = \binom{24}{4}$ we see that any four symbols agree on exactly one quad. Any fifth symbol meets this quad in an oval meeting the given oval in one point, so we have proved:

(5.11) Any five symbols have exactly one point in common.

We have therefore constructed the Steiner system $S(24,5,8)$ on the 24 symbols, where the octads are the points of Δ. Since two collinear points cannot belong to the same symbol, the corresponding octads are disjoint, and so the lines of Δ are trios of mutually disjoint octads. Each quad contains six ovals each of which gives four symbols, so the quads are the sextets, and we have proved that Δ is the M_{24} geometry of Example 2 in Section 1.

6. Groups

In this section we assume Δ admits a group satisfying the conditions of the Corollary, and we show that either Δ satisfies (Int) or $q=1$ and Example 4 of Sect. 1 is the only possibility. Firstly notice that finiteness and connectivity imply the number of lines per point in any quad is a constant $(q+1)$, and the group assumption implies the number of points per line is a constant $(s+1)$ in each quad, and hence a constant globally.

(6.1) Since the quads arise from Chevalley groups they have parameters (s,q) where $s=q^e$, $e=\frac{1}{2}, \frac{2}{3}, 1, \frac{3}{2}$ or 2.

Define two quads to be *skew* if they share a point p and represent skew lines of $\mathrm{Res}(p)$.

(6.2) **Lemma.** *The stabilizer S of a quad Q and a point $p \in Q$ is transitive on points x collinear with p and not in Q. Moreover S_x is transitive on quads containing px and skew to Q.*

Proof. Fix a line L on p and notice that in the $L_3(q)\times L_2(s)$ thereon, the subgroup of the second factor fixing p must fix the projective space $\mathrm{Res}(p)$ pointwise since it commutes with the $L_3(q)$ fixing L. Thus S_L is transitive on $L\setminus\{p\}$, and as S is obviously transitive on the set of lines on p not in Q, the first statement follows. Moreover in the $L_3(q)$ above, the subgroup fixing Q is a section of S_x, so the second statement follows.

(6.3) **Lemma.** *If x and y are collinear points of a quad Q, then $xy \in Q$.*

Proof. If false, then by virtue of the group induced on Q, x is collinear with all points of Q. Therefore by (6.2) the set $Q\setminus\{x\}$ of $s(1+q+sq)$ points is equal to the set of $s(q^3+q^2+q+1)$ points collinear with p, contradicting (6.1).

(6.4) **Corollary.** *Any two distinct quads meeting in a line have no further points in common.*

(6.5) Let x, y, X, Y be as in (Int). If x and y are collinear, let Z denote the set of quads containing xy. By (6.3) $X\cap Y\subset Z$, and since the $L_3(q)$ induced on Z may be regarded as a section of both the $L_4(q)$ induced on $\mathrm{Res}(x)$ and on $\mathrm{Res}(y)$, it follows that $X\cap Z$ and $Y\cap Z$ are subspaces of Z (regarded as the points of $PG_2(q)$ via $L_3(q)$) and hence of $\mathrm{Res}(x)$ and $\mathrm{Res}(y)$. Thus $X\cap Y$ is also a subspace, as required by (Int). Now if $d(x,y)\geqq 3$, it follows that $X\cap Y=\varnothing$, so if (Int) is not satisfied we have at least two quads containing two points (hence any two points) at distance 2. Assume therefore that (Int) is not satisfied and that $d(x,y)=2$.

(6.6) **Lemma.** *With the above assumption, x and y lie in exactly two quads.*

Proof. Let Q be a quad containing x and y, and let $H \cong SL_2(q)$ fix Q, x and y, and act naturally on the $q+1$ lines of Q on x. In Res(x) consider the action of H on the lines of $PG_3(q)$; it fixes exactly one other line $Q' \neq Q$, but for any other $R \neq Q, Q'$ there is some $h \in H$ such that R and R^h share a point, i.e. a line on x. If R contains y so does R^h, which by (6.4) is impossible. Therefore only Q and Q' can contain y, proving the lemma.

(6.7) **Proposition.** *If* (Int) *is not satisfied then* $q = 1$.

Proof. Let x and y be points with $d(x, y) = 2$. Assume $q \neq 1$, so there is a quad Q in Res(y) meeting exactly one of the two quads given by (6.6), and we let p denote the unique point of Q collinear with both x and y. Since (Int) is not satisfied there are points collinear with at least two points of a quad (see (6.5)), hence by (6.2) some line $L \neq px$ on x meets Q. Consider Res(x) and notice by (6.4) that the quad containing L and px is skew to Q in Res(p), hence by (6.2) each of the q^2 quads on px skew to Q contain a line on x meeting Q, so we obtain a set A of at least q^2+1 points of Q collinear with x. Now pick a point $z \in py$ and let $p', p'' \in A$ be collinear with z. Since the quads containing respectively $\{px, pz\}$, $\{p'x, p'z\}$ and $\{p''x, p''x\}$ are distinct, it follows by (6.6) that one of p' or p'' equals p, and moreover if $z = y$ or p, then $p' = p'' = p$. Therefore $q^2 + 1 \leqq |A| \leqq s$, contradicting (6.1).

It is now a simple matter to check that Example 4 of Sect. 1 is the only one possible.

7. Conclusion

Proof of the Theorem. Using (3.6), (4.2) and the construction given in Sect. 5, we are left with the case of quads having parameters (1, 1) in Proposition (4.2). It is easy to check that Example 3 of Sect. 1 is the only possibility here.

Proof of the Corollary. This follows from Theorem 1 and Sect. 6.

Finally we make one last comment on the geometry $\tilde{\Delta}$ constructed in (3.2). In the non-planar case, by adding the covering quads $\tilde{Q}$ we obtain a rank 3 geometry which in the M_{24} case is that obtained via "minimal 2-parabolic" subgroups as mentioned in Sect. 1. This and other geometries obtained in a similar fashion will be discussed in a forthcoming paper by Stroth and the author.

References

1. Biggs, N.L.: Algebraic Graph Theory. Cambridge: Cambridge University Press 1974
2. Brouwer, A.E.: The Uniqueness of the Near Hexagon on 759 points. In: Proceedings of a Conference (Pullman 1981) (to appear)
3. Buekenhout, F.: Diagrams for Geometries and Groups. J. Combinatorial Theory Ser. A**27**, 121–151 (1979)

4. Higman, D.G.: Invariant Relations, Coherent Configurations and Generalized Polygons. In: Combinatorics part 3: Combinatorial Group Theory (M. Hall; J.H. van Lint, eds.). Proceedings of an Advanced Study Institute (Breukelen 1974), pp. 27–43. Amsterdam: Mathematisch Centrum 1974
5. Ronan, M.A., Smith, S.D.: 2-Local Geometries for some Sporadic Groups. In: Proceedings of Symposia in Pure Mathematics **37**. The Santa Cruz Conference on Finite Groups (Santa Cruz 1979), pp. 283–289. Providence, Rhode Island: American Mathematical Society 1980
6. Shult, E.E., Yanushka, A.: Near n-gons and Line Systems. Geometriae Dedicata **9**, 1–72 (1980)
7. Tits, J.: Buildings of Spherical Type and Finite BN Pairs. Lecture Notes in Mathematics **386**. Berlin-Heidelberg-New York: Springer 1974
8. Tits, J.: A Local Approach to Buildings. In: Proceedings of the Coxeter Symposium (Toronto 1979) (to appear)

Received October 23, 1981; received in final form January 18, 1982

Math. Z. 180, 503–523 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Primitivity in Differential Operator Rings

Kenneth R. Goodearl[1] and Robert B. Warfield[2], Jr.

[1] Department of Mathematics, University of Utah, Salt Lake City, Utah 84112, U.S.A.
[2] Department of Mathematics, University of Washington, Seattle, Washington 98195, U.S.A.

In this paper we develop necessary and sufficient conditions for a differential operator ring (Ore extension) $R[\theta; \delta]$, over a commutative noetherian ring R with a derivation δ, to be a simple ring, a primitive ring, or a Jacobson ring. The main purpose is to study the primitivity of $R[\theta; \delta]$, and to obtain information about the kinds of simple modules this ring can have.

In previous work of Jordan, two different sufficient conditions for $R[\theta; \delta]$ to be primitive were found, for any right noetherian differential ring R. Part of our result is that his conditions are also necessary when R is commutative (i.e., one or the other must hold in order for $R[\theta; \delta]$ to be primitive). We also exhibit explicit faithful simple modules corresponding to each of these conditions (unlike Jordan's results, which in the second case only provide an existence theorem), and we obtain as much information as possible about the simple modules over $R[\theta; \delta]$.

The Jacobson condition was also studied by Jordan, who proved that whenever a right noetherian differential ring R is a Jacobson ring, then $R[\theta; \delta]$ is a Jacobson ring as well. However, easy examples show that it is not necessary for R to be Jacobson in order for $R[\theta; \delta]$ to be Jacobson. In the commutative case, we identify a certain family (possibly empty) of factor rings of R such that $R[\theta; \delta]$ is Jacobson if and only if each of these factor rings of R is Jacobson.

In the first section we collect results on ideals of a ring R closed under the action of a derivation. Many special cases of these results appear in the literature, but we have tried to state and prove them in the maximum possible generality. In Sect. II we obtain conditions for the rings $R[\theta; \delta]$, and certain localizations $R_c[\theta; \delta]$, to be simple rings. The results on simplicity of $R[\theta; \delta]$ were known when R is a domain, but not otherwise. In Sect. III we derive necessary and sufficient conditions for $R[\theta; \delta]$ to be primitive, and construct explicit faithful simple modules corresponding to these conditions. In the final section, we obtain necessary and sufficient conditions for $R[\theta; \delta]$ to be a Jacobson ring. The main technical point here is the study of the existence and "going up" properties of prime ideals P of $R[\theta; \delta]$ such that $R[\theta; \delta]/P$ is finitely generated as an R-module.

0025-5874/82/0180/0503/$04.20

All rings in this paper are associative with identity, and all modules are unital. If R is a ring, and δ a derivation on R, then the corresponding differential operator ring, or Ore extension, is denoted $R[\theta; \delta]$. As a left R-module, $R[\theta; \delta]$ is the free R-module with basis $\{1, \theta, \theta^2, \ldots\}$, and the multiplication is extended from R to $R[\theta; \delta]$ by the rule $\theta r = r\theta + \delta(r)$ (for $r \in R$). In particular, δ extends to an inner derivation of $R[\theta; \delta]$. On the occasions when we have need of this derivation, we just denote it by δ as well, so that $\delta(a) = \theta a - a\theta$ for all $a \in R[\theta; \delta]$; in particular, $\delta(\theta) = 0$. Also, δ extends uniquely to localizations of R. For instance, when R is commutative and X is a multiplicatively closed subset of R, we shall assume that δ has been extended to R_X (which is uniquely possible, via the quotient rule), and we shall identify $R_X[\theta; \delta]$ with the Ore localization $R[\theta; \delta]_X$.

The research of both authors was supported in part by grants from the National Science Foundation.

I. Differential Ideals

In this section we consider some of the basic relationships between δ-ideals and ordinary ideals in a differential ring, and between δ-prime ideals and prime ideals. Most of these results are quite general, but are only available in the literature under extra hypotheses, such as commutativity or noetherianness. For the sake of future usage, we derive the general cases here, although for the present paper the available literature would suffice with a few modifications.

Definition. By a *differential ring*, we mean a ring R (associative, with 1) equipped with a specified derivation δ. By a *differential F-algebra* (over a given field F), we mean an F-algebra equipped with a specified F-linear derivation δ. To avoid naming the derivation each time we refer to a differential ring, all derivations in this paper will be denoted δ.

Definition. Let R be a differential ring. A *δ-ideal* of R is any two-sided ideal I of R such that $\delta(I) \subseteq I$. If R is nonzero and the only δ-ideals in R are 0 and R, then R is said to be *δ-simple*. A *δ-prime ideal* of R is any proper δ-ideal P such that whenever I and J are δ-ideals of R with $IJ \subseteq P$, then either $I \subseteq P$ or $J \subseteq P$. If 0 is a δ-prime ideal of R, then R is said to be a *δ-prime ring*. Given any two-sided ideal K of R, we define

$$(K : \delta) = \{r \in R \mid \delta^n(r) \in K \text{ for all } n = 0, 1, 2, \ldots\},$$

which is the largest δ-ideal of R contained in K. Note that if K is a prime ideal, then $(K : \delta)$ is a δ-prime ideal.

Definition. Given a δ-prime differential ring R, we may define its characteristic in the same manner as for prime rings, as follows. In case R is torsion-free as a $\mathbb{Z}$-module, we say that R has *characteristic zero*. If not, there is a prime integer p such that the set $X = \{x \in R \mid px = 0\}$ is nonzero. Then pR and X are δ-ideals of R satisfying $(pR)X = 0$, hence $pR = 0$ because R is δ-prime. Obviously p is uniquely determined, and we say that R has *characteristic p* in this case.

Definition. For a differential ring R, the set $S=\{s\in R|\delta(s)=0\}$ is a subring of R, called the *subring of constants* of R. In case R is commutative and δ-simple, S is a field. [Namely, given a nonzero element $s\in S$, note that sR is a nonzero δ-ideal of R, whence $sR=R$ and s is invertible in R. Applying δ to the equation $ss^{-1}=1$ then shows that $\delta(s^{-1})=0$, so that $s^{-1}\in S$.] Note that S has the same characteristic as R. Thus a commutative δ-simple differential ring of characteristic zero is a $\mathbb{Q}$-algebra, while a commutative δ-simple differential ring of characteristic $p>0$ is a $(\mathbb{Z}/p\mathbb{Z})$-algebra.

Now consider a commutative δ-prime differential ring R, and let S be its subring of constants. Given a nonzero element $s\in S$, note that the set $X=\{x\in R|sx=0\}$ is a δ-ideal of R. Since sR is also a δ-ideal, and $(sR)X=0$, we must have $X=0$ (by δ-primeness), so that s is a non-zero-divisor in R. Therefore S is a domain and R is torsion-free as an S-module. Note that S and R have the same characteristic.

We first derive some information about those prime ideals P in a differential ring R for which R/P has characteristic zero. The first result was proved for commutative noetherian differential $\mathbb{Q}$-algebras by Seidenberg [14, Theorem 1], and for commutative differential rings by Carlson and the first author [3, Lemma 1.3]. For noncommutative differential $\mathbb{Q}$-algebras, the result appears in Dixmier [5, Lemma 3.3.2]; as our slightly more general version uses basically the same proof, we just provide an outline of the method.

Proposition 1.1. *Let R be a differential ring, and let P be a prime ideal of R. If R/P has characteristic zero, then $(P:\delta)$ is a prime ideal of R. Thus if P is a minimal prime ideal of R and R/P has characteristic zero, then P is a δ-ideal.*

Proof. If $(P:\delta)$ is not prime, choose $x, y\in R-(P:\delta)$ such that $xRy\subseteq(P:\delta)$. Let s and t be the smallest nonnegative integers such that $\delta^s(x)\notin P$ and $\delta^t(y)\notin P$, and choose $z\in R$ such that $\delta^s(x)z\delta^t(y)\notin P$. Expanding $\delta^{s+t}(xzy)$ by Leibnitz' Rule, all terms lie in P except for the term $\binom{s+t}{s}\delta^s(x)z\delta^t(y)$, which lies outside P because R/P has characteristic zero. But then $\delta^{s+t}(xzy)\notin P$ and hence $xzy\notin(P:\delta)$, a contradiction.

Therefore $(P:\delta)$ must be a prime ideal. In case P is minimal, $(P:\delta)=P$, whence $\delta(P)\subseteq P$. □

We next modify Proposition 1.1 to show that in a differential ring with no $\mathbb{Z}$-torsion, all the minimal prime ideals are δ-ideals. It obviously suffices to prove that for a ring R with no $\mathbb{Z}$-torsion, R modulo any minimal prime ideal has characteristic zero. Note that this is easy in case R is commutative noetherian, since then the minimal prime ideals of R are associated primes of R. A similar proof can be given in the noncommutative noetherian case using affiliated primes.

Definition. Recall that an *m-system* in a ring R is a nonempty subset $X\subseteq R$ such that given any $x, y\in X$, there exists $r\in R$ with $xry\in X$. We shall need the fact that any two-sided ideal of R which is maximal with respect to being disjoint from an m-system must be prime [12, Lemma 4].

Lemma 1.2. *If R is a ring with no $\mathbb{Z}$-torsion, then R/P has characteristic zero for every minimal prime ideal P of R.*

Proof. Set $X=\{nx \mid n\in\mathbb{N} \text{ and } x\in R-P\}$, and note that X is an m-system in R. Moreover, $0\notin X$ because R has no $\mathbb{Z}$-torsion. Consequently, there exists a two-sided ideal Q in R which is maximal with respect to being disjoint from X, and Q is prime. Now $R-P\subseteq X$ and so $Q\subseteq P$, hence $Q=P$ by the minimality of P. Therefore X is disjoint from P, proving that R/P has characteristic zero. □

Proposition 1.3. *Let R be a differential ring, and let N be the prime radical of R. If either R or R/N has no $\mathbb{Z}$-torsion, then all minimal prime ideals of R are δ-ideals, and N is a δ-ideal.*

Proof. In either case, Lemma 1.2 shows that R/P has characteristic zero for any minimal prime ideal P of R, whence Proposition 1.1 shows that P is a δ-ideal. As N is the intersection of the minimal prime ideals of R, it too is a δ-ideal. □

Part of the right noetherian case of Proposition 1.3 was proved by Jordan in [9, Lemma 2.1]; namely, that if R has no $\mathbb{Z}$-torsion, then N is a δ-ideal. The following corollary is implicit in Jordan's work [9, Lemma 2.1, Theorem 2.2].

Corollary 1.4. *Let R be a right noetherian differential ring, and let I be a δ-prime ideal of R. If R/I has characteristic zero, then I is a prime ideal of R.*

Proof. After factoring out I, we may assume that $I=0$, so that R is a δ-prime ring with no $\mathbb{Z}$-torsion. Since R is right noetherian, it has only finitely many minimal prime ideals, say $P_1, \ldots, P_n$. The intersection of these P_i is the prime radical of R, which is nilpotent, hence $(P_1 P_2 \ldots P_n)^k=0$ for some positive integer k. By Proposition 1.3, each P_i is a δ-ideal of R. Thus since R is δ-prime we conclude that some $P_i=0$, hence 0 is a prime ideal. □

The final proposition of this section is essentially a corollary of a result of Jordan [9, Theorem 2.2].

Proposition 1.5. *Let R be a right noetherian differential ring. If I is a δ-prime ideal of R, then there exists a prime ideal P in R such that $(P\colon \delta)=I$ and I contains a power of P. Moreover, if R/P has characteristic zero, then $P=I$.*

Proof. The first statement follows from applying [9, Theorem 2.2] to the differential ring R/I. Then P/I is a minimal prime ideal of R/I. If R/P has characteristic zero, then P/I is a δ-ideal by Proposition 1.1, whence P is a δ-ideal, and so $P=(P\colon \delta)=I$. □

We close this section with an example showing that Proposition 1.5 may fail when the noetherian hypothesis is removed. We conjecture that there should also exist a counterexample to the non-noetherian version of Corollary 1.4.

Example 1.6. There exists a commutative differential ring R (of characteristic 2) containing a δ-prime ideal I such that $I\neq(P\colon \delta)$ for all prime ideals P of R.

Proof. Choose a field F of characteristic 2, let $x_1, x_2, \ldots$ be independent commuting indeterminates, and set $R=F[x_1, x_2, \ldots]$. Define an F-linear derivation δ on R so that $\delta(x_n)=x_{n+1}$ for all n. Let I be the ideal of R generated by $x_1^2, x_2^2, x_3^2, \ldots$. As F has characteristic 2, we have $\delta(x_n^2)=0$ for all n, whence I is a δ-ideal of R. The only prime ideal of R that contains I is the maximal ideal M generated by $x_1, x_2, \ldots$. Since M is a δ-ideal, $(M:\delta)=M \neq I$. Therefore $I \neq (P:\delta)$ for all prime ideals P of R.

Thus it only remains to show that I is δ-prime. Set $S=R/I$, and set $y_n=x_n+I$ for all n. Then S is a commutative differential F-algebra with an F-linear derivation δ such that $\delta(y_n)=y_{n+1}$ for all n, and we must show that S is a δ-prime ring. Note that S has a basis consisting of 1 together with all monomials of the form

$$y_{n(1)}y_{n(2)}y_{n(3)}\cdots y_{n(k)}$$

where $n(1)<n(2)<n(3)<\ldots<n(k)$.

Claim I: If A is a nonzero ideal of S, then A must contain a monomial of the form $y_p y_{p+1}\cdots y_q$.

Choose a nonzero element $a\in A$ which is a linear combination of as few nonzero monomials as possible, and write $a=\alpha_1 a_1+\ldots+\alpha_m a_m$ for some nonzero $\alpha_i \in F$ and some distinct nonzero monomials a_i. Suppose that $m>1$. Since nonzero monomials in S are square-free, there must be some y_n which is a factor of one of a_1, a_2 but not the other. Thus one of $y_n a_1$, $y_n a_2$ is zero while the other is not. But then $y_n a$ is a nonzero element of A which is a linear combination of at most $m-1$ nonzero monomials, contradicting the minimality of m. Thus $m=1$, so that $a=\alpha_1 a_1$, hence the monomial $a_1=\alpha_1^{-1}a$ lies in A. Now

$$a_1=y_{p(1)}y_{p(2)}\cdots y_{p(k)}$$

for some indices $p(1)<p(2)<\ldots<p(k)$. If there are any positive integers n such that $p(i)<n<p(i+1)$ for some i, multiply a_1 by the corresponding y_n. Therefore A must contain the monomial

$$y_{p(1)}y_{p(1)+1}y_{p(1)+2}\cdots y_{p(k)},$$

proving Claim I.

Claim II: If A is a δ-ideal of S that contains a monomial $y_p y_{p+1}\cdots y_q$, then A also contains a monomial of the form $y_{p+1}y_{p+2}\cdots y_r$.

If $q=p$, then we are given $y_p\in A$, whence $y_{p+1}=\delta(y_p)\in A$ as well. Thus we may assume that $q>p$. We proceed by induction on $s=q-1, q-2, \ldots, p$ to show that the monomials

$$a_s=y_p y_{p+1}\cdots y_s y_{s+2}y_{s+3}\cdots y_{2q-s}$$

all lie in A. Observing that $\delta(y_p y_{p+1}\cdots y_q)=a_{q-1}$, we obtain $a_{q-1}\in A$ to start with. Now suppose that $a_s\in A$ for some s with $q-1\geqq s>p$. Computing $\delta(a_s)$, we find that all but two terms contain squares and so vanish, leaving only the monomials

$$y_p y_{p+1}\cdots y_{s-1}y_{s+1}y_{s+2}\cdots y_{2q-s}$$

$$y_p y_{p+1}\cdots y_s y_{s+2}y_{s+3}\cdots y_{2q-s-1}y_{2q-s+1}.$$

As a result, $y_{2q-s+1}\delta(a_s)=a_{s-1}$, showing that $a_{s-1}\in A$. Thus the induction works, hence $a_p\in A$. Computing $\delta(a_p)$ and multiplying by y_{2q-p+1}, we find that the monomial $y_{p+1}y_{p+2}\cdots y_{2q-p+1}$ lies in A, proving Claim II.

Given nonzero δ-ideals A and B in S, apply Claim I to obtain monomials

$$y_p y_{p+1}\cdots y_q\in A \quad \text{and} \quad y_r y_{r+1}\cdots y_s\in B.$$

After sufficiently many successive applications of Claim II to B, we can obtain a monomial $y_t y_{t+1}\ldots y_u$ in B such that $t>q$. Thus AB contains the monomial

$$y_p y_{p+1}\cdots y_q y_t y_{t+1}\cdots y_u,$$

hence $AB\neq 0$.

Therefore S is indeed a δ-prime ring, so that I is a δ-prime ideal. □

II. Simplicity

Our study of primitivity of the rings $R[\theta;\delta]$ rests heavily on conditions for certain localizations $R_c[\theta;\delta]$ to be simple rings. Specifically, we require necessary and sufficient conditions on a commutative noetherian differential ring R for there to be a non-zero-divisor $c\in R$ such that $R_c[\theta;\delta]$ is simple. We first consider conditions for $R[\theta;\delta]$ to be a simple ring, where R is assumed to be commutative but not necessarily noetherian. It is clearly necessary for R to be δ-simple (since all δ-ideals of R induce two-sided ideals of $R[\theta;\delta]$), and if the characteristic of R is zero, then it is well-known that this condition is sufficient, provided that $\delta\neq 0$ [2, Theorems 1,3,4; 4, Theorem 3.2]. For commutative domains of positive characteristic, the criterion we obtain is proved in [4, Theorem 3.2a]. We first need a bound on the dimensions of solution spaces of homogeneous linear differential equations over commutative δ-simple differential rings. The method of proof is essentially the same as that used by Amitsur to prove the corresponding result over differential division rings [1, Theorem 1].

Proposition 2.1. *Let R be a commutative differential ring. Assume that R is δ-simple, and let F be the subfield of constants of R. Let $a_0,\ldots,a_{n-1}\in R$, and set*

$$X=\{x\in R\mid \delta^n(x)+a_{n-1}\delta^{n-1}(x)+\ldots+a_1\delta(x)+a_0x=0\}.$$

Then $\dim_F(X)\leqq n$.

Proof. Set $Y=\{(x,\delta(x),\ldots,\delta^{n-1}(x))\mid x\in X\}$, which is an F-subspace of R^n. As vector spaces over F, we have $Y\cong X$, hence it suffices to prove that $\dim_F(Y)\leqq n$. Since R is commutative, R^n cannot contain a free R-module of rank $n+1$. Thus we shall be done after proving the following statement: If $y_1,\ldots,y_k\in Y$ are linearly independent over F, then they are linearly independent over R as well.

Write each $y_i=(x_i,\delta(x_i),\ldots,\delta^{n-1}(x_i))$ for some $x_i\in X$. Set

$$J=\{(r_1,\ldots,r_k)\in R^k\mid r_1y_1+\ldots+r_ky_k=0\};$$

then our problem is to show that $J=0$. We first show that $\delta(J)\subseteq J$.

Given $(r_1, \ldots, r_k) \in J$, we have $\sum_{i=1}^{k} r_i \delta^j(x_i) = 0$ for $j = 0, \ldots, n-1$. Using the fact that each $x_i \in X$, we compute that

$$\sum_{i=1}^{k} r_i \delta^n(x_i) = \sum_{i=1}^{k} r_i \left(-\sum_{j=0}^{n-1} a_j \delta^j(x_i) \right) = -\sum_{j=0}^{n-1} a_j \left(\sum_{i=1}^{k} r_i \delta^j(x_i) \right) = 0.$$

Thus for $j = 0, \ldots, n-1$ we have

$$\sum_{i=1}^{k} r_i \delta^{j+1}(x_i) = 0 \quad \text{and} \quad \sum_{i=1}^{k} r_i \delta^j(x_i) = 0.$$

Consequently,

$$\sum_{i=1}^{k} \delta(r_i) \delta^j(x_i) = \sum_{i=1}^{k} (\delta(r_i) \delta^j(x_i) + r_i \delta^{j+1}(x_i)) = \delta \left(\sum_{i=1}^{k} r_i \delta^j(x_i) \right) = 0$$

for each $j = 0, \ldots, n-1$, whence $(\delta(r_1), \ldots, \delta(r_k)) \in J$. Thus $\delta(J) \subseteq J$.

Now if $k = 1$, then we just have $y_1 \neq 0$ and J is a proper δ-ideal of R, hence $J = 0$ and thus $\{y_1\}$ is linearly independent over R.

If $k > 1$, then by induction we may assume that $y_2, \ldots, y_k$ are linearly independent over R. Set

$$I = \{r_1 \in R \mid \text{there exists a } k\text{-tuple } (r_1, \ldots, r_k) \text{ in } J\},$$

and note that I is a δ-ideal of R. The linear independence of $y_2, \ldots, y_k$ over R forces any nonzero $(r_1, \ldots, r_k)$ in J to have $r_1 \neq 0$. Thus if $J \neq 0$, then $I \neq 0$, whence $I = R$. Therefore there exists a k-tuple $(1, s_2, \ldots, s_k)$ in J. Applying δ, we find that $(0, \delta(s_2), \ldots, \delta(s_k))$ lies in J, whence each $\delta(s_i) = 0$, and so $s_i \in F$. However, the relation

$$y_1 + s_2 y_2 + \ldots + s_k y_k = 0$$

then contradicts the linear independence of the y_i over F. Thus $J = 0$, so that $y_1, \ldots, y_k$ are indeed linearly independent over R. □

Proposition 2.1 will be applied in cases where $X = R$, which cannot occur in characteristic zero unless $\delta = 0$, as follows.

Lemma 2.2. *Let R be a δ-simple commutative differential ring, and let F be the subfield of constants of R. If R has characteristic zero, then R is a domain, and R is transcendental over F. Thus $\dim_F(R) = \infty$ unless $\delta = 0$.*

Proof. As R has characteristic zero, Proposition 1.3 shows that any minimal prime ideal P of R is a δ-ideal, whence $P = 0$ by δ-simplicity. Therefore R is a domain.

If R is not transcendental over F, then we may choose an element x in $R - F$ that is algebraic over F. Let n be the degree of x, and note that $n > 1$. Thus

$$\alpha_0 + \alpha_1 x + \ldots + \alpha_n x^n = 0$$

for some $\alpha_i \in F$, with $\alpha_n \neq 0$. Applying δ, we obtain

$$(\alpha_1 + 2\alpha_2 x + \ldots + n\alpha_n x^{n-1}) \delta(x) = 0,$$

hence $\alpha_1 + 2\alpha_2 x + \ldots + n\alpha_n x^{n-1} = 0$ (because $\delta(x) \neq 0$). But $n\alpha_n \neq 0$, so this contradicts the assumption that x has degree n. Therefore R is transcendental over F. □

Theorem 2.3. *Let R be a commutative differential ring, and let F be the subring of constants of R. Then $R[\theta; \delta]$ is a simple ring if and only if R is δ-simple (whence F is a field) and* $\dim_F(R) = \infty$.

Proof. Let $T = R[\theta; \delta]$.

First assume that T is simple. As IT is a two-sided ideal of T for each δ-ideal I of R, it is clear that R must be δ-simple. Consequently, F is a field, and we note that F is central in T. Note also that $1, \theta, \theta^2, \ldots$ are linearly independent over F, so that $\dim_F(T) = \infty$. Since T is simple, $T/\theta T$ is a faithful right T-module, hence T embeds in $\mathrm{End}_F(T/\theta T)$. Thus $\mathrm{End}_F(T/\theta T)$ must be an infinite-dimensional F-algebra, whence $\dim_F(T/\theta T) = \infty$. Since $T/\theta T \cong R$ (as right R-modules), we conclude that $\dim_F(R) = \infty$.

Conversely, suppose that R is δ-simple and that $\dim_F(R) = \infty$. If T has a nontrivial two-sided ideal J, then $J \cap R$ is a proper δ-ideal of R, so $J \cap R = 0$. Let n be the minimal order for nonzero elements of J, then $n > 0$ because $J \cap R = 0$. Since J is closed under δ (that is, under the inner derivation induced by θ), the set

$$I = \{a \in R \mid a_0 + a_1\theta + \ldots + a_{n-1}\theta^{n-1} + a\theta^n \in J \text{ for some } a_0, \ldots, a_{n-1} \in R\}$$

is a nonzero δ-ideal of R. Consequently, $I = R$, and hence J contains an element of the form

$$b = b_0 + b_1\theta + \ldots + b_{n-1}\theta^{n-1} + \theta^n,$$

where $b_0, \ldots, b_{n-1} \in R$.

Given $r \in R$, we have $br - rb$ in J with order less than n, hence $br - rb = 0$ by the minimality of n. Writing $br - rb$ with left-hand coefficients, and collecting the terms not involving θ, we find that

$$\delta^n(r) + b_{n-1}\delta^{n-1}(r) + \ldots + b_1\delta(r) = 0.$$

As this holds for all $r \in R$, Proposition 2.1 says that $\dim_F(R) \leqq n$, which contradicts our hypotheses.

Therefore T must be simple. □

The following example shows that Theorem 2.3 is non-vacuous in the case of positive characteristic. Our construction is the characteristic p analog of the construction used in an example of McConnell [11, Theorem 2.1].

Example 2.4. Given a prime number p, there exists a differential field K of characteristic p such that K is infinite-dimensional over its subfield of constants.

Proof. Choose an infinite field k of characteristic p, and choose elements $\lambda_1, \lambda_2, \ldots$ in k which are linearly independent over $\mathbb{Z}/p\mathbb{Z}$. Let $x_1, x_2, \ldots$ be independent commuting indeterminates, and set

$$K = k(x_1, x_2, \ldots); \qquad F = k(x_1^p, x_2^p, \ldots).$$

Define a k-linear derivation δ on K so that $\delta(x_n)=\lambda_n x_n$ for all n. Because of characteristic p, we have $\delta(x_n^p)=0$ for all n, hence every element of F is a constant. We claim that F is exactly the subfield of constants of K, which will complete the example, because $\dim_F(K)=\infty$.

Now set $R=k[x_1, x_2, \ldots]$ and $S=k[x_1^p, x_2^p, \ldots]$. Note that all elements of K can be written in the form a/b^p for some $a, b \in R$ (with $b \neq 0$), and that $b^p \in S$ because of characteristic p. In addition, $\delta(a/b^p)=0$ if and only if $\delta(a)=0$. Consequently, to prove that F is the subfield of constants of K, it suffices to show that S is the subring of constants of R.

We may view R as a vector space over k and δ as a k-linear transformation on R. Consider a monomial

$$v=x_1^{n(1)}x_2^{n(2)}\cdots x_m^{n(m)}.$$

Then v is an eigenvector for δ, with eigenvalue

$$\lambda(v)=n(1)\lambda_1+n(2)\lambda_2+\ldots+n(m)\lambda_m.$$

Hence R is spanned by eigenvectors for δ, and is therefore (uniquely) a direct sum of eigenspaces. Note that $\lambda(v)=0$ if and only if p divides $n(i)$ for all i. The uniqueness of the eigenspace decomposition of R shows that S is actually the 0-eigenspace of δ in R, which is what we wished to prove. □

We now turn to the question of simplicity for a differential operator ring $R_c[\theta; \delta]$, where R is a commutative noetherian differential ring and c is a non-zero-divisor in R. [Recall that R_c is the ring obtained from R by inverting c.] According to Theorem 2.3, $R_c[\theta; \delta]$ is simple if and only if R_c is δ-simple and R_c is infinite-dimensional over its subfield of constants; we wish to translate these conditions into conditions intrinsic to R. The second condition corresponds to R having infinite rank over its subring of constants (which is a domain), as we prove in Proposition 2.6. In characteristic zero, this always holds unless $\delta=0$, as follows.

Lemma 2.5. *Let R be a δ-prime commutative noetherian differential ring of characteristic zero, and let D be the subring of constants of R. Then R is a domain, and R has infinite rank as a D-module if and only if $\delta \neq 0$.*

Proof. In view of Corollary 1.4, R must be a domain. Let F and K be the quotient fields of D and R. If $\delta \neq 0$, then according to Lemma 2.2, K is infinite-dimensional over its subfield of constants, which contains F, hence $\dim_F(K)=\infty$. Therefore R has infinite rank as a D-module. □

Proposition 2.6. *Let R be a commutative noetherian differential ring, and assume that there is a non-zero-divisor c in R such that R_c is δ-simple. Then R is δ-prime, and the subring D of constants of R is a domain. The ring $R_c[\theta; \delta]$ is a simple ring if and only if R has infinite rank as a D-module.*

Proof. The δ-simplicity of R_c clearly implies that R is δ-prime, whence D is a domain. In case R has characteristic zero, Lemma 2.2 and Theorem 2.3 show that $R_c[\theta; \delta]$ is simple if and only if $\delta \neq 0$ on R_c. This happens if and only if $\delta \neq 0$ on R, which by Lemma 2.5 occurs if and only if R has infinite rank as a D-module.

Thus we may assume that R has characteristic $p>0$. As R_c may also be obtained from R by inverting the element c^p, which satisfies $\delta(c^p)=0$, there is no loss of generality in assuming that $\delta(c)=0$. Let F be the subfield of constants of R_c. As $c\in D$, it is easily checked that $F=D_c$. Consequently, F is the quotient field of D, and $R_c\cong R\otimes_D F$, so R has infinite rank as a D-module if and only if R_c is infinite-dimensional over F. By Theorem 2.3, this happens if and only if $R_c[\theta;\delta]$ is simple. □

Lemma 2.7. *Let c be a non-zero-divisor in a commutative noetherian ring R. If R_c is a $\mathbb{Q}$-algebra, then all but finitely many prime integers are invertible in R.*

Proof. If p is a prime integer not invertible in R, then $1/p$ is an element of R_c-R such that $p(1/p)\in R$, which shows that R_c/R has p-torsion. Since $1/p\in c^{-n}R$ for some $n\in\mathbb{N}$, we see that $c^{-n}R/R$ has p-torsion. If $k\in\mathbb{N}$ is chosen minimal with respect to the property that $c^{-k}R/R$ has p-torsion, then $c^{-k}R/c^{-k+1}R$ also has p-torsion (since a nonzero element x in $c^{-k}R/R$ satisfying $px=0$ cannot lie in $c^{-k+1}R/R$). Now multiplication by c^k induces an isomorphism of $c^{-k}R/c^{-k+1}R$ onto R/cR, so R/cR has p-torsion.

Thus R/cR has p-torsion for any prime integer p not invertible in R. Now the $\mathbb{Z}$-torsion subgroup of R/cR is the direct sum of its p-torsion subgroups (over all prime integers p), and these p-torsion subgroups of R/cR are all ideals. Since R/cR is a noetherian ring, all but a finite number of these summands must be zero, so R/cR can only have p-torsion for finitely many p. Therefore only finitely many prime integers can be non-invertible in R. □

Proposition 2.8. *Let R be a commutative noetherian differential ring, and let c be a non-zero-divisor in R. Then R_c is δ-simple if and only if there is a positive integer n such that c^n is contained in all nonzero δ-prime ideals of R.*

Proof. If R_c is not δ-simple, then it contains a nontrivial δ-ideal J. Choose a maximal ideal M of R_c which contains J, and set $P=(M:\delta)$; then P is a δ-prime ideal of R_c, and $J\subseteq P$, so $P\neq 0$. Consequently, $P\cap R$ is a nonzero δ-prime ideal of R, and since $P\neq R_c$ no power of c can lie in $P\cap R$.

Conversely, suppose that R_c is δ-simple. Then R_c is either a $\mathbb{Q}$-algebra or a $(\mathbb{Z}/q\mathbb{Z})$-algebra for some prime integer q. In the first case, Lemma 2.7 shows that all but finitely many prime integers are invertible in R, while in the second case all prime integers except q are invertible in R. Thus there exists a positive integer n which is divisible by all prime integers that are not invertible in R.

If I is a nonzero δ-prime ideal of R, then since I_c is a nonzero δ-ideal of R_c, we must have $I_c=R_c$, which implies that I contains a power of c. According to Proposition 1.5 there is a prime ideal P in R such that $(P:\delta)=I$. Since $I\subseteq P$ and I contains a power of c, we obtain $c\in P$. If R/P has characteristic zero, then $P=I$ (Proposition 1.5 again), so that $c\in I$ and consequently $c^n\in I$. If R/P has positive characteristic p, then as $\delta(c^p)=pc^{p-1}\delta(c)$ we find that all derivatives of c^p lie in P, whence $c^p\in I$. In addition, p is not invertible in R, hence $p\mid n$ and so $c^n\in I$.

Therefore all nonzero δ-prime ideals of R contain c^n, as desired. □

Definition. Following Jordan [10], we say that a differential ring R is a *δ-G-ring* provided that R is δ-prime and the intersection of its nonzero δ-prime ideals is nonzero. This is parallel to the definition of a *G-ring*, meaning a prime ring R such that the intersection of the nonzero prime ideals of R is nonzero. In the commutative case, such an R is a domain and so is called a *G-domain.* Equivalently, a commutative ring R is a G-domain if and only if there is a non-zero-divisor $c \in R$ such that R_c is a field.

Proposition 2.9. *Let R be a commutative noetherian differential ring, and let D be the subring of constants of R. Then R is a δ-G-ring if and only if there exists a non-zero-divisor $c \in R$ such that R_c is a δ-simple ring. In case R is δ-prime with positive characteristic, then R is a δ-G-ring if and only if D is a G-domain.*

Proof. If R_c is δ-simple, where c is a non-zero-divisor in R, then R is δ-prime, and it is clear from Proposition 2.8 that R is a δ-G-ring. Conversely, if R is a δ-G-ring, then the intersection I of the nonzero δ-prime ideals of R is a nonzero δ-ideal. As R is δ-prime, I cannot be nilpotent, hence I is not nil. Thus we may choose a non-nilpotent element $c \in I$. If R has characteristic zero, then R is a domain (by Corollary 1.4), so c is a non-zero-divisor. If R has characteristic $p>0$, then c^p is a nonzero element of the domain D, whence c^p is a non-zero-divisor in R (because R is torsion-free as a D-module, as noted in Section I), and again c is a non-zero-divisor. Finally, since R_c has no nonzero δ-prime ideals (from the choice of c), we conclude that R_c must be δ-simple (namely, if J were a nontrivial δ-ideal of R_c, then for any prime ideal P of R_c containing J, the ideal $(P:\delta)$ would be a nonzero δ-prime ideal of R_c).

Now assume that R is δ-prime with characteristic $p>0$. If R is a δ-G-ring, then there is a non-zero-divisor $c \in R$ such that R_c is δ-simple. There is no loss of generality in replacing c by c^p, so we may assume that $c \in D$. As in Proposition 2.6, it follows that D_c is a field (namely the subfield of constants of R_c), whence D is a G-domain.

Conversely, if D is a G-domain, then there is a nonzero element $c \in D$ such that D_c is a field. Since R is torsion-free as a D-module, c is a non-zero-divisor in R. If K is a nonzero δ-ideal in R, then since R is δ-prime, K is not nilpotent and thus not nil, so there is some $x \in K$ with $x^p \neq 0$. This shows that $K \cap D \neq 0$, hence K contains a unit of R_c, forcing $K_c = R_c$. Therefore R_c is δ-simple, so that R is a δ-G-ring. □

Theorem 2.10. *Let R be a commutative noetherian differential ring, and let D be the subring of constants of R. Then there exists a non-zero-divisor $c \in R$ such that $R_c[\theta;\delta]$ is a simple ring if and only if R is a δ-G-ring and R has infinite rank as a D-module.*

Proof. Propositions 2.6, 2.8, 2.9. □

III. Primitivity

In this section we investigate conditions for a differential operator ring $R[\theta;\delta]$ over a commutative noetherian differential ring R to be (right) primitive. (As $R[\theta;\delta]$ is isomorphic to its opposite ring, we need make no distinction between right and left primitivity.) In the process we obtain at least the start of

a picture of what kinds of simple modules $R[\theta;\delta]$ can have. Faithful simple modules over $R[\theta;\delta]$ can arise from either of two rather different sources. One source, for which the results of the previous section are a preparation, is the occurrance of a non-zero-divisor $c \in R$ such that $R_c[\theta;\delta]$ is a simple ring. The other source, which we deal with first because it is much more straightforward, is a modification of a result of Jordan [10, Theorem 1].

Definition. Following Jordan [10], we say that a differential ring R is (right) *δ-primitive* provided there is a maximal right ideal of R that contains no nonzero δ-ideals of R.

Proposition 3.1. *Let R be a commutative noetherian differential ring, let P be an ideal of R, and set $T=R[\theta;\delta]$. Then T/PT is a simple right T-module if and only if P is a maximal ideal of R but not a δ-ideal, and R/P has characteristic zero. In any case, the annihilator of the module T/PT is the ideal $(P:\delta)T$.*

Proof. If P is a maximal ideal but not a δ-ideal, and R/P has characteristic zero, then an argument of Hart [7, Lemma 2.4] shows that PT is a maximal right ideal of T, so T/PT is a simple module.

Conversely, assume that T/PT is simple. If J is a proper ideal of R containing P, then JT is a proper right ideal of T containing PT, so $JT=PT$ and $J=P$. Thus P is a maximal ideal. If R/P has positive characteristic, then by [6, Lemma 13], $R/(P:\delta)$ is an artinian local differential ring. However, $T/(P:\delta)T$, being isomorphic to $[R/(P:\delta)][\theta;\delta]$, is a non-artinian ring. Since $R/(P:\delta)$ is artinian, there is a chain

$$J_0=(P:\delta)<J_1<\ldots<J_n=R$$

of ideals in R such that each $J_i/J_{i-1}\cong R/P$. Then each of the factors $J_iT/J_{i-1}T$ is isomorphic to T/PT. Since $T/(P:\delta)T$ is not artinian, this contradicts the simplicity of T/PT. Therefore R/P must have characteristic zero. If P were a δ-ideal, then PT would be an ideal of T, and T/PT would not be artinian, so it is also necessary that P not be a δ-ideal.

In general, it is clear that the annihilator I of T/PT must contain $(P:\delta)T$. Given an operator $x=x_0+x_1\theta+\ldots+x_n\theta^n$ in I (with all $x_i\in R$), we have $(1+PT)x=0$ and so $x\in PT$, whence all $x_i\in P$. As the operator

$$\delta(x)=\delta(x_0)+\delta(x_1)\theta+\ldots+\delta(x_n)\theta^n=\theta x-x\theta$$

also lies in I, all $\delta(x_i)\in P$. Similarly, $\delta^k(x)\in I$ for all k, hence $\delta^k(x_i)\in P$ for all k,i. Thus each $x_i\in(P:\delta)$, so that $x\in(P:\delta)T$. Therefore $I=(P:\delta)T$. □

Corollary 3.2. *Let R be a commutative noetherian differential ring with no $\mathbb{Z}$-torsion. If $\delta\neq 0$ and R is δ-primitive, then $R[\theta;\delta]$ is primitive.*

Proof. By assumption, R has a maximal ideal M that contains no nonzero δ-ideals. In particular, $(M:\delta)=0$, hence R is δ-prime. If p is any prime integer, then pR is a nonzero δ-ideal of R and so $pR\not\subseteq M$. Thus R/M has characteristic zero.

If $M=0$, then as $\delta\neq 0$ the ring $R[\theta;\delta]$ must be simple (e.g., [8, Theorem 3], or Lemma 2.2 and Theorem 2.3), hence primitive. If $M\neq 0$, then since $(M:\delta)$

$=0$ we see that M is not a δ-ideal. In this case, Proposition 3.1 shows that T/MT is a faithful simple right T-module, so that T is primitive. □

Proposition 3.3. *Let R be a commutative noetherian differential ring, and set $T=R[\theta;\delta]$. Assume that there exists a non-zero-divisor $x\in R$ such that R_x is a δ-simple ring and $\delta(x)^m\in xR$ for some $m\in\mathbb{N}$. Then the module $A=T/(\theta x-1)T$ is a simple right T-module, and $A\cong R_x$ as an R-module. Moreover, the localization map $A\to A_x$ is an isomorphism.*

Proof. On localizing by inverting x, we see that

$$A_x\cong T_x/(\theta x-1)T_x=T_x/(\theta-x^{-1})T_x.$$

The latter module may be obtained by equipping R_x with a right T_x-module multiplication $*$ under which $r*\theta=(x^{-1}-\delta)(r)$ for all $r\in R_x$. Let us use B to denote R_x made into a right T_x-module in this way. We first show that B (hence A_x) is a simple T-module, and then that the natural map $A\to A_x$ is an isomorphism.

Since the ideals of R_x are closed under multiplication by x^{-1}, we see that the T_x-submodules of B are just the δ-ideals of R_x. Thus B is at least simple as a T_x-module (because R_x is δ-simple). Now if C is a nonzero T-submodule of B, then C_x is a nonzero T_x-submodule of B and so $C_x=B$, hence in particular $1=cx^{-n}$ for some $c\in C$ and $n\in\mathbb{N}$. Thus $x^n\in C$; consequently, $x^k\in C$ for all integers $k\geqq n$. We show by induction (going down) that $x^k\in C$ for all integers k, which will show that $C=B$, since B is generated an R-module by the elements x^k.

Assuming that some $x^k\in C$, we calculate that

$$x^k*\theta=(x^{-1}-\delta)(x^k)=x^{k-1}(1-k\delta(x)).$$

Since $x^k*\theta\in C$ and the element $1+k\delta(x)+\ldots+k^{m-1}\delta(x)^{m-1}$ is in R, the submodule C also contains the element

$$(x^k*\theta)(1+k\delta(x)+\ldots+k^{m-1}\delta(x)^{m-1})=x^{k-1}(1-k^m\delta(x)^m).$$

By hypothesis, $\delta(x)^m\in xR$, so $x^{k-1}k^m\delta(x)^m\in x^kR\subseteq C$, and we conclude that $x^{k-1}\in C$. This completes the induction, thereby proving that $C=B$. Thus B is a simple T-module.

Therefore A_x is a simple T-module as well. Since the natural map $f\colon A\to A_x$ is a nonzero T-module homomorphism, it must be surjective. It remains to show that f is injective, that is, that A has no x-torsion. Thus for any $a,b\in T$ satisfying $ax=(\theta x-1)b$, we need to show that $a\in(\theta x-1)T$. As x is a non-zero-divisor, $(\theta x-1)b$ has order at least 1 provided $b\neq 0$, so if $a\in R$ the only possibility is $a=0$.

Now suppose that a has order $n>0$. Then b has order $n-1$, hence we may write

$$b=b_0+\theta b_1+\ldots+\theta^{n-1}b_{n-1}$$

for some $b_i\in R$. Since

$$ax=(\theta x-1)b=\theta^n xb_{n-1}+\theta^{n-1}(xb_{n-2}-(n-1)\delta(x)b_{n-1}-b_{n-1})+[\text{lower terms}],$$

we see that x must divide $xb_{n-2}-(n-1)\delta(x)b_{n-1}-b_{n-1}$. Then

$$(1-(1-n)\delta(x))b_{n-1}$$

is divisible by x, and hence so is the element

$$(1-(1-n)^m\delta(x)^m)b_{n-1}=\left(\sum_{i=0}^{m-1}(1-n)^i\delta(x)^i\right)(1-(1-n)\delta(x))b_{n-1}.$$

As $\delta(x)^m$ is divisible by x (by hypothesis), it follows that b_{n-1} must be divisible by x. Consequently, $b=tx+b'$ for some $t,b'\in T$ where b' has order less than $n-1$. Then

$$(a-(\theta x-1)t)x=(\theta x-1)b',$$

so $a-(\theta x-1)t$ has order less than n, hence $a-(\theta x-1)t$ lies in $(\theta x-1)T$ by induction, and thus $a\in(\theta x-1)T$.

Therefore f is indeed an isomorphism of A onto A_x, proving that A is a simple T-module, and that $A\cong R_x$ as R-modules. □

Corollary 3.4. *Let R be a commutative noetherian differential ring. If there exists a non-zero-divisor $c\in R$ such that $R_c[\theta;\delta]$ is a simple ring, then $R[\theta;\delta]$ is a primitive ring.*

Proof. Set $T=R[\theta;\delta]$ and $x=c^2$. Then $T_x=T_c$ is a simple ring, so in particular R_x is δ-simple. As $\delta(x)=2c\delta(c)$, we see that $\delta(x)^2\in xR$. By Proposition 3.3, the module $A=T/(\theta x-1)T$ is a simple right T-module, and the localization map $A\to A_x$ is an isomorphism. As T_x is simple, A_x is a faithful T_x-module, whence A is a faithful T-module (because T embeds in T_x). Therefore T is primitive. □

Corollary 3.5. *Let R be a commutative noetherian differential ring, and let D be the subring of constants of R. If R is a δ-G-ring and R has infinite rank as a D-module, then $R[\theta;\delta]$ is a primitive ring.*

Proof. Theorem 2.10 and Corollary 3.4. □

It is possible to derive Corollary 3.5 from a result of Jordan [10, Theorem 2], but that theorem does not indicate where to find the faithful simple module. Corollaries 3.2 and 3.5 provide two situations in which the differential operator ring over a commutative noetherian differential ring can be primitive. We shall prove that, conversely, primitivity only occurs in these two situations.

Proposition 3.6. *Let R be a commutative noetherian differential ring, let D be the subring of constants of R, and set $T=R[\theta;\delta]$. Let A be a simple right T-module which is not of the form T/PT for any ideal P of R, and suppose that A is faithful as an R-module. Then there exists a non-zero-divisor $c\in R$ such that R_c is δ-simple and A_c is finitely generated faithful as an R_c-module. Moreover, A is a faithful T-module if and only if R has infinite rank as a D-module.*

Proof. Since R is commutative noetherian, A has an R-submodule which is isomorphic to R/P for some prime ideal P of R. Then A is an epimorphic image of T/PT, so according to our hypotheses, $A\cong T/M$ where M is a

maximal right ideal of T properly containing PT. Now $T(P:\delta)\subseteq PT$ and so $A(P:\delta)=0$, hence $(P:\delta)=0$ because A_R is faithful. In particular, R is δ-prime, and D is a domain.

As M properly contains PT, we may choose an operator $y=y_0+y_1\theta+\ldots+y_n\theta^n$ in M with all $y_i\in R$ but $y_n\notin P$. We wish to arrange in addition for y_n to be a non-zero-divisor in R. If R has characteristic zero, then R is a domain (Corollary 1.4), and this is automatic. Otherwise, R has positive characteristic p, and after replacing y by y^p (whence n is replaced by pn and y_n by y_n^p), we may assume that $\delta(y_n)=0$. In this case it follows from the δ-primeness of R that y_n is a non-zero-divisor.

Thus there is no loss of generality in assuming that the element $c=y_n$ is a non-zero-divisor in R. Now the right ideal M_c in the ring T_c contains the monic operator yc^{-1}, so it follows that A_c (being isomorphic to T_c/M_c) is a finitely generated R_c-module. In addition, since $c\notin P$ and A_c has a submodule isomorphic to $(R/P)_c$, we see that $A_c\neq 0$. Consequently, A_c is a simple T_c-module, and the natural map $A\to A_c$ must be an embedding, whence A_c is faithful as an R-module, and thus also faithful as an R_c-module.

If I is any proper δ-ideal of R_c, then $A_cI\neq A_c$ (because A_c is finitely generated faithful over R_c), so A_cI is a proper T_c-submodule of A_c, hence $A_cI=0$ and thus $I=0$. Therefore R_c is δ-simple.

If R has infinite rank over D, then T_c is simple according to Proposition 2.6, so A_c must be a faithful T_c-module, and consequently A is a faithful T-module.

Conversely, assume that A_T is faithful. Then T is primitive, so $\delta\neq 0$. Now if R has characteristic zero, R has infinite rank over D by Lemma 2.5, hence we may assume that R has characteristic $p>0$. After replacing c by c^p, we may also assume that $c\in D$. Then D_c equals the subring of constants of R_c, so that D_c is a field. As A embeds in A_c, we see that A_c is faithful as a T-module, and thus also faithful as a T_c-module, hence T_c embeds in the D_c-endomorphism ring of A_c. Since T_c is infinite-dimensional over D_c, it follows that A_c must be also, and therefore R_c is infinite-dimensional over D_c. But $R_c\cong R\otimes_D D_c$, so this shows that R has infinite rank over D. □

We can now give criteria for the primitivity of $R[\theta;\delta]$. It is convenient to separate the characteristic zero and characteristic p cases.

Theorem 3.7. *Let R be a commutative noetherian differential ring with no $\mathbb{Z}$-torsion. Then $R[\theta;\delta]$ is primitive if and only if $\delta\neq 0$ and R is either δ-primitive or a δ-G-ring.*

Proof. Set $T=R[\theta;\delta]$. If T is primitive, then obviously $\delta\neq 0$. If there is a faithful simple *induced* right T-module (i.e., a module of the form T/MT, for an ideal M of T), then Proposition 3.1 shows that R is δ-primitive. If there is a faithful simple right T-module which is not induced, then Proposition 3.6 shows that there is a non-zero-divisor $c\in R$ such that R_c is δ-simple. In this case we conclude from Proposition 2.9 that R is a δ-G-ring.

Conversely, if $\delta\neq 0$ and R is δ-primitive, then T is primitive by Corollary 3.2. If $\delta\neq 0$ and R is a δ-G-ring, then R has infinite rank over its subring of constants (Lemma 2.5), and Corollary 3.5 shows that T is primitive. □

Theorem 3.8. *Let R be a commutative noetherian differential ring with $\mathbb{Z}$-torsion, and let D be the subring of constants of R. Then $R[\theta; \delta]$ is primitive if and only if*

(a) *R is δ-prime.*

(b) *D is a G-domain.*

(c) *R has infinite rank as a D-module.*

Proof. In any case, the ring $T = R[\theta; \delta]$ is prime, hence it (and R as well) must have positive characteristic. Proposition 3.1 shows that a simple right T-module cannot be induced, so we infer from Proposition 3.6 that if T is primitive, then there must be a non-zero-divisor $c \in R$ such that R_c is δ-simple and (c) holds. According to Proposition 2.9, the δ-simplicity of R_c implies that (a) and (b) hold as well. Conversely, if (a) and (b) hold, then Proposition 2.9 shows that R is a δ-G-ring. Assuming (c) holds in addition, T is primitive by Corollary 3.5. □

We may combine Theorems 3.7 and 3.8 as follows.

Theorem 3.9. *Let R be a commutative noetherian differential ring, and let D be the subring of constants of R. Then $R[\theta; \delta]$ is primitive if and only if one of the following conditions holds:*

(a) *R has no $\mathbb{Z}$-torsion, $\delta \neq 0$, and R is δ-primitive.*

(b) *R is a δ-G-ring and R has infinite rank as a D-module.*

Proof. Theorems 3.7 and 3.8, Lemma 2.5, Proposition 2.9. □

Note that case (b) of Theorem 3.9 occurs if and only if there exists a non-zero-divisor $c \in R$ such that $R_c[\theta; \delta]$ is a simple ring (Theorem 2.10).

Example 3.10. Given a prime number p, there exists a commutative noetherian differential domain R of characteristic p such that $R[\theta; \delta]$ is primitive but not simple.

Proof. Let K be the differential field constructed in Example 2.4. Choose an indeterminate x, set $R = K[\![x]\!]$, and extend the derivation from K to R according to the rule

$$\delta\left(\sum_{n=0}^{\infty} \alpha_n x^n\right) = \sum_{n=0}^{\infty} \delta(\alpha_n) x^n.$$

Since R is a domain, it is δ-prime. The subring D of constants of R is just $F[\![x]\!]$, where F is the subfield of constants of K, so D is a G-domain. As $\dim_F(K) = \infty$, we see that R has infinite rank as a D-module. Therefore $R[\theta; \delta]$ is primitive, by Theorem 3.8. However, $R[\theta; \delta]$ is not simple, because R is not δ-simple. □

IV. Jacobsonianism

Jordan proved in [9, Theorem 3.5] that if R is a right noetherian differential ring which is a Jacobson ring (i.e., all prime factor rings are semiprimitive), then $R[\theta; \delta]$ is a Jacobson ring. The converse is not true in general, and in

this section we find necessary and sufficient conditions on a commutative noetherian differential ring R for $R[\theta;\delta]$ to be Jacobson. What difficulties there are all arise in characteristic p, and the chief technical difficulty concerns the existence and "going up" properties of prime ideals P in $R[\theta;\delta]$ for which $R[\theta;\delta]/P$ is finitely generated over R. We begin by stating our final result.

Theorem 4.1. *Let R be a commutative noetherian differential ring, and let $\mathfrak{T}$ be the set of those δ-prime ideals I of R such that R/I has finite rank as a module over its subring of constants. Then $R[\theta;\delta]$ is a Jacobson ring if and only if R/I is Jacobson for all $I\in\mathfrak{T}$.* □

We recall from Corollary 1.4 that if I is a δ-prime ideal of R and R/I has characteristic zero, then R/I is a domain. In this case, either the derivation vanishes on R/I or R/I has infinite rank over its subring of constants (Lemma 2.5), so $I\in\mathfrak{T}$ if and only if $\delta(R)\subseteq I$. This gives us an easier statement when R is a $\mathbb{Q}$-algebra, using the fact that a ring is Jacobson if and only if its prime homomorphic images are.

Corollary 4.2. *Let R be a commutative noetherian differential $\mathbb{Q}$-algebra. Then $R[\theta;\delta]$ is a Jacobson ring if and only if $R/\delta(R)R$ is a Jacobson ring.* □

In particular, we obtain the following generalization of the commutative case of [10, Theorem 4].

Corollary 4.3. *If R is a commutative noetherian differential $\mathbb{Q}$-algebra and $\delta(R)R=R$, then $R[\theta;\delta]$ is a Jacobson ring.* □

These results are actually very easy to prove for $\mathbb{Q}$-algebras, and a brief look at the proof indicates why we proceed as we do for the general case. If R is a commutative noetherian differential $\mathbb{Q}$-algebra, and P is a prime ideal in the ring $T=R[\theta;\delta]$, then the δ-prime ideal $I=P\cap R$ of R is actually prime. If $P=IT$, then $J(T/P)=0$ by [9, Corollary 2.5]. If $P>IT$, then we let $Q=P/IT$ and identify T/IT with $(R/I)[\theta;\delta]$, noting that $Q\cap(R/I)=0$ and $Q\neq 0$. Such an ideal cannot exist unless $\delta(R)\subseteq I$, since if K is the quotient field of R/I and δ is non-trivial on R/I, then $K[\theta;\delta]$ is simple. Hence, to show that $R[\theta;\delta]$ is Jacobson, we need only consider ideals I containing $\delta(R)$. In this case, Jordan's theorem shows that our condition implies that $R[\theta;\delta]$ is Jacobson, and the fact that if I is such an ideal then R/I is a homomorphic image of T shows that the condition is necessary.

The proof in the general case follows similar lines until the final step, where we cannot conclude that if R/I is of finite rank over its subring of constants, then R/I is a homomorphic image of $R[\theta;\delta]$. Instead, we are able to show that $R[\theta;\delta]$ has a homomorphic image which is faithful and finitely generated as an (R/I)-module, which turns out to be sufficient to carry out the argument. We now turn to the lemmas which develop the properties of those ideals in T whose factors are finitely generated over R.

Lemma 4.4. *Let R be a commutative noetherian differential ring, and set $T=R[\theta;\delta]$. Assume that R is δ-prime, and let D be its subring of constants, so that D is a domain and R is a torsion-free D-module. Then the following conditions are equivalent:*

(a) *As a D-module, R has finite rank. (In case R has characteristic zero, this happens only if* $\delta=0$.)

(b) *There exists a prime ideal Q in T such that* $Q\cap R=0$ *and T/Q is finitely generated as a left R-module.*

(c) *There exists a nonzero two-sided ideal Q in T such that* $Q\cap R=0$.

Proof. (b) $\Rightarrow$ (c): Since ${}_RT$ is not finitely generated, Q must be nonzero.

(c) $\Rightarrow$ (a): First assume that R has characteristic zero, so that R is a domain (by Corollary 1.4). If $\delta\neq 0$, then δ extends to a nonzero derivation on the quotient field K of R, and the ring $K[\theta;\delta]$ is simple (e.g., [8, Theorem 3], or Lemma 2.2 and Theorem 2.3). It follows that if I is any nonzero two-sided ideal of T, then the ideal KI of $K[\theta;\delta]$ is all of $K[\theta;\delta]$, which means that $I\cap R\neq 0$. This shows that if R has characteristic zero, then (c) implies that $\delta=0$.

Now suppose that R has characteristic $p>0$. We let $S=D-\{0\}$, so that D_S is the quotient field of D, and we will show that R_S is δ-simple. The δ-primeness of R implies that R_S is at least δ-prime. Thus if I is a nonzero δ-ideal of R_S, then I is not nilpotent, and therefore not nil. As a result, $x^p\neq 0$ for some $x\in I$. Now $\delta(x^p)=0$, and we note that D_S is the subring of constants of R_S, so $x^p\in D_S$. It follows that x^p is invertible in R_S, whence $I=R_S$. Therefore R_S is indeed δ-simple.

The δ-primeness of R implies that all elements of S are non-zero-divisors in R and thus also in T, so that the localization map $T\to T_S$ is injective. In particular, $QT_S\neq 0$. Since $Q\cap R=0$, we know that Q is disjoint from S, whence $QT_S\neq T_S$. Thus, T_S is not a simple ring. Since R_S is δ-simple, Theorem 2.3 shows that R_S must be finite-dimensional over D_S, from which we conclude that R has finite rank over D.

(a) $\Rightarrow$ (b): We make R into a left T-module in the usual way, with a module multiplication $*$ such that $\theta * r=\delta(r)$ for all $r\in R$. Let J be the annihilator in T of the module ${}_TR$, so that J is a two-sided ideal of T and $J\cap R=0$. We will show that ${}_R(T/J)$ is finitely generated, assuming that (a) holds. This will suffice to prove (b), for if Q is a two-sided ideal of T containing J and maximal with respect to the property that $Q\cap R=0$, then it is easy to verify that Q is prime, and ${}_R(T/Q)$ is finitely generated.

We next note that if t is a monic operator in T, then T/Tt is finitely generated free as a left R-module (of the same rank as the order of t). To prove that ${}_R(T/J)$ is finitely generated, therefore, it will suffice to find a finite set $\{t_1,\ldots,t_n\}$ of monic operators in T such that $J\supseteq\bigcap Tt_i$. This we do as follows.

For any $y\in R$, the ideals $Ry+R\delta(y)+\ldots+R\delta^i(y)$ form an increasing sequence, which must stabilize. Consequently, for some positive integer m and elements $r_i\in R$, we have

$$\delta^m(y)=r_0y+r_1\delta(y)+\ldots+r_{m-1}\delta^{m-1}(y).$$

Let $t(y)=\theta^m-(r_0+r_1\theta+\ldots+r_{m-1}\theta^{m-1})$; then $t(y)$ is a monic operator in T such that $t(y)*y=0$. We now use the fact that R is of finite rank over D to choose a maximal D-independent subset $\{y_1,\ldots,y_n\}$ in R, and we let $I=\bigcap_{i=1}^{n}Tt(y_i)$. If $r\in R$, then there is some nonzero $d\in D$ such that rd is a D-linear

combination of the elements y_i, so that if $x \in I$, then $x*rd=0$, or, $(x*r)d=0$. This means that $x*r$ is a D-torsion element of R; hence $x*r=0$, since R is a torsion-free D-module. Therefore, $x*R=0$, so $x \in J$. This shows that $I \subseteq J$, and therefore that ${}_R(T/J)$ is finitely generated, which completes the proof of the lemma. □

Proposition 4.5. *Let R be a commutative noetherian differential ring, and let I_0 and I be δ-prime ideals of R with $I_0 \subseteq I$. Set $T=R[\theta;\delta]$, and let Q_0 be a prime ideal of T such that $Q_0 \cap R = I_0$. If ${}_R(T/Q_0)$ is finitely generated, then there exists a prime ideal Q in T such that $Q_0 \subseteq Q$ and $Q \cap R = I$.*

Proof. By Proposition 1.5, there is a prime ideal P in R such that $(P:\delta)=I$. As $Q_0 \cap R \subseteq P$, no elements of $R-P$ lie in Q_0, hence $Q_0 T_P$ is a proper two-sided ideal of T_P. Now $T_P/Q_0 T_P$ is a nonzero finitely generated left R_P-module, so $P(T_P/Q_0 T_P) \neq T_P/Q_0 T_P$, by Nakayama's Lemma. Thus $PT_P + Q_0 T_P \neq T_P$, and consequently $IT_P + Q_0 T_P \neq T_P$. As a result,

$$(IT_P + Q_0 T_P) \cap R_P \subseteq PR_P.$$

Setting $J=IT+Q_0$, we infer that J is a two-sided ideal of T such that $I \subseteq J \cap R \subseteq P$. But $J \cap R$ is a δ-ideal of R, so we must have $J \cap R = I$, because $(P:\delta)=I$.

Let $Q_1, \ldots, Q_n$ be the prime ideals of T minimal over J. Then $(Q_1 Q_2 \cdots Q_n)^k \subseteq J$ for some positive integer k, and consequently

$$[(Q_1 \cap R)(Q_2 \cap R) \cdots (Q_n \cap R)]^k \subseteq J \cap R = I.$$

Since I is δ-prime, it follows that $Q_j \cap R \subseteq I$ for some j. On the other hand, $I \subseteq J \subseteq Q_j$, hence $Q_j \cap R = I$. As $Q_0 \subseteq J \subseteq Q_j$, the proof is complete. □

We can regard Proposition 4.5 as a "going up" theorem. Notice that even if $\delta=0$, the going up result may fail for prime ideals which do not satisfy the finite generation condition. For example, if $R=K[\![t]\!]$ (for some field K and indeterminate t), and $\delta=0$, then the ideal $Q_0=(\theta t-1)R[\theta]$ is a maximal ideal of $R[\theta]$ even though $Q_0 \cap R=0$.

Corollary 4.6. *Let R be a commutative noetherian differential ring, and set $T=R[\theta;\delta]$. If M is a maximal two-sided ideal of T such that ${}_R(T/M)$ is finitely generated, then $R/(M \cap R)$ is a δ-simple ring.*

Proof. Let $I_0 = M \cap R$, which is a δ-prime ideal of R, and choose a maximal δ-ideal I of R containing I_0. Then I is δ-prime, and Proposition 4.5 provides us with a prime ideal Q in T such that $M \subseteq Q$ and $Q \cap R = I$. As M is maximal, $Q=M$, and hence $I=I_0$, so that I_0 is a maximal δ-ideal of R. Therefore R/I_0 is δ-simple. □

Proof of Theorem 4.1. Set $T=R[\theta;\delta]$. We suppose first that R/I is Jacobson for all $I \in \mathfrak{T}$, and show that T is Jacobson. Let Q be a prime ideal of T, and let $I=Q \cap R$. From [9, Corollary 2.5], we obtain $J(T/IT)=0$, so there is no problem if $Q=IT$. If $Q>IT$, then by applying Lemma 4.4 to the ring R/I, we find that $I \in \mathfrak{T}$, so that R/I is Jacobson. In this case, T/IT is Jacobson by [9, Theorem 3.5], so $J(R/Q)=0$. Therefore T is a Jacobson ring.

To prove the converse, we suppose that T is Jacobson and that $I \in \mathfrak{T}$, and show that R/I is Jacobson. Without loss of generality, we may assume that $I=0$, so that R is δ-prime and is torsion-free of finite rank over its subring of constants.

First suppose that R has characteristic zero. Then according to Lemma 2.5, $\delta=0$, so that T is an ordinary polynomial ring over R. This implies that R is a homomorphic image of T, and therefore that R is Jacobson.

Now suppose that R has positive characteristic p, and that P is a prime ideal in R. Let $H=(P:\delta)$, and note that H is a δ-prime ideal of R. We infer from [6, Lemma 13] that P/H is nilpotent, and so P is the prime radical of H. According to Lemma 4.4, there is a prime ideal Q_0 in T such that $Q_0 \cap R=0$ and ${}_R(T/Q_0)$ is finitely generated. From Proposition 4.5, we obtain a prime ideal Q in T such that $Q \cap R=H$ and $Q_0 \subseteq Q$, so that, in particular, ${}_R(T/Q)$ is finitely generated. We will use the fact that $J(T/Q)=0$ to show that $J(R/P)=0$.

Since ${}_R(T/Q)$ is finitely generated, T/Q is isomorphic to a subring of the endomorphism ring of a finitely generated module over a commutative ring. Thus T/Q satisfies a standard polynomial identity. Consequently, it follows from Kaplansky's Theorem that all primitive ideals of T/Q are maximal [13, Chapter II, Corollary 1.3]. Now if $\mathfrak{M}$ is the set of maximal two-sided ideals of T containing Q, then $\mathfrak{M}$ includes all primitive ideals containing Q, and thus (since T is Jacobson) we have $\bigcap \mathfrak{M}=Q$.

For each $M \in \mathfrak{M}$, let $P(M)$ be the prime radical of the ideal $M \cap R$ in R. Since M is a maximal two-sided ideal of T and ${}_R(T/M)$ is finitely generated (because $Q \subseteq M$), Corollary 4.6 shows that $R/(M \cap R)$ is δ-simple. Consequently, $M \cap R=(N:\delta)$ for any maximal ideal N of R which contains $M \cap R$. Since R has positive characteristic, it follows from [6, Lemma 13] that $R/(M \cap R)$ is an artinian local ring, and hence that $P(M)$ is a maximal ideal of R. Also, since the p-th. powers of elements of $P(M)$ generate a δ-ideal, it follows from the δ-simplicity of $R/(M \cap R)$ that $x^p \in M \cap R$ for all $x \in P(M)$.

Let $K=\bigcap_{M \in \mathfrak{M}} P(M)$. Then K is a semiprime (in fact, semiprimitive) ideal of R, and

$$K \supseteq \bigcap_{M \in \mathfrak{M}} (M \cap R)=Q \cap R=H,$$

so $K \supseteq P$ (because P is the prime radical of H). On the other hand, given $x \in K$, we have $x^p \in M \cap R$ for all $M \in \mathfrak{M}$, so $x^p \in H$, and thus $x \in P$. Therefore $K=P$, which proves that $J(R/P)=0$. This completes the proof that R is a Jacobson ring, and the proof of Theorem 4.1. □

References

1. Amitsur, S.A.: A generalization of a theorem on linear differential equations. Bull. Amer. Math. Soc. **54**, 937-941 (1948)
2. Amitsur, S.A.: Derivations in simple rings. Proc. London Math. Soc. (3) **7**, 87-112 (1957)
3. Carlson, R.C., Goodearl, K.R.: Commutants of ordinary differential operators. J. Differential Equations **35**, 339-365 (1980)
4. Cozzens, J.H., Faith, C.: Simple Noetherian Rings. Cambridge: Cambridge University Press 1975

5. Dixmier, J.: Enveloping Algebras. Amsterdam: North-Holland 1977
6. Goodearl, K.R.: Global dimension of differential operator rings III. J. London Math. Soc. (2) **17**, 397–409 (1978)
7. Hart, R.: Krull dimension and global dimension of simple Ore-extensions. Math. Z. **121**, 341–345 (1971)
8. Jacobson, N.: Pseudo-linear transformations. Ann. of Math. (2) **38**, 484–507 (1937)
9. Jordan, D.A.: Noetherian Ore extensions and Jacobson rings. J. London Math. Soc. (2) **10**, 281–291 (1975)
10. Jordan, D.A.: Primitive Ore extensions. Glasgow Math. J. **18**, 93–97 (1977)
11. McConnell, J.C.: Representations of solvable Lie algebras V: On the Gelfand-Kirillov dimension of simple modules. Preprint
12. McCoy, N.H.: Prime ideals in general rings. Amer. J. Math. **71**, 823–833 (1949)
13. Procesi, C.: Rings with Polynomial Identities. New York: Dekker 1973
14. Seidenberg, A.: Differential ideals in rings of finitely generated type. Amer. J. Math. **89**, 22–42 (1967)

Received November 9, 1981; received in final form March 3, 1982

Math. Z. 180, 525-534 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

The Theory and Structure of Dual *JB*-Algebras

Leslie J. Bunce

Department of Mathematics, Reading University, Reading RG6 2AX, England

Introduction

There is a highly developed and well-known theory (originated by Kaplansky [17]) of the important class of dual C^*-algebras. In this paper we introduce a theory of dual rings for *JB*-algebras and obtain numerous characterizations of dual *JB*-algebras. Also introduced is a sub-class of dual *JB*-algebras, called compact *JB*-algebras, in terms of which dual C^*-algebras can be characterized, providing the connection between the respective theories.

The reader is referred to [2, 8, 19, 20] for properties of dual C^*-algebras. The general theory of *JB*-algebras can be found in [3–7, 12–15, 21, 22, 30]. The standard reference for the algebraic theory of Jordan algebras is [16].

§ 0. Preliminaries

We recall that a *JB-algebra* is a real Jordan Banach algebra for which the norm obeys: $\|a^2-b^2\| \leqq \max(\|a^2\|, \|b^2\|)$, $\|a^2\| = \|a\|^2$. The second dual, A^{**} of a *JB*-Algebra A is a *JBW*-algebra (i.e. a *JB*-algebra with predual), in the Arens product, which contains A, [14, 21], and we write $\tilde{A}$ for the *JB*-subalgebra of A^{**} generated by A and the unit, 1, of A^{**}.

Let A be a *JB*-algebra, and let a, b, c be elements of A. The *Jordan triple product* $\{a, b, c\}$, is defined by $\{a, b, c\} = (a \circ b) \circ c + a \circ (b \circ c) - (a \circ c) \circ b$. L_a denotes the left (right) multiplication operator on A and $U_a = 2L_a^2 - L_{a^2}$. $C(1, a, b)$ means the *JB*-subalgebra of $\tilde{A}$ generated by $1, a, b$, which, by [30, Proposition 2.1], can be identified with a *JC*-algebra. When $a = b$, this is an associative algebra and can be realised as the self-adjoint part of a C^*-algebra, by [6, Proposition 2.3]. The set $\sigma(a) = \{\lambda \in \mathbb{R};\ a - \lambda 1$ is invertible in $C(1, a)\}$ is the *spectrum* of a.

We say that a projection p of A^{**} is *open* if there is a norm closed quadratic ideal J of A, whose weak* closure in A^{**}, equals $U_p(A^{**})$, equivalently, there is an increasing net (a_α) in J^+ with $a_\alpha \uparrow p$, in the sense of [6]. The results of [12, 13] imply a bijective order preserving correspondence from the norm closed quadratic ideals of A onto the open projections of A^{**}.

0025-5874/82/0180/0525/$02.00

A JB-algebra A is said to be *exceptional* if each factor representation of A is onto M_3^8 (the hermitian 3×3 matrices over the Cayley numbers), where we have used the term factor in the sense given in [6]. Every JB-algebra A possesses a norm closed Jordan ideal J such that J is an exceptional JB-algebra and A/J is isomorphic to a JC-algebra. In this respect J is unique and called the exceptional ideal of A [6, Theorem 9.5] (the non-unital version is obtained by passing from A to $\tilde{A}$).

Let A be a JB-algebra and let ϕ be in $P(A)$, the set of pure states of A. By the terms $s(\phi)$, $c(\phi)$, we mean respectively, the *support* and *central support* of ϕ in A^{**} as defined in [6]. The set $\{s(\phi)\colon \phi\in P(A)\}$ consists of all the minimal projections in A^{**} [3]. Each ϕ in $P(A)$ gives rise to a Jordan homomorphism $\pi_\phi\colon A\to c(\phi)\circ A^{**}$, $c(\phi)$ being a minimal central projection in A^{**} [6]. The *primitive ideal space* of A is defined to be the set $\mathrm{Prim}(A)=\{\ker\pi_\phi\colon \phi\in P(A)\}$. Some elementary properties of primitive ideals are describe, briefly, in [10]. By considering (pure) state extensions of A to $\tilde{A}$ for a non-unital JC-algebra A, Størmers proof [26, Theorem 7.1] can be adapted to show that the quadratic ideal $L_\phi=\{a\in A\colon \phi(a^2)=0\}$ is maximal norm closed if and only if ϕ is pure, and that in fact there is a bijective correspondence between the pure states of A and the maximal norm closed quadratic ideals of A given by $\phi\to L_\phi$. That the exceptional JB-algebra also enjoys this property is easier to see, following, for example, from [10, Lemma 4.3] and 29, Lemma 3.2. That the result is true for an arbitrary JB-algebra A is now more or less immediate from the fact that for each ϕ in $P(A)$, $A/\ker\pi_\phi$ is isomorphic to a JC-algebra or to M_3^8. The proof of [31, 3.13.6 (ii)$\Rightarrow$(iii)] then goes through with minor modifications to show that the set $\{1-s(\phi);\ \phi\in P(A)\}$ constitute the maximal ($\neq 1$) open projections of A in A^{**}. In summary, J is a maximal norm closed quadratic ideal of A if and only if it is of the form $J=U_{1-e}(A)^{**}\cap A$ where e is a minimal projection in A^{**} and when J is of this form the w^*-closure of J in A^{**} equals $U_{1-e}(A^{**})$.

§ 1. Dual JB-algebras

Let A be a JB-algebra. For each subset S of A define,

$$q(S)=\{a\in A;\ U_s(a^2)=0 \text{ for every } s \text{ in } S\}.$$
$$S^0=\{a\in A;\ a\circ S=\{0\}\}.$$

Definition. A JB-algebra A is *dual* if $(J^0)^0=J$ for every norm closed quadratic ideal J of A.

The following fact will be used repeatedly. The proof is elementary, and we give brief details only.

Lemma 1.1. *Let A be a JB-algebra. For each norm closed quadratic ideal J and each subset S of A we have*

(i) *$q(S)$ is a norm closed quadratic ideal*

(ii) $q(J)=J^0$.

Proof. (i) For x, y in $q(S)$ and s in S, $(x+y)^2 \leqq 2(x^2+y^2)$, hence $0 \leqq U_s((x+y)^2) \leqq 2U_s(x^2)+2U_s(y^2)=0$.

Also for each a in A,

$$0 \leqq U_s((U_x(a))^2) = U_s U_x U_a(x^2) \leqq U_s(x^2) \, \| U_a(x^2) \| = 0.$$

where we have used the Jordan identity $(U_b(c))^2 = U_b U_c(b^2)$ [16, (62), page 37].

(ii) Let y be in J. If x is an element of $q(J)$, then $U_y(x^2)=0$. By working in $C(1, x, y)$ simple calculations show that $y \circ x = 0$. On the other hand, using the Jordan identity $4(b \circ c)^2 = 2b \circ U_c(b) + U_c(b^2) + U_b(c^2)$ [6, page 21], we have that, for z in J^0, $U_y(z^2)=0$.

The concept of the multiplier algebra of a C^*-algebra has been extended to *JB*-algebras by Edwards [15]. In particular for each C^*-algebra (resp. *JB*-algebra) A there exists a largest C^*-subalgebra (resp. *JB*-subalgebra) $M(A)$ of A^{**} which contains A as a twosided ideal (resp. Jordan ideal). $M(A)$ is called the *multiplier algebra* of A in A^{**}.

Lemma 1.2. *Let A be a JB-algebra. If $M(A)=A^{**}$, then A is Dual.*

Proof. Let $M(A)=A^{**}$. Observe that all projections of A^{**} are open. Indeed, if p is a projection in A^{**} there is an increasing set (a_α) in A^+ such that $a_\alpha \uparrow p$, by the proof of [16, Theorem 9(i)]. With $J = U_p(A^{**}) \cap A$ we see that the a_α are in J, by hereditary considerations, [12–14]. Hence $\bar{J}^{\omega *} = U_p(A^{**})$. Suppose now that I is any norm closed quadratic ideal of A and let p be the open projection of I. Then $I = U_p(A)$, since $U_p(A) \subseteq A$. Then $U_{1-p}(A) \subseteq I^0$, whereas for each x in I^0 we have $x \circ p = 0$, and so $I^0 \subseteq U_{1-p}(A)$. The desired result now follows by symmetry.

Whenever A is a JC-algebra, the C^*-algebra generated by A is denoted by $[A]$.

Lemma 1.3. *Let A be a JC-algebra. If $[A]$ is a dual C^*-algebra, then A is a dual JB-algebra.*

Proof. Identify $[A]$ with its image in its universal representation space $B(H)$, say. Then $[A]^{**}$ and A^{**} can be identified with the weak operator closures in $B(H)$, $[A]^-$ and A^-, respectively; moreover $A = [A] \cap A^-$ [24]. If $[A]$ is a dual C^*-algebra, then $M([A]) = [A]^{**}$ [8, Theorem 5.5]. So, $A \circ A^- \subset [A][A]^- \subset [A]$. Hence, $A \circ A^- \subset [A] \cap A^- = A$, so A is a dual *JB*-algebra, by Lemma 1.2.

We recall, from [10], that an element x of a *JB*-algebra A is said to be *abelian* if it is positive and $U_x(A)$ is an associative algebra, and that A is said to be a *Type* I *JB-algebra* if each *JB* quotient of A contains a non-zero abelian element (see [10] for details). A *JB*-algebra is said to be *simple* if it contains no non-trivial norm closed Jordan ideals.

Corollary 1.4. *Simple Type* I *JB-algebras are dual.*

Proof. Let A be a simple Type I *JB*-algebra. If $A = M_3^8$, then A is dual by infinite dimensionality and Lemma 1.2. Thus, by the structure theorem [6, Theorem 9.3] (and its non-unital version if necessary) we can suppose that A is

a JC-algebra and further that A is irreducibly embedded in some $B(H)$, H a complex Hilbert space (take any irreducible representation of $[A]$). Then, by [10, Theorem 5.6], either $A \cap C(H) \neq 0$ or A is an infinite dimensional spin factor. In the first case (by simplicity) $A \subseteq C(H)$ and A is dual by Lemmas 1.2, 1.3. In the second case, when A is a spin factor, we again see that A is dual, by Lemma 1.2, since A^{**} also has the structure of a spin factor and $A = A^{**}$ by [26, Lemma 5.1] or [28, Corollary, p. 1060].

The $JB-(\infty)$ sum of a family of JB-algebras (A_λ), written $(\sum A_\lambda)_0$, is the JB-algebra of all sequence (x_λ) $(x_\lambda \in A_\lambda)$ vanishing at infinity, with the supremum norm.

For future reference we include the following lemma, its proof is similar to that of Kaplansky's for C^*-algebras (see [20, 4.10.25], for example) and we omit the details.

Lemma 1.5. *If (A_λ) is a family of dual JB-algebras then $(\sum A_\lambda)_0$ is a dual JB-algebra.*

§ 2. Annihilator JB-algebras

Definition. An *annihilator JB-algebra* is a JB algebra A for which J^0 is non-zero for each proper, norm closed, quadratic ideal J of A.

Proposition 2.1. *Let A be an annihilator JB-algebra and let M be a maximal norm closed quadratic ideal of A. Then $M = U_{1-p}(A)$ where p belongs to A and is a minimal projection of A^{**}.*

Proof. By assumption and Lemma 1.1, M^0 is a non-zero norm closed quadratic ideal of A. Choose an element x in $(M^0)^+$ with $\|x\| = 1$. Let I be the norm closure of $U_{1-x}(A)$ in A. Note that I is a quadratic ideal of A, moreover observe that $I \neq A$. Indeed, if $I = A$ then $U_{1-x}(\tilde{A})$ is norm dense in $\tilde{A}$. But then there is an element a in $\tilde{A}$ for which $U_{1-x}(a)$ is invertible in $\tilde{A}$, and so $1-x$ is invertible in $\tilde{A}$ [16, p. 52], which is impossible. Thus, there is a non-zero element y in I^0, by assumption. Note that $y \in q(\{z\})$, a quadratic ideal by Lemma 1.1, where $z = x - x^2$, and that $M \subseteq q(\{z\})$. If $z \neq 0$, we have $M = q(\{z\})$, from which it is easy to see that $y = 0$, a contradiction. Therefore, x is a projection in A and we see that $M = U_{1-x}(A)$. By the remarks made in the least paragraph of Sect. 0, or by the fact that the above proof shows that x is the only element of norm 1 in M^0, x is a minimal projection in A. This completes the proof.

We say that a projection p in the JB-algebra A is *1-dimensional* if $\dim U_p(A) = 1$. Such projections are clearly minimal in A^{**}. Also, for any subset S of A we say that the smallest Jordan (resp. quadratic) ideal of A containing S is the Jordan (resp. quadratic) ideal generated by S.

Corollary 2.2. *Let A be an annihilator JB-algebra. Then*

(i) *A contains the minimal projections of A^{**}.*

(ii) *The Jordan ideal generated by the 1-dimensional projections of A is dense in A.*

(iii) *A contains a familly of simple norm closed Jordan ideals (J_λ) whose sum is dense in A. In fact $A \simeq (\sum_\lambda J_\lambda)_0$.*

(iv) *A is a Type* I *JB-algebra.*

Proof. (i) Let e be a minimal projection in A^{**}. Then, by the remarks in § 0, $M = U_{1-e}(A^{**}) \cap A$ is a maximal norm closed quadratic ideal of A. Proposition 2.1 implies that $M = U_{1-e}(A)$ and that e belongs to A.

(ii) Let J be a norm closed quadratic ideal of A which contains all 1-dimensional projections. If J is proper then there is a 1-dimensional projection e of A with $J \subseteq U_{1-e}(A)$, by Proposition 2.1. It follows that $e=0$, a contradiction.

(iii) Let ϕ belong to $P(A)$. By (i), $s(\phi)$ lies in A. Hence $J_\phi = (c(\phi) \circ A^{**}) \cap A$ is non-zero; it is simple because $c(\phi)$ is a minimal central projection. The first statement now follows from (ii), from which the second statement is deduced by standard C^*-algebra techniques.

(iv) This is immediate from (i) and (ii), since the minimal projections of A^{**} are abelian elements of A.

Lemma 2.3. *Let A be a JB-algebra. The following conditions on A are equivalent.*

(i) $\mathrm{Prim}(A)$ *is discrete*

(ii) *A is isomorphic to the $JB-(\infty)$ sum of the family $\{A/P;\ P \in \mathrm{Prim}(A)\}$, and A/P is a simple JB-algebra for each P in* $\mathrm{Prim}(A)$

(iii) *$A \simeq (\sum A_\lambda)_0$ where (A_λ) is a family of simple JB-algebras.*

Proof. This is similar to the usual C^*-proof (see [20, 4.9.24]).

Corollary 2.4. *Let A be a JB-algebra. The following conditions on A are equivalent.*

(i) *A is an annihilator JB algebra*

(ii) *A is a dual JB-algebra*

(iii) *A is a Type* I *JB-algebra and* $\mathrm{Prim}(A)$ *is discrete.*

Proof. (ii)$\Rightarrow$(i) is immediate from definitions.

(i)$\Rightarrow$(ii) follows from Corollary 2.2(ii), (iii), Corollary 1.4 and Lemma 1.5.

(ii)$\Rightarrow$(iii) is a consequence of Corollary 2.2. (iv) and Lemma 2.3.

(iii)$\Rightarrow$(ii) can be seen by combining Lemma 2.3 with Corollary 1.4 and then using Lemma 1.5.

Proposition 2.5. *Let A be a dual JB-algebra. Then $M(A) = A^{**}$.*

Proof. By Corollary 2.3(iii) we can choose a family (J_λ) of simple Type I Jordan ideals of A whose sum is norm dense in A. Lemma 1.3 and Corollary 1.4 actually show that $M(J_\lambda) = J_\lambda^{**}$ for each λ. In particular each J_λ is a Jordan ideal of A^{**}. Let b belong to A, there is a sequence (b_n) in A for which $\|b_n - b\| \to 0$ and each b_n is of the form $x_{\lambda_1} + \ldots + x_{\lambda_n}$ where $x_{\lambda_i} \in J_{\lambda_i}$. For each a in A^{**}, $a \circ b_n \in A$ and thus $a \circ b_n \in A$. Hence, $M(A) = A^{**}$.

Corollary 2.6. *Each JB-subalgebra and each JB-quotient of a dual JB-algebra is a dual JB-algebra.*

Proof. Let A be a dual JB-algebra. If J is a norm closed Jordan ideal of A then A/J is seen to be dual on application of Corollary 2.4(i)$\Leftrightarrow$(iii).

Suppose that B is a JB-subalgebra of A. By Proposition 2.5, $M(A)=A^{**}$. Considering $B^{**}\subseteq A^{**}$ one sees easily (as in [8, Corollary 2.3]) that $M(B)=B^{**}$ and thus that B is dual by Lemma 1.3.

§ 3. Ideals and Associate Subalgebras

By combining Corollary 2.6 with the statement and proof of Lemma 1.2 one sees at once that for a JB-algebra A all projections in A^{**} are open and that each norm closed quadratic ideal of A is of the form $U_p(A)$ for some projection p of A^{**}. Actually, these conditions separately characterise dual JB-algebras.

Proposition 3.1. *Let A be a JB-algebra. Then A is dual if and only if it satisfies one of the two conditions below*

(i) *All projections of A^{**} are open*

(ii) *Each norm closed quadratic ideal of A is of the form $U_p(A)$ for some projection p of A^{**}.*

Proof. In view of the preceding remarks we need only show that A is dual if it satisfies any one of the conditions (i) and (ii). Suppose that A satisfies (i). Then certainly the minimal projections of A^{**} are open, but this means that they lie in A. Hence A is dual by Corollary 2.4 and the proof of Corollary 2.2.

Suppose now that A satisfied (ii), and let M be a maximal norm closed quadratic ideal of A. By assumption and the remarks in the final paragraph of Sect. 0, $M=U_{1-e}(A)$ where e is a minimal projection of A^{**}. Choose an element a in A such that $\|U_e(a^2)\|=1$. From the Jordan identities [16, p. 37] $(U_x(y))^2=U_xU_u(x^2)$; $U_{\{x,y,x\}}=U_xU_yU_x$ it follows that $U_a(e)$ is a minimal projection of A^{**}, and that the positive element $(U_a(1-e))^2=U_aU_{1-e}(a^2)$ lies in A.

By spectral theory, $a^2-U_a(e)=U_a(1-e)$ lies in A, and hence so does $U_a(e)$. As in the proof of Corollary 2.2(ii), A therefore contains a norm closed Jordan ideal which is also a simple dual JB-algebra. Let (J_λ) be a maximal family of norm closed Jordan ideals in A of this description. Then $J=\sum_\lambda J_\lambda$ is norm dense in A, for if not there is a minimal projection f in A^{**} with $J\subseteq U_{1-f}(A)\subseteq A$, by the initial part of the above argument, which implies that $U_f(J)=0$; and choosing an element x in A^+ (not in J) with $U_f(x)\neq 0$ we see that, as above, $U_x(f)$ generates a non-zero simple dual Jordan ideal of A contradicting the maximality of the (J_λ). Therefore, as in the proof of Corollary 2.2(iii) $A\simeq(\sum_\lambda J_\lambda)_0$, and A is dual by Lemma 2.3. This completes the proof.

Remarks 3.2. (a) The argument used by Berglund in [8, Theorem 5.6] carries over to JB-algebras with virtually no change to show that the JB-algebra A is a Jordan ideal of A^{**} if and only if $L_x\colon A\to A$ is a weakly compact operator on A, for each x in A. A similar argument shows that for each x in A the operator $U_x\colon A\to A$ is weakly compact if and only if A is a quadratic ideal of A^{**}, it is

easy to see that this is equivalent to A being a Jordan ideal of A^{**} (by an argument similar to that used in Proposition 3.1 for example).

(b) The associative JB-algebra A can be realised as the self adjoint part of a commutative C^*-algebra [6, Proposition 2.3], [22]. In fact A is isomorphic (as a JB-algebra) to $C(X)$ (resp. $C_0(X)$) of continuous real valued functions (resp. vanishing at ∞) on some compact (resp. locally compact) Hausdorff space when A is unital (resp. non-unital); we can take $X = \mathrm{Prim}(A)$.

Theorem 3.3. *Let A be a JB-algebra. Then conditions below are equivalent.*

(a) *A is dual*

(b) *A is annihilator*

(c) $M(A) = A^{**}$

(d) $\mathrm{Prim}(A)$ *is discrete and A is a Type* I *JB-algebra*

(e) *Each norm closed quadratic ideal of A is of the form $U_p(A)$ for a projection p in A^{**}*

(f) $A \simeq (\sum_\lambda J_\lambda)_0$ *where the (J_λ) is a family of simple Type* I *Jordan ideals of A*

(g) $A \simeq (\sum_\lambda A_\lambda)_0$ *where the (A_λ) is a family of dual JB-algebras*

(h) *For each x in A, $L_x\colon A \to A$ is a weakly compact operator*

(i) *For each x in A, $U_x\colon A \to A$ is a weakly compact operator*

(j) *Each maximal associative JB-subalgebra of A has discrete spectrum*

(k) *For each x in A, every non-zero point of $\sigma(x)$ is isolated.*

Proof. The proof of the equivalence of (a)–(i) is contained in § 1, § 2 and the remarks 3.2(a). That (a) implies (j) is immediate from Corollaries 2.4 and 2.6. In view of the remarks 3.2(b) the scheme of implications (k)$\Leftrightarrow$(j)$\Rightarrow$(a) can now be obtained as in [19, Theorem 1] with minor alterations. (To complete the proof of (j)$\Rightarrow$(a), the Jordan analogue of [19, Theorem 1(iii)$\Rightarrow$(i)], one uses the method of Corollary 2.2(iii) and Corollary 2.4(i)$\Rightarrow$(ii).)

§ 4. Compact JB-algebras

In [2], Alexander considered C^*-algebras with the property that, for each x in A, the operator on A defined by $a \mapsto xax$ is compact. It was established that these objects, called compact C^*-algebras, are precisely the dual C^*-algebras.

Definition. A JB-algebra A is a *compact JB-algebra* if U_x is a compact operator for each x in A.

Proposition 4.1. *Each compact JB-algebra is a dual JB-algebra.*

Proof. Let A be a compact JB-algebra and let B be a maximal associative subalgebra of A. From [6] and [22] we know that B can be realised as the self adjoint part of a commutative C^*-algebra; moreover, it is obvious that B is a compact algebra. Thus it is seen, from [2, Corollary 8.3], that B is a dual JB-algebra. Hence A is a dual JB-algebra, by Theorem 3.1.

A *spin factor representation* of a JB-algebra A is a Jordan homomorphism $\pi\colon A \to M$ onto a spin factor M. By [28, Corollary, p. 1060] this is equivalent to the requirement that πA be strongly dense in M.

Corollary 4.2. *Let A be a JB-algebra. The following statements are equivalent*

(i) *A is a compact JB-algebra*

(ii) *A is a dual JB-algebra with no infinite dimensional spin factor representations*

(iii) *$A \simeq (\sum_{\lambda} A)_0$ where each A_λ is a simple compact JB-algebra.*

Proof. It is obvious that any JB-quotient of a compact JB-algebra is a compact JB-algebra and that an infinite dimensional spin factor, being unital, cannot be compact. Thus (i) implies (ii), by Proposition 4.1. Assume (ii). It is immediate from the definitions that any JC-subalgebra of a compact C^*-algebra is a compact algebra. Thus, by [2, Corollary 8.3] and the statement and proof of Corollary 1.4 it is seen directly from Lemma 2.3 and Corollary 2.4 that condition (iii) follows.

The proof of the implication (ii)$\Rightarrow$(i) is similar to [2, Lemmas 3.5, 3.6] and will not be repeated for this reason.

Recall that a JC-algebra $A \subseteq B(H)_{sa}$ is said to be *reversible* if whenever $a_1, \ldots, a_n$ belong to A then so does $a_1 \ldots a_n + a_n \ldots a_1$. A simple JC-algebra which is not isomorphic to a spin factor is reversible [10, Lemma 4.3], [29, Lemma 3.2]. Corollary 4.2, together with Lemma 1.4, implies that a compact JC-algebra is isomorphic to the $JB-(\infty)$ sum of a family (A_λ) of simple reversible JC-algebras (A_λ) where $A_\lambda \subseteq C(H_\lambda)$ - the compact operators on some (complex) Hilbert space H_λ, and hence to a reversible JC-subalgebra of $C(H)$, where $H = \bigoplus_{\lambda} H_\lambda$, the Hilbert space direct sum.

For each cardinality k let us denote by V_k the spin factor of orthonormal dimension $k+1$. In addition, let us denote by $C_0(X, A)$, where X is a locally compact Hausdorff space and A a JB-algebra, the continuous A-valued functions on X vanishing at infinity.

Theorem 4.3. *Let A be a JB-algebra, then A is dual if and only if A is isomorphic to a JB-algebra of the form $C_0(X, M_3^8) \oplus (\sum_{k \in K} C_0(S_k, V_k))_0 \oplus C$, where K is a set of cardinalities, X, S_k are discrete and C is a reversible compact JC-algebra.*

Proof. This follows by combining the above remarks with Corollary 4.2 and Theorem 3.3(a)$\Leftrightarrow$(f).

From the observations made above it is easy to see that a compact JC-algebra is isomorphic to a JC-algebra which generates a dual C^*-algebra. However, there is a sharper result which characterizes compact JC-algebras rather neatly.

Recall that we denote by $[A]$ the C^*-algebra generated by the JC-algebra A.

Proposition 4.4. *Let A be a JC-algebra. Then A is a compact JC-algebra if and only if $[A]$ is a dual C^*-algebra.*

Proof. Suppose that A is a compact JC-algebra. Identify $[A]$ with its image in its universal representation space, $B(H_u)$ say. As in [23], we may identify

$[A]^{**}$ with $[A^-]$ and A^{**} with $A^-([A]^-$, A^- representing the weak operator closures in $B(H_u)$). Suppose first that A is simple. Then A^{**} is a factor as can be seen from Theorem 3.3(a)⇒(c), and Corollary 4.2. If $[A]^{**}$ is a factor it is well-known that $[A]$ is (simple) dual. If $[A]^{**}$ is not a factor then [26, Corollary 3.5] shows that $[A]=I+J$ where I, J are simple (norm closed) ideals of $[A]$. As $[A]$ is a Type I C^*-algebra, by [10, Proposition 4.5], I and J must be dual, as therefore must be $[A]$.

Suppose now that A is an arbitrary compact JC-algebra. By Theorem 3.3(a)⇒(f), there is a family of simple norm closed Jordan ideals (J_λ) in A such that $\sum J_\lambda$ is norm dense in A. Moreover, each $J_\lambda=e_\lambda \cdot A$ for some minimal central projection e_λ of A^{**}. Using [27, Proposition 2], one sees that the e_λ are in the centre of $[A]^{**}$. Therefore, the $[J_\lambda]=e_\lambda[A]$ and form an orthogonal family of two sided ideals, [24], which are dual C^*-algebras by the argument above. Finally, it is routine to verify that $\sum_\lambda [J_\lambda]$ is norm dense in $[A]$. Therefore $[A]\simeq(\sum_\lambda [J_\lambda])_0$ and so is dual.

The converse is clear from Lemma 1.3, Corollary 4.2 and [10, Lemma 4.2].

Concluding remarks. One can take the analogy with C^*-algebras to the limit of considering the JB-version of completely continuous (c.c.) C^*-algebras, [17]. By defining completely continuous JB-algebras as being that class of JB-algebras for which $L_x\colon A\to A$ is a compact operator for each x in A and using techniques similar to those already displayed here, one shows that the JB-algebra A is completely continuous if and only if it is isomorphic to the JB $-(\infty)$ sum of (simple) finite dimensional JB-algebras, and that when A is a JC-algebra this is equivalent to $[A]$ being a completely continuous C^*-algebra.

Finally, one of the reasons why dual JB-algebras might be considered to be of interest is that (as in the case for C^*-algebras) they provided an essentially algebraic characterization of Type I JB-algebras, as can be seen in the statement, the proof of which is now more or less immediate from [10, Theorem 5.6].

Corollary 4.5. *A JB-algebra A is a Type* I *JB-algebra if and only if A/P contains a non-zero (norm closed) dual Jordan ideal for each $P\in\mathrm{Prim}(A)$.*

Acknowledgement. The author is pleased to acknowledge the advice and encouragement given to him by his Ph.D. Supervisor, Professor J.D.M. Wright. The paper of Harris [32] contains certain generalisations of some of our results. We are grateful to the referee for pointing this out to us. The results of [32] combined with those of Wright [30] provide alternative arguments at a number of points. As an example, Theorems 3.3 and 5.9 of [32] give directly, the implications Corollary (ii)⇒(iii), and Theorem 3.3(k)⇒(f), which may be used to replace the relevant part of the argument in the latter half of the proof of Theorem 3.3.

References

1. Akemann, C.A., Pedersen, G.K., Tomiyama, J.: Multipliers in C^*-Algebras. J. Functional Analysis **13**, 277-301 (1973)
2. Alexander, J.A.: Compact Banach algebras. Proc. London Math. Soc. (3) **18**, 1-18 (1968)

3. Alfsen, E.M., Shultz, F.W.: A non-commutative Spectral theory and Jordan algebras. Proc. London Math. Soc. (3) **38**, 497–516 (1979)
4. Alfsen, E.M., Shultz, F.W.: State Spaces of Jordan Algebras. Acta. Math. **140**, 155–190 (1978)
5. Alfsen, E.M., Shultz, F.W., Hanche-Olsen, H.: State spaces of C^*-algebras. Acta Math. **144**, 267–305 (1980)
6. Alfsen, E.M., Shultz, F.W., Størmer, E.: A Gelfand Neumark Theorem for Jordan algebras. Advances in Math. **28**, 11–56 (1978)
7. Behncke, H.: Hermitian Jordan Banach Algebras. J. London Math. Soc. (2) **20**, 306–312 (1977)
8. Berglund, M.C.F.: Ideal C^*-algebras. Duke J. Math. **40**, 241–257 (1973)
9. Bonsall, F.F., Duncan, J.: Complete Normed Algebras. Berlin-Heidelberg-New York: Springer 1973
10. Bunce, L.J.: Type I JB-algebras. Quart. J. Math. Oxford, Ser. 2 (To appear)
11. Dixmier, J.: C^*-Algebras. Amsterdam: North Holland 1977
12. Edwards, C.M.: Ideal theory in JB-algebras. J. London Math. Soc. (2) **16**, 507–513 (1977)
13. Edwards, C.M.: On the facial structure of a JB-algebra. J. London Math. Soc. (2) **19**, 335–344 (1979)
14. Edwards, C.M.: On the centres of hereditary JBW-algebras of a JBW-algebra. Math. Proc. Cambridge Philos. Soc. **85**, 317–325 (1979)
15. Edwards, C.M.: Multipliers of JB-algebras. Math. Ann. **249**, 265–272 (1980)
16. Jacobson, N.: Structure and Representations of Jordan Algebras. Amer. Math. Soc. Colloquium Publications **39**. Providence R.I.: American Mathematical Society 1968
17. Kaplansky, J.: Dual Rings. Ann. Math. (2) **49**, 689–701 (1948)
18. Kaplansky, I.: The structure of certain operator algebras. Trans. Amer. Math. Soc. **70**, 219–255 (1951)
19. Ogasarawa, T., Yoshinaga, K.: A characterization of Dual B^*-algebras. J. Sci. Hiroshima Univ. Ser. A1 **18**, 179–182 (1954)
20. Rickart, C.E.: General Theory of Banach Algebras. Princeton, N.Y.: Van Nostrand 1960
21. Shultz, F.W.: On normed Jordan algebras which are Banach Dual Spaces. J. Functional Analysis **31**, 360–376 (1979)
22. Smith, R.R.: On non-unital Jordan Banach Algebras. Math. Proc. Cambridge Philos. Soc. **82**, 375–380 (1977)
23. Størmer, E., Effros, E.: Jordan algebras of self adjoint operators. Trans. Amer. Math. Soc. **127**, 312–315 (1967)
24. Størmer, E.: On the Jordan structure of C^*-algebras. Trans. Amer. Math. Soc. **120**, 438–447 (1965)
25. Størmer, E.: Jordan algebras of Type I. Acta. Math. **115**, 165–184 (1968)
26. Størmer, E.: Irreducible Jordan algebras of self adjoint operators. Trans. Amer. Math. Soc. **130**, 153–166 (1968)
27. Topping, D.: Jordan algebras of self adjoint operators. Mem. Amer. Math. Soc. **53**, (1965)
28. Topping, D.: An isomorphism invariant for spin factors. J. Math. Mech. **15**, 1055–1063 (1965)
29. Upmeier, H.: Derivation Algebras of JB^*-algebras. Manuscripta Math. **30**, 199–214 (1979)
30. Wright, J.D.M.: Jordan C^*-algebras. Michigan Math. J. **24**, 291–302 (1977)
31. Pedersen, G.K.: C^*-Algebras and their automorphism groups. London-New York: Academic Press 1979
32. Harris, L.A.: A Generalization of C^*-Algebras. Proc. Lond. Math. Soc. (3) **42**, 331–361 (1981)

Received July 15, 1981; received in final form March 18, 1982

Math. Z. 180, 535–544 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Geometrically Defined Transfers, Comparisons

Erik Kjær Pedersen

Department of Mathematics, Odense Universitet, Campusvej 55, DK-5230 Odense M, Denmark

§ 0. Introduction

There are a number of geometrically defined transfer homomorphisms relating finiteness obstructions, Whitehead torsions, concordances, surgery obstructions etc. in a bundle. It is the purpose of this paper to relate these various homomorphisms. As an example of the kind of results we obtain consider a bundle $F \to Y \to X$ of finite complexes. If two of these satisfy Poincaré duality then the third does as announced by Quinn in [16]. Since then a proof has appeared by Gottlieb [5]. However in both accounts Poincaré torsion is disregarded. The Poincaré torsion of a finite Poincaré Duality pair $(X, \partial X)$ denoted $\tau(X)$ is defined to be the torsion of the chain homotopy equivalence $[X] \cap - : C^{\#}(X) \to C_{\#}(X, \partial X)$ (see § 3 for sign convention). We prove

Theorem A. *Let $F \to (Y, \partial Y) \to (X, \partial X)$ be a bundle of Poincaré complexes. Let f^*: $Wh(\pi_1 X) \to Wh(\pi_1 Y)$ be the Whitehead transfer homomorphism of Anderson* [1] *(see* § 1.II *for definition) then*

$$\tau(Y) = (-1)^{|F|} \cdot f^*(\tau(X)) + \chi(X) \cdot \tau(F).$$

Here $|F|$ denotes the dimension of F and $\chi(X)$ the Euler characteristic of X.

Remark. The assumption that $(Y, \partial Y)$ is a Poincaré pair is actually redundant. In course of computing $\tau(Y)$ we actually prove $(Y, \partial Y)$ is a Poincaré pair thus generalizing the results of Quinn and Gottlieb to the relative case.

We start by introducing the homomorphisms we want to consider.

§ 1. Geometric Transfer Homomorphisms

I. The L-group Transfer. Let $F \to E \to B$ be a topological bundle with fibre F a manifold and w_B a double cover of B. We obtain a double covering w_E of E as follows: Let $\pi_1 F \to \mathbb{Z}/2\mathbb{Z}$ be the orientation homomorphism and $H(F, *)$ base-

0025-5874/82/0180/0535/$02.00

point preserving homotopy equivalences of F. Then $\pi_1 F \to \mathbb{Z}/2\mathbb{Z}$ factors $\pi_1 F \to \pi_0 H(F,*) \to \mathbb{Z}/2\mathbb{Z}$ and $\pi_1 F \to \pi_0 H(F,*)$ factors $\pi_1 F \to \pi_1 E \to \pi_0 H(F,*)$ so there is a canonical extension of $\pi_1 F \to \mathbb{Z}/2\mathbb{Z}$ to $\pi_1 E \to \mathbb{Z}/2\mathbb{Z}$. We add this homomorphism to the composite $\pi_1 E \to \pi_1 B \xrightarrow{w_B} \mathbb{Z}/2\mathbb{Z}$ to obtain a homomorphism $w_E\colon \pi_1 E \to \mathbb{Z}/2\mathbb{Z}$. This homomorphism defines an isomorphism class of $\mathbb{Z}/2\mathbb{Z}$-bundles over E. We choose one such bundle and denote this by w_E. Consider an element of $L^a(\pi_1 B; w_B)$ where a stands for s, h or p, simple, finite or projective surgery problems (we may also consider surgery problems with certain restrictions on torsions or finiteness obstructions). This element is represented [21, Ch. IX] by a degree 1 normal map with a reference map to B

$$\begin{array}{ccccc} \nu_M & \longrightarrow & \nu_X & & \\ \downarrow & & \downarrow & & \\ M & \xrightarrow{f} & X & \xrightarrow{\varphi} & B \end{array}$$

where f is a degree 1 map with respect to fundamental classes $[X]$ and $[M]$ in homology with respect to an integral local coefficient system given by w_B. Pulling back everything to E defines a transfer homomorphism

$$\tau_L\colon L^a_n(\pi_1 B, w_B) \to L^{a'}_{n+|F|}(\pi_1 E, w_E)$$

where a may or may not be equal to a'. This is the Wall group transfer homomorphism as defined by Quinn and Wall [15, 21] with the modifications that were observed to be necessary by Farrell and Hsiang [6]. It is essentially only computable in case of a product bundle or more generally in case the fundamental groups behave as in a product bundle [12]. The case of S^1-bundles will be described in a forthcoming joint work with Munkholm [11].

II. The Whitehead Torsion Transfer. Consider a bundle $F \to E \to B$ of finite complexes. Anderson [1] defines a transfer homomorphism

$$\tau_1\colon Wh(\pi_1 B) \to Wh(\pi_1 E)$$

such that if

$$\begin{array}{ccc} F & \xrightarrow{f} & F_1 \\ \downarrow{\scriptstyle i} & & \downarrow \\ E & \xrightarrow{g} & E_1 \\ \downarrow & & \downarrow \\ B & \xrightarrow{h} & B_1 \end{array}$$

is a commutative diagram of bundles with f, g and h homotopy equivalences, then the Whitehead torsions are related by

$$t(g) = \tau_1(t(h)) + i_*(t(f)) \cdot \chi(B)$$

where $\chi(B)$ is the Euler characteristic of B. When no confusion is possible we will some times denote τ_1 by p^* where $p\colon E\to B$.

III. The $\tilde{K}_0$-finiteness Obstruction Transfer. Let $F\to E\to B$ be a fibration of finitely dominated spaces. Ehrlich [5] defines a transfer homomorphism

$$\tau_0\colon \tilde{K}_0(\mathbb{Z}\pi_1 B)\to K_0(\mathbb{Z}\pi_1 E)$$

relating the finiteness obstructions of F, E and B according to the formulae

$$\sigma(E)=\tau_0(\sigma(B))+i_*(\sigma(F))\,\chi(B).$$

As in the *Wh*-case we will denote τ_0 by p^* when there is no chance of confusion, $p\colon E\to B$.

IV. A K_{-i} Transfer. Let $F\to E\to B$ be a bundle of finite complexes. It is proved by Munkholm and Pedersen [10] that the diagram

$$\begin{array}{ccc} Wh(\pi_1 E\times\mathbb{Z}) & \longrightarrow & \tilde{K}_0(\mathbb{Z}\pi_1 E) \\ \uparrow{\scriptstyle \tau_1} & & \uparrow{\scriptstyle \tau_0} \\ Wh(\pi_1 B\times\mathbb{Z}) & \longrightarrow & \tilde{K}_0(\mathbb{Z}\pi_1 B) \end{array}$$

is commutative.

Here the horizontal maps are the Bass-Heller-Swan homomorphisms [3] and the vertical maps are the *Wh*- and $\tilde{K}_0$-transfers.

Since by definition $K_{-n}(\mathbb{Z}\pi)$ is the intersection of various inclusions of $\tilde{K}_0(\mathbb{Z}(\pi\times\mathbb{Z}^n))$ in $Wh(\pi\times\mathbb{Z}^{n+1})$ this implies there is a welldefined homomorphism

$$\tau_{-n}\colon K_{-n}(\mathbb{Z}\pi_1 B)\to K_{-n}(\mathbb{Z}\pi_1 E)$$

associated with this bundle such that the diagram

$$\begin{array}{ccc} Wh(\pi_1 E\times\mathbb{Z}^{n+1}) & \longrightarrow & K_{-n}(\mathbb{Z}\pi_1 E) \\ \uparrow & & \uparrow{\scriptstyle \tau_{-n}} \\ Wh(\pi_1 B\times\mathbb{Z}^{n+1}) & \longrightarrow & K_{-n}(\mathbb{Z}\pi_1 B) \end{array}$$

commutes. The left hand vertical map is the *Wh*-transfer associated with $F\to E\times T^{n+1}\to B\times T^{n+1}$. The diagram associated with the finiteness obstruction transfer of the bundle $F\to E\times T^n\to B\times T^n$

$$\begin{array}{ccc} \tilde{K}_0(\mathbb{Z}(\pi_1 E\times\mathbb{Z}^n)) & \longrightarrow & K_{-n}(\mathbb{Z}\pi_1 E) \\ \uparrow & & \uparrow \\ K_0(\mathbb{Z}(\pi_1 B\times\mathbb{Z}^n)) & \longrightarrow & K_{-n}(\mathbb{Z}\pi_1 B) \end{array}$$

will also be commutative. Thus τ_{-n} will inherit all properties known of the $\tilde{K}_0$-transfer e.g. the composite $f_*\cdot\tau_{-n}$ is multiplication by a generalized Euler characteristic, see [14], and τ_{-n} only depends on the homotopy fundamental group data of the bundle [12].

V. The Concordance Transfer. Consider a bundle of closed manifolds $F \to E \to B$. Transfers of concordance spaces have been defined by Burghelea and Lashof, Hatcher and others [4, 8]. These transfers are maps of concordance spaces, which are topological groups, so we get induced homomorphisms on π_0 of these spaces, the isotopy classes of concordances. For a special class of concordances these isotopy classes are completely described by Anderson and Hsiang [2] as follows: Let $h: M \times \mathbb{R}^{n+2} \times I \to M \times \mathbb{R}^{n+2} \times I$ be a concordance, bounded in the $\mathbb{R}^{n+2}$ factor (with usual metric). Then the isotopy class of h is completely determined by an invariant in $K_{-n}(\mathbb{Z}\pi_1 M)$ for $n \geqq 0$ and in $Wh(\pi_1 M)$ for $n = -1$. We have the following definition: Let

$$\begin{array}{ccc} E \times \mathbb{R}^{n+2} \times I & \xrightarrow{k} & E \times \mathbb{R}^{n+2} \times I \\ \downarrow & & \downarrow \\ B \times \mathbb{R}^{n+2} \times I & \xrightarrow{h} & B \times \mathbb{R}^{n+2} \times I \end{array}$$

be a commutative diagram of bounded concordances. There is a homomorphism

$$\tau_c: K_{-n}(\mathbb{Z}\pi_1 B) \to K_{-n}(\mathbb{Z}\pi_1 E)$$

relating the concordance invariants of h and k, i.e. $\tau_c(\sigma(h)) = \sigma(k)$. We need to show that τ_c is everywhere defined and welldefined. That it is everywhere defined follows from e.g. [8] and welldefinedness will be shown in the next section.

§ 2. The Concordance Transfer

Let $F \to E \to B$ be a bundle of manifolds. We wish to show that the concordance transfer is welldefined

Theorem 2.1. *Assume we have a commutative diagram of bounded concordances*

$$\begin{array}{ccc} E \times \mathbb{R}^{n+2} \times I & \xrightarrow{k} & E \times \mathbb{R}^{n+2} \times I \\ \downarrow & & \downarrow \\ B \times \mathbb{R}^{n+2} \times I & \xrightarrow{h} & B \times \mathbb{R}^{n+2} \times I. \end{array}$$

Then $\sigma(k) = \tau_{-n}(\sigma(h))$ *where* $\tau_{-n}: K_{-n}(\mathbb{Z}\pi_1 B) \to K_{-n}(\mathbb{Z}\pi_1 E)$ *is the homomorphism IV.*

This shows that the concordance transfer is welldefined since $\tau_c = \tau_{-n}$.

Proof of Theorem 2.1. The invariant $\sigma(h)$ of Hsiang and Anderson is computed as follows: Wrapping around a torus [17] produces a concordance

$$\bar{h}: B \times T^{n+1} \times \mathbb{R} \times I \to B \times T^{n+1} \times \mathbb{R} \times I$$

such that the cover of $\bar{h}$, $\tilde{h}$ agrees with h on a open subset of $\mathbb{R}^{n+1}$ crossed with $B \times \mathbb{R} \times I$. Choosing c sufficiently big, the space between $B \times T^{n+1} \times O \times I$ and $\bar{h}(B \times T^{n+1} \times c \times I)$ is an h-cobordism with a torsion on $Wh(\pi_1 B \times \mathbb{Z}^{n+1})$. Now Hsiang and Anderson show this invariant actually lies in the summand $K_{-n}(\mathbb{Z}\pi_1 B)$ and completely determines the isotopy class of the concordance. However, wrapping around a torus is functorial, so we may obtain a commutative diagram

$$\begin{array}{ccc} \bar{k}\colon E \times T^{n+1} \times \mathbb{R} \times I & \longrightarrow & E \times T^{n+1} \times \mathbb{R} \times I \\ \downarrow & & \downarrow \\ \bar{h}\colon B \times T^{n+1} \times \mathbb{R} \times I & \longrightarrow & B \times T^{n+1} \times \mathbb{R} \times I \end{array}$$

and it is now evident that the invariants become related by the *Wh*-transfer which restricts to the K_{-n}-transfer.

We obtain the following

Corollary 2.2. *Let $M \to E \to B$ be a bundle with M a closed manifold. Then the Wh-transfer of $M \to E \times T^{n+2} \to B \times T^{n+2}$ preserves the K_{-i} summands of the Wh-groups of the fundamental groups.*

We also obtain the result that all known information on *Wh*- and K_0-transfers hold for concordance transfers as well. One of these results is that the concordance transfer only depends on fundamental froup data. This result has also been proved by Burghelea and Lashof [4] by a geometrical argument. We however further obtain that the various computational results for K_0- and *Wh*-transfers [10, 12, 14] extend to the concordance transfer.

§ 3. Poincaré Torsion

Let $(X; \partial_1 X, \partial_2 X)$ be a finite Poincaré triple in the sense of Wall [20]. The homotopy equivalence

$$[X] \cap -\colon C^{\#}(X, \partial_1 X) \to C_{\#}(X, \partial_2 X)$$

of based chain complexes over $\mathbb{Z}\pi_1 X$ has a well defined torsion $\tau(X, \partial_1 X)$, the Poincaré torsion. We need a sign convention in computing this torsion, and we choose the degrees of the cochain complex i.e. we think of $C_i(X, \partial_2 X)$ as degree $|X| - i$. Clearly $\tau(X, \partial_1 X)$ only depends on the simple homotopy type of $(X; \partial_1 X, \partial_2 X)$. The Poincaré torsion in a bundle is determined by the following

Theorem 3.1. *Let $F \to (Y, \partial Y) \to (X, \partial X)$ be a bundle of finite complexes. Then*

$$\tau(Y) = (-1)^{|F|} \cdot f^*(\tau(X)) + \chi(X) \cdot \tau(F)$$

where f^* is the Wh-transfer of § 2.II.

To prove this theorem we need a number of lemmas and various results about different ways to compute torsion. When we have an exact sequence $0 \to A \to B \to C \to 0$ of based chain complexes we shall allow ourselves to think of

$0 \to B \to C \to sA \to 0$ as a short exact sequence of based chain complexes even though C should really be replaced by the simply homotopy equivalent mapping cylinder of $B \to C$. We use * to denote the duality in the Whitehead group as well as in modules.

Lemma 3.2. *Let f: $A \to B$ be a homotopy equivalence of based chain complexes with torsion τ. The torsion of f^*: $B^* \to A^*$ is then $-\tau^*$.*

Proof. trivial.

Lemma 3.3. *Let $(X, \partial_1 X, \partial_2 X)$ be a Poincaré triple. Then*

$$\tau(X, \partial_1 X) = -(-1)^{|X|} \cdot \tau(X, \partial_2 X)^*$$

Proof. The dual of

$$C^\#(X, \partial_1 X) \to C_\#(X, \partial_2 X)$$

is

$$C^\#(X, \partial_2 X) \to C_\#(X, \partial_1 X).$$

We get the factor $(-1)^{|X|}$ since we change the way we count degrees by $|X|$.

Lemma 3.4. *Let $(X; \partial_1 X, \partial_2 X)$ be a Poincaré triple. Then*

$$\tau(X, \partial_2 X) = \tau(X) - i_* \tau(\partial_2 X).$$

Proof. Noting that $\partial\partial_1 X = \partial_1 X \cap \partial_2 X = \partial\partial_2 X$ we have a diagram of short exact sequences

$$\begin{array}{ccccccccc} 0 & \longleftarrow & C^\#(\partial_2 X, \partial\partial_2 X) & \longleftarrow & C^\#(X, \partial_1 X) & \longleftarrow & C^\#(X, \partial X) & \longleftarrow & 0 \\ & & \downarrow{\scriptstyle [\partial_2 X] \cap -} & & \downarrow{\scriptstyle [X] \cap -} & & \downarrow{\scriptstyle [X] \cap -} & & \\ 0 & \longleftarrow & s(C_\#(\partial_2 X)) & \longleftarrow & C_\#(X, \partial_2 X) & \longleftarrow & C_\#(X) & \longleftarrow & 0 \end{array}$$

giving $\tau(X, \partial_1 X) = \tau(X, \partial X) + i_* \tau(\partial_2 X, \partial\partial_2 X)$. Using Lemma 3.3 and dualizing we get the result.

We note that in case $\partial_2 X = \partial X, \partial_1 X = \varnothing$ we obtain the usual

$$\tau(\partial X) = \tau(X) - \tau(X, \partial X) = \tau(X) + (-1)^{|X|} \tau(X)^*.$$

We also need the following Mayer-Vietories type result.

Lemma 3.5. *Let $(A; \partial_1 A, \partial_2 A)$ and $(B; \partial_1 B, \partial_2 B)$ be Poincaré triples with $\partial_1 A = \partial_1 B = A \cap B$. Then $(A \cup B, \partial_2 A \cup \partial_2 B)$ is a Poincaré pair and $\tau(A \cup B) = \tau(A) + \tau(B) - \tau(A \cap B)$.*

Proof. We have Mayer-Vietories short exact sequences

$$\begin{array}{ccccccccc} 0 & \leftarrow & C^\#(A \cap B, \partial(A \cap B)) & \leftarrow & C^\#(A \cup B, \partial(A \cup B)) & \leftarrow & C^\#(A, \partial A) \oplus C^\#(B, \partial B) & \leftarrow & 0 \\ & & \downarrow & & \downarrow & & \downarrow & & \\ 0 & \leftarrow & sC_\#(A \cap B) & \leftarrow & C_\#(A \cup B) & \leftarrow & C_\#(A) \oplus C_\#(B) & \leftarrow & 0 \end{array}$$

giving

$$\tau(A\cup B,\partial(A\cup B))=\tau(A\cap B,\partial(A\cap B))+\tau(A,\partial A)+\tau(B,\partial B).$$

Now use Lemma 3.3 and dualize.

Lemma 3.6. *Let F be a Poincaré duality space. Then $\tau(F\times D^m)=\tau(F)$ and $\tau(F\times S^m)=(1+(-1)^m)\cdot\tau(F)$.*

Proof. It is easy to see that $\tau(F\times D^m)=\tau(F)$. It now follows from Lemma 3.5 that $\tau(F\times S^m)=\tau(F)+\tau(F)-\tau(F\times S^{m-1})$ which completes the proof using induction.

Lemma 3.7. *Let $(A,\partial A)$ be a Poincaré duality pair which is the union of $(B,\partial B)$ and $D^m\times F$, where $(B,\partial B)$ is a Poincaré duality pair and F is a Poincaré space and*

$$B\cap D^m\times F=\partial B\cap S^{m-1}\times F=S^{i-1}\times D^{m-i}\times F.$$

Then

$$\tau(A)=\tau(B)+(-1)^i\tau(F).$$

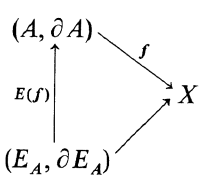
Proof. Follows trivially from Lemma 3.5 and 3.6.

The main application of this lemma is

Proposition 3.8. *Let $F\to(E,\partial E)\to(W,\partial W)$ be a bundle with F a Poincaré space and W a manifold. Then $\tau(E)=\chi(W)\cdot\tau(F)$ (where $\chi(W)$ is the Euler characteristic of W).*

Proof. W has a handlebody decomposition and the bundle restricted to a handle is trivial. Now use Lemma 3.7.

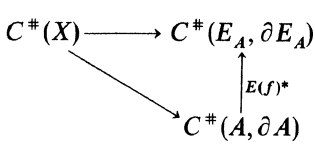
Let $(A,\partial A)\to X$ be a map of CW-complexes with homotopy fibre of homotopy type (D^m,S^{m-1}). We may use Stasheff's construction [19] to construct a pair of CW-complexes $(E_A,\partial E_A)$ and a homotopy commutative diagram



such that $E(f)$ is a homotopy equivalence and such that the inverse image of the interior of all D^n in X under $(E_A,\partial E_A)\to X$ is homeomorphic to $D^n\times(D^m,S^{m-1})$. Even though $(E_A,\partial E_A)\to X$ is not a fibration we do have a Thom isomorphism

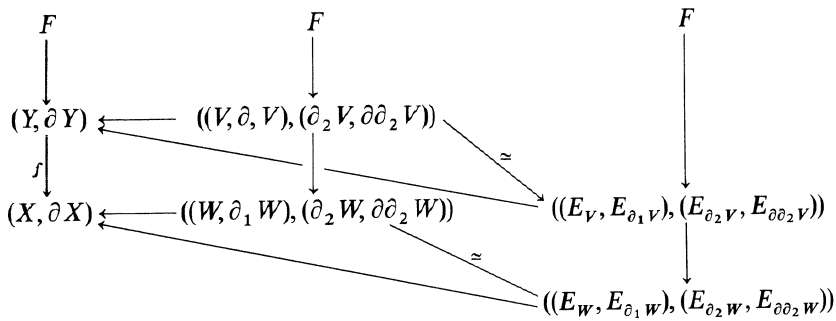
$$C^*(X)\to C^*(E_A,\partial E_A).$$

Assume X,A and ∂A are finite complexes. Then $(E_A,\partial E_A)$ is a finite pair and the Thom isomorphism $C^*(X)\to C^*(E_A,\partial E_A)$ is clearly simple, sending a basis to a basis. The diagram

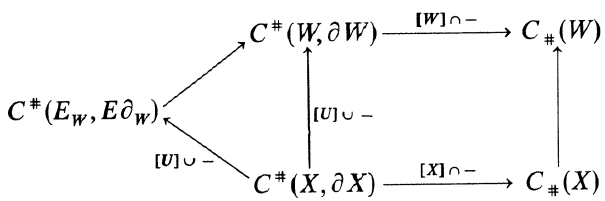


shows that the torsion of the Thom isomorphism $C^{\#}(X)\to C^{\#}(A,\partial A)$ is the torsion of $E(f)^{\#}$. We are now ready for

Proof of Theorem 3.1. Consider the following diagram

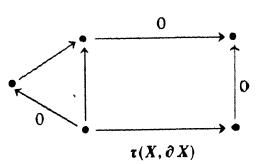


which we proceed to explain. We embed $(X,\partial X)$ in halfspace H^n, some big n such $X\cap\mathbb{R}^{n-1}=\partial X$. Now W is a regular neighborhood of X in H^n meeting $\mathbb{R}^{n-1}$ regularly in $\partial_2 W=W\cap\mathbb{R}^{n-1}$. We define $\partial_1 W=$closure $(\partial W-\partial_2 W)$, so $\partial_1 W\cap\partial_2 W=\partial\partial_1 W=\partial\partial_2 W$. According to Spivak [18] the homotopy fibre of $(W,\partial_1 W)\to X$ and $(\partial_2 W,\partial\partial_2 W)\to\partial X$ is (D^m,S^{m-1}) where $m=n-|X|$. We pull everything back to total space level to get $((V,\partial_1 V),(\partial_2 V,\partial\partial_2 V))$ and note that Stasheff's construction $((E_W,E_{\partial_1 W}),(E_{\partial_2 W},E_{\partial\partial_2 W}))$ pulled back to total space level gives Stasheff's construction on $((V,\partial_1 V),(\partial_2 V,\partial\partial_2 V))$. We have a Thom isomorphism $H^*(X,\partial X)\to H^*(W/\partial W)$ which at chain level (with $\mathbb{Z}\pi_1 X$-coefficients) is given by cup product with some element $[U]$ representing the Thom class. We get the homotopy commutative diagram



where $E_{\partial W}=E_{\partial_1 W}\cup E_{\partial_2 W}$.

The torsions of the various homotopy equivalences are as indicated in the diagram



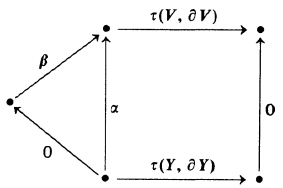
We conclude that the torsion of $C^{\#}(E_W,E_{\partial W})\to C^{\#}(W,\partial W)$ is $(-1)^m\cdot\tau(X,\partial X)$ (the factor $(-1)^m$ because of the degree shift in the Thom isomorphism). It follows from Lemma 3.2 that the torsion of $C_{\#}(W,\partial W)\to C_{\#}(E_W,E_{\partial W})$ is

$$-(-1)^m\tau(X,\partial X)^*=(-1)^m\cdot(-1)^{|X|}\cdot\tau(x).$$

By definition of the *Wh*-transfer homomorphism the torsion of $C_\#(V,\partial V)\to C_\#(E_V,E_{\partial V})$ where $E_{\partial V}=E_{\partial_1 V}\cup E_{\partial_2 V}$ is $f^*((-1)^m(-1)^{|X|}\cdot\tau(X))$. In the homotopy commutative diagram at totalspace level

$$\begin{array}{ccccc} & & C^\#(V,\partial V) & \longrightarrow & C_\#(V) \\ & \nearrow & \uparrow & & \uparrow \\ C^\#(E_V,E_{\partial V}) & & & & \\ & \searrow & C^\#(Y,\partial Y) & \longrightarrow & C_\#(Y) \end{array}$$

the torsions are



where $-\beta^*=f^*((-1)^m\cdot(-1)^{|X|}\cdot\tau(X))$.
We get $\alpha=(-1)^m\cdot\beta$ and

$$\alpha+(-1)^m\tau(V,\partial V)=\tau(Y,\partial Y)$$

using Proposition 3.8 to prove $\tau(V)=\chi(X)\cdot\tau(F)$ and dualizing we get

$$\tau(Y)=(-1)^{|F|}\cdot f^*(\tau(X))+\chi(X)\cdot\tau(F).$$

§ 4. Comparing *L*-Groups and *Wh*-Transfers

We obtain the following result on *L*-group transfers: Let $M\to E\to B$ be a bundle, M a closed manifold of $\dim n\cdot\pi_1(E)=\pi,\pi_1 B=\rho$.

Theorem. *There are commutative diagrams of Rothenberg exact sequences*

$$* \qquad \begin{array}{ccccc} \to L^s_i(\rho) & \longrightarrow & L^h_i(\rho) & \longrightarrow & \hat H^i(\mathbb{Z}_2;Wh\rho)\to \\ \downarrow & & \downarrow & & \downarrow \\ \to L^s_{i+n}(\pi) & \longrightarrow & L^h_{i+n}(\pi) & \longrightarrow & \hat H^{i+n}(\mathbb{Z}_2;Wh\pi)\to \end{array}$$

$$** \qquad \begin{array}{ccccc} \to L^h_i(\rho) & \longrightarrow & L^p_i(\rho) & \longrightarrow & \hat H^i(\mathbb{Z}_2,\tilde K_0(\mathbb{Z}[\pi]))\to \\ \downarrow & & \downarrow & & \downarrow \\ \to L^h_{i+n} & \longrightarrow & L^p_{i+n}(\rho) & \longrightarrow & \hat H^{i+n}(\mathbb{Z}_2,\tilde K_0(\mathbb{Z}[\pi]))\to \end{array}$$

where the vertical maps are (induced by) the Wall-group, K_0- and $(-1)^n$ times the Wh-transfer.

Proof. According to [13] $\hat H^i(\mathbb{Z}_2,\tilde K_0(\mathbb{Z}\pi))$ is represented by surgery problems on finitely dominated Poincaré complexes with finite boundary. It follows then

from the definition of the $\tilde{K}_0$-transfer that $\hat{H}^i(\mathbb{Z}_2, \tilde{K}_0(\mathbb{Z}\rho)) \to \hat{H}^{i+n}(\mathbb{Z}_2, \tilde{K}_0(\mathbb{Z}\pi))$ is induced by the $\tilde{K}_0$-transfer. To see that $\hat{H}^i(\mathbb{Z}_2, Wh\rho) \to \hat{H}^{i+n}(\mathbb{Z}_2, Wh\pi)$ is induced by $(-1)^n$ times the Wh-transfer we now use the similar geometric representation of $\hat{H}^i(\mathbb{Z}_2, Wh\rho)$ as surgery problems on simple Poincaré complexes with finite boundary, and Theorem 3.1.

Final Remarks. Consider an orientable bundle $S^1 \to E \to B$, $\pi_1 B = \rho$, $\pi_1 E = \pi$ both finite. Then the $\tilde{K}_0$-transfer is trivial [10] so we get a transfer homomorphism $L^p(\mathbb{Z}\rho) \to L^h(\mathbb{Z}\pi)$. This and similar situations pose some interesting problems concerning what happens in the Rothenberg exact sequence.

References

1. Anderson, D.R.: The Whitehead torsion of a fiber homotopy equivalence. Michigan Math. J. **21**, 171–180 (1974)
2. Anderson, D.R., Hsiang, W.C.: The functors K_{-i} and pseudoisotopies of polyhedra. Ann. of Math. (2) **105**, 201–233 (1977)
3. Bass, H., Heller, A., Swan, R.G.: The Whitehead group of a polynomial extension. Inst. Hautes Études Sci. Publ. Math. **22**, 61–80 (1964)
4. Burghelea, D., Lashof, R.: Geometric Transfer and homotopy type of the automorphism groups of a manifold. Trans. Amer. Math. Soc. **269**, 1–38 (1982)
5. Ehrlich, K.: Fibrations and a transfer map in algebraic K-theory. J. Pure Appl. Algebra **14**, 131–136 (1979)
6. Farrell, F.T., Hsiang, W.C.: Rational L-groups of Bieberbach groups. Comment. Math. Helv. **52**, 89–109 (1977)
7. Gottlieb, D.H.: Poincaré Duality and Fibrations. Proc. Amer. Math. Soc. **76**, 148–150 (1979)
8. Hatcher, A.E.: Concordance spaces, Higher simple homotopy theory, and Applications. Proceedings (Stanford 1976), pp. 3–21. Proceedings of Symposia in Pure Mathematics **32**, Part 1. Providence, Rhode Island: American Mathematical Society 1978
9. Munkholm, H.J., Pedersen, E.K.: On the Wall finiteness obstruction for the total space of certain fibrations. Trans. Amer. Math. Soc. **261**, 529–545 (1980)
10. Munkholm, H.J., Pedersen, E.K.: Whitehead transfers for S^1-bundles, an Algebraic Description. Comment. Math. Helv. **56**, 404–430 (1981)
11. Munkholm, H.J., Pedersen, E.K.: L-group transfers for S^1-bundles, an Algebraic Description. Preprint
12. Pedersen, E.K.: Universal Geometric Examples for Transfer maps in Algebraic K- and L-theory. J. Pure Appl. Algebra **22**, 179–192 (1981)
13. Pedersen, E.K., Ranicki, A.: Projective Surgery Theory. Topology **19**, 239–254 (1980)
14. Pedersen, E.K., Taylor, L.R.: The Wall finiteness obstruction for a fibration. Amer. J. Math. **100**, 887–896 (1978)
15. Quinn, F.S.: A Geometric Formulation of Surgery. Ph.D. thesis. Princeton University 1969
16. Quinn, F.S.: Surgery on Poincaré and Normal spaces. Bull. Amer. Math. Soc. **78**, 262–267 (1972)
17. Siebenmann, L.C.: A total Whitehead torsion obstruction. Comment. Math. Helv. **45**, 1–48 (1970)
18. Spivak, M.: Spaces satisfying Poincaré duality. Topology **6**, 77–102 (1967)
19. Stasheff, J.: A classification theorem for fibre spaces. Topology **2**, 239–246 (1963)
20. Wall, C.T.C.: Poincaré complexes I. Ann. of Math. (2) **86**, 213–245 (1970)
21. Wall, C.T.C.: Surgery on compact manifolds. London-New York: Academic Press 1970

Received September 3, 1981

Math. Z. 180, 545-551 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On the Spherical Derivative of a Meromorphic Function with a Valiron Deficient Value

Sakari Toppila[1] and Jörg Winkler[2]

[1] Department of Mathematics, University of Helsinki, SF-00100 Helsinki 10, Finland
[2] Fachbereich Mathematik, Technische Universität Berlin, D-1000 Berlin 12, Federal Republic of Germany

1. Introduction and Results

Let f be meromorphic in the plane. We denote

$$\rho(f(z))=\frac{|f'(z)|}{1+|f(z)|^2}$$

and

$$\mu(r,f)=\sup\{\rho(f(z)):\ |z|=r\}.$$

We shall employ the usual notation of Nevanlinna theory. We prove the following

Theorem 1. *Let f be a transcendental meromorphic function in the plane. Then*

$$\limsup_{r\to\infty}\frac{r\,\mu(r,f)\log r}{T(r,f)}\geqq\frac{\Delta(\infty,f)}{2(1-\Delta(\infty,f))} \tag{1.1}$$

if $0\leqq\Delta(\infty,f)<1$, *and*

$$\limsup_{r\to\infty}\frac{r\,\mu(r,f)\log r}{T(r,f)}=\infty \tag{1.2}$$

if $\Delta(\infty,f)=1$.

The following theorem shows that (1.1) is sharp.

Theorem 2. *Given any Δ, $0<\Delta<1$, there exists a transcendental meromorphic function f such that $\Delta(\infty,f)=\Delta$ and that*

$$\limsup_{r\to\infty}\frac{r\,\mu(r,f)\log r}{T(r,f)}=\frac{\Delta}{2(1-\Delta)}. \tag{1.3}$$

The following result of [6] shows that (1.2) is sharp.

0025-5874/82/0180/0545/$01.40

Theorem A. *Given any increasing function $\varphi(r)$ such that $\varphi(r)\to\infty$ as $r\to\infty$, then there exists a transcendental meromorphic function f such that $\Delta(\infty,f)=1$ and that*

$$r\,\mu(r,f)\log r = O(\varphi(r)\,T(r,f)) \quad \text{as} \quad r\to\infty. \tag{1.4}$$

Furthermore, we prove

Theorem 3. *Let f be a transcendental meromorphic function such that $\Delta(\infty,f)>0$. Then*

$$\limsup_{r\to\infty}\frac{r\,\mu(r,f)}{n(r/2,a,f)}>0 \tag{1.5}$$

for any complex value a.

2. A Lemma

We denote

$$L(r,f)=\min\{|f(z)|\colon |z|=r\}.$$

The following lemma is proved in [6].

Lemma A. *Let f be meromorphic in $|z|<R$, $0<R\leqq\infty$. If there exist r_0 and r, $0<r_0\leqq r<R$, such that*

$$L(r_0,f)\leqq 1 \tag{2.1}$$

and

$$m(r,f)\geqq 16, \tag{2.2}$$

then there exists a point a lying in

$$\min\{r_0, re^{-\pi}\}<|z|<r$$

such that

$$|a|\,\rho(f(a))\geqq\frac{A\,m(r,f)}{\pi+\log(r/r_0)} \tag{2.3}$$

where $A>0$ is an absolute constant.

3. Proof of Theorem 1

Let f be a transcendental meromorphic function such that $\Delta=\Delta(\infty,f)>0$.

Case 1. We consider first the case that there exist $d>0$, $k>1$ and a sequence r_n, $r_n\to\infty$ as $n\to\infty$, such that for any n, $L(r_n,f)\leqq 1$, and that

$$m(R_n,f)\geqq d\,T(R_n,f) \tag{3.1}$$

for some R_n satisfying $r_n\leqq R_n\leqq kr_n$. For large values of n, it follows from Lemma A and (3.1) that there exists z_n lying in

$$k^{-1}e^{-\pi}R_n<|z|<R_n$$

such that

$$|z_n|\,\rho(f(z_n)) \geqq \frac{A\,d\,T(R_n,f)}{\pi+\log k}, \tag{3.2}$$

and since $T(R_n,f)\geqq T(|z_n|,f)$, we deduce that

$$\limsup_{r\to\infty}\frac{r\,\mu(r,f)\log r}{T(r,f)}=\infty$$

in this case.

Case 2. We suppose now that the case 1 does not occur. Then there exist sequences R_n and r_n, $R_n\to\infty$ as $n\to\infty$, such that

$$m(R_n,f)=(\Delta+o(1))\,T(R_n,f) \quad \text{as } n\to\infty, \tag{3.3}$$

$L(R_n,f)>1$, $L(r_n,f)=1$, $|f(z)|>1$ for all z satisfying $r_n<|z|<R_n$,

$$\lim_{n\to\infty} R_n/r_n=\infty \tag{3.4}$$

and that

$$m(r_n,f)=o(T(r_n,f)) \quad \text{as } n\to\infty. \tag{3.5}$$

Let k_n be defined by the equality

$$k_n\log k_n=\log(R_n/r_n),$$

and let n be so big that $k_n>100$. We denote $t_n=R_n/k_n$, and let a_s be the zeros of f. Applying the Poisson-Jensen formula for $|z|=t_n$, we get

$$\log|f(z)|\geqq\frac{k_n-1}{k_n+1}\,m(R_n,f)-\sum_{|a_s|<R_n}\log\left|\frac{R_n^2-\bar{a}_s z}{R_n(z-a_s)}\right|.$$

Here, since $n(R_n,0)=n(r_n,0)$,

$$\begin{aligned}\log\left|\frac{R_n^2-\bar{a}_s z}{R_n(z-a_s)}\right| &\leqq \log\frac{2R_n}{t_n}\leqq 2\log k_n\\ &\leqq(2/k_n)\,(k_n\log k_n)=(2/k_n)\log(R_n/r_n)\\ &\leqq(2/k_n)\log(R_n/|a_s|),\end{aligned}$$

and we deduce that

$$\log|f(z)|\geqq\frac{k_n-1}{k_n+1}\,m(R_n,f)-\frac{2+o(1)}{k_n}\,N(R_n,0)$$

on $|z|=t_n$. This implies together with (3.3) and the fact that $k_n\to\infty$ as $n\to\infty$, that

$$\log L(t_n,f)\geqq(\Delta+o(1))\,T(R_n,f) \quad \text{as } n\to\infty. \tag{3.6}$$

We choose z_n lying on $|z|=r_n$ such that $|f(z_n)|=1$. We denote by $\varepsilon(d)$ each function satisfying

$$\varepsilon(d)\to 0 \quad \text{as } d\to 0.$$

We have, for $d>0$,

$$\begin{aligned}\log|f(z_n(1+d))| &= \log|f(z_n)+f'(z_n)\,d\,z_n+d\,\varepsilon(d)| \\ &\leqq d\,|f'(z_n)\,z_n|+d\,\varepsilon(d).\end{aligned} \tag{3.7}$$

On the boundary of the annulus

$$G_n=\{z\colon r_n<|z|<t_n\}$$

we have

$$\log|f(z)|\geqq\frac{(\log L(t_n,f))\log(|z|/r_n)}{\log(t_n/r_n)}, \tag{3.8}$$

and from the superharmonicity of $\log|f(z)|$ we deduce that (3.8) holds for all z lying in G_n. For $d>0$ we have

$$\log\frac{|z_n(1+d)|}{r_n}=d+d\,\varepsilon(d),$$

and we deduce from (3.8) that

$$\log|f(z_n(1+d))|\geqq\frac{d\log L(t_n,f)}{\log(t_n/r_n)}+d\,\varepsilon(d),$$

which together with (3.7) and (3.6) implies that

$$\rho(f(z_n))=\frac{|f'(z_n)|}{2}\geqq\frac{(\Delta+o(1))\,T(R_n,f)}{2r_n\log(R_n/r_n)} \tag{3.9}$$

as $n\to\infty$.

It follows from (3.3) that

$$T(R_n,f)=\frac{1}{1-\Delta+o(1)}N(R_n,\infty)\quad\text{as } n\to\infty,$$

and since $N(t,\infty)$ is an increasing and convex function of $\log t$, we deduce from (3.9) that

$$\begin{aligned}\mu(r_n,f)&\geqq\frac{\Delta N(R_n,\infty)}{2(1-\Delta+o(1))\,r_n\log R_n}\\ &\geqq\frac{\Delta N(r_n,\infty)}{2(1-\Delta+o(1))\,r_n\log r_n}.\end{aligned}$$

This together with (3.5) gives

$$\frac{r_n\mu(r_n,f)\log r_n}{T(r_n,f)}\geqq\frac{\Delta}{2(1-\Delta+o(1))}\quad\text{as } n\to\infty$$

which completes the proof of Theorem 1.

4. Proof of Theorem 3

Let f be as in the proof of Theorem 1, and let a be any complex value. In the case 1 we get (1.5) from (3.2) and the fact that

$$\begin{aligned} T(R_n, f) &\geqq (1+o(1))\, N(R_n, a) \\ &\geqq (1+o(1))\, n(R_n/2, a) \log 2 \\ &\geqq (1+o(1))\, n(|z_n|/2, a) \log 2 \quad \text{as } n \to \infty. \end{aligned}$$

In the case 2 we get (1.5) from (3.9) and the fact that in this case

$$T(R_n, f) \geqq (1+o(1))\, n(r_n, a) \log(R_n/r_n) \quad \text{as } n \to \infty.$$

Theorem 3 is proved.

5. Proof of Theorem 2

Let Δ, $0<\Delta<1$, be given. We set $k=1/(1-\Delta)$, and for any n, an integer s_n is chosen such that

$$k^n \leqq s_n < k^n + 1. \tag{5.1}$$

We choose a sequence r_n such that $r_0 = 1$ and for $n \geqq 1$,

$$r_n = \exp\exp\exp r_{n-1}. \tag{5.2}$$

We set

$$f(z) = \sum_{n=1}^{\infty} (1 - z/r_n^{(-1)^n s_n}) \tag{5.3}$$

Let n be odd and $r_n \leqq r \leqq r_{n+1}$. It follows from (5.2) and (5.1) that

$$\begin{aligned} N(r, 0) &= (1+o(1))\, n(r_n, 0) \log r \\ &= \frac{(1+o(1))\, k^{n-1} \log r}{1-k^{-2}}. \end{aligned} \tag{5.4}$$

Similarly, we get

$$\begin{aligned} N(r_n, \infty) &= \frac{(1+o(1))\, k^{n-2} \log r_n}{1-k^{-2}} \\ &= (1/k + o(1))\, N(r_n, 0). \end{aligned} \tag{5.5}$$

These estimates imply that

$$\begin{aligned} \frac{m(r_n, f)}{T(r_n, f)} &= 1 - \frac{N(r_n, \infty)}{T(r_n, f)} \geqq 1 - (1+o(1)) \frac{N(r_n, \infty)}{N(r_n, 0)} \\ &\geqq 1 - 1/k + o(1) = \Delta + o(1), \end{aligned}$$

and we deduce that $\Delta(\infty, f) \geqq \Delta$. If n is even then (5.4) holds for $1/f$, and we get for all values of n,

$$T(r,f) \geqq (1+o(1)) \max\{N(r,0), N(r,\infty)\} \geqq \frac{(1+o(1))\, k^{n-1} \log r}{1-k^{-2}} \tag{5.6}$$

for $r_n \leqq r \leqq r_{n+1}$.

Let n be large and

$$k^{2n} r_n \leqq |z| \leqq 2\sqrt{r_{n+1}}.$$

Since

$$\left|\frac{f'(z)}{f(z)}\right| = \left|\sum_{p=1}^{\infty} \frac{(-1)^p s_p}{z-r_p}\right|$$

and for $p \leqq n$

$$\left|\frac{s_p}{z-r_p} - \frac{k^p}{z}\right| = \left|\frac{(s_p - k^p)\, z + k^p r_p}{z(z-r_p)}\right| \leqq \frac{2}{|z|},$$

we get

$$\left|\frac{f'(z)}{f(z)}\right| \leqq |z|^{-1} (k^n - k^{n-1} + k^{n-2} - \ldots) + O(n/|z|) \leqq \frac{(1+o(1))\, k^n}{|z|\,(1+k^{-1})}. \tag{5.7}$$

Since

$$\rho(f(z)) \leqq \frac{|f'(z)|}{2\,|f(z)|},$$

we deduce from (5.6) and (5.7) that

$$\frac{r\,\mu(r,f) \log r}{T(r,f)} \leqq \frac{(1+o(1))\, k^n (1-k^{-1})(1+k^{-1})}{2(1+k^{-1})\, k^{n-1}} \leqq \frac{k-1+o(1)}{2} = \frac{\Delta + o(1)}{2(1-\Delta)} \tag{5.8}$$

for $k^{2n} r_n \leqq r \leqq 2\sqrt{r_{n+1}}$.

Let n be even and

$$\sqrt{r_n} \leqq |z| \leqq 2k^{2n} r_n.$$

We get from (5.2) and (5.1)

$$\begin{aligned} \log|f(z)| &\leqq n(r_{n-1}, 0) \log|z| - (1+o(1))\, n(r_{n-1}, \infty) \log|z| + s_n \log|1 - z/r_n| \\ &\leqq \left(\frac{k^{n-2}}{1-k^{-2}} - \frac{k^{n-1}(1+o(1))}{1-k^{-2}}\right) \log|z| \\ &\leqq -4 \log|z| \end{aligned}$$

if n is large enough. This implies that if

$$2\sqrt{r_n} \leqq |z_0| \leqq k^{2^n} r_n,$$

then

$$\rho(f(z_0)) \leqq |f'(z_0)| = \left| (2\pi)^{-1} \int_{|z-z_0|=1} \frac{f(z)\,dz}{(z-z_0)^2} \right| \leqq |z_0|^{-2}.$$

Considering $1/f$ instead of f if n is odd, we deduce that

$$\mu(r,f) \leqq r^{-2} \tag{5.9}$$

if

$$2\sqrt{r_n} \leqq r \leqq k^{2^n} r_n$$

and n is sufficiently large.

Combining (5.8) and (5.9), we deduce that

$$\frac{r\,\mu(r,f)\log r}{T(r,f)} \leqq \frac{\Delta + o(1)}{2(1-\Delta)}$$

for all large values of r. This implies, together with Theorem 1 and the fact that $\Delta(\infty,f) \geqq \Delta$, that

$$\frac{\Delta}{2(1-\Delta)} \leqq \frac{\Delta(\infty,f)}{2(1-\Delta(\infty,f))} \leqq \limsup_{r\to\infty} \frac{r\,\mu(r,f)\log r}{T(r,f)} \leqq \frac{\Delta}{2(1-\Delta)},$$

and we conclude that $\Delta(\infty,f) = \Delta$ and that (1.3) holds. Theorem 2 is proved.

References

1. Anderson, J.M., Clunie, J.: Slowly growing meromorphic functions. Comment. Math. Helv. **40**, 267–280 (1966)
2. Clunie, J., Hayman, W.K.: The spherical derivative of integral and meromorphic functions. Comment. Math. Helv. **40**, 117–148 (1966)
3. Lehto, O.: The spherical derivative of meromorphic functions in the neighbourhood of an isolated singularity. Comment. Math. Helv. **33**, 196–205 (1959)
4. Lehto, O., Virtanen, K.I.: On the behaviour of meromorphic functions in the neighbourhood of an isolated singularity. Ann. Acad. Sci. Fenn., Ser. A I **240**, 1–9 (1957)
5. Toppila, S.: On the growth of the spherical derivative of a meromorphic function. Ann. Acad. Sci. Fenn. (to appear)
6. Toppila, S.: On the spherical derivative of a meromorphic function with a deficient value. In: Proceedings of the V-th Romanian-Finnish Seminar on Complex Analysis (to appear)
7. Winkler, J.: Zur Existenz ganzer Funktionen bei vorgegebener Menge der Nullstellen und Einsstellen. Math. Z. **168**, 77–85 (1979)

Received September 9, 1981

Math. Z. 180, 553–565 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

On the Prime Ideal of an Ordering of Higher Level

Thomas C. Craven*

Department of Mathematics, University of Hawaii, Honolulu, Hawaii 96822, USA

1. Introduction

In this paper we continue our investigation of the properties of orderings of higher level in semilocal rings that was begun in [6]. Roughly speaking, an ordering of level n has the same relationship with 2^n-th powers that ordinary orderings have with squares. The next section is primarily devoted to a computation of the prime ideal associated with an ordering; this is a unique prime ideal $\mathfrak{p}$ of the ring A such that the ordering extends uniquely to the field of fractions $A(\mathfrak{p})$ of $A/\mathfrak{p}$. We show that for an integral domain, the ideal $\mathfrak{p}$ can be computed in terms of a certain valuation subring of the field of fractions. This provides a very powerful technique for reducing problems involving higher level orderings of rings to known results for fields. As an example, it is shown that if every ordering of A has level one, then A satisfies *SAP* (cf. [15]). Section 3 contains applications to Witt rings of higher level which were defined in [6]. Many of the results of [7] are extended from level one to higher level. The main theorems state that if A is a noetherian regular local ring with 2 a unit and fraction field F, then the kernel of the induced mapping between the Witt rings of level n is nil and every ordering of A extends to an ordering of F of the same level.

We shall now recall some of the basic notation and definitions from [6] which will be needed for this paper. Let A be a connected semilocal ring with no residue field having two elements; that is, A is a commutative ring with 1 which has no nontrivial idempotents and only finitely many maximal ideals $\mathfrak{m}_1, \ldots, \mathfrak{m}_g$, and also $|A/\mathfrak{m}_i| \neq 2$ for $i = 1, \ldots, g$. We denote the group of units of A by A^*. In [6] we defined higher level Witt rings $W_n(A)$, $n = 1, 2, \ldots$ so that $W_1(A)$ is the usual Witt ring of nondegenerate symmetric bilinear forms. Let $\mathfrak{C}_n$ denote the group of 2^n-th roots of unity in the complex numbers and let $\mathfrak{A}_n$ denote the ring of integers of the cyclotomic field $\mathbb{Q}(\mathfrak{C}_n)$. A group homomorphism $\sigma\colon A^* \to \mathfrak{C}_n$ which extends to a ring homomorphism $\sigma\colon W_n(A) \to \mathfrak{A}_n$ is

* Partially supported by National Science Foundation Grant MCS 79-00318

0025-5874/82/0180/0553/$02.60

called a *signature of level n*. We shall call the subgroup $P(\sigma)=\mathrm{Ker}(\sigma: A^*\to\mathfrak{C}_n)$ an ordering of level n. We then have $A^*/P(\sigma)\cong\mathfrak{C}_m$ for some $m\leqq n$ and we say that σ (or $P(\sigma)$) has *exact level m*. In general, we shall speak of *signatures* and *orderings of higher level* without specifying n. Orderings of level n are in one-to-one correspondence with the minimal, nonmaximal prime ideals of $W_n(A)$ via $P(\sigma)$ corresponds to $\mathrm{Ker}(\sigma: W_n(A)\to\mathfrak{A}_n)$. The homomorphism σ, however, is only determined up to an automorphism of $\mathfrak{C}_n$ [6, Proposition 1.5]. Given a ring homomorphism $\psi: A\to B$, one can speak of the extension of a signature σ of A to a signature σ' of level n of B. This means that if $\psi_n: W_n(A)\to W_n(B)$ is the induced Witt ring homomorphism, then $\sigma'\psi_n=\sigma$, or equivalently, $\psi^{-1}(P(\sigma'))\cap A^*=P(\sigma)$. Using these homomorphisms, we can always restrict an ordering or signature on B to one on A.

Since an ordering $P(\sigma)$ only contains units of A, we often need a larger set containing nonunits as well. Corresponding to an ordering $P(\sigma)$ of exact level n, we define

$$Q(\sigma)=\{\textstyle\sum\lambda_i^{2^n}a_i \mid \lambda_i\in A, \sum\lambda_i A=A, a_i\in P(\sigma)\}.$$

If $\omega\in A^*$ is chosen so that its image generates $A^*/P(\sigma)$, then we can form the set $S=\bigcup_{i=1}^{2^n}\omega^i Q(\sigma)$. In all known cases, and in particular for all local rings [6, Theorem 2.13], the complement of S in A is a prime ideal which we denote by $\mathfrak{p}(P)$, or $\mathfrak{p}(\sigma)$ if σ is a signature corresponding to P, and refer to it as the prime ideal associated with P or σ.

For the purposes of this paper, all rings will be assumed to satisfy the following conditions even though they will not again be explicitly stated:

(a) They are connected, semilocal and have no residue fields with only two elements.

(b) All signatures have associated prime ideals (i.e. the complement of S above is a prime ideal).

(c) Any ordering P is coinitial in the corresponding set Q in the sense that if $x\in Q$, then there exist $y\in P$ and $z\in Q\cup\{0\}$ such that $x=y+z$.

We wish to point out that some of these conditions can be relaxed or eliminated for certain results, but they are all required for the proofs of some of the most important theorems. Conditions (b) and (c) always hold for local rings [6, Theorem 2.13, Lemma 2.11] and for semilocal rings when the level is one [10, Appendix B]. No examples are known in which they do not hold.

We conclude this section with a result which shows the connection between orderings of level n and 2^n-th powers.

Proposition 1.1. *Let Q be a subset of A such that Q is maximal with respect to the following properties:* (a) *Q is closed under addition and multiplication;* (b) *Q contains all 2^n-th powers of units;* (c) *Q contains all elements of the form $\sum\lambda_i^{2^n}a_i$ with $a_i\in Q\cap A^*$ and $\sum\lambda_i A=A$; and* (d) *$-1\notin Q$. Then there exists a signature σ of A such that $Q=Q(\sigma)\cup\mathfrak{p}(\sigma)$.*

Proof. Let $P=Q\cap A^*$, a subgroup of A^* since Q is closed under multiplication and inverses arise from $a^{-1}=(a^{-1})^{2^n}a^{2^n-1}\in Q\cap A^*$ whenever $a\in Q\cap A^*$. The

proof now continues exactly as in the proof of parts (2) through (5) of [6, Theorem 3.4]. One concludes that there exists a homomorphism σ: $A^* \to \mathfrak{C}_n$ with kernel equal to P such that σ extends to a ring homomorphism σ: $W_n(A) \to \mathfrak{U}_n$. Thus σ is a signature and $P(\sigma)=P$. The definition of $Q(\sigma)$ shows that it is contained in Q. If $k \leqq n$ is the exact level of σ (i.e. k is minimal such that $A^{*2^k} \subset P$), then we know that A is partitioned into the sets $Q(\sigma)$, $\omega Q(\sigma)$, $\omega^2 Q(\sigma), \ldots, \omega^{2^k-1} Q(\sigma)$, $\mathfrak{p}(\sigma)$, where ωP generates A^*/P. Since $-1 \notin Q$ and $Q(\sigma) \subset Q$, the set Q cannot intersect the sets $\omega^i Q(\sigma)$, $i=1, \ldots, 2^k-1$. Thus $Q \subset Q(\sigma) \cup \mathfrak{p}(\sigma)$. Since Q is maximal and $Q(\sigma) \cup \mathfrak{p}(\sigma)$ satisfies conditions (a) through (d) [6, Theorem 2.14], we see that $Q = Q(\sigma) \cup \mathfrak{p}(\sigma)$.

2. The Associated Prime Ideal

In this section we shall develop methods for computing the prime ideal associated with an ordering of higher level. Such computations have already been done by Knebusch for the level one case [9] and our results are similar, though the proofs are necessarily more complicated. Recall that all rings are assumed to satisfy the conditions stated in §1.

Proposition 2.1. *Let σ be a signature of A with associated prime ideal $\mathfrak{p}$ and exact level m. Let φ: $A \to B$ be a ring homomorphism. If σ extends to a signature on B of exact level m, then* $\operatorname{Ker}\varphi \subset \mathfrak{p}$.

Proof. By [6, Theorem 3.4], the signature σ extends to a signature of level m on B if and only if

$$0 \notin S_m(A,B) = \{\textstyle\sum \lambda_i^{2^m} \varphi(a_i) \mid \lambda_i \in B, \sum \lambda_i B = B, a_i \in P(\sigma)\}.$$

Assume we can find $x \in \operatorname{Ker}\varphi$ such that $x \notin \mathfrak{p}$. Replacing x by x^{2^m} if necessary, we may also assume that $x \in Q(\sigma)$; that is, there exist elements $\lambda_i \in A$, $p_i \in P(\sigma)$ such that $\sum \lambda_i A = A$ and $x = \sum \lambda_i^{2^m} p_i$. But then $\sum \varphi(\lambda_i) B = B$, so we obtain $0 = \varphi(x) = \sum \varphi(\lambda_i)^{2^m} \varphi(p_i)$ in $S_m(A,B)$, a contradiction.

Definition. Let φ: $A \to B$ be a ring homomorphism and let τ be a signature of B. We say that A is *cofinal* in B with respect to τ (or with respect to $P(\tau)$) if for every $b \in Q(\tau)$, there exist $c \in Q(\tau)$ and $a \in P(\tau|A)$ such that $\varphi(a) = b + c$.

Remark 2.2. (a) In the previous definition, it suffices to specify only that $a \in A$. Then a can be modified as follows: Let $\mathfrak{m}_1, \ldots, \mathfrak{m}_g$ be the maximal ideals of A. By the Chinese Remainder Theorem, there exists an element $\lambda \in A$ such that $\lambda \equiv 0$ modulo $\mathfrak{m}_i$ for each $\mathfrak{m}_i$ not containing a and $\lambda \equiv 1$ modulo $\mathfrak{m}_i$ for each $\mathfrak{m}_i$ containing a. Then $\lambda^{2^n} + a \in A^*$ and

$$\varphi(\lambda^{2^n} + a) = b + (\varphi(\lambda)^{2^n} + c) \quad \text{with} \quad \varphi(\lambda)^{2^n} + c \in Q(\tau).$$

Thus we may assume that a is a unit. Since $\varphi(a) \in Q(\tau) \cap B^* = P(\tau)$, we have $a \in P(\tau|A)$.

(b) If A and B are fields and τ is an ordering of level one, this agrees with the well-known meaning of cofinal. For rings and level one orderings, this definition agrees with that of Knebusch [9].

Example 2.3. The unique (cf. [6, Example 4.1]) signature of the power series ring $A=\mathbb{R}[[x]]$ is not cofinal in either of the ordinary orderings of $\mathbb{R}((x))$. (Take $b=x^{-2}$ in the definition.) On the other hand, A is cofinal in $\mathbb{R}$ under the natural quotient map. (Take $a=b+1$, $c=1$ in the definition.) The next lemma shows that this behavior is determined by the fact that the associated prime ideal of the signature is the maximal ideal of A.

Lemma 2.4 (Compare [9, Lemma 5.2]). *Let A be a domain with field of fractions F. Let τ be a signature of F and let $\mathfrak{p}$ be the prime ideal of A associated with $\sigma=\tau|A$. Let S be a multiplicative subset of A. Then the following two statements are equivalent:*

(i) $S\cap\mathfrak{p}=\varnothing$.

(ii) *A is cofinal in $S^{-1}A$ with respect to the restriction of τ.*

Proof. Assume that (ii) holds. If $S\cap\mathfrak{p}\neq\varnothing$, then, replacing s by s^{2^n} if necessary, we may assume there exists an element $s\in S\cap\mathfrak{p}\cap P(\tau)$. For all elements $a\in A$, we have $s^{-1}-a=s^{-1}(1-as)\in P(\tau)$ since $s^{-1}\in P(\tau)$ and $1-as\in 1+\mathfrak{p}\subset Q(\sigma)$ [6, Theorem 2.14]. But then A cannot be cofinal in $S^{-1}A$: in the definition, take $b=s^{-1}$, so $-c=b-a\in P(\tau)$, hence c cannot lie in $Q(\tau)$.

Conversely, assume $\mathfrak{p}\cap S=\varnothing$. Let $s\in S$, $b\in A$ with $s^{-1}b\in P(\tau)=Q(\tau)$. Let $\omega\in A^*$ be an element such that $\omega P(\sigma)$ generates $A^*/P(\sigma)$. Without loss of generality we may assume $\omega\in S$. Since $S\cap\mathfrak{p}=\varnothing$ we have $s\in\bigcup\omega^iQ(\sigma)$, say $s\in\omega^jQ(\sigma)$. Replacing s by $\omega^{-j}s$ and b by ω^jb, we may assume $s\in Q(\sigma)$ and $b\in P(\tau)$ since $s^{-1}b\in P(\tau)$. In fact, $b\in Q(\sigma)\cup\mathfrak{p}$ since $b\notin\omega^iQ(\sigma)$ for $i\not\equiv 0\,(\mathrm{mod}\,2^n)$, where n is the exact level of σ. Now write $s=\sum\lambda_i^{2^n}p_i$ with $p_i\in P(\sigma)$. By our assumption that $Q(\sigma)$ is coinitial in $P(\sigma)$, we may assume $\lambda_1=1$. Let $x=p_1^{-1}\in A$. Then

$$x-s^{-1}=\frac{\sum\lambda_i^{2^n}p_i-p_1}{p_i\sum\lambda_i^{2^n}p_i}$$

lies in $P(\tau)$ since both the numerator and denominator do. Multiplying by b gives $bx-bs^{-1}\in P(\tau)$, hence $a=bx\in A$ shows that A is cofinal in $S^{-1}A$ by Remark 2.2(a).

Proposition 2.5. *Let A be a domain with field of fractions F. Let τ be a signature of F. Then $\mathfrak{p}(\tau|A)=0$ if and only if A is cofinal in F with respect to τ. If these hold, then τ is the only extension of $\tau|A$ to F.*

Proof. The first statement follows immediately from the previous lemma. The uniqueness of τ follows from the fact that $A(\mathfrak{p})=F$ when $\mathfrak{p}=\mathfrak{p}(\tau|A)=0$, so the signature on A extends uniquely [6, Theorem 2.14].

Lemma 2.6. *Let σ be a signature of A and let $\mathfrak{a}$ be an ideal of A contained in $\mathfrak{p}(\sigma)$. Let τ be the unique extension of σ to $A/\mathfrak{a}$ induced by restriction from $A(\mathfrak{p}(\sigma))$ (i.e. use the canonical homomorphisms*

$$W_n(A)\to W_n(A/\mathfrak{a})\to W_n(A/\mathfrak{p}(\sigma))\to W_n(A(\mathfrak{p}(\sigma))).$$

Then $\mathfrak{p}(\tau)=\mathfrak{p}(\sigma)/\mathfrak{a}$.

Proof. Trivial.

Theorem 2.7 (Compare [9, Theorem 5.6]) *Let $\varphi: A \to B$ be a ring homomorphism. Let τ be a signature of B and let $\sigma = \tau | A$. Set $\mathfrak{q} = \mathfrak{p}(\tau)$, $\mathfrak{p} = \varphi^{-1}(\mathfrak{q})$ and let $\bar{\tau}$ be the extension of τ to $B(\mathfrak{q})$. Then the following statements are equivalent:*

(i) $\mathfrak{p} = \mathfrak{p}(\sigma)$.

(ii) *$A/\mathfrak{p}$ is cofinal in $B(\mathfrak{q})$ with respect to $\bar{\tau}$.*

(iii) *$A/\mathfrak{p}$ is cofinal in $B/\mathfrak{q}$ with respect to $\bar{\tau}$.*

(iv) *A is cofinal in B with respect to τ.*

(v) *For each $b \in Q(\tau)$, there exist elements $\lambda_i \in B$, $c_i \in P(\tau)$, $\mu \in \mathfrak{p}(\tau)$ and $a \in Q(\sigma)$ such that $b + \sum \lambda_i^{2^n} c_i + \mu = \varphi(a)$, where n is the exact level of τ.*

Proof. Note first that we always have $\mathfrak{p} \subset \mathfrak{p}(\sigma)$ by Proposition 2.1 applied to $A/\mathfrak{p}$. Let $\bar{\sigma} = \bar{\tau} | (A/\mathfrak{p})$. The previous lemma says that $\mathfrak{p}(\bar{\sigma}) = \mathfrak{p}(\sigma)/\mathfrak{p}$. In view of this, the equivalence of (i), (ii) and (iii) follows from Proposition 2.5. The implication (iv) $\Rightarrow$ (v) is immediate from the definition.

Next we shall show that (v) implies (iii). Let $b \in B$ such that $\bar{b} \in Q(\bar{\tau} | B/\mathfrak{q})$. Then $b \in Q(\tau)$, so (v) implies that there exist $c \in Q(\tau)$ and $a \in P(\sigma)$ such that $\varphi(a) = b + c$. Thus $\bar{\varphi}(\bar{a}) = \bar{b} + \bar{c}$ with $\bar{c} \in Q(\bar{\tau} | B/\mathfrak{q})$ and $\bar{a} \in P(\bar{\sigma})$. Therefore $A/\mathfrak{p}$ is cofinal in $B/\mathfrak{q}$.

Finally, we show that (iii) implies (iv). For $b \in Q(\tau)$, the fact that $A/\mathfrak{p}$ is cofinal in $B/\mathfrak{q}$ implies that there exist elements $c \in Q(\tau)$, $a \in A$ and $\mu \in \mathfrak{p}(\tau)$ such that $\varphi(a) = b + c + \mu$. But $c + \mu \in Q(\tau)$ by [6, Theorem 2.14], and so A is cofinal in B by Remark 2.2(a). This concludes the proof of the theorem.

Our most useful result in this section is an application of the preceding theorem which will relate the associated prime ideal to valuations. Using the next theorem, one can obtain many results by reducing problems concerning orderings for rings to the corresponding situation in fields, where the answers are known.

Let F be a field and let τ be a signature of higher level on F. For any subfield k of F, we denote by $\mathfrak{D}(\tau, k)$ the convex hull of k in F with respect to τ: that is, $\mathfrak{D}(\tau, k) = \{a \in F \mid r + a \in P(\tau)$ and $r - a \in P(\tau)$ for some $r \in k \cap P(\tau)\}$. This is a valuation ring of F with maximal ideal $\mathfrak{m}(\tau, k) = \{a \in F \mid r + a \in P(\tau)$ and $r - a \in P(\tau)$ for all $r \in k \cap P(\tau)\}$ [1, Chap. I, Lemma 5]. We shall generally denote $\mathfrak{D}(\tau, \mathbb{Q})$ by $\mathfrak{D}_0$ and $\mathfrak{m}(\tau, \mathbb{Q})$ by $\mathfrak{m}_0$.

Let $\mathfrak{D}$ be a valuation subring of F. We say that $\mathfrak{D}$ (or its associated valuation v) is *compatible* with a signature τ (or with the ordering $P(\tau)$) if the following equivalent conditions hold [1, Chap. I, §2]:

(i) $1 + \mathfrak{m} \subset P(\tau)$, where $\mathfrak{m}$ is the maximal ideal of $\mathfrak{D}$.

(ii) $\mathfrak{D} = \mathfrak{D}(\tau, k)$ for some subfield k of F.

(iii) $\{a + \mathfrak{m} \mid a \in P(\tau) \cap \mathfrak{D}\}$ is an ordering of the residue field $\mathfrak{D}/\mathfrak{m}$.

(iv) $\mathfrak{D}$ is convex in F with respect to τ.

If $\bar{\tau}$ is the signature on the residue field $\mathfrak{D}/\mathfrak{m}$ corresponding to the ordering in (iii), then its restriction to $\mathfrak{D}$ is $\tau | \mathfrak{D}$ in the sense of §1 and homomorphisms of Witt rings. All valuation rings compatible with a given signature τ contain $\mathfrak{D}_0 = \mathfrak{D}(\tau, \mathbb{Q})$, so these rings are totally ordered by inclusion. Moreover, any subring of F containing $\mathfrak{D}_0$ is a valuation ring compatible with τ.

Theorem 2.8. *Let A be a domain with field of fractions F. Let τ be a signature of F and $\sigma=\tau|A$. Set $\mathfrak{D}=A\mathfrak{D}_0$ and let $\mathfrak{m}$ be its maximal ideal. Then $\mathfrak{p}(\sigma)=\mathfrak{m}\cap A$. The signature $\bar{\tau}$ of $\mathfrak{D}/\mathfrak{m}$ extends the signature $\bar{\sigma}$ of $A(\mathfrak{p}(\sigma))$.*

Proof. By the corollary to Theorem 10 in [1, Chap. I], the ring $\mathfrak{D}$ is the convex hull of any maximal subfield of $\mathfrak{D}$. On the other hand, the induced ordering of $\mathfrak{D}_0/\mathfrak{m}_0$ is archimedean (of level one), so $\mathfrak{D}_0$ is the convex hull of $\mathbb{Z}$. It follows that $\mathfrak{D}$ is the convex hull of A. Thus Theorem 2.7 implies that $\mathfrak{p}(\sigma)=\mathfrak{m}\cap A$. Moreover, σ has a unique extension to $\bar{\sigma}$ on $A(\mathfrak{p}(\sigma))$ and this equals the restriction of $\bar{\tau}$ defined on $\mathfrak{D}/\mathfrak{m}$.

This theorem can be applied to an arbitrary semilocal ring A by using the sequence of homomorphisms $A\to A/\mathfrak{p}\to A(\mathfrak{p})$. Our proof of the next lemma illustrates this, but first we need some additional background information.

Let A be a semilocal ring satisfying our usual conditions, and let X be the set of all signatures of level one of A. We shall adopt the terminology of [9] for fans and associated concepts; for further information on fans and their importance, see [1] and [4] for fields and [9] for rings. A subgroup T of A^* is called *saturated* if there is a subset Y of X such that $T=\{a\in A^*|\sigma(a)=1 \text{ for all } \sigma\in Y\}$. Given any subset S of A^*, there is a minimal saturated subgroup containing it and it consists of all units of the form $\sum t_i a_i$ where each t_i is a sum of squares and each a_i is a product of elements of S [9, Theorem 3.2]. A saturated subgroup T is a *fan domain* if and only if for every $a\in A^*$ with $a\notin -T$ we have $(T+aT)\cap A^*=T\cup aT$. A fan V of A is a subset of X consisting of all (level one) signatures σ such that $P(\sigma)\supset T$ for some fixed fan domain T. Since signatures are equivalent to orderings in level one, we shall also think of a fan as a set of orderings. Knebusch has shown that the fans of A are all induced from the residue class fields $A(\mathfrak{p})$ [9, Theorem 7.4]. More precisely, if a fan V has at least four elements, then all of the signatures in V have the same associated prime ideal which we shall denote by $\mathfrak{p}(V)$, and the set of induced signatures on $A(\mathfrak{p}(V))$ forms a fan of that field.

Lemma 2.9. *Let V be a fan of A with associated prime ideal $\mathfrak{p}$. Let $\overline{V}$ be the induced fan of $A(\mathfrak{p})$ and let v be the finest valuation of $A(\mathfrak{p})$ compatible with all orderings of $\overline{V}$. If $A(\mathfrak{p})$ has an ordering P' of exact level n compatible with v, then the restriction P of P' to A is an ordering with associated prime ideal $\mathfrak{p}$ and exact level n.*

Proof. Assume first that A is a domain and $\mathfrak{p}=0$. It follows from Bröcker's characterization of fans in fields [4, Theorem 2.7] that any ordering compatible with v induces an archimedean ordering on the residue field of v. Thus we may denote the valuation ring of v by $\mathfrak{D}_0=\mathfrak{D}(\tau,\mathbb{Q})$ where τ is any signature compatible with v. By Theorem 2.8, all such signatures have the same associated prime ideal. Since the prime is zero for those in the fan V, it is also zero for P. If A is not a domain, we can apply the above argument to $A/\mathfrak{p}$ and use Lemma 2.6 to conclude that P has associated prime ideal $\mathfrak{p}$ in A. Finally, P has exact level n by [6, Theorem 2.14] since its extension to $A(\mathfrak{p})$ has exact level n.

It is well known that the set X of level one signatures of A has a natural topology on it defined by taking sets of the form $\{\sigma\in X|a\in P(\sigma)\}$, $a\in A^*$, as a

subbasis for the open sets. This is equivalent to the Zariski topology on the set of prime ideals $\{\mathrm{Ker}\sigma | \sigma \in X\}$. One says that A satisfies the *strong approximation property* (SAP) if for any clopen subset Y of X, there exists an element $a \in A^*$ such that $Y = \{\sigma \in X | a \in P(\sigma)\}$. This is a very strong condition and many equivalent properties are known. Some of these, especially related to semilocal rings, can be found in [15]. Our next theorem generalizes results of [1] for fields in which every ordering has level one. Before stating it, we recall some notation from [6]. The set $\Sigma_n(A)$ consists of all elements of the form $\sum \lambda_i^{2^n}$ such that all λ_i lie in A and $\sum \lambda_i A = A$ together with elements x such that $xy = z$ where y and z are such sums of 2^n-th powers. If A is a field or if $n = 1$, these latter elements are already sums of 2^n-th powers.

Theorem 2.10. *Consider the following conditions on A.*

(i) *Every ordering of A has level one.*

(ii) $\Sigma_1(A) = \Sigma_n(a)$ *for every n.*

(iii) *Every fan has at most two elements; if a fan has two elements with associated prime ideal $\mathfrak{p}$, then the unique extensions of the two orderings to $A(\mathfrak{p})$ induce the same valuation on $A(\mathfrak{p})$ and they induce different orderings on the residue class field.*

(iv) *A satisfies* SAP.

(v) *Every fan has at most two elements.*

The first three conditions are equivalent, the last two are equivalent and the first three imply the last two; i.e., (i) $\Leftrightarrow$ (ii) $\Leftrightarrow$ (iii) $\Rightarrow$ (iv) $\Leftrightarrow$ (v).

Proof. The equivalence of (iv) and (v) is well known; see, for example, [13, Theorem 7.4]. The implication (iii) $\Rightarrow$ (v) is trivial. The equivalence of (i) and (ii) follows from [6, Theorem 3.7] which shows that $A^* \cap \Sigma_n(A) = \bigcap P$, where P ranges over all orderings of level n. Finally, we must prove the equivalence of (i) and (iii). Assume first that (iii) holds and that there is an ordering P of exact level n greater than one with associated prime ideal $\mathfrak{p}$. Let P' be the extension of P to $A(\mathfrak{p})$. Let v be the finest valuation compatible with P', let Γ_v be the value group and k_v be the residue field. From Corollary 2.6 of [2] we obtain a bijection between orderings of $A(\mathfrak{p})$ compatible with v which induce the same (archimedean) ordering on k_v as P' does and characters of Γ_v, where for $m > 0$, the character maps onto 2^m-th roots of unity if and only if the ordering has exact level m. Let χ be the character corresponding to P'. Then the characters χ^{2^n} and $\chi^{2^{n-1}}$ correspond to orderings of level one on $A(\mathfrak{p})$ compatible with v and inducing the same ordering on the residue field. This contradicts condition (iii).

Conversely, assume that (i) holds. Assume A has a fan V which does not satisfy (iii) and let $\overline{V}$ be the induced fan on $A(\mathfrak{p})$, where $\mathfrak{p} = \mathfrak{p}(V)$. Let v be the finest valuation of $A(\mathfrak{p})$ compatible with $\overline{V}$. By hypothesis, if V has exactly two orderings, then their extensions in $\overline{V}$ induce the same ordering on the residue class field of v. If V has four or more elements, then $\overline{V}$ does also and it follows from [4, Theorem 2.7] that there are at least two orderings in $\overline{V}$ which induce the same ordering on the residue class field of v. Let P_0', P_1' denote two such orderings of $A(\mathfrak{p})$. Then, if Γ_v denotes the value group of v, we have

$|\Gamma_v/v(P_0' \cap P_1')| = 2$ and it follows from [2, Corollary 2.6] that $A(\mathfrak{p})$ has an ordering of level greater than one compatible with v. Applying Lemma 2.9, we obtain a contradiction of (i).

Note that if, for each prime ideal $\mathfrak{p}$ of A, the orderings of $A(\mathfrak{p})$ are all of level one, then A also has only orderings of level one. The converse is not true. For example, the ring $A = \mathbb{R}[[x]]$ has only one signature [6, Example 4.1], thus necessarily of level one, while $A(0) = \mathbb{R}((x))$ has orderings of all levels [1, p. 43].

3. Witt Rings

In the first section we defined Witt rings of higher level. In this section we shall apply the results of the previous section to the homomorphism ψ_n: $W_n(A) \to W_n(B)$ induced by a ring homomorphism ψ: $A \to B$. We shall be especially interested in the situation where A is a semilocal domain and B is its field of fractions. Many results for the level one case can be found in [7]. Since we have only presentations of the Witt rings to work with, rather than an extensive theory of forms as in level one, the problems tend to be much more difficult to solve. For this reason, most of our results are for reduced higher level Witt rings where our theory of orderings can be applied to help analyze the structure of the rings. For any ring R, we shall denote the nilradical by $\mathrm{Nil}R$. We shall call a ring A *real* if it has an ordering of any level. This is equivalent to -1 not being a sum of squares [6, Theorem 3.5 and Corollary 3.6]. Otherwise we shall say that A is nonreal.

Recall that for $n > m$, there is a canonical ring homomorphism $\varphi_{n,m}$: $W_n(A) \to W_m(A)$ with the property that if σ_n: $W_n(A) \to \mathfrak{A}_n$ and σ_m: $W_m(A) \to \mathfrak{A}_m$ are signatures, then $\varphi_{n,m}^{-1}(\mathrm{Ker}\,\sigma_m) = \mathrm{Ker}\,\sigma_n$ if and only if $P(\sigma_n) = P(\sigma_m)$ [6, Theorems 1.7, 1.8].

Lemma 3.1. *Let ψ: $A \to B$ be a ring homomorphism and let σ be a signature of A of exact level m with $m \leqq n$, n an integer. The following statements are equivalent:*

(i) *σ extends to a signature σ' of B of level n.*

(ii) *$P(\sigma)$ extends to an ordering of B of level n.*

(iii) *The kernel of ψ_n: $W_n(A) \to W_n(B)$ is contained in* $\mathrm{Ker}\,\sigma$.

Proof. The equivalence of (i) and (ii) is trivial. Assume that σ has an extension σ': $W_n(B) \to \mathfrak{A}_n$. Then we have $\sigma = \sigma'\psi_n$, from which we see that $\mathrm{Ker}\,\psi_n \subset \mathrm{Ker}\,\sigma$. Conversely, assume that $\sigma(\mathrm{Ker}\,\psi_n) = 0$. Let R be the image of ψ_n in $W_n(B)$. Then there exists a homomorphism φ: $R \to \mathfrak{A}_n$ such that $\varphi\psi_n = \sigma$. Let $\mathfrak{P} = \mathrm{Ker}\,\varphi$. Then $\mathfrak{P}$ is a minimal prime ideal of R, so there exists a minimal prime ideal $\mathfrak{P}'$ of $W_n(B)$ such that $\mathfrak{P}' \cap R = \mathfrak{P}$ [3, Proposition 16, p. 74]. Then the corresponding signature σ': $W_n(B) \to \mathfrak{A}_n$ with kernel $\mathfrak{P}'$ is an extension of σ.

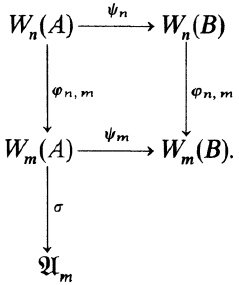
Proposition 3.2. *Let A be a real ring, let ψ: $A \to B$ be a ring homomorphism and let $m \leqq n$ be positive integers. The following two statements are equivalent:*

(i) *Every ordering of level m of A extends to an ordering of exact level less than or equal to n of B.*

(ii) $\varphi_{n,m}(\mathrm{Ker}\,\psi_n)$ *generates a nil ideal in* $W_m(A)$.

In particular, every ordering of level m of A extends to an ordering of level m of B if and only if $\mathrm{Ker}\,\psi_m$ *is nil.*

Proof. First we recall that for any m, the kernels of the signatures on $W_m(A)$ are precisely the minimal, nonmaximal prime ideals of $W_m(A)$ [6, Corollary 1.6]. Thus $\bigcap \mathrm{Ker}\,\sigma$, where σ ranges over all signatures of level m, is the nilradical of $W_m(A)$, hence is the set of all nilpotent elements of $W_m(A)$. Next we apply Lemma 3.1 to the following commutative diagram:



First assume that every ordering of level m extends to an ordering of level at most n on B. Then

$$\mathrm{Ker}\,\psi_n \subset \bigcap_{\sigma} \varphi_{n,m}^{-1}(\mathrm{Ker}\,\sigma) \qquad (\sigma \text{ a signature of level } m).$$

By the comment above, $\varphi_{n,m}(\mathrm{Ker}\,\psi_n)$ is contained in the nilradical of $W_m(A)$, so it generates a nil ideal.

Conversely, assume that $\varphi_{n,m}(\mathrm{Ker}\,\psi_n)$ generates a nil ideal in $W_m(A)$. Then $\varphi_{n,m}(\mathrm{Ker}\,\psi_n)$ is contained in $\mathrm{Ker}\,\sigma$ for each signature σ of level m, hence $\mathrm{Ker}\,\psi_n$ is contained in the inverse image of those prime ideals in $W_n(A)$. By Lemma 3.1, each ordering of level m extends to an ordering of B of level n. This completes the proof.

In Theorem 3.20 and the following example of [2], it is shown that there exist fields A and B and a signature of A which extends to signatures of all exact levels greater than or equal to some fixed one. Thus, given m in the previous proposition, there may not exist an n making the conditions hold, even though all signatures extend.

Proposition 3.3. *Let* $\psi: A \to B$ *be a ring homomorphism and assume that the kernel of the induced homomorphism* $\psi_n: W_n(A) \to W_n(B)$ *is nil. Then on the reduced rings,*

$$\bar{\psi}_n: W_n(A)/\mathrm{Nil}\,W_n(A) \to W_n(B)/\mathrm{Nil}\,W_n(B)$$

is injective.

Proof. If A is nonreal, then $W_n(A)$ and $W_n(B)$ each have a unique prime ideal [12, Proposition 3.16] so the result is trivial. Now assume that A is real. We

must show that if $\psi_n(x) \in \mathrm{Nil}\, W_n(B)$, then $x \in \mathrm{Nil}\, W_n(A)$. Let $\mathfrak{P}$ be a minimal prime ideal of $W_n(A)$. By Proposition 3.2 the corresponding signature extends to B, so there exists a minimal prime ideal $\mathfrak{Q}$ of $W_n(B)$ such that $\psi_n^{-1}(\mathfrak{Q}) = \mathfrak{P}$. By hypothesis, $\psi_n(x) \in \mathfrak{Q}$, so $x \in \mathfrak{P}$. Therefore x lies in every minimal prime ideal of $W_n(A)$, hence in $\mathrm{Nil}\, W_n(A)$.

Proposition 3.4. *Let A be a valuation ring with field of fractions F. Then every signature of A extends to F with the same exact level.*

Proof. Assume the contrary. By [6, Theorem 3.4], there exists a signature σ of exact level n on A and elements $\lambda_i \in F^*$, $a_i \in P$ such that $0 = \sum \lambda_i^{2^n} a_i$. Clearing denominators, we may assume all λ_i lie in A. If $\lambda_i \in A^*$ for any i, then we obtain $0 \in Q(\sigma)$, which is impossible. Therefore all λ_i lie in the unique maximal ideal of A. That is, if v denotes the valuation, then $v(\lambda_i) > 0$ for each i. Choose j so that $v(\lambda_j) = \min_i v(\lambda_i)$. Then $0 = \sum_i (\lambda_j^{-1}\lambda_i)^{2^n} a_i$ where $\sum \lambda_j^{-1}\lambda_i A = A$ and $v(\lambda_j^{-1}\lambda_i) \geqq 0$ for all i. Thus again we obtain the contradiction $0 \in Q(\sigma)$.

Corollary 3.5. *Let A be a valuation ring with field of fractions F. Then*

(i) $\mathrm{Ker}[W_n(A) \to W_n(F)]$ *is nil for all n.*

(ii) *The reduced map $W_n(A)/\mathrm{Nil}\, W_n(A) \to W_n(F)/\mathrm{Nil}\, W_n(F)$ is injective.*

(iii) *If A is a discrete valuation ring with maximal ideal $\mathfrak{m}$ and $\mathfrak{p}(\sigma) = \mathfrak{m}$ for some signature σ of A of exact level n, then there exist exactly $n+1$ orderings of F of level n extending $P(\sigma)$.*

Proof. Part (i) follows from Propositions 3.2 and 3.4. Part (ii) follows from (i) and Proposition 3.3. For (iii), let $t \in A$ such that $\mathfrak{m} = At$. Any extension τ of σ to F is clearly determined by the value of $\tau(t)$. By [6, Theorem 3.4], any choice in $\mathfrak{C}_n$ works. By [6, Proposition 1.5], two extensions τ_1 and τ_2 are equivalent (i.e., $P(\tau_1) = P(\tau_2)$) if and only if $\tau_1(t)$ and $\tau_2(t)$ have the same order, so that the same subgroup of $\mathfrak{C}_n$ is generated by them. Thus there are exactly $n+1$ different extensions of $P(\sigma)$ for the $n+1$ orders $1, 2, 2^2, \ldots, 2^n$.

Remark 3.6. (a) The special case of the corollary above in which $n = 1$ can be found in [11, Example 2.13].

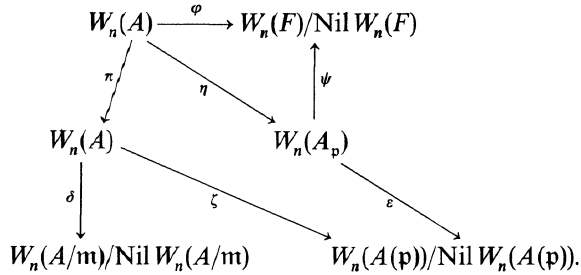
(b) For a rank one discrete valuation ring, the cokernel of the homomorphism $W_n(A) \to W_n(F)$ has been computed by Freeland [8, Theorem 3.14].

(c) For the situation in (iii) above, Proposition 2.5 implies that A cannot be cofinal in F with respect to the extensions of σ.

Proposition 3.7 (Compare [7, Proposition 2.1]). *Let A be a noetherian regular local ring with maximal ideal $\mathfrak{m}$ and $2 \in A^*$. Then the kernel of φ: $W_n(A) \to W_n(F)/\mathrm{Nil}\, W_n(F)$ is contained in the kernel of $W_n(A) \to W_n(A/\mathfrak{m})/\mathrm{Nil}\, W_n(A/\mathfrak{m})$.*

Proof. We induct on the Krull dimension of A. If $\dim A = 0$, then $A = F$ so the result is trivial. Suppose $\dim A > 0$ and let t be a member of a regular system of parameters for A [16, p. 301]. Let $\mathfrak{p} = At$. Then $\bar{A} = A/\mathfrak{p}$ is a regular local ring with $\dim \bar{A} = \dim A - 1$ [16, Theorem 26, p. 303 and Corollary 2, p. 291], and

thus $\mathfrak{p}$ is a prime ideal. Also, the localization $A_{\mathfrak{p}}$ is a regular local ring of dimension one, hence a discrete valuation ring [14, 4.7–4.8, pp. 73–76]. Now consider the following commutative diagram:



Let $x \in \mathrm{Ker}\,\varphi = \mathrm{Ker}\,\psi\eta$. Then $\eta x \in \mathrm{Ker}\,\psi$. By Corollary 3.5, we have $\eta x \in \mathrm{Nil}\, W_n(A_{\mathfrak{p}})$ since $A_{\mathfrak{p}}$ is a valuation ring. Therefore ηx is nilpotent and so $\varepsilon\eta x$ is equal to zero; that is, $\pi x \in \mathrm{Ker}\,\zeta$. We can then apply the induction hypothesis to ζ to conclude that $\pi x \in \mathrm{Ker}\,\delta$, so that $x \in \mathrm{Ker}\,\delta\pi$ as desired.

Corollary 3.8. *Let A be a noetherian regular semilocal domain. Then the kernel of φ: $W_n(A) \to W_n(F)/\mathrm{Nil}\, W_n(F)$ is equal to*

$$\mathfrak{K} = \bigcap_{\mathfrak{p} \in \mathrm{Spec}\, A} \mathrm{Ker}[W_n(A) \to W_n(A(\mathfrak{p}))/\mathrm{Nil}\, W_n(A(\mathfrak{p}))].$$

Proof. Since zero is a prime ideal of A, we have $\mathrm{Ker}\,\varphi \supset \mathfrak{K}$. Let $\mathfrak{p}$ be an arbitrary prime ideal of A and let $x \in \mathrm{Ker}\,\varphi$. Let ψ denote the homomorphism $W_n(A) \to W_n(A_{\mathfrak{p}})$. Then

$$\psi(x) \in \mathrm{Ker}[W_n(A_{\mathfrak{p}}) \to W_n(F)/\mathrm{Nil}\, W_n(F)],$$

which is contained in

$$\mathrm{Ker}[W_n(A_{\mathfrak{p}}) \to W_n(A(\mathfrak{p}))/\mathrm{Nil}\, W_n(A(\mathfrak{p}))]$$

by Proposition 3.7. But then

$$x \in \mathrm{Ker}[W_n(A) \to W_n(A(\mathfrak{p}))/\mathrm{Nil}\, W_n(A(\mathfrak{p}))].$$

Since $\mathfrak{p}$ was arbitrary, x lies in $\mathfrak{K}$.

Theorem 3.9 (Compare [7, Theorem 2.4]). *Let A be a noetherian regular local ring with $2 \in A^*$. Then the kernel of φ: $W_n(A) \to W_n(F)$ is nil.*

Proof. If A is nonreal, then $\mathrm{Ker}\,\varphi$ is contained in the unique prime ideal of $W_n(A)$, so the theorem is clear. Now assume A is real and let $x \in \mathrm{Ker}\,\varphi$. We shall show that x lies in every minimal prime ideal $\mathfrak{P}$ of $W_n(A)$. Let $\mathfrak{p}$ be the prime ideal of A associated with the signature having kernel $\mathfrak{P}$. This signature extends uniquely to $A(\mathfrak{p})$, so there exists a minimal, nonmaximal prime ideal $\mathfrak{Q}$ of $W_n(A(\mathfrak{p}))$ such that $\mathfrak{P}$ is the inverse image of $\mathfrak{Q}$ under the homomorphism $W_n(A) \to W_n(A(\mathfrak{p}))$. By Corollary 3.8, the image of x in $W_n(A(\mathfrak{p}))$ is nilpotent, hence lies in $\mathfrak{Q}$. Therefore we have $x \in \mathfrak{P}$, completing the proof.

An application of this theorem together with Proposition 3.2 gives us the following theorem on extension of orderings.

Theorem 3.10. *Let A be a noetherian regular local ring with $2 \in A^*$ and field of fractions F. Then every signature of A extends to F with the same exact level.*

Corollary 3.11. *Let A be a noetherian regular local ring with $2 \in A^*$ and assume that the field of fractions F has a unique minimal real valuation ring. Then all orderings of A have the same associated prime ideal.*

Proof. By hypothesis the ring $\mathfrak{D}_0(\tau, \mathbb{Q})$ is independent of the signature τ, so the corollary follows immediately from Theorem 2.8 and the fact that all signatures of A extend to F.

Example 3.12. Let F be a superpythagorean field; i.e., the Witt ring of F is an integral group ring. Then F has either exactly one or exactly two places into the real numbers [5, Corollary 6]. If it has only one such place, then F has a unique minimal valuation ring, so the conclusion of the corollary above applies to any subring of F satisfying the stated hypotheses. An example of such a field is the field of iterated Laurent series in any number of variables over a Euclidean field.

Example 3.13. Let A be the ring $\mathbb{R}[x, y]/(x^2+y^2)$ localized at the maximal ideal $\mathfrak{m}$ generated by x and y. The map $A^* \to (A/\mathfrak{m})^* = \mathbb{R}^* \to \{\pm 1\}$ is a signature of level one on A but the field of fractions F has no orderings (of any level). More generally, the ring $A = (K[x, y]/(x^2+y^2))_{\mathfrak{m}}$ will have orderings of all levels if the field K does, defined via $A^* \to K^* \to \mathfrak{C}_n$, but none can extend to the field of fractions of A since it is not formally real. The same holds for the n variable case $A = (K[x_1, \ldots, x_n]/(\sum x_i^2))_{\mathfrak{m}}$, where $\mathfrak{m}$ is generated by $x_1, \ldots, x_n$; if $n \geqq 5$, this ring is a noetherian unique factorization domain [3, Exercise 5, p. 564]. This shows the need for the hypothesis that A be regular in our applications to Witt rings above.

References

1. Becker, E.: Hereditarily-pythagorean fields and orderings of higher level. Monografias de Matemática **29**. Rio de Janeiro: Instituto de Mathemática Pura e Aplicada, 1978
2. Becker, E., Harman, J., Rosenberg, A.: Signatures of fields and extension theory. Preprint
3. Bourbaki, N.: Elements of mathematics, commutative algebra. Reading, Mass.: Addison-Wesley 1972
4. Bröcker, L.: Characterization of fans and hereditarily pythagorean fields. Math. Z. **151**, 149–163 (1976)
5. Brown, R.: Superpythagorean fields. J. Algebra **42**, 483–494 (1976)
6. Craven, T.: Orderings of higher level and semilocal rings. Math. Z. **176**, 577–588 (1981)
7. Craven, T., Rosenberg, A., Ware, R.: The map of the Witt ring of a domain into the Witt ring of its field of fractions. Proc. Amer. Math. Soc. **51**, 25–30 (1975)
8. Freeland, M.: Abstract Witt rings: signatures of arbitrary level. Ph. D. Thesis. Cornel University, Ithaca, N.Y. 1979
9. Knebusch, M.: On the local theory of signatures and reduced quadratic forms. Abh. Math. Sem. Univ. Hamburg **51**, 149–195 (1981)

10. Knebusch, M.: Real closures of commutative rings I. J. Reine Angew. Math. **274/275**, 61–89 (1975)
11. Knebusch, M., Rosenberg, A., Ware, R.: Signatures on semilocal rings. J. Algebra **26**, 208–250 (1973)
12. Knebusch, M., Rosenberg, A., Ware, R.: Structure of Witt rings and quotients of abelian group rings. Amer. J. Math. **94**, 119–155 (1972)
13. Marshall, M.: The Witt ring of a space of orderings. Trans. Amer. Math. Soc. **258**, 505–521 (1980)
14. Northcott, D.: Ideal theory. Cambridge: Cambridge University Press 1953
15. Rosenberg, A., Ware, R.: Equivalent topological properties of the space of signatures of a semilocal ring. Publ. Math. Debrecen **23**, 283–289 (1976)
16. Zariski, O., Samuel, P.: Commutative algebra, Vol. II. Princeton, New Jersey: Van Nostrand 1960

Received September 30, 1981; received in final form January 22, 1982

Math. Z. 180, 567–571 (1982)

Mathematische Zeitschrift
© Springer-Verlag 1982

Second-Dual Liminal C^*-Algebras

Helge Elbrønd Jensen

Mathematical Institute, Technical University of Denmark, Dk-2800 Lyngby, Denmark

Introduction

In the papers [2], [3] there has been given classifications of special types of C^*-algebras. The results in [2] have been used in [4] to extend the classification from [2] to the case, where each point in the spectrum $\hat{A}$ of the algebra A in question has both a finite neighbourhood and a finite closure. It is common to the work done so far in this area that one always puts some restriction on the algebras A, which ensures, that a point $x \in \hat{A}$ can only be in the closure of a set $S \subset \hat{A}$ if x is in the closure of some point in S.

In this paper we shall consider a class of C^*-algebras, where this condition is not satisfied in general. More precisely, we will give a complete description of all separable, liminal C^*-algebras A with countable spectrum, for which there exists a dual ideal $I \subseteq A$, such that A/I is dual. Algebras satisfying this last condition have been called second-dual by Behncke, and we shall use this notation. We use [5] as a general reference to C^*-algebra theory.

1. The Invariants

We shall prefer first to describe the concepts which turn out to determine the algebras of the type mentioned in the introduction. Let T denote a countable, locally quasi-compact, scattered T_1-space, which equals the second derived set of itself. This means, that if O denotes the set of isolated points in T, then each point in $T \setminus O$ is isolated in the relative topology in $T \setminus O$.

Further we shall consider a function q: $T \to N \cup \{\infty\}$, and a set of functions $\mathbf{F} = (F_t)$, $t \in T \setminus O$, where F_t: $T \to N_0$.

The system $(T, q, \mathbf{F})$ will hereafter be called a *second-dual system* if the following conditions are satisfied.

(1.1) $F_t(t) = 1$; $F_t(w) = 0$ for $w \in T \setminus (O \cup \{t\})$.

(1.2) $N_t = \{w \in T \mid F_t(w) > 0\}$ is open and quasi-compact for each $t \in T \setminus O$.

0025-5874/82/0180/0567/$01.00

(1.3) Suppose, that we for each $t\in T\setminus O$ have numbers $n_t\in N$ with $n_t\leqq q(t)$. Then for each $t\in T\setminus O$ there exists an open neighbourhood $N_t'\subseteq N_t$ of t, such that

$$\sum n_t\cdot F_t(u)\leqq q(u),\quad u\in O,$$

where the summation is over all t, for which $u\in N_t'$.

Our next task is, corresponding to each second-dual system, to construct a certain C^*-algebra. During this construction the meaning of the involved concepts will be clear.

Let $(T,q,\mathbf{F})$ be a second-dual system. For each $t\in T$ we choose a Hilbert space H_t with $\dim H_t=q(t)$. Let

$$B=\prod_{t\in O}L(H_t),$$

where $L(H_t)$ is the algebra of bounded linear operators in H_t, and let $I\subseteq B$ be the restricted product of the algebras $K(H_t)$, $t\in O$, where $K(H_t)$ is the compact operators on H_t. For each $s\in O$ we denote by π_s the representation, which sends an element $(a_t)\in B$ onto a_s.

For $i\in N$ and $t\in T\setminus O$ we put $n_t^i=\min\{i,q(t)\}$. Consider first (1.3) with $n_t=n_t^1$, and let N_t^1 be the open sets in (1.3) corresponding to these n_t. Then we can choose pairwise orthogonal projections $e_t^1\in B$, $t\in T\setminus O$, such that $\dim\pi_s(e_t^1)=F_t(s)$ for all $t\in T\setminus O$ and $s\in N_t^1$. Next we consider (1.3) for $n_t=n_t^2$ with corresponding open neighbourhoods N_t^2. We remark, that because of (1.2) the sets N_t^1 and N_t^2 both equal N_t except for at most a finite number of points in O. Hence we may assume, that $N_t^2\subseteq N_t^1$. Then, corresponding to the points $t\in T\setminus O$ with $n_t^2>n_t^1$, we can choose pairwise orthogonal projections $e_t^2\in B$, all orthogonal to each e_t^1, such that $\dim\pi_s(e_t^2)=F_t(s)$ for all $t\in T\setminus O$ and all $s\in N_t^2$. In this way we may proceed with $n_t=n_t^3$ and so on.

The result is a system of pairwise orthogonal projections (e_t^i), $t\in T\setminus O$, $i=1,\ldots,q(t)$ with the property, that for each $t\in T\setminus O$ and each $i\in\{1,\ldots,q(t)\}$, we have $\dim\pi_s(e_t^i)=F_t(s)$ for all $s\in N_t$ except at most a finite number of points in O. We can further assume, that if $q(u)=\infty$ for $u\in O$, then the projection space of $\sum\pi_u(e_t^i)$, where the summation is over all t and i for which u belongs to N_t^i, has infinite codimension.

Let t be fixed in $T\setminus O$. Then for each $i=2,\ldots,q(t)$ the projections $\pi_s(e_t^1)$ and $\pi_s(e_t^i)$ has the same dimension for all $s\in N_t$ except maybe for a finite number of points in O. Therefore we can choose partial isometries $p_t^i\in B$, such that $e_t^1-(p_t^i)^*(p_t^i)$ and $e_t^i-(p_t^i)(p_t^i)^*$ are both projections belonging to I.

For each $x\in B$, let $[x]$ denote the corresponding element in B/I. The elements $[e_t^i]$, $[p_t^i]$, $t\in T\setminus O$, $i=1,\ldots,q(t)$, uniquely determines a dual subalgebra of B/I isomorphic with the restricted product of $K(H_t)$, $t\in T\setminus O$. The counterimage in B of this algebra will be called the C^*-algebra determined by $(T,q,\mathbf{F})$ and denoted $A(T,q,\mathbf{F})$.

During this construction we make certain choices of projections and partial isometries. But it will be clear from the arguments in the next section, that different allowed choices give isomorphic algebras.

Theorem 1.1. *The algebra $A(T,q,\mathbf{F})$ is a liminal, second-dual C^*-algebra with spectrum T.*

Proof. By construction A is a second-dual C^*-algebra, and clearly the points in $\hat{A}$ and T are in one to one correspondance and A is liminal. The only problem is, if corresponding points, which are not isolated, have the same neighbourhoods.

Let t be such a point. The set N_t is an open, quasi-compact neighbourhood of t in T. On the other hand, the set $\{\pi\in\hat{A}\,|\,\|\pi(e_t^1)\|\geqq 1/2\}$ is an open, quasi-compact neighbourhood of t in $\hat{A}$, and by construction this set can only differ from N_t by a finite number of points in O. From this it easily follows, that each neighbourhood of t in T is also a neighbourhood of t in $\hat{A}$, and vica versa.

2. The Isomorphisms

First we shall determine the isomorphism classes of the model algebras and then prove, that each liminal, second-dual algebra is isomorphic with one of the models.

Theorem 2.1. *Two algebras $A_1=A(T_1,q_1,\mathbf{F}^1)$ and $A_2=A(T_2,q_2,\mathbf{F}^2)$ are isomorphic, if and only if there exists a homeomorphism $\varphi\colon T_1\to T_2$ such that*

(1) $q_2(\varphi(t))=q_1(t)$ *for all* $t\in T_1$.

(2) *For all $t\in T_1$ we have $F^2_{\varphi(t)}(\varphi(u))=F_t^1(u)$ for all $u\in O_1$ except for at most a finite number of points.*

Proof. O_1 (resp. O_2) of course denotes the set of isolated points in T_1 (resp. T_2). Referring to the construction, the algebra A_1 (resp. A_2) is acting on a Hilbert space $H_1=\prod_{t_1\in O_1} H^1_{t_1}$ (resp. $H_2=\prod_{t_2\in O_2} H^2_{t_2}$), and is determined by a system $e^i_{t_1}$, $p^i_{t_1}$ (resp. $f^i_{t_2}$, $r^i_{t_2}$) of projections and partial isometries.

Suppose first, that there exists a homeomorphism with the properties (1), (2). We can then put $T_1=T_2=T$, $O_1=O_2=O$, $H_1=H_2=H$, $q_1=q_2$. Let $t\in T\setminus O$.

For $s\in O$ we have $\dim\pi_s(e_t^1)=\dim\pi_s(f_t^1)=F_t^1(s)=F_t^2(s)$ except for at most a finite number of points s. Hence we can change e_t^1 and f_t^1 by projections in I (cf. the notation in the construction) to smaller projections, such that $\pi_s(e_t^1)$ and $\pi_s(f_t^1)$ has the same dimension for all $s\in O$. Next, we can in the same way - for $i=2,\ldots,q(t)$ - change e_t^i, f_t^i, p_t^i, r_t^i by elements in I, such that $\dim\pi_s(e_t^i)=\dim\pi_s(f_t^i)$, $\dim\pi_s((p_t^i)(p_t^i)^*)=\dim\pi_s(r_t^i)(r_t^i)^*)$ for all $s\in O$. Doing this for all $t\in T\setminus O$ we obtain two new systems (denoted as the old ones) of projections and partial isometries, which again defines the algebras A_1 and A_2, since the changes has only been by elements in I. And the properties needed in the construction will still be satisfied. Then we can define isomorphisms U_t, $t\in T\setminus O$, of H_t onto itself, such that $U=\oplus U_t$ takes e_t^i onto f_t^i and p_t^i onto r_t^i for $t\in T\setminus O$, $i=1,\ldots,q(t)$. The corresponding automorphism on $B=\prod_{s\in O} L(H_s)$ therefore takes A_1 onto A_2.

Suppose next, that A_1 and A_2 are isomorphic by an isomorphism Φ. This isomorphism defines a homeomorphism $\varphi: T_1 \to T_2$, which of course satisfies (1). Let $t_1 \in T_1 \setminus O_1$ and put $t_2 = \varphi(t_1)$. If I_1 (resp. I_2) is the greatest dual ideal in A_1 (resp. A_2) then Φ takes A_1/I_1 onto A_2/I_2, and especially takes $K(H_{t_1}^1)$ onto $K(H_{t_2}^2)$. Hence the images of $\Phi(e_t^1)$ and $f_{t_2}^1$ in A_2/I_2 are equivalent projections in this algebra. By the same arguments as in [1], there exists then a partial isometry $p \in A_2$, such that $\Phi(e_t^1) - p^*p$ and $f_{t_2}^1 - pp^*$ are both projections in I_2. Therefore $\dim \pi_s(\Phi(e_t^1)) = \dim \pi_s(f_{t_2}^1)$ for all $s \in O_2$ except at most for a finite number of points. Since Φ must be spatial this proves (cf. the construction), that also (2) is satisfied.

Theorem 2.2. *Each separable, liminal, second-dual C*-algebra A is isomorphic with a model algebra $A(T, q, \mathbf{F})$.*

Proof. Let T be the spectrum of A, and let π_t be an irreducible representation corresponding to $t \in T$. If H_t is the Hilbert space for π_t, and O is the set of isolated points in T, we can think of A as realized on the Hilbert space $\prod_{s \in O} H_s$, such that the restricted product of $K(H_s)$, $s \in O$, is the treatest dual ideal in A. If I is this ideal, then further A/I is isomorphic with the restricted product of $K(H_t)$, $t \in T \setminus O$. We put $q(t) = \dim H_t$ for all $t \in T$. Let $(\tilde{e}_t^i, \tilde{p}_t^i)$, $i = 1, \ldots, q(t)$ be a system of matrix units in $K(H_t)$ for each $t \in T \setminus O$. By the same arguments as in [1], the system of projections $\{\tilde{e}_t^i\}$, $t \in T \setminus O$, $i = 1, \ldots, q(t)$ can be lifted to a system of orthogonal projections $\{e_t^i\}$ in A, and the system $\{\tilde{p}_t^i\}$ can be lifted to a system $\{p_t^i\}$ of partial isometries in A, such that $e_t^1 - (p_t^i)^*(p_t^i)$, $e_t^i - (p_t^i)(p_t^i)^*$ are both projections in I for all $t \in T \setminus O$ and $i = 2, \ldots, q(t)$.

For any $t \in T \setminus O$ we define F_t: $T \to N_0$ by putting $F_t(t) = 1$, $F_t(u) = 0$ for $u \in T \setminus (O \cup \{t\})$ and $F_t(s) = \dim \pi_s(e_t^1)$ for $s \in O$. We put $\mathbf{F} = (F_t)$, $t \in T \setminus O$.

Let us consider the system $(T, q, \mathbf{F})$ and prove, that (1.1)–(1.3) is satisfied. That the values of each F_t are finite follows from the fact, that $\pi_s(e_t^1)$ is a compact projection for each $s \in O$. The condition (1.1) is satisfied by definition, and (1.2) follows as in the proof of Theorem 1.1. Concerning (1.3) suppose, that we for each $t \in T \setminus O$ have numbers $n_t \in N$ with $n_t \leqq q(t)$. The projections $e_t^1, \ldots, e_t^{n_t}$ are orthogonal, and for each $i = 1, \ldots, n_t$ we have $\dim \pi_s(e_t^i) = F_t(s)$ for all $s \in O$ except for at most a finite number of points. Recalling the definition of the set N_t we can therefore find an open neighbourhood $N_t' \subseteq N_t$ of t, such that $\dim \pi_s(e_t^i) = F_t(s)$ for all $s \in N_t' \cap O$ and all $i = 1, \ldots, n_t$. Consider now a point $s \in O$. Then of course $\sum \dim \pi_s(e_t^i) \leqq q(s)$, where the summation is over all $t \in T \setminus O$ and all $i = 1, \ldots, q(t)$. Especially the inequality holds, if the summation is over all t with $s \in N_t'$ and all $i = 1, \ldots, n_t$. But for such a point t we have $\sum_{i=1}^{n_t} \dim \pi_s(e_t^i) = n_t \cdot F_t(s)$. Hence (1.3) is satisfied.

It is straightforward to see, that by changing, if necessary, the elements in the system (e_t^i, p_t^i) by elements in I, we can obtain, that this system has all the properties, which are used in the construction of the algebra $A(T, q, \mathbf{F})$. Hence A is isomorphic with this algebra, which completes the proof.

In most of the previous work on characterization of C*-algebras with countable spectrum, it has been a main tool, that if A is an algebra of the type

under discussion, and I is an ideal in A for which A/I is elementary, then A/I can be lifted.

This property is in general not shared by the algebras discussed in this paper. With the notation of Sect. 1 it is for example clear, that if $q(s)<\infty$ for all $s\in O$, and if A/I is isomorphic with K_∞ (the compact operators on a separable, infinite-dimensional Hilbert space), then A/I cannot be lifted. And such a liminal, second-dual C^*-algebra does exist. One can just put $T=N\cup\{\infty\}$ with the usual topology, put $q(n)=n$ for $n\in N$, $q(\infty)=\infty$, and $F_\infty(t)=1$ for all $t\in T$. Then $A(T,q,\mathbf{F})$ has the required properties.

References

1. Behncke, H., Leptin, H.: C^*-algebras with a two point dual. J. Functional Analysis **10**, 330–335 (1972)
2. Behncke, H., Leptin, H.: Classification of C^*-algebras with a finite dual. J. Functional Analysis **16**, 241–257 (1974)
3. Elliott, G.A.: On totally ordered groups, and K_0. In: Ring Theory Proceedings (Waterloo 1978), pp. 1–49. Lecture Notes in Mathematics **734**. Berlin-Heidelberg-New York: Springer 1979
4. Jensen, H.E.: On some C^*-algebras with countable spectrum. Mitt. Math. Gesellsch. Hamburg (to appear)
5. Pedersen, G.K.: C^*-algebras and their Automorphism Groups. London-New York-San Francisco: Academic Press 1979

Received January 1, 1981; received in final form March 11, 1982